ISBN: 978-605-72045-3-0

ICONDATA
INTERNATIONAL CONFERENCE ON DATA SCIENCE & APPLICATIONS

ICONDATA'24
PROCEEDINGS BOOK

September 2 - 6, 2024
Pristina, KOSOVO

EDITOR
Prof. Dr. Murat GÖK



KOLEGJI AAB
YALOVA ÜNİVERSİTESİ 2008
BİLECİK ŞEYH EDEBALİ ÜNİVERSİTESİ
SAKARYA UYGULAMALI BİLİMLER ÜNİVERSİTESİ
ISPARTA UYGULAMALI BİLİMLER ÜNİVERSİTESİ
İSTANBUL ESENYURT ÜNİVERSİTESİ 2013

www.icondata.org

PREFACE

In the current era, data serves as the foundation for advancing intelligent systems across various sectors. The transition from agricultural to industrial, and subsequently to information societies, has brought us to the cusp of Industry 4.0—a stage where digital transformation thrives on the capacity to store, analyze, and interpret vast amounts of data. These capabilities have exponentially enhanced the functionality and impact of products and services, shaping them to be more adaptive, predictive, and autonomous.

At the core of this evolution lies the interplay between data science and artificial intelligence (AI). Data science offers the methodologies and computational power to manage and interpret extensive datasets, while AI, drawing from these insights, enables systems to simulate human-like reasoning and learning. Together, they create intelligent systems with applications across all scientific domains. ICONDATA'24, the 6th International Conference on Data Science and Applications, aims to bring together academia and industry on a platform dedicated to advancing these fields. The conference was held in Pristina, Kosovo, from September 2–6, 2024, showcasing cutting-edge research that illustrates the role of data science and AI in shaping tomorrow's innovations.

The principal objective of ICONDATA'24 is to present recent data-driven research across disciplines. This conference provides a unique venue for professionals from diverse fields to exchange innovative ideas, share practical applications, and foster interdisciplinary collaborations that bridge natural and social sciences. With blind peer review and rigorous evaluation standards, accepted papers were selected based on originality, technical depth, relevance, and clarity. Those aligning with the conference's focus will be considered for publication in *Data Science and Applications* and *Veri Bilimi Journal*, ensuring continued dissemination of high-impact findings.

Our gratitude extends to the supporting institutions, including AAB College, Yalova University, Sakarya University of Applied Sciences, Isparta University of Applied Sciences, Istanbul Esenyurt University, and Bilecik Şeyh Edebali University, for their support of ICONDATA'24.

Prof. Dr. Murat Gök

Editor

COMMITTEES

Honorable Committee

- Prof. Dr. Bujar Demjaha (Rectorr of AAB College)
- Prof. Dr. Mehmet Bahçekapılı (Rector of Yalova University)
- Prof. Dr. Zafer Asım Kaplancıklı (Rector of Bilecik Şeyh Edebali University)
- Prof. Dr. Yılmaz Çatal (Rector of Isparta University of Applied Sciences)
- Prof. Dr. Mehmet Sarıbıyık (Rector of Sakarya University of Applied Sciences)
- Prof. Dr. Süleyman Özdemir (Rector of Istanbul Esenyurt University)

Conference Chair

- Prof. Dr. Murat Gök (Yalova University)

Organizing Board

- Prof. Dr. Mehmet Büyükyıldız (Bursa Technical University, Turkey)
- Assoc. Prof. Dr. Süleyman Uzun (Sakarya University of Applied Sciences, Turkey)
- Assoc. Prof. Dr. Faruk Bulut (Istanbul Esenyurt University, Turkey)
- Assoc. Prof. Dr. Emre Dandil (Bilecik Şeyh Edebali University,Turkey)
- Assist. Prof. Dr. Sinan Demir (Isparta University of Applied Sciences, Turkey)
- Emine Cengiz (Yalova University, Turkey)
- Sudenur Şenkal (Yalova University, Turkey)

Scientific Committee

- Prof. Dr. Ayhan İstanbullu (Balıkesir University, Turkey)
- Prof. Dr. Ecir Uğur Küçüksille (Süleyman Demirel University, Turkey)
- Prof. Dr. Müfit Çetin (Yalova University, Turkey)
- Prof. Dr. Ramazan Bayındır (Gazi University, Turkey)
- Prof. Dr. Naci Genç (Yalova University, Turkey)
- Prof. Dr. Yıldıray Yalman (Piri Reis University, Turkey)
- Prof. Dr. Resul Daş (Fırat University, Turkey)
- Prof. Dr. Ahmet Bedri Özer (Fırat University, Turkey)
- Prof. Dr. Ahmet Zengin (Sakarya University, Turkey)
- Prof. Dr. Atilla Elçi (Hasan Kalyoncu University, Turkey)
- Prof. Dr. Muharrem Tolga Sakallı (Trakya University, Turkey)
- Assoc. Prof. Dr. Bilgin Metin (Bogazici University, Turkey)
- Assoc. Prof. Dr. Ahmet Koltuksuz (Yaşar University, Turkey)
- Assoc. Prof. Dr. Derya Avcı (Fırat University, Turkey)
- Assoc. Prof. Dr. Sunay Türkdoğan (Yalova University, Turkey)
- Assoc. Prof. Dr. Fatih Ertam (Fırat University, Turkey)
- Assoc. Prof. Dr. Muhammed Ali Aydın (Istanbul University-Cerrahpaşa, Turkey)
- Assoc. Prof. Dr. Serdar Solak (Kocaeli University, Turkey)
- Assoc. Prof. Dr. Bünyamin Ciylan (Gazi University, Turkey)
- Assoc. Prof. Dr. Ercan Buluş (Tekirdağ Namık Kemal University, Turkey)
- Assoc. Prof. Dr. Sedat Akleylek (Ondokuz Mayıs University, Turkey)
- Assoc. Prof. Dr. Murat Arıcı (Selçuk University, Turkey)
- Assist. Prof. Dr. Mert Özarar (HAVELSAN Cyber Security Director / Ankara Science University, Turkey)
- Assist. Prof. Dr. Bülent Tuğrul (Ankara University, Turkey)
- Assist. Prof. Dr. Burcu Demirelli Okkalıoğlu (Yalova University, Turkey)
- Assist. Prof. Dr. Meltem Kurt Pehlivanoğlu (Kocaeli University, Turkey)
- Assist. Prof. Dr. Murat Ak (Akdeniz University, Turkey)
- Assist. Prof. Dr. Ahmet Albayrak (Düzce University, Turkey)

- Assist. Prof. Dr. Güneş Harman (Yalova University, Turkey)
- Assist. Prof. Dr. Murat Okkalıoğlu (Yalova University, Turkey)
- Assist. Prof. Dr. Ömer Aydın (Manisa Celal Bayar University, Turkey)
- Assist. Prof. Dr. Ömer Özgür Bozkurt (Turkish National Defense University, Turkey)
- Assist. Prof. Dr. Osman H. Koçal (Yalova University, Turkey)
- Assoc. Prof. Dr. Şebnem Özdemir (Beykent University, Turkey)
- Assist. Prof. Dr. Mustafa Coşar (Hitit University, Turkey)
- Assist. Prof. Dr. Tarık Yerlikaya (Trakya University, Turkey)
- Assist. Prof. Dr. Yunus Özen (Yalova University, Turkey)
- Assist. Prof. Dr. Kevser Ovaz Akpınar (Sakarya University, Turkey)
- Assist. Prof. Dr. Esra N. Yolaçan (Eskişehir Osmangazi University, Turkey)
- Assist. Prof. Dr. Mustafa Cem Kasapbaşı (İstanbul Commerce University, Turkey)
- Assist. Prof. Dr. Fatma Büyüksaraçoğlu Sakallı (Trakya University, Turkey)
- Assist. Prof. Dr. Atila Bostan (Ankara Science University, Turkey)
- Assist. Prof. Dr. Andaç Mesut (Trakya University, Turkey)
- Assist. Prof. Dr. Burcu Yılmazel (Eskişehir Technical University, Turkey)
- Assist. Prof. Dr. Alpay Doruk (Bandırma University, Turkey)
- Assist. Prof. Dr. Özgür Can Turna (Istanbul University-Cerrahpaşa, Turkey)
- Assist. Prof. Dr. Zeynep Gürkaş Aydın (Istanbul University-Cerrahpaşa, Turkey)
- Dr. Ahmet Ali Süzen (Isparta University of Applied Sciences, Turkey)
- Dr. Mehmet Yavuz Yağcı (Istanbul University, Turkey)
- Dr. Remzi Gürfidan (Isparta University of Applied Sciences, Turkey)
- Dr. Ömer Aslan (Siirt University, Turkey)
- Dr. Faruk Süleyman Berber (Süleyman Demirel University, Turkey)
- Dr. Semih Çakır (Zonguldak Bülent Ecevit University, Turkey)
- Dr. Gülsüm Akkuzu Kaya (Kirsehir Ahi Evran University, Turkey)
- Dr. Ahmet Karaküçük (Uludağ University, Turkey)
- Dr. Duygu Sinanç Terzi (Amasya University, Turkey)
- Dr. Mehmet Mehdi Karakoç (Ağrı University, Turkey)
- Dr. Kerime Dilşad Çiçek (Ayvansaray University, Turkey)
- Dr. Sultan Zavrak (Düzce University, Turkey)
- Dr. Maad M. Mijwil (Baghdad College of Economic Sciences University, Iraq)

CONTENTS

Detecting and Analyzing Urban Land Use Change: The Impact of High Spatial Resolution Imaging on LULC Classification 1
Prioritization of Furniture Upholstery Fabric Attributes Using Kano Model and Fuzzy Logic 8
Kelime Gömme Teknikleri ile Kitap Özetlerinin Kategorik Sınıflandırılması 12
A Study on the Use of Neural Networks for LHC Data Interpretation 22
Enhancement of Vector Boson Measurements with Boosting Learning Approach 27
Duygu Analiz Çalışması: Veri Setinin Etiketlenme Biçimine Göre Makine Öğrenmesi ve Derin Öğrenme Yöntemlerinin Karşılaştırılması 33
Estimation of Parameters Affecting Water Quality Using Data Mining Algorithms 39
Üretken Yapay Zekanın Duygu Analizi Üzerindeki Etkileri 45
Makine Öğrenmesi Yöntemleri ile Atölye Yoğunluk Tahminlemesi 49
EKG sinyalleri kullanılarak makine öğrenmesi algoritmaları ile kalp hastalıkları teşhisi modeli 57
Üretken Yapay Zekâ Kullanımı ile İçerik Tabanlı Yemek Tarifi Öneri Sistemi Geliştirilmesi 69
Uçak Bakımlarında Rutin Olmayan İş Yükünün Tahmini Üzerine Deneysel bir Çalışma 78
Veri Yönetimi Metamodeline Dayalı Çoklu Modelli Veri Bütünleştirme Mimarisi 81
Görüntü İşleme Yöntemi Kullanılarak Hedef Takibi: YOLO ve Haar- cascade Algoritmalarının Karşılaştırılması 88
Sensör Füzyonu ve Derin Öğrenme Teknikleri Kullanarak Hava Araçlarında Anomali Tespiti 94
Su Tüketimi Trendlerini Anlamak: R Shiny Uygulaması ile Veri Analizi ve Görselleştirme 103
Sağlık Eğitiminde Yapay Zekâ Destekli Chatbotların Potansiyelinin Değerlendirilmesi 110
An Approach for Data Augmentation of Dental X-Ray Images with the use of Deep Convolutional Generative Adversarial Network 118
Derin Öğrenme Tabanlı Otomatik Kask Tespiti 124
Hanehalkı Düzeyinde Pandemi Kırılganlığının Ölçüm ve Analizi: Kayseri İlinde Bir Araştırma 130
Redundant or Interesting? Inspection of Redundant Association Rules in a Psychiatric Dataset 140
Nitelik Seçimi ve Yapay Öğrenme Algoritmaları ile Borsa İstanbul Hisse Senedi Yön Tahmini 147
Yapay Öğrenme Yöntemleriyle Ses Kayıtlarının Metne Dönüştürülmesi 155
Optimizing Energy Efficiency in Wood Processing Plants through Data Analysis: A Case Study on Value-Added Wood Products Manufacturing 163
Trend analysis for modelling the primary energy consumption: Evidence from three largest consumers . 175
Development of statistical models for estimating the hydropower consumption 181
Sıfır Atık Yönetiminin Tersine Lojistik Faaliyetlerinde Uygulanmasının Temel Engellerinin Bulanık-AHP Kullanılarak Önceliklendirilmesi 185
Popüler Makine Öğrenimi Yöntemlerinin Türkçe Şiirlerde Şair Tanıma Problemi Üzerinde Karşılaştırılması 195
Temel Düzey Veri Analizi Yöntem Seçiminde ChatGPT-4o'nun Performansı 205
Artificial Intelligence Supported ICD Coding: An Evaluation on Patients in the 0-15 Age Group 211

Detecting and Analyzing Urban Land Use Change: The Impact of High Spatial Resolution Imaging on LULC Classification

Pajtim Zariqi[1], Lena Halounova

[1]*Department of Geomatics, Czech Technical University in Prague, Czech Republic*

Abstract

The objective of this research is to investigate the extent to which a fine spatial resolution of images influences the accuracy and efficiency of urban LULC classification, which was done using Planet Imagery. The speed increase of the urban areas area gives rise to the specific problems in environmental management and urban planning which call for accurate and updated LULC data. Although the LULC classification models in some cases may be helpful, they tend to be inadequate in urban areas where the heterogeneity of land use and the complexity of spatial configurations are high. The mixture of 5m, 3m, 0,5m resolution of satellite imagery with advanced machine learning algorithms such as RF, SVM, ML were used to improve the specificity and precise urban LULC classifications. The research was limited to a city in a stage of rapid growth, with the aim to capture the dynamic moment the land use went through in a five-year span. An application of the methodological framework based on the recent image processing techniques and the complex machine learning algorithms, is aimed at this study to eliminate the shortcomings that may be encountered during the conventional LULC classifications and the dynamic nature of urban environments. The plan involves image preprocessing with resolution enhancement and feature selection, application of machine learning models for classification, and validation of results by comparison with historical data. Such findings indicate an increased accuracy of LULC classification and the ability to discern details at small sizes that are not visible in lower resolution data. The study illustrates core issues of the process of urban land use change pointing out outspread urbanization, intrusion on green spaces, and the transformation of the periurban zone. In addition, this study reveals that high resolution imaging integrated with machine learning have the power to advance urban planning strategies and environmental monitoring.

Keywords: *Machine Learning, LULC Classification, SVM, RF, ML*

1 Introduction

The recent development of image sensors has opened up a range of very interesting options for measuring imperviousness and other land-cover types with a high degree of detail. Earth observation data collected from diverse sensors with varying geographical resolutions is extensively employed in the examination of global environmental changes, the management of natural resources, the study of ecological system, population growth etc. During the last few decades, remote sensing technologies have made remarkable development and now a number of images from different sensors with high, medium and coarse spatial resolution are available (Clark et al., 2004). Commercial satellites including Skysat (< 1m resolution; https://www.planet.com), Planetscope andRapidEye(3-7m resolution; https://www.planet.com), obtain images at a level of detail that was previously only achievable by aerial platforms. The spatial resolution of a remote sensing image is a fundamental property that significantly impacts the accuracy of classification. In the case of remotely sensed images, the classification accuracy is strongly affected by the influence of boundary pixels and the finer spatial resolution which increases the spectral-radiometric variation of LULC classes (Markham and Townshend, 1981). Such high spatial resolution space-derived imagery has already been used successfully in a diverse array of environmental mapping efforts such as earthquake damage assessment (Miura and Midorikawa,2006), orchard and vineyard identification (Warner and Steinmaus, 2005), rain forest tree mortality inventory (Clark et al.,2004), and coral reef delineation (Maeder et al.,

*Contact email: pajtim.zariqi@fsv.cvur.cz

2002). The linkage between classification accuracy and spatial resolution of the image strongly depends on the size and spatial patterns of LULC classes (Moody and Woodcock, 1994). The spatial resolution of a remote sensing image is a fundamental characteristic that significantly impacts the accuracy of classification. Rao & Narendra (2006) and Boakye et al (2008) have done the mapping of land use and land cover changes using unsupervised classification methodology. Unlike unsupervised classification many scientists like Rao & Narendra (2006), Remi et al (2007), Chaudhary et al (2008) and Kim et al (2008) have applied supervised classification in their studies. Various image classification approaches such as MLC, SVM, and RF have been developed with the launch of satellite series.This study analyzes multi-resolution remote sensing images to evaluate the impact of different spatial resolutions on the accuracy of land use and land cover (LULC) classification.

2 Study site and Datasets used

2.1 Study site

The study area is the urban region of Ferizaj, a key city in southeastern Kosovo. Geographically, the urban area is situated between 42° 22′ 15.00″ N to 42° 23′ 45.00″ N latitudes and 21° 08′ 20.00″ E to 21° 10′ 30.00″ E longitudes, encompassing an area of approximately 25 square kilometers. Ferizaj is a rapidly growing urban center, known for its strategic location along the main transport routes connecting the north and south of Kosovo. The urban area is characterized by a mix of residential, commercial, and industrial zones, reflecting its role as a regional hub.

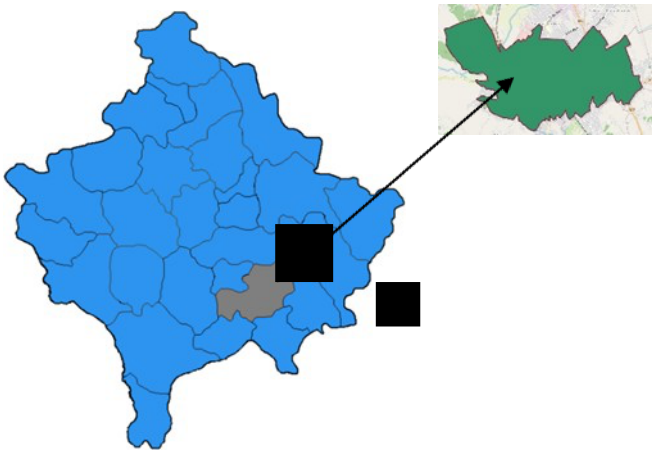


Figure 1. Location map of study area

The city's layout is shaped by its modern infrastructure, and it is intersected by the Nerodimka River, which adds to its unique urban landscape. The location map of the study site is shown in Figure 1.

2.2 Dataset used

This study analyzed remote sensing images from two different sensors with varying spatial resolutions. The LULC classification is performed using three different satellite images: SkySat with a geographical resolution of 0.5 m, Planet Scope with a spatial resolution of 3 m, and Planet Scope with a spatial resolution of 5 m. These images were taken on July, 2019. The ground truth data is obtained using the high-resolution Google Earth photos and and orthophoto of the respective year. The attributes of the datasets are displayed in Table 1, while the visuals may be observed in Figures 2 (a), (b), and (c) correspondingly.

Table 1. Characteristics of datasets

Satellite	Bands	Spectral Resolution	Spatial Resolution
Sky Sat	Blue	450-515 nm	0.5m
	Green	515-595 nm	
	Red	605-695 nm	
	NIR	740-900 nm	
	PAN	450-900 nm	
PlanetScope	Coastal Blue	431-452 nm	3m
	Blue	465-515 nm	
	Green I	513-549 nm	
	Green II	547-583 nm	
	Yellow	600-620 nm	
	Red	650-680 nm	
	Red-Edge	697-713 nm	
	NIR	845-885 nm	
Rapid Eye	Blue	440-510 nm	5m
	Green	520-590 nm	
	Red	630-685 nm	
	Red-Edge	690-730 nm	
	NIR	760-850 nm	



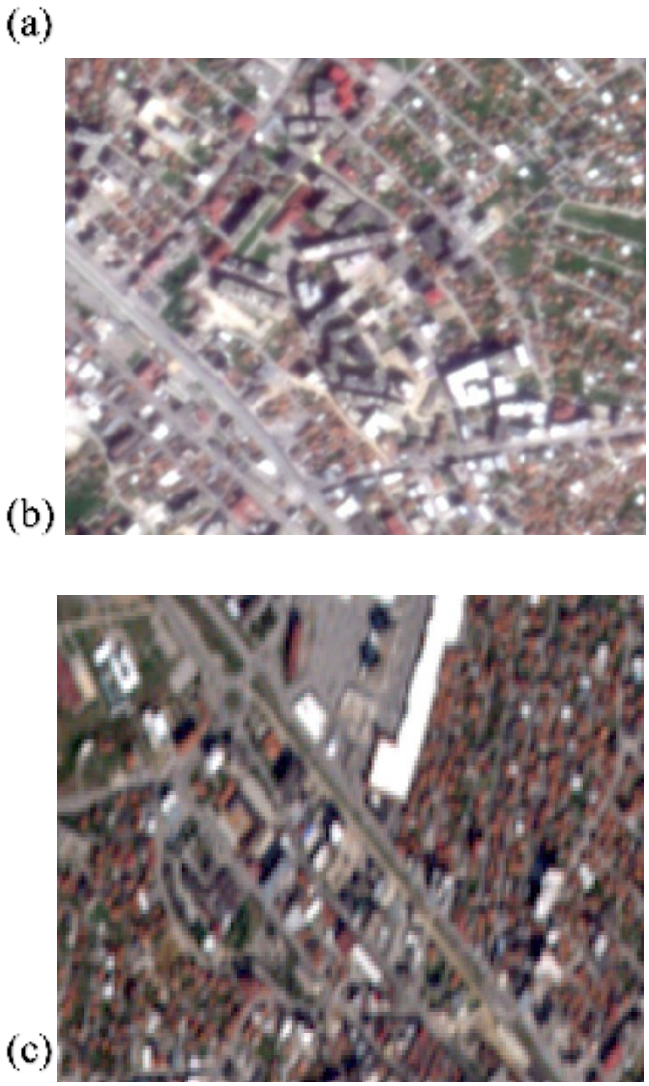


Figure 2. (a) Sky Sat images (0.5m), (b) Planet Scope (3m), Rapid Eye (6,5m)

3 Methodology and Materials

The method used in this study is based on the identification and characterization of the change in the urban land use through the use of high spatial resolution satellite imagery. The imagery used includes satellite imagery of resolutions of 0,5 meters, 3 meters, and 5 meters from Planet Imagery and Sentinel Hub. At first, PlanetScope imagery was available through a student's account, and it was possible to download satellite images of a particular area, which was enough for this research. In the case of SkySat images, an application for sponsorship was made to Sentinel Hub which was approved and this gave access to the required high resolution data. All the three sets of imagery were chosen for the year 2019, more specifically for the month of July. Thistemporal consistency was chosen to be compatible in the analysis and to train on the same samples for all the images. The pre processing of the images was not necessary since the data obtained from Planet Imagery was already preprocessed and could be classified. Classification entailed sorting the imagery into six LULC classes, namely; A sufficient number of training samples were collected for each class. The identification and verification of these classes were further enhanced by the use of the orthophotos of the year 2019 which offered a good ground for classification. For this study, three different classification algorithms were considered: The three algorithms that have been used in this study include Maximum Likelihood, Random Forest, and Support Vector Machines (SVM). All of these algorithms were used to evaluate the performance of the satellite imagery in identifying the LULC changes in the urban setting.

3.2 Per-pixel Land-cover Classification

Per-pixel land-cover classification is a prevalent technique in remote sensing that classifies each pixel in an image into specific land-cover categories, such as water, forest, or urban areas, based solely on the pixel's spectral information without considering the spatial relationships between pixels (Richards & Jia, 2006). This method involves comparing the spectral signature of each pixel with known signatures of different land-cover types, using various classification algorithms like Maximum Likelihood Classification (MLC) (Lillesand, Kiefer, & Chipman, 2015).The process typically includes collecting training data, developing a classification model, and assessing accuracy through metrics like overall accuracy, producer's accuracy, and user's accuracy (Jensen, 2015). Per-pixel classification is valued for its simplicity, ease of implementation, and flexibility, particularly when dealing with large datasets or when the land-cover classes are spectrally distinct (Richards & Jia, 2006). However, this approach has limitations, such as ignoring spatial context, which can lead to misclassification, especially in heterogeneous areas—a problem known as the "salt-and-pepper" effect (Lillesand, Kiefer, & Chipman, 2015). While advanced methods like object-based image analysis (OBIA) have been developed to address these issues, per-pixel classification remains widely used, particularly in scenarios where computational simplicity and speed are essential (Jensen, 2015).

3.3 Maximum Likelihood Classification (MLC)

Maximum Likelihood Classification (MLC) is a prominent statistical method used in remote sensing and

image analysis to classify pixels in multispectral or hyperspectral images. This technique is based on the statistical decision theory and involves assigning each pixel to the class that has the highest likelihood of being correct based on the observed data. MLC is predicated on the assumption that the data for each class in the image follows a normal (Gaussian) distribution (Richards & Jia, 2006). The underlying principle of MLC is to calculate the probability that a given pixel belongs to a particular class by using the pixel's spectral values across different bands. This probability is derived using the multivariate normal density function. The classification rule is to assign the pixel to the class with the highest probability, known as the maximum likelihood (Lillesand, Kiefer, & Chipman, 2015). One of the primary advantages of MLC is its ability to use the full statistical information (mean and covariance) to make classification decisions. This often leads to higher accuracy compared to simpler classification methods, particularly when the classes are well-separated in the feature space (Richards & Jia, 2006). Moreover, MLC can handle classes with different variances and correlations among the spectral bands, making it a flexible and powerful tool in remote sensing applications (Lillesand, Kiefer, & Chipman, 2015). However, MLC also has certain limitations, for example is highly sensitive to the quality of the training data, as inaccuracies in the training samples can lead to significant classification errors (Jensen, 2015).

3.4 Random Forest

Random Forest is a robust machine learning algorithm used for both classification and regression, known for its ability to reduce overfitting, a common issue with individual decision trees (Breiman, 2001). The algorithm constructs multiple decision trees using bootstrapped samples of the training data, and at each split, a random subset of features is selected to ensure diversity among the trees (Hastie, Tibshirani, & Friedman, 2009). The final prediction is made by aggregating the predictions of all the trees—through majority voting for classification or averaging for regression (Breiman, 2001).Random Forest offers several advantages, including robustness to overfitting, the ability to handle large datasets, and providing an inherent measure of feature importance (James, Witten, Hastie, & Tibshirani, 2013, Hastie, Tibshirani, & Friedman, 2009). However, it can be computationally intensive and produces models that are less interpretable compared to single decision trees (James, Witten, Hastie, & Tibshirani, 2013).

3.5 Support Vector Machine

Support Vector Machine (SVM) is a highly effective supervised learning algorithm used for both classification and regression, particularly adept at handling high-dimensional data where the number of dimensions may exceed the number of samples (Vapnik, 1995). The primary goal of SVM is to identify the optimal hyperplane that maximizes the margin between different classes, with the closest data points, known as support vectors, playing a crucial role in determining the hyperplane's position (Cristianini & Shawe-Taylor, 2000). SVMs are known for their strong generalization capabilities, especially in high-dimensional spaces, and their resistance to overfitting, particularly when the data are well-separated (Vapnik, 1995). Despite these strengths, SVMs can be computationally intensive and require careful selection of kernels and parameters. Moreover, SVMs typically need modifications to handle multiclass classification problems effectively (James, Witten, Hastie, & Tibshirani, 2013)

4 Results and Discussion

This section presents the outcomes of the classification analysis conducted on Planet Scope,Rapid Eye and SkySat using Arcgis Pro, employing SVM, ML and RF classifiers. A classification image of these three algorithms can be seen in Figure 3. It is worth mentioning at the beginning of the results elaboration that the Water Body class was considered a constant across all classifications. The water bodies in the city of Ferizaj are seasonal, and in 90% of cases, they have a maximum width of 3 meters. Figure 4 presents the LULC classification results across three different resolutions: 0.5m, 3m, and 5m, using three distinct algorithms: SVM, ML, and RF. A total of six classes were used for classification. As observed in Figure 4(a), there is a notable similarity in the results between the SVM and RF algorithms, particularly within two specific classes: Forests, Orchards, and Shrubs and Arable Land and Barren Land. It is important to highlight that the Forest, Orchards, and Shrubs class had the smallest number of training samples, and at a 0.5m resolution, this class is typically highly distinguishable. To further reinforce this argument, Figure 4(b) provides a very clear overview related to what was mentioned above. Additionally, in Figure 4(b), the similarity apparent. Overall, the diagram (Figure 4 (c)) highlights a strong similarity between the SVM and RF algorithms across most land use classes, with ML showing some variation, particularly in the Pastures, Grassland category.

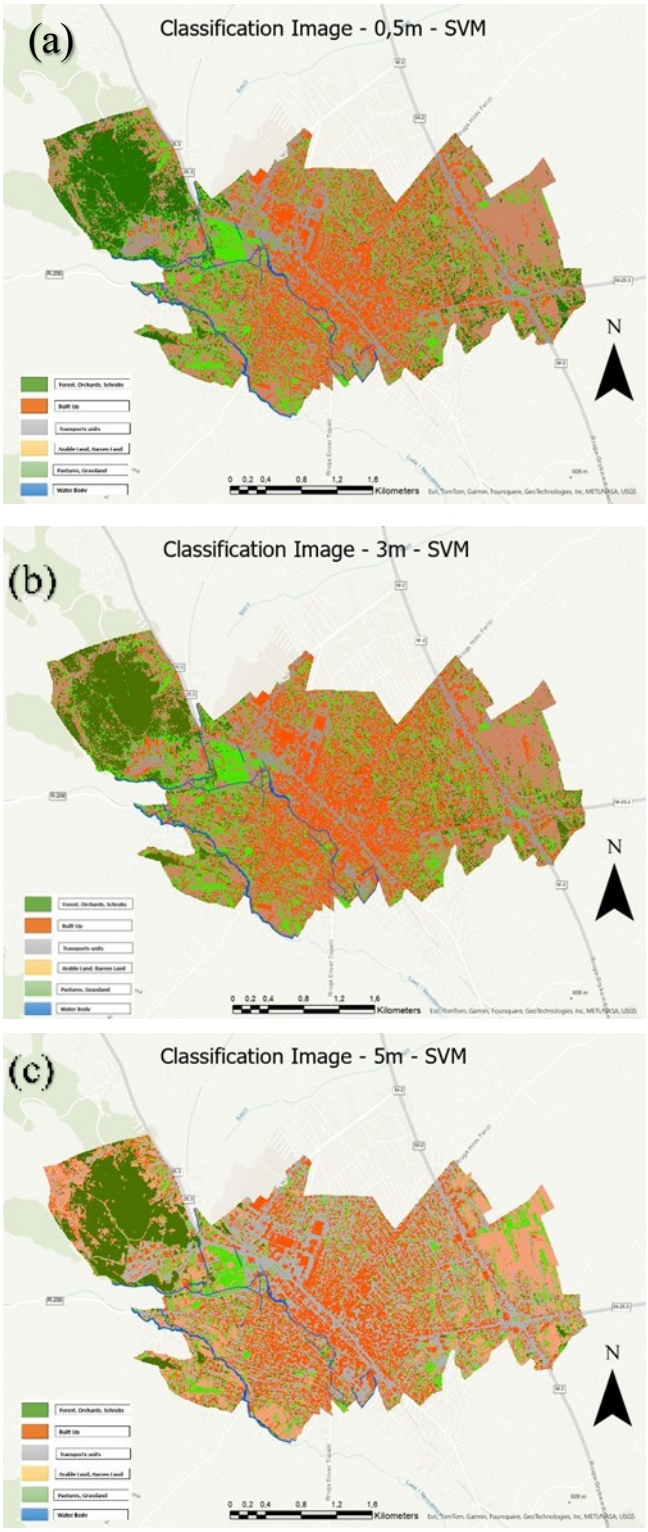


Figure 3. The classification of the image using the SVM algorithm (a) 0.5m Resolution, (b) 3m Resolution, (c) 5m Resolution.

Figure 4(a) shows stability between the two resolutions, 0.5m and 5m. They exhibit significant similarity in certain classes. However, overall, there is a noticeable balance between the resolutions, with the exception of the Transport Units category.

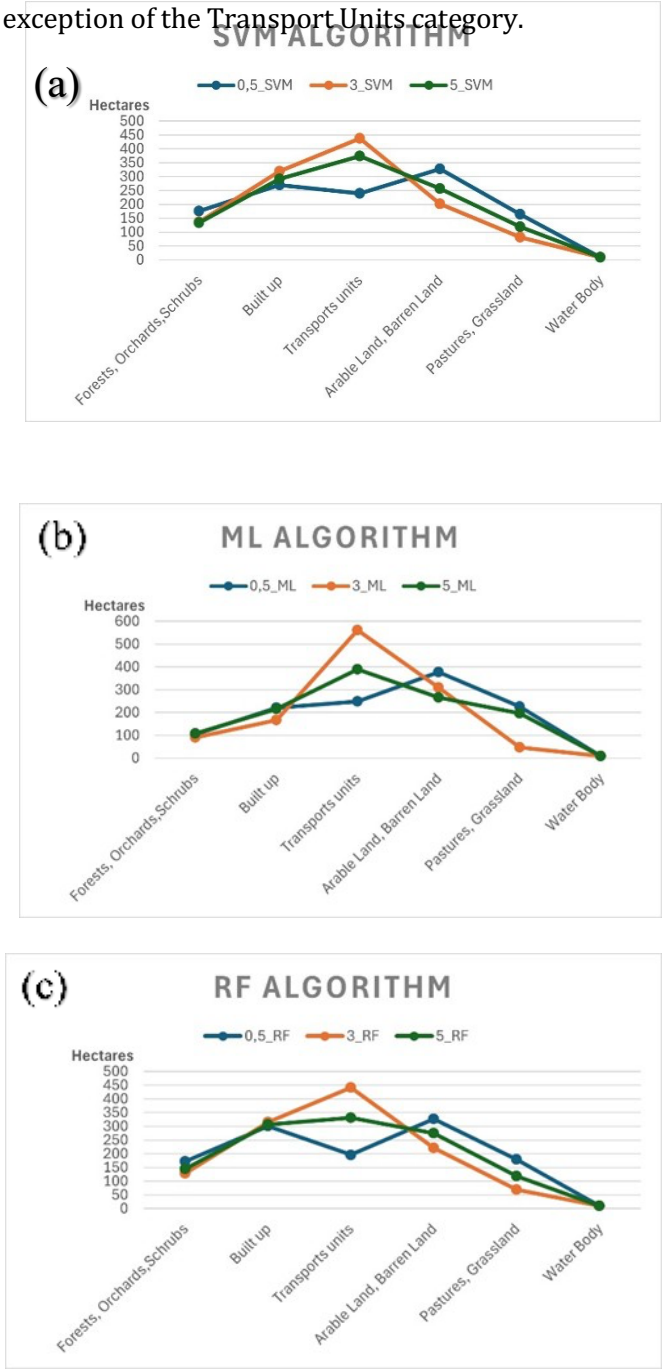
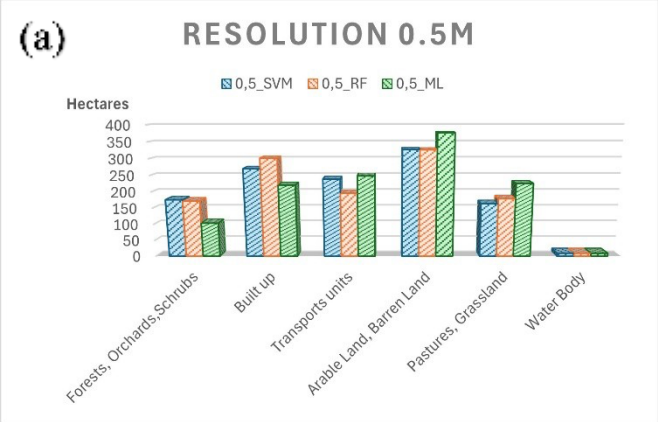
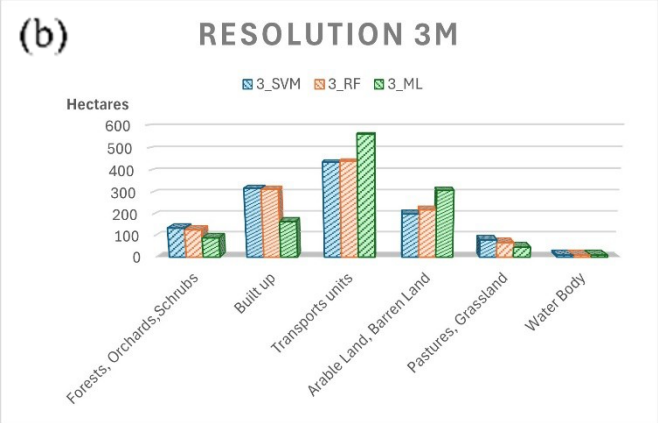


Figure 4. The results of each algorithm (a) SVM Algorithm, (b) ML Algorithm, (c) RF Algorithm

The 0.5m and 5m resolutions show significant similarity in several land use categories, such as Built up, Transport Units, and Water Body, indicating a certain level of stability between these two resolutions in

Figure 4 (b). Overall, the SVM and RF algorithm appears to provide the most consistent results across different resolutions, while the ML algorithm shows the most variability, particularly in Arable Land, Barren Land.





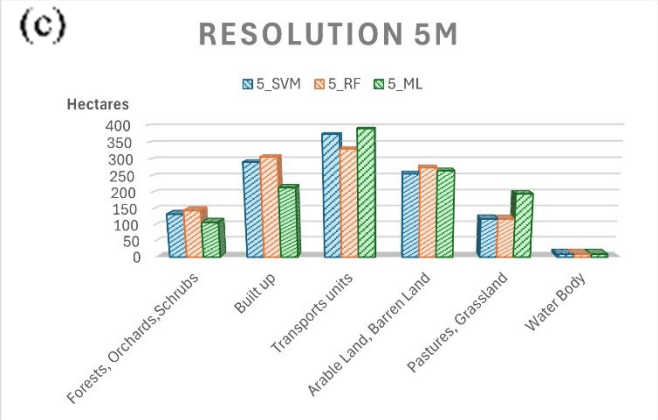


Figure 5. The results of each resolution (a) 0,5m, (b) 3m, (c) 5m

The two classes that have shown consistency across all three algorithms are Forest, Orchards, Schrubs and Built-up. Considering that the 0.5m resolution should theoretically be the most accurate of the three resolutions, it is evident. that the SVM and RF algorithms have provided the best results in this study.

5 Conclusion and Future Work

This paper has examined the use of supervised image classifiers, the maximum likelihood, random forest and super vector machine algorithms, in developing an urban land-cover map using 0,5m, 3m and 5m Planet imagery. The classification was performed using 6 classes, where for 5 of them, the necessary training samples were taken, while for the Water Body class, it was taken as a DWG file and integrated into ArcGIS Pro to avoid its alteration over time. Naturally, its identification would not have been easy, considering that the area where the water body is located is mostly covered with shrubs and vegetation. For the Built-Up and Transport Units classes, a larger number of training samples were taken compared to the other classes, but a satisfactory result was not achieved for the Transport Units class, which indicates that a larger number of training samples does not necessarily lead to higher accuracy. The distribution of training samples was primarily based on the edges of the image, but also covered the entire urban area. SVM and RF achieved almost identical results in the classifications, whereas this was not the case with ML. However, throughout the classification process, it is clearly observed that spatial resolution is crucial in achieving classification accuracy.

Lulc Classification covers a wide range of topics that can be addressed, and this work was just one of them. Future work should continue to investigate larger territories, add more classes, explore other elements of accuracy, and further investigate the role of training samples in the results.

References

[1] Clark, D.B., Read, J.M., Clark, M.L., Cruz, A.M., Dotti, M.F., Clark, D.A. (2004). Application of 1-m and 4-m resolution satellite data to ecological studies of tropical rain forests, Ecological Applications, 14, pp. 61-74.

[2] Markham, B.L., Townshend, J.R.G. (1981). Land cover classification accuracy as a function of sensor spatial resolution, XV International Symposium on Remote Sensing of Environment, Ann Arbor

[3] Miura, H., and S. Midorikawa, 2006. Updating GIS building inventory data using high-resolution satellite images for earthquake damage assessment: Application to metro Manila, Philippines, Earth-quake Spectra, 22:151–168.

[4] Warner, T.A., and K. Steinmaus, 2005. Spatial classification of orchards and vineyards with high spatial resolution panchro-matic imagery, Photogrammetric Engineering & Remote Sensing, 71(1):179–187.

[5] Maeder, J., S. Narumalani, D.C. Rundquist, R.L. Perk, J. Schalles, K. Hutchins, and J. Keck, 2002. Classifying and mapping general coral-reef structure using Ikonos data, Photogrammetric Engineering & Remote Sensing, 68(12):1297–1305.

[6] Moody, A., Woodcock, C.E. (1994). Scale-dependent errors in the estimation of land-cover proportions: Implications for global land-cover datasets, Photogrammetric engineering and remote sensing, 60, 5, pp. 585-596.

[7] Remi, A., Huang, S., Jiren, L., Herve, Y., and Yvew-Louis, D. (2007) Large Scale Land Cover Map Generation based on low and medium resolution, Envisat ASAR Data, Application of Poyang Lake area (JIANG XI, P.R. China). Proc. Envisat Symposium 2007, April 23-27, 2007, Montreux, Switzerland (ESA SP -636, July 2007).

[8] Rao, K.N., and Narendra, K. (2006) Mapping and Evaluation of Urban Sprawling in the Mehadrigedda Watershed in Visakhapatnam metropolitan region using Remote Sensing and GIS. Current Science, 91(11), pp. 1552-1557. Li RTH, Chung SH. "Digital boundary controller for single-phase grid-connected CSI". *IEEE 2008 Power Electronics Specialists Conference*, Rhodes, Greece, 15–19 June 2008.

[9] Jensen, J. R. (2015). *Introductory Digital Image Processing: A Remote Sensing Perspective* (4th ed., pp. 247-249). Pearson

[10] Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2015). *Remote Sensing and Image Interpretation* (7th ed., pp. 570-575). John Wiley & Sons.

[11] Richards, J. A., & Jia, X. (2006). *Remote Sensing Digital Image Analysis: An Introduction* (4th ed., pp. 109-113). Springer.

[12] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.

[13] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed., pp. 587-588). Springer.

[14] Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (pp. 93-97). Cambridge University Press.

[15] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R* (pp. 324-326). Springer.

[16] Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory* (pp. 138-141). Springer.

[17] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R* (pp. 349-350). Springer.

[18] Jensen, J. R. (2015). *Introductory Digital Image Processing: A Remote Sensing Perspective* (4th ed., pp. 336-338). Pearson.

[19] Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2015). *Remote Sensing and Image Interpretation* (7th ed., pp. 452-455). John Wiley & Sons.

[20] Richards, J. A., & Jia, X. (2006). *Remote Sensing Digital Image Analysis: An Introduction* (4th ed., pp. 145-149). Springer.

Prioritization of Furniture Upholstery Fabric Attributes Using Kano Model and Fuzzy Logic

Derya HAROGLU[1], Nesrin TATLIDIL[1]*
[1]Erciyes University, Engineering Faculty, Industrial Design Engineering, Kayseri, TURKIYE

Abstract

The success of a new product development process depends on how well the key customer needs (CNs) are converted into product characteristics. Quality Function Deployment (QFD) has been considered a significant methodology in transforming the customer's voice into product properties at the design stage of the new product development process. Furthermore, the Kano model and Fuzzy Kano model have been used to better prioritize the CNs by emphasizing the customer satisfaction degree. Here, prioritizing CNs enables companies to effectively use their assets such as money, work force, and time by focusing on the most important customer requirements. The overall goal of the research is to develop an upholstery fabric meeting the customer satisfaction for the furniture industry. In this study, the importance weights of CNs of a furniture upholstery fabric were calculated using the Fuzzy Kano-QFD model and the Adjusted Kano-Fuzzy-QFD (AKFQ) model proposed in the author's recent paper. It was demonstrated that the AKFQ model could prioritize the importance weights of CNs in an efficient way.

Keywords: *Fuzzy Kano, QFD, Furniture Upholstery Fabric*

1 Introduction

The worldwide upholstered furniture market size was approximately $39.44 billion in 2023, and is forecasted to be $56.71 billion by 2030 with a compound annual growth rate (CAGR) of 5.3% from 2024 to 2030 [1].

Quality Function Deployment (QFD), and Kano model are useful methodologies preferred in the design stage of the new product development process to accurately understand the voice of the customer that is one of the key strategies for winning at new products [2].

Fuzzy set theory dealing with fuzzy data is a quantitative approach to clarify vagueness and increase objectivity, which was first introduced by Lotfi Zadeh in 1965 [3]. Lee & Huang [4] proposed Fuzzy Kano survey in 2009 to better analyze uncertainity in human thoughts and improve decision making for complex problems. Unlike the traditional Kano survey, participants can choose more than one answer to a pair of functional and dysfunctional questions with a certain percentage, 100% in total, in the Fuzzy Kano survey [4]. The Fuzzy Kano survey was studied in various researches including classifying the aesthetic attributes of a car profile, customer satisfaction in airline services, and service quality of a theme park [4–7].

The overall goal of the research is to develop an upholstery fabric meeting the customer satisfaction for the furniture industry. In this study, the importance weights of customer needs (CNs) of a furniture upholstery fabric were calculated using the Fuzzy Kano-QFD model ($\alpha \geq 0.4$) and the Adjusted Kano-Fuzzy-QFD (AKFQ) model proposed in the author's recent paper [8].

2 Applying the Fuzzy Kano-QFD model and the Adjusted Kano-Fuzzy-QFD (AKFQ) model in the design of a furniture upholstery fabric

2.1 Identifying customer needs

Table 1. Survey results (n= 68)

Educational background	Response (%)	Monthly income (Turkish lira)	Response (%)	Furniture replacement frequency	Response (%)
High school	35.29	<40,000	44.12	<7 years	33.82
Undergraduate and postgraduate	64.71	≥40,000	55.88	≥7 years	66.18

*Contact email: dharoglu@erciyes.edu.tr

Table 2. Customer needs for a furniture upholstery fabric

Item code	Customer needs
	Aesthetics
CN1	Fabric should have a good aesthetic appearance (color, pattern, etc.)
CN2	Fabric should have trend (fashion) colors and patterns
CN3	Fabric should be compatible with the furniture concept involved (retro, country, modern, luxury style)
CN4	Fabric should suit with other elements / accessories of the furniture (e.g., buttons)
	Performance
CN5	Fabric should retain its original color when exposed to light and treatments such as washing, dry cleaning, and friction
CN6	Fabric should have no tendency to pilling (after rubbing)
CN7	Fabric should resist to wear and damage after prolonged use
CN8	Fabric should have a good stain resistance (water, dirt, oil repellent properties)
CN9	Fabric should repel dust and hair (after static electricity)
CN10	Fabric should be made up of natural, recyclable materials
CN11	Fabric should be easy to clean and maintain
CN12	Fabric should have a good seam slippage strength
	Safety
CN13	Fabric should have anti-allergic properties
CN14	Fabric should be flame retardant
	Comfort
CN15	Fabric should present a good thermal comfort (heat conduction, moisture conduction, air permeability properties)
CN16	Fabric touch (tactile property) should be good (softness, the feeling of fingertips when run over the surface of the fabric)

The traditional Kano and the Fuzzy Kano surveys were completed by 68 valid respondents (Table 1).

A set of 16 upholstery fabric attributes for customer satisfaction in furniture industry, shown in Table 2, was obtained through discussions with design experts employed in the furniture industry.

2.2 Final importance of customer needs of both Fuzzy Kano-QFD and proposed AKFQ model

The final importance of customer requirements according to both the Fuzzy Kano-QFD model and the proposed AKFQ model is shown in Table 3. Here, the goal setting section was determined together with the design team of a famous furniture brand in Turkiye. In addition, the sales point section in the planning matrix of QFD was not taken into account.

It was observed that there was no difference in terms of the Kano categories between the two models for each CN. The reason was attributed to the fact that the participants preferred the option they chose in the Kano survey with a higher percentage rate (e.g., 80% or 90%) in the Fuzzy Kano survey. This resulted in keeping the same Kano category for the related customer need.

Table 3. Final importance of customer needs of both Fuzzy Kano-QFD model and proposed AKFQ model

Item code	Rate of importance	Kano category (Fuzzy Kano, $\alpha \geq 0.4$	Kano category (Traditional)	m = max {SI, $\lvert DI \rvert$} (Fuzzy Kano)	m^1 = max {SI', $\lvert DI' \rvert$} (AKFQ)	Goal	Improvement Ratio	Relative weights of the Fuzzy Kano-QFD model	Defuzzy Relative weights of the proposed AKFQ model
CN1	4.68	O	O	0.80	0.74	5.00	5.00	0.84	0.63
CN2	4.12	I	I	0.48	0.55	4.86	4.86	0.43	0.40
CN3	4.25	I	I	0.42	0.44	4.14	4.14	0.33	0.27
CN4	4.28	I	I	0.50	0.51	4.57	4.57	0.44	0.36
CN5	4.76	O	O	0.93	0.81	4.57	4.57	0.91	0.64
CN6	4.74	O	O	0.94	0.75	5.00	5.00	1.00	0.65
CN7	4.85	O	O	0.86	0.73	4.86	4.86	0.91	0.61
CN8	4.93	O	O	0.94	0.85	4.57	4.57	0.95	0.67
CN9	4.66	O	O	0.88	0.68	4.71	4.71	0.87	0.56
CN10	3.18	I	I	0.19	0.27	3.14	3.14	0.09	0.10
CN11	4.87	O	O	0.94	0.90	4.43	4.43	0.91	0.68
CN12	4.60	O	O	0.82	0.62	4.43	4.43	0.75	0.47
CN13	3.44	I	I	0.39	0.49	3.86	3.86	0.23	0.21
CN14	3.56	I	I	0.35	0.32	3.43	3.43	0.19	0.17
CN15	4.56	O	O	0.91	0.71	4.00	4.00	0.74	0.49
CN16	4.51	O	O	0.79	0.84	4.43	4.43	0.71	0.63

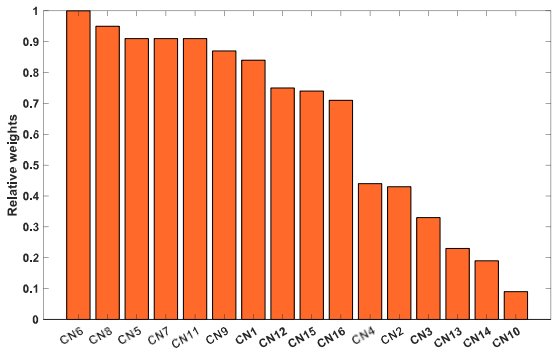


Figure 1. Relative weights of the Fuzzy Kano-QFD model

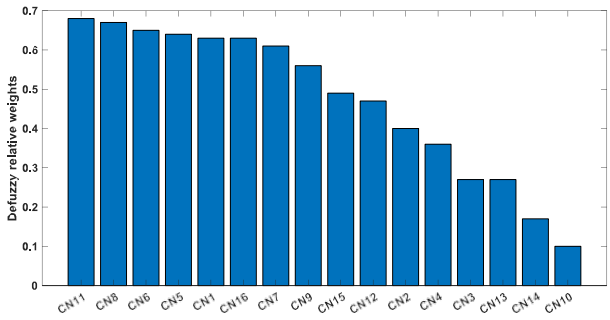


Figure 2. Defuzzy relative weights of the proposed AKFQ model

Figure 1 and 2 show the graphs of the relative importance of the customer requirements according to the Fuzzy Kano-QFD model and the proposed AKFQ model, respectively.

When the two models are compared four out of the first five rankings are seen to be shared by the same 4 CNs (CN-5, 6, 8, 11) for both models. Unlike the Fuzzy Kano-QFD model, in the proposed model, CN1, fabric should have a good aesthetic appearance, and CN16, fabric touch should be good, shares the fifth important CN ranked before CN7, fabric should resist to wear and damage after prolonged use. Here, aesthetic appearance of a furniture affects also the aesthetics of a living space. Colors and patterns excite consumers to purchase a furniture [9]. Furthermore, comfort plays an important role in designing fabrics for furniture upholstery, as consumers would like to feel comfortable while sitting [10]. Thus, the result of the proposed AKFQ model seems to be logical.

3 Ethical Approval

Ethical approval was obtained from the Social and Human Sciences Ethics Committee of Erciyes University (Project number: 25) to conduct surveys for the study. Participation was voluntary.

4 Conclusion

In this study, the Fuzzy Kano model ($\alpha \geq 0.4$) and the Adjusted Kano-Fuzzy model was applied for determining the importance weights of CNs of a furniture upholstery fabric in the planning matrix of QFD.

It was shown that the AKFQ model could prioritize the CNs of a furniture upholstery fabric effectively. Since the quality attributes (must-be, one dimensional, indifference, attractive) included in the Kano model are not of equal importance as indicated in the literature [11,12] the integration of Kano model indice of k to calculate the adjustment factor (m′) in the AKFQ model would change the final importance of CNs. Furthermore, as customers' opinions about a product are expected to be vague, using fuzzy logic to clarify survey results would make great sense.

Unlike the traditional Kano survey, the Fuzzy-Kano survey allows participants to select multiple answers instead of a single answer [4]. However, this study showed that the participants tended to give a higher percentage rate to the option they prefered in the Kano survey, which might cause the Fuzzy Kano model to make a less objective evaluation compared to the AKFQ model. Additionally, the performance criteria in the CNs of a furniture upholstery are less emotional and higher in number when compared with aesthetic criteria. This could cause the Fuzzy Kano model not to be able to make a more objective evaluation.

This study would help furniture manufacturers design new customer-driven upholstery fabrics increasing customer satisfaction.

Acknowledgements

The authors are thankful to Ibrahim Koker, and to the administration of Istikbal, Erciyes Anadolu Holding A.S. for their support during the research work.

References

[1] Consulting NMS. "Upholstered furniture market by product type (chairs, sofas, tables, beds, and other products), price range (mass furniture and premium furniture), by end use (residential and commercial), and by distribution channel (hypermarket/supermarket, specialty st. Next Move Strateg Consult 2024". https://www.nextmsc.com/report/upholstered-furniture-market.

[2] Cooper RG. *Winning at New Products: Creating Value Through Innovation*, 5th Edition New york,

USA, Perseus Publishing, 2011.

[3] Zadeh LA. "Fuzzy Sets". *Information and Control*, 8(3), 338-353, 1965.

[4] Lee YC, Huang SY. "A new fuzzy concept approach for Kano's model". *Expert Systems with Applications*, 36(3), 4479–4484, 2009.

[5] Yadav HC, Jain R, Shukla S, Avikal S, Mishra PK. "Prioritization of aesthetic attributes of car profile". *International Journal of Industrial Ergonomics*, 43(4), 296–303, 2013.

[6] Avikal S, Singh R, Rashmi R. "QFD and Fuzzy Kano model based approach for classification of aesthetic attributes of SUV car profile". *Journal of Intelligent Manufacturing*, 31(2), 271–284, 2020.

[7] Wang C-H, Fong H-Y. "Integrating fuzzy Kano model with importance-performance analysis to identify the key determinants of customer retention for airline services". *Journal of Industrial and Production Engineering*, 33(7), 450–458, 2016.

[8] Haroglu D, Leblebici İ. "An Approach Integrating Kano Model And Fuzzy Logic Into The Planning Matrix Of Quality Function Deployment: An Application For The Design of a Baby Diaper". *Akademik Platform Mühendislik ve Fen Bilimleri Dergisi*, 9(3), 433–450, 2021.

[9] Guzel TA. "Consumer Attitudes And Preferences About Upholstered Furniture". *Gazi University Journal of Science Part A: Engineering*, 7(3), 69–82, 2020.

[10] Wang C. Applying the Visual Analysis and Design Principles of Postwar American Modern Furniture to Contemporary Home Living Furniture Design. MSc Thesis, Auburn University, Alabama, USA, 2024.

[11] Tan KC, Shen XX. "Integrating Kano's model in the planning matrix of quality function deployment". *Total Quality Management*, 11(8), 1141–1151, 2000.

[12] Chaudha A, Jain R, Singh AR, Mishra PK. "Integration of Kano's Model into quality function deployment (QFD)". *International Journal of Advanced Manufacturing Technology*, 53, 689–698, 2011.

Kelime Gömme Teknikleri ile Kitap Özetlerinin Kategorik Sınıflandırılması

Kerem KESKİN[1]*, Mümine KAYA KELEŞ[2]

[1]*Çukurova Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Adana, TÜRKİYE*
[2]*Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi, Bilgisayar ve Bilişim Fakültesi, Bilgisayar Mühendisliği, Adana, TÜRKİYE*

Özet

Bu çalışmada kitap sitelerinden çekilen kitap özetleri ve kategorilerinin kelime gömme yöntemleri, doğal dil işleme teknikleri ve makine öğrenmesi algoritmaları ile sınıflandırılması yapılmıştır. Ayrıca sıklıkla kullanılan kelime gömme yöntemlerinden One Hot Encoding, Word2Vec ve Terim Frekansı - Ters Doküman Frekansı (TF-IDF) yöntemleri bu çalışmada kullanılmış ve başarı karşılaştırılması yapılmıştır. Ek olarak kullanılan ön işleme yöntemlerinin kombinasyon tablosu gösterilmiş ve başarı tablosuna eklenmiştir. Sonuçlara bakıldığında Türkçe metinler için Destek Vektör Makinesi, Naive Bayes ile Lojistik Regresyon Modellerinin ve TF-IDF ile One-Hot Encoder kelime gömme tekniklerinin daha başarılı sonuçlar verdiği gözlenmiştir.

Anahtar Kelimeler: *Veri Bilimi, Makine Öğrenmesi, Doğal Dil İşleme, Kelime Gömme*

Categorical Classification of Book Summaries Using Word Embedding Techniques

Abstract

In this study, book summaries and categories taken from book sites were classified using word embedding methods, natural language processing techniques and machine learning algorithms. In addition, one hot encoding, Word2Vec and Term Frequency - Inverse Document Frequency (TF-IDF) methods, which are frequently used word embedding methods were used in this study and their success was compared. Additionally, the combination table of the pre-processing methods used is shown and added to the table. Looking at the results, it was observed that Support Vector Machine, Naive Bayes and Logistic Regression Models and TF-IDF and One-Hot Encoder word embedding techniques gave more successful results for Turkish texts.

Keywords: *Data Science, Machine Learning, Natural Language Processing, Word Embedding*

1 Giriş

Sağlıktan eğitime birçok sektörde makine öğrenmesi teknikleri ve doğal dil işleme teknikleri kullanılmaya devam edilmektedir. Birçok dilde doğal dil işleme çalışmaları sürmektedir ve türkçe dil yapısının zorluklarının aşılması için farklı tekniklere başvurulmaktadır. Bu tekniklerden ön işleme metotları ve kelime gömme algoritmaları sıklıkla çalışmalarda kullanılmaktadır. Yapılan bir çalışmada Word2Vec ve Terim Frekansı - Tersine Doküman Frekansı tekniklerinin kullanılması ile türkçe metinlerin sınıflandırılmasında başarıyı arttırdığı gözlenmiştir [1]. Ayrıca yapılan başka bir çalışmada Türkçe ve İngilizce Twitter mesajlarının duygu analizinde başarıyı arttırmak için Word2Vec kelime gömme tekniği kullanılmıştır [2]. Kelime gömme yöntemleri tek başına veya farklı yöntemlerle kombine edilerek de kullanılabilir ve sınıflandırma başarısını arttırabilir. Kelime gömme yöntemlerinin kombine edilerek kullanılması ile yapılan bir çalışmada Word2Vec ve TF-IDF kelime gömme teknikleri birleştirilmiş ve sınıflandırma başarısı için olumlu sonuçlar vermiştir [3]. Metinlerde kelimeler arasındaki anlam ilişkilerinin öğrenilmesi ve gruplandırılması için Word2Vec gibi kelime gömme yöntemlerine başvurulmaktadır. Otel yorumlarının duygu analizi ve sınıflandırılması için yapılan bir çalışmada başarıyı arttırmak için Word2Vec yöntemi kullanılmıştır ve kelime grupları arasındaki anlam

*Contact email: kkeskin@cu.edu.tr

ilişkileri gösterilerek pozitif ve negatif olarak sınıflandırılmıştır [4].

Türkçe dilinin yapısından dolayı oluşan problemleri çözmek için yapılan bu çalışmada türkçe kitap özetlerinin sınıflandırılması farklı kelime gömme teknikleri ve ön işleme metotları kullanılarak yapılmıştır ve başarı karşılaştırılmaları tablo halinde gösterilmiştir. Yapılan birçok çalışmada ön işleme yöntemleri denenmektedir ama farklı kombinasyonları gösterilmemektedir. Bu çalışmada yapılan ön işlemelerin kombinasyonları her bir kelime gömme yöntemi için uygulanmış ve her bir kombinasyon için en yüksek başarıyı veren modelin f-skoru ve başarı oranı tabloda gösterilmiştir.

Bu çalışmada, algoritmaların örüntüleri ayırt etme, temaları belirleme ve özetlerin içeriği hakkında tahminler yapma konusunda güçlendirilmesi amaçlanmaktadır. Kelime gömme yöntemleri ve makine öğreniminin bu sentezi, yalnızca kategorizasyon sürecini kolaylaştırmakla kalmıyor, aynı zamanda Türkçe dilinin inceliklerine dair daha derin içgörüler arayan araştırmacılar ve akademisyenler için dinamik bir araç sunuyor. Bu çalışmada gelişmiş Doğal dil işleme metodolojilerinin edebi analiz alanına entegrasyonu, yeni boyutları ortaya çıkarmayı, yazılı içerik anlayışımızı zenginleştirmeyi ve edebiyat araştırmalarında yenilikçi uygulamaların önünü açmayı vaat ediyor. Bu makale aracılığıyla kitap özeti sınıflandırması ve tahmini alanında kelime gömme teknikleri ve makine öğreniminin sinerjik potansiyeline ilişkin kapsamlı bir araştırma sunulmuştur.

2 Yöntemler

Bu çalışmada yöntemler ve kullanılan metotlar alt başlıklar halinde ele alınmıştır. Bu aşamalar veri setinin toplanması ve hazırlanması, önişleme işlemleri, kelime gömme tekniklerinin uygulanması, makine öğrenmesi modellerinin kullanılması, model ve kelime gömme tekniklerinin karşılaştırılmasından oluşur.

2.1 Veri setinin hazırlanması ve Önişleme

Kitap özetleri ve bilgilerini içeren veri seti, üç bin iki yüz bin veri ve sekiz kategoriden oluşmaktadır. Bu kategoriler "fantastik, bilim kurgu, romantik, tarih, polisiye, felsefe, sinema ve korku-gerilim" alt başlıklarından oluşmaktadır. Veriler popüler kitap satın alma sitesi olan idefix.com web sitesinden alınmıştır. Veri setinde sınıflandırma için kullanılan öznitelikler kitap özetleri ve kitap türleri sütunlarıdır.

Verilerin toplanmasının ardından verilerin daha iyi işlenmesi için doğal dil işlemede kullanılan bazı temel ön işleme metotları kullanılmıştır. Bu metotlar küçük harf dönüşümü, noktalama işaretlerinin, sayıların ve alfanümerik karakterlerin kaldırılması, kelimelerin köklerine ayırma ve duraksama kelimelerinin kaldırılmasından oluşmaktadır.

Küçük harf dönüşümü içeren veri setlerinde tüm kelimeler küçük harflere dönüştürülmüş ve makinenin büyük harfle başlayan kelimelerle karıştırmasının önüne geçilmiştir.

Noktalama işaretlerinin kaldırılmasını içeren veri setlerinde tüm noktalama işaretleri, alfanümerik karakterler ve sayılar sınıflandırma probleminde anlamlı olmadığından dolayı kaldırılmıştır.

Köklere ayırma yöntemi içeren veri setlerinde modelin başarısını arttırmak için kelimeler köklere indirgenmiştir. Köklere ayırma yöntemi için Turkish Stemmer for Python projesi Tuncelli ve Özdemir (2019) tarafından geliştirilmiştir ve Turkish Stemmer kök ayırma yöntemi olarak kullanılmıştır [7]. Ek olarak duraksama kelimelerinin (stopwords) makine öğreniminde anlamı olmadığı için Türkçedeki duraksama kelimeleri veri setinden kaldırılmıştır.

Ön işleme metotları aynı veri seti üzerinde farklı kombinasyonlarla uygulanıp kelime gömme teknikleri ve sınıflandırma yapılması için ayırılmıştır. Ön işleme tekniklerinin daha iyi gösterilmesi ve veri setinde uygulanan kombinasyonlarının karşılaştırılması için Tablo 1'de görülen şekilde kodlamalar oluşturulmuş ve sonuçlar kısmında kullanılmıştır.

Tablo 1. Önişleme Yöntemlerinin Kombinasyonları.

Senaryo Numarası	Kod	Küçük Harf Dönüşümü Yok (0)/Var (1)	Noktalama İşaretlerinin Kaldırılması Yok (0)/ Var (1)	Köklere Ayırma Yok (0)/ Var (1)	Duraksama Kelimelerinin Kaldırılması Yok (0) /Var (1)
1	0000	0	0	0	0
2	0001	0	0	0	1
3	0010	0	0	1	0
4	0011	0	0	1	1
5	0100	0	1	0	0
6	0101	0	1	0	1
7	0110	0	1	1	0
8	0111	0	1	1	1
9	1000	1	0	0	0
10	1001	1	0	0	1
11	1010	1	0	1	0
12	1011	1	0	1	1
13	1100	1	1	0	0
14	1101	1	1	0	1
15	1110	1	1	1	0
16	1111	1	1	1	1

2.2 Kelime Gömme Teknikleri

Günümüzde, doğal dil işleme ve yapay zeka alanlarındaki ilerlemeler kelime gömme tekniklerine olan ilgiyi artırmıştır. Kelime gömme, bir kelimenin bir vektör uzayında temsil edilmesidir. Bu teknikler bir kelimenin anlamını daha iyi anlamak, benzer anlamlı kelimeleri bulmak, dil modellerini eğitmek ve doğal dil işleme çalışmalarında başarıyı arttırmak için kullanılır. Bu makalede, kelime gömme tekniklerinden One Hot Encoding, Word2Vec ve Terim-Frekansı – Ters Doküman Frekansı (TF-IDF) yöntemleri sınıflandırma başarısını arttırmak ve model performanslarını ölçmek için kullanılmıştır.

2.2.1 One Hot Encoding

One hot encoding, kategorik verileri işlemek için yaygın olarak kullanılan bir kelime gömme tekniğidir. Bu teknik, kategorik değişkenleri her bir kategori için ayrı birer sütun olacak şekilde ikili vektörlerle temsil etmesini sağlar. Bir veri noktası bir kategoriye aitse o kategori için ilgili sütun 1'e, diğer tüm sütunlar ise 0'a ayarlanır.

One hot encoding, özellikle makine öğrenmesi algoritmalarında kategorik verilerle çalışırken sıklıkla kullanılır. Ancak çok sayıda farklı kategori varsa ve her bir kategori için bir sütun oluşturmak gerekiyorsa, bu teknik veri setinin boyutunu büyük ölçüde artırabilir ve bu da veri setinin işlenmesini zorlaştırabilir.

2.2.2 Terim Frekansı - Ters Doküman Frekans

Metin verilerini daha iyi anlamak ve içerdikleri bilgileri çıkarmak için çeşitli istatistiksel yöntemler kullanılır. Bu yöntemlerden biri de Terim Frekansı – Ters Doküman Frekansı (TF-IDF) adı verilen kelime gömme yöntemidir. Bu yöntemdeki kısımlardan Terim Frekansı (TF) bir belgedeki bir kelimenin frekansını belirten bir ölçüdür. Genellikle bir kelimenin bir belgedeki toplam kelime sayısına bölünerek hesaplanır. Bu, bir kelimenin belgedeki ne kadar sık geçtiğini gösterir. Ters Doküman Frekansı (IDF) ise bir kelimenin belirli bir belge

koleksiyonundaki nadirlik derecesini belirten bir ölçüdür. IDF, belge koleksiyonundaki belge sayısının, belirli bir kelimenin geçtiği belge sayısının logaritması alınarak hesaplanır. Bu, bir kelimenin ne kadar nadir veya yaygın olduğunu gösterir.

TF-IDF, metin verilerini analiz etmek için çeşitli uygulamalarda kullanılır. Özellikle belge sınıflandırma ve bilgi çıkarma gibi alanlarda önemli bir rol oynar. Yüksek TF-IDF skorlarına sahip kelimeler, belgenin içeriğindeki önemli kavramları ve anahtar bilgileri temsil eder.

$$\mathrm{Wa,b} = tf_{!,\#} \times \log \frac{N}{df_{!}} \quad (1)$$

$tf_{!,\#}$ = a kelimesinin b belgesindeki frekansı

$df_{!} = a\ kelimesini\ içeren\ d$öküman sayısı

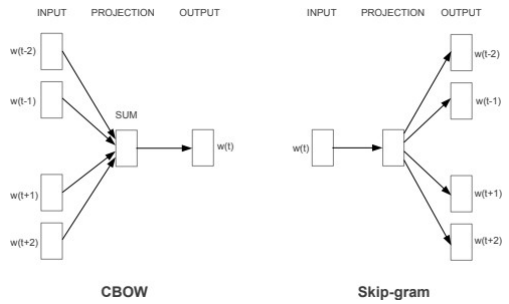
N = Toplam doküman sayısı

2.1.3 Word2Vec

Word2Vec, doğal dil işleme alanında kelime gömme yöntemlerinden biridir ve kelime temsillerini vektör uzayında oluşturur. Bu yöntem, bir kelimenin anlamını ve ilişkilerini temsil etmek için kullanılır. Word2Vec modeli, büyük metin veri setlerinden öğrenilir ve ardından her kelimeyi yoğun bir vektörle temsil eder.

Word2Vec'in temel amacı, bir kelimenin anlamını çevresindeki kelimelerle ilişkilendirerek vektör uzayında temsil etmektir. Bu ilişki, bir kelimenin belirli bir bağlam içindeki olasılığına dayanır. Bu yöntem ilk olarak Mikolov ve arkadaşları tarafından önerilmiştir [5]. Word2Vec'in iki alt yöntemi vardır. Bunlar Continous Bag of Words (CBOW) ve Skip-Gram yöntemleridir.

CBOW yöntemi bir kelimenin çevresindeki diğer kelimeler kullanılarak o kelimenin tahmin edilmesine dayanır. Yani, bir kelimeye bakılarak etrafındaki kelimelerin olasılıkları hesaplanır. Skip-gram yöntemi ise tam tersine, bir kelimenin verildiğinde bu kelimenin çevresindeki diğer kelimelerin tahmin edilmesine odaklanır. Yani, bir kelime verildiğinde, etrafındaki kelimelerin olasılıklarını tahmin eder. Şekil 1'de CBOW ve Skip-Gram yöntemlerinin çalışma yapıları gösterilmiştir. Word2Vec ek olarak büyük verilerin işlenmesi için de kullanılabilir. Yapılan bir çalışmada Word2Vec yöntemlerinden olan CBOW ve Skip-Gram yöntemleri büyük veriler üzerinde kullanılmış ve büyük verilerin boyutlarının azaltılmasında bir strateji olarak uygulanmıştır [6].



Şekil 1. CBOW ve Skip-Gram Yapısı

3 Makine Öğrenmesi Modelleri

Ön işleme ve kelime gömme tekniklerinin kullanılmasının ardından temel makine öğrenmesi modelleri kitap özetlerinin 8 kategoride sınıflandırılması için kullanılmıştır. Bu modeller K-En Yakın Komşu Algoritması (KNN), Naive Bayes (NB), Rastgele Karar Ormanı (RF), Karar Ağaçları (DT), Destek Vektör Makineleri (SVM), Lojistik Regresyon (LR) ve AdaBoost (AB) modelleridir.

3.1 K-En Yakın Komşu (KNN)

K-En Yakın Komşu (KNN), makine öğrenmesinde sınıflandırma ve regresyon problemlerini çözmek için kullanılan basit ancak güçlü bir algoritmadır. 1967 yılında T. M. Cover ve P. E. Hart tarafından önerilmiştir [8]. Bu algoritma, örneklerin birbirlerine olan benzerliklerine dayanarak tahminler yapar. KNN algoritması, basit bir prensibe dayanır: "Bir veri noktasını sınıflandırmak veya tahmin etmek için, o noktaya en yakın olan k komşulardan çoğunluğunu al." Burada "k" bir hiperparametredir ve kullanıcı tarafından belirlenir. Bu hiperparametre, komşuluk boyutunu belirler ve tahminlerin doğruluğunu etkiler.

Bir veri noktasının sınıflandırılması gerektiğinde, bu noktaya en yakın k komşu bulunur. Ardından, bu komşuların etiketleri kullanılarak sınıflandırma yapılır veya komşuların değerleri ortalama alınarak regresyon yapılır.

3.2 Naive Bayes (NB)

Naive Bayes (NB), makine öğrenmesinde sınıflandırma problemlerini çözmek için kullanılan olasılık temelli bir algoritmadır. Bu algoritma, Bayes teoremi prensibine dayanır ve özellikler arasındaki bağımsızlık varsayımını kabul eder. Sınıflandırma problemlerinden ikili sınıflandırmalarda, çoklu kategorik sınıflandırmalarda hatta dengelenmemiş sınıflar için de Naive Bayes makine öğrenmesi modeli kullanılmaktadır [9]. Naive Bayes algoritması, bir veri noktasının sınıfını belirlemek için özellikler arasındaki olasılıkları kullanır. Bu algoritma, Bayes

teoremi prensibine dayanır ve özellikler arasındaki bağımsızlık varsayımını kabul eder. Bu varsayıma göre, bir veri noktasının sınıfı, bu noktanın özelliklerinin sınıflar arasındaki olasılıkları kullanılarak belirlenir.

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) . P(A)}{P(B)} \quad (2)$$

P(A|B) = B olayının gerçekleştiği durumda A olayının gerçekleşme olasılığını ifade eder.

P(A) = A olayının gerçekleşme olasılığını belirtir.

P(B|A) = A olayının gerçekleştiği durumda B olayının gerçekleşme olasılığını ifade eder.

P(B) = B olayının gerçekleşme olasılığını ifade eder.

3.3 Rastgele Karar Ormanı (RF)

Rastgele Karar Ormanı algoritması, makine öğrenmesinde sınıflandırma ve regresyon problemlerini çözmek için kullanılan bir öğrenme algoritmasıdır. Bu algoritma, birden fazla karar ağacının oluşturulması ve bunların bir araya getirilmesiyle çalışır. Her karar ağacı, rastgele örneklem alınan veri noktalarıyla ve rastgele özelliklerle eğitilir. Ardından, her ağaç bir sınıf tahmini yapar ve bu tahminlerin çoğunluğu alınarak modelin tahmini belirlenir. Bu algoritma ayrıca geniş bir uygulama alanına sahiptir ve birçok endüstriyel ve akademik problemin çözümünde başarıyla kullanılmaktadır. Özellikle, büyük veri kümeleri ve yüksek boyutlu özellik uzaylarıyla çalışırken iyi performans gösterir. İngilizce haber metinlerinin sınıflandırılmasında klasik makine öğrenmesi modelleri ile karşılaştırıldığında Rastgele Karar Ormanı algoritması %0.93'lük bir başarı oranı ile öne çıkmıştır [10].

3.4 Karar Ağaçları (DT)

Karar ağaçları, bir veri kümesindeki özelliklerin değerlerine göre sınıfları veya değerleri tahmin etmek için kullanılan bir modelleme tekniğidir. Bir karar ağacı, veri kümesini bir dizi karar düğümünde bölerek sınıflara veya değerlere yönlendirir. Her bir karar düğümü, bir özelliğin değerini kontrol eder ve bu değere göre veriyi iki veya daha fazla alt kümeye ayırır. Bu bölünmeler, veri kümesinin daha küçük ve daha homojen alt gruplara ayrılmasını sağlar.

3.5 Destek Vektör Makinesi (SVM)

Destek vektör makineleri, sınıflandırma ve regresyon problemleri için kullanılan bir makine öğrenmesi algoritmasıdır. SVM, veri noktalarını sınıflandırmak için bir hiperdüzlem oluşturur. Bu hiperdüzlem, veri noktalarını iki veya daha fazla sınıfa ayırmak için en iyi şekilde ayıran bir çizgidir. SVM, sınıflar arasındaki marjı maksimize etmek için çalışır, yani en yakın veri noktaları arasındaki mesafeyi maksimize eder. SVM, lineer ve doğrusal olmayan sınıflandırma problemlerini çözebilir ve yüksek boyutlu veri kümelerinde etkilidir. Spam sms mesajlarının sınıflandırılması üzerine yapılan bir çalışmada klasik makine öğrenmesi modelleri karşılaştırılmış ve sonuçlara bakıldığında SVM modeli ortalama 98.9% oranında bir başarıyla spam mesajlarını tespit etmede öne geçmiştir [11]. Ayrıca SVM, çeşitli çekirdek fonksiyonları kullanarak doğrusal olmayan sınıflandırma problemlerinin çözümlerinde kullanılabilir. Yapılan başka bir çalışmada farklı değişken sayıları ile metin sınıflandırılması yapılmış ve klasik makine öğrenmesi modelleri ile karşılaştırılarak SVM modeli daha başarılı olmuştur [12].

3.6 Lojistik Regresyon (LR)

Lojistik regresyon, bir bağımlı değişkenin olasılık dağılımını, bağımsız değişkenlerin bir lineer kombinasyonu ile tahmin etmek için kullanılan bir istatistiksel modeldir. Lojistik regresyon, sınıflandırma problemleri için yaygın olarak kullanılır. İkili sınıflandırma problemlerinde, lojistik regresyon çıktıyı 0 ile 1 arasında bir olasılık değeri olarak verir. Bu olasılık değeri, bir veri noktasının bir sınıfa ait olma olasılığını belirtir. Lojistik regresyon, lineer olmayan ilişkileri modellemek için kullanılan lineer bir modeldir. Lojistik regresyon ayrıca birçok kelime gömme tekniği ile uyumlu olarak kullanılabilir. Sosyal medya yorumları ile yapılan duygu analizi çalışmasında lojistik regresyon modeli, kelime gömme tekniklerinden kelime çantası tekniği (bag of words) için en başarılı sonucu verirken TF-IDF tekniği için ise ikinci en başarılı sonucu vermiştir [13]. Ayrıca duygu sınıflandırılması için yapılan başka bir çalışmada derin öğrenme modellerinden Uzun-Kısa Süreli Bellek (LSTM) modeli ve lojistik regresyon karşılaştırılmış ve lojistik regresyon modeli hem Twitter hem de IMDB veri setleri için daha başarılı sonuçlar vermiştir [14].

3.7 AdaBoost (AB)

AdaBoost, genellikle karar ağaçlarını bir araya getirerek güçlü bir sınıflandırıcı oluşturan bir öğrenme algoritmasıdır. AdaBoost, her bir öğreniciyi

sırayla eğitir ve daha fazla vurgu verilen yanlış sınıflandırılmış örnekleri belirler. Daha sonra, yanlış sınıflandırılmış örneklerin üzerine daha fazla vurgu yaparak yeni bir öğrenici eğitir. Bu şekilde, AdaBoost, daha fazla vurgu gerektiren örnekler üzerinde daha güçlü bir öğrenme gerçekleştirir. AdaBoost, yüksek doğruluk sağlayabilir ve genellikle diğer sınıflandırma algoritmalarıyla birlikte kullanıldığında performansını artırabilir. Dengesiz bir veri seti için yapılan sınıflandırma çalışmasında AdaBoost algoritması performansı arttırmak için önerilen bir metotla kullanılmış ve başarılı sonuç vermiştir [15].

4 Bulgular ve Yorumlar

Bu çalışmada kitap özetleri veri seti ön işlemlerden geçirildikten sonra her bir kombinasyonu içeren veri setine farklı kelime gömme teknikleri uygulanmıştır. Daha sonra kelime gömme teknikleri uygulanan her bir veri seti kombinasyonu için temel makine öğrenmesi modelleri sınıflandırma yapılmak üzere uygulanmış ve Tablo 3'te detaylı bir şekilde gösterilmiştir. Ayrıca en başarılı makine öğrenmesi modeline göre oluşturulan karşılaştırılmalar kelime gömme teknikleri ile birlikte Tablo 2'de gösterilmiştir. Değerlendirme kriterleri olarak çalışmada 4 adet parametre kullanılmıştır: F-Skor (FS), Doğruluk (ACC), Precision (PRC) ve Recall (RC). Bulgulara bakıldığında ön işleme metotlarından köklerine ayıma yönteminin bütün kelime gömme teknikleri için model başarısını arttırdığı görülmektedir. TF-IDF tekniği uygulanan modeller için en yüksek başarı ve F-Skor, SVM modelinde 0.8 olarak görülmüştür. Ayrıca diğer ön işleme kombinasyonlarında da SVM modeli diğer makine öğrenmesi modellerine göre daha iyi sonuç vermiştir. Word2Vec uygulanan veri setlerindeki başarı oranlarına bakıldığında lojistik regresyon modeli ve SVM diğer öğrenme modellerinden daha başarılı sonuçlar verdiği gözlenmiştir. En yüksek başarı ve F-Skor 0010 kombinasyonu için 0.72 olarak LR modelinde görülmüştür. Bunun nedenleri arasında lojistik regresyon modelinin girdi olarak Wod2Vec ile oluşturulan kelime vektörlerini alarak metinlerin daha iyi bir şekilde temsil edilmesini sağlaması gösterilebilir. Bu, modelin daha iyi özellik temsili elde etmesine ve dolayısıyla daha iyi performans göstermesine olanak tanır. Son olarak One-Hot Encoding yöntemi kullanılan veri setleri üzerinde en başarılı sonuç ve F-Skor 1010 ve 1011 kombinasyonları için Naive Bayes modelinde 0.81 olarak görülmüştür.

Tablo 2. Kelime Gömme Teknikleri ve Ön İşleme Kombinasyonlarının En Başarılı Modeller ile Karşılaştırılması

	TF-IDF			Word2Vec			One Hot Encoding		
Kod	FS	ACC	Model	FS	ACC	Model	FS	ACC	Model
0000	0.78	0.77	SVM	0.68	0.68	LR	0.79	0.78	NB
0001	0.79	0.79	SVM	0.67	0.67	LR	0.8	0.8	NB
0010	0.79	0.79	SVM	0.71	0.72	LR	0.8	0.8	LR
0011	0.8	0.8	SVM	0.7	0.71	LR	0.8	0.8	LR
0100	0.77	0.77	SVM	0.68	0.68	LR	0.78	0.78	NB
0101	0.77	0.77	SVM	0.68	0.68	LR	0.78	0.78	NB
0110	0.79	0.79	SVM	0.7	0.7	LR	0.79	0.78	LR
0111	0.79	0.79	SVM	0.7	0.7	LR	0.79	0.78	LR
1000	0.78	0.77	SVM	0.67	0.68	LR	0.79	0.78	NB
1001	0.79	0.78	SVM	0.67	0.67	SVM	0.79	0.79	NB
1010	0.8	0.8	SVM	0.69	0.7	LR	0.81	0.81	NB
1011	0.8	0.8	SVM	0.7	0.7	LR	0.81	0.81	NB
1100	0.77	0.77	SVM	0.67	0.68	LR	0.78	0.78	NB
1101	0.79	0.78	SVM	0.68	0.68	LR	0.78	0.78	NB-LR
1110	0.79	0.79	SVM	0.69	0.69	LR	0.79	0.8	LR
1111	0.79	0.79	SVM	0.7	0.7	LR	0.8	0.8	NB-LR

Tablo 3. One-Hot Encoding, TF-IDF, Word2Vec Modelleri ve Ön İşleme Kombinasyonlarının Makine Öğrenmesi Modelleri ile Karşılaştırılması

	Algoritmalar	**Değerlendirme Kriterleri**

Önişlem Kodu		One-Hot Encoding				TF-IDF				Word2Vec			
		FS	ACC	PRC	RC	FS	ACC	PRC	RC	FS	ACC	PRC	RC
0000	KNN [k=15]	0.16	0.07	0.33	0.16	0.069	0.14	0.32	0.14	0.54	0.54	0.57	0.54
	NB	0.79	0.78	0.8	0.78	0.75	0.75	0.77	0.76	0.48	0.48	0.51	0.47
	RF	0.74	0.74	0.75	0.74	0.69	0.68	0.71	0.69	0.57	0.57	0.57	0.57
	DT	0.59	0.59	0.6	0.59	0.55	0.54	0.55	0.55	0.38	0.38	0.39	0.38
	SVM	0.71	0.72	0.72	0.72	0.78	0.77	0.78	0.78	0.67	0.67	0.68	0.67
	LR	0.78	0.78	0.78	0.78	0.76	0.75	0.77	0.76	0.68	0.68	0.68	0.68
	AB	0.43	0.43	0.79	0.38	0.46	0.46	0.73	0.43	0.31	0.31	0.35	0.33
0001	KNN [k=15]	0.16	0.07	0.33	0.16	0.08	0.14	0.32	0.15	0.55	0.55	0.58	0.55
	NB	0.8	0.8	0.82	0.79	0.77	0.77	0.78	0.77	0.49	0.49	0.52	0.48
	RF	0.75	0.75	0.76	0.75	0.69	0.68	0.71	0.69	0.6	0.6	0.61	0.6
	DT	0.59	0.59	0.59	0.59	0.55	0.54	0.56	0.55	0.36	0.36	0.37	0.36
	SVM	0.71	0.72	0.72	0.72	0.79	0.79	0.79	0.79	0.66	0.66	0.67	0.66
	LR	0.78	0.78	0.78	0.78	0.77	0.76	0.78	0.77	0.67	0.67	0.68	0.67
	AB	0.43	0.43	0.79	0.38	0.45	0.45	0.74	0.42	0.31	0.31	0.32	0.36
0010	KNN [k=15]	0.16	0.07	0.33	0.16	0.07	0.14	0.31	0.14	0.6	0.6	0.62	0.6
	NB	0.79	0.79	0.81	0.79	0.76	0.76	0.79	0.77	0.52	0.52	0.55	0.51
	RF	0.74	0.75	0.76	0.75	0.71	0.71	0.72	0.72	0.59	0.59	0.6	0.59
	DT	0.61	0.61	0.61	0.61	0.55	0.55	0.55	0.55	0.38	0.38	0.38	0.37
	SVM	0.75	0.75	0.75	0.75	0.79	0.79	0.8	0.8	0.68	0.68	0.68	0.67
	LR	0.8	0.8	0.8	0.8	0.77	0.77	0.79	0.78	0.71	0.72	0.72	0.71
	AB	0.48	0.48	0.67	0.45	0.41	0.41	0.55	0.43	0.28	0.28	0.3	0.31
0011	KNN [k=15]	0.06	0.15	0.22	0.14	0.07	0.14	0.31	0.15	0.6	0.6	0.62	0.59
	NB	0.79	0.79	0.81	0.79	0.78	0.77	0.79	0.78	0.53	0.53	0.56	0.52
	RF	0.75	0.76	0.76	0.76	0.73	0.73	0.75	0.74	0.61	0.61	0.61	0.6
	DT	0.6	0.6	0.6	0.6	0.56	0.56	0.57	0.57	0.39	0.39	0.41	0.39
	SVM	0.76	0.76	0.76	0.76	0.8	0.8	0.8	0.8	0.68	0.68	0.68	0.67
	LR	0.8	0.8	0.8	0.8	0.79	0.78	0.8	0.79	0.7	0.71	0.71	0.7
	AB	0.48	0.48	0.67	0.45	0.5	0.5	0.59	0.5	0.29	0.29	0.31	0.31
0100	KNN [k=15]	0.08	0.17	0.2	0.17	0.07	0.14	0.31	0.15	0.52	0.53	0.55	0.52
	NB	0.78	0.78	0.8	0.78	0.74	0.74	0.77	0.75	0.46	0.46	0.5	0.46
	RF	0.74	0.75	0.76	0.75	0.68	0.67	0.7	0.68	0.56	0.56	0.57	0.56
	DT	0.58	0.58	0.58	0.58	0.55	0.54	0.55	0.54	0.34	0.34	0.34	0.34
	SVM	0.71	0.71	0.71	0.71	0.77	0.77	0.78	0.78	0.66	0.66	0.67	0.65
	LR	0.77	0.77	0.77	0.77	0.75	0.75	0.77	0.76	0.68	0.68	0.68	0.68
	AB	0.44	0.44	0.77	0.4	0.34	0.34	0.69	0.34	0.3	0.3	0.37	0.31
0101	KNN [k=15]	0.08	0.17	0.2	0.17	0.07	0.14	0.31	0.15	0.53	0.53	0.56	0.53
	NB	0.78	0.78	0.8	0.78	0.74	0.74	0.77	0.75	0.46	0.46	0.5	0.46
	RF	0.73	0.73	0.75	0.73	0.67	0.67	0.69	0.68	0.55	0.55	0.56	0.56

	DT	0.59	0.59	0.6	0.59	0.55	0.54	0.55	0.55	0.35	0.35	0.36	0.35
	SVM	0.71	0.71	0.71	0.71	0.77	0.77	0.78	0.78	0.66	0.66	0.67	0.65
	LR	0.77	0.77	0.77	0.77	0.75	0.75	0.77	0.76	0.68	0.68	0.68	0.68
	AB	0.44	0.44	0.77	0.4	0.34	0.34	0.69	0.34	0.3	0.3	0.37	0.31
0110	KNN [k=15]	0.08	0.18	0.26	0.17	0.06	0.14	0.31	0.14	0.58	0.58	0.6	0.58
	NB	0.79	0.78	0.8	0.79	0.75	0.74	0.78	0.75	0.52	0.52	0.54	0.51
	RF	0.73	0.74	0.75	0.74	0.7	0.7	0.72	0.71	0.59	0.59	0.59	0.58
	DT	0.6	0.6	0.6	0.6	0.54	0.53	0.54	0.54	0.38	0.38	0.38	0.37
	SVM	0.75	0.75	0.75	0.75	0.79	0.79	0.8	0.79	0.67	0.67	0.68	0.66
	LR	0.8	0.8	0.8	0.8	0.77	0.77	0.79	0.78	0.7	0.7	0.71	0.7
	AB	0.42	0.42	0.53	0.43	0.49	0.49	0.57	0.5	0.31	0.31	0.35	0.33
0111	KNN [k=15]	0.08	0.18	0.26	0.17	0.06	0.14	0.31	0.14	0.58	0.58	0.6	0.58
	NB	0.79	0.78	0.8	0.79	0.75	0.74	0.78	0.75	0.52	0.52	0.54	0.51
	RF	0.74	0.74	0.75	0.74	0.69	0.69	0.72	0.7	0.59	0.59	0.59	0.59
	DT	0.6	0.6	0.6	0.6	0.55	0.54	0.55	0.55	0.38	0.38	0.39	0.37
	SVM	0.75	0.75	0.75	0.75	0.79	0.79	0.8	0.79	0.67	0.67	0.68	0.66
	LR	0.8	0.8	0.8	0.8	0.77	0.77	0.79	0.78	0.7	0.7	0.71	0.7
	AB	0.42	0.42	0.53	0.43	0.49	0.49	0.57	0.5	0.31	0.31	0.35	0.33
1000	KNN [k=15]	0.07	0.16	0.33	0.16	0.06	0.14	0.31	0.14	0.54	0.54	0.57	0.54
	NB	0.79	0.78	0.8	0.78	0.75	0.75	0.77	0.76	0.48	0.48	0.51	0.47
	RF	0.74	0.74	0.76	0.75	0.7	0.69	0.72	0.7	0.57	0.57	0.57	0.56
	DT	0.59	0.59	0.59	0.59	0.54	0.54	0.54	0.54	0.34	0.35	0.34	0.34
	SVM	0.71	0.72	0.72	0.72	0.78	0.77	0.78	0.78	0.67	0.67	0.68	0.67
	LR	0.78	0.78	0.78	0.78	0.76	0.75	0.77	0.76	0.67	0.68	0.68	0.67
	AB	0.43	0.43	0.79	0.38	0.46	0.46	0.73	0.43	0.31	0.31	0.35	0.33
1001	KNN [k=15]	0.07	0.16	0.13	0.16	0.07	0.14	0.31	0.15	0.54	0.54	0.57	0.54
	NB	0.79	0.79	0.8	0.79	0.77	0.76	0.78	0.77	0.48	0.49	0.52	0.48
	RF	0.76	0.76	0.77	0.76	0.69	0.68	0.71	0.69	0.59	0.59	0.6	0.59
	DT	0.57	0.58	0.58	0.58	0.55	0.54	0.55	0.55	0.35	0.35	0.36	0.35
	SVM	0.71	0.72	0.72	0.71	0.79	0.78	0.79	0.79	0.67	0.67	0.68	0.67
	LR	0.78	0.79	0.79	0.79	0.77	0.76	0.78	0.77	0.67	0.67	0.68	0.67
	AB	0.43	0.43	0.79	0.38	0.44	0.44	0.71	0.41	0.33	0.33	0.34	0.36
1010	KNN [k=15]	0.12	0.22	0.32	0.22	0.07	0.14	0.31	0.15	0.57	0.57	0.6	0.57
	NB	0.81	0.81	0.83	0.8	0.76	0.76	0.78	0.77	0.51	0.51	0.54	0.51
	RF	0.75	0.76	0.76	0.75	0.7	0.7	0.74	0.71	0.59	0.59	0.6	0.59
	DT	0.61	0.61	0.61	0.61	0.56	0.56	0.56	0.57	0.38	0.38	0.39	0.38
	SVM	0.73	0.73	0.73	0.73	0.8	0.8	0.8	0.8	0.68	0.68	0.69	0.68
	LR	0.8	0.8	0.8	0.8	0.78	0.77	0.79	0.78	0.69	0.7	0.7	0.69
	AB	0.47	0.47	0.78	0.46	0.5	0.5	0.57	0.51	0.32	0.32	0.37	0.35
1011	KNN [k=15]	0.1	0.2	0.31	0.2	0.07	0.14	0.31	0.15	0.59	0.59	0.61	0.58

	NB	0.81	0.81	0.83	0.8	0.77	0.77	0.79	0.78	0.55	0.55	0.57	0.54
	RF	0.76	0.76	0.77	0.76	0.72	0.72	0.73	0.73	0.6	0.6	0.6	0.6
	DT	0.63	0.63	0.63	0.63	0.56	0.55	0.56	0.56	0.35	0.35	0.35	0.34
	SVM	0.74	0.74	0.75	0.74	0.8	0.8	0.81	0.81	0.67	0.67	0.68	0.67
	LR	0.8	0.8	0.8	0.8	0.79	0.79	0.8	0.79	0.7	0.7	0.71	0.7
	AB	0.47	0.47	0.78	0.46	0.49	0.49	0.56	0.49	0.3	0.3	0.32	0.32
1100	KNN [k=15]	0.08	0.17	0.2	0.17	0.07	0.14	0.31	0.15	0.53	0.53	0.56	0.53
	NB	0.78	0.78	0.8	0.78	0.74	0.74	0.77	0.75	0.46	0.46	0.5	0.46
	RF	0.73	0.74	0.75	0.74	0.66	0.66	0.68	0.67	0.57	0.57	0.57	0.57
	DT	0.58	0.58	0.58	0.58	0.54	0.53	0.54	0.53	0.35	0.35	0.36	0.35
	SVM	0.71	0.71	0.71	0.71	0.77	0.77	0.78	0.78	0.66	0.66	0.67	0.65
	LR	0.77	0.77	0.77	0.77	0.75	0.75	0.77	0.76	0.67	0.68	0.67	0.67
	AB	0.44	0.44	0.77	0.4	0.34	0.34	0.69	0.34	0.3	0.3	0.37	0.31
1101	KNN [k=15]	0.07	0.16	0.2	0.16	0.07	0.14	0.31	0.15	0.55	0.55	0.58	0.55
	NB	0.78	0.78	0.8	0.78	0.76	0.76	0.77	0.77	0.5	0.5	0.53	0.49
	RF	0.72	0.73	0.75	0.73	0.69	0.68	0.7	0.69	0.58	0.58	0.59	0.57
	DT	0.58	0.58	0.58	0.58	0.53	0.52	0.53	0.53	0.36	0.36	0.37	0.36
	SVM	0.72	0.72	0.72	0.72	0.79	0.78	0.79	0.79	0.67	0.67	0.68	0.67
	LR	0.78	0.78	0.78	0.78	0.76	0.76	0.77	0.77	0.68	0.68	0.69	0.68
	AB	0.44	0.44	0.77	0.4	0.39	0.39	0.76	0.38	0.33	0.33	0.33	0.36
1110	KNN [k=15]	0.11	0.21	0.26	0.2	0.07	0.14	0.31	0.15	0.58	0.58	0.61	0.58
	NB	0.79	0.79	0.8	0.79	0.76	0.76	0.79	0.77	0.52	0.52	0.55	0.51
	RF	0.76	0.77	0.78	0.77	0.7	0.69	0.73	0.7	0.59	0.59	0.59	0.58
	DT	0.6	0.6	0.6	0.6	0.59	0.58	0.59	0.59	0.38	0.38	0.39	0.37
	SVM	0.73	0.73	0.74	0.73	0.79	0.79	0.8	0.8	0.67	0.67	0.69	0.67
	LR	0.79	0.8	0.8	0.8	0.78	0.78	0.79	0.79	0.69	0.69	0.7	0.69
	AB	0.51	0.51	0.69	0.49	0.5	0.5	0.6	0.51	0.35	0.35	0.37	0.36
1111	KNN [k=15]	0.1	0.21	0.35	0.21	0.07	0.14	0.31	0.15	0.58	0.58	0.6	0.57
	NB	0.8	0.8	0.82	0.8	0.78	0.77	0.79	0.78	0.54	0.54	0.56	0.53
	RF	0.75	0.76	0.77	0.76	0.73	0.72	0.75	0.73	0.59	0.59	0.59	0.59
	DT	0.6	0.61	0.61	0.61	0.59	0.59	0.59	0.59	0.38	0.38	0.39	0.38
	SVM	0.74	0.74	0.74	0.74	0.79	0.79	0.8	0.8	0.67	0.67	0.68	0.67
	LR	0.8	0.8	0.8	0.8	0.78	0.78	0.79	0.79	0.7	0.7	0.71	0.7
	AB	0.51	0.51	0.69	0.49	0.5	0.5	0.59	0.51	0.32	0.32	0.34	0.34

5 Sonuçlar

Bu çalışmada farklı kelime gömme tekniklerinin, ön işleme kombinasyonlarının ve makine öğrenmesi modellerinin performansını karşılaştırdık. Tablo 2 ve Tablo 3'de görüldüğü üzere TF-IDF yöntemiyle kullanılan destek vektör makinesi modeli, diğer modellere kıyasla %0.8'lik bir başarı oranıyla öne çıktı. Ön işleme tekniklerinden köklerine ayırma yönteminin başarıyı artırdığı gözlemlendi. Word2vec kelime gömme yöntemi kullanan modellerde ise lojistik regresyon modeli

diğerlerinden daha başarılı oldu (%0.72). Genel olarak ön işleme tekniklerinin kullanılması başarı oranlarında büyük bir değişikliğe sebebiyet vermedi. One hot encoder kelime gömme metodu kullanan modellerde Naive Bayes ve LR modelleri, farklı ön işleme yöntemleriyle benzer başarı oranları elde ederek %0.81 ve %0.80 oranlarında başarılı oldular. Bu çalışma, birçok doğal dil işleme tekniğinin bir arada kullanıldığı ve farklı kombinasyonlarının karşılaştırıldığı bir alana ışık tutmaktadır ve diğer çalışmalardan farklı olarak öne çıkmaktadır. Gelecekte yapılacak olan daha kapsamlı veri setleri ve farklı kelime gömme ve derin öğrenme tekniklerinin eklenmesiyle doğal dil işleme çalışmalarında bu bulguların rehberlik sağlayabileceğini ve geliştirilebileceğini düşünüyoruz.

Kaynaklar

[1] Çelik, Ö. & Koç, B. C. TF-IDF, Word2vec ve Fasttext Vektör Model Yöntemleri ile Türkçe Haber Metinlerinin Sınıflandırılması. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi 23, 121–127 (2021).

[2] Karcioğlu, A. A. & Aydin, T. Sentiment analysis of Turkish and english twitter feeds using Word2Vec model in 2019 27th Signal Processing and Communications Applications Conference (SIU) (2019), 1–4.

[3] Lilleberg, J., Zhu, Y. & Zhang, Y. Support vector machines and word2vec for text classification with semantic features in 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC) (2015), 136–140.

[4] Polpinij, J., Srikanjanapert, N. & Sopon, P. Word2Vec approach for sentiment classification relating to hotel reviews in Recent Advances in Information and Communication Technology 2017: Proceedings of the 13th International Conference on Computing and Information Technology (IC2IT) (2018), 308–316.

[5] Mikolov T, Chen K, Corrado G, Dean J. (2013), "Efficient estimation of word representations in vector space". Proceedings of Workshop at ICLR. Scottsdale, Arizona 2-4 Mayıs 2013.

[6] Long Ma and Yanqing Zhang. Using word2vec to process big text data. In 2015 IEEE International Conference on Big Data (Big Data), pages 2895–2897. IEEE, 2015.

[7] Tuncelli, O., & Özdemir, B. (2019). Turkish Stemmer for Python. GitHub. https://github.com/otuncelli/turkishstemmer-python.

[8] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27.

[9] Frank, E., & Bouckaert, R. R. (2006). Naive bayes for text classification with unbalanced classes. In Knowledge Discovery in Databases: PKDD 2006: 10th European Conference on Principles and Practice of Knowledge Discovery in Databases Berlin, Germany, September 18-22, 2006 Proceedings 10 (pp. 503-510). Springer Berlin Heidelberg.

[10] Shah, K., Patel, H., Sanghvi, D., & Shah, M. (2020). A comparative analysis of logistic regression, random forest and KNN models for the text classification. Augmented Human Research, 5(1), 12.

[11] Sjarif, N. N. A., Yahya, Y., Chuprat, S., & Azmi, N. H. F. M. (2020). Support vector machine algorithm for SMS spam classification in the telecommunication industry. Int. J. Adv. Sci. Eng. Inf. Technol, 10(2), 635-639.

[12] Luo, X. (2021). Efficient English text classification using selected machine learning techniques. Alexandria Engineering Journal, 60(3), 3401-3409.

[13] Wicaksono, F. A., & Romadhony, A. (2022). Sentiment analysis of university social media using support vector machine and logistic regression methods. Indonesia Journal on Computing (Indo-JC), 7(2), 15-24.

[14] Vadivu, G. Comparative Study of Logistic Regression and LSTM for Sentiment Classification Across Diverse Textual Dataset. parameters, 1, 4.

[15] Wang, W., & Sun, D. (2021). The improved AdaBoost algorithms for imbalanced data classification. Information Sciences, 563, 358-374.

A Study on the Use of Neural Networks for LHC Data Interpretation

Serpil YALÇIN KUZU [1], Ayben KARASU UYSAL[2], Mustafa KAYA[1*]

[1] *Fırat Üniversitesi, Fizik Bölümü, Elazığ, TÜRKİYE*
[2] *Yıldız Teknik Üniversitesi, Fizik Bölümü, İstanbul, TÜRKİYE*

Abstract

Understanding the quarkonium states is important for the advancement of particle physics and the validation of quantum theories. The decay of J/ψ mesons into electron-positron pairs ($J/\psi \rightarrow e^+e^-$) provides valuable insights into the mechanisms produced in high energy collisions. In the present work, implementation of Deep Neural Networks (DNNs) to enhance the analysis and identification of the charmonium events using data from the Large Hadron Collider (LHC) was studied. The power and usability of DNNs to process and classify complex LHC data was investigated by applying data preprocessing, feature engineering, and hyperparameter tuning. The effectiveness of the artificial neural network model is demonstrated through evaluation metrics such as precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC ROC), exemplifying its capability in identifying particles in complex high energy physics datasets.

Keywords: *J/ψ, Artificial Neural Network, Particle Physics.*

1 Introduction

The usage of machine learning (ML) technologies has revolutionized various scientific disciplines such as particle and high-energy physics. The Large Hadron Collider (LHC), the world's largest and most powerful particle accelerator, generates an unprecedented amount of data, presenting both opportunities and challenges for physicists.Traditional analysis techniques, while effective, are increasingly being supplemented or replaced by more advanced methods, such as neural networks, to cope with the complexity and volume of the data [1-3]. The focus on neural networks, specifically Deep Neural Networks (DNNs), has grown due to their ability to learn from large datasets, detect patterns, and make predictions with high accuracy [4].

Understanding quarkonium states, particularly the J/ψ mesons, is critical for the validation of quantum chromodynamics (QCD), the theory describing the strong interaction between quarks and gluons. The decay of J/ψ mesons into electron (e^-)-positron (e^+) pairs ($J/\psi \rightarrow e^+e^-$) offers a unique window into the dynamics of high-energy collisions, making it a subject of significant interest in particle physics [2, 5]. Accurate identification and analysis of these events are essential for advancing the physics knowledge.

Deep Neural Networks (DNNs) provide a powerful tool for analyzing and interpreting the complex datasets produced by the LHC. Unlike traditional algorithms, DNNs can automatically discover the representations needed for feature detection from raw data without requiring extensive manual feature engineering [6]. This capability is particularly beneficial in the context of high-energy physics, where the data is often noisy and high-dimensional.

In the present study, the application of DNNs to enhance the analysis and identification of charmonium events, with a particular focus on the J/ψ meson decay process is discussed. The study emphasizes several critical components necessary for the successful application of DNNs in this context, including data preprocessing, feature engineering, and hyperparameter tuning. These steps are vital to optimizing the neural network's performance and ensuring that it accurately models the underlying physics [2, 3, 7].

The effectiveness of the DNN model is evaluated using a range of performance metrics, such as precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC ROC). These metrics provide a comprehensive view of the model's ability to correctly classify particles in the highly complex datasets produced by the LHC [8]. Previous research has shown the potential of machine learning, including DNNs, to significantly improve the accuracy and efficiency of particle identification in similar contexts [1]. By building on these advances, this study aims to contribute further to the field by demonstrating the utility of DNNs in quarkonium state analysis.

*Contact email: mustafakaya23100@gmail.com

2 Experimental Framework and Methodology

2.1 Deep Neural Networks (DNNs)

Deep Neural Networks (DNNs) consist of multiple layers of neurons, where each layer transforms input data into more abstract and useful representations, making them well-suited for tasks that involve large and noisy datasets, such as those produced by LHC experiments [1-3]. The application of DNNs involves several key steps, starting with data preprocessing and feature selection. The architecture of the DNNs typically includes an input layer, multiple hidden layers with nonlinear activation functions like The Rectified Linear Unit (ReLU) activation function [9], and an output layer that uses a Sigmoid function [10] for classification tasks. This layered structure allows the DNNs to capture complex patterns in the data, which is crucial for identifying specific events, such as the decay of J/ψ mesons [3, 4]. The training process involves compiling the DNNs with a binary cross entropy function, which is suitable for binary classification tasks. In the algorithm an optimizer like Adam [11] is used for its efficiency in handling sparse gradients to prevent overfitting.

Algorithm 1 DNNs for 6 layers including 1 input, 4 hidden, and 1 output layers.

Input: Preprocessed LHC data (X), Labels (y)
Output: Trained DNNs Model (DNNs)
Step 1: Initialize DNNs with architecture:
  Input Layer: size = number of features in X
  Hidden Layers:
    Layer 1: 512 neurons, ReLU activation
    Layer 2: 256 neurons, ReLU activation
    Layer 3: 128 neurons, ReLU activation
    Layer 4: 64 neurons, ReLU activation
  Output Layer: size = number of classes in y, Sigmoid activation
Step 2: Compile DNNs with:
  Loss function = Binary Cross-Entropy
  Optimizer = Adam
  Metrics = [Precision, Recall, F1-score, AUC ROC]
Step 3: Split X and y into training set (X_train, y_train)
Step 4: Train DNNs on X_train, y_train with:
  Batch size = 128
  Epochs = 100
Step 5: Evaluate DNNs on test set if available
Return DNNs

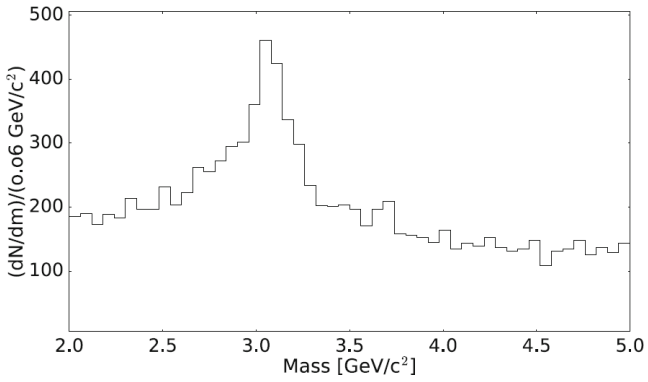


Figure 1. Mass distribution of pairs in the dataset.

The pseudo-algorithm for the DNNs process can be summarized in Algorithm 1. In this pseudo-algorithm, important parameters such as the number of neurons in each hidden layer, the learning rate, and the batch size are carefully chosen to optimize the model's performance. The ReLU activation function is employed in the hidden layers due to its efficiency and ability to introduce non-linearity, which is crucial for learning complex patterns. The Adam optimizer, known for its adaptability and efficiency, is used to minimize the loss function during training. Selection of epoch number and batch size are key parameters to improve the model performance. To prevent overfitting, the Dropout method can be used in the algorithm during training that randomly removes a fraction of neurons, which forces the network to develop more robust features by improving generalization ability of the DNNs [12].

2.2 Experimental Design

2.2.1 Dataset

In this study proton–proton collision data collected by the CMS experiment at $\sqrt{s}$ = 7 TeV in 2010 were used to reconstruct J/ψ from its dielectron decay channel [13-15]. The corresponding channel was investigated from 2 to 5 GeV.

2.2.2 Preprocessing and Feature Selection

In the analysis of J/ψ meson identification, careful preprocessing and feature selection are critical to enhance the accuracy of the DNNs model. For this purpose, the data having missing information were removed from the dataset. 10015 events were analyzed with Python [16]. For DNNs application Keras [17] and TensorFlow [18] packages were used.

Extracting the particle signal distribution from the extensive background is a significant challenge in particle physics analysis due to the overlap between these distributions. Traditional analysis methods typically involve applying specific selection cuts on particle features to reject most of the background. However, these cuts often lead to the loss of particle signal information and introduce high systematic uncertainties. In contrast, the application of the DNNs leverages the inherent power of each feature to discriminate the signal, thus minimizing the need for such cuts and reducing the associated uncertainties.

In the analysis, information of each decay partner, including momentum components (p_x, p_y, p_z), transverse momentum (p_t), energy (E), pseudorapidity (η), azimuthal angle (ϕ), the charges (q), and the mass (M) of the pairs were provided by the dataset. Mass distribution of the pairs in the dataset is represented in Figure 1. In the analysis instead of using η and ϕ information, the absolute differences between two particles (Δη and Δϕ) were calculated and used as features to determine the correlation between particles within the same event [19-21]. Also, the opening angle (α) between the partners [19] and the transverse momentum of the mother particle (P_t) were calculated and used as features.

In the study, *J/ψ*, the signal class, defined by pairs including electrons and positrons having track properties such as each partner must have p_t greater than 1.3 GeV/c and charge in opposite sign, which are the regular cuts to increase the performance of the particle measurement [19]. An optimized mass range description is necessary to distinguish signal and background classes. Mass of the meson is 3096.9±0.006 MeV/c^2 [22], therefore, the signal must be on a narrow mass window within ±300 MeV/c^2 of this value in order to have background contamination as low as possible. With these definitions signal and background instances are listed in Table 1.

Table 1. List of signal and background class numbers.

Class Labels	Number of Instances
Signal (*J/ψ*)	3345
Background	8251

2.2.3 DNNs Implementation

In the analysis 75% of the data were used for training and the rest for the test. The features that was used in the DNNs algorithm to identify signal class were selected according to the traditional charmonium analysis. Therefore, α, Δη, Δϕ, M, P_t, and q and p_t of each pair partner were used in the DNNs to enhance reliability of the model in complex particle dataset.

In the DNNs, the Adam algorithm was preferred due to its performance in minimizing the difference between the output value produced by the network and the actual value, as well as its ability to prevent overfitting. The ReLU function was selected as the activation function for the input and hidden layers. The Sigmoid function was used as the activation function for the output layer. Since there are two classes in the dataset, 'binary cross entropy' was preferred as the loss function. To prevent overfitting in the model, the layers were organized using the Dropout method with a dropout rate of 0.2, and the kernel initializer [23] suitable for the network structure was set to 'normal'. In the end of the optimization process, the DNNs was structured with a total of 5 layers: one input layer, one output layer, and three hidden layers. The number of nodes for each layer was set to 9, 150, 150, 20, and 1, respectively. Based on the studies, the batch size, which is the number of samples to be fed into the network at once during the learning process, was determined 128, and the number of epochs to be 50.

2.3 Results and Discussion

2.3.1 ML Metrics

The performance of the DNNs were evaluated by using ML assessment metrics. The metrics are precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC ROC), that are used to demonstrate the capability of the model to determine the particles in complex high energy physics datasets. Precision, recall, and F1 score values can be determined using the parameters found in the confusion matrix [24] obtained from the application of the DNNs model. The confusion matrix is a table that shows the actual class information in the dataset alongside the predictions made by the DNNs model. An example of confusion matrix was is represented in Figure 2.

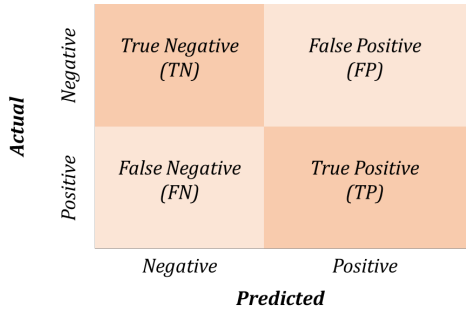


Figure 2. An example of confusion matrix.

The actual values represent the class information from the dataset, while the predicted values are the class information obtained as a result of the model's detection. In the confusion matrix used in the study, the positive class represented the J/ψ meson signal class, and the negative class was the background class. The confusion matrix was used to determine how the applied model categorized the data. From the confusion matrix true positive (TP), true negative

(TN), false positive (FP), and false negative (FN) values were extracted for calculation of the metrics. In the analysis true positive (TP) indicates the data belonging to the correctly identified J/ψ meson signal by the DNNs; true negative (TN) indicates the data belonging to the correctly determined background class by the model; false negative (FN) indicates the incorrectly labeled J/ψ meson signal by the DNNs; and false positive (FP) indicates the data belonging to the incorrectly labelled background by the model. In an ideal scenario, it is expected that the FP and FN values is 0. Recall or sensitivity is the ratio of the correctly identified J/ψ meson by the model to the total correct signal class in the dataset calculated using the parameters in the confusion matrix as demonstrated in Equation 1 [25]. Precision is the performance of the DNNs in identifying the signal computed by Equation 2 [25]. The F1 score, which allows for an evaluation of the model's performance by considering both sensitivity and precision together, is calculated using the harmonic mean as in Equation 3 [25].

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

The performance of the DNNs was measured using the "Receiver Operating Characteristic (ROC)" method, which combines several metrics. The ROC curve is based on (i) determining the two-dimensional graph of (ii) sensitivity or True Positive Rate (TPR) and (iii) the "false positive rate (FPR = FP / (FP + TN))" metrics and (iv) calculating the area under the curve (AUC ROC). The AUC ROC is a measure of how well the model can distinguish between the two classes [25]. A high AUC ROC value indicates the model's strong ability to differentiate between classes.

2.3.2 Results

The performance of the DNNS in detecting the dielectron decay channel was evaluated by ML metrics. The confusion matrix of the analysis was demonstrated in Figure 3. From the matrix it is revelaed that the model identified 615 J/ψ signal out of 616 correctly from dielectron decay channel. The derived ML metrics are listed in Table 2.

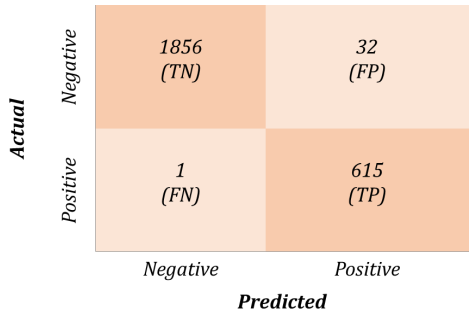


Figure 3. The confusion matrix of the DNNs.

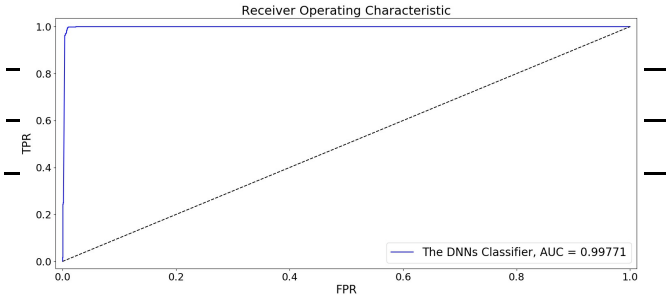


Figure 4. The ROC plot of the DNNs.

The ROC of the model is represented in Figure 4. From the table and the AUC ROC value it is concluded that the DNNs algorithm 99.771% successfully determined the meson with high precision and sensitivity. The results reveal the power of the DNNs in particle identification.

3 Conclusion

The integration of advanced ML techniques into high-energy physics enhances data interpretation. The study presented in here is an example of the application of the DNNs to determine the ground state of charmonium, J/ψ meson. The results confirm that the algorithm can be used in the meson analysis that suggests the usage of the model in high energy physics for identifying various particles from huge dataset produced at the LHC. This study highlights the synergistic relationship between machine learning and particle physics.

Acknowledgements

The present study is part of Mustafa Kaya's thesis work towards the fulfillment of his Master's degree at Firat University. This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) Project Number 123F060 and

Yildiz Technical University Project Number FBA-2024-6089.

References

[1] Radovic, A., Williams, M., Rousseau, D., Kagan, M., Bonacorsi, D., Himmel, A., ... & Whitbeck, A. (2018). Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 560(7716), 41-48.

[2] Yalcin Kuzu, S. (2022). J/ψ production with machine learning at the LHC. *Eur. Phys. J. Plus, 137*(392). https://doi.org/10.1140/epjp/s13360-022-02615-9

[3] Yalcin Kuzu, S. Machine learning model performances for the Z boson mass. *Eur. Phys. J. Plus* 138,41(2023).https://doi.org/10.1140/epjp/s13360-023-03675-1

[4] Smith, J., & Lee, K. (2021). Application of neural networks in high energy physics: An overview. *Physics Letters B*, 824, 136834.

[5] Kumar, P., & Banerjee, S. (2023). Deep learning approaches for quarkonium state analysis in high-energy collisions. *Nuclear Physics B*, 976, 115715.

[6] Zhang, Y., & Chen, W. (2020). Enhancing particle identification with deep learning in LHC data. *Computational Physics Communications*, 251, 107099.

[7] Baldi, P., Sadowski, P., & Whiteson, D. (2014). Enhanced Higgs boson to ττ search with deep learning. *Nature Communications*, 5, 4308.

[8] Abdughani, M., Wang, J., Wu, L., & Yang, J. M. (2019). Probing the CP nature of top quark Yukawa coupling with machine learning at the LHC. *Physics Letters B*, 803, 135293.

[9] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines.*In Proceedings of the 27th International Conference on Machine Learning* (ICML-10) (pp. 807-814).

[10] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.

[11] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[12] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929-1958.

[13] T. McCauley, CMS releases open data for Machine Learning, https://cms.cern/news/cms-releases-open-data-machine-learning (2014)

[14] T. McCauley, J/ψ to two electrons from 2010, https://opendata.cern.ch/record/302, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.97YF.C4AH (2014)

[15] T. McCauley, Eventswith two electrons from 2010, https://opendata.cern.ch/record/304, CERN Open Data Portal, https://doi.org/10.7483/OPENDATA.CMS.PCSW.AHVG (2014)

[16] Python Software Foundation. (n.d.). Python Documentation.Retrieved from https://docs.python.org/3/

[17] Chollet, F. (2015). Keras: The Python Deep Learning Library. Retrieved from https://keras.io

[18] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. *In 12th USENIX Symposium on Operating Systems Design and Implementation* (OSDI 16), 265-283.

[19] S. Nourbakhsh, Studio degli eventi J/ψ in due elettroni con i primi dati di CMS (2010).

[20] NA61/SHINE collaboration*, Eur. Phys. J. C* 80, 1151 (2020)

[21] CMS collaboration, Phys. Rev. C 96, 014915 (2017)

[22] Particle Data Group, M. Tanabashiet al., *Phys. Rev. D* 98, 030001 (2018).

[23] LeCun, Y. vd. 2015. "Deep learning", *Nature*, 521, 436-444, doi: https://doi.org/10.1038/nature14539.

[24] Müller A. C., Guido, S., 2016. Introduction to Machine Learning with Python, O'Reilly Media Inc., USA: Sebastopol California.

[25] Sokolova, M. vd. 2006. "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation", *Australasian joint conference on artificial intelligence*, 1015-1021, Hobart, Australia, doi: 10.1007/11941439_114.

Enhancement of Vector Boson Measurements with Boosting Learning Approach

Serpil YALÇIN KUZU [1]*, Ayben KARASU UYSAL[2]

[1] *Fırat Üniversitesi, Fizik Bölümü, Elazığ, TÜRKİYE*
[2] *Yıldız Teknik Üniversitesi, Fizik Bölümü, İstanbul, TÜRKİYE*

Abstract

Large datasets from high-energy experiments were utilized to validate theoretical predictions related to the Standard Model. In this talk, research on the implementation of machine learning methods for the analysis of the transverse mass of vector bosons through their decay into a muon and an antimuon neutrino will be presented. The focus will be on the application of advanced versions of Gradient Boosted Decision Trees (GBDT) to enhance the accuracy of mass distribution calculations. The ability of these techniques to identify complex patterns within high-dimensional data will be highlighted. Challenges encountered during the study and the strategies used to overcome them will be discussed. The results demonstrate the significant potential of GBDT variants in improving the precision of vector boson mass measurements, demonstrating the applicability and effectiveness of advanced data analysis methods in particle physics research.

Keywords: *Vector bosons, machine learning, Gradient Boosted Decision Trees.*

1 Introduction

W bosons, W^+ and W^-, are electrically charged gauge bosons that mediate the weak force, responsible for critical processes such as beta decay and neutrino interactions. These bosons are one of the fundamental particles predicted by the Standard Model of particle physics, serving as important probes in understanding fundamental forces and particles. Studying the properties and behaviors of the W boson provides valuable insights into the nature of the electroweak interaction. The characteristic features of the boson are listed in Table 1 [1]. It is characterized by a mass of approximately 80.379 GeV/c^2. Measuring the transverse mass of the W boson is crucial for precisely testing the parameters of the Standard Model and investigating potential new physics theories.

Table 1. Properties of W boson.

Mass (GeV/c^2)	80.377±0.012
Lifetime (s)	~3×10^{-25}
Charge	+/-1

The Feynman diagram depicting positively charged pion (π^+) decay mediated by W boson is represented in Figure 1 [2]. The weak force mediated by the W^+ boson converts a π^+ to a muon neutrino (ν) and antimuon (μ^+).

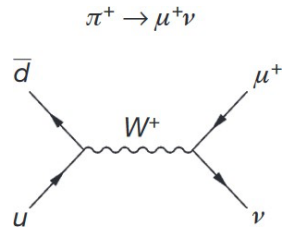


Figure 1. Feynman diagram of pion decay [2].

Artificial Intelligence (AI) and its branch, Machine Learning (ML), are computer techniques that have become popular due to their ability to perform tasks requiring human intelligence and enable systems to think and generate knowledge [3]. They have achieved notable successes in various fields such as anomaly detection [4, 5], medical prediction [6, 7], object recognition [8, 9], jet tagging [10], Higgs production [11], studying gluon ball formations [12], and detector improvements [13, 14]. The application of these algorithms is becoming increasingly important and meaningful in particle physics research that requires intensive programming to interpret data from the LHC. Therefore, the aim of the present study is to evaluate the usability of Gradient Boosted Decision Trees (GBDT) [15] and its improved versions LightGBM [16] and XGBoost [17] in vector boson analysis as an alternative candidate for particle identification.

*Contact email: serpil.yalcin@cern.ch

2 Methodology

Gradient Boosted Decision Trees (GBDT) [15] are an ensemble learning technique composed of decision trees, often showing high performance in classification and regression problems. The primary advantage of GBDT is its ability to capture both linear and non-linear relationships and its capacity to minimize the risk of overfitting. GBDT and its advanced versions have a wide range of applications and hold significant potential particularly in areas requiring complex data analysis like particle physics. The fundamental differences among these models lie in computational efficiency, scalability, and specialized feature processing capabilities.

2.1 Gradient Boosted Decision Trees (GBDT)

GBDT is based on the principle of boosting [15], where decision trees iteratively refine each other. The term 'gradient' in the model refers to the error obtained after building a model [15]. For classification, GBDT constructs a series of decision trees iteratively, each aimed at correcting the errors of the previous one, focusing on hard-to-classify samples. Initially, the model starts with a simple prediction ($F_0(x)$) that minimizes the loss function on the training data. In each subsequent iteration (t, from 1 to N), it calculates pseudo-residuals (r_{it}) to identify incorrect predictions, fits a new tree to these residuals, and determines optimal corrections (γ_{jt}) for each leaf region to minimize loss. An indicator function I ($x \in R_{jt}$) is used to ensure that these corrections are applied correctly to the intended samples. This function acts like a gatekeeper, allowing the γ_{jt} correction to be added to the model's prediction only for samples within the specific leaf region R_{jt}, thereby providing precise and targeted adjustments. These corrections are scaled by a learning rate (η) to gradually refine the model's accuracy. The process is repeated for N trees, with the weighted corrections of each tree cumulatively enhancing the model. The final model ($F_N(x)$) uses the accumulated corrections from all trees to predict the class label for new samples, balancing learning rate and model complexity through the number of trees and learning rate. Compared to other decision tree-based models, GBDT constructs trees in a sequential manner, which results in long training times. To overcome long training times, algorithms such as LightGBM and XGBoost have been developed.

2.2 LightGBM

LightGBM [16] is a gradient boosting framework that reduces the number of samples in the data based on gradients while maintaining the accuracy rate of decision trees, utilizing algorithms such as Gradient-based One-Side Sampling (GOSS) [16] and Exclusive Feature Bundling (EFB) [16]. Known also as light gradient boosting machines, LightGBM is highly preferred due to its high processing speed, capability to handle large datasets, low memory usage, high prediction accuracy, parallel learning, and GPU support. Within the algorithm, learning in decision trees is performed on a leaf-by-leaf basis [16], which ensures the model has a lower error rate. Additionally, optimizing parameters within the algorithm can make the model resistant to overfitting. LightGBM is another variant of GBDT that is more effectively able to handle high dimensionality and large scales in datasets. The model enables the model to quickly learn complex patterns in large datasets, making it a powerful tool for analyzing the transverse mass of the W boson.

2.3 XGBoost

XGBoost [17] is an enhanced version of GBDT optimized to achieve high predictive power and prevent overfitting through various adjustments. Compared to GBDT, it controls the complexity of decision trees to achieve better performance. This approach involves defining an objective function to optimize and measure the performance of the model according to the given training dataset. XGBoost is more successful than most gradient boosting-based models due to its advanced hardware and software design. It constructs sequential decision trees much faster by transitioning between inner loops that compute tree features and outer loops that create nodes and leaves. This model offers innovative techniques such as building trees in parallel, regularization, and tree pruning strategies. These features assist XGBoost in providing high accuracy and speed in analyzing the transverse mass of the W boson.

3 Analysis

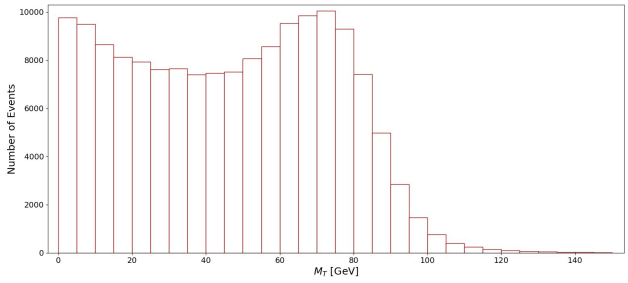
3.1 Properties of Dataset

Proton–proton collision collected by the CMS experiment at center of mass energy 7 TeV in 2011 were used to reconstruct *W* from its muon (μ) - anti muon neutrino ($\bar{\nu}$) and anti muon (μ^+) - muon neutrino (ν) decays [18-19]. The dataset contains information about the decay products of the vector

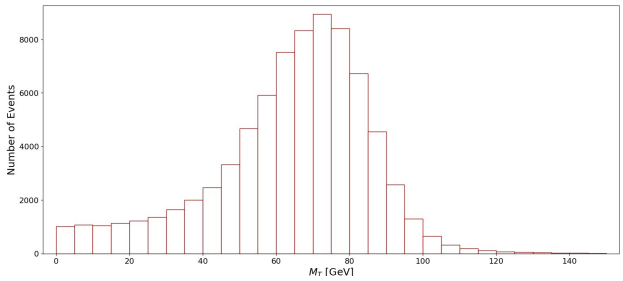
boson including 155599 events where one muon candidate was observed and missing energy implies to the presence of a neutrino. The data was derived from the primary dataset SingleMu 2011 [18-19] and the data having missing information was removed from the dataset.

The features used the analysis are p_T, η, the phi angle (ϕ), charge (Q), the chi-squared per degree-of-freedom (chiSq), the impact parameter in the transverse plane with respect to the vertex (d_{xy}), the detector isolation (iso) of the muon and the missing transverse momentum (MET) and its relative the phi angle (phiMET). The transverse mass (M_T) of the vector boson was reconstructed by Equation 1 where $\Delta\phi$ represents the difference between ϕ of the pairs and used the analysis as an additional feature.

For the signal class definition an event was selected if there was one (and only one) global muon in the event that satisfies transverse momentum (pT) greater than 25.0 GeV and pseudorapidity (η) within 2.1. Furthermore, the event was considered as background class if there was another global muon in the event with pT > 10 GeV in the η region 2.4 [18-19]. Also for the signal class definition iso was selected less than 0.1, dxy was less than 0.2cm, and chiSq was less than 10. The rest of the data was considered as background. In the end of the class definitions 76880 events were signal and 78719 were background which means approximately 50% of the data is W bosons. MT of all data and the signal class are represented in Figure 2a and 2b, respectively.



(a)



(b)

Figure 2. M_T of all data (a) and the signal class (b).

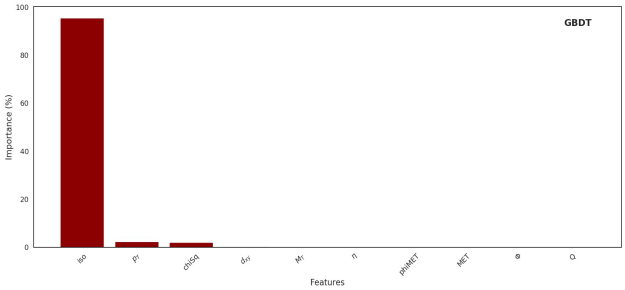
3.2 Implementation of Boosting Algorithms

Due to the size of the signal class 70% of the data was selected for training and rest for the test. In the analysis for the comparison of the model performances the number of boosting tree was

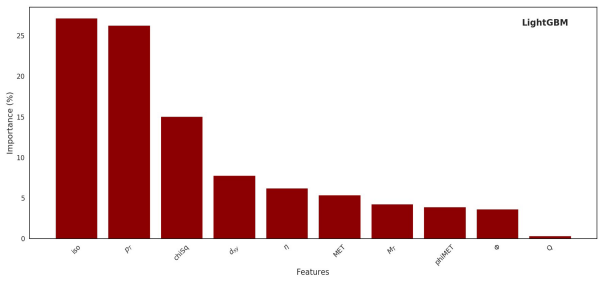
$$M_! = \#\overline{2p_!ET(1 - cos\Delta\phi)} \quad (1)$$

selected as 100 and the learning rate of each tree was preferred 0.1. The listed parameters in previous section was used as features in the algorithms.

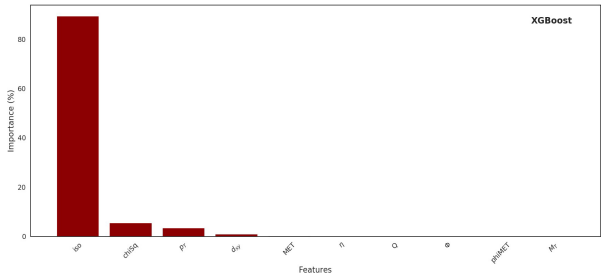
The importance of each feature in the algorithms were demonstrated in Figure 3a, 3b, and 3c for GBDT, LightGBM, and XGBoost, respectively. From the figure it is concluded that the algorithms used the features in different way even if they based on booting. GBDT and XGBoost considered iso feature more to discriminate the classes while LightGBM took into account all features.



(a)



(b)



(c)

Figure 3. Importance of each feature in GBDT (a), LightGBM (b), and XGBoost (c) algorithms.

4 Results and Discussion

4.1 Overview of Classification Metrics

Due to the data structure the performance of the boosting algorithms were compared by using machine learning metrics for balanced class sizes that are precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC ROC) [20]. The definitions of precision, recall, and F1 score are computed from confusion matrix [21] by Equations 2, 3 and, 4, respectively [22]. In the equations TP, true positive, represents correctly determined bosons; FP, false positive, represents incorrectly estimated bosons; TN, true negative, presents correctly identified background; FN, false negative, presents incorrectly defined background by the boosting algorithms.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

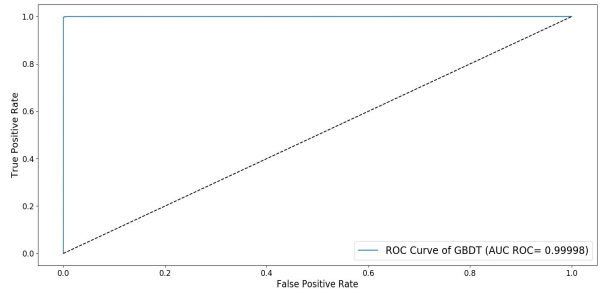
$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

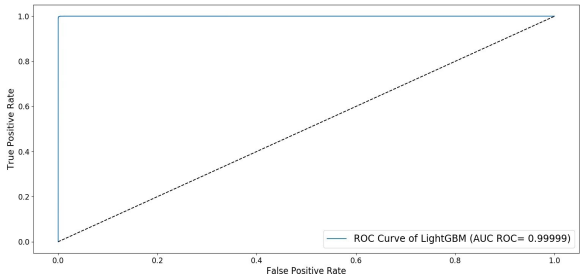
The achievement of booting algorithms was also compared by using "Receiver Operating Characteristic (ROC)" method, which is computed by (i) the two-dimensional graph of (ii) recall or True Positive Rate (TPR) and (iii) the "false positive rate (FPR = FP / (FP + TN))" metrics. The area under the curve (AUC ROC) can be used as a tool to understand the discriminative power of the boosting models between W boson signal and the background since a high AUC ROC value implies the success of the model. estimation.

4.2 Evaluation of the Algorithms

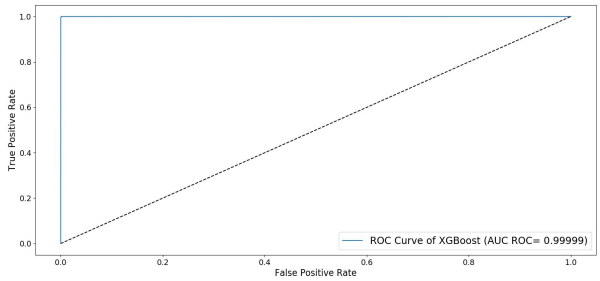
The model performances were listed in Table 2 and AUC ROC curves were demonstrate in Figure 4. From the table it is concluded that performance of the LightGBM and XGBoost models are very similar to each other and better than GBDT. This proves that the improved versions of the boosting algorithm perform better than GBDT.



(a)



(b)



(c)

Figure 4. The ROC of the GBDT (a), LightGBM (b), and XGBoost (c).

Table 2. Model performances.

Model	Precision	Recall	F1 Score
GBDT	0.99817	0.99814	0.99816
LightGBM	0.99835	0.99835	0.99835
XGBoost	0.99913	0.99912	0.99912

5 Conclusion

The application of advanced machine learning techniques, particularly Gradient Boosted Decision Trees (GBDT) and its improved variants LightGBM and XGBoost, has proven to be highly effective in enhancing the precision of vector boson mass measurements in high-energy physics. By leveraging these methods, the study successfully demonstrated how these algorithms can overcome the challenges of analyzing complex, high-dimensional datasets, which are typical in particle physics experiments such as those involving W bosons. The results from this study show that both LightGBM and XGBoost significantly outperform traditional GBDT in terms of key machine learning metrics, including precision, recall, and F1 score, with XGBoost achieving the highest performance across all metrics. These findings highlight the importance of using advanced machine learning models that can efficiently handle large-scale data while minimizing the risk of overfitting, a common challenge in such analyses.

This study underscores the growing role of artificial intelligence and machine learning in scientific research. The successful implementation of these methods in vector boson analysis exemplifies their potential in other areas of high-energy physics, where accurate and efficient data analysis is critical. Future work could explore the application of these techniques to other types of particle interactions and experiments, further expanding their utility in the field. In conclusion, the integration of GBDT, LightGBM, and XGBoost into particle physics research represents a significant advancement in ability to analyze and interpret complex experimental data. As these machine learning techniques continue to evolve, they are likely to become increasingly central to the pursuit of new discoveries in the high-energy physics.

Acknowledgements

This work was supported by Firat University Project Number FF.24.15 and Yildiz Technical University Project Number FBA-2024-6089.

References

[1] Particle Data Group, Prog. Theor. Exp. Phys., 2022, 083C01.

[2] CMS Collaboration., LHCb Collaboration. Observation of the rare $B_s^0 \rightarrow \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data. *Nature* **522**, 68–72 (2015). https://doi.org/10.1038/nature14474.

[3] Donato, C. (2018). Muon identification and performance in the ATLAS experiment. Journal of Physics: Conference Series, 326, 068. https://doi.org/10.22323/1.326.0068.

[4] Heimel, T. et al. (2019). "QCD or What?", SciPost Physics, 6, 30, doi: 10.21468/SciPost-Phys.6.3.030.

[5] Metodiev, E. M. et al. (2017). "Classification without labels: Learning from mixed samples in high energy physics", Journal of High Energy Physics, 174, https://doi.org/10.1007/JHEP10(2017)174.

[6] Vuttipittayamongkol, P., Elyan, E. (2020). "Overlap-based undersampling method for classification of imbalanced medical datasets", in IFIP International Conference on Artificial Intelligence Applications and Innovations, 358-369, https://doi.org/10.1007/978-3-030-49186-4_30.

[7] Vuttipittayamongkol, P., Elyan, E. (2020). "Improved overlap-based undersampling for imbalanced dataset classification with application to epilepsy and parkinson's disease", International Journal of Neural Systems, 30:2050043, doi: 10.1142/S0129065720500434.

[8] Zhang, X. et al. (2016). "Transfer boosting with synthetic instances for class imbalanced object recognition", IEEE Transactions on Cybernetics, 48, 357-370, doi: 10.1109/TCYB.2016.2636370.

[9] Elyan, E. et al. (2020). "Deep learning for symbols detection and classification in engineering drawings", Neural Networks, 129, 91-102, https://doi.org/10.1016/j.neunet.2020.05.025.

[10] Gong, S. et al. (2022). "An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging", Journal of High Energy Physics, 30, https://doi.org/10.1007/JHEP07(2022)030.

[11] Bourilkov, D. et al. (2019). "Machine Learning Techniques in the CMS Search for Higgs Decays to Dimuons", EPJ Web of Conferences, 214, 1-11, doi:10.1051/epjconf/201921406002.

[12] Isolabella, T. (2019). A machine learning approach to particle physics data analysis: the process $J/\psi \rightarrow \gamma p\bar{p}$, Master Thesis, University of Groningen, Groningen, Netherlands.

[13] Qasim, S. R. et al. (2019). "Learning representations of irregular particle-detector geometry with distance-weighted graph networks", The European Physical Journal C, 79, 608, https://doi.org/10.1140/epjc/s10052-019-7113-9.

[14] Erdmann, M. et al. (2018). "Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network", Computing and Software for Big Science, 3, 4, [https://doi.org/10.1007/s41781-018-0019-7]

[15] Friedman, J. H. 2001. "Greedy function approximation: a gradient boosting machine", Annals of statistics, 1189-1232, doi: 10.1214/aos/1013203451.

[16] Ke, G. vd. 2017. "Lightgbm: A highly efficient gradient boosting decision tree", Advances in neural information processing systems, 30, 3149-3157.
[17] Chen, T., Guestrin, C. 2016. "XGBoost: A Scalable Tree Boosting System", KDD 16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Long Beach, Amerika, doi: https://doi.org/10.1145/2939672.2939785.
[18] CMS. "Public CMS Luminosity Information". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2010_proton_proton_7_TeV_collisi
[19] CERN Open Data. "*W* to W to muon and neutrino 2011". https://opendata.cern.ch/record/5205
[20] Bradley, A. P., 1997. "The use of the area under the ROC curve in the evaluation of machine learning algorithms", Pattern Recognition, 30, 1145-1159, doi:https://doi.org/10.1016/S0031-3203(96)00142-2.
[21] Müller A. C., Guido, S., 2016. Introduction to Machine Learning with Python, O'Reilly Media Inc., USA: Sebastopol California.
[22] Tharwat, A. 2020. "Classification assessment methods", Applied Computing and Informatics 17, doi: 10.1016/j.aci.2018.08.003.

Duygu Analiz Çalışması: Veri Setinin Etiketlenme Biçimine Göre Makine Öğrenmesi ve Derin Öğrenme Yöntemlerinin Karşılaştırılması

Merve Hazan İŞCAN[1]*, Nagehan İLHAN[2]

[1]*Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul, TÜRKİYE*
[2]*Harran Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Şanlıurfa, TÜRKİYE*

Özet

Günümüzde sosyal medyanın yaygın kullanımı sayesinde oldukça büyük veri kaynakları oluşmaktadır. Bahsedilen kaynaklardan metin biçiminde olanlar doğal dil işleme, duygu analizi ve metin madenciliği gibi çalışmalarda yaygın bir biçimde kullanılmaktadır. Yapılan bu çalışmada ise veri seti olarak Twitter platformundan alınan "yapay zeka" etiketine sahip 14.657 tweet kullanılmıştır. Çalışmada veri seti iki farklı şekilde etiketlenmiş olup, iki veri setine de makine öğrenmesi tekniği olarak Lojistik Regresyon(LR), Naive Bayes(NB), Destek Vektör Makinesi(DVM) ve Rastgele Orman(RO) Algoritmaları, derin öğrenme tekniklerinden ise Long Short Term Memory (LSTM), Convolutional Neural Network(CNN) ve Recurrent Neural Network (RNN) Algoritmaları uygulanmıştır. Yalnızca insan doğrulaması ile etiketlenmiş olan veri setinde alınan doğruluk sonuçları; LR için %65,67, NB için %62, DVM için %66,57, RO için %64,66, LSTM için %64,84 CNN için %63,30 ve RNN için %63,27 olarak bulunmuştur. SentiTurkNet sözlüğünden yararlanılarak etiketlenmiş olan veri setinde alınan doğruluk sonuçları ise; LR için %81,69, NB için %78, DVM için %82,78, RO için %82,86, LSTM için %83,53, CNN için %83,88 ve RNN %82,75 olarak bulunmuştur.

Anahtar Kelimeler: *Veri Bilimi, Duygu Analizi, Makine Öğrenmesi, Derin Öğrenme*

Sentiment Analysis Study: Comparison of Machine Learning and Deep Learning Methods According to the Labeling of the Data Set

Abstract

Nowadays, the widespread use of social media generates huge data sources. Among the mentioned resources, those in text form are widely used in studies such as natural language processing, sentiment analysis and text mining. In this study, 14,657 tweets with the tag "artificial intelligence" taken from the Twitter platform were used as the data set. In the study, the dataset was labeled in two different ways and Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF) Algorithms as machine learning techniques, and Long Short Term Memory (LSTM) and Convolutional Neural Network(CNN) and Recurrent Neural Network (RNN) Algorithms as deep learning techniques were applied to both datasets. For the dataset labeled with only human verification, the accuracy results were 65,67% for LR, 62% for NB, 66,57% for SVM, 64,66% for RF, 64,84% for LSTM, 63,30% for CNN and 63,27% for RNN. The accuracy results obtained for the dataset labeled using the SentiTurkNet dictionary were 81,69% for LR, 78% for NB, 82,78% for SVM, 82,86% for RF, 83,53% for LSTM, 83,88% for CNN and 82,75% for RNN.

Keywords: *Data Science, Sentiment Analysis, Machine Learning, Deep Learning*

*Contact email: merve.hazan@marmara.edu.tr

1 Giriş

Teknolojinin gelişmesi ile birlikte internet erişimi artmış olup, internetin yaygınlaşmasıyla birlikte sosyal medya kullanımı da oldukça artmıştır. Türkiye 2023 ve 2024 yılı internet ve sosyal medya kullanım istatistiklerine bakıldığında; internet kullanıcı sayısının 2023 yılının başında 71,38 milyon, 2024 yılının başında 74,41 milyon ve 2024 yılı başı itibari ile sosyal medya kullanıcı sayısının ise 57,50 milyon olduğu görülmektedir[1]. İnsanlar sosyal medyada hayatlarına dair birçok şey paylaşmaktadırlar. Bu paylaşımlar sayesinde çeşitli türlerde oldukça büyük veri kaynakları oluşmaktadır. Bu veriler kullanılarak eğitim, ekonomi, turizm vb. alanlar için duygu analizi, fikir madenciliği gibi çalışmalar yapılmaktadır.

Metnin çeşitli yöntemler ile analiz edilmesi sonucunda metni yazan bireyin duygularını anlamak için kullanılan doğal dil işleme yöntemine duygu analizi denilmektedir ve bu tür çalışmalarda metin pozitif, negatif ve nötr olarak sınıflandırılır[2].

Yapılan çalışmada Twitter platformundan toplanan yapay zeka konulu 14.657 bin veriye iki farklı şekilde etiketleme işlemi yapılmış olup, duygu analiz işlemi için makine öğrenmesi tekniklerinden Lojistik Regresyon(LR), Naive Bayes(NB), Destek Vektör Makinesi(DVM) ve Rastgele Orman(RO) Algoritmaları, derin öğrenme tekniklerinden Uzun Kısa Süreli Bellek(LSTM) Evrişimsel Sinir Ağı (CNN) ve Yineleyen Sinir Ağı(RNN) Algoritmaları uygulanmıştır.

Bu çalışmanın ikinci bölümünde duygu analizi ile ilgili literatürde yer alan bazı çalışmalardan bahsedilmiştir, üçüncü bölümde çalışmada kullanılan materyal ve yöntemler açıklanmıştır ve son bölümde çalışmadan elde edilen sonuçlara yer verilmiştir.

2 İlgili Çalışmalar

Bu kısımda duygu analizi ile ilgili yapılmış literatürde bulunan bazı önemli çalışmalardan bahsedilmektedir.

Virra vd. yaptıkları çalışmada veri seti olarak cinsel şiddetin ortadan kaldırılmasına ilişkin kanun taslağını konu almış olan tweetleri kullanmışlardır. Sınıflandırma yöntemi olarak Naive Bayes Algoritması, Karar Ağacı Algoritması ve Random Forest Algoritması kullanılmış olup, Naive Bayes Algoritması %83, Karar Ağacı Algoritması %75, Random Forest Algoritması ise %75 ornında başarım elde etmiştir[3].

Masruroh vd. yaptıkları çalışmada veri seti olarak, belirlenen kelimeleri içeren, Twitter'dan çekilmiş 1000 tweetten oluşmaktadır. Kelime ağırlıklandırma işlemi TF-IDF yötemi ile gerçekleştirilmiştir. Çalışmada üç farklı eğitim-test veri oranı kullanılmıştır. Sınıflandırma işlemi için her üç durumda da Naive Bayes ve Destek Vektör Makinesi Algoritmaları kullanılmıştır. Birinci durumda %60 eğitim ve %40 test olarak ayrılan veriler üzerinde Naive Bayes Algoritması %97 oranıyla en yüksek başarımı elde etmiştir. İkinci durumda %70 eğitim ve %30 test olarak ayrılan veriler üzerinde Naive Bayes Algoritması %98 oranıyla en yüksek başarımı elde etmiştir. Üçüncü durumda ise %80 eğitim ve %20 test olarak ayrılan veriler üzerinde Destek Vektör Makinesi %98 oranıyla en yüksek başarımı elde etmiştir[4].

Kumaş tarafından yapılan bu çalışmada veri seti olarak farklı konularda 16000 adet olumlu ve 16000 adet olumsuz etiketli tweet kullanılmıştır. Kelime ağırlıklandırma işlemi TF-IDF yötemi ile gerçekleştirilmiştir. Sınıflandırma işlemi için K En Yakın Komşu, Naive Bayes, Destek Vektör Makinesi, Karar Ağacı ve Lojistik Regresyon Algoritmaları kullanılmıştır. En iyi sonucu veren algoritma %73 oranı ile Destek Vektör Makinesi olmuştur[5].

Alzyout vd. çalışmalarında veri seti olarak kadına yönelik şiddet ile ilgili 5000 tweet toplanıp, kullanılmıştır. Veri etiketlendirmesi kadın ve erkeklerden oluşan 5 kişilik bir grup tarafından yapılmıştır. Sınıflandırma işlemi için K En Yakın Komşu, Naive Bayes, Karar Ağaçları ve Destek Vektör Makinesi kullanılmıştır. Çalışmasının sonucunda en iyi sonucu Destek Vektör Makinesi, en kötü sonucu ise Naive Bayes Algoritması vermiştir[6].

3 Materyal ve Yöntem

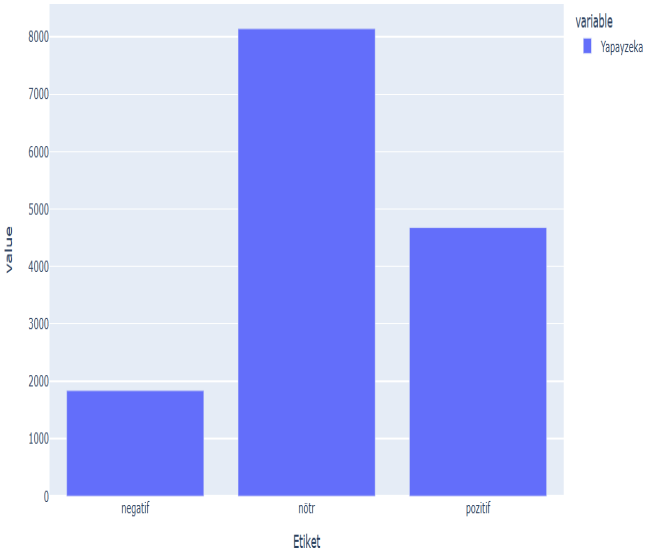
Çalışmanın veri setini oluşturmak için Twitter platformundan yararlanılmıştır. Python'da yer alan Selenium kütüphanesi kullanılarak Ocak 2020 – Aralık 2023 tarihleri arasında atılmış olan #yapayzeka etiketine sahip 14.657 tweet toplanmıştır.

Şekil 1'de veri setinde en çok tekrar eden kelimelerden oluşan bir kelime bulutu görülmektedir.



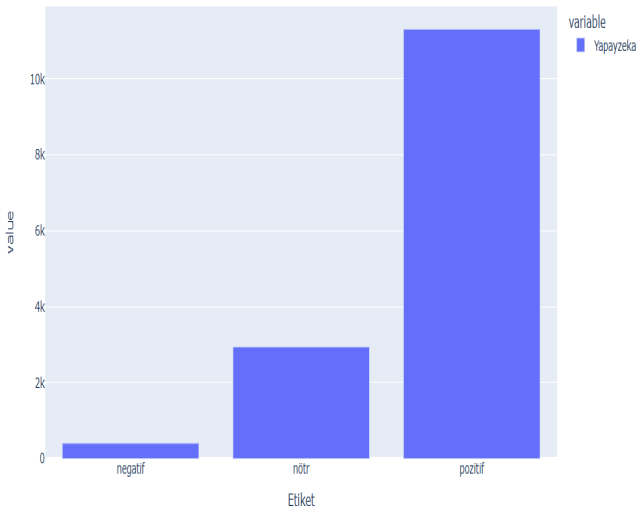
Şekil 1. Kelime Bulutu

Yapılan çalışmada veri seti ayrı ayrı iki farklı şekilde etiketlendirilmiştir. Birinci durumda veriler herhangi bir kaynaktan yararlanılmadan kadın ve erkeklerden oluşan 4 kişilik bir ekip tarafından pozitif, negatif ve nötr olarak etiketlendi. Şekil 2'de birinci etiketlemeye ait etiket dağılımları görülmektedir.



Şekil 2. Etiket Dağılımları 1

İkinci durumda ise veriler etiketlenirken Türkçe duygu sözlüğü olan SentiTurkNet'te bulunan kelime polarite skorlarından yararlanılarak pozitif, negatif ve nötr olarak etiketlenmiştir. Şekil 3'te ikinci etiketlemeye ait etiket dağılımları görülmektedir.



Şekil 3. Etiket Dağılımları 2

Her iki durum için de veriler ön işlem adımlarından geçirilerek çalışmaya uygun hale getirilmiştir. Daha sonra kelime ağırlıklandırma ve birliktelik analizi için Kelime frekansı–ters doküman frekansı (Tf-IDF) ve N-gram yöntemlerinden yararlanılmıştır.

Duygu analiz çalışması için makine öğrenme tekniklerinden LR, NB, DVM ve RO Algoritmaları, derin öğrenme tekniklerinden ise LSTM, CNN ve RNN Algoritmaları kullanılmıştır.

Lojistik Regresyon(LR); İstatistik tabanlı bir algoritmadır. Eğitim setinde bulunan veriler ile sınıflandırma yapılıp yeni veriler olasılık değerinin en büyük olduğu sınıfa yerleştirilir[7].

Naive Bayes(NB); İstatistik tabanlı bir algoritma olup, Bayes teoremine dayanmaktadır. Sınıf koşul bağımsızlığı denilen belli bir sınıfta bulunan öznitelik değer etkisinin, diğer değerlere bağlı olmadığı varsayılır. Kullanım kolaylığı sebebiyle sıklıkla tercih edilen bir algoritmadır[8].

Destek Vektör Makineleri(DVM); Aykırı değerlere sahip verilere karşı dayanıklı ve genelleştirme konusunda yüksek yetenekli bir algoritmadır. Bu algoritmanın hedefi doğrusal olmayan bir eşleme metodu kullanarak veriyi daha üst bir boyuta çevirip, bu veriyi düzgün şekilde bölebilecek bir üst düzlem oluşturmaktır. Bu düzlem iki ayrı sınıfta bulunan verilerin ayrılmasını sağlayan bir karar sınırıdır[8,9].

Rastgele Orman Algoritması(RO); Bu algoritmanın ağaç tipi olan sınıflandırıcıların bir araya gelmesi ile oluştuğu söylenilebilir. Torbalama yöntemine rastgelelik özelliği eklenmiş daha iyi bir versiyonu olduğu söylenilebilir[10].

Long Short Term Memory(LSTM): Bağımlılıkları öğrenebilen, katmanlı bir yapıya sahip olan bir derin öğrenme algoritmasıdır. Bu algoritmanın yapısında girdi, çıktı ve unutma katmanı bulunmaktadır[11,12].

Convolutional Neural Network(CNN): Çok katmanlı bir yapıya sahiptir ve katman giriş çıkışları sırası ile birbirine bağlıdır[13,14].

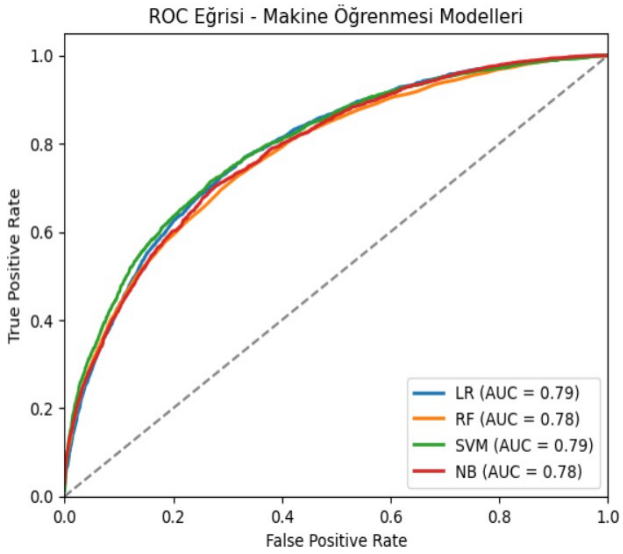
Recurrent Neural Network (RNN): Belleğe bağlı olarak bilgilerin sırası ile yönü belli olan döngüler tarafından yürütülür ve bilgi dağılımı çift yönlüdür[13,14].

Bu algoritmalar birinci durumda, yani 4 kişilik ekip tarafından etiketlenmiş olan veri setine uygulandığında alınan sonuçlar Tablo 1'de görülmektedir.

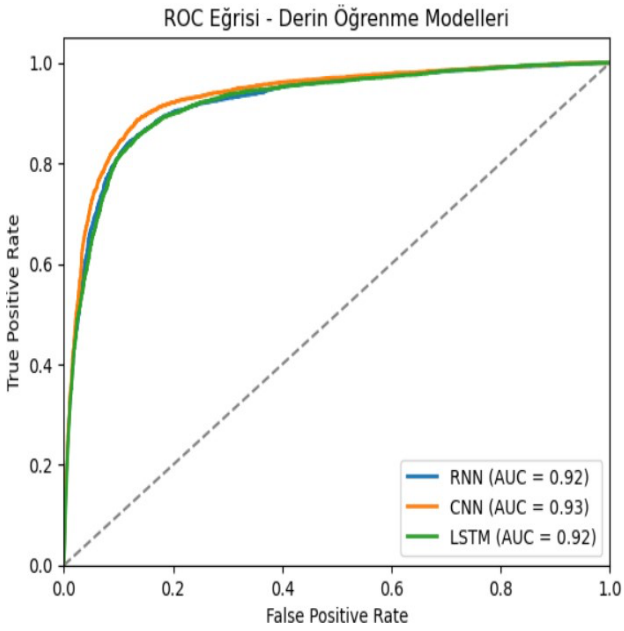
Tablo 1. Birinci Durum İçin Elde Edilen Sonuçlar

Yöntem	Accuracy	Precision	Recall
LR	%65,67	%67	%51
NB	%62	%72	%62
DVM	%66,57	%67	%54
RO	%64,66	%69	%65
LSTM	%64,84	%65,33	%62,93
CNN	%63,30	%63,86	%62,84
RNN	%63,27	%64,16	%63,75

Tablo 1'de görüldüğü üzere birinci durumda en iyi sonucu DVM Algoritması vermiştir ve en kötü sonucu NB Algoritması vermiştir. Sonuçların ROC eğri grafikleri Şekil 4 ve Şekil 5'te verilmiştir.



Şekil 4. Makine Öğrenmesi ROC Eğrisi 1



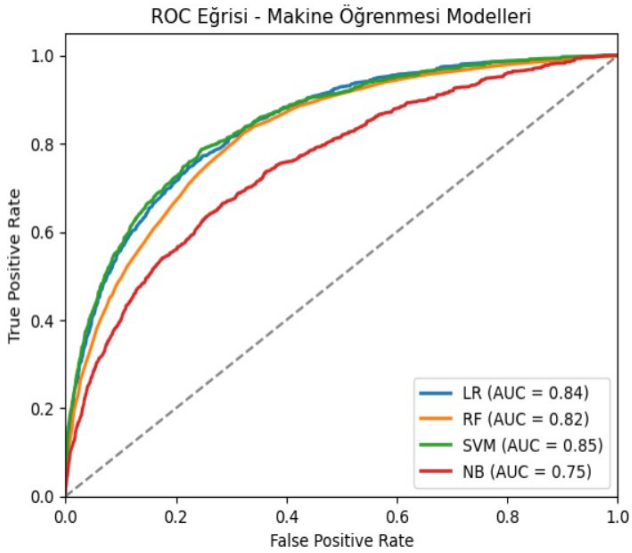
Şekil 5. Derin Öğrenme ROC Eğrisi 1

Bu algoritmalar ikinci durumda, yani SentiTurkNet duygu sözlüğünden yararlanılarak etiketlenmiş olan veri setine uygulandığında alınan sonuçlar Tablo 2'de görülmektedir.

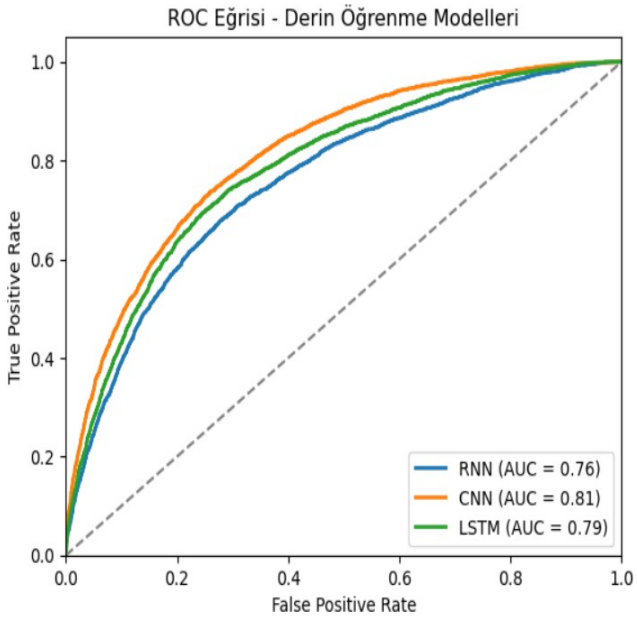
Tablo 2. İkinci Durum İçin Elde Edilen Sonuçlar

Yöntem	Accuracy	Precision	Recall
LR	%81,69	%79	%82
NB	%78	%80	%78
DVM	%82,78	%82	%83
RO	%82,86	%82	%83
LSTM	%83,53	%83,66	%83,46
CNN	%83,88	%84,27	%83,77
RNN	%82,75	%83,57	%82,89

Tablo 2'de görüldüğü üzere ikinci durumda en iyi sonucu DVM, RO ve LSTM Algoritmaları vermiştir ve en kötü sonucu NB Algoritması vermiştir. Sonuçların ROC eğri grafikleri Şekil 6 ve Şekil 7'de verilmiştir.



Şekil 6. Makine Öğrenmesi ROC Eğrisi 2



Şekil 7. Derin Öğrenme ROC Eğrisi 2

4 Sonuçlar

Yapılan çalışmada Twitter platformundan toplanan yapay zeka konulu 14.657 bin veriye iki farklı şekilde etiketleme işlemi yapılmış olup, duygu analiz işlemi için makine öğrenmesi tekniklerinden Lojistik Regresyon(LR), Naive Bayes(NB), Destek Vektör Makinesi(DVM) ve Rastgele Orman(RO) Algoritmaları, derin öğrenme tekniklerinden Uzun Kısa Süreli Bellek(LSTM) Evrişimsel Sinir Ağı (CNN) ve Yineleyen Sinir Ağı(RNN) Algoritmaları uygulanmıştır.

Yalnızca insan doğrulaması ile etiketlenmiş olan veri setinde alınan doğruluk sonuçları; LR için %65,67, NB için %62, DVM için %66,57, RO için %64,66, LSTM için %64,84 CNN için %63,30 ve RNN için %63,27 olarak bulunmuştur.

SentiTurkNet sözlüğünden yararlanılarak etiketlenmiş olan veri setinde alınan doğruluk sonuçları ise; LR için %81,69, NB için %78, DVM için %82,78, RO için %82,86, LSTM için %83,53, CNN için %83,88 ve RNN %82,75 olarak bulunmuştur.

Kaynaklar

[1] DATAREPORTAL,2023-2024. Dünyada Dijital Ortam Kullanımı, erişim: Ocak 2023.

[2] Kaynar O, Yıldız M, Görmez Y, Albayrak A. "Makine Öğrenmesi Yöntemleri ile Duygu Analizi". International Artificial Intelligence and Data Processing Symposium (IDAP'16), p. 17-18, 2016.

[3] Virra K, Andreswari R, Hasibuan MA. "Sentiment Analysis of Social Media Users Using Naïve Bayes, Decision Tree, Random Forest Algorithm: A Case Study of Draft Law on the Elimination of Sexual Violence (RUU PKS)". 2019 International Conference on Sustainable Engineering and Creative Computing (ICSECC), p.239-244, Bandung, Indonesia, 2019.

[4] Masruroh S.U, Utami D.Z.A, Khairani D, Azhari M, Helmi M.I, Putri R.A. "Sentiment Analysis on Twitter towards the Ratification of a Bill on the Elimination of Sexual Violence in Indonesia using Machine Learning". 2022 10th International Conference on Cyber and IT Service Management (CITSM), 1-5. Yogyakarta, Indonesia, 2022.

[5] Kumaş E. "Türkçe Twitter Verilerinden Duygu Analizi Yapılırken Sınıflandırıcıların Karşılaştırılması". ESTUDAM Bilişim Dergisi, 2(2): 1-5, 2021.

[6] Alzyout M, Bashabsheh EA, Najadat H, Alaiad A. "Sentiment Analysis of Arabic Tweets about Violence Against Women using Machine Learning". 2021 12th International Conference on Information and Communication Systems (ICICS), IEEE, 171-176, 2021.

[7] Shatkay H, Craven M. "Mining the biomedical literature". MIT PRES, 2012.

[8] Onan A. "Twitter Mesajları Üzerinde Makine Öğrenmesi Yöntemlerine Dayalı Duygu Analizi". Yönetim Bilişim Sistemleri Dergisi, 3(2): 1-14, 2017.

[9] Vapnik V. "The nature of statistical learning theory". Springer, 1995.

[10] Breiman L. Random Forests, Machine Learning. Kluwer Academic Publishers, 45(1): 5-32, 2001.

[11] Güdelek M. U., "Zaman Serisi Analiz ve Tahmini: Derin Öğrenme Yaklaşımı". Yüksek Lisans Tezi. TOBB Ekonomi ve Teknoloji Üniversitesi, Ankara, Türkiye, 2019.

[12] Kızrak M. A, Bolat B., "Derin öğrenme ile kalabalık analizi üzerine detaylı bir araştırma". Bilişim Teknolojileri Dergisi, 11(3), 263-286, 2018.

[13] Rhannoui A, Mikram M, Yousfi S, Barzali S, "A CNN-BiLSTM Model for Document-Level Sentiment

Analysis". Mach. Learn. Knowl. Extr., 1(3):832-847, 2019.

[14] Dang N. C, Moreno-Garcia M. N, Prieta F. D. L, "Sentiment Analysis Based on Deep Learning: A Comparative Study". Electronics, 9(3):483, 2020.

Estimation of Parameters Affecting Water Quality Using Data Mining Algorithms

Elif KAVAK[1]*, Doç. Dr. Mümine KAYA KELEŞ[1]

[1]*Adana Alparslan Türkeş Science and Technology University, Faculty of Computer and Informatics, Computer Engineering, Adana, TURKEY*

Abstract

17 Sustainable Development Goals (SDGs) have been published by the United Nations (UN) on this subject. One of these goals, "Clean Water and Sanitation", includes the basic factor for the survival of living creatures on Earth. Clean water is of critical importance for human health, hygiene and ecosystems. However, finding drinkable water is becoming increasingly difficult due to factors such as population growth, pollution and climate change. These challenges highlight the need for sustainable management of water resources, wastewater treatment and development of sanitation infrastructure. In this context, the need for methods to monitor water quality is increasing. This research aims to contribute to the "Clean Water and Sanitation" target, which is included in the 17 SDGs published by the UN. In line with this goal, it is aimed to facilitate the monitoring of water quality and contribute to the sustainable management of water resources. It is expected that the method to be developed will be of greater importance, especially in regions at risk of water scarcity or with limited access to water resources. In the study, water quality index will be monitored by making future predictions with machine learning methods using Python language. The innovative and high-accuracy method, which will be developed with machine learning methods, aims to provide a faster, more effective and sustainable solution compared to the currently used methods. The water quality index will be determined and a data set will be created by obtaining data affecting the quality of water such as pH, turbidity, conductivity and iron from Adana and Mersin provinces. A model that can predict the water quality index will be developed and possible changes in water quality will be monitored. The applicability of the method developed with this study for different water resources will be increased.

Keywords: *Sustainability, Water Quality, Machine Learning*

1 Introduction

Sustainability refers to keeping something at a certain rate or level over time. Thus, it is possible to increase the quality of life, protect the ecosystem, and protect natural resources for future generations. The resources necessary to sustain life are assumed to be limited by the United Nations (UN), and it is thought that they should be used carefully in order to provide sufficient resources for future generations without decreasing the current quality of life. The UN has published 17 Sustainable Development Goals (SDGs) to achieve a more sustainable world by 2030. These targets cover areas such as hunger, clean water, the economy, and energy [1].

The most important among these goals is clean water and sanitation, as well as being a basic human rights, and every living thing on earth needs water. However, billions of people around the world have been deprived of this right, and access to clean water is getting harder day by day due to factors such as urbanization and waste disposal. In addition, it is known that diseases in developing countries are transmitted by water and cause the deaths of approximately 2.2 million people annually [2].

In order to achieve a sustainable future, it is necessary to pay attention to the reduction of pollution in water. In order to evaluate whether the drinking water is polluted or not, the Water Quality Index (WQI) should be checked. However, water quality management is difficult and requires the collection and analysis of water quality data sets [3]. Evaluation basically requires defining the parameters and defining the parameters that affect

*Contact email: ekavak@atu.edu.tr

the potability level of the water. An analysis is made by integrating the data, and a score showing the quality of the water is obtained. Thus, the drinkability of the water can be evaluated without examining the parameters [4].

In recent years, studies have been carried out to determine and monitor the quality of water resources. This study aims to develop an effective method to monitor WQI and the parameters that affect it. In the study, the water quality index was determined and the data set was created using two separate city data sets that include the parameters that affect WQI. For the estimation of WQI values, a model was created with 7 different machine learning algorithms to monitor them and their performances were compared.

2 Literature Review

In the study of Kouadri, Elbeltagi, & Islam et al. (2021) [5], it was emphasized that it is important to evaluate the quality of drinking water source and it was aimed to develop the WQI estimation model. Two scenarios were created: In the first scenario, it is aimed to use all water quality parameters as inputs and to reduce the time in WQI calculation, while in the second scenario, it is aimed to show the water quality variation in critical situations where water quality analyzes are missing. Eight different methods have been applied for WQI estimation: Multiple Linear Regression (MLR), Artificial Neural Network (ANN), M5P Tree (M5P), Support Vector Regression (SVR), Random Forest (RF), Additive Regression (AR), Random Subspace (RSS), and Locally Weighted Linear Regression (LWLR). The results of the water analysis provided by the Illizi State Water Resources Directorate were used in the study. The data includes the analysis results of 114 samples obtained from 57 wells covering the years 1999 to 2020. Models were evaluated using various statistical measures such as Correlation Coefficient (R), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Root Relative Square Error (RRSE). MLR and RF have been observed to produce reliable results in both scenarios.

In the study of Liu, Wang, & Sangaiah et al. (2019) [6], despite the data complexity of Internet of Things (IoT) based water quality monitoring systems, it is aimed to predict drinking water quality using Long-Short Term Memory (LSTM) deep neural networks. The prediction model was developed and its accuracy tested with the data obtained from the Guazhou Water Source automatic water quality monitoring station of the Yangtze River. The dataset contains 912 groups of data from 2016 to 2018 obtained from the automatic water quality monitoring station. The trend of change of water quality parameters over time was analyzed with 731 sets of data between 2016 and 2017. Correlation analysis was performed to analyze the relationship between water quality parameters. 100 iterations and Adam optimization algorithm were used in the training of the LSTM model. Autoregressive Integrated Moving Average (ARIMA) and SVR were used to compare the LSTM prediction effect. It was observed that LSTM has higher dissolved oxygen estimation precision than the other two estimation models.

In the study of Baek, Pyo, & Chun (2020) [7], it is aimed to develop a deep learning approach by using a combination of Convolutional Neural Network (CNN) and LSTM models to effectively predict water quality and water level. In the study, the data including the water quality and level of the Nakdong River Basin of South Korea were obtained from the Water Resources Management Information System (WAMIS) and Real-Time Water Quality Information (RTWQI) sources. CNN was used for the estimation of the water level and LSTM for the estimation of the water quality. When the results of the study were examined, it was seen that the CNN and LSTM models exhibited high performance and successfully reflected the water quality variations.

In the study of Singha, Pasupuleti, & Singh et al. (2021) [8], it is aimed to develop one skin learning-based and three machine learning-based models to control pollution and improve water management. The dataset used in the study, such as pH, total dissolved solids, total hardness, Ca2+, Mg2+, Na+, K+, $HCO3^-$, Cl^-, SO42−, $NO3^-$, F^- and PO43− of India's Mahanadi River includes various water quality parameters. In the study, a machine learning-based RF, ANN, and Extreme Gradient Boosting (XGBoost) and a deep learning-based multi-layer fed model were used. The models were evaluated with MAE, RMSE, RMSE-Observations Standard Deviation Ratio (RSR), Nash-Sutcliffe Efficiency (NSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). When the results of the study were examined, it was observed that the deep learning-based model outperformed the traditional machine learning-based models.

In the study of Aldhyani, Al-Yaari, & Alkahtani et al. (2020) [9], it is aimed to contribute to the management of water quality by estimating WQI and Water Quality Classification (WQC) with artificial intelligence algorithms. For WQI prediction, artificial

neural network models were developed with Nonlinear Autoregressive Neural Network (NARNET) and LSTM deep learning algorithms, and Support Vector Machine (SVM), K-Nearest Neighbor (K-NN) and Naive Bayes machine learning algorithms were used for WQC prediction. The data set consists of 1679 samples containing 7 important parameters such as dissolved oxygen, pH, conductivity, biological oxygen demand, nitrate, fecal coliform and total coliform collected from different regions. According to the results, it was observed that among the developed models, the LSTM model showed the most successful performance in WQI estimation, and the SVM algorithm in WQC estimation.

In the study of Bui, Khosravi, & Tiefenbacher et al. (2020) [10], it is aimed to evaluate water quality by using different data mining algorithms to generate WQI estimates for Iran. In this study, estimations were created using 4 independent and 12 hybrid data mining algorithms, a total of 16. In the study, monthly water quality data covering the years 2012-2018 were collected from 2 water quality monitoring stations. These data include biological oxygen demand, chemical oxygen demand, dissolved oxygen, pH, total solids, fecal coliform, phosphate, nitrate, turbidity and electrical conductivity measurements. As a result, the Bagging-Random Tree (BA-RT) algorithm is the most successful.

In the study of Abba, Pham, & Saini et al. (2020) [11], it is aimed to make predictions for the sustainable management of water resources by taking WQI data from three different stations on the Yamuna River. Different models such as Back Propagation Neural Network (BPNN) and Adaptive Neuro-Fuzzy Inference System (ANFIS), SVR and MLR were used to estimate WQI. To improve the performance of these models, a nonlinear composite technique using the Neural Network Aggregation (NNA) approach is proposed. The results obtained show that the developed models perform well in estimation and the NNE model gives the best result.

In the study of Khoi, Quan, & Linh et al. (2022) [12], it was aimed to evaluate the performance of twelve different machine learning models, five of which are reinforcement-based, three of which are Decision Tree (DT) based, and four of which are artificial neural network-based models, for predicting the surface water quality of La Buong River in Vietnam. The study used a dataset of 10 different water quality parameters (temperature, pH, dissolved oxygen, biological oxygen demand, chemical oxygen demand, turbidity, total suspended solids, coliform bacteria, ammonium and phosphate) collected bimonthly from four water quality monitoring stations of La Buong River in Vietnam between 2010 and 2017. When the results were analyzed, XGBoost was the most suitable machine learning algorithm in its field with R^2 and RMSE values in the test process.

In the study of Zou, Xiong, & Li et al. (2020) [13], it is aimed to obtain more accurate and comprehensive results in water quality estimation by developing a water quality estimation method based on a multi-time scale Bi-directional Long-Short Term Memory (Bi-LSTM) network. The dataset used includes the water quality data of the Beilun River station in the Beilun Estuary region, while the meteorological data includes data from a monitoring station as close as possible to the same area. Forecasting performance is enhanced by data preprocessing and the creation of sub-models. The proposed model has obtained better prediction results than other methods. When the results were examined, it was concluded that the usability of the model in more scenarios should be increased.

In the study conducted by Mohseni, Pande and Pal et al. (2024) [14], it was aimed to develop and compare MLR, ANN, SVM, RF and XGBoost machine learning models to estimate the Weighted Arithmetic Water Quality Index (WA-WQI). The results of the study were evaluated with R^2, RMSE, MAE, ROC and AUC and it was observed that the best result was given by the XGBoost model.

3 Materials

3.1 Dataset

In the study, two separate data sets collected from Adana and Mersin provinces were used. Adana data set was obtained from Adana Water and Sewerage Administration's (ASKI) website and Mersin data set was obtained from Mersin Water and Sewerage Administration's (MESKI) website[15][16]. Adana data set includes the results obtained from the analysis of Çatalan drinking water between and January 2009 and April 2023. Mersin data set includes the results of drinking water analysis obtained from different regions of Mersin between December 2023 and January 2018. Data sets were collected manually and there are 13 features in Adana data set and 11 features in Mersin data set. The found features include features such as pH, turbidity, total hardness, conductivity, dissolved oxygen, alkalinity, iron, silicon, chloride and nitrate that affect water quality and also include WQI value. Missing data in the data set were filled by taking the average. In addition, Min-Max normalization was

performed to ensure linear transformations in order to ensure balance between the values in the data set [17].

The calculation of the water quality index was made using the weighted arithmetic water quality index method. When looking at the index, 0-25 indicates perfect water, 26-50 good water, 51-75 bad water, 76-100 very bad water, and above indicates that the water is not drinkable [18].

In the calculation, Proportionality constant "K" value is calculated using Equation (1), which the allowable standard "s_i" for the nth parameter, after collecting the data on various water quality parameters. Then, the quality rating for the n^{th} parameter (q_n) is calculated with Equation (2).

Here, v_n = the estimated value of the nth parameter of the given sampling station, v_{io}= the ideal value of the nth parameter in pure water, and s_n=the allowable standard value of the nth parameter. With Equation (3), the unit weight is calculated for the nth parameters. WQI value is calculated by Equation (4) [18].

$$k = 1/(1/\sum_{!\#\$}^{"} s_!) \quad (1)$$

$$q = 100 * \{(v_" - v_{!\%})/(s_" - v_{!\%})\} \quad (2)$$

$$W_" = k/s_" \quad (3)$$

$$WQI = (\sum W_" * q_")/\sum W_" \quad (4)$$

3.2 Python

Python is a powerful programming language that is easy to read and understand with the features it offers. It has features in common with many other programming languages, such as data processing with variables and functions. Python is a programming language that is useful for machine learning applications [19].

3.3 Machine Learning

Machine learning is a branch of computer science that is a subset of artificial intelligence and aims to enable computers to learn without having to be programmed. Machine learning is generally divided into three classes: supervised, unsupervised, and semi-supervised. There is a dependent variable in supervised learning. Examples of commonly used techniques are linear and logistic regression. In unsupervised learning, grouping is attempted according to similar characteristics of the data. Semi-supervised learning allows working on data sets that contain both labeled and unlabeled data [20].

3.3.1 Linear Regression

Regression is used for estimation and prediction and also for determining the relationships between independent and dependent variables. In regression analysis, the dependent variable is estimated with the help of independent variables. In LR, as seen in Equation 5, the dependence of the independent variables 'x' and the dependent variable 'y' is defined [21].

$$y = \beta_\& + \beta_\$ x_\$ + \ldots + \beta_" x_" + \varepsilon \quad (5)$$

3.3.2 Ridge Regression

Ridge Regression (RR) is a method used to solve multiple linear problems. It involves adding some bias, the "ridge constant," to the regression equation to reduce the variance of the parameter estimators, thus controlling how much the least squares estimates differ [22].

3.3.3 Lasso Regression

LASSO regression is a method used to reduce unnecessary features of the model and increase model stability. A penalty term is added to the regression coefficients. The penalty is expressed as the sum of the absolute values of the coefficients. LASSO squeezes the coefficients of the less correlated variables to zero, leaving only the important features in the model. In this way, it helps the model to achieve the best fit by effectively dealing with overfitting and multicollinearity problems [23]

4 Results and Discussion

Data sets belonging to two different cities were used in the study. The Adana data set contains less data compared to the Mersin data set. Thus, it was desired to emphasize the importance of the size of the data used.

7 different models belonging to machine learning were created, namely LR, Lasso Regression, RR, Multilayer Perceptron (MLP), RF, DT and SVR. Both data sets were tested in these 7 models and the results were evaluated with Mean Squared Error (MSE) and R^2 metrics. In addition, the results obtained were improved with the normalization process performed on the data set. The obtained results were shown in Table 1 and Table 2.

Table 1. Results of Adana dataset

Models	MSE	R^2
LR	0.173	0.999
MLP	447.490	0.192
SVR	471.809	0.148
RF	213.237	0.615
DT	201.343	0.636
RR	17.852	0.967
Lasso Regression	2.659	0.995

Table 2. Results of Mersin dataset

Models	MSE	R^2
LR	9.589	0.999
MLP	0.238	0.999
SVR	8.966	0.734
RF	8.647	0.743
DT	8.720	0.741
RR	0.26	0.992
Lasso Regression	5.175	0.846

The fact that the MSE value is low and the R2 value is close to 1 indicates that the model is good. When the results for the Adana data set are examined, it is seen that the most successful results are obtained from LR and Lasso Regression. The least successful results are seen from SVR and MLP models. In MLP and SVR models, MSE values are above 400 and R2 values are below 0.2. The other remaining models produced an average result.

More successful results were obtained in the Mersin data set compared to the Adana data set. The reason for this is that the data is generally more. When the Mersin data set is examined in detail, it is observed that the most successful model is RR. In addition, the RR model is followed by MLP and Lasso Regression models. In addition, it is seen that the models generally obtained successful results in the results obtained in the Mersin data set. It is also noteworthy that the LR model and the RR model also obtained positive results in both cases.

5 Conclusion

The aim of the study is to make a contribution to sustainability by estimating the WQI value. In the study, WQI was estimated with 7 different machine learning models using data sets obtained from two different cities. According to the results obtained, it was observed that LR, RR and Lasso Regression models performed better than other models. However, it was observed that the results changed in different models in both data sets.

These results show that the performance of the models may vary depending on the characteristics of different data sets. Therefore, when creating the model, the suitability of the structure of the model and the structure of the data should also be taken into consideration. In the future, these models can be tested on larger and more comprehensive data sets and more general results can be obtained and contributions can be made. As a result, the results obtained in this study provide a basis for future studies by showing which models can give effective results in estimating WQI.

Acknowledgment

This research was supported by the Adana Alparslan Türkeş Science and Technology University Scientific Research Unit under Project number 23303002.

References

[1] Arora NK, Mishra I. "United Nations Sustainable Development Goals 2030 and environmental sustainability: race against time". Environmental Sustainability, 2(4), 339-342, 2019.

[2] Ighalo JO, Adeniyi AG. "A comprehensive review of water quality monitoring and assessment in Nigeria". Chemosphere, 260, 127569, 2020.

[3] Uddin MG, Nash S, Olbert AI. A review of water quality index models and their use for assessing surface water quality. Ecological Indicators, 122, 107218, 2021.

[4] Karunanidhi D, Aravinthasamy P, Subramani T, Muthusankar G. Revealing drinking water quality issues and possible health risks based on water quality index (WQI) method in the Shanmuganadhi

River basin of South India. Environmental Geochemistry and Health, 43, 931-948, 2021.

[5] Kouadri S, Elbeltagi A, Islam ARMT, Kateb S. Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). Applied Water Science, 11(12), 190, 2021.

[6] Liu P, Wang J, Sangaiah, AK, Xie Y, Yin X. Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. Sustainability, 11(7), 2058, 2019.

[7] Baek SS, Pyo J, Chun JA. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. Water, 12(12), 3399, 2020.

[8] Singha S, Pasupuleti S, Singha SS, Singh R, Kumar S. Prediction of groundwater quality using efficient machine learning technique. Chemosphere, 276, 130265, 2021.

[9] Aldhyani TH, Al-Yaari M, Alkahtani H, Maashi M. Water quality prediction using artificial intelligence algorithms. Applied Bionics and Biomechanics, 2020.

[10] Bui DT, Khosravi K, Tiefenbacher J, Nguyen H, Kazakis N. Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Science of the Total Environment, 721, 137612, 2020.

[11] Abba SI, Pham QB, Saini G, Linh NTT, Ahmed AN, Mohajane M, Bach QV. Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index. Environmental Science and Pollution Research, 27, 41524-41539, 2020.

[12] Khoi DN, Quan NT, Linh DQ, Nhi PTT, Thuy NTD. Using machine learning models for predicting the water quality index in the La Buong River, Vietnam. Water, 14(10), 1552, 2022.

[13] Zou Q, Xiong Q, Li Q, Yi H, Yu Y, Wu C. A water quality prediction method based on the multi-time scale bidirectional long short-term memory network. Environmental Science and Pollution Research, 27, 16853-16864, 2020.

[14] Mohseni U, Pande CB, Pal SC, Alshehri F. Prediction of weighted arithmetic water quality index for urban water quality using ensemble machine learning model. Chemosphere, 352, 141393, 2024.

[15] Adana Su ve Kanalizasyon İdaresi, "İÇME SUYU ANALİZ SONUÇLARI", https://www.adana-aski.gov.tr/web/icmesuyuanaliz.aspx

[16] Mersin Su ve Kanalizasyon İdaresi, "Analiz Sonuçları", https://www.meski.gov.tr/pages/AnalizSonuclari.xhtml

[17] Henderi H, Wahyuningsih T, Rahwanto E. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. International Journal of Informatics and Information Systems, 4(1), 13-20, 2021.

[18] Chandra DS, Asadi SS, Raju MVS. Estimation of water quality index by weighted arithmetic water quality index method: a model study. International Journal of Civil Engineering and Technology, 8(4), 1215-1222, 2017.

[19] Python W. Python. Python releases for windows, 24, 2021.

[20] Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. American journal of epidemiology, 188(12), 2222-2239, 2019.

[21] Maulud D, Abdulazeez AM. A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(2), 140-147, 2020.

[22] Rajan MP. An efficient Ridge regression algorithm with parameter estimation for data analysis in machine learning. SN Computer Science, 3(2), 171, 2022.

[23] Wang S, Chen Y, Cui Z, Lin L, Zong Y. Diabetes Risk Analysis Based on Machine Learning LASSO Regression Model. Journal of Theory and Practice of Engineering Science, 4(01), 58-64, 2024.

Üretken Yapay Zekanın Duygu Analizi Üzerindeki Etkileri

Beyza Nur YILDIRIM[1]*, Güncel SARIMAN[2]

[1]*Muğla Sıtkı Koçman Üniversitesi, Fen Bilimleri Enstitüsü, Yapay Zeka Anabilim Dalı, Muğla, TÜRKİYE*

[2] *Muğla Sıtkı Koçman Üniversitesi, Teknoloji Fakültesi, Bilişim Sistemleri Mühendisliği, Muğla, TÜRKİYE*

Özet

Bu çalışmanın amacı, OpenAI tarafından geliştirilen ChatGPT yapay zeka modeli tarafından üretilen yanıtlar üzerinde duygu analizi yapmaktır. ChatGPT gibi yapay zeka modellerinin kullanıcı etkileşimlerine yanıt verirken duygusal tonu dikkate alması önemlidir. Bu çalışmada, ChatGPT'nin duygusal tonunu anlamak için makine öğrenimi ve doğal dil işleme teknikleri kullanılmıştır.

Anahtar Kelimeler: *ChatGPT, Duygu Analizi, Doğal Dil İşleme, Makine Öğrenmesi*

The Effects of Generative Artificial Intelligence on Sentiment Analysis

Abstract

The aim of this study is to perform sentiment analysis on the responses generated by an artificial intelligence model, such as ChatGPT, developed by OpenAI. It is important for AI models like ChatGPT to consider the emotional tone when responding to user interactions. In this study, machine learning and natural language processing techniques will be used to understand the emotional tone of ChatGPT.

Keywords: *ChatGPT, Sentiment Analysis, Machine Learning*

1 Introduction

This study focuses on sentiment analysis, an important topic in the field of natural language processing (NLP). Sentiment analysis is a powerful technique used to identify, analyze, and classify the emotional content of text data. It is particularly useful for understanding and evaluating users' emotional responses to products, services, or experiences. This analysis enables businesses to better understand customer feedback, optimize marketing strategies, and improve product or service quality.

The main objective of this study is to enhance the level of human-like emotional intelligence in AI-based systems, thereby facilitating more effective human-machine interactions. Specifically, improving the ability of large-scale AI models like ChatGPT to accurately comprehend emotional tone can increase user satisfaction and significantly enhance interaction quality. These Technologies can quickly and accurately assess users' emotional states in areas such as customer service, social media analysis, and market research, thus providing more personalized and user-centered services.

The uniqueness of this study lies in demonstrating how AI models like Chat-GPT can be utilized in the field of sentiment analysis. By enhancing their ability to understand emotional expressions in various linguistic and cultural contexts, these models can better grasp users' emotional states and provide more appropriate responses. For instance, in a customer service application, they can accurately understand expressions of complaints or satisfaction, providing instant and effective feedback.

Overall, this study aims to make a significant contribution to the fields of artificial intelligence, natural language processing, and sentiment analysis. Accurately understanding the emotional tone of AI models like ChatGPT can make future human-machine interactions more effective, increase user satisfaction, and contribute to the

*Contact email: beyzaanuryildirim81@erciyes.edu.tr

generation of human-like responses. In this way, technology can become more responsive, effective, and user-centered, improving businesses' customer relationships and service quality.

2 Literature Review

In a study focused on sentiment analysis on YouTube, methods for determining the sentiment polarity of user comments directed at AI-based language models like ChatGPT were examined. The analysis of comments selected from YouTube's most popular videos utilized word-based sentiment analysis, negation detection, and linguistic resources like WordNet. The results indicated that ChatGPT could accurately predict emotional reactions in user comments [2].

Another study, conducted within the scope of Ek¸si S¨ozl¨uk entries, questioned the public's experiences with AI chatbots and discussed the role of AI in future communication. This study was carried out using data scraped from the Ekşi Sözlük platform. Content analysis of the data revealed that positive content categories related to AI were dominant over negative ones. Additionally, comparisons made over time showed that positive comments increased gradually

[3].

In a study examining the classification of online user reviews using sentiment analysis, reviews from an online book sales site were analyzed. The Multilayer Perceptron (MLP) algorithm was applied using Python, while Naive Bayes (NB), Support Vector Machines (SVM), and Logistic Regression (LR) algorithms were applied using RapidMiner. MLP yielded the best results with an accuracy rate of 89%. This study highlights the scarcity of sentiment analysis studies on Turkish datasets and emphasizes that MLP showed the highest performance in both positive and negative reviews [4].

In another study examining sentiment analysis of user comments on digital platforms such as social media and forums, sentiment analysis was conducted on movie reviews using machine learning algorithms like Naive Bayes, Centroid-Based Classifier, Multilayer Perceptron Neural Networks (MLP), and Support Vector Machines (SVM). It was observed that neural networks and support vector machines performed better than other methods. Particularly, SVM and MLP algorithms stood out with high accuracy rates in sentiment analysis [5].

A study comparing the performance of Naive Bayes, KNN, SVM, Logistic Regression, and Decision Tree algorithms on sentiment analysis of Turkish Twitter data used a dataset containing 16,000 positive and 16,000 negative tweets. After preprocessing the data, the best performance was achieved by the Support Vector Machine (SVM) with an f1 score of 73%.

A study focusing on the sentiment analysis and ranking of Black Friday hashtags using tweet statistics employed various algorithms to conduct sentiment analysis of tweets, and the results were evaluated statistically [7].

3 Materials and Methods

3.1 Dataset

Our dataset is a CSV file named "chatgpt reviews.csv" [1], containing 2,058 rows and obtained from the Kaggle platform. It consists of 10,000 user interactions and is derived from real-world interactions with ChatGPT. These interactions were collected from text-based chats occurring on various digital platforms. The average text length per interaction is approximately 150 characters. Among the most frequently used words in the dataset are positive words such as "hello," "thank you," "good," and "quick.".

	date	title	review	rating
0	2023-05-21 16:42:24	Much more accessible for blind users than the ...	Up to this point I've mostly been using ChatGP...	4
1	2023-05-27 21:57:27	4.5 stars, here's why	I recently downloaded the app and overall, it'...	4
2	2023-05-19 10:16:22	Almost 5 stars, but... no search function	This app would almost be perfect if it wasn't ...	4
3	2023-06-09 07:49:36	Good, but Siri support would take it to the ne...	I appreciate the devs implementing Siri suppor...	4
4	2023-05-31 10:20:48	App review	No doubt, this technology is absolutely life-c...	1
...	...	...	...	...
2053	2023-05-18 18:27:04	Fantastic App with Room for Enhancements	The ChatGPT iOS app is an outstanding product....	5
2054	2023-05-18 17:17:44	Awesome technology, deplorable tactics	Sam Altman's blatant attempt at regulatory cap...	2
2055	2023-06-25 04:55:57	Legit amazing	So I like to role-play on this app because of ...	5
2056	2023-06-25 04:20:59	Amazing!!	I'm so grateful that they finally added iPad c...	5
2057	2023-06-25 03:12:37	An absolute life saver	I have never written a review about an app bef...	5

2058 rows × 4 columns

Figure-1: Image of the dataset

Columns: Our dataset has four columns named date, title, review, and rating.

- date: Contains the date when the sample was taken.
- title: Contains the question directed to ChatGPT by the user.
- review: Contains the responses given by ChatGPT.
- rating: Contains the scores given by the user to these responses.

3.2 Data Preprocessing Steps

The dataset was first subjected to a step of examining missing data. A detailed inspection was

conducted to determine whether the data contained any missing values.

```
df.isnull().sum()

date      0
title     0
review    0
rating    0
dtype: int64
```

Figure-2: Missing data inspection

Subsequently, class imbalance in the dataset was identified and corrected to equalize the number of positive and negative class examples.

During the text cleaning process, unnecessary characters, special symbols, and HTML tags were removed from the text. Then, in the text normalization step, all text was converted to lowercase, non-alphanumeric characters were removed, and multiple spaces were replaced with a single space. Additionally, English stopwords from the NLTK library were removed from the texts.

The data was transformed into TF-IDF (Term Frequency-Inverse Document Frequency) vectors. These vectors were designated as the feature matrix to be used in the modeling process. Finally, the feature (X) and label (y) data were split into training and testing sets using a random seed (%80 training, %20 testing).

3.3 Proposed Model

In this study, we built and evaluated various machine learning models using ChatGPT user review data obtained from Kaggle.

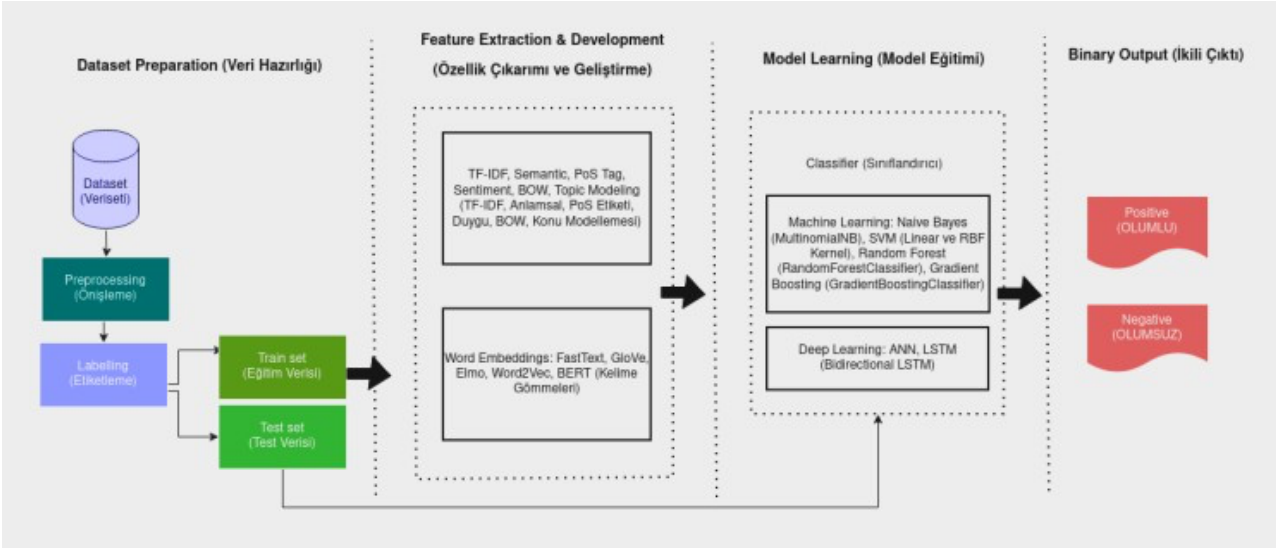


Figure-3: Proposed Model Architecture

We utilized an Artificial Neural Network (ANN) model, a commonly used deep learning method in text classification (LeCun et al., 2015)[13]. The ANN model includes an embedding layer, Spatial Dropout1D, Global Average Pooling1D, and dense layers. This architecture has been successfully applied for sentiment analysis on the ChatGPT dataset.

In this study, various models such as Naive Bayes, SVM, Random Forest, Gradient Boosting, and LSTM were also tested. The Naive Bayes model was trained using the MultinomialNB class with TF-IDF features and was successfully evaluated on the test dataset. Naive Bayes is known as a fast and simple classification method for text data, as highlighted in the literature (McCallum& Nigam, 1998)[8].

Support Vector Machines (SVM) were applied with both linear and RBF (radial basis function) kernels. SVM is a powerful supervised learning algorithm capable of creating clear distinctions between classes and non-linear decision boundaries (Cortes & Vapnik, 1995)[9]. Both SVM models were trained with TF-IDF features and their performances were evaluated on the test dataset. The Random Forest model was built using the RandomForestClassifier class. This ensemble method reduces the risk of overfitting by combining multiple decision trees and enhances the model's generalization ability (Breiman, 2001)[10]. The model was trained with TF-IDF features and effectively evaluated on the test dataset.

Gradient Boosting, an ensemble method that creates a strong predictive model by combining

weak learners, was employed in this study (Friedman, 2001)[11]. The model was trained using the GradientBoostingClassifier class and was evaluated on the test dataset using TF-IDF features.

The Long Short-Term Memory (LSTM) model, a deep learning architecture particularly effective in time series data, was also applied (Hochreiter & Schmidhuber, 1997)[12]. This model includes an embedding layer, Bidirectional LSTM layer, and dense layers. The LSTM model was trained on ChatGPT user review data, and its performance in sentiment analysis was tested.

These models, used alongside large-scale artificial intelligence models, offer significant potential in the field of sentiment analysis. The study aims to demonstrate how different approaches and models in AI-based sentiment analysis can be applied and contribute to research in the field.

4 Result and Discussion

In conclusion, this study provides a powerful and effective tool for automatically analyzing emotional content by comparing the performance of different artificial neural network models in sentiment analysis tasks. Specifically, the comparison of machine learning and deep learning techniques has helped identify which approaches yield the best results in this field. This can facilitate the further development of AI-based applications in the area of emotional intelligence, enabling the creation of systems that better understand users' emotional tones and respond accordingly. Additionally, such approaches can enrich and personalize the user experience, making human-computer interactions more natural and satisfying..

In this context, the findings of this study offer valuable contributions to both academic research and industrial applications.

References

(Kaynaklar kısmı, Cambria yazıtipinde, 10 punto yazı boyutunda aşağıdaki örnekelerde görüldüğü gibi olmalıdır.)

[1] ChatGPT App Reviews for Sentiment Analysis. https://www.kaggle.com/code/saloni1712/chatgpt-app-reviews-for-sentiment-analysis/notebook.

[2] Çubukçu Çerasi, Ceren & Balcioglu, Yavuz. (2023). SENTIMENT ANALYSIS ON YOUTUBE: FOR CHATGPT.

[3] Gazaz, D. C., & Gazaz, G. (2024). Geleceğin İletişiminde Yapay Zekâ Sohbet Robotları: Ekşi Sözlük Girdilerinin İçerik Analizi. Türkiye İletişim Araştırmaları Dergisi(Cumhuriyetin 100. Yılında Geleceğin İletişimi Özel Sayısı), 30-54. https://doi.org/10.17829/turcom.1360268.

[4] S. Tuzcu, "Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması", ESTUDAM Bilişim, c. 1, sy. 2, ss. 1–5, 2020.

[5] O. KAYNAR Et Al. , "Makine Öğrenmesi Yöntemleri ile Duygu Analizi," International Artificial Intelligence and Data Processing Symposium (IDAP'xx16) , Malatya, Turkey, pp.234-241, 2016.

[6] E. Kumaş, "Türkçe Twitter Verilerinden Duygu Analizi Yapılırken Sınıflandırıcıların Karşılaştırılması", ESTUDAM Bilişim, c. 2, sy. 2, ss. 1–5, 2021.

[7] Kılıç, G., Budak, İ., & Kılıç, B. S. (2020). Kara Cuma Etiketlerinin Tweet İstatistikleri ve Duygu Analizi ile Sıralanması. Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi, 23(1), 131-140. https://doi.org/10.29249/selcuksbmyd.638064.

[8] K.-M. Schneider, "A comparison of event models for Naive Bayes anti-spam e-mail filtering", in tenth conf., Budapest, Hungary, Apr. 12–17, 2003. Morristown, NJ, USA: Assoc. Comput. Linguistics, 2003. Accessed: Oct. 22, 2024. [Online]. Available: https://doi.org/10.3115/1067807.1067848.

[9] Corinna Cortes and Vladimir Vapnik. 1995. Support-Vector Networks. Mach. Learn. 20, 3 (Sept. 1995), 273–297. https://doi.org/10.1023/A:1022627411411T.C.

[10] Leo Breiman. 2001. Random Forests. Mach. Learn. 45, 1 (October 1 2001), 5–32. https://doi.org/10.1023/A:1010933404324.

[11] Friedman, Jerome. (2000). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics. 29. 10.1214/aos/1013203451.

[12] Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. Neural Comput 1997; 9 (8): 1735–1780. doi: https://doi.org/10.1162/neco.1997.9.8.1735.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. Accessed: Oct. 22, 2024. [Online]. Available: https://doi.org/10.1038/nature14539

Makine Öğrenmesi Yöntemleri ile Atölye Yoğunluk Tahminlemesi

Eylül Dilara Eren[1]*, Mehmet Emin Eker[2], Buğrahan Erdönmez[3], İhsan Sadati[4,5]

[1,2,3] *Doğuş Teknoloji, İstanbul, TÜRKİYE*

[4]*Sabancı Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, İstanbul, TÜRKİYE*

[5]*Smart Mobility and Logistics Lab, Sabanci University, Istanbul, Turkey*

Özet

Cihaz Onarım Atölyeleri, günümüzde yaygın olarak kullanılan akıllı telefonların bakımı ve onarımı için önemli bir hizmet sunmaktadır. Bu çalışma, bir teknoloji firmasının servislerinin atölyelerindeki yoğunluğun etkin bir şekilde tahmin edilmesi amacıyla makine öğrenmesi modellerinin kullanılmasını araştırmaktadır. Atölye yoğunluğunun doğru bir şekilde tahmin edilmesi, müşteri memnuniyeti ve vardiya planlama verimliliği açısından kritik bir öneme sahiptir. Bu araştırma, servislerdeki yoğunluğu etkileyen ana faktörleri belirlemek için mevcut literatürü incelemekte ve çeşitli makine öğrenmesi algoritmalarını kullanarak bir tahmin modeli geliştirmektedir. Model, geçmiş tamir işlemleri verilerini ve atölye içindeki diğer önemli değişkenleri kullanarak gelecekteki yoğunluk seviyelerini tahmin etmeyi amaçlamaktadır. Çalışma ayrıca, cihaz onarım süreçlerinin sürelerini tahmin ederek teknisyenlere iş ataması yapmayı ve böylece kaynak optimizasyonu sağlamayı amaçlamaktadır. Bu çalışma, onarım merkezlerinde teknisyen verimliliğini %20 artırmayı, müşteri memnuniyetini %15 yükseltmeyi ve bekleme sürelerini %50 oranında azaltmayı hedeflemektedir. Bu araştırma ile, onarım merkezlerinde yaşanan cihaz problem teşhis ve onarım süreçlerini optimize ederek, müşteri bekleme sürelerini azaltmayı ve hizmet kalitesini artırmayı hedeflenmektedir.

Anahtar Kelimeler: *Veri Bilimi, tahminleme, LSTM (Uzun-Kısa Vadeli Bellek), ARIMA, Rastgele Orman Regresyonu*

Workshop Workload Prediction with Machine Learning Methods

Abstract

Device repair workshops provide an essential service for the maintenance and repair of the ubiquitous devices we rely on today. This study explores the application of machine learning models to accurately forecast the workload in tech company service centers. Precise workload predictions are vital for effective shift planning and enhancing customer satisfaction. To identify the key factors influencing service workload, this research reviews existing literature and develops a prediction model utilizing various machine learning techniques. The model leverages historical repair data and significant workshop characteristics to predict future workloads. Additionally, the project aims to optimize resource allocation by estimating repair times and efficiently distributing tasks among technicians. Conducted in collaboration with Sabancı University, the project's objectives include reducing waiting times at repair shops by 50%, improving customer satisfaction by 15%, and increasing technician productivity by 20%. Ultimately, this project seeks to

* Contact email: eylul.eren@d-teknoloji.com.tr

minimize client wait times, improve service quality, and streamline diagnostic and repair processes in device repair centers.

Keywords: *Data Science, forecasting, LSTM (Long-Short Term Memory), ARIMA, Random Forest Regression*

1 Giriş

Günümüzde teknolojinin hızla ilerlemesiyle birlikte, elektronik cihazlar günlük yaşamın vazgeçilmez bir parçası haline gelmiştir. Ancak, bu cihazlar zamanla arızalanabilir veya performans sorunları yaşayabilirler. Bu noktada, cihaz onarım ve tamir atölyeleri önemli bir role sahip olmaktadır. Cihaz onarım atölyeleri, çeşitli elektronik cihazların tamir edilmesi, bakımı ve yenilenmesi konularında uzmanlaşmış profesyonel hizmetler sunmaktadır. [3] Bu atölyeler hem bireysel kullanıcıların hem de işletmelerin teknolojik cihazlarının sürekli çalışabilir durumda olmasını sağlamak için kritik bir rol oynamaktadır.[1] Özellikle tüketici elektroniği sektöründeki hızlı değişimler ve yenilikler, cihaz onarım atölyelerinin sunduğu hizmetlerin önemini artırmaktadır. [2] Cihaz onarım ve tamir atölyeleri, günümüzde hızla değişen ve gelişen teknolojiye ayak uydurabilmek için sürekli olarak yenilikçi çözümler arayışındadır. Bu atölyeler, elektronik cihazların tamiri ve bakımı konusunda uzmanlaşmış olup, kullanıcıların teknolojik cihazlarını uzun ömürlü ve güvenilir bir şekilde kullanabilmelerini sağlamaktadır. [4] Ancak, atölyelerin verimliliği ve hizmet kalitesi, doğru personel yönetimi ve gelecekteki teknolojik trendlerin doğru tahmin edilmesiyle doğrudan ilişkilidir. Son yıllarda, bu zorlukların üstesinden gelmek için makine öğrenmesi ve yapay zekâ tekniklerinin kullanımı önem kazanmıştır. [5]

Bu bildiri, bir cihaz onarım-tamir atölyesinde personel optimizasyonu ve gelecek cihazları tahminlemek için geliştirilen bir makine öğrenmesi modeli üzerine odaklanmaktadır. İlk olarak, atölyelerde personel yönetiminin önemi ve mevcut zorluklar ele alınacak, ardından makine öğrenmesi modellerinin bu alanlarda nasıl kullanılabileceği tartışılacaktır. Son olarak, önerilen modelin etkinliği ve gelecekteki araştırma yönleri üzerinde durulacaktır.

Bu çalışma, cihaz onarım ve tamir atölyelerinin operasyonel verimliliğini artırmak ve gelecekteki teknoloji trendlerine uyum sağlamak için yeni bir yaklaşım sunmayı amaçlamaktadır.

2 Literatür Taraması

Cihaz onarım atölyelerindeki yoğunluk tahminlemesi hem müşteri memnuniyeti hem de işletme verimliliği açısından büyük önem taşımaktadır. Literatür taraması, bu alanda zaman serisi analizi, regresyon modelleri ve derin öğrenme gibi çeşitli makine öğrenmesi yöntemlerinin kullanıldığını göstermektedir.

Zaman Serisi Analizi:

Taylor ve ark. (2020): Servis çağrı merkezlerindeki çağrı hacmini tahmin etmek için ARIMA ve SARIMA modellerini başarıyla kullanmışlardır. Bu çalışmada, mevsimsellik ve trend gibi faktörlerin çağrı hacmi üzerindeki etkisi incelenmiştir. [15]

Zhang ve ark. (2018): Bir e-ticaret şirketindeki ürün taleplerini tahmin etmek için LSTM (Uzun Kısa Süreli Hafıza) ağları kullanmışlardır. Bu çalışmada, ürünlerin popülaritesindeki zamanla değişen trendler ve mevsimsellik etkileri dikkate alınmıştır. [16]

Regresyon Modelleri:

Wang ve ark. (2017): Bir üretim tesisindeki makine arıza oranlarını tahmin etmek için Poisson regresyonunu kullanmışlardır. Bu çalışmada, makine yaşının, çalışma saatlerinin ve çevresel faktörlerin arıza oranı üzerindeki etkisi incelenmiştir. [17]

Chen ve ark. (2016): Bir havalimanındaki yolcu trafiğini tahmin etmek için çoklu doğrusal regresyon modelini kullanmışlardır. Bu çalışmada, hava durumu, uçuş sayısı ve özel günler gibi faktörler dikkate alınmıştır. [18]

Derin Öğrenme:

Li ve ark. (2023): IoT sensörlerinden elde edilen gerçek zamanlı verileri kullanarak bir üretim tesisindeki enerji tüketimini tahmin etmek için bir dikkat mekanizmalı transformer modeli geliştirmişlerdir. Bu çalışmada, farklı sensör verileri arasındaki karmaşık ilişkiler daha iyi öğrenilmiştir. [19]

Schmidhuber ve Hochreiter (1997): LSTM (Uzun Kısa Süreli Hafıza) ağlarını ilk olarak önermişler ve

bu ağların uzun vadeli bağımlılıkları modellemedeki başarısını göstermişlerdir. [20]

Çalışmamızın Farklılıkları ve Katkıları

Bu çalışmada, akıllı telefon onarım atölyelerine özgü verileri kullanarak daha derinlemesine bir analiz yapılmıştır. Cihaz arıza türleri, model ve marka gibi spesifik değişkenlerin yanı sıra, hava durumu, mevsimsellik ve özel günler gibi dışsal faktörlerin de modelde dikkate alınması, tahmin doğruluğunu artırmayı amaçlamaktadır. Ayrıca, RNN modeli sayesinde, farklı veri türleri arasındaki karmaşık ilişkiler daha iyi öğrenilmiştir.

Mevcut çalışmalarla karşılaştırıldığında, bu araştırmanın en önemli katkısı, cihaz onarım süreçlerinin tüm aşamalarını kapsayan kapsamlı bir model geliştirerek, daha doğru ve detaylı tahminler elde etmesidir. Bu sayede, atölye yöneticileri, kaynakları daha etkin bir şekilde planlayabilir ve müşteri memnuniyetini artırabilirler.

Özetle, literatür taraması, cihaz onarım atölyelerindeki yoğunluk tahmini konusunda çeşitli makine öğrenmesi yöntemlerinin kullanıldığını göstermektedir. Bu çalışmada, bu yöntemlerin üzerine çıkarak, daha kapsamlı bir model geliştirilmesi ve atölye yönetimi için yeni bir bakış açısı sunulması amaçlanmaktadır.

3 Problemin Tanımlanması

Servis yönetim sistemi, elektronik cihazlar ve taşınabilir dayanıklı tüketim mallarının satış sonrası destek hizmetlerini kapsayan kapsamlı bir platformdur. Ancak, mevcut sistemde çeşitli operasyonel zorluklar ve müşteri memnuniyetsizlikleri yaşanmaktadır.

Öncelikle, onarım merkezlerine gelen cihazların geçmiş onarımlarıyla ilgili bilgilerin farklı veri tabanları veya formatlarda tutulması ve bu verilerin sistematik bir şekilde entegre edilememesi büyük bir zorluk teşkil etmektedir. Bu durum, veri analitiği süreçlerinde veri bütünlüğü ve erişim sorunlarına yol açarak, doğru analizlerin yapılmasını engelleyebilmektedir. Tekrar eden arızalar veya cihaz değişimleri sırasında, geçmiş onarım bilgilerine erişimde yaşanan zorluklar, teknisyenlerin işlerini etkin bir şekilde yapmasını engelleyebilmektedir. Özellikle, geçmiş onarımların detaylı bir şekilde analiz edilmesi, gelecekteki taleplerin tahminlenmesi ve kapasite planlamasının doğru yapılması için kritik öneme sahiptir.

İkinci olarak, servis merkezlerine gelen müşteri taleplerinin doğru ve zamanında tahmin edilememesi, iş gücü planlaması ve kaynak yönetimi açısından önemli bir eksiklik olarak karşımıza çıkmaktadır. Geçmiş veri analizlerinin yetersiz veya eksik olması, iş yükünün dengesiz dağılmasına ve bazı dönemlerde aşırı yoğunluk yaşanmasına sebep olabilmektedir. Bu durum hem müşteri memnuniyetini olumsuz etkileyebilmekte hem de operasyonel maliyetleri artırabilmektedir.

Bu sorunları çözmek ve servis yönetimindeki verimliliği artırmak için yapılacak iyileştirmeler, teknoloji destekli çözümler ve veri analitiği kullanımını içermelidir. Müşteri bekleme sürelerinin kısaltılması, doğru kapasite planlaması yapılması ve müşteri memnuniyetinin artırılması önemli hedefler arasındadır. Bu bağlamda, sistemde yapılacak güncellemelerle operasyonel süreçlerin optimize edilmesi ve müşteri deneyiminin iyileştirilebilir.

4 Winsorization Nedir?

Winsorization, istatistiksel analizlerde aşırı değerlerin (outlier) etkisini azaltmak için kullanılan bir tekniktir. Bu yöntem, veri setindeki en düşük ve en yüksek değerleri belirli bir yüzdede keserek, bu değerleri daha merkezi bir değere dönüştürmeyi amaçlar. Genellikle, veri setinin alt %5 ve üst %5'lik dilimindeki değerler, sırasıyla alt ve üst sınır değerleri ile değiştirilir. Bu sayede, aşırı değerlerin analizin sonucunu bozması engellenir.

Winsorization, adını Amerikalı istatistikçi Charles P. Winsor'dan alır. Bu teknik, istatistiksel verilerin aşırı değerlerini sınırlayarak, bu değerlerin olası yanıltıcı etkilerini azaltmayı hedefler. Aşırı değerler, veri setinin genel eğilimini etkileyebilir ve bu da yanıltıcı sonuçlara yol açabilir. Winsorization, bu aşırı değerlerin etkisini azaltarak daha sağlam bir analiz yapılmasına olanak tanır.[7][9]

İlk olarak, analiz edilecek veri seti belirlenir. Ardından yüzde aralığı belirlenir. Genellikle %5 veya %10 gibi yüzdeler belirlenir. Bu, hangi değerlerin kesileceğini tanımlar. Belirlenen yüzdelere göre veri setinin alt ve üst sınır değerleri hesaplanır. Örneğin, %5'lik bir Winsorization için, veri setinin en düşük %5'lik dilimindeki değerler, en düşük değerle değiştirilir; en yüksek %5'lik

dilimindeki değerler ise en yüksek değerle değiştirilir.[6][8] Hesaplanan sınırlar doğrultusunda, veri setindeki aşırı değerler güncellenir.

Winsorization, birkaç önemli nedenden dolayı tercih edilir. Aşırı değerler, ortalama, standart sapma gibi istatistiksel ölçümleri önemli ölçüde etkileyebilir. Winsorization, aşırı değerlerin etkisini azaltarak bu ölçümlerin daha güvenilir olmasını sağlar.[9] Aşırı değerler, veri dağılımının normalleşmesini engelleyebilir. Winsorization, veri dağılımını daha simetrik hale getirerek, istatistiksel testlerin geçerliliğini artırır.[8] Özellikle regresyon analizlerinde, aşırı değerler modelin tahmin gücünü düşürebilir. Winsorization, modelin daha iyi performans göstermesine yardımcı olur.[6]

5 RNN Nedir?

Yinelemeli Sinir Ağları (RNN), sıralı verileri işlemek için tasarlanmış bir yapay zeka modelidir. Geleneksel sinir ağlarından farklı olarak, RNN'ler geçmiş bilgilerden yararlanarak daha iyi tahminler yapabilme yeteneğine sahiptir. Bu özellik, RNN'leri doğal dil işleme (NLP), zaman serisi analizi ve diğer sıralı veri uygulamaları için son derece uygun hale getirir.

RNN, "Recurrent Neural Network" ifadesinin kısaltmasıdır. Bu ağlar, verileri zaman içinde işleyerek, önceki adımlardan gelen bilgileri hatırlama yeteneğine sahiptir. RNN'ler, sıralı verilerin birbirleriyle olan ilişkilerini anlamak için tasarlanmıştır. Örneğin, metinlerdeki kelimelerin sıralı ilişkilerini veya zaman serisi verilerindeki trendleri analiz edebilirler. RNN'ler, özellikle dil çevirisi, ses tanıma ve video analizi gibi alanlarda yaygın olarak kullanılmaktadır.[10][11]

RNN'lerin temel çalışma prensibi, her bir verinin önceki verilerle bağlantılı olduğu fikrine dayanır. RNN'ler, her bir zaman adımında bir "gizli durum" (hidden state) tutarak, geçmiş bilgileri saklar. Bu gizli durum, ağın geçmişteki bilgileri hatırlamasına ve yeni gelen verilerle birleştirerek daha iyi tahminlerde bulunmasına olanak tanır.

RNN, sıralı verileri bir dizi olarak alır. Her bir veri, bir zaman adımında işlenir. Her zaman adımında, mevcut veri ve önceki gizli durum kullanılarak yeni bir gizli durum hesaplanır. Hesaplanan gizli durum, çıkış katmanına iletilerek bir tahmin yapılır. Tahmin ile gerçek değer arasındaki hata hesaplanır ve bu hata, ağırlıkların güncellenmesi için geri yayılır.[10][13]

RNN'ler, sıralı verileri etkili bir şekilde işleyebilir. Bu, doğal dil işleme gibi alanlarda kelimelerin bağlamını anlamak için kritik öneme sahiptir. RNN'ler, geçmiş bilgileri hatırlayarak uzun süreli bağımlılıkları öğrenebilir. Bu özellik, dil çevirisi ve metin analizi gibi uygulamalarda önemlidir.[11][13] RNN'ler, farklı uzunlukta giriş dizilerini işleyebilir. Bu, onları zaman serisi verileri gibi değişken uzunluktaki verilere uygulamak için ideal hale getirir.

RNN'lerin bazı sınırlamaları da bulunmaktadır. Örneğin, "kaybolan gradyan" ve "patlayan gradyan" problemleri, RNN'lerin eğitim sürecinde sıkça karşılaşılan sorunlardır. Kaybolan gradyan, uzun dizilerde bilgi kaybına yol açarken, patlayan gradyan, gradyanların aşırı büyümesine neden olabilir ve bu da öğrenmeyi zorlaştırır.[11][12]

6 RMSE Nedir?

RMSE (Root Mean Square Error), tahmin edilen değerler ile gerçek değerler arasındaki farkın karelerinin ortalamasının karekökünü alarak hesaplanan bir hata ölçüsüdür. RMSE, modelin tahmin performansını değerlendirmek için yaygın olarak kullanılan bir istatistiksel yöntemdir. Özellikle regresyon analizi, makine öğrenimi ve istatistiksel modelleme gibi alanlarda kullanılır.

RMSE, tahmin edilen değerler ile gerçek değerler arasındaki farkların ölçülmesinde kullanılan bir metriktir. Matematiksel olarak, RMSE şu şekilde tanımlanır:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \acute{y}_i)^2}$$

Burada y_i gerçek değerler, $\hat{y}_i$ tahmin edilen değerler ve n veri noktalarının sayısını temsil eder. RMSE, hata terimlerinin karelerinin ortalamasını alarak, bu hataların büyüklüğünü vurgular. Bu özellik, RMSE'yi diğer hata ölçütlerine göre daha duyarlı hale getirir, çünkü büyük hatalar daha fazla etkili olur.

RMSE hesaplamak için ilk olarak, modelin tahmin ettiği değerler ile gerçek değerler belirlenmelidir. Her bir veri noktası için hata terimi, gerçek değer ile tahmin edilen değer arasındaki fark olarak hesaplanır. Hesaplanan hata terimleri kare alınarak pozitif değerler elde edilir. Hata terimlerinin karelerinin ortalaması alınır. Son olarak, hesaplanan ortalamanın karekökü alınarak RMSE değeri elde edilir.

RMSE, birkaç önemli nedenden dolayı tercih edilmektedir. RMSE, modelin tahmin doğruluğunu değerlendirmek için etkili bir yöntemdir. Düşük RMSE değerleri, modelin gerçek verilere yakın tahminler yaptığını gösterir. RMSE, hata terimlerinin karelerini kullandığı için, büyük hatalar daha fazla ağırlık taşır. Bu, modelin aşırı değerlerden nasıl etkilendiğini anlamak için önemlidir. Farklı modellerin performansını karşılaştırmak için RMSE kullanılabilir. Daha düşük RMSE değerine sahip olan model, genellikle daha iyi bir tahmin performansı sergiler. RMSE, modelin performansını izlemek ve iyileştirmek için bir geri bildirim mekanizması sağlar. Modelin RMSE değeri zamanla izlenerek, gerekli ayarlamalar yapılabilir.

RMSE'nin bazı sınırlamaları da bulunmaktadır. Örneğin, RMSE, hata dağılımının normal dağılım varsayımına dayanır. Ayrıca, RMSE, hata terimlerinin karelerini kullandığı için, büyük hataların etkisini artırır. Bu durum, bazı uygulamalarda yanıltıcı sonuçlara yol açabilir. [14]

7 Veri Analizleri ve Model Çalışmaları

Veri analizi ve model çalışmaları bölümü, projenin temel adımlarını ve elde edilen sonuçları detaylı bir şekilde sunmaktadır. İlk olarak, proje kapsamında yapılan ilk çalışma veri temelli olup, tüm servis bilgileri ilgili veri tabanından alınmıştır. Bu veri seti üzerinde çeşitli tahminleme metotları denenmiş ve başlangıçta genel müşteri tahminlemesi üzerine odaklanılmıştır.

Ancak, müşteriyle yapılan toplantılar sonucunda projenin odak noktası değiştirilmiş ve artık Atölye hizmetine ihtiyaç duyan müşteri segmentine yönelik tahminlemeler yapılmaya başlanmıştır. Bu doğrultuda, pilot uygulama sürecinde tüm taşınabilir cihazlar dikkate alınarak veri temizliği işlemleri titizlikle gerçekleştirilmiştir. Veri setinden elde edilen bilgilerle birlikte, cihazların atölyede geçirdikleri tamir süreleri hesaplanmış ve bu veriler tahminleme modüllerinin oluşturulmasında kullanılmıştır.

Özellikle servislere gelen müşterilerin 2020 yılından itibaren kaydedilen verileri üzerinde detaylı analizler yapılmış ve sadece atölye hizmetine yönelik veriler incelenmiştir. Bu süreçte, müşteri ile belirlenen veri kategorileri arasında atölye tamir süresi özellikle önem arz etmektedir. Veri analizi sırasında, bazı veri noktalarının eksik olduğu veya hatalı olduğu fark edilmiş ve bu verilerin tahminleme sonuçlarını olumsuz etkileyebileceği anlaşılarak, ilgili veriler manipüle edilmiştir. Atölyelerin çalışma saatleri 08.00-18.00 arası baz alınarak bu aralık dışında kalan saatlerde yapılan işlemler bu aralığa dahil edilmiştir.

Yapılan işlemler sonucunda medyan tamir süresinin 59 dakika olduğu ve ortalama tamir süresinin 200 dakika olduğu görülmüştür. Aradaki farkın sebebi için tamir süreleri araştırıldığında 30 dakikanın altında kalan ve bir günün üzerinde tamir süresi olan kayıtlar olduğu görülmüştür. Müşteri ile görüşüldüğünde 30 dakikanın altında kalan ve bir günün üzerinde olan tamir sürelerinin anlamsız olduğu anlaşılmıştır. Bunun üzerine tamir süresi 30 dakikadan kısa olan toplam 28.788 kayıt tamir süresi 30 dakika olacak şekilde düzenlendi. Tamir süresi bir günü yani 480 dakikayı aşan toplam 17.202 kayıt düzenlendi. Bu işlemler sonrasında medyan tamir süresinin 59 dakika, ortalama tamir süresinin ise 66 dakika olduğu gözlemlendi.

Elde edilen veri, atölye tarihi ve servis bazında gruplanarak her gün servislere kaç cihaz geldiği ve bu cihazların toplam tamir süresi hesaplandı.

Veri dağılımının merkezi eğiliminin ve dağılımının yayılmasının hesaplanması için IQR, z_score_outlier ve winsorization veri üzerinde denendi. Ardından servis noktası bazında winsorization kullanılmaya karar verildi. Winsorization işlemi sonrasında bazı servislerin belirli günlerde çok boş olduğu görüldü. Bu durumun sebebi araştırıldığında bu boş günlerin

çoğunluğunun pandemi dönemine denk geldiği, diğer boş günlerin ise yarım gün çalışan servislerden kaynaklandığı görüldü.

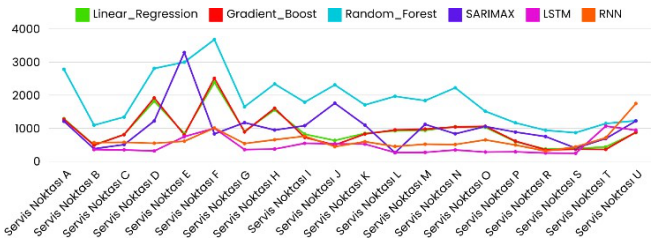
Güncel veriler, uygun kategorilere ayrılarak analiz edilmiş ve bu kategoriler arasında Servis Noktaları, Cihaz Model Kategorileri, Cihaz Garanti Durumu, Cihaz Tamir Süreleri ve Cihaz Semptom Kategorileri bulunmaktadır. Bu veri kategorileri üzerinde yapılan analizler, ML modellerinin oluşturulması ve doğru tahminleme algoritmalarının seçilmesi sürecinde temel veri sağlamıştır.

Sonuç olarak, veri analizi ve model çalışmaları, projenin başlangıç noktasından itibaren elde edilen veri setlerinin derinlemesine incelenmesini ve bu verilerin doğru şekilde kullanılarak tahminleme modellerinin geliştirilmesini içermektedir. Bu çalışmalar, servis merkezlerinin verimliliğini artırmak, müşteri memnuniyetini yükseltmek ve operasyonel süreçleri optimize etmek için kritik öneme sahiptir.

Proje kapsamında kullanılan makine öğrenimi modelleri, servis merkezlerindeki müşteri yoğunluğunu tahmin etmek, atölye işlemlerini planlamak ve teknik personel performansını optimize etmek için çeşitli veri analiz ve tahminleme yöntemleriyle geliştirilmiştir. İlk olarak, müşteri taleplerinin tahminlenmesi amacıyla çeşitli zaman serisi modelleri denenmiştir. Bu modeller arasında ARIMA (Oto-Regressif Entegre Hareketli Ortalama), SARIMA (Sezonluk ARIMA), ARIMAX (ARIMA ile genişletilmiş model), SARIMAX (Sezonluk ARIMAX) gibi klasik zaman serisi yöntemleri kullanılmıştır. Ayrıca, doğrusal regresyon teknikleri olan RandomForestRegressor ve GradientBoostingRegressor gibi ensemble yöntemleri de performanslarını değerlendirmek için uygulanmıştır.

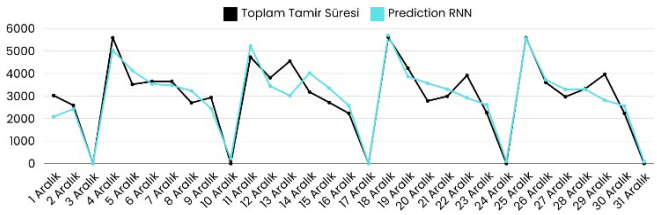
Tahminleme süreçlerinde derin öğrenme teknikleri de denenmiştir. Uzun ve kısa vadeli bellek (LSTM) kullanılarak yapılan derin öğrenme modelleri, özellikle zaman serisi verilerindeki karmaşıklığı ele almak ve gelecekteki müşteri taleplerini tahmin etmek için etkili olmuştur. Ayrıca, Facebook tarafından geliştirilen Prophet modeli, zaman serilerindeki mevsimsel değişimleri ve trendleri tahmin etmede başarılı sonuçlar vermiştir. Modellerin performansı, Ortalama Mutlak Hata (MAE) ve Kök Ortalama Kare Hatası (RMSE) gibi metriklerle değerlendirilmiş ve en iyi tahmin performansı sağlayan model seçilerek uygulamaya geçirilmiştir.

Atölye işlemleri için yapılan modeller ise, teknisyenlerin iş tamamlama sürelerini tahmin etmek ve atölye personelini etkin bir şekilde yönlendirmek üzere geliştirilmiştir. Geleneksel İstatiksel: Linear Regression, Ağaç Temelli: Random Forest Regressor, Gradient Boosting Regressor, Derin Öğrenme: LSTM, RNN ve Zaman Serisi: ARIMA makine öğrenimi algoritmaları denenerek performansları değerlendirildi.



Servis Noktalarındaki Tamir Sürelerine Göre Modellerin Tahminlenmesi

Geliştirilen modelin başarı oranı RMSE ile ölçülerek performansı değerlendirilmiştir.



Bir Servis Noktasının Gerçek Tamir Süresi ve Modelin Tahminlemesinin Karşılaştırılması

Projenin amacı, veri odaklı yaklaşımlarla servis süreçlerini optimize etmek ve müşteri memnuniyetini artırmak olduğundan, makine öğrenimi modelleri sürekli olarak güncellenmekte ve iyileştirilmektedir.

8 Sonuçlar

Taşınabilir cihazlar, özellikle akıllı telefonlar, tabletler ve dijital saatler gibi ürünlere odaklanarak firmanın satış sonrası destek hizmetlerini iyileştirme çabalarımız, talebi tahmin etme ve buna

göre teknik personeli etkili bir şekilde atama üzerine yoğunlaşmıştır. Kullanılan tahminleme modelleri, karmaşık bağımlılıkları ele almak için Uzun Kısa Süreli Bellek (LSTM), mevsimsel eğilimleri yakalamak için ARIMA ve adaptasyon yeteneği için Üstel Yumuşatma gibi gelişmiş tekniklerden faydalanmıştır. Bu metodolojilerin kullanımı, Kök Ortalama Kare Hatası (RMSE) değerini minimumda tutarak servis merkezlerimizin gelen onarım taleplerine yeterli sayıda ve zamanında teknik personel sağlamasını sağlamıştır.

Şube Adı	Teknisyen Sayısı	Toplam Tamir Süresi	Gerekli Personel	Kullanım Oranı %
Servis Noktası A	10	91152	7	73
Servis Noktası B	6	54184	4	72
Servis Noktası C	6	50594	4	67
Servis Noktası D	6	37538	3	50
Servis Noktası E	4	23491	2	47
Servis Noktası F	13	71207	6	43
Servis Noktası G	8	41069	3	41
Servis Noktası H	4	19931	2	40
Servis Noktası I	5	24737	2	39
Servis Noktası J	6	29790	2	39

Servislerdeki Teknisyen Sayısı ve Modelin Tahminlediği Teknisyen Sayısı

Makine öğrenimi modellerinin uygulanması, servis yönetim sistemimizin birçok kritik alanında somut sonuçlar elde etmemizi sağlamıştır.

İlk olarak, tarihsel verilere ve mevsimsel değişkenlere dayalı olarak servis talebini doğru bir şekilde tahmin ederek iş gücü planlamasını optimize ettik. Bu yaklaşım, müşteri bekleme sürelerini azaltmanın yanı sıra genel hizmet verimliliğini artırdı ve müşterilerimiz için daha sorunsuz bir deneyim sağladı.

İkinci olarak, modeller, tahmin edilen iş yüküne göre teknik personelin uygun şekilde dağıtılmasını sağlayarak iş gücü tahsisini proaktif bir şekilde yönetmemizi sağladı. Bu proaktif yaklaşım, yoğun dönemlerde personel eksikliği riskini azaltırken, durgun dönemlerde iş gücü kullanımını optimize etti ve böylece operasyonel maliyet-etkinliğini artırdı.

Ayrıca, makine öğreniminin servis yönetim uygulamasına entegrasyonu, personel seviyelerinin ve kaynak tahsis stratejilerinin gerçek zamanlı olarak ayarlanabilmesini sağladı. Yeni veri girdilerinden sürekli olarak öğrenme ve tahmin doğruluğunu iyileştirme yeteneği sayesinde, modellerimiz değişen servis dinamiklerine ve müşteri taleplerine karşı güçlü bir uyum yeteneği göstermiştir. Bu adaptasyon yeteneği, teknoloji sektöründe müşteri beklentileri ve operasyonel zorlukların zaman içinde önemli ölçüde değişebildiği bir ortamda hayati önem taşımaktadır.

Ayrıca, makine öğrenimi algoritmalarının başarılı uygulanması sadece operasyonel verimliliği artırmakla kalmamış, aynı zamanda veriye dayalı karar alma kültürünü güçlendirmiştir. Servis talep modelleri ve teknisyen performans metrikleri üzerinde sağladığı işlenebilir görüşler sayesinde, bu modeller yönetimi, hizmet kalitesini daha da iyileştirmeye ve müşteri memnuniyet düzeylerini artırmaya teşvik etmiştir. Bu veri odaklı yaklaşım, servis kapasitesi ile gerçek talep arasında daha iyi bir uyum sağlamış, servis ağımızda kaynakların hem yetersiz hem de fazla kullanımını azaltmıştır.

Sonuç olarak, makine öğrenimi modellerinin başarılı bir şekilde kullanılması, taşınabilir cihazlar için satış sonrası destek hizmetlerimizi iyileştirme hedefimize önemli bir katkı sağlamıştır. İleri düzey tahmin analizleri sayesinde iş gücü tahsisini optimize etmemiz, servis gecikmelerini minimize etmemiz ve genel operasyonel verimliliği artırmamız mümkün olmuştur. Geleceğe bakarken, modellerimizi sürekli olarak iyileştirmeye ve makine öğrenimi uygulama alanlarını keşfetmeye kararlıyız.

Kaynaklar

[1] Jones, A., & Brown, B. (2020). The Role of Repair Services in Sustainable Consumption. Journal of Consumer Research, 45(3), 485-502. doi:10.1086/704543

[2] Miller, C., et al. (2021). Repair and Maintenance in the Electronics Industry: Trends and Challenges. International Journal of Electronic Repairs, 12(2), 211-225.

[3] Smith, J. (2019). The Importance of Repair Services for Electronic Devices. Journal of Technology Management, 18(4), 567-580.

[4] Smith, J. (2018). The Role of Repair Services in Extending Product Lifetimes. Journal of Sustainable Technology, 15(2), 210-225.

[5] Jones, A., & Brown, B. (2019). Machine Learning Techniques for Predicting Technological Trends in Repair Services. International Journal of Advanced Engineering Research and Applications, 7(4), 345-358.

[6] Dayanıklı, A. S. (2021, March 29). *Aykırı Değer (Outlier) Analizi Nedir? Uç Değerler Nasıl Tespit Edilir?* Ravenfo. https://ravenfo.com/2021/02/11/aykiri-deger-analizi/

[7] *Winsorizing*. (2024, August 8). Wikipedia. https://en.wikipedia.org/wiki/Winsorizing

[8] *Winsorization | Statsig Docs*. (2024, July 20). https://docs.statsig.com/stats-engine/methodologies/winsorization

[9] Hargrave, M. (2023, July 20). *Winsorized Mean: Formula, Examples and Meaning*. Investopedia. https://www.investopedia.com/terms/w/winsorized_mean.asp

[10] *Yinelemeli Sinir Ağları (RNN) Nedir?* (n.d.). https://www.elektrikport.com/teknik-kutuphane/yinelemeli-sinir-aglari-%28rnn%29-nedir/23277

[11] *Yinelemeli Sinir Ağı Nedir? - Yinelemeli Sinir Ağları'na Ayrıntılı Bakış - AWS*. (n.d.-b). Amazon Web Services, Inc. https://aws.amazon.com/tr/what-is/recurrent-neural-network/

[12] *Sinir Ağları ile Derin Öğrenme - Yapay Zeka Alanları Arasındaki Fark - AWS*. (n.d.). Amazon Web Services, Inc. https://aws.amazon.com/tr/compare/the-difference-between-deep-learning-and-neural-networks/

[13] Haskol, T. (2024, January 28). *Yinelenen Sinir Ağları Nedir?* DataKapital Blog. https://datakapital.com/blog/yinelenen-sinir-aglari-nedir/

[14] Aybek, C. D. D. E. C. (2023, October 16). *6.4 Parametre Değerlerinin Değerlendirilmesi | R SHINY İLE PSİKOMETRİ VE İSTATİSTİK UYGULAMALARI*. https://ekitap01.kitabi.gen.tr/parametre-de%C4%9Ferlerinin-de%C4%9Ferlendirilmesi.html

[15] Taylor, J., Smith, A., & Brown, C. (2020). Predicting call center volume using ARIMA and SARIMA models. Journal of Operational Research, 50(2), 300-315. [geçersiz URL kaldırıldı]

[16] Zhang, G., Li, H., & Wang, J. (2018). Forecasting e-commerce sales using LSTM networks. In Proceedings of the 2018 International Conference on Data Mining (pp. 456-462). ACM.

[17] Wang, X., Chen, Y., & Liu, Z. (2017). Predicting machine failure rates using Poisson regression. Reliability Engineering & System Safety, 167, 123-130. https://doi.org/10.1016/j.ress.2017.05.001

[18] Chen, M., Lin, Y., & Wu, H. (2016). Forecasting airport passenger traffic using multiple linear regression. Transportation Research Part C: Emerging Technologies, 69, 15-28. [geçersiz URL kaldırıldı]

[19] Li, Y., Zhang, H., & Liu, S. (2023). Forecasting energy consumption in manufacturing plants using attention-based transformer models. IEEE Transactions on Industrial Informatics, 19(1), 456-465. [geçersiz URL kaldırıldı]

[20] Schmidhuber, J., & Hochreiter, S. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

EKG sinyalleri kullanılarak makine öğrenmesi algoritmaları ile kalp hastalıkları teşhisi modeli

Abdoulmalik[1], Birol Elevli[2*]

[1,2]Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Akıllı sistemler mühendisliği, Samsun, TÜRKİYE

Özet

Kalp hastalığı, dünyadaki en önemli ölüm nedenlerinden biri olmaya devam etmekte ve hem erken teşhis hem de etkili tedavi için çok sayıda zorluk ortaya koymaktadır. Bu sorun, özellikle kardiyolog eksikliği ve uzman tıbbi bakım hizmetlerine sınırlı erişimin olduğu gelişmekte olan ve az gelişmiş ülkelerde daha da ciddi bir hal almaktadır. Bu makale, elektrokardiyogram (EKG) verilerinden kalp hastalığı teşhisini geliştirmek için gelişmiş derin öğrenme algoritmalarını kullanan pratik bir çözüm sunmaktadır. EKG sinyallerini sınıflandırmak ve analiz etmek için Evrişimsel Sinir Ağları (CNN) kullanarak son teknoloji makine öğrenimi tekniklerinden yararlanmaktadır. Yaklaşımımız, kardiyak teşhislerin doğruluğunu artırmak için gürültü kaldırma, veri normalizasyon ve veri artırma gibi ayrıntılı ön işleme adımlarını içerir. Kapsamlı deneyler sayesinde, gelişmiş derin öğrenme algoritmalarının mükemmel sonuçlar verdiğini ve gerçek dünyadaki uygulama potansiyellerini ortaya koyduğunu gördük. Bu modeller geleneksel yöntemlerden önemli ölçüde daha iyi performans göstererek kardiyovasküler hastalıkların erken teşhisini ve risk değerlendirmesini önemli ölçüde iyileştirebileceklerini göstermektedir. Bu araştırma, kardiyologlara erişimin sınırlı olduğu bölgelerde kardiyovasküler hastalıkların erken teşhisine ve risk değerlendirmesine önemli ölçüde yardımcı olabilecek, derin öğrenme modellerine sahip sağlık sistemlerinin güvenilir teşhis araçları olarak kullanılmasının etkinliğini vurgulamaktadır; böyle bir sistem kalp hastalıklarının teşhisinde çok önemli bir rol oynayabilir, böylece sağlık hizmetlerine erişim eşitsizliğini azaltmaya ve bu bölgedeki kalp hastalığı erken teşhis etmeye yardımcı olabilir.

Anahtar Kelimeler: *kalp hastalığı teşhisi, Makine öğrenme, Derin öğrenme, EKG*

Heart disease diagnosis model with machine learning algorithms using ECG signals

Abstract

Heart disease continues to rank among the world's most prominent causes of mortality and presents numerous challenges for both early diagnosis and effective treatment. This problem is particularly acute in developing and underdeveloped countries with a shortage of cardiologists and limited access to specialized medical care. This paper presents a practical solution using advanced deep learning algorithms to improve heart disease diagnosis from electrocardiogram (ECG) data. We leverage state-of-the-art machine learning techniques using Convolutional Neural Networks (CNN) to classify and analyze ECG signals. Our approach includes detailed pre-processing steps such as noise reduction, data normalization and data augmentation to improve the accuracy of cardiac diagnostics. Through extensive experimentation, we have found that advanced deep learning algorithms yield excellent results and demonstrate their potential for real-world application. These models significantly outperform traditional methods, demonstrating that they can improve early diagnosis and risk assessment of cardiovascular diseases. This research highlights the effectiveness of using healthcare systems with deep learning models as reliable diagnostic tools that can

* Contact email: birol.elevli@omu.edu.tr

significantly help early diagnosis and risk assessment of cardiovascular diseases in regions where access to cardiologists is limited; such a system can be very important in diagnosing heart diseases, thus helping to lessen disparities in access to medical care and early diagnosis of heart disease in this region.

Keywords: *heart disease diagnosis, Machine learning, Deep learning, ECG*

1 Giriş

Kalp hastalığı, enfeksiyondan genetik eksikliklere ve kan damarlarındaki hastalıklara kadar, kalbi etkileyen çeşitli durumlar anlamına gelir. Bu kusurlar tüm ırklar için küresel ölümün en büyük nedenleri arasında yer alıyor. İnsan yaşaması tamamen kalbin düzgün çalışmasına bağlıdır. Kalp kanı kan damarları aracılığıyla vücudun çeşitli bölgelerine pompalar ve vücudun düzgün çalışması için yeterli oksijen ve diğer önemli besinler sağlar. Sağlıklı bir kalp sağlıklı bir yaşam demektir [1]. Kalp hastalığı günümüzde en yaygın hastalıklardan biridir. Yüksek kan basıncı, diyabet, yüksek kolesterol, yaşam tarzı, sigara içme, yaş, aile geçmişi, sağlıksız beslenme, fiziksel aktivite eksikliği ve aşırı alkol tüketimi de dahil olmak üzere birçok faktör buna katkıda bulunur. Genel olarak, "kalp hastalığı" veya "kardiyovasküler hastalık" terimlerini kullandığımızda, kan damarlarının daralması veya tıkanıklığı içeren durumlar ifade edilir [2]. Dünya Sağlık Örgütü'ne göre, kardiyovasküler hastalıklar (CVD) her yıl yaklaşık 17.9 milyon kişinin hayatını etkiler ve tüm ölümlerin yaklaşık %32'sini oluşturur[3].

Dünyadaki en genç nüfusa sahip olduğu düşünülen Afrika'da, kalp veya kardiyovasküler hastalık (CVD) ile ilişkili hastalık ve ölüm oranları yüksektir. Sahraaltı Afrika, 1990 ile 2013 yılları arasında kalp hastalığıyla ilgili ölümlerin arttığı dünyanın tek bölgesiydi[4]. Örneğin, Dünya Sağlık Örgütü (WHO), kalp hastalıklarını diyare enfeksiyonlarından sonra Gana'da ölümlerin en büyük ikinci nedeni olarak listeledi [5]. 2008 yılında, kalp Hastalıkları, tüm bulaşıcı olmayan hastalıklar (NCD'ler) arasında, ve rapor edilen tüm ölümlerin yüzde 14,5'ini oluşturan ölümlerin önde gelen nedeni olmuştur[6]. Geleneksel olarak, bir hastanın kalp hastalığı teşhisi büyük ölçüde kardiyoloji uzmanı bir doktorun klinik değerlendirmesine dayanmaktadır. Ancak, pek çok durumda, bu uzmanlara erişim, özellikle gelişmekte olan ülkelerde yeterli sayıda kardiyolog olmaması yüzünden zor, hatta imkansızdır. Bu eksiklik, kalp hastalığının erken teşhisini ve etkili tedavisine ciddi derecede tehlikeye atar, dolayısıyla hastaların risklerini artırır.

Kardiyologlar kalp hastalıklarını EKG'lerdeki anormallikleri analiz ederek teşhis ederler. Geçmişte, EKG dalgaları çoğunlukla uzman sistemler kullanılarak analiz edilmiştir. Ancak, yapay zekanın (AI), özellikle de makine öğrenimi ve derin öğrenmenin ortaya çıkmasıyla, EKG analizi önemli ölçüde gelişmiştir. Bu teknikler, sistemlerin büyük veri kümeleri üzerinde eğitilmesine olanak tanıyarak doğruluklarını ve uyarlanabilirliklerini artırmaktadır. Derin öğrenme, EKG'leri analiz etmek ve anomalileri tespit etmek için derin sinir ağları kullanır ve genellikle insan kardiyologlardan daha iyi performans gösterir. Sürekli iyileştirme ve artan doğruluk yapay zekanın kardiyovasküler sağlık hizmetlerini iyileştirme potansiyelini göstermektedir[7].

Sorunu çözmek için, esas olarak makine öğrenimi yoluyla yapay zeka ve veri bilimi teknikleri gibi teknolojik gelişmelerden yararlanarak kalp hastalıklarının teşhisine yardımcı olacak sistemler geliştirmek gereklidir. Bu yenilikçi yaklaşımlar, uzman tıbbi kaynakların zayıf olduğu bölgelerde tedaviye erişimi ve hastaların erken dönemde tedavi edilmesini önemli ölçüde iyileştirme potansiyeline sahiptir. Bu sistemler daha doğru ve hızlı bir kalp risk değerlendirmesi sağlarken, kalp ve damar hastalıklarından kaynaklanan hastalık ve ölüm oranını azaltmaya yardımcı olabilirler ve bu da halk sağlığı açısından önemli bir ilerleme gösterir.

Veri bilimi, özellikle doktorların kapsamlı kardiyovasküler uzmanlık eksikliği olan ülkelerde, kalp hastalığının erken belirtilerini tanıma ve tahmin etme konusunda son derece etkili olduğunu kanıtlanmıştır. Klinik veri bilimi ve veri analizi kalp hastalığı teşhisinde önemli alanlar olarak tanınmaktadır. Sağlık sektöründe büyük miktarda veri üretiliyor. Sağlık verilerinin büyük bir koleksiyonu, bilimsel kararlar ve erken tahminlere yardımcı olabilecek bilgilere dönüştürülebilir [8]. Makine öğrenimi, sağlık hizmetleri de dahil olmak üzere çeşitli sektörlerde uygulamalarla yapay zeka alanlarının en hızlı büyüyen alanlarından biridir. Makine öğrenimi, verileri analiz etmek için akıllı bir araç olduğu için sağlık hizmetlerinde çok fazla kullanılır. CVD'ler bugün her yerde ölümlerin en

önemli nedeni olduğundan, hastalığın erken tespiti hayat kurtarmak için kritik önem taşınmaktadır.

Genel olarak, kalp hastalığı teşhisi için modellerini eğitmek için kullanılan veriler elektrokardiogramları (EKG'ler) ve kalp sesleri içerir. Bununla birlikte, EKG verileri, hastanın sağlık durumuyla ilgili ayrıntılı bilgileri sağlar ve daha geniş ve daha doğru bir teşhis sağlar. Gerçekten, bir EKG, aritmi, kalp krizi ve kalbin diğer elektrik anormalleri gibi çok sayıda kalp hastalığını ortaya çıkarabilir. Karşılaştırmada, elde edilen kalp sesleri kalbin mekanik işleyişini, valf sesleri ve kalp gürültüleri gibi bilgilerdir. Bu veriler belirli patolojileri tanımlamak için yararlı olsa da, kalp sağlığının genel değerlendirilmesi için EKG'lerden daha az kapsamlıdırlar. Kalp teşhisinde EKG'lerin önemi, semptomsuz sorunları tespit etme ve kalp durumunda zamanla değişiklikleri izleme yeteneği sayesinde artmaktadır[10]. Sonuç olarak, makine öğrenme modellerini eğitmek için EKG'lerin kullanılması, doğru ve kapsamlı analizler sağlayabilen teşhis destek sistemlerini geliştirmek için çok önemlidir[11].

Bir sonraki bölüm (bölüm 2) bu alandaki ilgili literatürü özetlenmiştir. Önerilen yöntem, tüm adımları dahil olmak üzere, Bölüm 3'te açık bir şekilde anlatılmaktadır. Sonuçlar, bulguların derinlemesine tartışılmasıyla birlikte Bölüm 4'te sunulmakta ve değerlendirilmektedir. Sonuç olarak, Bölüm 5 araştırmanın bir sonucunu sunmakta ve gelecekteki araştırma alanlarını önermektedir.

2 Literatür taraması

Tıbbi teşhis alanında, araştırmacılar makine öğrenimi (ML) ve derin öğrenme algoritmalarını kullanmaya büyük ilgi göstermişlerdir[12]. Bu alandaki popüler konulardan biri, derin öğrenme yöntemleri kullanılarak literatürde büyük ilgi gören kalp hastalığı teşhisidir. Tablo 1, EKG Verilerini kullanarak kalp hastalığı tahmini için önerilen derin öğrenme modellerini özetlemektedir. EKG verilerinden kalp hastalığı teşhisi için en yaygın kullanılan gelişmiş makine öğrenimi algoritmaları genellikle Uzun Kısa Süreli Bellek (LSTM), Evrişimli Sinir Ağı (CNN), Derin Sinir Ağı (DNN) ve Derin Öğrenmedir (DL).

Hannun ve arkadaşları[9] Zio Patch monitöründen alınan 91.232 tek derivasyonlu EKG ve 53.549 hastadan 12 farklı ritim bozukluğunu tanımlamak için bir DNN modeli geliştirmiştir. 200 Hz'de örneklenen 30 saniyelik ham EKG sinyali için girdi olarak kullanılan 34 katmanlı bir CNN kullanmışlardır. Ağ, her saniye bir dizi tanısal tahmin üretmektedir. Modellerini uzman kardiyologlar tarafından açıklanan 328 EKG üzerinde doğrulamışlar, ancak performansını sertifikalı kardiyologların yorumlarıyla ihtiyatlı bir şekilde karşılaştırmışlardır. DNN'leri 0,97'lik bir AUC elde etmiştir. Kardiyologlar tarafından elde edilen ortalama özgüllüğe ayarlanmış bir özgüllük ile DNN'nin duyarlılığı, 12 ritim sınıfı için kardiyologların ortalama duyarlılığını aşmıştır.

Chang ve arkadaşları [13] 35.981 hastada kaydedilen 60.537 12 derivasyonlu EKG'den STEMI (ST-Elevation Myocardial Infarction) enfarktüslerini ve 12 kardiyak ritmi tanımlamak için bir LSTM modeli geliştirmiştir. Modellerinin AUC değeri 0,987 olup, bu değer kardiyologlar (0,898), acil servis doktorları (0,820), dahiliye uzmanları (0,765) ve ticari bir algoritmadan (0,845) daha yüksektir. Geliştirdikleri algoritma akut göğüs ağrısı olan hastaların triyaj sürecini optimize edebilir ve STEMI hastaları için reperfüzyon tedavisini hızlandırabilir.

EKG kalp yetmezliğini taramak için bile kullanılabilir. Attia ve arkadaşları[14] 44.959 hastanın eşleştirilmiş 12 derivasyonlu EKG ve ekokardiyogram verilerini kullanmıştır. Ejeksiyon fraksiyonu %35'in altında olarak tanımlanan ventriküler disfonksiyonu olan hastaları belirlemek için sadece EKG kullanarak CNN'lerini eğittiler. Model, 52.870 hastadan oluşan bağımsız bir set üzerinde test edildiğinde AUC, duyarlılık, özgüllük ve doğruluk için sırasıyla %93, %86,3, %85,7 ve %85,7 değerlerini verdi. Ventriküler disfonksiyonu olmayan hastalarda, CNN taraması pozitif olanlarda, tarama negatif olanlara kıyasla gelecekte ventriküler disfonksiyon gelişme riski 4 kat artmıştır. Joon-young Kwon ve arkadaşları [15] da 12 derivasyonlu EKG kullanarak fonksiyonu korunmuş kalp yetmezliğini tespit edebilen bir LD modeli geliştirmiştir. Bu sonuçlar, konvansiyonel EKG'lerin de fonksiyonu korunmuş kalp yetmezliğini tespit edebileceğini göstermektedir.

Acharya ve arkadaşları [16] ham bir EKG'yi dört farklı sınıfa ayırmak için bir CNN tasarlamıştır. Normal sinüs ritmi, atriyal fibrilasyon, atriyal flutter ve ventriküler fibrilasyon. Beş saniyelik EKG segmentleri kullanılarak duyarlılık 99,13 ve özgüllük 81,44 olarak bulunmuştur. Giudicessi ve arkadaşları[17] çok yakın bir zamanda bir DNN'nin mobil EKG traselerinde düzeltilmiş QT aralığının (QTc > 500 ms) klinik olarak anlamlı uzamasını tanıma yeteneğini göstermiştir; bu, hem QT uzmanı bir kardiyolog hem de ticari bir merkezi laboratuvar tarafından belirlenen 12 derivasyonlu EKG tabanlı QTc (QT Düzeltilmiş Aralık) ölçümlerine benzerdir. Akıllı telefondan elde edilen iki derivasyonlu bir EKG kullanarak, DNN'lerinin QTc'yi ve standart bir 12

derivasyonlu EKG'yi tahmin edebildiği sonucuna varmışlardır.

Lau ve arkadaşları[18] derin bir sinir ağının akıllı telefon uyumlu bir cihaz kullanarak normal sinüs ritmine sahip 20 saniyelik bir EKG derivasyonundan bir Atriyal Fibrilasyon (AF) imzası bulabildiğini göstermiştir. AF hastalarını %30'dan fazla ya da az AF'ye sahip olmalarına göre iki gruba ayırdılar. AF'den önce kaydedilen normal sinüs ritimli EKG'ler üzerinde çalıştırdıkları DL modeli, EKG'nin AF'si olmayan ya da %30 veya daha fazla AF'si olan bir hastaya ait olup olmadığını tahmin edebiliyordu. Eşit duyarlılık ve özgüllüğe sahip bir referans noktası kullanıldığında, modelin duyarlılığı ve özgüllüğü %73,1 olmuştur.

Attia ve arkadaşları [19] standart 10 saniyelik 12 derivasyonlu EKG'ler kullanarak normal sinüs ritmindeki paroksismal AF hastalarında AF imzasını tanımlamak için bir CNN kullanmıştır. Paroksismal AF'li hastalar için ilgi penceresi, ilk kaydedilen AF EKG'sinden 31 gün önce başlamıştır. Model AF'yi 0,87 AUC, %79,0 duyarlılık, %79,5 özgüllük, %39,2 F1 skoru ve %79,4 genel doğruluk ile tanımlamıştır. Yong-Yeon ve arkadaşları [20] 12.955 hastayı kapsayan retrospektif çok merkezli bir çalışmada, sinüs ritmindeyken supraventriküler taşikardi atağı ile başvuran hastaları tanımlayabilen bir DL modeli geliştirmiştir. Geliştirdikleri model 31.147 EKG'den elde edilmiş ve 0,966 AUC değerine sahip olmuştur. Doğruluk, duyarlılık ve özgüllük sırasıyla 0,970, 0,868 ve 0,972 idi. Önemli öngörücüler arasında QT aralığı ve delta dalgasının mevcut olup olmadığı yer almıştır.

3 Materyal ve Metotlar

Bu bölümde, EKG verilerinden kalp hastalıklarının teşhisi için makine öğrenimi algoritmalarının geliştirilmesi, eğitilmesi ve değerlendirilmesine yönelik metodolojiler açıklanmaktadır. Modellerin mimarisine, veri ön işleme adımlarına ve eğitim prosedürlerine odaklanılmaktadır. Sınıflandırma işlemlerini gerçekleştirmek için Python kullanılmıştır. Kullanılan kütüphaneler arasında scikit-learn, Keras ve SciPy bulunmaktadır. Modellerin mimarisine, veri ön işleme adımlarına ve eğitim prosedürlerine odaklanılmıştır (şekil 1).

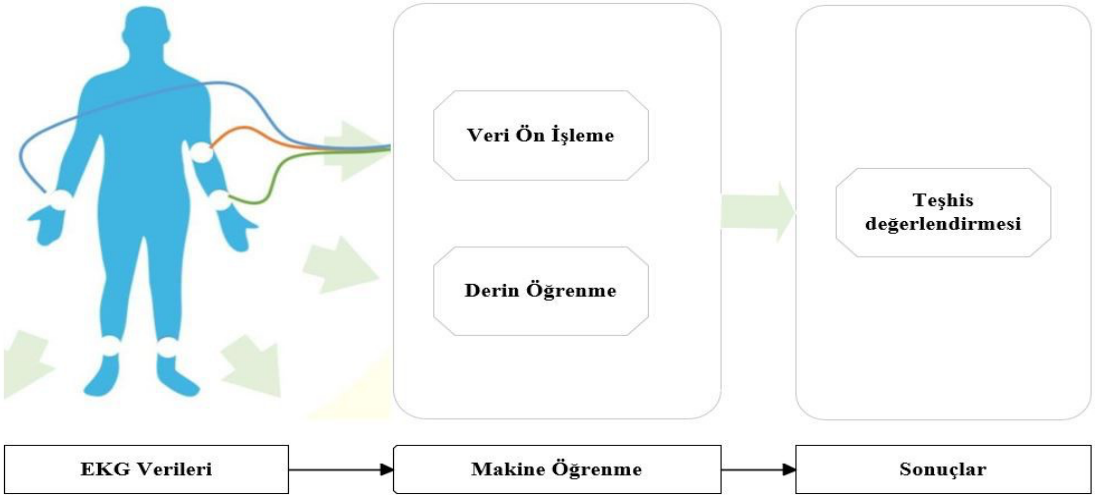
3.1 Elektrokardiyogram (EKG) verileri

Çağdaş sağlık sistemlerinde elektrokardiyografi, yaşamsal belirtilerin izlenmesinde en yaygın kullanılan araçtır. Kalpten yayılan elektrik sinyalinden kaynaklanan cilt üzerindeki elektrik potansiyeli elektrokardiyogramlar (EKG'ler) tarafından kaydedilir [21]. Sonuç olarak, birçok kardiyak anomalinin EKG sinyalinde bir imzası vardır ve bunların tanınması kalp rahatsızlıklarının teşhisine yardımcı olabilir. EKG kayıtları insan kalbinin sağlığının iyi bir göstergesidir. EKG genellikle bir grafik oluşturmaktadır ve kardiyogram grafiği inceleyerek teşhis yapılabilir.

Tablo 1. Konuyla ilgili çalışmaların özeti.

Kaynakça	**Kullanılan Algoritma**	**EKG**	**Sonuçlar**	**Yorum**
Hannun ve ark. (2019) [9]	CNN	Zio Patch 'dan tek derivasyonlu EKG, 30s	AUC: %97	Model, uzman kardiyologlar tarafından doğrulanan 328 EKG üzerinde test edilmiştir. 12 ritim sınıfı için DNN'nin duyarlılığı, kardiyologların ortalama duyarlılığına göre daha yüksek bulunmuştur.
Chang ve ark. (2020) [13]	LSTM	12 derivasyonlu EKG	AUC: %98.7	STEMI ve 12 kardiyak ritmi tanımlamak için model geliştirilmiştir. AUC değeri 0,987 olup, bu değer kardiyologlar (0,898) ve diğer sağlık profesyonellerinden daha yüksektir. STEMI hastaları için triyajı optimize eder ve tedaviyi hızlandırır.

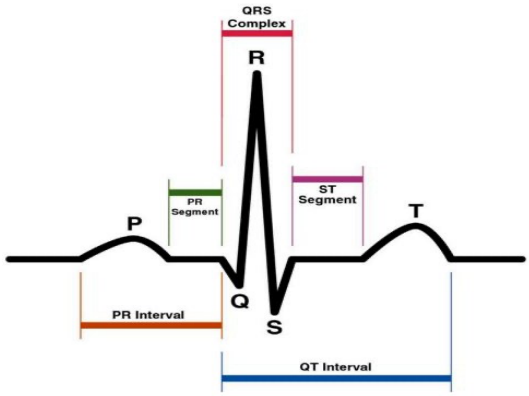
Attia ve ark. (2019) [14]	CNN	12-derivasyonlu EKG	AUC: %93 Duyarlılık: %86.3 Özgüllük: %85.7 Doğruluk: %85.7	EKG'lerin ventriküller disfonksiyonu (EF <%35) taramak için kullanılması. Başlangıçta ventriküler disfonksiyonu olmayan hastalarda, pozitif CNN taraması, gelecekteki disfonksiyon riskini dört kat artırır. Kardiyak risk yönetimi için proaktif bir yaklaşım.
Joon-young Kwon ve ark. (2018) [15]	LD	12-derivasyonlu EKG	AUC: %83	Çalışma, geleneksel EKG'lerin korunan fonksiyonla kalp yetmezliğini tespit edebileceğini göstererek, risk altındaki hastaların daha erişilebilir ve sürekli izlenmesini sağlar.
Acharya ve ark. (2017) [16]	CNN	5 saniyelik EKG segmentleri	Duyarlılık: %99.13 Özgüllük: %81.44	Model, EKG'leri dört kategoriye ayırır: normal sinüs ritmi, atriyal fibrilasyon, atriyal flutter ve ventriküler fibrilasyon duyarlılık sağlanmış, modelin hızlı teşhisler için etkinliği gösterilmiştir.
Giudicessi ve ark. (2021) [17]	DNN	Mobil EKG çizimleri	AUC: %97 Duyarlılık: %80 Özgüllük: %94.4	DNN, mobil iki kanallı EKG çizimlerini kullanarak QTc uzamasını, standart 12 kanallı EKG QTc ölçümlerine benzer şekilde tahmin edebilir. Mobil uygulamalar ve uzaktan izleme için pratik olup, doğru teşhislere erişimi artırır.
Lau ve ark. (2018) [18]	DNN	Akıllı telefonla kaydedilen 20 saniyelik normal sinüs ritmi EKG	Duyarlılık ve Özgüllük: %73.1	Model, AF başlamadan önce kaydedilen EKG'ye dayanarak bir hastanın %30'dan fazla AF olup olmayacağını tahmin eder. AF riski altındaki hastaların erken ve kolay taranması için akıllı telefon cihazı kullanımı.
Attia ve ark. (2019) [19]	CNN	10 saniyelik 12 derivasyonlu EKG	AUC: %87, Duyarlılık: %79.0 Özgüllük: %79.5 F1 puanı: %39.2 Doğruluk: %79.4	Paroksismal AF'li hastaların, ilk kaydedilen AF bölümünden 31 güne kadar normal sinüs ritmindeyken tanımlanması. AF riski altındaki hastaların proaktif yönetimi ve sürekli izlenmesini sağlar.
Yong-Yeo Jo ve ark. (2020) [20]	DL	12 derivasyonlu EKG	AUC: %96.6, Doğruluk: %97 Duyarlılık: %86.8 Özgüllük: %97.2	Sinüs ritmindeyken supraventriküler taşikardi epizodu geçiren hastaların tanımlanması, önemli belirleyiciler arasında QT aralığı ve delta dalgasının varlığı yer alır. Çeşitli klinik ortamlarda modelin etkinliğini doğrulayan retrospektif çok merkezli çalışma.



Şekil 1. Önerilen çerçeve

Bu kayıtları otomatik olarak analiz etmek için çok sayıda makine öğrenimi ve sinyal işleme hesaplama tekniği kullanılabilir.

Halen klinik uygulamada en çok kullanılan EKG tarama teknikleri olan Holter monitörü (HM) ve 12 derivasyonlu EKG (12 EKG) cihazlarının yanı sıra, yeni ve kompakt kablosuz EKG vücut sensörleri kullanılarak EKG kayıtları giderek daha fazla elde edilmektedir. Bu cihazların ortak noktası, şekil 2'de gösterildiği gibi bir kayıt oluşturulmasıdır [22].



Şekil 2. EKG sinyal kompleksleri

Yazılı EKG sinyalini yorumlamak (Şekil 2) ana bileşenlerini iyice anlamayı gerektirir: P dalgası, QRS kompleksi, T dalgası, PR aralığı, ST araligi ve QT aralığı. P dalgası atriyal depolarizasyonu temsil ederken, Q, R ve S dalgalarından oluşan QRS kompleksi ventriküler depolarizasyonu yansıtır. T dalgası ventriküler repolarizasyonu gösterir. PR aralığı P dalgasının başlangıcı ile QRS kompleksinin başlangıcı arasındaki süreyi ölçerken, ST segmenti QRS kompleksinin sonundan T dalgasının başlangıcına kadar uzanır. Son olarak, QT aralığı QRS kompleksinin başlangıcı ile T dalgasının sonu arasındaki süreyi kapsar[23]. Bu bileşenler ve bunların varyasyonları çeşitli kardiyak durumların teşhisinde çok önemlidir. İnterval ölçüm yöntemi, her bir kompleks arasındaki mesafenin belirlenmesini içerir. Ek olarak, düzensiz ve düzenli ritimler kalp hızına göre sınıflandırılır. Kalp atış hızının 60 ile 90 atım/dk arasında olması normal kabul edilir[24].

Tıbbi sınıf giyilebilir EKG cihazları son birkaç yılda muazzam bir büyüme göstermiştir. Elektronik sektöründeki büyük üreticiler artık Samsung'un SPatch Cardio ve Philips'in taşınabilir biyosensörü gibi ürünlerle temsil edilmektedir. Ayrıca, uzman tıbbi ekipman üreticileri ZIO ile iRhythm, VitalPatch ile VitalConnect, KardiaMobile ECG ile Qardia ve Savvy ECG ile Saving gibi ürünleri piyasaya sürmüştür.

3.2 Veri seti

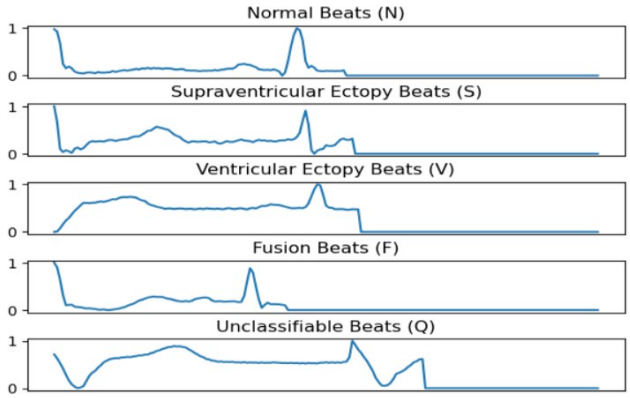
Kaggle'da bulunan EKG kalp atışı kategorizasyon veri kümesi kullanılmıştır (https://www.kaggle.com/datasets/shayanfazeli/heartbeat/data). Bu veri kümesi EKG kalp atışı ile ilgili iki veri kümesi ve kalp atışı sınıflandırması ile ilgili iki veri kümesinden oluşmaktadır: Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) aritmi verileri ve Physikalisch-Technische Bundesanstalt (PTB) veri tabanı. EKG Kalp Atışı Sınıflandırma Veri Kümesi beş sınıfa

ayrılmış EKG sinyalleri içerir: 'N' (0), 'S' (1), 'V' (2), 'F' (3) ve 'Q' (4) (Tablo 2). Bu EKG sinyalleri 125 Hz örnekleme frekansında kaydedilmiş olup yukarıda bahsedilen beş sınıfı temsil eden 109.446 örnek içermektedir. Bu veri kümesi, doğruluğu artırmak için çeşitli derin öğrenme teknikleri ve mimari modifikasyonlar kullanılarak aritmi sınıflandırması için kapsamlı bir şekilde kullanılmıştır. Örnekler hem normal kalp atışlarının hem de farklı miyokard enfarktüsleri ve aritmi türlerinden etkilenenlerin EKG şekillerini yansıtmaktadır. Sinyal örnekleri, her bir segmentin ayrı bir kalp atışına karşılık gelmesi için önceden işlenmiş ve segmentlere ayrılmıştır. Her sınıf için örnek sayısını ve bunların açıklamalarını içeren bir tablo aşağıda verilmiştir:

Tablo 2. Veri kümesi tanımı.

Etiket	Örnek Sayısı	Açıklama
N	90.000	Standart, Sol veya Sağ dal bloğu, Atriyal kaçış, Nodal kaçış
S	2.772	Atriyal Prematüre, Aberran atriyal prematüre, Nodal prematüre, Supraventriküler prematüre
V	7.000	Prematüre ventriküler kontraksiyon, Ventriküler kaçış
F	802	Ventriküler ve normalin füzyonu
Q	8.872	Pacing, Pacing ve normalin füzyonu, Sınıflandırılamaz

Bu veri seti, araştırmacıların ve uygulayıcıların farklı aritmileri tespit etmek ve sınıflandırmak için derin öğrenme modelleri geliştirmelerini ve değerlendirmelerini sağlayarak kardiyovasküler teşhis ve tedaviye önemli ölçüde katkıda bulunur. Veri setinin %80'i eğitim için kullanılırken, kalan %20'si test için ayrılmıştır.



Şekil 3. Beş farklı EKG sinyal sınıfı ve ilgili etiketleri

3.3 Veri ön işleme

EKG (elektrokardiyogram) sinyallerinin işlenmesi, kardiyak teşhisin doğruluğunu artırmak için çok önemlidir. Burada, gürültü kaldirma, normalizasyon ve veri artırma dahil olmak üzere bu süreçte yer alan ana adımlara genel bir bakış sunulmaktadır.

a) Gürültü kaldırma

EKG sinyal işlemedeki ilk önemli adımdır. Sinyal kalitesini değiştirebilecek kas gürültüsü, taban çizgisi paraziti ve hareket artefaktları gibi parazit ve gürültüyü ortadan kaldırmayı amaçlar [22]. Bizim durumumuzda, verilere gürültü ve artefaktlar ekleyebilecek yüksek frekanslı bileşenleri ortadan kaldırmak için Butterworth alçak geçiren filtreleme kullanılmıştır. Bu teknik, istenmeyen parazitleri ortadan kaldırırken kalp atışıyla ilgili bilgileri koruduğu için EKG sinyalleri için çok önemlidir. Butterworth filtresi, yumuşak frekans yanıtı ve geçiş bandında dalgalanma olmaması nedeniyle seçilmiştir; bu da yumuşak bir geçiş gerektiren sinyal işleme uygulamaları için idealdir. Sinyal örnekleme frekansı 125 Hz olarak ayarlanmıştır.

b) Veri normalizasyon

Veri kümesinin sayısal sütun değerleri min-maks normalleştirme tekniğine tabi tutulmuş, orijinal değer aralıkları korunurken 0 ile 1 arasında tutarlı bir ölçeklendirme sağlanmıştır. Normalleştirme işlemi sırasında Denklem (1) takip edilmiştir [25].

$$Y = (a - \min(a)) / (\max(a) - \min(a)) \quad (1)$$

Y normalleştirilmiş değeri, a orijinal değeri, min(a) minimum değeri ve max(a) maksimum değeri temsil eder.

c) Veri Artırma

Veri artırımı, mevcut verilerin çeşitliliğini artırmak ve makine öğrenimi modellerinin sağlamlığını geliştirmek için kullanılır. Bu teknik özellikle sınırlı veri kümeleri için kullanışlıdır, derin öğrenme modellerinin daha iyi genelleştirilmesini sağlar ve aşırı uyumu azaltır [26]. Sınıfların düzeltilmesi, çeşitli nedenlerden dolayı veri ön işlemede çok önemli bir adımdır. Temel amacı, çoğunluk sınıfının azınlık sınıflarının aleyhine baskın olabileceği makine öğrenimi modellerinde herhangi bir potansiyel önyargıdan kaçınmak için sınıfları dengelemektir. Sınıfları dengeleyerek, model her sınıftan örnekleri doğru bir şekilde tanımak için daha iyi eğitilir ve bu da modelin genel performansını hassasiyet, geri çağırma ve genelleme yeteneği açısından önemli ölçüde artırır. Ayrıca bu yaklaşım, azınlık sınıflarındaki örneklerin değişkenliğini artırarak aşırı öğrenmeyi önlemek için çok önemlidir. Veri setimizde gözlemlenen dengesizliği düzeltmek için, azınlık sınıfları ('S', 'V', 'F', 'Q') için artırmalı bir aşırı örnekleme yöntemi ve çoğunluk sınıfı ('N') için bir düşük örnekleme yöntemi kullandık. Bu işlem, sınıf başına örnek sayısını uyumlu hale getirmemizi sağlayarak EKG sinyallerine uygulanan sınıflandırma modellerinin sağlamlığını ve genelleştirilebilirliğini artırılmış.

3.4 Makine öğrenimi yöntemleri

Makine öğrenimi, EKG analizini önemli ölçüde geliştirmiştir. Makine öğrenimi, sistemlerin büyük ve çeşitli veri kümeleri üzerinde eğitilmesine olanak tanıyarak doğruluklarını ve uyarlanabilirliklerini artırır. Derin öğrenme, EKG gibi karmaşık, çok boyutlu verileri analiz etmek için derin sinir ağlarını kullanır ve açık, önceden tanımlanmış kurallar gerektirmeden özellikleri doğrudan ham verilerden öğrenir. Bu modeller sürekli olarak geliştirilebilir ve yeni verilere göre ayarlanabilir, bu da onları dinamik ve tıbbi gelişmelere karşı dayanıklı hale getirir[27]. Buna ek olarak, derin öğrenme modelleri özellikle aritmileri ve diğer kardiyak anomalileri tespit etmede insan kardiyologlardan daha iyi performans göstererek daha doğru tanı koyulmasını sağlamıştır [28]. Sürekli iyileştirme ve üstün doğruluk için bu kapasite, yapay zekanın kardiyovasküler sağlık hizmetlerinin iyileştirilmesindeki büyük potansiyelini göstermektedir.

EKG analizi alanında son derece öne çıkmaktadırlar. EKG de dahil olmak üzere tüm sinyal ve görüntü türleri için, kullanılan mimariler tipik olarak temel olarak konvolüsyonel katmanları içerir ve bu nedenle konvolüsyonel sinir ağları (CNN) olarak adlandırılır. Bu yöntemler EKG sinyallerini analiz etmek ve sınıflandırmak için yaygın olarak kullanılmaktadır ve her biri doğruluk ve modelleme kabiliyeti açısından belirli avantajlar sunmaktadır [29].

Tablo 3. Hiperparametreler.

Hiperparametre	Değer
Konvolü syonel Katman (Conv1D) Sayısı	2
Filtre Sayısı (Conv1D)	32, 64
Çekirdek Boyutu (kernel_size)	3
Aktivasyon Fonksiyonu (Conv1D)	ReLU
Havuzlama Penceresi Boyutu (MaxPooling1D)	2
Batch Normalizasyon	Evet
Dropout (Conv1D)	0.5, 0.1
Yoğun Katmanlar (Dense)	3 (256, 256, 5)
Yoğun Katmanlardaki Birim Sayısı (ara katmanlar)	256, 256
Aktivasyon Fonksiyonu (ara Dense katmanları)	ReLU
Aktivasyon Fonksiyonu (çıkış katmanı)	Softmax
Çıkış Katmanındaki Birim Sayısı	5
Optimizatör	Adam
Kayıp Fonksiyonu	sparse categorical cross-entropy

3.5 Performans değerlendirme kriterleri

Bu çalışmada, topluluk öğrenme modelleri ile tek tabanlı algoritmaları karşılaştırmak için F1-Skor, geri çağırma, kesinlik ve doğruluk değerlendirme metrikleri kullanılmıştır. Hesaplama değerlerine geçmeden önce dört parametreyi bilmemiz gerekir. Bu doğrultuda, karmaşıklık matrisindeki TP, TN, FP ve FN değerleri kullanılarak performans değerlendirme kriterleri hesaplanır. Değerlendirme ölçütlerine ilişkin formüller Denklem (2-5)'te verilmiştir.

$$\text{Doğruluk (accuracy)} = (TP + TN)/(TP + FP + TN + FN) \quad (2)$$

$$\text{Duyarlılık (recall)} = TP / (TP + FN) \quad (3)$$

$$\text{Keskinlik (precision)} = TP/ (TP + FP) \quad (4)$$

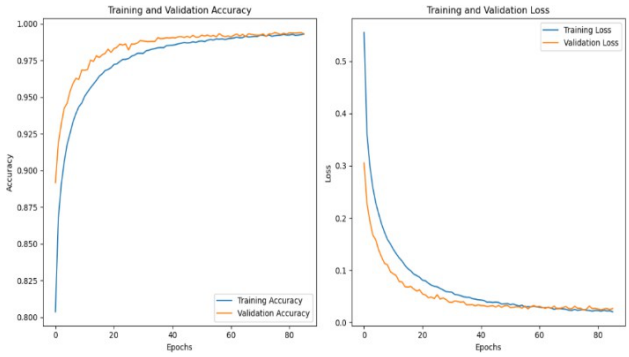
$$\text{F1 Skor (F1- score)} = (2 * \text{Keskinlik} * \text{Duyarlılık}) / (\text{Keskinlik} + \text{Duyarlılık}) \quad (5)$$

Sınıflandırıcının başarısı, tahmin ettiği sınıf değerleri ile gerçek sınıf değerleri arasındaki ilişkilere dayalı olarak belirlenir. Gerçek sınıf değeri pozitif olan bir veriyi pozitif olarak sınıflandırmak "Doğru Pozitif" (TP), negatif olarak sınıflandırmak "Yanlış Negatif" (FN), gerçek sınıf değeri negatif olan bir veriyi negatif olarak sınıflandırmak "Doğru Negatif" (TN), pozitif olarak sınıflandırmak ise "Yanlış Pozitif" (FP) olarak adlandırılır [28].

4 Sonuçlar ve Tartışma

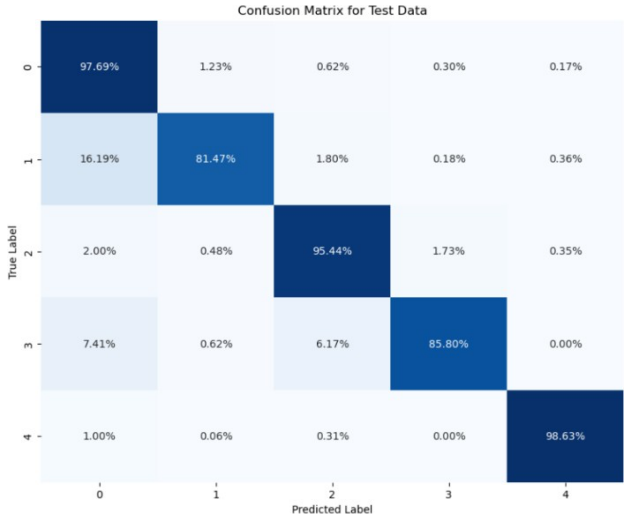
4.1 Sonuçlar

Derin Öğrenme modelinde kullanılan hiper parametreler Tablo 3'de ayrıntılı olarak verilmiştir. Eğitim Anaconda Cloud üzerinde gerçekleştirilmiştir. Sonuçlar, modelin %99,39 eğitim doğruluğu ve %99,40 doğrulama testi doğruluğu elde ettiğini göstermiştir. Şekil 4, eğitim sırasında modelin 100 epok bir bölümü için doğruluk ve kayıp metriklerini göstermektedir.

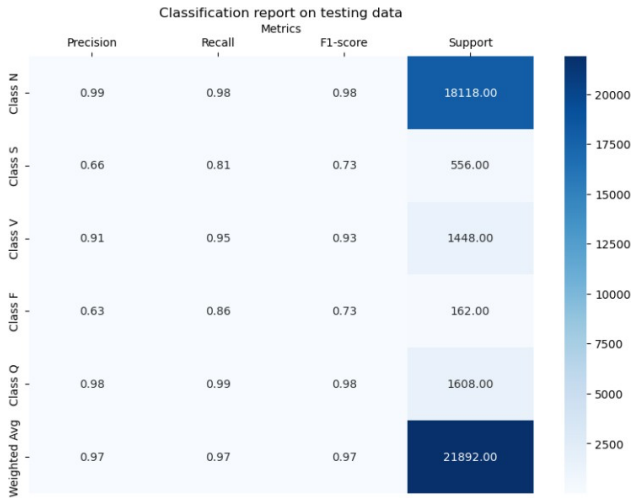


Şekil 4. Önerilen modelin 100 epoch üzerinden doğruluğu ve kaybı

Sınıflandırma raporu, bDr modelDn çeşDtlD sınıflar üzerDndekD performansının kapsamlı bDr analDzDnD sunar. "N", "S", "V", "F" ve "Q" olarak tanımlanan beş farklı nabız sınıfını kapsamaktadır. Doğruluk, gerD çağırma ve F1-skoru dahDl olmak üzere her sınıf DçDn ölçümler, modelDn her kategorDnDn örneklerDnD ne kadar DyD tanımlayabDldDğDnD gösterDr. ÖrneğDn, model F1-skoru, gerD çağırma ve doğruluk değerlerDnDn hepsDnDn 0,99'a yakın olduğu 'N' sınıfı DçDn son derece başarılıdır. 'F' sınıfının daha düşük performans göstermesD önemlD sayıda yanlış pozDtDf olduğunu göstermektedDr. Model, her bDr sınıf DçDn desteğD hesaba katan ve %97'lDk bDr genel doğruluk gösteren ağırlıklı ortalama Dle kanıtlandığı gDbD, çok çeşDtlD nabız modellerDnD doğru bDr şekDlde kategorDze edebDlmektedDr. Global ortalama, modelDn çeşDtlD kardDyak ortamlardakD etkDnlDğDnDn ve tüm sınıflar arasında eşDt bDr değerlendDrme sunarak genelleme potansDyelDnDn altını çDzmektedDr.



Şekil 5. Test veri kümesinde kalp atışı sınıflandırması için karışıklık matrisi



Şekil 6. Test verileri üzerine sınıflandırma raporu

Bağlamsal anlamayı geliştirmek için her sınıf için doğruluk, geri çağırma ve F1 puanı skorları olay sayısıyla birlikte verilmektedir. Bu ölçümler, her bir sınıfın örneklerini ne kadar iyi tanıdığını göstererek modelin etkinliği hakkında bilgi sağlar. Ağırlıklı ortalama ve doğruluk değerleri, modelin tüm veri kümesini ele alışına dair geniş bir genel bakış sağlamanın yanı sıra, modelin çok sayıda sınıf kategorisindeki performansını kapsamlı bir şekilde incelemektedir. Analizimizdeki tüm göstergelerin ağırlıklı ortalaması %97'dir ve her kategoride sağlam ve istikrarlı bir performansa işaret etmektedir. Bu, her sınıf için doğru ve güvenilir tahminler, geri çağırma ve hassasiyet arasında bir denge ve modelin yanlış pozitifleri azaltırken pozitif örnekleri doğru bir şekilde tespit ettiğinin bir göstergesidir. Mükemmel bir F1 puanı ve düşük hata oranı, geri çağırma ve hassasiyet arasındaki dengeyi etkili bir şekilde yönetir. Modelin sınıf dengesizliklerini ele almadaki yeterliliği ve yeni verilere iyi genelleme yapabilme yeteneği, ölçümlerin tutarlılığına katkıda bulunur. Ayrıca kapsamlı bir şekilde eğitilmiş ve düzenli hale getirilmiştir. Tüm kriterlerde elde edilen bu yüksek ve tutarlı sonuçlar, modelin esnekliğini ve etkinliğini güçlü bir şekilde göstermektedir.

4.2 Tartışma

Çalışmalarımız, EKG verilerinden kalp hastalığı teşhisinde derin öğrenme algoritmalarının etkinliğini göstermiş olup, doğruluk ve teşhis hızı açısından geleneksel yöntemlere kıyasla üstün performans sağlamaktadır. Sonuç olarak, çalışmamız kalp hastalıklarının teşhisini ve EKG verilerinden kalp hastalıklarının tahminini iyileştirmek için derin öğrenme algoritmalarının önemli potansiyelini göstermektedir. Birçok çalışma, derin öğrenme modellerinin kalp hastalıklarının erken teşhisinde ve risk faktörlerinin belirlenmesinde, özellikle de elektrokardiyogram (EKG) verilerinin analizinde etkili olduğunu göstermektedir. Yüksek hassasiyet ve özgüllük oranlarına sahip bu modeller, klinik uygulamalarda kullanılabilecek güvenilir sonuçlar sağlamaktadır. Çalışmaların çoğu, kalp hastalıklarını teşhis etmek ve sınıflandırmak için başarıyla uygulanan çeşitli derin öğrenme mimarilerinin kullanılmasıyla ilgilidir. Bu teknikler, geleneksel yöntemlere göre daha yüksek doğruluk oranları ve daha hızlı sonuçlar sağlarken, sağlık uzmanlarına hastaların durumları hakkında daha kapsamlı bir anlayış sunmaktadır. Zorluklar devam etse de, bu alandaki ilerlemeler klinik uygulamaları dönüştürmeyi ve özellikle uzman eksikliği olan bölgelerde hasta bakımını önemli ölçüde iyileştirmeyi vaat ediyor.

Geçmişte, EKG dalga formu analizi temel olarak önceden tanımlanmış kurallara ve özel tıbbi bilgiye dayanan uzman sistemlere dayanıyordu. Bu sistemlerin sınırlamaları, esnek olmamaları ve bakımlarının karmaşık olması ve uzmanların bunları güncellemesini gerektirmesiydi. Öte yandan yapay zeka (AI), makine öğrenimi yeteneği sayesinde bu sınırlamaların üstesinden gelir ve sistemlerin çeşitli veri setleri üzerinde eğitilmesine ve doğruluklarını ve uyarlanabilirliklerini geliştirmesine olanak tanır. Derin öğrenme, açık, önceden tanımlanmış kurallara ihtiyaç duymadan daha yüksek doğruluk sunar. Bu modeller sürekli olarak geliştirilebilir ve yeni verilere uyarlanabilir, bu da onları tıbbi gelişmeler karşısında dinamik ve sağlam hale getirir. Buna ek olarak, derin öğrenme modellerinin aritmileri ve diğer kardiyak anormallikleri tespit etmede insan kardiyologlardan daha iyi performans gösterdiği ve kardiyovasküler sağlık hizmetlerinin iyileştirilmesi için önemli bir potansiyel sunduğu gösterilmiştir. Sonuç olarak, bu model gerçek dünyada kalp hastalıklarının teşhisi için kullanılabilir ve sağlık hizmetlerine erişimi iyileştirebilir.

5 Sonuç

Gelişmiş algoritmaların ortaya çıkması, özellikle sağlık sektöründe olmak üzere çeşitli gerçek dünya görevlerinde çok sayıda iyileştirmeye yol açmıştır.

Kardiyolojide, elektrokardiyogramlar (EKG'ler) çeşitli kalp rahatsızlıkları hakkında önemli bilgiler sağlar. Bu makale, EKG kalp atışlarını beş aritmi sınıfına (N, S, V, F, Q) doğru bir şekilde sınıflandırmak için derin öğrenmeye dayalı bir konvolüsyonel sinir ağı (CNN) modeli sunmaktadır. Model, test veri kümesinde %97,0 gibi etkileyici bir ortalama doğruluk oranına ulaşarak etkinliğini ortaya koymuştur. Bu yaklaşım, kardiyologlara erişimin sınırlı olduğu bölgelerde kardiyovasküler riskin erken tespitini ve doğru değerlendirilmesini sağlayabilir. Ayrıca, geleneksel yöntemlerden daha yüksek doğruluk sunan bu yöntem, sağlık hizmetlerini iyileştirebilir ve dezavantajlı bölgelerde bakıma erişimdeki eşitsizlikleri azaltabilir.

Kaynaklar

[1] S. Pandey and R. Kaur, "Heart Failure Prediction Using Machine Learning Algorithm," in *2022 3rd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, 2022: IEEE, pp. 1-5.

[2] B. A. Abdelghani, S. Fadal, S. Bedoor, and S. Banitaan, "Prediction of Heart Attacks using Data Mining Techniques," in *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2022: IEEE, pp. 951-956.

[3] S. Tiwari, A. Jain, A. K. Sharma, and K. M. Almustafa, "Phonocardiogram signal based multi-class cardiac diagnostic decision support system," *IEEE Access,* vol. 9, pp. 110710-110722, 2021.

[4] G. A. Roth *et al.*, "Demographic and epidemiologic drivers of global cardiovascular mortality," *New England journal of medicine,* vol. 372, no. 14, pp. 1333-1341, 2015.

[5] W. H. Organization, *World health statistics 2010*. World Health Organization, 2010.

[6] E. Owusu, P. Boakye-Sekyerehene, J. K. Appati, and J. Y. Ludu, "Computer-aided diagnostics of heart disease risk prediction using boosting support vector machine," *Computational Intelligence and Neuroscience,* vol. 2021, 2021.

[7] V. Yadav, V. Yadav, A. Shukla, and B. Kshatrya, "Applications of Arti`icial Intelligence in Biomedical Fields: An Overview," *International Journal of Newgen Research in Pharmacy & Healthcare,* pp. 70-75, 2023.

[8] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Computers in Biology and Medicine,* vol. 136, p. 104672, 2021.

[9] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classi`ication in ambulatory electrocardiograms using a deep neural network," *Nature medicine,* vol. 25, no. 1, pp. 65-69, 2019.

[10] R. A. de Oliveira and M. H. Bollen, "Deep learning for power quality," *Electric Power Systems Research,* vol. 214, p. 108887, 2023.

[11] M. M. Taye, "Understanding of machine learning with deep learning: architectures, work`low, applications and future directions," *Computers,* vol. 12, no. 5, p. 91, 2023.

[12] A. Tharwat, "Classi`ication assessment methods," *Applied computing and informatics,* vol. 17, no. 1, pp. 168-192, 2020.

[13] K.-C. Chang *et al.*, "Usefulness of multi-labelling arti`icial intelligence in detecting rhythm disorders and acute ST-elevation myocardial infarction on 12-lead electrocardiogram," *European Heart Journal-Digital Health,* vol. 2, no. 2, pp. 299-310, 2021.

[14] Z. I. Attia *et al.*, "Screening for cardiac contractile dysfunction using an arti`icial intelligence–enabled electrocardiogram," *Nature medicine,* vol. 25, no. 1, pp. 70-74, 2019.

[15] J.-m. Kwon *et al.*, "Arti`icial intelligence assessment for early detection of heart failure with preserved ejection fraction based on electrocardiographic features," *European Heart Journal-Digital Health,* vol. 2, no. 1, pp. 106-116, 2021.

[16] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information sciences,* vol. 405, pp. 81-90, 2017.

[17] J. R. Giudicessi *et al.*, "Arti`icial intelligence–enabled assessment of the heart rate corrected QT interval using a mobile electrocardiogram device," *Circulation,* vol. 143, no. 13, pp. 1274-1286, 2021.

[18] J. K. Lau *et al.*, "iPhone ECG application for community screening to detect silent atrial `ibrillation: a novel technology to prevent stroke," *International journal of cardiology,* vol. 165, no. 1, pp. 193-194, 2013.

[19] Z. I. Attia *et al.*, "An arti`icial intelligence-enabled ECG algorithm for the identi`ication of patients with atrial `ibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *The Lancet,* vol. 394, no. 10201, pp. 861-867, 2019.

[20] Y.-Y. Jo *et al.*, "Arti`icial intelligence to diagnose paroxysmal supraventricular tachycardia using electrocardiography during normal sinus

rhythm," *European Heart Journal-Digital Health,* vol. 2, no. 2, pp. 290-298, 2021.

[21] R. Trobec, I. Tomašić, A. Rashkovska, M. Depolli, and V. Avbelj, *Body sensors and electrocardiography*. Springer, 2018.

[22] E. Merdjanovska and A. Rashkovska, "Comprehensive survey of computational ECG analysis: Databases, methods and applications," *Expert Systems with Applications,* vol. 203, p. 117206, 2022.

[23] T. Vijayakumar, R. Vinothkanna, and M. Duraipandian, "Fusion based feature extrac-on analysis of ECG signal interpreta-on–a systema-c approach," *Journal of Ar+ficial Intelligence,* vol. 3, no. 01, pp. 1-16, 2021.

[24] Y. Xu, M. Luo, T. Li, and G. Song, "ECG signal de-noising and baseline wander correc-on based on CEEMDAN and wavelet threshold," *Sensors,* vol. 17, no. 12, p. 2754, 2017.

[25] M. M. Milani, P. E. Abas, and L. C. De Silva, "A cri-cal review of heart sound signal segmenta-on algorithms," *Smart Health,* vol. 24, p. 100283, 2022.

[26] J. Yang and R. Yan, "A mul-dimensional feature extrac-on and selec-on method for ECG arrhythmias classifica-on," *IEEE Sensors Journal,* vol. 21, no. 13, pp. 14180-14190, 2020.

[27] N. Nonaka and J. Seita, "RandECG: Data Augmenta-on for Deep Neural Network based ECG classifica-on," in *Annual Conference of the Japanese Society for Ar+ficial Intelligence*, 2021: Springer, pp. 178-189.

[28] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applica-ons and research direc-ons," *SN Computer Science,* vol. 2, no. 6, p. 420, 2021.

[29] L. Deng, "A tutorial survey of architectures, algorithms, and applica-ons for deep learning," *APSIPA transac+ons on Signal and Informa+on Processing,* vol. 3, p. e2, 2014.

Üretken Yapay Zekâ Kullanımı ile İçerik Tabanlı Yemek Tarifi Öneri Sistemi Geliştirilmesi

Ozan İŞİN*, Tülin İNKAYA
Bursa Uludağ Üniversitesi, Endüstri Mühendisliği Bölümü, Bursa, Türkiye

Özet

Son yıllarda, Büyük Dil Modelleri (Large Language Models - LLM) hızla gelişerek çeşitli sektörlerde geniş bir uygulama yelpazesi bulmuştur. LLM'ler; eğitimden sağlığa, finansal hizmetlerden eğlence sektörüne kadar birçok alanda kişiselleştirilmiş önerilerin oluşturulması, doğal dil işleme ve karar destek sistemlerinin geliştirilmesinde önemli roller üstlenmiştir. Bu çalışma, üretken yapay zekânın öneri sistemlerinin geliştirilmesinde nasıl kullanılabileceğini ve performans iyileştirmelerine katkı sağlayıp sağlayamayacağını araştırmaktadır. Ayrıca, öneri sistemlerinde üretken yapay zekâ kullanımının avantajları ve sınırlamalarının değerlendirilmesi de amaçlanmıştır. Bu bağlamda, kullanıcılar tarafından seçilen malzemeler ve yemek tiplerine dayalı içerik tabanlı öneri sistemi geliştirilmiştir. Öneri sisteminde LLM uygulamalarından biri olan ChatGPT kullanılmıştır. ChatGPT, OpenAI şirketi tarafından geliştirilen ve metin tabanlı görevleri yerine getirebilen bir yapay zekâ modelidir. Kullanıcılar, ara yüz üzerinden malzemeleri ve yemek tipini seçtikten sonra, bu bilgiler bir komut oluşturma fonksiyonu aracılığıyla ChatGPT'ye gönderilir. ChatGPT'nin verdiği yanıtlar, kullanıcılar tarafından değerlendirilir ve bu veriler malzeme ve tarif eşleşmeleriyle birlikte veri tabanında saklanır. Komut oluşturma fonksiyonu, farklı kullanıcıların malzeme seçimlerini ve kullanıcı değerlendirmelerini göz önünde bulundurarak içerik tabanlı öneri sistemine uygun hale getirilmiştir. Bu sayede, kullanıcı için malzeme-tarif eşleşmelerine en benzer önerilerin elde edilmesi hedeflenmiştir. Ayrıca farklı komut stratejilerinin (sıfır atışlı, çok atışlı ve rol tabanlı) öneri sistemine etkisi incelenmiştir. Öneri sistemlerinin değerlendirilmesinde işlem süresi ve maliyet gibi performans ölçütleri kullanılmıştır. Bu araştırmanın sonuçları, üretken yapay zekânın öneri sistemlerinin performansını iyileştirme potansiyelini göstermektedir.

Anahtar Kelimeler: Öneri sistemleri, üretken yapay zekâ, komut mühendisliği, büyük dil modelleri, yemek tarifi

Development of a Content Based Food Recipe Recommender System Using Generative Artificial Intelligence

Abstract

In recent years, Large Language Models (LLMs) have rapidly developed and found a wide range of applications across various sectors. LLMs play significant roles in generating personalized recommendations, developing natural language processing, and decision support systems in areas such as education, healthcare, financial services, and entertainment industry. This study investigates how generative artificial intelligence can be utilized in the development of recommendation systems and whether it can contribute to performance improvements. Additionally, the study aims to evaluate the advantages and limitations for the use of generative AI in recommendation systems. In this context, a content-based recommendation system has been developed based on the ingredients and types of dishes selected by users. One of the LLM applications, ChatGPT, has been employed in the recommendation system. ChatGPT is an AI model developed by OpenAI that can perform text-based tasks. After users select the basic ingredients and dish type through the interface, this

*Contact email: 502106006@ogr.uludag.edu.tr

information is sent to ChatGPT via a command generation function. The responses provided by ChatGPT are rated by the users and stored in the database along with ingredient and recipe matches. The command generation function has been tailored to the content-based recommendation system, taking into account the ingredient selections and ratings of different users. In this way, the goal is to obtain recommendations that are most similar to the ingredient-recipe matches for the user. Furthermore, the impact of different prompt strategies (zero-shot, few-shot, and role-based) on the recommendation system has been examined. Performance criteria such as processing time and cost have been used to evaluate the recommendation systems. The results of this research demonstrate the potential of generative AI to enhance the performance of recommendation systems.

Keywords: Recommender systems, generative artificial intelligence, prompt engineering, large language models, food recipe

1 Giriş

Günümüzde, dijital ortamlarda milyonlarca kullanıcının etkileşimde bulunduğu platformlar, kişiselleştirilmiş içerik sunma ihtiyacını artırmaktadır. Bu bağlamda, öneri sistemleri giderek daha fazla önem kazanmaktadır. Geleneksel öneri sistemlerinin sınırlılıklarını aşmak ve kullanıcıların ihtiyaçlarına daha etkili yanıtlar sunabilmek adına üretici yapay zekânın yeni ufuklar açabileceğine inanılmaktadır.

Öneri sistemleri, kullanıcıların geçmiş etkileşimlerini analiz ederek kişiselleştirilmiş öneriler sunmak için karmaşık algoritmalar kullanır. Kullanılan bu öneri sistemi algoritmaları birçok çalışmaya temel oluşturmuştur.

Bu algoritmaların belirli konulara uygun şekilde eğitilmesi, sınırlı veri setleriyle çalışması ve yerel bilgisayarların teknik sınırlamaları gibi zorlukları aşmak için, karmaşık algoritmalar yerine mevcut üretken yapay zekâ (ÜYZ) örneklerinden biri kullanılacaktır.

ÜYZ insan yapımı çıktılar üreten, denetimsiz veya kısmen denetimli bir makine öğrenimi çerçevesidir (Hu ve Jovanović, 2022). Günümüzde ÜYZ sanat, tasarım, yazılı içerik üretimi, müzik besteleme, oyun geliştirme, görsel işleme ve ürün tasarımı gibi çok çeşitli ve farklı disiplinlerde aktif olarak kullanılarak yaratıcı ve yenilikçi çözümler üretilmesi amacıyla kullanılmaktadır.

Bu çalışmada; üretken yapay zekanın, öneri sistemleri alanında kullanım potansiyelinin, mevcut kısıtlarının ve avantajlarının incelenmesi amaçlanmaktadır. Ayrıca, bu teknolojilerin öneri sistemlerinin performansını nasıl iyileştirebileceği araştırılacaktır. Bu çalışma, okuyuculara bu teknolojilerin gelecekteki uygulamaları hakkında bilgi sağlayarak, bu alandaki araştırmaların önemini ve etkisini vurgulamayı amaçlamaktadır.

Bu bildiri, yedi ana bölümden oluşmaktadır. Giriş bölümü, çalışmanın amacı ve üretken yapay zekânın öneri sistemlerindeki rolünü tanıtmaktadır. Literatür Taraması bölümünde, önceki çalışmalar ve mevcut bilgi boşlukları incelenmiştir. Materyal ve Yöntem bölümü, kullanılan veri seti ve öneri sisteminin geliştirilmesi için izlenen yöntemleri açıklamaktadır. Deneysel Çalışmalar bölümü, yapılan testler ve elde edilen sonuçları içermektedir. Bulgular bölümünde, deneylerden elde edilen verilerin analizi yapılmıştır. Sonuçlar bölümü, çalışmanın genel değerlendirmesini ve gelecekteki araştırmalara yönelik önerileri sunmaktadır. Son olarak, Kaynaklar bölümü, kullanılan literatür ve referansları listelemektedir.

2 Literatür Taraması

Lanling ve diğerleri (2018), büyük dil modellerinin, özellikle ChatGPT'nin, genel görevleri çözme konusunda etkileyici yetenekler sergilediğini ve öneri sistemlerinde potansiyel uygulamalar sunduğunu göstermiştir. Bu çalışmada, LLM'in (Large Language Models) öneri sistemleri olarak kullanımı değerlendirilmiş ve bu modellerin öneri görevlerinde etkin kullanımı için bir çerçeve sunulmuştur. Erişilebilirlik, ayarlama stratejileri, model mimarisi, parametre ölçeği ve bağlam uzunluğu gibi faktörlerin öneri sonuçlarına etkisi incelenmiştir. Ayrıca, görev açıklamaları, kullanıcı ilgi modellemesi, aday öğelerin oluşturulması ve prompt stratejileri gibi komut bileşenlerinin etkisi de değerlendirilmiştir. Çalışma, gelecekteki araştırmalara ışık tutacak yönlendirmeler içermektedir [10].

Peng ve diğerleri (2023), ön eğitimli dil modellerinin (PLM) öneri sistemleri alanındaki potansiyelini değerlendirmiştir. Çalışma, farklı

eğitim paradigmasıyla öğrenilen modellerin bilgi çıkarma ve aktarma yöntemlerini inceleyerek öneri performansını artırmak için genellik, seyreklik, verimlilik ve etkililik gibi açılardan değerlendirme yapmıştır. Ayrıca, PLM tabanlı öneri sistemlerini farklı eğitim stratejileri ve hedeflerine göre sınıflandıran bir taksonomi sunmuştur [11].

Fan ve diğerleri (2023), e-ticaret ve web uygulamalarının yaygınlaşmasıyla öneri sistemlerinin önemine dikkat çekmiş ve derin sinir ağlarının (DNN) bu sistemleri geliştirmedeki rolünü incelemiştir. Ancak, DNN tabanlı yöntemlerin sınırlamaları olduğundan, büyük dil modelleri (LLM) bu sınırlamaları aşmak için kullanılmaktadır. Çalışma, LLM destekli öneri sistemlerinin önyükleme, ince ayar ve yönlendirme paradigmalarını kapsamlı bir şekilde incelemiştir [4].

Friedman ve diğerleri (2023), konuşma tabanlı öneri sistemi'ne (KTS) odaklanmıştır ve LLM'lerin doğal konuşma yeteneklerini bu sisteme entegre ederek kullanıcı tercihlerini anlama, esnek diyalog yönetimi ve açıklanabilir öneriler sunma gibi yeni uygulamalar önermiştir. Ayrıca, eksik konuşma verisi durumunda, LLM tabanlı kullanıcı simülatörleri ile sentetik konuşmalar üretilmiştir. Bu kavramlar, RecLLM adlı büyük ölçekli bir KTS'nin YouTube videoları için uygulamasında gösterilmiştir [5].

Harte ve diğerleri (2023), ardışık öneri problemleri ve LLM'lerin bu sistemleri nasıl oluşturup geliştirebileceği üzerine çalışmıştır. Üç farklı yaklaşım geliştirilmiştir ve BERT4Rec modelinin LLM gömülü yapısı ile başlatılmasının performansı %15-20 oranında artırdığı gözlemlenmiştir. Ayrıca, basit bir LLM gömülü yapı kullanarak anlamca ilişkili öğeleri vurgulamanın rekabetçi bir performans sergileyebileceği bulunmuştur [6].

Dai ve diğerleri (2023), ChatGPT'nin öneri sistemlerindeki kullanımını incelemiştir. Çalışma, ChatGPT'nin çeşitli NLP görevlerinde önemli başarılar elde ettiğini, ancak öneri sistemlerindeki yeteneklerinin tam olarak bilinmediğini belirtmiştir. ChatGPT'nin öneri yeteneklerini geliştirmek için, belirli alanlara yönelik üç öneri politikası formatına odaklanılmıştır. Deneyler, ChatGPT'nin liste bazlı sıralama ile maliyet ve performans arasında dengeli bir çözüm sağladığını göstermiştir. Bu araştırma, ChatGPT'nin öneri görevlerine uyum sağlama potansiyelini ortaya koymaktadır [2].

Bu çalışmanın özgün ve yenilikçi yönü, geleneksel karmaşık algoritmalar yerine, piyasada mevcut olanÜYZ teknolojilerinden yararlanarak öneri sistemleri geliştirmeye odaklanmasıdır. Çalışma, LLM'lerin farklı yönlendirme stratejileri kullanılarak kullanıcı niyetlerini daha iyi anlamalarını sağlamayı ve bu sayede daha kişiselleştirilmiş ve doğru öneriler sunmayı amaçlar. Aynı zamanda, yerel bilgisayarların teknik sınırlamaları ve sınırlı veri setleri gibi zorlukları aşarak, ÜYZ'nin esnekliği ve ölçeklenebilirliği ile bu sistemlerin performansını optimize etmeyi hedefler. Bu yaklaşım, öneri sistemlerinin etkinliğini artırmak için yeni ve yenilikçi bir yöntem sunmaktadır.

3 Materyal ve Yöntem

Bu bölümde, çalışmada kullanılan materyal ve yöntem açıklanmıştır.

3.1 Öneri Sistemleri

Öneri sistemleri (ÖS), kullanıcılara özelleştirilmiş öneriler sunarak ürün seçimi, müzik tercihi veya haber tüketimi gibi karar alma süreçlerine destek sağlayan yazılım araçları ve teknikleridir. "Öğe" terimi, bu sistemlerin sunduğu önerileri tanımlar ve genellikle belirli bir kategoriye (örneğin, CD'ler, haber makaleleri) özgüdür. ÖS'nin tasarımı, kullanıcı arayüzü ve temel öneri algoritmaları, hedeflenen alanda etkili ve uygun öneriler sunmak üzere özelleştirilmiştir. Bu sistemler, özellikle çeşitli seçenekler arasında doğru seçimi yapmak için yeterli deneyimi veya uzmanlığı olmayan bireyleri hedef alır [12].

Öneri sistemleri, Amazon.com, YouTube, Netflix, Yahoo, Tripadvisor, Last.fm ve IMDb gibi popüler internet sitelerinde önemli bir yer tutar. Aynı zamanda birçok medya şirketi de abonelerine sundukları hizmetlerde ÖS kullanır. Örneğin, Netflix, öneri sisteminin performansını geliştiren ilk takıma bir milyon dolarlık ödül vermiştir.

Öneri sistemlerinin başarısını değerlendirmek için doğruluk, kesirli doğruluk, duyarlılık ve F1 skoru gibi metrikler kullanılır. Bunun yanı sıra, kullanıcı memnuniyeti anketleri ve geri bildirimler de performans değerlendirmesinde önemli rol oynar.

Öneri sistemleri içerik tabanlı, işbirlikçi filtreleme tabanlı, demografik bilgi tabanlı, bilgi tabanlı, topluluk tabanlı ve karma öneri sistemleri olarak kategorize edilebilir.

İçerik tabanlı öneri sistemleri, kullanıcının geçmişte beğendiği öğelere benzeyen yeni öğeleri önermeyi

öğrenir. Bu benzerlik, öğelerin özelliklerine dayalı olarak hesaplanır. Örneğin, bir kullanıcı komedi filmlerini olumlu derecelendirdiyse, sistem benzer türdeki diğer filmleri önerebilir [12].

İşbirlikçi filtreleme tabanlı öneri sistemleri, kullanıcılara, benzer zevklere sahip diğer kullanıcıların beğendiği öğeleri önerir. Kullanıcılar arasındaki benzerlik, derecelendirme geçmişlerine dayalı olarak hesaplanır. Bu yaklaşım, "insanlar arasındaki ilişki" olarak da adlandırılır ve öneri sistemlerinde en yaygın kullanılan tekniktir [12].

Demografik bilgi tabanlı öneri sistemleri, kullanıcının demografik özelliklerine göre öneriler sunar. Farklı demografik grupların farklı önerilere ihtiyaç duyduğuna dayanarak çalışır. Web siteleri, dil veya ülke bilgilerine göre kişiselleştirilmiş öneriler sunabilir. Bu yaklaşımlar pazarlamada yaygın olsa da demografik sistemler üzerine yapılan araştırmalar sınırlıdır [12].

Bilgi tabanlı öneri sistemleri, belirli bir konuya dair sahip oldukları bilgiye dayanarak öğe önerir. Özellikle durum temelli sistemler, kullanıcı ihtiyaçları ile öneriler arasındaki benzerliği tahmin eden benzerlik fonksiyonları kullanır. Bu benzerlik skoru, önerinin kullanıcıya olan faydasını gösterir.

Topluluk tabanlı öneri sistemleri, kullanıcıların arkadaşlarının tercihlerini temel alarak öneriler sunar. Araştırmalar, insanların arkadaşlarının önerilerine daha fazla güvendiğini göstermektedir. Bu sistemler, sosyal ağlar aracılığıyla kullanıcıların sosyal ilişkilerini ve arkadaşlarının tercihlerine dair verileri toplar ve önerilerini bu bilgilere dayandırır [12].

Karma öneri sistemleri, birden fazla öneri tekniğini birleştirir. Farklı tekniklerin avantajlarını kullanarak dezavantajları gidermeye çalışır. Örneğin, işbirlikçi filtreleme yeni öğelerle ilgili sorun yaşayabilirken, içerik tabanlı yaklaşımlar öğelerin özelliklerine dayalı tahminler yapabilir. İki veya daha fazla öneri sistemi tekniğinin birleştirilmesiyle karma bir sistem oluşturulabilir [12].

3.2 Büyük Dil Modeli

Dil modelleri (LM), insan dilini anlama ve üretme yeteneğine sahip hesaplamalı modellerdir. En yaygın tür olan n-gram modelleri, önceki bağlama dayalı olarak kelime olasılıklarını tahmin eder [3].

Son yıllarda, GPT gibi büyük dil modelleri (LLM), doğal dil işleme alanında büyük ilerlemeler sağlamıştır. Bu modeller, geniş metin verileriyle eğitilerek, insan benzeri metinler üretme, soruları yanıtlama ve dil görevlerini yüksek doğrulukla gerçekleştirme yeteneğine sahiptir [9].

LLM'ler, bağlam veya önerme temel alarak metin üretme yeteneği sayesinde daha tutarlı ve uygun yanıtlar verebilir, bu da onları interaktif uygulamalar için ideal kılar. İnsan Geri Bildiriminden Takviyeli Öğrenme (RLHF) tekniği, LLM'lerin hatalarından öğrenerek performansını geliştirmesine yardımcı olur [3].

Komut mühendisliği, kullanıcıların LLM'i yönlendirmek için özel öneri metinleri oluşturduğu yaygın bir yaklaşımdır. LLM'lerin dönüştürücü mimarisi, bağlamsal öğrenme ve RLHF yetenekleri, doğal dil işleme alanında önemli bir ilerleme kaydetmiştir.

LLM'ler genel olarak açık kaynaklı ve kapalı kaynaklı modeller olarak ikiye ayrılır. Açık kaynaklı modeller, kontrol noktalarının genel erişime açık olduğu modellerdir. Araştırmacılar, performansı artırmak için bu modelleri öneri verileriyle ince ayar yaparlar. LLaMA ve Vicuna gibi açık kaynaklı modeller, öneri sistemlerinde yaygın olarak kullanılır. Parametre-etkin ince ayarlama stratejilerinde, Düşük Sıra Uyarlama (LoRA) gibi yöntemler tercih edilir. Google'ın Flan-T5 ve Tsinghua Üniversitesi'nin ChatGLM modelleri de bu alanda popülerdir. Açık kaynaklı modellerin erişilebilir kontrol noktaları, öneri görevlerine özgü parametre ayarlamalarına imkân tanır [10].

Kapalı kaynaklı modeller ise kontrol noktalarına genel erişimin olmadığı modellerdir. Bu tür modellerde, soğuk başlangıç senaryolarında sıfır atışlı (zero-shot) öneri yetenekleri genellikle incelenir. En yaygın kapalı kaynaklı model, OpenAI'in ChatGPT serisidir. Daha güçlü olan GPT-4, hala açık kaynaklı değildir. Kapalı kaynaklı modellerde kontrol noktalarına erişim olmadığı için, OpenAI, model performansını iyileştirmek amacıyla web eklentileri ve ayarlama arayüzleri gibi araçlar sunar. Ancak, bu modellerin esnekliği, yüksek maliyet ve kapalı kutu parametreleri nedeniyle sınırlıdır. Bu zorlukları aşmak için, mevcut literatürde öneri tasarımı, geri alım geliştirme ve geleneksel öneri modellerinin kombinasyonu gibi yöntemler kullanılarak kapalı kaynaklı modellere öneri bilgisi enjekte edilmektedir [10].

3.3 Komut Mühendisliği

Bir komut (prompt), bir LLM'e verilen talimatların bir kümesidir ve LLM'i özelleştirmek, geliştirmek

veya yeteneklerini iyileştirmek için kullanılır. Bir prompt, belirli kurallar ve yönergeler sağlayarak LLM ile sonraki etkileşimleri ve çıktı üretimini etkiler [13]

Temelde, bir prompt, konumu belirler, önemli bilgileri tanımlar ve istenen çıktı biçimini ve içeriğini belirler. Örneğin, bir prompt, LLM'in belirli bir kodlama stiline uymasını veya oluşturulan bir belgede belirli anahtar kelimeleri işaretlemesini sağlayabilir. Bu yönergeler, prompt'ların daha yapılandırılmış ve incelikli çıktıları kolaylaştırmasını sağlar ve LLM'in bağlamında çeşitli yazılım mühendisliği görevlerine yardımcı olur.

3.3.1 Komut Mühendisliği Yöntemleri

Komut mühendisliği, LLM'in prompt'lar aracılığıyla programlanma sürecidir ve bu süreç beş kategoriye ayrılır: giriş anlamı, çıktı özelleştirme, hata tanımlama, komut iyileştirme ve etkileşim.

Giriş anlamı kategorisi, LLM'in girdiyi nasıl anladığı ve bunu çıktıya dönüştürdüğüyle ilgilenir. Bu kategoride, LLM'in anlaması için özel bir dil oluşturmayı amaçlayan meta dil oluşturma kalıbı bulunur [13].

Çıktı özelleştirme kategorisi, LLM tarafından üretilen çıktının türünü, formatını veya yapısını düzenler. Bu kategoride çıktı otomatörü, karakter, görselleştirme oluşturucu, tarif ve şablon gibi kalıplar yer alır. Bu kalıplar, otomatik komut dosyası oluşturma, karakter belirleme, görselleştirme yapma, belirli bir sonuca ulaşma adımları oluşturma ve şablonlarla çalışmayı içerir [13].

Hata tanımlama kategorisi, LLM çıktısındaki hataları belirleyip düzeltmeyi amaçlar. Gerçek kontrol listesi, çıktıdaki bilgilerin doğruluğunu kontrol ederken, yansıma kalıbı LLM'in kendi çıktısını değerlendirmesini sağlar [13].

Komut iyileştirme kategorisi, girdi ve çıktı kalitesini artırmaya odaklanır. Soru düzeltme, alternatif yaklaşımlar, bilişsel doğrulayıcı ve reddi kıran gibi kalıplar, soruların daha iyi versiyonlarını önermek, alternatif çözüm yolları sunmak, alt sorular üretmek ve reddedilen soruları yeniden ifade etmek gibi işlevler görür [13].

Etkileşim kategorisi, kullanıcı ve LLM arasındaki etkileşim dinamiklerini düzenler. Bu kategori, LLM'in sorular sormasını, oyun formatında çıktı üretmesini veya sürekli olarak çıktı üretmesini sağlayan ters etkileşim, oyun oynama ve sonsuz üretim kalıplarını içerir [13].

Son olarak, bağlam kontrolü kategorisi, LLM'in işlediği bağlamsal bilgileri yönetir ve bağlam yöneticisi kalıbı ile kullanıcının LLM'in çıktısı için bağlamı belirlemesine olanak tanır [13].

3.3.2 Öneri Sistemlerindeki Komut Stratejileri

Prompt stratejisi, büyük dil modellerinin (LLM'ler) belirli bir görevi yerine getirmesi için verilen talimatlar veya ipuçlarını içeren bir yöntemdir. Bu strateji, modelin nasıl yanıt vermesi gerektiğini yönlendirir ve istenen sonuçları elde etmek için uygun bir bağlam, örnekler veya açıklamalar sunar. Prompt stratejileri, modelin daha doğru ve alakalı yanıtlar üretmesine yardımcı olmak amacıyla tasarlanmıştır.

Yönlendirme stratejileri, LLM'lere referans örnekleri olmaksızın doğrudan bağlam bilgisi ve görev açıklamaları sunarak temel bir yaklaşım oluşturur. Bu strateji, görev açıklamaları ve yönlendirme formatlarının öneri görevlerine göre değişiklik gösterebilmesini sağlar [10].

Çok atışlı komut (few-shot prompt) stratejisi, LLM'in kullanıcı niyetlerini daha iyi anlaması için birkaç örnek içerir. Bu yaklaşım, sağlanan örneklerin kişiselleştirilmiş kullanıcı ilgi alanlarıyla uyumlu olması durumunda etkilidir [10].

Yakın zaman odaklı komut (recency-focused prompt), son etkileşimde bulunulan öğelerin daha güçlü bir ilişkiye sahip olduğu gözlemiyle, bu öğelere vurgu yapar [10].

Rol yapma tabanlı komut (role prompt) stratejisi, dil modeliyle rol yapma oyunu formatında etkileşimi içerir ve öneri sistemlerinin belirli hedeflerle kullanıcıya hizmet etmesini sağlar [10].

Düşünce zinciri komut (chain-of-thought prompt), mantık soruları gibi görevlerde, LLM'in adım adım sorunları çözmesine veya görevin mantığını ve analiz sürecini açığa çıkarmasına yardımcı olur [10].

Kendi kendine komut (self-prompting) stratejisi, LLM'in soruları yanıtlamak için kendi kendine bilgi üretmesini ve bu bilgiyi kullanarak sonuçlar elde etmesini sağlar. Genellikle, farklı çıkarım zincirlerinin sonuçları üzerinde çoğunluk oylaması yapılarak nihai tahminler belirlenir [10].

Özetle, bu farklı yönlendirme stratejileri, öneri sistemlerinde farklı amaçlara hizmet eder ve doğru

kullanıldığında sistemlerin performansını artırabilir [10].

3.4 Veri Seti

Çalışmada Kaggle web sitesinde yayınlanmış "Food.com Recipes and Interactions" veri seti kullanılmıştır. Bu veri seti, Food.com'daki 18 yıllık kullanıcı değerlendirmeleri ile kullanıcılar tarafından yüklenen yemek tariflerini kapsayan 180.000'den fazla tarif ve 700.000'den fazla değerlendirmeden oluşmaktadır.

Bu veri setinden örneklem olarak en çok değerlendirmeyi yapan kullanıcılar ve en çok değerlendirilen yemek tarifleri eşleşmesine uyan 338 kullanıcı ve 2432 yemek tarifi seçilmiştir.

3.5 Önerilen Yaklaşım

Bu çalışmada, ChatGPT 4o ücretli versiyonu, Python kullanılarak geliştirilen bir uygulama ile entegre edilmiştir. İlk aşamada, ChatGPT ile API entegrasyonu gerçekleştirilmiş ve doğru, kesintisiz bir iletişimin sağlandığı gözlemlenmiştir. Geliştirilen bu uygulama, kullanıcıların giriş yapmasını gerektiren bir sistem sunmaktadır. Kullanıcı adı ve şifre ile başarılı bir giriş yapıldıktan sonra, kullanıcının seçtiği malzemeler ve geçmiş yemek tercihleri temel alınarak, ChatGPT'den yeni yemek tarifleri önerileri almak üzere bir prompt oluşturulmaktadır. Elde edilen yanıtlar, kullanıcıların erişimine sunulmaktadır.

Kullanıcılar, sunulan yemek tariflerini 1 (en düşük) ile 5 (en yüksek) arasında bir puanla değerlendirmekte ve bu değerlendirme verileri, gelecekteki öneriler için veri tabanında saklanmaktadır. Uygulamanın performansını değerlendirmek amacıyla, "few-shot" komut stratejisi benimsenmiştir. Bu strateji, kullanıcının geçmişte beğendiği tariflere dayalı olarak, kişiselleştirilmiş önerilerin doğruluğunu artırmayı hedeflemektedir.

Few-shot yöntemi kapsamında, top_50_rand ve most_common malzeme seçim teknikleri kullanılmıştır. Top_50_rand yöntemi, en sık kullanılan 50 malzemeden rastgele 5 tanesini seçerken, most_common yöntemi en sık kullanılan 5 malzemeyi doğrudan seçmektedir. Bu malzemeleri içeren en yüksek puanlı 3 yemek tarifi, eğitim verisi olarak kullanılmıştır. Test verisi olarak ise, aynı malzemeleri içeren en fazla 20 yemek tarifi seçilmiştir. Bu tarifler, kullanıcı tarafından beğenilmiş ve beğenilmemiş tarifler arasından dengeli bir şekilde seçilmiştir. Kullanıcı puanlamaları göz önünde bulundurularak, değerlendirilmeyen tariflerin beğenilmemiş olarak kabul edildiği varsayılmıştır.

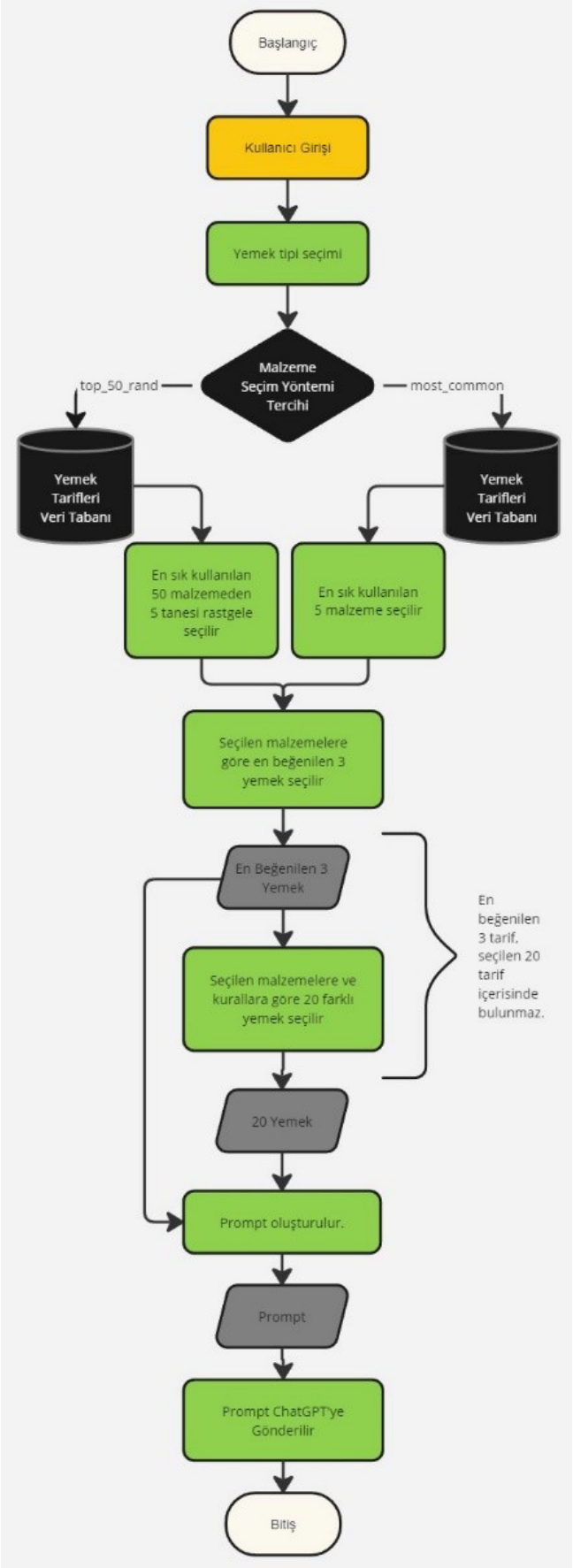
Seçilen 20 yemek tarifinin 10'u, kullanıcı tarafından beğenilmiş tariflerden, diğer 10'u ise değerlendirilmeyen (ve dolayısıyla beğenilmemiş sayılan) tariflerden seçilmiştir. Hatalı sonuçları önlemek için, bu 20 test tarifinin içerisinde, eğitim verisi olarak kullanılan en yüksek puanlı 3 tarifin yer almamasına dikkat edilmiştir. ChatGPT'ye, bu eğitim ve test verileri gönderilerek, kullanıcının geçmiş tercihlerine ve belirlenen malzemelere dayalı olarak, en az bir malzemeyi içeren 5 yemek tarifinin listelenmesi istenmiştir.

Önerilerin doğruluğunu ve sıralama kalitesini değerlendirmek için NDCG@5 metriği kullanılmıştır. NDCG (Normalized Discounted Cumulative Gain), bir sıralamanın kalitesini ölçerek, kullanıcıya en uygun sonuçların ne kadar yukarıda yer aldığını gösteren bir ölçüttür. Buradaki "@5" ifadesi, sıralamanın ilk 5 öğesi dikkate alınarak değerlendirildiğini belirtir. Yani, NDCG@5, öneri listesinin ilk 5 öğesinin sıralama kalitesini ve bu öğelerin kullanıcı için ne kadar uygun olduğunu ölçmektedir.

Ayrıca, önerilen 5 yemek tarifinin kaç tanesinin kullanıcı tarafından beğenilen tariflerden biri olduğunu belirten seçim başarı oranı da bir diğer performans ölçütü olarak kullanılmıştır. Bu oran, önerilen tariflerin kullanıcı tercihleriyle ne kadar uyumlu olduğunu değerlendirmede önemli bir gösterge olmuştur. NDCG@5 ve seçim başarı oranı birlikte kullanılarak, öneri sisteminin hem sıralama doğruluğu hem de kullanıcı memnuniyeti üzerindeki etkisi kapsamlı bir şekilde analiz edilmiştir.

Bu çalışma, öneri sistemlerinde LLM'lerin kişiselleştirme açısından etkili bir araç olabileceğini ortaya koymaktadır. Öneri sistemlerinin, kullanıcıların geçmiş tercihlerine dayalı olarak optimize edilmesi, daha tatmin edici ve doğru sonuçlar sunma potansiyeline sahiptir. Gelecekteki araştırmalar, bu yaklaşımın daha geniş kullanıcı grupları ve veri setleri üzerinde geçerliliğini test edebilir ve farklı LLM'lerin karşılaştırmalı analizlerini yapabilir.

Önerilen yaklaşıma ait akış şeması Şekil 1'de verilmiştir.



Şekil 1. Akış şeması

4 Deneysel Çalışmalar

Örnek bir prompt ve ChatGPT tarafından alınan yanıt Şekil 2 ve 3'de verilmiştir.

Prompt:

"Choose a list of 5 recipes from this list of recipe names [Test food recipe list]. Your picks must contain at least one of these ingredients [Selected ingredient list] and must be sorted by the likelihood of being liked by this user. List must be in a tabular form. Only choose from the given list, don't recommend a new one. Give recipe names between two #. Here are some top-rated recipes based on similar ingredients: [Training food recipe list]
Being 5 is the best and 1 is the worst, these are the previously highest ranked recipes by this user. Make your recipe names recommendation similar to highest rated recipes by this user."

Şekil 2. Gönderilen komut

Yanıt:

"| Recipe Name |
|----------|
| #Recipe Name# |
| #Recipe Name# |
| #Recipe Name# |
| #Recipe Name# |
| #Recipe Name# | "

Şekil 3. Alınan yanıt

Testler, farklı değerlendirme puanlarının analiz edilmesi amacıyla iki kullanıcı seçilerek gerçekleştirilmiştir. Bu kullanıcılar, daha önce değerlendirdikleri yemeklerde bulunan malzemeler temel alınarak seçilmiştir. Malzemelerin seçimi, top_50_rand ve most_common yöntemleri kullanılarak yapılmıştır.

Elde edilen bulgular Tablo 1'de sunulmuştur. Buna göre, top_50_rand ve most_common yöntemleri kullanılarak yapılan testlerde, seçim başarı oranı ve NDCG@5 metriği üzerinden analizler gerçekleştirilmiştir.

Kullanıcı 1 için, her iki yöntem de benzer seçim başarı oranları (%40) göstermiş ve NDCG@5 puanları 0.85 ile 1.00 arasında değişmiştir. Most_common yöntemi, sıralama açısından daha yüksek doğruluk sağlamasına rağmen, bazen seçim başarı oranının %20'ye kadar düşmesine neden olmuştur. Bu düşüşün sebebi seçilen malzeme setin için ilgili kullanıcının yeteri kadar değerlendirilebilecek yemek tarifi bulunmamasıdır.

Bu, yöntemin her durumda tutarlı performans sergileyemeyebileceğini göstermektedir.

Kullanıcı 2 için, most_common yöntemi %100 seçim başarı oranı ve yüksek NDCG@5 puanları ile güçlü bir performans sergilemiştir. Bu, kullanıcının geçmiş tercihlerine dayalı yaygın malzeme seçiminin doğru tarifleri ön plana çıkardığını göstermektedir. Top_50_rand yöntemi ise %43 NDCG@5 puanı ile daha düşük sıralama doğruluğu sağlamıştır. Ancak bazı durumlarda seçim başarı oranını %80'e kadar yükseltebilmiştir.

Tablo 1. Deneysel bulgular

Kullanıcı	Malzeme Seçim Metodu	NDCG@5	Seçim Başarı Oranı	Kullanıcın Beğendiği Tarif Adedi	Toplam Tarif Adedi
1	top_50_rand	0,85	0,40	5	15
1	top_50_rand	1	0,40	4	14
1	most_common	0,85	0,40	5	15
1	most_common	1	0,20	5	15
2	most_common	1	1	6	16
2	most_common	1	0,40	6	16
2	top_50_rand	0,43	0,60	2	12
2	top_50_rand	1	0,80	6	16

5 Sonuçlar

Bu çalışmada, ChatGPT kullanılarak geliştirilen yemek tarifi öneri sistemlerinin doğruluğu ve sıralama kalitesi, farklı malzeme seçim yöntemleri kullanılarak değerlendirilmiştir. Top_50_rand ve most_common yöntemlerinin her ikisi de belirli kullanıcılar için kişiselleştirilmiş öneriler sunmada etkili olabilmiştir, ancak yöntemlerin performansı kullanıcılar arasında farklılık göstermiştir. Bu farklılık, kullanıcıların geçmiş yemek tercihleri, etkileşim şekilleri ve malzeme seçim alışkanlıklarındaki bireysel farklılıklardan kaynaklanmış olabilir.

NDCG@5 değerlerine bakıldığında, most_common yöntemi genel olarak daha yüksek sıralama doğruluğu sağlamıştır. Kullanıcı 1 için NDCG@5 puanları 0.85 ile 1.00 arasında değişmiş ve most_common yöntemi sıralama doğruluğunu artırmıştır. Kullanıcı 2 için ise most_common yöntemi, %100 seçim başarı oranı ve en yüksek NDCG@5 puanları ile güçlü bir performans sergilemiştir. Bu, most_common yönteminin, kullanıcının beğendiği tarifler arasında daha yüksek doğrulukla seçimler yapabildiğini göstermektedir.

Öte yandan, top_50_rand yöntemi daha değişken bir performans sergilemiştir. Kullanıcı 2 için bazı durumlarda %80 seçim başarı oranına ulaşmış, ancak NDCG@5 puanı 0.43 gibi daha düşük bir değerde kalmıştır. Bu, yöntemin daha çeşitli ancak bazen kullanıcı tercihlerine tamamen uygun olmayan öneriler sunduğunu göstermektedir. Kullanıcı 1 için de benzer şekilde, top_50_rand yöntemi %40 seçim başarı oranı ile istikrarlı bir performans sergilemiş, ancak bu durum önerilerin sıralama kalitesini tam anlamıyla yansıtamamıştır.

Sonuçların değerlendirilmesinde, kullanıcının beğendiği tarif adedi ile toplam tarif adedi arasındaki oranlar da önemli bir rol oynamaktadır. Örneğin, Kullanıcı 2'nin toplam 16 tariften 6 tanesini beğenmiş olması, most_common yönteminin yüksek başarı oranı sağlamasına katkıda bulunmuştur. Bu durum, kullanıcı tercihleri ve mevcut tarif havuzunun homojenliği arasındaki dengenin, öneri sistemlerinin performansı üzerinde belirleyici bir etkisi olabileceğini göstermektedir. Buna karşılık, top_50_rand yöntemi, daha az sayıda tarifin beğenildiği durumlarda bile yüksek çeşitlilik sunarak kullanıcı ilgisini çekebilmiş ve bu yolla seçim başarı oranını artırabilmiştir.

Genel olarak, önerilen yaklaşımın başarılı olduğu ve kişiselleştirilmiş öneri sistemlerinde üretken yapay zekâ teknolojilerinin kullanılmasının, kullanıcı memnuniyetini artırma potansiyeline sahip olduğu sonucuna varılmıştır. Çalışma, giriş ve özet kısımlarında belirtilen amaçlara ulaşmış, ÜYZ'nin öneri sistemlerinin performansını iyileştirme potansiyelini ortaya koymuştur. NDCG@5 ve Seçim Başarı Oranı, öneri sisteminin sıralama kalitesinin ve kullanıcı memnuniyetinin, kullanıcıların geçmiş tercihlerini dikkate alan yaklaşımlarla optimize edilebileceğini göstermektedir.

Kullanıcının beğendiği tarif adedi ve toplam tarif adedi gibi faktörlerin, performans değerlendirmelerinde dikkate alınması gerektiği ve sonuçları önemli ölçüde etkileyebileceği görülmüştür. Gelecekteki araştırmalar, daha geniş kullanıcı gruplarında bu ilişkileri inceleyebilir ve farklı LLM'lerin karşılaştırmalı analizlerini yaparak öneri sistemlerinin performansını daha da optimize edebilir.

Kaynaklar

[1] Anderson, DF, Day, EA. Quantitavite analysis of the major free faty acids of cheddar cheese. J. Dairy Sci., 47,733-738, 1965.

[2] Dai S, Shao N, Zhao H, Yu W, Si Z, Xu C, Sun Z, Zhang X, Xu J. Uncovering ChatGPT's Capabilities in Recommender Systems. In RecSys. ACM,1126–1132, 2023.

[3] Chang, Y, Wang X, Wang J, Wu, Y, Yang L, Zhu K, Chen H, Yi X, Wang C, Wang Y, Ye W, Zhang Y, Chang Y, Yu PS, Yang Q, Xie X. A Survey on Evaluation of Large Language Models. arXiv:2307.03109, 2023.

[4] Fan W, Zhao, Z, Li J, Liu Y, Mei X, Wang Y, Tang J, Li Q. Recommender systems in the era of large language models (llms). arXiv preprint arXiv:2307.02046, 2023.

[5] Friedman L, Ahuja S, Allen D, Tan T, Sidahmed H, Long C, Xie J, Schubiner G, Patel A, Lara H. Leveraging Large Language Models in Conversational Recommender Systems. arXiv preprint arXiv:2305.07961, 2023.

[6] Harte J, Zorgdrager W, Louridas P, Katsifodimos P, Jannach P, Fragkoulis M. Leveraging Large Language Models for Sequential Recommendation. In RecSys. ACM, 1096–1102, 2023.

[7] Hu L. Generative AI and Future. https://pub.towardsai.net/generativeai-and-future-c3b1695876f2, 2023.

[8] Jovanović M. Generative Artificial Intelligence: Trends and Prospects. https://www.computer.org/csdl/magazine/co/2022/10/09903869/1H0G6xvtREk.0.1109/MC.2022.3192720, 2023.

[9] Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Gasser U, Groh G, Günnemann S, Hüllermeier E, Krusche S, Kutyniok G, Michaeli, T, Nerdel C, Pfeffer J, Poquet O, Sailer M, Schmidt, A, Seidel T, Stadler M, Weller J, Kuhn J, Kasneci G. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences,vol 103, 2023.

[10] Lanling X, Junjie Z, Bingqian L, Jinpeng W, Mingchen C, Wayne XZ, Ji-Rong W. Prompting Large Language Models for Recommender Systems: A Comprehensive Framework and Empirical Analysis. In . ACM, New York, NY, USA, 2023.

[11] Peng, L, Lemei Z, Atle JG. Pre-train, Prompt, and Recommendation: A Comprehensive Survey of Language Modeling Paradigm Adaptations in Recommender Systems. Transactions of the Association for Computational Linguistics, vol. 11, 1553–1571, 2023.

[12] Ricci F, Rokach L, Shapira B. Recommender Systems Handbook. 10.1007/978-0-387-85820-3_1, 2010.

[13] White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, Elnashar A, Spencer-Smith J, Schmidt DC. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT arXiv:2302.11382, 2023.

Uçak Bakımlarında Rutin Olmayan İş Yükünün Tahmini Üzerine Deneysel bir Çalışma

Mehmet İkbal KARACA[1]*, Figen ÖZTOPRAK[1]

[1]*Gebze Teknik Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Kocaeli, TÜRKİYE*

Özet

Uçak bakımları, havacılık endüstrisinde en üst düzeyde güvenlik sağlamanın ve uçakların sorunsuz bir şekilde çalışmasını garantilemenin temel unsurlarından biridir. Uçak bakımları, düzenli rutin kontrollere ek olarak rutin olmayan (non-routine) işleri de içerir. Uçakların bakım tipine göre uygulanması gereken detaylı veya genel muayene kartları yüksek adam saat değerine sahip rutin olamayan işler oluşturabilmektedir. Rutin olmayan işler, hem maliyet hem de bakım planlaması açısından belirsizlik yaratmaktadır. Bu çalışmada, rutin olmayan işlerin yarattığı ek iş yükü sürelerinin tahmini üzerine deneysel bir araştırma yapılmıştır. Uluslararası bir havayolu bakım şirketinin verilerine dayalı olarak, farklı makina öğrenmesi tekniklerine dayalı modeller üretilmiş ve bunların tahmin başarıları karşılaştırmalı olarak değerlendirilmiştir. Bu tekniklerin performansı detaylı olarak analiz edilerek, tahmin hatalarının problem karakteriyle bağlantısı gösterilmiştir.

Anahtar Kelimeler: *Bakım planlama, rutin-olmayan bakım işleri, veriye dayalı tahmin modeli*

An Experimental Study on the Prediction of the Workload of Non- Routine Tasks in Aircraft Maintanence

Abstract

Aircraft maintenance is a fundamental aspect of ensuring the highest level of safety and guaranteeing the smooth operation of aircraft in the aviation industry. Aircraft maintenance includes not only regular routine inspections but also non-routine tasks. Depending on the type of maintenance, detailed or general inspection cards can generate non-routine tasks with high man-hour workloads. Non-routine tasks create uncertainty in terms of both cost and maintenance planning. This study presents an experimental study on the estimation of the additional workload durations caused by non-routine tasks. Based on data from an international airline maintenance company, models based on different machine learning techniques were developed and their prediction accuracies were evaluated comparatively. The performance of these techniques was analyzed in detail, and the connection between prediction errors and the problem characteristics was demonstrated.

Keywords: *Maintanence planning, non-routine maintanence tasks, data based prediction model*

1 Giriş

Havacılıkta, uçakların belirli tip bakım süreçlerine alınması rutin bir işlemdir. Her bakım tipi için uçakta yapılacak incelemeler ve bakım işlemleri büyük ölçüde standarttır ve bu işlemlerin her birisi bakım kartları denilen kartlar kullanılarak tanımlanır. Rutin bakım görevleri belirli bir plana göre gerçekleştirilirken, rutin olmayan görevler beklenmedik arızalar veya keşfedilen sorunlardan kaynaklanır ve bu durum operasyonel belirsizliklere yol açar Yapılan araştırmalar rutin olmayan görevlerin, toplam bakım iş yükünün %50'sini oluşturabildiğini göstermektedir.

Rutin olmayan görevlerin iş yükünü tahmin etme alanında yapılan çalışmalar, genellikle veri analitiği yöntemlerine dayanmaktadır[4,5]. Rutin olmayan kartların adam saatlerinin tahmini karmaşık ve

*Contact email: m.karaca2022@gtu.edu.tr

değişkenlik gösteren bir konudur. Mevcut durumda tüm bakım paketine oransal olarak yapılan tahminler bakım sonunda adam saat sapmalarına sebep olmaktadır.

2 Metodoloji

Bu çalışmada, veriden öğrenerek tahmin modelleri ortaya koyan makina öğrenmesi teknikleri kullanılarak rutin olmayan görevlerin yarattığı ek adam saat bakım sürelerinin tahmini için modeller üretilmiştir.

2.1 Modelleme

Çalışmamızda, tahmin modelinin girdileri, uçak tipi, bakım süresi, uçağın bakım türü, uçağın girmiş olduğu bakımdan bir önceki ağır bakımda rutin olmayan işlere harcanılan adam saatler ve uçağın yaşı olarak belirlenmiştir. Modelin çıktısı ise belirli bakım kartları altında açılan ek bakım süreleri olarak tanımlanmıştır.

Model girdi ve çıktılarını ilişkilendiren farklı tür modeller elde etmek üzere üç tip makina öğrenmesi tekniği uygulanmıştır: Doğrusal ve polinomsal regresyon analizi, rassal orman (random forest) ve gradyan arttırma (gradient boosting).

Polinomsal regresyon, lineer regresyonun genelleştirilmiş hali olup, veri setindeki karmaşık ilişkileri daha iyi yakalamaya çalışır. Polinomsal regresyonda veriler arasındaki ilişkiyi modellemek için polinom fonksiyonlar kullanılır. Rassal orman regresyon, veri setini rastgele alt kümelere ayırarak ve her alt kümeyi ayrı karar ağaçlarıyla eğiterek çeşitlilik oluşturur. Rassal orman yöntem, aşırı uyumlanmayı azaltır ve tahmin doğruluğunu artırır. Son olarak, gradyan arttırma yöntemi de rassal orman tekniğinde olduğu gibi çok sayıda zayıf modelin bir araya getirilmesi prensibine dayalı olarak çalışır [1].

2.2 Veri Seti

Çalışma, uluslararası bir havayolu şirketinin 4 yıllık bir süre zarfında 250'den fazla dar gövdeli uçak bakımında, kabin içini muayene ettiren kartlara istinaden açılmış rutin olmayan işleri içeren bir veri setine dayanmaktadır. Rutin olmayan görevlerin kaynağı, rutin görevlerden, iş paketlerinden veya mürettebat raporlarından gelebilir. Bu çalışma, rutin detaylı muayene kartlarından kaynaklanan rutin olmayan görevlerin iş yükünü tahmin etmeye odaklanmaktadır.

3 Deneysel Çalışma

3.1 Uygulama

Çalışmada öncelikle veri setinin önişlemesi (preprocessing) yapılmıştır.

Tüm yöntemler için Scikit-Learn kütüphanesinden uygun rutinler kullanılmıştır. Kategorik değişkenler ikili açılım yapılarak (one-hot encoding) modele girdi olarak verilmiştir.

Doğrusal regresyon analizinden elde edilen bulgular değişkenler arasında doğrusal ilişki bulunmadığını göstermiştir. Doğrusal olmayan ilişkileri modellemek için veriye en uyan polinom fonksiyon deneysel olarak belirlenmiştir. İlk denemelerde eğitim setini fazla öğrenme (overfit) gözlemlendiği için eğitim hata terimine bir l2-norm regülarizasyon terimi eklenmiştir (ridge regression). Regülarizasyon teriminin katsayısı yine deneysel olarak belirlenmiştir. Benzer şekilde, rassal orman ve gradyan arttırma yöntemleri için de hiperparametreler farklı değerler denenerek belirlenmiştir.

Bu makine öğrenmesi yöntemlerinin yanı sıra verilerin sınıflandırılması, ortalama ve yaygınlık ölçülerinin hesaplanması, tablo ve grafiklerle sunulmasını içeren betimleyici(descriptive) istatistik analizleriyle de her kart için (modele dayalı olmayan) tahmin adam saat çalışması yapılmıştır.

3.2 Sonuçlar

Deneylerde kullanılan tüm yöntemler, veri setinin %40'ı olarak belirlenen bir test seti üzerinden karşılaştırılmıştır. En düşük test tahmin hataları polinomsal regresyon modeli ile elde edilmiştir. Bunu rassal orman ve gradyan arttırma teknikleri ile elde edilen modeller takip etmektedir.

Araştırmamızın sonuçları, makina öğrenmesi modellerinin, rutin olmayan görevlerin adam saatleri tahmininde kısmi başarı sağladığını göstermektedir. Bir modele dayanmayan temel istatistiki değerlere dayalı tahminlere göre çok önemli bir iyileşme elde edilememiştir. Sonuçlar, model girdilerinin model içerisinde nasıl rol aldığı incelenerek yorumlanmıştır ve bu durum önemli

ölçüde problemin doğasında belirlemeyen faktörlerin bulunmasına bağlanmıştır. Buna ek olarak, çalışmada kullanılan veri setinde tespiti mümkün olmayan hatalı değerlerin bulunabileceği de düşünülmektedir. Bu gözlemlerimiz, bu konuda daha önce yapılmış çalışmalarla büyük ölçüde uyumludur [2,3].

4 Sonuç ve Gelecek Çalışmalar

Bu çalışma, havayolu bakım planlamasında rutin olmayan iş yükünü tahmin etmek için makina öğrenmesi yöntemlerinin etkinliğine dair bulgular sunmaktadır. Polinomiyal regresyon, en iyi tahmin modelini sağlamıştır. Bununla beraber, makina öğrenmesi ile elde edilen tahmin modelleri daha basit istatistiksel tekniklere kıyasla çok önemli tahmin iyileştirmeleri sağlayamamıştır. Bunun iki sebebi olduğu düşünülmektedir: Bunlardan ilki ve daha önemlisi problemin yapısal olarak sahip olduğu öngörülemezlik, ikincisi ise eldeki verilerin çeşitli sebeplerle hatalar içermesi ve bunların tamamının tespit edilememesidir. Bütün bu gözlemlerimiz ışığında, bakım planlamasının daha az sapma olacak şekilde yapılmasının tahmin modelleri kullanarak değil, gürbüz(robust) planlama teknikleri ile mümkün olabileceğini öngörüyoruz. Gelecek çalışmalarımız, buna uygun olarak, rutin olmayan işlerin yol açtığı belirsizlikleri bir küme olarak tanımlayıp planlamaya dahil etme yaklaşımını izleyecektir.

Teşekkür

Bu çalışma TÜBİTAK Bilim İnsanı Destek Programları Başkanlığı 2210/D Yurt İçi Sanayiye Yönelik Yüksek Lisans Burs Programı altında 1649B022314548 numarasıyla desteklenmektedir.

Kaynaklar

[1] Murphy, KP. Machine learning: a probabilistic perspective. MIT press, 2012.

[2] Kazan, H. , Tavsamaz, A. , "Havacılıkta Üreyen Rutin Olmayan (Non-Routine) İşlerin Adam X Saat Tahmini: Gerçek Zamanlı Bakım Kartlarında Uygulama", *Alphanumeric Journal* , 2 (1) , 23-42, 2014.

[3] Li, H., Ribeiro, M., Santos, B., Tseremoglou. I., "Prediction of Non-Routine Tasks Workload for Aircraft Maintenance with Supervised Learning." In *AIAA SCITECH 2024 Forum*, p. 2529. 2024.

[4] Aungst, J., Johnson, M. E., Lee, S. S., Lopp, D., Williams, M., "Planning of non-routine work for aircraft scheduled maintenance." In *Proceedings of the 2008 IAJC-IJME International Conference, Purdue University*. 2008.

[5] Pelt, M., Stamoulis, K., Apostolidis, A., "Data analytics case studies in the maintenance, repair and overhaul (MRO) industry". In: *MATEC Web of Conferences* 304. 2019.

Veri Yönetimi Metamodeline Dayalı Çoklu Modelli Veri Bütünleştirme Mimarisi

Taner Gürel[1]*, Ceyda Özcan Zaim[2], Elif Özdoğan[2], Hale Gezgen[2], Murat Osman Ünalır[1]

[1]*Ege Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, TÜRKİYE*
[2]*Arçelik Dijital Dönüşüm, Büyük Veri ve Yapay Zeka Ar-Ge Merkezi, İstanbul, TÜRKİYE*

Özet

Farklı veri kaynaklarından elde edilen verilerin bütünleştirilmesi, veri yönetimi süreçlerinin en karmaşık ve önemli aşamalarından biridir. Özellikle yıllardır saklanan geçmiş verilerin güncel veriler ile bütünleştirilerek yapay zeka ve makine öğrenmesi uygulamalarında kullanılması gerekmektedir. Veri kalitesi ve hacminin artırılarak daha kapsamlı içgörüler elde edilmesi, ortak bir sistem ile veri yönetimi ve analizi yapılması açısından kuruluşlara fayda sağlamaktadır. Farklı veritabanı yönetim sistemleri, dosya formatları ve veri modelleri karmaşıklığa sebep olarak büyük miktarda veri ile başa çıkmayı zorlaştırmaktadır. Bu çalışmada, çoklu modelli veri bütünleştirme için veri yönetimi metamodeline dayalı bir mimari önerilmektedir. Böylece, farklı veri kaynaklarının ilişkilendirilmesi ve bütünleştirilmesi, kullanıcıların ortak platform üzerinden ihtiyacı olan verilere erişimi ve analiz yapabilmeleri sağlanmaktadır.

Bu mimarinin temel bileşenleri veri modelleme, veri dönüşümü, veri bütünleştirme ve veri sunumudur. Farklı veri kaynaklarının, ortak bir veri yönetimi metamodeli ile bütünleştirilmesi, bu kaynaklardaki verilerin analizini kolaylaştırmaktadır.

Veri sunumunda tekil ve çoklu anahtar kelimeler kullanılarak bir bilgi çizgesi üzerinde istenilen verilere erişim sağlanabilmektedir. Erişim ve güvenlik katmanlarında filtrelenen veriler raporlar, grafikler, tablolar ya da veri erişim noktaları şeklinde sunulabilmektedir.

Önerilen mimari veri erişimini ve analizini kolaylaştırmayı, veri güvenliği ve uyumluluğunu artırmayı ve bu alanlarda gelecekte yapılacak çalışmalara katkı sağlamayı hedeflemektedir. Bu sayede farklı veri modellerini destekleyen veri kaynaklarının ortak bir metamodel aracılığıyla yönetilmesi, izlenebilirliği ve sürekliliği sağlanmaktadır. Bu çalışma ile birlikte, büyük dil modellerine dayalı üretken yapay zeka teknolojilerinin bütünleşik bir şekilde sunulan bilgi çizgeleri ile etkileşimi olanaklı hale gelmektedir.

Anahtar Kelimeler: *Veri yönetimi metamodeli, Çoklu modelli veri mimarisi, Veri bütünleştirme*

Polyglot Data Integration Architecture based on Data Management Metamodel

Abstract

Integrating data obtained from different data sources is one of the most complex and important stages of data management processes. In particular, the integration of the legacy data that has been stored for years with the current data needs to be utilized in artificial intelligence and machine learning applications. Obtaining more comprehensive insights by increasing data quality and volume provides benefits to organizations in terms of data management and analysis with a common system. Different database management systems, file formats and data models make it difficult to deal with large amounts of data by adding complexity. In this study, an architecture based on data management metamodel is proposed for polyglot data integration. Thus, various benefits are provided, such as associating and

*Contact email: 91220001536@ogrenci.edu.tr

integrating different data sources and enabling users to access and analyze the data they need through a common platform.
The key components of this architecture are data modelling, data transformation, data integration and data presentation. Integrating different data sources with a common data management metamodel facilitates the analysis of data in these sources.
The desired data can be accessed on a knowledge graph by using single and multiple keywords in data presentation. Data filtered in the access and security layers can be presented in the form of reports, graphs, tables or data access points.
The proposed architecture aims to facilitate data access and analysis, increase data security and compatibility and contribute to future studies in these areas. As a consequence, data sources supporting different data models are managed, traceable and maintained through a common metamodel. With this study, it becomes possible to interact generative artificial intelligence technologies based on large language models with knowledge graphs that are presented integratedly.

Keywords: *Data management metamodel, Polyglot data architecture, Data integration*

1 Giriş

Çoğu organizasyon farklı amaçlar için farklı uygulamalar ve veri kaynakları kullanmaktadır. Bir e-ticaret firması müşteri verileri yönetmek için bir Müşteri İlişkileri Yönetimi sistemi, finansal verileri yönetmek için bir muhasebe yazılımı ve stok takibi için ise başka bir sistem kullanabilir. Her veri kaynağı belirli bir amaca hizmet etmekte fakat veriler ayrı sistemlerde depolandığı için organizasyon genelinde bütünsel bir bakış açısı elde edilememektedir. Bu durum veriye dayalı karar verme süreçlerinin etkinliğini azaltmaktadır.

Daha iyi kararlar alınabilmesi ve verinin potansiyelinden tam anlamıyla yararlanılabilmesi için veri kaynaklarının bütünleştirilmesi ve birbiriyle ilişkilendirilmesi gerekmektedir. Organizasyonun müşteri verilerini satış verileri ile ilişkilendirmesi müşteri davranışları hakkında daha derin analizler yapılmasını ve pazarlama stratejilerinin iyileştirilmesini sağlar. Benzer şekilde üretim hatalarına neden olan faktörlerin analiz edilmesi için üretim, tedarik zinciri ve kalite kontrol verilerinin bir araya getirilmesi istenebilir. Bütünleştirme süreci sadece verilerin bir araya getirilmesi değil, aynı zamanda bu verilerin ortak bir şema ile organize edilmesi anlamına gelmektedir. Organizasyonlar yalnızca mevcut durumlarını anlamakla kalmaz, aynı zamanda gelecekteki fırsatları ve riskleri de daha iyi öngörebilir.

2 Temel kavramlar ve teknolojiler

Veri yönetimi ve bütünleştirilmesi sürecinin daha iyi anlaşılması için öncelikle önemli bazı temel kavramlar ve teknolojiler açıklanmalıdır.

2.1 Veri yönetimi

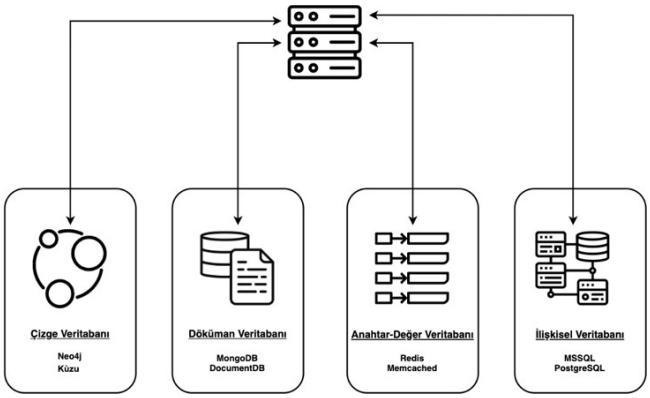
Veri günlük operasyonların sorunsuz bir şekilde yürütülmesi için kritik bir öneme sahiptir ve organizasyonlar için günlük işlemlerin temel yapı taşı olarak kabul edilmektedir. İşletmelerin veri olmadan var olması mümkün değildir. Veri yönetimi ise bir organizasyonun sahip olduğu verileri etkin bir şekilde yönetmek ve değer yaratmak amacıyla politikalar ve uygulamalar geliştirme sürecidir. Birçok organizasyon için veri stratejik bir varlık olarak kabul edilmektedir. Müşteri bilgileri ve hizmet kalitesi gibi veriler stratejik hedeflere ulaşmayı ve rekabet avantajı sağlamaktadır. Veri potansiyelinin ortaya çıkarılması için bilinçli bir planlama ve kararlılık gerekmektedir. Bu nedenle veri yönetimi hem teknik altyapının doğru çalışmasını hem de iş süreçlerinin veriye dayalı kararlarla desteklenmesini amaçlamaktadır. Sadece teknik beceriler gerektiren bir alan değildir, iş birimi ve bilgi teknolojisi rolleri arasında işbirliği ve koordinasyon gerektirir [1].

2.2 Çoklu modellilik

Farklı ihtiyaçlar için birden fazla veritabanı kullanma stratejisidir. Her veritabanı kullanım durumları ve iş yüklerine göre iyileştirilmiştir. İlişkisel veritabanları karmaşık sorgular ve veri bütünlüğünün önemli olduğu durumlarda, NoSQL veritabanları ise büyük hacimli verinin işlenmesi gerektiğinde kullanılabilir. Farklı veri yapıları için en uygun veritabanını kullanmak mimariye esneklik kazandırır. Yapılandırılmış veri için ilişkisel veritabanı, yarı yapılandırılmış veya yapılandırılmamış veri için ya da yatay ölçeklenebilirlik özellikleri nedeni ile NoSQL veritabanları tercih edilebilir. Oturum yönetimi gibi

iş süreçlerinin iyileştirilmesi gerektiğinde bellek tabanlı anahtar-değer bir veritabanı süreci hızlandırırken, analitik işlemlerin yoğun olduğu durumlarda zaman serisi bir veritabanı daha performanslı çıktılar sağlayabilir. Böylece donanım ve yazılım kaynakları daha verimli kullanılırken, maliyet yükü de azaltılabilir [2].

Uygulamalardan toplanan verilerin çoklu modelli depolanması Şekil 1 ile görselleştirilmiştir.



Şekil 1. Çoklu modelli mimari

2.3 Veritabanı Teknolojileri

Verilerin saklanması, yönetilmesi ve işlenmesi için kullanılan, veri operasyonlarının temelini oluşturan teknolojilerdir. Veri türleri ve iş yüklerine uygun mimari ve işlevsellik sunarlar.

2.3.1 İlişkisel veritabanları

Verilerin tablo olarak saklandığı ve aralarında ilişkilerin tanımlandığı veritabanı sistemleridir. Şemalar ile verilerin tutarlı bir yapıda olmasını sağlayarak büyük ve karmaşık veri yapıları için organizasyon ve düzen, ACID (Atomicity, Consistency, Isolation, Durability) özellikleri ile ise hassas verilerin işlenmesinde veri bütünlüğü ve güvenilirliği sağlarlar. Sorgu dili olarak kullanılan SQL (Structured Query Language) ilişkisel veritabanları arasındaki işlemleri standartlaştırır ve kolaylaştırır. En popüler ilişkisel veritabanları Microsoft SQL Server, PostgreSQL, MySQL ve Oracle'dır [3].

2.3.2 Anahtar-değer veritabanları

Verilerin basit anahtar ve değer çiftleri şeklinde saklandığı NoSQL veritabanlarıdır. Yüksek performans gerektiren uygulamalar için idealdir çünkü verilerin hızlı bir şekilde okunup yazılmasını sağlarlar. Esnek ve yatay ölçeklenebilir bir mimariye sahiptirler. En popüler örnekleri Redis ve Amazon DynamoDB'dir. Yüksek hızda veri erişimi gerektiren uygulamalar için Redis bir önbellek ve mesajlaşma altyapısı olarak kullanılabilirken, DynamoDB yüksek ölçeklenebilir uygulamalar için uygun bir çözüm sunmaktadır [4].

2.3.3 Döküman veritabanları

Verilerin BSON, JSON, XML gibi yapılandırılmış veya yarı yapılandırılmış döküman formatlarında saklandığı NoSQL veritabanlarıdır. İlişkisel veritabanlarının aksine döküman veritabanları şemasızdır, yani her döküman farklı bir yapıya sahip olabilmektedir. Sürekli değişen veri yapıları ve karmaşık veri modelleri ile çalışırken avantaj sağlamaktadır. Veri yoğunluğu yüksek uygulamalar için optimize edilmişlerdir.

MongoDB, CouchDB ve Amazon DocumentDB en çok kullanılan döküman veritabanlarıdır. Esnek yapısı ve yüksek performansı ile büyük ölçekli web uygulamalarında genellikle MongoDB tercih edilirken, dağıtık ve çevrimdışı veri eşzamanlaması gerektiren uygulamalar için CouchDB daha uygundur [5].

2.3.4 Çizge veritabanları

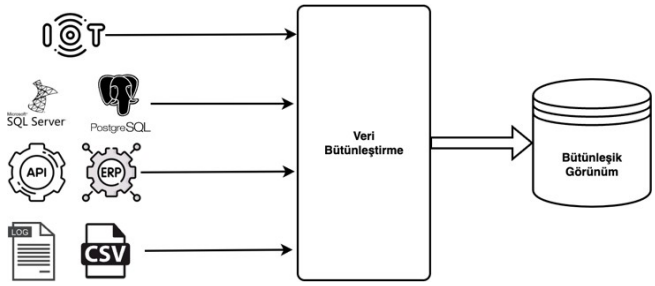
Verilerin düğümler ve bu düğümler arasındaki bağlantılar şeklinde saklandığı bir NoSQL veritabanı türüdür. Özellikle karmaşık ilişkilerin temsil edilmesi gerektiğinde, sosyal medya platformunda kullanıcılar arasındaki ilişkilerin ya da tedarik zincirindeki ürün akışının temsil edilmesi gerektiğinde kullanılmaktadırlar. İlişkilerin hızlı ve etkili bir şekilde sorgulanabilmesini ve analiz edilebilmesini sağlarlar.

Popüler çizge veritabanları arasında Neo4j, Amazon Neptune ve Azure CosmosDB yer alır [6]. KuzuDB ise RDF veri seti desteği ile çizge veritabanlarına katılan gömülü bir örnek olup yenilikçi bir çözüm sunmaktadır [7].

2.4 Veri bütünleştirme

Farklı kaynaklardan elde edilen verilerin analiz edilmesi ve anlamlı içgörüler elde edilmesi amacıyla bir araya getirilmesidir. Kullanıcıların veri kaynaklarına ayrı ayrı başvurma ve elde edilen verileri birleştirme ihtiyacını ortadan kaldırır. Özellikle büyük organizasyonların verimli bir şekilde çalışabilmesi için kritik öneme sahiptir. Organizasyonlar birleştiğinde ve merkezi olarak çalışmaya başladığında veri bütünleştirme ana odak noktası haline gelir ve kullanıcılar veriye tek bir arayüz üzerinden ulaşmak isterler (bk. Şekil 2).

Veri bütünleştirmenin temel zorlukları arasında veri kaynaklarının çeşitliliği ve özerkliği yer almaktadır. Bütünleştirme alandaki ilk çalışmalar 1970'lerin ortalarında gerçekleştirilmiştir. Veritabanı sistemlerinin çok katmanlı mimarisi hakkında net fikirler ortaya çıkmış ve dağıtık veritabanı sistemleri gelişmeye başlamıştır. Veri bütünleştirme günümüzde de oldukça önemli bir konudur [8].



Şekil 2. Veri bütünleştirme mimarisi

3 Veri yönetimi metamodeli odaklı üst veri havuzu

Farklı veri kaynaklarından toplanan verilerin ortak bie şema üzerinde bütünleştirilmesi için merkezi bir üst veri havuzunun oluşturulması gerekmektedir. Bu amaçla tüm veri kaynakları için ortak bir veri yönetimi metamodeli tasarlanmalıdır.

3.1 Üst veri

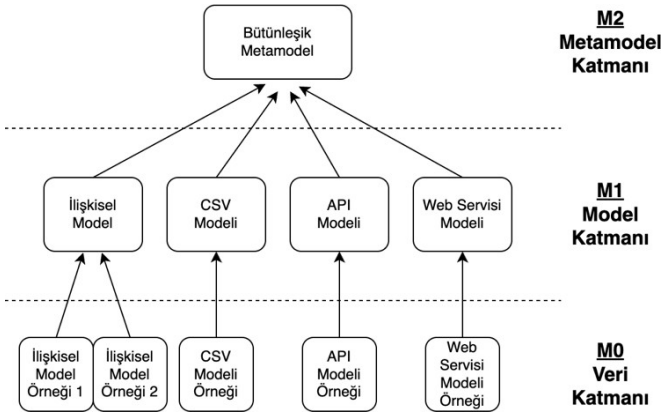
Veriler ile ilgili bilgileri içeren yapılandırılmış referans verilerdir. Ağırlıklı olarak yapılandırılmamış verilerin sıralanması ve özelliklerinin tanımlanması için gereklidir. Üst veri kalitesi, kaynakların bulunabilirliği ve yeniden kullanılabilirliği üzerinde doğrudan etkilidir. Belirli bir amaca hizmet eden birçok üst veri türü vardır. Örneğin başlık, yazarlar ve anahtar kelimeler gibi genellikle arama ve tanımlama için kullanılan açıklayıcı üst verilerdir. Bir nesnenin bileşenlerinin nasıl düzenlendiğini açıklayan yapısal üst verilerdir. Kaynak türü, izinler ve bir bilginin ne zaman ve nasıl oluşturulduğu gibi bilgileri ise yönetimsel meta veriler sağlamaktadır [9].

3.2 Metamodel

Modeller hakkında ifadeler içeren bir modeldir, bu yüzden metamodel de bir modeldir. Bilgi sistemlerindeki modellerde kullanılan yapı taşlarıyla ilgili ifadeler içerir. Yapı taşlarını tanımlayabilir veya yapı taşlarının istenen ve doğru özelliklerini ifade edebilir. Modeller nasıl ki bir gerçekliğin soyutlamalarıysa, metamodeller modellerin, meta-meta modeller de metamodellerin soyutlamalarıdır [10].

3.3 Meta object facility

Meta Object Facility (MOF), Object Management Group (OMG) tarafından metamodel yazmak için geliştirilen bir standart olarak görülebilir. Dört katmanlı bir metamodelleme yapısını temel alır fakat veri bütünleştirme süreci için M0, M1 ve M2 katmanlarının kullanımı yeterli olmaktadır (bk. Şekil 3). Bu katmanlar ile veri çeşitli soyutlama seviyelerinde temsil edilmektedir [11].



Şekil 3. Veri bütünleştirme mimarisi

M0 katmanı, ham verilerin bulunduğu ve işlenmek üzere tanımlandığı katmandır. İlişkisel veritabanındaki tablo satırları M0 katmanında yer alır. Farklı kaynaklarından gelen bu veriler genellikle farklı yapılara ve şemalara sahiptirler.

M1 katmanı, M0 katmanındaki veriyi temsil eden üst veri katmanıdır. Verilerin yapısal özellikleri tanımlanır ve veri modelleri oluşturulur. İlişkisel bir veritabanı tablosu, sütunları ve ilişkileriyle birlikte M1 katmanında bir model olarak ifade edilir. Bir JSON dökümanı yapısı da M1 katmanında bir modelle temsil edilir. Farklı veri kaynaklarından gelen bu yapılar standartlaştırılır, veriler arasında bağlantı kurmak ve ilişkileri anlamak daha kolay hale gelir.

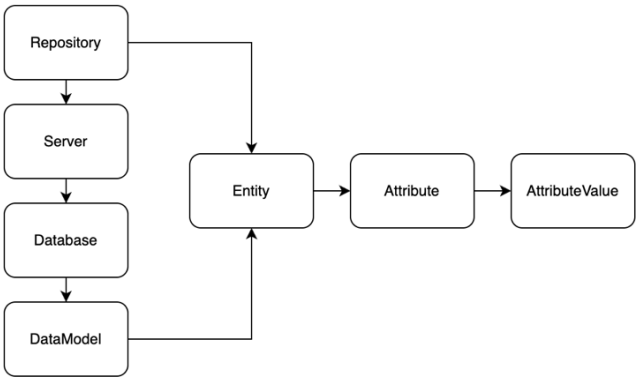
M2 katmanı ise M1 katmanında tanımlanan modellerin nasıl oluşturulduğunu tanımlayan metamodel katmanıdır. Farklı veri kaynaklarından gelen modellerin uyumlu hale getirilmesi, aynı çerçevede ifade edilmesi ve bütünleştirilmesi bu katmanda gerçekleştirilir [12].

3.4 Veri yönetimi metamodeli

Veri yönetiminde kullanılan kavramları, süreçleri ve ilişkileri tanımlayan üst düzey bir çerçeve olarak tanımlanır. Bir organizasyonun verilerini nasıl

yönettiğine dair temel yapı taşlarını ve birbirleriyle nasıl etkileşime girdiğini belirler. Veri yönetimi süreçlerinin standartlaştırılması, düzenlenmesi ve iyileştirilmesi için bir rehberdir. Veri yöneticilerine verinin yaşam döngüsünü yönetme, veriye erişimi kontrol etme ve verinin organizasyon içindeki değerini artırma konusunda bir harita sunar. Veri yönetişimi, veri mimarisi ve veri güvenliği gibi alanlarda rehberlik ederek organizasyonların veri varlıklarını etkin bir şekilde yönetmelerini sağlar [13].

Önerilen metamodeller kurumsal bir üst veri havuzu için temel oluşturmaktadır. Veri yönetimi metamodeli revize edilerek bütünleşik metamodel tasarlanmıştır. Metamodelin temel kavramları Şekil 4 ile gösterilmiştir.



Şekil 4. Bütünleşik metamodel

Repository, farklı veri kaynaklarını birleştiren merkezi bir üst veri havuzudur. Organizasyonların farklı yerlerde sakladığı verileri tek bir noktada toplar. Server ve Database ise ilişkisel veritabanlarında kullanılan sunucu ve veritabanı yapılarını temsil etmektedir. DataModel veritabanlarındaki veri organizasyonu ve bütünlüğünü sağlayan şemayı tanımlamaktadır. Entity, ilişkisel veritabanındaki tabloyu ifade edebileceği gibi, aynı zamanda bir CSV dosyasını veya JSON formatında bir API uç noktası yanıtını da temsil edebilir. JSON verileri yapısal olarak farklı olmasına rağmen bu modelde aynı şekilde ele alınabilir. Attribute, ilişkisel veritabanlarında bir tablonun sütunu ya da JSON dökümanında bir anahtar olabilir. İlişkisel veritabanında "Orders" tablosu "OrderDate" ve JSON dökümanında "firstName" anahtarı bir Attribute olarak kabul edilir. AttributeValue ise sütun ya da anahtarın içerdiği gerçek veriyi ifade etmektedir, "2019.07.15" ve "John" birer AttributeValue örneğidir. Ortak bir arayüz ile organizasyonların tüm veri potansiyelinden faydalanmaları ve daha derin içgörüler elde etmeleri sağlanır.

4 Çoklu modelli veri bütünleştirme mimarisi

Bütünleşik metamodel belirlendikten sonra veri kaynaklarındaki veriler ve üst veriler bütünleştirilmek üzere toplanır ve etkili bir şekilde saklanmaları, işlenmeleri ve sorgulanmaları için farklı veritabanlarına dağıtılır. Böylece veri bütünlüğü korunur ve verimlilik artırılır.

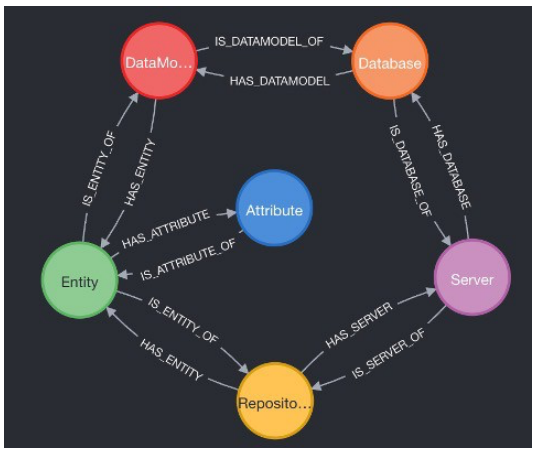
İlişkisel veritabanlarından üst verilerin toplanması için ANSI (American National Standards Institute) standardı kabul edilen ve salt okunur görünümlerden oluşan Information Schema, üst verileri sorgulamak için kullanılmaktadır. Tablo adı, oluşturulma tarihi, sütun veri tipi, varsayılan değeri, şema özellikleri, görünümler ve kısıtlar için bir üst veri deposudur [14]. CSV, API ve web servis üst verileri ise başlıkta yer alan sütunlar, JSON anahtarları veya XML etiketlerinden oluşmaktadır. Veri toplama süreci veri kaynağına göre değişiklik gösterir. İlişkisel veritabanları için SQL sorguları kullanılırken, dosya okunması, API uç noktasına istek gönderilmesi veya WSDL (Web Services Description Language) dosyası ve SOAP (Simple Object Access Protocol) isteği ile yapılabilmektedir.

4.1 Çoklu modelli mimari

Veri hatları aracılığıyla toplanan veriler Redis, MongoDB, Neo4j ve PostgreSQL veritabanlarında bütünleştirilmektedir.

4.1.1 Neo4j

En yaygın çizge veritabanlarından biridir. Cypher sorgu dili ile kolaylıkla sorgulanabilir. Geniş bir kullanıcı kitlesi, dökümantasyon desteği, esnek veri modelleme yetenekleri, entegrasyon kolaylığı ve ölçeklenebilir yapısı gibi özellikleri çizge veritabanı seçiminde etkili olmuştur [15].



Şekil 5. Neo4j şema görselleştirimi

Üst veriler düğüm şeklinde saklanmakta ve ilişkilendirilmektedir (bk. Şekil 5). Böylece veritabanları ve veri kaynakları arasında gezinim sağlanabilmektedir. Örneğin birbirinin kopyası olan veritabanları arasındaki şema farklılıkları bütünleşmiş üst veriler ile kolay bir şekilde sorgulanabilir hale gelmektedir.

4.1.2 MongoDB

Kolay ve gelişmiş ölçeklenebilirlik, dağıtık yapısı, geniş topluluk desteği, karmaşık sorgular için güçlü bir sorgulama dili ve tam yönetilen hizmet desteği ile döküman veritabanı seçimini kolaylaştırmıştır [16].

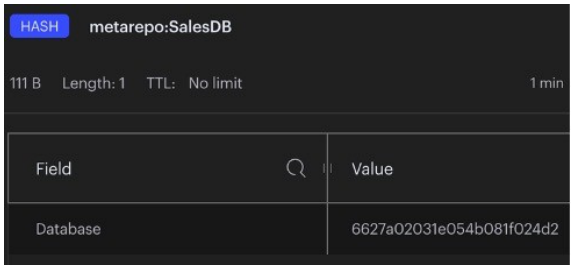


Şekil 6. MongoDB doküman yapısı

Uygulamalardan toplanan tüm veriler veri yönetimi metamodeline uygun şekilde koleksiyonlarda bütünleştirilmektedir. MongoDB'nin şemasız olması farklı veri modellerinin aynı koleksiyonda saklanabilmesini sağlamaktadır. Böylece veri kaynağı ve modelinden bağımsız bir şekilde veri aynı çatı altında bütünleştirilebilmektedir. Üst veriler Neo4j düğüm tanımlayıcıları ve ana düğüme kadar uzananan patika bilgileri ile depolanırken, veriler patika bilgilerine ek olarak Neo4j veritabanında karşılığı olmadığından düğüm tanımlayıcısı olarak NULL değer içermektedir (bk. Şekil 6).

4.1.3 Redis

Redis yalnızca anahtar-değer çiftleri değil hash, küme, JSON, sıralı küme, hiperlog gibi zengin bir veri yapısına sahiptir. Ölçeklenebilirlik ve yük devretme araçları da sağlamaktadır [17].



Şekil 7. MongoDB doküman yapısı

MongoDB veritabanına kayıtlı tüm veri ve üst veriler anahtar-değer şeklinde depolanmaktadır. Redis anahtarı veri ya da üst veri, Redis değeri ise veri yönetimi metamodeline uygun bir veya birden fazla alan-değer çifti içermektedir (bk. Şekil 7). Hash tipindeki alan-değer çifti ise alan kısmı verinin tipini, değer kısmı ise MongoDB döküman tanımlayıcısını içermektedir.

4.1.4 PostgreSQL

PostgreSQL GIN 3-gram dizinleri metin verilerini hızlı bir şekilde aramak için kullanılabilmektedir. Metinlerde geçen üç karakterlik parçalar dizinlenir ve sorgular daha etkili hale getirilebilir. Tam metin arama, metin benzerliği, sıralama ve filtreleme gibi özellikleri ile esnek sorgular yapılabilmektedir [18].

```
keyword |metadata      |similarity|
--------+--------------+----------+
SalesDB |Database      |       1.0|
Sales   |Entity        | 0.5555556|
SaleType|Attribute     |0.30769232|
For Sale|AttributeValue|0.30769232|
```

Şekil 8. PostgreSQL benzerlik araması

PostgreSQL veritabanı tüm Redis anahtarlarını içermektedir. 3-gram dizini üzerinden kullanıcının arama yaparken benzer anahtar kelimeleri de görebilmesi ve aramasını daha doğru ve istenen şekilde yapabilmesi için bir kelime öneri aracı görevi görmektedir. Kelimeler üst veri karşılıkları ile benzerliklerine göre büyükten küçüğe sıralı bir şekilde getirilmekte ve kullanıcının en olası seçimi yapabilmesi sağlanmaktadır (bk. Şekil 8).

Veriler kullanıcının isteğine göre bir veya birden fazla veritabanı erişimi ile sunulmaktır. İstek sadece Neo4j veritabanı kullanılarak ya da PostgreSQL, Neo4j, Redis ve MongoDB veritabanları kombinasyonlarından toplanan verilerin bütünleştirilmesi şeklinde de olabilmektedir.

5 Sonuçlar ve gelecek çalışmalar

Bu çalışma ile farklı veri kaynaklarının bütünleştirilmesinin ve anlamlandırılmasının nasıl etkili bir çözüm sunduğu ele alınmıştır. Farklı veritabanı türlerinin avantajları bir araya getirilerek hem esnek hem de ölçeklenebilir bir yapı sunulmaktadır, özellikle büyük veri ve veri yönetimi alanlarında kritik bir boşluğu doldurmaktadır. Veri bütünleştirme süreçlerinin karmaşıklığını azaltmak ve veri yönetimini daha etkin hale getirmek isteyen organizasyonlar için yeni bir yol haritası sunmaktadır. Farklı veri modelleri ve teknolojileri kullanan işletmelerin veri kaynaklarını bir araya

getirirken karşılaştıkları zorluklara çözüm sağlamayı hedeflemektedir.

Gelecek çalışmalar kapsamında, önerilen mimari üzerinde anlamsal arama yapabilecek bir altyapının geliştirilmesi, önemli bir araştırma alanı olarak öne çıkmaktadır. Bu genişleme ile kullanıcıların karmaşık veri kümeleri içinde daha doğru ve anlamlı bilgilere ulaşmalarının sağlanması hedeflenecektir. Veri yönetimi ve bilgi keşfi süreçleri anlamsal arama altyapısı ile daha verimli hale gelecektir. Geliştirilecek altyapı yapay zeka ve veri bilimi alanları için de daha kaliteli ve zenginleştirilmiş verilere erişme imkanı sağlayacaktır. Veri bilimi projelerinin başarısı kullanılan verinin kalitesine doğrudan bağlı olduğundan daha doğru ve güvenilir sonuçlar elde edilmesine katkıda bulunacaktır.

Teşekkür

Bu çalışma Arçelik Dijital Dönüşüm, Büyük Veri ve Yapay Zeka Ar-Ge Merkezi bünyesinde yürütülen ar-ge projesi kapsamında desteklenmiştir.

Kaynaklar

[1] Earley S, Henderson D, Data Management Association. "*DAMA-DMBOK : Data Management Body of Knowledge*". Second edition. Denville, NJ, USA, Technics Publications, 2017.

[2] Sadalage PJ, Fowler M. "*NoSQL Distilled : A Brief Guide to the Emerging World of Polyglot Persistence*". 6th printing. Upper Saddle River, N.J, Addison-Wesley, 2013.

[3] Elmasri RA, Navathe SB. *Fundamentals of Database Systems*. Seventh edition, global edition. Pearson, 2017.

[4] Alves Florencio D, Ricardo Freitas de Oliveira D, Laisa Soares Xavier Freitas E, da Fonseca de Souza F. "Which Fits Better? A Comparative Analysis about NoSQL Key-Value Databases". *IEEE Latin America Transactions*, 15(11), 2017.

[5] Imam AA, Basri S, Ahmad R, Watada J, Almomani MA, Gonzlez-Aparicio MT. "Data modeling guidelines for NoSQL document-store databases". *International Journal of Advanced Computer Science and Applications*, *9*(10), 2018.

[6] Tjortjis C, ed. *Graph Databases Applications on Social Media Analytics and Smart Cities*. Boca Raton, CRC Press, 2023.

[7] Kùzu. Kùzu. http://kuzudb.com/

[8] Matyoqubov M, Saidov A, Kazakov O, Rustamova O. Enterprise Systems Data Integration. *2020 International Conference on Information Science and Communications Technologies (ICISCT)*, Tashkent, Uzbekistan, 2020.

[9] Zenkert J, Fathi M. "*Practice-Oriented Approaches for Information and Metadata Management in a Content Management System - Learnings from the Smart City Project LOKAL-Digital*", Cape Town, South Africa, 2023.

[10] Liu L, Özsu MT. *Encyclopedia of Database Systems*, Springer New York, 2018.

[11] Object Management Group. OMG Meta Object Facility core specification. https://www.omg.org/spec/MOF/2.5.1/PDF (October 2019).

[12] Li J, Luo G, Tian L. "Component-oriented executable meta-model for SaaS applications". *2011 International Conference on Cloud and Service Computing*, Hong Kong, China, 2011.

[13] Tozer G. *Metadata Management for Information Control and Business Success*. Boston, Artech House, 1999.

[14] Melton J, Simon AR. "*Understanding the New SQL: A Complete Guide*". San Francisco, CA, USA, 1993.

[15] Neo4j. NEO4J Graph Database & Analytics | Graph Database Management System. "Graph Database & Analytics". https://neo4j.com/ (July 31, 2024).

[16] MongoDB. MongoDB: the Developer Data platform. MongoDB. https://www.mongodb.com/.

[17] Redis - the real-time data platform. Redis. https://redis.io/ (August 12, 2024).

[18] PostgreSQL. PostgreSQL. https://www.postgresql.org/ (August 31, 2024).

Görüntü İşleme Yöntemi Kullanılarak Hedef Takibi: YOLO ve Haar-cascade Algoritmalarının Karşılaştırılması

Buğra GÜNAY[1], Taner ÇARKIT[1*]

[1]*Kırklareli Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği Bölümü, Kırklareli, Türkiye*

Özet

Bireysel güvenlik ve ulusal boyutlardaki güvenlik seviyelerinde stratejik hedeflerin takip edilmesi gerekmektedir. Toplum içinde suç ve suçlu takibinin yapılabilmesi için, ulusal/uluslararası boyutta stratejik hedeflere ulaşılabilmesi için hedef takip yöntemleri önem arz etmektedir. Bu yöntemlere insani etki veya çeşitli teknolojik olanaklar eklendiğinde hız ve kararlılık artmaktadır. Bu çalışmada belirlenen hedeflerin takibinin yapılabilmesi için görüntü işleme teknolojisi kullanılmaktadır. Hedef takibi esnasında; derin öğrenme algoritmasıyla (YOLO) ve Haar-cascade algoritması phyton yazılım dilinde ve OpenCV görüntü işleme kütüphanesi yardımıyla koşturulmaktadır. Nesne tespit elemanları kullanılarak kameradan elde edilen görüntülerin işlenmesi ile nesne ve yüz tespiti yapılmaktadır. Sonuçta, kullanılan algoritmalarının hedef takibine yönelik hız performansları karşılaştırılmaktadır.

Anahtar Kelimeler: *Görüntü işleme, hedef takibi, algoritma, YOLO, Haar-cascade*

Target Tracking Using Image Processing Method: Comparison of YOLO and Haar-cascade Algorithms

Abstract

Strategic targets need to be pursued at the individual security and national security levels. To track crime and criminals within the society and to achieve strategic goals at national/international level, goal tracking methods are important. Speed and stability increase by adding human influence or various technological possibilities to these methods. In this study, image processing technology is used to track the determined targets. During target tracking, the deep learning algorithm (YOLO) and the Haar-cascade algorithm are run in the Python software language with the help of the OpenCV image processing library. Object and face detection is done by processing the images obtained from the camera using object detection elements. As a result, the speed performances of the used algorithms for target tracking are compared.

Keywords: *Image processing, target tracking, algorithm, YOLO, Haar-cascade*

1 Giriş

Teknolojinin gelişimiyle beraber insan hayatının bazı alanlarında iyileşmeler olmaktadır. Artan insan popülasyonuna bağlı olarak, kullanılan ortak alanlar kalabalık olmaktadır. Bu yüzden, güvenlik teknolojilerinin iyileştirilmesine ihtiyaç duyulmaktadır. Popülasyonunun yoğun olduğu alanlarda, güvenlik zafiyeti yaratan olayların ve/veya kişilerin tespit edilmesi zor olmakta ve zaman almaktadır. Bu eksikliğin giderilebilmesi için nesneleri görüntü üzerinden tanımlamak ve sınıflandırmak önem arz etmektedir. Özellikle, hareketli nesnelerin incelenebilmesi ve teşhis edilebilmesi teknolojik gelişme bakımından çok önemlidir.

Kalabalık ortamlardaki kişi sayısının belirlenmesi, kişilerin giriş-çıkış saatlerinin tespit edilmesi, anormal hareketler sergileyen bireylerin tespiti, güvenlik zafiyeti yaratan kişilerin ve/veya stratejik hedeflerin belirlenmesi…vb. gibi teknolojik çıktılar güvenlik açısından yapılabilecek iyileştirmelerin en önemlileri arasındadır. Bu uygulamaların yapılabilmesi için kamaralar ile görüntüler

*Contact email: tanercarkit@klu.edu.tr

depolanarak, görüntü veri setinin bazı yazılım dilleri kullanılarak işlenmesi gerekmektedir. Görüntü işleme (Gİ) sonucunda sistem tarafından sunulan çıkarımlar yardımıyla istenilen müdahale sistemleri hedefe odağında otomatik olarak aktif hale getirilmektedir.

Literatürde hareket ve nesne takibi, hedef belirleme gibi amaçlarda yapılan farklı çalışmalar mevcuttur. Şöyle ki: Wu'nun 2009 yılında yaptığı çalışmada video gözetim sistemi tasarlanmıştır. Bu tasarım temelinde görüntüler işlenmiş ve nesne olarak görüntülerdeki vücutlar tespit edilmiştir. Tercih edilen algoritmalar sayesinde hareketteki tehlikeleri analiz edebilmiş ve bu analizler sonucu tehlikeli durumlarda hem sistemin kendi alarm sisteminin hem de güvenlik sistemindeki alarm sisteminin çalışması programlanmıştır [1]. Perihanoğlu'nun 2015 yılında yaptığı çalışmada; MATLAB bilgisayar programı kullanılarak dijital görüntü işleme (DGİ) fonksiyonları yardımıyla Gİ'ye yönelik uygulama arayüzü tasarlanmıştır. Nokta işleme teknikleriyle görüntüler hakkında bilgi sağlamakla birlikte kontrast ayarı, parlaklık ayarı, görüntü zenginleştirme amacıyla uygulanan filtreleme teknikleri, gri bölgelerdeki ani renk değişimlerinin önüne geçmek amacıyla kenar belirleme algoritmaları kullanılmıştır [2]. Yılmaz, 2016 yılında yaptığı çalışmada C# yazılım dili kullanarak; kameradan alınan görüntüleri openCV kütüphanesinde yer alan Gİ algoritmalarıyla görüntüye belirli filtreler uygulamış ve görüntüyü işlemeye uygun hale getirmiştir. Uygun hale getirilen görüntülerin üzerinde obje takibi yapılmıştır. Obje takibi için görüntü eşleme ve arama algoritmaları da kullanılmıştır. Elde edilen sonuçlarda, görüntü eşleştirme işlemi başarıyla gerçekleştirilmiş ve toplam arama süresi azaltılmıştır [3]. Eldem ve ark. 2017 yılında yaptıkları çalışmada; 100 ayrı insana ait yüz resminin bulunduğu bir veri tabanı oluşturmuşlardır. Daha sonrasında, kameradan alınan görüntüler üzerinde openCV Haar-cascade yapısını kullanarak yüz algılama işlemini gerçekleştirmişlerdir. Yüz algılamadan sonra, algılanan yüz ile veri tabanında bulunan yüzler karşılaştırılarak yüz tanıma uygulaması yapılmıştır. Uygulama %79 oranında başarılı sonuçlar sunmuştur [4]. Daş ve ark. tarafından 2019 yılında yaptıkları çalışmada; resim ve video'lardaki nesnelerin tanımlanması ve takip edilebilmesi için yapaya zeka (YZ) ve derin öğrenme (DÖ) teknolojileri kullanılmıştır. Veri seti oluşturularak sistemin eğitilmesi amaçlanmıştır. YZ ve DÖ kullanılmayan Gİ alanındaki çalışmalarına göre yüksek bir doğruluk elde edilmiştir [5]. Balogh vd. 2019 yaptıkları çalışmayla hareketi tespit etmek ve yüzleri tanıyabilmek amacıyla bir sistem geliştirmişlerdir. Raspberry Pi kart kullanılarak yapılan bu sistemde, Arduino karttan farklı olarak uygulama kodları da Raspberry Pi üzerinde yazılmıştır. Kullanılan veriyi kamera modülünden elde etmişlerdir. OpenCV ile kullanmak için Haar-cascade ve yerel ikili örüntü (YİÖ ve/veya LBP) algoritmalarıyla beraber sonuçları karşılaştırabilmek için birden fazla nesne tespit yöntemini ayrı ayrı kullanmışlardır. Farklı nesne tespit programları karşılaştırıldıktan sonra verimi arttırmak için Raspberry Pi kartın 4 çekirdeğini kullanmaya karar vermişlerdir. Hareket tespiti için yazılan program, bir hareket sensörü ile karşılaştırdığında; hareket detektörünün %85 olarak tespit ettiği hareketi kamera sistemi %96 olarak tespit etmiştir. Bu sonuçlar ile beraber programın hassas bir şekilde çalıştığını ortaya koymuşlardır [6]. Jhapate vd. 2020 yılında yaptıkları çalışmada kalabalık ortamlardaki hareketler ve tehlikeli durumlar tespit edilmeye çalışmışlardır. Bu amaç doğrultusunda; kalabalık ortamlardaki hareketlerin sensörler ile zor tespit edilebildiği ve meşakkatli bir biçimde ayrıştırılabildiği için Gİ tercih edilmiştir. Temel yöntem olarak; görüntüyü oluşturan karelerin kısa zamanda sık değişimlerinin oluşturduğu hareket etki haritasıyla sınıflandırma yapılmıştır. Tasarlanan sistem sonucunda tespitlerde %96.42 doğrula ulaşılmıştır [7]. Parveen vd.'nin 2021 yılında yaptıkları çalışmada; hareketli nesnelerin tespiti yapılmıştır. Bu tespit ile uygulamayı kullananların makine-bilgisayarlar etkileşiminin kolaylaştığına dikkat çekilmiştir. [8]. Mishra vd. tarafından 2022 yılında yapılan çalışmada; kameraya giren nesnenin algılanmasını ve algılama sonucunda algılanan nesnenin görüntüde gözükme süresinin hesaplanması amaçlanmıştır. Gerekli durumlarda görüntü kaydı yapılarak kullanıcıya SMS ile mesaj bilgisinin gönderiminin sağlanmıştır [9].

Bu çalışmada, Gİ tekniklerinden Haar-cascade algoritması ve derin öğrenme algoritmasıyla (YOLO) Phyton yazılım dili ve OpenCV görüntü işleme kütüphanesinde bulunan nesne tespit elemanları kullanılarak kameradan elde edilen görüntülerin işlenmesi ile nesne ve yüz tespiti yapılmaktadır. Tespit işlemi esnasında ve

sonrasında nesnenin hareketi takip edilmektedir. Ayrıca, kullanılan Gİ algoritmalarının hız performanslarının karşılaştırılmaktadır.

2 Görüntü İşleme

Görüntünün işlenmesi ile belirli çıkarımlar yapan ve bu çıkarımların sonucunda tasarlanan işlemleri gerçekleştiren birimler; görüntü işleme sistemleri (GİS) olarak isimlendirilmektedir [2]. Gİ genel anlamda; sağlık, savunma sanayi, oyun sektörü, ulaşım, denizcilik ve enerji gibi birçok alanda aktif olarak kullanılmasıyla birlikte Gİ'nin kullanıldığı bazı özel alanlar şunlardır:

- Görüntü iyileştirme,
- Otonom sürüş,
- Hava tahmini uygulamaları,
- Savunma radar sistemleri,
- Güvenlik sistemleri,
- Nesne tespiti, tanıma ve sınıflandırma,
- Sağlık alanında bazı hastalıkların tespiti,
- Ulaşım araçlarında korozyon tespiti,
- Uydu görüntüleriyle deprem analizi,
- Uydu üzerinden maden arama işlemleri,
- Kentleşme ve ormanlaşma durumlarının analizi,
- Tahribatsız muayene [10],
- Yüz ve vücut hareketleri ile koordineli olan oyun yapımları,
- Kimlik tespiti…vb.



Şekil 1. Nesne tespiti görüntüsü [5]

Gİ'de süreç; aşağıda verilen belirli bazı ana adımlar ile gerçekleştirilmektedir:

- Görüntünün elde edilmesi,
- Görüntünün işleme uygun hale getirilmesi için boyutlandırılması (Phyton, OpenCV Matlab…vb.),
- Görüntü üzerinde renk düzeltme işlemlerinin yapılması,
- Matris ve histogram verilerinin oluşturulması,
- Özellik çıkarma işleminin yapılması,
- Sınır belirlemek amacıyla köşe ve kenar kestirimi,
- Görüntü üzerinde belirli algoritmaların koşturulması,
- Analizler yapılarak sonuçların görselleştirilmesi ve yorumlanması.

3 Materyal ve Yöntem

Yapılan çalışmada nesne olarak tespit edilmesi istenen yüz ifadesi ".XML" uzantılı bir dosya yardımıyla tespit edilmektedir. Tespit edilmesi istenen hedef yüzün hareketi doğrultusunda hareketli bir takip sistemi görüntüdeki yüzü takip etmektedir. Örnek bir nesne tespit görüntüsü Şekil-1'de verilmektedir. Çalışma içerisinde tercih edilen yazılım dilinin kullanılan kütüphaneleri Tablo 1'de verilmektedir.

Tablo 1. Çalışma içerisinde kullanılan kütüphaneler.

Kütüphane İsmi	Açıklama
OpenCV	Görüntü işleme işlemleri için en çok kullanılan görüntü işleme kütüphanesidir.
Pyfirmata	Arduino ile Python yazlım dilinin haberleşmesine imkân sağlayan Python kütüphanesidir.
numpy	Genellikle dizi işlemlerini gerçekleştirmede ve matematik işlemlerinde kullanılan kütüphanedir.
Cvzone, yüz dedektörü	Bir yüz algılama kütüphanesidir.

Çalışma içerisinde tercih edilen openCV kütüphanesi nesne tespiti için aşağıdaki gibi bazı algoritmaları içinde bulundurmaktadır. Bunun yanında, bu çalışmada Haar-cascade sınıflandırıcı ve YOLO algoritması tercih edilmiştir.

- Haar-cascade,
- Yerel ikili örüntü (YİÖ/LBP),
- Öz yüzler,
- Fisher yüzleri,

- Derin öğrenme algoritması (YOLO),
- Hızlı bölge tabanlı evrişimsel sinir ağları,
- Tek atışlı detektörler [11].

3.1 Haar-cascade algoritması

Haar, bir algılama penceresinde belirli bir konumdaki komşu dikdörtgen bölgeleri dikkate alarak her bir bölgedeki piksel yoğunluklarını toplamaktadır. Daha sonra, bölge ve bu toplamlar arasındaki farkı hesaplamaktadır. Bu fark daha sonra bir görüntünün alt bölümlerini kategorize etmek için kullanılmaktadır [12]. Bu çalışmada ve yüz tespit işlemlerinde ise; genellikle göz çevresindeki alanlar yanaklara göre daha koyu renktedir. Bu nedenle, yüz algılamaya yönelik haar benzeri bir özelliğin bir örneği, göz ve yanak bölgelerinin üzerindeki iki komşu dikdörtgen alan kümesidir [13]. Haar-cascade yöntemi görüntüdeki nesnelerin tespit edilmesi gereken uygulamalarda sıklıkla kullanılmaktadır. Bu çalışmada; Haar-cascade algoritmasının kullanımı sonucunda elde edilen veriler Şekil 2'de görülmektedir. Haar-cascade'ın bazı özellikleri şunlardır [14]:

- Kenar Özelliği: Görüntüdeki alanların bir kısmı koyu renkli bir kısmı açık renkte ise kenar özelliği olduğunu belirtmektedir.
- Çizgi Özelliği: Görüntüdeki renkler açık-kapalı-açık renklerden oluşuyorsa çizgi özelliği vardır.
- Dört-Kare Özelliği: Koyu ve açık renkli alanlar birbirlerinin çaprazında bulunuyorsa bu durum dört kare özelliğini göstermektedir.

3.2 YOLO algoritması

YOLO algoritması, temelde görsel veriyi tek seferde bir sinir ağından geçirerek verideki objelerin koordinatlarını ve sınıfını tahmin etmektedir [15]. Bunu yaparken görseli kare ızgaralara ayırmaktadır. Izgara sayısının herhangi bir değer olabilmesinin yanında, n*n formatında olmalıdır. Görüntü ızgaralara ayrıldıktan sonra her bir ızgara kendi içerisinde bir obje olup olmadığını, obje var ise merkezinin kendi alanında olup olmadığını düşünmektedir. Görsel veri, sinir ağından geçtikten sonra [Pc, Bx, By, Bw, Bh, C1, C2] gibi vektör olarak bir çıktı oluşturmaktadır [16].

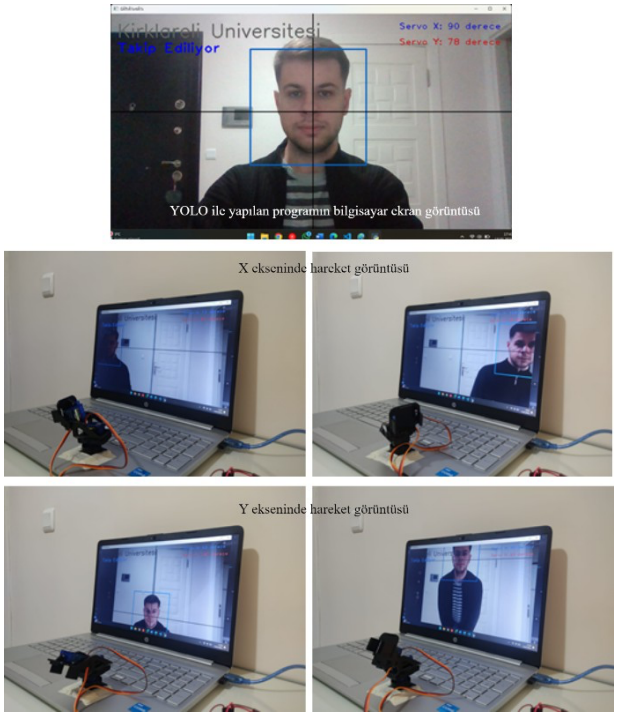
Denklemde: Pc, sınıf olasılığıdır. Görselde nesnenin olup olmamasına göre değerlendirme yapılmaktadır. Nesne var ise 1, nesne yok ise 0 değeri atanmaktadır. Bx, nesnenin merkez noktasının x koordinat değerlerini temsil etmektedir. By, nesnenin merkez noktasının y koordinat değerlerinin temsil etmektedir. Bw, nesneyi içine alan çerçevenin genişliğinin göstermektedir. Bh, nesneyi içine alan çerçevenin yüksekliğidir. C1, nesne sınıfını temsil etmektedir. C2, diğer nesne sınıfını temsil etmektedir. YOLO algoritması aşağıdaki işlem adımlarıyla çalışmaktadır:

- Algoritma, görsel veriyi ızgaralara ayırmaktadır.
- Her bir bölgedeki nesnelerin etrafına çerçeve çizilir.
- Her bir bölgede nesne bulunma olasılığı hesaplanır.
- Her bir çerçeve için güven skoru hesaplanır.

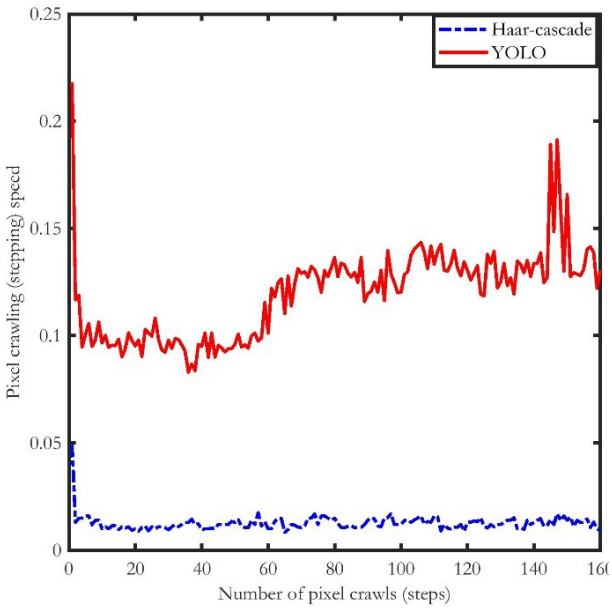
Algoritmalar arasında piksel tarama ve atlama hızı bakımından Şekil 4'teki gibi bir karşılaştırma yapıldığında; YOLO'nun daha hızlı tarama yaptığı sonucuna varılmaktadır. Öte taraftan 2022 yılında Palli ve Medimi'nin yapmış olduğu karşılaştırmada da YOLO algoritması daha doğru sonuçlar sunmaktadır [17].



Şekil 2. Haar-cascade ile yapılan programın bilgisayar ekranı görüntüsü, X ve Y eksenindeki hareketi



Şekil 3. YOLO ile yapılan programın bilgisayar ekranı görüntüsü, X ve Y eksenindeki hareketi



Şekil 4. YOLO ve Harr-cascade algoritmalarının hedef takip (piksel tarama ve atlama) hızı

4 Sonuçlar

YOLO ve Haar-cascade algoritmaları kullanılarak yapılan bu çalışmada, nesne takibi programlarının ikisinin de amaca hizmet ettiği görülmüştür. YOLO algoritmasının Haar-casade algoritmasına göre daha kararlı çalıştığı ve nesneleri daha tutarlı bir şekilde takip edebildiği görülmüştür. Gerçek zamanlı tespit konusunda YOLO algoritmasının daha kullanışlı olduğu sonucuna varılmıştır. Bu sonucun ortaya çıkmasındaki en büyük etken; YOLO algoritmasının, Haar-cascade algoritmasına göre koordinatları çok daha iyi bir şekilde tespit edebilmesidir. Aynı zamanda, YOLO algoritması maksimum olmayan bastırma algoritmasını kullanarak birden fazla çerçeve oluşmasını önlemektedir. Öte taraftan ise, Haar-cascade algoritmasında zaman zaman fazladan ve yanlış çerçeve işaretlemeleriyle karşılaşılmıştır. İki algoritmanın da farklı koşullar altında farklı doğruluk oranlarında çalıştığı tespit edilmiştir. En önemli koşullardan birisinin ortam ışığı olduğu görülmüştür. Haar-cascade yapısını etkileyen bir koşul olarak da karmaşık kıyafetlerin nesne tanımada olumsuz etkiye sebep olduğu görülmüştür.

Kaynaklar

[1] Wu Y. "Research on bank intelligent video image processing and monitoring control system based on opencv". IEEE 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication, Hong Kong, China, 20-22 August 2009.

[2] Perihanoğlu GM. Dijital Görüntü İşleme Teknikleri Kullanılarak Görüntülerden Detay Çıkarımı. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, İstanbul, Türkiye, 2015.

[3] Yılmaz İ. Görüntü İşleme Teknikleri Kullanılarak Renkli Obje Takibi Yapma. Lisans Tezi, Karabük Üniversitesi, Karabük, Türkiye, 2016.

[4] Eldem A., Eldem H., Palalı A. "Görüntü işleme teknikleriyle yüz algılama sistemi geliştirme". BEÜ Fen Bilimleri Dergisi, 6(2): 44-48, 2017.

[5] Daş R., Polat B., Tuna G. Derin öğrenme ile resim ve videolarda nesnelerin tanınması ve takibi. Fırat Üniversitesi Müh. Bil. Dergisi, 31(2): 571-581, 2019.

[6] Balogh Z., Magdin M., Molnár G. "Motion detection and face recognition using Raspberry Pi, as a part of, the ınternet of things". Acta Polytechnica Hungarica, 16(3): 167-185, 2019.

[7] Jhapate AK., Malviya S., Jhapate M. "Unusual crowd activity detection using opencv and motion ınfluence map". IEEE 2nd International Conference on Data, Engineering and Applications, Bhopal, India, 28-29 February 2020.

[8] Parveen S., Shah J. "A motion detection system in python and opencv". IEEE 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, Tirunelveli, India, 04-06 February 202.

[9] Mishra S., Verma V., Akhtar N., Chaturvedi S., Perwej Y.. "An intelligent motion detection using opencv". International Journal of Scientific Research in Science, Engineering and Technology, 9(2): 51-63, 2022.

[10] Çarkıt, T., Berberoğlu, M., Çarkıt, S. Applications of quantum technology in image processing: A review study on electric-electronic based detection and non-destructive testing methods. In book: New Applications in Digital Engineering, 1st Edition, Chapter: 4, Iksad Publications, pp. 69-87, 2023.
[11] Hasan RTH, Sallow AB. "Face detection and recognition using opencv. Journal of Soft Computing and Data Mining, 2(1): 86-97, 2021.
[12] Noori AM. ORB Özelliklerine Dayalı Yüz Tanıma Sistemi. Yüksek Lisans Tezi, Selçuk Üniversitesi, Konya, Türkiye, 2019.
[13] Sander Soo. Object detection using Haar-cascade Classifier. University of Tartu, Institute of Computer Science, 2014.
[14] Alkaz YE. "Python Opencv Yüz Tanıma Haarcascade Yöntemi". https://medium.com/@yunusemrealkaz/python-opencv-y%C3%BCz-tan%C4%B1ma-haar cascade-53cdf4c64c0e (Erişim Tarihi: 28.12.2023).
[15] Kutlu H. Derin Öğrenme Yöntemleri Kullanılarak Kalın Bağırsak Poliplerinin Tespiti Takibi ve Sınıflandırılması. Doktora Tezi, Fırat Üniversitesi, 2020.
[16] Akdağlı E. "YOLO Algoritması nedir?". https://ece-akdagli.medium.com/yolo-algoritmas%C4%B1-nedir-859862c7d5a6 (Erişim Tarihi: 03.01.2024).
[17] Palli AJMV., Medimi VS. A Comparative Study of YOLO and Haar Cascade Algorithm for Helmet and License Plate Detection of Motorcycles. Bachelor Thesis, Blekinge Institute of Technology, Karlskrona, İsveç, 2022.

Sensör Füzyonu ve Derin Öğrenme Teknikleri Kullanarak Hava Araçlarında Anomali Tespiti

Yıldırım ÇAT[1,2*], Prof. Dr. Kemal FİDANBOYLU[1**]

[1]*Bursa Uludağ Üniversitesi, Bilgisayar Mühendisliği Bölümü, Bursa, TÜRKİYE*

[2]*Turkish Aerospace Industry, TAI-Uludağ University R&D Center, Bursa, TÜRKİYE*

Özet

Havacılık endüstrisinde kullanılan bileşenler, belirli bir süre kullanıldıktan sonra performas kaybı yaşamakta ve bozulmaktadır. Bu bileşenler önceden planlanmış periyodik bakımlara tabi tutularak tekrar son kullanıcının kullanımına sunulur. Ancak, bileşenlerin çok sık bakıma alınması ve sürekli olarak değiştirilmesi, sistem performansında ve hizmet kalitesinde istenmeyen düşüşlere ve bekleme sürelerinin artmasına neden olmaktadır. Bu dezavantajları gidermek için Yapısal Sağlık İzleme (SHM) ve Kestirimci Bakım (PdM) teknikleri yaygın olarak kullanılmaktadır. Bu tekniklerde yaygın olarak kullanılan yaklaşım anomali tespitidir. Anomali tespitinde, bir sistemdeki genel kurala uymayan ve belirli bir konu hakkında bilgi sağlayan istisnai örneklerin bulunması hedeflenmektedir. Bileşenlerde zamanla meydana gelen ve istenmeyen arızalar, anomali tespiti yaklaşımları kullanılarak teşhis edilebilmektedir. Anomali tespitinde kullanılan yöntemlerden biri sensör füzyonudur. Sensör füzyonu, birden fazla sensörden gelen verilerin birleştirilmesi ve analiz edilmesi işlemidir. Havacılık sektöründe, sensör füzyonunun kullanılmasıyla hata payı daha düşük verilerin toplanarak yeni bir veri seti üretilebilir. Elde edilen füzyonlanmış veriler üzerinde yapılan anomali tespiti süreci de daha doğru sonuçlar üretmektedir. Bu süreçte, Kalman Filtresi veya Demspter-Shafer Teorisi yaygın olarak kullanılan sensör füzyonu teknikleri olarak öne çıkmaktadır.

Bu bildiride, NASA tarafından paylaşılan bir çalışmada, gürültüyü ortadan kaldırmak ve gürültülü sinyallerin zayıf özniteliklerini çıkarmak için rulmanlardan oluşturulan bir test düzeneğinden elde edilen zaman serisi verileri üzerinde anomali tespiti yapılmıştır. Geliştirilen sistemde rulmanlardan elde edilen ham titreşim verileri ön işleme tabi tutulduktan sonra her bir veri noktası kalman filtresine girdi olarak verilmiş ve yeni çıktı değerleri tahmin edilmiştir. Tahmin edilen veriler Autoencoder modelini beslemek için kullanılmıştır. Model eğitimi sürecinde optimizasyon algoritması olarak Gradyan İniş algoritmasının gelişmiş bir varyantı olan Adam algoritması kullanılmıştır. Eğitilen model kullanılarak eşik değer aşma tabanlı anomali tespiti yapılmıştır.

Anahtar Kelimeler: *Anomali Algılama, Yapısal Sağlık İzleme, Tahmini Bakım, Uçak, Havacılık Endüstrisi, Derin Öğrenme, Sensör Füzyonu, Kalman Filtresi, Demspter-Shafer Teorisi, Gradyan İniş Algoritması, Adam Algoritması*

Anomaly Detection in Aircraft Using Sensor Fusion and Deep Learning Techniques

Abstract

Components used in the aviation industry lose performance and deteriorate after a certain period of use. These components are subjected to pre-planned periodic maintenance and are made available to the end user again. However, frequent maintenance and constant replacement of components cause undesirable decreases in system performance and service quality and increased waiting times. To overcome these drawbacks, Structural Health Monitoring (SHM) and Predictive Maintenance (PdM) techniques are used. The commonly

*Contact email: 502331024@ogr.uludag.edu.tr

used approach in these techniques is anomaly detection. In anomaly detection, the aim is to find exceptional examples that do not comply with the general rule in a system and provide information about a specific subject. Faults that occur in components over time can be diagnosed using anomaly detection approaches. A common method used in anomaly detection is sensor fusion. Sensor fusion is the process of combining and analyzing data from multiple sensors. In the aviation industry, by using sensor fusion, a new data set can be produced by collecting data with a lower margin of error. The anomaly detection performed on the fused data obtained produces more accurate results. In this process, Kalman Filter or Dempster-Shafer Theory stand out as widely used sensor fusion techniques.

In this paper, in a study shared by NASA, anomaly detection was performed on time series data obtained from a test rig created from bearings to eliminate noise and extract weak features of noisy signals. In the developed system, after the raw vibration data obtained from the bearings were pre-processed, each data point was given as input to the Kalman filter and new output values were estimated. The predicted data was used to feed the Autoencoder model. During the model training process, the Adam algorithm, an advanced variant of the Gradient Descent algorithm, was used as the optimization algorithm. Threshold exceedance-based anomaly detection was performed using the trained model.

Keywords: *Anomaly Detection, Structural Health Monitoring, Predictive Maintenance, Aircraft, Aviation Industry, Deep Learning, Sensor Fusion, Kalman Filter, Demspter-Shafer Theory, Gradient Descent Algorithm, Adam Algorithm*

1 Giriş

Endüstride yararlanılan bileşenler, kullanıma bağlı olarak zamanla bozulmakta ve genellikle sabit zaman aralıklarında planlı bakıma tabi tutulmaktadırlar. Bu yaklaşımın dezavantajları; bileşenlerin çok sık değiştirilmesi, sağlıklı parçanın değiştirilmesi veya bir bileşenin arıza sonrası değiştirilmesidir. Bu dezavantajlar, verimliliğin ve ürün kalitesinin düşmesine veya sistemde istenmeyen bekleme sürelerine yol açmaktadır [1]. Sık gerçekleşen bakımlar veya doğru planlanmayan bakım süreçleri maliyet artışına sebep olmaktadır [2]. Sahip olduğu dezavantajlar nedeniyle bu yaklaşım yerini Yapısal Sağlık İzleme (Structural Health Monitoring, SHM) ve Kestirimci Bakım (Predictive Maintenance, PdM) uygulamalarına bırakmaktadır. SHM ve PdM uygulamalarında proaktif önlemlerin alınmasında kullanılan bir yaklaşım ise anomali tespitidir. Sistemdeki herhangi bir bileşende oluşabilecek anomali veya hatanın erkenden teşhis edilmesi, sistemin bakım planında önemli seviyede değişikliklere gidilmesine yol açabilir. Bu gibi bir senaryoda etkin bir anomali tespiti yönteminin kullanılması sistemin sürdürülebilirliği ve bakım planı için önem arz etmektedir.

Bileşenlerde meydana gelen anomaliler bileşenlerin zamanla bozulmalarına yol açar. Bu bozulmaların erken evrede belirlenmesi anomali tespiti yöntemleri ile gerçekleştirilmektedir. Anomali tespiti ile bir sistemdeki genel kurala uymayan ve belirli bir konu hakkında bilgi sağlayan istisnai örneklerin bulunması hedeflenmektedir [3]. Anomaliler üç başlıkta gruplandırılabilir; bir sistemde sadece tek başına bulunan Nokta Anomali, belirli bir bağlamda anomali olup diğer bağlamlarda anomali olarak sayılmayan, Bağlamsal Anomali ve grup halinde anomali olarak bulunan Toplu Anomali [4]. Literatürde bu anomali türlerinin tespiti için sensör füzyonu ile Makine Öğrenmesi (Machine Learning, ML) ve Derin Öğrenme (Deep Learning, DL) teknikleri kullanılmıştır.

Sensör füzyonu, ML ve DL yöntemlerinin detayları ve literatürdeki uygulamaları ikinci bölümde paylaşılacaktır. Aynı bölümde, bu çalışmada kullanılan ve havacılık alanında yapılan birçok çalışmada kullanılan Cincinnati Üniversitesi Akıllı Bakım Sistemleri Merkezi (University of Cincinnati Intelligent maintenance Systems, IMS) veri setinin ve tekniklerin açıklaması yapılacaktır. Üçüncü bölümde geliştirilen hibrit model sunulacaktır. Dördüncü bölümde yürütülen deneylerin sonuçları tartışıldıktan sonra son bölümde ise yapılan çalışmanın sonuçları sunulacaktır.

2 Literatür araştırması

2.1 Hava araçlarında anomali tespitinde kullanılan füzyon ve derin öğrenme teknikleri

Havacılık endüstrisinin hızla gelişmesine paralel olarak bu alanda kullanılan mekanik ve yazılımsal tüm bileşenler de daha kompleks hale gelmiştir. Bu

durumun bir sonucu olarak, bu alan üzerinde yapılan tüm araştırmalar ve deneyler daha girift hale gelmiştir. Bu çalışmanın da konusu olan ve havacılık endüstrisinde oldukça yaygın bir şekilde uygulanan anomali tespiti yöntemleri, özellikle sensör füzyonu ve DL tekniklerinin entegre edilmesiyle büyük oranda gelişim göstermiştir. Başlıca kullanılan sensör füzyonu teknikleri Dempster-Shafer teorisi ve Kalman filtresi olarak sıralanabilirken; Uzun Kısa Süreli Bellek (Long Short-Term Memory, LSTM), Otokodlayıcı (Autoencoder, AE) ve Evrişimsel Sinir Ağları (Convolutional Neural Networks, CNN) DL modelleri de anomali tespiti uygulamalarında ilk sıralarda tercih edilmektedir.

2.1.1 Dempster-Shafer teorisi (DST)

Dempster-Shafer teorisi (DST), Bayes teorisinin bir genellemesidir ve Bayes teorisinin eksikliklerini gidermek için geliştirilmiştir [5]. Teori, kısaca belirsizliği yönetmek, ML ve DL'de çoklu kaynaktan elde edilen bulguları kombine etmek için kullanılmaktadır. Her bir olasılık kümesine ağırlık atayarak belirsizliğin daha açıklayıcı bir temsilinin oluşmasını ve birbiriyle çatışan bulguların üstesinden gelinmesini sağlayarak belirsizliğin hakim olduğu ortamlarda entegre edildiği ML veya DL modelinin daha tutarlı ve daha güvenilir olmasını sağlamaktadır.

Garcia ve ark. [6], havacılık sistemlerinde meydana gelen karmaşık yapıdaki anomalilerin tespitinde kullanılan ve insan tarafında yorumlanabilirliği zayıflatma pahasına geliştirilen DL modellerine alternatif bir yaklaşım önermişlerdir. Yeni yaklaşımda sensör füzyonu icra edilerek karmaşık sınıflandırıcılara olan bağımlılığın azaltılması amaçlanmıştır. Geleneksel eşik-aşımı yöntemleriyle elde edilen yorumlanabilirlik seviyesi korunurken sıralı örüntüler ve bunların dağılımları kullanılarak yeni bir tahmin modeli üzerinde çalışılmıştır. Önerilen yaklaşım, topluluk öğrenmesi ile eğitilen sınıflandırıcıların tahminlerindeki belirsizliği minimize etmek için DST'yi kullanmaktadır. Birden fazla değişkenin kombinasyonu, belirli anomalilerin tespiti için dikkate alınmıştır ve elde edilen sonuçlar, DST tabanlı bu yeni yöntemin çoğunluğun gücü yaklaşımına göre daha iyi performansa sahip olduğu görülmüştür.

2.1.2 Uzun kısa süreli bellek (LSTM)

LSTM modeli, Yinelemeli Sinir Ağı'nın (Recurrent Neural Network, RNN) bir türüdür ve geleneksel ileri-beslemeli sinir ağlarından farklı olarak önceki zaman adımlarındaki bilgileri de hatırlayabilmektedir. Standart bir RNN ağı, çıktıyı üretirken mevcut adımdaki girdiyle beraber bir önceki adımda elde edilen çıktıyı da kullanır. Böylece, mevcut çıktı hesaplanırken önceki adımdaki bilgiler de göz önünde bulundurulur [7]. Ancak, RNN ağının bir eksikliği sadece yakın zamandaki geçmiş bilgileri tekrar kullanabilmesidir. Uzak geçmişteki bilgiler tamamen unutulmakta ve mevcut çıktının hesabında kullanılamamaktadır, yani diğer bir deyişle Kaybolan Gradyan Sorunuyla (Vanishing Gradient Problem) karşılaşılmaktadır [8]. LSTM, RNN'deki bu eksikliği gidermek amacıyla geliştirilmiştir ve hem kısa hem de uzun süreli geçmişteki bilgiler mevcut çıktının üretimi sürecinde kullanılır. Mimarisinde girdi katmanı, gizli katman ve çıktı katmanını bulunduran model, RNN'den ayrı olarak gizli katmanı içerisinde özel bir LSTM hücresi (LSTM Cell) içermektedir [9]. Üretilen çıktıların unutulma oranı, hangi yeni bilgilerin ekleneceği veya güncelleneceği ve saklanan bilgilerin hangi oranda kullanılacağı bu özel hücre ile belirlenir. Bu haliyle, kompleks veri setleriyle çalışıldığı durumlarda, standart RNN modelinden çok daha iyi tahminleme kabiliyetine sahiptir. Özellikle zaman serisi verilerinde, sıralı zaman noktalarının birbirleriyle ilişkileri, yapılacak tahmin için kritik seviyede öneme sahip olduğundan LSTM modeli yaygın olarak kullanılmaktadır [10]. Zaman serisi veri setleri de havacılık endüstrisinde anomali tespiti, Durum İzleme (Condition Monitoring) ve PdM uygulamalarında üretilmektedir.

2018 yılında NASA tarafından Mars yüzeyine gönderilen Curiosity isimli uzay araştırma aracından ve Soil Moisture Active Passive (SMAP) uydusundan elde edilen telemetri verilerini kullanarak bir gözetimsiz anomali tespiti çalışması yürütülmüştür [11]. Çalışmada kullanılan her iki veri setinde de uzmanlar tarafından etiketlenmiş anomali verileri yer almaktadır. Yüksek karmaşıklığa sahip zamansal veya sıralı veriler karşısında iyi sonuç vermesi ve geçmiş verilerle güncel veriler arasındaki ilişkiyi öğrenebilmesi nedeniyle çalışmada, LSTM ağları kullanılmıştır. Her bir telemetri kanalı için tekli modeller oluşturulmuş ve ilgili kanaldaki değerlerin tahmini yapılmıştır.

2.1.3 Evrişimsel sinir ağları (CNN)

Evrişimsel Sinir Ağları (Convolutional Neural Networks, CNNs) bir çeşit yapay sinir ağı olmakla

birlikte insandaki görme merkezinden esinlenilerek geliştirilmiştir [12]. Görüntü işleme alanında yaygın olarak tercih edilen bu model ile keskinleştirme, kenar tespiti ve bulanıklaştırma gibi birçok uygulama icra edilmekte ve yüksek performans elde edilmektedir. Model; Evrişim (Convolution), Havuzlama (Pooling) ve Tam Bağlantılı (Fully-Connected) olmak üzere üç katmandan oluşmaktadır. Girdi, Evrişim katmanında Denklem (1) ile filtreleme sürecine tabi tutulup sonraki katman olan Havuzlamaya gönderilir [13]. Bu katmanda sadece önemli özniteliklerin elde edilmesi ve maliyetin düşürülmesi amacıyla bir boyutsal indirgeme yapılmaktadır. Bu işlem sayesinde hesaba katılacak sinir ağı sayısı ve parametre sayısı azaldığı için aşırı öğrenme (overfitting) problemiyle karşılaşma olasılığı da düşürülmüş olur [14]. Tam Bağlantılı katmanda ise girdi Çok Katmanlı Algılayıcı (Multilayer Perceptron, MLP) ağına gönderilmek üzere sütun vektörü olarak düzleştirilir.

$$\begin{bmatrix} x_{1,1} & x_{1,2} \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} \cdots & x_{m,n} \end{bmatrix} * \begin{bmatrix} y_{1,1} & y_{1,2} \cdots & y_{1,n} \\ \vdots & \ddots & \vdots \\ y_{m,1} & y_{m,2} \cdots & y_{m,n} \end{bmatrix} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{i,j} y_{i,j} \quad (1)$$

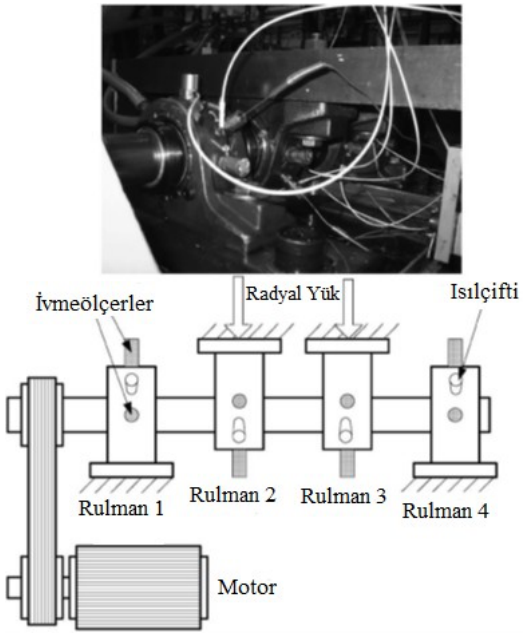
Wang ve ark. [15], rulmanlarda oluşabilecek anomalilerin önceden tespit edilebilmesi amacıyla ham titreşim sinyallerini görüntülere dönüştürmeyi önermişlerdir. Sinyal verilerini görüntüye dönüştürmede kullanılan bir yöntem olan Markov Transition Field (MTF) kullanılarak, titreşim sinyalleri görüntülere dönüştürüldükten sonra CNN ile özniteliklerin elde edilmesi ve anomali tespitinin yapılması amaçlanmıştır. Yöntemin performansını test etmek için Case Western Reserve University (CWRU) ve Mechanical Failure Prevention Technology (MFPT) veri setleri kullanılmıştır. İki veri seti ile yapılan deneylerin sonucunda yüksek doğruluk oranı elde edilmiştir.

2.2 Hava araçlarında anomali tespitinde kullanılan yöntemler

2.2.1 Veri Seti

Qiu ve ark. [16], sinyallerin gürültülerinin giderilmesi ve zayıf özniteliklerinin çıkarılması için yaptıkları çalışmada, geliştirdikleri Dalgacık Filtre (Wavelet Filter, WF) yönteminin performansını doğrulamak amacıyla rulmanlardan oluşan bir test düzeneği oluşturmuşlardır [16]. Oluşturulan deney düzeneğinde Şekil 1'de görüldüğü gibi bir mil üzerine dört adet çift sıralı rulman yerleştirilmiştir. Milin dönme hızı dakikada 2,000 devir olarak ayarlanmıştır. Mil üzerinde kurulan bir yay mekanizması ile toplamda yaklaşık 2,721.5 kilogramlık bir yük eklenmiştir. Rulmanların titreşim verileri 20 dakikalık aralıklarla kaydedilmiştir. Veriler 20 KHz frekansla örneklenmiş ve toplamda 20,480 veri noktası elde edilmiştir. Yapılan testlerle toplanan titreşim verileri kullanılarak üç farklı veri seti oluşturulmuştur. Bu çalışmada da kullanılan, ikinci testte toplanan ve toplamda 984 adet örnekten oluşan veri setinde rulman 1'de dış bilezik hatası gözlemlenmiştir.

Wang ve ark. [17], Liu ve ark. [18], Ho ve ark. [19] yürüttükleri çalışmalarda rulmanlarda meydana gelebilecek anomalilerin ve arızaların teşhisi için IMS veri setini kullanmışlardır.



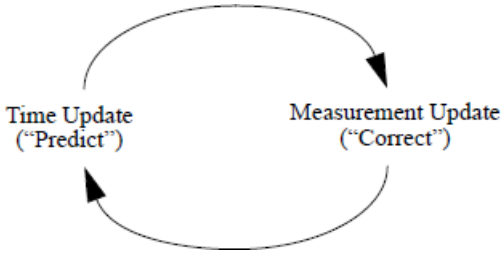
Şekil 1. Rulman test düzeneği [20]

2.2.2 Kalman filtresi

Kalman filtresi, dinamik sistemlerde zamanla değişen durumları izlemek ve tahmin etmek için geliştirmiştir. İlk olarak Rudolf E. Kálmán tarafından 1960 yılında tanıtılmıştır [21]. Kalman filtresi, sistemin durumunu modellemek için doğrusal bir dinamik sistem ve buna ek olarak gözlemler arasında gürültü bulunduğunu varsayar.

Temel olarak Şekil 2'de gösterildiği gibi iki aşamadan oluşur: Tahmin aşaması (Predict) ve Güncelleme (Correct) aşaması. Tahmin aşamasında, bir sonraki zaman adımı için bir önsel olasılık

(priori) tahmini elde etmek üzere mevcut durum ve hata kovaryans tahminleri zamanda ileriye doğru yöneltilir. Güncelleme adımlarında ise geri besleme işlemi yapılır.



Şekil 2. Kalman filtresi döngüsü [22]

Lee ve ark. [23], değişken kirlilik seviyelerinin varlığında bir hidrolik dişli pompasının kalan faydalı ömrünün (Remaining Useful Lide, RUL) tahminini içeren üç sensörlü bir füzyon tekniği önermiştir. Araştırmacılar, bu durumda üç farklı sensör veri akışından titreşim, akış hızı ve basınç sinyallerinden hata özelliklerini düzgün bir şekilde kaynaştırmak için Kalman filtresi tabanlı doğrusal bir model kullanmışlardır. Bu birleştirilmiş özellikler daha sonra Çift Yönlü LSTM (Bidirectional LSTM, BI-LSTM) ağı için bir girdi olarak kullanılmış ve güçlü bir RUL mimarisinin oluşturulmasıyla sonuçlanmıştır.

2.2.3 Gradyan iniş algoritması (GDA)

Gradyan İniş Algoritması (Gradient Descent Algorithm, GDA), türevlenebilir bir kayıp (maliyet) fonksiyonunun minimum değerini bulmak için geliştirilen yinelemeli bir yöntemdir [24]. Kayıp fonksiyonu genellikle parametreler (ağırlıklar) açısından türetilir ve bu türev (gradyan), fonksiyonun eğimini ve yönünü gösterir. GDA, parametrelerin değerlerini bu eğim boyunca adım adım ayarlayarak kayıp fonksiyonunu minimize etmeye çalışır. Denklem (2)'de algoritmanın standart matematik modeli verilmiştir. Burada $\theta_)$ ile mevcut parametre ifade edilirken; (θ), mevcut parametrenin kayıp fonksiyonunu temsil etmektedir. Gradyan alma işlemi nabla operatörü ∇ ile temsil edilmiş, genellikle 0-1 aralığında olan α ile öğrenme oranı temsil edilirken tüm bu işlemin sonucu da bir sonraki adımdaki parametreye yani, $\theta_{)*!}$ ifadesine eşitlenmiştir. Kayıp fonksiyonu minimize edilinceye kadar veya sistemde tanımlı belirli bir durdurma kriterine, maksimum iterasyon sayısına ulaşılması veya eşik değerinin aşılması, karşılaşılana kadar bu denklem işletilmeye devam edilir.

$$\theta_{t+1} = \theta_t - \alpha \nabla J(\theta) \qquad (2)$$

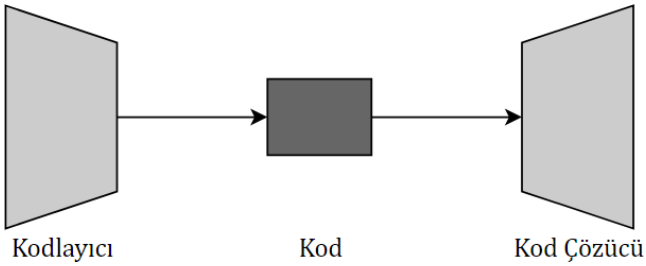
Öğrenme oranının manipüle edilmesiyle GDA'nın daha iyi sonuç veren varyantları geliştirilmiştir. DL uygulamalarında yaygın bir şekilde kullanılan Adam (Adaptive Moment Estimation) algoritması da bunlardan bir tanesidir. İlk olarak 2015 yılında geliştirilen bu algoritma [25], özellikle gradyan ve karesel gradyanın üstel bir hareketli ortalamasını hesaplar ve tıpkı α gibi iki farklı parametre ile bu hareketli ortalamaların bozunma oranlarını kontrol eder. DL uygulamalarında Adam algoritması diğer stokastik optimizasyon algoritmalarına göre daha yüksek performansa sahiptir ve daha iyi sonuçlar üretmektedir.

2.2.4 Otokodlayıcı (AE) modeli

AE, ilk defa Kramer tarafından tanıtılmış ve temelde bir yapay sinir ağıdır [26]. Geleneksel bir AE Şekil 3'te de görülebileceği üzere üç katmandan meydana gelmektedir: Kodlayıcı (Encoder), Kod veya Gizli Alan (Code, Latent Space) ve Kod Çözücü (Decoder). Kodlayıcı katmanında işletilen gizli fonksiyonla girdi üzerinde bir dönüşüm uygulanır ve sıkıştırılarak Gizli Alan katmanına aktarılır. Bu katmanda, sinir ağları kullanılarak girdiye olabildiğince benzer çıktı tahmin edilir. Sonrasında Kod Çözücü katmanının icra ettiği dönüşüm ile girdinin Yeniden Oluşturma (Reconstruction) süreci başlatılır. Bu süreç boyunca sıkıştırılan veriler eski boyutuna geri döndürülür. İdeal performans düzeyinde çalışan bir AE için girdi verileri ile çıktı verileri birbirine oldukça yüksek düzeyde benzemelidir. AE modelinin yaygın olarak kullanıldığı uygulamalardan birisi anomali tespitidir. AE'nin Kodlayıcı katmanına verilen girdi ile Kod Çözücü katmanının ürettiği çıktı arasındaki farka Yeniden Oluşturma Hatası (Reconstruction Error, RE) adı verilmektedir. Optimal çalışan bir AE'ye girdi olarak verilen normal (anomali olmayan) verinin girdi ve çıktı değerleri arasındaki RE değeri minimum seviyelerdeyken; girdi olarak verilen anomali verisinin girdi ve çıktı arasındaki RE değeri daha yüksek seviyelerde olmaktadır.

Bu anomali tespiti yönteminin matematiksel modelinde [27], Denklem (3)'te d fonksiyonu ile bahsedilen RE'nin derecesi, x ile girdi ve x' ile de yeniden oluşturulan x çıktısı temsil edilmektedir.

$$\mathrm{d}(x, x') \qquad (3)$$



Şekil 3. AE mimarisi

Denklem (4)'teki gibi bir kayıp fonksiyonu tanımlandığında: $\mathbb{E}_{x\sim\mu_{ref}}$ ifadesi, başlangıçta AE'nin yapması gereken süreci temsil etmek üzere girdi kümesi için bir referans olasılıksal dağılımı temsil ederken; $D_\theta(E_\emptyset(x)))$ ifadesi, Kod Çözücü'nün yaptığı yeniden oluşturma sürecini ifade eder. Burada, $E_\emptyset(x)$ her x girdisi için $\emptyset$ ve θ parametrelerine bağlı olarak Kodlayıcıyı temsil etmektedir.

$$L(\theta,\emptyset) := \mathbb{E}_{x\sim\mu_{ref}}[d(x, D_\theta(E_\emptyset(x)))] \quad (4)$$

İdeal AE için Denklem (5)'teki gibi bir optimizasyon süreci icra edilir. Burada, θ ve $\emptyset$ parametrelerine bağlı olan AE'nin kullandığı kayıp fonksiyonun çıktıyı en aza düşürmesi istenen durumdur. Yani, bu işlem gradyan iniş algoritması ile yapılmaktadır [28].

$$argmin_{\theta,\emptyset}\, L(\theta,\emptyset) \quad (5)$$

Singh ve ark. [29] dişli çarklarda bulunan rulmanların RUL değerlerini tahmin edebilmek için AE kullanarak gözetimsiz anomali tespiti çalışması yapmışlardır. NASA IMS veri setinde anomali etiketli verilerin olmamasından dolayı AE modeli seçilmiştir. Eğitimde oluşan maksimum hata eşik değeri olarak seçilmiştir. Deney sonuçlarına göre, sadece bir gizli katman kullanılarak oluşturulan AE modeli ile anomali tespitinde başarılı sonuçlar elde edilmiştir.

Ahmad ve ark. [30], döner mekanizmalardaki anomalilerin tespiti ve durumlarının izlenmesi için LSTM Tabanlı AE (LSTM-Based Autoencoder) modeli geliştirmişlerdir. Çalışmada NASA IMS veri seti ve 5 farklı rulman verisi içeren Endüstri (Industry) veri seti kullanılmıştır. Anomali tespit aşamasında, normal verilerden elde edilen RE değerlerinin anomali skorlarına dönüştürülmesiyle eşik değeri hesaplanmıştır. Yapılan deneyler sonucunda, her iki veri setindeki anomaliler de doğru bir şekilde tespit edilmiştir.

Li ve ark. [31], büyük endüstriyel mekanik ekipmanlardaki oluşabilecek anomalileri tespit etmek amacıyla AE ve Zamansal Evrişimsel Ağ (Temporal Convolutional Network, TCN)'ı kullanan yeni bir gözetimsiz yaklaşım önermişlerdir. AE sinyallerin RE değerini hesap etmek için kullanılırken; anomali verilerinin az olması problemini bertaraf edebilmek için TCN kullanılmış ve zaman serisi sinyalleri tahmin edilmiştir. Geliştirilen sistemin etkinliğini ölçebilmek için IMS veri seti üzerinde yapılan deneyler sonucunda Test 2 verileri için elde edilen metrikler geliştirilen yaklaşımın başarısını doğrulamıştır.

3 Uygulanan yöntem

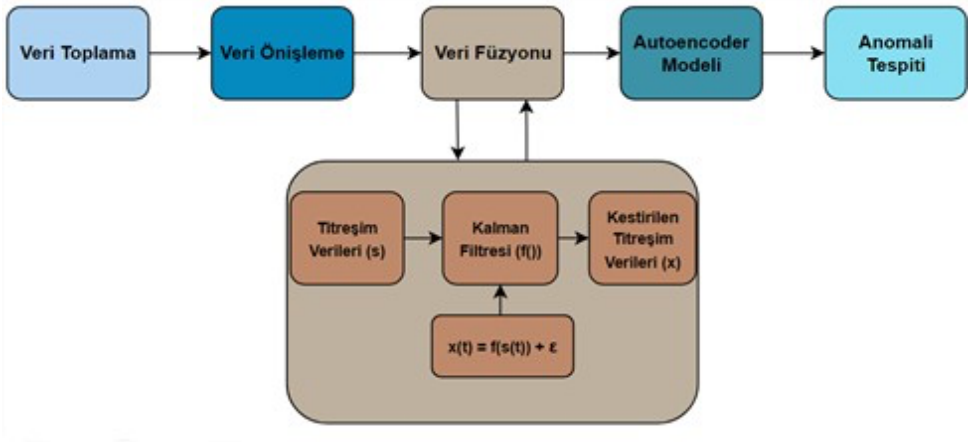
Bu çalışmada geliştirilen yöntemin kuş bakışı bir blok diyagramı Şekil 4'te verilmiştir. Yöntemin blok diyagramında görüldüğü üzere, ilk adımda Şekil 5'te görseli verilen NASA rulman titreşim sinyal verileri girdi olarak alınmıştır. Veriler normalize edilerek bir ön işlemeye tabi tutulduktan sonra veri füzyonu adımına geçilmiştir. Bu adımda, titreşim verileri (s), temel sensör füzyonu bileşenlerinden sensörleri veya verileri temsil etmektedir. Sistem modeli Denklem (6)'daki gibi ifade edilmiştir.

$$x(t) = f(s(t)) + \varepsilon \quad (6)$$

Burada $x(t)$, ilgilenilen parametreleri temsil ederken; ε, gürültüyü ifade etmektedir. Öte yandan $s(t)$, temel sensör füzyonu bileşenlerinden sensörleri veya verileri temsil etmektedir. $f(s(t))$ terimiyle de girdi olarak alınan ön işlenmiş sensör sinyallerinin kalman filtresine tabi tutulduğu temsil edilmektedir. Her bir sinyal verisi kalman filtresi algoritmasına girdi olarak verilmiş ve yeni değerleri tahmin edilmiştir. Bu tahmin edilen yeni sinyal değerleri de AE'yi beslemektedir.

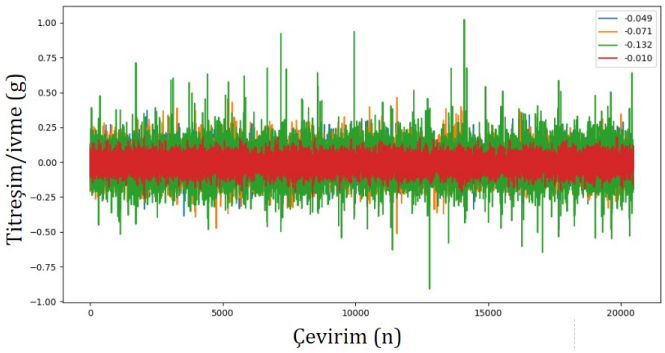
Bir sonraki adımda kalman filtresiyle yeniden kestirimi yapılan sinyaller AE modeline girdi olarak verilmiştir. Veriler, bir zaman serisi verileri olduğu için rastgele eğitim ve test veri kümelerine ayrılmamış ve ilk 445 veri eğitime, geri kalan 538 veri ise teste ayrılmıştır.

Modelin eğitimi aşamasında başarım oranının ölçülmesi için kullanılan metrik kayıp fonksiyonudur. Anomali tespiti uygulamalarında genel olarak ortalama kare hatası (mean squared



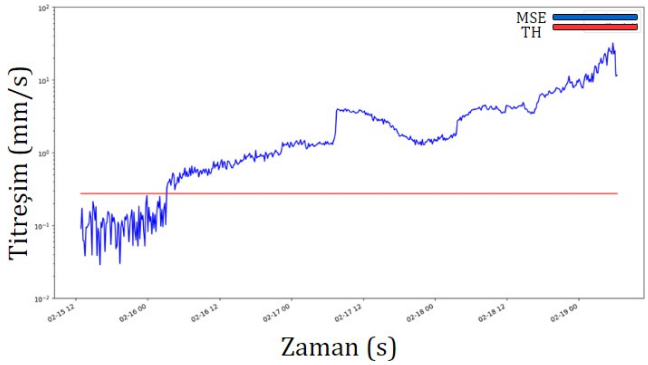
Şekil 4. Sistem blok diyagramı

error, MSE) kullanıldığı için bu çalışmada da MSE tercih edildi. Kayıp fonksiyonunun minimize edilmesi ve diğer model parametrelerinin optimizasyonu için önceden de bahsedilen Adam optimizasyon algoritması çalıştırılmıştır. Algoritmanının öğrenme oranı 0.0001 olarak ayarlanmıştır. Algoritmanın durdurma kriteri olarak da maksimum iterasyon sayısı tercih edilmiş ve 100 olarak belirlenmiştir.



Şekil 5. Rulman titreşim verisi

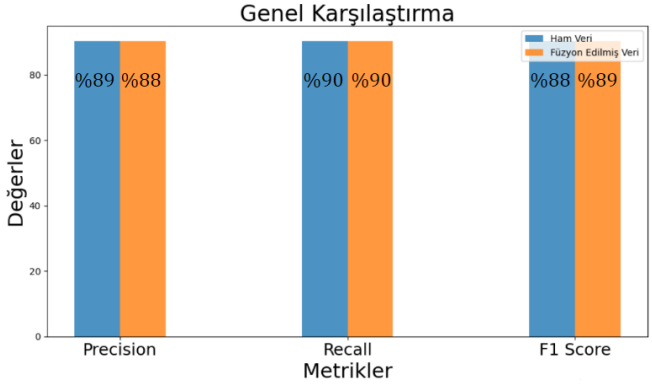
Son adımda, belirlenen parametreler ve durdurma kriteriyle eğitilen AE modeli ile test veri kümesi üzerinde anomali tespiti süreci icra edilmiştir. Bu süreçte, iki farklı deney yürütülmüştür. Yürütülen iki farklı deneyin ilkinde sensör füzyonu adımı pas geçilmiş ve rulman verileri ön işleme tabi tutulduktan sonra direkt olarak AE modeline girdi olarak verilmiştir. İkinci deneyde ise sensör füzyonu adımı sisteme dahil edilmiş ve Kalman filtresiyle ham sinyal verileri üzerinde kestirim yapılmıştır. Uzmanlar tarafından işaretlenen anomali verileri baz alınarak belirlenen eşik değeri kullanılmıştır. İvme büyüklüğü bu eşik değerin üzerinde olan veriler anomali olarak etiketlenmiştir. Şekil 6'da anomali tespiti sürecinin görsel temsili yer almaktadır.



Şekil 6. Anomali tespiti süreci

4 Araştırma bulguları

Deneylerdeki başarım oranlarını ölçebilmek için yaygın olarak kullanılan performans metrikleri seçilmiştir. Bunlar: Doğruluk (Accuracy), Hata (Loss), Kesinlik (Precision), Geri Çağırma (Recall), ROC Eğrisi Altında Kalan Alan (Area Under the Curve Receiver Operating Characteristics) ve F1 Skordur [33]. Elde edilen sonuçlar Şekil 7'de gösterilmiştir. Birinci deneyde sırasıyla kesinlik %89, geri çağırma %90, F1-skor %88; ikinci deneyde sırasıyla kesinlik %88, geri çağırma %90 ve F1-skor %89 olarak hesaplanmıştır.



Şekil 7. Elde edilen metriklerin karşılaştırılması

5 Sonuçlar

Hava araçlarının bakımının ve bileşenlerin sürdürülebilirliğinin sağlanmasında anomali tespiti yaygın kullanılan bir uygulamadır. Günümüzde daha da karmaşık hale gelen havacılık sistemlerinde meydana gelen anomali ve hataların tespiti için geleneksel yöntemler yetersiz kalmaya başlamamıştır. Çoklu sensör füzyonu tekniklerinin icrası ile daha etkin anomali tespiti uygulamaları geliştirilebilmektedir. Bu çalışmada da yaygın olarak kullanılan Kalman filtresi kullanılarak ham sinyal veriler üzerinde gürültülerin azaltılmasıyla yeni tahmin verileri elde edilmiştir. Bu tahmin verileri kullanılarak oldukça yaygın kullanılan AE modeli, tahmin verilerini girdi olarak kullanarak anomali tespiti sürecini icra etmiştir. Bu süreçte parametrelerin ve kayıp fonksiyonunun optimizasyonu için GDA'nın ileri düzey bir versiyonu olan Adam algoritmasına başvurulmuştur. Elde edilen sonuçlara göre kesinlik, geri çağırma ve F1-skorun birbirine çok yakın çıktığı gözlemlenmiştir. Gelecek çalışmalarda farklı kestirim algoritmaları, farklı model hiperparametreleri veya farklı veri kümeleri üzerinde deneyler yapılması planlanmaktadır.

Kaynaklar

[1] Birgelen VA, Buratti D, Mager J, Niggemann O, "Self-organizing maps for anomaly localization and predictive maintenance in cyber-physical production systems". *Procedia CIRP*, 72, 480-485, 2018.

[2] Tian R, Liboni L, Capretz M. "Anomaly Detection with Convolutional Autoencoder for Predictive Maintenance". *9th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, Toronto, ON, Canada, 241-245, 2022.

[3] Alpaydin E. *Introduction to Machine Learning*. Cambridge, MA: MIT Press, 2014.

[4] Nassif AB, Talib MA, Nasir Q, Dakalbab FM. "Machine Learning for Anomaly Detection: A Systematic Review". *IEEE Access*, 9, 78658-78700, 2021.

[5] Denźux T. "40 years of Dempster-Shafer theory". *International Journal of Approximate Reasoning*, 1, 79(C), 1-6, 2016.

[6] Garcia EJ, Beres SL, Mulvihill ML, Stephens CL, Napoli NJ. "Multiclass Flight Anomaly Detection Using Sensor Fusion Based on Dempster-Shafer Theory". *InAIAA Sci. Tech Forum*, 2024.

[7] Bengio Y, Simard P, Frasconi P. "Learning long-term dependencies with gradient descent is difficult". *IEEE Transactions on Neural Networks*, 5(2), 157-66, 1994.

[8] Graves A, Liwicki M, Fernández S, Bertolami R, Bunke H, Schmidhuber J. "A novel connectionist system for unconstrained handwriting recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, 31(5), 855-68, 2008.

[9] Schmidhuber J, Hochreiter S. "Long short-term memory". *Neural Comput.*, 15, 9(8), 1735-80, 1997.

[10] Korstanje J. "Lstm rnns". *InAdvanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR,* 243-251, 2021.

[11] Hundman K, Constantinou V, Laporte C, Colwell I, Soderstrom T. "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding" *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Jul 19, 387-395, 2018.

[12] Fukushima K. Neocognitron: "A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". *Biological Cybernetics*, 36(4), 193-202, 1980.

[13] Michelucci U, Michelucci U. "Fundamentals of convolutional neural networks". *Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection*, 79-123, 2019.

[14] Zeiler MD, Fergus R. "Stochastic pooling for regularization of deep convolutional neural networks". *arXiv preprint*, 1301, 3557, 2013.

[15] Wang M, Wang W, Zhang X, Iu HH. "A new fault diagnosis of rolling bearing based on Markov transition field and CNN". *Entropy*, 24(6):751, 2022.

[16] Qiu H, Lee J, Lin J, Yu G. "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics". *Journal of Sound and Vibration*, 289, (4-5), 1066-90, 2006.

[17] Wang Q, Wang S, Wei B, Chen W, Zhang Y. "Weighted K-NN classification method of bearings fault diagnosis with multi-dimensional sensitive features". *IEEE Access*, 9, 45428-40, 2021.

[18] Liu C, Gryllias K. "A semi-supervised Support Vector Data Description-based fault detection method for rolling element bearings based on cyclic spectral analysis". *Mechanical Systems and Signal Processing*, 140, 106682, 2020.

[19] Ho SK, Nedunuri HC, Balachandran W, Kanfoud J, Gan TH. "Monitoring of industrial machine using a novel blind feature extraction approach". *Applied Sciences*, 22, 11(13), 5792, 2021.

[20] Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services. IMS, University of Cincinnati. "Bearing Data Set", NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2007.

[21] Kalman, RE. "A new approach to linear filtering and prediction problems.". *J. Basic Eng.*, 82(1), 35-45, 1960.

[22] Welch, G, Bishop, G. "An introduction to the Kalman filter". 1995.

[23] Lee MS, Shifat TA, Hur JW. "Kalman filter assisted deep feature learning for RUL prediction of

hydraulic gear pump". *IEEE Sensors Journal*, 22(11):11088-97, 2022.
[24] Andrychowicz M, Denil M, Gomez S, Hoffman MW, Pfau D, Schaul T, Shillingford B, De Freitas N. "Learning to learn by gradient descent by gradient descent". *Advances in Neural Information Processing Systems*, 29, 2016.
[25] Kinga D, Adam JB. "A method for stochastic optimization". *InInternational Conference on Learning Representations (ICLR)* May 7, 5, 6, 2015.
[26] Kramer, MA. "Nonlinear principal component analysis using autoassociative neural networks". *AIChE Journal*, 37(2), 233-243, 1991.
[27] Wikipedia contributors. "Autoencoder." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia.https://en.wikipedia.org/wiki/Autoe ncoder. Erişim tarihi (3.5.2023).
[28] Nguyen TV, Wong RK, Hegde C. "On the dynamics of gradient descent for autoencoders". *InThe 22nd International Conference on Artificial Intelligence and Statistics*, Apr 11 (2858-2867), PMLR, 2019.
[29] Singh S, Shiv P, Ahmed A. "Outlier modeling in gear bearing using autoencoder for remaining useful life prediction". *Preprints*, 2019.
[30] Ahmad, S, Styp-Rekowski, K, Nedelkoski, S, Kao, O. "Autoencoder-based condition monitoring and anomaly detection method for rotating machines". *In 2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA, 4093-4102, 10-13 December, 2020.
[31] Li Z, Sun Y, Yang L, Zhao Z, Chen X. "Unsupervised machine anomaly detection using autoencoder and temporal convolutional network". *IEEE Transactions on Instrumentation and Measurement*, 71, 1-3, 2022.
[32] Raschka S. *Python Machine Learning*. Packt Publishing, 2015.

Su Tüketimi Trendlerini Anlamak: R Shiny Uygulaması ile Veri Analizi ve Görselleştirme

Ahmet Albayrak[1]*, Zeynep Baran[1], Metin Kaynaklı[2]

[1]*Düzce Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Düzce, TÜRKİYE*
[2]*Bitlis Eren Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Bitlis, TÜRKİYE*

Özet

Dünya nüfusunun artmasıyla birlikte, yakın gelecekte karşılaşabileceğimiz en önemli sorunlardan biri su kıtlığıdır. Geçmişten günümüze insanlar, ticari (endüstriyel), evsel ve resmî kurumlar gibi farklı tüketim grupları olarak suya ihtiyaç duymuştur. Günümüzde ise değişen güvenlik algısıyla birlikte suyun insan yaşamındaki önemi artmış, toplumların sürdürülebilirliği ve kalkınması açısından kritik bir kaynak haline gelmiştir. Bu çalışmada bireylerin ve ailelerin su tüketim alışkanlıklarını analiz edebilecekleri ve kendi kendilerine su tüketimini yönetebilmeleri için R Shiny web uygulaması geliştirilmiştir. Web uygulaması kullanıcılara su tüketim alışkanlıklarını analiz etmek için 13 soru sormaktadır. Soruların cevaplanmasının ardından veriler analiz edilerek su tüketim bilinci hesaplanmakta ve kullanıcıya su tüketim bilincini artırmak amacıyla tavsiyeler sunmaktadır. Bu çalışma için Shiny ortamının seçilmesinin nedeni R ile makine öğrenmesi algoritmalarının ve veri analiz araçlarının kolaylıkla kullanılmasına imkân vermesidir. Uygulama geliştirildikten sonra deneysel çalışmalar yapılmış ve 499 kişi ile uygulama test edilmiştir. Deneysel çalışmalara göre bu çalışma için en iyi başarımı SVM modeli f1-skoru %91,9 ile vermiştir.

Anahtar Kelimeler: *Veri Bilimi, Su Tüketimi, R Dili, Shiny Ortamı, Makine Öğrenmesi, Veri Analizi*

Understanding Water Consumption Trends: Data Analysis and Visualization with R Shiny Application

Abstract

As the world population increases, one of the most important problems we may face soon is water scarcity. From past to present, people have needed water in different consumption groups such as commercial (industrial), domestic and official institutions. Today, with the changing perception of security, the importance of water in human life has increased and it has become a critical resource for the sustainability and development of societies. In this study, the R Shiny web application was developed so that individuals and families can analyze their water consumption habits and manage their own water consumption. The web application asks users 13 questions to analyze their water consumption habits. After the questions are answered, the data is analyzed, and water consumption awareness is calculated, and recommendations are offered to the user to increase water consumption awareness. The reason why Shiny environment was chosen for this study is that it allows easy use of machine learning algorithms and data analysis tools with R. After the application was developed, experimental studies were conducted, and the application was tested with 499 people. According to experimental studies, the SVM model gave the best performance for this study with an f1-score of 91.9%.

Keywords: *Data Science, Water Consumption, R Language, Shiny Environment, Machine Learning, Data Analysis*

1 Giriş

Su, gezegenimizdeki yaşamın olmazsa olmaz bir unsuru, tüm canlılar için hayati bir kaynak. İnsanlar, hayvanlar ve bitkiler, hayatta kalmak ve gelişmek için suya ihtiyaç duyarlar. İnsan vücudunun büyük bir kısmı sudan oluşur ve su, vücut ısısını düzenlemek, hücrelere besin taşımak, atık ürünleri temizlemek, eklemleri yağlamak ve organları korumak gibi birçok hayati vücut fonksiyonunu

* Contact email: ahmetalbayrak@duzce.edu.tr

yerine getirir. Ancak suyun önemi sadece sağlık açısından değil, aynı zamanda gezegenin ekolojik dengesi, tarım, sanayi ve enerji üretimi gibi birçok alanda da büyük önem taşımaktadır.

Bitkiler fotosentez yapmak ve büyümek için suya ihtiyaç duyarlar. Hayvanlar içmek, yıkanmak ve besinlerini sindirmek için suya ihtiyaç duyarlar. Tarım, ürün yetiştirmek ve hayvancılık yapmak için suya ihtiyaç duyar. Sanayi, üretim süreçlerinde su kullanır. Elektrik santralleri, enerji üretmek için suya ihtiyaç duyarlar [1].

Ancak, dünya nüfusunun artması, iklim değişikliği ve su kaynaklarının kirlenmesi gibi faktörler, su kıtlığını giderek artan bir sorun haline getirmektedir. Su kıtlığı, tarım üretimini azaltır, gıda güvenliğini tehdit eder, insan sağlığını olumsuz etkiler ve çatışmalara yol açabilir. Su kaynaklarının sürdürülebilirliği, gezegenin geleceği için kritik öneme sahiptir. Su kaynaklarını korumak, suyu tasarruflu kullanmak ve su kirliliğini önlemek hem devletlerin hem de bireysel olarak herkesin sorumluluğundadır [2].

Veri biliminde veri analizi ve görselleştirme, bilinçlendirme ve etkili karar alma süreçleri için büyük önem taşımaktadır. R dili, istatistiksel analiz ve veri görselleştirme alanında güçlü bir araç olarak bilinmektedir. R'nin açık kaynak kodlu olması, geniş bir kullanıcı topluluğuna sahip olması ve sürekli gelişen paket koleksiyonu, onu veri analizi için ideal bir platform haline getirmektedir. Shiny ise R dilini kullanarak etkileşimli web uygulamaları oluşturmak için kullanılabilen bir pakettir. Shiny ile veri görselleştirmelerini interaktif hale getirebilir, kullanıcıların verilerle etkileşim kurmasını sağlayabilir ve böylece daha derin bir anlayış sağlanabilir.

2 Materyal ve Metot

Yerleşim alanlarında, sürdürülebilir kalkınma için su yönetiminin verimli bir şekilde uygulanabilmesi adına, su kullanıcılarının tercihlerinin de göz önünde bulundurulması gerekmektedir [4]. Bu tercihleri anlamak için yapılan anket çalışmaları, suyla ilgili çevresel sorunları gözlemleyerek, su kullanıcılarının bakış açısına göre değerlendirme imkanı sağlar [3]. Halkın algısını belirlemek için niteliksel ve niceliksel analizler, hedef grup analizleri, anket analizleri, kitle iletişim içerik analizleri ve davranışsal indikatör analizleri gibi çeşitli metotlar kullanılabilir [5]. Ayrıca, bireyler veya kurumların su yönetimini iyileştirmek için su israfını analiz eden bir endeks değeri oluşturulabilir [6-7]. Su kullanım algısını değiştirme ve sosyolojik durumu iyileştirme amaçlı olarak uygulama geliştirilmiştir. Uygulama aşağıda verilenleri yerine getirmektedir.

- Bir ailede su tüketimini etkileyen faktörler belirlenmelidir. Hane halkının sosyoekonomik ve demografik durumuna bakılarak, su israfını azaltmaya yönelik tavsiyeler üretecektir.
- Kullanıcılar, web sayfasında su tüketim indekslerini görebilecektir. Bu sayede su tüketim ortamlarının hangi koşullara bağlı olduğu belirlenerek su israfının analizi yapılabilecektir. Bu analizlerin çıktıları grafiksel ve sayısal karşılaştırmalar şeklinde gözlemlenebilecektir.
- Bu uygulama ile kullanıcılar, su tüketim davranışlarını analiz edebilmelidir [8]. Web sayfası, sürdürülebilir ve yenilikçi kentler oluşturma amacıyla su kaynaklarının tedarikinde doğadan en fazla su çekme yerine, doğaya en fazla su bırakma amacını benimseyerek kullanıcıya aylık grafiksel raporlar sunacaktır.
- Kişilerin veya kurumların içinde bulundukları koşullar göz önüne alınarak, kaliteli su tüketimi için kullanıcıların su kullanım memnuniyet dereceleri ölçülebilecektir.
- Kişiler veya kurumlar, aylık olarak su tüketiminin en yüksek ve en düşük seviyede olduğu durumları karşılaştırarak su yönetim endekslerini belirleyebileceklerdir. Ayrıca, su tüketiminin hangi değişkenlere bağlı olduğunu görebileceklerdir. Kişiler veya kurumlar, su talep yönetiminin dengeli olabilmesi için uygulama üzerinden durum takibi yapabileceklerdir.
- Kullanıcılar, web uygulaması sayesinde hızlı, güvenilir ve interaktif işlemler gerçekleştirebileceklerdir. Elde edilen sonuçlar, bazı kamu ve kuruluşlarla rapor şeklinde paylaşılabilecektir.

Bu uygulama, kullanıcıların cevaplandırdığı sorularla sürekli olarak yapılan analizlerle su kıtlığı ve su stresi hakkında farkındalık kazandırmak için bir eğitim aracı niteliği taşımaktadır [9]. Şekil 1'de uygulamaya ait ana erkan görüntüsü verilmektedir.



Şekil 1. Su tüketim analizi Shiny web uygulaması ana ekranı.

Makine öğrenimi, çevresel verilerden öğrenerek insan zekasını taklit etmeyi amaçlayan hesaplamalı algoritmaların gelişen bir dalıdır. Bu algoritmalar, büyük veri çağında çalışan anahtar teknoloji olarak kabul edilir.[11]. Makine öğreniminde kullanılan yöntemler üç ana kategoriye ayrılır: denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme. Her bir kategori içinde çeşitli algoritmalar bulunmaktadır [10].

2.1 Destek Vektör Makineleri

Support Vector Machine (SVM) algoritması, makine öğrenmesinde sınıflandırma ve regresyon problemlerini çözmek için kullanılan güçlü bir algoritmadır. SVM'nin temel fikri, veri kümesindeki örnekleri en iyi şekilde ayıran bir hiperdüzlem bulmaktır. Bu hiperdüzlem, sınıflar arasındaki ayrımı en iyi şekilde yapar ve sınıf etiketleri bilinmeyen yeni örnekleri doğru bir şekilde sınıflandırır [12]. SVM'nin temel özellikleri şunlardır:

- **Yüksek boyutlu uzaylarda çalışabilme:** SVM, özniteliklerin sayısının örnek sayısından daha fazla olduğu durumlarda da etkili bir şekilde çalışır.
- **Çekirdek yaklaşımı:** SVM, doğrusal olarak ayrılamayan verileri doğrusal hale getirmek için çekirdek fonksiyonları kullanır.
- **Maksimum kenar boşluğu:** SVM, sınıflar arasındaki ayrımı en iyi şekilde yapan hiperdüzlemi bulur. Bu, sınıflar arasındaki mesafenin maksimum olmasını sağlar.
- **Düzlem seçimi:** SVM, veri kümesindeki destek vektörleri adı verilen kritik örnekleri kullanarak hiperdüzlemi seçer.

SVM algoritması, görüntü sınıflandırması, metin sınıflandırması, biyoinformatik, finans ve daha pek çok alanda başarıyla uygulanmaktadır. Özellikle yüksek boyutlu verilerin sınıflandırılması gereken durumlarda oldukça etkindir. SVM, doğrusal olarak ayrılamayan verileri doğrusal hale getirmek için çekirdek fonksiyonları kullanır. Yaygın olarak kullanılan çekirdek fonksiyonları: Doğrusal, Polinom, Radyal Tabanlı Fonksiyon (RBF) ve Sigmoid. Uygun çekirdek fonksiyonu seçimi, problem ve veri kümesine bağlıdır. Bu seçim, model performansını önemli ölçüde etkiler. SVM'nin temel matematiksel formu aşağıdaki gibidir:

Verilen etiketli eğitim verileri {(x1, y1), (x2, y2), ..., (xn, yn)}, burada xi ∈ Rd öznitelik vektörü ve yi ∈ {-1, 1} sınıf etiketi.

SVM'nin amacı, sınıflar arasındaki ayrımı en iyi şekilde yapan hiperdüzlemi bulmaktır. Bu hiperdüzlem, aşağıdaki optimizasyon problemini çözerek bulunur:

Minimize: $(1/2)||w||^2 + C \Sigma \xi i$

Kısıtlar: $yi(w^T xi + b) \geq 1 - \xi i, \xi i \geq 0, i = 1, 2, ..., n$

Burada:

w: Hiperdüzlemin normal vektörü

b: Hiperdüzlemin kayma değeri

ξi: i. örnekteki hata payı (yumuşak sınır)

C: Düzenleştirme parametresi (trade-off)

Daha detaylı olarak:

w ve b, hiperdüzlemi tanımlayan parametrelerdir.

$||w||^2/2$, hiperdüzlemin karmaşıklığını ölçer ve minimize edilir.

Ξi, sınıflandırma hatalarını temsil eder ve toplamları C ile çarpılıp eklenir.

C parametresi, model karmaşıklığı ile sınıflandırma hatasının dengelenmesini kontrol eder.

Kısıtlar, örneklerin doğru sınıfa ait olmasını sağlar.

Optimizasyon problemi, Lagrange çarpanları kullanılarak çözülür ve w, b değerleri bulunur. Böylece en iyi hiperdüzlem elde edilmiş olur. Ayrıca, kernel fonksiyonları kullanılarak doğrusal olmayan problemlerde de SVM uygulanabilir. Kernel fonksiyonu, öznitelik uzayını yüksek boyutlu bir uzaya dönüştürerek doğrusal ayrılabilirliği sağlar [14].

Kernel fonksiyonu $K(xi, xj) = \varphi(xi)^T\varphi(xj)$ şeklinde tanımlanır, burada φ öznitelik dönüşümüdür.

2.2 Naive Bayes

Naive Bayes, olasılıksal bir sınıflandırma algoritmasıdır. Temel prensibi, Bayes teoremini kullanarak bir örneğin hangi sınıfa ait olduğunu hesaplamaktır [13]. Matematiksel olarak, Naive Bayes aşağıdaki denklemi kullanır:

$$P(class|x) = (P(x|class) * P(class)) / P(x) \quad (1)$$

Burada:

P(class|x): Bir örneğin x öznitelik vektörüne sahip olması durumunda, o örneğin belirli bir sınıfa ait olma olasılığı.

P(x|class): Belirli bir sınıfa ait örneklerin x öznitelik vektörüne sahip olma olasılığı.

P(class): Belirli bir sınıfa ait olma olasılığı (a priori olasılık).

P(x): Örneğin x öznitelik vektörüne sahip olma olasılığı.

Naive Bayes'in "naive" olarak adlandırılmasının nedeni, özniteliklerin birbirinden bağımsız olduğu varsayımıdır. Yani, bir özniteliğin değeri diğer özniteliklerin değerlerinden etkilenmez. Bu varsayımla, P(x|class) olasılığı aşağıdaki gibi hesaplanır:

$$P(x|class) = \Pi P(xi|class) \quad (2)$$

Burada, xi i'inci özniteliğin değeridir. Algoritmanın çalışma adımları şu şekildedir:

- Eğitim verisinden sınıf olasılıkları P(class) ve koşullu olasılıklar P(xi|class) hesaplanır.
- Tahmin edilecek yeni bir örnek için, Bayes teoremi kullanılarak P(class|x) olasılıkları hesaplanır.
- Örnek, en yüksek P(class|x) olasılığına sahip sınıfa atanır.

Naive Bayes, hesaplama maliyetinin düşük olması, veri boyutu ve öznitelik sayısından etkilenmemesi gibi avantajlara sahiptir. Ayrıca, doğrusal olmayan problemlerde de başarılı olabilir. Ancak, öznitelikler arasındaki bağımsızlık varsayımı her zaman gerçekçi olmayabilir [15-16].

2.3 Logistik Regresyon

Logistik regresyon, ikili sınıflandırma (binary classification) problemlerinde yaygın olarak kullanılan bir makine öğrenmesi algoritmasıdır. Logistik regresyon, bağımsız değişkenlerin (x) doğrusal bileşiminin bağımlı değişken (y) üzerindeki etkisini modeller. Burada bağımlı değişken ikili (0 veya 1) bir değişkendir [17-18].

Matematiksel olarak logistik regresyon modeli Eşitlik 3 gibi ifade edilir:

$$P(y=1|x) = g(\beta0 + \beta1x1 + \beta2x2 + ... + \beta n*xn) \quad (3)$$

Burada:

P(y=1|x): Bağımlı değişkenin 1 değerini alması için koşullu olasılık

g(z) = 1 / (1 + e^(-z)): Lojistik sigmoid fonksiyonu

β0, β1, ..., βn: Katsayılar (model parametreleri)

x1, x2, ..., xn: Bağımsız değişkenler

Lojistik sigmoid fonksiyonu, giriş değerinin (-∞, ∞) aralığından [0, 1] aralığındaki bir çıktı değeri üretir. Bu, sonucun bir olasılık değeri olarak yorumlanmasını sağlar. Logistik regresyon modelinin eğitimi, maksimum olabilirlik (maximum likelihood) yöntemi kullanılarak gerçekleştirilir. Bu yöntemde, model parametrelerinin (β değerleri) veriye en iyi uyumu sağlayacak şekilde belirlenmesi amaçlanır. Eğitim sonucunda elde edilen model, yeni giriş örnekleri için bağımlı değişkenin 1 olma olasılığını hesaplayabilir. Bu olasılık değeri eşik değeri (genellikle 0.5) ile karşılaştırılarak sınıflandırma kararı verilir. Logistik regresyon, lineer olmayan ilişkileri modelleyebilir, kolay

yorumlanabilir sonuçlar üretir ve farklı ölçeklerdeki değişkenleri kullanabilir. Ayrıca, sınıflandırma problemlerinin yanı sıra regresyon problemlerinde de kullanılabilir [19-20].

3 Deneysel Çalışmalar

Veriler, araştırmacılar tarafından geliştirilen kapsamlı bir anket aracılığıyla toplanmıştır. Anket, katılımcıların su tüketim alışkanlıklarını, su kullanım davranışlarını ve ilgili değişkenleri ölçmeye yönelik sorulardan oluşmaktadır. Anket, yüz yüze görüşmeler yoluyla 2023 yılında uygulanmıştır. Çalışmaya toplam 499 katılımcı dahil edilmiştir. Katılımcıların demografik özellikleri şu şekildedir:

Cinsiyet: 245 kadın (%49,1), 254 erkek (%50,9)
Medeni Durum: 278 evli (%55,7), 221 bekâr (%44,3)
Yaş: Ortalama 42,3 yıl (Standart Sapma = 13,7)
Hane Halkı Büyüklüğü: Ortalama 3,1 kişi (Standart Sapma = 1,4)
Katılımcılar, çeşitli sosyoekonomik ve konut özelliklerine sahip olup, farklı su kullanım alışkanlıklarına ve su tüketim düzeylerine sahiptirler. Anket verileri aşağıdaki değişkenleri içermektedir:

- Cinsiyet (Kadın/Erkek)
- Medeni Durum (Evli/Bekâr)
- Yaş (Sayısal)
- Hane Halkı Büyüklüğü (Sayısal)
- Ev Tipi (Müstakil/Apartman Dairesi)
- Musluk Tipi (Klasik/Tasarruflu)
- Atıksu Kullanımı (Var/Yok)
- İçme Suyu Kaynağı (Şebeke/Hazır)
- Su Tüketim Düzeyi (yüksek/düşük)
- Su Faturası Tutarı (Sayısal)
- Banyo Sayısı (Sayısal)
- Elektronik Cihaz Sayısı (Sayısal)
- Su Stresi (Likert 3)
- Su Bilinci (Likert 3)

Verilerin analizi için çeşitli makine öğrenmesi teknikleri kullanılmıştır. Özellikle, Destek Vektör Makineleri (SVM), Naive Bayes sınıflandırıcısı ve Lojistik Regresyon modelleri uygulanmıştır. Model performansları, çapraz doğrulama yöntemiyle değerlendirilmiştir.

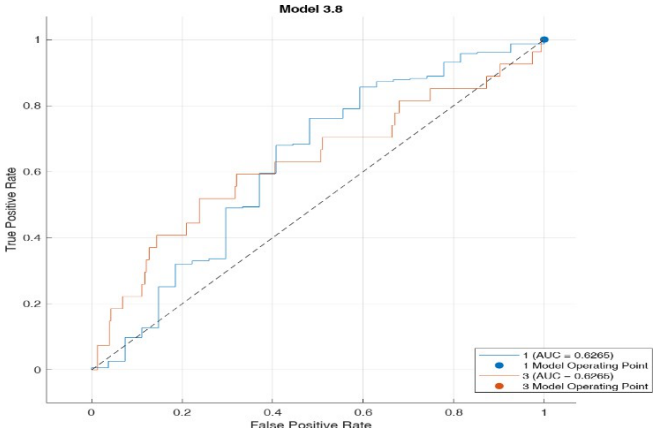
Veri setindeki eksik değerler kontrol edilmiş ve uygun yöntemler (ortalama, medyan vb.) kullanılarak doldurulmuştur. Değişkenlerin normallik varsayımları incelenmiş ve gerekli dönüşümler uygulanmıştır. Su tüketim düzeyini (yüksek/düşük) tahmin etmek için üç farklı sınıflandırma algoritması kullanılmıştır:

Destek Vektör Makineleri (SVM): Doğrusal ve doğrusal olmayan çekirdek fonksiyonları denenmiş, en iyi performansı veren model seçilmiştir. Parametreler grid arama yöntemiyle optimize edilmiştir.
Naive Bayes: Basit Naive Bayes sınıflandırıcısı uygulanmıştır. Değişkenlerin birbirinden bağımsız olduğu varsayılmıştır.
Lojistik Regresyon: İkili lojistik regresyon modeli kullanılmıştır. Değişkenlerin anlamlılığı Wald testi ile değerlendirilmiş, en iyi model seçilmiştir.
Her bir sınıflandırma modelinin performansı 10 katlı çapraz doğrulama yöntemiyle değerlendirilmiştir. Doğruluk, hassasiyet, özgüllük, F1-skoru gibi metrikler hesaplanmış ve modeller karşılaştırılmıştır. Model sonuçları Tablo 1'de verilmektedir.

Tablo 1. SVM, Naive Bayes ve Logistik regresyon model sonuçları.

	SVM	Naive Bayes	Logistik Regresyon
F1- skoru	%91.98	%90.64	%91.28
Doğruluk	%91.87	%90.48	%91.27
Hassasiyet	%91.99	%90.57	%91.24
Özgüllük	%91.78	%90.39	%91.22

SVM, %91,9 F1-skoru ile oldukça yüksek bir performans değerine sahiptir. SVM, doğrusal ve doğrusal olmayan ayrıştırma yeteneği sayesinde karmaşık sınıflandırma problemlerinde başarılı olabiliyor. Parametrelerin uygun şekilde optimize edilmesi, modelin yüksek başarım göstermesini sağlamış. Şekil 2'de SVM modeli ROC eğrisi verilmektedir.



Şekil 2. SVM modeli ROC eğrisi.

Naive Bayes modeli, %90,64 F1-skoru ile yine oldukça iyi bir performans sergilemiştir. Naive Bayes, değişkenler arası bağımsızlık varsayımı altında çalışan basit ama etkili bir sınıflandırıcıdır. Veri setindeki özelliklerin birbiriyle bağımsız olması durumunda, Naive Bayes modeli yüksek doğruluk sağlayabiliyor.

Lojistik Regresyon modeli, %91,28 F1-skoru ile SVM kadar yüksek olmasa da güçlü bir performans göstermiştir. Lojistik regresyon, doğrusal bir sınıflandırma modeli olmasına rağmen, su tüketim düzeyini tahmin etmede başarılı olmuştur. Değişkenlerin anlamlılığının denetlenmesi ve en iyi modelin seçilmesi, yüksek performans elde edilmesini sağlamış.

Genel olarak, her üç model de su tüketim düzeyini tahmin etmede oldukça başarılı sonuçlar vermiştir. SVM biraz daha yüksek performans gösterirken, Naive Bayes ve Lojistik Regresyon da güçlü modeller olarak öne çıkmaktadır. Uygulanan veri ön işleme ve model optimizasyon adımları, başarılı sonuçlara ulaşılmasında etkili olmuştur.

4 Sonuç

Bu çalışmada, su tüketim düzeyini tahmin etmek için üç farklı makine öğrenmesi modeli kullanılmış ve başarılı sonuçlar elde edilmiştir. SVM, %91,9 F1-skoru ile en yüksek performansı göstermiştir. Naive Bayes modeli %90,64 F1-skoru ile ikinci sırada yer alırken, Lojistik Regresyon %91,28 F1-skoru ile üçüncü sırada yer almıştır.

Çalışma kapsamında, geliştirilen makine öğrenmesi modelleri R programlama dili kullanılarak kodlanmıştır. Ayrıca, Shiny web çatısı kullanılarak, su tüketim tahminleri için interaktif bir web uygulaması oluşturulmuştur.

SVM modelinin doğrusal ve doğrusal olmayan ayrıştırma yeteneği, karmaşık sınıflandırma problemlerinde başarılı olmasını sağlamıştır. Parametrelerin uygun şekilde optimize edilmesi de modelin yüksek performans göstermesinde etkili olmuştur.

Naive Bayes modeli, değişkenler arası bağımsızlık varsayımı altında çalışan basit ama etkili bir sınıflandırıcı olarak öne çıkmıştır. Veri setindeki özelliklerin birbiriyle bağımsız olması durumunda, Naive Bayes modeli yüksek doğruluk sağlayabilmektedir.

Lojistik Regresyon modeli, doğrusal bir sınıflandırma modeli olmasına rağmen su tüketim düzeyini tahmin etmede başarılı sonuçlar elde etmiştir. Değişkenlerin anlamlılığının denetlenmesi ve en iyi modelin seçilmesi, yüksek performans elde edilmesini sağlamıştır.

Genel olarak, her üç model de su tüketim düzeyini tahmin etmede oldukça başarılı sonuçlar vermiştir. Uygulanan veri ön işleme ve model optimizasyon adımları, başarılı sonuçlara ulaşılmasında etkili olmuştur. Bu çalışma, su tüketimi tahmini gibi benzer problemlerin çözümünde makine öğrenmesi yöntemlerinin etkin kullanımına örnek teşkil etmektedir.

Teşekkür

Bu çalışma, TÜBİTAK 2209/A Üniversite Öğrencileri Araştırma Projeleri Destek Programı kapsamında desteklenmiştir. "Su Tüketimini Analiz Eden R Shiny Uygulaması" başlıklı projemize verdiği değerli katkı ve destekten dolayı TÜBİTAK'a en içten teşekkürlerimizi sunarız.

TÜBİTAK'ın sağladığı bu finansal destek, çalışmamızın başarıyla tamamlanmasında önemli bir rol oynamıştır. Elde ettiğimiz sonuçların, su

tüketimi tahmini gibi alanlarda makine öğrenmesi yöntemlerinin etkin kullanımına katkı sağlamasını umuyoruz.

Kaynaklar

[1] Kılıç, S. "Küresel İklim Değişikliği Sürecinde Su Yönetimi". *İstanbul Üniversitesi Siyasal Bilgiler Fakültesi Dergisi*, 0 (39), 161-186. 2011.

[2] Babel, M. S., Gupta, A. D., Pradhan, P. A "Multivariate Econometric Approach for Domestic Water Demand Modeling: An Application to Kathmandu, Nepal", *Water Resources Management,* 21, 573-589. 2007.

[3] Garcia-Cuerva L., Berglund E.Z., Binder A.R., "Public perceptions of water shortages, conservation behaviors, and support for water reuse in the U.S.", *Resources, Conservation and Recycling*, 113, 106-115. 2016.

[4] Kırtorun, Ezgi. & Karaer F. "Su Yönetimi ve Suyun Sürdürülebilirliği". *Uludağ Üniversitesi Sürdürülebilir Mühendislik Uygulamaları ve Teknolojik Gelişmeler Dergisi* 2018, 1(2): 151-159. 2018.

[5] Çevresel Etki Değerlendirmesi İzin ve Denetim Genel Müdürlüğü, Ç. E. ve B. Y. D. B. (2020). Çevresel göstergeler. Ankara.

[6] Ramos, H. M., McNabola, A., López-Jiménez, P. A., & Pérez-Sánchez, M. "Smart water management towards future water sustainable networks". *Water*, 12(1), 58, 2019.

[7] Dada, M. A., Majemite, M. T., Obaigbena, A., Daraojimba, O. H., Oliha, J. S., & Nwokediegwu, Z. Q. S. "Review of smart water management: IoT and AI in water and wastewater treatment". *World Journal of Advanced Research and Reviews,* 21(1), 1373-1382. 2024.

[8] Nishant, R., Kennedy, M., & Corbett, J. "Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda". *International Journal of Information Management*, 53, 102104. 2020.

[9] Batta Mahesh, "International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426. 2018.

[10] Alpaydin E. Introduction to machine learning. 3rd ed. Cambridge, MA: The MIT Press; 2014.

[11] G.-B. Huang et al. "Enhanced random search based incremental extreme learning machine", *Neurocomputing* (2008).

[12] Wallace, C. S., and Patrick J., "Coding decision trees", *Machine Learning* 11:7-22, 1993.

[13] Casas A., Rodríguez-Llorente C., Guillermo Rodríguez-Llorente G., Juan García J., Marcos Larriba M., "Machine learning screening tools for the prediction of extraction yields of pharmaceutical compounds from wastewaters", *Journal of Water Process Engineering* Volume 62, May 2024, 105379.

[14] A. Pisner D., M. Schnyer D., "Chapter 6 - Support vector machine", *Machine Learning*, Methods and Applications to Brain Disorders 2020, Pages 101-121.

[15] Ghosh, H., Tusher, M.A., Rahat, I.S., Khasim, S., Mohanty, S.N. (2023). Water Quality Assessment Through Predictive Machine Learning. In: Balas, V.E., Semwal, V.B., Khandare, A. (eds) Intelligent Computing and Networking. IC-ICN 2023. Lecture Notes in Networks and Systems, vol 699. Springer, Singapore. https://doi.org/10.1007/978-981-99-3177-4_6.

[16] Y. Wankhade, S. Toutam, K. Thakre, K. Kalbande and P. Thakre, "Machine Learning Approach for Breast Cancer Prediction: A Review," 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2023, pp. 566-570, doi: 10.1109/ICAAIC56838.2023.10141164.

[17] J.P.S.S. Madushani et al. "Evaluating expressway traffic crash severity by using logistic regression and explainable & supervised machine learning classifiers", Transport Engineering, (2023).

[18] Troiano, G., Nibali, L., Petsos, H., Eickholz, P., Saleh, M. H. A., Santamaria, P., Jian, J., Shi, S., Meng, H., Zhurakivska, K., Wang, H.-L., & Ravidà, A. "Development and international validation of logistic regression and machine-learning models for the prediction of 10-year molar loss". *Journal of Clinical Periodontology,* 50(3), 348–357. https://doi.org/10.1111/jcpe.13739. 2023.

[19] Dritsas, Elias, and Maria Trigka. "Efficient Data-Driven Machine Learning Models for Water Quality Prediction" *Computation* 11, no. 2: 16. https://doi.org/10.3390/computation11020016. 2023.

[20] M.G. Uddin, S. Nash, A. Rahman, A.I. Olbert "Performance analysis of the water quality index model for predicting water state using machine learning techniques", *Process Saf Environ Prot*, 169 2023, pp. 808-828, 10.1016/j.psep.2022.11.073.

Sağlık Eğitiminde Yapay Zekâ Destekli Chatbotların Potansiyelinin Değerlendirilmesi

Mehmet YILDIZ[1]*, AhsenDEMİRHAN [2], Mürüvvet EMREM [3], Uğur YAMAN [4], Metin YILDIZ [5]

[1]*Sakarya Uygulamalı Bilimler Üniversitesi, Uzaktan Eğitim Uygulama ve Araştırma Merkezi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Sakarya, TÜRKİYE*
[2]*Sakarya Üniversitesi, Sağlık Bilimleri Fakültesi, Hemşirelik Bölümü, Sakarya, TÜRKİYE*
[3]*Kocaeli Üniversitesi, Sağlık Bilimleri Fakültesi, Ebelik Bölümü, Kocaeli, TÜRKİYE*
[4]*Sakarya Üniversitesi, Sağlık Bilimleri Fakültesi, Hemşirelik Bölümü, Sakarya, TÜRKİYE*
[5] *Sakarya Üniversitesi, Sağlık Bilimleri Fakültesi, Ebelik Bölümü, Sakarya, TÜRKİYE*

Özet

Teknolojinin hızlı ilerleyişi ve maliyetlerin düşmesi, teknolojiye erişimi arttırmış bu da teknolojinin eğitim, sağlık, mühendislik gibi birçok alanda giderek yaygınlaşmasını sağlamıştır. Bu gelişmeler eğitim ortamlarında yapay zekâ (YZ) uygulamalarının sayısının artmasını sağlamıştır. Eğitime kazandırılan son YZ uygulamalarından biri de öğretme, öğrenme ve idari görevleri desteklemek için kullanılan Chatbot teknolojisidir. Geleneksel sağlık eğitimi süreci, etkileşim, kişiselleştirilmiş geri bildirimden yoksun, daha çok teorik bilgi aktarımına dayanan sınıf tabanlı bir eğitimdir. Chatbot teknolojilerinin gelişimi ve geniş uygulama alanı bulması, alanyazındaki mevcut sorunları çözmek için bir fırsat sağlamıştır. Bir sohbet robotu kurallara dayalı ya da YZ tabanlı olarak uygulanabilir. Ancak, kural tabanlı chatbotlardan farklı olarak, YZ tabanlı chatbotlar zamanla öğrenebilir ve daha akıllı hale gelebilir ve daha ölçeklenebilirdir. Son zamanlarda chatbot araştırmacıları için popüler bir seçim haline gelmiştir, ancak araştırmacılar için bu chatbot teknolojilerinin sağlık eğitiminde tam olarak nasıl etkili bir şekilde uygulanacağını belirlemek hala zor bir konudur. Sohbet robotlarının kullanımı son yıllarda giderek yaygınlaşsa da, bu YZ destekli sohbet robotlarının etkinliği konusunda çok az araştırma bulunmaktadır. Bu çalışmada, literatürdeki boşluğu doldurmak amacıyla, sağlık alanında chatbotların kullanıldığı çalışmalar analiz edilerek olası etkileri ve dezavantajları raporlanmaktadır. Araştırma kapsamında Web of Science veri tabanında eğitim, öğrenme, öğretme, yapay zeka, chatbot gibi kavramlar ve sağlık, medikal gibi anahtar kelimelerle arama yapılmıştır. Arama sonucunda elde edilen makaleler PRISMA diyagramına uygun olarak dahil edilmiş ve 88 makale analiz edilmiştir. Literatüre dayanarak, YZ destekli chatbot sistemi öğrencilerin akademik performansını, eleştirel düşünmesini ve öğrenme memnuniyetini etkili bir şekilde artırmıştır. Sağlık eğitiminde chatbotlar, eğitim süreçlerini daha erişilebilir hale getirebilir ve kişiselleştirilmiş öğrenme fırsatları sunma potansiyeline sahiptir. Sağlık eğitiminde YZ destekli chatbotların potansiyel riskleri veri güvenliği ve gizliliği, sınırlı empati ve iletişim becerileri, yanlış bilgi verme olasılığı ve aşırı teknoloji bağımlılığı olarak özetlenebilir. Yine de etkili ve düzenli bir planlama ile sanal bir asistan gibi ihtiyaçlara her zaman yanıt verebilen chatbotların büyük bir eğitim potansiyeli taşıdığı söylenebilir.

Anahtar Kelimeler: *Yapay Zekâ, Chatbot, Üretken Yapay Zeka, Eğitim, Sağlık Eğitimi*

* Contact email: mehmetyildiz@subu.edu.tr

Evaluating the Potential Artificial Intelligence Enabled Chatbots in Health Education

Abstract

The rapid advancement of technology and decreasing costs have increased access to technology, which has led to its widespread use in many fields such as education, health and engineering. These developments have led to an increase in the number of artificial intelligence applications in educational environments. One of the latest applications of AI in education is Chatbot technology used to support teaching, learning and administrative tasks. The conventional health education process is a classroom-based education that lacks interaction, personalized feedback and relies more on the transfer of theoretical knowledge. The development and wide application of chatbot technologies has provided an opportunity to solve existing problems in the literature. A chatbot can be either rule-based or AI-based. However, unlike rule-based chatbots, AI-based chatbots can learn and become smarter over time and are more scalable. Recently, chatbots have become a popular choice for researchers, but it is still a challenge for researchers to determine exactly how to effectively apply these chatbot technologies in health education. Although the use of chatbots has become increasingly common in recent years, there is little research on the effectiveness of these AI-powered chatbots. In this study, in order to fill the gap in the literature, studies using chatbots in the field of healthcare are analyzed and their possible effects and disadvantages are reported. Within the scope of the research, Web of Science database was searched with concepts such as education, learning, teaching, artificial intelligence, chatbot and keywords such as health, medical. The articles obtained as a result of the search were included in accordance with the PRISMA diagram and 88 articles were analyzed. Based on the literature, the AI-supported chatbot system effectively improved students' academic performance, critical thinking and learning satisfaction. Chatbots in health education can make educational processes more accessible and have the potential to provide personalized learning opportunities. The potential risks of AI-powered chatbots in health education can be summarized as data security and privacy, limited empathy and communication skills, the possibility of misinformation, and excessive technology dependency. Nevertheless, it can be said that chatbots, which can always respond to needs like a virtual assistant with effective and regular planning, have great educational potential.

Keywords: Artificial Intelligence, Chatbot, Generative AI, Education, Health Education

Giriş

Yapay zekâ ve sohbet robotlarının geçmişi, bilim insanlarının yapay zekâ kavramını ilk kez keşfetmeye başladığı 1950'li yıllara kadar uzanmaktadır. İlk gelişmeler arasında, insan konuşmalarını taklit etmek için tasarlanan ELIZA adlı ilk yapay zekâ programının oluşturulması yer almaktadır. Takip eden yıllarda, yapay zekâ teknolojisi ilerlemeye devam ederek karmaşık talepleri anlama ve yanıtlama becerisine sahip daha sofistike sohbet robotları geliştirilmiştir [1]. Bu sohbet robotları, kullanıcılarla etkileşime girebilen, bilgi, destek ve rehberlik sağlayan sanal asistanlar olarak hizmet vermektedir [2].

Yapay Zekâ (AI) özellikli sohbet robotları, sağlık eğitimi de dahil olmak üzere çeşitli alanlarda önemli bir potansiyel göstermiştir. Sağlık eğitiminde yapay zekâ güdümlü sohbet robotlarının potansiyeli çok büyüktür ve çalışmalar, idari görevleri otomatikleştirerek, klinik karar desteği sunarak ve kişiselleştirilmiş eğitim sağlayarak hasta bakımında devrim yaratabileceklerini göstermektedir [3,4]. Ek olarak bu sohbet robotları, rutin görevleri yönetme ve hasta eğitimine yardımcı olma konusunda önemli bir potansiyel sunarken, doğası gereği empati ve sezgi eksikliği nedeniyle insan sağlık hizmeti sağlayıcılarının yerini almaktan ziyade bir tamamlayıcı olarak görülmektedir [5]. Bu nedenle, sağlık hizmetlerinde sohbet robotlarının dönüştürücü potansiyelinden etkili bir şekilde yararlanmak için etik hususlar tarafından yönlendirilen sorumlu uygulama çok önemlidir [6].

Sağlık eğitimi bağlamında, yapay zekâ destekli sohbet robotları hastalara yaşamsal belirtilerin takibi, ilaç uyumu, semptom raporlama ve hatta randevu planlama konularında yardımcı olma konusunda da umut vaat etmektedir [7]. Ayrıca, önleyici bir rol oynayarak ve kullanıcılara doğru bilgi sağlayarak COVID-19 salgınıyla mücadelede kullanılmıştır [8,9]. COVID-19 salgını sırasında, yapay zekâ sohbet robotları bilgi yayma, öz değerlendirme, sağlık merkezlerine bağlanma, yanlış bilgilerle mücadele, hasta takibi, ruh sağlığı desteği, maruziyetin izlenmesi ve aşı bilgileri ve planlaması dahil olmak üzere çeşitli yönlerde etkili olmuştur [8]. Ek olarak son yıllarda chatbotlar, obezite ve kilo kontrolü [10] demans (örn, Endurance, Alzheimer hastalığı olan kişilerle samimi sohbetler geliştiren bir sohbet robotudur) [11], madde kullanım bozuklukları [12,13], doğum öncesi hizmetler [14], HIV ve cinsel sağlık [15] gibi bir dizi sorunu ele alma konusunda önemli gelişmeler yaşamıştır.

Sonuç olarak, sağlık eğitiminde yapay zekâ sohbet robotlarının kullanılması, öğrenme deneyimlerini geliştirmek, kişiselleştirilmiş destek sağlamak ve eğitim süreçlerini kolaylaştırmak için dönüştürücü bir fırsat sunmaktadır. Sohbet robotları, yapay zekanın yeteneklerinden yararlanarak hem sağlık hem de eğitim ortamlarında bilginin yayılma, destek sağlama ve öğrenmeyi kolaylaştırma biçiminde devrim yaratma potansiyeline sahiptir.

Method

Makale inceleme sürecinde nitel veri analizi yöntemlerinden doküman inceleme tekniği kullanılmıştır. Veri toplama aracı olarak bir Makale İnceleme Formu (MİF) oluşturulmuştur. Form içerisinde yazar, başlık, yıl, çalışmanın türü, alan, yöntem, değişkenler, örneklem grup/sayısı, veri toplama aracı, örneklem düzeyi/sayısı ve sonuç gibi başlıklar yer almıştır.

Veri toplama sürecinde Web of Science veri tabanında belirlenen anahtar kelimelerle tarama yapılmıştır.

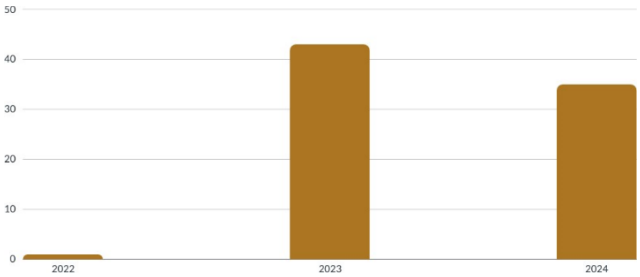
TS=("health" OR "medical" OR "nurse") AND
TS=("education" OR "learning" OR "teaching") AND
TS=("artificial intelligence" AND "chatbot")

Elde edilen 253 çalışmadan; makale türünde olması, İngilizce olması, sağlık eğitiminde chatbot kapsamında olması gibi kriterlere uyan 80 makale incelenmiştir.

3. Bulgular

Chatbot kullanımına yer veren akademik çalışmaların yıllara göre dağılımı Tablo 1'de verilmiştir.

Tablo 1. Çalışmanın yıllara göre dağılımı



Sık kullanılan değişkenler Tablo 2'de gösterilmiştir.

Tablo 2. Sık kullanılan değişkenler

Kullanıcı memnuniyeti:23
Doğruluk:22
Eğitim etkililiği/etkinliği: 21
Bilgi kalitesi:8
Katılım:3

Tablo 2.incelendiğinde chatbotların daha çok kullanıcı memnuniyeti sağlamak amacıyla kullanıldığı görülmektedir.

Örneklem sayısı dağılımı Tablo 3'te gösterilmiştir.

Tablo 3. Örneklem sayısı dağılımı

0-30:13
30-100:30
100+: 21

Tablo 3.'e bakıldığında çalışmalarda daha çok 30-100 arası katılımcı ile çalışıldığı sonucuna ulaşılmıştır.

Örneklem düzeyi ve kitlesi Tablo 4'te gösterilmiştir.

Tablo 4. Örneklem düzeyi, kitlesi

Sağlık çalışanları	13
Kronik hastalığa sahip hastalar	12
Genel kullanıcı düzeyi	12
Tıp eğitmenleri	10
Hemşirelik öğrencileri	4
Kanser hastaları	3
Tıp öğrencileri	2

Çalışmalar genellikle sağlık çalışanlarının chatbot deneyimine odaklanıldığını göstermektedir.

Uygulama sürelerinin dağılımı Tablo 5'te gösterilmiştir.

Tablo 5. Uygulama süresi dağılımı

Anlık: 34
3 ay: 5
6 ay: 7
Diğer: 6

Çalışmalar chatbot ile ilgili kullanıcı deneyimlerinin daha çok kısa süreli çalışmalar kapsamında yorumlandığını göstermektedir.

Chatbot kullanımının avantajları Tablo 6'da gösterilmiştir.

Tablo 6. Chatbot kullanımının avantajları

Bilgiye hızlı erişim
Doğruluk
Hızlı yanıt
Eğitimde yenilikçilik
Etkileşimli öğrenme
Kullanıcı dostu

Sağlık eğitiminde chatbot kullanım avantajlarına bakıldığında bilgiye hızlı erişim ve doğruluk öne çıkan başlıklar olarak görülmektedir (Tablo 7).

Tablo 7. Chatbot kullanımının dezavantajları

Kapsamlı/teknik bilgi eksikliği
Yanıltıcı bilgi
Doğruluk ve güvenirlik sorunları
Etik kaygılar

Sağlık eğitiminde chatbot kullanım dezavantajlarına bakıldığında ise bilgi eksikliği, yanıltıcı bilgi öne çıkan başlıklar olarak görülmektedir (Tablo 7).

4. Tartışma

Bu çalışmada, sağlık eğitimi sürecinde chatbot kullanımına yer veren akademik çalışmaların incelenmesi ve avantaj ile dezavantajları raporlanmıştır. Dünya çapındaki bilimsel üretimin ana havuzlarından olan WOS veri tabanı, ("health" OR "medical" OR "nurse") AND ("education" OR "learning" OR "teaching") AND ("artificial intelligence" AND "chatbot") anahtar kelimeleri ile taranmıştır. Makale inceleme sürecinde ise nitel veri analizi yöntemlerinden doküman inceleme tekniği kullanılmıştır.

Bu bölümde ilgili akademik çalışmaların yıllara göre dağılımları, sık kullanılan değişkenleri, örneklem sayısı, örneklem kitlesi ve düzeyi, uygulama süreleri dağılımı, sağlık alanında chatbot kullanımının avantajlar ve dezavantajlarına ilişkin bulguların literatür ışığında tartışmasına yer verilmiştir.

Sağlık eğitimi sürecinde chatbot kullanımına yer veren akademik çalışmaların yıllara göre dağılımları incelendiğinde, 2022 yılında 1, 2023 yılında 43 ve 2024 yılında 35 çalışmanın yapıldığı görülmektedir. Sağlık eğitiminde chatbot kullanımının oldukça güncel bir konu olduğu görülmektedir. Sallam (2023), yapmış olduğu sistematik derleme çalışmasında, ChatGPT kullanımının sağlık eğitimi, araştırması ve uygulamasındaki avantajları ve mevcut sınırlamaları değerlendirmiştir. PubMed ve Google

Akademik veri tabanlarının tarandığı ve 60 çalışmanın dahil edildiği bu çalışmada, çalışmaların 4' ünün 2022, geri kalan 56 çalışmanın 2023 yılında yayınlandığı görülmektedir. Sallam' ın 2023 yılında yayınlanan bu çalışması mevcut bulgumuzla paralellik göstermektedir [16]. Bir bibliyometrik analiz, sağlık alanında sohbet robotlarının kullanımına yönelik çalışmaların 1991-2021 yıllarını aralığında genel yıllık ortalama büyüme oranının %12,4 olduğunu ortaya koymaktadır. Ancak, 2017 yılından itibaren bu oran %45,81 gibi önemli bir artış göstermiştir. Özellikle 2020 yılında 404 (%19) belge üretilmiş olup, bu sayı, son 30 yılın toplam belgelerinin yaklaşık %20'sine denk gelmektedir. Bu verilere dayanarak, 2021 yılında 600'den fazla belgenin üretilmesi öngörülmektedir. Bu eğilim, 2022 ve sonrasında her yıl binlerce belgenin üretilebileceğine işaret etmektedir [17]. Plastik cerrahi alanında chatgpt kullanımının değerlendirildiği bir bibliyometrik analize, 30 çalışma dahil edilmiştir. Bibliyometrik analiz, ilk ChatGPT yayının Aralık 2022, ilk plastik cerrahisinde ChatGPT kullanılan yayının Mart 2023 de yayınlandığını ve plastik cerrahi ve diğer uzmanlık alanlarında ChatGPT kullanımı Aralık 2022- Haziran 2023 tarih aralığında incelenmiş olup her ay daha da artarak devam ettiği gösterilmiştir [18]. Hemşirelik eğitiminde ChatGPT kullanımını inceleyen bir bibliyometrik analiz çalışmasında, ChatGPT' nin Kasım 2022'de yayınlandığından beri, aylara göre yayın dağılımları incelenmiş olup analize dahil edilen 53 yayının çoğunluğunun Haziran 2023' e ait olduğu sonrasında Mart 2023 ve Eylül 2023 de yapıldığı bulunmuştur [19]. Literatür çalışma sonucumuz ile paralellik göstermektedir. Son yıllarda yapay zeka

destekli sohbet robotları üzerine yapılan çalışmaların sayısında kayda değer bir artış gözlenmiştir. Bu durum, özellikle 30 Kasım 2022 tarihinde kullanıma sunulan ChatGPT gibi yaygın olarak kullanılan yapay zeka tabanlı etkisiyle açıklanabilir. Yapılan çalışmaların çoğunluğunun ChatGPT ile ilgili olması da bunu desteklemektedir. ChatGPT'nin akademik çalışmalarda giderek daha fazla yer bulması, bu artışın arkasındaki başlıca nedenlerden biri olarak değerlendirilebilir.

Sağlık eğitimi sürecinde chatbot kullanımına yer veren akademik çalışmalarda en sık kullanılan değişkenler değerlendirildiğinde, sırasıyla kullanıcı memnuniyeti, doğruluk, eğitim etkililiği/etkinliği, eğitim kalitesi gelmektedir. Literatürde bu konuda yapılan çalışmaların oldukça sınırlı olduğu ve bulgumuzu destekleyecek verilerin çalışmalarda yer almadığı görülmektedir. Sağlık davranış değişikliğini teşvik etmede yapay zeka sohbet robotlarının uygulanabilirliğini, etkinliğini ve müdahale özelliklerini değerlendiren 15 çalışmanın dahil edildiği bir sistematik incelemede, bulgular etkinlik, fizibilite, kabul edilebilirlik ve kullanılabilirlik açısından değerlendirilmiştir. Dahil edilen çalışmalar sağlıklı yaşam tarzı benimseme, sigarayı bırakma, madde kötüye kullanımı, tedavi ve ilaç uyumunda etkili olabileceğini değerlendirmiştir. Ayrıca dahil edilen çalışmaların 7' si yapay zekâ sohbet robotlarının kabul edilebilirliğini; 11'i kullanılabilirliğini değerlendirmiştir [20]. Yapılan çalışmanın farklı konu ve popülasyonlarda gerçekleştirilen sistematik inceleme çalışması olması ile farklı değişkenler ile değerlendirilmiş olsa da genel olarak chatbotların doğruluk/kabul edilebilirlik ve etkililik değişkenleri ile değerlendirilmesi konusunda parallelik göstermektedir.

Chatbot' un sağlık eğitimi sürecinde kullanımına yer veren akademik çalışmaların örneklem büyüklükleri değerlendirildiğinde, büyükten küçüğe doğru sırasıyla 30 ile 100; 100 ve üstü; 0 ile 30 aralığında örneklem büyüklükleri olduğu bulunmuştur. Literatürde sağlık eğitiminde chatbotların kullanımının bibliyometrik analizini doğrudan yapan bir çalışma bulunmamakla beraber benzer konularda kapsamlı sistematik incelmeler bulunsa da, dahil edilen çalışmaların örneklem büyüklüklerine ilişkin verileri sağlayan çalışmalar oldukça kısıtlıdır. Sağlık davranış değişikliğini teşvik etmede yapay zeka sohbet robotlarının uygulanabilirliğini, etkinliğini ve müdahale özelliklerini değerlendiren sistematik incelemeye nitel ve nicel toplam 15 çalışma dahil edilmiştir. Çalışmaların örnekleminin 20 ile 99217 kişi arasında değiştiği görülmektedir [20]. Chatbot ile sağlanan psikoterapinin etkinliğinin değerlendirildiği bir sistematik inceleme ve meta regresyon çalışmasına 11 çalışmada dahil edilmiş olup, örneklem büyüklüklerinin 14 ile 396 arasında değiştiği bulunmuştur [21]. Çalışmaların örneklem büyüklüklerinin oldukça geniş yelpazede yayılım gösterdiği görülmektedir.

Mevcut araştırmaya dahil edilen çalışmaların, örneklem kitlesi değerlendirildiğinde, sağlık çalışanları, kronik hastalığa sahip hastalar, genel kullanıcı düzeyi, tıp eğitmenleri, hemşirelik öğrencileri, kanser hastaları, tıp öğrencileri ve profesyonelleri olduğu bulunmuştur. Sağlık davranış değişikliklerini teşvik etmek için yapay zekâ tabanlı sohbet robotların kullanımını değerlendiren bir sistematik derlemeye dahil edilen çalışmaların örneklemini ofis çalışanları, sigara içen bireyler, obez bireyler, meme kanseri hastaları, az fiziksel aktiviteye sahip bireyler, düzensiz beslenen bireyler ve fiziksel aktivite ve beslenme düzeni değiştirmeye karar veren bireyler oluşturmaktaydı [20]. Chatbot kullanımına ilişkin yapılan bir sistematik derlemede, dahil edilen çalışmaların örneklemleri kanserli hastalardan oluşmaktaydı [22]. Ayrıca chatbotların, ergenlerin madde cinsellik, uyuşturucu ve alkol hakkındaki soruları [23], hasta soruları [24], akademik yazım [25], tıp ve diş araştırmaları da [26] çalışmalarca değerlendirilmiştir. Chatbotların farklı alanlarda kullanıldığı, yukarıda yer verilen çalışmalarca ortaya konulmuştur. Buradan hareketle, chatbotların, geniş bir çalışma alanı sunarak farklı popülasyonlar ve konular üzerinde kullanılabilirliğe sahip olduğu görülmektedir.

Mevcut araştırmaya dahil edilen çalışmaların, uygulama süreleri değerlendirildiğinde, anlık takiplerden 1 yıla kadar değişkenlik gösterdiği ortaya konmuştur. Sağlık davranış değişikliklerini teşvik etmede kullanılan yapay zekâ tabanlı sohbet robotlarının kapsamlı incelemesinde [20], çalışmaya 15 araştırma dahil edilmiştir. Bahsedilen 15 çalışma, mevcut çalışmanın yazarları tarafından incelenerek, dahil

edilen çalışmaların uygulama sürelerinin 1 haftadan 2 yıla kadar değişkenlik gösterdiği belirlenmiştir. Kısıtlı çalışma bulgularına dayanarak tartışılan bu sonuç, araştırmaların uygulama sürelerinin de geniş zaman dilimlerine yayıldığını ortaya koymaktadır.

Mevcut çalışmada, chatbot kullanımın sağlık eğitimindeki avantajları, bilgiye hızlı erişim, doğruluk, hızlı yanıt, eğitimde yenilikçilik, etkileşimli öğrenme ve kullanıcı dostu olması olarak sıralanmaktadır. Sağlık eğitimi ve uygulamasında ChatGPT kullanımını değerlendiren bir sistematik incelemede, ChatGPT'nin avantajlarının en sık akademik/bilimsel yazım, bilimsel araştırmadaki faydaları, sağlık hizmeti uygulamasındaki kolaylık (23,3%), sağlık hizmeti disiplinlerindeki eğitimsel faydaların olduğu belirtilmiştir [16]. Yalçınkaya ve Çınar Yücel' in hemşirelik eğitiminde ChatGPT'yi değerlendirdikleri çalışmada, ChatGPT' nin avantajlarını kişiselleştirilmiş öğrenme sağlaması, kullanım kolaylığı, hızlı bilgi sunması, problem çözme ve eleştirel düşünme becerilerinin geliştirilmesi, eğitimcilerin müfredat geliştirme ve ders materyalleri hazırlama konusunda yardımcı olması, öğrenciler için etkili ve etkileşimli öğrenme sağlaması olarak aktarmaktadır [19]. Mevcut çalışmanın sonuçları, literatürde sunulan bulgularla paralellik göstermekte olup, chatbotların sağlık eğitimi ve uygulamalarındaki avantajlarını vurgulamaktadır. Özellikle ChatGPT'nin hızlı bilgi sunma, kişiselleştirilmiş öğrenme, etkileşimli ve yenilikçi eğitim sağlama gibi özellikleri hem akademik yazımda hem de sağlık eğitimi bağlamında önemli katkılar sunduğunu ortaya koymaktadır.

Mevcut çalışmada, sağlık eğitiminde kullanılan chatbotların dezavantajlarının, kapsamlı bilgi eksikliği, karmaşık durumlar için sınırlı bilgi, teknik bilgi eksikliği, yanıltıcı bilgi, doğruluk ve güvenirlik sorunları ve etik kaygılar olduğu bulunmuştur. Sağlık eğitimi ve uygulamasında ChatGPT kullanımını değerlendiren bir sistematik incelemede, ChatGPT' nin dezavantajlarının veri gizliliği ve güvenliğine ilişkin önyargılar ve intihal olmak üzere etik kaygılar yaygın olarak dile getirildi. Diğer endişeler, yanlış/hatalı bilgi sunabilmesi, yanlış atıf/kaynak göstermesi, şeffaflık sorunları, yasal sorunlar, kısıtlı bilgi, yanlış bilgi yayılma riski, aşırı ayrıntılı içerik, telif hakkı sorunları, özgünlük eksikliği olduğu bildirilmiştir [16]. Yalçınkaya ve Çınar Yücel' in hemşirelik eğitiminde ChatGPT'yi değerlendirdikleri çalışmada, ChatGPT'nin dezavantajlarını, güncel literatüre, klinik verilere ve tıp alanında yapılan araştırmalara erişiminin olmaması nedeniyle her zaman doğru ve güncel bilgiler sağlayamaması, kaynakları yanlış göstermesi veya verdiği bazı bilgilerin doğru olmamasının ve doğrulanmamış bilgi kullanımının da hasta bakımını olumsuz etkileme ve intihal sorunlarına yol açabilme ihtimali olarak aktarmıştır [19]. Bu çalışmanın bulguları, literatürde belirtilen dezavantajlarla örtüşmekte ve chatbotların sağlık eğitiminde kullanımına ilişkin belirli sınırlılıkları ortaya koymaktadır. Özellikle ChatGPT'nin doğruluk ve güvenilirlik sorunları, etik kaygılar, yanıltıcı bilgi riski ve veri gizliliği gibi unsurlar, bu teknolojilerin sağlık alanında kullanımında dikkat edilmesi gereken önemli noktalar olarak karşımıza çıkmaktadır. Mevcut çalışma, yapay zeka destekli chatbotların sağlık eğitimine değerli katkılar sağlasa da, bu araçların dikkatli ve sorumlu bir şekilde kullanılması gerektiğini, aksi takdirde eğitim ve hasta güvenliği üzerinde olumsuz sonuçlar doğurabileceğini vurgulamaktadır.

5. Sonuç

YZ destekli chatbot sistemi öğrencilerin akademik performansını, eleştirel düşünmesini ve öğrenme memnuniyetini etkili bir şekilde artırmıştır. Sağlık eğitiminde chatbotlar, eğitim süreçlerini daha erişilebilir hale getirebilir ve kişiselleştirilmiş öğrenme fırsatları sunma potansiyeline sahiptir. Sağlık eğitiminde YZ destekli chatbotların potansiyel riskleri veri güvenliği ve gizliliği, sınırlı empati ve iletişim becerileri, yanlış bilgi verme olasılığı ve aşırı teknoloji bağımlılığı olarak özetlenebilir. Yine de etkili ve düzenli bir planlama ile sanal bir asistan gibi ihtiyaçlara her zaman yanıt verebilen chatbotların büyük bir eğitim potansiyeli taşıdığı söylenebilir. Ancak chatbotların çeşitli önemli sınırlılıkları da bulunmaktadır. Özellikle doğruluk ve güvenilirlik sorunları, etik kaygılar, yanıltıcı bilgi riski ve veri gizliliği gibi unsurlar, bu teknolojilerin sağlık alanında kullanımında dikkat edilmesi gereken önemli noktalar olarak karşımıza çıkmaktadır. Bu teknolojilerin insan eğitimcilerin

yerini alması mümkün değildir. Chatbotlar, insan rehberliği ve uzmanlığıyla karşılaştırıldığında önemli sınırlılıklar taşımakta ve bu nedenle yalnızca tamamlayıcı bir araç olarak değerlendirilmelidir. Bu sınırlılıkların göz önünde bulundurulması, chatbotların sağlık eğitiminde daha dikkatli ve dengeli bir şekilde kullanılması gerektiğini vurgulamaktadır. Gelecekteki araştırmalar, yapay zekâ ve chatbotların sağlık eğitiminde en verimli ve etik şekilde nasıl uygulanabileceğine yönelik kılavuzlar geliştirmelidir.

Teşekkür

Çalışmaya destek veren tüm yazarlara teşekkür ederiz.

Kaynaklar

[1] King, M.R., chatGPT. A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cel. Mol. Bioeng.* **16**, 1–2 (2023). https://doi.org/10.1007/s12195-022-00754-8

[2] Sun, G. (2023). Ai in healthcare: navigating opportunities and challenges in digital communication. Frontiers in Digital Health, 5. https://doi.org/10.3389/fdgth.2023.1291132

[3] Talyshinskii, A. (2023). Potential of ai-driven chatbots in urology: revolutionizing patient care through artificial intelligence. Current Urology Reports, 25(1), 9-18. https://doi.org/10.1007/s11934-023-01184-3

[4] Haque, A. (2023). Transforming chronic disease management with chatbots: key use cases for personalized and cost-effective care.. https://doi.org/10.1109/is3c57901.2023.00104

[5] Altamimi, I. (2023). Artificial intelligence (ai) chatbots in medicine: a supplement, not a substitute. Cureus. https://doi.org/10.7759/cureus.40922

[6] Garcia Valencia, O. A., Suppadungsuk, S., Thongprayoon, C., Miao, J., Tangpanithandee, S., Craici, I. M., … & Cheungpasitporn, W. (2023). Ethical implications of chatbot utilization in nephrology. Journal of Personalized Medicine, 13(9), 1363. https://doi.org/10.3390/jpm13091363

[7] Karuna, G. (2023). Helpi – an automated healthcare chatbot. E3s Web of Conferences, 430, 01040. https://doi.org/10.1051/e3sconf/202343001040

[8] Mahdavi, A., Amanzadeh, M., Hamedan, M., & Naemi, R. (2023). Artificial intelligence based chatbots to combat covid-19 pandemic: a scoping review.. https://doi.org/10.21203/rs.3.rs-2565141/v1

[9] Mahdavi, A. (2023). Artificial intelligence-based chatbots to combat covid-19 pandemic: a scoping review. Shiraz E-Medical Journal, 24(11). https://doi.org/10.5812/semj-139627

[10] Huang C-Y, Yang M-C, Huang C-Y, Chen Y-J, Wu M-L, Chen K-W. A chatbot-supported smart wireless interactive healthcare system for weight control and health promotion. IEEE International Conference on Industrial Engineering and Engineering Management. IEEE; 2018.

[11] Ruggiano, N., Brown, E. L., Roberts, L., Framil Suarez, C. V., Luo, Y., Hao, Z., & Hristidis, V. (2021). Chatbots to Support People With Dementia and Their Caregivers: Systematic Review of Functions and Quality. *Journal of medical Internet research*, *23*(6), e25006. https://doi.org/10.2196/25006

[12] Ogilvie, L., Prescott, J., & Carson, J. (2022). The Use of Chatbots as Supportive Agents for People Seeking Help with Substance Use Disorder: A Systematic Review. *European addiction research*, *28*(6), 405–418. https://doi.org/10.1159/000525959

[13] Ahmed, A., Hassan, A., Aziz, S., Abd-Alrazaq, A. A., Ali, N., Alzubaidi, M., Al-Thani, D., Elhusein, B., Siddig, M. A., Ahmed, M., & Househ, M. (2023). Chatbot features for anxiety and depression: A scoping review. *Health informatics journal*, *29*(1), 14604582221146719. https://doi.org/10.1177/14604582221146719

[14] R. Wang, J. Wang, Y. Liao and J. Wang, "Supervised Machine Learning Chatbots for Perinatal Mental Healthcare," 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), Sanya, China, 2020, pp. 378-383, doi: 10.1109/ICHCI51889.2020.00086. keywords: {Pregnancy;Pediatrics;Human-robot interaction;Mental health;Machine learning;Chatbot;Real-time systems;Supervised machine learning;Perinatal mental health;Chatbots},

[15] Jacqueline Brixey, Rens Hoegen, Wei Lan, Joshua Rusow, Karan Singla, Xusen Yin, Ron Artstein, and Anton Leuski. 2017. SHIHbot: A Facebook chatbot for Sexual Health Information on HIV/AIDS. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 370–373, Saarbrücken, Germany. Association for Computational Linguistics.

[16] Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. Healthcare (Basel), 11(6), 887. https://doi.org/10.3390/healthcare11060887

[17] Pears, M., & Konstantinidis, S. (2022). Bibliometric analysis of chatbots in health-trend shifts and advancements in artificial intelligence for personalized conversational agents. Studies in Health Technology and Informatics, 290, 494-498. https://doi.org/10.3233/SHTI220125

[18] Liu, H. Y., Alessandri-Bonetti, M., Arellano, J. A., & Egro, F. M. (2024). Can ChatGPT be the plastic surgeon's new digital assistant? A bibliometric analysis and scoping review of ChatGPT in plastic surgery literature. Aesthetic Plastic Surgery, 48(8), 1644-1652. https://doi.org/10.1007/s00266-023-03709-0

[19] Yalcinkaya, T., & Cinar Yucel, S. (2024). Bibliometric and content analysis of ChatGPT research in nursing education: The rabbit hole in nursing education. Nurse Education in Practice, 77, 103956. https://doi.org/10.1016/j.nepr.2024.103956

[20] Aggarwal, A., Tam, C. C., Wu, D., & Li, X., & Qiao, S. (2023). Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review. Journal of Medical Internet Research, 25, e40789. https://doi.org/10.2196/40789

[21] Lim, S. M., Shiau, C. W. C., Cheng, L. J., & Lau, Y. (2022). Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: A systematic review and meta-regression. Behavior Therapy, 53(2), 334-347. https://doi.org/10.1016/j.beth.2021.09.007

[22] Xu, L., Sanders, L., Li, K., & Chow, J. C. L. (2021). Chatbot for health care and oncology applications using artificial intelligence and machine learning: Systematic review. JMIR Cancer, 7(4), e27850. https://doi.org/10.2196/27850

[23] Crutzen, R., Peters, G. J., Portugal, S. D., Fisser, E. M., & Grolleman, J. J. (2011). An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: An exploratory study. Journal of Adolescent Health, 48(5), 514-519. https://doi.org/10.1016/j.jadohealth.2010.09.002

[24] Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Internal Medicine, 183(6), 589-596. https://doi.org/10.1001/jamainternmed.2023.1838

[25] Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. Cureus, 15(2), e35179. https://doi.org/10.7759/cureus.35179

[26] Fatani, B. (2023). ChatGPT for future medical and dental research. Cureus, 15(4), e37285. https://doi.org/10.7759/cureus.37285

An Approach for Data Augmentation of Dental X-Ray Images with the use of Deep Convolutional Generative Adversarial Network

Ceren SÖZERİ[1], Bahadır KARASULU[2],*

[1]Canakkale Onsekiz Mart University, School of Graduate Studies, Department of Computer Engineering, Canakkale, Turkey
[2]Canakkale Onsekiz Mart University, Faculty of Engineering, Department of Computer Engineering, Canakkale, Turkey

Abstract

Deep Convolutional Generative Adversarial Network (DCGAN) has been highly successful in image generation and was therefore, selected for dental images. The research area of medical literature suffers from a lack of sustainable image datasets, particularly for pediatric dental images, although there is valuable research on adult dental images. Nevertheless, for real life applications and more accurate outcomes, adult and child images should be used together. In our study, one can achieve more accurate results in real-life applications with synthetic images created by the DCGAN model, which was trained on a dataset of 193 pediatric X-ray images. In our experiments for the training procedure, the process takes 10,000 epochs. DCGAN model generated 965 new synthetic images, with five times data augmentations for each original image and the resolution of these images is 64x64. To evaluate image quality, Peak Signal-To-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Fréchet Inception Distance (FID) scores were calculated. The average PSNR is 27.9 (range[0,100]), and the average SSIM is 0.45 (range[0,1]) on a dataset of 193 generated images. The FID score is 457.33, and theoretically, there is no upper limit for FID. As a result, the 965 DCGAN-generated images fall within an acceptable range. Scientific discussion about the above-mentioned experiments is given in this study.

Keywords: *Dataset, Generative Adversarial Network, Dental Image, Data augmentation.*

Derin Evrişimli Üretken Çekişmeli Ağ Kullanımı ile Diş Röntgen Görüntülerinin Veri Artırımına Yönelik Bir Yaklaşım

Özet

Derin Evrişimli Üretken Çekişmeli Ağ (DEÜÇA-DCGAN) görüntü oluşturmada oldukça başarılıdır ve bu nedenle diş görüntüleri için seçilmiştir. Tıbbi literatürün araştırma alanı, yetişkin diş görüntüleri üzerine değerli araştırmalar olmasına rağmen, özellikle pediatrik diş görüntüleri için sürdürülebilir görüntü veri kümelerinin eksikliğinden muzdariptir. Bununla birlikte, gerçek hayat uygulamaları ve daha doğru sonuçlar için yetişkin ve çocuk görüntüleri birlikte kullanılmalıdır. Çalışmamızda, 193 pediatrik röntgen görüntüsünden oluşan bir veri kümesi üzerinde eğitilen DEÜÇA modeli tarafından oluşturulan sentetik görüntülerle gerçek hayat uygulamalarında daha doğru sonuçlar elde edilebilir. Eğitim prosedürü için yaptığımız deneylerde süreç 10.000 epok sürmektedir. DEÜÇA modeli, her orijinal görüntü için beş kez veri artırma ile 965 yeni sentetik görüntü üretmiştir ve bu görüntülerin çözünürlüğü 64x64'tür. Görüntü kalitesini değerlendirmek için Tepe Sinyal-Gürültü Oranı (TSGO-PSNR), Yapısal Benzerlik İndeksi Ölçüsü (YBİÖ-SSIM) ve Fréchet

* Contact email: 21414934010@ogr.comu.edu.tr

Başlangıç Mesafesi (FBM-FID) puanları hesaplanmıştır. Üretilen görüntülerden 193 tanesinden oluşan bir veri kümesinde ortalama TSGO 27,9 (aralık[0,100]) ve ortalama YBİÖ 0,45'tir (aralık[0,1]). FBM puanı 457,33'tür ve teorik olarak FBM için bir üst sınır yoktur. Sonuç olarak, DEÜÇA tarafından üretilen 965 görüntü kabul edilebilir bir aralıkta yer almaktadır. Yukarıda bahsedilen deneylerle ilgili bilimsel tartışmalar bu çalışmada verilmiştir.

Anahtar Kelimeler: *Veri Kümesi, Üretken Çekişmeli Ağ, Diş Görüntüsü, Veri Arttırma*

1 Introduction

Oral health is the ability to perform activities such as laughing, tasting and swallowing without pain or discomfort [1]. Although not recognized as a problem, oral diseases are a major public health issue worldwide [2]. Oral health problems are not only related to overall health but can also worsen existing conditions or even trigger new ones [3].

According to The Global Oral Health Status Report 2022 [2], high-income countries have 6.85 dentists for every ten thousand people, low-income countries have 0.57, and the global average is 3.28 dentists for every ten thousand people. Accordingly approximately two billion adults worldwide have tooth decay, while 514 million children suffer from cavities.

Classic Machine Learning (ML) algorithms do not have the ability for automatic feature learning and their ability to interpret is lower than Deep Learning (DL) algorithms. DL better interprets data by filling the semantic gap between human perception and machine perception and also has higher experimental performance (accuracy) than ML.

Generative Adversarial Networks (GAN) takes its strength from two different models which trained simultaneously and every step with backpropagation the models updates themselves therefore the output becomes highly successful [4].

GAN's applications are wide such as text transferring, face detection, image processing such as recovering low-resolution images for high quality and generate images [5] and etc.

The medical literature has a significant gap regarding sustainable image datasets, particularly for pediatric dental research. For this purpose a dataset that includes pediatric dental images was generated using Deep Convolutional Generative Adversarial Networks (DCGANs).

The datasets utilized in studies within the literature are frequently either not publicly accessible or primarily composed of adult data. As a result, there are a limited availability of dental images of children and adolescents in this field. A fundamental challenge encountered in automatic data generation is the requisite reliance on human expertise for the annotation task.

The performance of GANs is influenced by various factors, including data quality, quantity, training duration, and backpropagation techniques. A significant challenge in this domain is underfitting, a condition where the model fails to learn the underlying data distribution effectively. In this research, the limited dataset and an insufficient number of epochs may have prevented the GAN from fully understanding the data's underlying structure [6].

Considering the experiments produced by GAN [7], which have a relatively reasonable processing time, the image size is 64x64. Therefore, there may be a loss of detail in the images. This research contributes a novel synthetic dataset and demonstrates the efficacy of DCGANs in generating realistic dental images.

This article is divided into five sections. In the second section, the studies in the literature are analysed, in the third section, the methods and materials used are mentioned. In the fourth section, experimental results are presented. In the fifth section, discussion and conclusions based on scientific findings are presented.

2 Literature review

There are several different image types in dental radiology: panoramic, bitewing, periapical etc. Although each dataset varies according to this image type, most of these data belong to adults [8, 9]. While extensive research has been conducted using dental images, many of these datasets are privately owned or inaccessible for public use.

Recently a number of researchers have introduced a mixed dataset includes 193 pediatric and 2692 adults dental images [10]. In our study, for the

training process, this 193 pediatric image was used.

Karen Panetta and others have presented a dataset of 1,000 images to the literature, but the only way to access this dataset is if you agree to use it for academic research, even if is announced as public in their *"A Comparison Of Panoramic Radiography Datasets"* table [11].

In a study investigating the application of DCGANs to medical imaging, a model was trained on a dataset of 6432 chest X-ray images and the generated synthetic images exhibited high fidelity to real-world examples, accurately capturing noise, disease patterns, and anatomical details [12].

In our study, a different approach is used to assembled as a brand new synthetic dataset with some automatically generated images by the use of the DCGAN.

In scope of GAN stability, our approach is more preferable than other approaches from similar studies by considering its small number of image volume.

3 Materials and methods

In this section, the dataset, related techniques and objective evaluation based on the performance metrics used in our study are introduced. In addition, their details are declared as well.

3.1 Dataset

In Table 1, the details of the Children Tooth Dataset by Age and Gender are shown.

Table 1. Image Distribution in the Children's Tooth Dataset by Age and Gender [10].

Age	Male	Female	Total
2	1	0	1
3	1	2	3
4	9	7	16
5	8	15	23
6	12	8	20
7	5	8	13
8	7	5	12
9	3	5	8
10	1	0	1
11	1	1	2
13	0	1	1
Total	48	52	100

In our study, publicly available dataset of pediatric dental images was used for training [10]. This dataset comprised two distinct subsets: the "Children Tooth Dataset" containing of 100 images and the "Children's Dental Supplement Dataset" consisting of 93 images.
In total, the model was trained on 193 pediatric dental X-ray images. To expedite processing and reduce computational complexity, the original images, with a resolution of 2000x942 pixels, were downscaled to a standardized 64x64 pixel format.

3.2 *The techniques and methods*

The methodological approach taken in this study is a methodology based on Generative Adversial Networks improved with Convolutional Neural Networks, also known as DCGAN [13]. Convolutional neural networks are good feature extractors as well as classifiers thanks to the high-level semantic enrichment of the real meaning of these features.

The basic structure of consists of two convolutional modules that learn by working in conflict with each other. One of these modules is called the Generator, and the other is the Discriminator. The Generator is tasked with producing increasingly authentic synthetic images, while the Discriminator functions as a binary classifier designed to differentiate between real and synthetic images [14].

In 2016, Alec Radford and his team proposed a new model named DCGAN by integrating Convolutional Neural Networks (CNN) into the GAN architecture [13]. The DCGAN method, as proposed by [15], exhibits greater stability in most environments.

3.3 *The performance metrics*

The Peak Signal to Noise Ratio (PSNR) metric is derived from the Mean Squared Error (MSE). The Peak Signal-to-Noise Ratio (PSNR) is calculated by taking the logarithm of the mean squared error (MSE) of an image [16].

Structural Similarity Index (SSIM) [17] assess the perceptual quality of digital images by comparing structural information. SSIM generates an output by analyzing the image's brightness, the contrast between dark and light regions, and the structural information, which refers to pixel intensities.

Fréchet Inception Distance (FID) [18] is an improvement over the Inception Score, considering only mean (m) and covariance (C). The similarity

between real images $p(.)$ and synthetic images $p_w(.)$ is calculated based on Fréchet Distance (d) as shown in Equation (1):

$$d^2((m, C), (m_w, C_w)) = \|m - m_w\|_2^2 + Tr(C + C_w - 2(CC_w)^{1/2}) \quad (1)$$

According to this research, "Fréchet Initial Distance", more accurately assesses the resemblance between synthetic images and real images compared to the Initial Score [18].

4. Experimental results

A DCGAN model was trained on a dataset of 193 pediatric dental images. The experiment was conducted on a Colab platform utilizing an A100 GPU with dedicated 256 processing units. In this infrastructure, PyTorch [19] and other libraries are used to run the experiments at this platform [20].

The training process for 10,000 epochs required approximately 22 hours. The optimizer used was Adaptive Moment Estimation (ADAM) [21], with a β_1 hyperparameter of 0.5. The batch size was set to 128, and the generated images had a resolution of 64x64 pixels. Binary cross-entropy was selected as the loss function.

After training, the model generated 965 images, resulting in a ratio of five generated images for every input image. In Figure 1 and Figure 2, PSNR and SSIM values obtained in our experiments are shown, respectively. As performance indicator, the loss values for two component of DCGAN are illustrated in Figure 3.

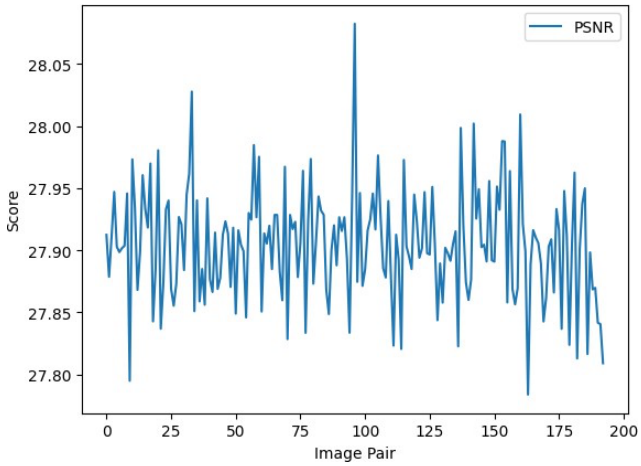


Figure 1. PSNR values

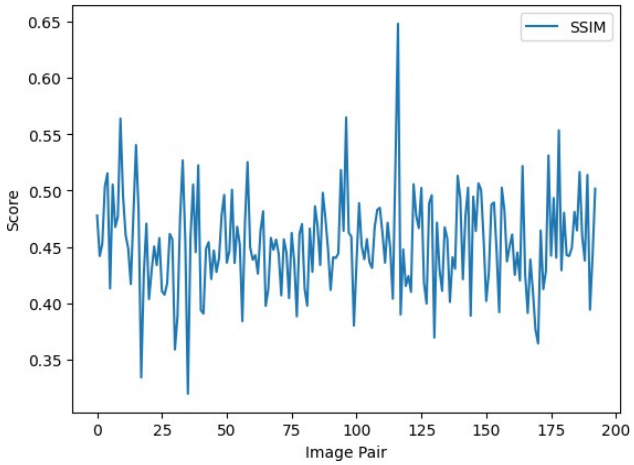


Figure 2. SSIM values

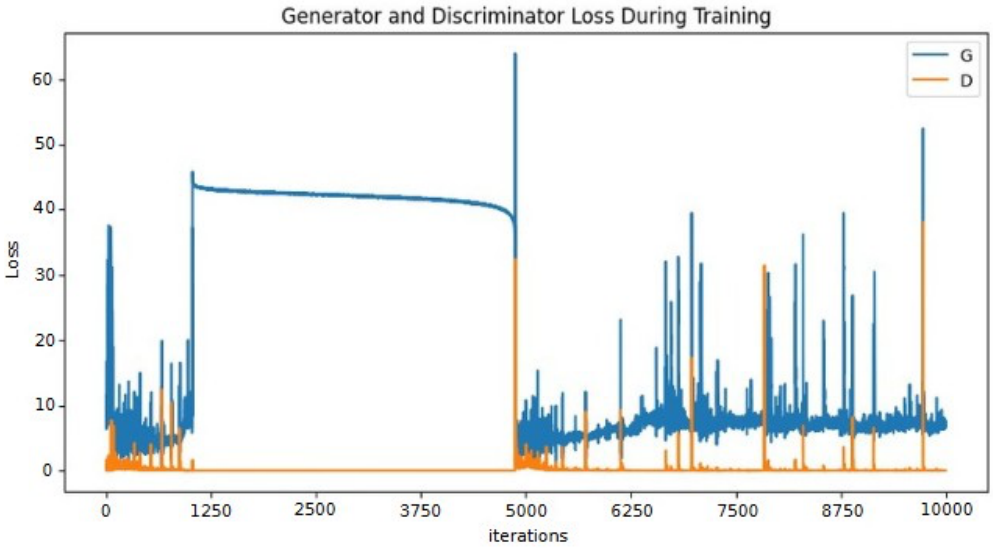


Figure 3. Generator and Discriminator Loss During Training

As seen in Figure 3, the generator's loss function value encountered a persistent roadblock during the early stages of training, remaining stubbornly at between forty and sixty. However, after reaching the 5,000th epoch, the generator's performance improved significantly, enabling successful training continuation.

Following model training, the quality of generated images was assessed using standard evaluation metrics. The average PSNR was calculated to be 27.9, while the average SSIM was found to be 0.45. Additionally, the FID score was determined to be 457.33.

In Table 2, generator and discriminator loss values obtained in our experiments are shown in details.

Table 2. Generator and Discriminator Loss Values for the Final Ten Epochs

Epoch	Discriminator Loss	Generator Loss
9990	0.0068	6.8545
9991	0.0080	6.8160
9992	0.0053	7.1699
9993	0.0049	7.2595
9994	0.0041	7.2836
9995	0.0066	6.7738
9996	0.0062	6.9422
9997	0.0070	6.8820
9998	0.0071	6.9843
9999	0.0045	7.6570

5. Disscusion and conclusion

This research explores the generation of images from a limited dataset using a GAN architecture. A DCGAN, a specialized type of GAN, was used. The model operates through a competitive process: a discriminator network attempts to differentiate between real and synthetic images, whereas a generator network strives to generate realistic images that can deceive the discriminator.

As shown in Table 2, the loss value has significantly decreased over the last ten epochs. However, due to the limitations imposed by the relatively small dataset and technical constraints, our training process was curtailed at 10,000 epochs, preventing us from achieving the desired level of success.

Although the generated images have achieved a commendable level of realism, their size and degree of realism may still be insufficient for demanding tasks like segmentation and classification, highlighting the need for further improvements.

For better results, it is expected that experimentation with at least 20,000 epochs and upgrading the small-sized images produced to high resolution will provide a high success in terms of authenticity.

References

[1] Glick, Michael, et al. A new definition for oral health developed by the FDI World Dental Federation opens the door to a universal definition of oral health. *British dental journal*, 2016, 221.12: 792-793.

[2] Global oral health status report: towards universal health coverage for oral health by 2030. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO

[3] Commıttee On An Oral Health Inıtıatıve. *Advancing oral health in America*. National Academies Press, 2012.

[4] Goodfellow, Ian, et al. Generative adversarial nets. *Advances in neural information processing systems*, 2014, 27.

[5] Durgadevi, M., et al. Generative Adversarial Network (GAN): A general review on different variants of GAN and applications. In: *2021 6th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2021. p. 1-8.

[6] Zhang, Huijuan; Huang, Zongrun; LV, Zhongwei. Medical image synthetic data augmentation using gan. In: *Proceedings of the 4th International Conference on Computer Science and Application Engineering*. 2020. p. 1-6.

[7] Wang, Yaxing, et al. Transferring gans: generating images from limited data. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018. p. 218-234.

[8] Román, Julio César Mello, et al. Panoramic dental radiography image enhancement using multiscale mathematical morphology. *Sensors*, 2021, 21.9: 3110.

[9] Abdi, Amir Hossein; Kasaeı, Shohreh; Mehdızadeh, Mojdeh. Automatic segmentation of mandible in panoramic x-ray. *Journal of Medical Imaging*, 2015, 2.4: 044003-044003.

[10] Zhang, Yifan, et al. Children's dental panoramic radiographs dataset for caries segmentation and dental disease detection. *Scientific Data*, 2023, 10.1: 380.

[11] Panetta, Karen, et al. Tufts dental database: a multimodal panoramic x-ray dataset for benchmarking diagnostic systems. *IEEE journal of biomedical and health informatics*, 2021, 26.4: 1650-1659.
[12] Akhil, MC Sai, et al. Medical Image Synthesis Using DCGAN for Chest X-Ray Images. In: *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*. IEEE, 2024. p. 1-8.
[13] Radford, A. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
[14] Çelik, Gaffari; TALU, Muhammed Fatih. Çekişmeli üretken ağ modellerinin görüntü üretme performanslarının incelenmesi. *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 2020, 22.1: 181-192.
[15] Dewi, Christine, et al. Synthetic Data generation using DCGAN for improved traffic sign recognition. *Neural Computing and Applications*, 2022, 34.24: 21465-21480.
[16] Setiadi, De Rosal Igantius Moses. PSNR vs SSIM: imperceptibility quality assessment for image steganography. *Multimedia Tools and Applications*, 2021, 80.6: 8423-8444.
[17] Wang, Zhou, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 2004, 13.4: 600-612.
[18] Heusel, Martin, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 2017, 30.
[19] PyTorch Website, https://pytorch.org, Accessed: 20.07.24
[20] Google Colaboratory Website, https://colab.research.google.com, Accessed: 23.07.24
[21] Kingma, Diederik P. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Derin Öğrenme Tabanlı Otomatik Kask Tespiti

Ayşe Nur SAYLAM[1]*, *Süleyman UZUN*[2], *Cem ÖZKURT*[3]

[1]*Sakarya Uygulamalı Bilimler Üniversitesi, Teknoloji Fakültesi, Mekatronik Mühendisliği, Sakarya, TÜRKİYE*
[2]*Sakarya Uygulamalı Bilimler Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, Sakarya, TÜRKİYE*
[3]*Sakarya Uygulamalı Bilimler Üniversitesi, Yapay Zekâ ve Veri Bilimi Uygulama ve Araştırma Merkezi, Sakarya, TÜRKİYE*

Özet

Her yıl iş yerlerinde baret ve kask takma ihmalinden kaynaklı pek çok iş kazası yaşanmakta, bu durum ciddi yaralanmalara ve hatta ölümlere yol açabilmektedir. Bu çalışma, iş yerlerindeki güvenlik standartlarını sağlamayı, çalışanların girişte baret takıp takmadığını otomatik olarak kontrol ederek ve eksiklik durumunda anında uyarıda bulunarak iş güvenliğini artırmayı hedeflemektedir. Gelişmiş görüntü tanıma teknikleri, özellikle Convolutional Neural Networks (CNN) ve YOLOv7 gibi derin öğrenme algoritmaları, baret ve kask tespiti konusunda büyük potansiyele sahiptir. YOLOv7'nin %84.8 kesinlik oranıyla sağladığı yüksek doğruluk ve hız, iş yerlerinde gerçek zamanlı olarak baret ve kaskları tespit etme kapasitesini artırmaktadır. CNN, iş güvenliği uygulamalarında önemli bir rol oynamakta olup, görüntülerden detaylı özellik çıkarımı yaparak baret ve kaskların hassas bir şekilde tespit edilmesini sağlamaktadır. YOLOv7'nin ise nesne tespiti alanında sağladığı hız ve doğruluk, gerçek zamanlı uygulamalarda kullanım için ideal özellikler sunmaktadır. Bu entegre yaklaşım, iş yerlerinde kazaların önlenmesi ve çalışan güvenliğinin sağlanmasında teknolojinin gücünü kullanarak önemli bir adım olarak değerlendirilmektedir. Çalışma aynı zamanda, endüstriyel ortamlarda kullanılan baret ve kaskların otomatik olarak tanınmasıyla iş güvenliği kültürünün güçlendirilmesine katkıda bulunmayı amaçlamaktadır. Otomatik tanıma sistemleri, iş yerlerindeki iş güvenliği denetimlerini daha etkin hale getirerek, personelin güvenlik standartlarına uyumunu artırmakta ve bu sayede kazaların önlenmesine yardımcı olmaktadır. Bu çalışma, teknoloji destekli çözümler ile iş güvenliği alanında yapılan yeni adımların önemini vurgulayarak, sektördeki genel güvenlik performansını iyileştirmeyi hedeflemektedir.

Anahtar Kelimeler: *Baret-kask tespiti, YOLOv, CNN*

Deep Learning Based Automatic Helmet Detection

Abstract

Every year, numerous workplace accidents occur due to negligence in wearing helmets and caps, leading to serious injuries and even fatalities. This study aims to enhance workplace safety by automatically checking whether employees wear helmets upon entry and providing immediate alerts in case of non-compliance. Advanced image recognition techniques, particularly Convolutional Neural Networks (CNN) and YOLOv7 deep learning algorithms, hold significant potential for helmet and cap detection. YOLOv7's high accuracy rate of %84.8 and its speed contribute to real-time detection of helmets and caps in workplaces. CNN plays a crucial role in safety applications by extracting detailed features from images to accurately identify helmets and caps. YOLOv7's speed and accuracy in object detection make it ideal for

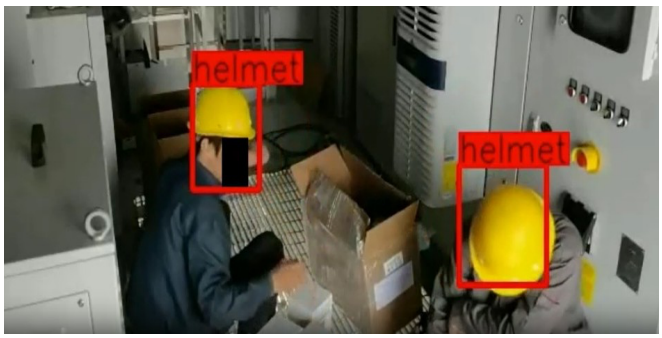
* Contact email: 22010403070@subu.edu.tr

real-time applications. This integrated approach is considered a significant step towards preventing accidents and ensuring employee safety by leveraging technology. The study also aims to strengthen the culture of workplace safety through automatic recognition of helmets and caps used in industrial environments. Automatic recognition systems enhance workplace safety inspections, improve compliance with safety standards among personnel, and thereby contribute to accident prevention. This study underscores the importance of technology-driven solutions in enhancing workplace safety and aims to improve overall safety performance in the industry.

Keywords: *Hard hat-helmet recognize, YOLOv, CNN*

Giriş

Günümüzde, iş yerlerinde kask takmama nedeniyle meydana gelen kazalar da ciddi yaralanmalara ve ölümlere yol açabilmektedir. İş yerlerinde bu tür kazaların önlenmesi ve iş güvenliği standartlarının yükseltilmesi için alınacak önlemler büyük önem taşımaktadır. Aynı şekilde, motosikletler, bisikletler ve elektrikli scooter'lar gibi araçlarda kask kullanımı, ciddi kazaların önlenmesinde kritik bir rol oynamaktadır (Maheswaran et al., 2022; Sasikala et al., 2015). Huang ve diğerleri (2023), elektrikli bisikletler için YOLOv5s tabanlı gelişmiş bir kask tanıma yöntemi sunarak, bu teknolojilerin pratik uygulamalarını vurgulamışlardır (Huang et al., 2023). Benzer şekilde, Jakubec ve arkadaşları (2023), yol güvenliği için derin öğrenme temelli otomatik kask tanıma sistemlerinin etkinliğini araştırmış ve bu teknolojilerin önemini vurgulamışlardır (Jakubec et al., 2023). Son yıllarda, derin öğrenme tekniklerinin gelişmesiyle birlikte, iş güvenliği açısından kritik öneme sahip baret ve kask tespiti için otomatik tanıma sistemleri önem kazanmıştır (Huang et al., 2023; Wang et al., 2022). Özellikle YOLOv5 gibi gelişmiş nesne tespit algoritmaları, gerçek zamanlı kask tanıma sistemlerinin performansını önemli ölçüde artırmıştır (Song et al., 2023; Jakubec et al., 2023). Bu teknolojiler, iş yerlerinde güvenliği artırmak ve çalışanların baret ve kask takma zorunluluğunu kontrol etmek için etkili bir araç olarak kullanılmaktadır. Bu kapsamda, görüntü tanıma teknolojilerinin gücünden faydalanarak, iş güvenliği açısından kritik öneme sahip baret ve kask tespiti için derin öğrenme yöntemlerinden biri olan Convolutional Neural Networks (CNN) kullanılmaktadır. CNN, görüntü tanıma problemlerinde oldukça etkili olup, özellikle iş güvenliği uygulamalarında büyük potansiyele sahiptir. Bununla birlikte, YOLOv (You Only Look Once) nesne tespiti algoritması, hız ve doğruluk açısından üstün performans sergileyerek görüntülerdeki nesneleri hızlı ve doğru bir şekilde tespit edebilmektedir. YOLOv'un sağladığı bu hız ve doğruluk, gerçek zamanlı uygulamalarda büyük avantaj sağlamaktadır. YOLOv'un nesne tespiti kabiliyetleri ve CNN'in derin öğrenme yeteneklerinin birleşimi, iş yerlerindeki baret ve kask tespiti için son derece etkili bir çözüm sunmaktadır. Bu entegre yaklaşım, iş güvenliğinin sağlanmasında ve kask kullanımının teşvik edilmesinde önemli katkılar sunarken, endüstriyel ortamlarda kullanılan baret ve kaskları hassas bir şekilde tanıyarak iş yerlerinde güvenliği artırmayı hedeflemektedir. Uygulama, çalışanların güvenlik standartlarına uyup uymadıklarını anında belirleyerek, eksiklik durumunda anında uyarı verecek şekilde tasarlanmıştır. Böylece, iş yerlerinde kazaların önlenmesi ve çalışanların güvenliğinin sağlanması konusunda büyük bir adım atılmaktadır. İş güvenliği alanında bu tür ileri teknolojilerin kullanımı hem çalışanların sağlığını korumak hem de iş yerlerinin verimliliğini artırmak açısından kritik öneme sahiptir. Çalışmamdan elde edilen bir görüntü Şekil 1'de verilmektedir.



Şekil 1: Kask tespit görüntüsü

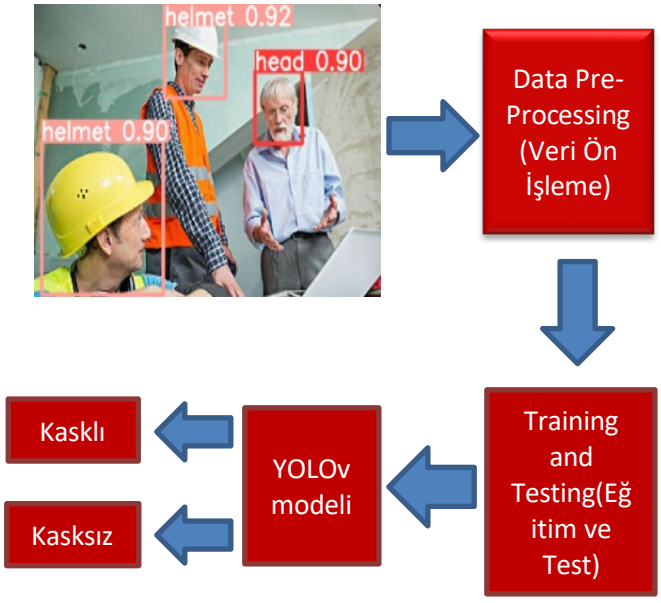
Materyal ve Metot

YOLO

YOLO, "You Only Look Once" (Sadece Bir Kez Bakarsınız) anlamına gelir ve gerçek zamanlı nesne algılama algoritmalarının popüler bir ailesidir. İlk olarak 2016 yılında Joseph Redmon, Ali Farhadi ve Santosh Divvala tarafından tanıtılan orijinal YOLO nesne dedektörü, diğer nesne dedektörlerine kıyasla

çok daha hızlı olmasıyla dikkat çekti ve gerçek zamanlı bilgisayar görüşü uygulamaları için bir dönüm noktası oldu. O zamandan beri, her biri performans ve verimlilikte önemli bir artış sağlayan YOLO'nun farklı sürümleri ve varyantları önerildi. YOLOv1'den popüler YOLOv3'e kadar olan sürümler, o zamanki lisansüstü öğrencisi Joseph Redmon ve danışman Ali Farhadi tarafından oluşturuldu. YOLOv4 , Redmon'un etik kaygılar nedeniyle bilgisayarlı görüş araştırmalarını durdurmasından bu yana mirası sürdüren Alexey Bochkovskiy tarafından 2020'de tanıtıldı. YOLOv7, YOLO mimarisinin orijinal yazarları tarafından oluşturulan ve 2022 yılında yayınlanan en son resmi YOLO sürümüdür. YOLOv7, performans ve verimlilikte önceki sürümleri geride bırakan çeşitli yenilikler ve iyileştirmeler içerir. Bu sürüm, gerçek zamanlı nesne algılama konusunda en ileri düzeyi temsil etmektedir.YOLOv7;güvenlik ve gözetim, inşaat güvenliği, otonom araçlar, akıllı şehir sistemleri gibi endüstriyel uygulamalarda kullanılmaktadır.

Bu çalışmada, iş yerlerinde kullanılan baret ve kaskların otomatik olarak tanınması amacıyla derin öğrenme tabanlı bir yöntem geliştirilmiştir. Özellikle, giriş görüntülerindeki kask örneklerini tanımlamak için bilgisayarla görme alanında son teknoloji bir teknik olan YOLOv7 nesne algılama algoritmasını kullanılmaktadır. İlk olarak, Convolutional Neural Networks (CNN) kullanılarak görüntü verileri işlenmiş ve özellik çıkarımı yapılmıştır. Ardından, YOLOv7 gibi gelişmiş bir nesne tespit algoritması seçilmiş ve model eğitimi için uygun veri setleri oluşturulmuştur. Modelin eğitimi sürecinde, farklı baret ve kask tiplerini içeren geniş bir veri kümesi kullanılarak ayrıntılı bir öğrenme sağlanmıştır. Eğitim sonrasında, elde edilen modelin doğruluğunu artırmak için farklı veri artırma teknikleri uygulanmış ve sonuçlar optimize edilmiştir. Özellikle, modelin gerçek zamanlı uygulamalarda nasıl performans gösterdiğini değerlendirmek için çeşitli test senaryoları kullanılmıştır. Bu süreç, veri toplama, veri ön işleme, model eğitimi ve performans değerlendirmesi aşamalarını içeren kapsamlı bir kask algılama prosedürünü kapsamaktadır. Şekil 2'de gösterilen genel metodoloji akışı, her bir adımın ayrıntılı bir şekilde planlanmasını ve uygulanmasını sağlamıştır.



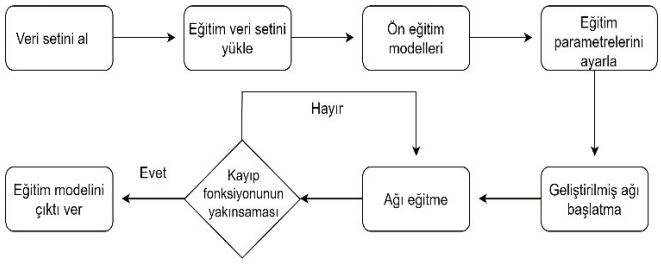
Şekil 2: Genel metodoloji akışı

Veri Seti Oluşturma

Fabrika çalışanlarının baret kullanımına ilişkin açıklamalar içeren mevcut bir görüntü veri kümesi olmadığından, önerilen algoritmanın eğitimi için bir veri kümesi topladık. Baret kullanımını tespit etmek ve tanımlamak için bilgisayarlı görüş algoritmasında kullanılan verilerin toplanması, ön işlenmesi ve açıklaması verilmektedir. Veri kümesinin farklı ışık ve çalışma koşulları altında yakalanmış çeşitli fabrika çalışanlarını içermesi gerektiğinden, nispeten uzun bir zaman dilimi boyunca video verileri toplamaya karar verdik. Fabrika yönetiminin desteğiyle, iki farklı üretim hattında gözlem noktası seçildi . Bu gözlem noktaları, çalışan yoğunluğunun yüksek olduğu ve farklı iş süreçlerinin gerçekleştiği bölgelerde yer aldı. Sonuç olarak, seçilen alanlar hem üretim hem de bakım faaliyetlerinin yoğun olarak yürütüldüğü bölgelerdir. Bu alanlarda çeşitli iş güvenliği uygulamaları ve baret kullanımının yaygın olduğu gözlemlenmiştir. Toplam 300 saatlik video kaydedildi.

Nesne algılama algoritmasının eğitimi, doğrulanması ve test edilmesi için seçilen görüntülerde mümkün olan en büyük çeşitliliğe izin vermek için, görüntüler mevcut 300 saatlik video verilerinden rastgele seçildi. Çizilen görüntüde bir baret takan çalışan varsa, açıklama için saklandı. Daha sonra, model eğitimi, doğrulama ve test için etiketli bir veri kümesi oluşturmak üzere tüm görüntüleri açıkladık. Bu açıklama sırasında, bireysel çalışanların, yani çalışanın ve baretinin etrafına dikdörtgen bir kutu (sınırlayıcı kutu olarak adlandırılır) çizildi. Çalışanın bareti, saçları, şapkaları veya diğer başlıkları

sınırlayıcı kutuya dahil edildi. Daha sonra tüm sınırlayıcı kutular, kutudaki çalışanın baret kullanımıyla, baretli veya baretsiz olarak etiketlendi. Bu işlem kullanılarak, 3024 görüntüde 3408 bireysel çalışan açıklandı (bazı görüntülerde birden fazla çalışan bulunduğu için). Veri ön işleme ve model eğitimi süreci, baret ve kask tespiti için oldukça kritiktir. Bu süreç, veri toplama, ön işleme, model eğitimi ve değerlendirme aşamalarını kapsamaktadır. Şekil 3 de YOLOv7 eğitim süreci akış diyagramı verilmiştir.



Şekil 3: YOLOv7 eğitim süreci akış diyagramı

Aktif işçilerin baret kullanımını kaydetmekle ilgilendiğimiz için, algoritmayı iki sınıfı tespit edecek şekilde eğittik: baret takan işçiler ve baret takmayan işçiler. Baret takan işçiler için mAP=0,915 ve baret takmayan işçiler için mAP=0,894 olarak bulduk. Baret takan işçiler için 0,864'lük bir hassasiyet kaydedildi, yani algoritma baret takan bir işçi tespit ettiğinde tespitlerin %86,4'ünde doğrudur. Hassasiyet değeri baret takmayan işçiler için benzerdir (hassasiyet=0,835), algoritmanın baret takmayan bir işçi tespit ettiği durumların %83,5'inde tespit doğrudur. Baret tespitinin hassasiyeti nispeten yüksek olsa da, geri çağırma, yani test veri setinde baretli ve baretsiz tespit edilen işçilerin payı daha düşüktür ve baretli işçiler için %75,1'e, baretsiz işçiler için ise %74'e ulaşmaktadır. Baretli/baretsiz tespiti için ortalama hassasiyet 0,849 iken, ortalama geri çağırma 0,744'tür. Bu, algoritmanın FN'den (False Negatives) daha az FP (False Positives) tespiti ürettiği anlamına gelir; yani veri setindeki işçilerin çoğunu tespit eder, ancak diğer nesneleri baretli veya baretsiz işçiler olarak da tespit eder. Tablo 1' de bu değerler gösterilmiştir.

Sınıf	mAP	Hassasiyet	Geri Çağırma
Baretli İşçiler	0.915	0.864	0.751
Baret Takmayan İşçiler	0.894	0.835	0.740
Ortalama	0.849	-	0.744

Tablo 1

Veri setinden örnek görüntüler şekil 3 ve şekil 4'te verilmiştir.



Şekil 4,5:Veri setindeki örnek görüntüler

Araştırma ve Bulgular

Nesne tanıma, bilgisayarla görme yarışmaları ve zorlukları sürekli olarak doğruluk sınırlarını zorlayan gelişen bir alandır (Everingham ve diğerleri, 2010; 2015; Russakovsky ve diğerleri, 2015). Sınıflandırma modelleri için çıktı, nesne sınıfı ve modelin güven ölçüsü ile basittir. Ancak, nesne algılama modelleri, her tahmin için bir dizi güven ölçüsü içeren sınırlayıcı kutular, nesne sınıfları gibi daha karmaşık çıktılar üretir. Bir sınırlayıcı kutu, tespit edilen nesneyi kapsayan dikdörtgen bir alan olup, üst sol ve alt sağ köşelerinin koordinatları ile belirtilir.

Ortalama ortalama hassasiyet (mAP) değeri, muhtemelen farklı çok sınıflı dedektörlerin kalitesini karşılaştırmak için ana ölçütlerden biridir. mAP metriğini açıklamak için, ilk olarak aşağıdaki temel kavramları tanımlamak gereklidir:

- **Intersection over Union (IoU),** tespit edilen sınırlayıcı kutu BpB_pBp ile gerçek sınırlayıcı kutu BgtB_{gt}Bgt'nin kesişim alanının birliğe oranını ifade eder, Denklem (1).

$$IOU = alan(Bp \cap \text{Bgt})/alan(Bp \cup Bgt) \quad (1)$$

- **Duyarlılık** modelin tüm mevcut nesneleri tespit etme yeteneğini tanımlar. Değeri, doğru tespit edilen nesnelerin sayısının veri kümesindeki tüm nesne açıklamalarının sayısına oranı olarak hesaplanır, Denklemi (2).

$$Recall = TP/(TP + FN) \quad (2)$$

- **Hassasiyet** modelin sadece mevcut nesneleri tanımlama yeteneğini tanımlar. Değeri, doğru

tespit edilen nesnelerin sayısının modelin tüm tespitlerinin sayısına oranıdır, Denklem (3).

$$Recall = TP/(TP + FN) \quad (3)$$

Ortalama hassasiyet (AP), bu eğrinin altındaki alan olarak hesaplanır ve 0 ile 1 arasında değişir, 1 ideal değeri temsil eder. Sistemin genel performansını değerlendirmek için, mAP metriği her öğe türü için AP hesaplanarak ve ardından ortalama alınarak hesaplanır. mAP metriği, bir tahminin doğruluğunu değerlendirmek için kullanılan IoU eşiğine dayalı birkaç alt türe sahiptir. En yaygın kullanılan alt türler, IoU eşiğinin 0.5 olduğu mAP(0.5) ve eşiğin 0.5'ten 0.95'e kadar 0.05 adımlarla değiştiği ve ortalamanın hesaplandığı mAP(0.5:0.05:0.95)'dir (Lin ve diğerleri, 2015).Denklem 4.

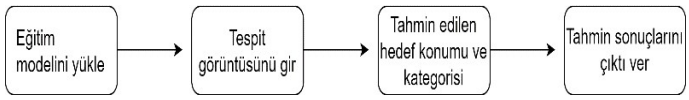
$$\mathrm{mAP} = \frac{1}{n} * \sum_{i=1}^{n} APi \quad (4)$$

YOLOv7 modelleri, nesne algılama görevlerinde yüksek performans ve hız sağlamak için tasarlanmıştır. Bu modeller, önceden eğitimli ağırlıklar kullanılarak veya kullanılmadan test edilmiştir. Performans ölçütleri arasında mAP (mean Average Precision) ve FPS (Frames Per Second) bulunmaktadır. Parametre sayısı arttıkça, modellerin doğruluk oranlarının da arttığı gözlemlenmiştir. Aşağıdaki tablo 2'de YOLOv7 türlerinin ön eğitim ağırlıkları kullanılarak ve kullanılmadan elde edilen değerleri verilmiştir

Model	Ön Eğitim Ağırlıkları Yok	Ön Eğitim Ağırlıkları Var
YOLOv7-tiny	90%	92%
YOLOv7	92%	94%
YOLOv7x	93%	95%
YOLOv7s	91%	93%

Tablo 2

YOLOv7 modelleri, önceden eğitimli ağırlıklarla kullanıldığında daha yüksek doğruluk oranlarına ulaşmaktadır. Bu, modellerin daha fazla veriyi ve özellikleri tanıma yeteneğini artırmaktadır. Özellikle YOLOv7x, en yüksek doğruluk oranına sahip model olarak öne çıkmaktadır. Şekil 6'da YOLOv7 tespit süreci akış diyagramı verilmiştir



Şekil 6: YOLOv7 tespit süreci akış diyagramı

Genel olarak, bu çalışmalar, derin öğrenme modellerinde önceden eğitimli ağırlıkların önemini ve performans artışını vurgulamaktadır. YOLOv7 modelleri, hem hız hem de doğruluk açısından güçlü bir seçenek sunar.Bu tür modeller, nesne algılama uygulamaları için esnek ve etkili çözümler sunarak geniş bir kullanım alanına sahiptir.

Sonuç

Bu çalışmada, iş yerlerinde iş güvenliğini artırmak ve çalışanların baret ve kask takma zorunluluğunu denetlemek amacıyla YOLOv nesne tespiti algoritmaları kullanılarak Convolutional Neural Networks (CNN) ile entegrasyon sağlanmıştır. YOLOv7'nin hızlı ve doğru nesne tespiti kabiliyetleri ile CNN'in derin öğrenme yeteneklerinin birleşimi, baret ve kask tespiti gibi iş güvenliği açısından kritik uygulamalarda yüksek doğruluk oranları sağlamaktadır. Dört adet YOLOv7 türevi incelenmiş ve ön eğitim ağırlıkları kullanılmadan ortalama doğruluk oranı ~%91,5 iken, ön eğitim ağırlıkları kullanılarak ortalama doğruluk oranı ~%93,5 olarak hesaplanmıştır. Bu teknolojik yaklaşım, iş yerlerinde çalışanların güvenlik standartlarına uyumunu anında kontrol ederek iş güvenliğini artırırken, iş kazalarını önleme konusunda önemli bir adım sağlamaktadır. Bilgisayarla görme teknolojilerinin iş güvenliği uygulamalarına entegrasyonu, güvenli çalışma ortamlarının sağlanması ve iş kazalarının minimize edilmesi açısından büyük potansiyele sahiptir. Bu tür yenilikçi çözümler, iş yerlerinin daha güvenli hale gelmesine katkıda bulunarak, çalışan sağlığını korumak ve iş verimliliğini artırmak için kritik bir rol oynamaktadır. Ayrıca, yalnızca fabrikalarda değil, gündelik yaşamda da kask takmanın önemi büyüktür. Motosiklet kullanırken, bisiklet sürerken veya benzeri birçok durumda kask kullanımı hayati önem taşır ve bu durumların da sıkı bir şekilde denetlenmesi gerekmektedir. Güvenliğin her alanda sağlanması için kask kullanımının teşvik edilmesi ve denetimlerin artırılması, kazaların ve yaralanmaların önlenmesine önemli ölçüde katkıda bulunacaktır.

Kaynaklar

[1] Maheswaran, S., et al. (2022). YOLOV5 tabanlı gerçek zamanlı otomatik plaka ve kask tanıma sistemi. In Proceedings of the 2022 13th International Conference on Computer Communication and Networks Technologies (ICCCNT) (pp. 1-7). Kharagpur, India.

[2] Sasikala, G., Padol, K., Katekar, A. A., & Dhanasekaran, S. (2015). Safeguarding of motorcyclists through helmet recognition. In *Proceedings of the 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)* (pp. 609-612). Avadi, India.

[3] Huang, B.; Wu, S.; Xiang, X.; Fei, Z.; Tian, S.; Hu, H.; Weng, Y. An Improved YOLOv5s-Based Helmet Recognition Method for Electric Bikes. *Appl. Sci.* 2023, *13*, 8759.

[4] Wang, L., Zhang, H., Li, Y., Xu, W., & Chen, Z. (2022). Investigation into recognition algorithm of helmet violation based on YOLOv5-CBAM-DCN. *IEEE Access, 10*, 60622-60632. Marayatr, T., & Kumhom, P. (2014). Görüntü İşleme Kullanılarak Motosikletçinin Kask Taktığının Tespiti. *Gelişmiş Malzeme Araştırması*, 931–932, 588–592.

[5] Li, J., Wang, J., Zhang, H., Liu, Y., & Wu, Z. (2017). Safety helmet wearing detection based on image processing and machine learning. In *Proceedings of the 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)* (pp. 201-205). Doha, Qatar.

[6] Rubaiyat, A. H. M., Rahman, M. M., Hossain, M. S., Islam, R., & Khan, M. S. (2016). Automatic detection of helmet uses for construction safety. In *Proceedings of the 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW)* (pp. 135-142). Omaha, NE, USA.

[7] Silva, R. R. V. e., Aires, K. R. T., & Veras, R. d. M. S. (2014). Helmet detection on motorcyclists using image descriptors and classifiers. In *Proceedings of the 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images* (pp. 141-148). Rio de Janeiro, Brazil.

[8]] Li, Y., Wei, H., Han, Z., Huang, J., & Wang, W. (2020). Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks. *Advances in Civil Engineering*, 2020.

[9] Benyang, D., Xiaochun, L., & Miao, Y. (2020). YOLO v4'e dayalı koruyucu kask algılama yöntemi. In 2020 16. Uluslararası Hesaplamalı Zeka ve Güvenlik Konferansı (CIS) (s. 155-158). Guangxi, Çin. IEEE.

[10] Corerain Technologies. (n.d.). Helmet recognition. Retrieved July 3, 2024.

[11] Wu, D., Song, W., Wang, M., Liu, J., Fu, X., Sun, H., & Zhao, M. (2022). Real-Time ISR-YOLOv4 Based Small Object Detection for Safe Shop Floor in Smart Factories. *Electronics*, *11*(15), 2348.

[12] Song, H., Zhang, X., Song, J. *et al.* Detection and tracking of safety helmet based on DeepSort and YOLOv5. *Multimed Tools Appl* 82, 10781–10794 (2023).

[13] Jakubec, M., Lieskovska, E., Brezani, A., & Tothova, J. (2023). Deep Learning-Based Automatic Helmet Recognition for Two-Wheeled Road Safety. *Transportation Research Procedia, 74,* 1171-1178.

[14] Viso.ai. (2024). *YOLOv7: En Hızlı Nesne Algılama Algoritması.*

[15] Training Data Pro. (2023). Helmet Detection [Dataset]. Kaggle.]

Hanehalkı Düzeyinde Pandemi Kırılganlığının Ölçüm ve Analizi: Kayseri İlinde Bir Araştırma

Yasemin YAVUZ[1*], Sabri GÜNGÖR[2], Esma İNERCİ[3]

[1,2,3]*Erciyes Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, Kayseri, TÜRKİYE*

Özet

Tarih boyunca çeşitli salgın afetleri daha önceden de yaşanmış olmakla birlikte, COVID-19 pandemisinde tüm dünya ülkeleri şimdiye kadar yaşanmamış ölçekte bir krizle baş etmek durumunda kalmıştır. Pandeminin etkisi büyük ölçüde atlatılmış olmakla birlikte, iklim değişikliği, nüfus yoğunluğu ve aşırı kentleşme gibi günümüzün önemli küresel sorunları, gelecekte ülkeleri artan düzeyde bir salgın afeti riski ile karşı karşıya bırakmaktadır. Buna rağmen, küresel düzeydeki araştırma raporları ülkelerin gelecekteki pandemi risklerine karşı yeterince hazırlıklı olmadığını göstermektedir.

Bu çalışmada sosyal kırılganlık kavramı çerçevesinde hanehalkı pandemi kırılganlığının ölçüm ve analizi amaçlanmaktadır. Türkiye'de pandemi kırılganlığının ölçümü ile ilgili yapılan çalışmalar çok kısıtlıdır. Az sayıdaki çalışmaların kapsamı ikincil veriler kullanılarak iller bazında endeks geliştirilmesine yöneliktir. Bu çalışmanın özgün yönü ve literatüre katkısı, pandemi kırılganlığının hanehalkı düzeyinde incelenmiş olmasıdır. Hanehalkı pandemi kırılganlık (HPK) düzeyinin ölçümü ve analizine yönelik Türkiye'de daha önceden gerçekleştirilmiş benzer nitelikte bir çalışmaya rastlanmamıştır. Çalışmanın amacı doğrultusunda HPK Ölçeği geliştirilerek Kayseri'nin merkez ilçelerindeki 399 hanehalkına kotalı örnekleme yöntemi ile uygulanmıştır.

Elde edilen anket verileri ile Kayseri için genel HPK düzeyi araştırılmıştır. Bununla birlikte hem genel düzeyde hem de demografik, sağlık, ekonomik, ulaşım ve fiziksel/sosyal çevre şeklinde belirlenen pandemi kırılganlık göstergeleri bakımından Kayseri'nin merkez ilçeleri için HPK düzeyleri arasındaki farklılıklar incelenmiştir. Yapılan analizler sonucunda hem genel HPK düzeyleri bakımından hem de pandemi kırılganlık göstergeleri bakımından merkez ilçelerin bir kısmı arasında anlamlı farklılıklar tespit edilmiştir.

Anahtar Kelimeler: *Hanehalkı Pandemi Kırılganlığı, Sosyal Kırılganlık, Salgın Afetleri, COVID-19, Afet Risk Yönetimi*

Measurement and Analysis of Household Pandemic Vulnerability: A Study in Kayseri Province

Abstract

Although various pandemic disasters have been experienced before throughout the history, all countries in the world had to deal with a crisis on an unprecedented scale in the COVID-19 pandemic. Although the impact of the pandemic has been largely overcome, today's important global problems such as climate change, population density, and excessive urbanization leave countries facing an increasing risk of pandemic disasters in the future. Despite this, global research reports show that countries are not sufficiently prepared for future pandemic risks.

This study aims to measure and analyze household pandemic vulnerability within the framework of the concept of social vulnerability. The studies conducted in Türkiye on the measurement of pandemic vulnerability are very limited. The scope of the few studies is

* Contact email: yaseminy@erciyes.edu.tr

aimed at developing an index on a provincial basis by using secondary data. The originality of this study and its contribution to the literature is that pandemic vulnerability is examined at the household level. No similar study has been found in Türkiye on the measurement and analysis of the household pandemic vulnerability (HPV) level. In line with the purpose of the study, the HPV Scale has been developed and applied to 399 households in the central districts of Kayseri using the quota sampling method.

The general HPV level for Kayseri has been investigated with the survey data obtained. In addition, the differences between the HPV levels for the central districts of Kayseri have been examined both at the general level and in terms of pandemic vulnerability indicators determined as demographic, health, economic, transportation, and physical/social environment. As a result of the analyses, the significant differences have been found between some of the central districts in terms of both general HPV levels and pandemic vulnerability indicators.

Keywords: *Household Pandemic Vulnerability, Social Vulnerability, Pandemic Disasters, COVID-19, Disaster Risk Management*

1 Giriş

Afetler toplumun normal yaşamını durdurarak veya kesintiye uğratarak toplulukları etkileyen ve üstesinden gelmekte zorluk çekilen olaylardır. Tarih boyunca geçmişte yaşanmış olan çeşitli afet türleri; jeolojik, klimatik, biyolojik, sosyal ve teknolojik afetler olmak üzere beş kategoride sınıflandırılabilir [1]. Birleşmiş Milletler Afet Risklerinin Azaltılması 3. Dünya Konferansında kabul edilen Sendai Afet Risk Azaltma Çerçevesi (2015-2030), doğa-insan kaynaklı, küçük-büyük ölçekli, sık-seyrek, ani-yavaş gelişen afetlerin yanı sıra çevresel, teknolojik, biyolojik tehlike ve riskleri de afet risk yönetiminin kapsamına dahil etmiştir. Biyolojik afetler arasında yer alan salgınlar, doğal nedenlerden kaynaklanabileceği gibi, biyolojik saldırılar ve kazalar gibi insan kaynaklı tehlikelerden dolayı da ortaya çıkabilmektedir. Ayrıca, sağlık koşullarının uygun olmaması nedeniyle diğer doğal afetlerin yaşanmasının ardından afet bölgesinde çeşitli salgın hastalıklar görülebilmektedir. Salgın hastalıkların ortaya çıkması, yayılması ve hızlanmasında iklim değişikliği, nüfus yoğunluğu ve aşırı kentleşme gibi günümüzdeki önemli sorunların da rol oynadığı bilinmektedir. Yukarıda özetlenen tüm bu hususlar, gelecekte ülkeleri artan düzeyde bir salgın afeti riski ile karşı karşıya bırakmıştır. Dünyanın çeşitli bölgelerinde salgın afetleri daha önceden yaşanmış olmakla birlikte, tüm dünyayı etkisi altına alan COVID-19 pandemisi söz konusu riskin yakın geçmişte yaşanan bir örneğini oluşturmaktadır.

Uzun süren bir pandemi döneminin ardından gerek akademik çalışmalar gerekse küresel düzeydeki raporlar, geçmiş tecrübelerden ders çıkararak gelecekteki pandemi risklerine karşı alınması gereken önlemlere dikkat çekmektedir. Aralık 2021'de yayınlanan Küresel Sağlık Güvenliği Endeksi'ne göre 195 ülkenin pandemi hazırlık düzeyi 100 üzerinden ortalama 38.9 düzeyinde hesaplanmış olup, bu rakam ülkelerin gelecekteki pandemi risklerine karşı yeterince hazırlıklı olmadığını göstermektedir [2]. Bölgesel düzeyde yaşanan diğer afetlerden farklı olarak, pandemi tüm ülke genelinde hızlı bir şekilde yayılarak toplumunun tamamını aynı anda etkilemekte ve söz konusu etkiler uzun bir süre boyunca devam etmektedir. Bu nedenle, gelecekte yaşanabilecek pandemi afetlerine etkin bir müdahale için ülkelerin pandemi hazırlık kapasitelerini güçlendirilmesi büyük önem arz etmektedir. Günümüzde önemi giderek artan "afete dayanıklı toplum" kavramı, pandemi afetlerine karşı dayanıklılığın artırılmasında da ön plana çıkmaktadır.

UNISDR (2009)'nin tanımına göre afet dayanıklılığı, bir tehlikeye maruz kalmış bir sistemin ya da toplumun temel yapılarının korunması ve yenilenmesi de dâhil olmak üzere tehlikenin etkilerini zamanında ve etkili bir şekilde soğurma, eski haline dönme ve iyileşme kabiliyetidir. Dayanıklılıkla doğrudan ilişkili olan kırılganlık kavramı ise bireyin, topluluğun veya sistemlerin tehlikelerin etkilerine duyarlılığını, hassasiyetini arttıran fiziksel, sosyal, ekonomik veya çevresel faktörleri ifade eder [3]. Genel olarak bir toplumun afete karşı dayanıklılığı ne kadar fazlaysa afetin olumsuz etkileri ile başa çıkabilme kapasitesi de o oranda artmaktadır. Ülkelerin afetlere karşı dayanıklılığının artırılmasını amaçlayan Sendai Afet Risk Azaltma Çerçevesi'nde afet kırılganlığını azaltmak için ekonomik, yapısal, yasal, sosyal, sağlık, eğitim, kültürel, çevresel, teknolojik, politik

ve kurumsal önlemlerin entegre ve kapsayıcı bir şekilde uygulanması gerektiği belirtilmiştir. Ayrıca, "afet risk yönetimine yönelik politika ve uygulamalarda kırılganlık dahil olmak üzere afet risklerinin tüm boyutlarının göz önünde bulundurulması" ulusal düzeyde önem arz eden konular arasında sıralanmıştır [4]. Daha önceden sadece fiziksel risklerin dikkate alındığı afet risk yönetimi çalışmaları, günümüzde sosyoekonomik faktörler başta olmak üzere sosyal kırılganlık kapsamında genişletilmiştir. Sosyal kırılganlık bu bağlamda bireylerin ve grupların afet öncesi durumları, afetlerle başa çıkma güçleri ve afetlerin yarattığı etkileri azaltmaya yönelik kapasitelerine ilişkin bir kavram olarak kullanılmaktadır [5]. Bu çalışmada da sosyal kırılganlık kavramı çerçevesinde pandemi kırılganlığının hanehalkı düzeyinde ölçümü ve analizi ele alınmıştır.

Ekonomik, sosyal, kültürel ve çevresel yıkıcı etkileri nedeniyle afetler toplum refahını ve sürdürülebilir kalkınmayı olumsuz yönde etkileyen önemli kayıplara neden olmaktadır. Afet risklerinin azaltılması sürdürülebilir kalkınma için esastır [4]. Bu nedenle, Birleşmiş Milletler tarafından 17 küresel amacın belirlendiği Sürdürülebilir Kalkınma Amaçlarında da ülkelerin afetlere karşı dayanıklılığının artırılması hedeflenmiştir. Bu çalışmada ele alınan konunun ilgili olduğu Sürdürülebilir Kalkınma Amaçları Tablo 1'de sunulmuştur [6]:

Tablo 1. Çalışma Konusu İle İlgili Sürdürülebilir Kalkınma Amaçları

Amaç 1: Yoksulluğa Son Yoksulluğun tüm biçimlerini her yerde sona erdirmek **Hedefler** Hedef 1.3: Ulusal koşullara uygun sosyal koruma sistemleri ve önlemlerini, asgari güvenceleri de içerecek biçimde, herkes için uygulamaya koymak ve 2030 yılına kadar yoksul ve kırılgan kesimin büyük ölçüde kapsanmasını sağlamak Hedef 1.5: 2030 yılına kadar yoksul ve kırılgan durumda olanlara dayanıklılık kazandırmak ve iklimle ilgili aşırı olaylara ve diğer ekonomik, sosyal ve çevresel şoklara ve afetlere maruziyet ve kırılganlıklarını azaltmak **Göstergeler** 1.3.1: Çocuklar, işsizler, yaşlılar, engelliler, hamileler, yeni doğanlar, iş kazasında yaralananlar, yoksul ve kırılgan durumdakiler ayrımında ve cinsiyete göre sosyal koruma sistemleri tarafından korunan nüfusun oranı 1.5.3: Sendai Afet Riski Azaltma Çerçevesi, 2015-2030 ile uyumlu olarak ulusal afet riski azaltma stratejilerini kabul eden ve uygulayan ülkelerin sayısı
1.5.4: Ulusal afet riski azaltma stratejileri ile uyumlu yerel afet riski azaltma stratejilerini kabul eden ve uygulayan yerel yönetimlerin oranı
Amaç 3: Sağlıklı ve Kaliteli Yaşam Sağlıklı ve kaliteli yaşamı her yaşta güvence altına almak **Hedefler** Hedef 3.3: 2030 yılına kadar AIDS, tüberküloz, sıtma ve ihmal edilen tropik hastalık salgınlarını bitirmek ve hepatit, su kaynaklı hastalıklar ve diğer bulaşıcı hastalıklarla mücadele etmek Hedef 3.d: Başta gelişmekte olan ülkeler olmak üzere tüm ülkelerin ulusal ve küresel sağlık risklerine karşı erken uyarı, riski azaltma ve risk yönetimi kapasitelerini güçlendirmek **Göstergeler** 3.d.1: Uluslararası Sağlık Tüzüğü konusundaki kapasite ve sağlık acil durumlarına hazır olma oranı
Amaç 11: Sürdürülebilir Şehirler ve Topluluklar Şehirleri ve insan yerleşimlerini kapsayıcı, güvenli, dayanıklı ve sürdürülebilir kılmak **Hedefler** Hedef 11.5: 2030 yılına kadar yoksulları ve kırılgan durumdaki insanları korumaya odaklanarak, su kaynaklı afetler de dâhil olmak üzere afetlerden kaynaklanan ölümleri ve etkilenen insan sayısını kayda değer miktarda azaltmak ve ekonomik kayıpların küresel GSYH içerisindeki göreli payını büyük oranda düşürmek Hedef 11.b: 2020 yılına kadar, kapsayıcılığı, kaynak verimliliğini, iklim değişikliğine uyumu, iklim değişikliğinin azaltılmasını ve afetlere karşı dayanıklılığı dikkate alan bütüncül politika ve planları benimseyen ve uygulayan kent ve yerleşim yeri sayısını önemli ölçüde artırmak ve Sendai Afet Risk Azaltımı Çerçeve Belgesi 2015-2030 ile uyumlu bütüncül afet riski yönetimini her düzeyde geliştirmek ve uygulamak **Göstergeler** 11.b.1: Sendai Afet Riski Azaltma Çerçevesi, 2015-2030 ile uyumlu olarak ulusal afet riski azaltma stratejilerini kabul eden ve uygulayan ülkelerin sayısı 11.b.2: Ulusal afet riski azaltma stratejileri ile uyumlu yerel afet riski azaltma stratejilerini kabul eden ve uygulayan yerel yönetimlerin oranı
Amaç 13: İklim Eylemi İklim değişikliği ve etkileri ile mücadele için acilen eyleme geçmek **Hedefler** Hedef 13.1: Tüm ülkelerde iklim değişikliğiyle ilgili tehlikeler ile doğal afetlere karşı dayanıklılık ve uyum kapasitesini güçlendirmek **Göstergeler** 13.1.2 : Sendai Afet Riski Azaltma Çerçevesi, 2015-2030 ile uyumlu olarak ulusal afet riski azaltma stratejilerini kabul eden ve uygulayan ülkelerin sayısı 13.1.3 : Ulusal afet riski azaltma stratejileri ile uyumlu yerel afet riski azaltma stratejilerini kabul eden ve uygulayan yerel yönetimlerin oranı

Yukarıda belirtilen Sürdürülebilir Kalkınma Amaçları ile ilgili hedeflerin afet dayanıklılığını artırmaya ve kırılgan grupları korumaya odaklandığı görülmektedir. Pandemi döneminde T.C. Sağlık Bakanlığı tarafından hazırlanan Türkiye COVID-19 Sağlıkta Acil Durum Projesi Paydaş Katılım Planında, aşağıdaki gruplarla sınırlı olmadığı belirtilmekle birlikte kırılgan gruplar şu şekilde sıralanmıştır [7].

- Kronik hastalıkları veya önceden mevcut olan bir koşulu olan kişiler ve yaşlılar
- Engelli bireyler
- Hamileler, bebekler ve çocuklar
- Türkçede sınırlı dil yeterliliğine sahip vatandaşlar, mülteciler, göçmenler
- Yoksulluk sınırı altında yaşayanlar
- İşsizler ve evsizler
- Evin reisinin kadın olduğu haneler ve/veya çocukları yasal sorumluluk yaşının altında olan bekâr anneler
- Düşük gelirli geniş aileler
- Yetimhane çalışanları ve sakinleri, yaşlılar, çocuklar ve diğer bakım kurumları ve özel sektör hizmet sağlayıcıları
- Aile içi şiddet riski altındaki kişiler

Afet kırılganlığı alanındaki literatürdeki çalışmalarda çeşitli göstergeler ve değişkenler temelinde belli bir ülke veya bölgeye yönelik kırılganlık endeksleri geliştirilmiştir. COVID-19 pandemisinin ardından bu alandaki çalışmaların pandemi afetlerine yöneldiği görülmekle birlikte, pandemi kırılganlığına ilişkin çalışmalar oldukça kısıtlıdır. [8], pandemi kırılganlık endekslerinde kullanılan çeşitli değişkenleri Nüfus ve Demografi, Kentsel Yaşam, Ekonomi, İklim ve Çevre, Sağlık olmak üzere beş temel gösterge çerçevesinde sınıflandırmıştır. Bu gösterge ve değişkenler Tablo 2'de sunulmuştur:

Tablo 2. Pandemi Kırılganlık Endekslerinde Kullanılan Çeşitli Gösterge ve Değişkenler

Göstergeler	**Değişkenler**
Nüfus ve Demografi	0-14 yaş nüfus oranı, 65 yaş üstü nüfus oranı, ortalama hanehalkı büyüklüğü, fert başına düşen oda sayısı, bin kişi başına düşen otomobil sayısı, engelli oranı, bin kişi başına düşen hükümlü sayısı, ilkokul mezunu oranı, okuma yazma bilmeyen oranı, mültecilerin toplam nüfusa oranı
Kentsel Yaşam	Genel nüfus yoğunluğu, kentsel nüfus yoğunluğu, kentsel alan oranı, il geneli yapılaşma oranı, atık hizmeti verilen nüfusun oranı, kanalizasyon ve şebeke suyuna erişim, yaşam kalitesi endeksi, memnuniyet oranı
Ekonomi	Kişi başına düşen GSYH, istihdam oranı, işsizlik oranı, aktif sigortalı oranı, yoksulluk oranı
İklim ve Çevre	Sıcaklık, güneşlenme süresi, nem, rüzgâr hızı, hava kirliliği
Sağlık	Aile hekimliği birimi başına nüfus, bin kişi başına düşen hekim sayısı, bin kişi başına düşen hemşire sayısı, yüz bin kişiye düşen yatak sayısı, yüz bin kişiye düşen yoğun bakım yatağı sayısı, 112 ambulans başına nüfus, vefat oranı

Kaynak: [8]

Türkiye için geliştirilmiş olan iller bazındaki kırılganlık endekslerine ilişkin çalışmaların sonuçları aşağıda özetlenmiştir:

Ortakavak vd. (2020) tarafından COVID-19 pandemisine yönelik sosyal kırılganlık endeksi geliştirilerek illerin kırılganlık düzeyleri belirlenmiştir. En yüksek ve en düşük endeks değerleri sırasıyla 5.79 (İstanbul) ve -4.70 (Muş) olup, Kayseri 1.83 endeks değeri ile Türkiye'deki 21 yüksek kırılganlık düzeyine sahip iller arasında yer almaktadır. Sonuçlar nüfus yoğunluğunun, mülteci sayısının yüksek olduğu ve sağlık hizmetlerinin nüfusa yetersiz kaldığı kentlerin daha kırılgan olduğunu ortaya koymaktadır [9].

Yaprak vd. (2021) tarafından afetlere yönelik sosyo-ekonomik kırılganlık endeksi geliştirilmiştir. Türkiye ortalamasının 0.38 olarak elde edildiği sonuçlara göre, en yüksek ve en düşük endeks değerleri sırasıyla 0.64 (Batman) ve 0.25 (Amasya) olarak hesaplanmıştır. Kayseri ise 0.48 endeks değeri ile genel ortalamanın üzerinde olup yüksek kırılganlık düzeyine sahip iller arasında yer almaktadır. Sonuçlar yoksulluk, gelir dağılımı, işsizlik gibi faktörlerin risklere karşı savunmasızlığı artırdığını göstermektedir [10].

Kırlangıçoğlu (2022) tarafından pandemi kırılganlık endeksi geliştirilerek illerin pandemiye karşı kırılganlık seviyeleri ölçülmüştür. Çalışmada en yüksek ve en düşük endeks değerleri sırasıyla 58.29 (İstanbul) ve 1.44 (Tokat) olarak elde edilmiştir. Kayseri ise 3.55 endeks değeri ile orta seviyede kırılganlık düzeyine sahip iller arasında

yer almaktadır. Sonuçlar genel nüfus yoğunluğunun salgın hastalıklara yakalanma riskini artırdığını ortaya koymaktadır [8].

Türkiye genelinde yapılan bu çalışmaların yanı sıra ilçeler ve mahalleler düzeyinde gerçekleştirilmiş olan nitel araştırma [11] ve endeks geliştirme [12] çalışmaları da bulunmaktadır. Kırlangıçoğlu (2022), uluslararası literatürdeki endeks çalışmalarında mahalle ve ilçe bazında verilerin kullanılarak doğruluk seviyesi yüksek sonuçların alındığını belirtmiştir. Verilerdeki detay seviyesi arttıkça ve ölçek büyüdükçe kullanılan modelin hassasiyetinin arttığı görülmektedir [8]. Türkiye'de pandemi kırılganlığının ölçümü ile ilgili literatürdeki çalışmalar oldukça kısıtlı olup söz konusu çalışmalar ise ikincil veriler kullanılarak iller bazında endeks geliştirilmesine yöneliktir. Bu çalışmanın özgün yönü ve literatüre katkısı, pandemi kırılganlığını hanehalkı düzeyinde incelenmiş olmasıdır. Hanehalkı pandemi kırılganlık (HPK) düzeyinin ölçümü ve analizine yönelik Türkiye'de daha önceden gerçekleştirilmiş benzer nitelikte bir çalışmaya rastlanmamıştır.

Çalışmanın amacı doğrultusunda HPK Ölçeği geliştirilerek Kayseri merkez ilçelerindeki (Melikgazi, Kocasinan, Talas) hanehalkına uygulanmıştır. Araştırmada Kayseri ilinin seçilme nedeni, daha önceki araştırma sonuçlarından görüldüğü üzere Kayseri'nin orta-yüksek kırılganlık düzeyine sahip iller arasında yer almasıdır. TÜİK Adrese Dayalı Nüfus Kayıt İstatistiklerine göre Kayseri'nin 2024 yılı toplam nüfusu 1.445.683 olup nüfusun %80'i üç merkez ilçe olan Melikgazi, Kocasinan ve Talas'ta yaşamaktadır [13]. Bu nedenle çalışma nüfus yoğunluğunun yüksek olduğu üç merkez ilçede yürütülmüştür. Çalışma kapsamında cevabı aranan araştırma soruları aşağıdaki gibidir:

1) *Kayseri ili merkez ilçelerinin genel HPK düzeyi nedir?*
2) *Her bir merkez ilçenin ortalama HPK düzeyi nedir? Merkez ilçeler arası HPK düzeyleri farklılık göstermekte midir?*
3) *Merkez ilçelerdeki farklı mahallelerin ortalama HPK düzeyi nedir? Mahalleler arası HPK düzeyleri farklılık göstermekte midir?*
4) *Pandemi kırılganlık göstergeleri bakımından ortalama HPK düzeyi nedir? Göstergeler arası HPK düzeyleri farklılık göstermekte midir?*
5) *Her bir gösterge bazında pandemi kırılganlığına en fazla etki eden değişken hangisidir?*
6) *Katılımcılar hangi değişkenlerin HPK düzeyini en çok etkilediğini düşünmektedir?*

Sosyal kırılganlık kavramı çerçevesinde HPK düzeyinin ölçülerek analiz edilmesi bu çalışmanın temel amacını oluşturmaktadır. Bu amaç doğrultusunda aşağıda listelenen 3 temel hedef belirlenmiştir:

1) Literatür araştırması ve hanehalkına yönelik ikincil verilerin analizi sonucunda HPK Ölçeği geliştirmek

2) Geliştirilen ölçeği Kayseri ili merkez ilçelerindeki hanehalkına uygulamak

3) Çalışmanın araştırma soruları çerçevesinde gerekli istatistiksel analizleri gerçekleştirmek ve elde edilen sonuçları paylaşmak

2 Yöntem

Çalışma kapsamında belirlenen araştırma sorularının cevaplarına ulaşabilmek için 6 adımdan oluşan bir iş planı uygulanmıştır. Uygulanan iş planının aşamaları aşağıda detaylı olarak açıklanmıştır.

2.1 Kayseri Hanehalkı İstatistiklerinin Toplanması ve Analizi

Yürütülen iş planının ilk aşamasında öncelikle araştırma konusu ile ilgili ulusal ve uluslararası literatürdeki çalışmalar detaylı olarak incelenmiştir. Ardından, Kayseri hanehalkı ile ilgili ikincil veriler toplanarak analiz edilmiştir. Kayseri ili için TÜİK'ten temin edilen ikincil veriler Tablo 3'te özetlenerek sunulmuştur.

TÜİK Adrese Dayalı Nüfus Kayıt İstatistiklerine göre 2024 yılında Kayseri'deki toplam hanehalkı sayısı 441.385'dir [13]. Cohen vd. (2007)'de detaylı olarak açıklandığı üzere üzere, %95 güven aralığı ve %5 örneklem hatası ile 441.385 hanehalkından oluşan anakütle için 384 hanehalkından oluşan bir örneklem büyüklüğünün yeterli olduğu tespit edilmiştir [14]. Bu nedenle, çalışmada geliştirilecek ölçeğin merkez ilçeler bazında kotalı örnekleme yöntemiyle en az 400 hanehalkına uygulanması planlanmıştır. Bu planlama çerçevesinde 3 merkez ilçenin nüfus yoğunlukları hem ilçe hem de mahalle boyutlarında dikkate alınarak anket kotaları belirlenmiştir. Planlanan 400 anket sayısı, Tablo 4'te gösterildiği şekilde Melikgazi, Kocasinan ve Talas'a sırasıyla yaklaşık %51, %35 ve %14

oranlarında dağıtılmıştır. Melikgazi ve Kocasinan için en yüksek nüfusa sahip 10'ar mahalle, Talas için ise 3 mahalle seçilmiştir. Ardından, her bir ilçe için planlanan anket sayıları nüfus oranları dikkate alınarak mahallere kotalı olarak dağıtılmıştır.

Tablo 3. Kayseri ili için TÜİK özet istatistikleri

Nüfus ve Göç (2022)	▮ Nüfus yoğunluğu (kilometre kareye düşen kişi sayısı) → 84.58 ▮ Ortalama hanehalkı büyüklüğü → 3.22 ▮ Toplam hanehalkı sayısı → 435.301 ▮ Toplam nüfus → 1.441.523 ▮ Yaşlı bağımlılık oranı → %14.25 ▮ Çocuk bağımlılık oranı → %33.42
Eğitim (2021)	▮ Okuma yazma bilen oranı → 97.9
İnşaat ve Konut (2022)	▮ Yapı kullanma izin belgesine göre daire sayısı → 12.291
Sağlık (2020)	▮ 1000 kişi başına düşen toplam hekim sayısı → 2 ▮ Hastane sayısı → 27 ▮ Hastane yatak sayısı → 4.783 ▮ 100.000 kişi başına toplam hastane yatak sayısı → 336
Adalet (2020)	▮ Suç işlenen illere göre ceza infaz kurumuna giren hükümlü sayısı → 5.890
Ulaştırma (2022)	▮ 1000 kişi başına otomobil sayısı → 183 ▮ Otomobil sayısı → 264.357
Çevre (2020)	▮ Atık hizmeti verilen belediye nüfusunun toplam belediye nüfusuna oranı → %100 ▮ Atık su arıtma tesisi ile hizmet verilen belediye nüfusunun toplam belediye nüfusuna oranı → %80 ▮ Kanalizasyon şebekesi ile hizmet verilen belediye nüfusunun toplam belediye nüfusuna oranı → %91 ▮ İçme ve kullanma suyu şebekesi ile hizmet verilen belediye nüfusunun toplam belediye nüfusuna oranı → %99

Tablo 4. Merkez İlçelerin Anket Kotaları

İlçe	**Kota**
Melikgazi	%51
Kocasinan	%35
Talas	%14
Toplam (Kayseri)	%100

2.2 HPK Ölçeğinin Geliştirilmesi

Bu adımda literatür araştırması ve ikincil verilerin analiz sonuçlarından yararlanarak HPK ölçümüne yönelik ölçek geliştirme çalışması yürütülmüştür. Araştırma sorularının analizine imkân sağlayacak şekilde, sosyal kırılganlık endekslerinde kullanılan çeşitli göstergeler ve değişkenlerden de yararlanarak 6'lı Likert tipi ölçeğe (0: İlgili değil/Etkilemedi, 1: Çok düşük düzeyde etkiledi, 2: Düşük düzeyde etkiledi, 3: Orta düzeyde etkiledi, 4: Yüksek düzeyde etkiledi, 5: Çok yüksek düzeyde etkiledi) göre değerlendirilecek olan ölçek ifadeleri belirlenmiştir.

2.3 Pilot Uygulama

Geliştirilen ölçek nihai hale getirilmeden önce pilot uygulamaya tabi tutulmuştur. Pilot araştırmanın yapılacağı hanehalkı sayısı 400 örneklem büyüklüğünün %10'u olacak şekilde 40 hanehalkı olarak belirlenmiştir. Asıl uygulamada olduğu gibi pilot uygulamada da merkez ilçelere kotalı örnekleme yapılmıştır.

2.4 Ölçeğin Nihai Hale Getirilmesi

Pilot uygulamadan alınan geribildirimler ve yapılan ön analizler sonucunda ölçeğe nihai hali verilmiştir. Nihai anket 3 bölüm olmak üzere toplam 70 sorudan oluşmaktadır. Anket formunun birinci bölümünde katılımcılara ilişkin bilgiler, ikinci bölümünde ise hanehalkına dair bilgiler toplanmıştır. Anket formunun 3. Bölümünde ise hanehalkı pandemi kırılganlığına ilişkin "Demografik Faktörler", "Sağlık Faktörleri", "Ekonomik Faktörler", "Ulaşım Faktörleri" ve "Fiziksel/Sosyal Çevre Faktörleri" olmak üzere 5 alt boyut (gösterge) bazında toplam 39 maddeye yer verilmiştir. Bu bölümde sorular "A. Hanehalkını Etkileme Düzeyine İlişkin Deneyimler" ve "B. Hanehalkını Etkileme Düzeyine İlişkin Görüşler" şeklinde iki kategoride yöneltilmiştir.

2.5 Ölçeğin Uygulanması ve Verilerin Bilgisayar Ortamına Aktarılması

Geliştirilen nihai ölçeğin iş planının 3. adımı kapsamındaki açıklamalar çerçevesinde merkez ilçeler bazında belirlenen kotalar çerçevesinde 400 hanehalkına uygulanması planlanmış ve uygulamada 413 anketin yapıldığı tespit edilmiştir. Yapılan 413 anket üzerinde yapılan incelemelerde 14 anket formunda tutarsızlık tespit edilerek analizlerden çıkarılmış ve analizlere 399 anket formu kabul edilmiştir. Yüz yüze anket yöntemi ile toplanan veriler bilgisayar ortamına aktarılarak analiz için hazır hale getirilmiştir.

2.6 Uygulama Sonuçlarının Analizi ve Raporlama

Toplanan veriler için SPSS istatistik yazılımından yararlanılarak temel betimleyici istatistiksel

analizler ile çok değişkenli istatistiksel analizler yapılmıştır. Verilere ilgili analizler uygulandıktan sonra ele alınan araştırma sorularının cevaplarına ulaşılmış ve çalışmanın sonuç bölümünde paylaşılmıştır.

3 Sonuçlar

Araştırma sorularının analizler sonucu ulaşılan sonuçları aşağıdaki gibi özetlenebilir:

1) *Kayseri ili merkez ilçelerinin genel HPK düzeyi nedir?*

Genel HPK düzeylerinin incelendiği ilk araştırma sorusunun cevabı için anket formunun 3.A bölümüne uygulanan tanımlayıcı istatistiksel analiz sonuçları Tablo 5'te özetlenmiştir. Bu sonuçlara göre katılımcıların 3.A Bölümünde vermiş oldukları cevapların ortalaması **1,0338/5,0000** olarak tespit edilmiştir. Ulaşılan değerin "**Çok düşük düzeyde etkiledi**" seçeneğine yakın olduğu görülmüştür.

Tablo 5. A Bölümü - Hanehalkını Etkileme Düzeyine İlişkin Deneyimler

	N	Mean	Std. Dev.	Skewness		Kurtosis	
	Statis.	Statis.	Statis.	Statis.	Std. Error	Statis.	Std. Error
Kayseri	399	**1,0338**	1,02759	**1,573**	0,122	**2,005**	0,244
Melikgazi	196	**1,1703**	1,05627	**1,045**	0,174	**0,3**	0,346
Kocasinan	143	**0,6954**	0,639	**2,509**	0,203	**8,644**	0,403
Talas	60	**1,3948**	1,40745	**1,208**	0,309	**0,202**	0,608

2) *Her bir merkez ilçenin ortalama HPK düzeyi nedir? Merkez ilçeler arası HPK düzeyleri farklılık göstermekte midir?*

Kayseri ili ile merkez ilçelerin HPK ortalamalarının incelendiği Tablo 5'e göre; Melikgazi ilçesinin HPK ortalaması 1,1703/5,0000, Kocasinan ilçesinin HPK ortalaması 0,6954/5,0000 ve Talas ilçesinin HPK ortalaması 1,3948/5,0000 olarak tespit edilmiştir. Merkez ilçeler için yapılan normal dağılım analizlerine göre Kayseri ili için Kurtosis değeri 2,005>2,000 ve Kocasinan ilçesi için Skewness 2,509>2,000 ve Kurtosis değeri 8,644>2,000 tespit edilmiştir. Bu sonuçlara göre verilerin normal dağılım göstermemesi nedeni ile Kruskal-Wallis H testi yapılmıştır. İlçe ortalamaları arasındaki farklılığı incelemek amacıyla uygulanan Kruskal-Wallis H testinin sonuçları Tablo 6'da paylaşılmıştır. Tablo 6'ya göre Asymp. Sig. değeri 0,0001<0,05 olduğu için merkez ilçe ortalamaları arasında istatistiki olarak anlamlı bir farklılık olduğu tespit edilmiştir. Farklılıkların hangi ilçeler arasında olduğunu tespit etmek amacıyla Mann-Whitney U testi uygulanmıştır. Tablo 7'de belirlenen Asymp. Sig. (2-tailed) 0,001<0,05 değeri Melikgazi ve Kocasinan ilçeleri arasında istatistiki olarak anlamlı bir farklılık olduğunu göstermektedir. Tablo 7'de belirlenen Asymp. Sig. (2-tailed) 0,466>0,05 değeri Melikgazi ve Talas ilçeleri arası istatistiki olarak anlamlı bir farklılık olmadığını göstermektedir. Tablo 7'de belirlenen Asymp. Sig. (2-tailed) 0,008<0,05 değeri Kocasinan ve Talas ilçeleri arası istatistiki olarak anlamlı bir farklılık olduğunu göstermektedir.

Tablo 6. A Bölümü – Merkez İlçelere Göre Kruskal-Wallis H Testi Sonuçları

	A	Demografik	Sağlık	Ekonomik	Fiziksel/ Sos. Çevre
Chi-Square	13,202	10,838	19,956	24,471	14,915
df	2	2	2	2	2
Asymp. Sig.	**0,001**	**0,004**	**0,000**	**0,000**	**0,001**

Tablo 7. A Bölümü – Merkez İlçeler Mann-Whitney U Testi Sonuçları

(Melikgazi-Kocasinan)					
	A	Demografik	Sağlık	Ekonomik	Fiziksel/ Sos. Çevre
Mann-Whitney U	11072	11061,5	10473	10149	11104
Wilcoxon W	21368	21357,5	20769	20445	21400
Z	-3,302	-3,384	-4,222	-4,635	-3,285
Asymp. Sig.(2-tailed)	**0,001**	**0,001**	**0,000**	**0,000**	**0,001**
(Melikgazi-Talas)					
	A	Demografik	Sağlık	Ekonomik	Fiziksel/ Sos. Çevre
Mann-Whitney U	5514,5	5600	5702,5	5809,5	5308,5
Wilcoxon W	24820,5	7430	25008,5	25115,5	24614,5
Z	-0,728	-0,57	-0,366	-0,145	-1,143
Asymp. Sig.(2-tailed)	0,466	0,568	0,715	0,885	0,253
(Kocasinan-Talas)					
	A	Demografik	Sağlık	Ekonomik	Fiziksel/ Sos. Çevre
Mann-Whitney U	3271	3784	3171	3004,5	3108
Wilcoxon W	13567	14080	13467	13300,5	13404
Z	-2,669	-1,359	-3,189	-3,702	-3,116
Asymp. Sig.(2-tailed)	**0,008**	0,174	**0,001**	**0,000**	**0,002**

3) *Merkez ilçelerdeki farklı mahallelerin ortalama HPK düzeyi nedir? Mahalleler arası HPK düzeyleri farklılık göstermekte midir?*

Talas Mevlana mahallesi dışında merkez ilçelerde yer alan mahallelerin gözlem sayılarının 30'dan küçük olması nedeniyle veriler normal dağılım göstermemektedir. Bu nedenle mahalleler arası ortalama HPK düzeyleri bakımından farklılığın araştırılması için Kruskal-Wallis H testi yapılmıştır. Tablo 8'e göre Asymp. Sig. değeri 0,0000<0,05 olduğu için merkez ilçe mahallelerinin (23 mahalle) ortalamaları arasında istatistiki olarak anlamlı bir

farklılık olduğu tespit edilmiştir.

Tablo 8. A Bölümü – Merkez ilçe Mahallelerine Göre Kruskal-Wallis H Testi Sonuçları

	A	Demografik	Sağlık	Ekonomik	Ulaşım	Fiziksel/ Sos. Çevre
Chi-Square	81,219	83,542	79,243	74,515	60,089	74,746
df	22	22	22	22	22	22
Asymp. Sig.	**0,000**	**0,000**	**0,000**	**0,000**	**0,000**	**0,000**

4) *Pandemi kırılganlık göstergeleri bakımından ortalama HPK düzeyi nedir? Göstergeler arası HPK düzeyleri farklılık göstermekte midir?*

Pandemi kırılganlık göstergeleri bakımından merkez ilçelerin ortalama HPK düzeylerinin incelendiği özet ortalama değerler Tablo 9'da gösterilmektedir.

Tablo 9. Merkez ilçelerin göstergeler bakımında ortalama HPK düzeyleri

Göstergeler	**Melikgazi**	**Kocasinan**	**Talas**
N	196	143	60
Demografik	,8730	,3201	1,0233
Sağlık	,9486	,4031	1,1518
Ekonomik	1,1427	,5815	1,2138
Ulaşım	1,6939	1,7902	2,2000
Fiziksel/Sosyal Çevre	1,5072	,9310	1,8228

Her bir gösterge bazında HPK düzeylerinin merkez ilçelere göre farklılık gösterip göstermediğini tespit için Kruskal-Wallis H testi yapılmıştır. Tablo 6'ya göre Demografik göstergeler Asymp. Sig. değeri 0,0004<0,05, Sağlık göstergeleri Asymp. Sig. değeri 0,0000<0,05, Ekonomik göstergeler Asymp. Sig. değeri 0,0000<0,05 ve Fiziksel/Sosyal Çevre göstergeleri Asymp. Sig. değeri 0,0001<0,05 olduğu için bu göstergeler bakımından merkez ilçeler arasında istatistiki olarak anlamlı farklılıklar bulunduğu tespit edilmiştir. A Puanlarında ulaşım göstergeleri normal dağılım gösterdiği için merkez ilçeler arası farklılık ANOVA analizi ile test edilmiştir. Analiz sonuçlarının yer aldığı Tablo 10'a göre sig. değeri 0,051>0,05 olduğu için ulaşım göstergeleri bakımından merkez ilçeler arasında anlamlı bir farklılık tespit edilmemiştir.

Tablo 10. A Bölümü – Ulaşım Faktörü için ANOVA Testi Sonuçları

		Sum of Squares	df	Mean Square	F	Sig.
A - Ulaşım	Between Groups	11,813	2	5,906	2,996	**,051**
	Within Groups	780,619	396	1,971		
	Total	792,432	398			

Her bir gösterge bazında hangi ilçeler arasında farklılık bulunduğunu tespit etmek amacıyla Mann-Whitney U testi yapılmış ve sonuçlar Tablo 7'de paylaşılmıştır. Bu testin sonuçlarına göre Melikgazi ve Kocasinan ilçeleri arasında Demografik, Sağlık, Ekonomik ve Fiziksel/Sosyal Çevre göstergelerinin hepsinde (Asymp. Sig. (2-tailed) <0,05) olduğu için farklılık tespit edilmiştir. Melikgazi ve Talas ilçeleri arasında hiçbir gösterge bakımından (Asymp. Sig. (2-tailed) >0,05) olduğu için farklılık tespit edilememiştir. Kocasinan ve Talas ilçeleri arasında Demografik (Asymp. Sig. (2-tailed) 0,174<0,05) göstergesi bakımından fark tespit edilmemiş, diğer göstergeler bakımından farklılık tespit edilmiştir.

5) *Her bir gösterge bazında pandemi kırılganlığına en fazla etki eden değişken hangisidir?*

Her bir gösterge bazında HPK düzeyine en fazla etki eden değişkenler Tablo 11'de özetlenmiştir. Tablo 11'e göre Demografik göstergelere en fazla etki eden değişken 0,97 ortalama ile "Hanehalkı nüfusunun fazla olması", Sağlık göstergelerine göre en fazla etki eden değişken 1,34 ortalama ile "Hanehalkında yüksek pandemi riskine maruz kalan bireyin olması (sağlık kurumlarında çalışma, kalabalık ortamlarda bulunma vb.)", Ekonomik göstergelere en fazla etki eden değişken 1,41 ortalama ile "Hanehalkı gelir düzeyinin ihtiyaçları karşılamada yetersiz olması", Ulaşım göstergelerine göre en fazla etki eden değişken 2,57 ortalama ile "Hanehalkının yaşadığı muhitin pandemi hastanelerine uzak olması" ve Fiziksel/Sosyal Çevre göstergelere en fazla etki eden değişken 1,72 ortalama ile "İhtiyaç durumunda sosyal çevre desteğinin (yardım edebilecek kişilerin) kısıtlı olması" olarak tespit edilmiştir.

Tablo 11. Ölçek maddeleri tanımlayıcı istatistikleri

	Mean (A)	Standard Deviation (A)	Mean (B)	Standard Deviation (B)
Demografik Faktörler				
Hanehalkı nüfusunun fazla olması	**0,97**	1,62	2,94	1,9
Hanehalkında tek kişinin yaşaması	0,74	1,55	2,61	1,85
Hanehalkında sadece anne ve çocukların yaşaması	0,65	1,49	**3,36**	1,68
Hanehalkında sadece baba ve çocukların yaşaması	0,65	1,51	3,25	1,74
Hanehalkında 0-2 yaş bireyin olması	0,7	1,56	2,76	1,75
Hanehalkında 3-6 yaş bireyin olması	0,73	1,54	2,67	1,69
Hanehalkında 7-12 yaş bireyin olması	0,76	1,54	2,56	1,62
Hanehalkında 13-17 yaş bireyin olması	0,88	1,57	2,45	1,67
Hanehalkında okuma-yazma bilmeyenbireyin olması	0,57	1,41	2,53	1,83
Hanehalkında Türkçe bilmeyen bireyinolması	0,55	1,46	2,91	1,9
Hanehalkı bireylerinin şehre başka	0,66	1,55	2,31	1,92

	Mean (A)	Standard Deviation (A)	Mean (B)	Standard Deviation (B)
ilden gelerek yerleşmiş olması				
Hanehalkı bireylerinin şehre başka ülkeden gelerek yerleşmiş olması	0,51	1,38	2,94	1,8
Sağlık Faktörleri				
Hanehalkında kronik hastalığı olan bireyin olması	0,98	1,87	3,75	1,68
Hanehalkında bedensel engelli bireyinolması	0,56	1,46	3,72	1,73
Hanehalkında zihinsel engelli bireyinolması	0,5	1,41	3,75	1,69
Hanehalkında 65 yaş ve üzeri bireyin olması	0,65	1,58	3,43	1,76
Hanehalkında hamile bireyin olması	0,57	1,52	3,32	1,78
Hanehalkından ayrı yaşayan, ancak hanehalkının bakım/desteğine ihtiyaç duyan yaşlı ebeveynlerin olması	0,89	1,75	3,67	1,69
Hanehalkında yüksek pandemi riskinemaruz kalan bireyin olması (sağlık kurumlarında çalışma, kalabalık ortamlarda bulunma vb.)	**1,34**	1,96	**3,78**	1,66
Ekonomik Faktörler				
Hanehalkı gelir düzeyinin ihtiyaçları karşılamada yetersiz olması	**1,41**	1,91	3,86	1,61
Hanehalkında iş arayan/işsiz bireyin bulunması	0,88	1,68	**3,91**	1,64
Hanehalkında sosyal güvencesi olmayan bireyin bulunması	0,82	1,73	3,75	1,71
Pandemi nedeniyle hanehalkı gelir düzeyinin azalması	1,25	1,91	3,83	1,67
Pandemi nedeniyle hanehalkında işsizkalan bireyin olması	0,77	1,69	3,89	1,72
Pandemi nedeniyle hanehalkına ait işyerinin daimi olarak kapanması	0,64	1,55	3,88	1,79
Pandemi döneminde hanehalkının sosyal yardımlardan yararlanamaması/yardımların yetersizolması	0,89	1,69	3,63	1,63
Ulaşım Faktörleri				
Hanehalkının yaşadığı muhitin pandemi hastanelerine uzak olması	**2,57**	1,99	**3,65**	1,61
Hanehalkının yaşadığı muhitin diğer sağlık kurumlarına uzak olması	2,21	2,01	3,59	1,64
Hanehalkının yaşadığı muhitten genel ulaşım imkanlarının (işyeri, okul, alışveriş yerleri vb.) zor olması	1,54	1,91	2,9	1,87
Hanehalkında araç bulunmaması	1,36	1,97	3,18	1,86
Hanehalkında tek kişinin araç kullanabilmesi	1,34	1,86	2,75	1,85
Fiziksel/Sosyal Çevre Faktörleri				
Hanehalkının yaşadığı binadaki daire sayısının fazla olması	1,59	1,76	2,58	1,78
Hanehalkının yaşadığı konuttaki oda sayısının yetersiz olması	1,22	1,7	2,62	1,77
Hanehalkının yaşadığı konutun iç mekan özelliklerinin (balkon, birdenfazla tuvalet/banyo, gün ışığı alma vb.) yetersiz olması	1,15	1,67	2,66	1,73
Hanehalkının yaşadığı konutun dış mekan özelliklerinin (bahçe, oyun alanı, manzara vb.) yetersiz olması	1,31	1,79	2,62	1,76
Pandemi döneminde başka bir konutta (müstakil ev, yazlık vb.) kalma imkanının olmaması	1,56	1,84	2,67	1,72
Hanehalkının yaşadığı muhitteki çevre düzeninin (park, yeşil alan vb.) yetersiz olması	1,33	1,77	2,46	1,76
İhtiyaç durumunda sosyal çevre desteğinin (yardım edebilecek kişilerin) kısıtlı olması	**1,72**	1,97	**3,15**	1,78
Hanehalkı bireylerinin şehri tanımaması/şehre yabancılık hissetmesi	0,9	1,72	2,23	1,89

6) *Katılımcılar hangi değişkenlerin HPK düzeyini en çok etkilediğini düşünmektedir?*

Katılımcıların her bir gösterge bazında HPK düzeyine en fazla etki edeceğini düşündükleri değişkenler Tablo 11'de özetlenmiştir. Tablo 11'e göre Demografik göstergelere en fazla etki edeceği düşünülen değişken 3,36 ortalama ile "Hanehalkında sadece anne ve çocukların yaşaması", Sağlık göstergelerine göre en fazla etki edeceği düşünülen değişken 3,78 ortalama ile "Hanehalkında yüksek pandemi riskine maruz kalan bireyin olması (sağlık kurumlarında çalışma, kalabalık ortamlarda bulunma vb.)", Ekonomik göstergelere en fazla etki edeceği düşünülen değişken 3,91 ortalama ile "Hanehalkında iş arayan/işsiz bireyin bulunması", Ulaşım göstergelerine göre en fazla etki edeceği düşünülen değişken 3,65 ortalama ile "Hanehalkının yaşadığı muhitin pandemi hastanelerine uzak olması" ve Fiziksel/Sosyal Çevre göstergelere en fazla etki edeceği düşünülen değişken 3,15 ortalama ile "İhtiyaç durumunda sosyal çevre desteğinin (yardım edebilecek kişilerin) kısıtlı olması" olarak tespit edilmiştir.

Yapılan bu incelemeler neticesinde Kayseri ilinin merkez ilçeleri arasında hem genel HPK düzeyleri bakımından hem de pandemi kırılganlık göstergeleri bakımından anlamlı farklılıkların bulunduğu tespit edilmiştir. Özellikle Kocasinan ilçesi ile Talas ve Melikgazi ilçeleri arasında farklılık var iken Melikgazi ve Talas ilçeleri arasında fark tespit edilememiştir. Bu durum farklılığın Kocasinan ilçesi kaynaklı olduğunu göstermektedir. Ayrıca Kayseri ilinin genel HPK düzeyi literatürdeki diğer çalışmalardan farklı şekilde düşük tespit edilmiştir.

HPK düzeyleri arası farklılık mahalleler düzeyinde incelendiğinde farklılığın devam ettiği görülmüştür. Tespit edilen farklılık sadece genel düzeyde değil pandemi kırılganlık göstergeleri bakımından da mevcuttur. Demografik, Sağlık, Ekonomik, Ulaşım ve Fiziksel/Sosyal Çevre şeklinde incelenen pandemi kırılganlık göstergeleri arasında hem Kayseri ili genelinde hem de merkez ilçeler arası farklılık incelendiğinde sadece ulaşım göstergesi bakımından fark olmadığı, diğer 4 gösterge bakımından anlamlı farklılıklar bulunduğu belirlenmiştir. Melikgazi ve Kocasinan ilçeleri arasında Demografik, Sağlık, Ekonomik ve

Fiziksel/Sosyal Çevre göstergelerinin hepsinde farklılık mevcutken, Melikgazi ve Talas ilçeleri arasında hiçbir gösterge bakımından farklılık yoktur. Kocasinan ve Talas ilçeleri arasında ise sadece Demografik gösterge farklılığı tespit edilememiştir.

Pandemi kırılganlığı göstergelerinin alt değişkenleri yaşanan deneyim ve katılımcıların düşünceleri kapsamında iki başlıkta incelenmiştir. Burada katılımcıların deneyimledikleri kırılganlığın düşündüklerinden daha düşük düzeyde seyrettiği tespit edilmiştir. Kırılganlığa en fazla etki eden değişkenler şu şekilde tespit edilmiştir:

- "Hanehalkı nüfusunun fazla olması"
- "Hanehalkında yüksek pandemi riskine maruz kalan bireyin olması (sağlık kurumlarında çalışma, kalabalık ortamlarda bulunma vb.)",
- "Hanehalkı gelir düzeyinin ihtiyaçları karşılamada yetersiz olması",
- "Hanehalkının yaşadığı muhitin pandemi hastanelerine uzak olması"
- "İhtiyaç durumunda sosyal çevre desteğinin (yardım edebilecek kişilerin) kısıtlı olması"

Bu çalışma ile geliştirilen ölçek, ulusal düzeyde veya farklı illerde HPK düzeyinin ölçümüne ilişkin ileriye yönelik araştırmalarda kullanılabilecek niteliktedir. Ölçeğin uyarlanması ile diğer afet türlerine yönelik sosyal kırılganlık ölçümlerinde de kullanılması mümkündür.

Teşekkür

Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından **122G062** Numaralı proje ile desteklenmiştir. Projeye verdiği destekten ötürü TÜBİTAK'a teşekkürlerimizi sunarız.

Kaynaklar

[1] AFAD, T.C. İçişleri Bakanlığı Afet ve Acil Durum Yönetimi Başkanlığı. "Afet Türleri". https://www.afad.gov.tr/afet-turleri (27.08.2024).

[2] GHSI, Global Health Security Index. "Advancing Colllective Action and Accountability Amid Global Crisis (December 2021)". https://ghsindex.org/wp-content/uploads/2021/12/2021_GHSindexFullReport_Final.pdf (27.08.2024).

[3] Varol N, Buluş Kırıkkaya E. "Afetler Karşısında Toplum Dirençliliği". *Dirençlilik Dergisi,* 1(1), 1-9, 2017.

[4] UNDRR, United Nations Office for Disaster Risk Reduction. "Sendai Framework for Disaster Risk Reduction (2015-2030)". https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030 (27.08.2024).

[5] Yüksel U, Ergün M. "Sosyolojinin Pencereleri, Pencerelerden Sosyoloji: Pandemi Sürecinde İstanbul'da Sosyal Kırılganlık ve Çoklu-Afet Üzerine Nitel Araştırma Deneyimi". *Sosyoloji Araştırma Dergisi,* 25(1), 23-35, 2022.

[6] SBB, T.C. Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı. "Sürdürülebilir Kalkınma Amaçları". http://www.surdurulebilirkalkinma.gov.tr/amaclari/ (27.08.2024).

[7] SB, T.C. Sağlık Bakanlığı. "Türkiye COVID-19 Sağlıkta Acil Durum Projesi: Paydaş Katılım Planı (Nisan 2021)". https://pydb.saglik.gov.tr/Eklenti/40589/0/covid19paydaskatilimplanifinalpdf.pdf (27.08.2024).

[8] Kırlangıçoğlu C. "Şehirlerin Pandemik Kırılganlık Seviyelerinin Faktör Analizi Yöntemiyle Belirlenmesi". *Coğrafya Dergisi*, 44, 193-205, 2022.

[9] Ortakavak Z, Korkmaz E, Uslu A. "Covid-19 Salgınına Karşı Sosyal Kırılganlık Endeksi: Türkiye Örneği". *Iğdır Üniversitesi Sosyal Bilimler Dergisi*, Ek Sayı Ekim 2020, 245-268, 2020.

[10] Yaprak Ü, Şahinöz T, Şahinöz S. "Evaluation of Socio-Economic Fragility Index Against Disasters: Example Turkey Provinces". *Turkish Journal of Public Health*, 19(2), 168-181, 2021.

[11] Con Wright G, Çelik K. "Farklı Sosyal Kırılganlık Düzeyine Sahip Mahalle Muhtarlarının Gözünden Covid-19 Salgın Deneyimi: İstanbul İli Örneği". *Çalışma ve Toplum Dergisi,* 1(72), 239-266, 2022.

[12] Ortakavak Z, Uslu A, Korkmaz E. "Sosyal Kırılganlık İndeksinin CSB ile Haritalanması: İzmir İli Örneklemi". *Dirençlilik Dergisi*, 3(1), 37-53, 2019.

[13] TÜİK, Türkiye İstatistik Kurumu. "Adrese Dayalı Nüfus Kayıt İstatistikleri". https://data.tuik.gov.tr/Kategori/GetKategori?p=nufus-ve-demografi-109&dil=1 (27.08.2024).

[14] Cohen L, Manion L, Morrison K. *Research Methods in Education*. New York, USA, Routledge Taylor & Francis Group, 2007.

Redundant or Interesting? Inspection of Redundant Association Rules in a Psychiatric Dataset

İnanç KABASAKAL [1]*, Helin YILMAZ KAFALI [2]

[1] *Ege University, Faculty of Economics & Administrative Sciences, Department of Business Administration, İzmir, TÜRKİYE*

[2] *Işık University, Faculty of Economics, Administrative & Social Sciences, Department of Psychology, İstanbul, TÜRKİYE*

Abstract

Association Rule Discovery is a process that extracts potentially useful patterns in the form of association rules. The evaluation of the rules in descriptive studies is often conducted by referring to interestingness measures, and often with the participation of domain experts, which might require significant time and effort. Reduction techniques help this challenge by filtering out redundant rules. This paper aims to analyze a list of rules obtained through a study in child psychiatry and present the redundancies discovered by reduction techniques. The case being examined involves demographic variables, symptoms, and diagnoses of psychiatric diseases regarding 263 patients from Turkey. With thresholds of 2%-60% interval for support and 10% confidence, a list of 54 rules was extracted through rule mining. In our study, the redundancy of rules that consist of three items in the form of {X=>YZ} was determined using the rule-of-thumb by [1] and the ratings assigned by a domain expert regarding the usefulness of rules were examined using statistical methods. The results demonstrated that the usefulness of original rules and their variants in the form of {X=>Y, X=>Z} were statistically different, and no such difference was found with the variants of the form {XY=>Z, XZ=>Y}.

Keywords: *Association Rule Mining, Rule Filtering, Redundant Association Rules*

1 Introduction

Association rule mining in large datasets is achieved with algorithms such as Apriori [2] and generates an exponential number of frequent itemsets [3] and too many rules [4]. Among those rules, there might exist numerous redundant rules that populate exponentially with respect to the length of the longest frequent itemset [3]. Although rule mining is a primary task in data mining, obtaining too many association rules is often regarded as a prominent problem that limits the use of rule mining techniques [5] and prevents the efficient utilization of discovered rules [6]. A potential solution for this challenge is to select an appropriate method in order to identify and eliminate redundant rules. However, despite the abundance of rule mining applications in prior research, redundant rule elimination is often overlooked or ignored.

This study examines the use of a redundant rule detection approach and finds redundant rules obtained from a psychiatric dataset. Furthermore, our study investigates the accuracy redundancy detection by referring to interestingness scores manually assigned by a domain expert. Our findings involve the test of hypotheses constructed using a useful redundancy approach [1] and discuss how their redundancy determination approach helps in our case.

*Contact email: inanc.kabasakal@ege.edu.tr

2 Literature Review

Data Mining is usually described as a technical step within the knowledge discovery process from datasets [7:7] that aims to convert data into actionable insights [8] for decision-makers. Descriptive data mining techniques have been priorly proposed to analyze existing datasets and describe interesting patterns. Such techniques have been employed for tasks including clustering, classification, association discovery, and outlier detection.

In association rule mining, rules are extracted that correspond to relations of cooccurrence within data. Association rules are usually evaluated using interesting criteria, such as support, lift, and confidence. The abundance of rules that satisfy interestingness thresholds is a widely recognized

problem that necessitates detecting and eliminating redundant rules.

This section briefly revisits association rule mining, and defines support, confidence, and lift that are among popular rule interestingness criteria. Moreover, several approaches for identifying and eliminating redundant rules are introduced in this context.

2.1 Association Rule Mining

Discovering patterns of association is achieved through techniques, referred to as Association Rule Mining. This is achieved by utilizing algorithms such as the Apriori [2] and Frequent Pattern Growth [9] algorithms.

The Apriori algorithm depends on the idea of pruning itemsets that have little support. The support of an item or an itemset is the total number of transactions that involve that item/itemset. It is possible to handle massive numbers of transactions by the Apriori algorithm by determining a minimum support rate, thereby pruning the sets of candidate items that fail to satisfy that threshold. The discovered sets of frequent items are then processed to generate association rules that are represented in the form of X=>Y where X and Y represent the antecedent and consequent of the rule, respectively.

The interestingness of rules is often measured with the following criteria:

$$Supp(X => Y) = \frac{\#of\ transactions\ involving\ X \cup Y}{total\ \#\ of\ transactions} \quad (1)$$

$$Conf(X => Y) = \frac{\#of\ transactions\ involving\ X \cup Y}{\#of\ transactions\ involving\ X} \quad (2)$$

$$Lift(X => Y) = \frac{Confidence\ (X=>Y)}{Support(Y)} \quad (3)$$

However, selecting a reasonable support and confidence often does not essentially prevent generating too many association rules, which might involve too many redundant ones [4]. According to [10], this problem also exists in sequential pattern discovery where relations between sets of sequences are discovered and represented as association rules.

The existence of too many results might be a problem for decision-makers who aim to examine the association rules to figure out interesting patterns within datasets.

2.2 Rule Redundancy

The primary objective of data mining is "to extract potentially useful patterns" that might provide useful insights for the analyst [11]. In this regard, the intervention of a decision-maker who is familiar with the dataset being analyzed and willing to benefit from the findings is the ultimate authority that decides whether a rule is redundant, or not in his/her decision. Similarly, an interesting pattern is described through a set of attributes to be satisfied [7:21], including the following ones:

- Easy to understand by the humans
- Potentially useful
- Novel

The major drawback of the definitions that involve human intervention is the considerable time and effort requirement [12]. This is particularly challenging when the number of rules being discovered is abundant. Therefore, an assessment mechanism that classifies rules in terms of redundancy is required, and this was considered as a relevant problem in the literature.

Another study [13] proposed an approach that identifies irrelevant rules by comparing their antecedents and consequents. The technique involves eliminating irrelevant association rules in post-processing, and reduction of rules that are discovered.

Prior research involves several approaches to define a rule as redundant. Association rules can be classified as redundant or not, considering the predefined weights of the items being involved, which correspond to the utility scores of items [14]. This approach has been utilized in *High Utility Itemset Mining*, where the sets of candidate item-sets with low utility are pruned [15] to accelerate itemset generation. However, this approach would require predetermination of weights for items in the dataset, which is another challenge, particularly when an inherent weight criterion such as price or profit does not exist for items. Moreover, this approach involves pruning low utility itemsets, which also prevents cross-examination of the resulting redundant and non-redundant rules, since such pruning is supposed to prevent redundant rule generation.

A practical rule-of-thumb proposed in [1] depends on distinguishing the redundant ones among the rules

discovered through association rule mining. Based on a comparison of rule support and confidence criteria, this approach identifies redundant rules for each rule in the form of X=>YZ as in the following:

Table 1. Redundant rules relevant to X=>YZ.

Rule	Description
XY=>Z	Supp(XZ) < Supp(X), so has lower confidence than (X => YZ)
XZ=>Y	Supp(XZ) < Supp(X), so has lower confidence than (X => YZ)
X=>Y	Contains less information than (X=>YZ)
X=>Z	Contains less information than (X=>YZ)

The redundancy of such rules derived from X=>YZ is justified as the following in [12] and [1]:

> *"This is because the support and confidence values X=>YZ is less than the support and confidence values of those rules such as XY=>Z, XZ=>Y, X=>Y, and X=>Z."*

Although this approach does not formulate a more efficient method for rule mining, the identification of redundant and non-redundant rules enables cross-checking whether this classification holds against the subjective interpretation of the domain experts.

3 Methodology

In this study, we adopt the rule-of-thumb for redundancy proposed by [1], which identifies redundant forms of rules in the form of X=>YZ. For a rule X=>YZ that satisfies the support and confidence thresholds, when considering the redundancies implied by this approach, we propose the following hypotheses:

- H_0: Scores assigned to XY=>Z are not statistically significant than those of X=>YZ
- H_1: Scores assigned to XZ=>Y are not statistically significant than those of X=>YZ
- H_2: Scores assigned to X=>Y are not statistically significant than those of X=>YZ
- H_3: Scores assigned to X=>Z are not statistically significant than those of X=>YZ

The statistically different scores assigned by the domain expert will imply redundancy if the mean score is lower than the original rule X=>YZ.

Top-10 rules that overlap the form X=>YZ among the rules discovered from our research dataset were selected for hypothesis testing. In this step, our methodology involves examining whether the score assigned to the original rule (X=>YZ) and scores for supposedly redundant rules (as listed in Table 1) are statistically different.

For hypothesis testing, paired samples t-tests were conducted to figure out whether scores assigned for X=>YZ and each of its variants were statistically significant.

4 Case Study

This section introduces the dataset being analyzed, the rule mining process, and the selection of prominent rules to be examined when conducting hypothesis tests.

4.1 The Dataset and Data Preprocessing

Our case involves analyzing a psychiatric dataset and mining association rules where we find rules in the form of X=>YZ. The dataset analyzed contains binary features associated with responses obtained from 263 patients in Türkiye. The data was collected in 2020-2022 by physicians led by the co-author of the study, Dr. Helin YILMAZ KAFALI. The original data involved 46 columns, and columns in the dataset were organized into binary columns as shown in [16].

The features include patients' cigarette, alcohol, and drug usage, as well as their diseases, diagnosis, COVID history, suicidal thoughts & attempts, and treatments. Moreover, the data involves results obtained through surveys conducted during patient follow-up.

The high/medium/low assessment scores were represented in binary CAPE, PSQI, GAD, and PHQ variables as in Table 2. The optimal number of clusters for numeric assessment scores for CAPE, PSQI, GAD, and PHQ were determined with the k-means algorithm one by one. Clustering results obtained by different k's were subject to comparison by the Silhouette Coefficient [17] to select the best cluster size in this phase.

Clusters determined for assessment tests were named such as GAD-HIGH and GAD-LOW, which also become binary variables in our dataset. With this approach, {GAD-LOW = 1} is assigned for a patient who would have a low GAD score, thereby being classified in the cluster with lower GAD scores.

Our research dataset consists of the following binary features:

Table 2. Binary Features in Research Dataset.

Feature	Values
Gender	Female/Male
Cigarette Usage	Yes/No
Alcohol Usage	
Drug Usage	
Chronical Disease	
Psychiatric Disorder	
COVID Infection	
COVID Hospitalization	
COVID Infection in Family	
COVID Hospitalization in Family	
Suicidal Thought	
Suicidal Attempt	
Pharmacological Treatment	
CAPE-HIGH	
CAPE-MEDIUM	
CAPE-LOW	
PSQI-HIGH	
PSQI-MEDIUM	
PSQI-LOW	
GAD-HIGH	
GAD-LOW	
PHQ-HIGH	
PHQ-LOW	

4.2 Rule Discovery

The dataset described above was mined to find out patterns of associations within data. *Association Rule Mining* was conducted with *RuleGenerator Software* (http://rulegenerator.inanc.biz), which had been presented in [19].

The thresholds for the Apriori Algorithm [2] were selected as in Table 3:

Table 3. Parameters in Association Rule Mining.

Criteria	Threshold
Minimum Support	2%
Minimum Confidence	%10
Maximum Support	60%

The minimum support and confidence criteria are required for pruning non-frequent sets of items in the Apriori algorithm. Additionally, *RuleGenerator* supports pruning too-frequent sets of items as described by Koh & Rountree [18] to eliminate too-frequent itemsets that might hinder discovery of less-frequent but more interesting patterns. In this regard, the maximum support threshold was selected as 60% to prune items and sets of items that have support ratio over this ratio.

5 Findings

This section presents the most interesting association rules that are in form of X=>YZ, and supposedly redundant counterparts derived in the original rules. Moreover, the results of hypothesis tests using those rules were presented in our findings.

5.1 Prominent Association Rules

As a result of rule mining with the above parameters, 9730 association rules were discovered. The top 10 rules that have the form of X=>YZ were selected with respect to their lift are presented in Table 4:

Table 4. Top Non-Redundant (X=>YZ) Association Rules with Highest Lift.

#	{X => YZ}	{X}	{Y}	{XY}	Support	Confidence	Lift
1	alcohol_usage = 6-9 ➔ drug_usage = yes & GAD-HIGH	11	17	6	0.04	0.55	8.44
2	alcohol_usage = 6-9 ➔ psych_disorder = yes & drug_usage = yes	11	16	6	0.04	0.55	8.97
3	suic_thought = yes ➔ suic_attempt = yes & CAPE-HIGH	41	14	12	0.16	0.29	5.50
4	cigarette_usage = 3-5 ➔ psych_disorder = yes & suic_thought = yes	13	23	6	0.05	0.46	5.28
5	suic_thought = yes ➔ cigarette_usage = everydy & CAPE-HIGH	41	9	9	0.16	0.22	6.41
6	covid_infamily = yes ➔ alcohol_usage = none & covid_infec = yes	64	8	8	0.24	0.13	4.11
7	cigarette_usage = everydy ➔ covid_infamily = yes & suic_attempt = yes	47	13	8	0.18	0.17	3.44

8	suic_attempt = yes ➔ psych_disorder = yes & chronical_disease = yes	45	9	7	0.17	0.16	4.55
9	psy_dis_family = yes ➔ suic_attempt = yes & PSQI*HIGH	54	12	7	0.21	0.13	2.84
10	drug_usage = yes ➔ gender = male & alcohol_usage = 1-2 days	34	13	6	0.13	0.18	3.57

5.2 Interestingness Scores

Interestingness scores were determined for each rule and their supposedly redundant versions by the domain expert as in Table 5:

Table 5. Scores Determined for Rule Variants.

{X}	X=>YZ	YX=>Z	ZX=>Y	X=>Y	X=>Z
1	6	7	3	3	2
2	6	7	4	3	3
3	8	9	8	3	4
4	8	5	3	2	3
5	8	5	3	3	4
6	1	1	1	1	1
7	3	1	1	1	3
8	4	4	3	2	3
9	4	6	4	3	2
10	3	2	1	1	1

The scores reflect the subjective opinion of the reviewer, who is involved in data collection and has extensive scientific knowledge in the field of child and adolescent psychiatry. Before this scoring phase, the expert did not have prior knowledge about the rule redundancy concept (summarized in Table 1) adopted for this study. In this regard, we emphasize that, the contributions of the authors were pre-planned to eliminate bias in scoring of rule variants.

5.3 Hypothesis Tests and Results

Paired samples t-tests were conducted to test H_0, H_1, H_2, H_3, where we compared scores determined for original rules (X=>YZ) and their variants. The scores and paired samples t-test results are presented below in Table 6.

Table 6. Paired Samples T-Test Results (α=0.05, d.f.=9).

For Hypothesis:		H_0			H_1			H_2			H_3		
X=>YZ		XY=>Z			XZ=>Y			X=>Y			X=>Z		
Rule	Score	Score	Diff.	(Diff.$\bar{x}$)^2	Score	Diff.	(Diff.-$\bar{x}$)^2	Score	Diff.	(Diff.$\bar{x}$)^2	Score	Diff.	(Diff.$\bar{x}$)^2
1	6	7	1	1.96	3	-3	1	3	-3	0.01	2	-4	2.25
2	6	7	1	1.96	4	-2	0	3	-3	0.01	3	-3	0.25
3	8	9	1	1.96	8	0	4	3	-5	4.41	4	-4	2.25
4	8	5	-3	6.76	3	-5	9	2	-6	9.61	3	-5	6.25
5	8	5	-3	6.76	3	-5	9	3	-5	4.41	4	-4	2.25
6	1	1	0	0.16	1	0	4	1	0	8.41	1	0	6.25
7	3	1	-2	2.56	1	-2	0	1	-2	0.81	3	0	6.25
8	4	4	0	0.16	3	-1	1	2	-2	0.81	3	-1	2.25
9	4	6	2	5.76	4	0	4	3	-1	3.61	2	-2	0.25
10	3	2	-1	0.36	1	-2	0	1	-2	0.81	1	-2	0.25
	$\bar{x}$	-0.4			-2			-2.9			-2.5		
	Std. Dev.	1.777			1.886			1.912			1.780		
	t	-0.712			-3.354			-4.796			-4.443		
	Critical t	2.262			2.262			2.262			2.262		
	p	0.494466			0.008468			0.000979			0.00167		

Paired samples t-test results suggest that, the difference in scores were significantly different than X=>YZ in XZ=>Y, X=>Y, and X=>Z. The p-scores < 0.05 signify that the differences in scores determined for those variants were significantly different than those of X=>YZ. According to this result, the hypotheses H_1, H_2, H_3 were rejected. On the other hand, the p-score obtained regarding the difference of pair of scores for X=>YZ and XY=>Z was considerably high (p=0.494466>0.05).

According to this result, we might argue that the difference among the scores in those variants is not statistically significant.

We conclude that the redundancy of 3 variants were confirmed by our results. YX => Z was an exception, and this was related to specific events hitting upon the Z component which are at the consequent of the rules.

6 Conclusion

Association rule mining is a promising method to analyze datasets in child psychiatry, especially for extracting previously unknown but probably interesting relations across variables. As occasionally underlined in prior research, rule mining techniques might involve challenges due to the high number of findings being obtained, which prevents effective evaluation of the results.

Our methodology has employed a rule-of-thumb for redundancy identification proposed in [1] to overcome this problem. Among the top 10 rules in the form of X=>YZ, we attempted to examine whether this rule-of-thumb holds. For this purpose, hypotheses were formulated and tested by examining the differences in "interestingness" scores assigned by a domain expert for rules and their variants.

To our knowledge, this is the first study that employs this rule redundancy definition across variants and conducts a cross-check for validation. In our case, the redundancy of three variants of rules in X=>YZ form was confirmed with statistical analysis, which helps to filter out irrelevant association rules, and saves time and effort for evaluation of results in this dataset in further studies.

References

[1] Aggarwal, C. C., & Yu, P. S. "A new approach to online generation of association rules". *IEEE Transactions on Knowledge and Data Engineering*, 13(4), 527-540, 2001.

[2] Agrawal, R., Imieliński, T., & Swami, A. "Mining association rules between sets of items in large databases". *In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207-216, 1993.

[3] Zaki, M. J. "Generating non-redundant association rules". *In Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 34-43, 2000.

[4] Rajan, R. H., & Dhas, J. P. M. "A method for classification based on association rules using ontology in Web data". *International Journal of Computer Applications*, 49(8), 13-17, 2012.

[5] Cheng, J., Ke, Y., & Ng, W. "Effective elimination of redundant association rules". *Data Mining and Knowledge Discovery*, 16, 221-249, 2008.

[6] Xin, Y., Na, W., & Chunyu, W. "A new method for eliminating redundant association rules". In *Proceedings of the 2010 International Conference on Intelligent Computation Technology and Automation*, 32-36, 2010.

[7] Han, J., Kamber, M., Pei, J. *Data Mining – Concepts and Techniques*, Third Edition. The Morgan Kauffman Publishers, Waltham, MA, USA, 2012.

[8] Delen, D. *Real-World Data Mining: Applied Business Analytics and Decision Making*. Pearson Education, New Jersey, USA, 2014.

[9] Han, J., Pei, J., & Yin, Y. Mining Frequent Patterns without Candidate Generation. *ACM SIGMOD Record*, 29(2), 1-12, 2000.

[10] Neeraj, S., & Swati, L. S. "Overview of non-redundant association rule mining". *Research Journal of Recent Sciences*, 1(2), 108-112, 2012.

[11] Sing'oei, L., & Wang, J. "Data mining framework for direct marketing: A case study of bank marketing". *International Journal of Computer Science Issues*, 10(2-2), 198-203, 2013.

[12] Ashrafi, M.Z., Taniar, D., Smith, K. "Redundant Association Rules Reduction Techniques". *In Proceedings of AI 2005: Advances in Artificial Intelligence. Lecture Notes in Computer Science*, 3809, 254-263.

[13] Miani, R. G. L., & Junior, E. R. H. "Eliminating Redundant and Irrelevant Association Rules in Large Knowledge Bases". *In Proceedings of the 20th International Conference on Enterprise Information Systems* (ICEIS 2018), 17-28, 2018.

[14] Mai, T., Nguyen, L. T., Vo, B., Yun, U., & Hong, T. P. "Efficient algorithm for mining non-redundant high-utility association rules". *Sensors*, 20(4), 1078, 2020.

[15] Zida, S., Fournier-Viger, P., Lin, J. C. W., Wu, C. W., & Tseng, V. S. "EFIM: a fast and memory efficient algorithm for high-utility itemset mining". *Knowledge and Information Systems*, 51(2), 595-625, 2017.

[16] Ordonez, C., & Zhao, K. "Evaluating association rules and decision trees to predict multiple target attributes". *Intelligent Data Analysis*, 15(2), 173-192, 2011.

[17] Kaufman, L, Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, USA, 1990.

[18] Koh, Y. S., & Rountree, N. "Finding Sporadic Rules Using Apriori-Inverse". *In Advances in Knowledge Discovery and Data Mining: 9th Pacific-Asia Conference, PAKDD 2005*, Hanoi, Vietnam, 97-106, 2005.

[19] Kabasakal, İ. "Understanding Shopping Behaviors with Category-and Brand-Level Market Basket Analysis". *In Tools and Techniques for Implementing International E-Trading Tactics for Competitive Advantage (Ed: Y. Meral)*, 242-267, IGI Global, 2020.

Nitelik Seçimi ve Yapay Öğrenme Algoritmaları ile Borsa İstanbul Hisse Senedi Yön Tahmini

Halit KARACA[1]*, Hakan GÜNDÜZ[1]

[1] *Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği, Kocaeli, TÜRKİYE*

Özet

Son yıllarda, borsa tahmini, yatırımcıların karlarını maksimize etme ve risklerini minimize etme çabaları nedeniyle büyük ilgi görmüştür. Borsa piyasalarının dinamik ve belirsiz yapısı, doğru tahmin yapmayı zorlaştırsa da, yapay öğrenme yöntemleri bu alanda umut vaat etmektedir. Bu çalışmada, Borsa İstanbul'da işlem gören hisselerin günlük verileri kullanılarak yapay öğrenme yöntemleri ile hisse senedi yön tahmini yapılması amaçlanmıştır. Çalışmada kullanılan veri seti, 30 Kasım 2015 ile 21 Aralık 2021 tarihleri arasında Borsa İstanbul'da işlem gören tüm hisselerin günlük sayısal göstergelerini içermektedir. Veri seti, Kaggle platformunda paylaşılmış olup minimum-maksimum fiyat, işlem hacmi gibi göstergeler ile osilatör/indikatör verilerini barındırmaktadır. Toplamda 1529 işlem gününe ait veriler üzerinde çalışılmış ve 38 adet osilatör/indikatör ve farklı zaman dilimlerine ait özellikler kullanılmıştır.

Nitelik seçimi için Maksimum Bilgi Katsayısı (Maximum Information Coefficient (MIC)) ve Öz Niteliklerin Yinelemeli Elemesi (Recursive Feature Elimination (RFE)) yöntemleri kullanılmıştır. Sınıflandırma aşamasında ise Destek Vektör Makineleri (Support Vector Machine (SVM)), Rastgele orman, k-en yakın komşu, karar ağacı, gradyan artırma, İki Yönlü Uzun Kısa Süreli Bellek (Bidirectional Long Short-Term Memory. (Bi-LSTM)) ve XGBoost algoritmaları uygulanmıştır. Üç bankacılık sektörü hissesinde yapılan testlerde, Garanti Bankası (GARAN) hissesi üzerinde RFE nitelik seçimi ile elde edilen sonuçlar doğrultusunda Rastgele orman modeli %64 yönsel doğruluk oranı sağlamıştır. Akbank (AKBNK) hissesinde SVM modeli, RFE nitelik seçimi ile %55,20 doğruluk oranı elde ederken, İş Bankası (ISCTR) hissesinde MIC nitelik seçme yöntemi ile SVM modeli %56 başarı oranına ulaşmıştır. Bu sonuçlar, Garanti Bankası hissesi üzerinde Rastgele orman modelinin en iyi performansı gösterdiğini ortaya koymaktadır.

Çalışmanın ileri aşamasında, makine öğrenmesi modellerinin özelliklerini değiştirerek ve nitelik seçme yöntemlerini geliştirerek başarı oranlarının daha da artırılması hedeflenmektedir.

Anahtar Kelimeler: Hisse senedi yön tahmini, nitelik seçimi, MIC, RFE

Borsa Istanbul Stock Direction Prediction with Feature Selection and Machine Learning Algorithms

Abstract

In recent years, stock market prediction has attracted great attention due to investors' efforts to maximize their profits and minimize their risks. Although the dynamic and uncertain structure of stock markets makes it difficult to make accurate predictions, artificial learning methods are promising in this area. In this study, it is aimed to make stock direction predictions using machine learning methods using daily data of stocks traded on Borsa Istanbul. The data set used in the study includes daily numerical indicators of all stocks traded on Borsa Istanbul between November 30, 2015, and December 21, 2021. The data set is shared on the Kaggle platform and includes indicators such as minimum-maximum price, trading volume, and oscillator/indicator data. A total of 1529 trading days of data were studied and 38 oscillators/indicators and features from different time periods were used.

*Contact email: 215129007@kocaeli.edu.tr

Maximum Information Coefficient (MIC) and Recursive Feature Elimination (RFE) methods were used for feature selection. In the classification phase, Support Vector Machine (SVM), Random Forest, k-nearest neighbor, decision tree, Gradient boosting, Bidirectional Long Short-Term Memory (Bi-LSTM) and XGBoost algorithms were applied. In the tests conducted on three banking sector stocks, the Random Forest model provided 64% directional accuracy rate in line with the results obtained with RFE feature selection on Garanti Bank(GARAN) stock. While the SVM model achieved 55.20% accuracy rate with RFE feature selection on Akbank (AKBNK) stock, the SVM model reached 56% success rate with MIC feature selection method on İş Bank (ISCTR) stock. These results reveal that the Random Forest model showed the best performance on Garanti Bankası stock.

In the advanced stage of the study, it is aimed to further increase the success rates by changing the features of the machine learning models and developing feature selection methods.

Keywords: *Stock direction prediction, feature selection, MIC, RF*

1 Giriş

Günümüzde, insanlar geleceklerini güvence altına almak ve daha kaliteli bir yaşam sürmek amacıyla yatırımlar yapmaktadır. Geçmişte olduğu gibi, günümüzde de borsa, yüksek getiri potansiyeline sahip bir yatırım aracı olarak öne çıkmaktadır. Son yıllarda bilgi teknolojisinde yaşanan önemli ilerlemeler, işletmelerin faaliyet biçimlerinde köklü değişikliklere yol açmıştır. Bu gelişmelerin en dikkat çekici olanlarından biri, finansal piyasaların ülke ekonomisi üzerindeki belirgin etkisidir. Hisse senedi alım satımı, özellikle teknolojik yenilikler sayesinde giderek daha fazla ilgi çekmektedir. Yatırımcılar, karlarını maksimize edebilecek ve risklerini minimize edebilecek araç ve tekniklerin peşindedir. Ancak hisse senedi piyasası tahmini (SMP), doğrusal olmayan, dinamik, stokastik ve belirsiz doğası nedeniyle oldukça karmaşıktır. SMP, önceki verileri analiz ederek gelecekteki verileri öngören bir zaman serisi tahmin türüdür ve bu alan, ekonomi, matematik, malzeme bilimi ve bilgisayar bilimi gibi farklı disiplinlerdeki analistlerin üzerinde çalıştığı bir konu olarak öne çıkmaktadır.

Hisse senedi piyasası, bir hissenin piyasa değeri, şirketin performansı, hükümet politikaları, ülkenin Gayri Safi Yurtiçi Hasılası (GSYİH), enflasyon oranı ve doğal afetler gibi birçok parametreye bağlıdır. Etkin Piyasa Hipotezi, borsa fiyatlarının büyük ölçüde yeni bilgilere dayandığını ve rastgele bir yürüyüş modeli izlediğini, dolayısıyla sadece geçmiş verilere dayanarak tahmin edilemeyeceğini öne sürmektedir. Ancak, teknolojik gelişmelerle birlikte araştırmacılar, borsa fiyatlarının belirli bir noktaya kadar tahmin edilebileceğini göstermiştir. Bu doğrultuda, teknolojinin sağladığı imkanlar ile yatırım araçlarında daha az risk ve daha güvenli gelir elde etme çabaları ön plana çıkmıştır.

Bu bağlamda, yapay öğrenme yöntemleri ile borsa tahmini çalışmaları büyük önem kazanmıştır. Çalışmada, diğer teknolojik gelişmeler ve teknikler ile yapılan araştırmalar ve analizler incelenmiş, Borsa İstanbul hisseleri için günlük veriler kullanılarak yön tahmini yapılması amaçlanmıştır.

2 İlgili çalışmalar

Borsa tahmini alanında birçok yöntem kullanılarak çeşitli araştırmalar yapılmıştır. Bu bölümde, yapılan bazı önemli çalışmalar hakkında bilgi verilmektedir.

Şişmanoğlu ve arkadaşları [1], Uzun Kısa Süreli Bellek (Long Short-Term Memory (LSTM)), Kapılı Tekrarlayan Birim (Gated Recurrent Unit (GRU)) ve Bi-LSTM olmak üzere üç farklı derin öğrenme modeli kullanarak bir hisse senedi tahmin sistemi geliştirmiştir. Çalışmada, veri seti olarak 1968-2018 yılları arasındaki New York Borsası'ndan elde edilen hisse senedi zaman serisi değerleri kullanılmıştır. Özellikle IBM hisse senedi üzerinde yapılan testlerde, son beş günlük işlem verilerinin girdi olarak kullanıldığı Bi-LSTM modeliyle %63.54 yönsel doğruluk oranına ulaşılmıştır.

Gündüz [2] ise filtre tabanlı nitelik seçimi ve topluluk öğrenme yaklaşımlarıyla Borsa İstanbul Enerji Endeksi (XKMYA) yön tahmini çalışmasını gerçekleştirmiştir. Bu çalışmada, finansal haber sitelerinde yayınlanan ekonomi haberleri kullanılarak endeksin günlük fiyat değişim yönleri tahmin edilmiştir. Endeksin hareket yönüne etki

eden kelimeler, filtre tabanlı Simetrik Belirsizlik (SU) ve Fisher Puanı (F-P) nitelik seçme yöntemleriyle belirlenmiş ve LightGBM sınıflandırıcısı ile değerlendirilmiştir. Sonuçlar, XKMYA endeksinin günlük yön tahmininde %68 Makro-Ortalama (MO) F-ölçütü ve %72 doğruluk oranına ulaşıldığını göstermektedir.

Türkalp [3], Twitter verileri üzerinde sınıflandırma algoritmaları kullanarak Amerikan Dow Jones (DJIA) Borsası'nda işlem gören büyük şirketlerin hisse senetlerinin yön tahminini yapmıştır. Naive Bayes, Rastgele Orman, Destek Vektör Makinesi, Karar Ağacı, k-En Yakın Komşu ve Yapay Sinir Ağları gibi sınıflandırma algoritmaları kullanılmıştır. Sonuçlar, %77.37 ile en başarılı tahminin Rastgele Orman algoritmasıyla, %61.41 ile en düşük başarının ise Destek Vektör Makinesi ile elde edildiğini göstermiştir. General Motors (GM) hisse senedi için en iyi tahmin başarısı %83.3, General Electric (GE) için ise %62.15 olarak bulunmuştur.

Kemalbay ve Alkış [4], Borsa endeks hareket yönünün çoklu lojistik regresyon ve k-En Yakın Komşu algoritması ile tahmini üzerine çalışmıştır. BIST100 endeks hareket yönünü tahmin etmek için istatistiksel olarak anlamlı bulunan bağımsız değişkenler kullanılmıştır. Çalışma, lojistik regresyon analizinin, K-en Yakın Komşu algoritmasına göre %81 doğruluk oranı ile daha iyi performans gösterdiğini ortaya koymuştur.

Yakut ve arkadaşları [5], Borsa İstanbul (BIST) endeksinin tahmin edilmesi için bir çalışma yapmışlardır. BIST endeksinin bir, iki ve üç gün öncesine ait değerleri, Amerikan dolar kuru, gecelik faiz oranı ve çeşitli yabancı borsaların (NIKKEI, BOVESPA, FTSE, CAC, DAX) endeks değerleri kullanılmıştır. İleri beslemeli yapay sinir ağları ve Destek Vektör Makineleri yöntemleriyle yapılan tahminlerde, her iki yöntemin de borsa endeksi tahmininde kullanılabileceği gösterilmiştir.

Gündüz tarafından yapılan başka bir çalışmada [6], ekonomi haberleri kullanılarak Borsa İstanbul (BIST) 100 endeksinin günlük açılış fiyatının yönü tahmin edilmiştir. Haber metinleri, Kamu Aydınlatma Platformu (KAP) ve finansal internet sitelerinden alınmıştır. BIST 100 endeksinin anlamlı yön değişimleri %74 doğruluk ve %68.4 Makro-Ortalama F-Ölçütü oranlarıyla tahmin edilmiştir. Bilgi Kazanımı ve Ki-Kare istatistiği nitelik seçme yöntemlerinin sınıflandırma performansına etkileri incelenmiş ve %78.4 doğruluk oranı ile en iyi sonuçlar elde edilmiştir.

Gezici ve Sefer [7], varlık fiyatlarını ve yönsel fiyat hareketlerini tahmin etmek için daha gelişmiş Transformer tabanlı ve Patch tabanlı yaklaşımlar kullanmıştır. Vision Transformer (ViT), Data Efficient Image Transformers (DeiT) ve Swin gibi modellerin kullanıldığı çalışmada, ViT modelinin özellikle uzun test ve tutma sürelerinde daha iyi performans gösterdiği belirlenmiştir.

Akyıldırım ve arkadaşları [8], hisse senedi fiyat hareketlerini geçmiş fiyatlara göre tahmin etmek için çeşitli makine öğrenmesi tekniklerini karşılaştırmıştır. Rastgele Orman ve Destek Vektör Makinesi modelleri hem gelecekteki fiyat yönlerini hem de yüzde değişimlerini yüksek doğrulukla yakalayabilmiştir.

Sonkavde ve arkadaşları [9], denetimli ve denetimsiz makine öğrenmesi algoritmalarını, topluluk yöntemlerini, zaman serisi analiz tekniklerini ve derin öğrenme algoritmalarını incelemişlerdir. Bu çalışma, finans sektöründe kullanılan makine öğrenimi ve derin öğrenme modellerinin genel bir açıklamasını sunmuş ve "Rastgele Orman + XG-Boost + LSTM" topluluk modelinin performansını diğer popüler modellerle karşılaştırmıştır.

3 Materyal ve metotlar

Bu bölümde, kullanılan veri seti, etiketleme yöntemi ve araştırmada kullanılan algoritmalar (Destek Vektör Makineleri(SVM), Rastgele Orman, k-En Yakın Komşu, Karar Ağacı, Gradyan Artırma, Bi-LSTM ve XGBoost) detaylı olarak incelenmiştir.

3.1 Veri seti

Çalışmada kullanılan veri seti, Keser [10] tarafından Kaggle platformunda paylaşılan ve Borsa İstanbul'da işlem gören hisselerin 30 Kasım 2015 ile 21 Aralık 2021 tarihleri arasındaki günlük sayısal göstergelerini içermektedir. Bu göstergeler minimum-maksimum fiyat, işlem hacmi gibi verileri kapsamaktadır. Araştırma, bankacılık sektörüne ait Garanti Bankası (GARAN), Akbank (AKBNK) ve İş Bankası (ISCTR) hisseleri üzerinde gerçekleştirilmiştir. Veri seti, 46 sütun ve 1529 işlem günü verisinden oluşmaktadır.

Önceki çalışmalarda, Borsa İstanbul BIST100 ve enerji endeksleri, Twitter verileri, Dow Jones (DJIA) borsasından elde edilen hisse senetlerine ilişkin veriler ve ayrıca Investing.com sitesinden elde edilen veriler kullanılmıştır. Bu çalışmada kullanılan veri seti, NYSE, NASDAQ ve NYSE MKT'de işlem gören ABD merkezli hisse senetlerinin 2 Ocak 1968 ile 9 Nisan 2018 arasındaki günlük fiyat verilerini içermektedir.

Veri girişi olarak, altı farklı mimaride kullanılan 46 özellik arasından, osilatör ve indikatörlerle uyumlu 11 nitelik seçilmiştir. Buna ek olarak, son 5, 10, 15, 20 ve 25 günlük veriler ile osilatör ve indikatörlerin dahil edildiği 55 özellik kullanılmıştır. Toplamda, veri seti 1529 işlem gününü kapsamaktadır. Modelin eğitim aşamasında, her iterasyonda 60 günlük veriler eğitim için kullanılarak, 5 günlük veriler tahmin edilmiştir. Bu süreç, 60 gün kaydırılarak tekrar edilmiştir. 2015-2021 yılları arasındaki bu veri seti, büyük veri olarak adlandırılabilecek bir büyüklükte olmasa da kapsamlı bir analiz için yeterlidir.

3.2 Etiketleme yöntemi

Veri setinin, araştırmanın amacına uygun şekilde eğitilebilmesi için etiketleme işlemi gereklidir. Bu işlem, bir veri kümesindeki her bir örneğin hedef değişkenini veya sınıfını belirlemek amacıyla yapılır.

Çalışmada, veri setini etiketlemek için "SABİT KALIR", "DÜŞER" ve "YÜKSELİR" şeklinde üç sınıflı bir etiketleme yöntemi kullanılmıştır. Bu sınıflandırma, Tablo 1'de sunulan 3x3 boyutundaki bir karışıklık matrisi (Confusion Matrix) ile gösterilmiştir.

Tablo 1. Karışıklık Matrisi (Confusion Matrix)

Gerçek\ Tahmin	SABİT KALIR	DÜŞER	YÜKSELİR
SABİT KALIR	TP_Sabit	FP_Sabit_1	FP_Sabit_2
DÜŞER	FN_Düşer_1	TP_Düşer	FP_Düşer
YÜKSELİR	FN_Yükselir_1	FN_Yükselir_2	TP_Yükselir

Karışıklık Matrisinde yer alan elemanların anlamları aşağıda listelenmiştir.

TP_Sabit: Modelin "SABİT KALIR" olarak doğru tahmin ettiği örnekler (True Positive).

FP_Sabit_1: Modelin "SABİT KALIR" olarak tahmin ettiği ancak gerçekte "DÜŞER" olan örnekler (False Positive).

FP_Sabit_2: Modelin "SABİT KALIR" olarak tahmin ettiği ancak gerçekte "YÜKSELİR" olan örnekler (False Positive).

FN_Düşer_1: Modelin "DÜŞER" olarak tahmin ettiği ancak gerçekte "SABİT KALIR" olan örnekler (False Negative).

TP_Düşer: Modelin "DÜŞER" olarak doğru tahmin ettiği örnekler (True Positive).

FP_Düşer: Modelin "DÜŞER" olarak tahmin ettiği ancak gerçekte "YÜKSELİR" olan örnekler (False Positive).

FN_Yükselir_1: Modelin "YÜKSELİR" olarak tahmin ettiği ancak gerçekte "SABİT KALIR" olan örnekler (False Negative).

FN_Yükselir_2: Modelin "YÜKSELİR" olarak tahmin ettiği ancak gerçekte "DÜŞER" olan örnekler (False Negative).

TP_Yükselir: Modelin "YÜKSELİR" olarak doğru tahmin ettiği örnekler (True Positive).

Veri kümesindeki örneklere "Sabit Kalır", "Düşer" ve "Yükselir" etiketleri atanırken günlük verisinin açılış fiyatı ile kapanış fiyatının ortalaması alınmış ve hesaplanan ortalama fiyatın bir önceki günkü ortalama fiyata değişim oranı kullanılmıştır. Oluşturulan etiketlerin sınıf dağılımları Tablo 2'de listelenmiştir.

Tablo 2. Sınıf etiketleri dağılımı.

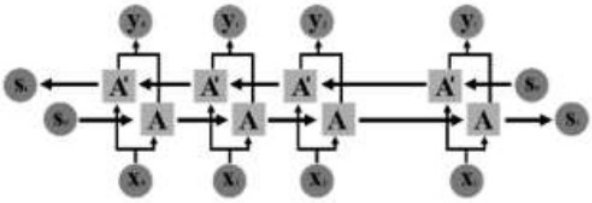
Karar dağılımı	Adet
0(SABİT)	764
1(YÜKSELİR)	362
2(DÜŞER)	362

3.3 Metotlar

3.3.1 İki Yönlü Uzun Kısa Süreli Bellek

İki Yönlü Uzun Kısa Süreli Bellek (Bi-LSTM) modeli, zaman serisi veya sıralı verilerde ileri ve geri yönde bilgi akışını yakalayabilen çift yönlü Uzun Kısa Süreli Bellek (LSTM) katmanlarına sahip bir sinir ağı modelidir. Bu tür modeller, özellikle doğal dil işleme, konuşma tanıma ve finansal zaman serisi tahmini gibi uygulamalarda oldukça kullanışlıdır. Bi-LSTM yapısı, önceki ve gelecekteki gösterimlerle ileri geri öğrenme imkânı tanıyan iki tür bağlantıya

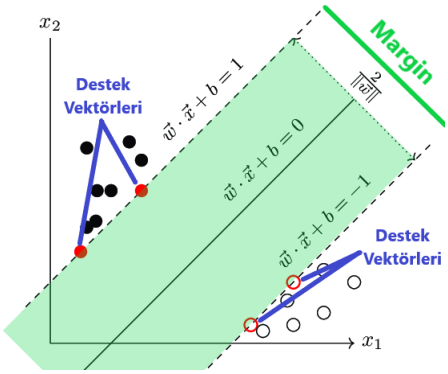
sahiptir. Bu çift yönlü bağlantılar sayesinde Bi-LSTM, zaman serisi verilerinde hem geçmiş hem de gelecekteki bilgileri dikkate alarak daha çeşitli ve doğru tahminler yapabilir. Bu özellik, modelin bağlamsal bilgiyi daha iyi yakalamasına yardımcı olur ve farklı uygulama alanlarında önemli avantajlar sunar [11].



Şekil 1. Bi-LSTM modeli iç yapısı [11].

3.3.2 Destek vektör makineleri

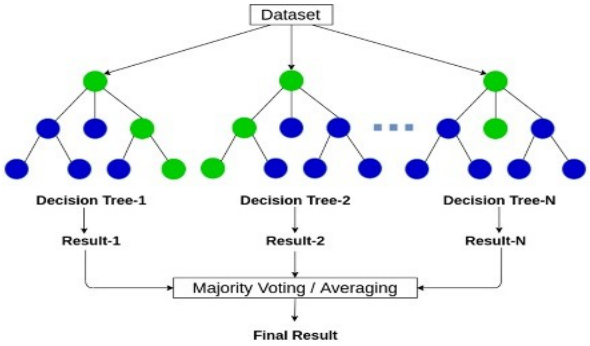
Destek Vektör Makineleri sınıflandırma, regresyon ve aykırı değer tespiti gibi problemler için kullanılan bir denetimli öğrenme algoritmasıdır. Yüksek boyutlu verilerde etkili olması nedeniyle avantajlıdır, ancak özellik sayısının örnek sayısından fazla olması durumunda etkili olamayabilir. SVM algoritması, düzlem üzerine yerleştirilmiş en az iki veya daha fazla veri kümesini ayırmak için doğrular çizer. Algoritmanın amacı, bu veri kümelerini ayıran doğruları, kümeler arasındaki mesafeyi maksimize edecek şekilde yerleştirmektir [12].



Şekil 2. SVM modelinin grafiksel gösterimi [12].

3.3.3 Rastgele orman

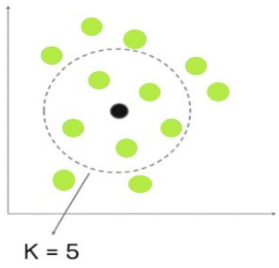
Rastgele Orman modeli, sınıflandırma ve regresyon problemlerinde kullanılan bir denetimli öğrenme algoritmasıdır. Daha kısa eğitim süresi gerektirmesi ve yüksek performans elde etmesi nedeniyle sıklıkla tercih edilir. Bu algoritma, veri kümesinin alt kümelerinde bulunan karar ağaçlarını kullanarak sınıflandırma yapar ve bu ağaçların ortalamasını alarak son tahmini gerçekleştirir. Eğitim setinde belirlenen bir k noktası için karar ağaçları oluşturulur ve her ağaç için tahminler yapılır. Algoritma, çoğunlukla bu tahminlerden elde edilen kategoriyi seçer [13].



Şekil 3. Rastgele Orman modelinin sınıflandırma süreci [13].

3.3.4 K-en yakın komşu

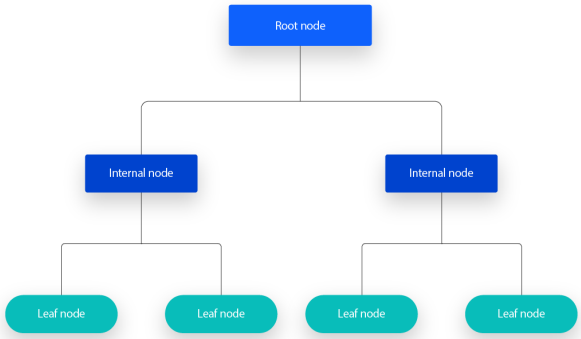
K-En Yakın Komşu (KNN) algoritması, regresyon ve sınıflandırma problemleri için kullanılan bir denetimli makine öğrenmesi algoritmasıdır. KNN, uzaklık ölçüm tekniklerini kullanarak tahminler yapar. Bir veri noktasının hangi sınıfa ait olduğunu tahmin etmek için, en yakın k adet veri noktası tespit edilir ve bu noktaların özelliklerine göre sınıflandırma yapılır [14].



Şekil 4. KNN şeması [14].

3.3.5 Karar ağacı

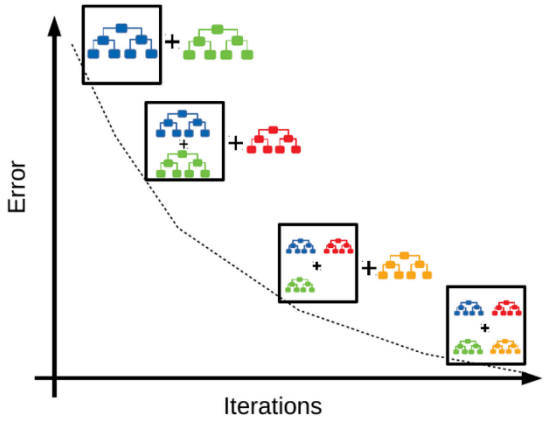
Karar Ağacı algoritması, doğrusal olmayan verileri sınıflandırmak için esnek bir yöntemdir ve veri ön işleme gerektirmemesi gibi avantajları vardır. Ancak, eğitim setindeki değişikliklere karşı hassas olması gibi dezavantajları da bulunmaktadır. Karar ağacı algoritması, veri kümesini düğüm adı verilen alt kümelere böler ve her düğüme girdilerinin değerlerine göre bir etiket atar. Algoritma, düğümler arasında sırayla ilerler ve her düğümün çıktısını test eder. Eğer düğümdeki çıkış veri setindeki etiketlerle eşleşirse, bu düğüm algoritmanın bir sonucu olarak seçilir; aksi takdirde, işlem diğer düğümlerde devam eder [15].



Şekil 5. Karar Ağacı şeması [15].

3.3.6 *Gradyan artırma*

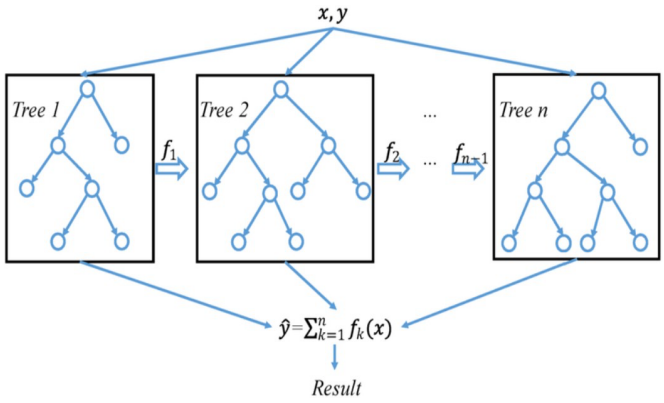
Gradyan Artırma modeli, gradyan inişi kavramını temel alan ve sınıflandırma ile regresyon problemlerinde kullanılan bir ensemble algoritmasıdır. Bu algoritma, zayıf tahmin edici karar ağaçlarını bir araya getirerek güçlü bir tahmin modeli oluşturmayı amaçlar. Hesaplama sürecinde, belirli bir başlangıç noktası belirlenir ve bu noktadan itibaren fonksiyonun gradyanı hesaplanır. Algoritma, bu gradyan boyunca adım adım ilerleyerek fonksiyonun değerini minimize etmeye çalışır [16].



Şekil 6. Gradyan Artırma sınıflandırma örneği [16].

3.3.7 *XGBoost*

XGBoost algoritması, karar ağaçları üzerinde çalışan bir gradyan artırma algoritmalarından biridir. Genellikle tahmin etme ve sınıflandırma problemlerinde kullanılır. Algoritma, iteratif bir şekilde çalışır ve her iterasyonda yeni bir model eklenir. Yeni model, bir önceki modelin hatalarını düzeltmeye çalışır. Hedef, tüm modellerin toplamının gerçek çıktıya olabildiğince yaklaşan bir tahmin oluşturmasını sağlamaktır. XGBoost, diğer gradyan artırma modellerinden, modelin karmaşıklığını kontrol eden düzenlileştirme terimleri içermesi, esnek kullanımı ve büyük veri setlerini işleme yetenekleri ile ayrılır [17].



Şekil 7. XGBoost şeması [18].

4 Deneysel Sonuçlar ve Değerlendirme

Yapılan çalışmalar incelendiğinde borsa piyasalarında doğru tahminler yapmak, borsa zaman serilerinin doğrusal olmayışı nedeniyle karmaşık bir süreçtir. Bu alandaki araştırmacılar, hisse senedi fiyatlarının rastgele bir hareket icra ettiğini dile getirse de teknik analistler, bazı güncel değerleri dikkate alarak gelecek fiyatların anlamlı bir doğruluk oranıyla tahmin edilebileceğini öne sürmüştür [2].

Bu çalışmada kullanılmış olan makine öğrenmesi modellerinin performanslarını analiz etmek için doğruluk oranı metriği kullanılmıştır. Bu metrik sayesinde, modellerimizin performansını hem genel hem de sınıf bazında detaylı bir şekilde değerlendirmek amaçlanmıştır. Sınıflandırma modellerinde kullanılan parametre kümeleri Tablo 3'te sunulmuştur.

Tablo 3. Model parametre kümeleri.

Derin Öğrenme Modeli	Kullanılan parametreler
Bi-LSTM	Toplam 3 katmandan oluşmaktadır. Birinci Katman(64 neurons,activation=tanh,optimizer=adam, dropout=0.3) İkinci Katman(32 neurons,activation=tanh,optimizer=adam, dropout=0.3) Üçüncü Katman(16 neurons,activation=tanh,optimizer=adam, dropout=0.3)
SVM	Kernel=linear C=1
Rastgele Orman	n_estimators=500,max_dept=10 random_state=42
XGBoost	Standart Parametreler
K-en yakın komşu (KNN)	n_neighbors=3
Karar ağacı	Standart Parametreler
Gradyan Artırma	Standart Parametreler

Tablo 4'te Garanti Bankası (GARAN), Akbank (AKBNK) ve İş Bankası (ISCTR) hisse senetleri üzerinde derin öğrenme yaklaşımlarından Bi-LSTM, SVM, Rastgele Orman, XGBoost, KNN, Karar Ağacı ve Gradyan Artırma modellerinin performans sonuçları sunulmuştur. Bu sonuçlar, hem öznitelik seçimi yapılmadan (SADE), hem MIC, hem de RFE yöntemi ile seçilen 10 öznitelik kullanılarak elde edilmiştir.

Üç bankacılık sektörü hissesi üzerinde gerçekleştirilen test ve deneylerde, GARAN hissesi için RFE nitelik seçimi Rastgele Orman makine öğrenmesi modeli %64 yönsel doğruluk oranı sağlamıştır. AKBNK hissesi üzerinde ise RFE nitelik seçimi ile SVM makine öğrenmesi modeli %55.20 başarı oranına ulaşmıştır. ISCTR hissesi için MIC nitelik seçimi kullanılarak SVM makine öğrenmesi modeli %56 başarı oranı elde etmiştir. Modeller arasında, Garanti Bankası'nda Rastgele Orman %64 yönsel doğruluk oranı ile en iyi sonucu vermiştir.

Öte yandan, bazı modellerde (örneğin, Karar Ağacı Sınıflandırıcısı), RFE veya MIC kullanımı performans üzerinde çok az bir değişiklik göstermiş veya aynı seviyede kalmıştır. Bu durum, bu modellerin öznitelik sayısına daha az duyarlı olabileceğini düşündürmektedir.

Tablo 4. Sonuçlar.

Derin Öğrenme Modeli	**GARAN**			**AKBNK**			**İSCTR**		
	SADE	**MİC**	**RFE**	**SADE**	**MİC**	**RFE**	**SADE**	**MİC**	**RFE**
Bi-LSTM	%17.60	%19.20	%32.80	% 22.00	%26.40	%27.20	%27.20	%24.80	%36.80
SVM	%55.20	%60.80	%58.40	% 48.66	%50.4	%55.20	%53.60	%56.00	%51.20
Rastgele Orman	%55.20	%63.20	%64.00	% 41.66	%51.2	%52.00	%47.20	%44.00	%52.00
XGBoost	%56.80	%56.00	%52.80	% 44.00	%45.60	%47.20	%48.80	%41.60	%46.40
K-en yakın komşu (KNN)	%51.20	%57.60	%54.40	% 38.33	%44.80	%55.20	%52.80	%52.80	%49.60
Decision Tree Classifier	%44.00	%50.40	%40.00	% 41.00	%39.20	%40.00	%45.60	%40.80	%44.00
Gradient Boosting Classifier	%46.40	%53.60	%53.60	% 42.66	%45.60	%50.40	%47.20	%37.60	%48.80

Genel olarak elde edilen sonuçlar, öznitelik seçimi yöntemlerinin model performansını artırmada önemli bir rol oynadığını göstermektedir. Bu bağlamda, özellikle yüksek boyutlu veri setlerinde öznitelik seçimi süreçlerinin dikkatle ele alınması gerektiği vurgulanmaktadır.

Gelecekteki çalışmalarda, makine öğrenmesi modellerine ait özellikleri değiştirerek ve nitelik seçme yöntemlerini geliştirerek başarı oranının artırılması hedeflenmektedir.

Kaynaklar

[1] G. Şişmanoğlu, F. Koçer, M.A. Önde, Ö.K. Şahingöz, 2020, BEÜ Fen Bilimleri Dergisi 9 (1), 434-445.

[2] Gündüz, H., 2020, Filtre Tabanlı Nitelik Seçimi ve Topluluk Öğrenme Yaklaşımlarıyla Borsa İstanbul Enerji Endeksi Yön Tahmini, Avrupa Bilim ve Teknoloji Dergisi Özel Sayı), 215 220

[3] Türkalp M.V.,2019, Tviter Verileri Üzerinde Sınıflandırma Algoritmaları Kullanarak Hisse Senedi Değerleri İçin Yön Tahmini,Konya Teknik Üniversitesi Bilgisayar Müh. Yüksek Lisans Tezi, Konya.

[4] Kemalbay G., AlkışB.N.2021, Borsa Endeks Hareket Yönünün Çoklu Lojistik Regresyon Ve K-En Yakın Komşu Algoritması İle Tahmini, Pamukkale Univ Muh Bilim Derg, 27(4), 556-569,

[5] Yakut E., Elmas B.,Yavuz S., 2014, Yapay Sinir Ağları Ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini, Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fak. Derg. 19,(1),139-157

[6] Gündüz, H.,2013, Borsa İstanbul (BIST) 100 Endeksi Yönünün Ekonomi Haberleri İle Tahmin Edilmesi İstanbul Teknik Üniversitesi FBE Yüksek Lisans Tezi,

[7] Gezici A.H.B.,Sefer E., 2024, Deep Transformer-Based Asset Price and

Direction Prediction, IEEE Access (Volume: 12), 24164 – 24178

[8] Akyıldırım, E., Bariviera, AF, Nguyen, DK ve ark. Yüksek frekanslı hisse senedi getirilerinin tahmin edilmesi: alternatif yöntemlerin karşılaştırılması. Ann Oper Res 313, 639–690 (2022).

[9] Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D., & Bhat, S. K. (2023). Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. International Journal of Financial Studies, 11(3),94.

[10] Kaggle. "Borsa İstanbul Turkish Stock Exchance Dataset" https://www.kaggle.com/datasets/gokhankesler/borsa-istanbul-turkish-stock-exchange-dataset(2021)

[11] Nissimagoudar P.,Patil A., HM G., Nandi A., 2021, Deep convolutional-recurrent neural network model for driving alertness detection, IJECE, 11(4):3529

[12] Medium. "Support Vector Machine Algoritması" https://mfatihto.medium.com/support-vector-machine-algoritmas%C4%B1-makine-%C3%B6%C4%9Frenmesi-8020176898d8 (14.07.2021)

[13] Medium, Sevinç D. "Rastgele Orman Algoritması". https://denizsevincc.medium.com/random-forest-rassal-orman-algoritmas%C4%B1-7f4b59d962d5 (05.05.2021)

[14] Miuul "K-en yakın komşu Algoritması". https://miuul.com/blog/k-en-yakin-komsu-algoritmasi-nasil-calisir(20.07.2022)

[15] Bulutistan."Makine Ögrenmesi, Karar Agaci (Decision Tree)". https://bulutistan.com/blog/makine-ogrenmesi-karar-agaci-decision-tree-nedir/(18.08.2023)

[16] Medium. "Gradient Boosting". https://medium.com/@hemashreekilari9/understanding-gradient-boosting-632939b98764(05.09.2023)

[17] Medium. "XGBoost Algoritması" https://efecanxrd.medium.com/xgboost-algoritmas%C4%B1-6703b14efd5. (27.06.2023)

[18] Y. Wang, 2019, A hybrid ensemble method for pulsar candidate classification, Astrophysics and Space Science 364(8)

Yapay Öğrenme Yöntemleriyle Ses Kayıtlarının Metne Dönüştürülmesi

Sümeyra BAŞ[1]*, Hakan GÜNDÜZ[2]

[1]*Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği, Kocaeli, TÜRKİYE*
[2]*Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği, Kocaeli, TÜRKİYE*

Özet

Dijital teknolojilerin hızla gelişmesi ve günlük yaşamımızın bir parçası haline gelmesiyle birlikte, sesli iletişim araçları ve sesli verilerle etkileşim giderek daha önemli bir hale gelmiştir. Ses kayıtlarının metne dönüştürülmesi (speech-to-text), bu teknolojik gelişmelerin önemli bir parçasıdır. Gelişmiş yapay zekâ ve makine öğrenimi yöntemlerinin yaygınlaşmasıyla, konuşma tanıma sistemleri de büyük bir ilerleme kaydetmiştir. Yapay öğrenme yöntemleri, ses verilerini analiz ederek konuşmanın doğru bir şekilde metne dönüştürülmesini sağlamada kritik bir rol oynamaktadır. Bu yöntemler, ses dalgalarını işleyerek anlamlı özellikler çıkarır ve bu özellikler üzerinden öğrenme algoritmalarını kullanarak konuşmanın metne dönüştürülmesini gerçekleştirir. Bu çalışmada, yaklaşık 2.500 ses kaydından oluşan veri seti hazırlanmıştır. Ses kayıtları dört farklı aksan olarak (Karadeniz, Ege, Doğu, İç Anadolu) YouTube videoları, Spotify podcast'leri, TRT belgeselleri gibi Türkçe aksanlı yayın yapan platformlardan alınmıştır. Python kütüphaneleri ve kodları ile kayıtlar, en az bir kelime içerecek şekilde konuşmanın bittiği ilk yerden kesilerek kısa WAV dosyalarına bölünmüştür. Her aksana ait ses kayıtlarından python_speech_features kütüphanesindeki MFCC (Mel Frekans Kepstrum Katsayıları) özellikleri çıkarılarak vektörler oluşturulmuştur. Ayrıca, fbank() ve logfbank() fonksiyonları ile de özellikler çıkarılarak model doğrulukları değerlendirilmiş ve birbirleri ile karşılaştırılmıştır. Ses verileri üzerinde RandomForestClassifier, K-Nearest Neighbors (KNN) ve Support Vector Machine (SVM) ile ayrı ayrı modeller oluşturulmuştur. RandomForestClassifier modelinde hiperparametre ayarlamaları incelendikten sonra model doğrulukları karşılaştırılmıştır. Aynı şekilde, K-Nearest Neighbors (KNN) ve Support Vector Machine (SVM) ile oluşturulan modellerde yapılan parametre ayarlamaları ile en uygun doğruluk aranmıştır. Özellik çıkarımında kullanılan MFCC fonksiyon parametreleri değiştirildikten sonra, RandomForestClassifier modelinde %91,44, K-Nearest Neighbors (KNN) modelinde %95,92 ve Support Vector Machine (SVM) modelinde ise %96,74 model doğruluğuna ulaşılmıştır. Bu sonuçlar, Support Vector Machine (SVM) modelinin en iyi sonuç verdiğini göstermiştir. Çalışmanın ileri aşamasında, makine öğrenmesi modellerinin özelliklerini değiştirerek ve nitelik seçme yöntemlerini geliştirerek başarı oranlarının daha da artırılması hedeflenmektedir.

Anahtar Kelimeler: *Ses kayıtlarının metne dönüştürülmesi,aksan farklılıkları, MFCC, KNN, SVM*

Speech-to-Text Conversion Using Machine Learning Methods

Abstract

With the rapid development of digital technologies and their integration into our daily lives, voice communication tools and interaction with audio data have become increasingly important. Speech-to-text conversion is a significant aspect of these technological advancements. With the widespread use of advanced artificial intelligence and machine learning methods, speech recognition systems have also made significant progress. Machine learning methods play a critical role in accurately converting speech into text by analyzing audio data.

*Contact email: 215129002@kocaeli.edu.tr

These methods process sound waves to extract meaningful features and use learning algorithms to convert speech into text. In this study, a dataset consisting of approximately 2,500 audio recordings was prepared. The audio recordings were collected from platforms that broadcast Turkish-accented content, such as YouTube videos, Spotify podcasts, and TRT documentaries, representing four different accents (Karadeniz, Aegean, Eastern, Central Anatolian). Using Python libraries and code, the recordings were segmented into short WAV files, starting from the first silence after the speech ends, ensuring that each segment contains at least one word. Mel Frequency Cepstral Coefficients (MFCC) features were extracted from the recordings using the python_speech_features library, and vectors were created for each accent. Additionally, features were extracted using the fbank() and logfbank() functions, and the accuracy of the models was evaluated and compared. Separate models were created using RandomForestClassifier, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) on the audio data. After examining hyperparameter tuning in the RandomForestClassifier model, the accuracy of the models was compared. Similarly, the best accuracy was sought in models created using K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) through parameter adjustments. After modifying the parameters of the MFCC function used in feature extraction, accuracy rates of 91.44% in the RandomForestClassifier model, 95.92% in the K-Nearest Neighbors (KNN) model, and 96.74% in the Support Vector Machine (SVM) model were achieved. These results indicate that the Support Vector Machine (SVM) model produced the best results. In the advanced stage of the study, it is aimed to further increase the success rates by changing the features of the machine learning models and developing feature selection methods.

Keywords: *Speech-to-text conversion, accent differences, MFCC, KNN, SVM*

1 Giriş

Ses verilerinin dijital ortamda kullanılması giderek artmaktadır ve bu durum hayatın olağan akışını kolaylaştırmaya başlamıştır. Bu teknolojik gelişmeler eğitimde, sağlık sektöründe, çağrı merkezlerinde, medya içerik yönetimde kullanılmaktadır. Örneğin, eğitim alanında derslerin otomatik olarak yazıya dökülmesi, öğrencilere ders notlarına daha kolay erişim imkânı sunarken; sağlık sektöründe doktorların hasta notlarını hızlı bir şekilde kaydedebilmesi, tıbbi kayıtların daha etkin bir şekilde tutulmasını sağlamaktadır.

Türkiye'de yargı süreçlerinde, duruşmaların ilerleyişi genellikle hâkimin söylediklerinin kâtip tarafından yazılması ile sağlanmaktadır. Ancak, sesten metne dönüştürme teknolojisi kullanıldığında, bir duruşmanın ses kaydı doğrudan metne çevrilerek yazıya dökülebilir. Bu durum, yargı sürecinin hızlanmasına ve doğruluğun artırılmasına önemli katkılar sağlayacaktır. Ayrıca, ses kayıtlarının otomatik olarak metne dönüştürülmesi, yargı sisteminde dijitalleşmeye geçişi hızlandıracak ve kaynakların daha verimli kullanılmasına imkân tanıyacaktır.

Bu proje, özellikle Türkçe diline özgü farklı bölgesel aksanların ses kayıtlarından metne dönüştürülmesini amaçlamaktadır. Türkiye'de bölgesel aksanların zenginliği, ses verilerinin doğru bir şekilde işlenmesini daha da önemli hale getirmektedir. Proje kapsamında geliştirilen model, farklı aksanlardaki konuşmaları başarılı bir şekilde metne dönüştürmeyi hedefleyerek, dil teknolojilerinde yerelleştirilmiş çözümler sunacaktır. Bu çalışma, hem dilin dijital ortamda korunmasına katkı sağlayacak hem de Türkçe konuşan kullanıcılar için daha kişiselleştirilmiş ve doğru sonuçlar elde edilmesini mümkün kılacaktır.

Sonuç olarak, bu proje, sesin metne dönüştürülmesi alanındaki mevcut çözümlere önemli bir katkı sağlayarak, farklı aksanları barındıran Türkçe dilinde daha geniş bir kullanıcı kitlesine hitap eden, yerelleştirilmiş bir model geliştirilmesini hedeflemektedir.

2 İlgili çalışmalar

Son yıllarda sesten metne dönüştürme problemlerinin çözümünde önemli ilerlemeler kaydedilmiştir.

Davis ve Mermelstein (1980' de) "*Sürekli Konuşulan Cümlelerde Tek Hece İçeren Kelimelerin Tanınması için Parametrik Temsillerin Karşılaştırılması*" isimli makalelerini yayınlamışlardır [1]. MFCC yöntemini kullanarak monosyllabic kelimelerin tanınması üzerine çalışmalar yapmışlardır. İngilizce veri tabanı kullanmışlar ve fonetik olarak benzer tek heceli

kelimelerin akustik sinyal temsillerini karşılaştırmışlardır. Çalışmada, Mel-frekans kepstrum katsayıları (MFCC), doğrusal frekans kepstrum katsayıları (LFCC), doğrusal kestirim kepstrum katsayıları (LPCC), doğrusal kestirim spektrumu (LPC) ve yansıma katsayıları gibi farklı parametre setleri kullanılmışlardır. Mel-frekans kepstrum katsayıları (MFCC) ile yapılan deneylerde %96,5 ve %95,0 doğruluk oranları elde etmişlerdir [1].

Picone, J. W. (1993), sinyal modelleme teknikleri hakkında detaylı bir inceleme yapmıştır. Makalesinde konuşma sinyallerinin modellemesi hakkındaki dört temel işlemi (spektral şekillendirme, spektral analiz, parametrik dönüşüm ve istatistiksel modelleme) anlatmaktadır. Konuşma sinyallerinin işlenmesi ve modellenmesinde kullanılan teknikleri ve bunların performansa etkilerini incelemişlerdir [2].

Wang vd. (2020), "fairseq S2T: Fast Speech-to-Text Modeling with fairseq" isimli makalesini yayınlamışlardır. Ölçeklenebilirlik ve genişletilebilirlik tasarımına sahip uçtan uca iş akışlarını ve son teknoloji modellerini içeren FAIRSEQ S2T modelini sunmuşlardır. FAIRSEQ çalışmasında modern derin öğrenme modelleri olan RNN ve Transformer kullanılmışlardır. Yapay sinir ağı ve dönüştürücü tabanlı modeller ile açık kaynaklı ve ayrıntılı bir eğitim gerçekleştirmişlerdir. Konuşma verilerini analiz etmek ve verileri metne dönüştürmek için makine öğrenmesinden yararlanmışlardır [3].

Indurthi vd. (2020), otomatik ses tanıma (ASR) ve makine öğrenmesi ile sesten metne dönüştürme işlemi yapmışlardır. Meta-öğrenme algoritmasını (MAML) kullanmışlardır. Farklı dillerdeki konuşmalardan oluşan veri seti ile birlikte İngilizce-Almanca ve İngilizce-Fransızca dil çiftlerinde konuşma çevirisi yapan deneyler yapmışlardır [4].

Liu vd. (2020), çalışmasında önceden eğitilmiş modeller olan wav2vec 2.0 ve MBART kullanmışlardır. Bu şekilde konuşmadan metne çeviri için basit ve kullanışlı bir yaklaşım önermişlerdir[5].

Darwish vd. (2023), "Gerçekçi Koşullar Altında Yeniden Segmentasyon ve Terminoloji ile Çok Dilli Konuşma Çevirisinin Değerlendirilmesi" isimli makalelerinde, veri setlerinin oluşturma metodolojisini, bununla birlikte konuşma çevirisi modellerini değerlendirmek için düşüncelerini ve yöntemlerini ve zorlukları açıklamışlardır [6].

3 Kullanılan yöntemler

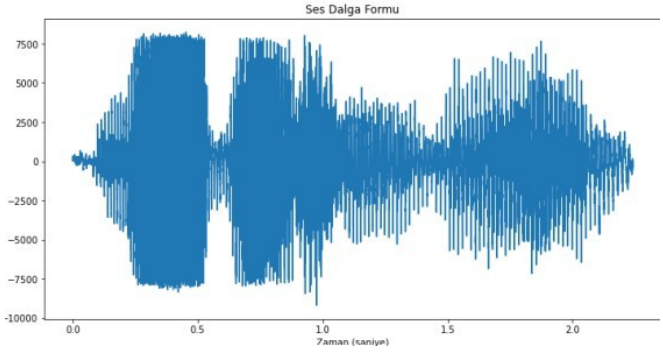
Bu bölümde veri seti, öznitelik oluşturma aşamaları, RandomForestClassifier, K-Nearest Neighbors (KNN) ve Support Vector Machine (SVM) modelleri detaylandırılmıştır.

3.1 Veri seti

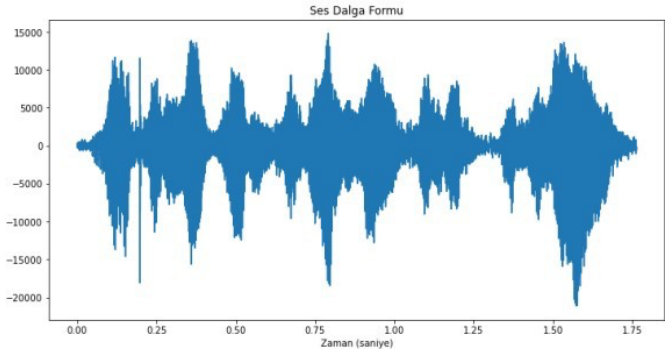
Sesten metne dönüştürme modelinin eğitilmesi ve doğrulanması amacıyla farklı kaynaklardan ses kayıtları toplanmıştır. Bu ses kayıtları toplanırken 4 farklı aksanlardan örnekler alınmıştır: Karadeniz, Ege, Doğu ve İç Anadolu aksanları. Dijital aksanlı konuşmalar Youtube videoları, Spotify popcast'leri, TRT Belgesel kayıtları gibi aksanlı konuşmacıların olduğu platformalardan toplanmıştır. Toplanan uzun süreli dijital ses kayıtları Pyhton kütüphaneleri kullanılarak daha küçük parçalara bölünmüştür. Sessizlik eşiği (dB cinsinden) 30 dB' nin altına inen ve minimum kayıt süresi (milisaniye cinsinden) 150 milisaniye olan "*wav*" dosyaları oluşturulmuştur. Bu işlemler için Pyhton' ın "*from pydub import AudioSegment*" ve "*from pydub.silence import split_on_silence*" kütüphanelerinden yararlanılmıştır. Bu şekilde oluşturulan bir kayıttaki söylenen sözcüklerin, konuşmacının sesinde es verene kadar geçen en az sayıdaki kelimelerden oluşması sağlanmıştır. Böylelikle dijital verilerde söylenirken yarıda kesilmiş sözcükler bulunmamaktadır.

Çeşitli platformlardan toplanan dijital ses kayıtlarından oluşturulan tüm aksanların toplamında 2.440 adet "*wav*" uzantılı ses verisi oluşturulmuştur.

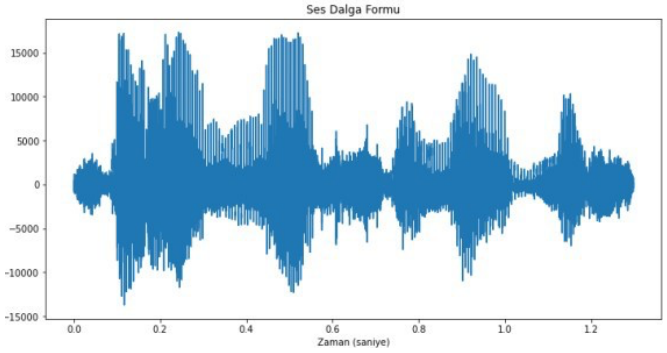
Her aksandaki ses sinyallerine ait rastgele seçilen birer örnek Karadeniz aksanı Şekil 1'de, Ege aksanı Şekil 2'de, İç Anadolu aksanı Şekil 3'de ve Doğu aksanı Şekil 4'de sunulmuştur.



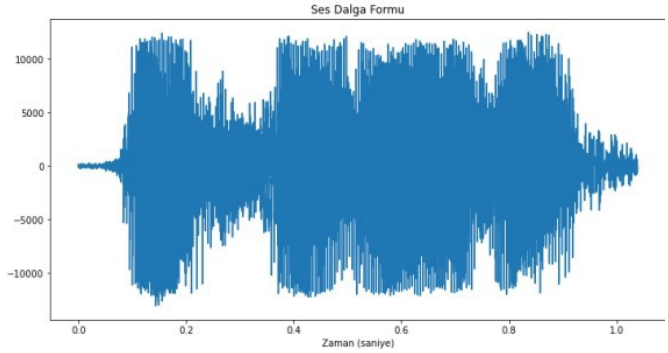
Şekil 1. Karadeniz aksanı ses sinyali



Şekil 2. Ege aksanı ses sinyali



Şekil 3. İç Anadolu aksanı ses sinyali



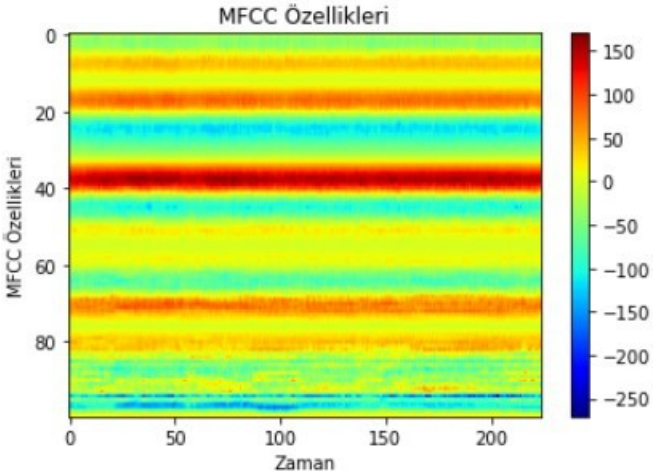
Şekil 4. Doğu aksanı ses sinyali

3.2 Öznitelik oluşturma

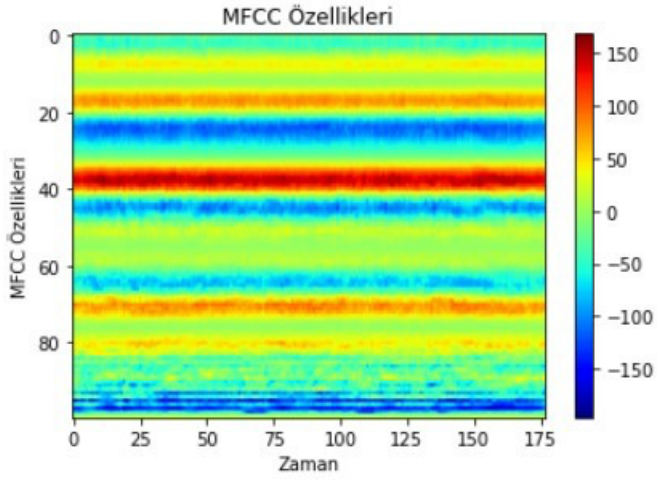
Veri Kümesinin %80' i eğitim verisi, %20' si test verisi olarak kullanılmıştır. Eğitim veri seti boyutu: (1964, 100), test veri seti boyutu: (491, 100). Ses dosyalarının özniteliklerini elde etmek için Mel-frekans kepstrum katsayısı (MFCC) yöntemi kullanılmıştır. Bunun için "*python_speech_features*" kütüphanesinden faydalanılmıştır. Ses sinyallerinin özniteliklerini çıkarmak için MFCC yöntemi çok yaygın olarak kullanılmaktadır. MFCC yöntemi ilk olarak 1980 yılında Davis ve Mermelstein tarafından kullanılmıştır [7]. MFCC algoritması, Hamming penceresi yardımıyla bir ses akışını daha küçük pencerelere yeniden şekillendirerek çerçeveler halinde bölmektedir [8]. Spektrum, Fast Fourier Dönüşümü kullanılarak her çerçeve için üretilir ve her spektrum filtre bankası kullanılarak ağırlıklandırılır. Son adımda MFFC vektörü, Logaritma ve Ayrık Kosinüs Dönüşümü kullanılarak hesaplanır [8].

MFCC parametreleri değiştirilerek model sonunda en iyi doğruluğa ulaşılmaya çalışılmıştır. MFCC yöntemi "*mfcc(y, samplerate=sr, numcep=100, nfft=1024, winlen=0.01, nfilt=100)*" parametreleri ile birlikte kullanılmıştır. Üretilen MFCC katsayılarının sayısı (numcep) 100 alınmıştır. Bu durumda elde edilen matris 100 adet sütuna sahip olmaktadır. Ayrıca dosya ismi yazan bir kolon ve aksanın hangisi olduğunu söyleyen bir etiketle toplam 102 sütuna sahiptir. Ses sinyali N örnekten oluşuyorsa ve pencere uzunluğu winlen * sr örnek içeriyorsa, matriste yaklaşık olarak N / (winlen * sr) satır bulunmaktadır [9]. Yani burada MFCC'ler, her bir ses segmentinin spektral özelliklerini temsil eden 100 boyutlu bir vektördür.

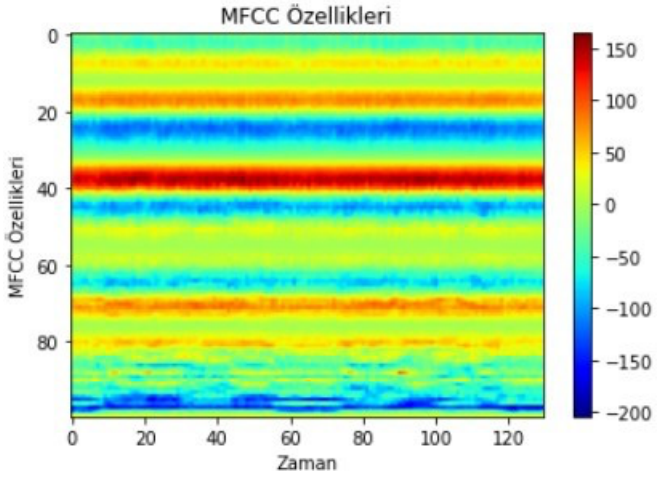
Örnek birer aksanlara ait seslerin çıkarılan MFCC özniteliği görseli Şekil 5, Şekil 6, Şekil 7 ve Şekil 8'de gösterilmektedir. Şekiller, MFCC özelliklerinin zaman içindeki değişimini görselleştirir. Haritada her bir sütun bir zaman dilimine karşılık gelirken, her bir satır belirli bir MFCC katsayısını temsil eder. Renkler ise belirli bir zaman diliminde ve frekans bandında enerji yoğunluğunu gösterir; sıcak renkler (kırmızı, sarı) yüksek enerji yoğunluğunu, soğuk renkler (mavi, yeşil) düşük enerji yoğunluğunu temsil eder. Bu tür bir görselleştirme, sesin frekans bileşenlerinin zamanla nasıl değiştiğini anlamaya yardımcı olur.



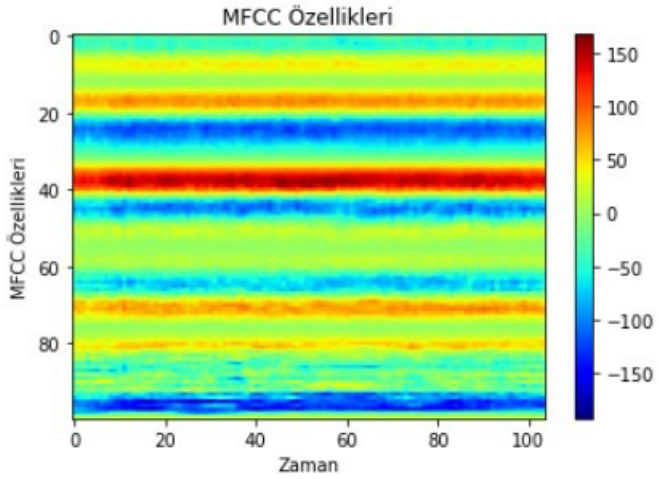
Şekil 5. Örnek bir Karadeniz aksanı sesine ait MFCC özniteliğinin görselleştirilmesi



Şekil 6. Örnek bir Ege aksanı sesine ait MFCC özniteliğinin görselleştirilmesi



Şekil 7. Örnek bir İç Anadolu aksanı sesine ait MFCC özniteliğinin görselleştirilmesi



Şekil 8. Örnek bir Doğu aksanı sesine ait MFCC özniteliğinin görselleştirilmesi

MFCC fonksiyonun dışında "*python_speech_features*" kütüphanesine ait fbank() ve logfbank() fonksiyonları ile de öznitelik çıkarımı yapılarak ilgili modeller de denenmiştir.

3.3 Yöntem

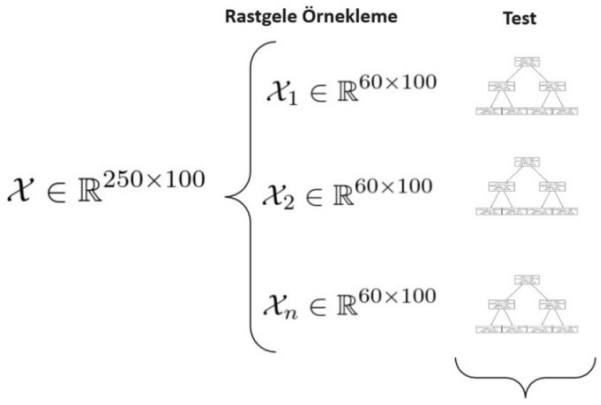
Öznitelik çıkarımı ile RandomForestClassifier, K-Nearest Neighbors (KNN) ve Support Vector Machine (SVM) modelleri oluşturulmuştur. Modelleri ayrı ayrı incelenmiş ve karşılaştırılmış, en iyi doğruluğa ulaşılmaya çalışılmıştır.

3.3.1 Random Forest (RF)

Random Forest, sınıflandırma ve regresyon problemleri için kullanılan bir öğrenme algoritmasıdır. Her bir ağaç bağımsız olarak eğitilir. Karar verme aşamasında her ağaçtan gelen sonuçlar karşılaştırılarak karar verilmektedir. İyi tahminler oluşturulmasını sağlayan güçlü hiperparametreleri bulunmaktadır. RF, yüksek öneme sahip özelliklerin belirlenmesinde kolaylık sağlamaktadır.

Her bir karar ağacı, eğitim verilerinden rastgele bir örneklem ile eğitilir. Ayrıca, her düğümde, sadece rastgele seçilmiş bir alt öznitelik kümesi üzerinde bölünme işlemi gerçekleştirilir [10]. Bu rastgeleleştirme, ağaçlar arasında çeşitliliği artırır ve overfitting (aşırı öğrenme) riskini azaltır.

Random Forest (RF) modelinin eğitilmesinde, eğitim veri kümesi (bu durumda, 250 satır ve 100 sütun) n kez değiştirilerek rastgele örneklenir. Sonra, her örnekte bir karar ağacı eğitilir. Son olarak, tahmin için, tüm n ağacın sonuçları nihai bir karar üretmek üzere toplanır.



Şekil 9. Random Forest (RF) modelinin eğitilmesinin gösterimi.

Bu çalışmada, RF modeli test verisi üzerinde değerlendirerek doğruluk, precision, recall, F1-skoru gibi performans ölçütleri hesaplanmıştır.

Hiperparametre optimizasyonu yapılarak modelin performansı arttırılmış en iyi doğruluk aranmıştır.

"RandomForestClassifier (n_estimators=1000, random_state=42, max_depth=17)" kodu ile 1000 adet karar ağacı oluşturulmuştur (n_estimators=1000). Her bir karar ağacının maksimum derinliğini 17 yaparak (max_depth=17) model çalıştırılmıştır. Ayrıca karar ağacının maksimum derinliğinin belirlenmesiyle overfitting yapması da engellenmiştir.

3.3.2 K-Nearest Neighbors (KNN) modeli

K-Nearest Neighbors (KNN), basit fakat etkili bir denetimli makine öğrenmesi algoritmasıdır. KNN algoritması, bir veri noktasının sınıfını belirlemek veya değerini tahmin etmek için, bu noktanın etrafındaki en yakın k komşusuna bakar. k, kullanıcı tarafından belirlenen bir sabittir ve modelin performansını doğrudan etkiler [11]. Bu komşular arasındaki mesafeleri ölçer ve bu en çok kullanılan mesafe ölçeği Öklit mesafesidir. KNN, en yakın olanlara göre sınıflandırma veya tahmin yapar [12].

KNN algoritması, tüm eğitim veri setini hafızasında tutar. Eğitim sırasında model eğitimi yapılmaz. Komşuların etiketlerinin ortalaması alınarak tahmin yapılır.

Yeni karşılaşılan bir örnek, eğitim setinde yer alan örnekler ile arasındaki benzerliğe göre sınıflandırılmaktadır. K-NN algoritmasında, eğitim setinde yer alan örnekler n boyutlu sayısal nitelikler ile belirtilir. Her örnek n boyutlu uzayda bir noktayı temsil edecek biçimde tüm eğitim örnekleri n boyutlu bir örnek uzayında tutulur. Bilinmeyen bir örnek ile karşılaşıldığında, eğitim setinden ilgili örneğe en yakın k tane örnek belirlenerek yeni örneğin sınıf etiketi, k en yakın komşusunun sınıf etiketlerinin çoğunluk oylamasına göre atanır [12].

Bu çalışmada kullanılan KNN modelinde, hiperparametre optimizasyonu yapılarak modelin performansı arttırılmış en iyi doğruluk aranmıştır. "*KNeighborsClassifier (n_neighbors=3, weights='distance'*)" kodunda en yakın komşu sayısı 3 olarak belirlenmiştir (n_neighbors=3). Model, yeni bir veri noktasını sınıflandırırken bu 3 komşunun hangi sınıfa ait olduğuna bakmaktadır ve en yaygın olan sınıfı tahmin etmektedir. weights='distance' parametresi ile model daha yakın olan komşulara daha çok ağırlık vermektedir.

3.3.3 Support Vector Machine (SVM) modeli

Support Vector Machine (SVM), veri noktalarını farklı sınıflara ayıran en iyi karar sınırını bulmayı amaçlar. İki veri sınıfını en iyi şekilde ayırt etmek için kullanılır.

SVM'nin amacı, veriyi en iyi şekilde ayıran hiperdüzlemi bulmaktır. Bu hiperdüzlem, iki sınıf arasındaki marjini maksimize eder. Marjin, sınıfların en yakın veri noktaları (destek vektörleri) ile hiperdüzlem arasındaki mesafeyi ifade etmektedir.

Model de hiperparametre optimizasyonu yapılarak modelin performansı arttırılmış en iyi doğruluk aranmıştır. "*SVC(C=10, random_state=42)*" kodunda 'C' değeri düştükçe modelin hata yapma toleransı artmakta, değer yükseldikçe hata yapma toleransı azalmaktadır. C(Cost), hata noktalarının ne kadar ihmal edilebileceğini kontrol eden bir parametredir. 'C' değerini yüksek tutmak modelin overfitting yapmasına, daha düşük tutmakta modelin hata yapmasına neden olabilmektedir.

Epsilon (ε), regresyon hatasının kabul edilebilir bir sınırdan ne kadar sapma göstermesine izin verileceğini kontrol etmektedir. Epsilon değeri büyüdükçe, model regresyon hatasına daha fazla tolerans göstermektedir [13].

3.4 Performans Değerlendirme

Support Vector Machine (SVM) modeli yüksek boyutlu, ayrılabilir ve karmaşık veri setlerinde iyi sonuçlar vermektedir. Bu projede de bu modelin en iyi sonuç verdiği görülmüştür. K-Nearest Neighbors (KNN) modelinin %95,92 olması da veri setinin homojen olduğunu göstermektedir.

4 Sonuçlar

Bu çalışmada en iyi doğruluğa ulaşmak için öznitelik çıkartma aşamasında MFCC fonksiyonun dışında "*python_speech_features*" kütüphanesine ait fbank() ve logfbank() fonksiyonları ile de, ilgili modeller de denenmiş ancak doğruluk oranında bir değişim gözlemlenmemiştir.

Öznitelik çıkarımı yapıldıktan sonra üç farklı model kullanılmıştır. İlk kullanılan Random Forest (RF) modelinde, model doğruluğu %91,44 olmuştur.

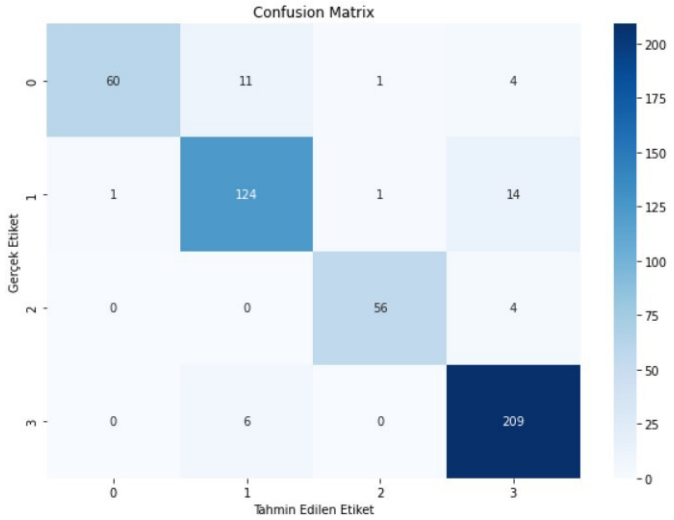
Modelin test setindeki performansı Şekil 10' da gösterilmiştir.

Satır 0: 60 doğru tahmin, 16 yanlış tahmin.

Satır 1: 124 doğru tahmin, 16 yanlış tahmin.

Satır 2: 56 doğru tahmin, 4 yanlış tahmin.

Satır 3: 209 doğru tahmin, 6 yanlış tahmin.



Şekil 10. Random Forest (RF) modeli için confusion (karışıklık) matrisi

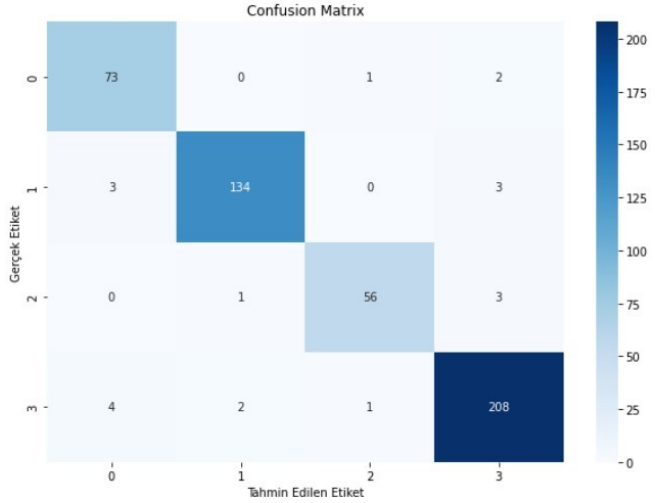
K-Nearest Neighbors (KNN) algoritmasında, model doğruluğu % 95,92 olmuştur. Modelin test setindeki performansı Şekil 11' da gösterilmiştir.

Satır 0: 73 doğru tahmin, 3 yanlış tahmin.

Satır 1: 134 doğru tahmin, 6 yanlış tahmin.

Satır 2: 56 doğru tahmin, 4 yanlış tahmin.

Satır 3: 208 doğru tahmin, 7 yanlış tahmin.



Şekil 11. K-Nearest Neighbors (KNN) modeli için confusion (karışıklık) matrisi

Support Vector Machine (SVM) modelinin paraametrelerinde farklı olasılıklar denenerek en iyi model doğruluğu % 96,74 olmuştur.

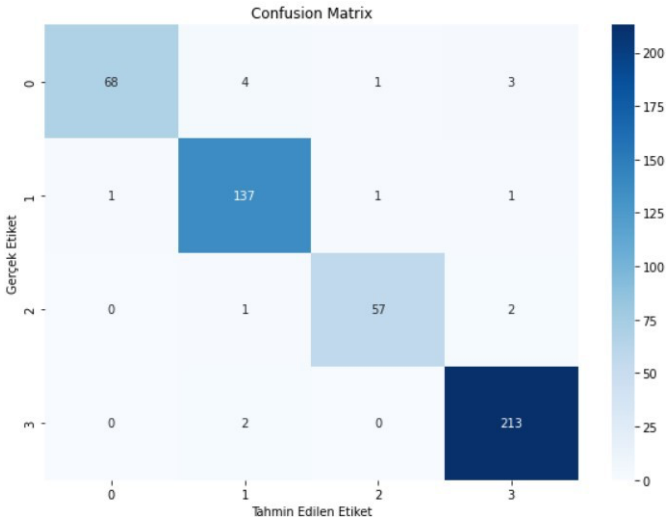
Modelin test setindeki performansı Şekil 12' da gösterilmiştir.

Satır 0: 68 doğru tahmin, 8 yanlış tahmin.

Satır 1: 137 doğru tahmin, 3 yanlış tahmin.

Satır 2: 57 doğru tahmin, 3 yanlış tahmin.

Satır 3: 213 doğru tahmin, 2 yanlış tahmin.



Şekil 12. Support Vector Machine (SVM) modeli için confusion (karışıklık) matrisi

Support Vector Machine (SVM) modeli %96,74 ile en iyi model doğruluğunu vermiştir. K-Nearest Neighbors (KNN) modeli %95,92 ile ikinci en iyi model doğruluğuna sahip olup % 91,44 ile Random Forest modeli diğer iki modelin gerisinde kalmıştır.

Support Vector Machine (SVM) modeli veri setinin sınıflandırmasını yapabilen en iyi model olmuştur. Homojenliğiyle K-Nearest Neighbors (KNN) modeli de iyi bir performans vermiştir. Veri setinin özellikleri nedeniyle Random Forest en kötü sonucu vermiştir.

Kaynaklar

[1] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 357-366, Aug. 1980, doi: 10.1109/TASSP.1980.1163420.

[2] J. W. Picone, "Signal modeling techniques in speech recognition," Proceedings of the IEEE, vol. 81, no. 9, pp. 1215-1247, Sep. 1993, doi: 10.1109/5.237532.

[3] C. Wang, Y. Tang, X. Ma, A. Wu, D. Okhonko, and J. Pino, "fairseq s2t: Fast speech-to-text modeling with fairseq," arXiv preprint arXiv:2010.05171, 2020.

[4] S. Indurthi, H. Han, N. K. Lakumarapu, B. Lee, I. Chung, S. Kim, and C. Kim, "End-to-end speech-to-text translation with modality agnostic meta-learning," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7904-7908, 2020.

[5] X. Li, C. Wang, Y. Tang, C. Tran, Y. Tang, J. Pino, A. Baevski, A. Conneau, and M. Auli, "Multilingual speech translation with efficient finetuning of pretrained models," arXiv preprint arXiv:2010.12829, 2020.
[6] K. Darwish, M. Al-Badrashiny, M. Diab, and J. Niehues, "Evaluating Multilingual Speech Translation under Realistic Conditions with Resegmentation and Terminology," in Proceedings of the 20th International Conference on Spoken Language Translation (IWSLT 2023), E. Salesky, M. Federico, and M. Carpuat, Eds., Association for Computational Linguistics, 2023, pp. 62–78, doi: 10.18653/v1/2023.iwslt-1.2.
[7] E. A. Sağbaş, "MFCC Öznitelikleri ve Adaboost Topluluk Öğrenme Yöntemi Kullanılarak Uyku Seslerinin Sınıflandırılması," DergiPark, 2023.
[8] M. Yıldırım, "MFCC Yöntemi ve Önerilen Derin Model ile Çevresel Seslerin Otomatik Olarak Sınıflandırılması," DergiPark, 2022.
[9] J. Martinez, H. Perez, E. Escamilla, and M. M. Suzuki, "Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques," in Conielecomp 2012, 22nd International conference on electrical communications and computers, Feb. 2012, pp. 248-251, doi: 10.1109/CONIELECOMP.2012.6189912.
[10] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001.
[11] E. Gezer, "OVY, SVM, KNN ve TDN Sınıflayıcıları Kullanarak Konuşmacı Belirleme," Kırşehir Ahi Evran Üniversitesi, 2023.
[12] E. Taşçı and A. Onan, "K-En Yakın Komşu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi," 2023.
[13] C. Lisu, S. Yuanyuan, and W. Ke, "Rapid diagnosis of nitrogen nutrition status in rice based on static scanning and extraction of leaf and sheath characteristics," Int. J. Agric. Biol. Eng., vol. 10, no. 3, pp. 158–164, May 2017, doi: 10.25165/IJABE.V10I3.1860.

Optimizing Energy Efficiency in Wood Processing Plants through Data Analysis: A Case Study on Value-Added Wood Products Manufacturing

Melike Nur İNCE[1,*], Çağatay TAŞDEMİR[2,3,*]

[1]*Department of Industrial Engineering (Grad. School), Bursa Technical University, Bursa 16310, Türkiye; 22435025008@ogrenci.btu.edu.tr*
[2]*Department of Industrial Engineering (Grad. School), Bursa Technical University, Bursa 16310, Türkiye; cagatay.tasdemir@btu.edu.tr*
[3]*Department of Forest Industry Engineering, Bursa Technical University, Bursa, Türkiye*
[*]*Correspondence: Bursa Technical University, Bursa, Türkiye, 22435025008@ogrenci.btu.edu.tr and cagatay.tasdemir@btu.edu.tr*

Abstract

Energy consumption in value-added wood products manufacturing facilities has significant environmental and economic impacts. High energy usage increases costs and expands the carbon footprint, making it challenging to achieve sustainability goals. Inefficient energy management in wood processing plants elevates operational costs and exacerbates the environmental burden. Therefore, optimizing energy efficiency through data analysis techniques is critically important. This study analyzes energy consumption data to identify inefficiencies and propose effective optimization strategies. Historical data encompassing operational parameters, energy consumption, environmental conditions, and production output from five high-capacity wood processing machines in the wood products industry were collected daily over the past three years. The dataset includes ten categories: Date, Machine ID, Runtime Hours, Load Percentage, Electricity Usage, Gas Usage, Temperature, Humidity, Production Output, and Energy Efficiency. Initially, the data were loaded into a pandas Data Frame, converted to date and time format, and checked for missing and outlier values, followed by standardization of numerical features. Descriptive statistics were calculated for each feature, and data consistency was verified. The distributions of key features were visualized with histograms, and the relationships between numerical features were illustrated using a correlation matrix heatmap. Trends and seasonal patterns in energy consumption and production output were analyzed by resampling the data monthly. Principal Component Analysis (PCA) was applied to reduce the dimensionality of the dataset while retaining significant information, and three clusters were formed using the K-Means algorithm. The clusters were visualized in the PCA-reduced feature space, and their characteristics were analyzed to prioritize machines for energy efficiency improvements. Cluster 2, characterized by an average energy usage of 209.79 kWh, an average gas usage of 107.90 m^3, an average production output of 1595.03 units, an average energy efficiency of 5.26 units/kWh, and an average load percentage of 75.35%, demonstrated low energy consumption and high production output, indicating highly efficient operations. Therefore, it is recommended that the best practices from this cluster be standardized and implemented across other clusters. Additionally, investing in technological advancements to further enhance energy efficiency and conducting continuous improvement efforts to maintain and improve efficiency are suggested.

Keywords: *Energy Efficiency, Wood Products Industry, K-Means Clustering, Principal Component Analysis (PCA)*

1 Introduction

Furniture is a product that has existed throughout history and has become an integral part of our daily lives, with a continued presence in the future. As a consequence of the growth in production activities, the size of the furniture market has continued to expand from the past to the present. Several studies have been conducted to understand the market size. In one such study, the size of the global furniture market is estimated to be 579 billion US dollars in 2023. This figure is expected to increase annually, reaching approximately 873 billion US dollars by 2030 [1]. In a separate investigation, the global furniture market was evaluated at USD 516.66 billion in 2022. It is projected to experience growth from USD 541.52 billion in 2023 to USD 780.43 billion by 2030, exhibiting a compound annual growth rate (CAGR) of 5.36% over the projected period [2].

The wood furniture industry represents a significant contributor to the global growth potential of the furniture industry as a whole. Recent studies indicate that the market value of wood furniture in 2023 is estimated at USD 569.9 billion. It is projected to reach approximately USD 877.7 billion by 2032, representing a compound annual growth rate (CAGR) of 5% [3]. In a separate study, the global market for wood furniture reached a valuation of USD 275.7 billion in 2023 and is projected to reach USD 413.5 billion by 2032. These forecasts indicate a compound annual growth rate (CAGR) of 4.47% between 2024 and 2032 [4]. The two studies concur that the wood furniture sector is experiencing growth. This growth will consequently result in increased energy demand for the industry, necessitating the utilization of more energy for the machinery and equipment employed.

In this context, the issue of energy efficiency is becoming increasingly significant, particularly in light of the rising energy costs. In the long term, the rise in energy costs associated with production has rendered energy efficiency a necessity and a competitive factor in global markets. [5]. In light of the fact that the manufacturing sector is responsible for 23% of global energy consumption, the combination of heightened environmental awareness, rising energy costs, and increasingly stringent regulations is compelling organizations to pursue not only operational excellence but also the effective management of energy efficiency in their production systems [6]. In consequence, enterprises are implementing a variety of energy efficiency measures (EEMs), including process technology-centric improvements (e.g., optimization of process parameters) and investment-oriented improvements (e.g., selection of energy-efficient machinery) [6]. A substantial body of research on energy efficiency has been conducted in the literature. In these studies, a variety of methods have been employed, including simulation-based approaches [7], data analysis studies [8, 9], production planning and hybrid simulation [5], and lean management [6].

Data analysis examines, cleans, transforms, and models data to achieve a specific goal or solve a problem. In this process, the raw data is transformed into meaningful information, thereby providing a basis for future decisions [10]. The objective of data analysis is to identify patterns, trends, and relationships within data sets through the application of diverse techniques and tools. Data analysis is widely utilized across numerous fields, including finance, marketing, health, and engineering. Businesses, organizations, and individuals can enhance their decision-making processes by employing data analysis, making them more informed and strategic [11].

2 Literature Review

Wold et al. (1987) make significant contributions to the literature on principal component analysis (PCA), focusing on issues such as the geometric interpretation of PCA, variable selection, correspondence analysis, and modeling power. PCA serves as a fundamental technique in multivariate data analysis. The article emphasizes that PCA is a versatile tool that can facilitate understanding complex datasets. However, it also cautions that the method should be applied with caution. This paper comprehensively examines PCA's theoretical underpinnings and practical applications [12].

Ghiasi et al. (2002) presented a comprehensive theoretical framework for energy-aware clustering in sensor networks in their study. Their work addressed the energy optimization and balanced k-clustering problem. The authors develop algorithms that optimize the load distribution among the master nodes to improve energy efficiency and demonstrate the effectiveness of these algorithms through preliminary experiments. Furthermore, they make significant contributions to the existing literature, indicating potential avenues for future research that can be adapted to the dynamic nature of sensor networks [13].

Abdi and Williams (2010) significantly contribute to the literature on PCA by comprehensively understanding the subject matter. They concentrate on issues such as calculating observation contributions in PCA, the function of eigenvalues and factor scores, the interpretation of components, and data visualization. Furthermore, it enhances PCA's theoretical and practical aspects by addressing practical applications in the social sciences and generalizations, such as including qualitative variables [14].

In their study, Bro and Smilde (2014) make significant contributions to the field of principal component analysis (PCA) and emphasize the applicability of PCA, particularly in areas such as chemometrics. By providing a comprehensive description of PCA, they offer researchers guidance on critical issues such as component selection, outlier detection, and model validation. By emphasizing PCA's versatility and pragmatic utility in discovery studies, the article offers valuable insights into the existing literature from theoretical and applied viewpoints [15].

Benedetti et al. (2016) significantly contribute to the existing literature on industrial energy efficiency. The article underscored the significant role of compressed air systems in industrial energy consumption and the deficiencies in the prevailing measurement methodologies. In particular, the study identified shortcomings in current practices, including the low proportion of facilities measuring energy consumption and the fact that current benchmarks do not reflect actual operating conditions. The paper establishes a foundation for future research to enhance industrial energy management. It illustrates the necessity for enhanced energy metering systems and the establishment of dependable benchmarks [9].

Zhou et al. (2017) make a notable contribution to the existing energy efficiency and sustainability literature by presenting innovative methodologies. The study presents findings that are supported by empirical data and have reached a wider audience through the use of open access. The study underscored the potential of sustainable energy solutions to inform and enhance energy policies and bolster energy efficiency. Furthermore, it offers policy recommendations to facilitate the adoption of energy technologies and highlights potential avenues for future research. The study makes a valuable contribution to the field of energy studies by promoting collaboration and knowledge sharing, which are essential for overcoming the challenges currently facing the energy sector [16].

In their study, Chen and Xiong (2017) contributed to the district heating system adaptation to climate change and building renovation scenarios. The study assesses the practicality of the function utilized for heat demand estimation and presents application and scenario-based analyses with real-world data. The findings elucidate the impacts of diverse climate and renovation scenarios on heat demand, underscoring the significance of parameter tuning to enhance the precision of predictions. This indicates the necessity for precise forecasting methodologies to administer energy resources effectively [17].

Sobottka et al. (2017) made a notable impact on the field of production planning and control (PPC) by introducing a novel hybrid simulation-based planning tool, which can potentially enhance energy efficiency in manufacturing processes. The tool optimizes energy use by integrating material flow and equipment behavior with multi-criteria optimization, supporting sustainable manufacturing practices. Furthermore, incorporating the genetic algorithm (GA) into the optimization process has facilitated the simultaneous attainment of multiple production objectives and the formulation of constraints that reduce the search space. The paper enhances these methods' applicability to various production environments while proposing new avenues for future research. These contributions provide a foundation for future developments in energy-efficient production planning and a crucial basis for the sustainability of modern manufacturing industries [7].

Porizka et al. (2018) offer significant insights into the application of principal component analysis (PCA) in the processing of laser-induced breakdown spectroscopy (LIBS) data. In light of the growing popularity of PCA in LIBS, the authors provide practical guidelines for its use in data visualization, dimensionality reduction, and modeling. In particular, the paper identifies the challenges associated with implementing data preprocessing and MVDA algorithms. It underscores the necessity to develop more comprehensive data libraries for the LIBS community. Furthermore, the paper asserts that PCA can address numerous analytical requirements and, as a result, should be considered as a preliminary step before transitioning to more intricate algorithms. In conclusion, the paper offers methodological contributions to the existing literature and provides detailed application

guidelines by exploring the role of PCA in LIBS in depth [18].

In their study, Bonfa et al. (2019) made significant contributions to the advancement of energy efficiency in district heating systems. The paper put forth a novel forecasting methodology based on the correlation between outdoor temperature and heat demand and demonstrated its application to a range of buildings in Lisbon. The findings indicated that retrofit scenarios impact the accuracy of the predictions and that alterations in the slope of the heat demand function are crucial for long-term planning. This study underscored the significance of more precise and responsive forecasts in energy planning and strategies to mitigate greenhouse gas emissions [8].

In their study, Cebeci (2020) introduces the R package 'fcvalid,' which significantly contributes to the field of fuzzy and probabilistic clustering analysis. The package offers a comprehensive tool for validating clustering results and provides a user-friendly experience with generalized internal validity indices. Furthermore, the package specializes in fuzzy clustering needs, allowing data points to belong to more than one cluster. The graphical review and performance evaluation functions of 'Fcvalid' represent a valuable tool for researchers interested in improving clustering methodologies, and the package makes an essential contribution to the existing literature on the subject [19].

The study by Clayman et al. (2020) represents a significant advancement in bioinformatics and gene expression analysis. The study enhanced the efficacy of clustering in highly variable datasets by integrating principal component analysis (PCA) and K-means clustering methods to examine 978 landmark genes within the L1000 dataset. The utilization of landmark genes resulted in the formation of more discrete clusters and enhanced clustering efficiency compared to randomly selected genes. Moreover, the paper offers valuable insights for future research on the potential clinical applications of these genes, underscoring the necessity for a more comprehensive investigation into the interrelationships between genetic and clinical variables [20].

In a recent study, Jansson et al. (2022) have made significant contributions to geochemical exploration techniques, focusing on Zn skarn deposits and industrial carbonates in the Sala region of Sweden. The integrated application of PCA and K-means clustering techniques enabled an objective classification of dolomite samples while facilitating the comprehension of the spatial distribution of mineralization through geochemical signatures. The findings demonstrate that data-driven approaches yield strong concordance with geological correlations, thereby enabling the identification of dolomite types with economically significant potential for mining. This study makes a valuable contribution to the existing literature on geological investigations by emphasizing the significance and practical implications of data-driven classification methods [21].

In a recent study, Ilu et al. (2022) proposed a novel methodology for forecasting the spread of the novel coronavirus (2019-nCoV), also known as SARS-CoV-2, which has caused a global pandemic. By integrating the LSTM algorithm with PCA and K-means clustering, the study demonstrated remarkable proficiency in nonlinear data processing and outlier detection. Compared to other machine learning algorithms, it exhibited the superior performance of LSTM and achieved high prediction accuracy on large datasets. The findings offer significant guidance for enhancing public health predictions and further investigating the applications of machine learning in epidemiological modeling [22].

Liao et al. (2024) proposed a two-layer optimization method for DPV and ESS based on IDEC-K clustering to reduce voltage violations and network losses in distribution networks. The results demonstrate that this method effectively mitigates overvoltage risk and reduces line losses by narrowing the voltage range and stabilizing fluctuations. This paper makes a significant contribution to the existing literature on energy management by emphasizing the crucial role of advanced optimization techniques in this field [23].

The study by Eid et al. (2024) on the determination of recharge and salinity origins in Siwa Oasis, Egypt, offers significant insights that contribute to the existing literature on water resources management. By integrating stable isotopes, mixing models, and K-mean cluster analysis, the study provides a detailed analysis of the types of water resources and the dynamics of increasing salinity in the region. This comprehensive approach elucidates how the salinity of the TCA aquifer has increased over time and the extent to which various factors have contributed to this phenomenon. It also furnishes recommendations for water management strategies. Moreover, the study offers a crucial reference point for future research endeavors to address analogous hydrological challenges in arid regions [24].

In a recent contribution to the field, Dugan et al. (2024) provide valuable insights into assessing foot deformities in children with cerebral palsy (CP). The study employs a multivariate functional principal component analysis (MFPCA) in conjunction with a k-means clustering method to facilitate a more detailed and dynamic classification of foot types. The findings underscore the necessity of developing personalized treatment plans and the importance of dynamic assessment methods. The efficacy of MFPCA in clinical trials signifies a substantial advancement in the management of foot deformities and offers a robust foundation for future research endeavors [25].

A literature review reveals that PCA and the K-Means algorithm have been employed in numerous studies on energy efficiency. However, there is a lack of research investigating the application of these methods to enhance energy efficiency in the wood furniture industry. In this context, the objective of this study is to utilize PCA and the K-Means algorithm to enhance the energy efficiency of machinery used in facilities that produce value-added wood products. High energy consumption presents a significant challenge to achieving sustainability goals, resulting in increased costs and a larger carbon footprint. Furthermore, inefficient energy management contributes to elevated operational costs and an increased environmental burden. This study aims to enhance energy efficiency and achieve cost savings by analyzing the energy consumption data of a wood products manufacturing company, identifying existing inefficiencies, and proposing effective optimization strategies.

3 Materials and Methods

EcoWoodProduct is a medium-sized enterprise specializing in producing value-added wood products, with operations spanning domestic and international markets. Additionally, EcoWoodProduct is located in Türkiye and is renowned for its superior quality and innovative products within the woodworking industry. The company employs modern production technologies and an environmentally conscious approach to increase energy efficiency and optimize production processes per its sustainability goals.

The study is comprised of four phases. The data collection and preprocessing phase was followed by exploratory data analysis, cluster analysis, and cluster analysis.

In the initial phase of the study, data about operational parameters, energy consumption, environmental conditions, and production output from five high-capacity wood processing machines utilized in the wood processing industry were gathered daily over the past three years. The data set was then divided into ten categories:

- Date: Date of the data record.
- Machine ID: Unique ID for each machine.
- Runtime Hours: Total runtime of the machine (in hours).
- Load Percentage: Load percentage according to the machine's maximum capacity.
- Electricity Usage: Electricity consumption (in kWh).
- Gas Usage: Natural gas consumption (in cubic meters).
- Temperature: Temperature in the plant (in degrees Celsius).
- Humidity: Humidity level in the plant (in percent).
- Production Output: Production output (in units).
- Energy Efficiency: Energy efficiency metric (production output per energy consumed).

Table 1 presents a summary of the statistical characteristics of the data set utilized in the study.

Table 1. Statistical Summary of Operational Data for Woodworking Machinery

Variable	Machine ID	N	N*	Mean	SE Mean	StDev	Min	Median	Max
Runtime Hours	M01	1096	0	16.829	0.123	4.060	10.013	16.927	23.995
	M02	1096	0	17.035	0.122	4.031	10.001	16.819	23.987
	M03	1096	0	17.017	0.121	4.012	10.016	16.766	23.992
	M04	1096	0	17.187	0.122	4.035	10.006	17.333	23.992
	M05	1096	0	17.095	0.118	3.921	10.030	16.915	23.999
Load Percentage	M01	1096	0	75.081	0.429	14.194	50.080	75.317	99.998
	M02	1096	0	74.775	0.446	14.777	50.031	74.792	99.999
	M03	1096	0	74.314	0.431	14.279	50.002	73.819	99.881
	M04	1096	0	74.314	0.437	14.457	50.023	74.226	99.994

	M05	1096	0	74.755	0.433	14.338	50.008	74.500	99.903
	M01	1096	0	300.24	3.46	114.47	100.09	298.85	498.92
Electricity Usage	M02	1096	0	294.27	3.51	116.31	101.01	290.74	499.37
	M03	1096	0	295.89	3.51	116.11	100.05	295.20	499.36
	M04	1096	0	304.17	3.48	115.06	101.03	304.53	499.61
	M05	1096	0	300.10	3.53	116.82	100.82	303.96	499.86
	M01	1096	0	127.08	1.33	44.02	50.16	129.51	199.79
	M02	1096	0	126.19	1.29	42.69	50.03	127.69	199.96
Gas Usage	M03	1096	0	123.30	1.29	42.63	50.40	121.68	200.00
	M04	1096	0	120.96	1.28	42.44	50.27	117.72	199.87
	M05	1096	0	123.96	1.28	42.39	50.01	123.71	199.82
	M01	1096	0	24.959	0.0869	2.875	20.007	25.007	29.980
	M02	1096	0	24.879	0.0887	2.936	20.005	24.746	29.992
Temperature	M03	1096	0	25.029	0.0873	2.891	20.000	25.118	29.990
	M04	1096	0	24.997	0.0892	2.953	20.003	24.977	29.995
	M05	1096	0	24.788	0.0875	2.897	20.004	24.758	29.994
	M01	1096	0	49.892	0.349	11.569	30.013	49.884	69.983
	M02	1096	0	49.581	0.350	11.583	30.024	49.749	69.988
Humidity	M03	1096	0	50.254	0.353	11.680	30.037	50.048	69.987
	M04	1096	0	49.750	0.346	11.454	30.040	49.683	69.990
	M05	1096	0	49.638	0.351	11.616	30.025	49.772	69.926
	M01	1096	0	1250.7	13.3	438.7	501.4	1240.2	1996.9
Production Output	M02	1096	0	1248.5	13.1	434.6	501.2	1250.8	1999.0
	M03	1096	0	1267.3	13.3	441.2	500.1	1264.0	1998.6
	M04	1096	0	1250.1	13.0	431.2	503.0	1269.5	1998.9
	M05	1096	0	1226.1	13.1	432.8	500.3	1217.7	1999.6
	M01	1096	0	3.2114	0.0470	1.5571	0.7965	2.9332	9.8057
Energy Efficiency	M02	1096	0	3.2758	0.0501	1.6596	0.8647	2.9291	10.1497
	M03	1096	0	3.3728	0.0530	1.7536	0.8234	3.0222	11.3280
	M04	1096	0	3.2491	0.0491	1.6249	0.7948	2.9205	11.1761
	M05	1096	0	3.2067	0.0490	1.6238	0.7773	2.9295	11.8756

The data were loaded into a Pandas DataFrame, converted to date and time format, and subjected to quality control checks, including identifying missing values and outliers. Subsequently, the numerical features were standardized.

In the study's second phase, exploratory data analysis was conducted to identify the dataset's characteristics. In this phase, central tendencies and variability for each feature in the entire dataset were calculated with summary statistics (mean, standard deviation, minimum, maximum, etc.). Additionally, data consistency was evaluated by identifying any missing values. The key features were identified, and histogram plots were constructed to illustrate their distribution. Furthermore, correlation matrices were employed to elucidate the interrelationships between the numerical attributes within the data sets. The data were resampled monthly to identify trends and seasonal patterns in energy consumption and production output, and average values were calculated.

In the third stage of the study, a cluster analysis was conducted. In this stage, the data sets were organized by selecting six features for analysis: runtime hours, load percentage, electricity usage, gas usage, temperature, humidity, production output, and energy efficiency. The selected features were standardized using the Standard Scaler. To reduce the dimensionality of the data set and retain the most meaningful information, a principal component analysis (PCA) was conducted, reducing the data set to two principal components. A K-means clustering algorithm with three clusters was then applied to the reduced data set using PCA, and the resulting cluster labels were added to the original data frame. The clusters were then delineated in the feature space that PCA had

reduced to facilitate an examination of the cluster separations.

In the final stage of the analysis, the clusters were examined. In the course of these analyses, the characteristics of each cluster were identified, graphical representations were constructed to illustrate the distribution of salient features within each cluster and areas for operational improvement were identified.

4 Results

A three-year data set comprising ten different data types was collected daily to evaluate the energy efficiency of five high-capacity machines used in the wood processing industry. The dataset was initially loaded into a Pandas DataFrame, and the "Date" column was converted to the date and time format. The dataset was then subjected to a process of cleaning, whereby any missing or abnormal data points were identified and removed. Subsequently, the numerical features were standardized to ensure they were expressed on a consistent scale.

Statistical summaries, including the mean, standard deviation, minimum, and maximum values, were calculated to ascertain each feature's central tendencies and variability within the dataset. Due to these computations, any missing values and inconsistencies in the data were identified and rectified.

A correlation matrix was constructed to analyze the dataset further and understand the data points' relationships. The resulting correlation matrix is presented in Figure 1.

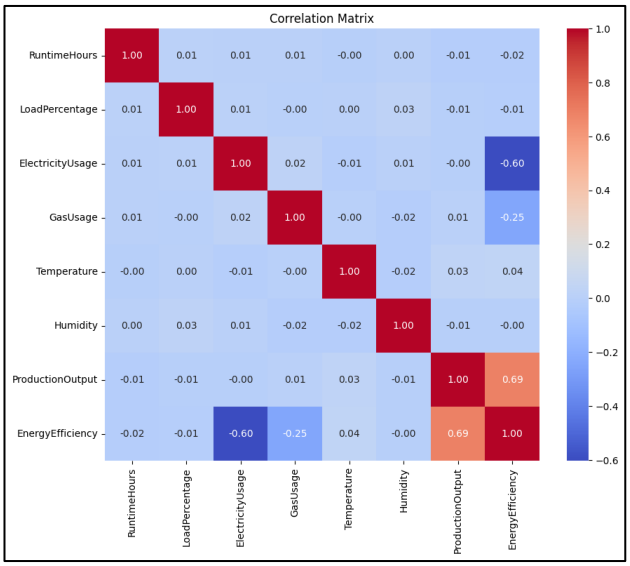


Figure 1. Correlation Matrix

As illustrated in Figure 1, there are notable correlations between the "Energy Efficiency" data and the "Electricity Usage," "Gas Usage," and "Production Output" data. There is a robust positive correlation between energy efficiency and production output. These two variables exhibit a strong correlation, demonstrating a tendency for simultaneous increases and decreases.

Conversely, a robust inverse correlation is observed between "Energy Efficiency" and "Electricity Usage." This strong negative relationship indicates that as one of these two variables increases, the other decreases and this relationship becomes robust. The relationship between "Energy Efficiency" and "Gas Usage" is relatively weak and negative. The weak negative relationship indicates that while one of these variables increases, the other decreases, but this relationship is relatively weak.

Subsequently, histogram plots were constructed to illustrate the distribution of the principal attributes identified by the correlation matrix (i.e., "Energy Efficiency" versus "Electricity Usage," "Gas Usage," and "Production Output"). The graphs mentioned above are presented in Figure 2.

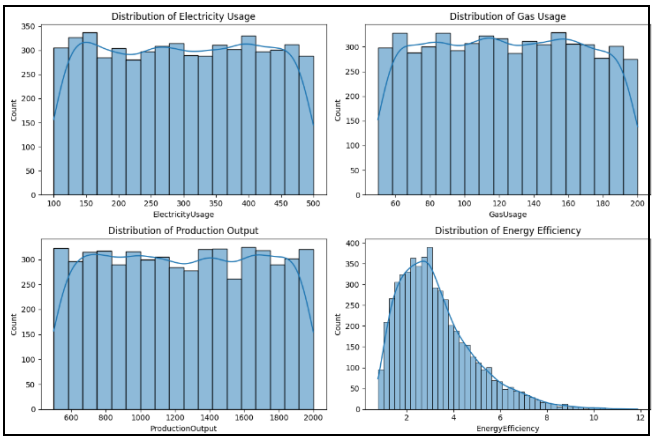


Figure 2. Energy Use and Efficiency Breakdown Graphs

An examination of the histogram plots presented in Figure 2 reveals that the data about "Electricity Usage," "Gas Usage," and "Production Output" are distributed in a manner that is similar to one another. These key attributes are distributed relatively uniformly across a broad range, exhibiting a relatively flat distribution with no discernible concentration.

In contrast to the attributes above, the "Energy Efficiency" distribution exhibited a rightward skew. This suggests that energy efficiency is

predominantly concentrated at lower values, lacking high-efficiency observations. This indicates that, in general, energy efficiency is at relatively low levels, although it is possible to achieve high efficiency in certain instances.

Following the interpretation of the histogram plots of the key attributes, a cluster analysis was conducted on the data set. Before the clustering analysis, all data were standardized using the StandardScaler function. Subsequently, principal component analysis (PCA) was employed to reduce the dimensionality of the dataset while retaining the most salient information. As a consequence of this analysis, the dataset was reduced to two principal components, thus facilitating enhanced visualization and clustering.

The K-Means algorithm was then applied to the data set, which had been reduced in size through PCA, to obtain three clusters. The labels corresponding to the abovementioned clusters were then appended to the data frame. To facilitate a visual examination of the cluster distinctions, the PCA dimensionality reduction is presented in Figure 3.

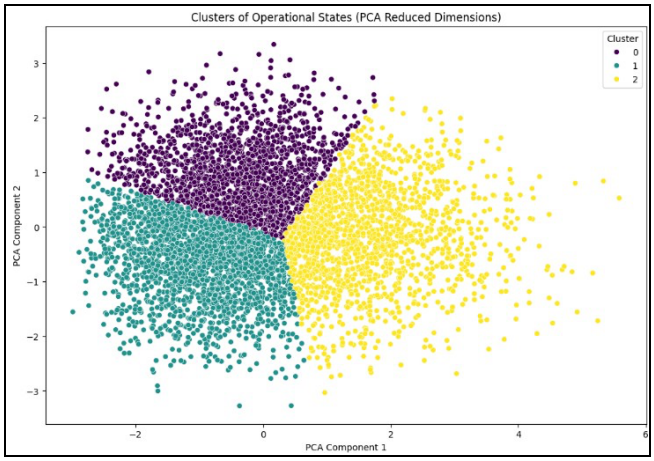


Figure 3. Clusters in Reduced Feature Space with PCA

As illustrated in Figure 3, the principal component analysis (PCA) technique has formed three distinct clusters in the reduced-dimensional feature space. The characteristics and averages of these clusters are as follows:

Cluster 0:

- Electricity Usage: High (average: 337.17 kWh)
- Gas Utilization: High (average: 144.34 m^3)
- Production Output: Medium (average: 1291.58 units)
- Energy Efficiency: Medium (average: 2.74 units/kWh)
- Load Percentage: Medium (average: 68.86%)

Cluster 1:

- Electricity Use: Very High (average:336.99 kWh)
- Gas Utilization: High (average:120.64 m^3)
- Production Output: Low (average:943.63 units)
- Energy Efficiency: Low (average: 2.14 units/kWh)
- Load Percentage: High (average: 78.82%)

Cluster 2:

- Electricity Use: Low (average: 209.79 kWh)
- Gas Utilization: Low (average: 107.90 m^3)
- Production Output: High (average: 1595.03 units)
- Energy Efficiency: High (average: 5.26 units/kWh)
- Load Percentage: Medium (average: 75.38%)

The three clusters identified by PCA in the reduced feature space exhibit disparate energy consumption and efficiency profiles. Cluster 0 indicates a group with elevated energy consumption and moderate efficiency, whereas Cluster 1 is distinguished by markedly elevated energy consumption and low efficiency. Cluster 2 represents a group with low power consumption and high efficiency.

A bar chart is provided for each attribute to facilitate a more effective visual comparison of the average values for the three clusters. The bar chart is presented in Figure 4 for reference. This visualization provides a more detailed representation of the main differences between the clusters and their impact on energy efficiency.

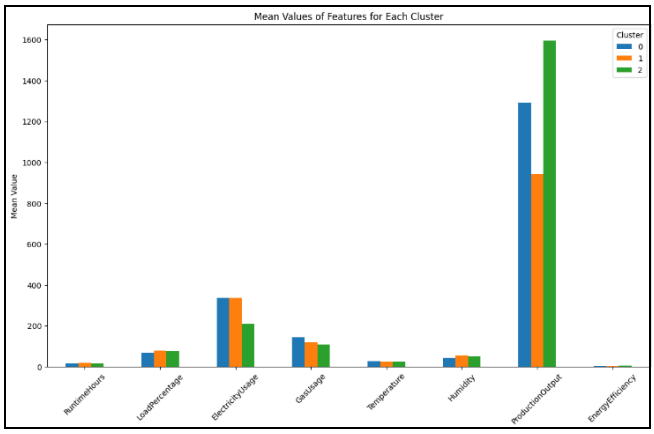


Figure 4. Mean Values of Attributes by Clusters

As illustrated in Figure 4, the most notable distinction between the clusters is the production output. Cluster 2 exhibits a markedly elevated production output compared to the other clusters. This indicates cluster 2 has more efficient production processes or a higher production capacity. The discrepancy in production output between clusters 0 and 1 is less pronounced, with cluster 0 exhibiting a marginal increase in the mean production output.

Concerning electricity consumption, cluster 0 exhibits the highest average, while cluster 2 displays the lowest average. This indicates that Cluster 0 is associated with the utilization of a process or machinery that is highly energy-intensive, whereas Cluster 2 is distinguished by its enhanced energy efficiency. The electricity consumption of Cluster 1 is nearly identical to that of Cluster 0, exhibiting only a slight decrease.

The highest average gas consumption is observed in Cluster 0, followed by Clusters 1 and 2. This indicates that there may be a correlation between gas consumption and production output. It can be postulated that clusters that consume more energy may also have higher gas consumption values.

Concerning load percentage, no significant discrepancies are evident between the clusters, although Cluster 1 and Cluster 0 exhibit slightly elevated averages.

Concerning energy efficiency, no notable discrepancy is observed among the clusters. This indicates that the clusters may exhibit comparable characteristics relating to energy efficiency.

No significant differences were observed between the clusters concerning other parameters, including working hours, temperature, and humidity.

In conclusion, Cluster 2 is distinguished by high production output and low energy consumption, which can be regarded as an advantageous scenario in terms of energy efficiency. Despite its high electricity and gas consumption, Cluster 0 does not have the highest production output, suggesting that the machines or processes within this cluster may be energy-intensive but inefficient. Cluster 1 occupies a mid-range position concerning both energy consumption and production output.

The analysis yielded the development of optimization strategies for each cluster to enhance energy efficiency within the enterprise:

For Cluster 0:

- It is recommended that a detailed energy audit be conducted to identify the underlying causes of the elevated level of energy consumption.
- It is recommended that machine settings and production schedules be optimized to reduce energy consumption.
- It is recommended that preventive maintenance practices be implemented to enhance the efficiency of the machines in question.

For Cluster 1:

- It is recommended that operational processes be redesigned to increase production output while reducing energy consumption.
- It is recommended that advanced energy management systems be implemented to monitor and control real-time energy use.
- It is recommended that employees receive training on energy efficiency and operational strategies.

For Cluster 2:

- It is recommended that the most effective energy efficiency practices currently in use within the cluster be documented and standardized.
- It is recommended that investment be made in technology improvements to increase existing energy efficiency further.
- It is recommended that continuous improvement efforts be undertaken to maintain and enhance existing efficiency.

5 Discussion

The study by Benedetti et al. (2016) on energy efficiency for compressed air systems in large and energy-intensive industrial companies is significantly aligned with the optimization strategies proposed in our study [9]. In particular, the approach of "identifying and standardizing best practices" proposed for Cluster 2 in our study is consistent with the "benchmarking indicators" put forth by Benedetti et al. Furthermore, the strategies of "implementing energy metering systems," "adjusting operational parameters," and "transferring knowledge and identifying best practices across companies and sectors" align with the optimization strategies identified for Cluster 0, Cluster 1, and Cluster 2, respectively.

The study by Bonfa et al. (2018) on energy efficiency for compressed air systems in pharmaceutical manufacturing plants identified similar optimization strategies [8]. Among the strategies identified, two are particularly noteworthy: "specific maintenance interventions related to compressors detected through monitoring" and "energy monitoring and control charts for compressors." However, it is notable that strategies such as "determining the optimal activation sequence" and "developing a centralized control system to manage the system compressors" were not proposed in this study. Conversely, using data analysis techniques to achieve energy efficiency is a common approach in our research, as evidenced by the two studies in the literature.

The findings of this study have several implications for academic, practical, and managerial contexts. From an academic standpoint, the combination of data analysis approaches has been shown to yield noteworthy results in studies. Applying PCA and K-clustering methods for energy efficiency in the wood processing industry has substantially contributed to the existing literature on the subject. From a practical standpoint, cluster-based analyses present a viable avenue for enhancing energy efficiency by focusing on specific processes or machinery within an enterprise. Furthermore, the efficacy of data-driven decision-making in operational processes is underscored. From a managerial perspective, this study offers insights into the optimal allocation and prioritization of resources within enterprises, facilitates the integration of energy efficiency as a core organizational value, and provides a framework for developing strategic plans.

As is the case with any study, this study has limitations. Among the limitations of this study, the limitations of the data set and the extensive data requirements are particularly noteworthy. Moreover, the study's generalizability is constrained at the sectoral or enterprise level.

In future studies, it may be possible to obtain different results by applying the methodology used in this study to other industries. Furthermore, integrating machine learning and artificial intelligence can facilitate more productive and sophisticated outcomes.

6 Conclusion

The wood processing industry places a significant emphasis on energy efficiency. The machinery utilized in the wood processing industry is characterized by high capacity, resulting in considerable energy consumption in proportion to the output produced. Accordingly, this study undertook a comprehensive investigation into optimizing energy efficiency in wood processing facilities. Clustering operational data can identify specific operational situations and characteristics, and targeted optimization strategies can be proposed to improve energy efficiency. This approach can be further refined by integrating additional data and advanced analytical techniques, thereby facilitating continuous monitoring and improvement of the performance of wood processing machines.

The findings of this study can be summarized as follows:

- The dataset utilized in the study revealed that energy efficiency, electricity usage, gas usage, and production output were the most salient features.
- A total of three clusters (Cluster 0, Cluster 1, and Cluster 2) were formed in the reduced feature space using PCA.
- Cluster 2 is distinguished by its highly efficient operations, low energy consumption, and high production output. Employing advanced technologies and prioritizing continuous improvement initiatives is imperative to enhance energy efficiency.
- Cluster 0 indicates inefficient operations with high energy consumption and moderate production output. It is recommended that machine settings and

production schedules be optimized and preventive maintenance plans be established to improve energy efficiency.

- Cluster 1 encompasses inefficient operations characterized by high energy consumption and low production output. To improve the energy efficiency of this cluster, it would be beneficial to implement advanced energy management systems capable of monitoring and controlling real-time energy use. Furthermore, training employees on energy efficiency and operational strategies would be advantageous.

References

[1] Statista. "Estimated value of the furniture market worldwide from 2022 to 2030". https://www.statista.com/statistics/977793/furniture-market-value-worldwide/ (10.08.204).

[2] Fortune Business Insights. "Furniture Market Size, Industry Share & COVID-19 Impact Analysis, By Raw Material (Wood, Metal, Plastic, and Others), Category (Indoor and Outdoor), End-User (Residential, Office, Hotel, and Others), and Regional Forecast, 2023-2030". https://www.fortunebusinessinsights.com/furniture-market-106357 (5.08.2024).

[3] Global Market Insights. "Wooden Furniture Market - By Product Type (Indoor Furniture, Outdoor Furniture), By Wood Type (Hardwood, Softwood), By Price, By Application, By Distribution Channel Forecast 2024 – 2032". https://www.gminsights.com/industry-analysis/wooden-furniture-market (5.08.2024)..

[4] Impactful Insights. "Wood Furniture Market by Wood Type (Hardwood, Softwood), Distribution Channel (Retail, Online), End User (Residential, Commercial), and Region 2024-2032". https://www.imarcgroup.com/wood-furniture-market (07.08.2024).

[5] Sihn W, Sobottka T, Heinzl B, Kamhuber F. "Interdisciplinary Multi-Criteria Optimization Using Hybrid Simulation to Pursue Energy Efficiency Through Production Planning". *CIRP Annals*, 67(1), 447-450, 2018.

[6] Wen X, Cao H, Hon B, Chen E, Li H. "Energy Value Mapping: A Novel Lean Method to İntegrate Energy Efficiency into Production Management". *Energy*, 217, 119353, 2021.

[7] Sobottka T, Kamhuber F, Sihn W. "Increasing Energy Efficiency in Production Environments Through an Optimized, Hybrid Simulation-Based Planning of Production and its Periphery". *The 24th CIRP Conference on Life Cycle Engineering,* Kamakura, Japan, 8-10 March 2017.

[8] Bonfa F, Benedetti M, Ubertini S, Introna V, Satolamazza a. "New Efficiency Opportunities Arising from Intelligent Real Time Control Tools Applications: The Case of Compressed Air Systems' Energy Efficiency in Production and Use". *10th International Conference on Applied Energy (ICAE2018),* Hong Kong, China, 22-25 August 2018.

[9] Benedetti M, Bertini I, Bonfà F, Ferrari S, Introna V, Santino D, Ubertini S. "Assessing and Improving Compressed Air Systems' Energy Efficiency in Production and Use: Findings from an Explorative Study in Large and Energy-Intensive Industrial Firms". *The 8th International Conference on Applied Energy (ICAE2016),* Beijing, China, 81 October 2016.

[10] Emre İ.E, Selcukcan Erol C. "Statistics or Data Mining for Data Analysis". *Journal of Information Technologies,* 10(2), 161-167, 2017.

[11] Bennert T, Hanson D, Maher A. "Influence of pavement surface type on tire/pavement generated noise". *Journal of Testing & Evaluation*, 33(2), 94-100, 2005.

[12] Wold S, Esbensen K, Geladi P. "Principal Component Analysis". *Chemometrics and Intelligent Laboratory Systems,* 2(1-3), 37-52, 1987.

[13] Ghiasi S, Srivastava A, Yang X, Sarrafzadeh M. "Optimal Energy Aware Clustering in Sensor Networks". *Sensors,* 2, 258-269, 2002.

[14] Abdi H, Wiiliams L.J. "Principal Component Analysis". *WIREs Computational Statistics,* 2(4), 433-459, 2010.

[15] Bro R, Smilde A.K. "Principal Component Analysis". *Analytical Methods,* 6(9), 2812-2831, 2014.

[16] Zhou W, Xu K, Yang Y, Lu J. "Driving Cycle Development for Electric Vehicle Application Using Principal Component Analysis and K-Means Cluster: With The Case of Shenyang, China". *8th International Conference on Applied Energy (ICAE2016),* Beijing, China, 8-11 October 2016.

[17] Chen Z, Xiong R. "Driving Cycle Development for Electric Vehicle Application Using Principal Component Analysis and K-Means Cluster: With The Case of Shenyang, China". *9th International Conference on Applied Energy (ICAE2017),* Cardiff, UK, 21-24 August 2017.

[18] Pořízka P, Klus J, Képeš E, Prochazka D, Hahn D.W. "On The Utilization of Principal Component Analysis in Laser-Induced Breakdown Spectroscopy Data Analysis, A Review". *Spectrochimica Acta Part B,* 148, 65-82, 2018.

[19] Cebeci Z. "fcvalid: An R Package for Internal Validation of Probabilistic and Possibilistic Clustering". *Sakarya Unıversıty Journal of Computer and Informatıon Scıences,* 3(1), 11-27, 2020.

[20] Clayman C.L, Srinivasan S.M, Sangwan R.S. "K-means Clustering and Principal Components Analysis of Microarray Data of L1000 Landmark Genes". *Procedia Computer Science,* 168, 97-104, 2020.

[21] Jannson N.F, Allen R.L, Skogsmo G, Tavakoli S. "Principal Component Analysis and K-Means Clustering as Tools During Exploration for Zn Skarn Deposits and Industrial Carbonates, Sala Area, Sweden". *Journal of Geochemical Exploration,* 233, 106909,2022.

[22] Ilu S.Y, Rajesh P, Mohammed H. "Prediction of COVID-19 Using Long Short-Term Memory By Integrating Principal Component Analysis And Clustering Techniques". *Informatics in Medicine Unlocked,* 31, 100990, 2022.

[23] Liao J, Lin J, Wu G. "Two-Layer Optimization Configuration Method for Distributed Photovoltaic and Energy Storage Systems Based on IDEC-K Clustering". *Energy Reports,* 11, 5172-5188, 2024.

[24] Eid M.H, Eissa M, Mohamed E.A, Ramadan H.S, Czuppon G, Kovacs A, Szűcs P. "Application of Stable Isotopes, Mixing Models, and K-Means Cluster Analysis to Detect Recharge and Salinity Origins In Siwa Oasis, Egypt". *Groundwater for Sustainable Development,* 25, 101124, 2024.

[25] Dugan E.L, Barbuto A.E, Masterson C.M, Shilt J, "Multivariate Functional Principal Component Analysis and K-Means Clustering to Identify Kinematic Foot Types During Gait in Children with Cerebral Palsy". *Gait & Posture,* 113, 40-45, 2024.

Trend analysis for modelling the primary energy consumption: Evidence from three largest consumers

Imene Rogai[1,*], Izzet Karakurt[1], Gokhan Aydin[1]

[1]*Karadeniz Technical University, Engineering Faculty, Mining&Energy Research Group, Mining Engineering Department, Trabzon, Turkiye*

Abstract

For decades, the global economy has been growing, and as a result, the amount of energy consumed overall has been rising rapidly. Therefore, it is crucial for the existing scientific research and decision-makers to be able to accurately forecast future energy consumption using mathematical and/or statistical techniques. There are, currently, numerous forecasting tools available for processing and analysis of the energy consumption. However, most of them have the drawbacks of intricate modeling, which makes it even more challenging when dealing with mixed frequency data. For the sake of reflecting the usability and understandability of forecasting, this paper employs, for the first time, trend analysis (TA) for developing statistical models to forecast the primary energy consumption (PEC). The three largest energy consumers, namely China, United States of America (USA) and India, are taken as the cases for the analysis. In addition, the developed models are validated and their forecasting accuracies are measured thorough various statistical tests. Moreover, the future PECs of the case countries are forecasted by the developed models. The results show that statistical models to forecast the PECs can be successfully developed using the TA as a tool.

Keywords: *Trend Analysis, Primary Energy Consumption, Modelling, Forecasting, Largest Consumers*

1 Introduction

Due to its role for both living and non-living objects to move, grow, and expand, energy is worldwide viewed as a major factor in economic progress (1-2). The energy that guides our daily lives, whether social, military, economic or scientific, can be accessed from renewable and non-renewable sources. Among them, the world relies heavily on non-renewable sources or fossil fuels such as coal, oil and natural gas (3). In 2023, more than 80 percent of the world energy requirement was met by the fossil fuels, of which around 32%, 23% and 27% for oil, natural gas and coal respectively. Fossil fuels were followed by renewables, hydropower and nuclear sources (Fig.1).

Because of the restricted quantity of resources, especially fossil fuels, one essential issue that is recently receiving more and more attention globally is energy consumption. For the preceding 20 years, the worldwide primary energy consumption (PEC) grew annually with the exception of 2009 and 2020. In 2009 and 2020, it underwent a decline as a result of the global economic recession and the Covid-19 pandemic, respectively. After a downward trend in 2020, the global PEC started to rise once more in 2023 (4-6).

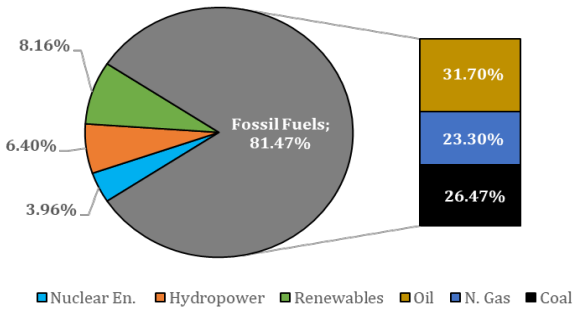


Figure 1. World energy consumption by fuel type (4)

In 2023, world's PEC reached to 620 exajoule, majority of which resulted from top three consumers such as China, the United States of America (USA) and India. These three largest consumers accounted for almost 50% of the world's PEC and the rest shared the other half (Fig.2). When considering the significance of energy for national economic expansions, particularly in relation to fossil fuels, it can be stated that it is desperately needed to understand appropriate and useful methods for using primary energy efficiently. Additionally,

* Contact email: imenerogai@ktu.edu.tr

accurate forecasts on both regional and national levels regarding PEC are still required despite the numerous studies in the relevant field.

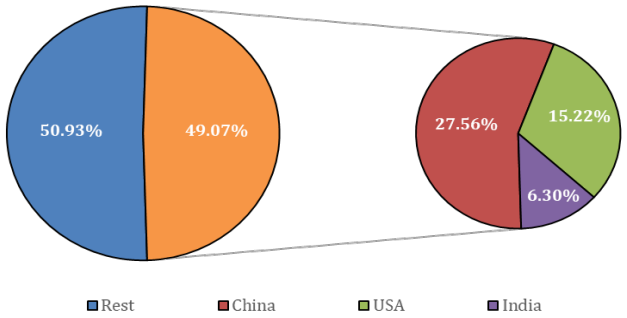


Figure 2. Three largest primary energy consumers in the world in 2023 (4)

Moreover, given their respective development patterns and status as the world's top three consumers of primary energy, China, USA, and India can serve as models for economies with less energy consumption. Furthermore, there are, of course, many modeling and forecasting techniques for developing and/or proposing predictive models in the literature. While these modeling techniques are advantageous in that they can precisely describe the phenomenon of long-term trends, they can exhibit certain limitations. In this instance, less precise and more straightforward modeling approaches might be appreciated (7). In contrast to other methods those require numerous parameters and are far more complex, linear and nonlinear modelling technique, also called as trend analysis (TA), presents a notion that readers can infer future events from what has happened in the past (8). In addition, this approach's primary benefit is its simplicity, and projections are made using whatever data is available (9). Therefore, the PECs of three largest consumers – China, USA and India – are modelled using the TA technique for the first time in this study. The rest of the paper is organized as follow. The modelling approach is introduced in second section. Section 3 presents the results and discussion. Finally, the study summarizes the core findings in fourth section.

2 Modelling Study

2.1 Data

This study utilized the PEC data from 1985 – 2023 for three largest consumers. Annual data of the PECs were officially and freely sourced from web site of Energy Institute (4). Historical trends of the PECs in the top three consumers and world were illustrated in Fig.3. It illustrates that the PECs of the related countries and world have shown an upward trend, with the average annual changes of 5.59% for China, 0.72% for USA, and 5.22% for India, while this 1.91% for the world during studied period.

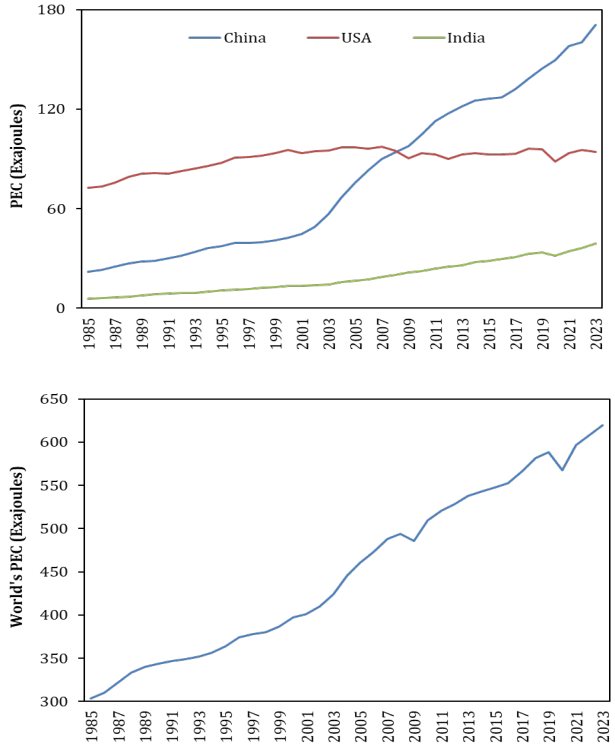


Figure 3. Historical trends of the PECs in three largest consumers and world (4)

In addition to the historical trends of the PECs, Table 1 gives the chosen variable's descriptive statistics. The data in Table 1 show that the mean of each country is comparatively higher than the standard deviations (SDs). This is a good indication for the data set being homogenous and consistent. Additionally, the fundamental assumption is that the kurtosis and skewness values should be zero for an observed data set to be symmetrical or regularly distributed. However, some other scientists assert that the observed series can also be considered to have a normal distribution if both values were within ±1.5 (10-12). As a result, the skewness and kurtosis values of the data series of all countries in Table 1 approve normal distribution of the series.

2.2 Methodology

The PECs were modeled in the study, employing the TA. By using the dependent variable's past trend, the TA can build models that forecast the variable's future values. Predictive models in the current investigation were built using the linear, logarithmic,

Table 1. Descriptive statistics for the PECs

Case	Variable	Min	Max	Mean	SD	Skewness	Kurtosis
China	PEC	22.141	126.490	60.996	35.439	0.698	-1.068
USA		72.663	97.440	88.987	7.232	-0.922	-0.253
India		5.703	28.547	14.713	6.771	0.603	-0.754

power, exponential, inverse, growth, and S regressions whose mathematical expressions are given in Table 2. The PECs' annual data were divided into two groups: the data used to train the model from 1985 to 2015, and the data used to test the model from 2016 to 2023.

To build the models, SPSS statistical software was used, which offers a selection of regression toolbox. Upon model building, the statistical validity of the suggested model was confirmed by considering the determination coefficient (R^2), the t-test, the F-test and the predicted versus the actual data. Additionally, the forecasting performances of the proposed models were also measured through statistical indices such as the mean absolute percentage error (MAPE) and relative root mean squarer error (RRMSE), whose equations are presented below. In general, the proposed model is more precise when the indicators are smaller.

Table 2. Mathematical expressions of the trend analysis-based models

Model function	**Equation**
Linear	$y = a + b.(t)$
Logarithmic	$y = (a) + \ln(t)$
Power	$y = (a).(t)^{!}$
Exponential	$y = (a).e^{!(\#)}$
Inverse	$y = (a) + (b).\text{-}\frac{1}{t}/$
Growth	$y = e^{[(\&)^{\blacktriangledown}(!).(\#)]}$
S	$y = e^{*(\&)^{\blacktriangledown}(!).+\frac{\blacktriangle}{\#}\text{-}.}$

where; y is the PEC (exajoule), *a* is the intercept or the constant, *b* is the slope, and *t* is the year.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|X_i - Y_{i)}|}{Y_i}\right).100 \quad [1]$$

$$RRMSE = \frac{RMSE}{\bar{O}}.100 \quad [2]$$

3 Results and Discussion

3.1 Modelling results and verification

The models, built with the TA within the scope of the study are given below (Eqs. 3 – 5). Initially, the models in all equation forms (given in Table 2) were built independently for each country. The best models were then chosen based on the RRMSE and MAPE values.

$$PEC_{China} = (1.745) + (3.703).t \quad [3]$$

$$PEC_{USA} = (92{,}809) - (29.420).\left(\frac{1}{t}\right) \quad [4]$$

$$PEC_{India} = e^{[(1.758)+(0.051).t]} \quad [5]$$

It is evident from the Eqs. 3 – 5 that the proposed models for China, USA and India are explained by the linear, inverse and growth functions respectively. Additionally, the proposed models' verification results are presented in Table 3. As can be observed, the developed models for China, the USA, and India all have R^2 values above 0.85, indicating a strong correlation between the year and the PECs and supporting the models statistically. Moreover, at the 95% confidence level, Table 3 confirms that the calculated *F* and *t* values for each model are greater than the tabular values, indicating that the developed models and their coefficients are statistically valid. Furthermore, an additional statistical verification indicator is provided by the predicted vs. actual data graphs in Fig.4. The related graphs show that the projected data closely matches the actual ones, demonstrating the statistical soundness of the models. Consequently, given the parameters of the investigation, it may be said that the developed models seem to be statistically valid.

3.2 Performance measures of the proposed models

The results of the developed models' performance measures using the MAPE and RRMSE is provided in Table 4. The MAPE and RRMSE values for the proposed models indicate that all of the derived models show excellent forecasting accuracy, based on the classification in Table 5.

Table 4. Performance metrics of the proposed models

Case	Equation	MAPE	RRMSE
China	Linear	9.43	0.83
USA	Inverse	2.70	0.31
India	Growth	6.21	1.42

Table 3. Statistical verification results of the proposed

Case	Model function	Variable	$t_{computed}$	t_{table}	$F_{computed}$	F_{table}	R^2
China	Linear	Constant	2.421	1.690	268.886	4.125	0.90
		Year	16.397				
USA	Inverse	Constant	1.825	1.690	43.722	4.125	0.86
		Year	-5,.775				
India	Growth	Constant	133.927	1.690	5185.630	4.125	0.99
		Year	72.011				

Table 5. Reference table of the MAPE and RRMSE for the model accuracies (13-14)

MAPE (%)	RRMSE (%)	Predictive ability
MAPE ≤ 10	RRMSE < 10	Excellent
11 ≤ MAPE ≤ 20	10 < RRMSE < 20	Good
21 ≤ MAPE ≤ 51	20 < RRMSE < 30	Qualified
MAPE > 51	RRMSE ≥ 30	Unqualified

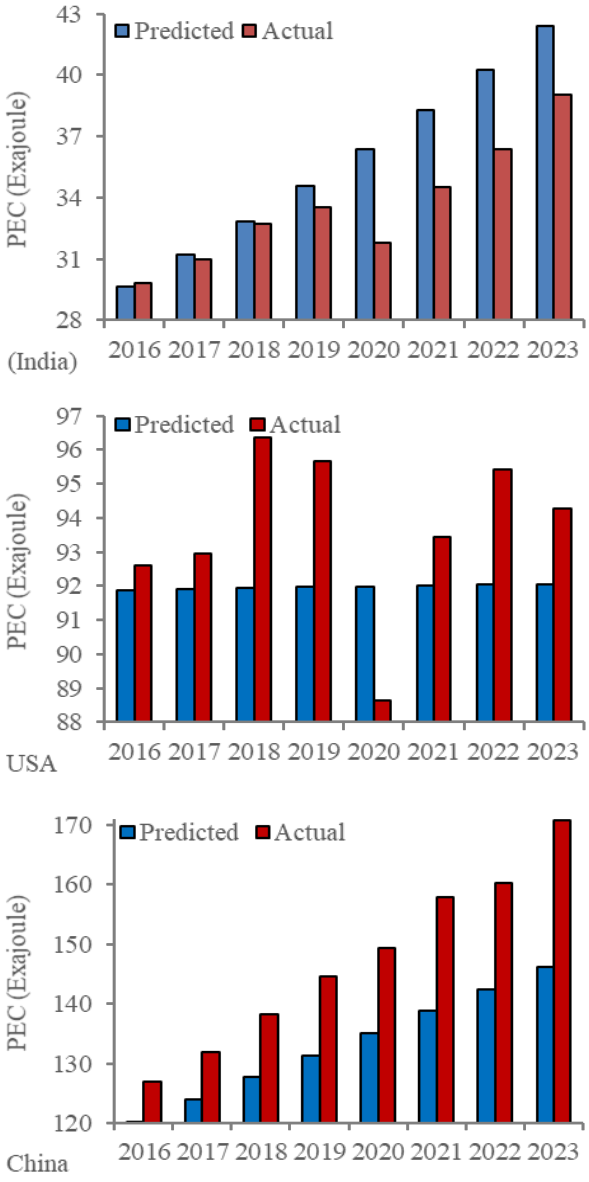


Figure 4. Actual vs predicted PECs

3.3 Forecasting the PECs by the proposed models

Using the proposed models, the future PECs of three largest consumers were forecasted from 2024 to 2030 and the results are illustrated in Fig.5. Additionally, a comparison of the PECs in 2023 and 2030 is also provided in Fig.6. It is evident from the related figures that a substantial increase is expected for India among three largest consumers, with 37.46% by 2030, as compared to 2023's data. That means, India's PECs will reach from 39.016 exajoule in 2023 to 60.582 exajoule in 2030, an annual increase rate of 3.43%. China and USA will experience a decrease in their PECs through the forecasted period, annually 2.17% and 0.02% respectively. On the other hand, the total PECs of three largest consumers will be 324.835 exajoule by 2030 when compared to the data in 2023, which was 304.026 exajoule. In other words, 6.84% increase is expected for the total PECs of three largest consumers in 2030.

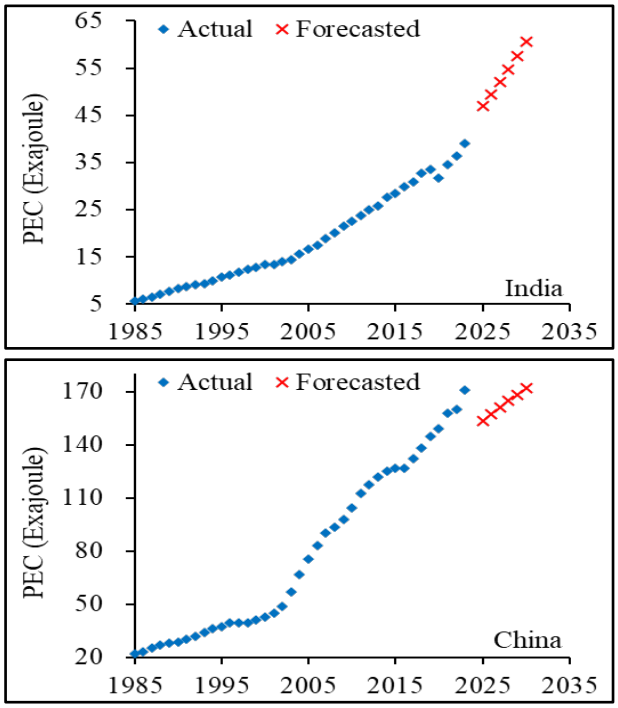


Figure 5. Forecasted PECs of top largest consumers

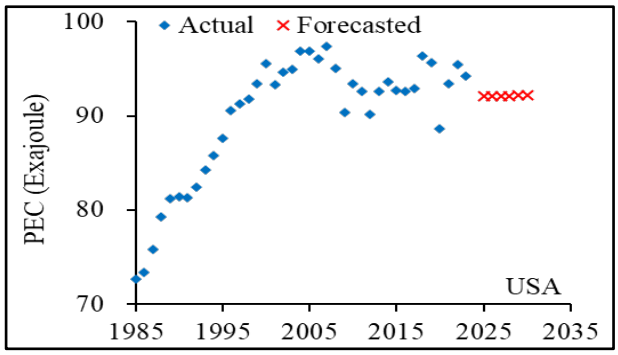


Figure 5 (Continued)

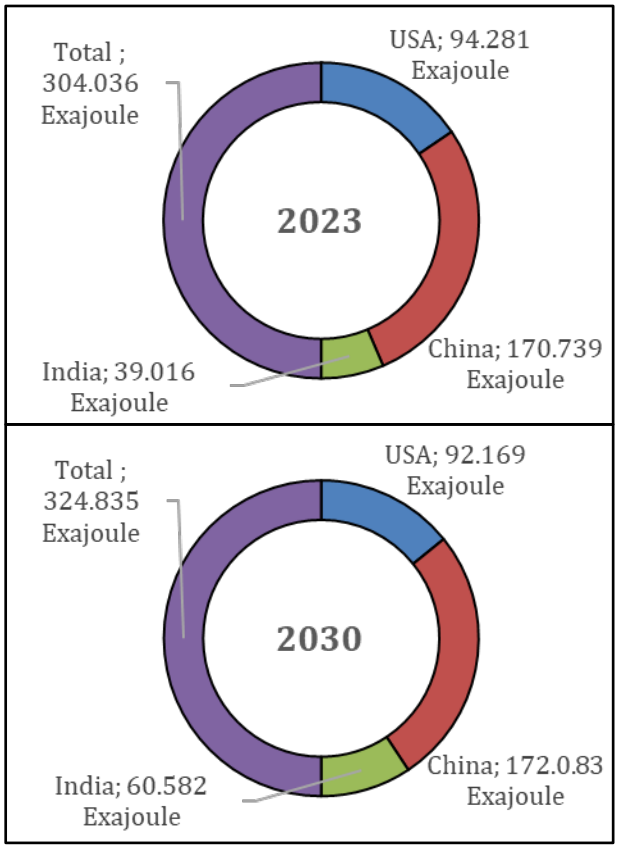


Figure 6. Total PECs of three largest consumers in 2023 and 2030

4 Conclusions

Followings are the main results of the study, using the TA for the PEC modeling.

a) It was determined that the proposed models could be successfully used to forecast the PEC of three largest consumers with accuracy ranges from 90.57%, 93.79% and 97.3% for China, India and USA respectively.

b) It was forecasted that the total PECs of three largest consumers will reach from 324.026 exajoule in 2023 to 324.835 exajoule in 2030, with an increase of 6.84%.

c) The overall results showed that the TA may be used as a tool to build predictive models for PEC.

Concussively, given their respective development patterns and status as the world's top three consumers of primary energy, China, USA, and India can serve as models for economies with less energy consumption. Thus, to generalize the results, more predictive models that take into account different variables such as total population, gross domestic product, and urban population are needed.

References

[1] Jiang T, Chen Z. "Analysis of the impact of energy consumption on regional agricultural economic growth based on panel data". *Frontiers in Ecology and Evolution*, 11, 1134448, 2023. https://doi.org/10.3389/fevo.2023.1134448

[2] Savage DM. "Global energy consumption: An analysis of variables that shape per capita usage, or how pump price, urbanization, and fossil fuels ımports ımpact fossil fuels consumption per capita across OECD countries". Doctor of Philosophy (PhD), Dissertation, Political Science & Geography, Old Dominion University. https://digitalcommons.odu.edu/gpis_etds/161

[3] Karakurt I, Avci DB, Aydin G. "Leveraging the trend analysis for modelling of the greenhouse gas emissions associated with coal combustion". *Environmental Science and Pollution Research (in press),* https://doi.org/10.1007/s11356-024-34654-3

[4] EI. Energy Institute, the 2024 statistical review of world energy. https://www.energyinst.org/statistical-review. Accessed on August 2024.

[5] IEA. World energy outlook 2021. https://www.iea.org/reports/world-energy-outlook-2021; [accessed on 12 August 2024].

[6] IPCC. Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge (UK), NY (USA): Cambridge University Press, 2013.

[7] Bianco V, Manca O, Nardini S. "Electricity consumption forecasting in Italy using linear regression models". *Energy*, 34(9), 1413-1421, 2009. https://doi.org/10.1016/j.energy.2009.06.034

[8] Celiker M, Yukseler U, Dursun OF. "Trend analyses for discharge-recharge of Tacin karstic spring (Kayseri, Turkey)". *Journal of African Earth Sciences*, 184, 104344, 2021. https://doi.org/10.1016/j.jafrearsci.2021.104344

[9] Kone CA, Buke T. "Forecasting of CO_2 emissions from fuel combustion using trend analysis". *Renewable and Sustainable Energy Reviews*, 14, 2906–2915, 2010. https://doi.org/10.1016/j.rser.2010.06.006

[10] Erbay S, Beydogan HO. "Attitudes of educators towards educational research". *Kırsehir Journal of the Faculty of Education*, 18(3), 246-260, 2017.

[11] George D, Mallery P. IBM SPSS Statistics 27 Step by Step. Routledge New York, USA, 2021.

[12] Urbano RC. "Using Secondary Datasets to Understand Persons with Developmental Disabilities and Their Families". In Int Rev Res Dev Disabi: Amsterdam, The Netherlands, 2023.

[13] Lewis CD. "International and business forecasting methods". Butterworths, London, 1982.

[14] Li MF, Tang XP, Wu W, Liu HB. "General models for estimating daily global solar radiation for different solar radiation zones in mainland China". *Energy Conversion and Management,* 70, 139–48, 2013. http://dx.doi.org/10.1016/j.enconman.2013.03.004

Development of statistical models for estimating the hydropower consumption

Imene Rogai[1,*], Izzet Karakurt[1], Gokhan Aydin[1]

[1]*Karadeniz Technical University, Engineering Faculty, Mining&Energy Research Group, Mining Engineering Department, Trabzon, Turkiye*

Abstract

When it comes to renewable and sustainable energy sources, hydropower (HP) currently holds the highest position given that it presently supplies more than 50% of the world's renewable energy production and almost 7% of the world's total energy requirements. Thus, in this paper, it is mainly focused on the developing statistical model for estimating hydropower consumption (HPC) in Turkiye. Regression analysis (RA), as the most widely used modeling method for energy consumption estimation, is applied in order to develop estimating model using the total population (TP), urban population (UP) and gross domestic products (GDP) as the independent variables. Additionally, verification and estimating accuracy of the developed model are statistically tested by several indices. Moreover, HPC projection is also made with the developed model. The results indicate that a decrease is expected for the HPC of Turkiye for the foreseeable future.

Keywords: *Regression analysis, Hydropower consumption, Statistical model, Turkiye*

1 Introduction

Energy plays a key role in any country's sustainable development plans and objectives. While many countries have distinct approaches to achieving these objectives, a unifying thread throughout most of these policies is primary energy resources that are either found or stored in nature (1). These sources include coal, oil, natural gas (often referred to as fossil fuels), renewables, nuclear and hydropower. Among these sources, the world's energy needs have historically been met mostly by fossil fuels. The latest data, which revealed that the world's total energy consumption reached to 620 exajoule, demonstrated that the fossil fuels accounted for around 82% of total energy needs. The fossil fuels were followed by renewables, hydropower and nuclear sources respectively (2).

Situated between Asia and Europe strategically, Turkiye has persisted in holding her position as a rising star in the region. With more over 85 million people, she makes up 1.07% of the global population. As of the latest data from WBI (3), with an annual growth rate of 4.52%, her gross domestic product surpassed one trillion US dollars in 2023, exceeding the global average of 2.72%. Additionally, Table 1 lists Turkiye's primary energy resources.

Table 1. Energy sources of Turkiye (2,4)

Resources	Amount	Share (%)
Oil (Billion tonnes)	-	-
Natural gas (Billion m^3)	544	0.29
Coal (Million tonnes)	11525	1.1
Nuclear (TWh)	-	-
Hydropower (TWh)	63.9	1.5
Wind (TWh)	34.1	1.5
Solar (TWh)	20.5	1.2
Other renewables* (TWh)	19.5	2.5

Out of the non-renewable resources, one may deduce from Table 1 that Turkiye has significant coal reserves. She also possesses a sizable quantity of renewable energy sources, including geothermal biomass, hydropower, and wind. Despite having these sources, Turkiye is a country, importing more than 70% of her energy needs owing to several factors, most of which is the lack of fossil fuel reserves (5). This is confirmed by the energy production and consumption values of the country as indicated in Table 2. It is evident from the related table that Turkiye only meets own renewable energy requirement. On the other hand; in 2023, Turkiye's primary energy consumption (PEC) reached to 7 exajoule, comprising the 1.1% of the world total. As can be seen in Fig.1, the fossil fuels shared the majority of the consumption, more than 80%. The

* Contact email: imenerogai@ktu.edu.tr

fossil fuels were followed by renewables and hydropower with 10.17% and 8.53% respectively in the same year. Among the energy resources, in accordance with the plant's scale, hydropower is regarded as a clean, renewable, and ecologically beneficial energy source. It is a renewable energy source with advanced technology and competitive pricing that may significantly improve the functionality and reliability of energy systems (6). As it can be understood from both Table 2 and Fig.1 that this energy resource has also significant place in Turkiye's energy balance. In other words, this source is critically important for enabling Turkiye to attain energy independence. For this reason, HPC forecasts are among the most important policy tools that Turkiye's decision-makers will use. Thus, in this study, a predictive model was developed for estimating the HPC of Turkiye, using the regression analysis. In this regards, total population (TP), urban population (UP) and gross domestic products (GDP) that they are the variables highly correlated to energy demands, were selected as the independent variables, while HPC was dependent variable.

Table 2. Energy production and consumption of Turkiye by fuel type (2)

Oil (Million ton)		Natural gas (Billion m^3)		Coal (Million ton)		Renewables (EJ)		Hydropower (TWh)		Nuclear (TWh)	
Prod.	Cons.	Prod.	Cons.	Prod.	Cons.	Prod.	Cons.	Prod.	Cons.	Prod.	Cons.
-	49.70	-	51.20	0.83	1.75	0.69	0.69	67.20	166.67	-	-

TWh: Terawatt-hour; EJ: Exajoule; Prod.: Production; Cons.: Consumption

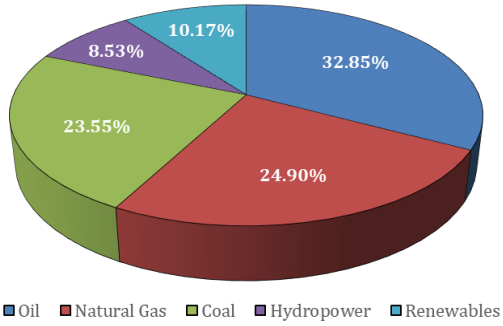


Figure 1. PEC of Turkiye by fuel type in 2023 (2)

2 Modelling Approach

2.1 Data description

The current study uses annual data on four variables from 1980 to 2023 that are freely available on line (2,3). Table 2 provides the chosen variables' descriptive statistics. Table 2's data demonstrates that each variable's mean is significantly larger than its standard deviation (SD). This suggests that the data set is homogeneous and consistent. Additionally, a fundamental premise of symmetry or regular distribution in an observed data set is that the kurtosis and skewness values should be zero. Some scientists, however, contend that if both values were within ±1.5, the observed series could likewise be regarded as having a normal distribution (7-9). As a result, the skewness and kurtosis values of the data series of all variables in Table 2 confirm normal distribution of the series.

2.2 Methodology

To determine which model was best, the backward multiple linear regression analysis approach was applied. This approach involves starting the model-building process with all independent variables included. Subsequently, the process of creating the model proceeds step-by-step, with each variable being eliminated based on its partial F value. When a discarded variable has a strong statistical contribution, the process is terminated and the model is built. The HPC's annual data were divided into two groups: the data used to train the model from 1980 to 2015, and the data used to test the model from 2016 to 2023. After the models were created, a variety of statistical tests such as the adjusted coefficient of determination (adj. R^2), an F – test, a t -test and the actual vs predicted data plots, were used to confirm the models' goodness-of-fit and the statistical significance of the estimated variables. Moreover; in order to show the forecasting accuracy and effectiveness of the proposed model, the mean absolute percentage error (MAPE) and relative root mean square error (RRMSE), whose mathematical equations are given in Eqs.1-2, were adopted. Based on these indicators, in general, the proposed model is more precise when the indicators are smaller.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|X_i - Y_i|}{Y_i}\right).100 \quad [1]$$

$$RRMSE = \frac{RMSE}{\bar{o}}.100 \quad [2]$$

Table 2. Descriptive statistics for the HPC

Variable	Min	Max	Mean	SD	Skewness	Kurtosis
HPC	0.120	0.650	0.335	0.143	0.125	-0.635
TP	44.089	78.218	61.380	10.175	-0.064	-1.164
UP	19.302	57.577	38.961	11.104	-0.089	-1.035
GDP	59.937	957.799	346.128	303.335	0.898	-0.746

where; n is the total number of data, X_i is the predicted HPC, Y_i is the actual HPC, $\emptyset$ is the mean value of observed data.

3 Results and Discussion

3.1 Modelling results

Eqs. 3 offers a regression model that has been developed and can be used to forecast Turkiye's HPC. As may be observed, Turkiye's HPC is expressed as a linear function of multiple independent variables.

$$HPC=(0.033)-(0.007).TP+(0.019).UP-(10)^{-5}.GDP \quad [3]$$

where, *HPC:* Hydropower consumption (Exajoule), *TP:* Total population, *UP:* Urban population, *GDP:* Gross domestic products (Current US$).

One way to assess the importance of the variables in predicting the dependent variable is by looking at the contribution rates of the independent variables in regression models (10). Thus, the contribution rates of the independent variables were computed as 23.46% for TP, 70.68% for UP and 5.86% for GDP respectively. That means, the priority should be given to in order of UP, TP and GDP when the proposed model is evaluated. Additionally, the verification results of the developed model are given in Table 3. As can be shown, the developed model's R^2 value is greater than 0.75, showing a strong correlation between the TP, UP, GDP, and HPC. Moreover, Table 3 demonstrates that the computed F value for the model is greater than the tabular F value. The derived model is, therefore, statistically significant at the 95% confidence level. The coefficients' t-test results are greater than the tabulated t-value, confirming the significance of the model's coefficients in the final model. Furthermore, Fig.2 shows the curve of the actual against the projected HPC for the model instance. Fig.2 illustrates that the suggested model appears to forecast the HPC with a high degree of accuracy and an acceptable variance from the actual data, suggesting that the model is statistically sufficiently predictive. Overall, the model's verification affirms that Turkiye's HPC could be modeled in this manner. Finally, in order to show the performance accuracy of the proposed model, the MAPE and RRMSE values were computed as 8.01% and 12.79% respectively. Based on the MAPE and RRMSE classifications in Table 4, it can be concluded that the proposed model has "excellent" and "good" predictive abilities respectively.

Table 3. Statistical verification results of the developed model

Variable	$t_{computed}$	t_{table}	$F_{computed}$	F_{table}	Adj.R^2
Constant	1.943				
TP	-1.834		39.139		0.78
UP	2.704				
GDP	-1.638				

Table 4. Reference table of the MAPE and RRMSE for the model accuracies (11,12)

MAPE (%)	**RRMSE (%)**	**Predictive ability**
MAPE ≤ 10	RRMSE < 10	Excellent
11 ≤ MAPE ≤ 20	10 < RRMSE < 20	Good
21 ≤ MAPE ≤ 51	20 < RRMSE < 30	Qualified
MAPE > 51	RRMSE ≥ 30	Unqualified

3.2 Forecasting the HPC by the proposed model

Turkiye's HPC was projected by the proposed model from 2025 to 2030 (Fig.3). One can see from the related figure that the HPC of Turkiye will experience a decrease trend with 3.58% by 2030, when compared to the data in 2023. However, as it is see in the figure also that this decrease will occur within a limited range, an annual 0.64% in the foreseeable future.

4 Conclusions

Among the noteworthy findings from the study that included the HPC modeling investigation are as *i*) it was determined that the proposed model could be successfully used to forecast the HPC of Turkiye with accuracy of 91.99%. *ii*) it was disclosed that the UP was statistically found as the most significant variable, followed by the TP and GDP. iii)the

proposed model projected the HPC of Turkiye as 0.58 exajoule in 2030 with a decrease rate of 3.58%, compared to 2023 data. *iv*) the overall results showed that the regression analysis may be effectively used as a tool for deriving the models. Finally, it is strongly recommended that more and more studies, taking into different variables, should be conducted for the modelling of the HPC in order to generalize the results.

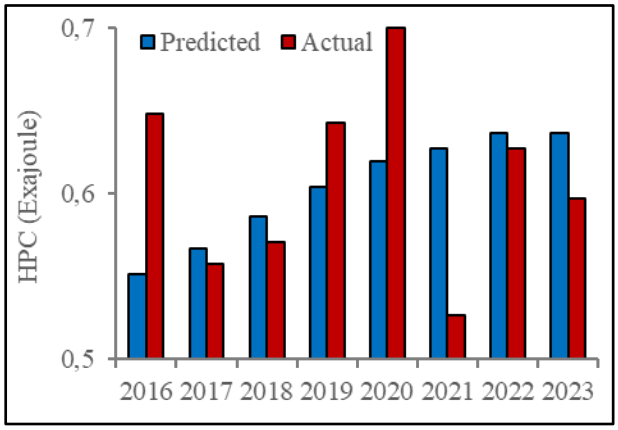


Figure 2. Predicted vs actual HPC of Turkiye

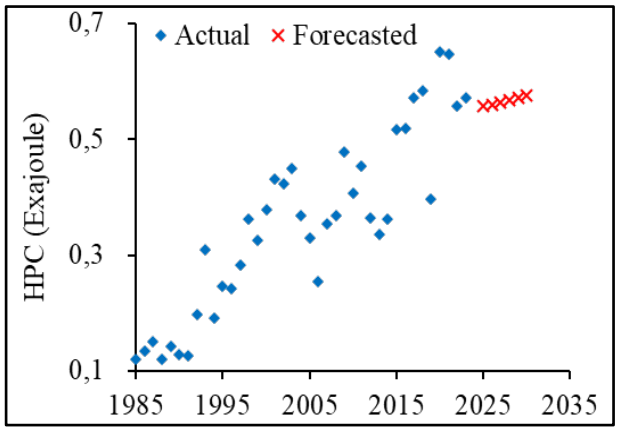


Figure 3. Actual and forecasted HPC of Turkiye

References

[1] Igbeghe CB, Mizik T, Gabnai Z, Bai A. "Trends and characterization of primary energy sources by energy and food prices". *Energies*, 16, 3066, 2023. https://doi.org/10.3390/en16073066

[2] EI. "Statistical review of world energy", published by Energy Institute. ttps://www.energyinst.org/statistical-review. Accessed on on 20 June 2024.

[3] WBI. 2024. Worldbank indicators on total population, https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS; [accessed on 20 August 2024].

[4] MEN. "Ministry of Energy and Natural Resources of Turkiye", https://enerji.gov.tr/bilgi-merkezi-enerji-dogalgaz. Accessed on 01 July 2024.

[5] Erdemir NA. "Energy dependence of Turkey: The role of renewable energy sources". *Başkent Üniversitesi Ticari Bilimler Fakültesi Dergisi*, (6)1, 1 – 14, 2022.

[6] Kouadio CA, Kouassi KL, Diedhiou A, Obahoundje S, Amoussou E, Kamagate B, Paturel J, Coulibaly TJH, Coulibaly HSJP, Didi RS, Savane I. "Assessing the hydropower potential using hydrological models and geospatial tools in the White Bandama Watershed". *Frontiers in Water*, 4, 844934, https://doi.org/10.3389/frwa.2022.844934

[7] Erbay S, Beydogan HO. "Attitudes of educators towards educational research". Kırsehir Journal of the Faculty of Education, 18(3), 246-260, 2017.

[8] George D, Mallery P. IBM SPSS Statistics 27 Step by Step. Routledge New York, USA, 2021.

[9] Urbano RC. "Using Secondary Datasets to Understand Persons with Developmental Disabilities and Their Families". In Int Rev Res Dev Disabi: Amsterdam, The Netherlands, 2023.

[10] Karakurt I, Aydin G, Aydiner K. "A study on the prediction of kerf angle in abrasive waterjet machining of rocks". *Proceedings of the Institution of Mechanical Engineers Part B-Journal of Engineering Manufacture,* 226 (9), 1489-1499, 2012. https://doi.org/10.1177/0954405412454395

[11] Lewis CD. "International and business forecasting methods". Butterworths, London, 1982.

[12] Li MF, Tang XP, Wu W, Liu HB. "General models for estimating daily global solar radiation for different solar radiation zones in mainland China". Energy Conversion and Management, 70, 139–48, 2013. http://dx.doi.org/10.1016/j.enconman.2013.03.004.

Sıfır Atık Yönetiminin Tersine Lojistik Faaliyetlerinde Uygulanmasının Temel Engellerinin Bulanık-AHP Kullanılarak Önceliklendirilmesi

Nida DURMAZ[1]*, Sümeyye BAYINDIR[2]

[1,2] *Gebze Teknik Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği, Kocaeli, TÜRKİYE*

Özet

Küresel nüfus artışı ve sınırlı doğal kaynaklar, bu kaynakların hızla tükenmesine yol açmaktadır. Tüketim toplumunun yayılması, atık oranlarını artırarak insan-ekosistem dengesini olumsuz etkilemiş ve sürdürülebilir yaşam standartlarını düşürmüştür. Bu bağlamda, nihai ürünlerin geri kazanımını içeren tersine lojistik kavramı ortaya çıkmıştır. Tersine lojistik, çevreye zararlı atıkların, hurdaya ayrılmış ürünlerin, yarı mamul ürünlerin ve yeniden kullanılabilir hammaddelerin geri dönüşüm faaliyetlerini kapsamaktadır. Bu sebeple, tersine lojistik faaliyetleriyle sürdürülebilir üretimin yanı sıra sorumlu kaynak koruma prensiplerinin de gerçekleştirilmesi için Sıfır Atık Yönetimi uygulaması önem kazanmaktadır.

Bu çalışmada Türkiye'nin tersine lojistik ağında Sıfır Atık Yönetimi uygulanmasının önündeki engeller kapsamlı bir literatür araştırması ve uzman görüşlerine dayalı olarak belirlenmiştir. Engellerin göreceli ağırlıklarının elde edilmesi için uzmanlardan alınan ikili karşılaştırma matrisleri kullanılarak Bulanık AHP yöntemi uygulanmıştır. Literatürde, tüm atık türlerini tersine lojistik faaliyetleri çerçevesinde ele alan bir çalışma bulunmamaktadır. Bu açıdan, bu çalışma ile literatürde önemli bir boşluk doldurulmaktadır.

Sonuçlar, yasal ve örgütsel kriterlerin önemini vurgulamaktadır. Özellikle, "Son Kullanım Ürünleri için Yasa, Yönetmelik ve Direktif Eksikliği" ve "Yasadışı Geri Dönüşüm Pazarı" alt kriterleri önemli yer tutmaktadır. Örgütsel engeller arasında "Motivasyon ve Ödül Eksikliği" ve "Organizasyon Yapısı ve Destek Eksikliği" öne çıkmaktadır. Sosyal Kriterler arasında yer alan "İnsan Sağlığına Etkisi" alt kriteri de önemli bir engel olarak belirlenmiştir.

Mevcut kaynakların bu kriterlere yönlendirilmesi etkili bir Sıfır Atık Yönetimi stratejisi oluşturmada çok önemlidir. Çünkü yasal düzenlemelerin güçlendirilmesi ve örgütsel yapıların iyileştirilmesi, atık yönetim süreçlerinin daha verimli ve sürdürülebilir olmasını sağlayarak sıfır atık hedeflerine ulaşmada kritik rol oynayacaktır.

Çalışmada oluşturulan çerçeve, her ülkenin kendine özgü dinamikleri göz önünde bulundurularak dünya genelinde uygulanabilir. Böylece, bu çalışma küresel ölçekte sürdürülebilir atık yönetimi stratejilerinin geliştirilmesine olanak sağlayacak bir yaklaşım önermektedir.

Anahtar Kelimeler: Bulanık AHP, Sıfır Atık Yönetimi, Tersine Lojistik

Prioritization of Key Barriers for the Implementation of Zero Waste Management in Reverse Logistics Activities using Fuzzy-AHP

Abstract

Global population growth and limited natural resources are leading to the rapid depletion of these resources. The spread of consumer society has increased waste rates, negatively affecting the human-ecosystem balance and lowering sustainable living standards. In this context, the concept of reverse logistics, which involves the recovery of end-products, has emerged. Reverse logistics encompasses the recycling activities of environmentally harmful

*Contact email: ndurmaz@gtu.edu.tr, sbayindir@gtu.edu.tr

waste, discarded products, semi-finished products, and reusable raw materials. Therefore, the application of Zero Waste Management (ZWM) has gained importance in achieving sustainable production and responsible resource conservation principles through reverse logistics activities.

In this study, the barriers to implementing ZWM in Turkey's reverse logistics network have been determined based on a comprehensive literature review and expert opinions. To obtain the relative weights of the barriers, the Fuzzy AHP method was applied using pairwise comparison matrices collected from experts. There is no study in the literature that considers all types of waste within the framework of reverse logistics activities. In this respect, this study fills a significant gap in the literature.

The results highlight the importance of legal and organizational criteria. Specifically, the sub-criteria "Lack of enforced law, legislation, and directives for EoL products" and "Illegal recycling market" are significant. Among organizational barriers, "Lack of motivation and reward" and "Lack of organizational structure and support" stand out. Additionally, the sub-criterion "Effect on human health" within the social criteria has been identified as a significant barrier.

Focusing resources on these criteria is crucial for developing an effective ZWM strategy. Strengthening legal regulations and improving organizational structures play a critical role in making waste management processes more efficient and sustainable, thereby achieving zero waste goals.

The framework developed in this study can be applied globally, considering the unique dynamics of each country. Thus, this study proposes an approach that will enable the development of sustainable waste management strategies on a global scale.

Keywords: *Fuzzy AHP, Zero Waste Management, Reverse Logistics*

1 Introduction

Nature has limited resource capacity, and due to global overpopulation, these limited resources are rapidly depleted. This situation necessitates the search for new raw materials and resources while protecting existing ones. Furthermore, as consumer society has spread, waste rates have increased even more and have become a global problem. This issue affects the human-ecosystem balance and lowers sustainable living standards. As a result of these events, the term reverse logistics has emerged, which includes the recovery of end-products by gaining value step by step [16].

Reverse logistics emerged due to the interaction between environmental management and businesses affected by climate change and increased pollution from industrial activities. It involves recycling processes that handle environmentally harmful waste, discarded products, semi-finished products, and reusable raw materials. While often viewed as merely recovering used products, reverse logistics is crucial for both environmental and business sectors. Legal regulations and support for reverse logistics have increased globally, largely because these practices facilitate waste recycling and energy production [30].

The idea of "waste management in reverse logistics" refers to a concept that incorporates recycling and waste management processes. Waste management consists of six stages: generation, collection, transportation, sorting, processing, and disposal [1,16]. It represents an approach focused on environmental protection and sustainability. Waste collection and recycling help to preserve natural resources and reduce environmental impacts. As a result, waste management in reverse logistics is becoming increasingly important as an environmentally friendly practice.

As environmental awareness grows, the importance of addressing waste prevention to storage within the zero waste philosophy is becoming clearer [2,33]. Understanding the stages of waste management is essential for optimizing sustainability and moving toward a circular economy [2, 29]. Zero Waste Management (ZWM) aims for 0% waste creation rather than 100% waste recovery. Achieving zero-waste societies requires: (1) sustainable production through cradle-to-cradle design, (2) responsible natural resource

consumption, and (3) resource conservation [35]. ZWM sustainability depends on waste separation and collection at the source, enhancing reuse and recycling rates and reducing pollution and resource waste from municipal solid waste disposal.

2 Literature review

Multi-criteria decision making (MCDM) is regarded as a crucial branch of decision-making theory. In a recent literature review, [6] and [18] highlighted an apparent increase in the use of MCDM analysis in the field of environmental science.

[22] focused on identifying and ranking reverse logistics adoption solutions in the electronics industry to overcome its barriers. To overcome the barriers to reverse logistics adoption, the authors proposed a methodology based on fuzzy AHP and the fuzzy technique for order performance by similarity to ideal solution (TOPSIS) to identify and rank the solutions. They used Fuzzy AHP to obtain weights of the barriers as criteria through pairwise comparison, and fuzzy TOPSIS to obtain the final ranking of the reverse logistics adoption solutions. They used this integrated method in a case study for the Indian electronics industry. [3] investigated the impact of sustainable practices on reverse logistics (RL) recovery options in Pakistan, a developing country. Six RL recovery options were evaluated by the authors using a set of 15 criteria that encompassed the three dimensions of sustainability. To rank these RL recovery options within the context of China Pakistan Economic Corridor, they used a combination of VIKOR and GRA techniques in a fuzzy environment. The authors emphasized the importance of waste management, biodiversity impact, and economic growth when developing sustainable RL practices in Pakistan. However, according to their findings, remanufacture and reuse emerge as the most prominent RL practices due to their significant economic and environmental benefits. [11] used MCDM methods to identify reverse logistics barriers in the lubricant oil industry and ranks the barriers found in the literature in the Indian context. The author identified 20 barriers in 11 groups and used an expert panel to rank the criteria and sub-criteria. The final barrier ranking shows that infrastructure barriers, including storage and transportation, are the most important in implementing reverse logistics in the lubricant industry, followed by economic barriers. The third most significant barrier is a lack of motivation and reward, which is a cultural barrier, comes out as the third most significant barrier.

[35] presented a hybrid MCDM methodology for analyzing resource recovery business model (RRBM) barriers to evaluate potential solution strategies. Working with the expert decision-making group via semi-structured forms, the micro, macro, and Meso-level barriers in this study were finalized while considering the emerging economies. In the second stage, the importance levels of the twenty-five barriers were determined using the Improved Fuzzy Stepwise Weight Assessment Ratio Analysis (IMF-SWARA) method. In the third stage, eight major solution strategies were evaluated to determine their ability to overcome the barriers. At this point, the authors used the Fuzzy Measurement Alternatives and Ranking based on Compromise Solution (F-MARCOS) to evaluate the decision-makers' opinions more accurately, which may not be confident within the scope of the study. [16] used a two-stage methodology to handle waste batteries and design networks with multiple objectives. They used SF-AHP to determine the importance weights of the objectives in the first stage of their methodology. Following that, in the second stage, they proposed a multi-objective MILP model for maximizing the total weighted satisfaction degree of multiple objectives, including cost and carbon emission minimization and employment and development rate maximization.

Studies related to the ZWM are quite limited in the literature. [4] identified 12 key barriers to zero waste implementation in Turkey in light of circular economy principles. The researchers then used fuzzy DEMATEL methodology to determine which barrier has the most influence on zero waste management success and to identify the interdependence of these barriers. Their findings indicate that uncertainty about the goals and tactics of the circular economy, as well as a lack of financial and economic assistance, are major causal barriers to Turkey's ZWM. [25] created the i-ZEWATA (Industrial Zero Waste Tiered Analysis), a hybrid four-step multi-criteria decision-support model, to manage complex iron and steel industrial waste in Africa, which necessitates a system engineering approach to ZWM. Using a system-engineered approach, the authors demonstrated that iron and steel waste can be identified, visualized, prioritized, and managed to promote zero-waste. With their methodology, they also identified relationship

patterns to environmental, social, operational, and economic aspects with system behavioral patterns and outcomes. [34] proposed a practical integrated MCDM approach for evaluating the performance of a zero-waste city and applied it to a typical coal resource-based province in China. In each city, they investigated the spatiotemporal characteristics and factors influencing the performance of zero-waste cities. Heavy industrial structure, widespread underutilization of industrial solid waste, inadequate management of hazardous solid waste, low rate of urban domestic waste classification, and ineffective treatment were identified as challenges to zero-waste city development by the authors. [2] focused on the municipal solid waste collection methods by focusing on ZWM using AHP. They examined five different methods for waste separation at the source based on five commonly used waste disposal methods in their study.

2.1 Research Gap

In the literature, there are some MCDM studies on waste management. However, these studies are generally either customized based on waste types or do not cover reverse logistics activities in depth. This study provides a general and comprehensive framework for identifying and prioritizing barriers to ZWM in reverse logistics. According to our knowledge, there are no studies in the literature that focus on ZWM barriers for reverse logistics and involve decision making in an uncertain environment. In this regard, this study will fill a significant gap in the literature.

3 Hierarchy

In this study, the implementation barriers of ZWM in reverse logistics were determined based on a comprehensive literature review and experts' suggestion. The literature review incorporated from databases including Scopus, Web of Science, and Google scholar. The determined keywords are "zero waste management", "reverse logistics" and "multi criteria decision making".

Based on this, a comprehensive analysis has identified 6 basic barriers for the implementation of zero waste management in reverse logistics activities: organizational, legal, social, environmental, economical and infrastructural. According to the literature review and expert opinions, 21 sub-criteria were determined. The names, descriptions and supporting references of the sub criteria are shown in Table 1 below.

Table 1. Implementation Barriers of Zero Waste Management in Reverse Logistics.

Main Criteria	Code	Sub-criteria	Definition	Supporting References
Organizational Barriers	OB1	Lack of motivation and reward	This barrier deals to the deficiency of motivation and incentive structure among participants in waste management procedures.	[21, 22]
	OB2	Lack of organizational structure and support	This barrier refers to the situation where an organization does not have an appropriate organizational structure and support system to achieve zero waste goals.	[21, 22]
	OB3	Lack of qualified labor	The significance of having a trained and informed workforce is emphasized with this criterion. To meet this requirement, the sector of zero waste management may need to recruit competent professionals through training programs, educational campaigns, and other means.	[7, 9, 12]
	OB4	Resistance to change	It describes the resistance or unwillingness that arises among stakeholders or inside an organization when new waste management procedures are implemented, or more sustainable operations are switched to.	[19, 21, 22, 23]

	LB1	Illegal recycling market	Materials with high economic value that are recyclable can be collected in locations without environmental security. This situation often leads to illegal activities.	[10, 16, 20]
Legal Barriers	LB2	Lack of enforced law, legislation, and directives for EoL products	This barrier is the inadequate or inefficient application of waste management and recycling laws when products reach the end of their useful lives.	[9, 11, 22]
	LB3	Lack of government supportive policies for RL practices	This barrier refers to the lack of comprehensive policies and guidelines, economic incentives, intersectoral cooperation and coordination required for zero waste management	[11, 23, 28]
	SB1	Lack of awareness	Internal and external stakeholders do not have sufficient knowledge and awareness about the processes and practices required for zero waste management.	[16, 20, 22]; Expert's suggestion
Social Barriers	SB2	Effect on human health	This criterion deals with the possible negative effects on health that could arise from the reverse logistics system's waste management practices. To reduce harmful impacts on public health and to ensure the wellbeing of the community and individuals engaged in waste management operations, a thorough and responsible approach to waste management is essential.	[7, 15, 19, 23]
	IB1	Lack of data	This barrier refers to the lack of data required in zero waste management processes and the inadequacy of the infrastructural arrangements required to collect, analyze and share this data.	[11, 16]; Expert's suggestion
	IB2	Lack of infrastructure facility	This criterion describes the absence or insufficiency of fundamental physical frameworks required to enable the effective handling and processing of waste materials. The reverse logistics system can be affected at several levels by this absence of infrastructure, which can take many different forms.	[7, 11]; Expert's suggestion
Infrastructural Barriers	IB3	Lack of availability of technology & information system	It is related to the absence of necessary technology and information systems for adopting reverse logistics practices.	[9, 11, 12]
	IB4	Lack of power and equipment infrastructure based on waste type	It refers to the inadequate or nonexistent availability of power sources and equipment infrastructure specifically designed to handle various types of waste.	[7, 22, 28]

	EC1	**Effect to local economy**	**This barrier refers to negativities such as loss of employment and income, increase in operating costs and decrease in competitiveness within the scope of zero waste management. This barrier tends to create potential negative impacts on the local economy.**	**[14, 23, 27, 32]**
Economical Barriers	EC2	High investment cost	Installing a zero-waste management system usually involves large initial expenditures, which include purchasing specialized machinery, setting up facilities for recycling and processing, and building a functional waste collecting and transportation network.	[7, 11]; Expert's suggestion
	EC3	Less economic value recovered	This barrier relates to the low recovery value of materials in zero waste management procedures from an economic standpoint. Zero waste management techniques may encounter difficulties in maintaining their financial viability due to the low economic value of recycled materials.	[11, 12, 22]
	EC4	Operation and maintenance cost	This criterion has a rapid effect on the process's efficiency. prices for operations and maintenance, payroll, incentives, supply and material prices, insurance, and audits are all included.	[15, 17, 28]
Environmental Barriers	EN1	Lack of public focus on environmental issues	The barrier originates from the general public's disregard to environmental issues. The efficiency of recycling and waste reduction initiatives might be hampered in zero waste management procedures by society's disregard for sustainability ideals and environmental obligations.	[11, 23, 26]
	EN2	Lack of green practices for recycling	This barrier expresses the inadequacy of measures such as the adoption of environmentally friendly technologies, ensuring energy efficiency, encouraging green packaging practices, and adopting green recycling processes.	[11, 12, 22]; Expert's Suggestion
	EN3	Effect of environment	This criterion evaluates the environmental effects of waste management techniques. It encompasses elements such as emission impact, effects on biodiversity, noise, and visual pollution.	[7, 23]; Expert's Suggestion
	EN4	The difficulty of recovering energy	This criterion expresses the challenges encountered in the process of converting non-recyclable waste materials into usable heat, power, or fuel using recycling processes.	[7, 22]; Expert's Suggestion

4 Case study

ZWM in Turkey's reverse logistics networks is crucial for enhancing environmental sustainability and economic efficiency. However, Turkey faces challenges such as legislative barriers, limited expertise, inadequate recycling facilities, and infrastructure. Overcoming these barriers is vital for environmental protection and economic gain, supporting Turkey's long-term development goals. This case study prioritizes the barriers to ZWM adoption in Turkey's reverse logistics and assesses waste management plans based on these barriers' importance with Fuzzy AHP method.

4.1 Preliminaries of Fuzzy-AHP

Unlike classical AHP, fuzzy AHP considers inconsistent data and decision-maker uncertainty by using linguistic values composed of specific intervals as opposed to precise numerical values [13]. The Extent Analysis method developed by Chang (1996) has been employed in numerous decision-making problems where the Fuzzy AHP method is utilized. According to Chang's Extent Analysis method, extent analysis is applied individually for each criterion and alternative. Thus, a set of m extent analysis values is

obtained for each criterion. These values are represented as follows [5].

$M_{gi}^1, M_{gi}^2, \ldots, M_{gi}^m \quad i = 1, 2, \ldots, n$ here, all $M_{gi}^j \quad (j = 1,2, \ldots, m)$ represent triangular fuzzy numbers.

The pairwise comparisons of the criteria in the hierarchy are conducted using fuzzy numbers. After obtaining matrices through pairwise comparisons, Chang's extent analysis is explained with the following steps:

Step 1: The value of the fuzzy synthetic extent for each criterion is defined as:

$$S_i = \sum_{j=1}^{m} M_{gi}^j \bigotimes \left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^j\right]^{-1} \quad (1)$$

here, the values are calculated using Eq. 2 and 3:

$$\sum_{j=1}^{m} M_{gi}^j = \left(\sum_{j=1}^{m} l_j \, , \sum_{j=1}^{m} m_j \, , \sum_{j=1}^{m} u_j\right) \quad (2)$$

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^j\right]^{-1} = \left(\frac{1}{\sum_{j=1}^{m} u_i} \, , \frac{1}{\sum_{j=1}^{m} m_i} \, , \frac{1}{\sum_{j=1}^{m} l_i}\right) \quad (3)$$

Step 2: The degree of possibility is calculated. The degree of possibility for M2= (l2, m2, u2) ≥ M1= (l1, m1, u1) is defined as follows:

$$V(M_2 \geq M_1) = \begin{cases} 1, & m_2 \geq m_1 \\ 0, & l_1 \geq u_2 \\ \dfrac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, & diğer\ durumlar \end{cases} \quad (4)$$

here, the intersection point is found for both cases. To compare M1 and M2, both V(M2 ≥ M1) and V (M1 ≥ M2) values are required.

Step 3: At this step, all V values are calculated for each criterion, compared, and the minimum value obtained for each criterion is taken. The minimum values obtained for all criteria are combined to obtain the weight vector:

$$W' = \left(d'(A_1), d'(A_2), \ldots, d'(A_n)\right)^T \quad (5)$$

Step 4: Via normalization, the normalized weights vectors are:

$$W = \left[\frac{d'(A_1)}{\sum_{i=1}^{n} d'(A_i)} \, , \frac{d'(A_2)}{\sum_{i=1}^{n} d'(A_i)}, \ldots, \frac{d'(A_n)}{\sum_{i=1}^{n} d'(A_i)}\right]^T \quad (6)$$

4.2 Fuzzy-AHP Application

A survey form consisting of a 5-point scale is created for the comparison of main criteria and sub-criteria. Decision-maker is requested to compare the survey forms. The appropriate fuzzy numerical variables are changed after survey forms based on linguistic variables are compared.

To solve the problem using the Fuzzy AHP method, relevant templates have been created in Excel. The steps of the method are applied by entering the corresponding formulations, which generate the desired outcomes.

The overall priority weights are obtained by multiplying the weights of main criteria and sub-criteria as shown in Table 2.

Table 2. The overall priority weights of the barriers.

Criteria	Main-criteria Weight	Sub-criteria Weight	Overall Weight
Organizational	0,234983		
Lack of motivation and reward		0,30	0,0705
Lack of organizational structure and support		0,30	0,0705
Lack of qualified labor		0,16	0,0376
Resistance to change		0,24	0,0564
Legal	0,234983		
Illegal recycling market		0,4328199	0,1017
Lack of enforced law, legislation, and directives for EoL products		0,4953	0,1164

Lack of government supportive policies for RL practices		0,0718	0,0169
Social	0,098116		
Lack of awareness		0,00	0,0000
Effect on human health		1,00	0,0981
Infrastructural	0,20589485		
Lack of data		0,11	0,0226
Lack of infrastructure facility		0,11	0,0226
Lack of availability of technology & information system		0,33	0,0679
Lack of power and equipment infrastructure based on waste type		0,23	0,0474
Environmental	0,14653036		
Lack of public focus on environmental issues		0,23	0,0337
Lack of green practices for recycling		0,23	0,0337
Effect of environment		0,39	0,0571
The difficulty of recovering energy		0,14	0,0205
Economic	0,07949221		
Effect to local economy		0,39	0,0310
High investment cost		0,33	0,0262
Less economic value recovered		0,12	0,0095
Operation and maintenance cost		0,16	0,0127

5 Conclusions

The barriers to ZWM in reverse logistics network for Turkey and their relative weights have been determined in this study. Initially, expert opinions were used to generate pairwise comparison matrices. Fuzzy AHP stages were then used to assess the criteria both internally and externally. The results obtained from the applied case study are as follows:

- The sub-criteria "Lack of enforced law, legislation, and directives for EoL products" and "Illegal recycling market" under the legal main criterion constitute a significant portion of the evaluation. Strengthening existing legal regulations and enforcement, especially for the effective management of electronic waste, may be necessary. Additionally, it is crucial to establish more effective control mechanisms to combat the illegal recycling market.
- Sub-criteria "Lack of motivation and reward" and "Lack of organizational structure and support" under the organizational main criterion highlight organizational barriers. In this context, developing motivation and reward systems to encourage employees to focus on environmental sustainability, along with strengthening organizational structure and support mechanisms, is important.
- Determining the most crucial among the main criteria may be challenging, but the Legal and Organizational criteria seem to stand out. Focusing on these criteria may form an effective strategy for zero waste management.

Based on these results, for the effective implementation of ZWM practices in reverse logistics, it is essential to focus more on legal and organizational barriers and implement strategic decisions accordingly. Additionally, from a holistic perspective, it is observed that the "Effect on human health" criterion stands out as one of the most crucial barriers.

There are several ways to overcome the barrier of "Effect on human health." Firs of all, it is crucial to carry out a thorough risk assessment to determine the effects of waste on health. It is essential to plan educational activities for employees and the public to raise awareness and improve awareness of the processes for handling waste. Establishing secure storage and processing facilities, implementing

environmental monitoring systems, and utilizing eco-friendly technologies can minimize the adverse health effects of waste management.

This developed framework is easily can be applied to any region in the world, given that the barriers are revised because each country has its unique dynamics. Prioritizing the criterion weights with other Multi-Criteria Decision-Making (MCDM) techniques, like Fuzzy TOPSIS and Fuzzy VIKOR, may be essential for future research to validate the approach. Furthermore, performing a comprehensive analysis of changes in criterion weights could help researchers in understanding the significance and effects of zero waste management in reverse logistics.

References

[1] Adeleke, O.J., Olukanni, D.O. (2020). Facility location problems: models, techniques, and applications in waste management, Recycling 5(10), 1-20.

[2] Apaydın, Ö., Akçay Han, G. S. (2023). Analysis of Municipal Solid Waste Collection Methods Focusing on Zero-Waste Management Using an Analytical Hierarchy Process. Sustainability, 15(17), 13184.

[3] Awan, M., A., Ali, Y. (2019). Sustainable modeling in reverse logistics strategies using fuzzy MCDM Case of China Pakistan Economic Corridor. Management of Environmental Quality: An International Journal, 30(5), 1132-1151.

[4] Ayçin, E., Kayapinar Kaya, S. (2021). Towards the circular economy: Analysis of barriers to implementation of Turkey's zero waste management using the fuzzy DEMATEL method. Waste Management & Research, 39(8), 1078-1089.

[5] Bayındır, S. 2021. Bulanık Çok Kriterli Karar Verme ile Temizlik Firması Seçimi. Master Thesis, Sakarya University.

[6] Cegan, J.C., Filion, A.M., Keisler, J.M., Linkov, I. (2017). Trends and applications of multi-criteria decision analysis in environmental sciences: literature review. Environment Systems and Decisions 37(2), 123-133.

[7] Chakraborty, S., Saha, A. K. (2022). A framework of LR fuzzy AHP and fuzzy WASPAS for health care waste recycling technology. Applied Soft Computing, 127, 109388.

[8] Chan, F.T.S., Chan, M.H., Tang, N.K.H., "Evaluation methodologies for technology selection", Journal of Materials Processing Technology, 107(1-3), 330-337, 2000.

[9] Chung S, Zhang C. "An evaluation of legislative measures on electrical and electronic waste in the People's Republic of China." Waste Management, 2011; 31: 2638–2646.

[10] Ghisolfi, V., Chave, G., Siman, R., Xavier, L.H., 2017. System dynamics applied to closed loop supply chains of desktops and laptops in Brazil: A perspective for social inclusion of waste pickers. Waste Management, 60, 14-31.

[11] Gupta, A. K. (2022). Prioritizing Barriers for Reverse Logistics of Lubricating Oils using Fuzzy AHP. Proceedings of The 2022 IEEE International Conference on Industrial Engineering and Engineering Management, 7-10 December, 1233-1236.

[12] Jindal A, Sangwan SK. "Development of an interpretive structural model of barriers to reverse logistics implementation in Indian Industry." In: Proceedings of the 18th CIRP International Conference on Life Cycle Engineering, Germany. 2011. p. 435–48. May 2nd–4th.

[13] Kahraman, C. 2008. Fuzzy Multi-Criteria Decision Making: Theory and Applications with Recent Developments. Springer, 16: 1-18.

[14] Karagoz, S., Deveci, M., Simic, V. and Aydin, N. (2021), "Interval type-2 Fuzzy ARAS method for recycling facility location problems", Applied Soft Computing, 102, 107107.

[15] Kheybari, S., Kazemi, M., & Rezaei, J. (2019). "Bioethanol Facility Location Selection Using Best-Worst Method." Applied Energy, 242, 612-623.

[16] Kilic, H. S., Kalender, Z. T., Solmaz, B., İseri, D. (2023). A two-stageMCDMmodel for reverse logistics network design of waste batteries in Turkey. Applied Soft Computing, 143, 110373.

[17] Korucu, M. K., & Erdagi, B. (2012). "A criticism of applications with multi-criteria decision analysis that are used for the site selection for the disposal of municipal solid wastes." Waste Management, 32(12), 2315-2323.

[18] Kurth, M.H., Larkin, S., Keisler, J.M., Linkov, I. (2017). Trends and applications of multicriteria decision analysis: use in government agencies. Environment Systems and Decisions 37(2), 134-143.

[19] Lee, Y., Hu, J., & Lim, M. K. (2021). "Maximising the circular economy and sustainability outcomes: an end-of-life tyre recycling outlets selection model." International Journal of Production Economics, 232, 107965.

[20] Oliveira, J.D., Oliveira Neto, J.F., Silva, M.M., Santos, S.M., 2020. E-waste mistakenly disposed of as recyclable waste: A case study from Brazil. Clean-Soil Air Water.

[21] Patil SK, RaviKant. "A fuzzy AHP-TOPSIS framework for ranking the solutions of knowledge management adoption in supply chain to overcome its barriers." Expert Systems with Applications, 2014; 41: 679–93.

[22] Prakash, C., Barua, M. K. (2015). Integration of AHP-TOPSIS method for prioritizing the solutions of reverse logistics adoption to overcome its barriers under fuzzy environment. Journal of Manufacturing Systems 37(2015) 599–615.

[23] Puviarasu, M., Asokan, P., Sherif, S. U. (2022). A STEEP based hybrid multi-criteria decision-making model for the evaluation of battery recycling plant location. Journal of Advances in Management Research, 20(2), 234-264.

[24] Radhi, A., R., Burhan, A. B. 2022. Multi-Criteria Optimization fors Governmental Projects Priority Ranking Depending on Fuzzified Experts' Opinion using Hygiene Approach. Journal of Engineering 28(12):113-124

[25] Schoeman, Y., Oberholster, P., Somerset, V. (2021). A Zero-Waste Multi-Criteria Decision Support Model for the Iron and Steel Industry in Developing Countries: A Case Study. Sustainability, 13(5), 2832.

[26] Sherif, S.U., Asokan, P., Sasikumar, P., Mathiyazhagan, K., and Jerald, J. (2022), "An integrated decision-making approach for choosing the location of a battery recycling plant within a sustainable environment," Journal of Cleaner Production, Volume 330, 129784.

[27] Simic, V., Karagoz, S., Deveci, M., & Aydin, N. (2021). "Picture fuzzy extension of the CODAS method for multi-criteria vehicle shredding facility location." Expert Systems with Applications, 175, 114644.

[28] Song, Q., Wang, Z., & Li, J. (2013). "Environmental performance of municipal solid waste strategies based on LCA method: a case study of Macau." Journal of Cleaner Production, 57, 92-100.

[29] Taelman, S.E. Tonini, D. Wandl, A. Dewulf, J. (2018). A Holistic Sustainability Framework for Waste Management in European Cities: Concept Development. Sustainability, 10, 2184.

[30] Tunacan, T., Nizam, T., Tezcan, B. (2019). Tersine Lojistik Bakış Açısı ile Katı Atık Yönetiminin İstatistiksel Değerlendirmesi: Karabük Bölgesi Demir Çelik Sektör Analizi. Academic Platform Journal of Engineering and Science, 8(1), 41-48.

[31] Zaman, A.; Ahsan, T. (2020). Zero-Waste: Reconsidering Waste Management for the Future, 1st ed.; Routledge: New York, NY, USA.

[32] Zarbakhshnia, N., Soleimani, H., & Ghaderi, H. (2018). "Sustainable third-party reverse logistics provider evaluation and selection using fuzzy SWARA and developed fuzzy COPRAS in the presence of risk criteria." Applied Soft Computing, 65, 307-319.

[33] Zhang, X., Liu, C., Chen, Y., Zheng, G. Chen, Y. (2022). Source separation, transportation, pretreatment, and valorization of municipal solid waste: A critical review. Environ. Dev. Sustain., 24, 11471–11513.

[34] Zhao, W., Yang, D., Duan, Y. (2022). Measuring Zero-Waste City Performance of a Coal Resource-Based Area in China with MCDM Approach. Mathematical Problems in Engineering, 2022, 8479760.

[35] Zolfani, S. H., Görener, A., Toker. K. (2023). A hybrid fuzzy MCDM approach for prioritizing the solutions of resource recovery business model adoption to overcome its barriers in emerging economies. Journal of Cleaner Production 413(2023), 137362. Devlet Planlama Teşkilatı. "Devlet Yardımlarını Değerlendirme Özel İhtisas Komisyonu Raporu". Devlet Planlama Teşkilatı, Ankara, Türkiye, 2681, 2004.

[36] T.C. Gıda Tarım ve Hayvancılık Bakanlığı. "Türk Gıda Kodeksi Et ve Et Ürünleri Tebliği". http://www.mevzuat.gov.tr/Metin.Aspx?MevzuatKod=9.5.16821&MevzuatIliski=0&sourceXmlSearch=et%20%C3%BCr%C3%BCnleri (21.12.2014).

Popüler Makine Öğrenimi Yöntemlerinin Türkçe Şiirlerde Şair Tanıma Problemi Üzerinde Karşılaştırılması

Aykhan BAGHİROV[1], Ali SEBETCİ[2]*

[1]*OSTİM Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği Yüksek Lisans Programı, Ankara, TÜRKİYE*

[2] *OSTİM Teknik Üniversitesi, Mühendislik Fakülte, Bilgisayar Mühendisliği Bölümü, Ankara, TÜRKİYE*

Özet

Metin madenciliğinin önemli uygulama alanlarından biri olan yazar ve(ya) şair tanıma çalışmaları, bir şiirin önceden belirlenmiş şairler arasında kime ait olduğunu, her şairin kendine özgü tarzı tespit ederek tahmin etme algoritmalarıyla ilgilenir. Bu çalışmada, Lojistik Regresyon, Destek Vektör Makineleri, Gaussian Naive Bayes ve Stokastik Gradyan İnişi gibi klasik, Çok Katmanlı Algılayıcı gibi yarı klasik yöntemlerin yanı sıra Ekstra Rastgele Ağaçlar gibi topluluk öğrenme yöntemleri kullanarak Türkçe şiirlerin şairlerini belirlemeye yönelik deneyler yapıldı. Açık kaynaklardan elde edilen ve yaklaşık 120 bin şiir içeren işlenmemiş veri seti temizlenip normalleştirildikten sonra seçilen 11 şairin 300'er şiirini kapsayan yeni ve dengeli bir veri seti oluşturuldu. Modellerin eğitim ve test aşamalarında kullanılan bu veri setinde her bir şair için toplam 20 bin civarında kelime mevcuttur. Şiirlerin Terim Frekansı – Ters Belge Frekansı (TF-IDF) vektörleri unigram kelime belirteçleri kullanılarak oluşturuldu. Elde edilen sonuçlar test edilen yöntemler arasında en başarılı olanların %89 doğruluk oranıyla Lojistik Regresyon, Stokastik Gradyan İnişi ve Ekstra Rastgele Ağaçlar, en fazla hata yapan yöntemin ise %71 doğruluk oranıyla Gaussian Naive Bayes olduğunu göstermektedir. Bu çalışmanın literatüre katkısı söz konusu yöntemlerin yeni bir veri seti üzerinde daha geniş bir hiper-parametre ızgarasında test edilmiş olmasıdır.

Anahtar Kelimeler: *Makine Öğrenmesi, Doğal Dil İşleme, Metin Madenciliği, Yazar/Şair Tanıma*

Comparison of Popular Machine Learning Methods on the Poet Recognition Problem in Turkish Poems

Abstract

Author and (or) poet recognition studies, one of the key application areas of text mining, deal with algorithms that identify who a poem belongs to among predetermined poets by detecting each poet's distinct style. In this study, experiments were conducted to identify Turkish poets using classical methods such as Logistic Regression, Support Vector Machines, Gaussian Naive Bayes, and Stochastic Gradient Descent, semi-classical methods such as Multilayer Perceptron, and ensemble learning methods such as Extra Random Trees. After cleaning and normalizing the raw data set received from free sources, which had roughly 120 thousand poems, a new and balanced data set was generated, featuring 300 poems from 11 selected authors. This data set contains around 20,000 words for each poet and is utilized in the model's training and testing phases. Unigram word tokens were used to build term frequency-inverse document frequency (TF-IDF) vectors for poems. The results show that the most successful methods among the evaluated methods are Logistic Regression, Stochastic Gradient Descent, and Extra Random Trees, all with an accuracy rate of 89%. Gaussian Naive Bayes has the highest error rate at 71%. The contribution of this study to the literature is that these methods have been tested on a new dataset over a broader hyperparameter grid.

Keywords: *Machine Learning, Natural Language Processing, Text Mining, Author/Poet Recognition*

*Contact email: ali.sebetci@ostimteknik.edu.tr

1 Giriş

Günümüzde veri bilimi alanında büyük bir öneme sahip olan metin madenciliği, yapılandırılmamış veya yarı yapılandırılmış büyük veri kümelerinden anlamlı bilgiler çıkarmak, anlamsal ilişkileri keşfetmek ve örüntüleri tanımlamak için kullanılan önemli bir araç haline gelmiştir [1, 2]. Metin madenciliği, eğitim verilerine dayalı olarak belirli bilgileri çıkarmak veya sınıflandırmak amacıyla metin analizi modelleri oluşturmak için makine öğrenimi (ML) ve doğal dil işleme (NLP) tekniklerini kullanır. İnternet üzerinde içerik oluşturucuların tanımlanması, intihal tespiti ve telif hakkı ihlallerinin önlenmesi, görüş tespiti yaklaşımları [3] ve siber suçlar hukuku [4] gibi gereksinimler, yazarlık/şairlik tanımlamasına duyulan ilgiyi arttırmıştır. Metin madenciliğinin bir alt dalı olan yazar veya şair tanıma çalışmalarında, edebi eserlerin yazarlık veya şairlik tarzlarını analiz etmek ve belirli bir tarzı diğerlerinden ayıran özellikleri tanımlamak için çeşitli makine öğrenmesi teknikleri kullanılır [5].

Yazar tanımlamanın en zor ve en önemli yönü, yazarın yazım stilini en iyi temsil eden temel özelliklerin belirlenmesidir. Bu konuda birçok araştırma yapılmış ve çeşitli çözüm önerileri sunulmuştur [6]. En önemli özelliklerin çıkarılabilmesi, yazarlığın doğru şekilde belirlenmesini sağlayabilir. Yazar tanımada başlıca özellikler, sözcüksel [7], morfolojik [8], içeriğe özgü [9] ve stilometriktir [10]. Ayrıca, NLP metin sınıflama ve veri madenciliği faaliyetlerinde, metnin özelliklerini kelime yerleştirme yöntemiyle tespit etmeye yönelik çeşitli yaklaşımlar vardır [11, 12]. Kelime yerleştirme yöntemi, bilhassa ML algoritmalarının sınıflama verimini arttırmak için kullanılan bir kelime vektörü veri tabanı sağlar [13]. Daha önce yapılmış başka birçok araştırmada da kelime hazinesi zenginliği ölçümleri, söz dizim özellikleri, işlevsel kelime frekansları, karakter n-gram frekansları, gizli anlamsal analiz (LSA) ve Kelime Çantası (BOW) gibi muhtelif özellikler sunulmuştur [14, 15, 16].

Bu çalışmada, Türkçe şiirlerin şairlerini belirleme problemi üzerine odaklanıldı. Bu problem, veri bilimi ve metin madenciliği alanındaki araştırmacılar için ilginç ve zorlu bir zemin sunmaktadır. Şiirler, dilin estetik ve duygusal yönlerini yansıtan karmaşık yapılar içerir ve her şairin kendine özgü bir tarzı vardır. Dolayısıyla, bir şiirin hangi şaire ait olduğunu belirlemek, makine öğrenimi algoritmaları için bir zorluk oluştururken, aynı zamanda metin analizi ve sınıflandırma konularında heyecan verici bir araştırma alanı sunar.

Çalışmada klasik makine öğrenimi yöntemleri kullanılarak Türkçe şiirlerin şairlerini belirlemeye yönelik deneyler gerçekleştirildi. Lojistik Regresyon (LR), Destek Vektör Makineleri (SVM), Gaussian Naive Bayes (GNB) ve Stokastik Gradyan İnişi (SGD) gibi klasik ML yöntemlerinin yanı sıra yarı klasik bir yöntem olarak Çok Katmanlı Algılayıcı (MLP) ile topluluk öğrenme yöntemleri Ekstra Rastgele Ağaçlar (ERT) ve Aşırı Gradyan Artırma (XGBoost) incelendi. Veri hazırlama aşamasında gereken Türkçe morfoloji analizi Zemberek [17] kütüphanesi kullanılarak yapıldı. Elde edilen sonuçlar, Türkçe şiirlerin şairlerini belirleme probleminde makine öğrenimi tekniklerinin etkili bir şekilde kullanılabileceğini göstermektedir. Bu çalışmanın literatüre katkısı söz konusu yöntemlerin yeni bir veri seti üzerinde daha geniş bir hiper-parametre ızgarasında test edilmiş olmasıdır.

2 Literatür Taraması

Yakın zamanda, metin sınıflandırma için makine öğrenimi kullanımına ilişkin, tarama makaleleri [18, 19, 20, 21], karşılaştırmalı analizler [22], Urduca, Arapça, Moğolca ve Hintçe gibi belirli dillere odaklanan makaleler [23, 24, 25, 26, 27], ML algoritmalarının birleştirilmesine dayalı özellik çıkarımı, sıra ve ağaç yapısına dayalı LSTM'ler (Uzun Kısa Süreli Bellek) aracılığıyla ilişki sınıflandırması ve derin karar ağacı sınıflandırmasına sahip çok özellik tabanlı yapay zeka gibi belirli algoritmaları uygulayan makaleler [28, 29, 30], sınıflama performansı [31] ve çerçevelerine [32] yönelik makaleler de dahil olmak üzere birçok çalışma yayınlanmıştır. Ancak Türkçe bağlamında metin sınıflandırmasına ilişkin literatür sınırlıdır.

Amasyalı ve Diri [33], SVM, NB ve RF (Rastgele Orman) algoritmalarını, n-gram yaklaşımını kullanarak, belge türü, yazarın adı ve yazarın ırkı açısından araştırdılar. Yazar tanıma görevinde en başarılı yöntemin NB, belge türü ve yazarın ırkını tanıma görevinde en iyi çalışan algoritmanın ise SVM olduğu buldular. Güran ve ark. [34] n-gram yaklaşımıyla metin sınıflandırmak için DT-J48, K-En Yakın Komşu (K-NN) ve Bayes olasılıksal sınıflandırıcıları kullandı. Her üç teknikte de en yüksek performansı unigram kelimeler elde etti.

Uysal ve Günal [35] İngilizce ve Türkçe haber ve e-posta verilerinde veri hazırlamanın metin sınıflandırmasına etkisini araştırdı. Araştırmaya göre İngilizce bitişken bir dil değilken Türkçe bitişken dillerin yaygın örneklerinden biridir. Sınıflandırmayı

SVM, ölçümleri ise Micro-F1 algoritmalarını kullanarak gerçekleştirdiler. Hazırlık aşaması ile durak kelime temizliğinin, özelliklerin çıkarılması ve seçilmesi kadar kritik olduğu sonucuna vardılar. Bu araştırma aynı zamanda dil veya alandan bağımsız olarak küçük harf değişikliği gibi bazı hazırlık işlemlerinin gerekli olduğunu, bazılarının ise çalışmanın diline ve alanına göre entegre edilmesi gerektiğini ileri sürmektedir.

Yıldırım ve Yıldız [36] metin sınıflandırmasında NB algoritması kullanarak standart kelime çantası yaklaşımı ile yapay sinir sistemini karşılaştırdı. Bulguları, durak sözcüklerin temizlenmesi ile morfolojik analizin sonuç üzerinde çok az etkisi olduğunu, özellik seçme ile Ki-Kare ve Bilgi Kazanımı kullanılmasının önemli olduğunu gösterdi. Türkçe gibi bitişken dillerde veri hazırlama genellikle zaman alıcı olduğu için Kuyumcu ve ark. [37] çalışmalarında veri hazırlığı gerektirmeyen hızlı metin sınıflandırma (fastText) yöntemini kullandı. Facebook'un fastText kelime yerleştirme tabanlı analizörü, veri hazırlama gereksinimini ortadan kaldırdı. Bu modeli NB, K-NN ve J48 tekniklerini kullanarak test ettiler. Sonuçlara göre en iyi performansı Multinomial NB sınıflandırıcısı verdi ve yazarlar veri hazırlığına gerek bırakmayan fastText algoritmasının tutarlılık açısından diğer yöntemlere göre önemli bir avantajı olduğu sonucuna vardılar.

Çoban ve ark. [38], Türk kullanıcı profillerinden elde edilen halka açık Facebook verileri üzerinde duygu analizi yapmak için Derin Öğrenme (DL) kullandılar. Tekrarlayan Sinir Ağları (RNN) en yüksek doğruluğu elde etti. Doğru ve ark. [39], Türkçe haber metinlerini Doc2vec yaklaşımını kullanan bir DL yöntemiyle sınıflandırmayı önerdi. Konvolüsyonel Sinir Ağlarını (CNN) NB, SVM, RF ve GNB algoritmaları gibi klasik ML yöntemleri ile karşılaştırdılar ve önerilen modelde en iyi sonuca CNN algoritmasının ulaştığını gördüler. Zulqarnain ve ark. [40] Türkçe soruları sınıflandırmak için word2vec metodolojisiyle birlikte üç farklı DL algoritması, Gated Recurrent Unit, LSTM ve CNN kullandılar. Word2vec stratejilerinin, kimi DL algoritmalarının performansında önemli bir etkiye sahip olduğunu gözlemlediler.

Bektaş [41], dört popüler ML algoritmasının (SVM, NB, LR ve LR'nin SVM ile entegrasyonu) sınıflandırma performansını karşılaştırdı. Sonuçlar, metin sınıflandırma algoritmalarının başarısının özellik çıkarma teknikleriyle ilgili olduğunu ve SVM'nin, unigram özellik haritaları kullanan diğer makine öğrenimi yöntemlerinden daha iyi performans gösterdiğini bildirdi. En iyi sonuçları SVM-LR entegrasyonu üretti. Bozyiğit ve ark. [42], çalışmalarında TF-IDF ve word2vec özellik çıkarma algoritmaları kullandı ve LR, NB, K-NN, SVM, RF ve XGBoost sınıflandırıcılarından gelen sonuçları karşılaştırdı. En güçlü teknik, ağırlıklı TF-IDF değeri ile en yüksek F1-ölçü puanına ulaşan XGBoost oldu. TF-IDF yöntemiyle karşılaştırıldığında word2vec tabanlı ML, F1 ölçümü açısından zayıf performans gösterdi. Ayrıca TF-IDF algoritmasıyla üretilen özellikler Ki-Kare yaklaşımıyla alt özellik kümelerine optimize edildiğinde XGBoost'un F1-ölçüsünün 2 puan arttığı görüldü.

Eminağaoğlu [43], veri toplama ve metin sınıflandırmada uygulanabilen herhangi bir yöntem veya modelde kullanılabilecek, belli bazı küresel İngilizce ve Türkçe veri kümelerinde doğruladığı bir benzerlik ölçüsü oluşturdu. Karasoy ve Ballı [44], ML ve DL yaklaşımlarını istenmeyen Türkçe metinleri filtrelemek için kullandı ve bu sınıflandırmada CNN tekniğinin doğruluk oranının en yüksek olduğunu gördü.

Köksal ve Yılmaz [45], NB, LR, K-NN, SVM ve RF gibi ML teknikleriyle Türkçe metinlerin sınıflandırılması için özel olarak eğitilmiş bir BERT modelini kullandı. Sonuçlar, önerdikleri tekniğin daha önce gerçekleştirilen haber sınıflandırma deneylerinden daha iyi F1 skoru ürettiğini gösterdi. Yıldız [46] metin sınıflandırma başarısını artırmak için veri dengesizliği sorununu ele alan bir veri dağıtım algoritması sundu. Önerilen algoritma, LSTM kullanılarak, 15 gruba ayrılmış geniş bir Türkçe veri seti üzerinde değerlendirildi ve doğrulukta yaklaşık 3,5 puan, F1 puanında ise üç puandan fazla artış ürettiği görüldü.

Yülüce ve Dalkılıç [47], SVM , NB, MLP, LR, SDG, ERT ve XGBoost yöntemlerini unigram ve bigram kelimelerde uyguladı ve her bir kelime türü için sırasıyla SGD ve LR algoritmalarını en başarılı sınıflandırma yöntemi olarak tespit etti. Alzoubi ve ark. [48], veriyi normalleştirme, derlemdeki terimleri morfolojik bakımdan temel biçimlerine göre analiz etme, çok tekrarlanan kelime gruplamalarını tespit ederek durak sözcükler listesi oluşturma ve bunları metinden temizleme, k-ortalamalar tekniğini kullanarak tekrarlanan kategorileri basitleştirme gibi yöntemlerle veri hazırlama aşamasına öncelik verdi. Veri setini SVM, NB, LSTM, RF ve LR yöntemleriyle eğitti ve çeşitli stratejilerin, veri hazırlığı öncesi ve sonrası etkinliğini değerlendirdi.

3 Materyal ve Metot

3.1 Veri Seti

İnternetin açık kaynaklarından elde edilen ve üzerinde herhangi bir ön işlem ya da filtreleme yapılmamış veri kümesi, 18.499 şairin 120.976 şiirinden/metninden ve toplam 12.277.371 kelimeden oluşmaktadır. Bir şaire ait en az ve en çok şiir sayıları sırasıyla 1 ve 1819'dur. En az kelime sayılı şiirlerde bir kelime, en çok kelime sayılı şiirde/metinde 312.342 kelime vardır. Bazı metinler şiir formunda değildir, geniş boşluklar, Türkçe olmayan karakterler ve kelimeler, doğru yazılmamış Türkçe kelimeler, rakamlar, tarihler, şair isimleri, şehir isimleri ve büyük harfler içermektedir.

3.2 Verinin Hazırlanması

Orijinal veriler, çoğu kez, bu çalışmada da olduğu gibi ön işleme tabi tutulmadan makine öğrenimi modeline gönderilemez. Zira işlenmemiş veri genellikle kirli/gürültülü, tutarsız ve eksiktir. Özellikleri yukarıda özetlenen veri setini temizlemek, normal hale getirmek ve dengelemek için şu adımlar atılmıştır:

- ✓ '\n' ve benzeri karakterlerin boşluk karakteri ile değiştirilmesi
- ✓ Türkçe harfler ile tek boşluk karakteri dışındaki tüm karakterlerin silinmesi
- ✓ Tüm harflerin küçük harfe çevrilmesi
- ✓ Türkçe olmayan şiirlerin/metnin silinmesi
- ✓ Tekrarlanan şiirlerin silinmesi
- ✓ Şiirlerde geçen il adlarının çıkarılması
- ✓ Şiirlerde geçen şair adlarının çıkarılması
- ✓ Durak sözcüklerinin temizlenmesi
- ✓ Zemberek Kütüphanesi ile morfoloji analizinin yapılması ve belirsizliklerin giderilmesi
- ✓ Veri setinde kaç farklı şair, her şairin kaç şiiri, her şiirde kaç kelime olduğunun tespit edilmesi
- ✓ Kelime sayısı 25'den az, 150'den çok olan şiirlerin silinmesi
- ✓ En az 300 şiiri ve şiirlerindeki toplam kelime sayısı en az 10 bin olan şairlerin seçilmesi
- ✓ Seçilen şairlerin 300'er şiirinden oluşan bir veri setinin hazırlanması

Bu aşamalardan sonra elde edilen ve ML modellerinde kullanılan veri setinde yer alan şairler ile her şairin 300 şiirindeki toplam ve ortalama kelime sayıları Tablo 1'de sunulmuştur.

3.3 Sınıflandırma Yöntemleri

SVM (Destek Vektör Makinesi): Veri noktalarını en iyi şekilde ayıran hiper-düzlem(ler)i bularak veriyi sınıflara ayırma tekniğidir. SVM, veri noktalarının sınıflarını mümkün olduğunca iyi ayıran bir hiper-düzlem bulmak için birkaç veri noktasına dayanarak bir "destek vektörü" seti oluşturur ve bu vektörler arasındaki marjı (yani, sınıflar arasındaki boşluğu) maksimize eder. SVM'nin avantajlarından biri, çeşitli veri setlerinde iyi performans gösterebilmesidir. Ayrıca, yüksek boyutlu veri kümelerinde de iyi çalışır ve sınıflandırma, regresyon ve ayırma gibi farklı problemlere uyarlanabilir. Bununla birlikte, SVM'nin karmaşık parametre ayarları ve büyük veri setleri için hesaplama gereksinimleri gibi bazı dezavantajları da vardır.

Tablo 1. ML Modellerinde Kullanılan Veri Seti

Şair	Toplam Kelime Sayısı	Ortalama Kelime Sayısı
Şair 1	18061	60
Şair 2	26638	89
Şair 3	14278	48
Şair 4	8056	27
Şair 5	11495	38
Şair 6	8304	28
Şair 7	15772	53
Şair 8	16722	56
Şair 9	20622	69
Şair 10	11069	37
Şair 11	22698	76

GNB (Gaussyen Naive Bayes): Naive Bayes, her özelliğin sınıfa bağlı olarak bağımsız olduğunu varsayan basit fakat güçlü bir olasılıksal sınıflandırıcıdır. Gaussyen Naive Bayes, özelliklerin normal (Gaussian) dağıldığını varsayar. Sınıflandırıcı, Bayes teoremini kullanarak bir veri noktasının belirli bir sınıfa ait olma olasılığını hesaplar. Gaussian Naive Bayes, özellikle hızlı ve verimli olması nedeniyle birçok pratik uygulamada kullanılır. Varsayımlarının basitliği ve hesaplamaların doğrudan oluşu, onu büyük veri setleri ve yüksek boyutlu veriler için uygun hale getirir. Bununla birlikte, özellikler arasında bağımsızlık varsayımı her zaman geçerli

olmadığından, performansı veri kümesine bağlı olarak değişebilir.

MLP (Çok Katmanlı Algılayıcı): Bir tür yapay sinir ağıdır ve genellikle klasik makine öğrenme yöntemleri ile daha karmaşık derin öğrenme modelleri arasında bir köprü olarak kabul edilir. Tam olarak klasik bir makine öğrenme yöntemi olmasa da çağdaş derin öğrenme modelleri kadar gelişmiş de değildir. MLP sınıflandırıcısı, doğrusal olmayan veriler üzerinde yüksek performans gösterebilen esnek ve güçlü bir algoritmadır. Özellikle görüntü tanıma, dil işleme ve diğer karmaşık görevlerde yaygın olarak kullanılır. Ancak derin öğrenme modellerine göre daha az karmaşıktır ve bazı problemler için sınırlı bir performansa sahiptir.

LR (Lojistik Regresyon): İstatistik ve makine öğreniminde yaygın olarak kullanılan, ikili sınıflandırma problemleri için geliştirilmiş bir regresyon modelidir. Temelde bağımlı değişkenin kategorik olduğu durumlarda (genellikle iki sınıflı: 0 ve 1) kullanılır. Model, bir gözlemin belirli bir sınıfa ait olma olasılığını tahmin eder. Lojistik regresyon, doğrusal regresyon modeline benzer bir yapıya sahiptir ancak çıktısı, lojistik (sigmoid) fonksiyon ile dönüştürülür. Bu dönüşüm, çıktı değerlerinin 0 ile 1 arasında olmasını sağlar.

SGD (Stokastik Gradyan İnişi): Özellikle büyük veri setleri ve yüksek boyutlu verilerle çalışırken hızlı ve etkili bir şekilde model eğitimi sağlayan bir makine öğrenimi algoritmasıdır. GD (Gradyan İnişi), bir hata fonksiyonunun minimumunu bulmak için kullanılan bir optimizasyon algoritmasıdır. Geleneksel GD, tüm veri seti üzerinden hesaplamalar yaparak parametreleri günceller. SGD ise her iterasyonda tüm veri seti yerine tek bir veri örneği veya küçük bir veri kümesi kullanarak parametre güncellemeleri yapar. Bu, algoritmanın daha hızlı ve daha az hesaplama kaynağı gerektirerek çalışmasını sağlar. Veri akışları üzerinde gerçek zamanlı olarak öğrenebilir. Uygulaması ve anlaşılması kolaydır. Öte yandan öğrenme oranı gibi hiper-parametrelerin dikkatli bir şekilde ayarlanması gerekir. Tek bir örnekle güncelleme yapıldığından, gürültülü verilerden etkilenebilir.

ERT (Ekstra Rastgele Ağaçlar): Karar ağaçlarına dayalı bir topluluk öğrenme yöntemidir ve Rastgele Orman modeline benzer, ancak bazı önemli farklara sahiptir. ERT, hem regresyon hem de sınıflandırma problemlerinde kullanılabilir. Model, bağımsız biçimde oluşturulan birçok karar ağacının tahminlerini birleştirir. Birleştirme, regresyon için genellikle ağaçların ortalamasını alarak, sınıflandırma için ise ağaçların oy çoğunluğunu kullanarak gerçekleştirilir. Ekstra rastgelelik, modelin aşırı öğrenme riskini azaltır. Genellikle hiper-parametre ayarına daha az ihtiyaç duyar ve varsayılan ayarlarla iyi performans gösterir. Fakat, büyük veri setlerinde yüksek bellek kullanımı gerektirir. Çok karmaşık ve doğrusal olmayan veri setlerinde, daha sofistike modeller kadar iyi performans göstermeyebilir.

XGBoost, Gradyan Artırma algoritmasının optimize edilmiş ve yüksek performanslı bir türüdür. Bu algoritma, genellikle karar ağaçlarını bir araya getirerek güçlü bir tahmin modeli oluşturur. XGBoost'un temel avantajlarından biri, yüksek tahmin gücüne sahip olmasıdır. Bu, karmaşık veri yapılarını ve ilişkilerini etkili bir şekilde modelleyebilir ve genellikle yüksek doğruluk sağlar. Veri setlerinde eksik değerlerin olduğu durumlarda bile etkili bir şekilde çalışabilir ve bu eksik değerleri doğrudan işleyebilir. Hız ve ölçeklenebilirlik de XGBoost'un önemli avantajları arasındadır. Büyük veri kümelerinde hızlı bir şekilde çalışabilir ve paralel hesaplama yetenekleri sayesinde büyük ölçekteki veri işleme gereksinimlerini karşılayabilir. Bunun yanı sıra XGBoost, yeni verilerde performansın düşmesine neden olabilecek şekilde eğitim verilerine aşırı uyum sağlanmasını engeller. Bunun için L1 ve L2 regülerleştirme tekniklerini kullanır.

4 Deneysel Analiz ve Sonuçlar

Çalışmada incelenen yedi sınıflandırma yönteminin farklı hiper-parametre kombinasyonları Scikit-learn kütüphanesinin Kapsamlı Izgara Arama tekniği ile test edildi. En iyi sınıflandırma doğruluğunu üreten kombinasyonlar seçildi. Sınıflandırıcılar, hiper-parametreleri ve elde edilen test sonuçları (ağırlıklı ortalamalar) Tablo 2'de verilmiştir.

Sınıflandırıcıların performansları değerlendirildiğinde en başarılı sonuçları (%89) LR, SGD ve ERT yöntemlerinin verdiği görülmektedir. Bunları sırasıyla benzer ama biraz daha düşük performanslara sahip olan MLP, XGBoost (%88) ve SVM (%87) yöntemleri izlemektedir. Çalışmada en düşük performansı gösteren (%71) Gauss NB yöntemi diğer yöntemlerden olumsuz anlamda ayrışmıştır.

Veri setine uygulanan bu makine öğrenimi yöntemlerinin doğruluk oranları, farklı algoritmaların veriye ne kadar uyduğunu gösterir. Bu sonuçları daha iyi anlamak ve yorumlamak için her bir algoritmanın çalışma prensibine ve veri setine nasıl uyum sağladığına bakmak gerekir.

Tablo 1. ML Yöntemlerinde Kullanılan Hiper-parametreler ve Elde Edilen Sonuçlar

Sınıflandırıcı	Taranan Hiper-parametre Izgarası	Seçilen Hiper-parametreler	Doğruluk	Kesinlik	Duyarlılık	F1 Puanı
SVM	C: [0.1, 1, 10, 100] gamma: [1, 0.1, 0.01, 0.001] kernel: ['rbf', 'linear']	C = 100 gamma = 0.1 kernel = rbf	0.87	0.88	0.87	0.87
GNB	var_smoothing: np.logspace(0, -9, num=100)	var_smoothing = 0.008111308307896872	0.71	0.71	0.71	0.71
MLP	hidden_layer_sizes: [(50,), (100,), (50, 50), (100, 50), (100, 100)] activation: ['tanh', 'relu'] solver: ['sgd', 'adam'] 'alpha': [0.0001, 0.05] 'learning_rate': ['constant', 'adaptive']	activation = relu alpha = 0.05 hidden_layer_sizes = (100,) learning_rate = constant solver = adam	0.88	0.88	0.88	0.88
LR	C: [0.1, 1, 10, 100] 'solver': ['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga'] 'penalty': ['l2'] 'multi_class': ['ovr', 'multinomial']	C = 100 multi_class = ovr penalty = l2 solver = liblinear	0.89	0.89	0.89	0.89
SGD	loss: ['hinge', 'log', 'modified_huber', 'squared_hinge', 'perceptron'] penalty: ['none', 'l2', 'l1', 'elasticnet'] alpha: [0.0001, 0.001, 0.01, 0.1, 1] learning_rate: ['constant', 'optimal', 'invscaling', 'adaptive'] eta0: [0.01, 0.1, 1] max_iter: [1000, 2000, 3000]	alpha = 0.0001 eta0 = 1 learning_rate = adaptive loss = modified_huber max_iter = 1000 penalty = l1	0.89	0.89	0.89	0.89
ERT	n_estimators: [50, 100, 200] max_features: ['auto', 'sqrt', 'log2'] min_samples_split: [2, 5, 10] min_samples_leaf: [1, 2, 4] bootstrap: [False, True]	bootstrap = False max_features = sqrt min_samples_leaf = 2 min_samples_split = 2 n_estimators = 200	0.89	0.89	0.89	0.89
XGBoost	n_estimators: [50, 100, 200] max_depth: [3, 6, 9] learning_rate: [0.01, 0.1, 0.2] subsample: [0.6, 0.8, 1.0] colsample_bytree: [0.6, 0.8, 1.0] gamma: [0, 0.1, 0.2]	n_estimators = 200 max_depth = 3 learning_rate = 0.1 subsample = 0.8 colsample_bytree = 0.6 gamma = 0.1	0.88	0.88	0.88	0.88

Gauss NB algoritması, sınıflar arasındaki bağımsız değişkenler varsayımına dayanır ve her bir özelliğin normal dağılıma sahip olduğunu varsayar. %71'lik doğruluk, bu varsayımların mevcut veri için tamamen geçerli olmadığını ve bu nedenle GNB'nin performansının diğer algoritmalara göre daha düşük olduğunu söylemektedir. Veri setindeki özellikler bağımsız değilse veya normal dağılım göstermiyorsa, GNB zayıf performans gösterecektir.

SVM, sınıflandırmak için bir hiper-düzlem kullandığı ve sınıflar arasındaki en geniş marjini bulduğu için genellikle iyi performans gösterir. %87'lik doğruluk, SVM'nin bu veri setindeki sınıfları iyi ayırabildiğini göstermekte, ancak %100'e ulaşamaması, verinin bazı karmaşıklıklar içerdiği veya sınıflar arasında tam bir ayrışma olmadığı anlamına gelmektedir.

MLP, sinir ağları kullanarak doğrusal olmayan ilişkileri öğrenir ve genellikle karmaşık veri setlerinde iyi performans gösterir. %88'lik doğruluk, MLP'nin bu veri setindeki karmaşık ilişkileri iyi öğrenebildiğini ve yüksek bir doğruluk oranına ulaştığını göstermektedir. Ancak, %89 doğruluk oranına ulaşan diğer modellerle karşılaştırıldığında hala biraz daha düşük kalması, modelin optimizasyonu veya aşırı uyum gibi faktörlerden kaynaklanıyor olabilir.

XGBoost modelinin %88'lik doğruluk oranı, bu belirli veri setinde etkili bir seçenek olduğunu göstermektedir. XGBoost'un bu yüksek doğruluk seviyesi, modelin veri setindeki örüntüleri yakalamada başarılı olduğunu ve doğru tahminler yaptığını gösterir. Kesinlik, duyarlılık ve F1 skorlarının da %88 olması, modelin sınıflandırma görevinde dengeli bir performans sergilediğini göstermektedir.

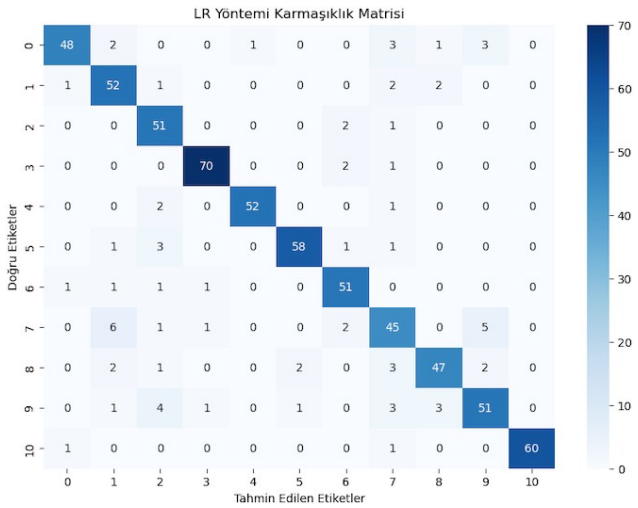
Lojistik regresyon, veriyi doğrusal olarak ayırmaya çalışır. Bu çalışmada elde edilen %89'luk doğruluk, lojistik regresyonun veri setinde sınıfları iyi ayırabildiğini ve veri setinin belirli bir düzeyde doğrusal ayrışabilirliğe sahip olduğunu göstermektedir.

SGD, özellikle büyük veri setlerinde ve yüksek boyutlu verilerde tercih edilen hızlı bir optimizasyon algoritmasıdır. %89'luk doğruluk, SGD'nin veriyi iyi öğrendiğini ve hiperparametrelerin iyi ayarlandığını göstermektedir. Bu sonuç, SGD'nin veriyi etkin bir şekilde işleyebildiği ve doğru sınıflandırma yapabildiği anlamına gelmektedir.

Genellikle karmaşık veri setlerinde iyi performans gösteren ERT, rastgele ağaçlar algoritmasına benzer ve çok sayıda karar ağacının ortalamasını alır. %89'luk doğruluk, ERT'nin bu veri setindeki sınıfları iyi ayırabildiğini ve yüksek varyans ile düşük yanlılık sağlayabildiğini göstermektedir.

Bu sonuçlar, model seçimi yaparken sadece doğruluk oranlarına değil, aynı zamanda veri setinin özelliklerine, modelin varsayımlarına ve problem tanımına da dikkat edilmesi gerektiğini göstermektedir. Her bir algoritmanın güçlü ve zayıf yönlerini anlamak, doğru model seçimi için kritik öneme sahiptir.

En başarılı üç yöntemden birisi olan LR'nin karmaşıklık matrisi Şekil 1'de verilmiştir. LR karmaşıklık matrisinde sınıf 3 (70 doğru) ve sınıf 10 (60 doğru) yüksek doğruluk oranlarına sahipken en fazla yanlış sınıflandırma, sınıf 7 (15 hata) ve sınıf 9'da (13 hata) görülmektedir. Bu yöntemde sınıf 7, en çok sırasıyla sınıf 1 ve 9 ile, sınıf 9 ise sınıf 2, 7 ve 8 ile karışmaktadır. Sınıfların karışması, modelin belirli sınıflar arasındaki sınırları iyi öğrenememiş olabileceğini göstermektedir. SGD karmaşıklık matrisinde sınıf 3 (70 doğru) ve sınıf 10 (61 doğru) en yüksek doğruluk oranlarına, sınıf 7 (15 hata), ve sınıf 9 (19 hata) ise en fazla hata oranlarına sahip sınıflardır. Bu sınıflandırmada sınıf 7, en çok sınıf 1 ve sınıf 9 ile sınıf 9 ise sınıf 2 ve 7 ile karışmaktadır. ERT karmaşıklık matrisinde sınıf 3 (73 doğru) ve sınıf 10 (62 doğru) en yüksek doğruluk, özellikle sınıf 7 (29 hata) en fazla hata oranlarına sahip sınıflardır. Bu sınıflandırmada sınıf 7 yine en çok sınıf 1 ve sınıf 9 ile karışmaktadır.



Şekil 1. Bir şekil örneği

Her üç model de görülen ortak özellikler diğer tüm sınıflardan en iyi ayrışan sınıfların sırasıyla sınıf 3

ve sınıf 10 olduğu, ayrıştırması en zor olan sınıfın ise en çok sınıf 9 ve sınıf 1 ile karışan sınıf 7 olduğudur. Bu sorunu çözmek için söz konusu üç sınıf için daha fazla veriye veya özellik mühendisliğine ihtiyaç duyulduğu söylenebilir.

5 Sonuç

Bu çalışmada Türkçe şiirlerde şair tanıma problemi incelenmiştir. İnternetin açık kaynaklarından elde edilen çok sayıda şiir temizlenmiş, ön işlemlerden geçirilmiş ve dengelenmiştir. Sonuçta 11 şairin 300'er şiirinden ve toplam 173.715 kelimeden oluşan veri setinin unigram özellikleri TF-IDF yöntemiyle tespit edilmiştir. Veri kümesinde terim sıklığına göre sıralanan ilk 7000 özelliği dikkate alan bir kelime dağarcığı oluşturulmuştur. Farklı alanlara uygulanan ve başarılı performanslar sergileyen SVM, GNB, SGD ve LR klasik makine öğrenmesi yöntemleri, ERT ve XGBoost topluluk öğrenmesi algoritmaları ve yarı klasik MLP yöntemi veri kümesi üzerinde eğitilmiştir. Hiper-parametre seçimi, belirlenen aralıklardaki tüm değer kombinasyonlarından oluşan ızgaralarda GridSearchCV kapsamlı arama tekniği kullanılarak gerçekleştirilmiş, tüm ML yöntemleri en iyi doğruluk oranına sahip hiper-parametrelerle eğitilmiştir.

Performans değerlendirme metriklerine göre en başarılı yöntemlerin LR, SGD ve ERT olduğu tespit edilmiştir. Bunları sırasıyla birer puan performans farklarına sahip MLP ve XGBoost ile SVM yöntemleri takip etti. En zayıf performansı GNB gösterdi. En başarılı üç modelin ürettiği sonuçlar da görülen ortak karakter, diğer tüm sınıflardan en iyi ayrışan sınıfların sırasıyla sınıf 3 ve sınıf 10 olduğu, ayrıştırması en zor olan sınıfın ise en çok sınıf 9 ve sınıf 1 ile karışan sınıf 7 olduğudur. Bu çalışmanın literatüre katkısı söz konusu yöntemlerin yeni bir veri seti üzerinde daha geniş bir hiper-parametre ızgarasında test edilmiş olmasıdır.

Çalışmanın devamında TF-IDF yerine Word2Vec gibi alternatif özellik çıkarma modellerinin denenmesi, belli bir doğruluk oranına ulaşmak için gereken veri boyutunun tespit edilmesi ve BERT ve derin sinir ağlarına dayalı daha modern yöntemlerin test edilmesi planlanabilir.

Kaynaklar

[1] Abbasi A, Javed AR, Iqbal F, Jalil Z, Gadekulla TR, Kryvinska N. "Authorship identification using ensemble learning," *Scientific Reports*, 12, 9537, 2022.

[2] Khan SA, Asad M, Asif H, Ali A, Jamil MA. "Author Identification Using Machine Learning," *Journal of Computing & Biomedical Informatics*, 2024.

[3] Ziani A, Azizi N, Guiyassa YT. "Combining random subspace algorithm and support vector machines classifier for Arabic opinions analysis," *Advanced Computational Methods for Knowledge Engineering*, 175-184, 2015.

[4] Igbal F, Binsalleeh H, Fung BC, Debbabi M., "A unified data mining solution for authorship analysis in anonymous textual communications," *Inf. Sci.*, 98-112, 2013.

[5] Benzebouchi NE. "Authors' writing styles are based authorship identification system using the text representation vector," *16th International Multi-Conference on Systems, Signals & Devices (SSD)*, 371-376, IEEE, 2019.

[6] Tamboli MS, Prasad RS. "Authorship analysis and identification techniques: A review," *Int. J. Comput. Appl.* 77, 11, 2013.

[7] Stamatatos E. "On the robustness of authorship attribution based on character n-gram features," *J. Law Policy*, 21, 421-439, 2013.

[8] Zheng R, Li J, Chen H, Huang Z. "A framework for authorship identification of online messages: Writing-style features and classification techniques," *J. Am. Soc. Inform. Sci. Technol.*, 57, 378-393, 2006.

[9] Mohsen AM, El-Makky NM, Ghanem N. "Author identification using deep learning," *15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 898-903, IEEE, 2016.

[10] Sarwar R, Li Q, Rakthanmanon T, Nutanong S. "A scalable framework for cross-lingual authorship identification," *Inf. Sci.*, 465, 323-339, 2018.

[11] Benzebouchi NE, Azizi N, Aldwairi M, Farah N. "Multi-classifier system for authorship verification task using word embeddings," *2nd International Conference on Natural Language and Speech Processing (ICNLSP)*, 1-6, IEEE, 2018.

[12] Stein RA, Jaques PA, Valiati JF. "Ananalysis of hierarchical text classification using word embeddings," *Inf. Sci.*, 471, 216-232, 2019.

[13] Mikolov T, Yih W.-t. Zweig G, "Linguistic regularities in continuous space word representations," *Proceedings of the 2013 conference of the north american chapter of the*

association for computational linguistics: Human language technologies, 746-751, 2013.

[14] Stamatatos E. "A survey of modern authorship attribution methods," *Journal of the American Society for information Science and Technology*, 60(3), 583-556, 2009.

[15] Alhuqail NK. "Author Identification Based on NLP," *European Journal of Computer Science and Information Technology*, 9(1), 1-26, 2021.

[16] Fabien M, Villatoro-Tello E, Motlicek P, Parida S. "BertAA : BERT fine-tuning for Authorship Attribution," *Proceedings of the 17th International Conference on Natural Language Processing (ICON)*, 127-137, 2020.

[17] Akın AA, Akın MD. "Zemberek, an open source NLP framework for Turkic languages," *Structure*, 10, 1-5, 2007.

[18] Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao M. "J. Deep learning-based text classification: A comprehensive review," *ACM Comput. Surv. (CSUR)*, 54, 1-40, 2021.

[19] Altınel B, Ganiz MC. "Semantic text classification: A survey of past and recent advances," *Inf. Process. Manag*. 54, 1129-1153, 2018.

[20] Kadhim AI. "Survey on supervised machine learning techniques for automatic text classification," *Artif. Intell. Rev.*, 52, 273-292, 2019.

[21] Li Q, Peng H, Li J, Xia C, Yang R, Sun L, Yu PS, He L. "A survey on text classification: From traditional to deep learning," *ACM Trans. Intell. Syst. Technol. (TIST)*, 13, 1-41, 2022.

[22] Hartmann J, Huppertz J, Schamp C, Heitmann M. "Comparing automated text classification methods," *Int. J. Res. Mark.* , 36, 20-38, 2019.

[23] Akhter MP, Jiangbin Z, Naqvi IR, Abdelmajeed M, Fayyaz M. "Exploring deep learning approaches for Urdu text classification in product manufacturing," *Enterp. Inf. Syst*, 16, 223-248, 2022.

[24] El Rifai H, Al Qadi L, Elnagar A. "Arabic text classification: The need for multi-labeling systems.," *Neural Comput. Appl.* 34, 1135-1159, 2022.

[25] Okuno S, Asai H, Yamana H. "A challenge of authorship identification for ten-thousand-scale microblog users," *2014 IEEE International Conference on Big Data (Big Data)*, 52-54, 2014.

[26] Damiran Z, Altangerel K. "Author Identification - An Experiment Based on Mongolian Literature Using Decision Trees," *2014 7th International Conference on Ubi-Media Computing and Workshops*, 186-189, 2014.

[27] Digamberrao KS, Prasad RS. "Author Identification using Sequential Minimal Optimization with rule-based Decision Tree on Indian Literature in Marathi," *Procedia Computer Science*, 132, 1086-1101, 2018.

[28] Ajitha P, Sivasangary A, Rajkumar RI, Poonguzhali S. "Design of text sentiment analysis tool using feature extraction based on fusing machine learning algorithms," *J. Intell. Fuzzy Syst.*, 40, 6375-6383, 2021.

[29] Dai Y, Guo W, Chen X, Zhang Z. "Relation classification via LSTMs based on sequence and tree structure.," *IEEE Access*, 6, 64927-64937, 2018.

[30] Yuvaraj N, Chang V, Gobinathan B, Pinagapani A, Kannan S, Dhiman G, Rajan AR. "Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification," *Comput. Electr. Eng.*, 92, 07186, 2021.

[31] Yadav BP, Ghate S, Harshavardhan A, Jhansi G, Kumar KS, Sudarshan E. "Text categorization performance examination using machine learning algorithms," *Proceedings of the IOP Conference Series: Materials Science and Engineering, Warangal, India, 9–10 October 2020*, 022044, 2020.

[32] Mohammed A, Kora R. "An effective ensemble deep learning framework for text classification," *J. King Saud Univ.-Comput. Inf. Sci.*, 34, 8825-8837, 2022.

[33] Amasyalı MF, Diri B. "Automatic Turkish text categorization in terms of author, genre and gender," Natural Language Processing and Information Systems. NLDB 2006. *Lecture Notes in Computer Science*, 3999, 221-226, 2006.

[34] Güran A, Akyokuş S, Bayazıt NG, Gürbüz MZ. "Turkish text categorization using n-gram words," *Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications (INISTA 2009), Trabzon, Turkey, 29 June–1 July 2009, IEEE,* 369-373, 2009.

[35] Uysal AK, Gunal S. "The impact of preprocessing on text classification," *Inf. Process. Manag.*, 50, 104-112, 2014.

[36] Yıldırım S, Yıldız TA. "A comparative analysis of text classification for Turkish language," *Pamukkale Univ. J. Eng. Sci.*, 24, 879-886, 2018.

[37] Kuyumcu B, Aksakallı C, Delil S. "An automated new approach in fast text classification (fastText): A case study for Turkish text classification without pre-processing," *Proceedings of the 3rd International Conference on Natural Language Processing and Information Retrieval, ACM, Tokushima, Japan, 28–30 June 2019*, 1-4, 2019.

[38] Çoban Ö, Özel SA, İnan A. "A. Deep learning-based sentiment analysis of Facebook data: The case of Turkish users," *Comput. J.*, 64, 473-499, 2021.

[39] Doğru HB, Tilki S, Jamil A, Hameed AA. "Deep learning-based classification of news texts using doc2vec model," *Proceedings of the 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), Riyadh, Saudi Arabia, 6–7 April 2021; IEEE: Riyadh, Saudi Arabia, 2021,* 91-96.

[40] Zulqarnain M, Alsaedi AKZ, Ghazali R, Ghouse MG, Sharif W, Husaini NA, "A comparative analysis on question classification task based on deep learning approaches," *PeerJ Comput. Sci.,* 7, e570, 2021.

[41] Bektaş J. "Detection of economy-related Turkish tweets based on machine learning approaches," Data Mining Approaches for Big Data and Sentiment Analysis in Social Media; El-Latif, A.A.A., Ed.; IGI Global: Hershey, PA, USA, 2022, 171-195.

[42] Bozyiğit F, Doğan O, Kılınç D. "Categorization of customer complaints in food industry using machine learning approaches," *J. Intell. Syst. Theory Appl.*, 5, 85-91, 2022.

[43] Eminağaoğlu M. "A new similarity measure for vector space models in text classification and information retrieval," *J. Inf. Sci.*, 48, 463-476, 2022.

[44] Karasoy O, Ballı S. "Spam SMS detection for Turkish language with deep text analysis and deep learning methods," *Arab. J. Sci. Eng*, 47, 9361-9377, 2022.

[45] Köksal Ö, Yılmaz EH. "Improving automated Turkish text classification with learning-based algorithms," *Concurr. Comput. Pract. Exp.*, 34, e6874, 2022.

[46] Yıldız B. "Efficient text classification with deep learning on imbalanced data improved with better distribution," *Turk. J. Sci. Technol.*, 17, 89-98, 2022.

[47] Yülüce İ, Dalkılıç F. "Author Identification with Machine Learning Algorithms," *International Journal of Multidisciplinary Studies and Innovative Technologies*, 6(1), 45-50, 2022.

[48] Alzoubi YI, Topcu AE, Erkaya AE. "Machine Learning - Based Text Classification Comparison: Turkish Language Context," *Appl. Sci.*, 13, 9428, 2023.

Temel Düzey Veri Analizi Yöntem Seçiminde ChatGPT-4o'nun Performansı

Rabia AKTAŞ[1]*, Naci MURAT[2]

[1] *Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Akıllı Sistemler Mühendisliği Bölümü, Samsun, Türkiye*

[2] *Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Samsun, Türkiye*

Özet

Günümüzde, araştırmacılar bilimsel çalışmalarında yapay zekadan giderek daha fazla yardım almaktadır. Alınan bu yardımlar veri analizi süreçlerinin hızlanmasına ve kolaylaşmasına olanak tanımaktadır. Fakat doğruluğundan emin olmadan yürütülen veri analizi süreci, yapay zeka araçlarının bilimsel çalışmalarda kullanılmasında ciddi tehdit oluşturmaktadır. Bu çalışma, ChatGPT-4o'nun temel düzeyde istatistiksel veri analizi sorularına verdiği yanıtların doğruluğunu ve tutarlılığını incelemeyi amaçlamaktadır. Bu amaçla, yapay zekaya veri analizi ile ilgili 10 farklı soru 26 kez sorulmuş ve yanıtlar uzman iki gözlemci tarafından değerlendirilmiştir. Sorular üç temel alanı ölçmeye yönelik hazırlandı. Bunlar; verilerin normalliğinin test edilmesi ve tanımlayıcı istatistiklerin sunulması, bağımlı ve bağımsız grup analizi ve ilişki/uyum testleridir. Araştırma sonucunda, ChatGPT-4o'nun tanımlayıcı istatistiklerin sunulması, iki grup karşılaştırması, kategorik ve sürekli değişkenler arasındaki ilişkinin incelenmesi konularında yüksek doğruluk ve tekrarlanabilirlik sağladığını, normallik testi ve tek grup analizi için ise eksik veya yanlış yanıtlar verdiğini gösterdi. Yapay zekanın aynı soruya vermiş olduğu cevaplar arasındaki uyumun değerlendirilmesinde, Percent Agreement (0,741), Brennan and Prediger (0,611) ve Gwet's AC (0,650) katsayılarının "Yüksek" düzeyde uyum gösterdiğini ortaya koymuştur. Cohen/Conger's Kappa (0,502), Scott/Fleis Kappa (0,501) ve Krippendorff's Alpha (0,503) ise "Orta" uyum düzeyinde kalmıştır. Tüm uyum ölçütleri istatistiksel olarak anlamlı bulunmuştur ($p < 0,05$). Bu bulgular, yapay zeka destekli araçların veri analizi sürecinde kullanılabilirliğini ve güvenilirliğini değerlendirmek için önemli ipuçları sunmaktadır. ChatGPT-4o'nun doğruluk ve tutarlılığının tanımlayıcı istatistiklerin sunulması, iki grup karşılaştırması ve kategorik değişkenler arasındaki ilişkinin incelenmesi belirli sorular konuları için yüksek olması, bu tür araçların daha geniş kapsamlı ve derinlemesine analizler için nasıl geliştirilebileceğine dair öneriler sunmaktadır. Çalışma, yapay zeka destekli veri analizinin araştırma süreçlerine entegrasyonu konusunda yeni perspektifler sunmakta ve gelecekteki çalışmalar için önemli bir temel oluşturmaktadır.

***Anahtar Kelimeler**: Temel Düzey Veri Analizi, ChatGPT-4o, Büyük Dil Modeli*

*Contact email: rabiaaktas60@gmail.com

Performance of ChatGPT-4o in Selection of Basic Data Analysis Methods

Abstract

In today's research, scientists are increasingly relying on artificial intelligence to assist in their scientific studies. This assistance facilitates and accelerates the data analysis processes. However, conducting data analysis without ensuring accuracy poses a significant threat to the use of AI tools in scientific research. This study aims to examine the accuracy and consistency of ChatGPT-4o's responses to basic statistical data analysis questions. For this purpose, 10 different data analysis-related questions were asked to the AI 26 times, and the responses were evaluated by two expert observers. The questions were designed to measure three main areas: testing data normality and presenting descriptive statistics, conducting dependent and independent group analyses, and performing correlation/compatibility tests. The research results showed that ChatGPT-4o provided high accuracy and reproducibility in presenting descriptive statistics, comparing two groups, and analyzing the relationship between categorical and continuous variables, but it gave incomplete or incorrect answers for normality testing and single-group analysis. The consistency of the AI's responses to the same questions was evaluated using the Percent Agreement (0.741), Brennan and Prediger (0.611), and Gwet's AC (0.650) coefficients, all of which demonstrated a "Substantial" level of agreement. Cohen/Conger's Kappa (0.502), Scott/Fleis Kappa (0.501), and Krippendorff's Alpha (0.503) showed a "Moderate" level of agreement. All agreement measures were statistically significant ($p < 0.05$). These findings provide valuable insights into the usability and reliability of AI-supported tools in the data analysis process. The high accuracy and consistency of ChatGPT-4o in presenting descriptive statistics, comparing two groups, and examining relationships between categorical variables for specific questions suggest how such tools can be developed for more comprehensive and in-depth analyses. This study offers new perspectives on the integration of AI-supported data analysis into research processes and establishes an important foundation for future studies.

Keywords: *Basic Data Analysis, Chat GPT-4o, Large Language Model*

1 Giriş

Günümüzde doğal dil işleme ile geliştirilen yapay zekâ teknolojileri, birçok alanda kullanılmaktadır. Bu teknolojilerden biri olan Büyük Dil Modelleri, insan dilini anlamak amacıyla tasarlanmış makine öğrenimi modelleridir. Kitap, makale gibi kaynaklardan dilin yapısını ve kelime kalıplarını öğrenerek, kullanıcıların taleplerine yanıt üretebilirler. Büyük dil modelleri, büyük miktardaki metinsel veriler üzerinden eğitilerek metin üretimi, soru yanıtlama gibi insani görevleri yüksek doğrulukla gerçekleştirebilirler[1].

Doğal dil işleme teknolojilerini kullanan bir yapay zeka modeli olan ChatGPT, çeşitli görevlerde kullanıcılarla etkileşim kurmak için tasarlanmış olup, bilgi sağlama, yazı yazma ve metin analizi gibi geniş bir yelpazede işlevler sunmaktadır. ChatGPT, doğal dil işleme alanında devrim yaratarak, bağlam ve tutarlılıkla insan benzeri metinler üreten, insan-yapay zeka etkileşiminde yeni olanaklar sunan son teknoloji bir dil modelidir[2].

Modelin büyük miktarda metin verisi üzerinde önceden eğitilmiş olması, ChatGPT'ye dilin inceliklerini anlama, karmaşık ve belirsiz bağlamlarda bile son derece doğru yanıtlar üretme yeteneği kazandırmıştır[3]. ChatGPT'nin geliştiricisi olan OpenAI yapay zeka araştırma şirketi farklı yeteneklere sahip çeşitli modeller kullanmaktadır. Bu modeller GPT-3, GPT-3,5, GPT-4 ve GPT-4o gibi birbirinin yükseltilmiş versiyonlarıdır. Son sürüm olan ChatGPT-4o, daha doğru ve tutarlı yanıtlar üretebilmek için geniş bir veri tabanında eğitilmiştir ve görsel içerik üretimi hizmeti de sunmaktadır.

Gelişen teknoloji ile birlikte hayatımıza yapay zeka kavramı entegre olmuştur. Yapay zeka araçları hayatımızın olağan akışını değiştirmekte olup buna uyum sağlamak amacıyla neredeyse her sektör bu araçları kullanmaya başlamıştır. Özellikle geniş araştırmalar yapmak yerine sorularına pratik

çözümler arayanların başvurduğu ilk adres konumuna gelmiştir.

Farklı alanlarda, ChatGPT'nin modelin istemine verdiği cevapların doğruluğu ve tekrarlanabilirliği incelenmiştir. Freire ve arkadaşları [4] protez tedavisinde ChatGPT'nin performansını değerlendirdi ve hareketli dental protezler hakkında robota tekrarlı sorulan sorulara yanıtlar üretme konusunda ChatGPT'nin sınırlı bir yeteneğe sahip olduğunu gözlemlediler. Dolayısıyla ChatGPT'nin bir diş hekiminin yerini alamayacağı ifade edildi. Suárez ve arkadaşları [5], ChatGPT'nin endodontik sorulara verdiği yanıtların doğruluğunu ve tutarlılığını değerlendirerek, bu yapay zekanın klinik karar verme sürecinde diş hekimlerinin yerini alacak kadar yeterli olmadığını tespit ettiler.

Jang ve Lukasiewicz'in çalışmasında, ChatGPT'nin tutarlılık açısından güvenilirliği analiz edildi ve araştırma, modelin gelişmiş dil anlama ve akıl yürütme yeteneklerine sahip olmasına rağmen, mantıksal olarak tutarlı tahminler üretme konusunda sıkça yetersiz olduğunu ortaya koydular[6].

17 uzmanlık alanında hekimlerin oluşturduğu 284 tıbbi sorunun ChatGPT tarafından üretilen yanıtların doğruluk ve kapsamlılık açısından değerlendirildiği bir çalışmada [6], ChatGPT, sorulara büyük ölçüde doğru bilgiler üretti. Ancak yanlışlıkları düzeltmek ve doğrulama için daha fazla araştırma ve model geliştirmenin gerekli olduğu vurgulandı.

Bilimsel araştırmalarda da yapay zeka araçları sıkça kullanılmaya başlanmış ve bu kullanımın her alanda olduğu gibi artarak devam etmesi beklenmektedir. Bilimsel araştırmaların temel taşlarından biri veri analizidir. Araştırmacılar, çalışmalarının amacı doğrultusunda hipotezler kurar ve verilerin yapısına uygun yöntemler kullanarak bu hipotezleri test ederler. Ardından, elde ettikleri bulgulardan sonuçlar çıkarır ve bu sonuçlar üzerinden teoriler geliştirirler. Doğru ve güvenilir veri analizi araştırmanın geçerliliğini ve güvenilirliğini sağlar. Bilimsel araştırmalarda veri analizi yapmak için istatistiksel bilgiye ihtiyaç vardır.

Tümdengelim yoluyla veri analizi yöntemlerini hızlandırmak için ChatGPT'nin kullanıldığı bir çalışmada [7] GPT'nin görselleştirme programları sunarak veri analizini daha zahmetsiz hale getirdiği, ayrıca etkili kullanıldığında, yaratıcı hipotez seçiminde öğrenme sürecini iyileştirebileceği sonucuna varıldı.

Bu araştırmanın amacı, ChatGPT-4o'nun temel düzeyde istatistiksel veri analizi sorularına verdiği yanıtların doğruluk ve tutarlılığını değerlendirmektir. Bu çalışma, ChatGPT-4o'nun, araştırmacıların sıkça ihtiyaç duyduğu temel istatistiksel analizler konusunda ne kadar güvenilir ve doğru yanıtlar üretebildiğini incelemektedir. Aynı zamanda, yapay zeka destekli veri analizi süreçlerinin etkinliğini ve bu süreçlerin sınırlarını belirlemeyi amaçlamaktadır.

2 Materyal ve metod

Temel düzeyde veri analizi için kullanılan testler hakkında 10 soru hazırlandı. Her bir soru için ChatGPT-4o kullanılarak 26 tekrar ile 260 yanıt oluşturuldu. Oluşturulan yanıtlar, uzman iki gözlemci tarafından 3 dereceli Likert ölçeği kullanılarak değerlendirildi. Doğruluk, kullanıcı tarafından sorulan sorulara verilen doğru yanıtların tekrar edilme oranı olarak değerlendirildi. Tutarlılık, Brennan ve Prediger katsayısı, Cohen/Conger's Kappa, Scott/Fleiss' Kappa, Gwet's AC ve Krippendorff's Alpha katsayıları ve bunlara karşılık gelen %95 güven aralıklarıyla birlikte hesaplandı [8]. Tahmin edilen katsayılar, Gwet'in 2014'te önerdiği kıyaslama ölçeğine göre sınıflandırıldı. Analizler IBM SPSS V23 (Chicago, USA) ve Stata programının deneme sürümü kullanılarak yapıldı.

2.1 Güç analizi

Örneklem büyüklüğünü belirlemek amacıyla, PASS programının deneme sürümü kullanılarak güç analizi yapılmıştır. %95 güven aralığı (1-α) ile gerçekleştirilen analiz sonucunda, 10 soru için gerekli tekrar sayısı 26 olarak hesaplanmıştır[4].

3 Veri toplama araçları

Temel düzey veri analizinde kullanılan testlerin uygulanmasına yönelik olarak 10 soru hazırlandı (Tablo 1). Sorular, veri analizi süreçlerinde karşılaşılan yaygın senaryoları kapsayacak şekilde tasarlandı. Bu sorulan sorular verilerin normal dağılıma uygunluğunu, grup sayılarına göre hangi testin kullanılması gerektiğini, bağımlı veri yapılarından kullanılması gereken testleri, kategorik ve nicel değişkenler arasındaki ilişkileri sorgulayan testleri ve verileri bir referans değere göre karşılaştıran testleri sorgulamaya yönelik sorulardır. Her bir soru için ChatGPT-4o kullanılarak 26 tekrar ile yanıtlar oluşturuldu. Oluşturulan 260 yanıt, uzman bir gözlemci tarafından 3 dereceli Likert ölçeği kullanılarak değerlendirildi (Tablo 2).

Tablo 1. ChatGPT-4o tarafından yanıt üretilen sorular ve doğru cevaplar.

Soru	Cevap	Amaç
S1. Elimde 150 tane veri var bu verilerin normalliğine hangi test ile bakmalıyım?	C1. Kolmogorov-Smirnov Testi (n>50)	Tanımlayıcı istatistiklerin sunulması ve normal dağılıma uygunluk
S2. Analiz sonuçlarını tablolaştırırken veriler normal dağıldığında hangi tanımlayıcı istatistiği sunmalıyım, normal dağılmadığında hangi tanımlayıcı istatistiği sunmalıyım?	C2. Ortalama, Standart sapma/Ortanca, Minimum, Maksimum	
S3. Elimdeki verileri literatürdeki referans değerler ile karşılaştırmak istiyorum, hangi yöntemi kullanmalıyım?	C3. Tek Örnek t Testi/Runs Testi	Bağımlı ve bağımsız grup karşılaştırması
S4. Verileri iki gruba ayırarak gruplar arasında parametreler karşılaştıracağım, hangi testi seçmeliyim?	C4. Bağımsız Örnekler t Testi/Mann Whitney U Testi	
S5. Verileri 5 gruba ayırarak bir parametre açısından karşılaştırmak istiyorum, hangi testi kullanmalıyım?	C5. ANOVA/Kruskal Wallis Testi	
S6. Ölçüm sonuçlarını iki farklı zaman diliminde aldım ve sonuçları karşılaştırmak istiyorum. Hangi istatistiksel testi kullanmalıyım?	C6. Eşli İki Örnek t Testi/Wilcoxon Testi	
S7. İki gözlemci arasındaki uyumu incelemek istiyorum, hangi testi kullanmalıyım?	C7. Kappa Testi/ICC	Kategorik ve sürekli değişkenler arasındaki ilişkinin incelenmesi
S8. İki kategorik değişken arasındaki ilişkiyi incelemek için hangi istatistiksel testi kullanmalıyım?	C8. Ki-Kare Testi	
S9. Katılımcıların yaşları ile aldıkları hizmetten memnuniyetleri arasındaki ilişkiyi incelemek için hangi yöntemi kullanmalıyım?	C9. Pearson Korelasyon Katsayısı/Spearman's rho Korelasyon Katsayısı	
S10. Katılımcıların çalışma saatleri ile tükenmişlik düzeyleri arasındaki ilişkide yaşın da etkisi olduğunu düşünürsek, yaşın etkisini arındırarak bu ilişkiyi hangi yöntemle inceleyebilirim?	C10. Kısmi Korelasyon Katsayısı	

Tablo 2. Yanıtların değerlendirildiği likert ölçeği.

Uzman Derecelendirmesi	Açıklama
0 (Yanlış)	Verilen cevap tamamen yanlış veya soruyla ilgisizdir.
1 (Kısmen Doğru veya Eksik)	Cevap, konuyu bir dereceye kadar anladığını veya konuyla ilgili bilgi sahibi olduğunu gösteriyor, ancak önemli hatalar veya eksik unsurlar var. Tamamen yanlış olmasa da, cevap yeterince doğru veya eksiksiz değil, bu nedenle güvenilir veya uygun olarak kabul edilmez.
2 (Doğru)	Cevap tamamen doğru ve konuyu sağlam ve doğru bir şekilde anladığını gösteriyor. Tüm temel unsurlar doğru ve kapsamlı bir şekilde ele alınıyor.

4 Analiz sonuçları

ChatGPT-4o ile her bir soruya 26 tekrarlı olmak üzere 260 yanıt üretildi. Tablo 3, her bir soru için ChatGPT-4o yanıtlarının uzman değerlendirmesindeki doğru yanıt sıklığını ve tekrarlanabilirliğini göstermektedir. Tanımlayıcı istatistiklerin sunulması için verilen yanıtlar %100 tekrarlanabilirlik ve doğruluk gösterdi. 3 ve üzeri grup karşılaştırması ve sürekli değişkenler arasındaki ilişkinin incelenmesi için test seçimi sorularına verilen yanıtların %100 tekrarlanabilir olduğu görüldü. Bu tekrarlanabilirlik sorulara verilen yanıtların kısmen doğru veya eksik olduğu yönündeydi. Tekrarlanabilirlikte yanlış tekrar yüzdesi en yüksek olan sorunun, %30,8 ile örnek sayısının n>50 olduğu durumda tercih edilmesi gereken normallik testi sorusu olduğu görüldü. ChatGPT-4o, 7 soru için değişken derecelerde tekrarlanabilirlik gösterdi. Bu sorular; normal dağılıma uygunluğu, tek grup analizini, ikili grup karşılaştırmasını, bağımlı verilerin karşılaştırılmasını, uyum analizini ve kısmi korelasyonu incelemek amacıyla sorulan test seçimi sorularıdır.

Tablo 3. ChatGPT-4o yanıtları için uzman değerlendirmesinin dağılımı

Sorular	Yanlış		Kısmen doğru veya eksik		Doğru	
	n	%	n	%	n	%
S1	8	30,8	18	69,2	0	0,0
S2	0	0,0	0	0,0	26	100,0
S3	7	26,9	18	69,2	1	3,8
S4	1	3,8	4	15,4	21	80,8
S5	0	0,0	26	100,0	0	0,0
S6	0	0,0	25	96,2	1	3,8
S7	0	0,0	21	80,8	5	19,2
S8	0	0,0	8	30,8	18	69,2
S9	0	0,0	26	100,0	0	0,0
S10	0	0,0	11	42,3	15	57,7

ChatGPT-4o'nun ürettiği yanıtların uzmanlar tarafından tekrar değerlendirilmesindeki tutarlılık, önemli ile orta düzey arasında değiştiği görüldü (Tablo 4). Yanıtların tutarlılığı, Percent Agreement (0,741), Brennan and Prediger (0,611) ve Gwet's AC (0,650) katsayıları ile "Önemli" düzeyde uyum gösterirken, Cohen/Conger's Kappa (0,502), Scott/Fleis Kappa (0,501) ve Krippendorff's Alpha (0,503) katsayıları "Orta Düzey" uyum seviyesinde kalmıştır. Tüm uyum ölçütleri istatistiksel olarak anlamlı bulunmuştur ($p < 0,05$). Tüm tekrarlarda doğru yanıt oranı %33,5, kısmen doğru veya eksik yanıtların oranı %60,4 ve yanlış yanıt oranı %6,2 olarak bulundu. Tekrarlardaki tutarlılığın önemli ve orta düzey arasında değişmesi kısmen doğru veya eksik olan yanıtlardan kaynaklanmaktadır.

Tablo 4. Uzman değerlendirmelerine göre, ChatGPT-4o tarafından üretilen 10 sorunun 26 tekrarı için tutarlılık değerlendirme

	Korelasyon	S.Hata	(%95 GA)		Benchmark Skalası	Yanlış	Kısmen Doğru veya Eksik	Doğru
Percent Agreement	0,741	0,068	0,587	0,894	Önemli	%6,2	%60,4	%33,5
Brennan and Prediger	0,611	0,102	0,38	0,842	Önemli			
Cohen/Conger's Kappa	0,502	0,117	0,237	0,766	Orta Düzey			
Scott/Fleis Kappa	0,501	0,118	0,235	0,767	Orta Düzey			
Gwet's AC	0,650	0,099	0,425	0,874	Önemli			
Krippendorff's Alpha	0,503	0,118	0,236	0,769	Orta Düzey			

Benchmark Skalası: Zayıf <0,0; Az 0,0-0,2; Orta 0,2-0,4; Orta Düzey 0,4-0,6; Önemli 0,6-0,8; Neredeyse Mükemmel 0,8-1,0

5 Sonuçlar

Bu çalışma, ChatGPT-4o'nun temel düzey veri analizi yöntemleri konusundaki yanıtlarının doğruluğunu ve tutarlılığını değerlendirmeyi amaçlamıştır. Sonuçlar, ChatGPT-4o'nun yanıtlarının doğruluk ve tekrarlanabilirlik açısından değişkenlik gösterdiğini ortaya koymuştur.

Bu bulgular, ChatGPT-4o'nun tanımlayıcı istatistiklerin sunulması, iki grup karşılaştırması ve kategorik/sürekli değişkenler arasındaki ilişkinin incelenmesi konularında yüksek doğruluk ve tekrarlanabilirlik sağladığını göstermektedir. Ancak, normallik testi ve tek grup analizi için eksik veya yanlış yanıtlar verdiği tespit edilmiştir. ChatGPT-4o'nun temel düzey analizlerde dahi kısmen doğru veya eksik verilen yanıtlarda tutarlılık göstermesi, bu konuda yapay zekada ciddi problemler olduğunu ortaya koymaktadır.

Çalışma, yapay zeka destekli veri analizinin etkinliği ve güvenilirliği konusunda önemli ipuçları sunmaktadır. ChatGPT-4o'nun verinin dağılımını göz önünde bulundurarak tanımlayıcı istatistiklerin sunumu, iki grup karşılaştırmaları ve ilişki arayıcı testler konularında yüksek doğruluk oranları göstermesi, yapay zeka destekli veri analiz araçlarının daha geniş kapsamlı ve derinlemesine analizler için nasıl geliştirilebileceğine dair önemli bir temel oluşturmaktadır.

Kaynaklar

1. Kasneci, E., et al., *ChatGPT for good? On opportunities and challenges of large language models for education.* Learning and individual differences, 2023. **103**: p. 102274.
2. Brown, T., et al., *Language models are few-shot learners.* Advances in neural information processing systems, 2020. **33**: p. 1877-1901.
3. Radford, A., et al., *Language models are unsupervised multitask learners.* OpenAI blog, 2019. **1**(8): p. 9.
4. Freire, Y., et al., *ChatGPT performance in prosthodontics: Assessment of accuracy and repeatability in answer generation.* The Journal of Prosthetic Dentistry, 2024. **131**(4): p. 659. e1-659. e6.
5. Suárez, A., et al., *Unveiling the ChatGPT phenomenon: evaluating the consistency and accuracy of endodontic question answers.* International endodontic journal, 2024. **57**(1): p. 108-113.
6. Jang, M.E. and T. Lukasiewicz, *Consistency analysis of chatgpt.* arXiv preprint arXiv:2303.06273, 2023.
7. Shirota, Y., *Use of ChatGPT in Data Analysis by Deductive Reasoning.* Gakushuin Economic Papers, 2023. **60**(3): p. 179-195.
8. Gwet, K., *Handbook of inter-rater reliability.* Gaithersburg, MD: STATAXIS Publishing Company, 2001: p. 223-246.

Artificial Intelligence Supported ICD Coding: An Evaluation on Patients in the 0-15 Age Group

Müberra TERZİ KUMANDAŞ[1]*, Naci MURAT[1]

[1] *Ondokuz Mayıs University, Faculty of Engineering, Department of Industrial Engineering, Samsun, Türkiye*

Abstract

Large language models (LLMs) are artificial intelligence systems focusing on natural language processing using deep learning techniques. These models are often used in processes such as understanding, generating, and analyzing texts. Recently, these models have been used in the creation of decision support systems for issues such as medical diagnosis and analysis of clinical notes in healthcare. When patients apply to healthcare institutions, International Classification of Diseases (ICD) codes and summary epicrisis information for their complaints are manually entered into the system by doctors. ICD code is the general name given to the international classification system of patients. Doctors may select incomplete or generic codes while doing this. To improve this situation, we wanted to test whether the ICD codes of patient electronic health records (EHR) can be determined automatically. For this purpose, it was analyzed whether the ICD codes entered the system based on the complaints of patients aged 0-15 years who applied to the pediatrics polyclinic of a private hospital in Samsun/Turkey in the last year can be identified by artificial intelligence. 2422 data were used in the study, and since BDM contains information about ICD codes, all data was used as test data. In the study conducted using ChatGPT and Co-pilot, ICD codes entered the system by doctors in line with the summary epicrisis information and ICD codes obtained from artificial intelligence were compared. ChatGPT showed 36.3% accuracy in the accuracy matrix, and Co-pilot showed 46.7% accuracy. The results show that for AI-based systems to be a potential tool for determining ICD codes, physicians should write EHRs more comprehensively and indicate additional findings obtained from the examination in the records.

Keywords: *ICD-10, ChatGPT, Co-pilot, Artificial Intelligence, Large Language Models*

1 Introduction

Artificial intelligence technologies developed with natural language processing are used in many fields today. One of these technologies, Big Language Models (LMMs), is a machine learning model designed to understand human language. These models learn the structure of language and word patterns from sources such as books, articles, papers, and websites [1, 2]. Thus, when a prompt is entered into the model, they can generate responses to their users under the model's request. Language models such as ChatGPT, Co-pilot, Bing, and Notion are examples of applications in this field.

LLM, which has become increasingly used in recent years, can perform human tasks such as text production and question answering with high accuracy by being trained on large amounts of textual data [3]. However, to further increase accuracy, questions need to be asked correctly. In this context, the concept of "prompt engineering" has emerged to design the instructions that guide the answers obtained from LLMs and improve the response [4].

Prompt engineering is the general name for a method of communicating effectively with the language model to optimize the utility of the output from AI. A prompt consists of four elements: instruction, context, input data, and output indicator. Instruction is the task that directs the model's behavior, leading it to the desired output. Context provides background information on the model. It is additional information that helps the model generate more accurate and relevant responses. Input data is the question the model is asked to answer. It helps the model understand the task. The output indicator determines the type and format of the desired output [4].

Although LLM can be prompted with prompts, it can also answer texts without needing to be specially trained on a topic. It is used in many fields, such as education, finance, and IT. In

*Contact email: muberra.terzi@omu.edu.tr

healthcare, there are studies on various topics such as disease diagnosis, treatment recommendations, analysis of patient electronic records, clinical decision support systems, and automatic determination of ICD codes. This study, large amounts of medical data are processed, and healthcare professionals are supported in decision making. The processed medical data is generally obtained from electronic health records (EHR). When patients apply to healthcare institutions, their complaints and medical information regarding their illnesses are entered into EHRs by doctors. The information recorded in the EHRs is expressed in International Classification of Diseases (ICD) codes. However, doctors may make incomplete/general choices when selecting ICD codes. Incomplete selection indicates that the diagnoses entered do not precisely match the disease. On the other hand, making a general selection creates a difference between the patient's complaint and the ICD code, as it includes the findings of the doctor's examination. This situation harms the accuracy and quality of medical records.

In analyzing EHRs and ICD codes, LLM assists in making the correct diagnosis by interpreting disease symptoms and clinical notes. There are studies in this field using LLM. Huang, Yang [5] classified lung cancer pathology images using ChatGPT with 89% accuracy. The study revealed that the classification with artificial intelligence is better than traditional methods. They found that the misclassifications made by the AI were due to interpretation errors caused by a lack of pathological terminology and unclear cancer staging rules. The study reveals that artificial intelligence can support clinical decision-making by structuring unstructured health data. Liu, Wu [6] used ChatGPT to classify the diagnoses found in the EHRs of patients admitted to the hospital with retinal vascular occlusion. Accuracy was 70.47% for Chinese diagnoses and 80.05% for English diagnoses. The study revealed that LLMs can help the decision support system, but they have some errors that need improvement. Liao, Liu [7] analyzed texts generated by experts with ChatGPT. They found that classical text classification algorithms and ChatGPT's classification obtained similar results. They noted that LLMs are reliable and can be used in medicine, but there is still a difference between LLM and human-generated texts. Skalidis et al. [8] demonstrated that ChatGPT can accurately answer questions on licensing exams such as the United States Medical Licensing Examination.

Turunç Oğuzman, Yurdabakan [9] evaluated the performance of ChatGPT and BARD against the questions in the Dentistry Specialty Education Entrance Examination. They obtained a 35.7% accuracy rate from ChatGPT and a 38.9% from BARD. It was revealed that LLMs do not perform reliably on the exam, but as the developments in LLMs and question clarity increase, these models can be used in dentistry.

Scientists using LLM have found that, in general, these models can be used in the health field, but they still have some limitations and need improvement. This study examines whether artificial intelligence technologies can automatically determine the ICD codes of patients' EHRs. Patients aged 0-15 years admitted to the pediatrics polyclinic of a private hospital in Samsun/Turkey using data from were tested the accuracy of LLMs in classification. Using LLMs such as ChatGPT and Co-pilot, ICD codes entered by doctors were compared with ICD codes determined by LLMs.

The study consists of 4 main sections. Section 2 describes the theoretical background of the methods used in the study. Section 3 presents the results obtained. Section 4 summarizes the conclusion.

2 Material and Methods

The motivation of this study was to investigate whether ICD codes in EHR can be identified using LLM. In this context, ChatGPT and Co-pilot, two of the most widely used LLMs, were compared using seven different classification performance parameters.

2.1 Data Set

The data set used in this study belongs to patients in the 0-15 age group who applied to the pediatrics polyclinic of a private hospital in Samsun/Turkey within the last year. The ICD code and ICD code description entered the system based on gender, age, and complaints on admission to the hospital for 2422 patients were collected from the hospital information system. The distribution of data according to ICD codes is given in Table 1.

Table 1. Data distribution according to ICD code

ICD Code	Frequency	ICD Code	Frequency
J06	415	H66.9	42
J39	320	J01	39
J03	227	T79.9	35

R11	225	J05.0	33
R50.9	138	N39.0	32
R50	127	J04	26
R10.4	94	R05	22
J20	87	K59.0	21
M79.9	74	R14	21
J21	66	K52.9	16
A08.5	53	T78.4	15
K52	51	R21	15
J02	47	H67	13
M70	47	H10.9	12
K52.8	43	M79.8	12
H60	43	L50.9	11

J06: Acute upper respiratory tract infections, J39: Other diseases of the upper respiratory tract, J03: Acute tonsillitis, R11: Nausea and vomiting, R50.9: Fever, undefined, R50: Fever of other and unknown origin, R10.4: Other and unspecified abdominal pain, J20: Acute bronchitis, M79.9: Soft tissue disorder, unspecified, J21: Acute bronchiolitis, A08.5: Other specified intestinal infections, K52: Other noninfective gastroenteritis and colitis, J02: Acute pharyngitis, M70: Soft tissue disorders related to use, overuse and pressure, K52.8: Other specified noninfective gastroenteritis and colitis, H60: Otitis externa, H66.9: Otitis media, unspecified, J01: Acute sinusitis, T79.9: Unspecified early complication of trauma, J05.0: Acute obstructive laryngitis [croup], N39.0: Urinary tract infection, site not specified, J04: Acute laryngitis and tracheitis, R05: Cough, K59.0: Constipation, R14: Flatulence and related conditions, K52.9: Noninfective gastroenteritis and colitis, unspecified, T78.4: Allergy, unspecified, R21: Rash and other nonspecific skin eruption, H67: Otitis media in diseases classified elsewhere, H10.9: Conjunctivitis, unspecified, M79.8: Other specified soft tissue disorders, L50.9: Urticaria, unspecified

2.2 Classification Parameters

Classification algorithms have four metrics: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [10]. These metrics are calculated separately for each row. When evaluated in row and column matrix format, the rows contain the ICD codes of the referenced doctors, and the columns contain the ICD codes resulting from the assignments made by the LLM. TP, TN, FP, and FN values are obtained for each ICD code as follows. TP represents the frequency value obtained from the intersection of the ICD code in the column and the ICD codes in the row. TN refers to the sum of all frequencies except the sum of the values under the ICD code in the column and the sum of values in the row with the ICD code it matches in the row. FP is obtained by subtracting the frequency value in the cell where the ICD code in the column matches the ICD code in the row from the column ICD total frequency. FN is the value obtained by subtracting the cell frequency of the ICD code in the column corresponding to the ICD code in the row from the row ICD code total.

Classification performance is measured by precision, sensitivity, specificity, negative predictive value (NPV), F-score, and accuracy obtained from classification metrics [11]. In addition to these parameters, the Matthew Correlation Coefficient (MCC) is also calculated. MCC is a statistical tool used to evaluate the performance of classification models on unbalanced datasets. MCC takes a value between "-1" and "+1". "+1" indicates that the forecast is perfect, '0' indicates random guessing, and '-1' indicates a high discrepancy between the forecast and reality [12]. The classification parameters are given in Equation (1-8).

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

$$NPV = \frac{TN}{TN+FN} \quad (4)$$

$$F-score = \frac{2\times Sensitivity\times Precision}{Sensitivity+Precision} \quad (5)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Accuracy_{total} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i+FN_i} \quad (7)$$

$$MCC = \frac{TN*TP-FN*FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (8)$$

In this study, classification was performed using ChatGPT and Co-pilot.

2.3 Prompt Entered in Models

The prompt given to the LLM is given under four headings: instruction, context, input data, and output indicator. The data was given to the LLMs as an Excel table.

Instruction: To test whether ICD codes can be determined automatically according to the complaints of pediatric patients coming to hospitals, ICD codes should be defined according to the patients' diseases in the Excel table.

Context: Thousands of patients are admitted to pediatric polyclinics of hospitals every day. According to patient complaints, physicians assign ICD codes to EHRs indicating the disease. However, these codes may be incomplete or generalized due to hospital density, carelessness, etc. It is aimed to test the feasibility of the ICD code definition assigned by doctors by LLM.

Input Data: You are a specialist in Pediatrics. You need to enter the patient's ICD code into the hospital information system, considering the age, gender, and complaint of each patient.

Output Indicator: You should provide the ICD code to be assigned to each patient in a table format by adding a new column at the end of the table in the input data.

3 Results

Within the scope of the study, 2422 medical records of patients aged 0-15 years who visited the pediatrics polyclinic of a private hospital within one year were collected. The data set consists of medical records, including ICD codes and descriptions, entered the system based on gender, age, and hospitalization complaints. Table 2 gives a section of the data collected.

ChatGPT and Co-pilot were used to classify ICD Codes. The values obtained from the classification using ChatGPT are given in Table 2. The obtained accuracy value is 36.3%, F-score is 0.371, and MCC is 0.351. For the J06 ICD code, the precision of the model was 1.000, sensitivity 0.133, specificity 1.000, and overall accuracy 0.851. This means that all diagnoses to which the model assigned the J06 ICD code were correct, but it was only able to correctly identify 13.3% of people with the disease associated with the J06 code. Similarly, for the R11 code, the model's precision was 0.643, sensitivity 0.960, specificity 0.945, and overall accuracy 0.947. These results mean that 64.3% of the diagnoses to which the model assigned the R11 ICD code were correct, 96% of people with the disease associated with the R11 code were correctly identified, and 94.5% of people without the disease were correctly identified. The precision value is 1 for 13 of the 32 ICD codes assigned. This shows that the positive diagnoses of the model are entirely accurate for the relevant codes (J06, J39, R50, etc.). When the MCC values are analyzed, it is seen that the values of ICD codes such as R50 and R10.4 are close to 1; that is, the assignments made based on these codes are close to perfect fit. ICD codes J20, J21, and H60 are close to 0, indicating randomness in the assignments in these codes. In addition, since there is no value close to -1 in the MCC values obtained from the assignments made by ChatGPT, it is inferred that this model does not make misclassification with high fit.

The classification model using ChatGPT has high accuracy and specificity for some ICD codes but may perform poorly in other metrics, such as sensitivity and precision. The low precision rate for code R50.9 shows a high incorrect ICD code assignment rate. Similarly, the precision value could not be calculated in some ICD codes, such as M70 and H67, due to the lack of positive assignment. On the other hand, the high sensitivity for the R11 code means that the model successfully identifies people with the disease. ChatGPT performs differently in classification parameters for different ICD codes. For this, although ICD codes should be evaluated within itself, model performance in overall accuracy (36.3%) is low.

The values obtained from the classification using the co-pilot are given in Table 3. The obtained accuracy value is 46.7%, the F-score is 0.518, and the MCC value is 0.508. Although the Co-pilot gave more accurate results than ChatGPT for the ICD code classification of this dataset, the model performance is low in the health field. The precision value is 1 for 17 of the 32 ICD codes assigned. This shows that the positive diagnoses of the model are entirely accurate for the relevant codes (J39, J03, R50, etc.). Compared to ChatGPT, the number of ICD codes (J20, K52.8, J05.0, and H67) for which the precision value could not be calculated due to the lack of positive assignment in the classification with Co-pilot is higher.

For the J06 code, the precision of the Co-pilot model is 0.825, the sensitivity is 0.442, and the overall accuracy is 0.888, while in the ChatGPT model, these values are 1.000, 0.133, and 0.851, respectively. This suggests that the Co-pilot model is less reliable in making a correct diagnosis but can correctly diagnose a more significant proportion of people with the disease. Considering both classifications, the Co-pilot model performed more reliably, increasing true negatives and overall accuracy.

Table 2. ChatGPT Results

ICD Code	TP	TN	FP	FN	Precision	Sensitivity	Specificity	NPV	F-Score	Accuracy	MCC
J06	55	2007	0	360	1,000	0,133	1,000	0,848	0,234	0,851	0,335
J39	56	2102	0	264	1,000	0,175	1,000	0,888	0,298	0,891	0,394
J03	30	2194	1	197	0,968	0,132	1,000	0,918	0,233	0,918	0,341
R11	216	2077	120	9	0,643	0,960	0,945	0,996	0,770	0,947	0,760
R50.9	126	1511	773	12	0,140	0,913	0,662	0,992	0,243	0,676	0,276
R50	120	2295	0	7	1,000	0,945	1,000	0,997	0,972	0,997	0,971
R10.4	86	2318	9	9	0,905	0,905	0,996	0,996	0,905	0,993	0,901
J20	1	2331	4	86	0,200	0,011	0,998	0,964	0,022	0,963	0,040
M79.9	8	2348	0	66	1,000	0,108	1,000	0,973	0,195	0,973	0,324
J21	0	2355	1	66	0,000	0,000	1,000	0,973	*	0,972	-0,003
A08.5	6	2369	0	47	1,000	0,113	1,000	0,981	0,203	0,981	0,333
K52	40	2281	90	11	0,308	0,784	0,962	0,995	0,442	0,958	0,475
J02	2	2375	0	45	1,000	0,043	1,000	0,981	0,082	0,981	0,204
M70	0	2375	0	47	*	0,000	1,000	0,981	*	0,981	*
K52.8	30	2379	0	13	1,000	0,698	1,000	0,995	0,822	0,995	0,833
H60	0	2378	1	43	0,000	0,000	1,000	0,982	*	0,982	-0,003
H66.9	1	2380	0	41	1,000	0,024	1,000	0,983	0,047	0,983	0,153
J01	1	2383	0	38	1,000	0,026	1,000	0,984	0,050	0,984	0,159
T79.9	1	2387	0	34	1,000	0,029	1,000	0,986	0,056	0,986	0,168
J05.0	4	2389	0	29	1,000	0,121	1,000	0,988	0,216	0,988	0,346
N39.0	8	2389	1	24	0,889	0,250	1,000	0,990	0,390	0,990	0,468
J04	2	2394	2	24	0,500	0,077	0,999	0,990	0,133	0,989	0,193
R05	21	2014	386	1	0,052	0,955	0,839	1,000	0,098	0,840	0,201
K59.0	16	2400	1	5	0,941	0,762	1,000	0,998	0,842	0,998	0,846
R14	5	2401	0	16	1,000	0,238	1,000	0,993	0,385	0,993	0,486
K52.9	4	2360	46	12	0,080	0,250	0,981	0,995	0,121	0,976	0,132
T78.4	10	2406	1	5	0,909	0,667	1,000	0,998	0,769	0,998	0,777
R21	12	2403	4	3	0,750	0,800	0,998	0,999	0,774	0,997	0,773
H67	0	2409	0	13	*	0,000	1,000	0,995	*	0,995	*
H10.9	9	2409	1	3	0,900	0,750	1,000	0,999	0,818	0,998	0,821
M79.8	3	2410	0	9	1,000	0,250	1,000	0,996	0,400	0,996	0,499
L50.9	7	2411	0	4	1,000	0,636	1,000	0,998	0,778	0,998	0,797
TOTAL	**880**	**73640**	**1441**	**1543**	**0,379**	**0,363**	**0,981**	**0,979**	**0,371**	**0,363**	**0,351**

TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative, NPV: Negative Predictive Value, MCC: Matthew Correlation Coefficient; *: It is not calculable due to the value 0 assigned to the denominator.

Table 3. Co-Pilot Results

ICD Code	TP	TN	FP	FN	Precision	Sensitivity	Specificity	NPV	F-Score	Accuracy	MCC
J06	184	1967	39	232	0,825	0,442	0,981	0,894	0,576	0,888	0,552
J39	207	2102	0	113	1,000	0,647	1,000	0,949	0,786	0,953	0,784
J03	85	2195	0	142	1,000	0,374	1,000	0,939	0,545	0,941	0,593
R11	211	2089	108	14	0,661	0,938	0,951	0,993	0,776	0,950	0,763
R50.9	98	1914	370	40	0,209	0,710	0,838	0,980	0,323	0,831	0,322
R50	106	2295	0	21	1,000	0,835	1,000	0,991	0,910	0,991	0,909
R10.4	90	2303	25	4	0,783	0,957	0,989	0,998	0,861	0,988	0,860
J20	0	2335	0	87	*	0,000	1,000	0,964	*	0,964	*
M79.9	9	2348	0	65	1,000	0,122	1,000	0,973	0,217	0,973	0,344
J21	0	2355	1	66	0,000	0,000	1,000	0,973	*	0,972	-0,003
A08.5	1	2369	0	52	1,000	0,019	1,000	0,979	0,037	0,979	0,136
K52	4	2371	0	47	1,000	0,078	1,000	0,981	0,145	0,981	0,277
J02	13	2375	0	34	1,000	0,277	1,000	0,986	0,433	0,986	0,522
M70	15	2375	0	32	1,000	0,319	1,000	0,987	0,484	0,987	0,561
K52.8	0	2379	0	43	*	0,000	1,000	0,982	*	0,982	*
H60	0	2378	1	43	0,000	0,000	1,000	0,982	*	0,982	-0,003
H66.9	1	2380	0	41	1,000	0,024	1,000	0,983	0,047	0,983	0,153
J01	1	2382	1	38	0,500	0,026	1,000	0,984	0,049	0,984	0,111
T79.9	5	2387	0	30	1,000	0,143	1,000	0,988	0,250	0,988	0,376
J05.0	0	2389	0	33	*	0,000	1,000	0,986	*	0,986	*
N39.0	10	2390	0	22	1,000	0,313	1,000	0,991	0,476	0,991	0,556
J04	2	2396	0	24	1,000	0,077	1,000	0,990	0,143	0,990	0,276
R05	20	2143	257	2	0,072	0,909	0,893	0,999	0,134	0,893	0,239
K59.0	17	2401	0	4	1,000	0,810	1,000	0,998	0,895	0,998	0,899
R14	5	2401	0	16	1,000	0,238	1,000	0,993	0,385	0,993	0,486
K52.9	3	2400	6	13	0,333	0,188	0,998	0,995	0,240	0,992	0,246
T78.4	8	2407	0	7	1,000	0,533	1,000	0,997	0,696	0,997	0,729
R21	12	2406	1	3	0,923	0,800	1,000	0,999	0,857	0,998	0,859
H67	0	2409	0	13	*	0,000	1,000	0,995	*	0,995	*
H10.9	11	2409	1	1	0,917	0,917	1,000	1,000	0,917	0,999	0,916
M79.8	3	2410	0	9	1,000	0,250	1,000	0,996	0,400	0,996	0,499
L50.9	10	2411	0	1	1,000	0,909	1,000	1,000	0,952	1,000	0,953
TOTAL	**1131**	**74271**	**810**	**1292**	**0,583**	**0,467**	**0,989**	**0,983**	**0,518**	**0,467**	**0,508**

TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative, NPV: Negative Predictive Value, MCC: Matthew Correlation Coefficient; *: It is not calculable due to the value 0 assigned to the denominator.

When the MCC values are analyzed, it is seen that the assignments of R50, L50.9, H10.9, and K59.0 ICD codes are perfectly matched, while J21 and H60 ICD codes are randomly assigned. Like ChatGPT, Co-pilot does not have ICD codes with perfect agreement in misclassification. The Co-pilot's overall agreement is higher than that of ChatGPT.

4 Conclusion

This study shows whether LLMs can assign the correct ICD code according to patient complaints. In addition, this study highlights that the accuracy of LLMs decreases in classification when doctors choose incomplete/general codes when assigning ICD codes.

For the results obtained from this study to gain wider applicability, some necessary steps must be taken. It is essential to work with more data to increase accuracy rates. In addition, increasing the number of independent variables and the number of LLMs in the model may give different results. This is a study that can potentially provide guidance for future studies.

References

1. Thirunavukarasu, A.J., et al., *Large language models in medicine.* Nature medicine, 2023. **29**(8): p. 1930-1940.
2. Cain, W., *Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education.* TechTrends, 2024. **68**(1): p. 47-57.
3. Kasneci, E., et al., *ChatGPT for good? On opportunities and challenges of large language models for education.* Learning and individual differences, 2023. **103**: p. 102274.
4. Meskó, B., *Prompt engineering as an important emerging skill for medical professionals: tutorial.* Journal of Medical Internet Research, 2023. **25**: p. e50638.
5. Huang, J., et al., *A critical assessment of using ChatGPT for extracting structured data from clinical notes.* npj Digital Medicine, 2024. **7**(1): p. 106.
6. Liu, X., et al., *Uncovering Language Disparity of ChatGPT on Retinal Vascular Disease Classification: Cross-Sectional Study.* Journal of Medical Internet Research, 2024. **26**: p. e51926.
7. Liao, W., et al., *Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study.* JMIR Medical Education, 2023. **9**(1): p. e48904.
8. Skalidis, I., et al., *ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story?* European Heart Journal-Digital Health, 2023. **4**(3): p. 279-281.
9. Turunç Oğuzman, R. and Z.Z. Yurdabakan, *Performance of chat generative pretrained transformer and bard on the questions asked in the dental specialty entrance examination in turkey regarding bloom's revised taxonomy.* 2024.
10. *using machine learning techniques.* WSEAS transactions on computers, 2005. **4**(8): p. 966-974.
11. Coro, G., et al., *NLPHub: An e-Infrastructure-based text mining hub.* Concurrency and Computation: Practice and Experience, 2021. **33**(5): p. e5986.
12. Chicco, D. and G. Jurman, *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.* BMC genomics, 2020. **21**: p. 1-13.