ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

Fikriye ERDOĞAN

TÜRKİYE'DE LİSANS DÜZEYİNDEKİ İSTATİSTİK DERS İÇERİKLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ

İSTATİSTİK ANABİLİM DALI

ADANA, 2019

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKİYE'DE LİSANS DÜZEYİNDEKİ İSTATİSTİK DERS İÇERİKLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ

Fikriye ERDOĞAN

YÜKSEK LİSANS TEZİ

İSTATİSTİK ANABİLİM DALI

Bu Tez 12/09/2019 Tarihinde Aşağıdaki Jüri Üyeleri Tarafından Oybirliği/Oyçokluğu ile Kabul Edilmiştir.

.....	.....	.....
Prof. Dr. Güzin YÜKSEL DANIŞMAN	Doç. Dr. Gülsen KIRAL ÜYE	Prof. Dr. Nazif ÇALIŞ ÜYE

Bu Tez Enstitümüz İstatistik Anabilim Dalında hazırlanmıştır.
Kod No:

Prof. Dr. Mustafa GÖK
Enstitü Müdürü

Bu Çalışma Ç. Ü. Araştırma Projeleri Birimi Tarafından Desteklenmiştir.
Proje No: FYL201810542

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZ

YÜKSEK LİSANS TEZİ

TÜRKİYE'DE LİSANS DÜZEYİNDEKİ İSTATİSTİK DERS İÇERİKLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ

Fikriye ERDOĞAN

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI

Danışman : Prof. Dr. Güzin YÜKSEL
Yıl: 2019, Sayfa: 93
Jüri : Prof. Dr. Güzin YÜKSEL
: Doç. Dr. Gülsen KIRAL
: Prof. Dr. Nazif ÇALIŞ

Günümüzde teknolojinin gelişmesi ile birlikte veri miktarlarında ve çeşitliliğinde hızlı artışların olduğu gözlenmektedir. Veri artışları büyük ve karmaşık veri kümelerini oluşturmaktadır. Veri madenciliği, büyük ve karmaşık olan bu veri kümelerini veri madenciliği yöntemleri ile kolay bir şekilde analiz ederek yararlı olan anlamlı bilgilerin ortaya çıkarılmasını sağlamaktadır. Bu tez çalışmasında, Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümü bulunan diğer üniversitelerin lisans programlarında yer alan derslerin ders içerikleri karşılaştırılmıştır ve veri madenciliği yöntemleri ile analizleri yapılmıştır. Yapılan analizlerde; istatistik bölümlerinin birbirleri ile olan ilişkileri, istatistik bölümlerinin başarılarına etki eden ve istatistik bölümlerinde birlikte alınması gereken derslerin veri madenciliği ile belirlenmesi amaçlanmıştır. Uygulamada veri madenciliği yöntemlerinden kümeleme analizi, birliktelik kuralları analizi ve karar ağacı analizi kullanılmıştır. İstatistik bölümleri bulunan üniversitelerin birbirleri ile olan benzerlikleri kümeleme analizi ile gösterilmiştir. İstatistik bölümlerinin ders programlarında hangi derslerin birlikte alınması gerektiğini belirlemek için birliktelik kuralları analizi kullanılmıştır. İstatistik bölümlerinin başarılarına etki eden derslerin bulunması ise karar ağacı analizi ile yapılmıştır. Analizler için KNIME ve WEKA veri madenciliği programları kullanılmıştır.

Anahtar Kelimeler: Veri Madenciliği, İstatistik, Ders İçeriği, KNIME, WEKA

ABSTRACT

MSc THESIS

ANALYSIS OF THE COURSE CONTENTS OF STATISTICS AT BACHELORS LEVEL IN TURKEY WITH DATA MINING TECHNIQUES

Fikriye ERDOĞAN

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF STATISTICS

Supervisor : Prof. Dr. Güzin YÜKSEL
Year: 2019, Pages: 93
Jury : Prof. Dr. Güzin YÜKSEL
: Assoc. Prof. Dr. Gülsen KIRAL
: Prof. Dr. Nazif ÇALIŞ

Today with developing technology it is noticed that there is increase in the data quantity and diversity quickly. Depending on the data increase, huge and complicated datasets are formed. Data mining can easily analyze these huge and complicated datasets with data mining methods to search out helpful information.In this thesis; The course contents of the Statistics Department of Çukurova University and the Statistics Departments of other Universities were compared and analyzed using the data mining method. In the findings obtained from the analyzes, the relations of the statistical sections were examined. As a result, it is aimed to obtain the factors that affect the success of statistics departments and the courses that should be taken together in statistics departments by using data mining method. In the analysis of data, clustering analysis, association rules analysis and decision tree analysis were used. The common courses of the statistical departments were examined by cluster analysis. Association rules analysis was used to determine which courses should be taken together in the curriculum of the departments of statistics. Detecting the courses that affect the success of statistical departments was made by decision tree analysis. KNIME and WEKA data mining programs were used for analysis.

Key Words: Data Mining, Statistics, Course Contents, KNIME, WEKA

GENİŞLETİLMİŞ ÖZET

Teknolojinin gelişmesi ile veri miktarlarında ve çeşitliliğinde hızlı bir artış olduğu görülmektedir. Verilerdeki artışlar büyük ve karmaşık veri kümelerinin oluşmasına neden olmaktadır. Bu veri kümelerinin klasik yöntemler ile analizlerinin yapılması oldukça zordur. Veri madenciliği yöntemleri büyük ve karmaşık veri kümelerini kolay bir şekilde işleyerek veri madenciliği analizi ile yararlı olan anlamlı bilgilerin ortaya çıkarılmasını sağlamaktadır. Büyük ve karmaşık veri kümelerinde; verilerin birbirleri ile olan ilişkileri, verilerden oluşan modeller ve örüntüler veri madenciliği analizi ile belirlenmektedir. Veri madenciliği sağladığı kolaylıklar nedeniyle çeşitli alanlarda yaygın olarak kullanılmaktadır.

Veri madenciliğinin yaygın olarak kullanıldığı alanlardan biri de eğitim alanıdır. Eğitim alanlarında, öğrencilerin başarı durumlarının belirlenmesi, ders içeriklerinin incelenmesi ile ders programlarının düzenlenmesi gibi daha etkili eğitim ortamlarının oluşturulmasını sağlayacak analizler veri madenciliği ile yapılmaktadır.

Bu tez çalışmasında, ülkemizde bulunan istatistik bölümüne sahip 24 üniversitenin ders içerikleri incelenmiştir. Çukurova Üniversitesi İstatistik Bölümünün lisans programında yer alan dersler ve diğer istatistik bölümü bulunan üniversitelerin isimleri Excel dosyası olarak kaydedilmiştir. Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümüne sahip diğer üniversitelerin lisans programlarındaki derslerin ders içerikleri karşılaştırılarak sonuçları Excel dosyasına yazılmıştır. Analizler 3 farklı veri madenciliği yöntemi ile yapılmıştır. Veri madenciliği analizleri için KNIME ve WEKA programları kullanılmıştır.

Üniversitelerin istatistik bölümlerinin birbirilerine olan benzerliklerini belirlemek için kümeleme analizi, istatistik bölümlerinde birlikte bulunması gereken derslerin tespit edilmesi için birliktelik kuralları analizi ve istatistik bölümlerinin başarılarına etki eden derslerin bulunması için karar ağacı analizi

yapılmıştır. İlk analizde, kümeleme analizi yöntemlerinden biri olan hiyerarşik kümeleme analizi yöntemi kullanılarak istatistik bölümü bulunan üniversitelerin birbirleriyle olan benzerlikleri belirlenmiştir. Kümeleme sonucunda Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümüne sahip diğer 19 üniversitenin birbirlerine benzerlik gösterdiği ifade edilmiştir.

İkinci analizde, birliktelik kuralları analizi için istatistik bölümlerinde birlikte alınması gereken derslerin tespit edilmesinde Apriori algoritması kullanılmıştır. Apriori algoritmasında birliktelik kuralları düşük destek, yüksek güven ve kaldıraç (lift) değerlerine göre oluşturulmuştur. Bu birliktelik kurallarına göre istatistik bölümlerinde bulunması gereken önemli derslerin Olasılığa Giriş, Matematik 1, Kombinatorik, Aktüerya Modelleri, Lineer Cebir 1 ve Lineer Cebir 2 dersleri olduğu belirlenmiştir.

Üçüncü analizde ise, karar ağacı analizi için istatistik bölümlerinin başarılarına etki eden derslerin belirlenmesi amacıyla J48 (C4.5) algoritması kullanılmıştır. J48 (C4.5) algoritması analiz sonucunda sınıflandırma modelinin değerlendirilmesi için ölçüler ve grafikler oluşturmuştur. Bu ölçüler ve grafikler incelenerek sınıflandırmanın doğru tahmin edilip edilmediği belirlenmiştir. Analiz sonucunda karar ağacından oluşturulan karar kuralları incelendiğinde; istatistik bölümlerinin başarılarına etki eden derslerin Lineer Cebir 1, Aktüerya Modelleri, Lineer Modeller, Olasılığa Giriş, Lineer Cebir 2 ve Regresyon Analizi olduğu ifade edilmiştir.

Bu çalışmadaki analiz sonuçlarından faydalanılarak; üniversitelerin başka bölümleri için veri madenciliği uygulamaları yapılabilir. Uygulamalar sonucunda; üniversitelerin birbirleri ile olan ilişkileri, bölümlerde bulunması gereken dersler gibi bölümlerin başarılarını etkileyecek durumlar veri madenciliği analizleri ile belirlenmektedir. Ayrıca, yeni açılacak üniversitelerdeki bölümlerin eğitim programlarının hazırlanması ve bölümlerin başarılarına etki eden derslerin incelenmesi gibi araştırma konularının da analizleri veri madenciliği yöntemleri ile yapılabilmektedir.

TEŞEKKÜR

Tez çalışmam boyunca bilgilerini ve deneyimlerini benimle paylaşan, lisans eğitimimden itibaren desteğini ve emeğini üzerimden hiç esirgemeyen, her konuda bana hoşgörülü ve anlayışlı bir şekilde davranan, güler yüzünü ve içtenliğini daima bana gösteren değerli danışman hocam Prof. Dr. Güzin YÜKSEL'e sonsuz sevgilerimi ve teşekkürlerimi sunarım.

Bilgi ve tecrübesiyle bana her türlü konuda destek ve emek veren, lisans eğitimimden itibaren değerli fikirleri ile hayata karşı bakış açımı değiştirmemde bana yardımcı olan, her zaman bana sabırlı ve anlayışlı bir şekilde davranan, sevgisini ve samimiyetini bana daima hissettiren kıymetli hocam Doç. Dr. Deniz ÜNAL'a sonsuz teşekkürlerimi sunarım.

Her zaman yanımda olan, hoşgörü ve desteğini hiçbir zaman benden esirgemeyen değerli aileme sonsuz saygı ve sevgilerimle.

İÇİNDEKİLER **SAYFA**

ÖZ I
ABSTRACT I
GENİŞLETİLMİŞ ÖZET III
TEŞEKKÜR V
İÇİNDEKİLER VI
ÇİZELGELER DİZİNİ VIII
ŞEKİLLER DİZİNİ X
SİMGELER VE KISALTMALAR XII
1. GİRİŞ 1
2. ÖNCEKİ ÇALIŞMALAR 5
3. VERİ MADENCİLİĞİ 9
3.1. Veri Madenciliği Kısa Tarihi 9
3.2. Veri Madenciliği Tanımları 10
3.3. Veri Madenciliği Uygulama Alanları 11
3.4. Veri Madenciliği ve İstatistik 12
3.5. Veri Madenciliği Yöntemleri 12
3.5.1. Denetimli Öğrenme 12
3.5.1.1. Sınıflandırma 13
3.5.1.2. Regresyon 16
3.5.2. Denetimsiz Öğrenme 16
3.5.2.1. Kümeleme Analizi 17
3.5.2.2. Birliktelik Kuralları Analizi 19
3.6. Veri Madenciliği Yazılımları 21
4. MATERYAL VE YÖNTEM 23
4.1. Veri Madenciliği Programları 23
4.1.1. KNIME 23
4.1.2. WEKA 25

4.2. İstatistik Bölümleri Ders İçeriklerinin Veri Madenciliği Yöntemleri ile Analizi 26
4.2.1. Hiyerarşik Kümeleme Yöntemi 27
4.2.2. Apriori Algoritması 28
4.2.3. C4.5 (J48) Algoritması 29
5. UYGULAMA VE ANALİZ 31
5.1. Araştırmanın Tanımı 31
5.2. Verilerin Toplanması ve Hazırlanması 32
5.3. URAP İndeksi 33
5.4. İstatistik Bölümü Derslerine Göre Üniversitelerin Kümeleme Analizi İle Belirlenmesi 33
5.4.1. Veri Dönüşüm İşlemleri 33
5.4.2. Kümeleme Analizinin Uygulanması 37
5.4.3. Kümeleme Analizinin Sonuçları 41
5.5. İstatistik Bölümü Derslerinin Birbirleri İle Olan İlişkilerinin Birliktelik Kuralları Analizi İle Belirlenmesi 42
5.5.1. Veri Dönüşüm işlemleri 42
5.5.2. Birliktelik Kuralları Analizinin Uygulanması 44
5.5.3. Birliktelik Kuralları Analizinin Sonuçları 50
5.6. İstatistik Bölümlerinin Başarılarına Etki Eden Derslerin Karar Ağacı Analizi ile Belirlenmesi 50
5.6.1. Veri Dönüşüm işlemleri 50
5.6.2. Karar Ağacı Analizinin Uygulanması 51
5.6.3. Karar Ağacı Analizinin Sonuçları 64
6. SONUÇLAR VE ÖNERİLER 67
KAYNAKLAR 71
ÖZGEÇMİŞ 79
EKLER 80

ÇİZELGELER DİZİNİ **SAYFA**

Çizelge 5.1. Üniversitelerin Kümeleme Sonuçları 39
Çizelge 5.2. Kümelenen Üniversitelerin Renklendirilmesi 41
Çizelge 5.3. J48 Algoritması Sonuçları (Orijinal Veri Kümesi)..... 53
Çizelge 5.4. J48 Algoritması Sonuçları (Yeniden Örneklendirilmiş Veri Kümesi) 55

ŞEKİLLER DİZİNİ **SAYFA**

Şekil 4.1. KNIME Analiz Platformu 24
Şekil 4.2. WEKA Giriş Ekranı 25
Şekil 4.3. WEKA Kullanıcı Ara Yüzü Görüntüsü 26
Şekil 5.1. Excel Reader (XLS) Düğümü 34
Şekil 5.2. Create Byte Vector Düğümü 34
Şekil 5.3. Byte Vector Distances Düğümü 34
Şekil 5.4. Hierarchical Clustering (DistMatrix) Düğümü 35
Şekil 5.5. Hierarchical Cluster Assigner Düğümü 35
Şekil 5.6. Hierarchical Cluster View Düğümü 36
Şekil 5.7. Number To String Düğümü 36
Şekil 5.8. Column Filter Düğümü 36
Şekil 5.9. Color Manager Düğümü 37
Şekil 5.10. Excel Writer (XLS) Düğümü 37
Şekil 5.11. Kümeleme Analizi İçin Düğümlerin Çalıştırılması 38
Şekil 5.12. Kümelenen Üniversitelerin Dendogram Grafiği 40
Şekil 5.13. ARFF Reader Düğümü 42
Şekil 5.14. Create Bit Vector Düğümü 43
Şekil 5.15. Association Rule Learner Düğümü 43
Şekil 5.16. Association Rule Learner (Borgelt) Düğümü 44
Şekil 5.17. Birliktelik Kuralları Analizi İçin Düğümlerin Çalıştırılması 44
Şekil 5.18. WEKA Kullanıcı Ara Yüzü Görüntüsü (Orijinal Veri Kümesi) 52
Şekil 5.19. WEKA Kullanıcı Ara Yüzü Görüntüsü (Yeniden Örneklendirilmiş Veri Kümesi) 54
Şekil 5.20. Çok İyi Sınıf Etiketine Ait ROC Grafiği 58
Şekil 5.21. Orta Sınıf Etiketine Ait ROC Grafiği 58
Şekil 5.22. Kötü Sınıf Etiketine Ait ROC Grafiği 58
Şekil 5.23. İyi Sınıf Etiketine Ait ROC Grafiği 59

Şekil 5.24. Çok Kötü Sınıf Etiketine Ait ROC Grafiği 59
Şekil 5.25. Çok İyi Sınıf Etiketine Ait PRC Grafiği..... 60
Şekil 5.26. Orta Sınıf Etiketine Ait PRC Grafiği..... 60
Şekil 5.27. Kötü Sınıf Etiketine Ait PRC Grafiği..... 60
Şekil 5.28. İyi Sınıf Etiketine Ait PRC Grafiği 61
Şekil 5.29. Çok Kötü Sınıf Etiketine Ait PRC Grafiği 61
Şekil 5.30. J48 Algoritması Sonucunda Oluşan Karar Ağacı..... 62

SİMGELER VE KISALTMALAR

SQL	: Structured Query Language
SPSS	: Statistical Package for the Social Sciences
ÖİOS	: Öğrenci İşleri Otomasyon Sistemi
ÖSYM	: Ölçme, Seçme ve Yerleştirme Merkezi
CHAID	: Chi-squared Automatic Interaction Detection
QUEST	: Quick Unbiased Efficient Statical Tree
C&RT	: Classification and Regression Tree
SAS	: Statistical Analysis Software
MATLAB	: Matrix Laboratory
YÖK	: Yükseköğretim Kurulu
AKTS	: Avrupa Kredi Transfer Sistemi
URAP	: University Ranking By Academic Performance
KNIME	: Konstanz Information Miner
ARFF	: Attribute Relation File Format
WEKA	: Waikato Environment For Knowledge Analysis
CSV	: Comma Separated Values
MAE	: Mean Absolute Error (Ortalama Mutlak Hata)
RMSE	: Root Mean Squared Error (Kök Hata Kareler Ortalaması)
RAE	: Relative Absolute Error (Bağıl Mutlak Hata)
RRSE	: Root Relative Squared Error (Bağıl Karesel Hatanın Karekökü)
MCC	: Matthews Correlation Coefficient (Matthews Korelasyon Katsayısı)
ROC	: Receiver Operating Characteristic (Alıcı İşlem Karakteristiği)
PRC	: Precision Recall Curve (Kesinlik Anma Eğrisi)
ID3	: Iterative Dichotomiser 3
C5.0	: See5
NB	: Naive Bayes

KDH : Kendini Düzenleyen Haritalama

AIS : Agrowal Imielinski Swami

SETM : Set-Oriented Mining

FPGrowth : Frequent Patterns-Growth

Eclat : Equivalence Class Transformation

Kxen : Knowledge Extraction Engine

Keel : Knowledge Extraction Based on Evolutionary Learning

MDS : Multi Dimesional Scaling

1. GİRİŞ

Son yıllarda bilgisayar teknolojisindeki gelişmeler ile veri miktarlarında ve çeşitliliğinde hızlı bir artış olduğu gözlenmektedir. Bu veri artışları büyük ve karmaşık veri kümelerini oluşturmaktadır. Büyük ve karmaşık olan veri kümelerinde bulunan bilgilere ulaşmak hem zor hem de zaman almaktadır. Ulaşılması zor olan ve zaman alan bu bilgilerin oluşturulması veri madenciliği ile sağlanmaktadır. Veri madenciliği büyük ve karmaşık veri kümelerindeki yararlı olan anlamlı bilgilerin ortaya çıkarılmasını kolay bir şekilde yapmaktadır.

Veri madenciliği analizleri, elde bulunan veri kümelerinin durumuna ve verilerin kullanım amacına göre veri madenciliği yöntemleri ile yapılmaktadır. Veri madenciliği yöntemlerini uygulamadan önce analiz için önemli olacak değişkenlerin belirlenmesi gerekmektedir. Değişkenler belirlendikten sonra veri madenciliği yöntemleri ile analizler gerçekleştirilmektedir. Böylece, veri madenciliği ile doğru bilgilere hızlı ve güvenilir bir şekilde ulaşılmaktadır.

Veri madenciliğinin kullanımı belirli uygulama alanlarıyla sınırlandırılamaz. Veri madenciliğini verinin üretilip kayıt altına alındığı her alanda kullanmak mümkündür. Sağlık, endüstri, mühendislik, pazarlama, bankacılık ve eğitim alanları veri madenciliğinin yoğun olarak kullanıldığı başlıca uygulama alanlarıdır (Han ve Kamber, 2006: Gülen 2014).

Eğitim, veri madenciliğinin yaygın olarak kullanılan alanlarından biridir. Eğitim alanlarında yapılan analizlerde; öğrencilerin başarı durumlarının tespit edilmesi, ders içeriklerinin incelenmesi ile ders programlarının hazırlanması gibi eğitimin kalitesini arttıracak konular veri madenciliği ile gerçekleştirilmektedir.

Veri madenciliği yöntemlerinin eğitim alanında kullanılmasıyla birlikte ortaya çıkan, literatürdeki adıyla eğitsel veri madenciliği (educational data mining), öğrencilere ve okullara ait çoklu verinin işlenmesine olanak sağlayan ve kullanım alanı oldukça yaygın olan bir yöntemler topluluğudur. Özellikle eğitim alanındaki kullanımı, eğitimcilere ve eğitim planlamacılarına farklı açılardan ışık tutarak

eğitim stratejilerinin belirlenmesinde ihtiyaca yönelik çözümler sunabilmektedir. Eğitsel veri madenciliğini, eğitim ortamından elde edilen verilerin daha önceden bilinmeyen ya da oluşturulmamış yapısını keşfetmek için yöntem geliştirme ve geliştirilen yöntemlerin öğrencileri daha iyi tanımada kullanan gelişmekte olan bir disiplin alanı olarak tanımlamak mümkündür (Bilen ve ark., 2014).

Bu tez çalışmasında, ülkemizde bulunan 24 istatistik bölümünün ders içerikleri incelenmiştir. Çukurova Üniversitesi İstatistik Bölümü lisans programlarında bulunan derslerin ders içerikleri karşılaştırılarak veri madenciliği ile analizleri yapılmıştır. Bu analizler 3 farklı veri madenciliği yöntemi ile gerçekleştirilmiştir.

İlk analizde, istatistik bölümü bulunan üniversitelerin birbirine olan benzerliklerini bulmak için kümeleme analizi yöntemlerinden seçilen hiyerarşik kümeleme analizi yöntemi kullanılarak analiz yapılmıştır. Kümeleme analizi istatistik bölümleri derslerine göre birbirilerine benzer olan üniversitelerin hangileri olduğunu göstermiştir.

İkinci analiz olan birliktelik kuralları analizi için istatistik bölümlerinin eğitim programlarında bulunan derslerin hangilerinin birlikte alınması gerektiğini tespit etmek amacıyla Apriori algoritması kullanılarak analiz gerçekleştirilmiştir. Apriori algoritmasında düşük destek, yüksek güven ve kaldıraç (lift) değerleri seçilerek birliktelik kuralları oluşturulmuştur. Bu birliktelik kurallarının incelenmesi ile istatistik bölümlerinde bulunması gereken dersler belirlenmiştir.

Üçüncü analizde ise istatistik bölümlerinin başarılarına etki eden derslerin belirlenmesi için J48 (C4.5) algoritması ile karar ağacı analizi yapılmıştır. J48 (C4.5) algoritması analiz sonucunda sınıflandırma performansını gösteren ölçüler ve grafikler oluşturmuştur. Sınıflandırma performansını gösteren ölçülerin ve grafiklerin incelenmesi ile sınıflandırma modelinin doğruluğu değerlendirilmiştir.

Sınıflandırma modelinin doğruluğu karar ağacı analizini iyi bir şekilde açıklamaktadır. Analiz sonucunda oluşan, karar ağacından oluşturulan karar

kuralları incelenerek istatistik bölümü bulunan üniversitelerin başarılarına etki eden derslerin hangileri olduğu tespit edilmiştir.

Kümeleme analizi ve birliktelik kuralları analizi için KNIME programı, karar ağacı analizi için ise WEKA programı kullanılmıştır.

Bu çalışmada yapılan analizlerde, istatistik bölümleri bulunan üniversitelerin birbirleriyle olan ilişkileri, istatistik bölümlerinin başarılarına etki eden ve ders programlarında bulunması gereken derslerle ilgili sonuçlar elde edilmiştir. Analizler sonucunda ortaya çıkan bilgilerin kullanılması ile hem bölümlerin hem de üniversitelerin başarılarının artması amaçlanmaktadır. Bu analiz sonuçlarından yararlanılarak üniversitelerin başka bölümleri için veri madenciliği uygulamaları yapılabilir.

Yeni açılacak üniversitelerdeki bölümlerin eğitim programlarının hazırlanması ve bölümlerin başarılarına etki edecek derslerin belirlenmesi gibi araştırma konuları veri madenciliği yöntemleri ile analiz edilebilmektedir.

2. ÖNCEKİ ÇALIŞMALAR

Kılıçer (2018) Türkiye'deki bilgisayar mühendisliği lisans programlarının ders içerikleri ile Erasmus Programıyla gidilen Avrupa ülkelerindeki bilgisayar mühendisliği lisans programlarının ders listelerini belirleyerek dersleri karşılaştırmıştır. İncelediği ders içeriklerini WEKA programında Naive Bayes ve J48 algoritmaları ile analiz ederek Türkiye'deki bilgisayar mühendisliği bölüm derslerinin eksiklerini tespit etmiştir.

Güven (2016) Türk Üniversitelerindeki Bilgisayar Mühendisliği Bölümleri müfredatlarında yer alan derslerin ders içeriklerini inceleyerek bir eğitsel veri madenciliği uygulaması gerçekleştirmiştir. Analizleri kümeleme, birliktelik kuralları ve karar ağacı veri madenciliği yöntemleri ile yapmıştır. Çalışmasında KNIME programını kullanmıştır.

Delavari ve arkadaşları (2004) üniversitelerin daha verimli, etkili ve doğru eğitim süreçlerine sahip olmalarını kolaylaştırmak amacıyla veri madenciliği teknolojisini kullanarak yükseköğretim sistemlerindeki karar verme süreçlerinin iyileştirebilir olduğunu açıklamışlardır.

Al-Radaideh ve arkadaşları (2006) veri madenciliği süreçlerini, derslerdeki öğrenci performansını etkileyebilecek özellikleri inceleyerek yüksek eğitim sisteminin kalitesini arttırmak amacıyla öğrenci verilerini analiz etmişlerdir.

Güvenç (2001) çalışmasında yükseköğrenimdeki öğrenci ders performansları ve ders seçimi ile ilgili olarak bilgi tabanlı bir sistem geliştirmiştir. Ham veri şeklindeki mevcut öğrenci kayıtları verisine ilişkilendirme kurallarını, karar ağaçlarını ve istatistiksel analizlere dayalı veri madenciliği tekniklerini uygulayarak öğrenci ve üniversite yönetimi için yararlı bilgiler ortaya çıkarmıştır.

İnan (2003) Selçuk Üniversitesi'nin öğrenci işleri otomasyonundan elde ettiği veriler üzerinden üniversitenin geleceği ile ilgili plan yapmak için karar ağacı ve birliktelik kuralları veri madenciliği yöntemlerini kullanmıştır. Oracle 8i veri tabanı üzerinde bulunan verilerden analiz edilmeye uygun olanlarını seçerek SQL

Server Analysis Services veri madenciliği aracı ile karar ağacı ve birliktelik kuralları analizleri sonuçlarından anlamlı bilgiler ortaya çıkarmıştır.

Bölükbaşı (2005) akademik başarıyı etkileyen faktörleri belirlemek için İstanbul Teknik Üniversitesi İşletme Fakültesi 2. 3. ve 4. sınıf lisans öğrencilerine bir anket uygulamıştır. Anket sonuçlarını karar ağacı tekniği ile analiz ederek işletme öğrencilerinin başarılarına etki eden faktörleri tespit etmiştir.

Romero ve Ventura (2013) yaptıkları çalışmada eğitimsel veri madenciliğini ve uygulamalarını incelemişlerdir.

Gülçe (2010) Pamukkale Üniversitesi Bilgi İşlem Daire Başkanlığı sunucuları üzerinden öğrenci işleri otomasyon sisteminde (ÖİOS) bulunan verileri Microsoft karar ağaçları, Microsoft Naive Bayes, Microsoft birliktelik kuralları ve Microsoft sinir ağları veri madenciliği algoritmaları ile analiz etmiştir. Analizleri Microsoft SQL Server 2008, Microsoft Visual Studio 2008, Microsoft Analysis Services 2008 ve Microsoft Reporting Services 2008 veri tabanı yönetim sistemlerinde yaparak veri madenciliği modelleri ile öğrencilerin akademik ortalamalarını, ders başarılarını ve üniversitenin gelişmesine yönelik bilgileri ortaya çıkarmıştır.

Baepler ve Murdoch (2010) yükseköğretimdeki akademik analitik ve veri madenciliği konularını incelemişlerdir.

Akça (2014) Gazi Üniversitesi öğrencilerine, Fen Fakültesini seçmelerini etkilediği düşündüğü öğrencilerin not ortalamaları, nüfusa kayıtlı olduğu ili, doğum yeri, tercih sırası, cinsiyeti, lise türü ve ÖSYM puanı gibi çeşitli değişkenler ile akademik başarı puanları arasındaki ilişkiyi belirlemek amacıyla Gazi Üniversitesi Fen Fakültesi öğrencilerine ait 2010 yılı verilerini kullanarak veri madenciliği çalışması yapmıştır. SPSS Clementine 12 programında CHAID, QUEST, C&RT ve C5.0 veri madenciliği algoritmalarını kullanarak analizi gerçekleştirmiştir.

Vialardi ve arkadaşları (2009) öğrencilerin akademi ile ilgili karar almalarına yardımcı olmak için veri madenciliği yöntemlerine dayanan bir öneri sisteminin kullanılmasını incelemişlerdir. Lima Üniversitesi'ndeki sistem

mühendisliği verilerini analiz ederek öğrenciler için bir öneri sistemi oluşturmuşlardır.

Kumar ve Chadha (2011) yaptıkları çalışmada büyük veri kümelerinden faydalı bilgiler çıkarmak için eğitim kurumlarındaki veri madenciliği uygulamalarını kullanmışlardır. Stratejik bilgi üreterek eğitim sistemini destekleyebilecek çeşitli veri madenciliği tekniklerini yükseköğrenim kurumlarında uygulanması gerektiğini önermişlerdir.

3. VERİ MADENCİLİĞİ

3.1. Veri Madenciliği Kısa Tarihi

Veri madenciliği, kavramsal olarak 1960'lı yıllarda, bilgisayarların veri analiz problemlerini çözmek için kullanılmaya başlamasıyla ortaya çıktı. O dönemlerde, bilgisayar yardımıyla, yeterince uzun bir tarama yapıldığında, istenilen verilere ulaşmanın mümkün olacağı gerçeği kabullenildi. Bu işleme veri madenciliği yerine önceleri veri taraması (data dredging), veri yakalanması (data fishing) gibi isimler verildi (Öğüt, 2009).

1960'lı yıllarda ortaya çıkan bu kavramın teknolojinin ilerlemesi ile birlikte veri madenciliği uygulamalarındaki kullanımlarının günümüze kadar olan gelişimi aşağıdaki şekilde açıklanmıştır.

1970'lerde İlişkisel Veri Tabanı Yönetim Sistemleri uygulamaları kullanılmaya başlanmıştır. Bilgisayar uzmanları bununla beraber basit kurallara dayanan uzman sistemler geliştirerek, basit anlamda makine öğrenimini sağlamışlardır. 1980'lerde veri tabanı yönetim sistemleri yaygınlaşmış ve bilimsel alanlarda uygulanmaya başlanmıştır. Bu yıllarda şirketler; müşteriler, rakipler ve ürünler ile ilgili verilerden oluşan veri tabanları oluşturmuşlardır. Bu veri tabanlarının içerisinde çok büyük miktarlarda veri bulunmaktadır ve bunlara SQL veri tabanı sorgulama dili ya da benzeri diller kullanarak ulaşılabilmektedir (Savaş ve ark., 2012).

1990'larda bilgisayar mühendisleri, geleneksel istatistiksel yöntemlerinin yerine algoritmik bilgisayar modülleri ile veri analizinin değerlendirilebileceğini ifade ederek, veri madenciliği ismini kullanmışladır. Bu yıllarda veri tabanlarındaki veri miktarları katlanarak arttığı için, büyük miktardaki veri içinden yararlı bilgilere nasıl ulaşılması gerektiği üzerinde düşünülmeye başlanmış ve veri madenciliği için ilk yazılım gerçekleştirilmiştir. 2000'li yıllardan itibaren veri madenciliği sürekli gelişmiş ve geniş bir yelpazede uygulanmaya başlanmıştır (Boyacı, 2017).

3.2. Veri Madenciliği Tanımları

Veri madenciliği için literatürde çeşitli tanımlar bulunmaktadır. Bu tanımlardan bazıları aşağıdaki şekilde açıklanmıştır:

- ✓ Veri madenciliği veri tabanı teknolojisi, istatistik, yapay zekâ (artificial intelligence), makine öğrenimi (machine learning), örüntü tanımlama (pattem recognition) ve veri görselleştirmesi (data visualization) gibi pek çok teknik alan arasında köprü görevi gören çok disiplinli bir alandır (Özekes, 2003).
- ✓ Kittler ve Wang (1999), veri madenciliğini oldukça tahminci anahtar değişkenlerin binlerce potansiyel değişkenden izole edilmesini sağlama yeteneği olarak tanımlamışlardır.
- ✓ Veri madenciliği; veri ambarında bulunan büyük miktardaki verileri, istatistiksel ve matematiksel tekniklerin yanı sıra patern tanıma teknolojileri de kullanılarak, faydalı yeni korelasyonlar, paternler ve trendler keşfetme sürecidir (Larose, 2005).
- ✓ Veri madenciliği, veri içerisinde gizli kalmış, önceden bilinemeyen ve potansiyel olarak faydalı olan anlamlı bilginin çıkartılmasıdır (Flawley ve ark., 1992).
- ✓ Veri madenciliğinin genel bir tanımı yapılacak olursa, büyük hacim ve farklı tipteki veriden anlamlı bir takım desenler/örüntüler oluşturmayı, veri arasındaki ilişkileri keşfetmeyi ve veriden bilgi elde etmeyi amaçlayan bir alandır. Tıpkı bir maden keşfi yapar gibi veri yığınları arasından da "değerli, anlamlı" olan bilginin bulunmasını sağlar (Emre ve Erol, 2017).
- ✓ Veri Madenciliği; geniş veri tabanlarından bilgi çıkartabilmek amacıyla makine öğrenmesi, örüntü tanıma, istatistik, görselleştirme gibi alanların tekniklerini bir araya getiren disiplinler arası bir alandır (Cabena ve ark., 1998).

3.3. Veri Madenciliği Uygulama Alanları

Veri madenciliği, teknolojinin gelişmesi ile birlikte artan verileri kolay bir şekilde analizlerde kullanılabilir duruma getirerek veri madenciliği yöntemleriyle anlamlı bilgilerin ortaya çıkarılmasını sağlamaktadır. Analizlerde sağladığı kolaylıklar nedeniyle çeşitli alanlarda kullanılmaktadır.

Veri madenciliğinin kullanıldığı başlıca alanlar şunlardır (Yalçın, 2019):

- Pazarlama yönetimi
- Müşterilerin alışveriş alışkanlıklarının tespiti
- Kredi skorlama
- Reklam kampanyalarının etkinliğini artırma
- Müşteri sadakati ve yeni müşteri kazanımı
- Pazar sepeti analizi
- Dolandırıcılık tespiti ve risk yönetimi
- Sinyal işleme (telekomünikasyon)
- Biyoloji (DNA sıra işleme)
- Tıp
- Fiyatlandırma
- Tedarik zinciri analizi
- Ulaştırma
- Tarım
- Hava durumu tahmini
- Eğitim - öğretim
- Sosyal ağ analizi
- Bilgi güvenliği
- Silahlı kuvvetler güvenlik uyarıları
- Şehircilik ve planlama
- Coğrafi bilgi sistemleri

3.4. Veri Madenciliği ve İstatistik

Büyük veri kümeleri içinde gizli kalmış örüntüleri ve önceden bilinmeyen anlamlı bilgileri ortaya çıkarmak için veri madenciliği uygulamalarında birçok disiplinden yararlanılmaktadır. İstatistik (statistics), veri tabanı teknolojisi (database technology), makine öğrenmesi (machine learning), örüntü tanıma, veri görselleştirme (data visualisation) ve yapay zeka gibi disiplinlerin hepsi veri madenciliğinde rol oynamaktadırlar. Bu disiplinler arasında kesin sınırlar çizmek mümkün olmadığı gibi, veri madenciliği ile bu disiplinler arasında da sınır belirlemek mümkün değildir. Sınırlarda bir kişinin veri madenciliği problemi, bir başkasının istatistik, veri tabanı ya da makine öğrenmesi problemi olabilmektedir (Hand ve ark., 2001).

Kuşkusuz ki bu disiplinler içinde istatistik biliminin yeri diğerlerinden farklıdır. İstatistikte kullanılan tahmin yöntemleri bilgisayarlarda otomatikleştirilmiştir. Bunun sonucunda istatistik tabanlı veri madenciliği teknikleri ortaya çıkmıştır. Veri madenciliği, istatistiksel süreçleri etkili bir biçimde otomatikleştirir. Bu sayede son kullanıcının yükünü biraz da olsa hafifletir. Veri madenciliği ve istatistikte ortak amaç, karmaşık verilerden anlamlı bilgiler elde etmektir. En büyük farkları ise; veri madenciliğinin bilgisayar teknolojisi ve birtakım algoritmalar sayesinde çok büyük miktardaki veri kümelerine hitap ediyor olmasıdır. Fakat bilinmelidir ki bu algoritmaların temelini istatistiksel yöntemler oluşturmaktadır (Atılgan, 2011).

3.5. Veri Madenciliği Yöntemleri

Veri madenciliği yöntemleri denetimli (supervised) ve denetimsiz (unsupervised) olmak üzere iki gruba ayrılmaktadır.

3.5.1. Denetimli Öğrenme

Denetimli öğrenme örnekten öğrenme olarak da isimlendirilen denetimli öğrenimde, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre

ayrılarak, her sınıf için çeşitli örnekler verilir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir (Akpınar, 2000). Veri madenciliği yöntemlerinden sınıflandırma ve regresyon denetimsiz öğrenme olarak gösterilmektedir.

3.5.1.1. Sınıflandırma

Sınıflandırma veri madenciliğinin en temel işidir. Girdilerin önceden belirlenen niteliklere göre sınıflara atanmasıdır. Eldeki verilerin bir sınıfa dahil olup olamayacağı, ya da hangi sınıfa dahil olacağı belirlenir. Yani isminden de anlaşıldığı gibi sınıfın belirlenmesidir. Sınıflandırmada girdi gözlem ve örneklerden oluşmuş eğitim setidir. Çıktı modelin gözlem sonucunda belirlenmiş olduğu sınıf etiketidir. Sınıflandırma iki aşamada gerçekleşen bir işlemdir. Bu aşamalar eğitim ve modelin testidir. Eğitim aşaması, eğitim kümesinin kullanılarak modelin oluşturulmasıdır. Test ise modelin kesinliğinin kontrol edilmesine karşılık gelir. Modelin kesinliğini kontrol etmek için: test örneklerinden sınıfı iyi bilenenin model tarafından tahminini sonucu karşılaştırılır. Girdilere göre çıktıları üreten model ile sonradan sınıf etiketi bilinmeyen ya da kayıp olan örneklerin sınıfı tahmin etmesinde kullanılır (Emel ve Taşkın, 2005).

Veri madenciliği uygulamalarında kullanılan başlıca sınıflandırma teknikleri şunlardır; karar ağaçları, yapay sinir ağları, genetik algoritmalar, k-en yakın komşu, Naive Bayes ve destek vektör makineleridir.

Karar Ağaçları: Veri madenciliğinde karar ağaçları, kurulmasının ucuz olması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri ve güvenilirliklerinin iyi olması nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahip tahmin edici bir tekniktir (Linoff ve Berry, 1999).

Karar ağacı karar düğümleri, dallar ve yapraklardan oluşur (Han ve Kamber, 2000). Karar düğümü, gerçekleştirilecek test belirtir. Bu testin sonucu ağacın veri kaybetmeden dallara ayrılmasına neden olur. Her düğümde test ve

dallara ayrılma işlemleri ardışık olarak gerçekleşir ve bu ayrılma işlemi üst seviyedeki ayrımlara bağımlıdır. Ağacın her bir dalı sınıflama işlemini tamamlamaya adaydır. Eğer bir dalın ucunda sınıflama işlemi gerçekleşemiyorsa, o dalın sonucunda bir karar düğümü oluşur. Ancak dalın sonunda belirli bir sınıf oluşuyorsa, o dalın sonunda yaprak vardır. Bu yaprak, veri üzerinde belirlenmek istenen sınıflardan biridir. Karar ağacı işlemi kök düğümünden başlar ve yukarıdan aşağıya doğru yaprağa ulaşana dek ardışık düğümleri takip ederek gerçekleşir (Ayık ve ark., 2007) .

Karar ağaçları analizinde; ID3 (Iterative Dichotomiser 3), C4.5 (J48), C&RT (Classification and Regression Tree), CHAID (Chi-squared Automatic Interaction Detection), QUEST (Quick Unbiased Efficient Statical Tree) ve C5.0 (See5) algoritmaları kullanılmaktadır.

Yapay Sinir Ağları: İnsan beyninin özelliklerinden esinlenerek; öğrenme yoluyla yeni bilgiler türetebilme, keşfetme ve oluşturma gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacıyla geliştiren bilgisayar sistemleridir. Bu yeteneklerin gerçekleştirilmesi geleneksel programlama yöntemleriyle çok zordur. Bu nedenle, yapay sinir ağlarını mümkün olmayan olaylar veya programlanması çok zor olan olaylar için geliştirilmiş bilgi işlemeyle ilgilenen bir bilgisayar bilim dalı olduğu söylenebilir (Öztemel, 2003).

Genetik Algoritmalar: Biyolojideki kalıtım, seçim, çaprazlama veya mutasyon gibi tekniklerden esinlenerek arama veya optimizasyon problemini tam ya da yaklaşık olarak tahmin eden sezgisel bir optimizasyon tekniğidir. Genetik Algoritmalar, kısıtların belli olmadığı veya arama uzayının çok büyük olduğu, değişken sayısının çok fazla olduğu, çok boyutlu fonksiyonlardan oluşan karmaşık problemlerde basit şekilde uygulanabilir ve mantıklı sonuçlar verebilir (Kevran, 2009).

K-En Yakın Komşu: Sınıflandırma problemini çözen denetimli öğrenme yöntemleri arasında yer almaktadır. Yöntemde; sınıflandırma yapılacak verilerin öğrenme kümesindeki normal davranış verilerine benzerlikleri hesaplanarak; en

yakın olduğu düşünülen k verinin ortalamasıyla, belirlenen eşik değere göre sınıflara atamaları yapılmaktadır. Önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş olmasıdır. Yöntemin performansını k en yakın komşu sayısı, eşik değer, benzerlik ölçümü ve öğrenme kümesindeki normal davranışların yeterli sayıda olması kriterleri etkilemektedir (Çalışkan ve Soğukpınar, 2008). Bu metotta eğitim örnekleri n-boyutlu sayısal niteliklerle tanımlanır. Her örnek n-boyutlu uzayda bir noktayı gösterir. Bu yolla, bütün eğitim örnekleri n-boyutlu örnek uzayda depolanır (Kalıkov, 2006). K en yakın komşu sınıflayıcısı, bilinmeyen örneğe en yakın k eğitim örneğini bulur. Yakınlık, X= (x_1,x_2,….,x_n) ve Y= (y_1,y_2,…….,y_n) gibi iki nokta arasındaki Öklid uzaklığı ile ifade edilmektedir (Boyacı, 2017).

Naive Bayes (NB): Temeli Bayes teorisine dayanan, verileri istatistiksel sınıflandırma yöntemlerinden biridir. Veri madenciliği sınıflandırma algoritmalarından olan Bayes, uygulanabilirliği ve hızlı hesaplama performansı ile araştırmacılar tarafından öne çıkan bir algoritmadır. Sınıflandırılacak olayları birbirinden bağımsız olarak ele almaktadır (Olgun ve Özdemir, 2012).

NB algoritmasının uygulanmasında bir takım kabuller yapılmaktadır. Bunların en önemlisi niteliklerin birbirinden bağımsız olmasıdır. Nitelikler birbirini etkilediği taktirde, burada olasılık hesaplamak zorlaşacağı için niteliklerin hepsinin aynı derecede önemli olduğu kabul edilmektedir (Çalış, 2013).

NB, sürekli veri ile çalışmaz. Bu nedenle sürekli değerleri içeren bağımlı ya da bağımsız değişkenler kategorik hale getirilmelidir. NB, modelin öğrenilmesi esnasında, her çıktının öğrenme kümesinde kaç kere meydana geldiğini hesaplamaktadır. Bulunan bu değer, öncelikli olasılık olarak adlandırılır. NB aynı zamanda her bağımsız değişken / bağımlı değişken kombinasyonunun meydana gelme sıklığını bulmaktadır. Bu sıklıklar öncelikli olasılıklarla birleştirilmek suretiyle tahminde kullanılır (Akbulut, 2006).

Destek Vektör Makineleri: Destek vektör makinesi algoritması Boser ve arkadaşları (1992) tarafından keşfedilmiş, parametrelerinde lineer olan fakat giriş

bileşenlerinde lineer bağımlılıklarla sınırlı olmayan fonksiyonlar olarak tanımlanmıştır. Algoritma özellik sayısı kadar boyutlu bir uzayda tüm verileri yerleştirir ve sonrasında sınıfları birbirinden en iyi ayıran hiper düzlemi bularak sınıflandırma işlemi yapar (Yalçın, 2019).

3.5.1.2. Regresyon

Regresyon analizi, bu yöntem değerler arasındaki ilişkileri tahmin etmek için tercih edilen istatistiksel bir yöntemdir. Dikkat edilmesi gereken nokta, bağımsız değişkenlerin bağımlı değişkenleri açıklama oranının yüksek olması beklenir. Büyük verilerin olduğu uygulamalarda neredeyse yansız sonuç elde edilebilir. Bir diğer ifade ile bu yöntem ölçüm sonucunda elde edilen veri noktalarına mümkün olduğu kadar yakın geçecek olan bir fonksiyon eğrisi bulmaya yarar (Sezgin ve Çelik, 2013). Regresyon modelinde bağımsız değişkenlerden hareketle çıktı/bağımlı değişken denilen sonuç değerlerinin ilişkilendirecek yapıyı kurgulayarak en iyi tahmine ulaşılmak hedeflenilir. Elde edilen sonuç kullanılırken güven sınırları belirtilir. Bu yöntemde girdiler birden fazla bulunmaktadır. Bu girdilerin sonuç üzerindeki katkıları ise tek tek değerlendirilerek en az katkısı olan girdi göz ardı edilebilir (Özcan, 2014).

3.5.2. Denetimsiz Öğrenme

Denetimsiz öğrenme, verilerdeki önceden bilinmeyen yapı ve ilişkileri bulmak için kullanılan yöntemleri ifade etmektedir. Denetimsiz öğrenmenin amacı, veri içindeki benzer örnek gruplarını bulmak ve veriler hakkında daha fazla bilgi edinmek için verilerdeki temel yapıyı veya dağılımı modellemektir (Akay, 2018).

Kümeleme ve birliktelik kuralları analizleri denetimli öğrenme veri madenciliği yöntemleri olarak gösterilmektedir.

3.5.2.1. Kümeleme Analizi

Çok değişkenli istatistik yöntemlerinden biri olan ve çok sayıda ve karmaşık yapıda veri içeren bir alandaki verileri gruplandırmak ve oluşan grupları karşılaştırmak amacıyla kullanılan kümeleme analizi, kullanımının kolay ve sonuçlarının anlaşılır olması nedeniyle sıkça başvurulan bir yöntemdir. Veri madenciliğinin bir alt türü olan bu yöntemde veriler uzaklık ve benzerliklerine göre kümelere ayrılarak, kümeler arasındaki farklılıklar ve nedenleri üzerinde durulur. Birbirine en çok benzeyen birimler aynı kümelerde toplanırlar. Bu tür kümeler kendi içlerinde homojen diğer kümelerle heterojen durumdadırlar (Tekin, 2015).

Kümeleme analizinde amaç verileri birbirleriyle benzer alt kümelere ayırmaktır. Verilerin hangi kümelere ayrılacağı, hatta kaç değişik küme oluşturulacağı verilerin birbirine olan benzerliğine ve uzaklığına göre belirlenmektedir. Verilerin benzerliği ile kastedilen şey ise aralarındaki mesafenin ölçülmesi ve değerlendirilmesidir. Bu veri tabanında bulunan her bir kaydın diğer bir kayıtla olan benzerliği ya da diğer kayıtlara uzaklığı olduğu gibi oluşturulan gerçek ve aday kümeler arasındaki benzerlik ve uzaklığı da içermektedir. En çok kullanılan uzaklık ölçütleri Öklid, Manhattan, Minkowski; benzerlik ölçütleri ise Dice, Jaccard, Cosine, Overlap'dir (Pfitzner ve ark., 2009).

Veri madenciliği uygulamalarında verileri benzerliklerine göre kümelemek için yaygın olarak kullanılan bazı kümeleme yöntemleri şunlardır: hiyerarşik kümeleme, k-means, k- medoids, kohonen ağları, bulanık c-ortalamalar.

Hiyerarşik Kümeleme: Hiyerarşik kümelemede küme sayısı önceden belli değildir. Bu yöntemde mümkün küme yapıları bir ağaç şeklinde gösterilir. Bu ağaca dendogram adı verilir. Dendogramın oluşturulmasında iki yaklaşım söz konusudur. Bunlardan birincisi tüm gözlemleri birer küme olarak dikkate alıp gözlemleri birleştirerek daha az sayıda küme elde etme yöntemidir. Bu yöntem, birleştirici hiyerarşik kümeleme (agglomerative hierarchical clustering) yöntemi olarak adlandırılır. İkincisi ise, tüm gözlemleri tek bir küme olarak ele alıp bu kümeye alt kümeleri de daha alt kümelere ayırma şeklinde uygulanan yöntemdir.

Bu yöntem, ayrıştırıcı hiyerarşik kümeleme (divisive hierarchical clustering) yöntemi olarak adlandırılmaktadır (Altunkaynak, 2017).

K-Means: 1967'de J. B. MacQueen tarafından ilk kez yayınlanan k-means algoritması, kümeleme algoritmaları içinde yaygın olarak kullanılan en popüler algoritmalardan biridir (Ghoson 2010: Bargavi 2017).

Nesneler arasındaki uzaklığa dayalı tipik bir kümeleme algoritmasıdır ve mesafeler arasındaki benzerlik ölçüsüne dayanır (Wu ve ark., 2018). Nesneler arasındaki mesafe uzaklığı aşağıdaki Öklid uzaklığı formülü ile hesaplanır (Joyendri, 2017):

$$dij = \sum_{k=1}^{p} \{x_{ik} - x_{jk}\}^2 \quad (3.1.)$$

Burada dij = nesnenin ile j arasındaki mesafesini, p= verinin boyutunu, x_{ik}= Boyut k için cisim i'nin koordinatı, x_{jk}= k boyutunda j nesnesinin koordinatını temsil eder. Bu algoritma random seçilen k adet merkez nokta ile başlar. Veri setindeki her bir nokta kendisine en yakın olan merkez noktaya ait kümeye atanır. Merkez değeri, kendine ait olan noktaların ortalaması alınarak hesap edilir. Bu işlem merkez değeri değişmeyene kadar tekrar edilir. Merkez değerinde bir hareketlenme olmadığında işlem sonlandırılır (Sarıman, 2011).

K-Medoids: K-Medoids algoritması, istatistiksel analizde en iyi bilinen bölme tabanlı kümeleme yöntemlerinden biridir. Bu yöntem gerçek veri setlerindeki gürültülü veriye karşı dayanıklıdır ve daha önce herhangi bir sınıflandırma bilgisi bulunmayan verinin bölümlendirilmesinde kullanılabilir (Modak, 2019). Veri setindeki nesnelerin çeşitli yapısal özelliklerine göre medoid olarak adlandırılan k adet temsilci nesneyi bulma esasına dayanan bir tekniktir. Medoid (temsilci nesne) kümedeki tüm nesnelerin en merkezi konumudur (Hamzaoui ve ark., 2018).

Kohonen Ağları: 1981'de Teuvo Kohonen tarafından geliştirilmiş

kohonen ağları, diğer adı ile KDH (Kendini Düzenleyen Haritalama), yapay sinir ağlarının bir türüdür (Kohonen, 2013). Çok boyutlu verinin tek veya iki boyutlu uzayda görsel sunumu amaçlanarak ortaya atılan bu teknik; kümeleme, sınıflandırma, görselleştirme ve soyutlama gibi görevler için geliştirilmiştir (Bikari ve ark., 2017). Veri setinde bilinmeyen örüntülerin ortaya çıkartılmasında kullanılan, geniş uygulama alanına sahip veri madenciliği yöntemlerinden biridir. KDH, büyük ölçekli veri setlerindeki doğal küme sayısını bulmada başarılı, denetimsiz iki katmanlı, giriş ve çıkış nöronlarından oluşan özel yapay sinir ağıdır. Bu teknikte veri setindeki değişken sayısı, giriş nöronlarının sayısını belirler. Çıkış nöronlarının her biri ise bir kümeyi simgeler. Çıkış katmanındaki nöronların dizilimi topolojik komşuluk açısından önemlidir. Deneme – yanılma yoluyla karar verilen dizilim doğrusal, altıgen, küp veya dikdörtgensel biçiminde olabilir. Uygulamada en çok dikdörtgensel ve altıgen dizilimler tercih edilmektedir (Taşkın ve Emel, 2010: Özçalıcı, 2016).

Bulanık C- Ortalamaları: Bulanık C-Ortalamalar uygulanması kolay, büyük verileri kümeleyebilme, her bir veri noktasının tam olarak bir küme veya birden çok küme içerisine yerleştirebilme gibi avantajları olan, geleneksel kümeleme yaklaşımlarından farklı olarak kararlılık oranları açısından daha iyi çıkışa ve yüksek sınıflandırma testi verimliliğine sahip (Anamisa ve ark., 2018), hızlı ve iyi sonuçlar üreten gözetimsiz bir kümeleme yöntemidir (Zare ve Koch, 2018). Bu teknik, veri noktaları arasındaki mesafeye dayalı analiz için kullanılır. Kümeler, veri noktaları arasındaki mesafeye göre oluşur ve her küme için küme merkezleri oluşturulur. Bir veri kümesi, her küme için veri kümesindeki merkez noktasına sahip n adet kümeye gruplanır ve her veri gruplandığı küme içindeki nesnelerle yüksek düzeyde benzerdir (Ghosh ve Dubey, 2013).

3.5.2.2. Birliktelik Kuralları Analizi

Birliktelik kuralları, veri tabanında mevcut olan parametreler arasındaki ilişkileri saptamak için kullanılan etkili bir algoritmadır. Ayrıca bir birliktelik

kuralı, veri tabanındaki spesifik durumların senkronize bir şekilde meydana gelmeleriyle bağlantılı bir olasılık ifadesidir ve genelden özele bir arama yöntemi kullanır. Bir birliktelik kuralı A→B şeklinde gösterilir. Bu ifade ile A'nın (öncül) meydana gelmesi durumunda B'nin (sonuç) de meydana gelmesi durumu anlatılmaktadır (Agrawal ve ark., 1993). Makine öğrenimi literatüründe birçok farklı kural bulma algoritması mevcutsa da, birliktelik kuralı algoritmaları, çok büyük veri kümelerinde nispeten verimli bir şekilde çalışmak üzere özel olarak tasarlanmıştır (Hand ve ark., 2001).

Birliktelik kuralları analizinde birçok birliktelik kuralları oluşmaktadır. Bu birliktelik kurallarının hangilerinin güvenilir, sık kullanılabilir veya anlamlı olduğunu belirlemek için destek (support), güven (confidence) ve kaldıraç (lift) değerleri ölçüt olarak kullanılmaktadır.

Destek (Support), bir seçenek veya seçenek kümesinin tüm gözlemler içerisindeki oranını gösterir. X ve Y seçenek kümesi için; X ve Y'nin seçenek kümesinin görülmesi sayısı N(X,Y) ve gözlem sayısı n olmak üzere destek oranının formülü aşağıdaki şekildedir.

$$P(X, Y) = N(X, Y) / n \qquad (3.2.)$$

Birliktelik kurallarının oluşturulabilmesi için minimum destek değeri D'nin önceden belirlenmesi gerekir. $0 \leq D \leq 1$ aralığındaki bu değer sıfıra yaklaştıkça birlikteliklerin sayısı artar, 1'e yaklaştığında birliktelik bulmak zorlaşacağından birlikteliklerin sayısı azalır. Uygulamada genelde 0.30 alınır. Güven (Confidence), X seçeneğini seçenlerin Y seçeneğini de seçme olasılığıdır. Yani, X verilmişken Y'nin koşullu olasılığını ifade eder ve aşağıdaki gibi hesaplanır.

$$P(Y/X) = P(X, Y) / P(X) \qquad (3.3.)$$

Burada P(X,Y) X ve Y seçeneğinin destek değerini, P(X) ise X seçeneğinin destek değerini göstermektedir.

Güven kriteri (G) $0 \leq G \leq 1$ aralığında yer alır. Uygulamada genelde 0.50 veya üzerinde bir değer alınır (Altunkaynak, 2017).

Kaldıraç (Lift): Birliktelik kurallarının ne kadar ilginç olduğunu gösteren değerdir. Lift değerlerinin 1'den büyük olması birliktelik kurallarının ilginç olduğu anlamına gelmektedir. Lift değerinin formülü aşağıdaki gibi hesaplanmaktadır (Ateş ve Karabatak, 2017).

$$\text{Lift } (X \rightarrow Y) = P(Y/X) / P(Y) \qquad (3.4.)$$

Burada, P(Y/X) X verilmişken Y'nin koşullu olasılığını, P(Y) ise Y seçeneğinin destek değerini göstermektedir.

Sonuç olarak, birliktelik kuralları analizlerinde kullanıcı tarafından belirlenen destek (support), güven (confidence) ve kaldıraç (lift) değerlerine göre birliktelik kuralları oluşturulur. Bu değerlerin altında kalan birliktelik kuralları ise anlamsız olarak ifade edilmektedir. Birliktelik kuralları analizinde Apriori, AprioriTid, AIS, SETM, FP-Growth ve Eclat gibi algoritmalar kullanılmaktadır.

3.6. Veri Madenciliği Yazılımları

Veri madenciliği analizlerinin yapılması için ticari ve açık kaynak kodlu (ücretsiz) yazılımlar kullanılmaktadır. Uygulamalarda büyük ve karmaşık veri kümeleri analize uygun hale getirilerek veri madenciliği yazılımları ile analizlerin yapılması sağlanmaktadır. Böylece bu yazılımlar sayesinde analizler kolay bir şekilde ve kısa sürede yapılmaktadır.

Veri madenciliği analizlerinde kullanılan bazı ticari ve açık kaynak kodlu (ücretsiz) yazılımlar aşağıda gösterilmiştir.

Ticari yazılımlar: IBM SPSS Modeler (SPSS Clementine), SAS Enterprise Miner, KXEN, SQL Server, Matlab ve Oracle.

Açık kaynak kodlu (ücretsiz) yazılımlar: RapidMiner, WEKA, R, Orange, KNIME, Tanagra ve KEEL.

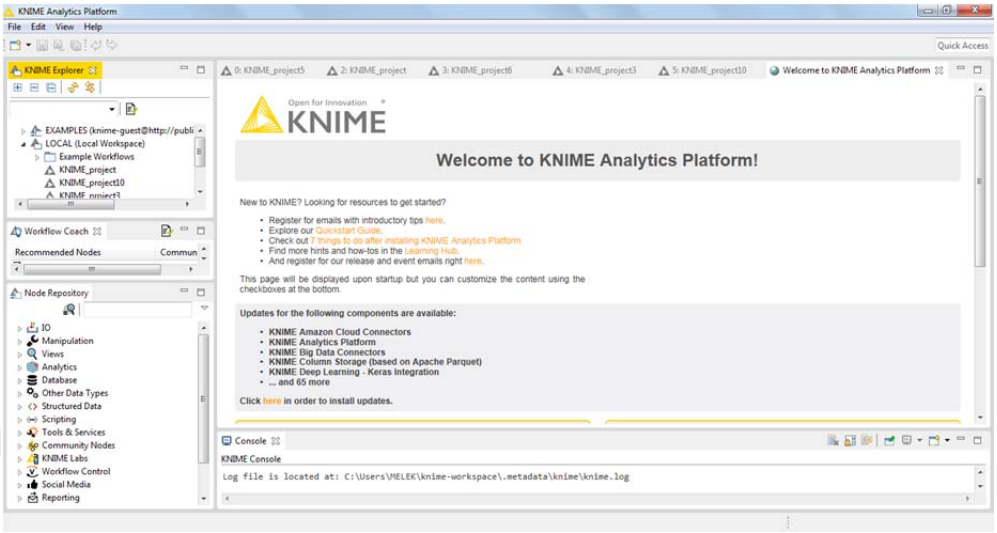
4. MATERYAL VE YÖNTEM

4.1. Veri Madenciliği Programları

Tez çalışmasında istatistik ders içerikleri için yapılan kümeleme ve birliktelik kuralları analizlerinde KNIME programı, karar ağaçları analizinde ise WEKA 3.8 programı kullanılmıştır.

4.1.1. KNIME

KNIME (Konstanz Information Miner) adlı veri madenciliği geliştirme yazılımı Konstanz Üniversitesi görsel veri madenciliği araştırma grubu tarafından Eclipse Rich Client Platformu üzerinde geliştirilmiştir. KNIME, modüler ve görsel veri akış sistemi geliştirme ortamı, eğitim ve araştırma amaçlı ortak çalışma ortamlarını içermektedir. Veri madenciliği yöntemlerinden destek vektör makinaları, bayes, multi dimesional scaling (MDS) gibi yöntemleri içerir. KNIME, Regresyon, Korelasyon ve Korelasyon Filtresi gibi istatistik tabanlı yöntemlerin veri akış tasarımında kullanılmasına imkân sağlar. Diğer yazılımlara göre en zengin görselleştirme araçlarına sahiptir. Scatter Plot, Parallel Coordinates, Box Plot ve Histogram gibi birçok görselleştirme araçları ve JFreeChart bileşeni sayesinde çok daha ayrıntılı Java tabanlı görselleştirme düğümleri bulunmaktadır (Bilgin, 2009).



Şekil 4. 1. KNIME Analiz Platformu

KNIME temel versiyonu, veri ön işleme ve temizleme, analizler ve veri madenciliği de dahil scatter plots, parallel coordinates ve bir çok interaktif görüntüleme gibi 100'den fazla iş parçasını G/Ç bilgisi olarak birleştirir. WEKA'nın veri madenciliğinde kullanılan en iyi bilinen bütün analiz modülleri ve R-scripts'in eklentilerinin çalıştırılmasına, istatistik kütüphanelerinin kullanılmasını sağlar. KNIME modüler API özelliği ile Eclipse platformunu temel alır ve kolayca genişleyebilir. Bu modülerlik ve genişleyebilirlik özelliği KNIME'nin öğretim ve araştırma özelliklerinin yanında ticari olarakta kullanılmasını sağlar. KNIME çok fazla fonksiyonelliklere sahiptir.

G/Ç: Dosyalardan ya da veri tabanlarından veri alış verişi yapar.

Veri Manipülasyon: Filtreleyerek veri ön işleme, gruplama, pivot, kovalama, normalleştirme, toplama, karıştırma, örnekleme, bölümleme gibi işlemleri yapar.

Görüntüleme: Veriyi ve sonuçlarını birçok görüntüleme aracı ile verinin keşfinin sağlanmasını kolaylaştırır.

Madencilik: Demetleme, karar ağaçları, kural oluşturma, ilişki kuralları,

sinir ağları, destek vektör makineleri gibi uygulamaları gerçekleştirir (Tekerek, 2011).

4.1.2. WEKA

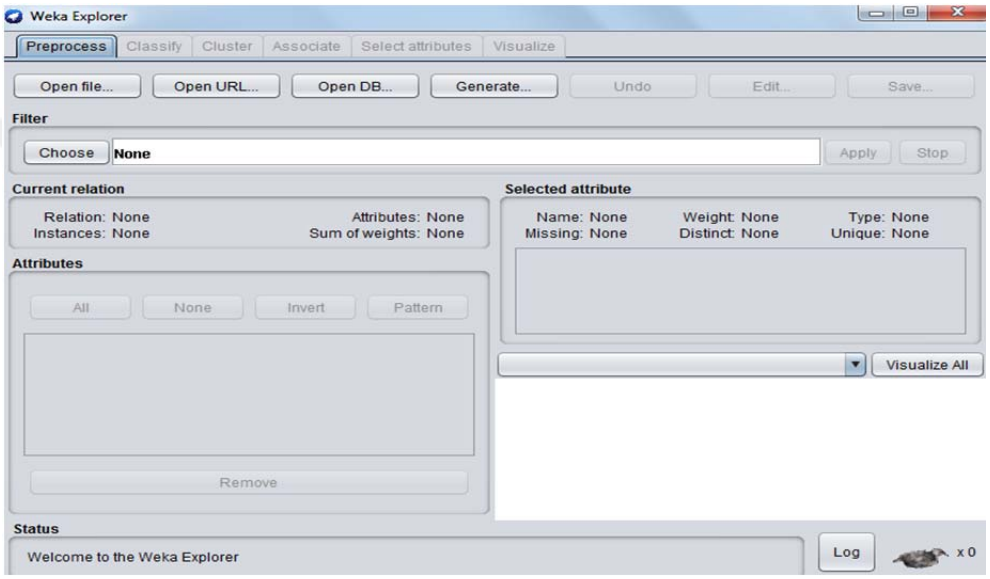
WEKA, veri madenciliği yapmak amacıyla, Java tabanlı, açık kaynak kodlu, Waikato Üniversitesi'nde (Yeni Zelanda) geliştirilmiş ve makine öğrenme algoritmalarından oluşan bir programdır (Dean, 2014). Seçilen algoritmalar herhangi bir veri setine doğrudan uygulanabileceği gibi Java kodları çağırılarak da uygulanabilir. Program veri ön işleme, sınıflandırma, kümeleme, ilişki kuralları, özellik seçimi ve görselleştirme için hazır araçlar içerir. Eklentilerin yüklenmesiyle çeşitli 3 boyutlu görüntüleme ve programlama dilleri de kullanılabilir (Yalçın, 2019).



Şekil 4. 2. WEKA Giriş Ekranı

WEKA'da analiz edilecek veriyi sisteme yüklemek için Şekil 4.2'deki Explorer butonuna tıklanır. Ardından ekrana Şekil 4.3'te gösterilen Preprocess sekmesinden Open file butonuna tıklayarak analiz edilecek veri dosyası sisteme yüklenir. Eklenen veri setindeki bilgilere göre kaç adet attribute (özellik) ve satır

olduğu Current relation bölümünden görülebilmektedir. Ayrıca her attribute ait istatistiksel bilgi ve veri tiplerine ait bilgiler Selectes attribute bölümünde görüntülenir. Veri dosyası (ARFF, CSV) programa yüklendikten sonra WEKA ile veri ön işleme yapılabilir. Ön işlemden geçirilmiş veri yazılım içerisinde hazır olarak bulunan algoritmalar ile analiz edilir (Armutlu, 2018).



Şekil 4. 3. WEKA Kullanıcı Ara Yüzü Görüntüsü

4.2. İstatistik Bölümleri Ders İçeriklerinin Veri Madenciliği Yöntemleri ile Analizi

Tez çalışmasında, ülkemizde bulunan istatistik bölümüne sahip 24 üniversitenin ders içerikleri incelenmiştir. Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümüne sahip diğer üniversitelerin lisans programlarındaki derslerin ders içerikleri karşılaştırılmıştır ve 3 farklı veri madenciliği yöntemi ile analizler yapılmıştır. Üniversitelerin istatistik bölümlerinin birbirilerine olan benzerliklerini belirlemek için hiyerarşik kümeleme analizi yöntemi kullanılmıştır. KNIME programından seçilen hiyerarşik kümeleme düğümlerinin çalıştırılması ile

kümeleme analizi yapılmıştır. İstatistik bölümleri derslerine göre birbirilerine benzer olan üniversitelerin kümeleri analiz sonucunda gösterilmiştir.

İstatistik bölümlerinin ders programlarında yer alan derslerin hangilerinin birlikte alınması gerektiğini belirlemek için Apriori algoritması kullanılmıştır. Apriori algoritmasının kullanılması için KNIME programından seçilen düğümler ile birliktelik kuralları analizi gerçekleştirilmiştir. İstatistik bölümlerinde bulunması gereken dersler analiz sonucunda tespit edilmiştir.

İstatistik bölümlerinin başarılarına etki eden derslerin bulunması için ise J48 (C4.5) algoritması kullanılmıştır. WEKA programından seçilen J48 (C4.5) algoritması ile karar ağacı analizi yapılmıştır. Analiz sonucunda istatistik bölümlerinin başarılarına etki eden dersler belirlenmiştir.

4.2.1. Hiyerarşik Kümeleme Yöntemi

Hiyerarşik (hierarchical) ya da aşama sıralı kümeleme yöntemlerinde işleyişin kolay anlaşılabilmesi için ağaç diyagram (dendogram) örneğinden yararlanılmaktadır. Kümeleme sürecinin başlangıcında her birey bir kümedir, süreç sonunda ise tüm bireyler bir kümede toplanır. Bu işleyiş daha ayrıntılı bir biçimde aşağıdaki dört adımlı bir algoritma ile ifade edilebilir:

1.n tane birey, n tane küme olmak üzere işleme başlanır.

2.En yakın iki küme birleştirilir.

3.Küme sayısı bir indirgenerek yinelenmiş uzaklıklar matrisi bulunur.

4. 2 ve 3. adımlar n–1 kez tekrarlanır.

Bu algoritmaya dayalı sekiz farklı hiyerarşik teknikten söz edilmektedir. Bu teknikler, sırasıyla en yakın komşuluk, en uzak komşuluk, ortalama bağlantı, McQuitty bağlantı, Ward (minimum varyans), küresel ortalama bağlantı, ortanca ve Lance Willims teknikleridir (Terlemez, 2008).

Hiyerarşik kümeleme yöntemi birleştirici (agglomerative) kümeleme ve ayrıştırıcı (divisive) kümeleme olmak üzere iki grupta incelenmektedir. Bu tez

çalışmasında birleştirici (agglomerative) kümeleme yöntemi kullanılmıştır. Birleştirici (agglomerative) kümeleme: Kaufman ve Rousseeuw (1990) tarafından sunulmuştur. Aşağıdan yukarı inşa yapısı izler. Başlangıçta her bir nesneyi bağımsız bir küme olarak görür ve her adımda bu kümelerden benzer özellik gösterenleri birleştirir. Birleştirme işlemi bir sonlandırma koşulu sağlanana kadar sürer (Dinçer, 2006). KNIME programında yapılan hiyerarşik kümeleme analizi için veriler arasındaki uzaklık kosinüs benzerliği ile ölçülmüştür. Kosinüs benzerliği ölçüsünün açıklaması aşağıdaki gibi yapılmıştır.

Kosinüs benzerliği, iki doküman arasındaki benzerliği karşılaştırmakta kullanılır. "n" boyutlu iki vektör arasındaki açının bulunmasıyla elde edilir (Yüksel ve ark., 2010). Kosinüs benzerliğinin formülü aşağıdaki gösterilmiştir (Kaya ve Özel, 2014).

$$S = \frac{A \cdot B}{\| A \| \| B \|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2 \times \sum_{i=1}^{n} (B_i)^2}} \quad (4.1.)$$

4.2.2. Apriori Algoritması

Birliktelik kuralı için en çok bilinen strateji Apriori'dir (Liao ve Wen, 2007). Algoritmanın ismi, yaygın nesnelerin önsel bilgilerini kullanmasından yani bilgileri bir önceki adımdan almasından "önceki (prior)" anlamında aprioridir (Döşlü, 2008). Agrawal ve Srikant (1994) tarafından geliştirilen Apriori algoritması, veri madenciliği tarihinde birliktelik kurallarının çıkarılması konusunda elde edilmiş büyük bir başarıdır. Apriori algoritması veri tabanındaki verileri tekrarlayarak kaydeder ve her kayıttan sonra geniş veri kümelerini oluşturur. İşlemleri indirgemek için aday veri kümeleri için sadece destek seviyeleri hesaplanır (Liao ve Wen, 2007). Apriori algoritmasında k öğeli sık geçen öğe küme adayları, (k-1) öğeli sık geçen öğe kümelerinden faydalanılarak bulunur.

Ancak bu algoritma veri tabanının pek çok kere taranmasını gerektirmektedir. Takipteki taramalarda bir önceki taramada bulunan sık geçen öğe kümeleri aday kümeleri adı verilen yeni potansiyel sık geçen öğe kümelerini üretmek için kullanılır. Aday kümelerin destek değerleri tarama sırasında hesaplanır ve aday kümelerinden minimum destek metriğini sağlayan kümeler o geçişte üretilen sık geçen öğe kümeleri olur. Sık geçen öğe kümeleri bir sonraki geçiş için aday küme olurlar. Bu süreç yeni bir sık geçen öğe kümesine rastlanıncaya kadar devam eder (Bilen, 2009).

Apriori algoritmasında sık geçen öğe kümeleri seçilen destek eşik değerine göre birliktelik kuralları belirlemektedir. Bu kurallar için güven değerleri hesaplanarak belirlenen güven eşik değerine göre birliktelik kuralları oluşturulmaktadır.

4.2.3. C4.5 (J48) Algoritması

C4.5 algoritması, yine ID3'ün tasarımcısı olan Quinlan (1993) tarafından geliştirilmiştir. Temel olarak entropiye dayalı bu algoritmanın ID3'e göre en büyük farkı sayısal veriler üzerinde de çalışabilmesi ve eksik verileri de işleyebilme yeteneğidir. Yine bununla birlikte C4.5, ağaç üzerinde oluşan gereksiz yaprakları budama (prunning) özelliğine sahiptir. ID3 algoritmasının değişkenleri birçok alt bölüme ayırması sırasında yaşanılan aşırı öğrenme durumunun önüne geçilebilmesi için C4.5 algoritmasında ID3'ten daha farklı bir 'kazanım' oranı kullanılmaktadır (Dunham, 2003: Silahtaroğlu 2008).

C4.5 algoritmasının çalışma şekli aşağıdaki gibi açıklanmaktadır.

- $\sum_{i=1}^{n} p_i \log_2 p_i$ formülüne göre veri kümesindeki tüm sütunların entropisi hesaplanır (Shannon, 1948).
- Veri kümesindeki her bir sütunun entropisi, sınıfın entropisine bölünerek her bir sütunun kazanım oranı hesaplanır.

- Kazanım oranı en yüksek olan kök düğüm olarak atanır.

Kök düğüm belirlendikten sonra yukarıda açıklanan işlemler kalan veriler için tekrardan yapılmaktadır ve bütün veriler bitene kadar bu işlemlere devam edilmektedir. C4.5 algoritması için yapılan işlemler tüm verilere uygulandıktan sonra analiz sonucunda karar ağacı oluşturulmaktadır. Bu karar ağacından yararlanılarak elde edilen karar kuralları ile veri kümesinde bulunan kayıtlar sınıflandırılmaktadır.

5. UYGULAMA VE ANALİZ

5.1. Araştırmanın Tanımı

Üniversitelerin aynı bölümlerinin ders içerikleri incelendiğinde, derslerin birbirinden farklı olduğu görülmüştür. Bu yüzden, farklı üniversitelerin aynı bölümündeki ders içeriklerinin birbirleriyle olan benzerlikleri, bu benzerlikler ile üniversitelerin birbirleriyle ne kadar ilişkili olduğu merak edilen araştırma konularından biri olmuştur. Bunlara ek olarak, yeni açılacak üniversitelerin ders planlanması hazırlanırken hangi derslerin bulunması gerektiği de araştırma konusu ile yakından ilişkilidir. Bu durumlara çözüm bulmak amacıyla uygulamada, derslerin birbiriyle olan ilişkilerini belirlemek için üniversitelerin aynı bölümündeki eğitim programlarında bulunan derslerin ders içerikleri incelenecektir.

Tez çalışmasında, Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümü bulunan diğer üniversitelerin lisans programlarındaki derslerin ders içerikleri karşılaştırılmıştır. Karşılaştırılan ders içeriklerinin sonuçları bölüm dersleri dosyası olarak adlandırılmıştır ve Excel dosyası olarak kaydedilmiştir. Bölüm dersleri dosyasına çeşitli veri madenciliği yöntemleri uygulanarak 3 farklı analiz yapılmıştır. Analizlerde KNIME ve WEKA programları kullanılmıştır.

KNIME programından kümeleme ve birliktelik kuralları analizleri için seçilen düğümlerin bölüm dersleri dosyasına uygulanması ile analizler yapılmıştır. WEKA programında ise bölüm dersleri dosyası için J48 (C4.5) algoritması kullanılarak karar ağacı analizi gerçekleştirilmiştir.

Birinci analizde, istatistik bölümü bulunan üniversitelerin birbirleriyle olan benzerliklerini bulmak için kümeleme analizi yöntemlerinden biri olan hiyerarşik kümeleme analizi yöntemi kullanılmıştır. İkinci analizde, istatistik bölümlerinin eğitim programlarındaki hangi derslerin birlikte alınması gerektiğini tespit etmek için birliktelik kuralları analizi algoritmalarından biri olan Apriori algoritması ile analiz gerçekleştirilmiştir. Üçüncü analizde ise istatistik bölümlerinin başarılarına

etki eden derslerin hangilerinin olduğunu belirlemek amacıyla karar ağacı analizi algoritmalarından biri olan J48 (C4.5) algoritması kullanılmıştır. Böylece yeni açılacak üniversitelerdeki istatistik bölümlerinin ders planlanması hazırlanırken hangi derslerin bulunması gerektiği ve bu derslerin içeriklerinin nasıl olacağı hakkında bilgilerin bulunması amaçlanmıştır.

5.2. Verilerin Toplanması ve Hazırlanması

Veri toplama işlemi için üniversitelerin istatistik bölümlerini belirlemek amacıyla lisans programlarının bilgisinin yer aldığı YÖK'ün Lisans Atlası ve Yüksek Öğretim Bilgi Yönetim Sistemi web sayfaları referans olarak alınmıştır. Referans alınan YÖK'ün Lisans Atlas'ında yer alan tüm devlet ve vakıf üniversiteleri tez çalışmasına dâhil edilmiştir. Uygulamada üniversitelerin istatistik bölümlerinin ve ders içeriklerinin tespit edilmesi 2 aşamada yapılmıştır.

Birinci aşamada, Çukurova Üniversitesi İstatistik Bölümü'nün ders programında yer alan dersler Excel dosyasına girilmiştir. Referans alınan YÖK'ün Lisans Atlas'ında tespit edilen istatistik bölümü bulunan 25 üniversitenin isimleri belirlenerek Excel'e kaydedilmiştir. Bu üniversitelerin, istatistik bölümlerinin ders içeriklerine erişmek için her üniversitenin kendi web sayfasına giriş yapılarak ders programlarındaki ders içerikleri incelenmiştir. İstatistik bölümlerinin ders içerikleri incelenirken üniversitelerin web sayfalarındaki AKTS bilgi paketinden yararlanılmıştır. Ancak, AKTS bilgi paketinde sıkıntı bulunan Kırıkkale Üniversitesi İstatistik Bölümü tez çalışmasından çıkarılmıştır.

İkinci aşamada ise istatistik bölümlerinin ders içerikleri incelenerek Excel'de iki tablo oluşturulmuştur. Birinci tabloya istatistik bölümü bulunan 24 üniversitenin isimleri ile Çukurova Üniversitesi İstatistik Bölümü'nün ders programında yer alan dersler yazılmıştır. Tabloya bölüm dersleri yazıldıktan sonra, Çukurova Üniversitesi İstatistik Bölümü'nün ders içeriği ile diğer üniversitelerin istatistik bölümlerinin ders içerikleri karşılaştırılmıştır. Bu tabloya, derslerin içerikleri aynı ise 1, aynı değil ise 0 yazılmıştır. İkinci tabloya ise ders içerikleri

aynı fakat ders isimleri farklı olan dersler eklenmiştir. Derslerin eş değerlik tablosu Ek 1'de gösterilmiştir.

5.3. URAP İndeksi

URAP Araştırma Laboratuvarı 2009 yılında Orta Doğu Teknik Üniversitesi Enformatik Enstitüsü bünyesinde kurulmuştur. URAP' ın amacı yükseköğretim kurumlarını akademik başarıları doğrultusunda değerlendirebilmek için bilimsel metodlar geliştirmek ve yapılan çalışmaların sonuçlarını kamuoyu ile paylaşmaktır (URAP, 2018).

URAP indeksi üniversitelerin başarı sıralamalarını belirlerken, üniversitelerdeki öğretim üyelerinin bulunduğu yılda yayınladığı makale puanını, doktora mezun öğrenci puanını, öğretim üyelerine düşen öğrenci sayılarının puanlarını, bilimsel doküman sayısının toplam puanını ve öğretim üyelerinin toplamda yaptığı atıf puanını kullanmaktadır.

2018-2019 yılları arasında URAP tarafından açıklanan Türkiye'deki üniversitelerin akademik performansına göre başarı sıralamaları tablosu Ek 2'de gösterilmiştir.

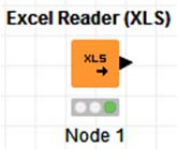
5.4. İstatistik Bölümü Derslerine Göre Üniversitelerin Kümeleme Analizi İle Belirlenmesi

5.4.1. Veri Dönüşüm İşlemleri

Uygulama için KNIME programının düğüm kutusu (Node Repository) panelinden kümelemede kullanılması gereken düğümler iş akışı ekranına yerleştirilmiştir. Bu düğümler uygun bağlantı noktaları ile bağlanmıştır ve her bir düğüm ayrı ayrı konfigüre edilerek çalıştırılmıştır. Kümeleme için seçilen düğümlerin açıklamaları aşağıda verilmiştir.

1. Excel Reader

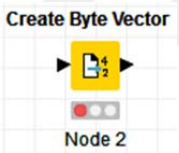
Excel Reader düğümü sayısal, karakter ya da ikili türünden oluşan veri dosyalarını okumaktadır.



Şekil 5. 1. Excel Reader (XLS) Düğümü

2. Create Byte Vector

Create Byte Vector düğümü veri kümesinden seçilen her bir satırı bir bit vektöre dönüştürmektedir.



Şekil 5. 2. Create Byte Vector Düğümü

3. Byte Vector Distances

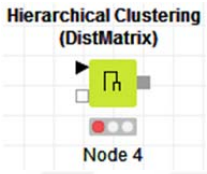
Byte Vector Distances düğümü bir verinin diğer verilere olan uzaklığını ölçmektedir. Bu düğüm, Öklid, Minkowski, Cosine ya da Manhattan uzaklıklarından birini kullanarak uzaklık hesaplamaktadır.



Şekil 5. 3. Byte Vector Distances Düğümü

4. Hierarchical Clustering (DistMatrix)

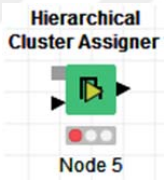
Hierarchical Clustering (DistMatrix) düğümü veri olarak uzaklık vektörünü seçmektedir. Bu düğüm Tek bağlantı, Tam bağlantı ya da Ortalama bağlantı yöntemlerinden birini kullanarak kümeler arasındaki mesafeyi belirlemektedir.



Şekil 5. 4. Hierarchical Clustering (DistMatrix) Düğümü

5. Hierarchical Cluster Assigner

Hierarchical Cluster Assigner düğümü veriler için seçilen küme sayısına göre verileri kümelemektedir.



Şekil 5. 5. Hierarchical Cluster Assigner Düğümü

6. Hierarchical Cluster View

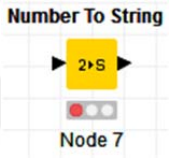
Hierarchical Cluster View düğümü kümelenmiş verileri görselleştirmek için ağaç yapısı şeklindeki dendogram grafiğini kullanmaktadır.



Şekil 5. 6. Hierarchical Cluster View Düğümü

7. Number To String

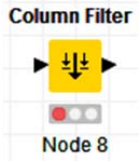
Number To String düğümü veri kümesindeki sütun değerlerini sayısal veri türünden karakter veri türüne dönüştürülmesini sağlamaktadır.



Şekil 5. 7. Number To String Düğümü

8. Column Filter

Column Filter düğümü veri tablosundaki istenmeyen sütunları filtrelemek için kullanılmaktadır.



Şekil 5. 8. Column Filter Düğümü

9. Color Manager

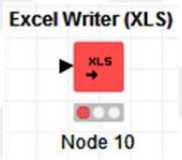
Color Manager düğümü verilerin farklı renklerde gösterilmesini sağlamaktadır.



Şekil 5. 9. Color Manager Düğümü

10. Excel Writer (XLS)

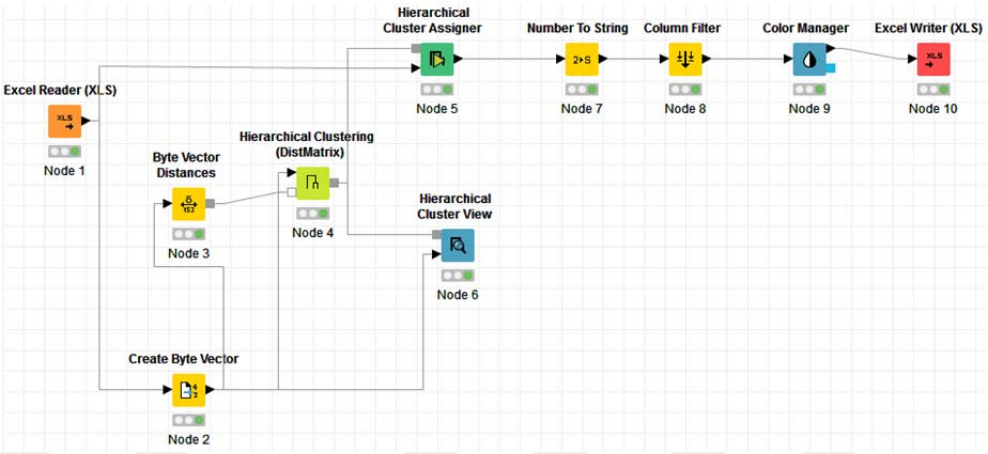
Excel Writer (XLS) düğümü veri dosyalarını Excel dosyası olarak kaydetmek için kullanılmaktadır.



Şekil 5. 10. Excel Writer (XLS) Düğümü

5.4.2. Kümeleme Analizinin Uygulanması

Kümeleme analizi için seçilen düğümler Şekil 5.11'deki gibi KNIME programının iş akışı ekranına yerleştirilip bağlantıları kurularak çalıştırılmıştır.



Şekil 5. 11. Kümeleme Analizi İçin Düğümlerin Çalıştırılması

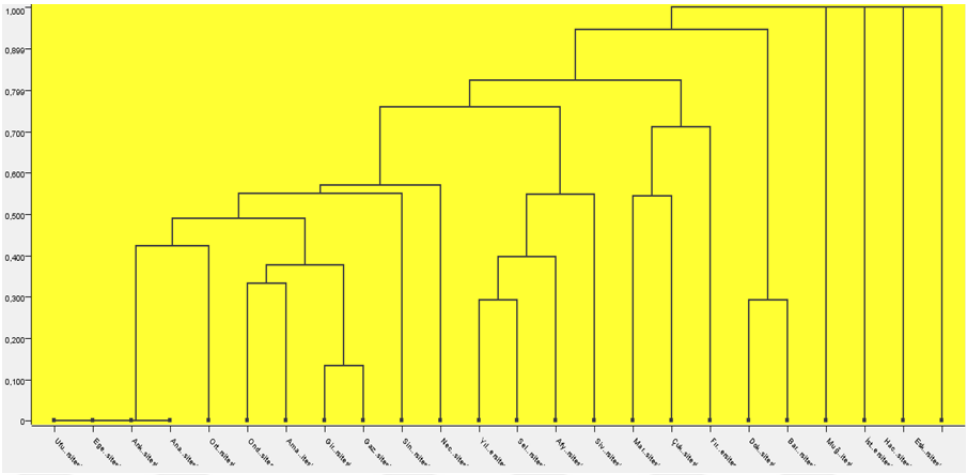
Uygulamada ilk olarak KNIME programının iş akışı ekranına yerleştirilen Excel Reader düğümü, bölüm dersleri dosyasını okumuştur ve Create Byte Vector düğümüne bağlanmıştır. Excel Reader düğümüne yüklenen bölüm dersleri verisinin her bir satırında bulunan üniversiteler Create Byte Vector düğümü tarafından bir bit vektöre dönüştürülmüştür. Sonra Create Byte Vector düğümüne Byte Vector Distances düğümü bağlanmıştır. Bu düğüm için uygun uzaklık ölçüsü Cosine uzaklığı seçilerek bit vektörler arasındaki uzaklık (üniversiteler arasındaki uzaklık) hesaplanmıştır. Byte Vector Distances düğümü ise Hierarchical Clustering (DistMatrix) düğümüne bağlanarak kümeler arasındaki uzaklığı bulmak için Ortalama Bağlantı yöntemini kullanmıştır. Daha sonra Hierarchical Clustering (DistMatrix) düğümü de hem Hierarchical Clustering Assigner düğümüne hem de Hierarchical Cluster View düğümüne bağlanmıştır. Hierarchical Clustering Assigner düğümü küme sayısının 5 seçilmesi ile üniversitelerin bölüm derslerine göre kümelenmesini sağlamıştır. Ayrıca, farklı küme sayıları da Hierarchical Clustering Assigner düğümünde kullanılmıştır. Farklı küme sayılarının analiz sonuçlarında çok sayıda üniversitenin bir küme içinde yer aldığı görülmüştür. Bu durum analizde oluşan kümelerin doğruluğunu azaltmaktadır.

Hierarchical Clustering Assigner düğümü çalıştırıldıktan sonra bölüm derslerinin bulunduğu tablonun son satırına Çizelge 5.1'de görülen Cluster Number sütunu eklenmiştir. Cluster Number sütunu, üniversitelerin hangi kümeye ait olduğu bilgisini göstermektedir.

Çizelge 5. 1. Üniversitelerin Kümeleme Sonuçları

Row ID	Lineer ...	Çok De...	Araştır...	Yöneyl...	Demografi	Lojistik ...	Ekonom...	İstatisti...	Cluster number
Çukurova Üniversitesi	1	1	1	1	1	1	1	1	3
Afyon Kocatepe Üniversitesi	1	0	0	0	0	0	0	0	3
Amasya Üniversitesi	0	1	0	0	0	0	0	0	3
Anadolu Üniversitesi	0	0	0	0	0	0	0	0	3
Ankara Üniversitesi	0	0	0	0	0	0	0	0	3
Bartın Üniversitesi	0	0	0	0	0	0	0	0	3
Dokuz Eylül Üniversitesi	0	0	0	0	0	0	0	0	3
Ege Üniversitesi	0	0	0	0	0	0	0	0	3
Eskişehir Osmangazi Üniversitesi	0	0	0	0	0	0	0	0	0
Fırat Üniversitesi	0	0	0	0	0	0	0	0	3
Gazi Üniversitesi	0	1	0	0	0	0	0	0	3
Giresun Üniversitesi	0	1	0	0	0	0	0	0	3
Hacettepe Üniversitesi	0	0	0	0	0	0	0	0	1
İstanbul Aydın Üniversitesi	0	0	0	0	0	0	0	0	2
Marmara Üniversitesi	0	0	0	0	0	0	0	0	3
Muğla Sıtkı Koçman Üniversitesi	0	0	0	0	0	0	0	0	4
Necmettin Erbakan Üniversitesi	1	1	0	0	0	0	0	0	3
Ondokuz Mayıs Üniversitesi	0	1	0	0	0	0	0	0	3
Orta Doğu Teknik Üniversitesi	0	0	0	0	0	0	0	0	3
Selçuk Üniversitesi	1	0	0	0	0	0	0	0	3
Sinop Üniversitesi	0	1	0	0	0	0	0	0	3
Sivas Cumhuriyet Üniversitesi	1	0	0	0	0	0	0	0	3
Ufuk Üniversitesi	0	0	0	0	0	0	0	0	3
Yıldız Teknik Üniversitesi	1	0	0	0	0	0	0	0	3

Hierarchical Clustering (DistMatrix) düğümünün diğer bağlandığı Hierarchical Cluster View düğümü ise kümeleme sonucunda birbirine benzerlik gösteren üniversitelerin dendogram grafiğini oluşturmuştur. İstatistik bölümleri derslerine göre kümelenen bu üniversitelerin dendogram grafiği Şekil 5.12'de gösterilmiştir.



Şekil 5. 12. Kümelenen Üniversitelerin Dendogram Grafiği

Hierarchical Cluster Assigner düğümüne ise Number To String düğümü bağlanmıştır. Bu düğüm bölüm dersleri sütunlarını ve kümeleme sonucunda oluşan Cluster Number sütununu, sayısal veri türünden karakter veri türüne dönüştürmüştür. Number To String düğümü de kümelenen üniversitelerin renklendirilmesini gösteren tabloda (Çizelge 5.2) bölüm derslerinin (Çizelge 5.1) gösterilmemesi için kullanılan Column Filter düğümüne bağlanmıştır. Column Filter düğümünün Color Manager düğümüne bağlanması ile kümelenen üniversiteler renklendirilmiştir ve Çizelge 5.2'de gösterilmiştir. Böylece aynı renkte olan üniversiteler aynı küme içinde gösterilmiştir.

Çizelge 5. 2. Kümelenen Üniversitelerin Renklendirilmesi

Row ID	S Cluster number
Çukurova Üniversitesi	3
Afyon Kocatepe Üniversitesi	3
Amasya Üniversitesi	3
Anadolu Üniversitesi	3
Ankara Üniversitesi	3
Bartın Üniversitesi	3
Dokuz Eylül Üniversitesi	3
Ege Üniversitesi	3
Eskişehir Osmangazi Üniversitesi	0
Fırat Üniversitesi	3
Gazi Üniversitesi	3
Giresun Üniversitesi	3
Hacettepe Üniversitesi	1
İstanbul Aydın Üniversitesi	2
Marmara Üniversitesi	3
Muğla Sıtkı Koçman Üniversitesi	4
Necmettin Erbakan Üniversitesi	3
Ondokuz Mayıs Üniversitesi	3
Orta Doğu Teknik Üniversitesi	3
Selçuk Üniversitesi	3
Sinop Üniversitesi	3
Sivas Cumhuriyet Üniversitesi	3
Ufuk Üniversitesi	3
Yıldız Teknik Üniversitesi	3

Son olarak Color Manager düğümüne Excel Writer (XLS) düğümü bağlanmıştır ve kümelenen üniversiteler Excel dosyası olarak kaydedilmiştir.

5.4.3. Kümeleme Analizinin Sonuçlar ı

Kümeleme analizi istatistik bölüm derslerine göre birbirine benzerlik gösteren üniversitelerin kümelenmesini sağlamıştır. Uygulamada hiyerarşik kümeleme analizi yöntemi kullanılmıştır. Bu analiz sonucunda; dendogram grafiği ve üniversitelerin benzerliklerine göre hangi kümede bulunduğunu gösteren tablo oluşturulmuştur. Dendogram grafiği istatistik bölümleri derslerine göre hangi üniversitelerin aynı kümede olduğunu göstermektedir. Küme sayısının 5 seçilmesi ile oluşan Çizelge 5.1 tablosunda üniversitelere ait küme numaraları 0, 1, 2, 3 ve 4 şeklinde Cluster Number sütununda gösterilmiştir. Küme numaraları 0, 1, 2 ve 4

olan kümelerin eleman sayıları toplamı 4'tür. Bu durum istatistik bölüm derslerine göre birbirine benzemeyen üniversite sayısının toplamda 4 olduğunu göstermektedir. Küme numarası 3 olan kümenin eleman sayısı ise 20'dir. Bu ifade ise istatistik bölüm derslerine göre birbirine benzerlik gösteren üniversite sayısının 20 olduğunu açıklamaktadır.

Sonuç olarak; Çizelge 5.2'de mor renkte gösterilen 3 numaralı kümede bulunan Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümüne sahip diğer 19 üniversitenin kümeleme analizi sonucunda birbirlerine benzerlik gösterdiği ifade edilmiştir.

5.5. İstatistik Bölümü Derslerinin Birbirleri İle Olan İlişkilerinin Birliktelik Kuralları Analizi İle Belirlenmesi

5.5.1. Veri Dönüşüm işlemleri

KNIME programının düğüm kutusu (Node Repository) panelinden seçilen düğümler, istatistik bölümü derslerine göre en sık birlikte alınan derslerin belirlenmesi için iş akışı ekranına yerleştirilmiştir. Bu düğümler uygun bağlantı noktaları ile bağlanmıştır ve her bir düğüm ayrı ayrı konfigüre edilerek çalıştırılmıştır. Birliktelik kuralları analizi için seçilen düğümlerin açıklamaları aşağıda verilmiştir.

1. ARRF Reader

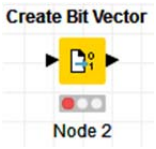
ARFF Reader düğümü ARFF türündeki veri dosyalarını okumaktadır.



Şekil 5. 13. ARFF Reader Düğümü

2. Create Bit Vector

Create Bit Vector düğümü veri kümesinden seçilen her bir satırı bir bit vektöre dönüştürerek, oluşturulan bit vektörleri veri kümesine sütun olarak eklemektedir.



Şekil 5. 14. Create Bit Vector Düğümü

3. Association Rule Learner

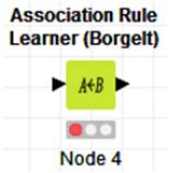
Association Rule Learner düğümü veri kümesindeki nesneler için öğe kümeleri tablosu oluşturmaktadır. Ayrıca bu tabloda öğe kümelerine ait destek ve güven değerleri bulunmaktadır.



Şekil 5. 15. Association Rule Learner Düğümü

4. Association Rule Learner (Borgelt)

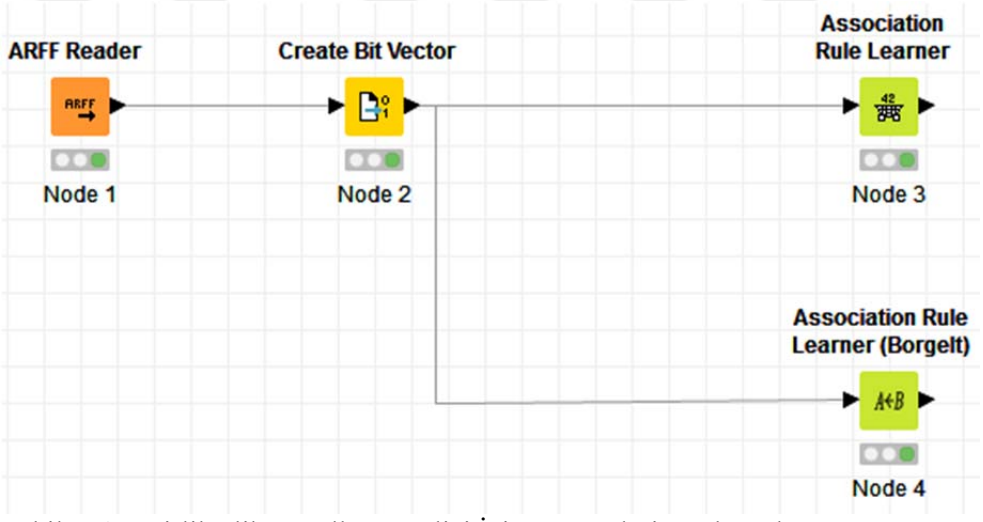
Association Rule Learner (Borgelt) düğümü Apriori algoritması kullanarak öğelerin birbiri ile ilişkilerini gösteren tablo oluşturmaktadır. Bu tablo, birliktelik kuralları ve birliktelik kurallarına ait ölçüleri göstermektedir.



Şekil 5. 16. Association Rule Learner (Borgelt) Düğümü

5.5.2. Birliktelik Kuralları Analizinin Uygulanması

Birliktelik kuralları analizi için seçilen düğümler Şekil 5.17'deki gibi KNIME programının iş akışı ekranına yerleştirilerek çalıştırılmıştır.



Şekil 5. 17. Birliktelik Kuralları Analizi İçin Düğümlerin Çalıştırılması

Uygulama için KNIME programının iş akışı ekranında ilk olarak kullanılan ARFF Reader düğümü, ARFF dosyasına dönüştürülmüş istatistik bölüm derslerinin bulunduğu veri kümesini okumuştur ve Create Bit Vector düğümüne bağlanmıştır. Bu düğümde her bir üniversite bir bit vektör haline dönüştürülerek, oluşturulan bit vektörler bölüm dersleri veri kümesine sütun olarak eklenmiştir. Sonra Create Bit Vector düğümüne iki düğüm bağlanmıştır. Bu düğüme bağlanan ilk düğüm Association Rule Learner düğümü olmuştur. Association Rule Learner düğümü

Itemset Mining, Output ve Association Rules olmak üzere üç özellikten oluşmaktadır. Bu düğümün birinci özelliği olan Itemset Mining seçenekleri aşağıdaki şekilde açıklanmıştır.

Column Containing Transactions; sık bulunan öğe kümeleri veya ilişkilendirme kuralları için yapılacak işlemleri içeren sütunu belirlemek amacıyla kullanılmaktadır. Minimum Support; veri kümesindeki öğelerin birlikteliklerinin ne kadar sık olduğunu göstermektedir. Underlying Data Structure; veri kümesindeki satır sayısı öğe sayısından büyük olduğunda ARRAY seçeneğini, veri kümesindeki satır sayısı öğe sayısından küçük olduğunda ise TIDList seçeneğini kullanmaktadır. Bu özellik kullanılan seçeneğe göre veri yapısını oluşturmaktadır.

Association Rule Learner düğümünde ikinci özellik olan Output özelliği şu şekilde açıklanmıştır. Itemset type; öğe kümesinin türünü belirlemektedir. Öğe kümesinin türü öğe kümelerinin verdiği bilgilere göre; kümeler hakkında en fazla bilgi veren seçenek Closed, kümeler için istediğimiz bilgi dışında da bilgi veren seçenek Free ve kümelerde bilgiyi gizleyen seçenek ise Maximal olmak üzere 3 farklı şekilde belirlenmektedir. Maximal Itemset Length; sonuç tablosunda ortaya çıkacak maksimum öğe kümelerinin sayısını göstermektedir.

Association Rule Learner düğümü için üçüncü özellik olan Association Rules özelliği ise öğelerin birlikte bulunma olasılıklarını gösteren Minimum Confidence özelliğinden oluşmaktadır. Bu seçenek uygulamada tercih edilmemiştir.

Uygulamada birliktelik kuralları analizi için Association Rule Learner düğümünde yapılan işlemler şu şekilde açıklanmaktadır:

Association Rule Learner düğümünde ilk olarak işlemleri içeren bit vektör sütunu belirlenmiştir ve öğe birlikteliklerinin destek değerleri 0.1-1.0 aralığında seçilerek hesaplanmıştır. Sonra istatistik bölüm derslerinin bulunduğu veri kümesindeki satır sayısı öğe sayısından küçük olduğundan dolayı veri yapısı için TIDList özelliği tercih edilmiştir. Öğe kümesinden en fazla bilginin alınması için de Closed seçeneği kullanılmıştır. Daha sonra 0.1-1.0 aralığında seçilen destek

değerlerine göre oluşturulacak öğe kümelerindeki ders sayıları 1,2 ve 3 olarak belirlenmiştir. Bu değerler sonucunda 1'li, 2'li ve 3'lü ders birlikteliklerinin bulunduğu öğe kümeleri oluşturulmuştur. Derslerin destek değerleri ile birlikte sadece 1 dersin bulunduğu 1'li öğe kümesi Ek 3'te, 2 dersli birlikteliklerin bulunduğu 2'li öğe kümesi Ek 4'de ve 3 dersli birlikteliklerin bulunduğu 3'lü öğe kümesi Ek 5'te gösterilmiştir.

Create Bit Vector düğümüne ikinci olarak Association Rule Learner (Borgelt) düğümü bağlanmıştır. Bu düğüm Item column ve Itemset setting özelliklerinden oluşmaktadır. Item column özelliği işlem yapılacak sütunu belirlemektedir. Itemset setting özelliğini oluşturan seçenekler ise şu şekilde tanımlanmıştır. Minimum Set Size; minimum küme sayısını göstermektedir. Minimum Support; minimum destek değerini, Minimum Rule Confidence ise; minumum güven değerini yani oluşturulacak kuralların birlikte bulunma olasılığını ifade etmektedir.

Analiz için, Association Rule Learner (Borgelt) düğümünde Apriori algoritması kullanılmıştır. Algoritmadaki destek ve güven değerlerine 0.1-1.0 aralığında farklı değer verilerek uygulama yapılmıştır. Uygulama sonucunda birliktelik kuralları elde edilmiştir. Bu kuralların iyi olabilmesi için düşük destek, yüksek güven ve kaldıraç (lift) değerlerine sahip olması gerekmektedir. Çünkü destek değeri azaldıkça birliktelik kurallarının sayısı arttığı için analizde düşük destek değerleri seçilmiştir. Yüksek güven değerleri, birliktelik kurallarının güvenilirliğinin artması için kullanılmıştır. Yüksek kaldıraç değerleri ise birliktelik kurallarının ne derece ilginç olduğunu gösterdiği için analizde yer almıştır. Bu yüzden birliktelik kuralları analizi, düşük destek değerlerine sahip olan ders birlikteliklerinin güven değerleri %70 ve üzeri alınarak, yüksek kaldıraç değerlerinin kullanılması ile yapılmıştır ve oluşturulacak öğe kümelerindeki ders sayıları 1,2 ve 3 olarak belirlenmiştir. Bu kümelerdeki ders sayıları kullanılarak birliktelik kuralları ortaya çıkarılmıştır. Analiz sonucunda oluşan derslerin birliktelik kuralları tablosu Ek 6'da gösterilmiştir. Birliktelik kuralları analizi

sonucunda 21 tane kural oluşturulmuştur. Ders birlikteliklerinin ne kadar sık olduğu destek değerleri, derslerin birlikte bulunma olasılığı olan güven değerleri ve birliktelik kurallarının ne kadar ilginç olduğunu gösteren kaldıraç değerleri Ek 6'da tablo şeklinde ifade edilmiştir. Dersler için oluşturulan birliktelik kuralları ve yorumları aşağıda açıklanmıştır.

1.Kural: İstatistiksel Sonuç Çıkarım → Olasılığa Giriş

İstatistiksel Sonuç Çıkarım dersini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %100'dür.

2.Kural: Matematik 2 → Olasılığa Giriş

Matematik 2 dersini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %100'dür.

3.Kural: Optimizasyon Teknikleri I → İstatistik İçin Matematik I

Optimizasyon Teknikleri I dersini alan öğrencilerin İstatistik İçin Matematik I dersini de alma olasılığı %100'dür.

4.Kural: Kombinatorik → Matematik 1

Kombinatorik dersini alan öğrencilerin Matematik 1 dersini de alma olasılığı %100'dür.

5.Kural: Kombinatorik → Olasılığa Giriş

Kombinatorik dersini alan öğrencilerin Olasılığa Giriş dersini de alma olasılğı %100'dür.

6.Kural: Aktüerya Modelleri → Matematik 1

Aktüerya Modelleri dersini alan öğrencilerin Matematik 1 dersini de alma olasılığı %100'dür.

7.Kural: Aktüerya Modelleri → Olasılığa Giriş

Aktüerya Modelleri dersini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %100'dür.

8.Kural: Lineer Cebir 1 → Lineer Cebir 2

Lineer Cebir 1 dersini alan öğrencilerin Lineer Cebir 2 dersini de alma olasılığı %75'tir.

9.Kural: Lineer Cebir 1 → Olasılığa Giriş

Lineer Cebir 1 dersini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %100'dür.

10.Kural: Matematik 1 → Olasılığa Giriş

Matematik 1 dersini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %80'dir.

11.Kural: Çok Değişkenli İstatistiksel Analiz 2 → Olasılığa Giriş

Çok Değişkenli İstatistiksel Analiz 2 dersini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %100'dür.

12.Kural: Lineer Cebir 2 → Olasılığa Giriş

Lineer Cebir 2 dersini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %83,3'tür.

13.Kural: Kombinatorik, Matematik 1 → Olasılığa Giriş

Kombinatorik ve Matematik 1 derslerini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %100'dür.

14.Kural: Kombinatorik, Olasılığa Giriş → Matematik 1

Kombinatorik ve Olasılığa Giriş derslerini alan öğrencilerin Matematik 1 dersini de alma olasılığı %100'dür.

15.Kural: Matematik 1, Olasılığa Giriş → Kombinatorik

Matematik 1 ve Olasılığa Giriş derslerini alan öğrencilerin Kombinatorik dersini de alma olasılığı %75'tir.

16.Kural: Aktüerya Modelleri, Matematik 1→ Olasılığa Giriş

Aktüerya Modelleri ve Matematik 1 derslerini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %100'dür.

17.Kural: Aktüerya Modelleri, Olasılığa Giriş → Matematik 1
Aktüerya Modelleri ve Olasılığa Giriş derslerini alan öğrencilerin Matematik 1 dersini de alma olasılığı %100'dür.

18.Kural: Matematik 1, Olasılığa Giriş → Aktüerya Modelleri
Matematik 1 ve Olasılığa Giriş derslerini alan öğrencilerin Aktüerya Modelleri dersini de alma olasılığı %75'tir.

19.Kural: Lineer Cebir 1, Lineer Cebir 2 → Olasılığa Giriş
Lineer Cebir 1 ve Lineer Cebir 2 derslerini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %100'dür.

20.Kural: Lineer Cebir 1, Olasılığa Giriş → Lineer Cebir 2
Lineer Cebir 1 ve Olasılığa Giriş derslerini alan öğrencilerin Lineer Cebir 2 dersini de alma olasılığı %75'tir.

21.Kural: Çok Değişkenli İstatistiksel Analiz 2, Lineer Cebir 2 → Olasılığa Giriş
Çok Değişkenli İstatistiksel Analiz 2 ve Lineer Cebir 2 derslerini alan öğrencilerin Olasılığa Giriş dersini de alma olasılığı %100'dür.

21 birliktelik kuralının ilginçlik derecesi incelendiğinde, derslerin birbirleri ile olan etkileşimlerini gösteren kaldıraç (lift) değerlerinin 1'den büyük olduğu görülmüştür. Kaldıraç değerlerinin 1'den büyük olması bu birliktelik kurallarının ilginç olduğu anlamına gelmektedir. Bu kurallar arasında en büyük lift değerine sahip olan kural en ilginç kural olarak adlandırılmaktadır. 21 birliktelik kuralı arasında en yüksek lift değerine sahip olan kural 3. kuraldır. Aynı zamanda bu kural {Optimizasyon Teknikleri I, İstatistik İçin Matematik I} öğe kümesinde bulunan derslerin birbirlerine olan etkisinin yüksek derecede olduğunu gösteren en ilginç kuraldır.

Birliktelik kuralları analizinde oluşturulan kuralların güçlü olması için güven ve kaldıraç değerlerinin yüksek olması gerekmektedir. Bu yüzden güçlü birliktelik kuralları elde etmek amacıyla uygulamada derslerin güven değeri %70 ve üzeri, kaldıraç değeri ise 1'den büyük seçilmiştir. Seçilen bu değerlere göre,

analiz sonucunda oluşan düşük destek değerlerine sahip olan ders birlikteliklerinin, yüksek güven ve kaldıraç değerlerine de sahip olmasından dolayı Ek 6'da gösterilen 21 birliktelik kuralı, güçlü kurallar olarak adlandırılmıştır.

5.5.3. Birliktelik Kuralları Analizinin Sonuçları

Birliktelik kuralları analizi istatistik bölümü bulunan üniversitelerde hangi derslerin birlikte alınması gerektiğini ortaya çıkarmıştır. Uygulamada Apriori algoritması kullanılmıştır. Bu algoritma düşük destek değerlerine sahip olan ders birlikteliklerinin güven değerlerini %70 ve üzeri alarak yüksek kaldıraç (lift) değerleri ile öğe kümeleri için derslerin birliktelik kurallarını ortaya çıkarmıştır. Analiz sonucunda 21 tane birliktelik kuralı oluşmuştur. Bu kurallar incelendiğinde; istatistik bölümünün lisans programlarında bulunması gereken önemli derslerin Olasılığa Giriş, Matematik 1, Kombinatorik, Aktüerya Modelleri, Lineer Cebir 1 ve Lineer Cebir 2 olması gerektiği tespit edilmiştir. Ayrıca, 21 birliktelik kuralının 17 tanesinde Olasılığa Giriş dersi yer almaktadır. Bu durum istatistik bölümü olan üniversitelerin ders programlarında bulunması gereken en önemli dersin Olasılığa Giriş dersi olacağını ifade etmektedir.

5.6. İstatistik Bölümlerinin Başarılarına Etki Eden Derslerin Karar Ağacı Analizi ile Belirlenmesi

5.6.1. Veri Dönüşüm işlemleri

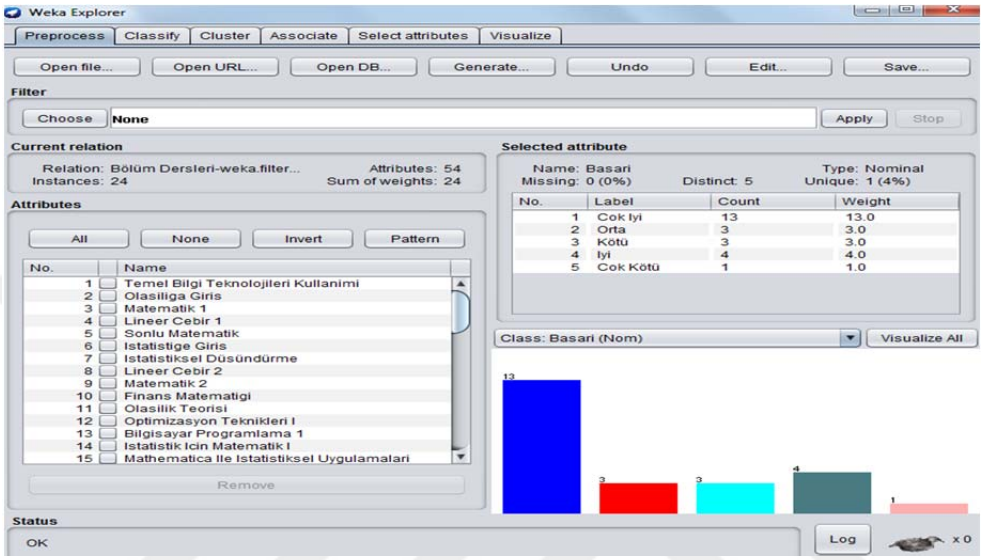
Çukurova Üniversitesi İstatistik Bölümü derslerinin kayıtlı olduğu Excel dosyasına, 2018 yılında URAP tarafından açıklanan Türkiye'deki üniversitelerin başarı sıralamaları yeni bir sütun olarak eklenmiştir. Üniversitelerin başarı sıralamaları kategorik değişken olarak Excel'e girilmiştir. Bu kategorik değişkenler "Çok İyi", "Orta", "Kötü", "İyi" ve "Çok Kötü" şeklinde 5 sınıfa ayrılarak Excel'e kaydedilmiştir. Bölüm dersleri veri kümesine üniversitelerin başarı durumunu gösteren sütunun eklenmesi ile istatistik bölümlerinin başarılarına etki eden derslerin hangilerinin olduğunu belirlemek için sınıflandırma yöntemlerinden biri

olan karar ağacı analizi kullanılmıştır. Bu analiz WEKA programında bulunan J48 (C4.5) karar ağacı algoritması ile gerçekleştirilmiştir. WEKA programı ARFF ya da CSV dosya türlerini kabul etmektedir. Bu yüzden öncelikle bölüm derslerinin kayıtlı olduğu Excel dosyası CSV türünden dosya haline dönüştürülmüştür. Sonra dönüştürülen bu dosya WEKA programına yüklenerek uygulamaya başlanmıştır. Bölüm dersleri CSV uzantılı dosya WEKA programına yüklendikten sonra WEKA panelinden seçilen J48 algoritması ve Cross Validation (Çapraz Geçerlilik) seçeneği ile karar ağacı analizi gerçekleştirilmiştir. Analiz sonucunda, sınıflandırmanın performansına ait ölçüler, bu ölçüler ile oluşturulan grafikler, sınıf tahminlerinin doğruluğu hakkında bilgi veren karışıklık matrisi (Confusion Matrix) ve sınıfların bulunduğu karar ağacı oluşturulmuştur. Analiz sonuçlarının değerlendirilmesi için sınıflandırma modelinin doğruluğu belirlenmiştir. Sınıflandırma modelinin doğruluğu belirlenirken sınıflandırmaya ait performans ölçülerinden yararlanılmıştır. Bu ölçüler incelendiğinde, karar ağacı analizi sonucunda iyi bir sınıflandırma yapılmadığı görülmüştür. Sınıf kategorilerinin dengesiz dağılımını ortadan kaldırmak amacıyla WEKA programında bulunan Resample (Yeniden Örneklendirme) özelliği kullanılmıştır. Resample (Yeniden Örneklendirme) özelliği veri kümesini yeniden örneklendirerek sınıf kategorilerini dengeli bir dağılım haline getirmiştir. Dengeli veri kümesi haline gelen bölüm dersleri için WEKA programında tekrardan karar ağacı analizi yapılmıştır. Yapılan analizin sonuçlarına göre; karar ağacı analizi istatistik bölümlerinin başarılarına etki eden dersleri iyi bir şekilde sınıflandırmıştır.

5.6.2. Karar Ağacı Analizinin Uygulanması

İstatistik bölümleri bulunan üniversitelerin 2018 yılında URAP indeksinde yayınlanan başarı sıralamalarına göre bölümlerin başarılarına etki eden derslerin hangileri olduğunu belirlemek için bölüm dersleri veri kümesine sınıflandırma yöntemlerinden biri olan karar ağacı analizi yöntemi uygulanmıştır.

Veri kümesindeki derslerin WEKA programındaki görünümü Şekil 5.18'de gösterilmiştir.



Şekil 5. 18. WEKA Kullanıcı Ara Yüzü Görüntüsü (Orijinal Veri Kümesi)

Uygulamada CSV dosyasına dönüştürülen bölüm dersleri dosyası WEKA programına yüklendikten sonra WEKA'nın Classify panelinde bulunan J48 karar ağacı algoritması ile model geçerliliğinin test edilmesini sağlayan Cross Validation (Çapraz Geçerlilik) seçeneği kullanılarak analize başlanmıştır. Seçilen bu algoritma ile karar ağacı analizi sonucunda; üniversitelerin başarılarına etki eden derslerin sınıflandırma performansını gösteren ölçüleri ve grafikleri, üniversitelerin başarılarının gösterildiği sınıf tahmin sayılarının doğruluğunu ifade eden karışıklık matrisi (Confusion Matrix) ve üniversitelerin başarılarına etki eden derslerin yer aldığı karar ağacı oluşturulmuştur. Analiz sonuçları sınıflandırma modelinin doğruluğu ile değerlendirilmiştir. Bu durumda sınıflandırma modelinin doğruluğu hakkında yorum yapabilmek için Çizelge 5.3'te gösterilen J48 algoritması sonuçları tablosundaki performans ölçüleri incelenmiştir.

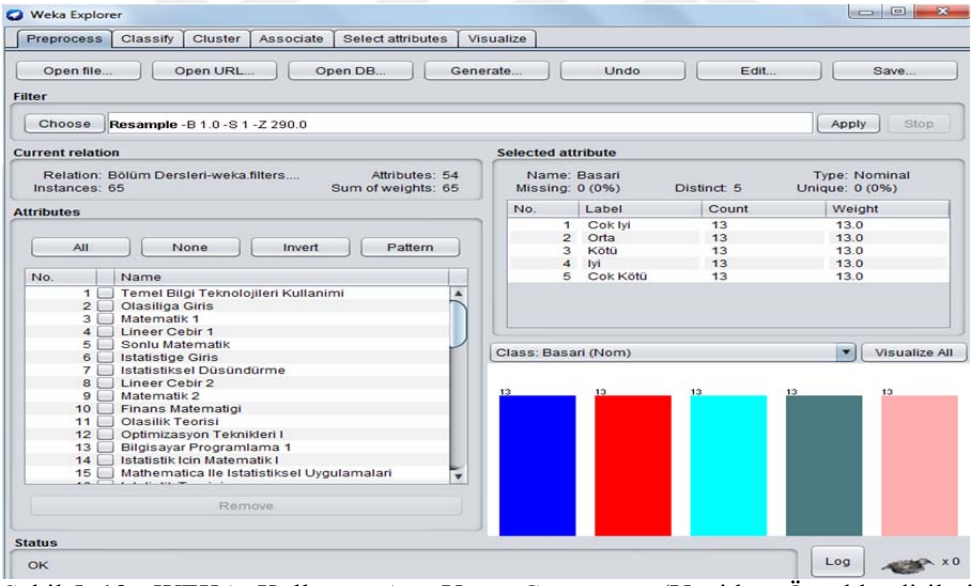
Çizelge 5. 3. J48 Algoritması Sonuçları (Orijinal Veri Kümesi)

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          12               50      %
Kappa statistic                         -0.0549
Mean absolute error                      0.2667
Root mean squared error                  0.3901
Relative absolute error                 96.6508 %
Root relative squared error            105.3547 %
Total Number of Instances               24

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,923    1,000    0,522      0,923    0,667      -0,192   0,325     0,480     Cok Iyi
                 0,000    0,000    ?          0,000    ?          ?        0,111     0,105     Orta
                 0,000    0,000    ?          0,000    ?          ?        0,286     0,143     Kötü
                 0,000    0,050    0,000      0,000    0,000      -0,093   0,300     0,182     Iyi
                 0,000    0,000    ?          0,000    ?          ?        0,109     0,042     Cok Kötü
Weighted Avg.    0,500    0,550    ?          0,500    ?          ?        0,280     0,323

=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
 12  0  0  1  0 |  a = Cok Iyi
  3  0  0  0  0 |  b = Orta
  3  0  0  0  0 |  c = Kötü
  4  0  0  0  0 |  d = Iyi
  1  0  0  0  0 |  e = Cok Kötü
```

Çizelge 5.3'teki tablo incelendiğinde; doğru sınıflandırılmış verileri ve doğru sınıflandırılan verilerin oranını (Correctly Classified Instances) ifade eden değerin %50, veriler arasındaki uyumun güvenilirliğini ölçen değerin (Kappa Statistic) ise -0.0549 olduğu görülmüştür. Bu iki ölçünün aldığı değerlerin analiz için düşük ve yetersiz olduğu tespit edilmiştir. Aynı zamanda bu durum, üniversitelerin başarı sıralamalarına göre istatistik bölümlerinin başarılarına etki eden derslerin kötü sınıflandırıldığını açıklamaktadır. Sınıflandırmanın değerlendirilmesinde; Correctly Classified Instances ve Kappa Statistic ölçü değerlerinin sınıf tahminlerini iyi açıklayamamasından dolayı diğer performans ölçüleri incelenmemiştir.

Analiz sonuçlarına bakıldığında; bölüm dersleri veri kümesinde bulunan "Çok İyi", "Orta", "Kötü", "İyi" ve "Çok Kötü" şeklindeki üniversitelerin başarılarını gösteren sınıf kategorilerinin sırasıyla aldığı 13, 3, 3, 4 ve 1

değerlerinin dengesiz dağılım göstermesinden dolayı sınıflandırma sonuçları iyi çıkmamıştır. Böyle dengesiz bir dağılım gösteren veri kümesi için WEKA programının giriş panelindeki Filter özelliğinden Resample (Yeniden Örneklendirme) seçeneği kullanılarak veri kümesinin dengeli bir dağılım haline gelmesi sağlanmıştır. Yeniden Örneklendirme özelliği veri kümesindeki sınıf kategorilerinin sırasıyla aldığı 13, 3, 3, 4 ve 1 olan değerlerini yeniden örneklendirme yöntemi ile sınıf kategorilerinin aldığı değerleri 13, 13, 13, 13 ve 13 şekline dönüştürerek veri kümesini dengelemiştir. Dengeli hale gelen bölüm dersleri dosyasına istatistik bölümlerinin başarılarına etki eden derslerin belirlemesi için WEKA programında tekrardan karar ağacı analizi uygulanmıştır. Dengeli duruma gelen veri kümesindeki derslerin görünümü Şekil 5.19'da gösterilmiştir.



Şekil 5. 19. WEKA Kullanıcı Ara Yüzü Görüntüsü (Yeniden Örneklendirilmiş Veri Kümesi)

Analiz sonuçlarında, üniversitelerin başarılarına etki eden derslerin sınıflandırma performansına ait ölçüleri ve grafikleri, üniversitelerin başarılarının gösterildiği sınıf kategorilerinin tahmin sayılarının doğruluğunu ifade eden

karışıklık matrisi (Confusion Matrix) ve üniversitelerin başarılarına etki eden derslerin bulunduğu karar ağacı oluşturulmuştur. Bu analizin değerlendirilmesi sınıflandırmaya ait performans ölçülerini gösteren Çizelge 5.4'te yapılmıştır.

Çizelge 5. 4. J48 Algoritması Sonuçları (Yeniden Örneklendirilmiş Veri Kümesi)

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          48               73.8462 %
Kappa statistic                          0.6731
Mean absolute error                      0.1325
Root mean squared error                  0.284
Relative absolute error                 41.2924 %
Root relative squared error             70.7779 %
Total Number of Instances               65

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,385    0,000    1,000      0,385    0,556      0,577    0,766     0,579     Cok Iyi
               0,846    0,038    0,846      0,846    0,846      0,808    0,976     0,849     Orta
               0,846    0,077    0,733      0,846    0,786      0,730    0,972     0,900     Kötü
               0,615    0,096    0,615      0,615    0,615      0,519    0,885     0,674     Iyi
               1,000    0,115    0,684      1,000    0,813      0,778    0,901     0,553     Cok Kötü
Weighted Avg.  0,738    0,065    0,776      0,738    0,723      0,683    0,900     0,711

=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
  5  2  0  2  4 |  a = Cok Iyi
  0 11  1  1  0 |  b = Orta
  0  0 11  2  0 |  c = Kötü
  0  0  3  8  2 |  d = Iyi
  0  0  0  0 13 |  e = Cok Kötü
```

Çizelge 5.4 tablosuna bakıldığında; doğru sınıflandırılmış verileri ve doğru sınıflandırılan verilerin oranını gösteren değerin %73.8462, veriler arasındaki uyumun güvenilirliğini ölçen değerin ise 0.6731 olduğu görülmüştür. Doğru sınıflandırma oranının yüksek, veriler arasındaki uyumun güvenilirliğini ifade eden değerin ise 1'e yakın olması analiz için büyük önem taşımaktadır. Çünkü başarıyı etkileyen derslerin doğru sınıflandırılması ve güvenilirliğinin yüksek olması karar ağacı analizini güçlendirmektedir.

Correctly Classified Instances ve Kappa Statistic ölçülerinin yanı sıra analiz sonucunda oluşan tabloda sınıflandırmayı etkileyen başka ölçülerde

bulunmaktadır. Tablodaki bu ölçüler aşağıdaki şekilde ifade edilmiştir.

MAE (Ortalama Mutlak Hata) ölçüsü, veri kümesindeki değerler ile sınıflandırma için tahmin edilen değerler arasındaki mutlak hatanın 0.1325 olduğunu göstermektedir. RMSE (Kök Hata Kareler Ortalaması) ölçüsü, veri kümesinde bulunan değerler ile sınıflandırmada tahmin edilen değerler arasındaki hata oranının 0.284 olduğunu ifade etmektedir. RAE (Bağıl Mutlak Hata) ölçüsü ise sınıflandırma tahminin sahip olduğu hata oranının %41.2924 olduğunu açıklamaktadır. İncelenen MAE, RMSE ve RAE ölçülerinin aldıkları değerlerin yapılan analize uygun olduğu belirlenmiştir. Ayrıca, bu ölçülerin sıfıra yakın değerler alması sınıflandırma modelinin tahmin yeteneğini arttırmıştır. Tabloda bulunan diğer bir ölçü ise RRSE (Bağıl Karesel Hatanın Karekökü) ölçüsüdür. RRSE, Correctly Classified Instances ölçüsü ile birbirine zıt olan bir ölçüdür. Analiz sonucunda oluşan derslerin doğru sınıflandırma oranı %73.8462'dir. RRSE ölçüsünün oranı ise %70.7779'dur. Bu durum doğru sınıflandırma oranının yüksek, RRSE ölçüsünün aldığı hata değerinin ise düşük olmasından dolayı sınıflandırmanın iyi bir şekilde yapıldığını ifade etmektedir. MAE, RMSE, RAE ve RRSE ölçüleri yorumlandıktan sonra tabloda bulunan sınıf kategorilerine ait performans ölçülerinin ve karışıklık matrisinin (Confusion Matrix) karar ağacı analizine olan etkisi incelenmiştir. İlk olarak tablodaki sınıf kategorilerine ait ölçüler sınıflandırma için aşağıdaki şekilde değerlendirilmiştir.

True Positive Rate (Gerçek Pozitif Oranı) ölçüsünün sınıf tahminleri için aldığı değerler analizde yapılan sınıflandırmanın iyi olduğunu açıklamıştır. J48 algoritması ile yapılan bu sınıflandırmada özellikle "Orta", "Kötü", "İyi" ve "Çok Kötü" sınıf kategorilerinin gerçek pozitif değerlerinin yüksek olduğu görülmüştür. Bu değerin yüksek olması üniversitelerin başarılarına etki eden derslerin tahminlerinin doğru yapıldığını göstermektedir.

False Positive Rate (Yanlış Pozitif Oranı) ölçüsü incelendiğinde; J48 algoritması ile yapılan sınıflandırmada sınıf kategorilerinin yanlış pozitif oranlarının düşük olduğu belirlenmiştir. Bu durum üniversitelerin başarılarına etki

eden derslerin tahmininde yapılan hataların az olduğunu ifade etmektedir.

Precision (Kesinlik) ölçüsü, sınıflandırmada pozitif olarak tahmin edilen durumların başarısını göstermektedir. Bu ölçü değerlerinin 1'e yaklaşması ile başarılı bir tahmin yapılmıştır.

Recall (Anma) ölçüsü sınıflandırmada pozitif durumların ne kadar başarılı tahmin edildiği durumu ifade etmektedir. Bu ölçünün değerlerinin 1'e yakın olması ile sınıf kategorileri başarılı bir şekilde tahmin edilmiştir.

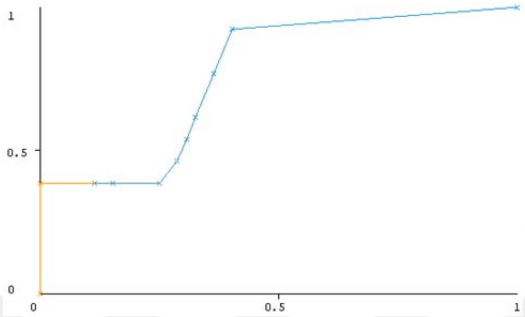
Precision (Kesinlik) ve Recall (Anma) ölçülerini ayrı ayrı incelemek analizin değerlendirilmesi için yeterli olmayabilir. Her iki ölçüyü birlikte değerlendirmek ve daha iyi sonuçlar almak için F-Measure (F ölçüsü) değeri kullanılmıştır. F-Measure (F ölçüsü) değeri Precision ve Recall ölçülerinin harmonik ortalamasının alınması ile bulunmaktadır. Sınıf kategorilerinin F değerlerinin yüksek ve 1'e yakın olması analizin iyi olduğunu göstermiştir.

MCC (Mathews Korelasyon Katsayısı) ölçüsü karışıklık matrisinden (Confusion Matrix) yararlanılarak elde edilmektedir. Bu ölçünün aldığı değerlerin 1'e yakın olması sınıflandırma başarısının doğru yapıldığını göstermektedir.

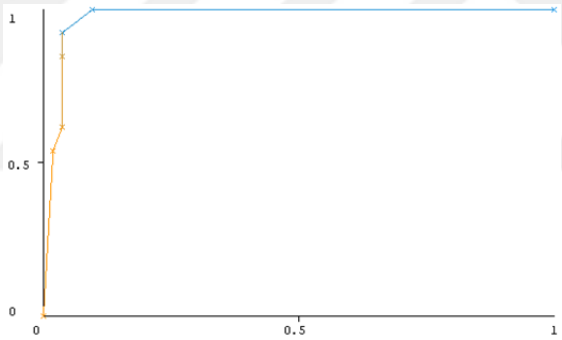
Sınıf kategorilerine ait ölçüler sınıflandırma için değerlendirildikten sonra Confusion Matrix (Karışıklık Matrisi) analiz için incelenmiştir. Karışıklık matrisi, sınıf tahminlerinin doğruluğu hakkında bilgi veren bir matristir. Analiz sonucunda oluşan karışıklık matrisinin köşegen elemanları doğru sınıflandırılan sınıf kategorilerinin sayılarını Çok İyi=5, Orta=11, Kötü=11, İyi=8 ve Çok Kötü=13 şeklinde göstermiştir. Bu matrisin köşegen elemanlarının toplamı doğru sınıflandırılan verilerin sayısını, köşegen dışındaki elemanların toplamı ise yanlış sınıflandırılan verilerin sayısını ifade etmektedir. Karar ağacı analizi sonucunda doğru sınıflandırılan verilerin sayısının 48, yanlış sınıflandırılan verilerin sayısının ise 17 olduğu görülmüştür.

Karışıklık matrisi ile tahminlerin doğruluğu incelendikten sonra ROC ve PRC grafikleri sınıflandırma için değerlendirilmiştir.

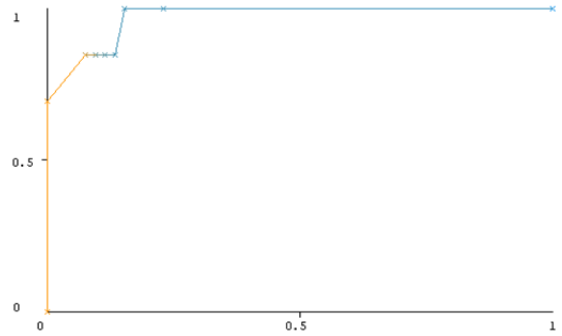
ROC (Alıcı İşletim Karakteristiği) grafiği, gerçek pozitif oranının yanlış pozitif oranına karşı oluşturulan bir grafiktir. Analiz sonucunda oluşan ROC grafikleri Şekil 5.20, Şekil 5.21, Şekil 5.22, Şekil 5.23 ve Şekil 5.24'te gösterilmiştir.



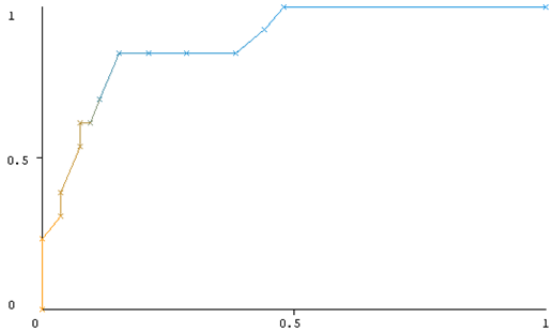
Şekil 5. 20. Çok İyi Sınıf Etiketine Ait ROC Grafiği



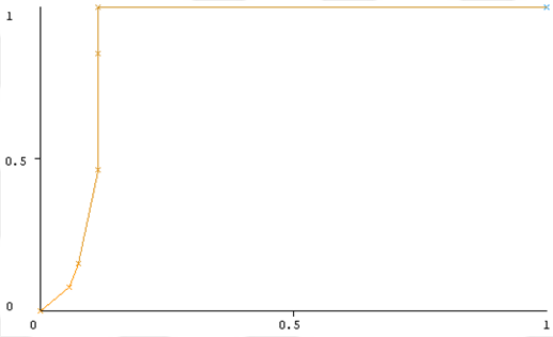
Şekil 5. 21. Orta Sınıf Etiketine Ait ROC Grafiği



Şekil 5. 22. Kötü Sınıf Etiketine Ait ROC Grafiği



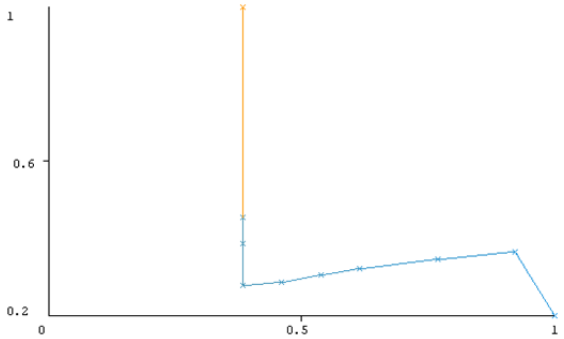
Şekil 5. 23. İyi Sınıf Etiketine Ait ROC Grafiği



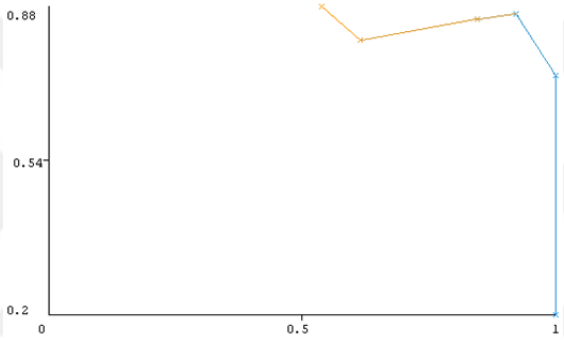
Şekil 5. 24. Çok Kötü Sınıf Etiketine Ait ROC Grafiği

ROC grafiklerindeki sınıf kategorilerinin aldığı değerlerin sol üst köşeye yakın olması yanlış sınıflandırılan pozitif değerlerin düşük oranlara sahip olduğunu ifade etmektedir. Bu durumda ROC ölçü değerleri analizdeki sınıflandırma başarısının iyi bir şekilde yapıldığını göstermektedir.

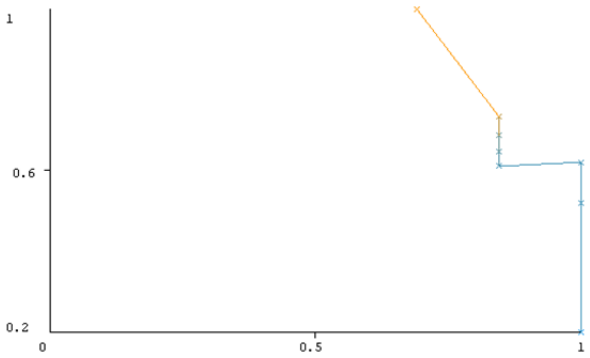
PRC (Kesinlik Anma Eğrisi) grafiği ise kesinlik ölçüsünün anma ölçüsüne karşı oluşturulan bir grafiktir. PRC grafiklerinin aldığı değerler y eksenine doğru bir yönelme gösteriyor ise gerçek pozitif durumlar, x eksenine doğru bir yönelme gösteriyor ise yanlış negatif durumlar tespit edilmektedir. Analiz sonucunda oluşan PRC grafikleri Şekil 5.25, Şekil 5.26, Şekil 5.27, Şekil 5.28 ve Şekil 5.29'da gösterilmiştir.



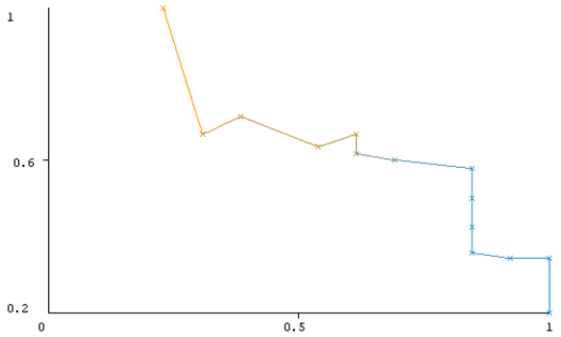
Şekil 5. 25. Çok İyi Sınıf Etiketine Ait PRC Grafiği



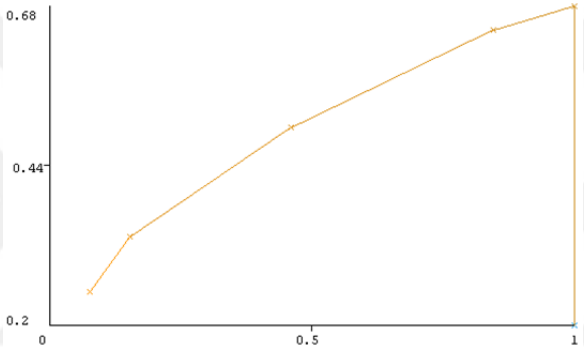
Şekil 5. 26. Orta Sınıf Etiketine Ait PRC Grafiği



Şekil 5. 27. Kötü Sınıf Etiketine Ait PRC Grafiği



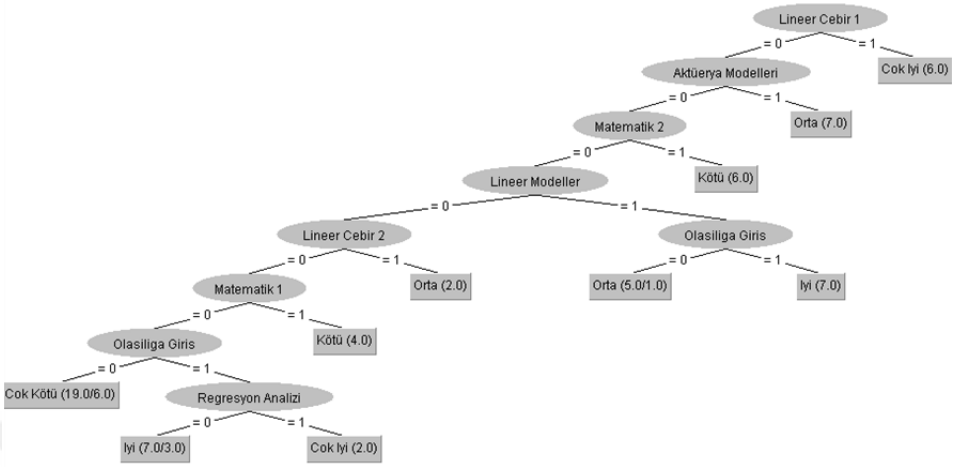
Şekil 5. 28. İyi Sınıf Etiketine Ait PRC Grafiği



Şekil 5. 29. Çok Kötü Sınıf Etiketine Ait PRC Grafiği

PRC grafikleri incelediğinde; "Orta", "Kötü" ve "İyi" sınıf kategorilerinin tahminlerinin gerçek pozitif değerleri, "Çok İyi" ve "Çok Kötü" sınıf kategorilerinin tahminlerinin ise yanlış negatif değerleri belirlenmiştir.

Çizelge 5.4'teki sınıflandırma performans ölçü değerlerinin analiz için uygun oldukları belirlendikten sonra üniversitelerin başarılarını etkileyen dersleri gösteren karar kurallarının oluşturulması için Şekil 5.30'daki karar ağacı tablosu incelenmiştir.



Şekil 5. 30. J48 Algoritması Sonucunda Oluşan Karar Ağacı

Karar ağacında sınıflandırma kök düğüm olan Lineer Cebir 1 dersi ile başlamıştır. Derslerin aldığı 0 ve 1 değerlerine göre düğümler dallara ayrılmıştır. Bu değerlere göre yukarıdan aşağıya doğru elde edilen dallar ile bölüm derslerini ifade eden ara düğümler ve üniversitelerin başarı durumunu gösteren yapraklar (sınıflar) oluşturulmuştur. Karar ağacı oluşturulduktan sonra kök düğümden yaprağa doğru aşağı inilerek karar kuralları yazılmıştır. Üniversitelerin başarılarını etkileyen dersler için oluşturulan karar ağacı kuralları ve yorumları aşağıda açıklanmıştır.

Kural 1: If Lineer Cebir 1 = 1 then SINIF = Çok İyi

Eğer Lineer Cebir 1 dersi bulunuyor ise üniversite "Çok İyi" sınıf kategorisinde yer almaktadır.

Kural 2: If Lineer Cebir 1 = 0 and Aktüerya Modelleri = 1 then SINIF = Orta

Eğer Lineer Cebir 1 dersi bulunmuyor, Aktüerya Modelleri dersi bulunuyor ise üniversite "Orta" sınıf kategorisinde yer almaktadır.

Kural 3: If Lineer Cebir 1 = 0, Aktüerya Modelleri = 0 and Matematik 2 = 1 then SINIF = Kötü

Eğer Lineer Cebir 1 ve Aktüerya Modelleri dersleri bulunmuyor, Matematik 2 dersi

bulunuyor ise üniversite "Kötü" sınıf kategorisinde yer almaktadır.

Kural 4: If Lineer Cebir 1 = 0, Aktüerya Modelleri = 0, Matematik 2 = 0, Lineer Modeller = 1 and Olasılığa Giriş = 1 then SINIF = İyi

Eğer Lineer Cebir 1, Aktüerya Modelleri ve Matematik 2 dersleri bulunmuyor, Lineer Modeller ve Olasılığa Giriş dersleri buluyor ise üniversite "İyi" sınıf kategorisinde yer almaktadır.

Kural 5: If Lineer Cebir 1 = 0, Aktüerya Modelleri = 0, Matematik 2 = 0, Lineer Modeller = 1 and Olasılığa Giriş = 0 then SINIF = Orta

Eğer Lineer Cebir 1, Aktüerya Modelleri, Matematik 2 ve Olasılığa Giriş dersleri bulunmuyor, Lineer Modeller dersi bulunuyor ise üniversite "Orta" sınıf kategorisinde yer almaktadır.

Kural 6: If Lineer Cebir 1 = 0, Aktüerya Modelleri = 0, Matematik 2 = 0, Lineer Modeller = 0 and Lineer Cebir 2 = 1 then SINIF = Orta

Eğer Lineer Cebir 1, Aktüerya Modelleri, Matematik 2 ve Lineer Modeller dersleri bulunmuyorsa, Lineer Cebir 2 dersi bulunuyor ise üniversite "Orta" sınıf kategorisinde yer almaktadır.

Kural 7: If Lineer Cebir 1 = 0, Aktüerya Modelleri = 0, Matematik 2 = 0, Lineer Modeller = 0, Lineer Cebir 2 = 0 and Matematik 1 = 1 then SINIF = Kötü

Eğer Lineer Cebir 1, Aktüerya Modelleri, Matematik 2, Lineer Modeller ve Lineer Cebir 2 dersleri bulunmuyorsa, Matematik 1 dersi bulunuyor ise üniversite "Kötü" sınıf kategorisinde yer almaktadır.

Kural 8: If Lineer Cebir 1 = 0, Aktüerya Modelleri = 0, Matematik 2 = 0, Lineer Modeller = 0, Lineer Cebir 2 = 0, Matematik 1 = 0 and Olasılığa Giriş = 0 then SINIF = Çok Kötü

Eğer Lineer Cebir 1, Aktüerya Modelleri, Matematik 2, Lineer Modeller, Lineer Cebir 2, Matematik 1 ve Olasılığa Giriş dersleri bulunmuyor ise üniversite "Çok Kötü" sınıf kategorisinde yer almaktadır.

Kural 9: If Lineer Cebir 1 = 0, Aktüerya Modelleri = 0, Matematik 2 = 0, Lineer Modeller = 0, Lineer Cebir 2 = 0, Matematik 1 = 0, Olasılığa Giriş = 1 and

Regresyon Analizi = 1 then SINIF = Çok İyi

Eğer Lineer Cebir 1, Aktüerya Modelleri, Matematik 2, Lineer Modeller, Lineer Cebir 2 ve Matematik 1 dersleri bulunmuyor, Olasılığa Giriş ve Regresyon Analizi dersleri bulunuyor ise üniversite "Çok İyi" sınıf kategorisinde yer almaktadır.

Kural 10: If Lineer Cebir 1 = 0, Aktüerya Modelleri = 0, Matematik 2 = 0, Lineer Modeller = 0, Lineer Cebir 2 = 0, Matematik 1 = 0, Olasılığa Giriş = 1 and Regresyon Analizi = 0 then SINIF = İyi

Eğer Lineer Cebir 1, Aktüerya Modelleri, Matematik 2, Lineer Modeller, Lineer Cebir 2, Matematik 1 ve Regresyon Analizi dersleri bulunmuyor, Olasılığa Giriş dersi bulunuyor ise üniversite "İyi" sınıf kategorisinde yer almaktadır.

Karar ağacı analizi sonucunda 10 tane karar kuralı oluşmuştur. Bu kurallar, istatistik bölümleri bulunan üniversitelerin başarılarına etki eden derslerin Lineer Cebir 1, Aktüerya Modelleri, Lineer Modeller, Olasılığa Giriş, Lineer Cebir 2 ve Regresyon Analizi olduğunu ifade etmektedir.

5.6.3. Karar Ağacı Analizinin Sonuçları

Karar ağacı analizi istatistik bölümleri bulunan üniversitelerin başarılarına etki eden derslerin belirlenmesini sağlamıştır. Uygulamada J48 (C4.5) algoritması kullanılmıştır. Bu algoritma sonucunda oluşan sınıflandırmaya ait performans ölçülerinden Correctly Classified Instances %50, Kappa Statistic ise -0.0549 değerini almıştır. İki ölçünün analiz için düşük ve yetersiz değerlere sahip olduğu tespit edilmiştir. Analiz sonuçlarının kötü olmasının nedeni üniversitelerin başarılarını gösteren sınıf kategorilerinin dengesiz dağılmasıdır. Veri kümesindeki bu dengesiz dağılım için WEKA programında bulunan Resample (Yeniden Örneklendirme) seçeneği kullanılarak veri kümesi dengeli bir dağılım haline getirilmiştir. Dengeli duruma gelen veri kümesi için WEKA'da tekrardan karar ağacı analizi yapılmıştır.

Karar ağacı analizi sonucunda, üniversitelerin başarılarına etki eden derslerin sınıflandırma performansını gösteren ölçüler incelenmiştir. Correctly

Classified Instances ölçüsü %73.8462, Kappa Statistic ölçüsü ise 0.6731 değerini almıştır. Bu iki ölçü yapılan ilk analiz ile karşılaştırıldığında; doğru sınıflandırılmış verileri ve doğru sınıflandırılan verilerin oranını gösteren değerin %50 'den %73.8462'ye, veriler arasındaki uyumun güvenilirliğini ölçen değerin ise -0.0549'dan 0.6731'e yükseldiği tespit edilmiştir. Yüksek değere sahip olan Correctly Classified Instances ve Kappa Statistic ölçü değerleri sınıflandırma modelinin tahmininin iyi yapıldığını ifade etmektedir. Sınıflandırma modelinin doğruluğu belirlendikten sonra Çizelge 5.4'teki sınıf kategorilerine ait performans ölçüleri incelenmiştir. Bu ölçüler üniversitelerin başarılarına etki eden derslerin tahmin doğruluğunun yüksek, sınıflandırmada yapılan hata oranlarının ise düşük olduğunu göstermektedir. Tabloda bulunan performans ölçüleri incelendikten sonra analiz sonucunda oluşan PRC ve ROC grafiklerinin aldıkları değerlerin sınıflandırmayı doğru bir şekilde ifade ettiği görülmüştür.

Sınıf kategorilerini gösteren ölçüler değerlendirildikten sonra Confusion Matrix (Karışıklık Matrisi) sınıflandırma tahmini için değerlendirilmiştir. Bu matrisin köşegen elemanları doğru sınıflandırılan sınıf kategorilerinin sayısının 48, köşegen dışındaki elemanları ise yanlış sınıflandırılan sınıf kategorilerinin sayısının 17 olduğunu göstermiştir. Sınıflandırma performansına ait ölçü değerlerinin analiz için uygun oldukları belirlendikten sonra karar ağacı analizinde üniversitelerin başarılarını etkileyen dersleri gösteren 10 karar kuralı oluşturulmuştur. Bu kurallara göre karar ağacında başarıya etki eden en önemli dersin Lineer Cebir 1 dersi olduğu görülmüştür. Karar ağacında başarıyı etkileyen diğer önemli derslerin ise Olasılığa Giriş ve Regresyon Analizi dersleri olduğu belirlenmiştir. İstatistik bölümlerine sahip üniversitelerin ders programlarında Lineer Cebir 1 dersi ya da Olasılığa Giriş ve Regresyon Analizi dersleri birlikte bulunuyor ise bu üniversitelerin "Çok İyi" üniversite kategorisinde yer aldığı görülmektedir.

Özet olarak; karar ağacı analizinde kullanılan J48 (C4.5) algoritması ile elde edilen performans ölçüleri ve grafikler incelenerek analizin başarılı bir şekilde yapıldığı görülmüştür. Böylece üniversitelerin başarılarına etki eden derslerin sınıflandırma tahminlerinin doğruluğunu iyi açıkladığı belirlenmiştir.

6. SONUÇLAR VE ÖNERİLER

Günümüzdeki teknolojik gelişmeler veri miktarlarında yoğun ve hızlı bir artışın olduğunu göstermektedir. Bu veri artışlarından dolayı büyük ve karmaşık veri kümeleri oluşmaktadır. Bu veri kümeleri için zor olan analizler veri madenciliği yöntemleri ile kolay bir şekilde yapılmaktadır. Veri madenciliği analizleri sonucunda büyük ve karmaşık verilerden anlamlı bilgiler ortaya çıkarılmaktadır. Ayrıca, verilerin birbirleri ile olan ilişkileri, verilerden oluşan modeller ve örüntüler veri madenciliği analizi ile belirlenmektedir. Veri madenciliği analizlerde sağladığı kolaylıklar nedeniyle geniş bir çalışma alanına sahiptir. Bu çalışma alanlarından biri olan eğitim alanında, öğrencilerin başarılarının belirlenmesi, ders içeriklerinin incelenmesi ile ders programlarının düzenlenmesi gibi eğitim kalitesinin arttırılmasını sağlayacak konular veri madenciliği ile analiz edilmektedir.

Bu tez çalışmasında; ülkemizde bulunan istatistik bölümüne sahip 24 üniversite incelenmiştir. İstatistik bölümlerinin lisans programlarında yer alan derslerin ders içerikleri kullanılarak farklı üniversitelerin istatistik bölümlerindeki ders içeriklerinin birbirleriyle olan benzerlikleri, bu benzerlikler ile üniversitelerin birbirleriyle ne kadar ilişkili olduğu, istatistik bölümlerinin ders programlarında bulunan derslerden hangi derslerin birlikte alınması gerektiği ve üniversitelerin başarılarına etki eden derslerin belirlenmesi için yapılan analizler veri madenciliği yöntemleri ile gerçekleştirilmiştir. Veri madenciliği analizlerinde kullanım kolaylığı sağlaması ve analiz sonuçlarının anlaşılabilir olması nedeniyle tez çalışmasında KNIME ve WEKA programları tercih edilmiştir. KNIME programından kümeleme ve birliktelik kuralları analizleri için seçilen düğümler çalıştırılarak uygulamalar yapılmıştır. WEKA programında ise J48 (C4.5) algoritması kullanılarak karar ağacı analizi gerçekleştirilmiştir.

Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümüne sahip diğer üniversitelerin lisans programlarında bulunan bölüm dersleri incelenmiştir. Bu

bölüm derslerinin ders içerikleri karşılaştırılarak veri madenciliği için 3 farklı analiz yöntemi kullanılmıştır.

Birinci analizde, istatistik bölümü bulunan üniversitelerin birbirleriyle olan benzerliklerinin belirlenmesi ve üniversitelerin kümelenmesi hiyerarşik kümeleme analizi ile gerçekleştirilmiştir. Kümeleme analizi sonucunda; dendogram grafiği ve kümelenen üniversitelerin hangi kümede yer aldığını gösteren tablo oluşturulmuştur. Dendogram grafiği, kümelenen üniversitelerin aralarındaki uzaklıkları inceleyerek birbirine benzerlik gösteren üniversitelerin belirlenmesini sağlamıştır. Küme sayısının 5 seçilmesi ile kümelenen üniversitelerin tablosu incelendiğinde; istatistik bölüm derslerine göre birbirine benzemeyen üniversite sayısının toplamda 4 olduğu ifade edilmiştir. Birbirine benzerlik gösteren üniversite sayısının ise 20 olduğu görülmüştür. Bu durum Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümüne sahip diğer 19 üniversitenin birbirlerine benzerlik gösterdiğini açıklamaktadır.

İkinci analizde, istatistik bölümlerinin eğitim programlarında yer alan derslerden hangi derslerin birlikte alınması gerektiğini belirlemek için birliktelik kuralları analizi algoritmalarından biri olan Apriori algoritması kullanılmıştır. Bu algoritma düşük destek değerine sahip ders birlikteliklerinin güven değerini %70 ve üzeri alarak analiz sonucunda 21 birliktelik kuralı oluşturmuştur. Bu kurallar incelenerek; istatistik bölümünün lisans programlarında Olasılığa Giriş, Matematik 1, Kombinatorik, Aktüerya Modelleri, Lineer Cebir 1 ve Lineer Cebir 2 derslerinin bulunması gerektiği tespit edilmiştir. Ayrıca, 21 birliktelik kuralının 17 tanesi; istatistik bölümü olan üniversitelerin ders programlarında yer alması gereken en önemli dersin Olasılığa Giriş dersi olduğunu ifade etmiştir.

Üçüncü analizde ise karar ağacı analizi algoritmalarından biri olan J48 (C4.5) algoritması ile istatistik bölümlerinin başarılarına etki eden dersler belirlenmiştir. Bu algoritma sonucunda oluşan sınıflandırma performansını gösteren ölçülerden Correctly Classified Instances %50, Kappa Statistic ise -0.0549 değerini almıştır ve bu değerlerin analizi iyi bir şekilde açıklamadığı belirlenmiştir.

Analiz sonuçlarının kötü olması üniversitelerin başarılarını gösteren sınıf kategorilerinin dengesiz dağılmasından kaynaklanmıştır. Veri kümesindeki bu dengesiz dağılım için WEKA programında bulunan Resample (Yeniden Örneklendirme) seçeneğinin kullanılması ile veri kümesi dengeli bir dağılım haline getirilmiştir ve dengeli duruma getirilen veri kümesine WEKA'da tekrardan karar ağacı analizi uygulanmıştır.

Karar ağacı analizi sonucunda oluşan, Correctly Classified Instances ölçüsü %73.8462, Kappa Statistic ölçüsü ise 0.6731 değerini almıştır. Correctly Classified Instances değerinin %50 'den %73.8462'ye, veriler arasındaki uyumun güvenilirliğini ölçen değerin (Kappa Statistic) ise -0.0549'dan 0.6731'e yükseldiği görülmüştür. Bu ölçüler sınıflandırma modelinin tahminin iyi yapıldığını ifade etmektedir. Bu iki ölçünün aldığı değerler ile birlikte sınıflandırma performansına ait diğer ölçüler ve grafikleri karar ağacı analizini iyi bir şekilde açıklamıştır. Sınıflandırma performansını gösteren ölçü değerlerinin analiz için uygun oldukları belirlendikten sonra karar ağacı tablosunda üniversitelerin başarılarını etkileyen derslerin ifade edildiği 10 karar kuralı oluşturulmuştur. Bu kurallar incelenerek; karar ağacında başarıya etki eden en önemli dersin Lineer Cebir 1 dersi olduğu tespit edilmiştir. Başarıyı etkileyen diğer önemli derslerin ise Olasılığa Giriş ve Regresyon Analizi dersleri olduğu görülmüştür. İstatistik bölümlerine sahip üniversitelerin ders programlarında Lineer Cebir 1 dersi ya da Olasılığa Giriş ve Regresyon Analizi dersleri birlikte yer alıyor ise bu üniversitelerin "Çok İyi" üniversite kategorisinde olduğu gözlemlenmiştir.

Veri madenciliği yöntemleri tez çalışmasında; üniversitelerin istatistik bölümlerinin birbirine olan benzerliklerini, istatistik bölümlerinde başarıya etki eden ve birlikte bulunması gereken derslerin belirlenmesini başarılı bir şekilde analiz etmiştir.

Yapılan bu analizler arasında karar ağacı analizi sonucunda elde edilen karar kuralları ile URAP indeksine göre üst sıralarda bulunan başarılı üniversitelerin istatistik bölümlerindeki eğitim modelleri, URAP indeksinde orta ve

alt sıralarda yer alan üniversitelerin istatistik bölümlerinin başarılarının artması için kullanılabilir. Bu karar kurallarının ve eğitim modellerinin istatistik bölümlerine uygulanması hem bölümlerin hem de üniversitelerin başarılı hale gelmesini sağlayacaktır.

Bu tez çalışmasında yapılan veri madenciliği analizleri sadece istatistik bölümleri bulunan üniversitelere uygulanmıştır. Çalışmadaki analiz sonuçlarından yararlanılarak üniversitelerin farklı bölümleri için analizler veri madenciliği yöntemleri ile yapılabilir. Yapılan analizler sonucunda; üniversitelerin birbirleri ile olan ilişkileri, bölümlerde bulunması gereken dersler, bölümlerin başarılarına etki eden dersler ve bu konuda araştırılmak istenen farklı durumlar veri madenciliği yöntemleri ile ortaya çıkarılmaktadır.

İstatistik bölümleri ile bu bölümlerin anlaşmalı olduğu Erasmus programına bağlı üniversitelerin ders içeriklerinin karşılaştırılarak istatistik lisans programlarındaki eksikliklerin tespit edilmesi ve tamamlanması, istatistik bölümlerinin öğrenci işleri otomasyonu üzerinden elde edilecek lisans öğrencilerinin aldıkları dersleri gösteren veriler ile yapılacak olan analizler sonucunda öğrencilerin başarılarına etki eden faktörlerin belirlenmesi, yeni açılacak üniversitelerdeki bölümlerin ders planlanması hazırlanırken ders programlarında hangi derslerin yer alması gerektiği ve ders içeriklerinin nasıl olacağı hakkındaki bilgilerin bulunması gibi araştırma konularının veri madenciliği analizleri ile yapılması önerilmektedir.

KAYNAKLAR

Agrawal, R., Imielinski, T., Swami, A., 1993. Mining Association Rules Between Sets of Items in Large Databases. SIGMOD '93 Proceedings of The 1993 ACM SIGMOD International Conference on Management of Data, USA, 207-216.

Agrawal, R., Srikant, R., 1994. Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the 20th International Conference on Very Large Data Bases, Santiago de Chile, 487-499.

Akay, E., Ç., 2018. Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi. Sosyal Bilimler Araştırma Dergisi, 7(2): 41-53.

Akbulut S., 2006. Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu. Gazi Üniversitesi Yüksek Lisans Tezi, Ankara, 91s.

Akça, F., 2014. Veri Madenciliği ile Fen Fakülteleri Öğrenci Profillerinin İncelenmesi: Gazi Üniversitesi Örneği. Gazi Üniversitesi Yüksek Lisans Tezi, Ankara, 68s.

Akpınar, H., 2000. Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. İstanbul Üniversitesi İşletme Fakültesi Dergisi, 29(1): 1:22.

Al Ghoson, A., M., 2010. Decision Tree Induction & Clustering Techniques in SAS Enterprise Miner, SPSS Clementine, and IBM Intelligent Miner-a Comparative Analysis. International Journal of Management & Information Systems, 14(3): 57-70.

Al-Radaideh, Q., A., Al-Shawakfa, E., M., Al-Najjar, M., I., 2006. Mining Student Data Using Decision Trees. In International Arab Conference on Information Technology (ACIT'2006), Jordan, 1-5.

Altunkaynak, B., 2017. Veri Madenciliği ve R Uygulamaları. Seçkin Yayıncılık, Ankara, 204s.

Anamisa, D., R., Yusuf, M., Syakur, M., A., 2018. Determination System of Food Vouchers for The Poor Based on Fuzzy C-Means Method. In Journal of Physics: Conference Series, 953(1): 1-7.

Armutlu, Ş., 2018. Veri Madenciliği ile Kütüphane Kullanımı ve Ders Başarısı Arasındaki İlişkinin İncelenmesi. Uşak Üniversitesi Yüksek Lisans Tezi, Uşak, 106s.

Ateş, Y., Karabatak, M., 2017. Nicel Birliktelik Kuralları için Çoklu Minimum Destek Değeri. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 29(2): 57-65.

Atılgan, E., 2011. Karayollarında Meydana Gelen Trafik Kazalarının Karar Ağaçları ve Birliktelik Analizi ile İncelenmesi. Hacettepe Üniversitesi Yüksek Lisans Tezi, Ankara, 88s.

Ayık, Y., Z., Özdemir, A., Yavuz, U., 2007. Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte ile İlişkisinin Veri Madenciliği Tekniği ile Analizi. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 10(2): 441-454.

Baepler, P., Murdoch, C., J., 2010. Academic Analytics and Data Mining in Higher Education. International Journal for The Scholarship of Teaching and Learning, 4(2), Article 17.

Bargavi, C., 2017. Analysis of Accuracy and Performance of Big Data Using Data Mining Techniques in Hadoop Mapreduce. Advances in Natural and Applied Sciences, 11(4): 169-175.

Bikari, S., Ünal, S., Deveci, F., G., 2017. Tüketicilerin Çevreye Yönelik Tutumları ve Kişilik Özellikleri Açısından Kohonen Ağları (Self-Organizing Map-Som) ile Bölümlendirilmesi. Uluslararası İktisadi ve İdari İncelemeler Dergisi, 22. UPK Ahmet Hamdi İslamoğlu Özel Sayısı, 145-162.

Bilen, H., 2009. Bankacılık Sektöründe Personel Seçimi ve Performans Değerlendirilmesine İlişkin Veri Madenciliği Uygulaması. Gazi Üniversitesi Yüksek Lisans Tezi, Ankara, 102s.

Bilen, Ö., Hotaman, D., Aşkın E., Ö., Büyüklü H., A., 2014. LYS Başarılarına Göre Okul Performanslarının Eğitsel Veri Madenciliği Teknikleriyle İncelenmesi: 2011 İstanbul Örneği. Eğitim ve Bilim, 39(172): 78-94.

Bilgin, T., T., 2009. Veri Akışı Diyagramları Tabanlı Veri Madenciliği Araçları ve Yazılım Geliştirme Ortamları. 11.Akademik Bilişim Konferansı, Şanlıurfa, 807-814.

Boser, B., E., Guyon, I., M., Vapnik, V., N., 1992. A Training Algorithm for Optimal Margin Classifiers. COLT '92 Proceedings of The Fifth Annual Workshop on Computational Learning Theory, USA, 144-152.

Boyacı, A., 2017. Öğretmenlerin Algılanan Örgütsel Destek ve Örgütsel Özdeşleme Düzeylerinin Veri Madenciliği ile Analizi. Hitit Üniversitesi Yüksek Lisans Tezi, Çorum, 80s.

Bölükbaşı, İ., 2005. Defining Critical Factors Affecting Student Success: A Data Mining Approach. İstanbul Teknik Üniversitesi Yüksek Lisans Tezi, İstanbul, 127s.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A., 1998. Discovering Data Mining: From Concept to Implementation. Prentice Hall, 517s.

Çalış, A., 2013.Veri Madenciliği Yaklaşımı ile Bireysel Müşterilerin Kredi Ödeme Performanslarının Değerlendirilmesi. Kocaeli Üniversitesi Yüksek Lisans Tezi, Kocaeli, 95s.

Çalışkan, S., K., Soğukpınar, İ., 2008. KxKNN: K-Means ve K En Yakın Komşu Yöntemleri ile Ağlarda Nüfuz Tespiti. EMO Yayınları, 120-124.

Dean, J., 2014. Big Data, Data Mining and Machine Learning. John Wiley & Sons, 265s.

Delavari, N., Shirazi, M., R., A., Beikzadeh, M., R., 2004. A New Model for Using Data Mining Technology in Higher Educational Systems. Information Technology Based Proceedings of The Fifth International Conference onHigher Education and Training, Turkey, 319-324.

Dinçer, E., Ş., 2006. Veri Madenciliğinde K-Means Algoritması ve Tıp Alanında Uygulanması. Kocaeli Üniversitesi Yüksek Lisans Tezi, Kocaeli, 101s.

Döşlü, A., 2008. Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi. Yıldız Teknik Üniversitesi Yüksek Lisans Tezi, İstanbul, 100s.

Dunham, M., H., 2003. Data Mining Introductory and Advanced Topics. Pearson Education, 315s.

Emel, G. G., Taşkın, Ç., 2005. Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. Sosyal Bilimler Dergisi, 6(2): 221-239.

Emre, İ., E., Erol, Ç., S., 2017. Veri Analizinde İstatistik mi Veri Madenciliği mi?. Bilişim Teknolojileri Dergisi, 10(2): 161-167.

Flawley, W., J., Piatetsky-Shapiro, G., Matheus, C., J., 1992. Knowledge Discovery in Databases: An Overview. AI Magazine, 13(3): 57-70.

Ghosh, S., Dubey, S., K., 2013. Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. International Journal of Advanced Computer Science and Applications, 4(4): 35-39.

Gülçe, G., 2010. Veri Ambarı ve Veri Madenciliği Teknikleri Kullanılarak Öğrenci Karar Destek Sistemi Oluşturma. Pamukkale Üniversitesi Yüksek Lisans Tezi, Pamukkale, 97s.

Gülen, Ç., Ö., 2014. Veri Madenciliği Teknikleri ile Üstün Yetenekli Öğrencilerin İlgi Alanlarının Analizi. Gazi Üniversitesi Doktora Tezi, Ankara, 160s.

Güven, B., Z., 2016. Türk Üniversitelerindeki Bilgisayar Mühendisliği Bölümleri Müfredatları Kullanılarak Veri Madenciliği Uygulaması Gerçekleştirilmesi. Maltepe Üniversitesi Yüksek Lisans Tezi, İstanbul, 68s.

Güvenç, E., 2001. Student Performance Assessment in Higher Education Using Data Mining. Boğaziçi Üniversitesi Yüksek Lisans Tezi, İstanbul, 157s.

Hamzaoui, Y., Amnai, M., Choukri, A., Fakhri, Y., 2018. Novel Clustering Method Based on K-Medoids and Mobility Metric. International Journal of Interactive Multimedia and Artificial Intelligence, 5(1): 29-33.

Han, J., Kamber, M., 2000. Data Mining Concepts and Techniques. Morgan Kaufmann, 550s.

Han, J., Kamber, M., 2006. Data Mining Concepts and Techniques. Morgan Kaufmann, 770s.

Hand D., Mannila H., Smyth P., 2001. Principles of Data Mining. The MIT Press, 546s.

İnan O., 2003. Öğrenci İşleri Veri Tabanı Üzerinde Veri Madenciliği Uygulamaları. Selçuk Üniversitesi Yüksek Lisans Tezi, Konya, 81s.

Joyendri, A., 2017. Strategi Customer Relationship Management Untuk Meningkatkan Loyalitas Pelanggan Dan Volume Penjualan Menggunakan Teknik Clustering K-Means. Telematika, 14(2): 75–82.

Kalıkov, A., 2006. Veri Madenciliği ve Bir E-Ticaret Uygulaması. Gazi Üniversitesi Yüksek Lisans Tezi, Ankara, 94s.

Kaufman, L., Rosseeauw, P., J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley and Sons Inc, 355s.

Kaya, M., Özel, A., S., 2014. Türkçe Dokümanlardaki Benzerliklerin Tespiti için Mevcut Yazılımların Karşılaştırılması ve Türkçe Karakter Kullanımı ile Kök Almanın Etkisinin İncelenmesi. Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 29(2): 115-130.

Kevran, M., M., 2009. Çoklu Sensör Konumlandırma Probleminin Genetik Algoritmalar ile Çözülmesi. Hava Harp Okulu Komutanlığı, Havacılık ve Uzay Teknolojileri Enstitüsü Yüksek Lisans Tezi, İstanbul, 114s.

Kılıçer, S., 2018. Veri Madenciliği ile Türkiye'deki ve Avrupa Birliği Ülkelerindeki Bilgisayar Mühendisliği Bölümleri Ders İçeriklerinin Karşılaştırılması. İstanbul Üniversitesi Yüksek Lisans Tezi, İstanbul, 103s.

Kittler, R., Wang, W., 1999. The Emerging Role for Data Mining. Solid State Technology, 42(11): 45s.

Kohonen, T., 2013. Essentials of the Self-Organizing Map. Neural Networks, 37: 52–65.

Kumar, V., Chadha, A., 2011. An Empirical Study of The Applications of Data Mining Techniques in Higher Education. International Journal of Advanced Computer Science and Applications, 2(3): 80-84.

Larose, D., T., 2005. Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, 217s.

Liao, S., H., Wen, C., H., 2007. Artificial Neural Networks Classification and Clustering of Methodologies and Applications–Literature Analysis from 1995 to 2005. Expert Systems with Applications, 32(1): 1-11.

Linoff, G., S., Berry, M., J., 1999. Mastering Data Mining: The Art and Science of Customer Relationship Management. John Wiley & Sons, 512s.

MacQueen, B., J., 1967. Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1(14): 281-297.

Modak, S., Chattopadhyay, T., Chattopadhyay, A., K., 2019. Unsupervised Classification of Eclipsing Binary Light Curves Through K-Medoids Clustering. Journal of Applied Statistics, 1-17. https://arxiv.org/pdf/1801.09406.pdf (Erişim tarihi: 1 Ağustos 2019).

Olgun, O., M., Özdemir, G., 2012. İstatistiksel Özellik Temelli Bayes Sınıflandırıcı Kullanarak Kontrol Grafiklerinde Örüntü Tanıma. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 27(2): 303-311.

Öğüt, S., 2009. Veri Madenciliği Kavramı ve Gelişim Süreci. Yeditepe Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 12s.

Özcan, C., 2014. Veri Madenciliğinin Güvenlik Uygulama Alanları ve Veri Madenciliği ile Sahtekârlık Analizi. İstanbul Bilgi Üniversitesi Yüksek Lisans Tezi, İstanbul, 90s.

Özçalıcı, M., 2016. Hisse Senetlerinin Özdüzenleyici Haritalarla Kümelendirilmesi: BIST50 Endeksinde Yer Alan Hisseler Üzerine Bir Uygulama. İstanbul Üniversitesi İşletme Fakültesi Dergisi, 45(1): 22-33.

Özekes, S., 2003. Veri Madenciliği Modelleri ve Uygulama Alanları. İstanbul Ticaret Üniversitesi Dergisi 3: 65-82.

Öztemel, E., 2003. Yapay Sinir Ağları. Papatya Yayıncılık, İstanbul, 232s.

Pfitzner, D., Leibbrandt, R., Powers, D., 2009. Characterization and Evaluation of Similarity Measures for Pairs of Clusterings. Knowledge and Information Systems,19(3): 361-394.

Quinlan, J., R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, 302s.

Romero, C., Ventura, S., 2013. Data Mining in Education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1): 12-27.

Sarıman, G., 2011. Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması. Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü Dergisi, 15(3): 192-202.

Savaş, S., Topaloğlu, N., Yılmaz, M., 2012. Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 11(21): 1-23.

Sezgin, E., Çelik Y., 2013. Veri Madenciliğinde Kayıp Veriler için Kullanılan Yöntemlerin Karşılaştırılması. 15. Akademik Bilişim Konferansı, Antalya, 611- 614.

Shannon, C., E., 1948. A Mathematical Theory of Communication. Bell System Technical Journal, 27(3): 379-423.

Silahtaroğlu, G., 2008. Kavram ve Algoritmalarıyla Temel Veri Madenciliği. Papatya Yayıncılık, İstanbul, 173s.

Taşkın, Ç., Emel, G., G., 2010. Veri Madenciliğinde Kümeleme Yaklaşımları ve Kohonen Ağları ile Perakendecilik Sektöründe Bir Uygulama. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 15(3): 395-409.

Tekerek, A., 2011. Veri Madenciliği Süreçleri ve Açık Kaynak Kodlu Veri Madenciliği Araçları. 13. Akademik Bilişim Konferansı, Malatya, 161-169.

Tekin, B., 2015. Temel Sağlık Göstergeleri Açısından Türkiye'deki İllerin Gruplandırılması: Bir Kümeleme Analizi Uygulaması. Çankırı Karatekin Üniversitesi İİBF Dergisi, 5(2): 389-416.

Terlemez, L., 2008. Eş İşlem Stratejisi Yöntemiyle İMKB'de Portföy Oluşturmada Veri Madenciliği Uygulaması. Anadolu Üniversitesi Doktora Tezi, Eskişehir, 152s.

URAP, http://tr.urapcenter.org/2018/. (2018). (Erişim Tarihi : 26 Mart 2018).

Vialardi, C., Bravo, J., Shafti, L., Ortigosa, A., 2009. Recommendation in Higher Education Using Data Mining Techniques. International Conference on Educational Data Mining (EDM), Spain, 190-199.

Wu, H., Yang, S., Huang, Z., He, J., Wang, X., 2018. Type 2 Diabetes Mellitus Prediction Model Based on Data Mining. Informatics in Medicine Unlocked, 10: 100-107.

Yalçın, L., 2019. Sağlık Sektöründe Veri Madenciliği. Milli Savunma Üniversitesi Yüksek Lisans Tezi, İstanbul, 91s.

Yüksel, E., M., Turna, C., Ö., Ertürk, A., M., 2010. Bilgiye Erişim Sistemlerinde Veri Arama ve Eşleştirme. 12. Akademik Bilişim Konferansı, Muğla. https://docplayer.biz.tr/13058826-Bilgiye-erisim-sistemlerinde-veri-arama-ve-eslestirme.html (Erişim Tarihi: 11 Ağustos 2019).

Zare, M., Koch, M., 2018. Groundwater Level Fluctuations Simulation and Prediction By ANFIS-and Hybrid Wavelet-ANFIS/Fuzzy C-Means (FCM) Clustering Models: Application to The Miandarband Plain. Journal of Hydro-Environment Research, 18: 63-76.

ÖZGEÇMİŞ

07.02.1992 yılında Adana/Kozan ilçesinde doğdu. İlk, orta ve lise eğitimini Adana'da tamamladı. 2010 yılında Çukurova Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü'nde lisans eğitimine başladı. Bir yıl ingilizce hazırlık eğitimi aldıktan sonra lisans eğitiminden 2015 yılında mezun oldu ve 2016 yılında İstatistik Bölümünde yüksek lisansa başladı.

EKLER

Ek 1. İstatistik Bölümü Derslerinin Eş değerlik Tabloları

Temel Bilgi Teknolojileri Kullanımı	Temel Bilgi Teknolojisi Kullanımı I, Temel Bilgi Teknolojisi Kullanımı II, Temel Bilgi Teknolojisi Kullanımı, Bilgisayar I, Bilgisayar II, Bilgisayar Programlama I, Bilgisayara Giriş.
Olasılığa Giriş	Olasılığa Giriş I, Olasılık I, Olasılık ve İstatistik I, Olasılık, Olasılık 1, Olasılık 2, Olasılık ve İstatistiğe Giriş I, Matematiksel İstatistik ve Olasılık 1.
Matematik 1	Matematik I, Matematik II.
İstatistik İçin Matematik I	İleri Analiz I, İleri Analiz II.
Varyans Analizi ve Deneysel Tasarımlar	Deney Tasarımı, İstatistiksel Yöntemler.
İstatistiksel Sonuç Çıkarım	Matematiksel İstatistik 2, Matematiksel İstatistik II.
İstatistiksel Kalite Kontrol	İstatistiksel Kalite Kontrolü, Kalite Kontrol.
Çok Değişkenli İstatistiksel Analiz 2	Çok Değişkenli İstatistik II, Çok Değişkenli İstatistiksel Analizler.
Lineer Cebir 1	Lineer Cebir I
Lineer Cebir 2	Lineer Cebir II
Matematik 2	Matematik II, İstatistik için Matematik II.
Optimizasyon Teknikleri I	Optimizasyon I, Optimizasyon II, Yöneylem Araştırması I
Regresyon Analizi	Doğrusal Modeller I
Aktüerya Modelleri	Aktüerya
Nümerik Analiz	Sayısal Analiz
Parametrik Olmayan İstatistik	Parametrik Olmayan İstatistiksel Yöntemler
Matris Kuramı	Lineer Cebir I
Lineer Modeller	Lineer Modeller I

Ek 2. URAP İndeksine Göre Üniversitelerin Başarı Sıralamaları

Sıralama	Üniversitelerin Adı
1	HACETTEPE ÜNİVERSİTESİ
2	ORTA DOĞU TEKNİK ÜNİVERSİTESİ
3	İSTANBUL ÜNİVERSİTESİ
4	İSTANBUL TEKNİK ÜNİVERSİTESİ
5	ANKARA ÜNİVERSİTESİ
6	KOÇ ÜNİVERSİTESİ
7	İHSAN DOĞRAMACI BİLKENT ÜNİVERSİTESİ
8	GAZİ ÜNİVERSİTESİ
9	EGE ÜNİVERSİTESİ
10	BOĞAZİÇİ ÜNİVERSİTESİ
11	GEBZE TEKNİK ÜNİVERSİTESİ
12	SABANCI ÜNİVERSİTESİ
13	BEZM-İ ÂLEM VAKIF ÜNİVERSİTESİ
14	YILDIZ TEKNİK ÜNİVERSİTESİ
15	ATATÜRK ÜNİVERSİTESİ
16	ERCİYES ÜNİVERSİTESİ
17	MARMARA ÜNİVERSİTESİ
18	İZMİR YÜKSEK TEKNOLOJİ ENSTİTÜSÜ
19	DOKUZ EYLÜL ÜNİVERSİTESİ
20	SELÇUK ÜNİVERSİTESİ
21	ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ
22	KARADENİZ TEKNİK ÜNİVERSİTESİ
23	AKDENİZ ÜNİVERSİTESİ
24	ÇUKUROVA ÜNİVERSİTESİ
25	FIRAT ÜNİVERSİTESİ
26	BURSA ULUDAĞ ÜNİVERSİTESİ
27	ONDOKUZ MAYIS ÜNİVERSİTESİ

Ek 2. Devam

28	SÜLEYMAN DEMİREL ÜNİVERSİTESİ
29	SAKARYA ÜNİVERSİTESİ
30	BAŞKENT ÜNİVERSİTESİ
31	GAZİANTEP ÜNİVERSİTESİ
32	İNÖNÜ ÜNİVERSİTESİ
33	ANADOLU ÜNİVERSİTESİ
34	ABDULLAH GÜL ÜNİVERSİTESİ
35	DİCLE ÜNİVERSİTESİ
36	İZMİR KÂTİP ÇELEBİ ÜNİVERSİTESİ
37	KOCAELİ ÜNİVERSİTESİ
38	İSTANBUL MEDENİYET ÜNİVERSİTESİ
39	ANKARA YILDIRIM BEYAZIT ÜNİVERSİTESİ
40	PAMUKKALE ÜNİVERSİTESİ
41	VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
42	ÇANKAYA ÜNİVERSİTESİ
43	MANİSA CELÂL BAYAR ÜNİVERSİTESİ
44	YEDİTEPE ÜNİVERSİTESİ
45	TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ
46	ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ
47	ACIBADEM MEHMET ALİ AYDINLAR ÜNİVERSİTESİ
48	BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
49	TOKAT GAZİOSMANPAŞA ÜNİVERSİTESİ
50	HATAY MUSTAFA KEMAL ÜNİVERSİTESİ
51	SİVAS CUMHURİYET ÜNİVERSİTESİ
52	DÜZCE ÜNİVERSİTESİ
53	NECMETTİN ERBAKAN ÜNİVERSİTESİ
54	MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ
55	KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ

Ek 2. Devam

56	RECEP TAYYİP ERDOĞAN ÜNİVERSİTESİ
57	KIRIKKALE ÜNİVERSİTESİ
58	ZONGULDAK BÜLENT ECEVİT ÜNİVERSİTESİ
59	TEKİRDAĞ NAMIK KEMAL ÜNİVERSİTESİ
60	İSTANBUL MEDİPOL ÜNİVERSİTESİ
61	ATILIM ÜNİVERSİTESİ
62	ÖZYEĞİN ÜNİVERSİTESİ
63	ADNAN MENDERES ÜNİVERSİTESİ
64	MERSİN ÜNİVERSİTESİ
65	HARRAN ÜNİVERSİTESİ
66	BALIKESİR ÜNİVERSİTESİ
67	HİTİT ÜNİVERSİTESİ
68	TRAKYA ÜNİVERSİTESİ
69	KÜTAHYA DUMLUPINAR ÜNİVERSİTESİ
70	BURSA TEKNİK ÜNİVERSİTESİ
71	AFYON KOCATEPE ÜNİVERSİTESİ
72	BAHÇEŞEHİR ÜNİVERSİTESİ
73	SİNOP ÜNİVERSİTESİ
74	YOZGAT BOZOK ÜNİVERSİTESİ
75	İSTANBUL BİLİM ÜNİVERSİTESİ
76	NİĞDE ÖMER HALİSDEMİR ÜNİVERSİTESİ
77	ADIYAMAN ÜNİVERSİTESİ
78	GALATASARAY ÜNİVERSİTESİ
79	ERZURUM TEKNİK ÜNİVERSİTESİ
80	AKSARAY ÜNİVERSİTESİ
81	DOĞUŞ ÜNİVERSİTESİ
82	KARABÜK ÜNİVERSİTESİ
83	ÇANKIRI KARATEKİN ÜNİVERSİTESİ

Ek 2. Devam

84	KADİR HAS ÜNİVERSİTESİ
85	NEVŞEHİR HACI BEKTAŞ VELİ ÜNİVERSİTESİ
86	YALOVA ÜNİVERSİTESİ
87	KAFKAS ÜNİVERSİTESİ
88	ORDU ÜNİVERSİTESİ
89	ERZİNCAN BİNALİ YILDIRIM ÜNİVERSİTESİ
90	İZMİR EKONOMİ ÜNİVERSİTESİ
91	MUNZUR ÜNİVERSİTESİ
92	BİNGÖL ÜNİVERSİTESİ
93	KIRŞEHİR AHİ EVRAN ÜNİVERSİTESİ
94	OSMANİYE KORKUT ATA ÜNİVERSİTESİ
95	GİRESUN ÜNİVERSİTESİ
96	YAŞAR ÜNİVERSİTESİ
97	GÜMÜŞHANE ÜNİVERSİTESİ
98	KASTAMONU ÜNİVERSİTESİ
99	BARTIN ÜNİVERSİTESİ
100	BİLECİK ŞEYH EDEBALİ ÜNİVERSİTESİ
101	BİTLİS EREN ÜNİVERSİTESİ
102	AMASYA ÜNİVERSİTESİ
103	IŞIK ÜNİVERSİTESİ
104	KARAMANOĞLU MEHMET BEY ÜNİVERSİTESİ
105	BAYBURT ÜNİVERSİTESİ
106	İSTANBUL BİLGİ ÜNİVERSİTESİ
107	İSTANBUL AREL ÜNİVERSİTESİ
108	BATMAN ÜNİVERSİTESİ
109	BURBUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
110	TÜRK HAVA KURUMU ÜNİVERSİTESİ
111	SİİRT ÜNİVERSİTESİ

Ek 2. Devam

112	MALTEPE ÜNİVERSİTESİ
113	ALTINBAŞ ÜNİVERSİTESİ
114	İSTANBUL TİCARET ÜNİVERSİTESİ
115	KİLİS 7 ARALIK ÜNİVERSİTESİ
116	İSTANBUL OKAN ÜNİVERSİTESİ
117	UŞAK ÜNİVERSİTESİ
118	AĞRI İBRAHİM ÇEÇEN ÜNİVERSİTESİ
119	KTO KARATAY ÜNİVERSİTESİ
120	ARTVİN ÇORUH ÜNİVERSİTESİ
121	İSKENDERUN TEKNİK ÜNİVERSİTESİ
122	TED ÜNİVERSİTESİ
123	UFUK ÜNİVERSİTESİ
124	HASAN KALYONCO ÜNİVERSİTESİ
125	MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
126	İSTANBUL ŞEHİR ÜNİVERSİTESİ
127	HAKKARİ ÜNİVERSİTESİ
128	IĞDIR ÜNİVERSİTESİ
129	PİRİ REİS ÜNİVERSİTESİ
130	İSTANBUL AYDIN ÜNİVERSİTESİ
131	İSTANBUL KÜLTÜR ÜNİVERSİTESİ
132	ÜSKÜDAR ÜNİVERSİTESİ
133	ANTALYA BİLİM ÜNİVERSİTESİ
134	ÇAĞ ÜNİVERSİTESİ
135	SAĞLIK BİLİMLERİ ÜNİVERSİTESİ
136	MUŞ ALPARSLAN ÜNİVERSİTESİ
137	ARDAHAN ÜNİVERSİTESİ
138	KIRKLARELİ ÜNİVERSİTESİ
139	HALİÇ ÜNİVERSİTESİ

Ek 2. Devam

140	İSTANBUL YENİ YÜZYIL ÜNİVERSİTESİ
141	BEYKENT ÜNİVERSİTESİ
142	ANKARA SOSYAL BİLİMLER ÜNİVERSİTESİ
143	İSTANBUL GELİŞİM ÜNİVERSİTESİ
144	İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ
145	ŞIRNAK ÜNİVERSİTESİ
146	MARDİN ARTUKLU ÜNİVERSİTESİ
147	İSTANBUL GEDİKLİ ÜNİVERSİTESİ
148	TOROS ÜNİVERSİTESİ
149	FATİH SULTAN MEHMET VAKIF ÜNİVERSİTESİ
150	ALANYA ALAADDİN KEYKUBAT ÜNİVERSİTESİ
151	BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ
152	İSTANBUL 29 MAYIS ÜNİVERSİTESİ
153	TÜRK-ALMAN ÜNİVERSİTESİ
154	AVRASYA ÜNİVERSİTESİ
155	NUH NACİ YAZGAN ÜNİVERSİTESİ
156	İSTANBUL ESENYURT ÜNİVERSİTESİ
157	NİŞANTAŞI ÜNİVERSİTESİ

Ek 3. Öğe Kümesinde 1 Ders Bulunan Birliktelik Tablosu

ITEMSET	SUPPORT	ITEM COUNT
Olasılığa Giriş	0,625	1
Matematik 1	0,208	1
Lineer Cebir 1	0,167	1
Lineer Cebir 2	0,25	1
Matematik 2	0,125	1
Optimizasyon Teknikleri I	0,125	1
İstatistik İçin Matematik I	0,125	1
Aktüerya Modelleri	0,125	1
Varyans Analizi ve Deneysel Tasarımlar	0,125	1
İstatistiksel Sonuç Çıkarım	0,125	1
İstatistiksel Kalite Kontrol	0,167	1
Kombinatorik	0,125	1
Lineer Modeller	0,25	1
Çok Değişkenli İstatistiksel Analiz 2	0,292	1

Ek 4. Öğe Kümesinde 2 Ders Bulunan Birliktelik Tablosu

ITEMSET	SUPPORT	ITEM COUNT
Olasılığa Giriş Matematik 1	0,167	2
Lineer Cebir 1 Olasılığa Giriş	0,167	2
Lineer Cebir 2 Olasılığa Giriş	0,208	2
Olasılığa Giriş Matematik 2	0,125	2
Olasılığa Giriş Aktüerya Modelleri	0,125	2
Olasılığa Giriş İstatistiksel Sonuç Çıkarım	0,125	2
Kombinatorik Olasılığa Giriş	0,125	2
Lineer Modeller Olasılığa Giriş	0,167	2
Olasılığa Giriş Çok Değişkenli İstatistiksel Analiz 2	0,292	2
Matematik 1 Aktüerya Modelleri	0,125	2
Kombinatorik Matematik 1	0,125	2
Lineer Cebir 2 Lineer Cebir 1	0,125	2
Lineer Modeller Lineer Cebir 2	0,125	2
Lineer Cebir 2 Çok Değişkenli İstatistiksel Analiz 2	0,125	2
İstatistik İçin Matematik I Optimizasyon Teknikleri I	0,125	2

Ek 5. Öğe Kümesinde 3 Ders Bulunan Birliktelik Tablosu

ITEMSET	SUPPORT	ITEM COUNT
Olasılığa Giriş Matematik 1 Aktüerya Modelleri	0,125	3
Kombinatorik Olasılığa Giriş Matematik 1	0,125	3
Lineer Cebir 2 Lineer Cebir 1 Olasılığa Giriş	0,125	3
Lineer Cebir 2 Olasılığa Giriş Çok Değişkenli İstatistiksel Analiz 2	0,125	3

Ek 6. Birliktelik Kuralları Tablosu

	Antecedent	Consequent	Support%	Confidence%	Lift
1.Kural	İstatistiksel Sonuç Çıkarım	Olasılığa Giriş	12,5	100	1,6
2.Kural	Matematik 2	Olasılığa Giriş	12,5	100	1,6
3.Kural	Optimizasyon Teknikleri I	İstatistik İçin Matematik I	12,5	100	8
4.Kural	Kombinatorik	Matematik 1	12,5	100	4,8
5.Kural	Kombinatorik	Olasılığa Giriş	12,5	100	1,6
6.Kural	Aktüerya Modelleri	Matematik 1	12,5	100	4,8
7.Kural	Aktüerya Modelleri	Olasılığa Giriş	12,5	100	1,6
8.Kural	Lineer Cebir 1	Lineer Cebir 2	12,5	75	3
9.Kural	Lineer Cebir 1	Olasılığa Giriş	16,667	100	1,6
10.Kural	Matematik 1	Olasılığa Giriş	16,667	80	1,28
11.Kural	Çok Değişkenli İstatistiksel Analiz 2	Olasılığa Giriş	29,167	100	1,6
12.Kural	Lineer Cebir 2	Olasılığa Giriş	20,833	83,3	1,333

Ek 6. Devam

13.Kural	Kombinatorik Matematik 1 Matematik 1	Olasılığa Giriş	12,5	100	1,6
14.Kural	Kombinatorik Olasılığa Giriş	Matematik 1	12,5	100	4,8
15.Kural	Matematik 1 Olasılığa Giriş	Kombinatorik	12,5	75	6
16.Kural	Aktüerya Modelleri Matematik 1	Olasılığa Giriş	12,5	100	1,6
17.Kural	Aktüerya Modelleri Olasılığa Giriş	Matematik 1	12,5	100	4,8
18.Kural	Matematik 1 Olasılığa Giriş	Aktüerya Modelleri	12,5	75	6
19.Kural	Lineer Cebir 1 Lineer Cebir 2	Olasılığa Giriş	12,5	100	1,6
20.Kural	Lineer Cebir 1 Olasılığa Giriş	Lineer Cebir 2	12,5	75	3
21.Kural	Çok Değişkenli İstatistiksel Analiz 2 Lineer Cebir 2	Olasılığa Giriş	12,5	100	1,6