ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Gözde ÖZSERT YİĞİT

PREDICTING THE ADMISSION DECISION OF A PARTICIPANT TO THE SCHOOL OF PHYSICAL EDUCATION AND SPORTS AT CUKUROVA UNIVERSITY BY USING DIFFERENT MACHINE LEARNING METHODS COMBINED WITH FEATURE SELECTION

DEPARTMENT OF COMPUTER ENGINEERING

ADANA, 2016

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

PREDICTING THE ADMISSION DECISION OF A PARTICIPANT TO THE SCHOOL OF PHYSICAL EDUCATION AND SPORTS AT CUKUROVA UNIVERSITY BY USING DIFFERENT MACHINE LEARNING METHODS COMBINED WITH FEATURE SELECTION

Gözde ÖZSERT YİĞİT

MSc THESIS

DEPARTMENT OF COMPUTER ENGINEERING

We certify that the thesis titled above was reviewed and approved for the award of degree of the Master of Science by the board of jury on

…….....	…….....	…….....
Assoc. Prof. Dr. M. Fatih AKAY	Assoc. Prof. Dr. Ramazan ÇOBAN	Asst. Prof. Dr. Şule ÇOLAK
SUPERVISOR	MEMBER	MEMBER

This MSc Thesis is written at the Department of Institute of Natural And Applied Sciences of Çukurova University.
Registration Number:

Prof. Dr. Mustafa GÖK
Director
Institute of Natural and Applied Sciences

This study was supported by Research Projects Unit Ç.Ü. Project Number: FYL-2015-3845.

Not: The usage of the presented specific declerations, tables, figures, and photographs either in this thesis or in any other reference without citiation is subject to "The law of Arts and Intellectual Products" number of 5846 of Turkish Republic

ABSTRACT

MASTER THESIS

PREDICTING THE ADMISSION DECISION OF A PARTICIPANT TO THE SCHOOL OF PHYSICAL EDUCATION ANS SPORTS AT CUKUROVA UNIVERSITY BY USING DIFFERENT MACHINE LEARNING METHODS COMBINED WITH FEATURE SELECTION

Gözde ÖZSERT YİĞİT

ÇUKUROVAUNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF COMPUTER ENGINEERING

Supervisor : Assoc. Prof. Dr. Mehmet Fatih AKAY
Year: 2016, Pages 129
Jury : Assoc. Prof. Dr. Mehmet Fatih AKAY
: Assoc. Prof. Dr. Ramazan ÇOBAN
: Asst. Prof. Dr. Şule ÇOLAK

The purpose of this thesis is to develop new hybrid admission decision prediction models by using different machine learning methods including Support Vector Machines (SVM), Multilayer Perceptron (MLP), Radial Basis Function (RBF) Network, TreeBoost (TB) and K-Means Clustering (KMC) combined with feature selection algorithms to investigate the effect of the predictor variables on the admission decision of a candidate to the School of Physical Education and Sports at Cukurova University. Three feature selection algorithms including Relief-F, F-Score and Correlation-based Feature Selection (CFS) have been considered. Experiments have been conducted on the datasets, which contain data of participants who applied to the School in 2006 and 2007. The datasets have been randomly split into training and test sets using 10-fold cross validation as well as different percentage ratios. The performance of the prediction models for the datasets has been assessed using classification accuracy, specificity, sensitivity, positive predictive value (PPV) and negative predictive value (NPV). The results show that a decrease in the number of predictor variables in the prediction models usually leads to a parallel decrease in classification accuracy.

Key Words: Machine learning, feature selection, physical ability test

ÖZ

YÜKSEK LİSANS TEZİ

ÇUKUROVA ÜNİVERSİTESİ BEDEN EĞİTİMİ VE SPOR YÜKSEKOKULUNA BAŞVURAN ADAYLARIN KABUL EDİLİP EDİLMEYECEĞİNİ NİTELİK SEÇME ALGORİTMALARI İLE BİRLEŞTİRİLMİŞ FARKLI MAKİNE ÖĞRENME YÖNTEMLERİ KULLANARAK TAHMİN ETME

Gözde ÖZSERT YİĞİT

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Danışman : Doç. Dr. M. Fatih AKAY
Yıl: 2016, Sayfa: 129
Jüri : Doç. Dr. M. Fatih AKAY
: Doç. Dr. Ramazan ÇOBAN
: Yrd. Doç. Dr. Şule ÇOLAK

Bu tezin amacı Çukurova Üniversitesi Beden Eğitimi ve Spor Yüksekokuluna başvuran adayların kabul edilip edilmeyeceğini nitelik seçme algoritmaları ile birleştirilmiş Destek Vektör Makineleri (SVM), Çok Katmanlı Algılayıcı (MLP), Radyal Tabanlı Fonksiyon Ağları, Ağaç Yükseltme ve K-Ortalama Sınıflandırma Algoritmaları kullanarak tahmin etmektir. Nitelik seçme algoritmaları olarak Relif-F, Korelasyon-tabanlı Nitelik Seçme (CFS) ve F-Score algoritmaları seçilmiştir. Veri kümeleleri 10 katlı çapraz doğrulama ve farklı yüzdesel oranlar kullanılarak eğitim ve test kümelerine ayrılmıştır. Deneyler sırasıyla 2006 ve 2007 yıllarında okula alınan adayların gerçek test sonuçları üzerinde yapılmıştır. Her iki veri seti üzerinde sınıflayıcıların sınıflandırma performansı; doğruluk, duyarlılık, özgünlük, pozitif öngörü değeri ve negatif öngörü değeri kullanılarak değerlendirilmiştir. Sonuçlar tahmin modellerindeki tahmin değişkenleri azaldıkça doğruluğun da paralel olarak azaldığını göstermektedir.

Anahtar Kelimeler: Makine öğrenme, nitelik seçme, fiziksel yetenek testi

DEDICATION

... to my husbandTalat

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Assoc. Prof. Dr. M. Fatih AKAY, for his supervision guidance, encouragements, patience, motivation, useful suggestions and his valuable time for this work.

I would like to thank members of the MSc thesis jury, Assoc. Prof. Dr. Ramazan ÇOBAN and Asst. Prof. Dr. Şule ÇOLAK, for their suggestions and corrections.

I would also like to thank Cukurova University Scientific Research Projects Center for supporting this work (Project no: FYL-2015-3845)**.**

Last but not the least, I would like to thank my husband Talat Yiğit and my family for their endless support and encouragements for my life and career.

CONTENTS **PAGE**

ABSTRACT I
ÖZ II
DEDICATION III
ACKNOWLEDGEMENTS IV
CONTENTS V
LIST OF TABLES VII
LIST OF FIGURES IX
LIST OF ABBREVIATONS XIX
l. INTRODUCTION 1
1.1. The Vertical Jump Test 1
1.2. The Coordination and Skill Test 2
1.3. 30-meter Dash Test 3
1.4. 20-meter ShuttleRun Test 3
1.5 Determining Point of Participant Admission 4
1.6. Previous Works 5
1.7. Feature Selection 7
1.8. Motivation, Purpose and Contributions of This Thesis 7
1.9. Overview of Dataset 9
2. OVERVIEW OF METHODS 11
2.1. Support Vector Machines 11
2.2. Multi-Layer Perceptrons 16
2.3. Radial Basis Function Network 19
2.4. TreeBoost 20
2.5. K-Means Clustering 21
2.6. Feature Selection Algortihms 23
3. DEVELOPMENT OF PREDICTION MODELS 27
3.1. SVM Model for Predicting Admission Decision 27
3.2. MLP Model for Predicting Admission Decision 30
3.3. RBF Network Model for Predicting Admission Decision 32

3.4. TB Model for Predicting Admission Decision 33
3.5. KMC Model for Predicting Admission Decision 34
3.6. Performance Metrics 36
4. RESULTS AND DISCUSSION 41
4.1. General Discussion on the Results 114
4.2. Discussion for Classification Accuracy on the 2006 Dataset 115
4.3. Discussion for Classification Accuracy on the 2007 Dataset 117
4.4. Discussion on Other Performance Criteria 118
5. CONCLUSION 123
REFERENCES 125
CURRICULUM VITAE 129

LIST OF TABLES **PAGE**

Table 1.1. Results of the SVM and MLP prediction models on the 2006 datasets.....6
Table 1.2. Results of the SVM and MLP prediction models on the 2007 datasets.....6
Table 1.3. Statistical analysis of each predictor variable for the 2006 dataset 9
Table 1.4. Statistical analysis of each predictor variable for the 2007 dataset 10
Table 3.1. The intervals for values of the parameters for the SVM model..... 30
Table 3.2. The intervals for values of the parameters for MLP models..... 30
Table 3.3. Important parameters of the RBF classification models 33
Table 3.4. Important parameters and their values of the TB classification models . 33
Table 3.5. Important parameters and their values of the K-means classification models 36
Table 3.6. Confusion Matrix 36
Table 4.1. Relief-F scores of the attributes for the 2006 dataset..... 41
Table 4.2. Relief-F scores of the attributes for the 2007 dataset..... 41
Table 4.3. F-scores of the attributes for the 2006 dataset 42
Table 4.4. F-scores of the attributes for the 2007 dataset. 42
Table 4.5. List of selected attributes for the 2006 dataset..... 42
Table 4.6. List of selected attributes for the 2007 dataset..... 43
Table 4.7. Admission decision models based on Relief-F for the 2006 dataset 43
Table 4.8. Admission decision models based on F-score for the 2006 dataset..... 43
Table 4.9. Admission decision model based on CFS for the 2006 dataset 44
Table 4.10. Admission decision models based on Relief-F for the 2007 dataset 44
Table 4.11. Admission decision models based on F-score for the 2007 dataset..... 44
Table 4.12. Admission decision model based on CFS for the 2007 dataset 44

LIST OF FIGURES **PAGE**

Figure 1.1. The vertical jump test 1
Figure 1.2. Structure of the coordination and skill test 2
Figure 1.3. 30-meter dash test 3
Figure 1.4. 20-meter shuttle run test 4
Figure 2.1. MLP Structure 16
Figure 2.2. The Relief-F algorithm 23
Figure 3.1. Flow chart of the SVM-based model.....29
Figure 3.2. Flow chart of the MLP classification model 31
Figure 3.3. The basic flow K-means 34
Figure 3.4. The flow chart of K-means algorithm. 35
Figure 4.1. Classification accuracies of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case) 45
Figure 4.2. Sensitivities of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case) 45
Figure 4.3. Specificities of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case). 46
Figure 4.4. PPV's of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case) 46
Figure 4.5. NPV's'sof prediction models using Relief-F on the 2006 dataset (10-fold cross validation case) 47
Figure 4.6. Classification accuracies of prediction models using Relief-F on the 2006 dataset (80-20% case) 47
Figure 4.7. Sensitivities of prediction models using Relief-F on the 2006 dataset (80-20% case) 48
Figure 4.8. Specificities of prediction models using Relief-F on the 2006 dataset (80-20% case) 48
Figure 4.9. PPV's of prediction models using Relief-F on the 2006 dataset (80-20% case) 49
Figure 4.10. NPV'sof prediction models using Relief-F on the 2006 dataset

(80-20% case) 49
Figure 4.11. Classification accuracies of prediction models using Relief-F on the 2006 dataset (70-30% case) 50
Figure 4.12. Sensitivities of prediction models using Relief-F on the 2006 dataset (70-30% case) 50
Figure 4.13. Specificities of prediction models using Relief-F on the 2006 dataset (70-30% case) 51
Figure 4.14. PPV's of prediction models using Relief-F on the 2006 dataset (70-30% case) 51
Figure 4.15. NPV'sof prediction models using Relief-F on the 2006 dataset (70-30% case) 52
Figure 4.16. Classification accuracies of prediction models using Relief-F on the 2006 dataset (60-40% case) 52
Figure 4.17. Sensitivities of prediction models using Relief-F on the 2006 dataset (60-40% case) 53
Figure 4.18. Specificities of prediction models using Relief-F on the 2006 dataset (60-40% case) 53
Figure 4.19. PPV's of prediction models using Relief-F on the 2006 dataset (60-40% case) 54
Figure 4.20. NPV'sof prediction models using Relief-F on the 2006 dataset (60-40% case) 54
Figure 4.21. Classification accuracies of prediction models using Relief-F on the 2006 dataset (50-50% case) 55
Figure 4.22. Sensitivities of prediction models using Relief-F on the 2006 dataset (50-50% case) 55
Figure 4.23. Specificities of prediction models using Relief-F on the 2006 dataset (50-50% case) 56
Figure 4.24. PPV's of prediction models using Relief-F on the 2006 dataset (50-50% case) 56
Figure 4.25. NPV'sof prediction models using Relief-F on the 2006 dataset (50-50% case) 57

Figure 4.26. Classification accuracies of prediction models using F-Score on the 2006 dataset (10-fold cross validation case). 57
Figure 4.27. Sensitivities of prediction models using F-Score on the 2006 dataset (10-fold cross validation case) 58
Figure 4.28. Specificities of prediction models using F-Score on the 2006 dataset (10-fold cross validation case) 58
Figure 4.29. PPV's of prediction models using F-Score on the 2006 dataset (10-fold cross validation case) 59
Figure 4.30. NPV'sof prediction models using F-Score on the 2006 dataset (10-fold cross validation case) 59
Figure 4.31. Classification accuracies of prediction models using F-Score on the 2006 dataset (80-20% case). 60
Figure 4.32. Sensitivities of prediction models using F-Score on the 2006 dataset (80-20% case). 60
Figure 4.33. Specificities of prediction models using F-Score on the 2006 dataset (80-20% case) 61
Figure 4.34. PPV's of prediction models using F-Score on the 2006 dataset (80-20% case) 61
Figure 4.35. NPV'sof prediction models using F-Score on the 2006 dataset (80-20% case) 62
Figure 4.36. Classification accuracies of prediction models using F-Score on the 2006 dataset (70-30% case) 62
Figure 4.37. Sensitivities of prediction models using F-Score on the 2006 dataset (70-30% case) 63
Figure 4.38. Specificities of prediction models using F-Score on the 2006 dataset (70-30% case) 63
Figure 4.39. PPV's of prediction models using F-Score on the 2006 dataset (70-30% case) 64
Figure 4.40. NPV'sof prediction models using F-Score on the 2006 dataset (70-30% case) 64

Figure 4.41. Classification accuracies of prediction models using F-Score on the 2006 dataset (60-40% case) 65
Figure 4.42. Sensitivities of prediction models using F-Score on the 2006 dataset (60-40% case) 65
Figure 4.43. Specificities of prediction models using F-Score on the 2006 dataset (60-40% case) 66
Figure 4.44. PPV's of prediction models using F-Score on the 2006 dataset (60-40% case) 66
Figure 4.45. NPV'sof prediction models using F-Score on the 2006 dataset (60-40% case) 67
Figure 4.46. Classification accuracies of prediction models using F-Score on the 2006 dataset (50-50% case) 67
Figure 4.47. Sensitivities of prediction models using F-Score on the 2006 dataset (50-50% case) 68
Figure 4.48. Specificities of prediction models using F-Score on the 2006 dataset (50-50% case) 68
Figure 4.49. PPV's of prediction models using F-Score on the 2006 dataset (50-50% case) 69
Figure 4.50. NPV'sof prediction models using F-Score on the 2006 dataset (50-50% case) 69
Figure 4.51. Classification accuracies of prediction models using CFS on the 2006 dataset 70
Figure 4.52. Sensitivities of prediction models using CFS on the 2006 dataset.....70
Figure 4.53. Specificities of prediction models using CFS on the 2006 dataset.....71
Figure 4.54. PPV's of prediction models using CFS on the 2006 dataset 71
Figure 4.55. NPV'sof prediction models using CFS on the 2006 dataset 72
Figure 4.56. Classification accuracies of prediction models using Relief-F on the 2007 dataset (10-fold cross validation case) 72
Figure 4.57. Sensitivities of prediction models using Relief-F on the 2007 dataset (10-fold cross validation case) 73

Figure 4.58. Specificities of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case) 73
Figure 4.59. PPV's of prediction models using Relief-F on the 2007 dataset (10-fold cross validation case) 74
Figure 4.60. NPV'sof prediction models using Relief-F on the 2007 dataset (10-fold cross validation case) 74
Figure 4.61. Classification accuracies of prediction models using Relief-F on the 2007 dataset (80-20% case) 75
Figure 4.62. Sensitivities of prediction models using Relief-F on the 2007dataset (80-20% case) 75
Figure 4.63. Specificities of prediction models using Relief-F on the 2006 dataset (80-20% case) 76
Figure 4.64. PPV's of prediction models using Relief-F on the 2007 dataset (80-20% case) 76
Figure 4.65. NPV'sof prediction models using Relief-F on the 2007 dataset (80-20% case) 77
Figure 4.66. Classification accuracies of prediction models using Relief-F on the 2007 dataset (70-30% case) 77
Figure 4.67. Sensitivities of prediction models using Relief-F on the 2007dataset (70-30% case) 78
Figure 4.68. Specificities of prediction models using Relief-F on the 2006 dataset (70-30% case) 78
Figure 4.69. PPV's of prediction models using Relief-F on the 2007 dataset (70-30% case) 79
Figure 4.70. NPV'sof prediction models using Relief-F on the 2007 dataset (70-30% case) 79
Figure 4.71. Classification accuracies of prediction models using Relief-F on the 2007 dataset (60-40% case) 80
Figure 4.72. Sensitivities of prediction models using Relief-F on the 2007 dataset (60-40% case) 80

Figure 4.73. Specificities of prediction models using Relief-F on the 2006 dataset (60-40% case) 81
Figure 4.74. PPV's of prediction models using Relief-F on the 2007 dataset (60-40% case). 81
Figure 4.75. NPV'sof prediction models using Relief-F on the 2007 dataset (60-40% case) 82
Figure 4.76. Classification accuracies of prediction models using Relief-F on the 2007 dataset (50-50% case) 82
Figure 4.77. Sensitivities of prediction models using Relief-F on the 2007 dataset (50-50% case) 83
Figure 4.78. Specificities of prediction models using Relief-F on the 2006 dataset (50-50% case) 83
Figure 4.79. PPV's of prediction models using Relief-F on the 2007 dataset (50-50% case) 84
Figure 4.80. NPV'sof prediction models using Relief-F on the 2007 dataset (50-50% case) 84
Figure 4.81. Classification accuracies of prediction models using F-Score on the 2007 dataset (10-fold cross validation case) 85
Figure 4.82. Sensitivities of prediction models using F-Scoreon the 2007dataset (10-fold cross validation case) 85
Figure 4.83. Specificities of prediction models using F-Score on the 2006 dataset (10-fold cross validation case) 86
Figure 4.84. PPV's of prediction models using F-Scoreon the 2007 dataset (10-fold cross validation case) 86
Figure 4.85. NPV'sof prediction models using F-Scoreon the 2007 dataset (10-fold cross validation case) 87
Figure 4.86. Classification accuracies of prediction models using F-Score on the 2007 dataset (80-20% case) 87
Figure 4.87. Sensitivities of prediction models using F-Scoreon the 2007dataset (80-20% case) 88

Figure 4.88. Specificities of prediction models using F-Score on the 2006 dataset (80-20% case) 88
Figure 4.89. PPV's of prediction models using F-Scoreon the 2007 dataset (80-20% case) 89
Figure 4.90. NPV'sof prediction models using F-Scoreon the 2007 dataset (80-20% case) 89
Figure 4.91. Classification accuracies of prediction models using F-Score on the 2007 dataset (70-30% case) 90
Figure 4.92. Sensitivities of prediction models using F-Scoreon the 2007dataset (70-30% case) 90
Figure 4.93. Specificities of prediction models using F-Score on the 2006 dataset (70-30% case) 91
Figure 4.94. PPV's of prediction models using F-Scoreon the 2007 dataset (70-30% case) 91
Figure 4.95. NPV'sof prediction models using F-Scoreon the 2007 dataset (70-30% case) 92
Figure 4.96. Classification accuracies of prediction models using F-Score on the 2007 dataset (60-40% case) 92
Figure 4.97. Sensitivities of prediction models using F-Scoreon the 2007dataset (60-40% case) 93
Figure 4.98. Specificities of prediction models using F-Score on the 2006 dataset (60-40% case) 93
Figure 4.99. PPV's of prediction models using F-Scoreon the 2007 dataset (60-40% case) 94
Figure 4.100. NPV'sof prediction models using F-Scoreon the 2007 dataset (60-40% case) 94
Figure 4.101. Classification accuracies of prediction models using F-Score on the 2007 dataset (50-50% case) 95
Figure 4.102. Sensitivities of prediction models using F-Scoreon the 2007 dataset (50-50% case) 95
Figure 4.103. Specificities of prediction models using F-Score on the 2006

dataset (50-50% case) 96
Figure 4.104. PPV's of prediction models using F-Scoreon the 2007 dataset (50-50% case) 96
Figure 4.105. NPV'sof prediction models using F-Scoreon the 2007 dataset (50-50% case) 97
Figure 4.106. Classification accuracies of prediction models using CFS on the 2007 dataset 97
Figure 4.107. Sensitivities of prediction models using CFSon the 2007dataset 98
Figure 4.108. Specificities of prediction models using CFS on the 2006 dataset 98
Figure 4.109. PPV's of prediction models using CFSon the 2007 dataset 99
Figure 4.110. NPV'sof prediction models using CFSon the 2007 dataset 99
Figure 4.111. Average classification accuracies of prediction models using Relief-F on the 2006 dataset 100
Figure 4.112. Average sensitivities of prediction models using Relief-F on the 2006 dataset 100
Figure 4.113. Average specificities of prediction models using Relief-F on the 2006 dataset. 101
Figure 4.114. Average PPV's of prediction models using Relief-F on the 2006 dataset 101
Figure 4.115. AverageNPV'sof prediction models using Relief-F on the 2006 dataset 102
Figure 4.116. Average classification accuracies of prediction models using F-score the 2006 dataset 102
Figure 4.117. Average sensitivities of prediction models using F-score on the 2006 dataset 103
Figure 4.118. Average specificities of prediction models using F-score on the 2006 dataset 103
Figure 4.119. Average PPV's of prediction models using F-score on the 2006 dataset 104
Figure 4.120. AverageNPV'sof prediction models using F-score on the 2006 dataset 104

Figure 4.121. Average classification accuracies of prediction models using CFS on the 2006 dataset 105
Figure 4.122. Average sensitivities of prediction models using CFS on the 2006 dataset 105
Figure 4.123. Average specificities of prediction models using CFS on the 2006 dataset 106
Figure 4.124. Average PPV's of prediction models using CFS on the 2006 dataset 106
Figure 4.125. AverageNPV'sof prediction models using CFS on the 2006 dataset 107
Figure 4.126. Average classification accuracies of prediction models using Relief-F on the 2007 dataset 107
Figure 4.127. Average sensitivities of prediction models using Relief-F on the 2007 dataset 108
Figure 4.128. Average specificities of prediction models using CFS on the 2007 dataset 108
Figure 4.129. Average PPV's of prediction models using Relief-F on the 2007 dataset 109
Figure 4.130. AverageNPV'sof prediction models using Relief-F on the 2007 dataset 109
Figure 4.131. Average classification accuracies of prediction models using F-score on the 2007 dataset 110
Figure 4.132. Average sensitivities of prediction models using F-score on the 2007 dataset 110
Figure 4.133. Average specificities of prediction models using F-scoreon the 2007 dataset 111
Figure 4.134. Average PPV's of prediction models using F-score on the 2007 dataset 111
Figure 4.135. AverageNPV'sof prediction models using F-score on the 2007 dataset 112
Figure 4.136. Average classification accuracies of prediction models usingCFS

on the 2007 dataset 112
Figure 4.137. Average sensitivities of prediction models using CFS on the 2007 dataset 113
Figure 4.138. Average specificities of prediction models using CFS on the 2007 dataset 113
Figure 4.139. Average PPV's of prediction models using CFS on the 2007 dataset 114
Figure 4.140. AverageNPV'sof prediction models using CFS on the 2007 dataset 114

LIST OF ABBREVIATIONS

CFS	:	Correlation-Based Feature Selection
DTS	:	30-meter Dash Test
GPA	:	Grade Point Average
KMC	:	K-Means Clustering
MLP	:	Multi-Layer Perceptrons
NSPE	:	National Student Placement Exam
NSSE	:	National Student Selection Exam
PATS	:	Physical Ability Test Score
RBF	:	Radial Basis Function
SA	:	Specialization Area at High School
SRTS	:	20-meter Shuttle Run Test Score
SKILL	:	The Coordination and Skill Test
SV	:	Support Vector
SVM	:	Support Vector Machines
TB	:	Tree Boost
VJTS	:	The Vertical Jump Test Score

1. INTRODUCTION

In order to admit a participant to the School of Physical Education and Sports at Cukurova University, the candidate has to be succesful in the physical test applied at the School. There are two parts in the physical ability test. Each of these parts contains two sub tests. The vertical jump test as well as the coordination and skill tests are applied in the first portion of the test. The second portion of the test comprises of the 30-meter dash test and 20-meter shuttle run test (Çukurova Üniversitesi Beden Eğitimi ve Spor Yüksekokulu, 2015). The details of each test is given Section 1.1 through Section 1.4.

1.1. The Vertical Jump Test

In this test, the participant waits for resetting of timing mat with weight equally balanced on both feet. After the mat is arranged, participant jumps vertically to reach the highest point he can and then he steps back on the mat. The score of the vertical jump test is calculated considering the time spent in the air. [Çukurova Üniversitesi Beden Eğitimi ve Spor Yüksekokulu, 2015). Figure 1.1 shows the vertical jump test.

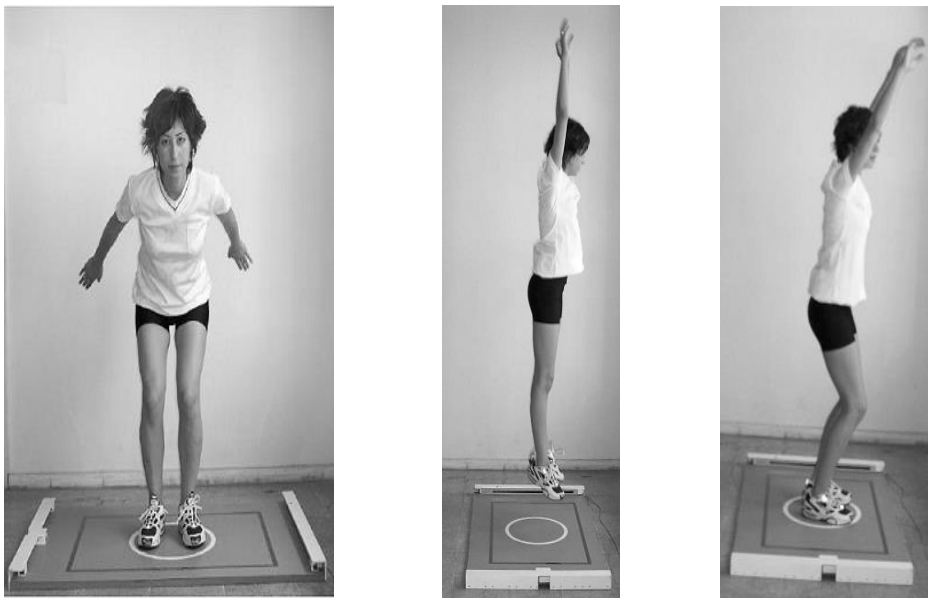


Figure 1.1.The vertical jump test

1.2. The Coordination and Skill Test

The coordination and skill test has several steps. This test starts with a front somersault. The participant gathers a ball, throws the ball up in the air and gets hold of the ball after passing across the horizontal barrier. After this step, he jumps over the balance tool and tries to keep stability while moving. Back somersault follows these steps. The other two steps are related with the participant's movement capabilities. The last part of the this test consists of sliding down and jumping over the obstacles. The setup of the coordination and skill test is shown in Figure 1.2.

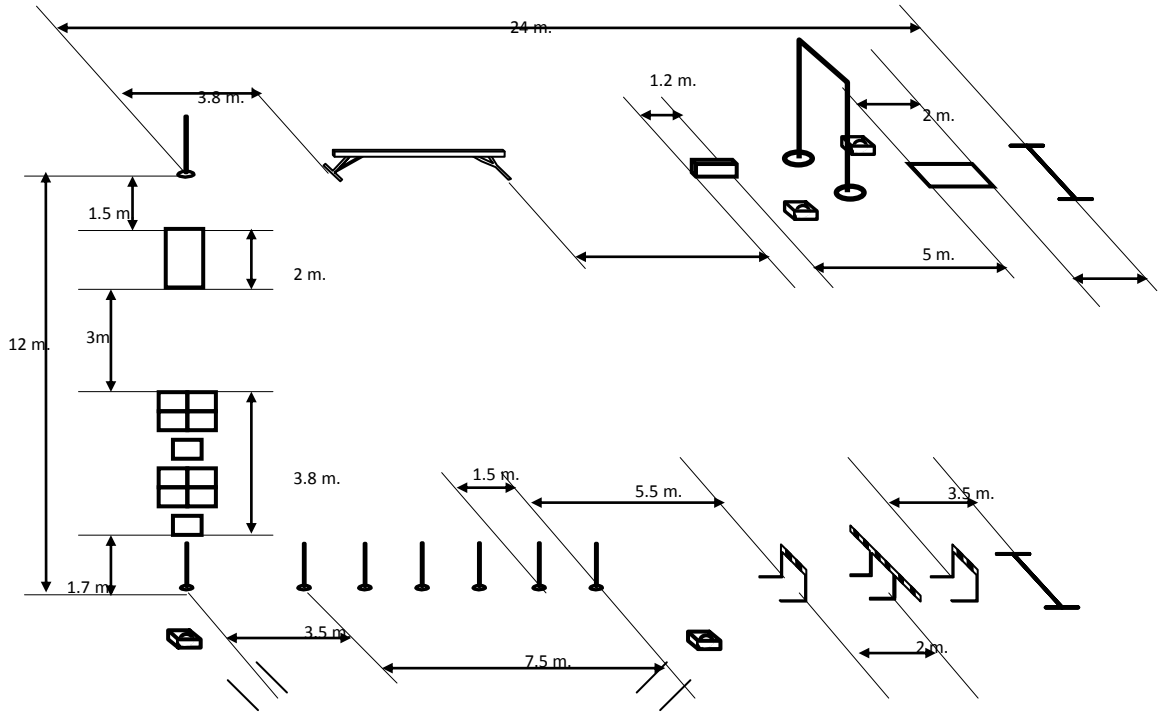


Figure 1.2. Structure of the coordination and skill test.

Each participant has two opportunities for the two tests of the first part. A participant's final score is calculated using the highest scores he achieves. Participants who are successful in these two tests are allowed to take part in the second section of the physical ability test (Çukurova Üniversitesi Beden Eğitimi ve Spor Yüksekokulu, 2015).

1.3. 30-meter Dash Test

30-meter dash test evaluates the ability of participant's to quickly gain speed on 30 meters. The test contains running one maximum sprint over 30 meters. At the beginning of the test, one foot should be one step ahead. The front foot has to be on the starting line. The participant should hold on to this position for two seconds and moving isnot permitted. The instructor should provide information for the acceleration of the participant and motivate to continue running hard through the finish line.

The time starts with the first movement of participant or if there exists time system when it is triggered, and finishes when the participant reaches the finish line or the finishing system is triggered (TopendSports, 2015). Each participant has two trials and the lowest completion time of the participant is recorded. Illustration of the 30-meter dash is given in Figure 1.3.



Figure 1.3. 30-meter dash test

1.4. 20-meter Shuttle Run Test

The shuttle run test is performed at a pitch which has a distance of 20 meters between the two end points. This testisalso called the 'beep' test. The participant stands behind the starting line and starts running when instructed. In the beginnig, the speed of participant is comparatively slow. Running between the

two lines should be continued in accordance with the recorded beeps. As the test progresses, speed of the participant is increased with every beep sound and the frequency of beeps is gradually increased. If the participant arrives the line before the beep sounds, he must hold up until the beep sounds before continuing. If the participant can not arrive before the beep sounds, he is given a warning and must keep on running to the line, then turns and attempts to get up to speed with the pace within two more 'beeps'. The test is completed when participant gets two failures in consecutive (TopendSports, 2015).

Illustration of 20-meter shuttle run test is shown in Figure 1.4.

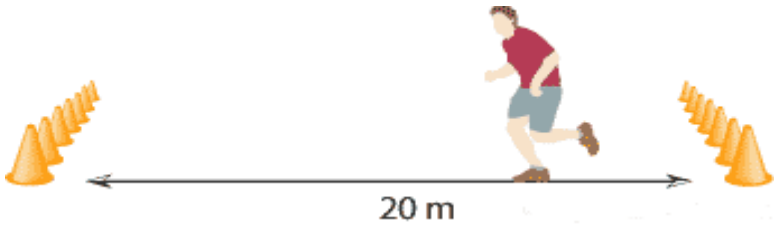


Figure 1.4. 20-meter shuttle run test

1.5. Determining a Participant's Admission Decision

A participant's admission decision is related with the participant's total scores from the physical ability test together with his National Student Selection Examination (NSSE) and National Student Placement Examination (NSPE) scores and Grade Point Average (GPA) at high school. The overall score of a participant who graduated from a sports branch at high school is calculated by Equation (1.1).

$$OVERALL\ SCORE = (PATS) + (0.52\ X\ GPA) + (0.36\ X\ NSPE) \quad (1.1.)$$

The overall score of a participant who graduated from another area at high school is calculated by Equation (2).

$$OVERALL\ SCORE = (PATS) + (0.16\ X\ GPA) + (0.47\ X\ NSPE) \quad (1.2.)$$

In Equation (1) and Equation (2), PATS is the physical ability test score. After the overall scores are calculated for all participants, the scores are sorted in descending order for each participant and a pre-defined number of participants are accepted to the School.

1.6. Previous Work

Developing admission decision prediction models has been an active reserach area for several years. In this regard, there exist a few studies in literature which have attempted to predict the admission decision of a candidate to the School of Physical Education and Sports of Cukurova University by using different machine learning methods.

The first study in this field was carried out by (Acikkar and Akay, 2008). Multilayer Perceptron (MLP) has been used to develop admission decision prediction models. Two datasets consisting of the real test results of participants in the years of 2006 and 2007 have been used. Several performance metrics including classification accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) have been reported for the developed prediction models. The authors concluded that MLP was a feasible tool in this application domain.

In a follow-up work by (Acikkar and Akay, 2009), Support Vector Machines (SVM) based admission decision prediction models have been developed. The same datasets were used to develop the admission decision prediction models. It was shown that SVM-based prediction models perform slightly better than MLP-based prediction models. It was concluded that the SVM classification can be a useful tool for this application area.

Detailed results of the SVM and MLP prediction models on the 2006 and 2007 datasets for (Açıkkar and Akay, 2009; Açıkkar and Akay, 2008) are given in Table 1.1 and Table 1.2, respectively.

Table 1.1 Results of the SVM and MLP prediction models for the 2006 dataset

Study	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV(%)
Açıkkar and Akay, 2009	97.94	97.50	98.00	96.00	99.05
Açıkkar and Akay, 2008	97.17	92.50	99.00	97.50	97.14

Table 1.2 Results of the SVM and MLP prediction models for the 2007 dataset

Study	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Açıkkar and Akay, 2009	93.12	85.00	97.42	94.92	92.99
Açıkkar and Akay, 2008	90.51	85.00	93.46	89.33	93.19

In (Akay et al., 2014) the authors used MLP on the same datasets to develop admission decision prediction models. Each dataset has been split into train and test sets using different ratios. More specifically, the datasets are proportionally grouped in 90-10%, 80-20%, 70-30%, 60-40% and 50-50% portions. For the 2006 dataset, the highest classification accuracy (i.e. 96.20%) has been obtained for the case of 80-20% split whereas for the 2007 dataset, the highest classification accuracy (i.e. 86.34%) has been obtained for the case of 80-20% split.

There are also other studies in literature which developed admission decision prediction models by using different machine learning methods. In (Abut et al., 2015), the authors used four classification methods including SVM, Logistic Regression, RBF Network and K-Means Clustering (KMC) on the 2006 dataset by employing different validation techniques. The results have shown that the performance of SVM employing 10-fold cross validation is superior compared to the performance of other methods. The reported value for the highest classification accuracy was 97.90%.

In (Turhan, 2015), the author used different machine learning algorithms including SVM, MLP, Logistic Regression, RBF Network, Single Decision Tree and KMC on the 2006 and 2007 datasets. Classification accuracy and several

other performance metrics have been used to assess the performance of the machine learning methods on the datasets. SVM using 10-fold cross validation achieves the highest accuracy with 97.90% and 91.45% for the 2006 and 2007 datasets, respectively. The ranking among the six classifiers in terms of achieved classification accuracy has been determined as SVM, Logistic Regression, MLP, RBF Network, Single Decision Tree and KMC.

1.7. Feature Selection

Feature selection is helpful in locating the discriminative features that are the most appropriate to predict the class. Feature selection is used in data mining and statistics. The basic approach of feature selection is to choose a subset of input variables by removing non-relevant features.

In theory, if one knew the full statistical distribution, the use of more features can yield better results. On the other hand, since using a large number of features consume memory and time, doing so may lead the algorithms become wasteful. Therefore, in the preprocessing step, it may be advantageous to pick the relevant and necessary features. Obviously, the favorable circumstances of utilizing feature selection may be enhancing comprehensibility and bringing down cost of data acquisition and handling. As a result of all the advantages, feature selection has interested much consideration inside the machine learning, artificial intelligent and data mining communities (Sun et al., 2011).

1.8. Motivation, Purpose and Contributions of This Thesis

There is only one study (Açıkkar et al., 2014) in literature that uses machine learning methods combined with a feature selection algorithm to develop hybrid admission decision prediction models for the School of Physical Education and Sports at Cukurova University. In (Açıkkar et al., 2014), MLP combined with a feature selection algorithm has been used to develop a prediction model to predict the admission decision. As a feature selection algorithm, Relief-F has been selected. Predictor variables are gender, NSSE and NSPE scores, GPA, area at

high school and scores from coordination and skill test, vertical jump test, 30-meter dash test and 20-meter shuttle run test. The results have shown that the model including all the predictor variables yields the best classification accuracies, independent of which activation function has been used at the output layer. Among the results obtained by using different activation functions, the double layered MLP model using the linear activation function has specialization yielded the best classification accuracy (i.e. 96.50%).

Apparently, more research is required with the help of several different machine learning methods combined with different feature selection algorithms in order to identify the effect of the predictor variables on the admission decision. The purpose of this thesis is to develop new hybrid admission decision prediction models by using different machine learning methods including SVM, MLP, RBF Network, TB and KMC combined with feature selection algorithms to investigate the effect of the predictor variables on the admission decision of a candidate to the School of Physical Education and Sports at Cukurova University. With the help of the prediction models that have been developed in this thesis, the participant can have an idea about the importance of each predictor variable and hence prepare for the physical test using an approprite training program.

In this thesis, two datasets, namely 2006 dataset and 2007 dataset, have been utilized. By using the Relief-F and F-score feature selection algorithms, ranking of the attributes has been calculated. Then, based on these ranking scores, several models have been developed by removing the attribute with the lowest score at a time. In contrast to Relief-F and F-score algorithms, the CFS algorithm gives a set of selected variables to develop a single model. The models have been evaluated using different machine learning methods including SVM, MLP, RBF, TB and KMC. For model testing, 10-fold cross validation and percentage splits of data have been used. The performances of the machine learning methods for two datasets have been assessed utilizing classification accuracy, specificity, sensitivity, PPV and NPV.

This thesis has two main contributions when compared to the studies in literature. First of all, this is the first thesis in literature that develops hybrid admission decision prediction models to the School of Physical Education and

Sports at Cukurova University using several machine learning methods combined with different feature selection algorithms. Secondly, by integrating feature selection algorithms into machine learning methods, this thesis yields discriminating the useful and redundant features for admission decision prediction.

1.9. Overview of Datasets

Two different datasetshave been used in this thesis. The datasets were provided by the School of Physical Education and Sports of Cukurova University. These datasets include data of participants who performed the physical ability tests in the years of 2006 and 2007. They contain nine attributes including gender, the scores from the NSSE and NSPE, GPA, the specialization area at high school, the scores from the vertical jump test, coordination and skill test, 30-meter dash test and 20-meter shuttle run test. The datasets also include two classes. These classes are represented with the values 0 and 1, where 0 means “reject” and 1 means “admit”. There are 143 subjects (87 males and 56 females) and 117 subjects (73 males and 44 females) in the datasets of 2006 and 2007, respectively.

The predictor variables and their statistics for each dataset are given in Table 1.3. and Table 1.4., respectively.

Table 1.3. Statistical analysis of each predictor variable for the 2006 dataset

Predictor Variable	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**
Gender	0	1	0.61	0.49
NSSE	165.34	259.89	205.88	19.48
NSPE	186.78	262.93	226.20	14.90
GPA	37.88	93.46	75.98	10.89
Specialization area	0	1	0.14	0.348
Vertical jump test score	23	59	39.87	8.076
Coordination and skill test score	25.37	46.13	30.67	3.36
30-meter dash test score	3.74	5.32	4.30	0.39

Table 1.4. Statistical analysis of each predictor variable for the 2007 dataset

Predictor Variable	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**
Gender	0	1	0.62	0.49
NSSE	184.21	274.59	217.07	20.69
NSPE	190.46	283.21	237.68	14.69
GPA	37.03	99.38	77.27	10.38
Specialization area	0	1	0.094	0.29
Vertical jump test score	24	65	50	9.39
Coordination and skill test score	25.47	39.73	29.86	3.06
30-meter dash test score	3.74	5.94	4.26	0.40

2. OVERVIEW OF METHODS

2.1. Support Vector Machines

SVM is related to statistical learning theory (Vapnik, 1999), which was introduced in 1992 (Boser, 1992). The SVM solves classification problems by utilizing an adaptable representation of the class limits by executing automatic complexity control. With the help of this feature, the problem of overfitting is reduced.

A. Linear SVM

Assume a training set of *N* data points, $S=\{x_k, y_k\}$ where $x_k \in R^n$ is an input vector and $y_k \in R$ is an output vector. The SVM problems are related with hyperplanes which separate the data. The equation of the hyperplanes are determined by a vector w and a bias b. Decision function for the hyperplanes is $w^t.x+b=0$. The margin of separation ρ can be made maximum by constructing the optimal hyperplane. The support vector method technique aims at building a classifier

$$f(x)=sign(w^t.x+b). \tag{2.1.}$$

The $\mathbf{w}$ and b parameters are limited with

$$\min_i |w.x_i+b| \geq 1. \tag{2.2.}$$

After the vectors are divided without any problem, and as a result of this dividing process, if the space between the nearest vector and the hyperplane is maximum, it is called to be divided by a hyperplane. Consequently, a dividing hyperplane in standard form has to fulfill the constraints given in (2.3),

$$y_i\left(w.x_i+b\right)\geq 1,\ i=1,2,...,n. \tag{2.3.}$$

A point $\mathbf{x}_i$ having the distance d from the hyperplane $\left(\mathbf{w},b\right)$ is,

$$d((w,b),xi)=\frac{y_i(x_i.w+b)}{\|w\|}\geq\frac{1}{\|w\|} \tag{2.4.}$$

ρ can be calculated as

$$\rho=\frac{2}{\|w\|}\ . \tag{2.5.}$$

Hence, SVM searches for a separating hyperplane by minimizing

$$\Phi\left(w\right)=\frac{1}{2}\left(w.w\right). \tag{2.6.}$$

$\Phi\left(w\right)$ in (2.6) can be minimized by performing the structural risk minimization principle,

$$\|w\|^2\leq c. \tag{2.7.}$$

h is the series of standard hyperplanes in space that has *n*-dimension and is limited by,

$$h\leq\min\left[\left(R^2c\right),d\right]+1, \tag{2.8.}$$

in which *R* is a hypersphere's radius surrounding all training vectors. As a result of this, minimizing (2.6) is equal to minimization of the upper bound.

The limitations of (2.3) can be reduced by presenting slack variables $\xi_i \geq 0,\ i=1,2,...,n$, therefore (2.3) can be rewritten as

$$y_i.\left(w.x_i+b\right)\geq 1-\xi_i,\ i=1,2,...,n. \tag{2.9.}$$

Under these circumstances, the problem of optimization becomes

$$\Phi\left(w,\xi\right)=\frac{1}{2}\left(w.w\right)+C\sum_{i=1}^{n}\xi_i \quad . \tag{2.10.}$$

In (2.10.) *C* is a user specified positive fixed constant. The saddle point of Lagrangian function is utilized in the solution of the problem given in (2.10).

$$L\left(w,b,\alpha,\xi,\gamma\right)=\frac{1}{2}\left(w.w\right)+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\alpha_i\left|y_i\left(w.x_i+b\right)-1+\xi_i\right|-\sum_{i=1}^{n}\gamma_i\xi_i \tag{2.11.}$$

In (2.11), $\alpha_i \geq 0,\ \xi_i \geq 0, i=1,2,...,n$ are Lagrange multipliers. (2.11) must be solved in terms of $\mathbf{w}$, b, and ξ_i. Classical Lagrangian duality empowers the first issue, turning (2.11) into a dual problem of it and this makes the solution easier. (2.12) shows the dual problem to be solved

$$\max_{\alpha}\left[\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j\left(x_i.x_j\right)\right] \tag{2.12.}$$

with constraints

$$\sum_{i=1}^{n}\alpha_i y_i=0,\ 0\leq\alpha_i\leq C,\ i=1,2,...,n. \tag{2.13.}$$

There is a unique solution for this classic quadratic optimization problem. In respect of the Kuhn-Tucker theorem of optimization theory (2.16.),

$$\alpha_i \left[y_i \left(w.x_i + b \right) - 1 \right] = 0, \; i = 1, 2, ..., n. \qquad (2.14.)$$

If $\mathbf{x}_i$ satisfy (2.15), then (2.14) will have non-zero Lagrange multipliers

$$y_i \left(\mathbf{w}.\mathbf{x}_i + b \right) = 1. \qquad (2.15.)$$

The points in (2.15.) are called support vectors (SV's). Small subset of the training vectors of the SV's determines the hyperplane. Therefore, if the optimal solution α_i^* does not take a value of zero, the classifier function can be represented as

$$f\left(x\right) = \text{sgn}\left\{ \sum_{i=1}^{n} \alpha_i^* y_i \left(x_i . x \right) + b^* \right\} . \qquad (2.16.)$$

In (2.16.) b^* is the solution of (14) for any non-zero α_i^* .

B. Non-linear SVM

The majority of the datasets can not be decently divided by a linear separating hyperplane. However, they can be linearly divided if mapped into a higher dimensional field by utilizing a nonlinear mapping. Therefore, $z = \phi(x)$ that converts the input vector x having a dimension d into a vector z having a dimension d' is defined and $\phi()$ is selected so that $\{\phi(x_i, y_i)\}$ (new training data) is divisible with a hyperplane.

The data points from the input space into some space of higher dimension are mapped by using the function

$$\varphi(.): R^n \to R^{nh}. \tag{2.17.}$$

Optimal function (2.9) transforms (2.18) using the same constraints,

$$\max_{\alpha}\left[\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j)\right] \tag{2.18.}$$

where

$$K(\mathbf{x}_i,\mathbf{x}_j) = \{\varphi(\mathbf{x}_i).\varphi(\mathbf{x}_j)\} \tag{2.19.}$$

is the kernel function.

The exponential kernel function is given by (2.20)

$$K(\mathbf{x}_i,\mathbf{x}_j) = \exp\left\{-\frac{|\mathbf{x}_i-\mathbf{x}_j|^2}{2\sigma^2}\right\}, \tag{2.20.}$$

whereas the polynomial kernel function is given by (2.21)

$$K(\mathbf{x}_i,\mathbf{x}_j) = (\mathbf{x}_i.\mathbf{x}_j+1)^q,\ q=1,2,... \tag{2.21.}$$

where the parameters σ and q in (2.20) and (2.21) have to be prearranged.

Then, the classifier function is specified as

$$f(\mathbf{x}) = \operatorname{sgn}\left\{\sum_{i=1}^{n}\alpha_i^* y_i K(\mathbf{x}_i,\mathbf{x}) + b^*\right\} \tag{2.22.}$$

and b^* is the solution of (2.22) for any non-zero α_i^*.

2.2. Multi-Layer Perceptron

An MLP is a type of artificial neural network model and also feed-forward process. This model maps a set of input data over a set of convenient outputs. An MLP includes multiple layers of nodes in a coordinated graph and each layer is completely connected to the following one. Each node is a neuron with a nonlinear activation function except the input nodes. This method uses a handled learning technique. This technique is named as back propagation in order to train the network. MLP is an alteration of the standard linear perceptron and can recognize inseparable data (Imrie and Durucan, 2000). An MLP has a form as given Figure 2.1.

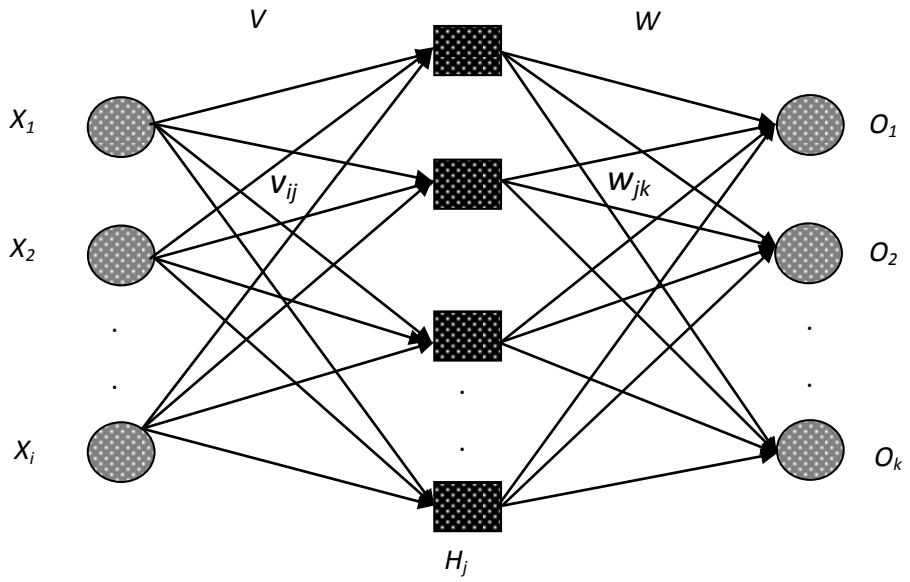


Figure 2.1. A typical MLP Structure

A data model consisting of the x_i values used in the input layer i^{th} is transmitted through the network towards *j* in the first hidden layer. The weighted outputs $w_{ji}x_i$ are received of the previous layer's units per hidden unit. The

outputs from this process are summed. After that, these values are to be turned into an output value using an activation function.

Activation function is $f^{(n)}(x)$ at layer n. The output unit of has two-layer.

$$out_k^{(2)} = f^{(2)}(\sum_J out_j^{(1)}.w_{jk}^{(2)}) = f^{(2)}(\sum_J f^{(1)}(\sum_I in_i w_{ij}^{(1)}).w_{jk}^{(2)}) \quad (2.23.)$$

If the activations at hidden layer are linear, (2.23.) is reduced to

$$out_k^2 = f^{(2)}(\sum_J out_j^{(1)}.w_{jk}^{(2)}) = f^{(2)}(\sum_J f^{(1)}(\sum_I in_i w_{ij}^{(2)}) . \quad (2.24.)$$

Nevertheless, (2.24) is equal to a network having one layer with weight $w_{ik} = \sum_j w_{ij}^{(1)} w_{jk}^{(2)}$. This network can not be used on non-linearly separable problems.

A. Non-Linear Activation/Transfer Function

The values of the logistic sigmoid function range from 0 to 1. The standard sigmoid is basically used with the hyperbolic tangent. It has the feature,

$$f(x) = \tanh(x) = 2Sigmoid(2x) - 1 \; f(-x) = -f(x), \quad (2.25.)$$

and its derivative is given by (2.26.)

$$f'(x) = 1 - f(x)^2 . \quad (2.26.)$$

B. Learning

The same steps are used for training $N-layer$ neural networks as the networks having a single layer. The network weights $w_{ij}^{(n)}$ are set to make the output cost function given in (2.27) minimum,

$$E_{SSE} = \frac{1}{2}\sum_{p}\sum_{J}(t\arg_j^p - out_j^{(N)p})^2, \qquad (2.27.)$$

or

$$E_{CE} = -\sum_{p}\sum_{J}[t\arg_j^p .\log(out_j^{(n)p}) + (1-t\arg_j^p).\log(1-out_j^{(N)p})]$$

(2.28.)

and once again this can be done by a series of gradient descent weight changes

$$\Delta w_{kl}^{(m)} = -\eta\frac{\partial E(\{w_{ij}^{(n)}\})}{\partial w_{kl}^{(m)}} \quad . \qquad (2.29.)$$

$out_j^{(N)}$ is the only output of the last layer. This becomes apparent in the error function E of the output. However, outputs of the final layer are related with all the weights of the previous layers. This learning algorithm automatically sets $out_j^{(n)}$ of the previous layers.

C. Training

Training for multi-layer networks performs in same way with networks having a single layer:

1. Assume that we have the series of training patterns given below in (2.30)

$$\{in_i^p, out_j^p : i = 1...ninputs, j = 1...noutputs, p = 1...npatterns\} \quad (2.30.)$$

2. Network is set up with N inputs, $N-1$ hidden layers of $nhidden^{(n)}$ nonlinear units, and N outputsin layer N. Each layer is completely connected with the previous layer by weight $w_{ij}^{(n)}$.
3. First weights are generated randomly in the range $[-smwt, +smwt]$.
4. The learning rate η and a convenient error function $E(w_{jk}^{(n)})$ are selected.
5. For each training pattern p, the weight update equation $\Delta w_{jk}^{(n)} = -\eta \partial E(w_{jk}^{(n)}) / \partial w_{jk}^{(n)}$ is applied to each weight $w_{jk}^{(n)}$.
6. Step 5 is repeated until the network's error function gives negligible error rates.

Algebraical statements need to be reproduced for the weight updates in order to be practical.

2.3. Radial Basis Function Network

RBF's appeared as a variant of artificial neural network in the last years of 1980. Nonetheless, their roots are settled in much more pattern recognition techniques. These techniques can be clustering, mixture models, potential functions, spline interpolation and functional approximation (Park and Sandberg, 1991).

RBF network includes several layers and the first layer has input neurons. These neurons feed the feature vectors into the network. The second one is the hidden layer that calculates the result of the main functions. The last layer is the output layer that calculates a linear combination of the main functions. These kinds of networks have the general estimation property (Park and Sanberg, 1991). Simple structures of these networks give a decreasing for training time and make possible learning in stages.

RBF has the feed-forward structure. It has one layer that consists of hidden units. These units are completely adjusted with the output units. (2.31) indicates the output units (ψ_j) from a linear combination of the main functions (Ghosh-Dastidar et al., 2008).

$$\psi_j(x) = K\left(\frac{\left\| x - c_j \right\|}{\sigma_j^2}\right) \tag{2.31.}$$

$f : R^n \to R^1$ (a function) is estimated by using an RBF network. $x \in R^n$ (a vector) is an input, $\psi(x, c_j, \sigma_j)$ is the j^{th} function with centre $c_j \in R^n$, width σ_j, and $w = (w_1, w_2, ..., w_m) \in R^M$ is the vector of linear output weights. M represents the main number function. The M centres $c_j \in R^n$ are connected to obtain $c = (c_1, c_2, ..., c_m) \in R^{nM}$. Lastly, the widths are connected to obtain $\sigma = (\sigma_1, \sigma_2, ..., \sigma_M) \in R^M$. The output of the network for $x \in R^n$ and $\sigma \in R^M$ is shown by (2.32).

$$F(x, c, \sigma, \omega) = \sum_{j=1}^{M} (\omega_j \psi(x, c_j, \sigma_j)) . \tag{2.32.}$$

Suppose that $y = (y_1, y_2, ..., y_n)$ is the weighted output vector and $(x_i, y_i) : i = 1, 2, ..., N$ is a series of training pairs. For each $c \in R^{nM}$, $\omega \in R^M, \sigma \in R^M$ and for random weights $\lambda_i, i = 1, 2, ..., N$, which are taken as positive numbers to point out importance of definite domains of the input space, set (2.33) (Wright et al., 2013).

$$E(c,\sigma,\omega)=\frac{1}{2}\sum_{i=1}^{N}\left[\lambda_i(y_i-F(x_i,c,\sigma,\omega))\right]^2 \quad (2.33.)$$

2.4. Tree Boost

Tree Boost is a machine learning method used on classification and regression problems, which creates an estimation model that has the shape of a group of low estimation models. Boosting Tree algorithm is calculated by using of boosting methods to regression trees. The common idea is to evaluate an order of simple trees, where every progressive tree is constructed for the estimation leftover of the former tree. This method will construct binary trees. It is assumed that the complexities of the trees are limited with only 3 nodes. These are two child nodes and a root node. Therefore, at each step of the boosting, a basic dividing of the data is specified, and the deviations of the noted values from the individual means are calculated. The following 3-node tree will be fitted to those residuals, in order to find alternate partition that will further decrease the residual variance for the data, given the former a group of trees (Hastie et al., 2009).
A boosted classifier (or regressor) has the form

$$H(x)=\sum_t \alpha_t h_t(x) \,.$$

(2.34.)

This form can be trained by greedily minimizing a loss function L; i.e. by optimizing scalar α_t and weak learner $h_t(x)$ at each iteration *t*. Before training begins, each data sample x_i is assigned a non-negative weight w_i (which is derived from L). After each iteration, misclassified samples are weighted more heavily thereby increasing the intensity of misclassifying them in next iterations. Regardless of the type of Boosting used (i.e. AdaBoost, LogitBoost, L2Boost, etc.), each iteration requires training a new weak learner given the sample weights (Appel et al., 2013).

2.5. K-Means Clustering

K-means method uses the idea of the division of data. Nevertheless, K-means requires a user to give off the number of clusters for beginning and sensibility to initial conditions. It keeps away from the basing on any beginning conditions and performs better, but it has a highly computational load (Xie et al., 2013).

The aim of this algorithm is to minimize an error function, given by (2.35)

$$J(V)=\sum_{i=1}^{c}\sum_{j=1}^{c_i}(\|x_i-v_j\|)^2 \qquad (2.35.)$$

where,

the Euclidean distance between x_i and v_j is represented by $\|x_i-v_j\|$,

the number of data points in i^{th} cluster is denoted by c_i ,

and finallythe number of cluster centers is designated by c .

K-means Clustering Algorithm

Assume that $X=\{x_1,x_2,x_3,...,x_n\}$ is a series of data points and $V=\{v_1,v_2,v_3,...v_n\}$ is a series of centers.

1. Let $'c'$ (center of cluster) be selected randomly.
2. Computation of the distance between each data point and cluster centers is carried out.
3. The minimum of all the cluster centers is given to the data point to the cluster center.
4. The new cluster center is recalculated using (2.36),

$$v_i=(1/c_i)\sum_{J=1}^{c_i}x_i \, . \qquad (2.36.)$$

In (37), 'c_i' designates the number of data points in i^{th} cluster.

5. The distance between new acquired cluster centers and each data point is recalculated.
6. If there was no data point to reallocate then stop, else the process is repeated from step 3.

2.6. Feature Selection Algorithms

A. Relief-F Algorithm

Relief-F is an instance-based algorithm: Instances are tried randomly and checked whether to have neighborhood of the same or different classes by this algorithm. The algorithm works on continuous and discrete class data. The parameters determine the number of neighbors to check as well as the number of instances to sample. (Witten et al., 2011).

The Relief-F algorithm is given in Figure 2.2.

Input: Gene variables and labels
Output : W for the gene rank
Set all weights $W := 0$;
foreach *Iteration n* do
  Randomly select an instance Ins_m :
  Find K nearest hits H;
  foreach *class c* ? $Label_m$ **do**
    from class c find K nearest misses M_c ;
  end
  foreach g_i **do**
    Update W_i ;
  end
end

Figure 2.2. The Relief-F Algorithm

B. F-score Algorithm

This algorithm is a main method that assesses the differences between two classes having real values (Wright et al., 2013). Each feature in dataset has a score value named F-score. These F-scores values are computed and then a threshold value is included by computing the average value of F-scores of all features. By the helping of threshold value, the features are selected from the entire dataset.

(2.37) is used to calculate the F-scores of every feature. If the threshold value is smaller than the F-score of any feature, that feature is added to the feature area (Jaganathan et al., 2012). Else, the feature is eliminated from the feature area. The F-score of the i^{th} feature is explained by (2.37) (Chen et al., 2003).

$$F(i)=\frac{(\bar{x}_i^{(+)}-\bar{x}_i)^2+(\bar{x}_i^{(-)}-\bar{x}_i)^2}{\frac{1}{n_+-1}\sum_{k=1}^{n_+}(x_{k,i}^{(+)}-\bar{x}_i^{(+)})^2+\frac{1}{n_--1}\sum_{k=1}^{n_-}(x_{k,i}^{(-)}-\bar{x}_i^{(-)})^2} \quad (2.37.)$$

In (2.37), $\bar{x}_i$ denotes the average of the i^{th} feature of the entire instances. The same of the positive and negative instances are given by given by $\bar{x}_i^{(+)}$ and $\bar{x}_i^{(-)}$, respectively. $x_{k,i}^{(+)}$ is the i^{th} feature of the k^{th} positive instance, whereas the same of the negative instance is given by $x_{k,i}^{(-)}$. x_k, $k=1,...,m$, are training vectors, n_+ and n_- are the number of positive and negative instances, respectively. The discrimination between the negative and positive series is demonstrated by the numerator, and the denominator describes the one within each of the positive and negative series. The feature having a bigger F-score is said to be more discriminative.

C. Correlation Feature Selection

In this algorithm, a subset of features is selected depending on a heuristic given by (2.38),

$$Merit_s = \frac{k\overline{r_{cf}}}{\sqrt{k+k(k-1)\overline{r_{ff}}}}. \qquad (2.38.)$$

In (2.38), $Merit_s$ shows how much value a subset containing k features has. The mean feature class correlation is designated by $\overline{r_{cf}}$ whereas the average feature-feature intercorrelation is given by $\overline{r_{ff}}$. The numerator gives a sign of the class a group of features and determines the redundance among them is determined by the denominator. The heuristic processes non-relevant features since they are weak estimators of the class. Excess attributes are differentiated as they are the most related with one feature (Hall et al., 1997).

3. DEVELOPMENT OF PREDICTION MODELS

In this thesis, by using the Relief-F and F-score feature selection algorithms, ranking of the attributes has been calculated. Then, based on these ranking scores, several models have been developed by removing the attribute with the lowest score at a time. In contrast to Relief-F and F-score algorithms, the CFS algorithm gives a set of selected variables to develop a single model. The models have been evaluated using different classification methods including SVM, MLP, RBF, TB and KMC. For model testing, 10-fold cross validation and different percentage splits of data have been used. Numerous metrics including classification accuracy, specificity, sensitivity, positive predictive value and negative predictive value have been used to assess the performance of the classifiers for the two datasets.

3.1. SVM Model for Prediction of Admission Decision

The value of accuracy of a SVM classification model depends on the value of C, type of the kernel function and parameters of this function.There is a tradeoff between reducingthe error related to training and reducing complexity of model and this is stated by parameter C. A small value of C will increase the quantity of training errorsand a large value of C will exhibit anattitude similar to that of a hard-margin SVM. The kernel parameters are importantin the sense that they act as a bridge between the input space and the high-dimensional feature space (Ji and Wang, 2007). In this thesis, the RBF kernel requiring the optimization of the parameter γ is selected for creatingthe SVM classification models.

There is no way to determine in advance which C and γ are the ideal for a classification problem. Hence some kind of parameter search has to be manipulated. The aim is to achieveoptimal values of the parameters (C and γ) so that the testing data can be estimated with minimum error by the regression model. There exist many methods that can be used fordeciding the values of the optimal parameters. The grid search is the most effective one for achieving the

optimum values of C and γ on median-sized problems. In this method, the parameters are differentiated by constant step-sizes through aninterval of values, and the success of each value of these parameters is evaluated. A cross validation within the grid search can be utilizedorder to develop the generalization capability of the SVM classification model. In the process of k-fold cross validation, the cross validation is recurred k times (this means k folds), with each of the k subsets utilizedprecisely once as the validation data. The k results from the folds then ought to be associated to generate a single prediction.

For building the prediction models, grid search employing 5-fold cross validation has been utilized to indicate the optimal values of C and γ. The flow chart of the SVM classification model for prediction of admission decision is shown in Figure 3.1. In the beginning, by using the Relief-F attribute evaluator combined with ranker search method, the score of each attribute has been calculated and the attributes have been sorted in descending order according to their scores. Dimensionality of the data set is reduced by one feature at a time (the attribute with the lowest score is removed) before being passed on to the SVM. The test and train subsets are preprocessed to yield a mean of zero and a variance of unity. This process allowedcreating new train and test subsets. Usingall values of (C, γ), the new training subset is processed using 5-fold cross validation, which in turnyield the computation of root mean squared errors for each prediction. The (C, γ) values that give the most minimal overall root mean squared erroris selected for training the train subset, which in turn generatedthe prediction mode. Finally, admission decision in the new test subset is predicted using the developedSVM prediction model.

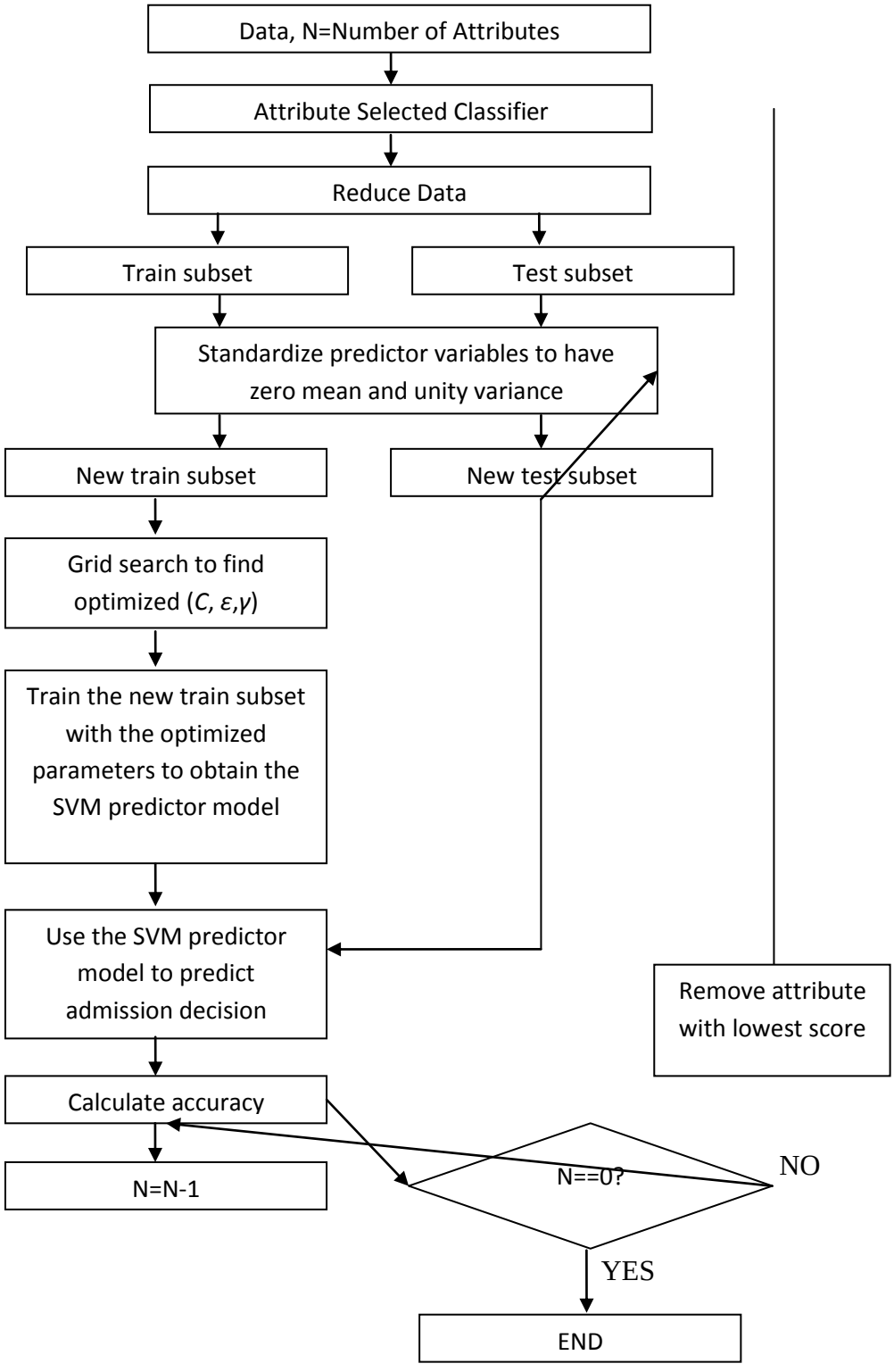


Figure 3.1. Flow chart of SVM-based model

The intervals for values of the parameters for the SVM models are given in Table 3.1.

Table 3.1. The intervals for values of the parameters for the SVM model

Parameter	**Value**
Cost (*C*)	[1, 100]
Gamma (γ)	[0.001- 50]
Kernel Function	RBF

3.2. MLP Model for Prediction of Admission Decision

Back-propagation is utilized to create MLP models for prediction of admission decision. The success of an MLP is related with the number of hidden layers and neuron numbers in each hidden layer. If the number of neuron that is utilized is not sufficient, less information will be acquired. On the contrary, the local minimum might enhance and the network mightcome close to a local minimum,hence the network's sensitive will decrease. Nevertheless, there is not a regulationfordetecting the number of neurons in a hidden layer. Generally, the optimal number is selected with try and error based on the difficulty of the problem.

In the hidden layer of the MLP models, the tansigmoid function is utilized, whereas in the output layer of MLP models, the pure-linear function is used. Levenberg-Marquardt algorithm is implementedto trainthe network.Other important parameters of the MLP classification models and their values are given in Table 3.2. Figure 3.2 shows the steps of utilizinga MLP for classification of prediction of admission decision.

Table 3.2 The intervals for values of the parameters for MLP models

Parameter	**Value**
Number of neurons in the hidden	[2-20]
Learning rate	(0-1)
Momentum	(0-1)
Maximum iteration	10000
Number of epochs	500

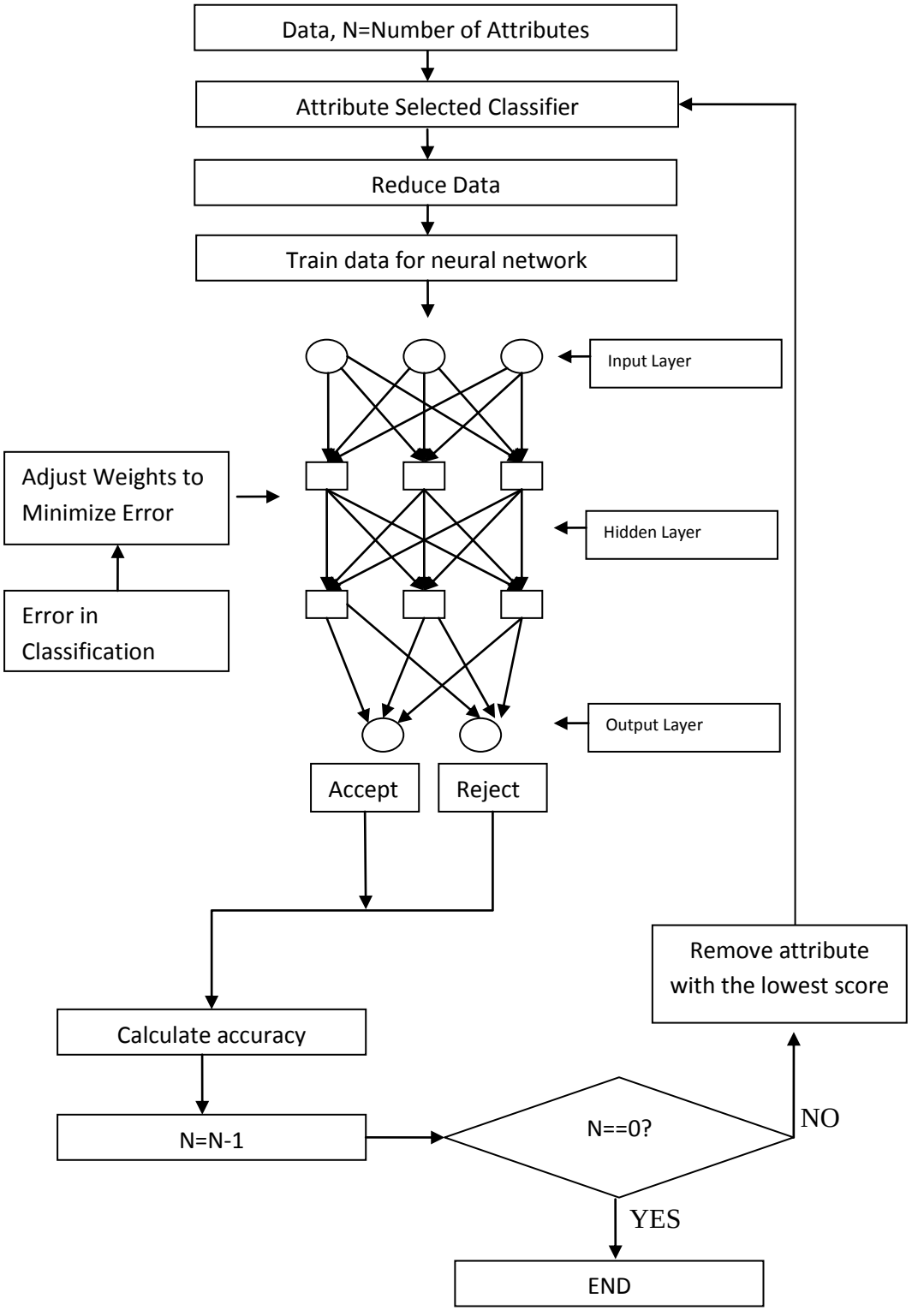


Figure 3.2. Flow chart of the MLP classification model

3.3. RBF Network Model for Prediction of Admission Decision

The success of RBF network relies on numerous factors. The neuron number of the hidden layer has to be specified before the parameter selection for the RBF network. After the neuron number of the hidden layer is selected, the success of the RBF depends on the choosing of the network parameters. The efficiency of training an RBF network is subjected toa form of functions, their quantity and locations, and the method utilized for learning the input-output mapping.The most commonlearning strategies (Nicolaos and Karayiannis, 2003) for RBF networks could be categorized as follows:

- The first one is selecting the RBF centers randomly from the training data.
- The second one is usinguncontrolledtechniques for selecting the RBF centers.
- The last one is usingcontrolledtechniques for selecting the RBF centers.

Some approaches have been presented, which include orthogonal least squares (Gomm and Yu, 2000), resource allocating in genetic algorithm in gradient decent. These approaches involve searching for a sub-optimal network in alower dimensional space. However, one common feature of the above methods isthey often lead to over-fitting, with a negative influence on the success of trained RBF neural networks. The biggest obstacle is iterative strategy that oftencannot reach an optimal result. To overcome these computational difficulties, thecomplexity of the network must be decreased, which needs an approach to the arranged network. The design of a controlled neuralnetwork may be maintained in a diversity of ways.

Other important parameters of the RBF classification models and their values are given in Table 3.3.

Table 3.3. Important parameters of the RBF classification models

Parameter	**Value**
Regularization parameter (λ)	[0.001 - 10]
Population size	200
Radius of the RBF Network	[0.001 - 400]
Maximal number of neurons	8

3.4. TB Model for Prediction Admission of Decision

TreeBoost methodsusually have a degree of accuracy that cannot be acquiredutilizing a large, single-tree model. TreeBoost models are often equal to or superior to any other predictive functions obtaining neural networks.TreeBoost models can process hundreds of possible predictor variables. Unrelated predictor variables are identified automatically and have no effect on the predictive model.TreeBoost utilizes the Huber M-regression loss function which makes it highly deprecating to outliers and misclassified conditions.The sophisticated and accurate method of alternate splitters is utilized for processing missing predictor values.The randomizationcomponent in this algorithm makes it exceptionallyimpervious to over fitting.Cross-validation and random-row-sampling methods can be utilized to assess the generalization of a TreeBoost model and make preparations over fitting.

Table 3.4. shows the important parameters and their values of the TB classification models.

Table 3.4. Important parameters and their values of the TB classification models

Parameter	**Value**
Shrinkage Factor	Automatic
Huber’s Quantile Cutoff	(0-1)
Influence Trimming Factor	(0-1)
Depth of Individual Tress	[1-5]

3.5. KMC Model for Prediction of Admission Decision

This algorithm is anexceptionally prevalent algorithm for uncontrolled clustering of feature vector data. It is named k-Means on the grounds thatthere are k clusters, and the average of the cluster substance decides the cluster centroid. This algorithm lives up to expectations generally well and it is extremely straightforward both to comprehend and to execute. The basic way to deal with this algorithm is in viewof centroids symbolizing the average of the current arrangement of feature vectors included within the cluster. The centroid is recalculated whenever an object gets into or out of a cluster. Picking the cluster for a given feature vector is taking into account the Euclidean distance of the feature vector and the accessible cluster centroids. The nearer the Euclidean distance, the higher the likelihood the feature vector belongs to the cluster. Figure 3.3 and figure 3.4 show the basic flow and the flowchart of the algorithm, sequentially.

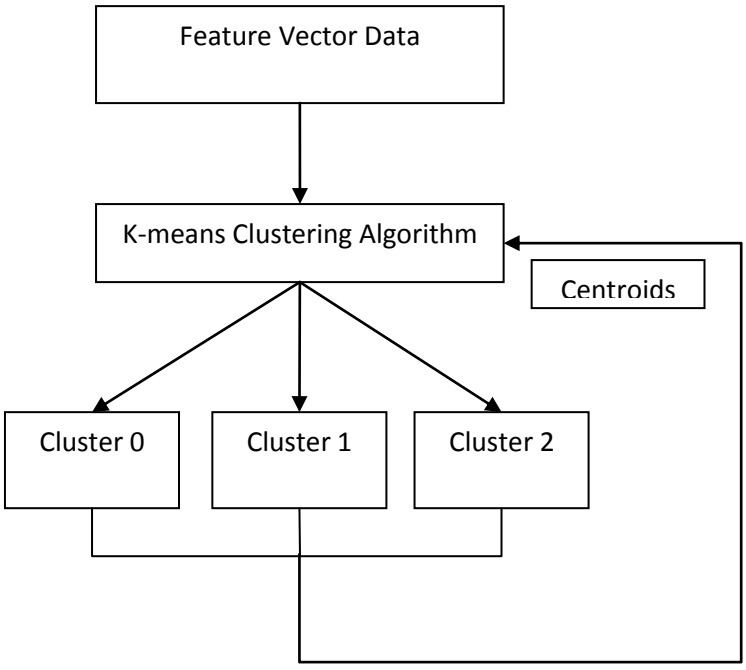


Figure 3.3. The basic flow K-means

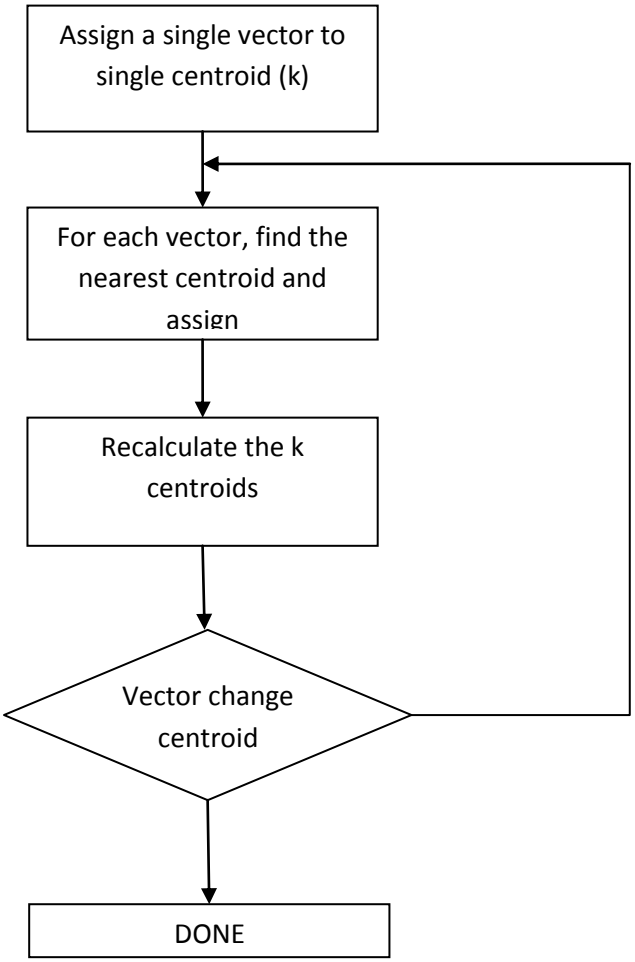


Figure 3.4. The flow chart of K-means algorithm

The number of clusters, k, shouldfurther be described a priori. Every cluster has arelated centroid feature vector which symbolizes the average of the prototype feature vectors included within the cluster. After that, take k of the prototype feature vectors and allocate one each to the accessible clusters. These k prototype feature vectors coul be mixed random, however in a perfect world, we would pick vectors that are disparate. The least complex methodologyis to select the first k feature vectors and allocate them to the clusters. With the clusters now including one prototype feature vector, we instate the centroid. Because the cluster includes only one prototype feature vector, the centroid is equavalent to the prototype feature vector.

The execution of a clustering algorithm may be influenced by the picked value of K. Thusly, as opposed to utilizing a single predefined K, a series of

values may be received. It is imperative for the quantity ofvalues thought to be sensibly expansive, to reflectthe particularattributes of the data sets. In the meantime, the selected values must be essentially smaller than the quantity of objects in thedata sets, which is the fundamental motivation for performing data clustering (Su, 2003).

Also, there have been two points in developing a K-Means clustering model in this study. The first one is specifying the optimal number of clusters to create and the other one is determining the center of each cluster. The value of these parameters is given in Table 3.5.

Table 3.5. Important parameters and their values of the K-means classification models

Parameter	**Value**
Range for the number of clusters	[2 - 200]
Search step	1

3.6. Performance Metrics

A confusion matrix includes information about real and estimated classifications done by a classification system. Execution of such frameworks is generallyassessedutilizing the data in the matrix, which is shown in Fig. 3.2.

Table 3.6. Confusion Matrix

		Predicted	
		Negative	Positive
Actual	Negative	**a**	**b**
	Positive	**c**	**d**

In Table 3.6.,

- *a*represents the number of correctly identificationof a negative instance.

- *b*represents the number of incorrectly identification ofa positive instance.
- *c*represents the number of incorrectly identification ofa negative instance.
- *d*represents the number of correctly identificationof a positive instance.

The accuracy of a measurement system representsnearless of measurements of a number to that number's actual (true) value. The accuracy (AC) is the ratio of the total number of predictions that were accurate. AC is identifiedutilizing(1):

$$AC=\frac{a+d}{a+b+c+d} \tag{3.1.}$$

The recall or true positive rate (TP) is the ratio of positive events that were accurately identified, as computedutilizing(2):

$$TP=\frac{d}{c+d} \tag{3.2.}$$

The false positive rate (FP) is the ratio of negatives events that were inaccurately classified as positive, as computedutilizing(3):

$$FP=\frac{b}{a+b} \tag{3.3.}$$

The true negative rate (TN) is defined as the ratio of negatives events that were classified accurately, as computedutilizing(4):

$$TN=\frac{a}{a+b} \tag{3.4.}$$

The false negative rate (FN) is the ratio of positives events that were inaccurately classified as negative, as computedutilizing(5):

$$FN = \frac{c}{c+d} \tag{3.5.}$$

Lastly, precision (P) is the ratio of the estimated positive events that were correct, as computedutilizing(6):

$$P = \frac{d}{b+d} \tag{3.6.}$$

Sensitivity (also called the true positive rate, or the recall rate) evaluates the ratio of actual positives which are accurately identified and is supplementary to the false negative rate. Sensitivity is computed utilizing (7):

$$Sensitivity = \frac{TP}{TP+FN} \tag{3.7.}$$

Specificity (also called the true negative rate) evaluatesthe ratioof negatives which are accurately identified and is supplementaryto the false positive rate. Specifity is computed utilizing (8):

$$Specifity = \frac{TN}{FP+TN} \tag{3.8.}$$

The positive and negative predictive values (PPV and NPV,sequantially) are the ratios of positive and negative results in statistics and identification tests that are true positive and true negative results (Fletcher and Susanne, 2005). The PPV and NPV determine the success of a identification test or other statistical measure. The PPV and NPV are computed utilizing (9) and (10):

$$PPV \frac{TP}{TP+FP} \tag{3.9.}$$

$$NPV = \frac{TN}{TN + FN} \tag{3.10.}$$

4. RESULTS AND DISCUSSION

In order to build the models for predicting the admission decision, three different feature selection algorithms including Relief-F, F-Score and CFS have been used. By using the Relief-F and F-score feature selection algorithms combined with a ranker search, the score of each attribute has been calculated and the attributes have been sorted in descending order according to their scores. Table 4.1 through Table 4.4 show the Relief-F scores and F-scores of each attribute for the dataset 2006 and 2007, respectively.

Table 4.1. Relief-F scores of the attributes for the 2006 dataset.

Variables	**Score**
30-meter dash test score	0.01755
NSPE	0.01212
Vertical jump test score	0.00686
NSSE	0.00523
Coordination and skill test score	0.00325
Gender	0.0
20-meter shuttle run test score	-0.00282
Specialization Area	-0.01099
GPA	-0.01509

Table 4.2. Relief-F scores of the attributes for the 2007 dataset.

Variables	**Score**
30-meter dash test score	0.02036
20-meter shuttle run test score	0.00858
NSPE	0.00735
GPA	0.00666
NSSE	0.00507
Vertical jump test score	0.0
Gender	-0.00164
Coordination and skill test score	-0.00466
Specialization Area	-0.0211

Table 4.3. F-scores of the attributes for the 2006 dataset.

Variables	**Score**
20-meter shuttle run test score	0.00405
Vertical jump test score	0.00369
Specialization Area	0.00257
30-meter dash test score	0.00118
NSPE	0.00097
Coordination and skill test score	0.00072
Gender	0.00065
GPA	0.00035
NSSE	0.00030

Table 4.4. F-scores of the attributes for the 2007 dataset.

Variables	**Score**
Specialization Area	0.00823
20-meter shuttle run test score	0.00602
Vertical jump test score	0.00100
NSPE	0.00097
30-meter dash test	0.00037
NSSE	0.00026
Coordination and skill test score	0.00024
GPA	-0.27302
Gender	-0.030277

The list of attributes selected by the CFS algorithm for the 2006 and 2007 datasets are given in Table 4.5 and Table 4.6, respectively.

Table 4.5. List of selected attributes for the 2006 dataset.

Variables
NSPE
GPA
Specialization Area
Vertical jump test score
30-meter dash test score
20-meter shuttle run test score

Table 4.6. List of selected attributes for the 2007 dataset.

Variables
NSPE
Specialization Area
20-meter shuttle run test score

Table 4.7 through Table 4.12 show the admission decision models and the predictor variables each model includes.

Table 4.7. Admission decision models based on Relief-F for the 2006 dataset.

Models	**Predictor Variables**
Model 1	Gender, NSSE, NSPE, GPA, SA, VJTS, CSTS, DTS, SRTS
Model 2	Gender, NSSE, NSPE, SA, VJTS, CSTS, DTS, SRTS,
Model 3	Gender, NSSE, NSPE, VJTS, CSTS, DTS, SRTS
Model 4	Gender, NSSE, NSPE, VJTS, CSTS, DTS
Model 5	NSSE, NSPE, VJTS, CSTS, DTS
Model 6	NSSE, NSPE, VJTS, Dash
Model 7	NSPE, Vertical, DTS
Model 8	NSPE, DTS
Model 9	DTS

SA: Specialization area, **VJTS**: Vertical jump test score, C**STS**: Coordination and skill test score

DTS: 30-m dash test score, **SRTS**: 20-meter Shuttle run test score

Table 4.8. Admission decision models based on F-score for the 2006 dataset.

Models	**Predictor Variables**
Model 10	Gender, NSSE, NSPE, GPA, SA, VJTS, CSTS, DTS, SRTS
Model 11	Gender, NSPE, GPA, SA, VJTS, CSTS, DTS, SRTS
Model 12	Gender, NSPE, SA, VJTS, CSTS, DTS, SRTS
Model 13	NSPE, SA, VJTS, CSTS, DTS, SRTS
Model 14	NSPE, SA, VJTS, DTS, SRTS
Model 15	SA, VJTS, DTS, SRTS
Model 16	SA, VJTS, SRTS
Model 17	VJTS, SRTS
Model 18	SRTS

SA: Specialization area, **VJTS**: Vertical jump test score, C**STS**: Coordination and skill test score

DTS: 30-m dash test score, **SRTS**: 20-meter Shuttle run test score

Table 4.9. Admission decision model based on CFS for the 2006 dataset.

Models	Predictor Variables
Model 19	NSPE, GPA, SA, VJTS, DTS, SRTS

SA: Specialization area, **VJTS**: Vertical jump test score, C**STS**: Coordination and skill test score

DTS: 30-m dash test score, **SRTS**: 20-meter Shuttle run test score

Table 4.10. Admission decision models based on Relief-F for the 2007 dataset.

Models	Predictor Variables
Model 20	Gender, NSSE, NSPE, GPA, SA, VJTS, CSTS, DTS, SRTS
Model 21	Gender, NSSE, NSPE, GPA, VJTS, CSTS, DTS, SRTS
Model 22	Gender, NSSE, NSPE, GPA, VJTS, DTS, SRTS
Model 23	NSSE, NSPE, GPA, VJTS, DTS, SRTS
Model 24	NSSE, NSPE, GPA, DTS, SRTS
Model 25	NSSE, GPA, DTS, SRTS
Model 26	NSSE, DTS, SRTS
Model 27	DTS, SRTS
Model 28	DTS

SA: Specialization area, **VJTS**: Vertical jump test score, C**STS**: Coordination and skill test score

DTS: 30-m dash test score, **SRTS**: 20-meter Shuttle run test score

Table 4.11. Admission decision models based on F-score for the 2007 dataset.

Models	Predictor Variables
Model 29	Gender, NSSE, NSPE, GPA, SA, VJTS, CSTS, DTS, SRTS
Model 30	NSSE, NSPE, GPA, SA, VJTS, CSTS, DTS, SRTS
Model 31	NSSE, NSPE, SA, VJTS, CSTS, DTS, SRTS
Model 32	NSSE, NSPE, SA, VJTS, DTS, SRTS
Model 33	NSPE, SA, VJTS, DTS, SRTS
Model 34	NSPE, SA, VJTS, SRTS
Model 35	SA, VJTS, SRTS
Model 36	SA, SRTS
Model 37	SRTS

SA: Specialization area, **VJTS**: Vertical jump test score, C**STS**: Coordination and skill test score

DTS: 30-m dash test score, **SRTS**: 20-meter Shuttle run test score

Table 4.12. Admission decision model based on CFS for the 2007 dataset.

Models	Predictor Variables
Model 38	NSPE, SA, SRTS

SA: Specialization area, **VJTS**: Vertical jump test score, C**STS**: Coordination and skill test score

DTS: 30-m dash test score, **SRTS**: 20-meter Shuttle run test score

Five machine learning methods including SVM, MLP, RBF Network, TB and KMC have been employed to predict the admission decision of a candidate to the School of Physical Education and Sports at Cukurova University. The

performance of SVM, MLP, RBF Network, TB and KMC models have been evaluated by using 10-fold cross validation and different percentage splits of data and calculating classification accuracy, specificity, sensitivity, PPV and NPV.

Figure 4.1 through Figure 4.110 show the classification accuracies, senstivities, specificities, PPV and NPV of all prediction models using Relief-F, F-Score and CFS on the 2006 and 2007 datasets.

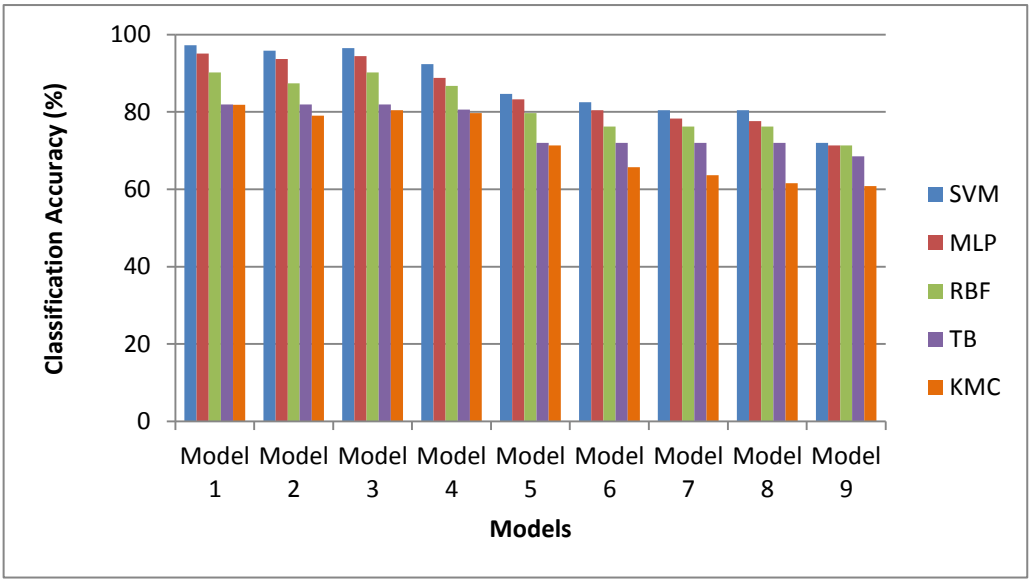


Figure 4.1. Classification accuracies of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case)

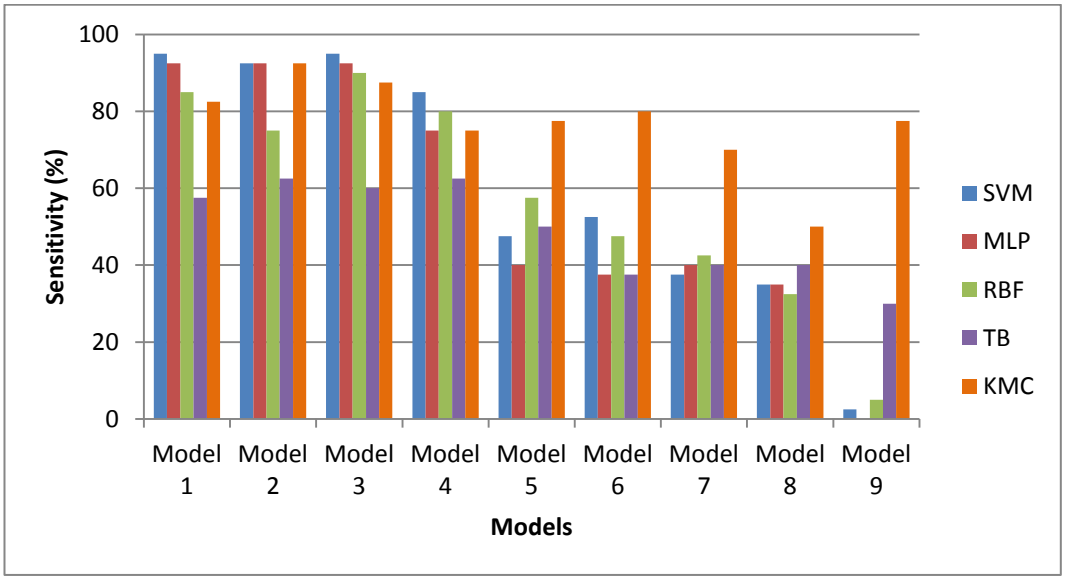


Figure 4.2. Sensitivities of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case)

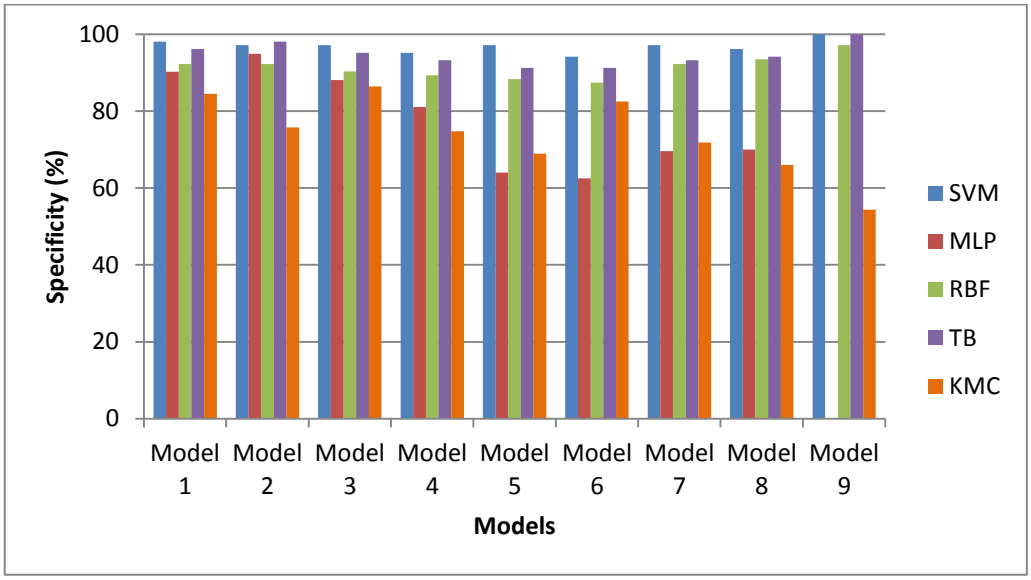


Figure 4.3. Specificities of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case)

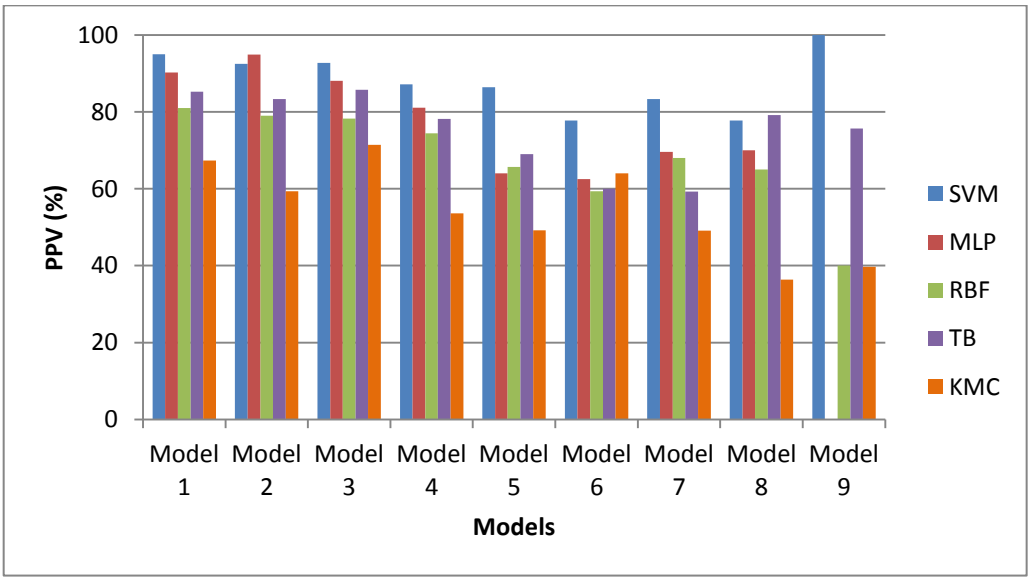


Figure 4.4. PPV's of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case)

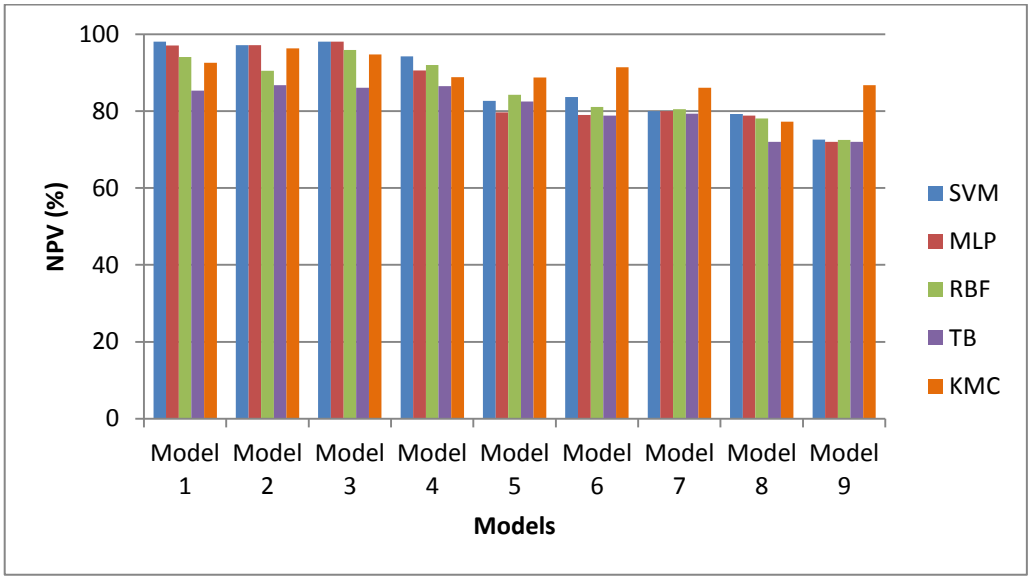


Figure 4.5. NPV's of prediction models using Relief-F on the 2006 dataset (10-fold cross validation case)

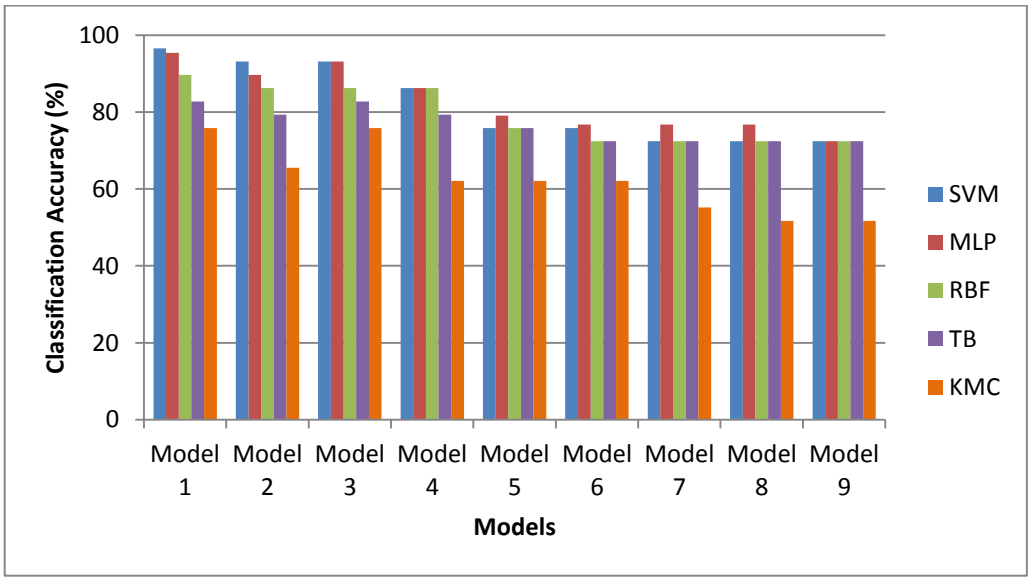


Figure 4.6. Classification accuraciesof prediction models using Relief-F on the 2006 dataset (Percentage Split (%20) case)

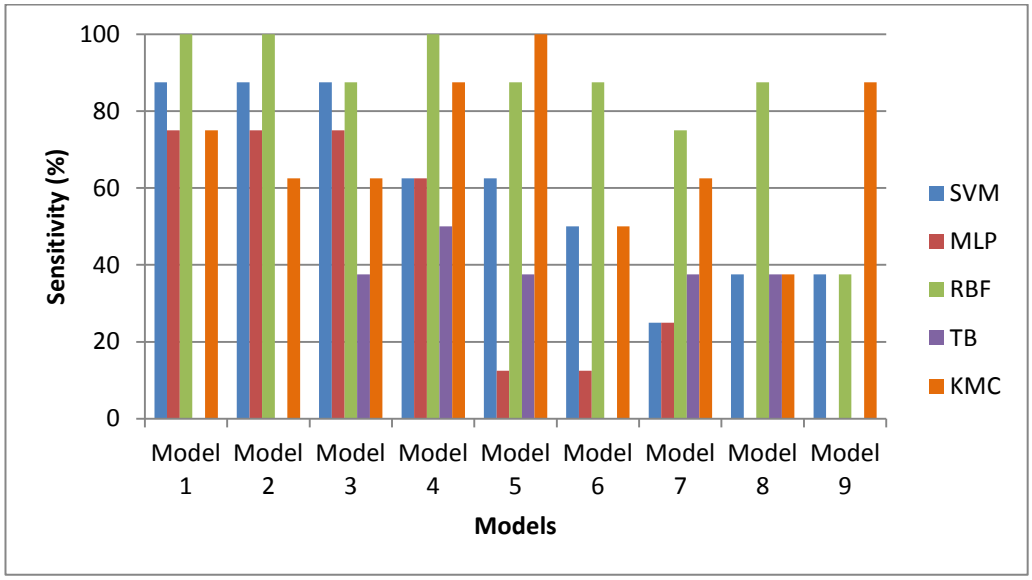


Figure 4.7. Sensitivities of prediction models using Relief-F on the 2006 dataset (Percentage Split (%20) case)

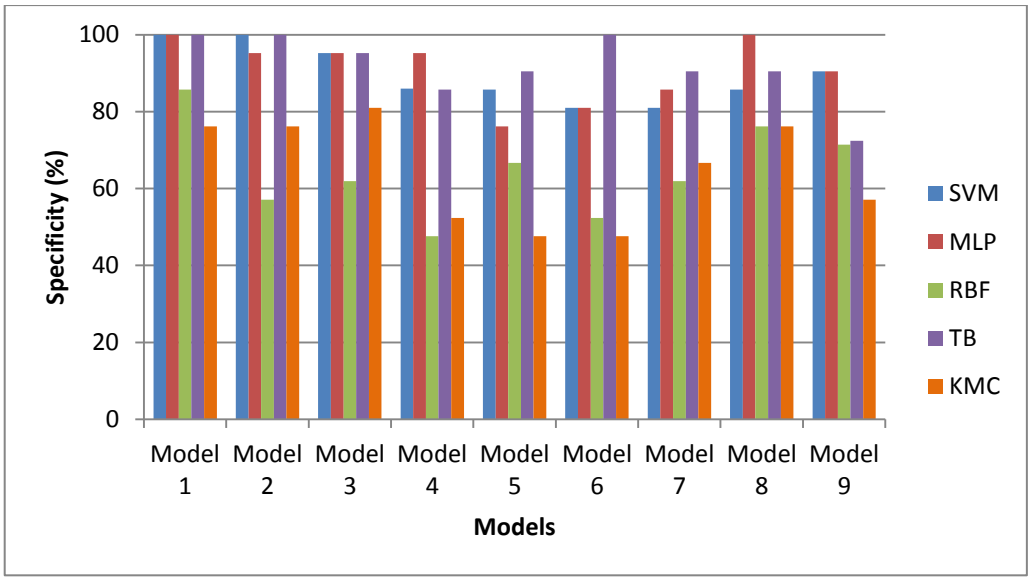


Figure 4.8. Specificities of prediction models using Relief-F on the 2006 dataset (Percentage Split (%20) case)

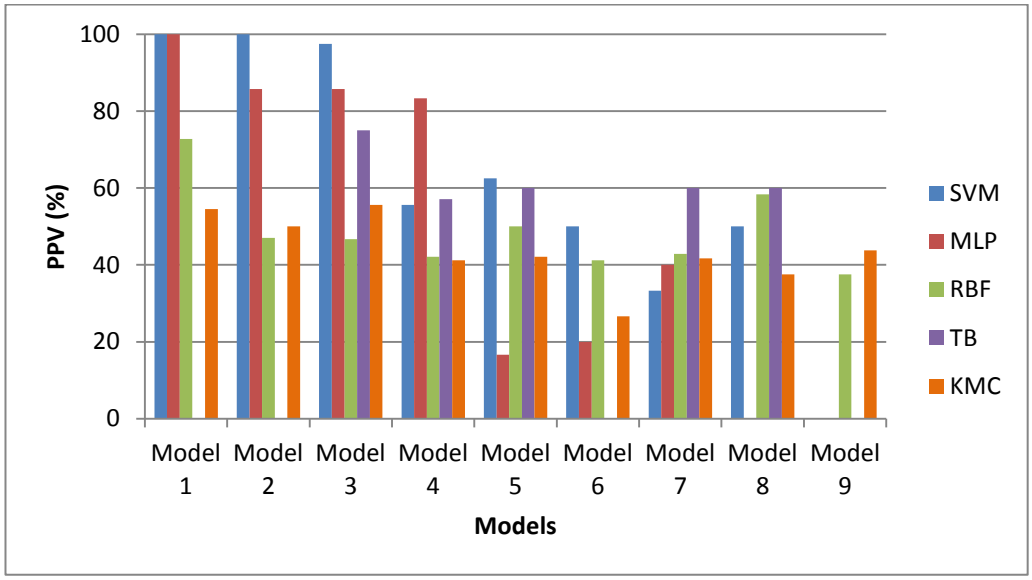


Figure 4.9. PPV's of prediction models using Relief-F on the 2006 dataset (Percentage Split (%20) case)

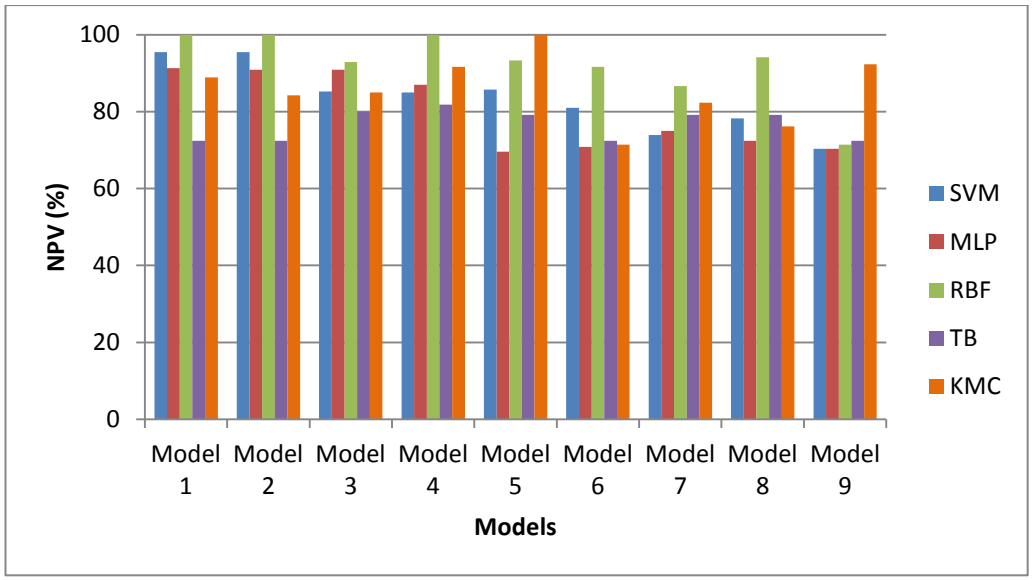


Figure 4.10. NPV'S of prediction models using Relief-F on the 2006 dataset (Percentage Split (%20) case)

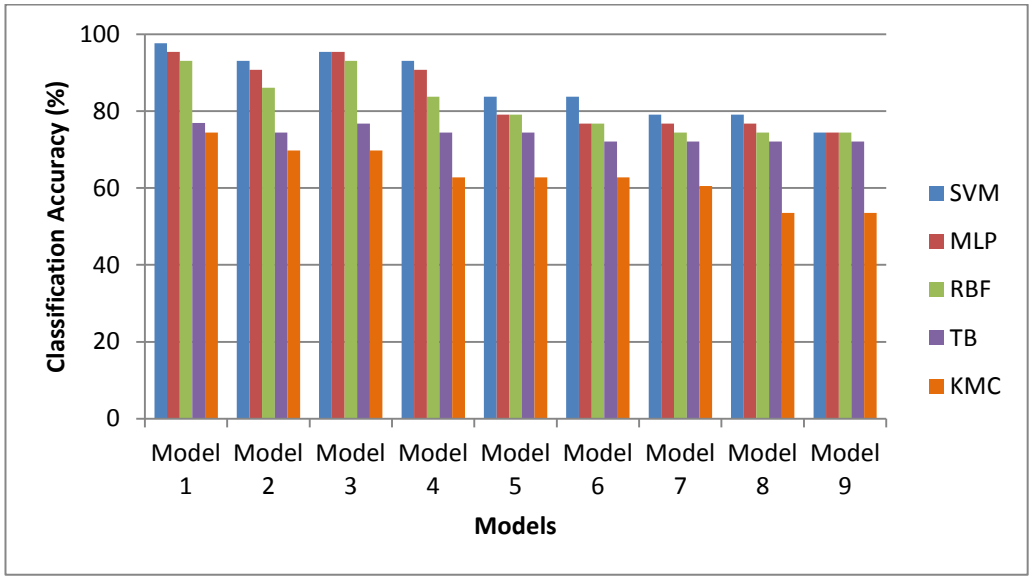


Figure 4.11. Classification accuracies of prediction models using Relief-F on the 2006 dataset (Percentage Split (%30) case)

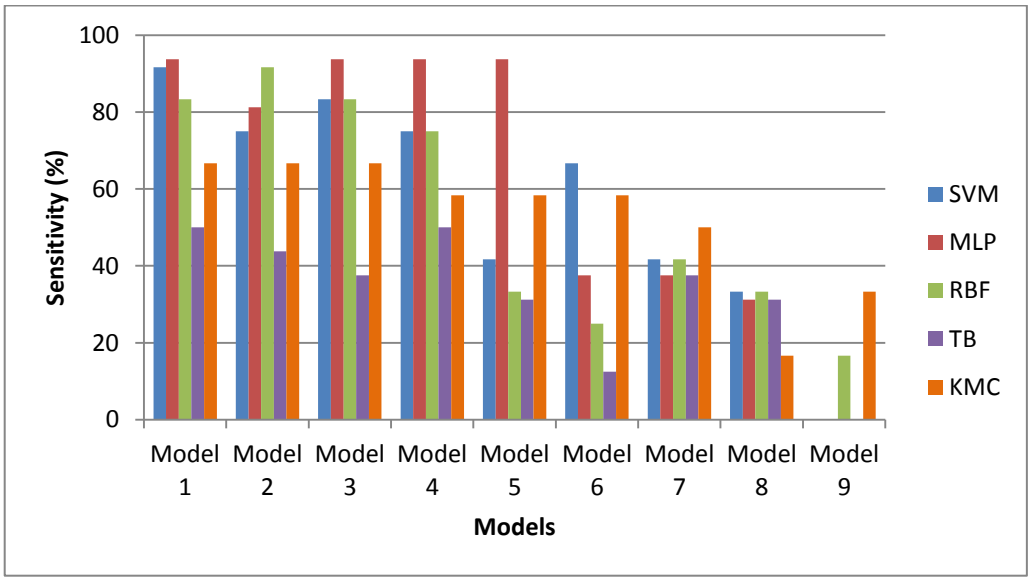


Figure 4.12. Sensitivities of prediction models using Relief-F on the 2006 dataset (Percentage Split (%30) case)

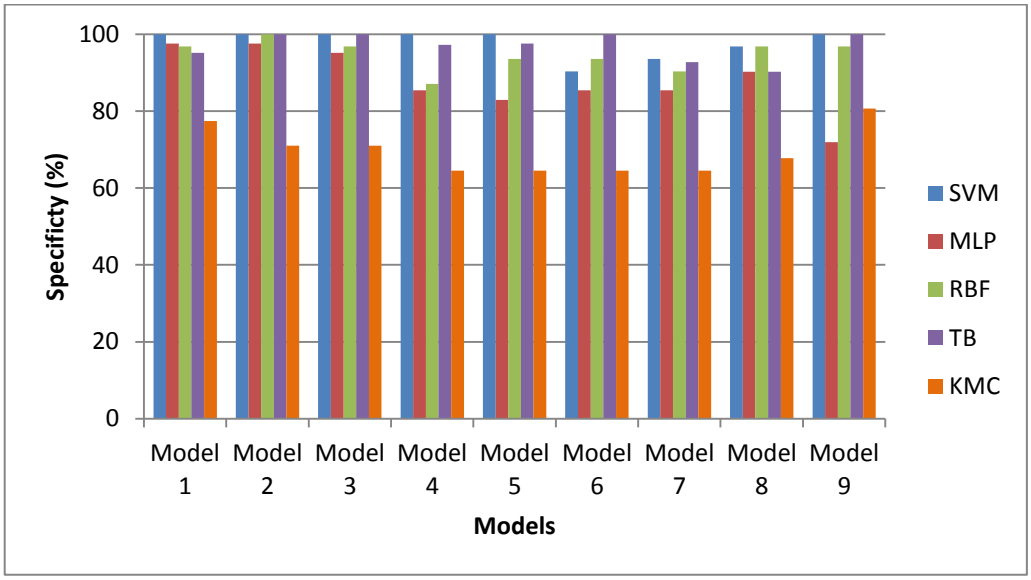


Figure 4.13. Specificities of prediction models using Relief-F on the 2006 dataset (Percentage Split (%30) case)

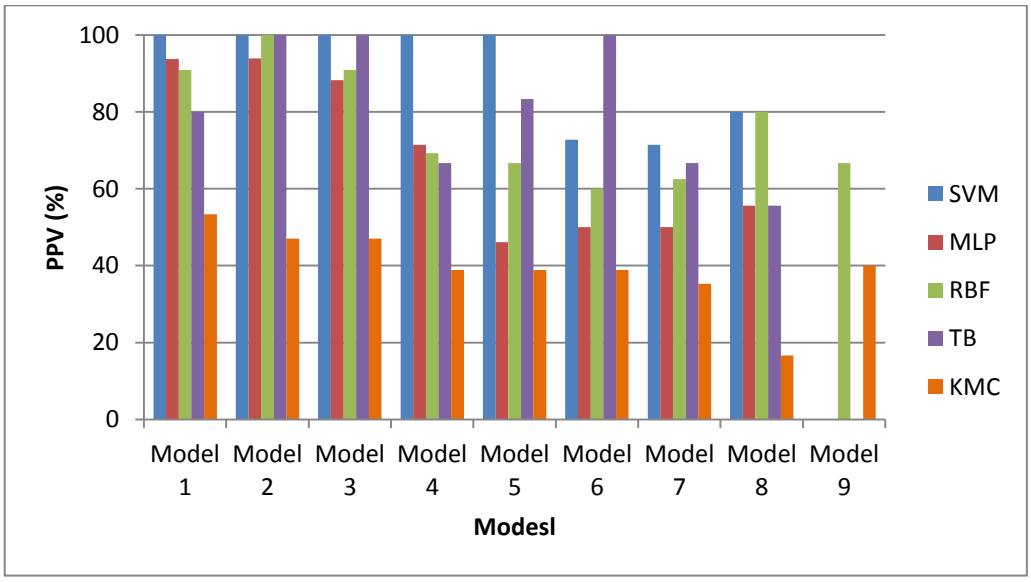


Figure 4.14. PPV's of prediction models using Relief-F on the 2006 dataset (Percentage Split (%30) case)

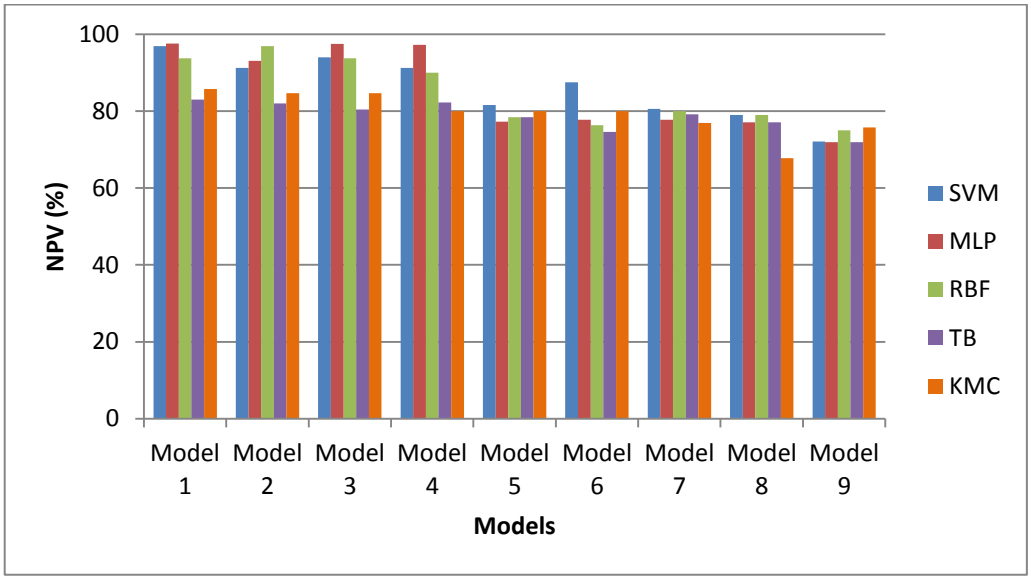


Figure 4.15. NPV's of prediction models using Relief-F on the 2006 dataset (Percentage Split (%30) case)

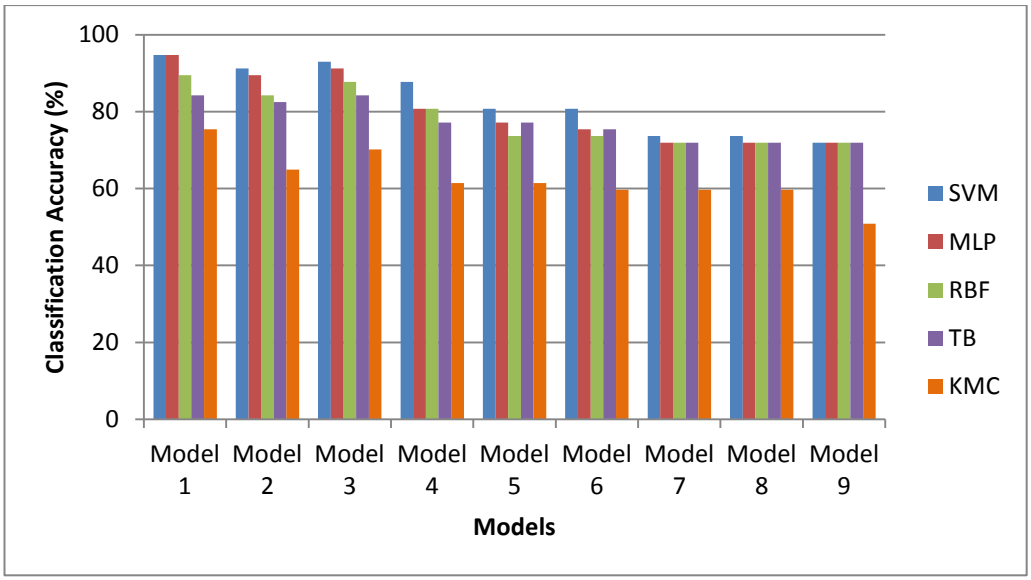


Figure 4.16. Classification accuracies of prediction models using Relief-F on the 2006 dataset (Percentage Split (%40) case)

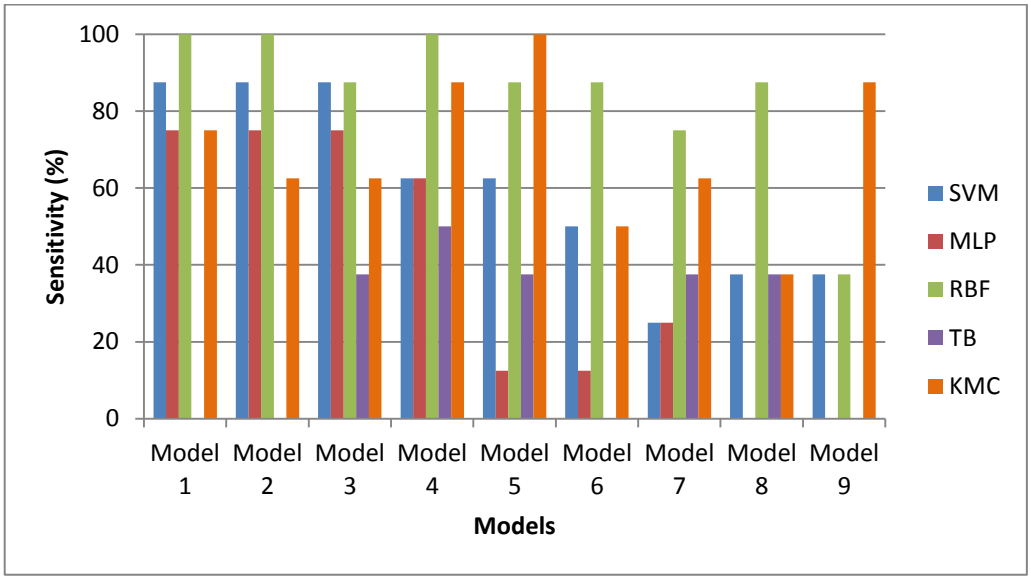


Figure 4.17. Sensitivities of prediction models using Relief-F on the 2006 dataset (Percentage Split (%40) case)

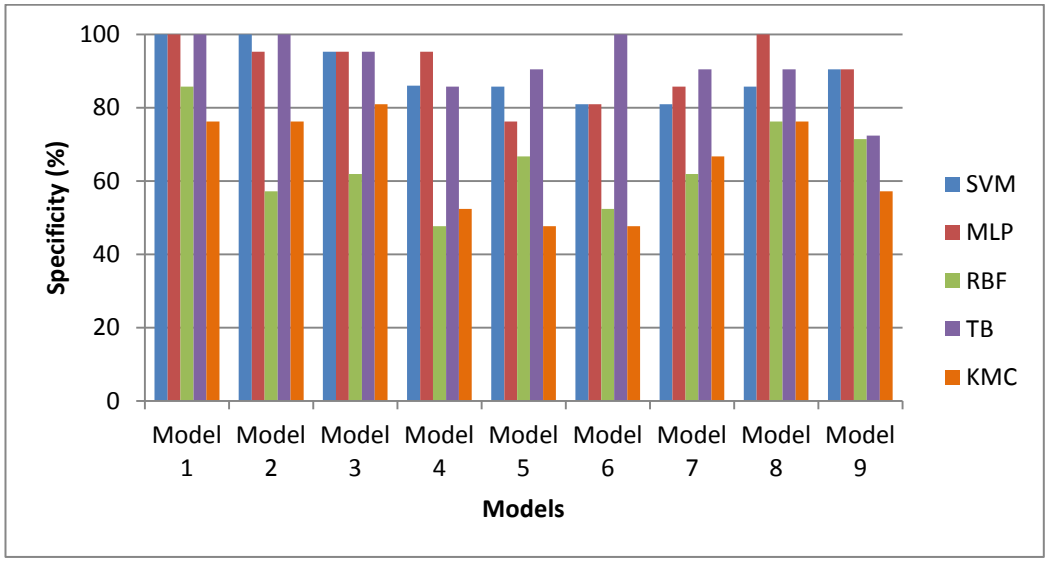


Figure 4.18. Specificities of prediction models using Relief-F on the 2006 dataset (Percentage Split (%40) case)

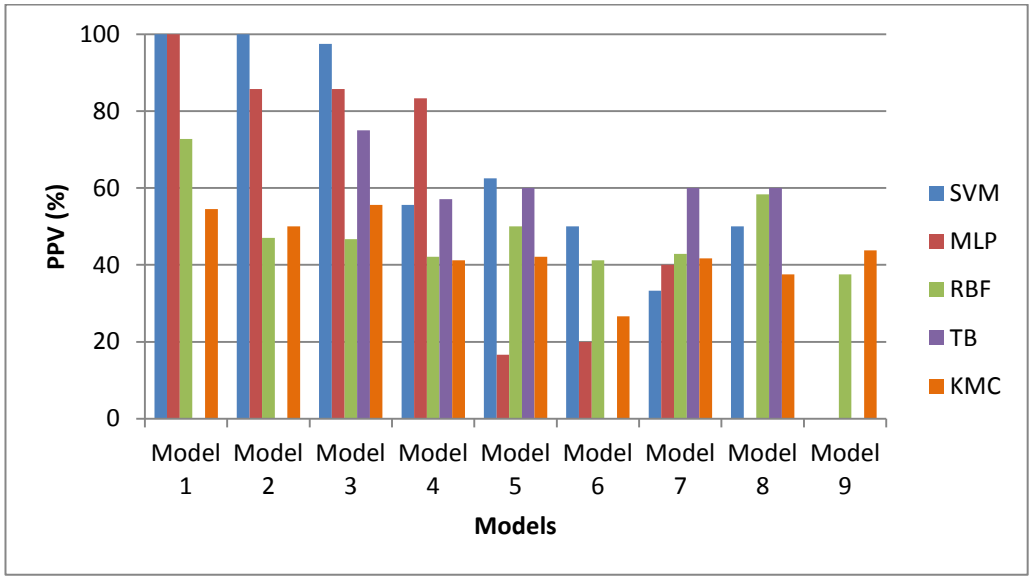


Figure 4.19. PPV's of prediction models using Relief-F on the 2006 dataset (Percentage Split (%40) case)

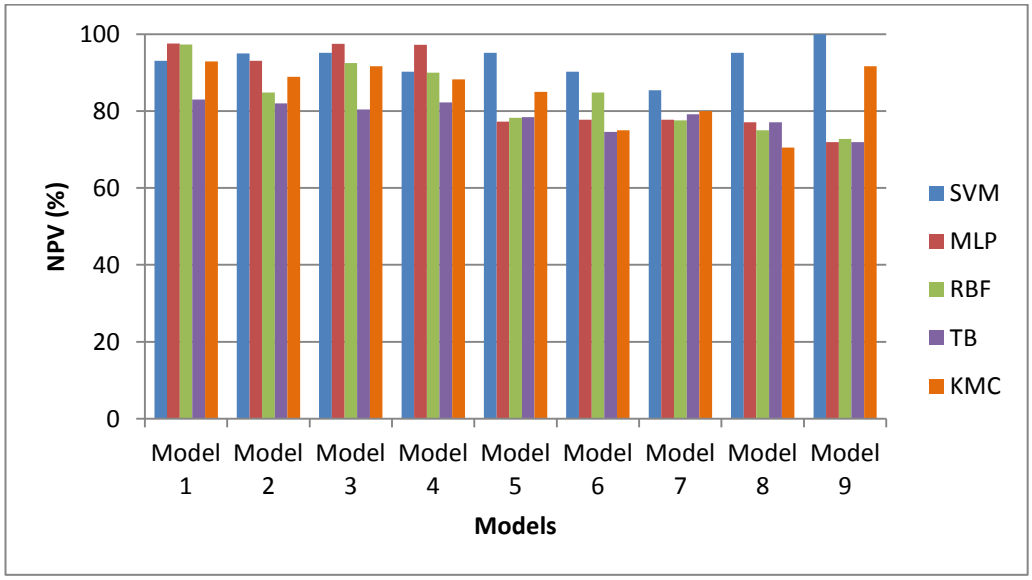


Figure 4.20. NPV's of prediction models using Relief-F on the 2006 dataset (Percentage Split (%40) case)

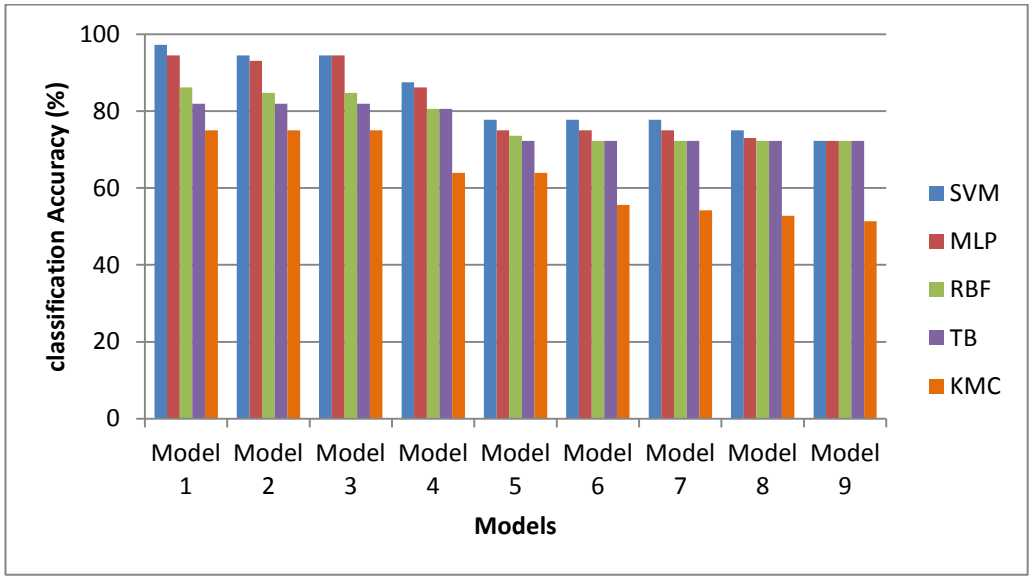


Figure 4.21. Classification accuracies of prediction models using Relief-F on the 2006 dataset (Percentage Split (%50) case)

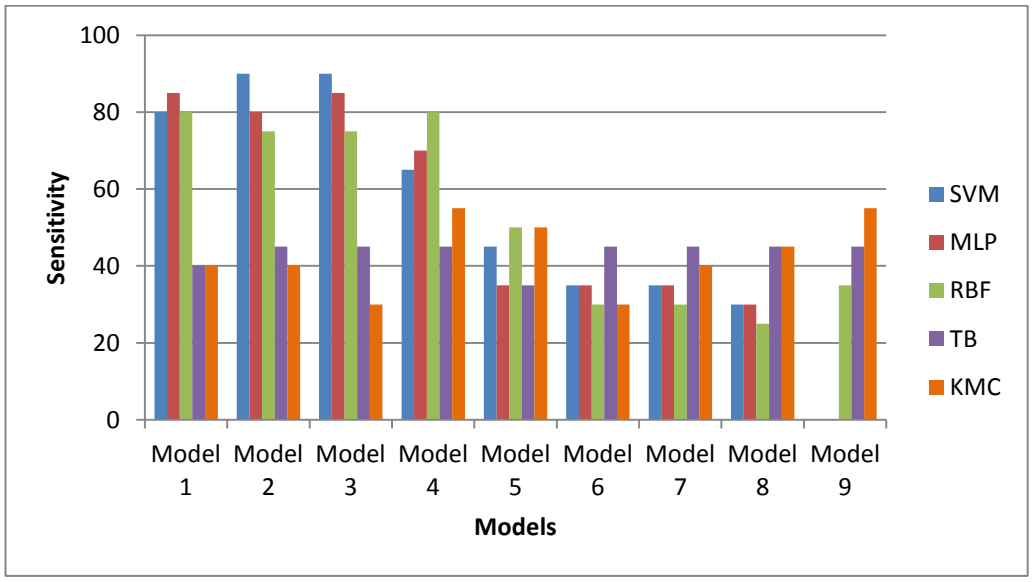


Figure 4.22. Sensitivities of prediction models using Relief-F on the 2006 dataset (Percentage Split (%50) case)

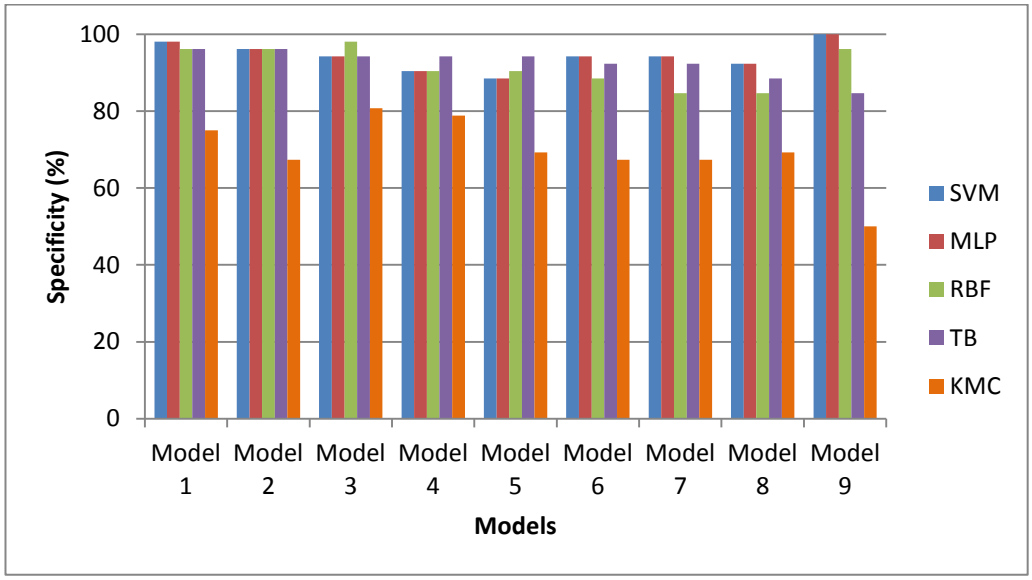


Figure 4.23. Specificites of prediction models using Relief-F on the 2006 dataset (Percentage Split (%50) case)

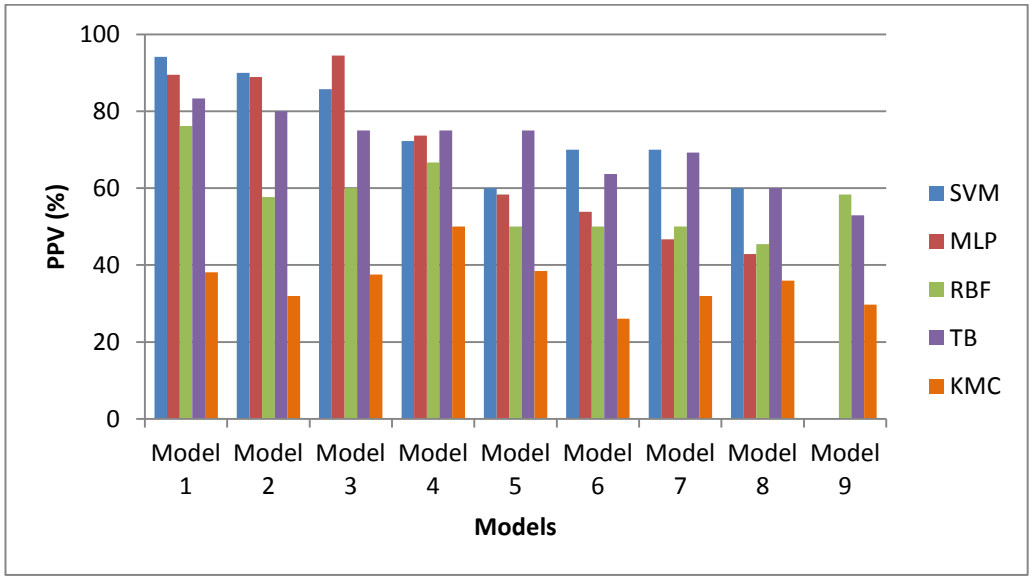


Figure 4.24. PPV'sof prediction models using Relief-F on the 2006 dataset (Percentage Split (%50) case)

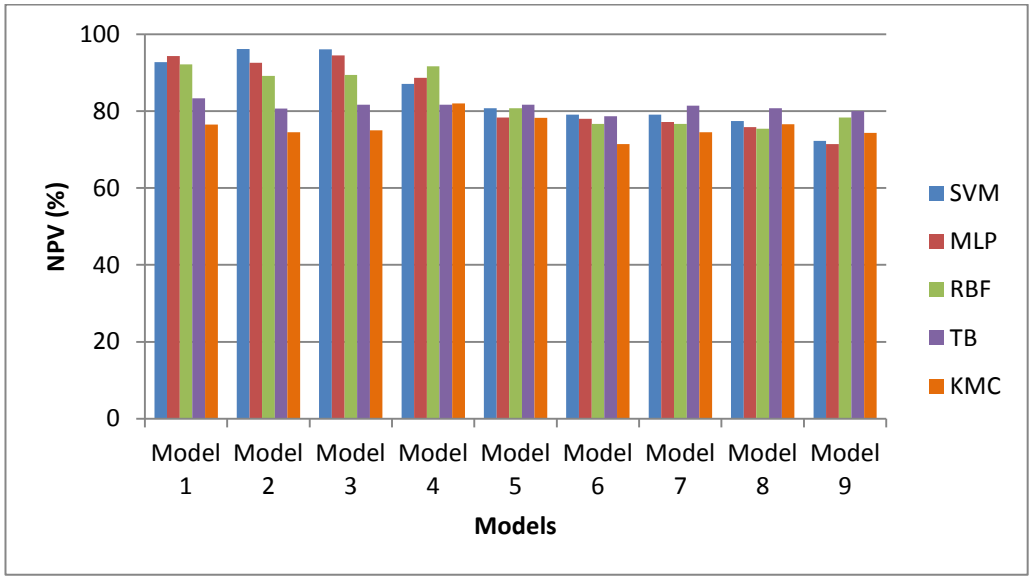


Figure 4.25. NPV'sof prediction models using Relief-F on the 2006 dataset (Percentage Split (%50) case)

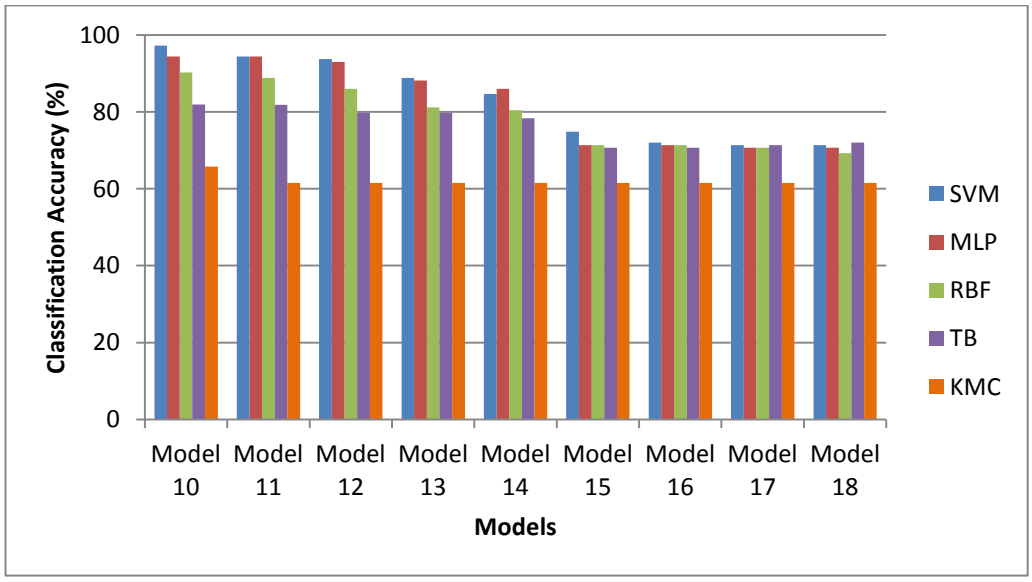


Figure 4.26. Classification Accuracies of prediction models using F-Score on the 2006 dataset (10-fold cross validation case)

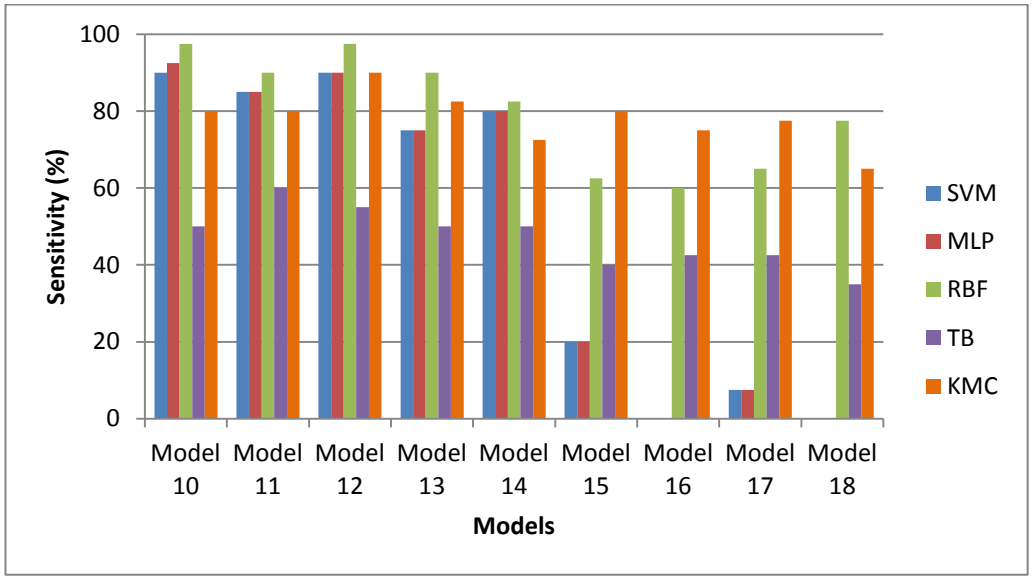


Figure 4.27. Sensitivities of prediction models using F-Score on the 2006 dataset (10-fold cross validation case)

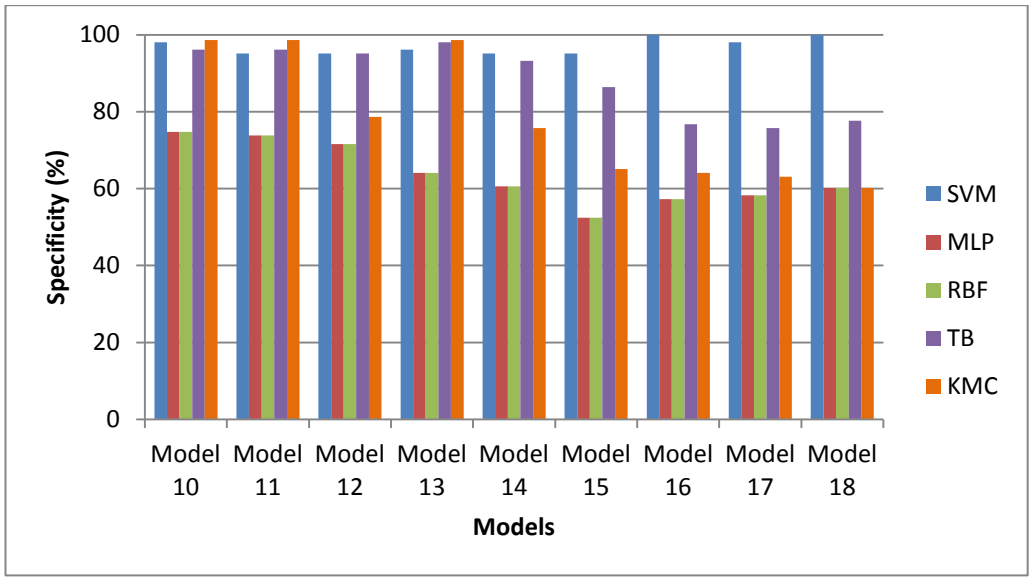


Figure 4.28. Specificities of prediction models using F-Score on the 2006 dataset (10-fold cross validation case)

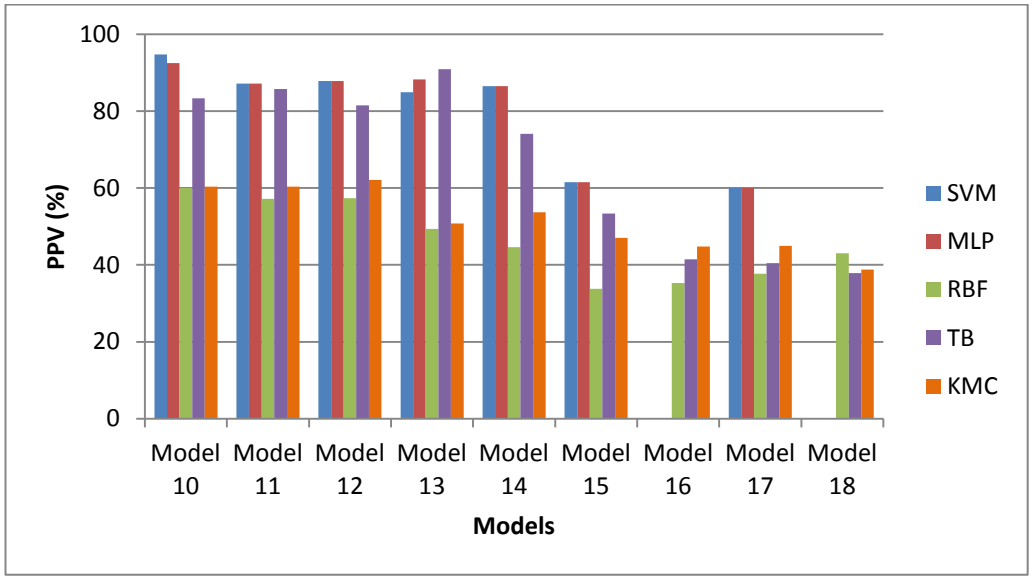


Figure 4.29. PPV'sof prediction models using F-Score on the 2006 dataset (10-fold cross validation case)

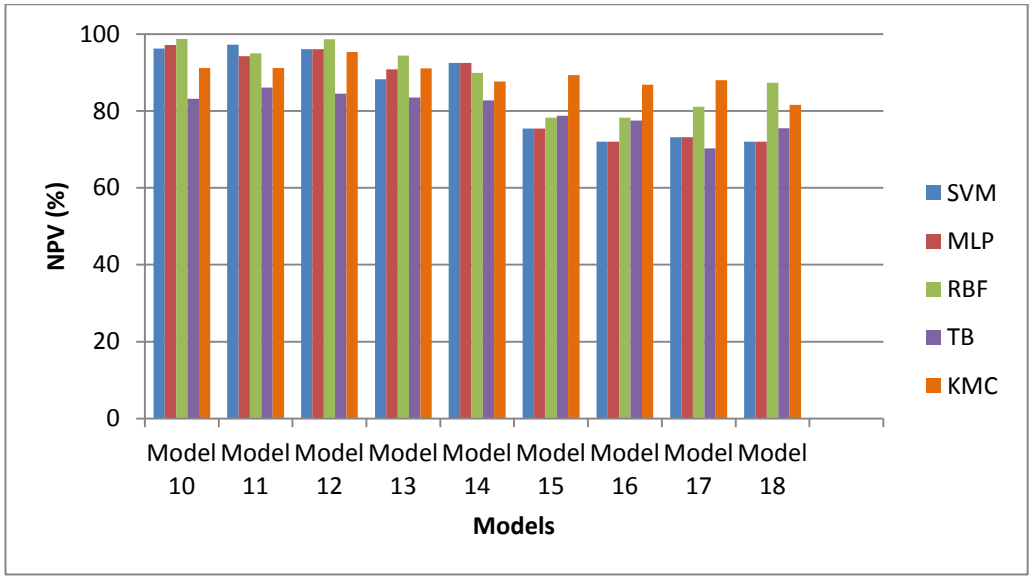


Figure 4.30. NPV'sof prediction models using F-Score on the 2006 dataset (10-fold cross validation case)

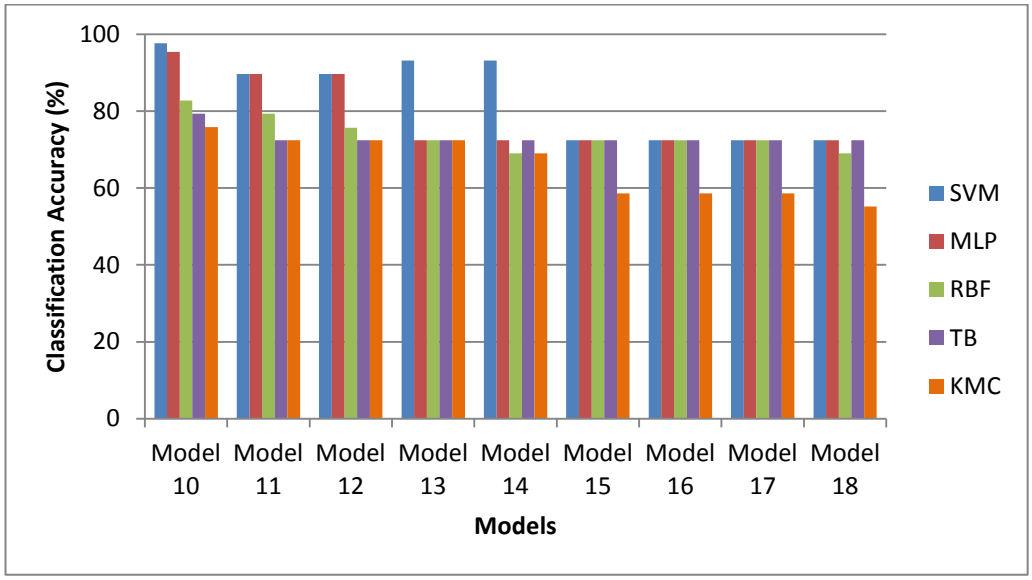


Figure 4.31. Classification accuracies of prediction models using F-Score on the 2006 dataset (Percentage split (%20) case)

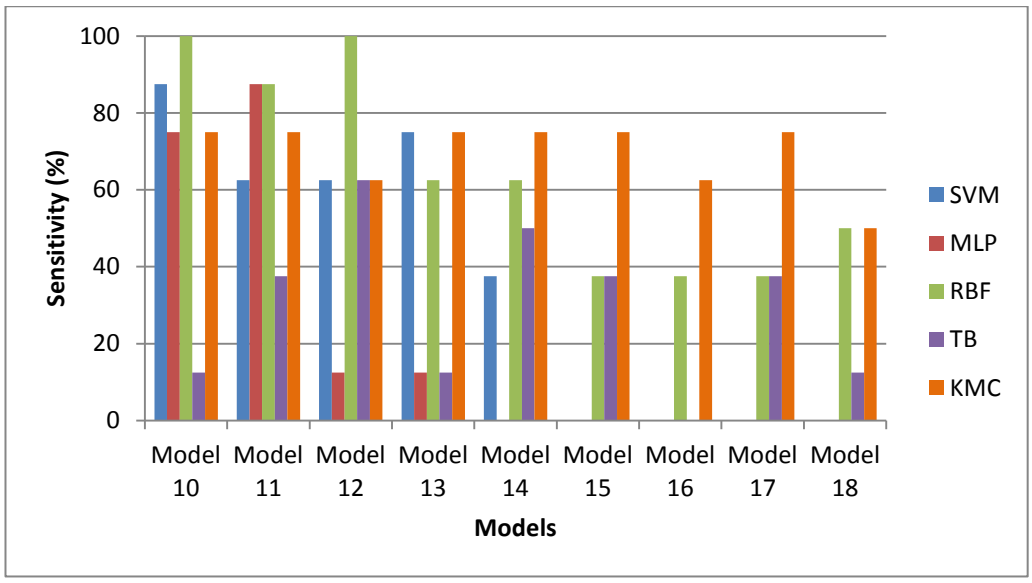


Figure 4.32. Sensitivities of prediction models using F-Score on the 2006 dataset (Percentage split (%20) case)

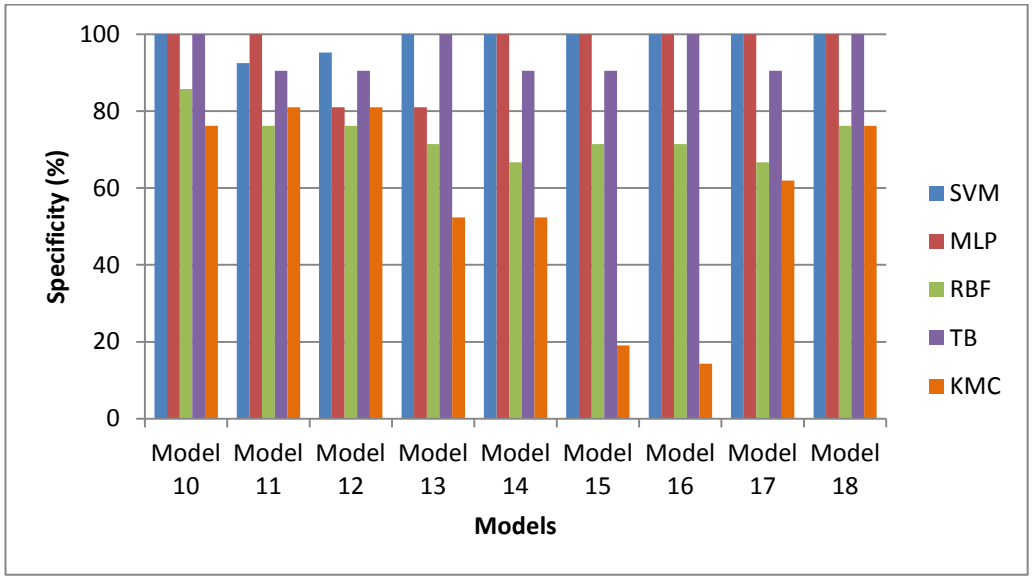


Figure 4.33. Specificities of prediction models using F-Score on the 2006 dataset (Percentage split (%20) case)

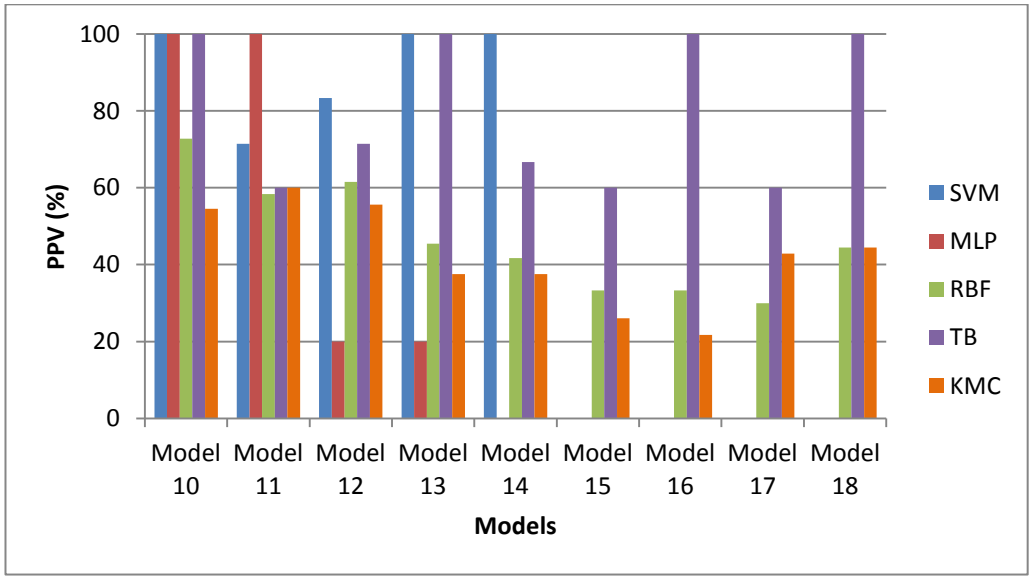


Figure 4.34. PPV'sof prediction models using F-Score on the 2006 dataset (Percentage split (%20) case)

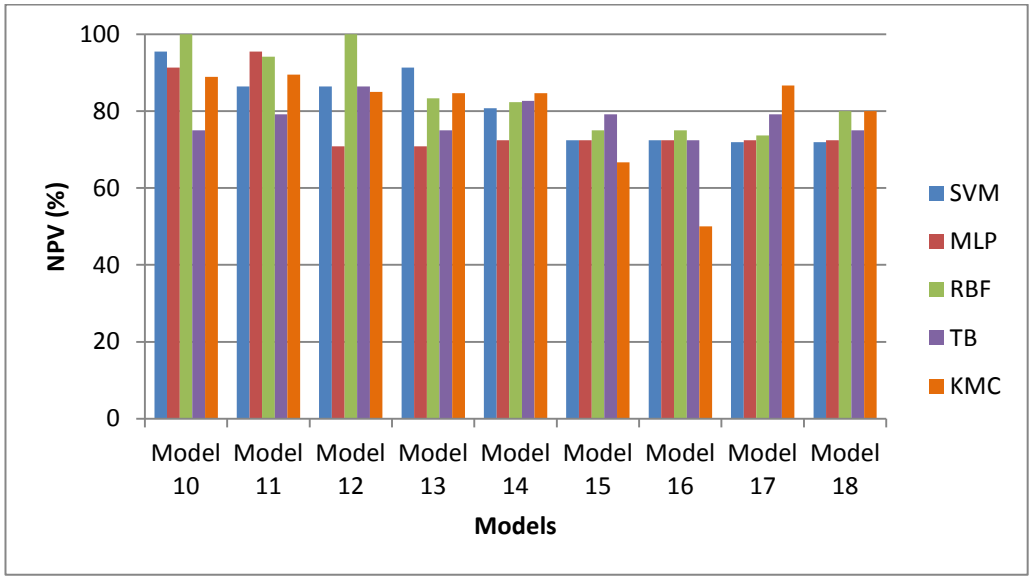


Figure 4.35. NPV'sof prediction models using F-Score on the 2006 dataset (Percentage split (%20) case)

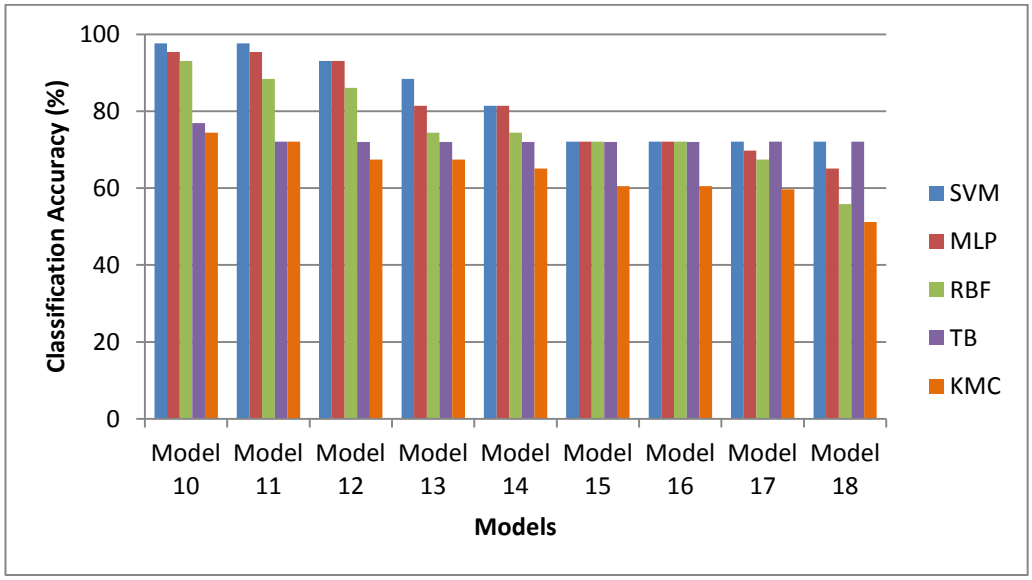


Figure 4.36. Classification accuraciesof prediction models using F-Score on the 2006 dataset (Percentage split (%30) case)

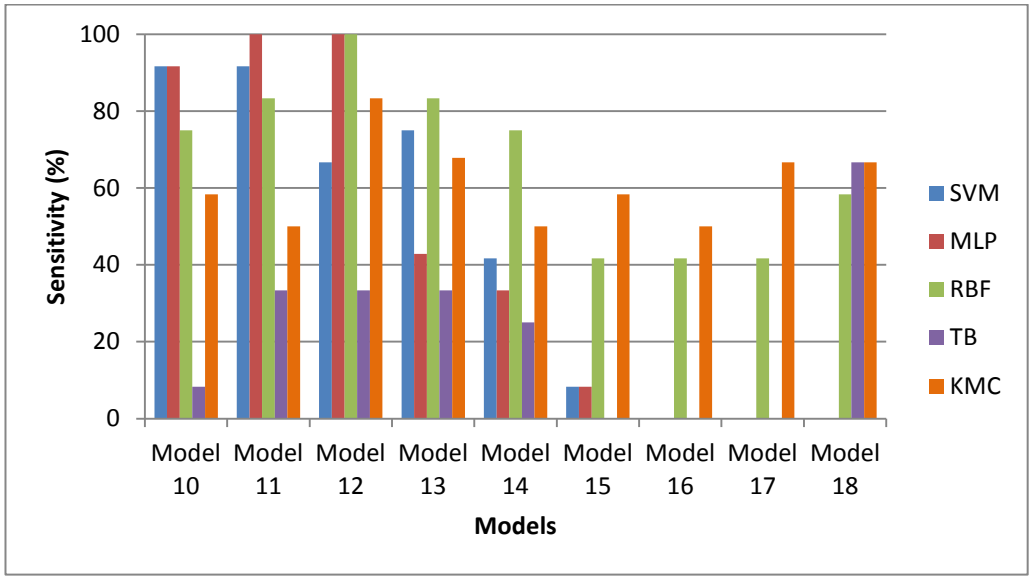


Figure 4.37. Sensitivities of prediction models using F-Score on the 2006 dataset (Percentage split (%30) case)

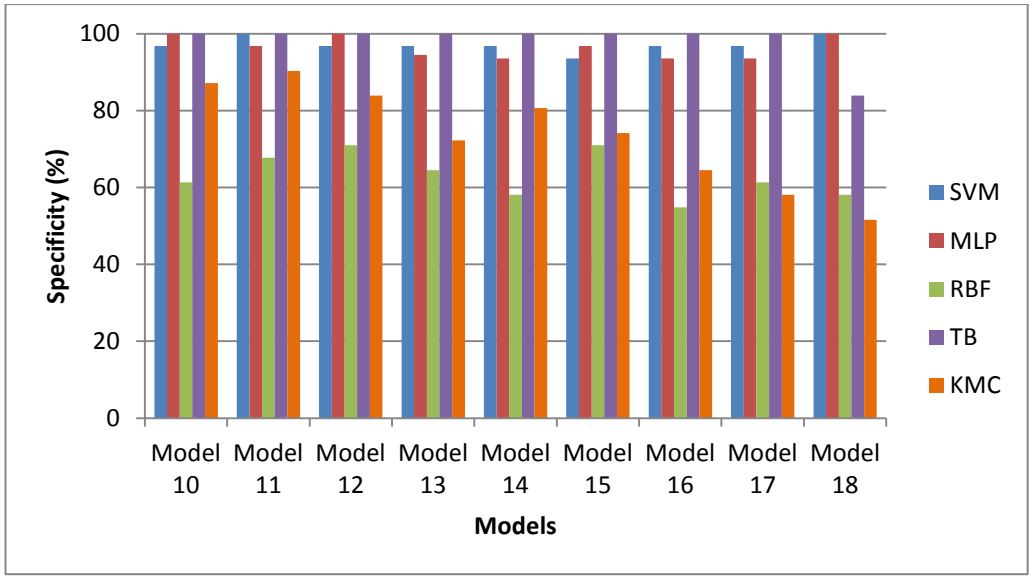


Figure 4.38. Specificities of prediction models using F-Score on the 2006 dataset (Percentage split (%30) case)

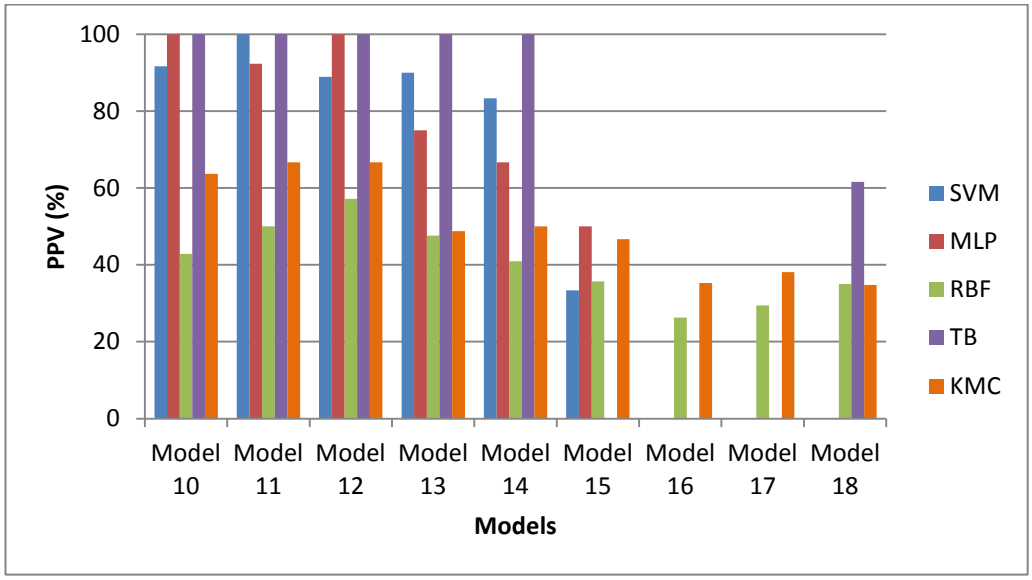


Figure 4.39. PPV's of prediction models using F-Score on the 2006 dataset (Percentage split (%30) case)

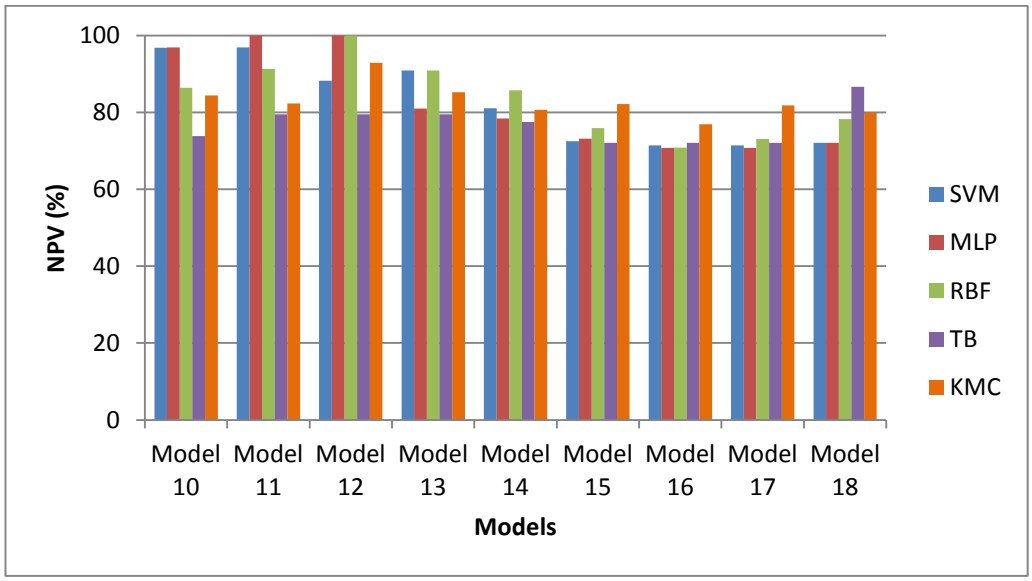


Figure 4.40. NPV's of prediction models using F-Score on the 2006 dataset (Percentage split (%30) case)

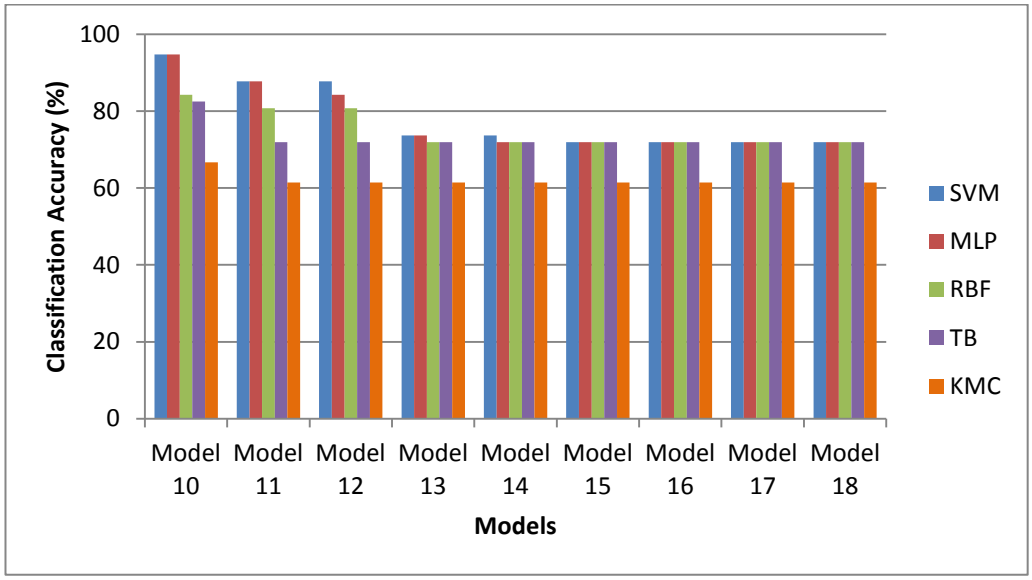


Figure 4.41. Classification accuracies of prediction models using F-Score on the 2006 dataset (Percentage split (%40) case)

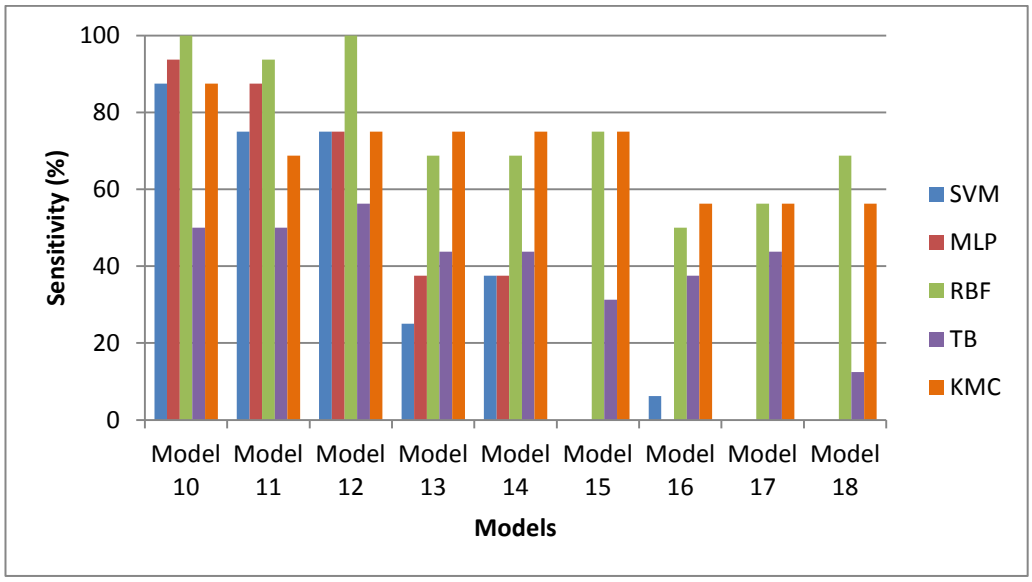


Figure 4.42. Sensitivities of prediction models using F-Score on the 2006 dataset (Percentage split (%40) case)

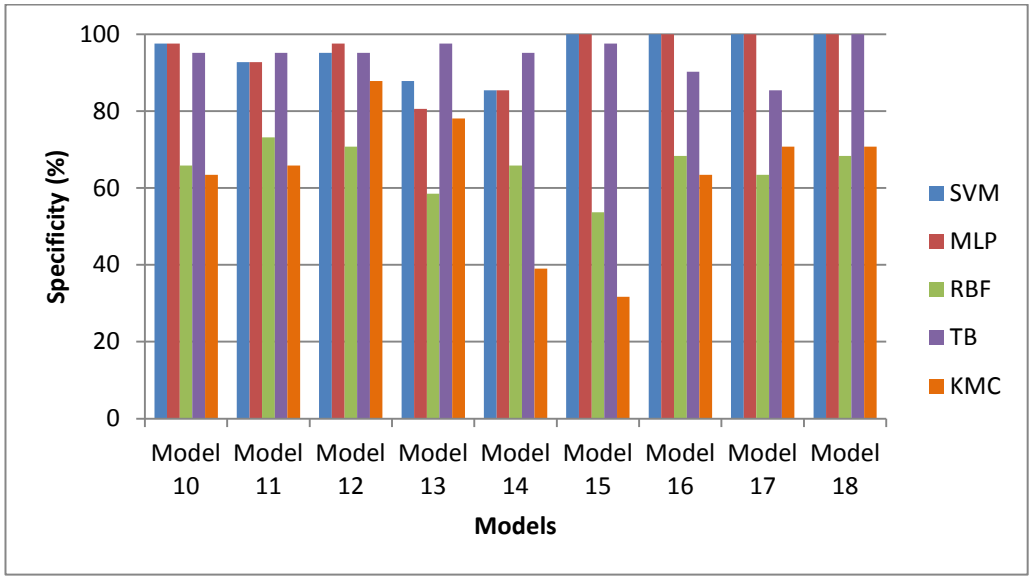


Figure 4.43. Specificities of prediction models using F-Score on the 2006 dataset (Percentage split (%40) case)

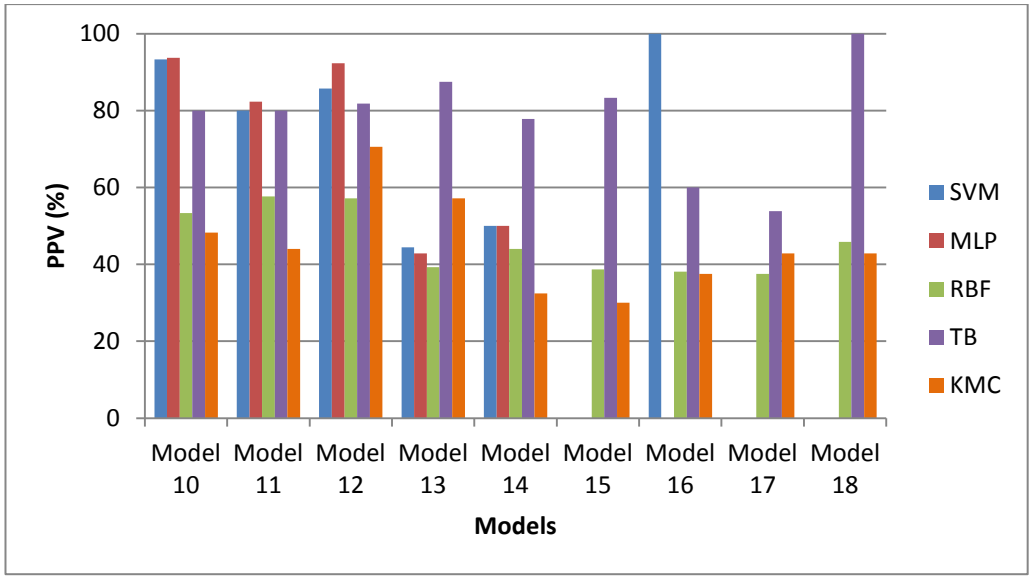


Figure 4.44. PPV's of prediction models using F-Score on the 2006 dataset (Percentage split (%40) case)

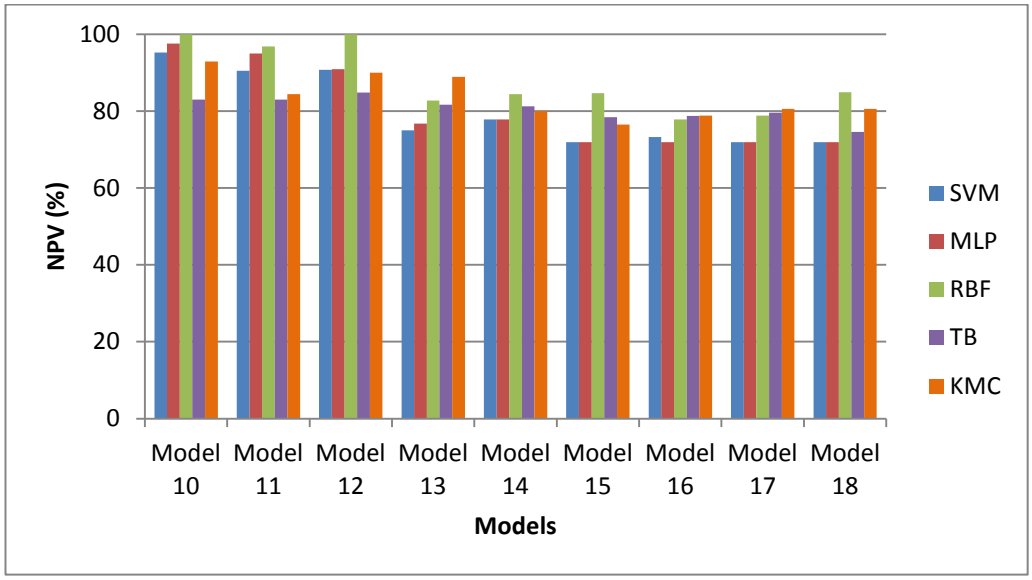


Figure 4.45. NPV's of prediction models using F-Score on the 2006 dataset (Percentage split (%40) case)

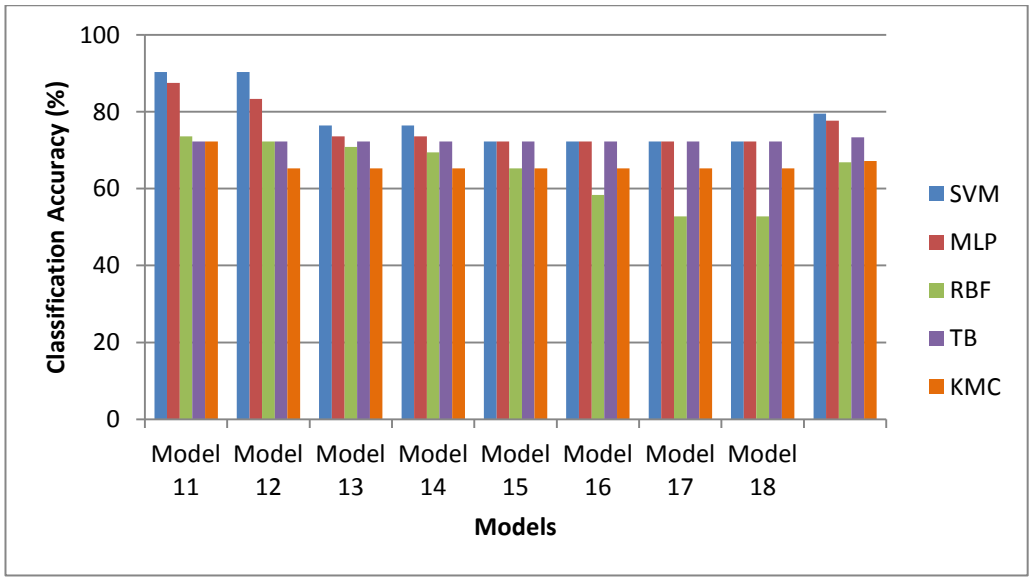


Figure 4.46. Classification accuracies of prediction models using F-Score on the 2006 dataset (Percentage split (%50) case)

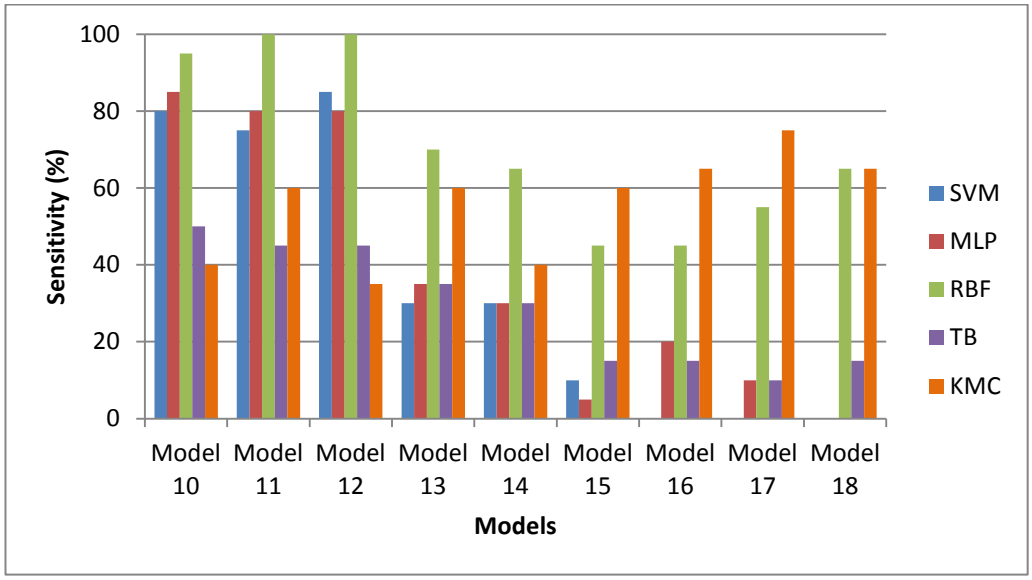


Figure 4.47. Sensitivites of prediction models using F-Score on the 2006 dataset (Percentage split (%50) case)

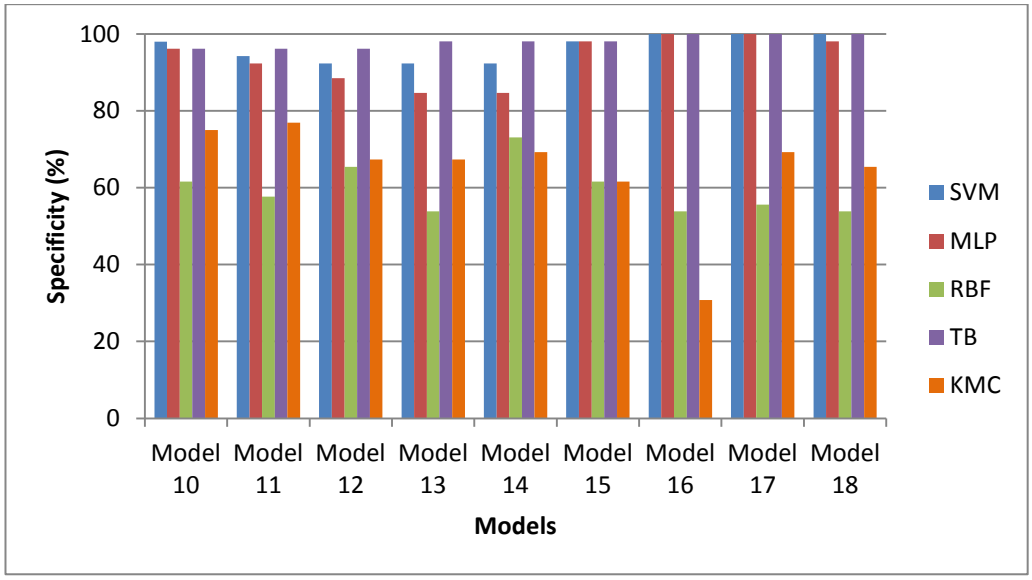


Figure 4.48. Specificities of prediction models using F-Score on the 2006 dataset (Percentage split (%50) case)

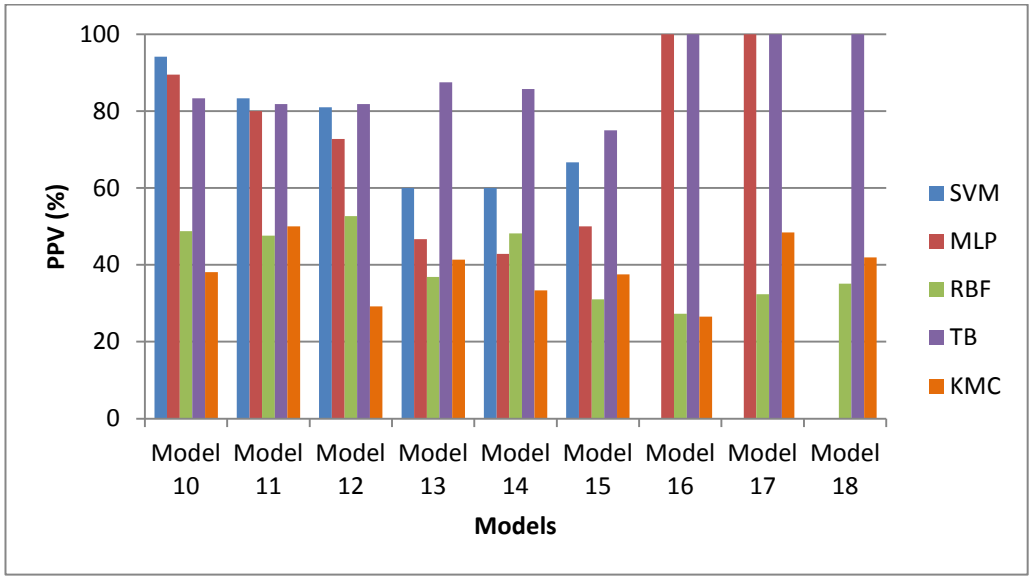


Figure 4.49. PPV's of prediction models using F-Score on the 2006 dataset (Percentage split (%50) case)

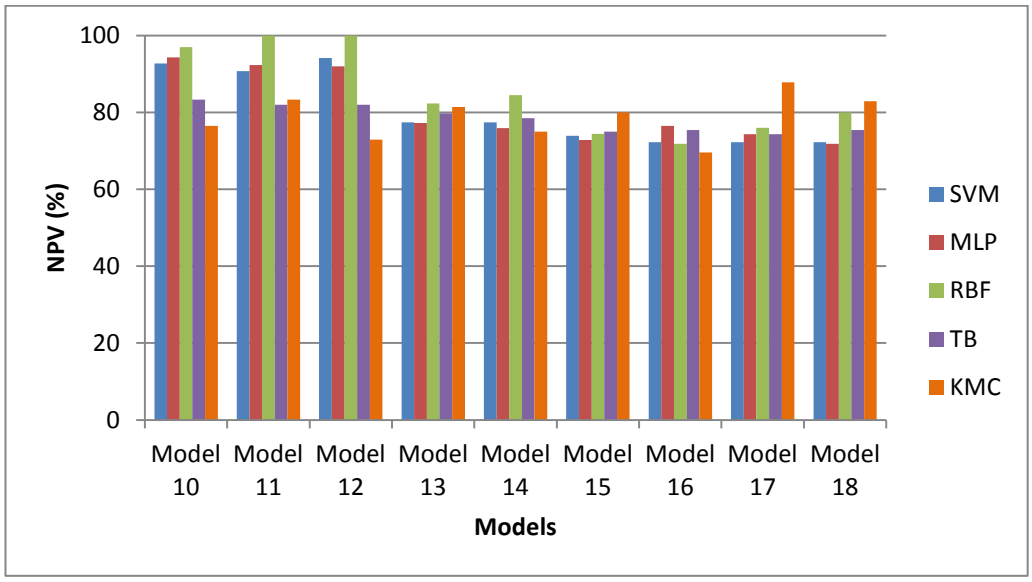


Figure 4.50. NPV's of prediction models using F-Score on the 2006 dataset (Percentage split (%50) case)

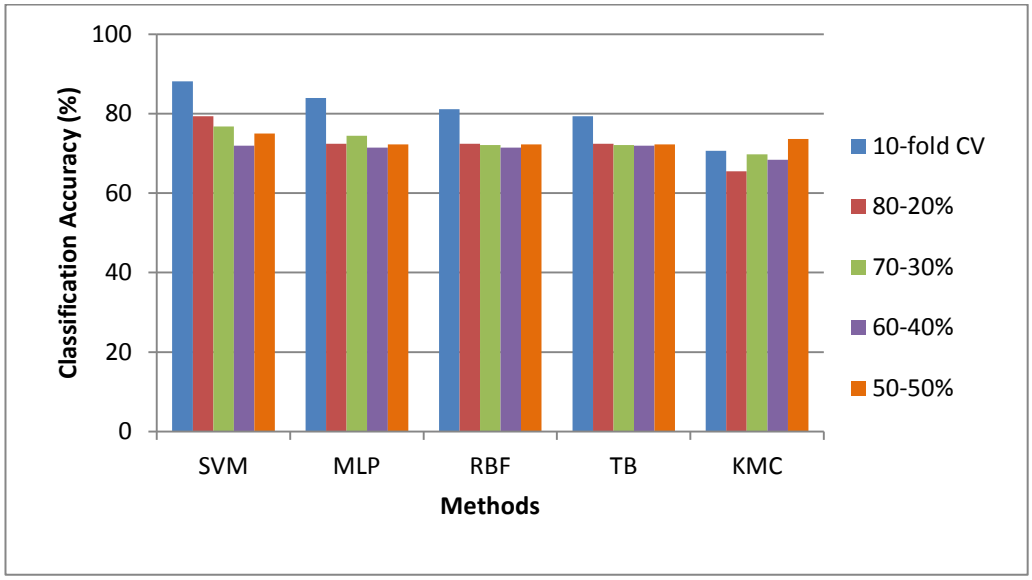


Figure 4.51. Classification accuracies of prediction models using CFS on the 2006 dataset

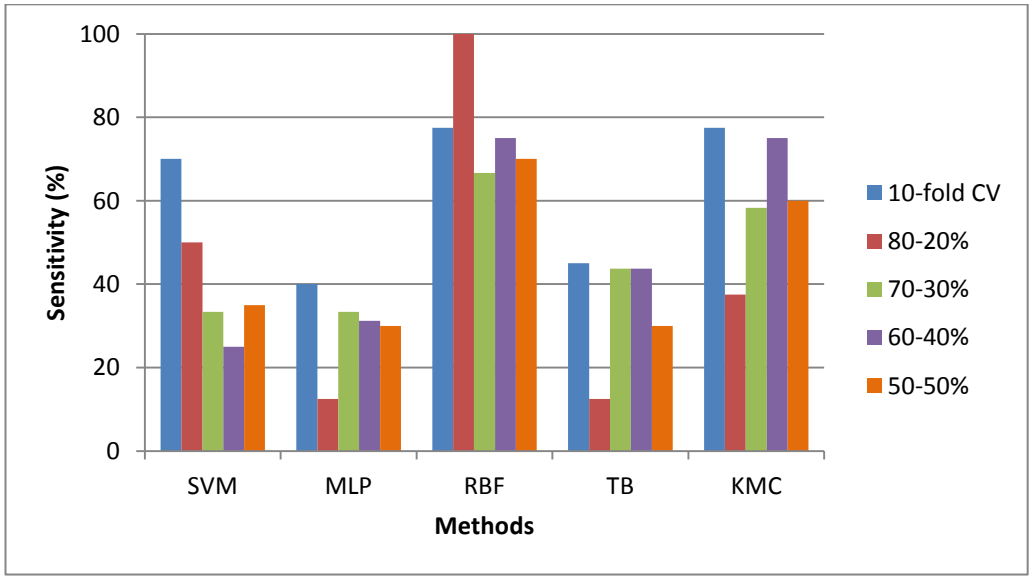


Figure 4.52. Sensitivities of prediction models using CFS on the 2006 dataset

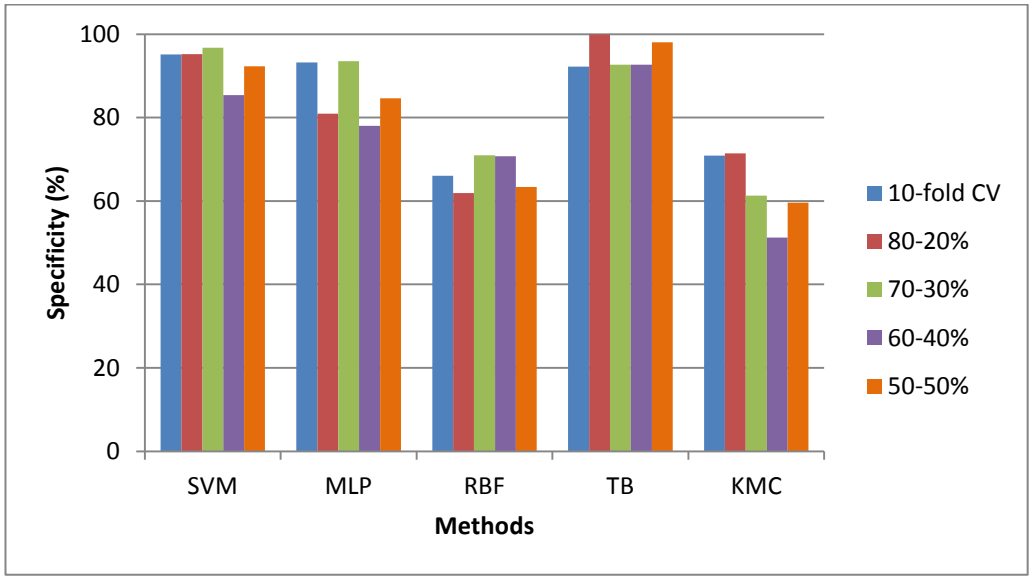


Figure 4.53. Specificities of prediction models using CFS on the 2006 dataset

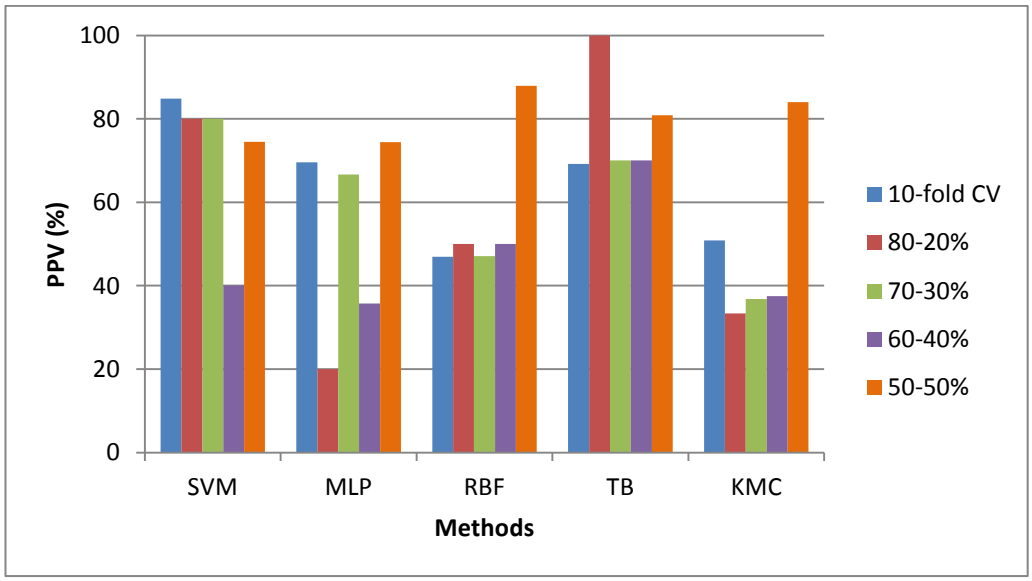


Figure 4.54. PPV's of prediction models using CFS on the 2006 dataset

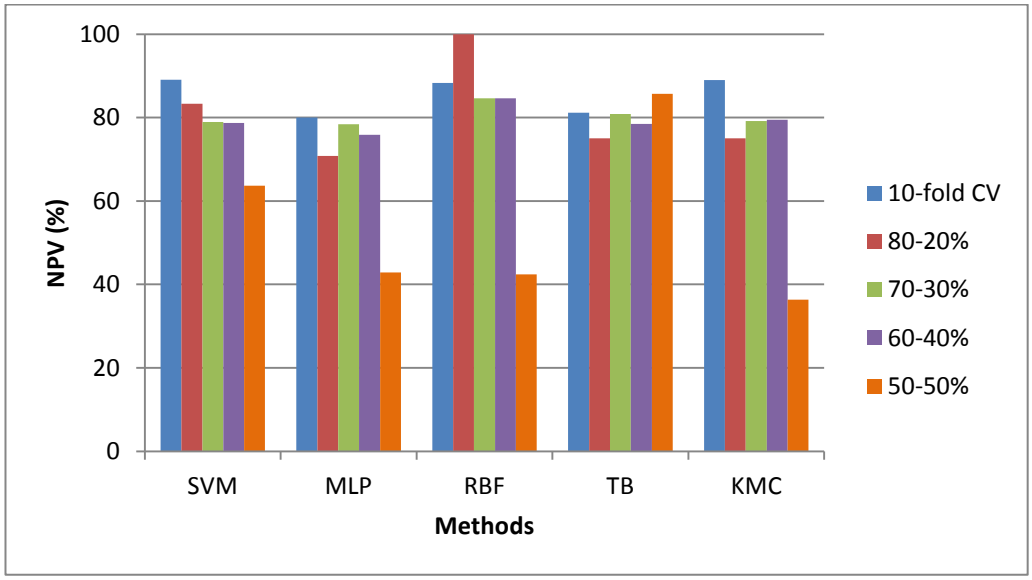


Figure 4.55. NPV's of prediction models using CFS on the 2006 dataset

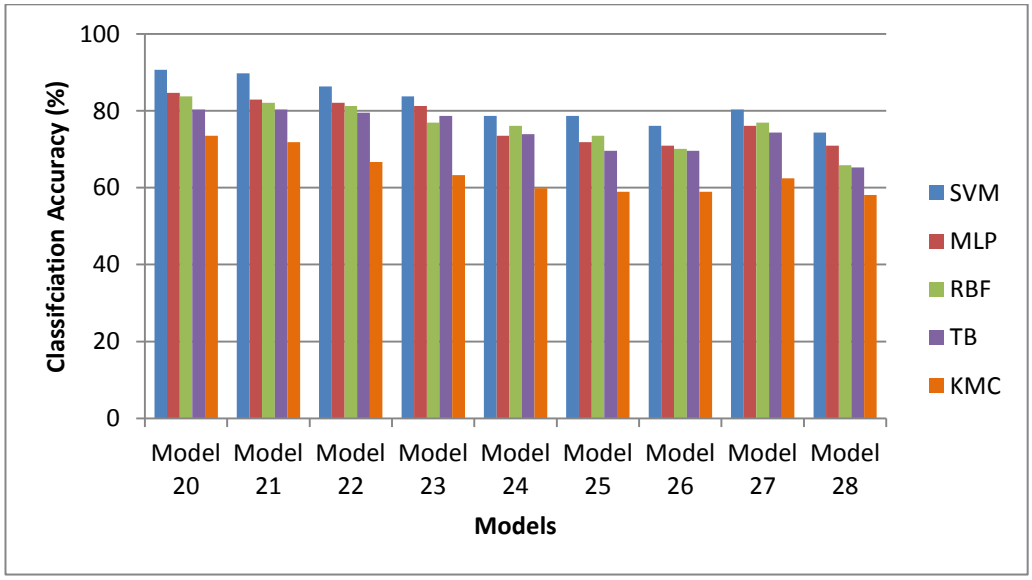


Figure 4.56. Classification accuracies of prediction models using Relief-F on the 2007 dataset (10-fold cross validation case)

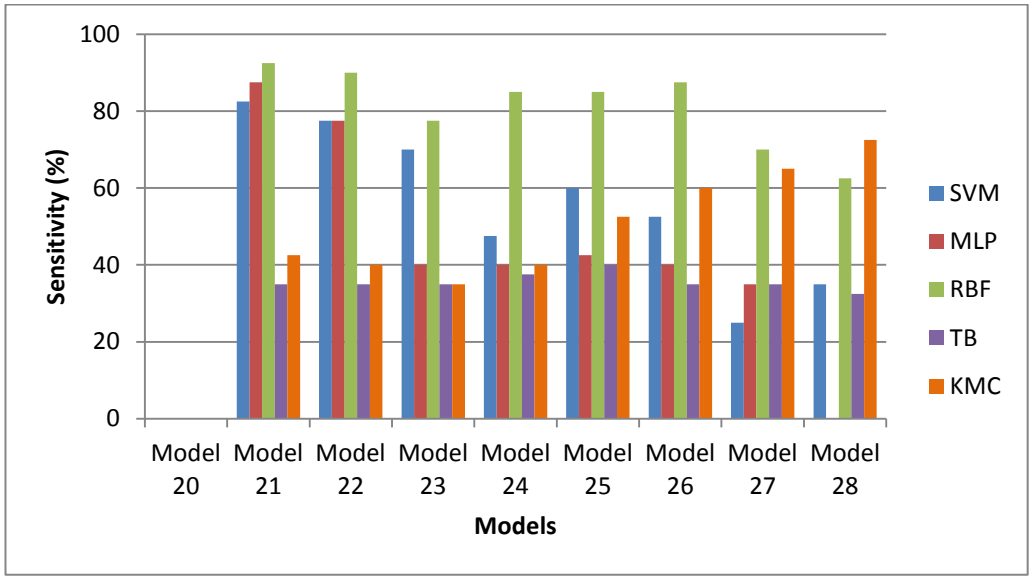


Figure 4.57. Sensitivities of prediction models using Relief-F on the 2007 dataset (10-fold cross validation case)

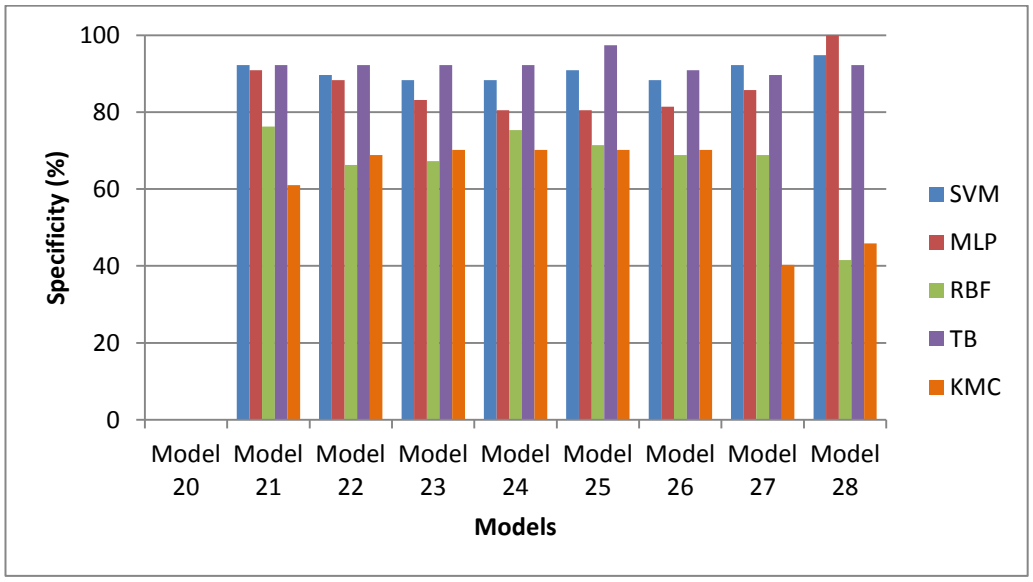


Figure 4.58. Specificities of prediction models using Relief-F on the 2007 dataset (10-fold cross validation case)

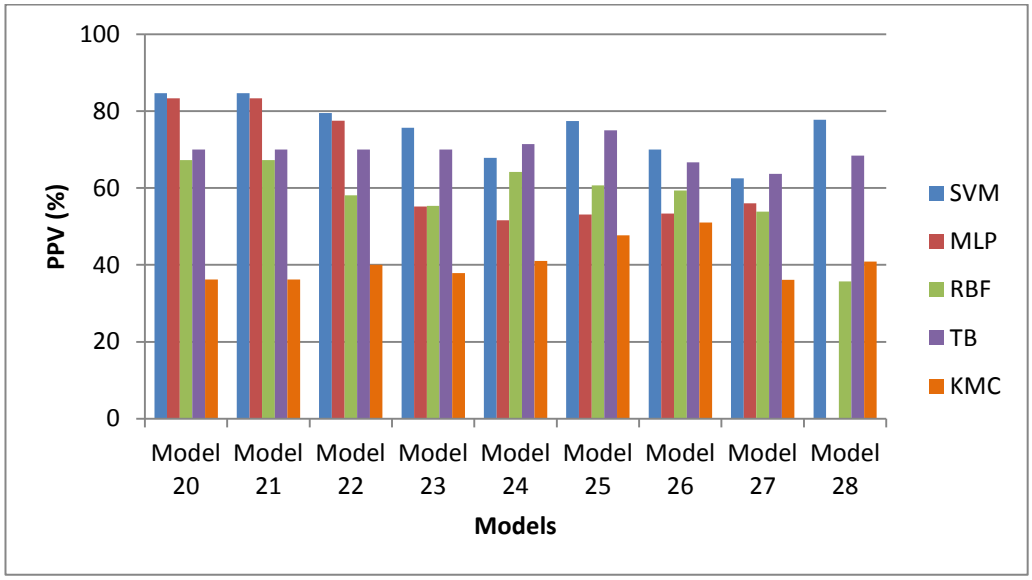


Figure 4.59. PPV'sof prediction models using Relief-F on the 2007 dataset (10-fold cross validation case)

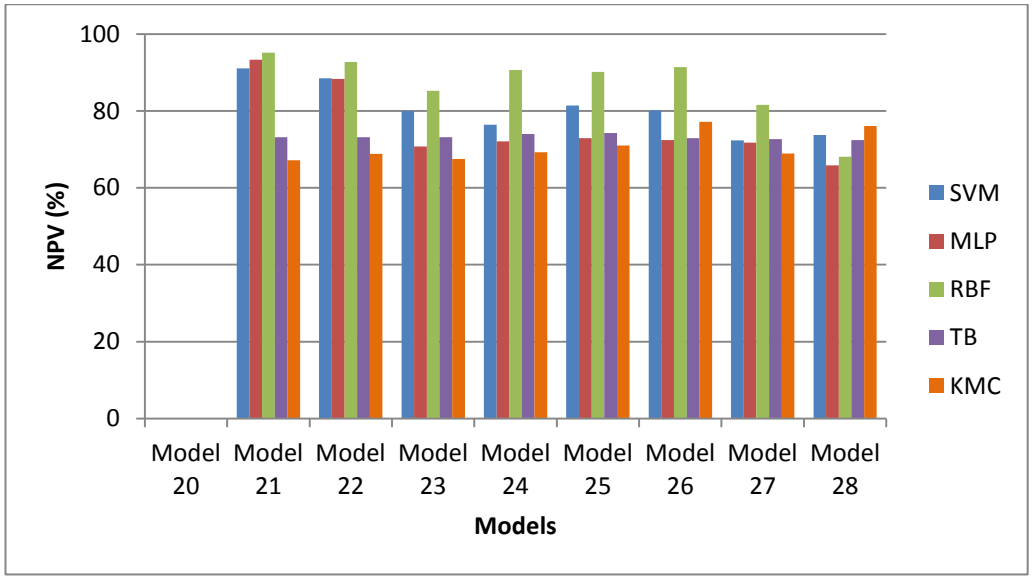


Figure 4.60. NPV'sof prediction models using Relief-F on the 2007 dataset (10-fold cross validation case)

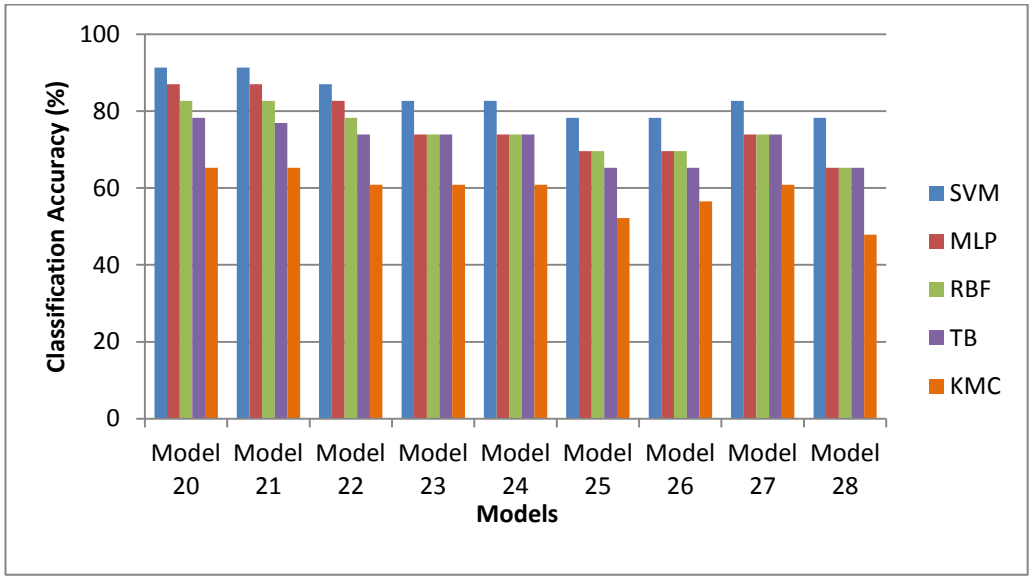


Figure 4.61. Classification accuracies of prediction models using Relief-F on the 2007 dataset (Percentage Split (%20) case)

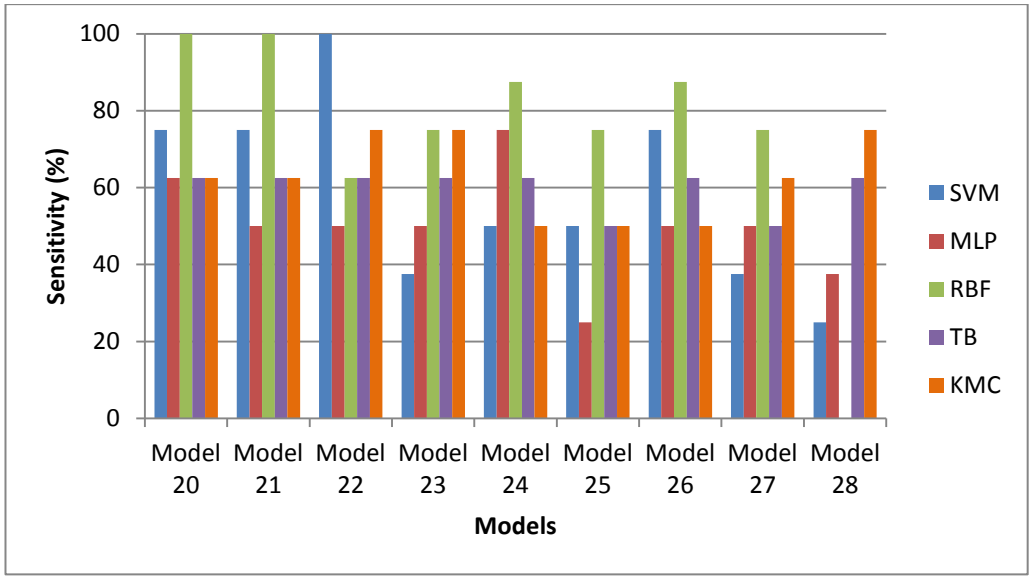


Figure 4.62. Sensitivities of prediction models using Relief-F on the 2007 dataset (Percentage Split (%20) case)

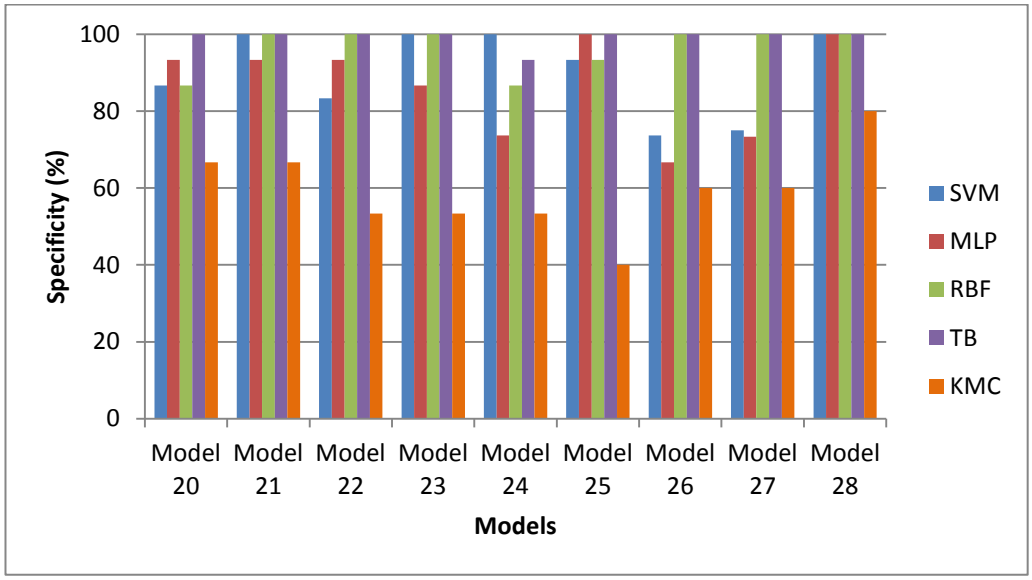


Figure 4.63. Specificities of prediction models using Relief-F on the 2007 dataset (Percentage Split (%20) case)

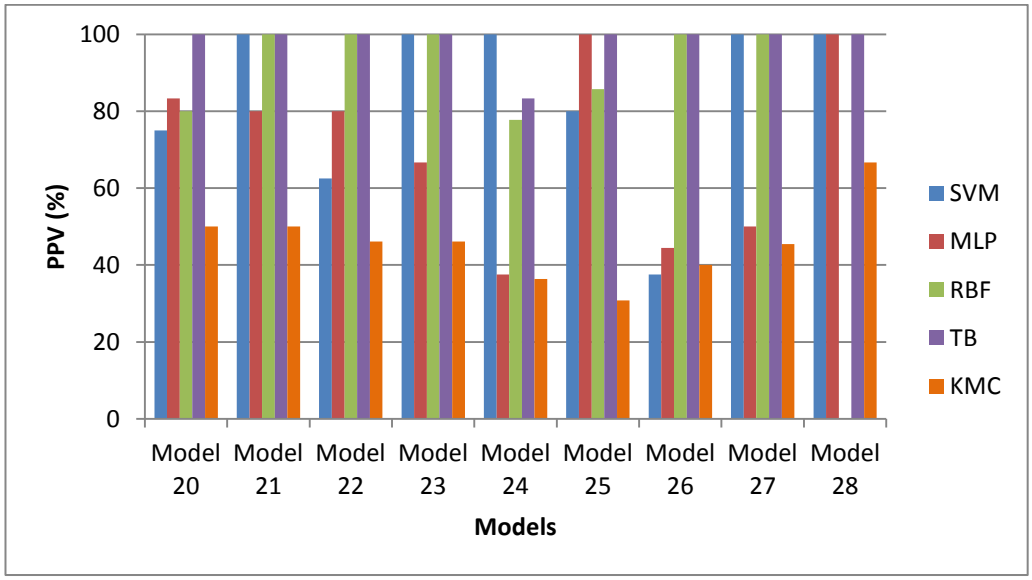


Figure 4.64. PPV'sof prediction models using Relief-F on the 2007 dataset (Percentage Split (%20) case)

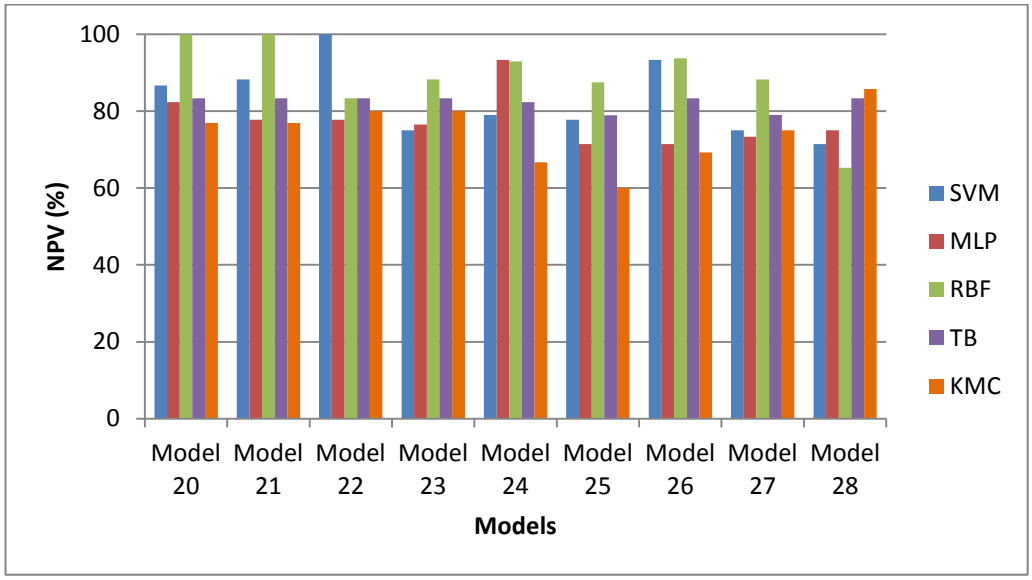


Figure 4.65. NPV'sof prediction models using Relief-F on the 2007 dataset (Percentage Split (%20) case)

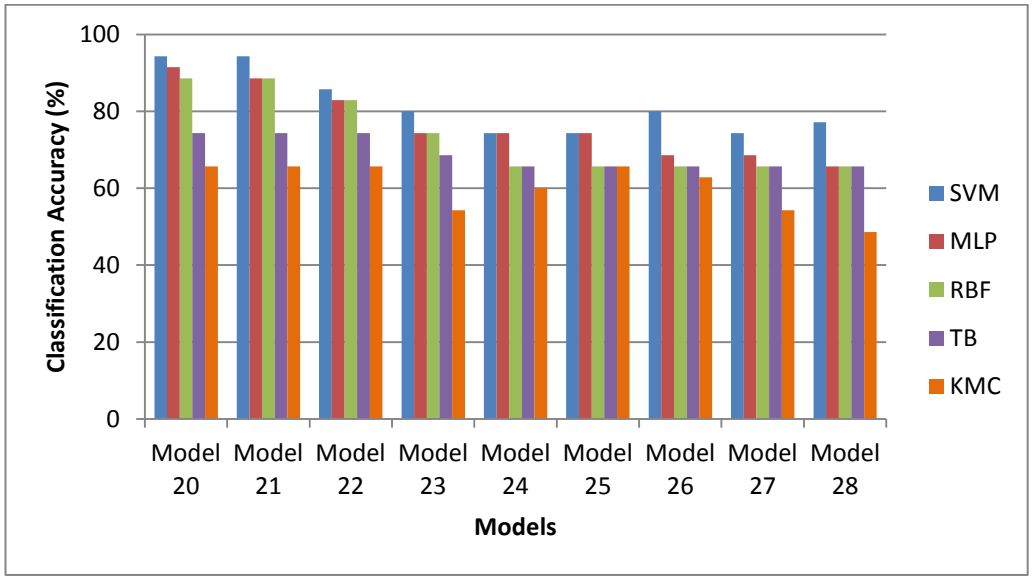


Figure 4.66. Classification accuracies of prediction models using Relief-F on the 2007 dataset (Percentage Split (%30) case)

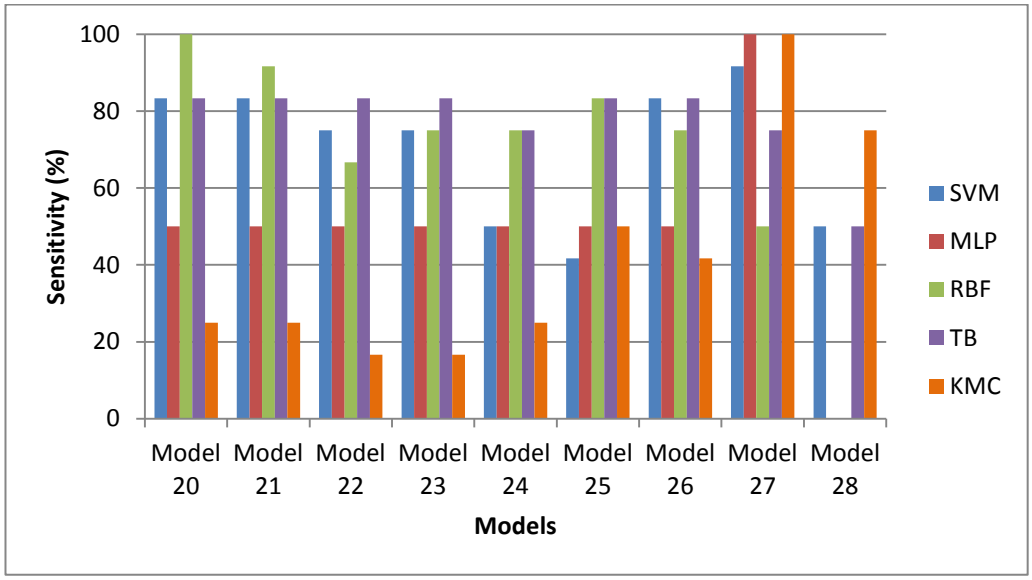


Figure 4.67. Sensitivities of prediction models using Relief-F on the 2007 dataset (Percentage Split (%30) case)

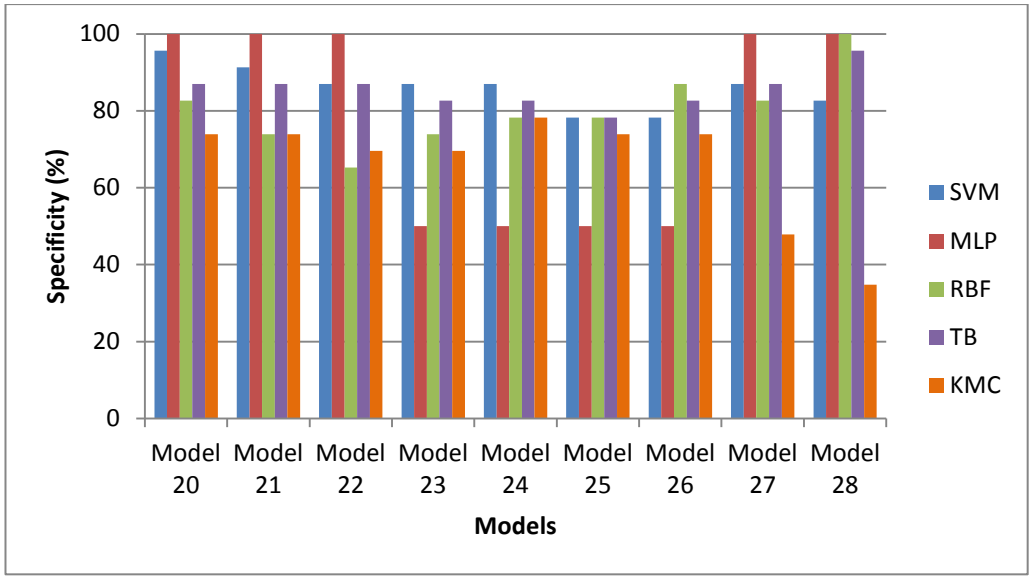


Figure 4.68. Specificities of prediction models using Relief-F on the 2007 dataset (Percentage Split (%30) case)

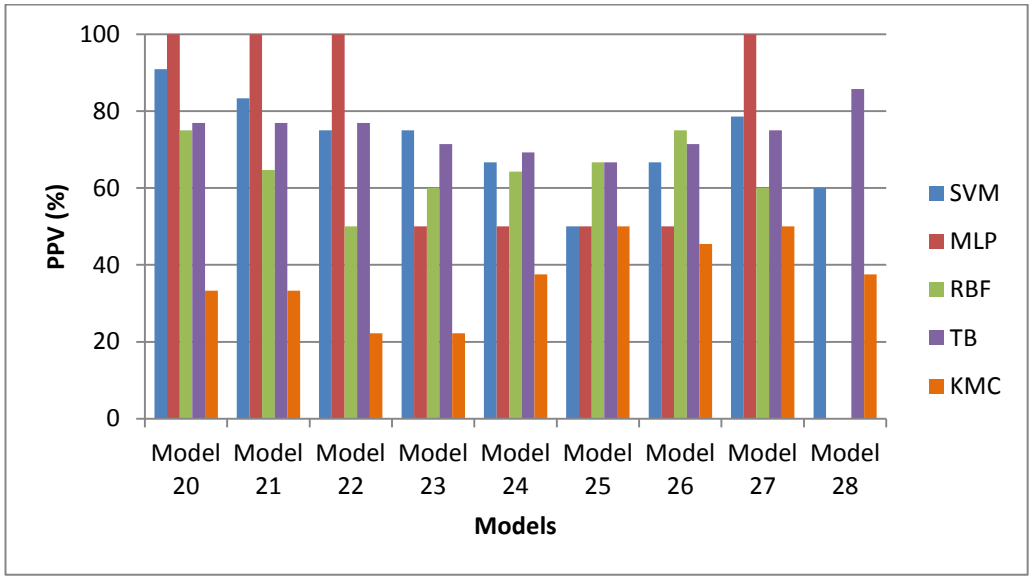


Figure 4.69. PPV'sof prediction models using Relief-F on the 2007 dataset (Percentage Split (%30) case)

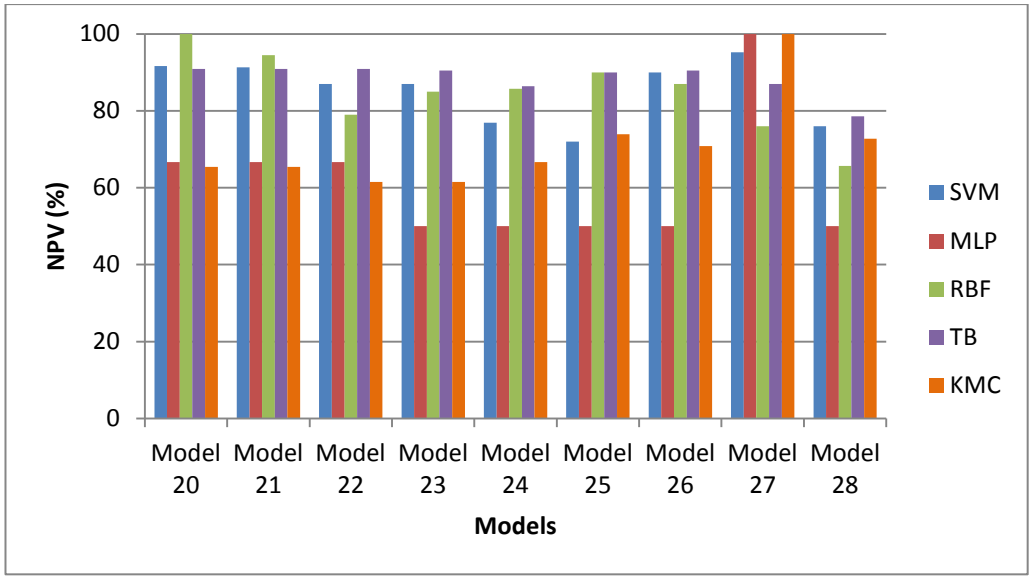


Figure 4.70. NPV'sof prediction models using Relief-F on the 2007 dataset (Percentage Split (%30) case)

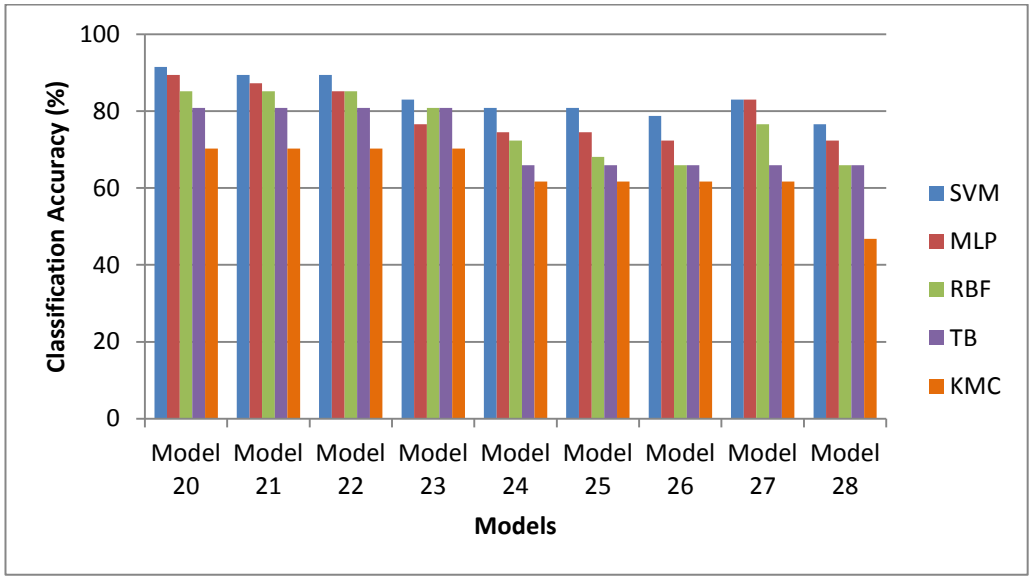


Figure 4.71. Classification accuracies of prediction models using Relief-F on the 2007 dataset (Percentage Split (%40) case)

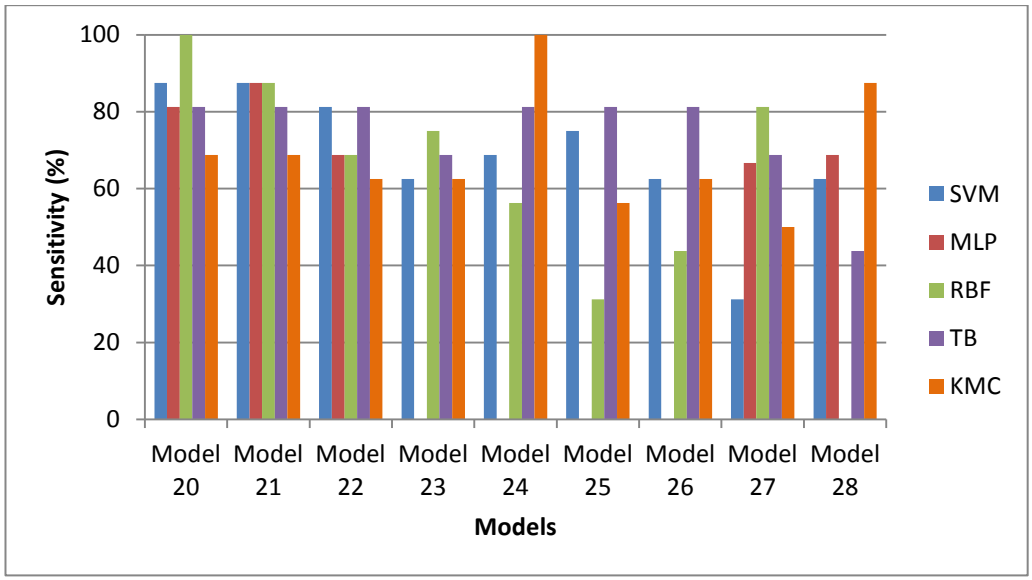


Figure 4.72. Sensitivities of prediction models using Relief-F on the 2007 dataset (Percentage Split (%40) case)

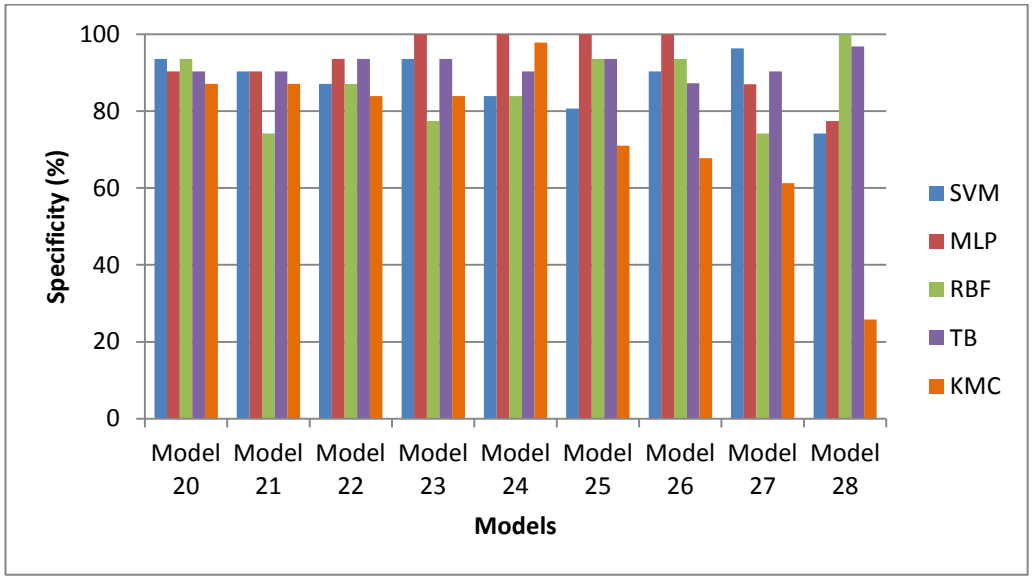


Figure 4.73. Specificities of prediction models using Relief-F on the 2007 dataset (Percentage Split (%40) case)

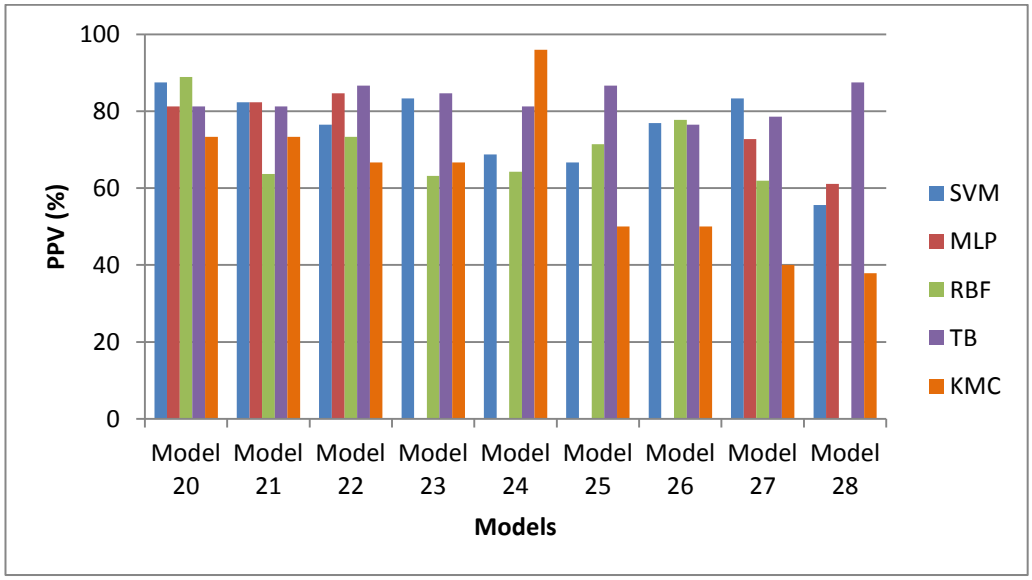


Figure 4.74. PPV'sof prediction models using Relief-F on the 2007 dataset (Percentage Split (%40) case)

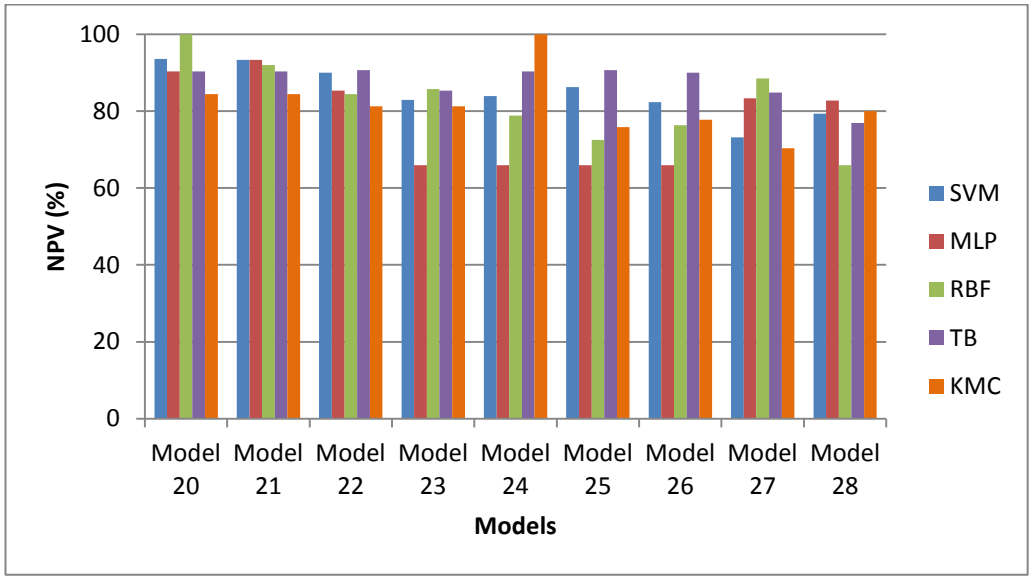


Figure 4.75. NPV'sof prediction models using Relief-F on the 2007 dataset (Percentage Split (%40) case)

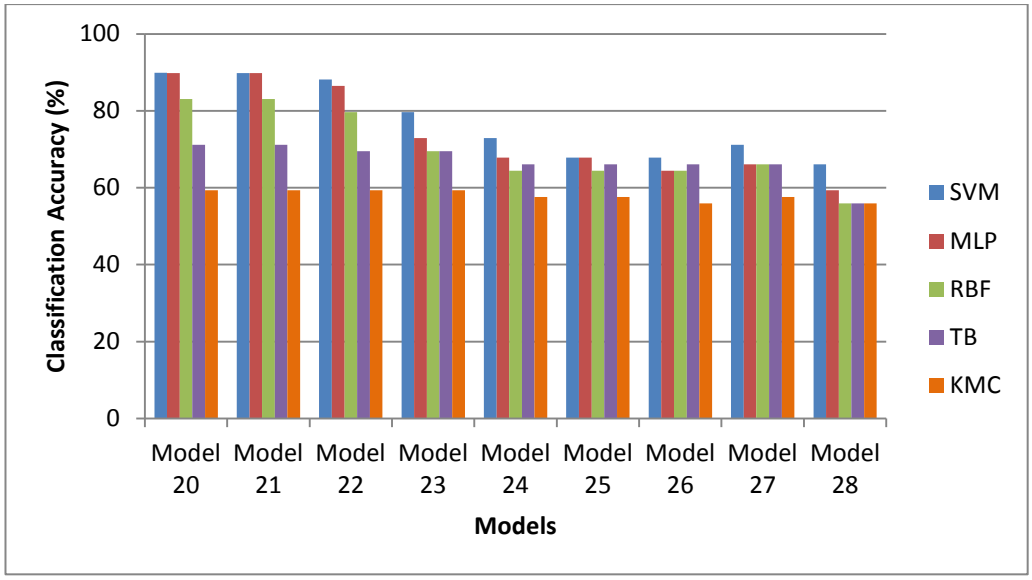


Figure 4.76. Classification accuracies of prediction models using Relief-F on the 2007 dataset (Percentage Split (%50) case)

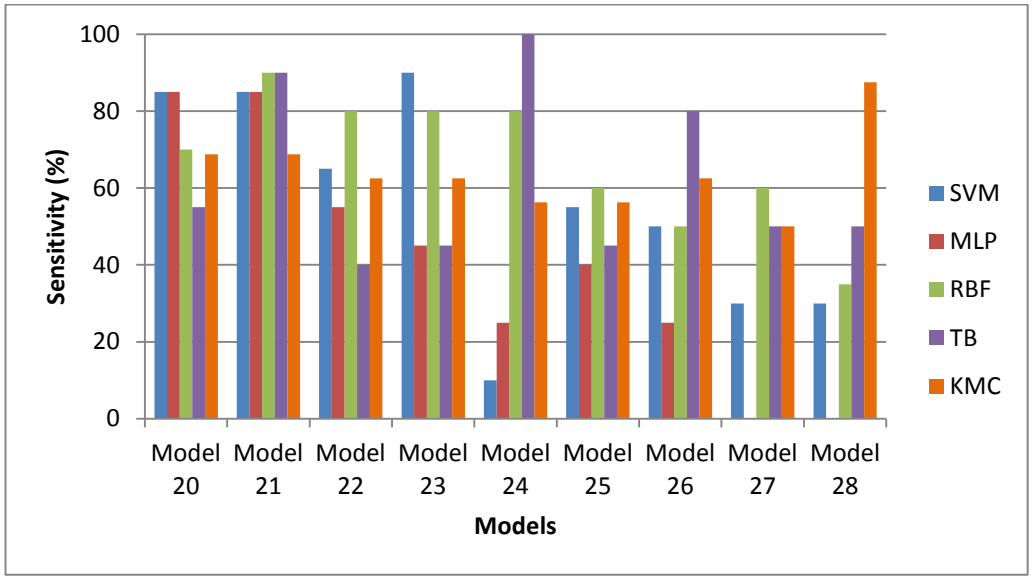


Figure 4.77. Sensitivities of prediction models using Relief-F on the 2007 dataset (Percentage Split (%50) case)

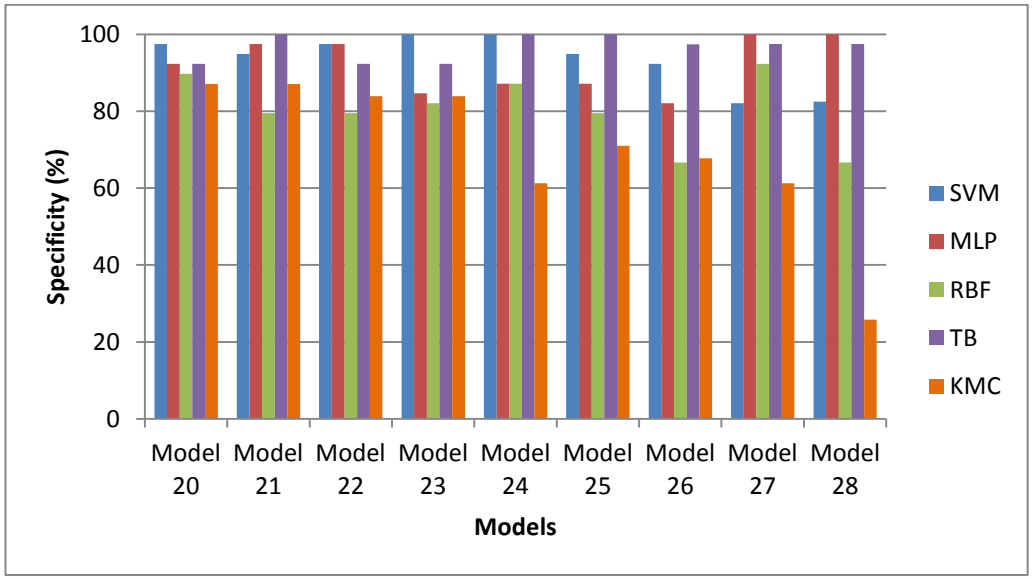


Figure 4.78. Specificities of prediction models using Relief-F on the 2007 dataset (Percentage Split (%50) case)

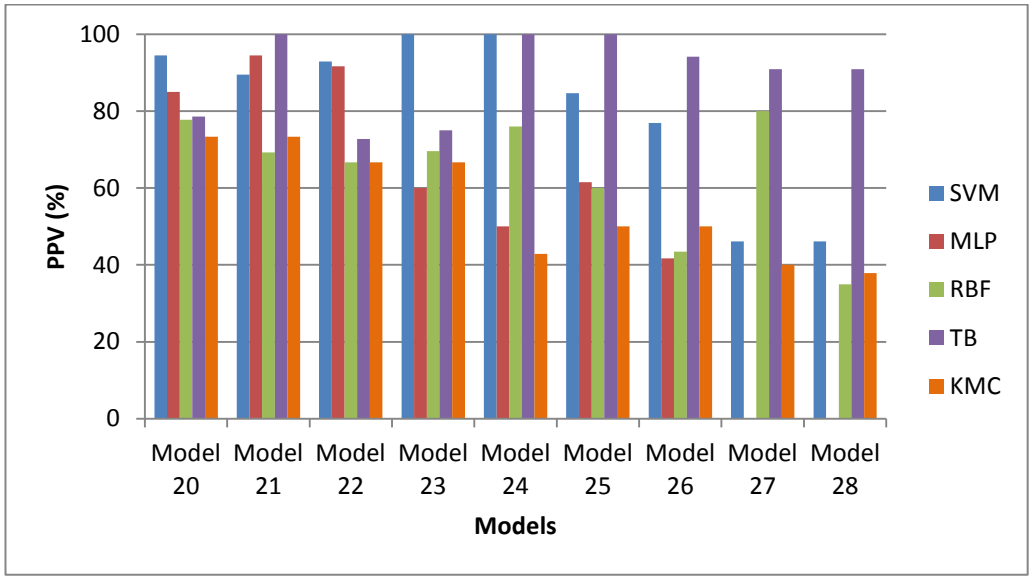


Figure 4.79. PPV'sof prediction models using Relief-F on the 2007 dataset (Percentage Split (%50) case)

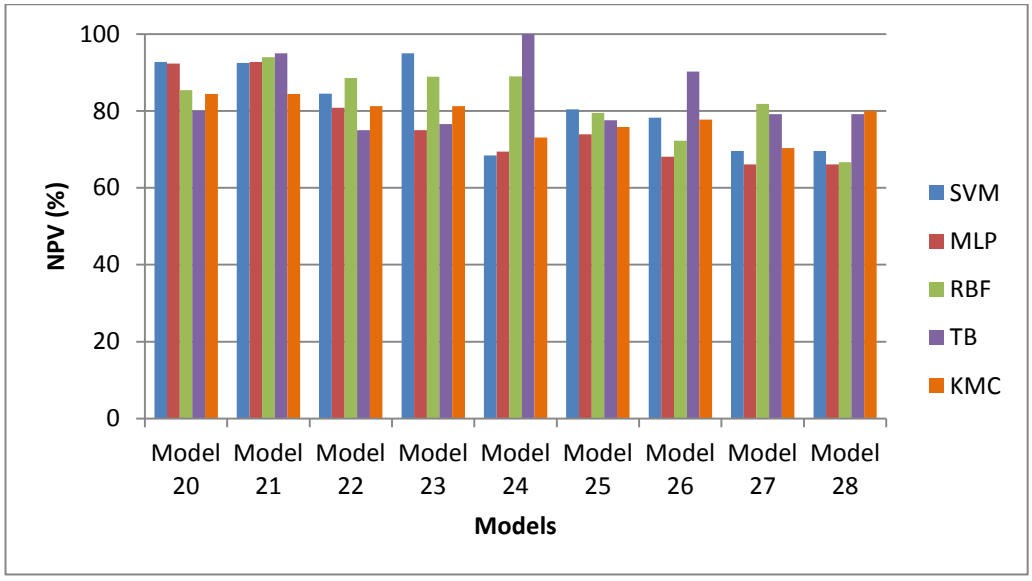


Figure 4.80. NPV'sof prediction models using Relief-F on the 2007 dataset (Percentage Split (%50) case)

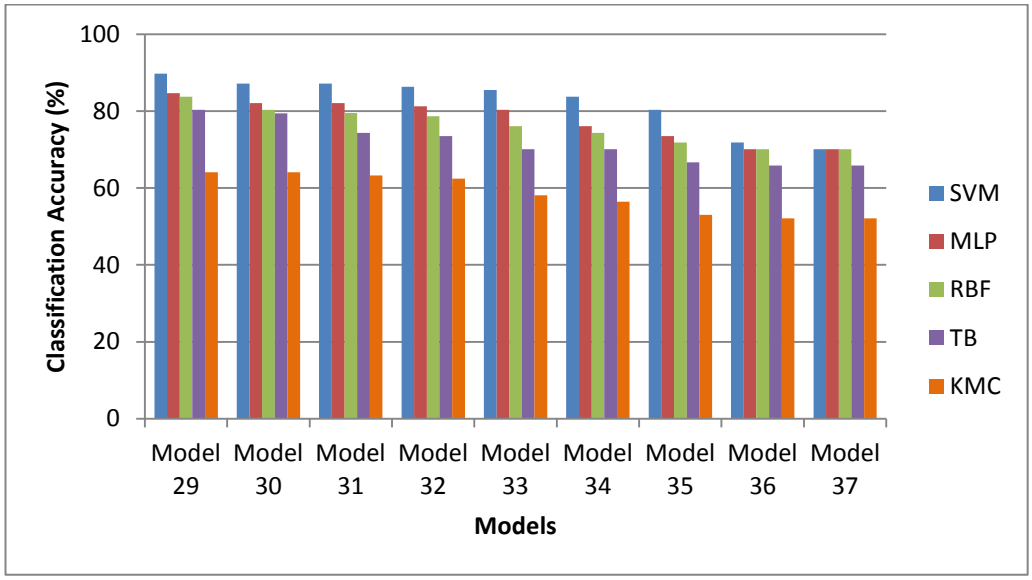


Figure 4.81. Classification accuracies of prediction models using F-Score on the 2007 dataset (10-fold cross validation)

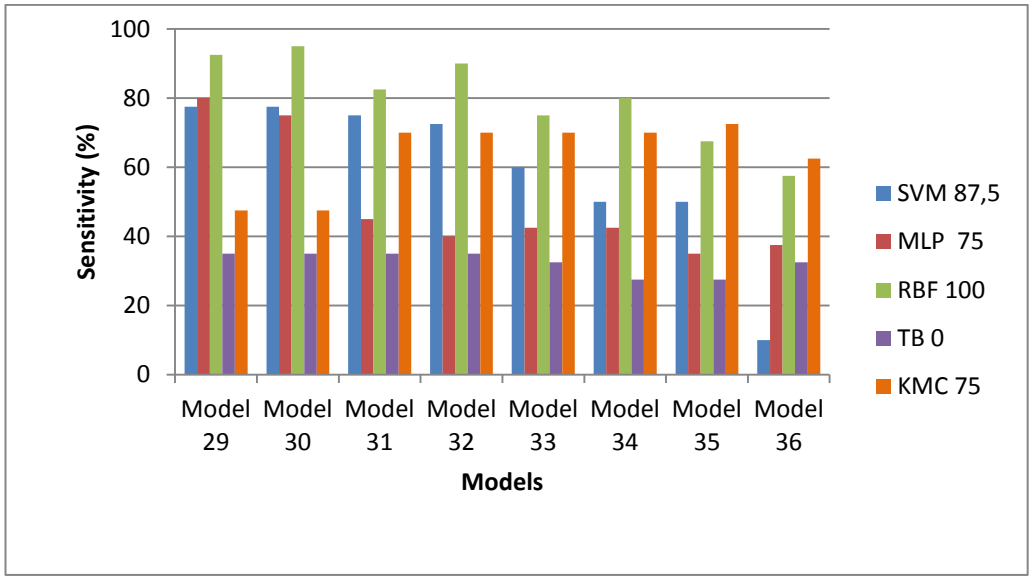


Figure 4.82. Sensitivities of prediction models using F-Score on the 2007 dataset (10-fold cross validation)

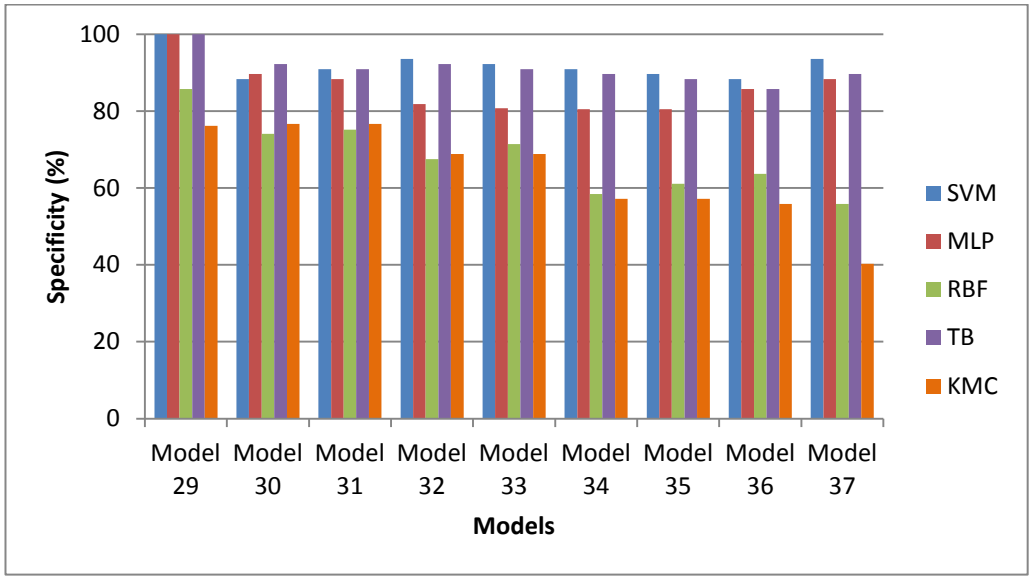


Figure 4.83. Specificities of prediction models using F-Score on the 2007 dataset (10-fold cross validation)

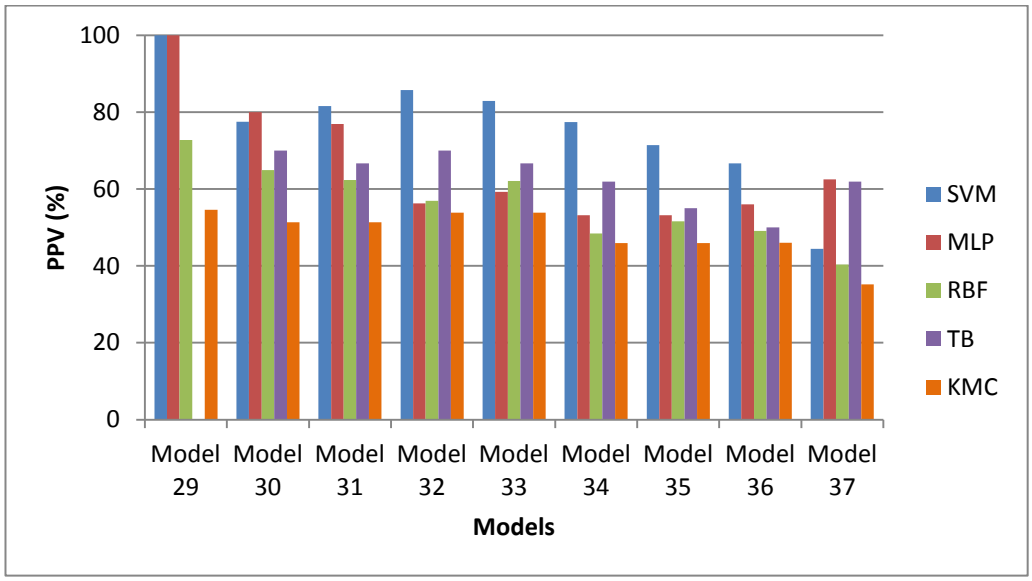


Figure 4.84. PPV's of prediction models using F-Score on the 2007 dataset (10-fold cross validation)

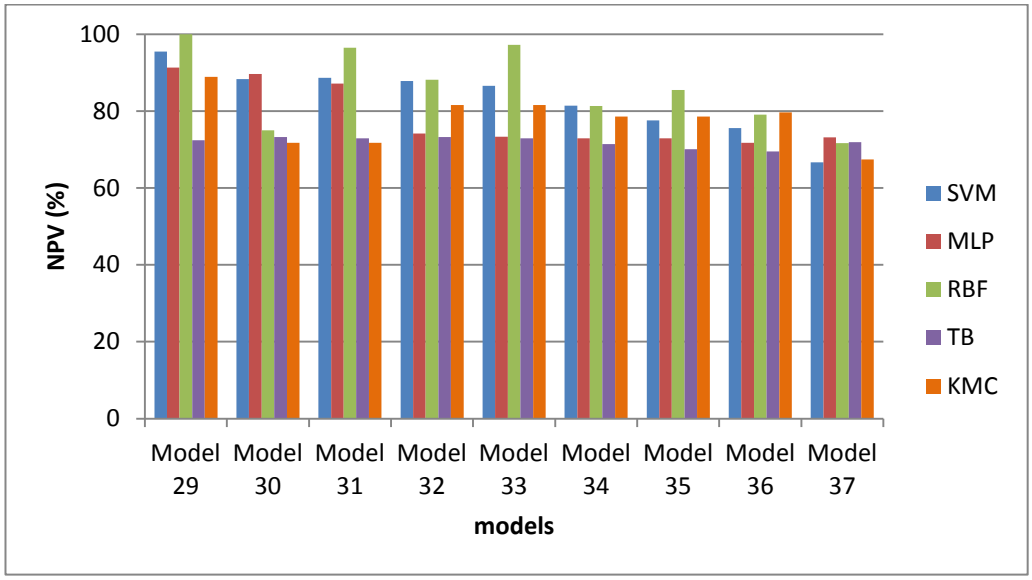


Figure 4.85. NPV's of prediction models using F-Score on the 2007 dataset (10-fold cross validation)

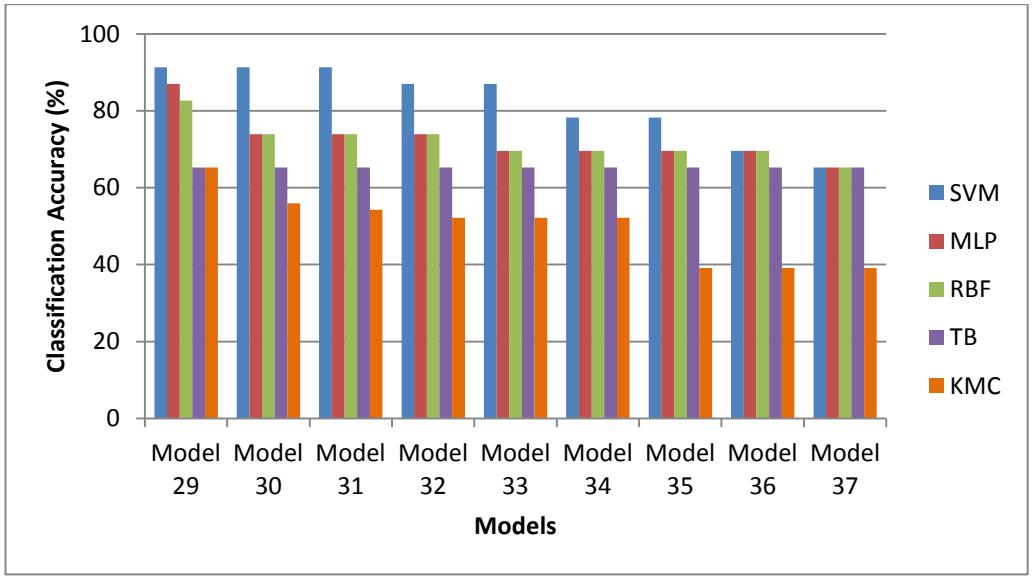


Figure 4.86. Classification accuracies of prediction models using F-Score on the 2007 dataset (Percentage Split (%20) case)

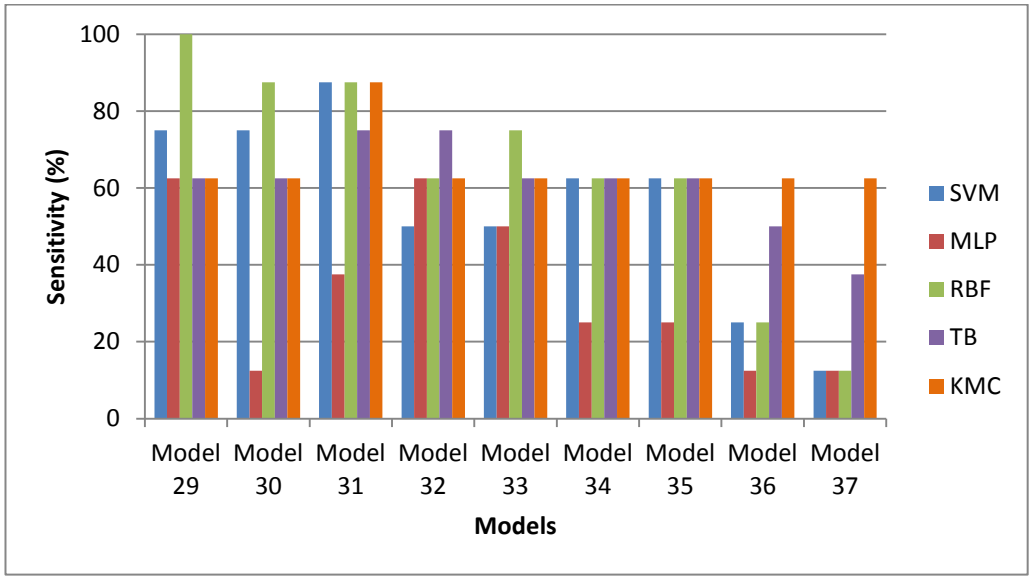


Figure 4.87. Sensitivities of prediction models using F-Score on the 2007 dataset (Percentage Split (%20) case)

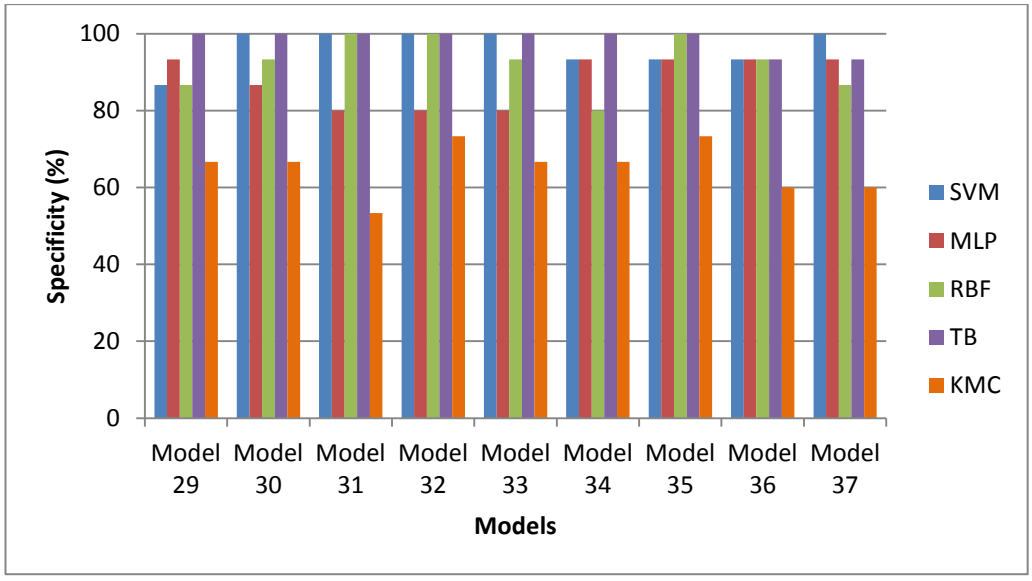


Figure 4.88. Specificities of prediction models using F-Score on the 2007 dataset (Percentage Split (%20) case)

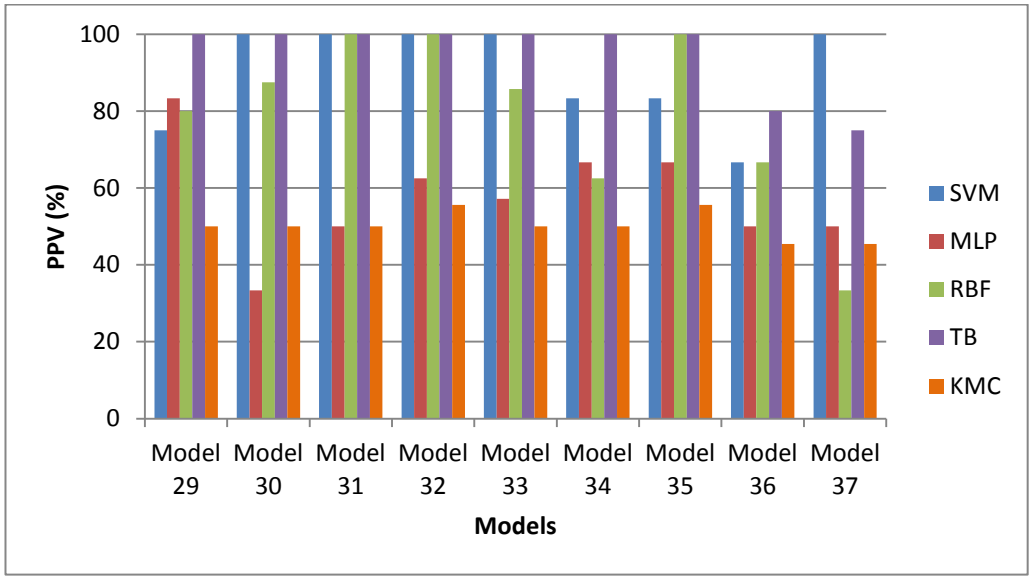


Figure 4.89. PPV's of prediction models using F-Score on the 2007 dataset (Percentage Split (%20) case)

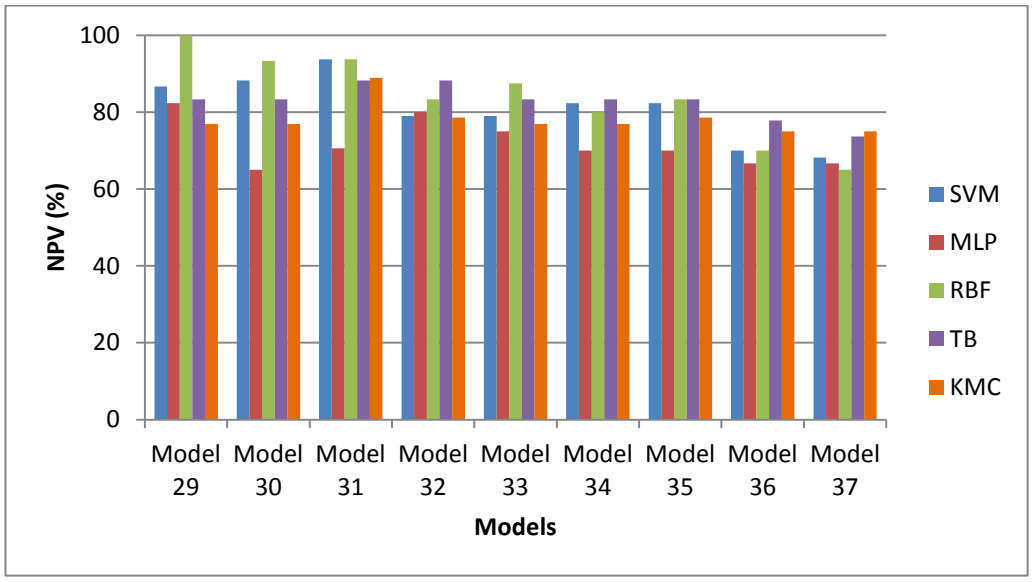


Figure 4.90. NPV's of prediction models using F-Score on the 2007 dataset (Percentage Split (%20) case)

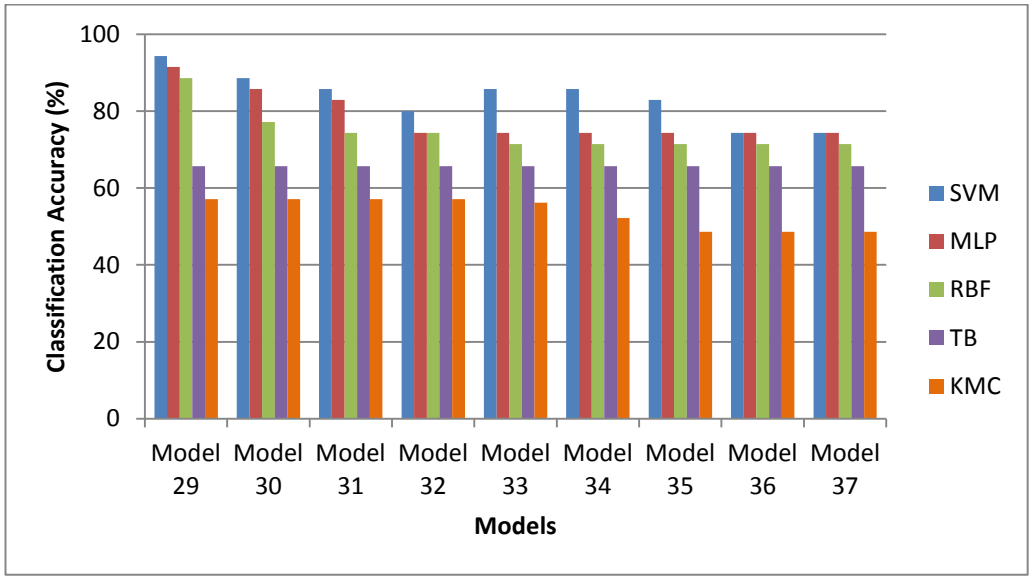


Figure 4.91. Classification accuracies of prediction models using F-Score on the 2007 dataset (Percentage Split (%30) case)

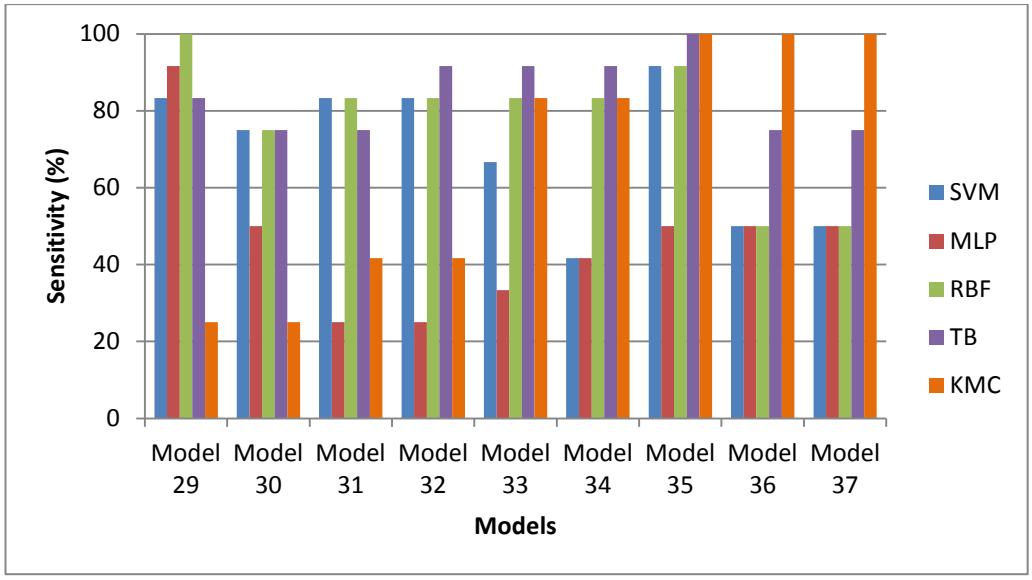


Figure 4.92. Sensitivities of prediction models using F-Score on the 2007 dataset (Percentage Split (%30) case)

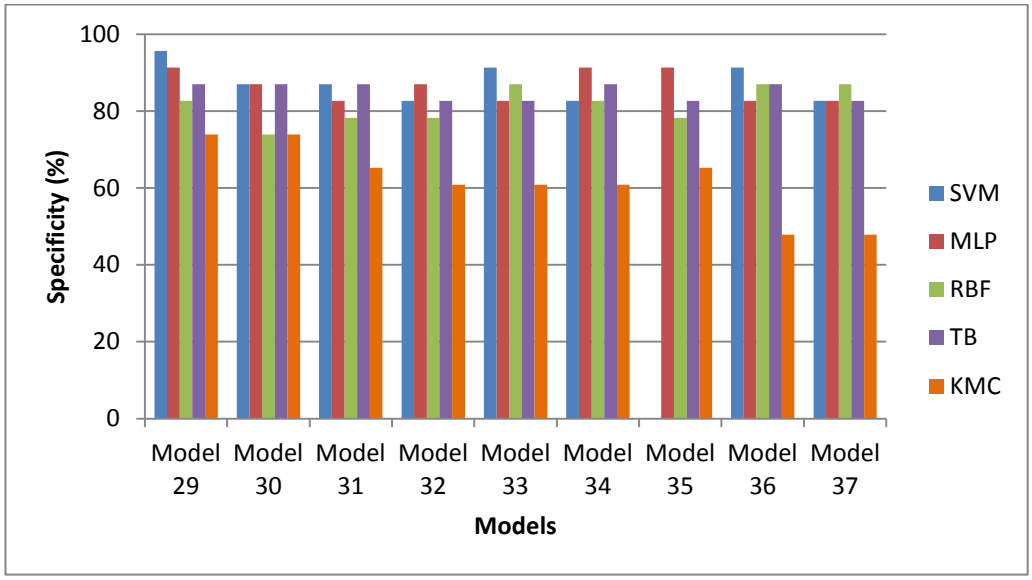


Figure 4.93. Specificitiesof prediction models using F-Score on the 2007 dataset (Percentage Split (%30) case)

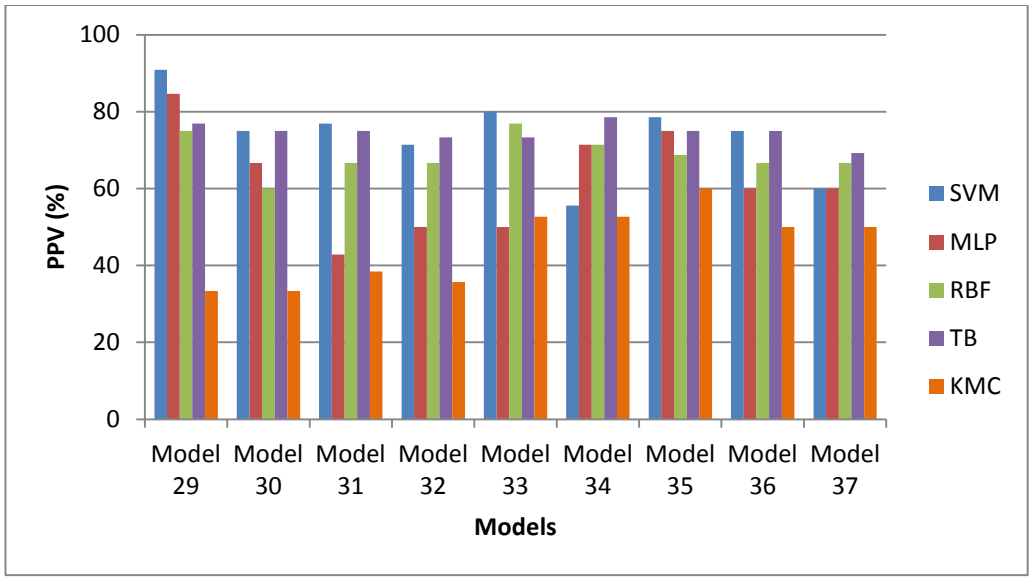


Figure 4.94. PPV's of prediction models using F-Score on the 2007 dataset (Percentage Split (%30) case)

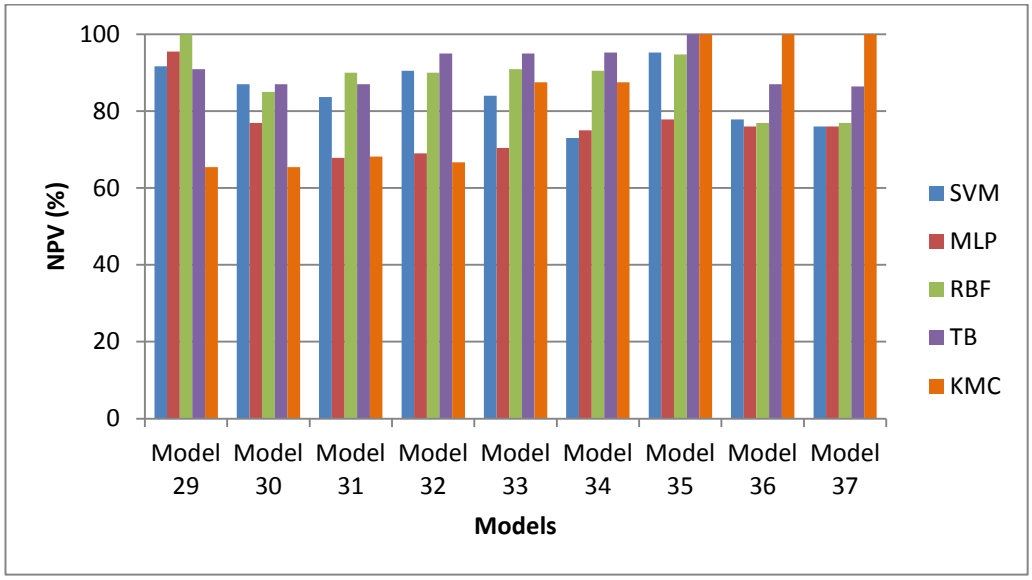


Figure 4.95. NPV's of prediction models using F-Score on the 2007 dataset (Percentage Split (%30) case)

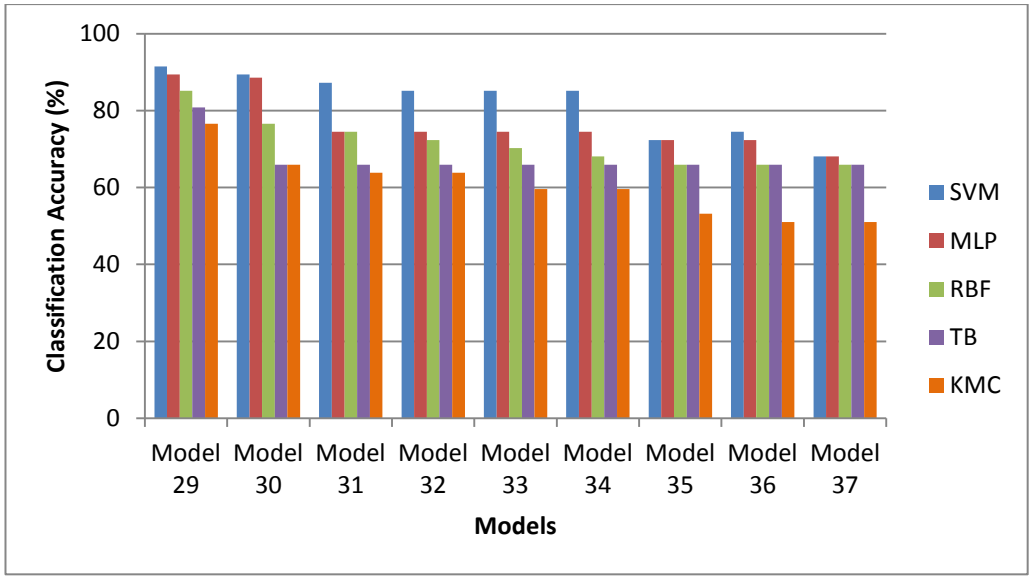


Figure 4.96. Classification accuracies of prediction models using F-Score on the 2007 dataset (Percentage Split (%40) case)

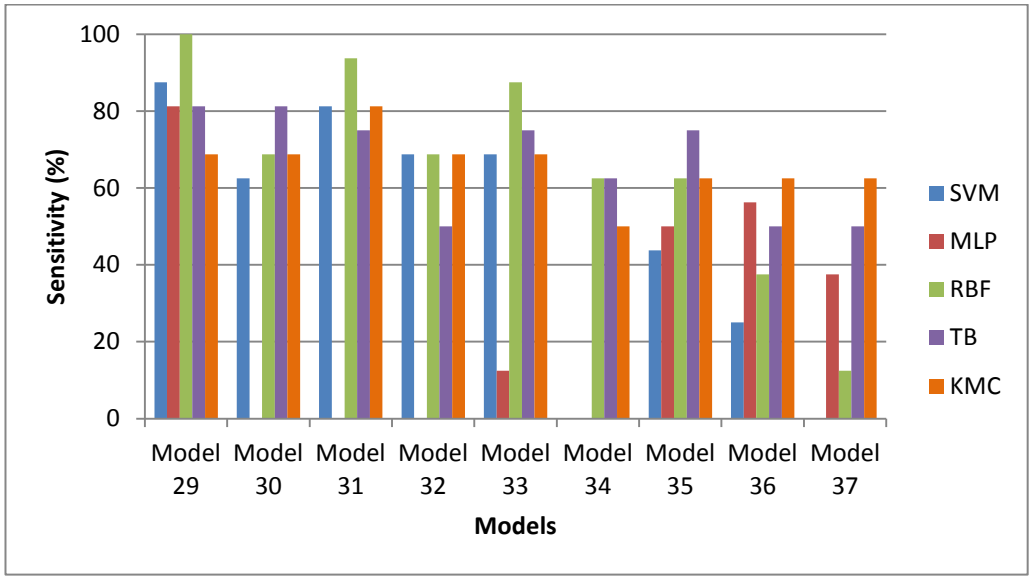


Figure 4.97. Sensitivities of prediction models using F-Score on the 2007 dataset (Percentage Split (%40) case)

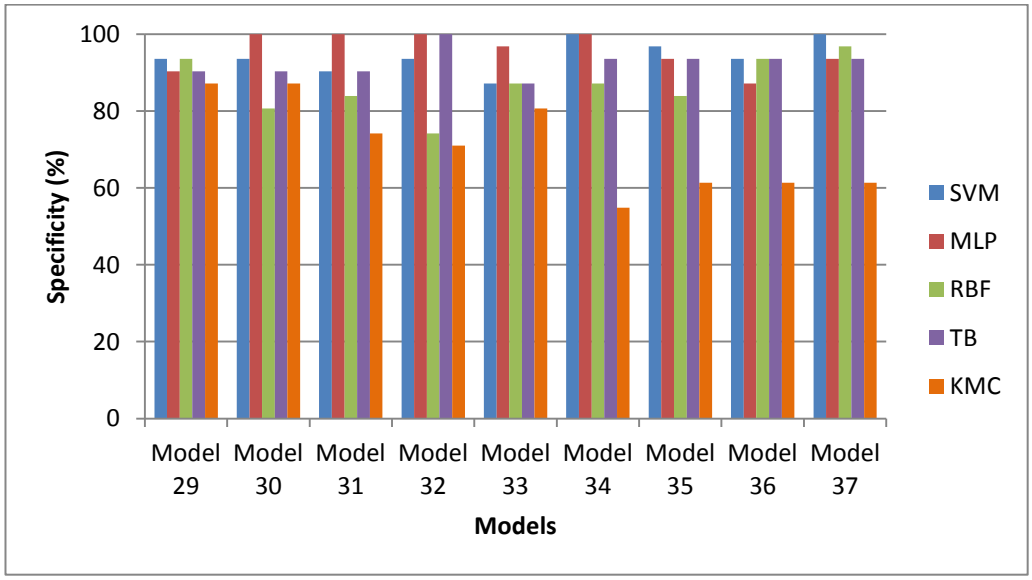


Figure 4.98. Specificities of prediction models using F-Score on the 2007 dataset (Percentage Split (%40) case)

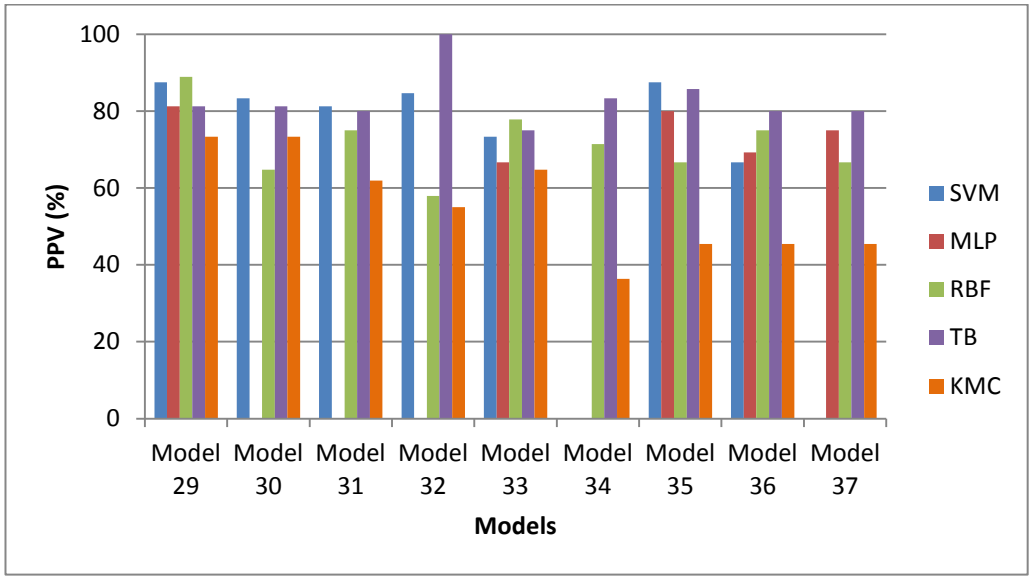


Figure 4.99. PPV's of prediction models using F-Score on the 2007 dataset (Percentage Split (%40) case)

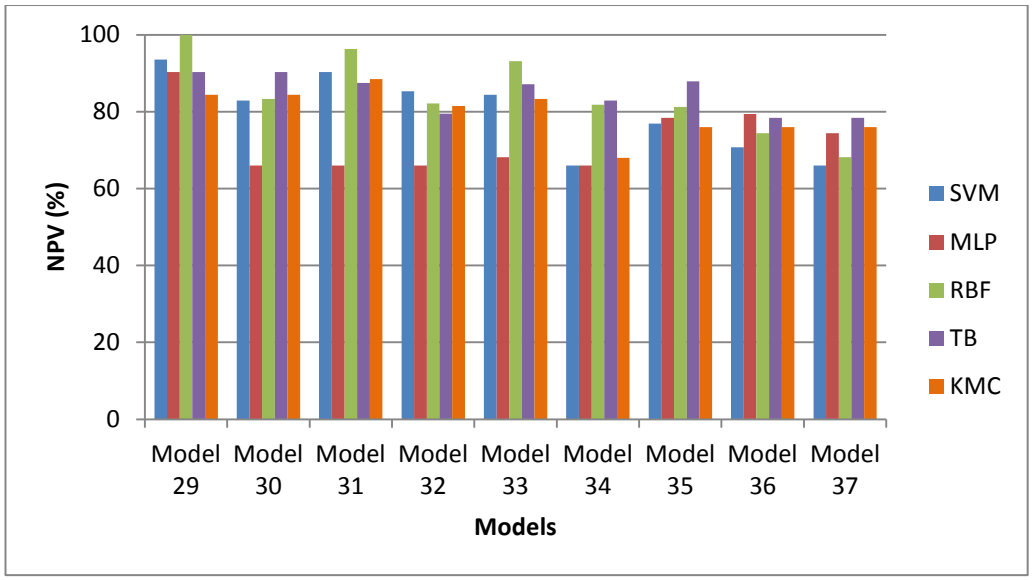


Figure 4.100. NPV's of prediction models using F-Score on the 2007 dataset (Percentage Split (%40) case)

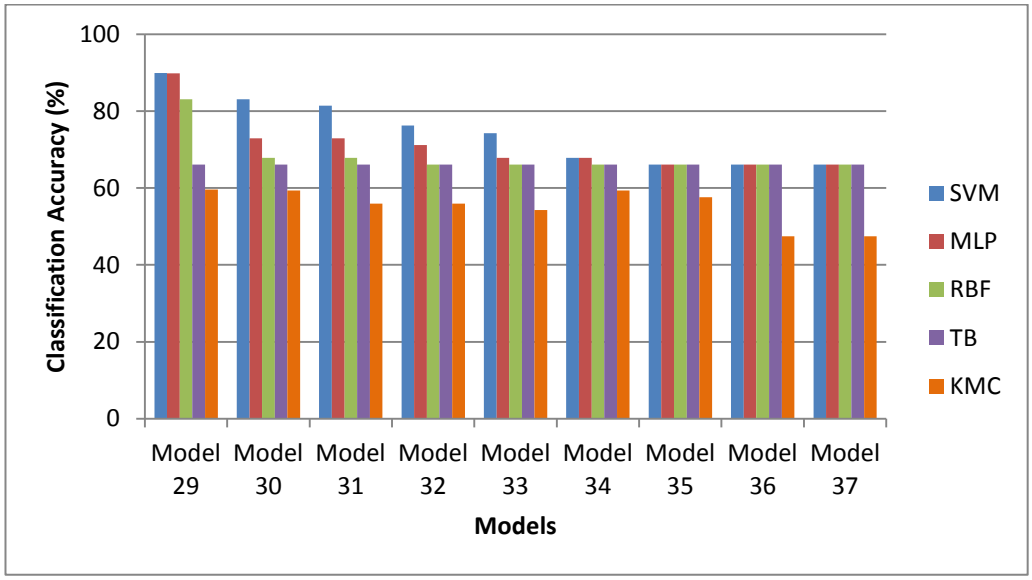


Figure 4.101. Classification accuracies of prediction models using F-Score on the 2007 dataset (Percentage Split (%50) case)

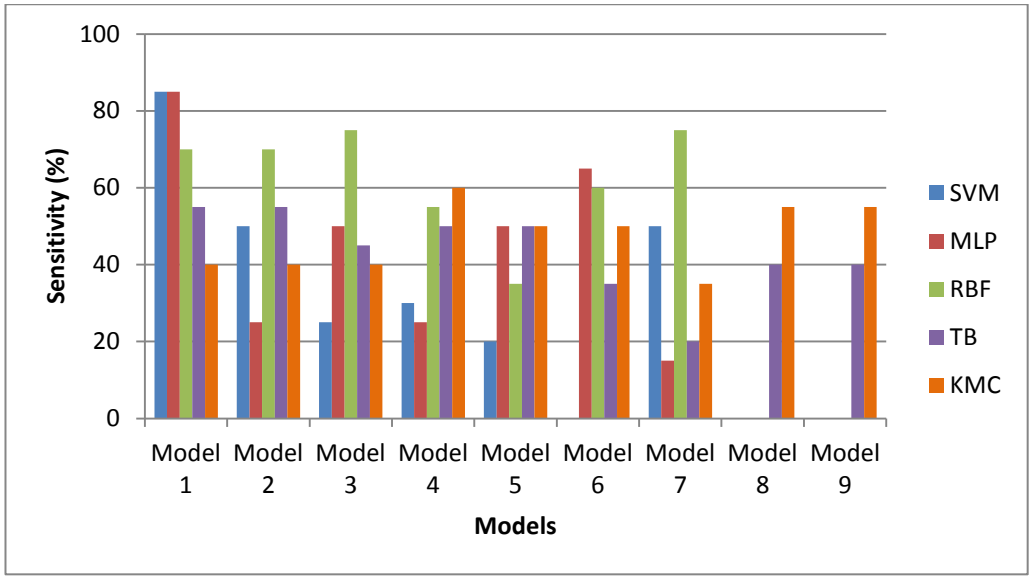


Figure 4.102. Sensitivities of prediction models using F-Score on the 2007 dataset (Percentage Split (%50) case)

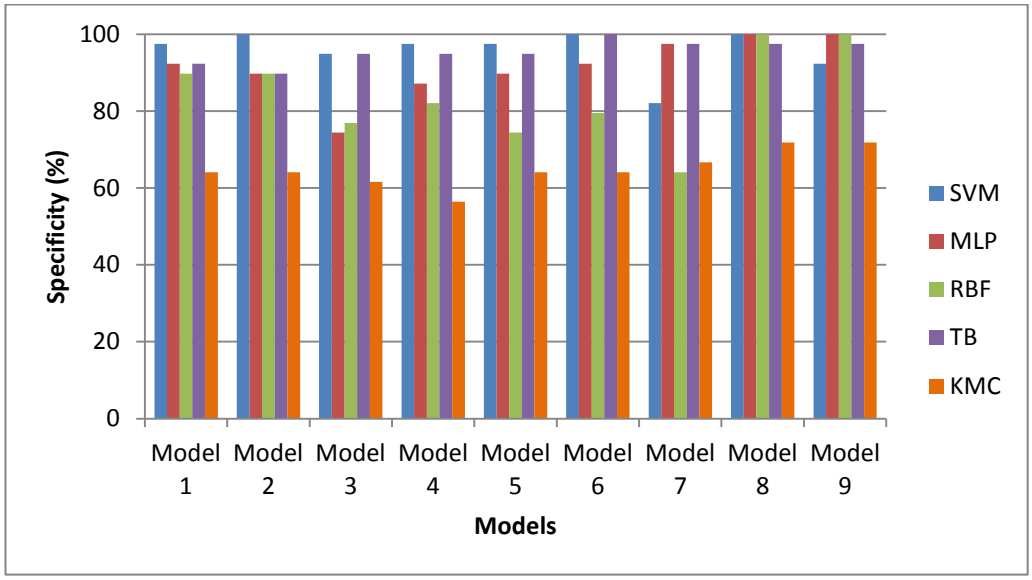


Figure 4.103. Specificities of prediction models using F-Score on the 2007 dataset (Percentage Split (%50) case)

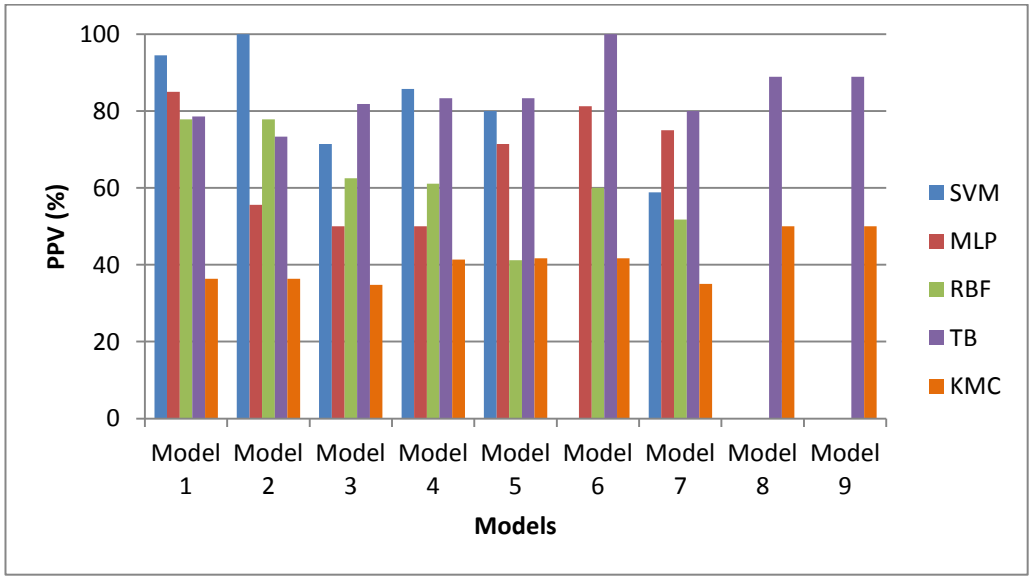


Figure 4.104. PPV's of prediction models using F-Score on the 2007 dataset (Percentage Split (%50) case)

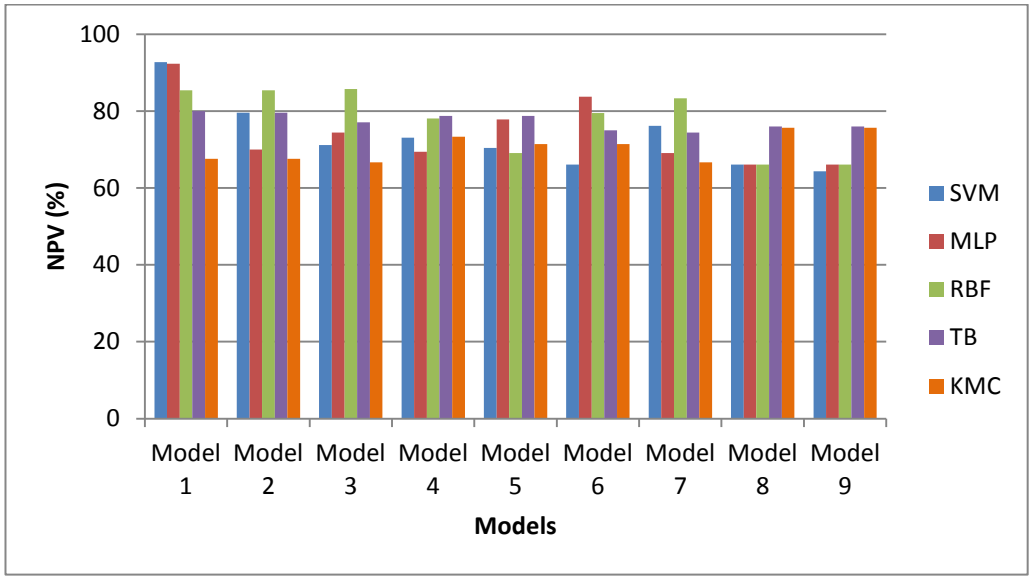


Figure 4.105. NPV's of prediction models using F-Score on the 2007 dataset (Percentage Split (%50) case)

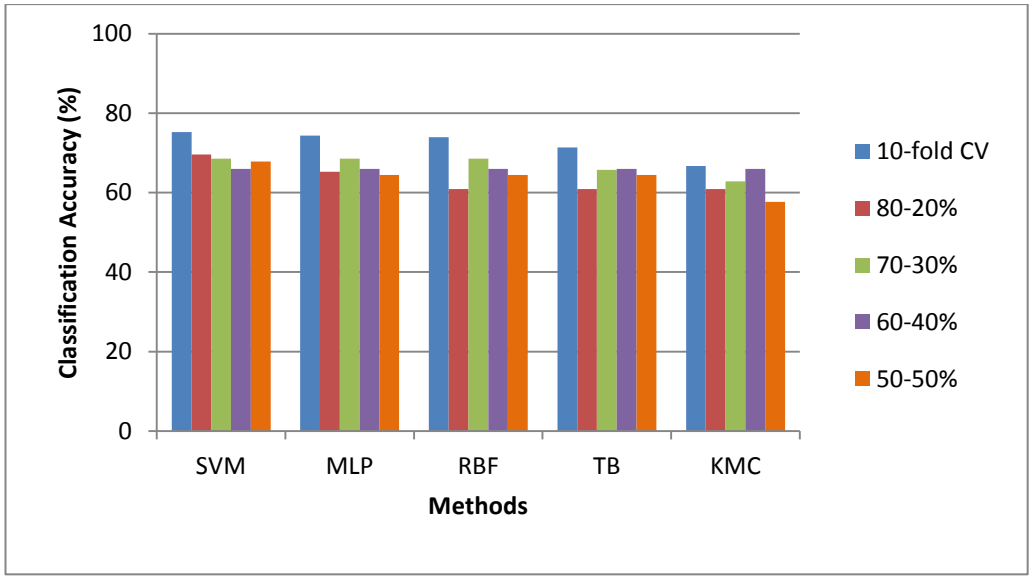


Figure 4.106. Classification accuracies of prediction models using CFS on the 2007 dataset

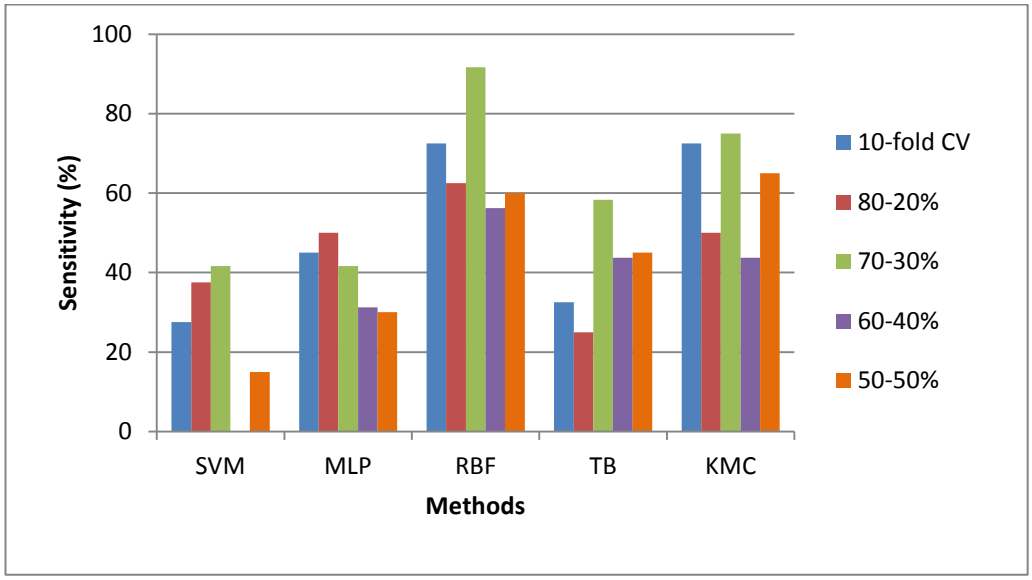


Figure 4.107. Sensitivities of prediction models using CFS on the 2007 dataset

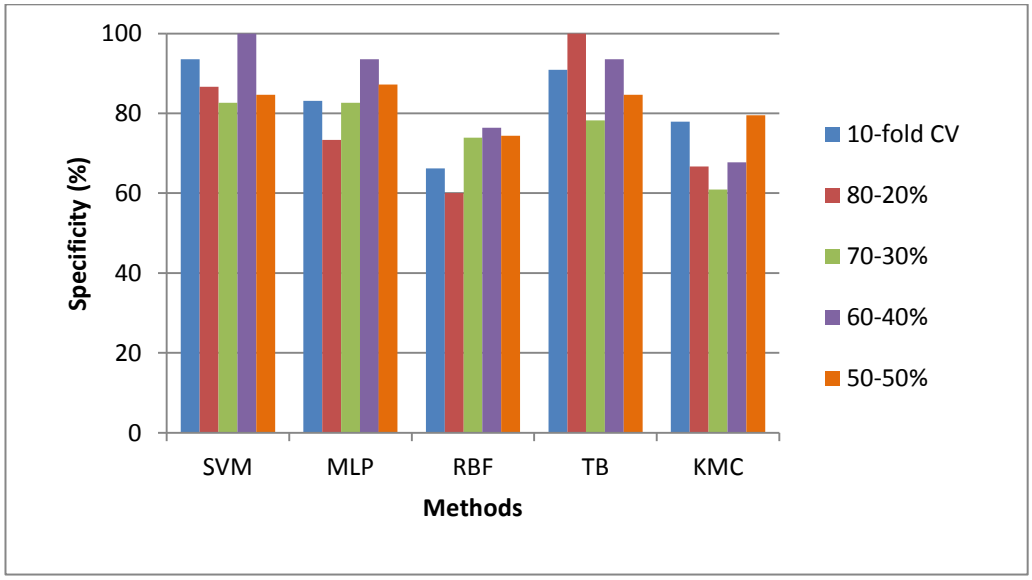


Figure 4.108. Specificities of prediction models using CFS on the 2007 dataset

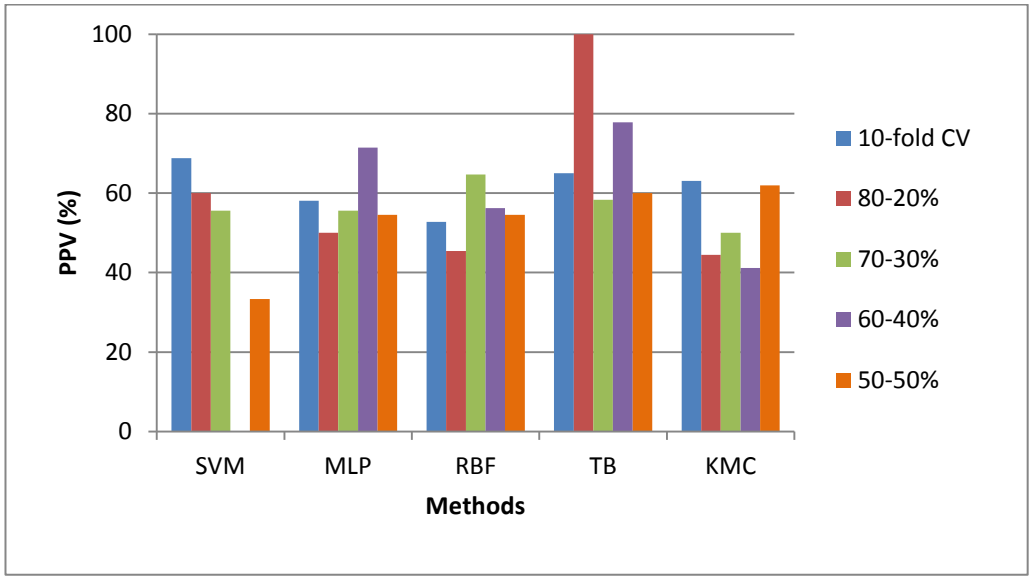


Figure 4.109. PPV's of prediction models using CFS on the 2007 dataset

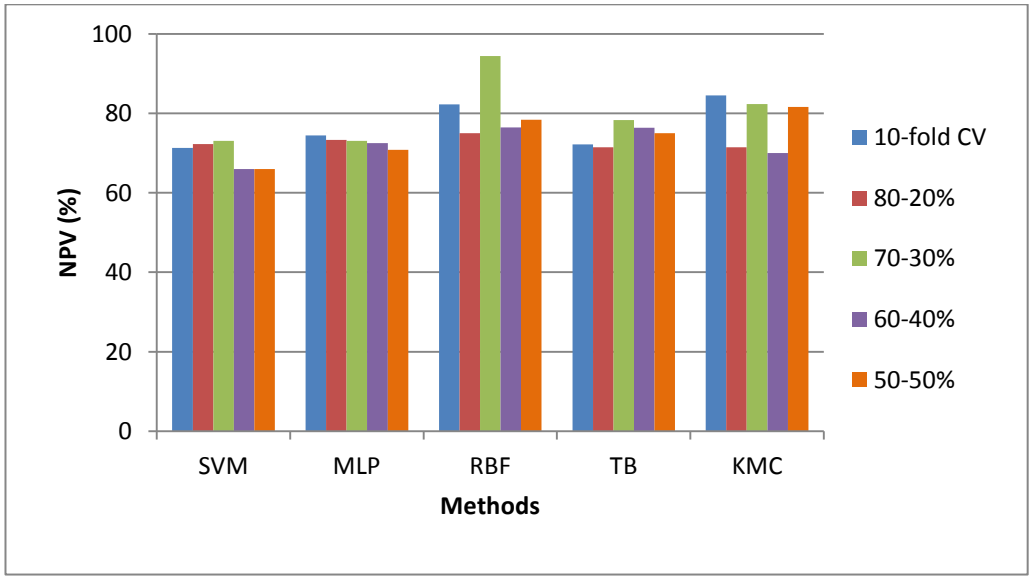


Figure 4.110. NPV's of prediction models using CFS on the 2007 dataset

Figure 4.111 through Figure 4.140 show the average classification accuracies, average senstivities, average specificities, average PPV's and average NPV's's of all prediction models using Relief-F, F-Score and CFS on the 2006 and 2007 datasets.

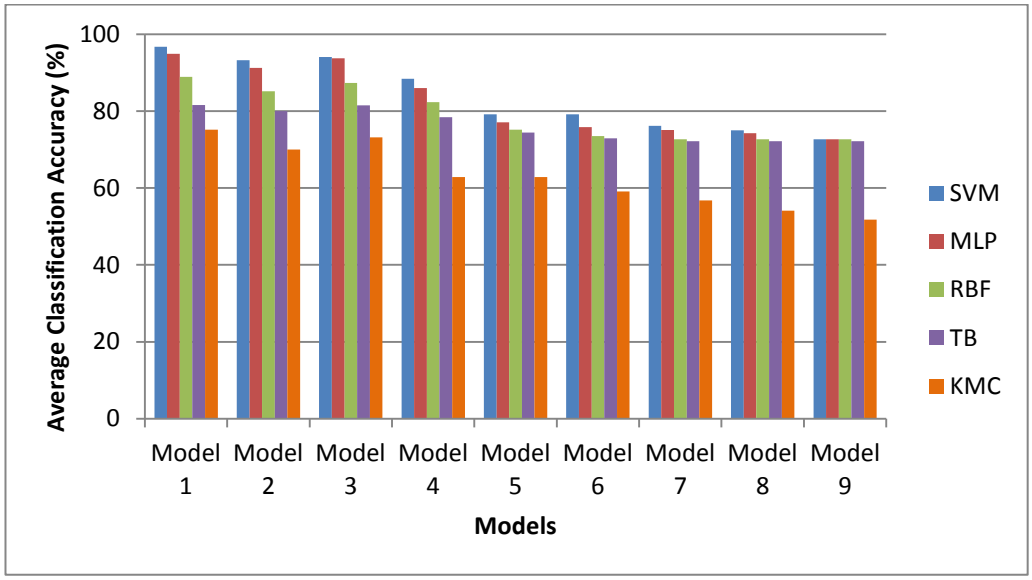


Figure 4.111. Average classification accuracies of prediction models using Relief-F on the 2006 dataset

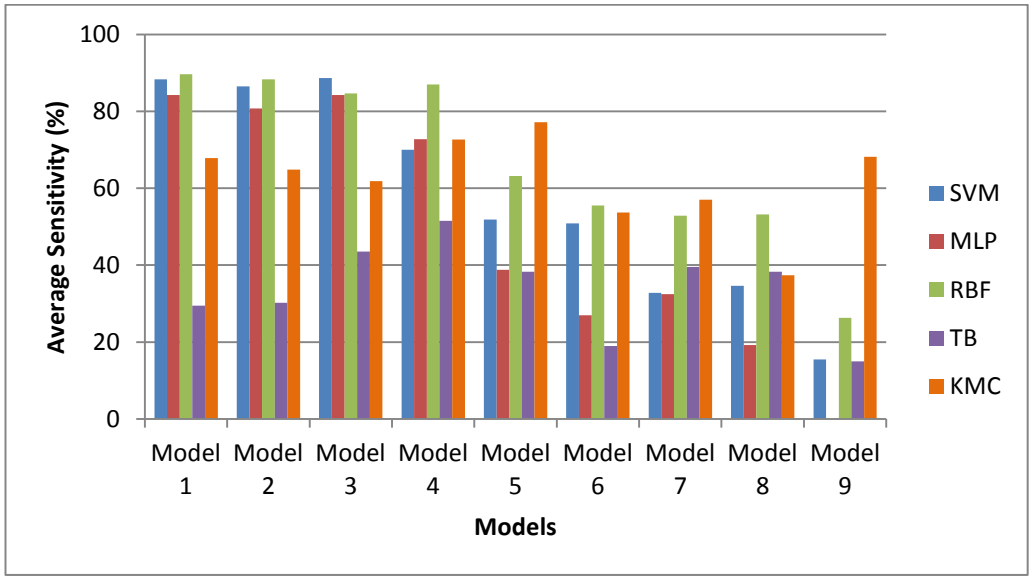


Figure 4.112. Average sensitivities of prediction models using Relief-F on the 2006 dataset

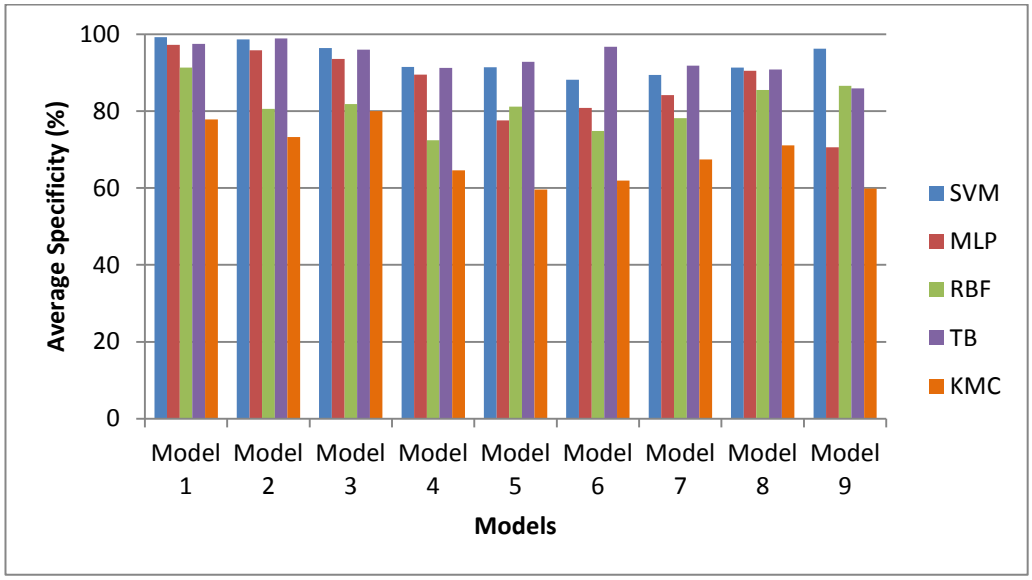


Figure 4.113. Average specificities of prediction models using Relief-F on the 2006 dataset

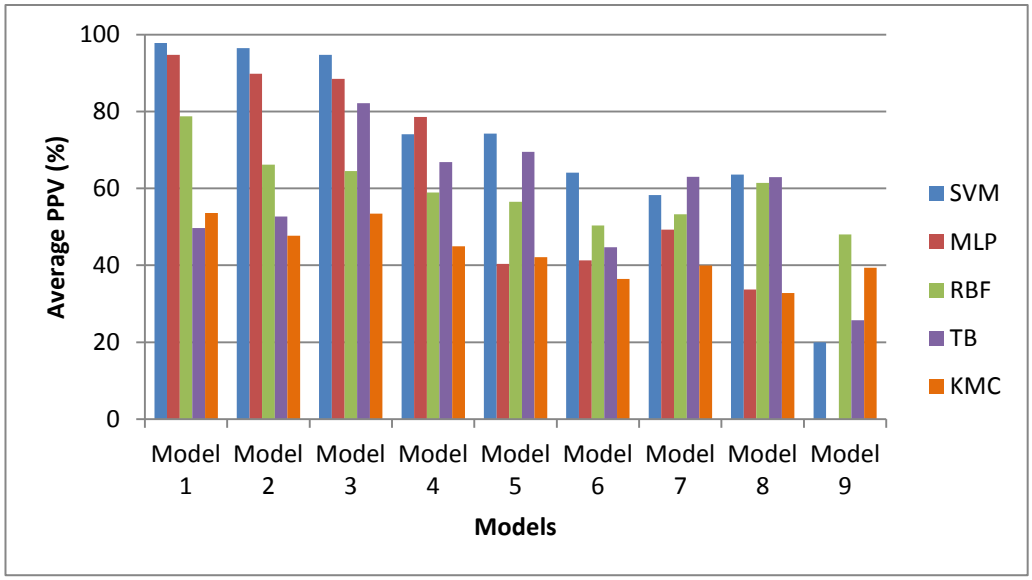


Figure 4.114. Average PPV's of prediction models using Relief-F on the 2006 dataset

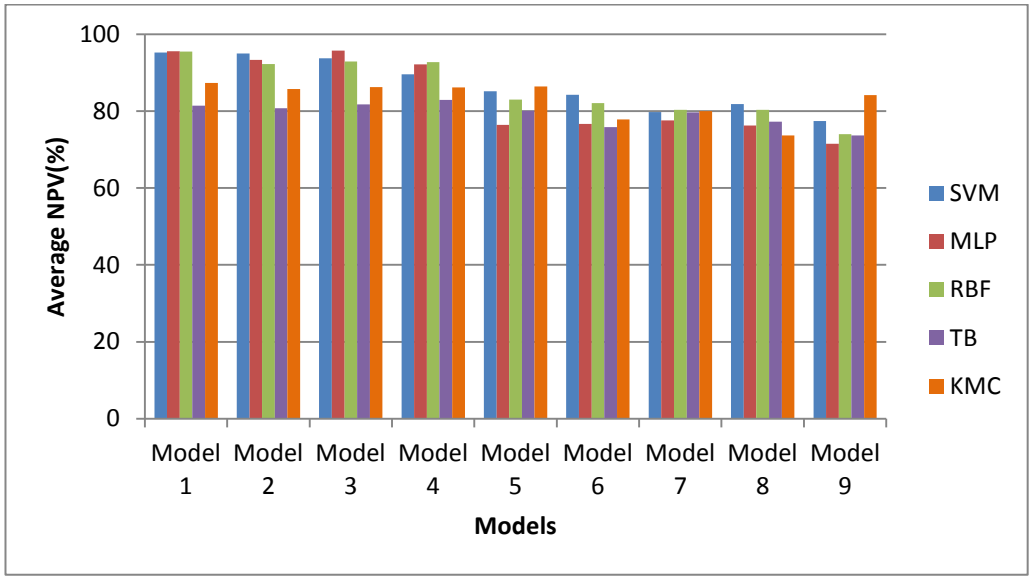


Figure 4.115. Average NPV's of prediction models using Relief-F on the 2006 dataset

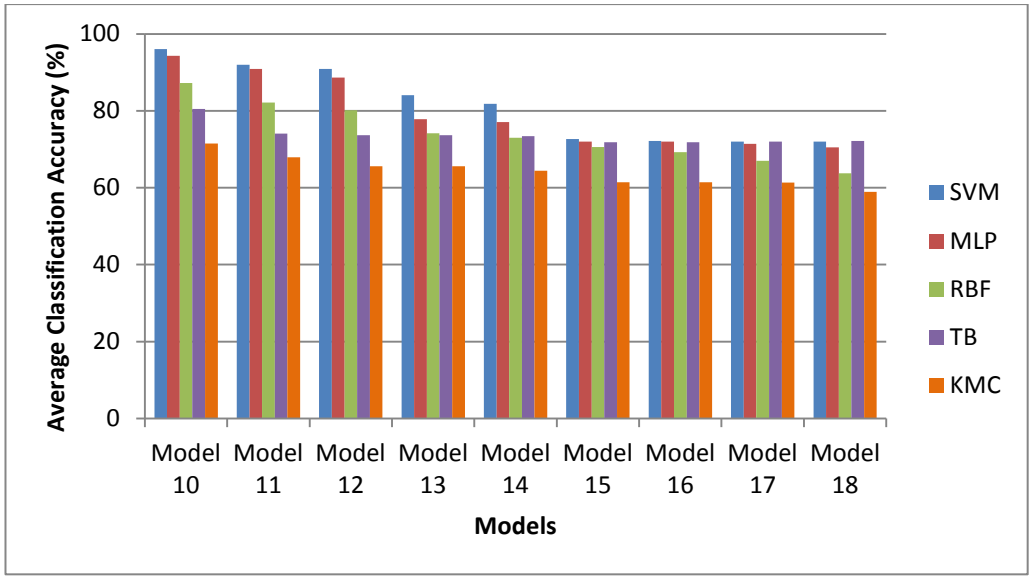


Figure 4.116. Average classification accuracies of prediction models using F-Score on the 2006 dataset

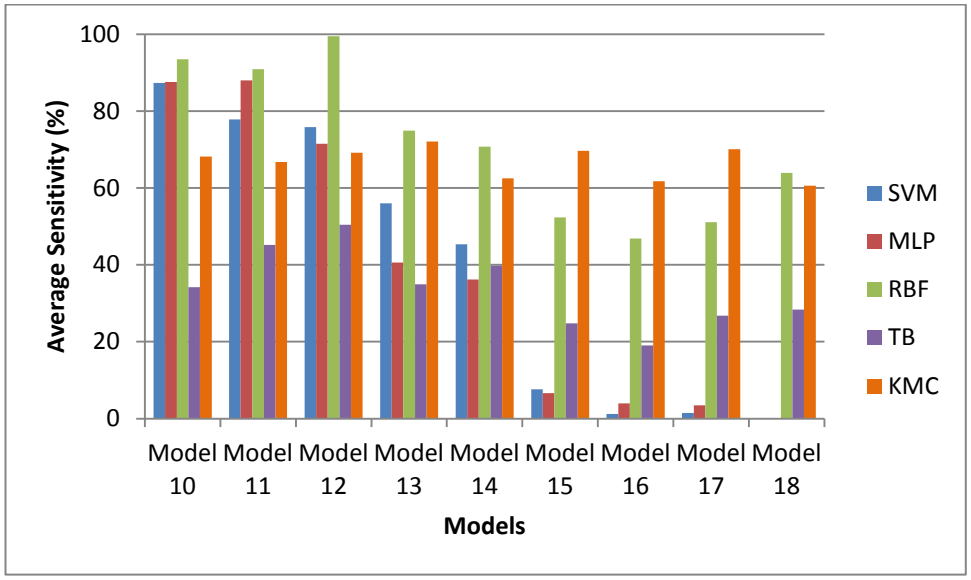


Figure 4.117. Average sensitivities of prediction models using F-Score on the 2006 dataset

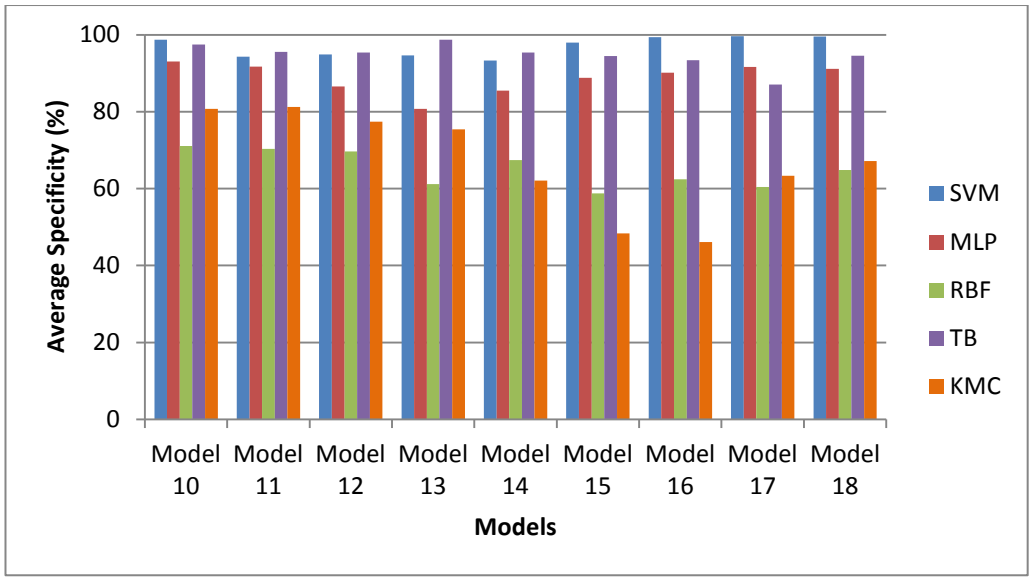


Figure 4.118. Average specificities of prediction models using F-Score on the 2006 dataset

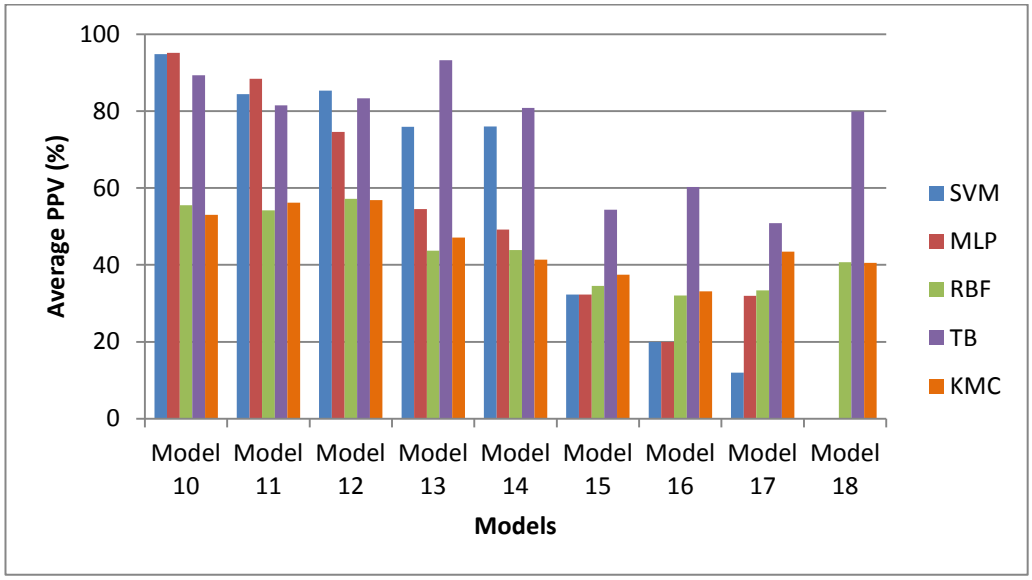


Figure 4.119. Average PPV's of prediction models using F-Score on the 2006 dataset

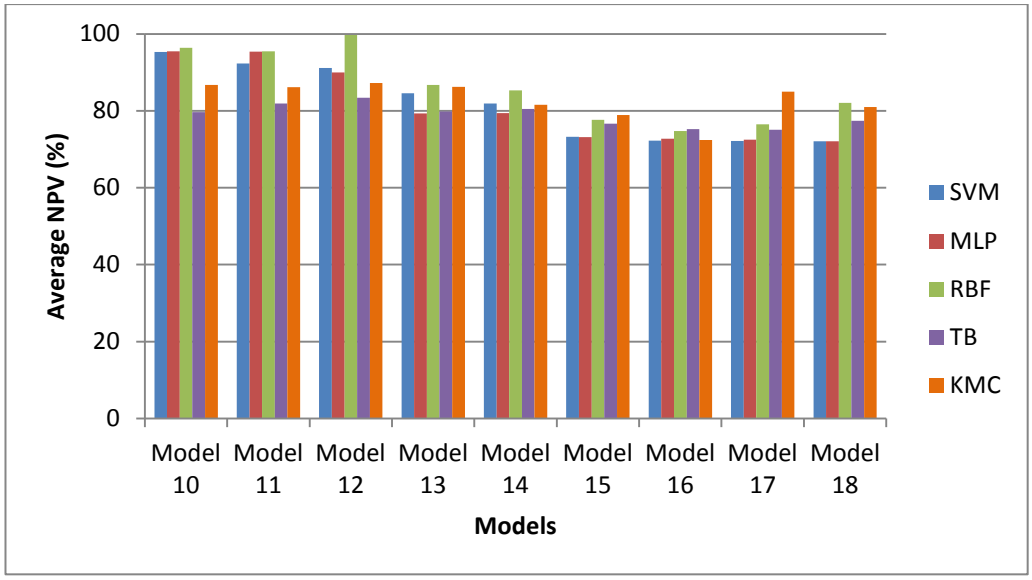


Figure 4.120. Average NPV's of prediction models using F-Score on the 2006 dataset

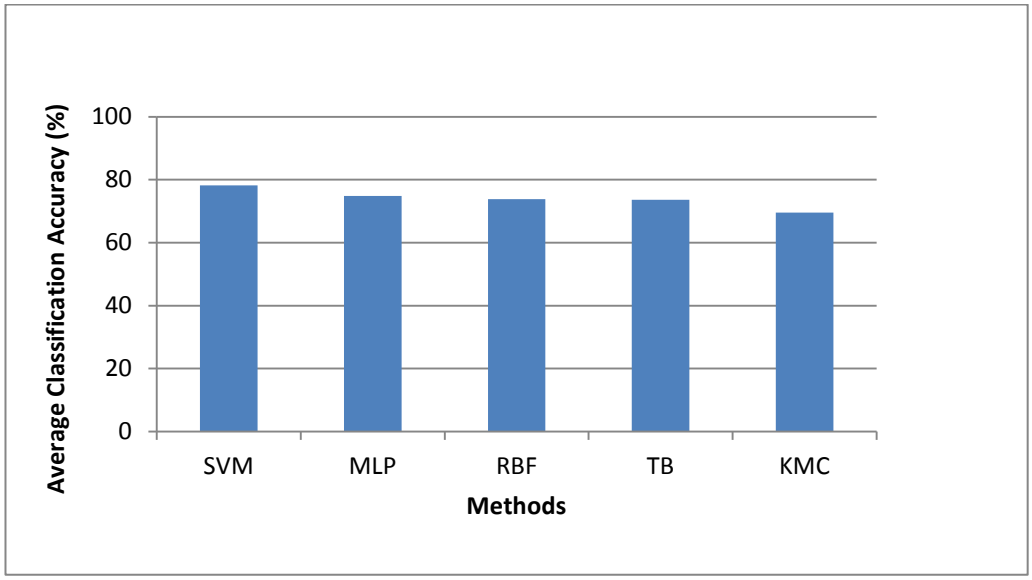


Figure 4.121. Average classification accuracies of prediction models using CFS on the 2006 dataset

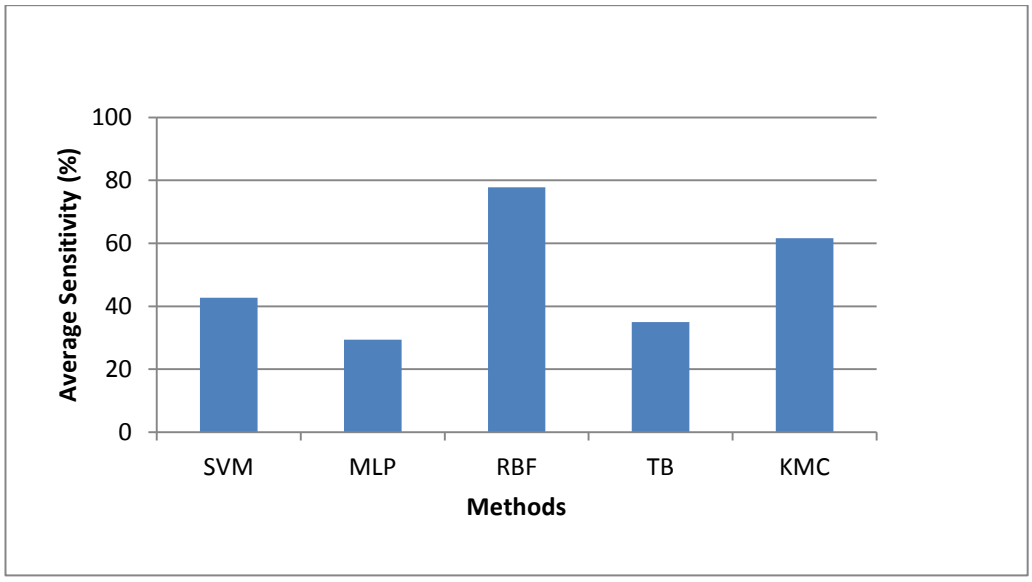


Figure 4.122. Average sensitivities of prediction models using CFS on the 2006 dataset

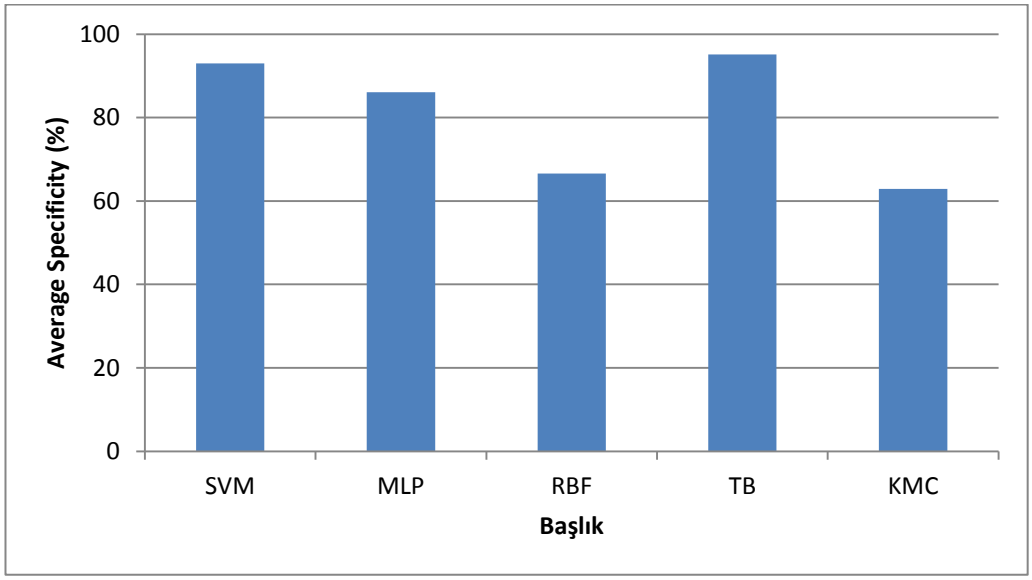


Figure 4.123. Average specificities of prediction models using CFS on the 2006 dataset

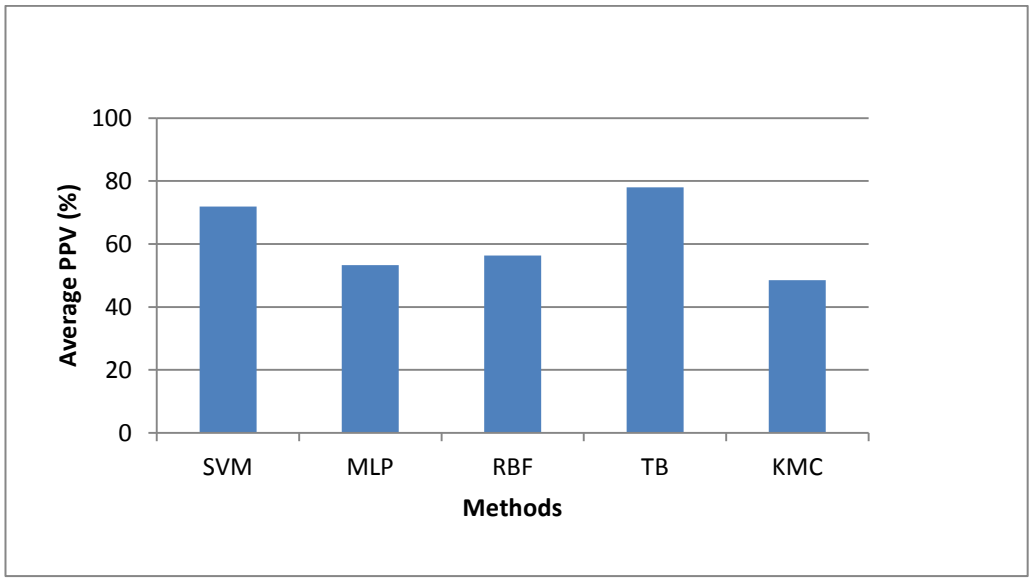


Figure 4.124. Average PPV's of prediction models using CFS on the 2006 dataset

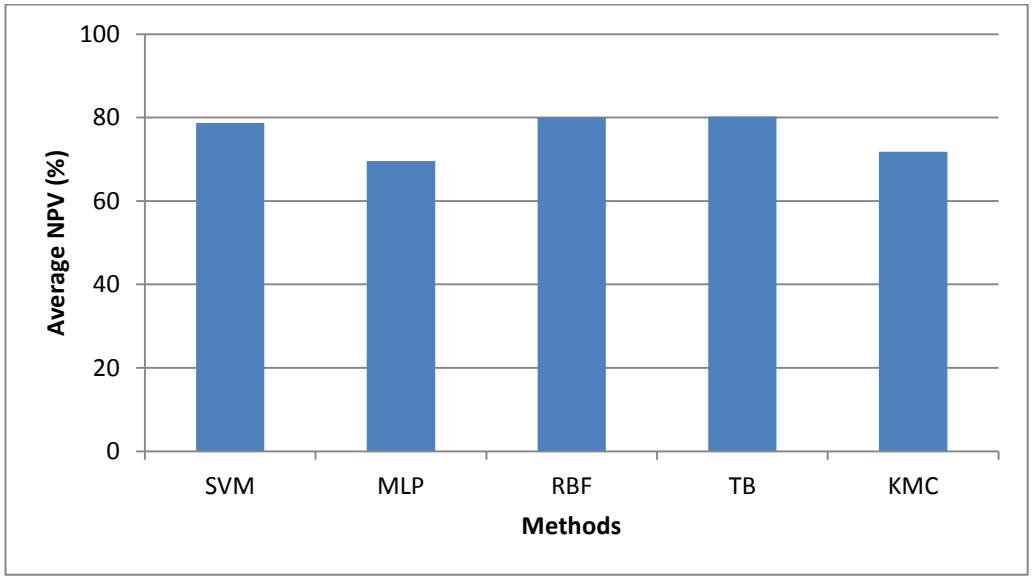


Figure 4.125. Average NPV's of prediction models using CFS on the 2006 dataset

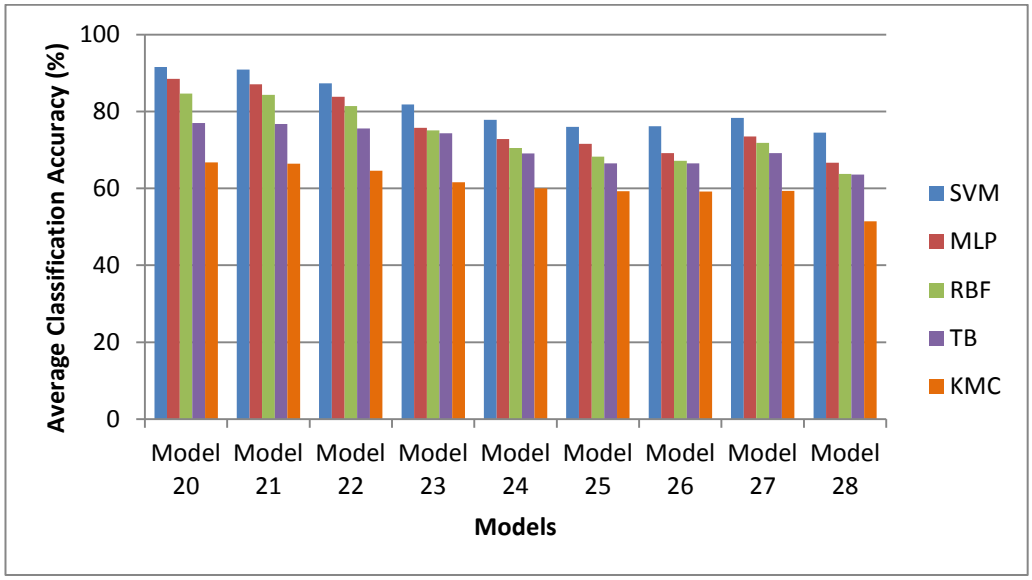


Figure 4.126. Average classification accuracies of prediction models using Relief-F on the 2007 dataset

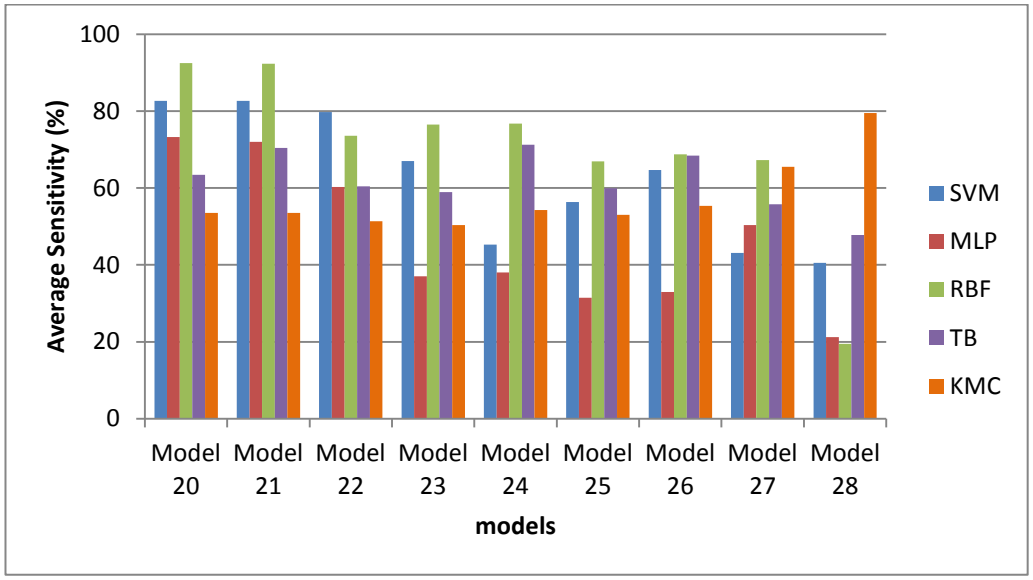


Figure 4.127. Average sensitivities of prediction models using Relief-F on the 2007 dataset

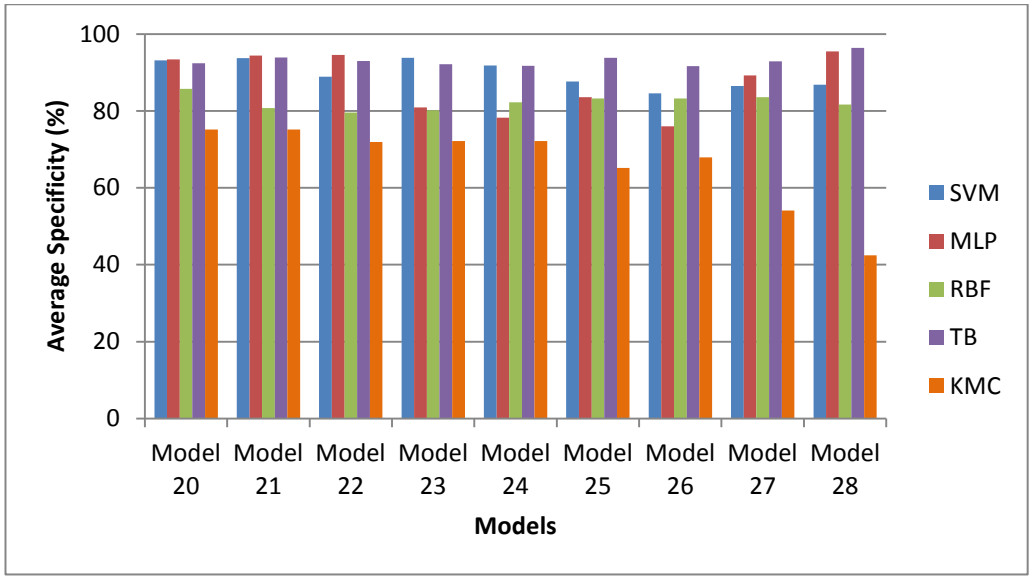


Figure 4.128. Average specificities of prediction models using Relief-F on the 2007 dataset

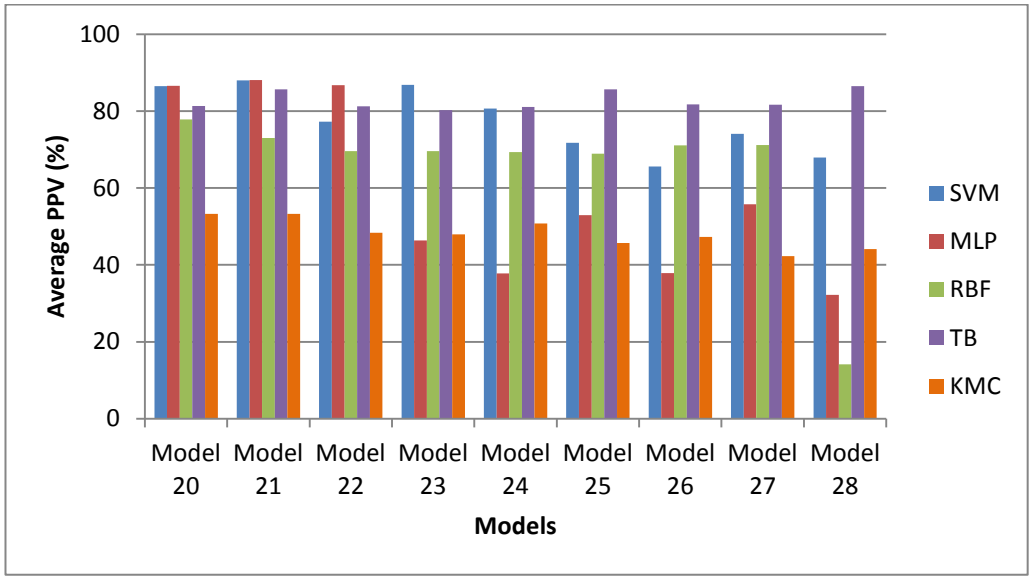


Figure 4.129. Average PPV's of prediction models using Relief-F on the 2007 dataset

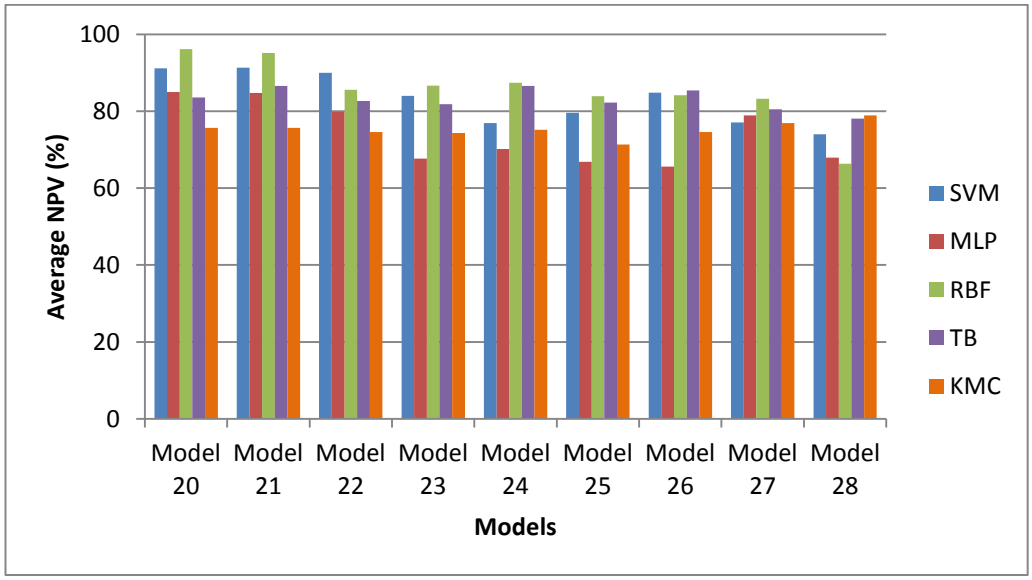


Figure 4.130. Average NPV's of prediction models using Relief-F on the 2007 dataset

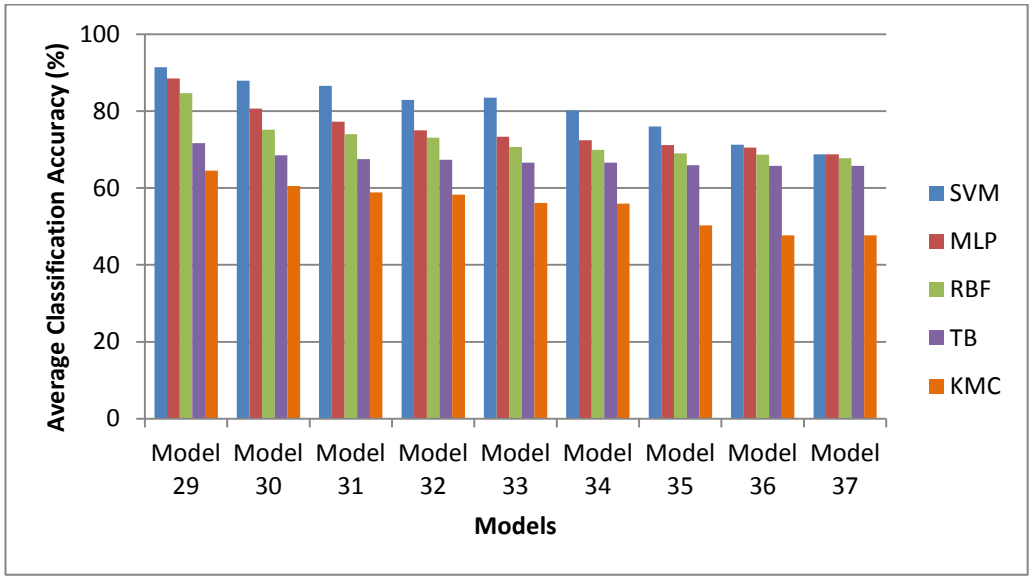


Figure 4.131. Average classification accuracies of prediction models using F-Score on the 2007 dataset

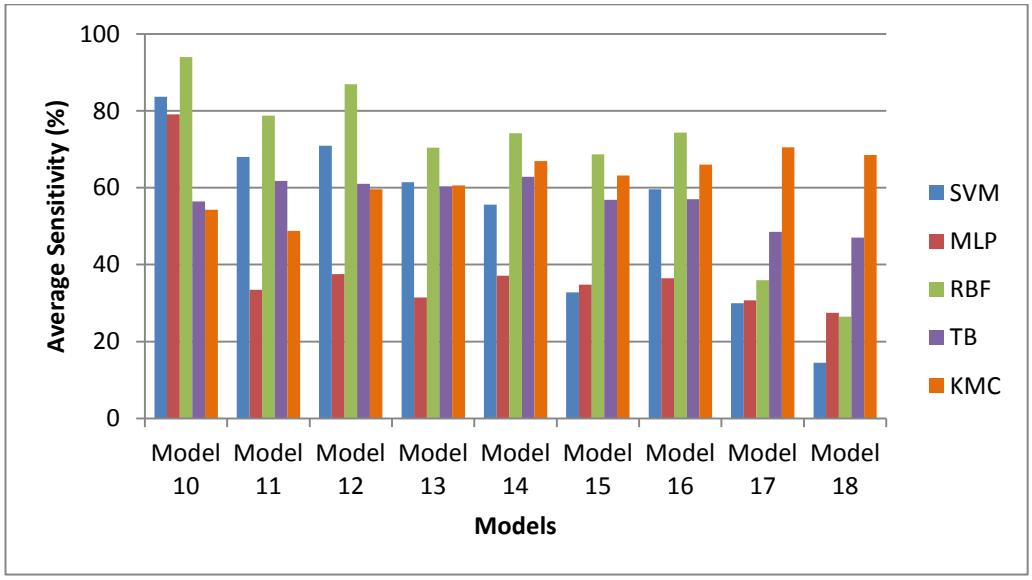


Figure 4.132. Average sensitivities of prediction models using F-Score on the 2007 dataset

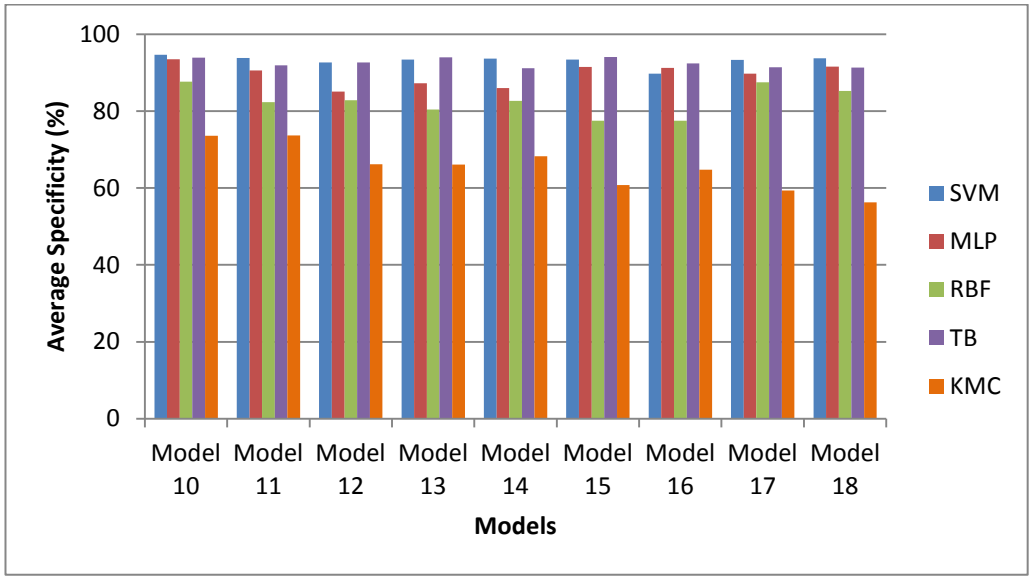


Figure 4.133. Average specificities of prediction models using F-Score on the 2007 dataset

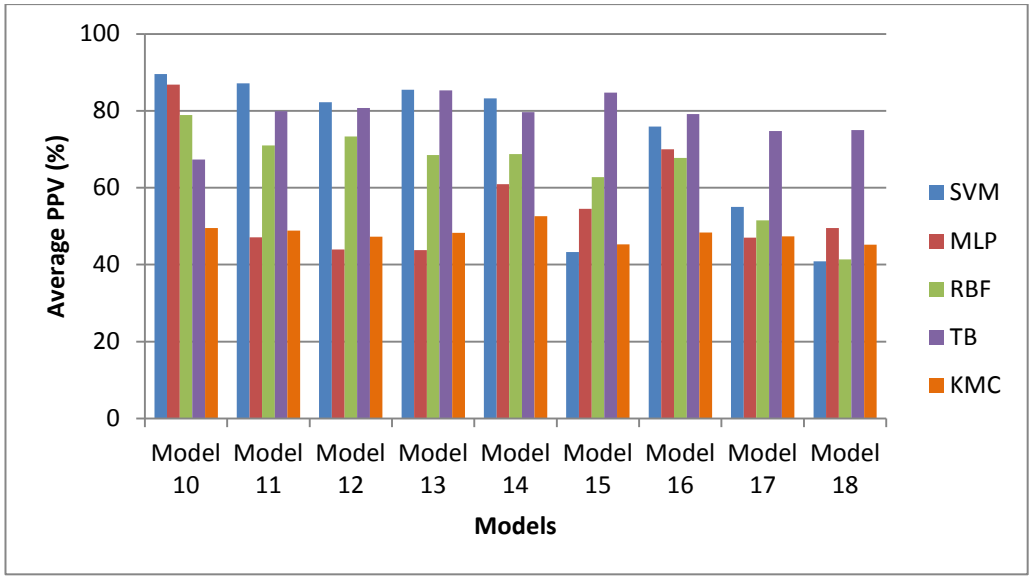


Figure 4.134. Average PPV's of prediction models using F-Score on the 2007 dataset

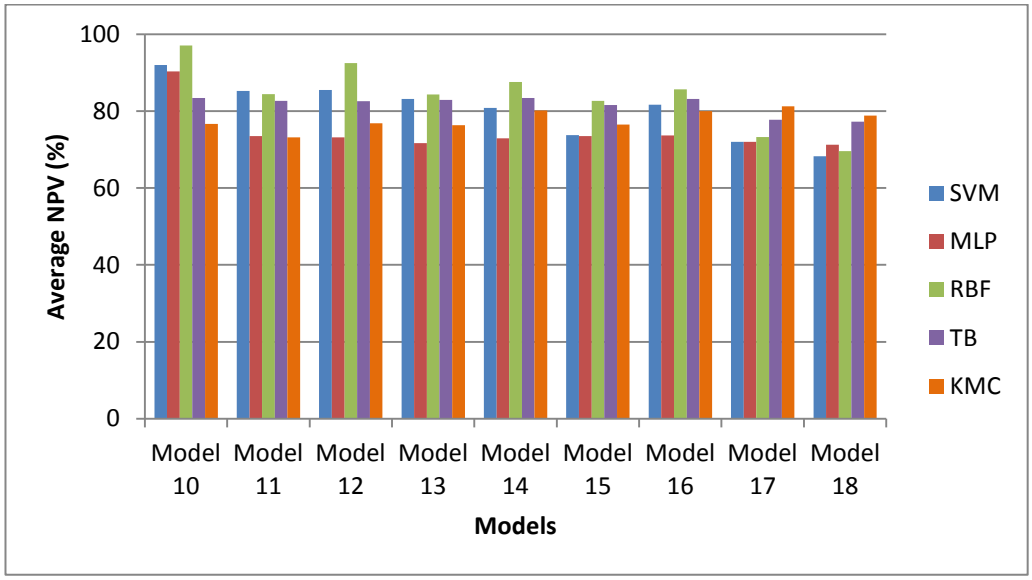


Figure 4.135. Average NPV's of prediction models using F-Score on the 2007 dataset

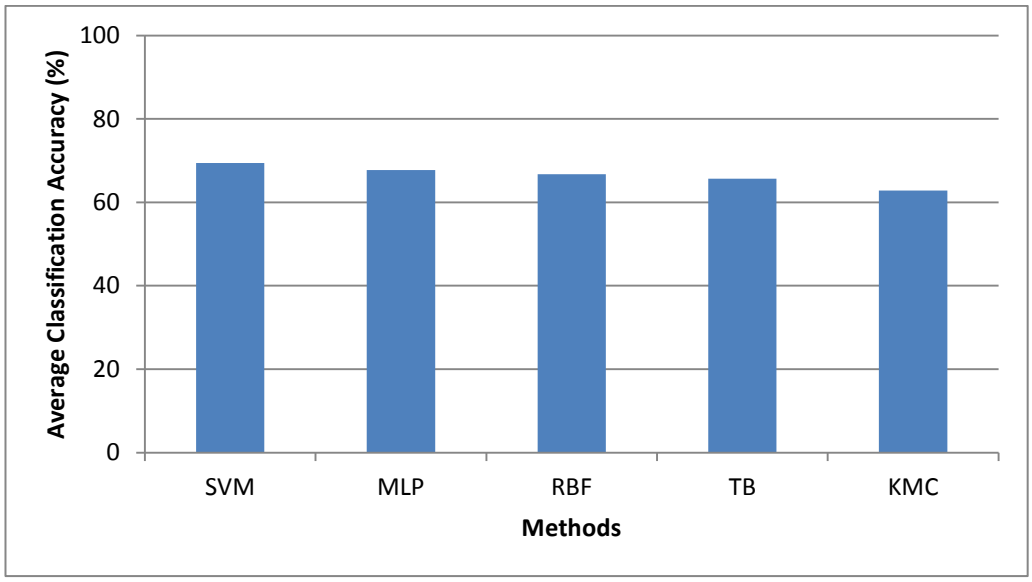


Figure 4.136. Average classification accuracies of prediction models using CFS on the 2007 dataset

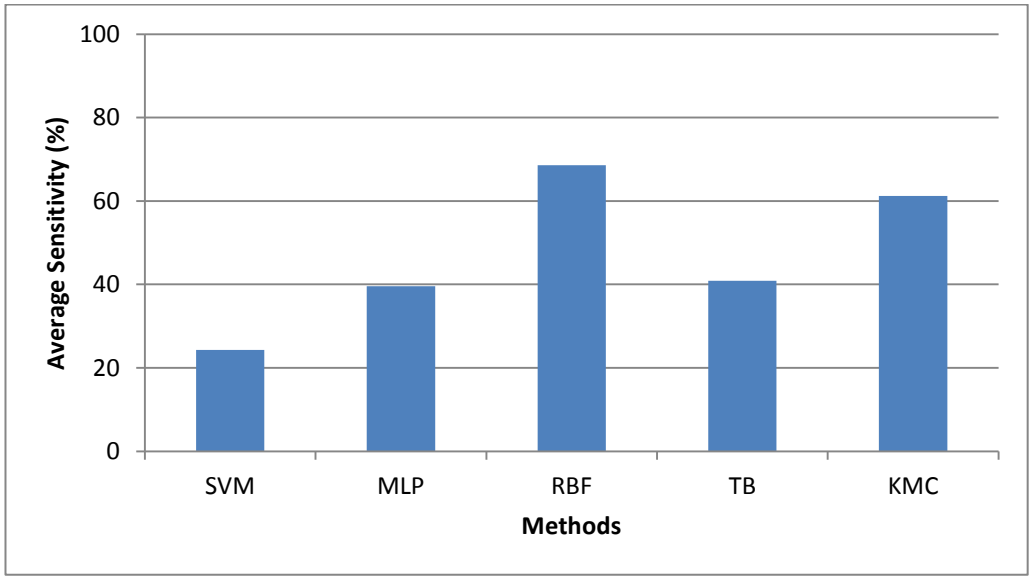


Figure 4.137. Average sensitivities of prediction models using CFS on the 2007 dataset

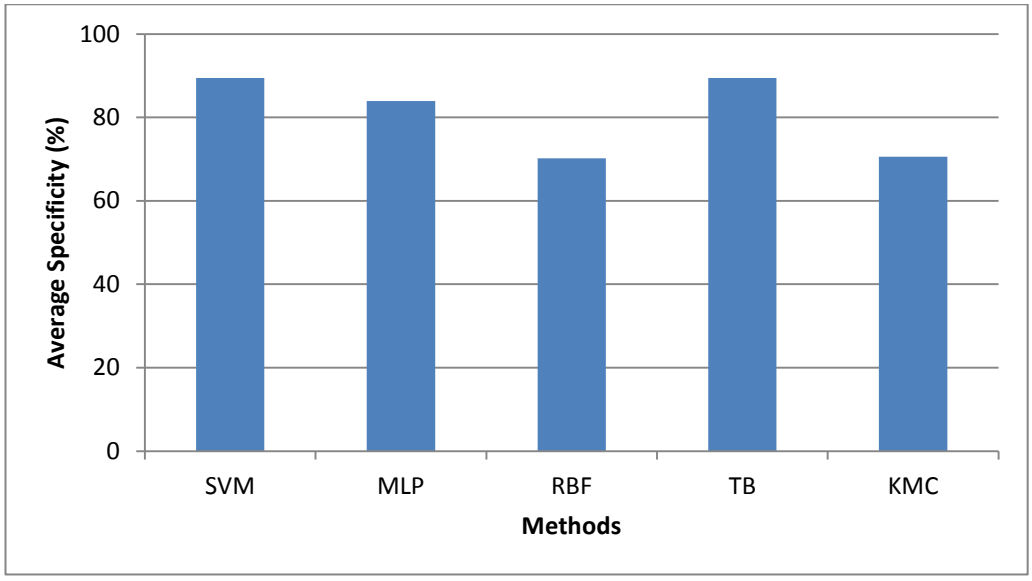


Figure 4.138. Average specificities of prediction models using CFS on the 2007 dataset

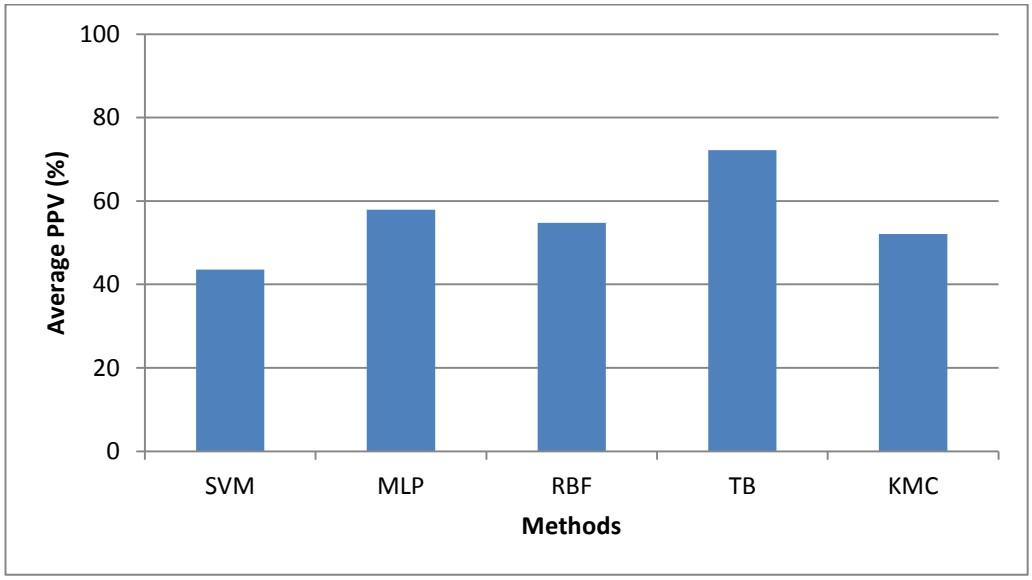


Figure 4.139. Average PPV's of prediction models using CFS on the 2007 dataset

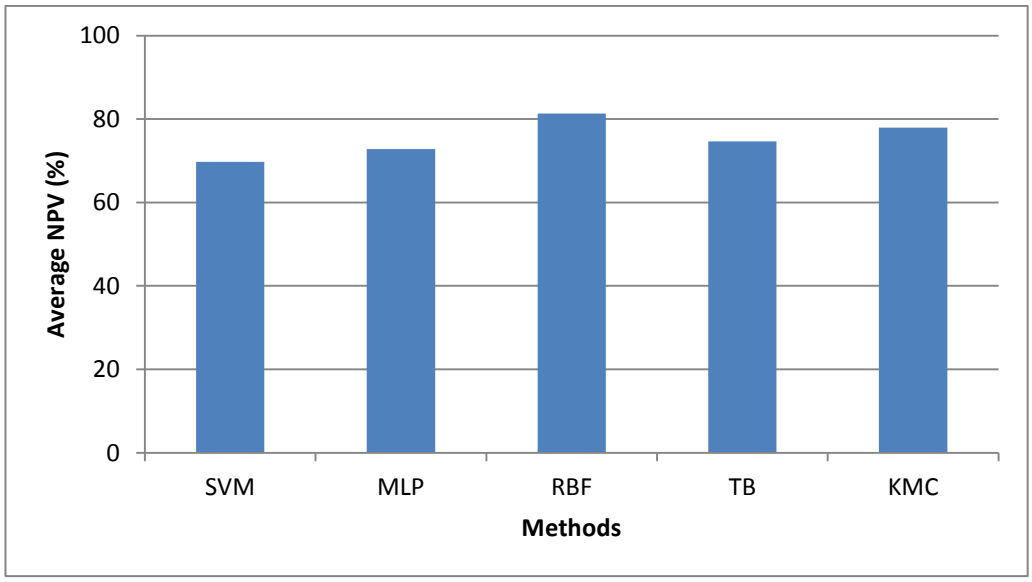


Figure 4.140. Average NPV's of prediction models using CFS on the 2007 dataset

4.1. General Discussion on the Results

- For both datasets, in general, SVM based prediction models show higher performance than the prediction models based on other machine learning methods regardless of the feature selection algorithm used and validation technique applied. The general ranking

among the machine learning methods in terms of classification acuuracy is SVM, MLP, RBF, TB and KMC.

- For both datasets, in general, prediction models based on the Relief-F feature selection algorithm exhibit higher performance than the models based on F-Score and CFS algorithms. In other words, compared with F-score and CFS, Relief-F is more successful in discriminating the useful and redundant variables for prediction of admission decision.
- For both datasets, in general, the prediction models having more predictor variables have better performance.
- The prediction accuracy of all methods usually decrease as the split of the data goes from 10-fold cross validation down to 50-50% split of the data. In other words, putting more data in the training set increases the learning capacity of the prediction model.
- The classification accuracies on the 2006 dataset are higher than those on the 2007 dataset for all methods. This is due to the existence of more outliers in the 2007 dataset.
- Investigation of average classification accuracies reveal that SVM, MLP, RBF and TB exhibit similar performance as the number of predictor variables diminishes in the prediction models.
- Although SVM based models show superior performance in terms of classification accuracy, the execution times of these models are higher than the ones of prediction models based on other machine learning methods. This is due to the usage of grid search algorithm in the training phase of SVM to find the optimal values of the SVM parameters.

4.2. Discussion for Classification Accuracy on the 2006 Dataset

- The prediction models including the gender, NSSE, NSPE, GPA, specialization area, vertical jump test score, coordination and skill test

score, 30-meter dash test score and20-meter shuttle run test score yield the highest accuracies for both Relief-F and F-score.

- For Relief-F based prediction models, the ones including the 30-meter dash test score give the lowest accuracies whereas the models including the 20-meter shuttle run test score give the lowest accuracies for F-Score based prediction models.
- The smallest and highest accuracies vary in the range of 71.33%-97.20%, 71.33%-95.35%, 71.33%-93.02%, 68.53%-84.21%, 50.88%-81.82%for the SVM-based, MLP-based, RBF-based, TB-based and KMC-based prediction models, respectively that applied Relief-F.
- The smallest and highest accuraciesvary in the range of 65.22%-94.29%, 65.22%-88.57%,65.22%-88.57%, 65.22%-80.34%,47.4%-76.60% for the SVM-based, MLP-based, RBF-based, TB-based and KMC-based prediction models, respectively that applied F-Score.
- For the 2006 dataset, the overall average accuracies for Relief-F based, F-score based and CFS based prediction models are 77.56%, 74.60% and 73.93%, respectively. For the 2007 dataset, the same overall average accuracies for the aforementioned models are 72.82%, 70.31% and 66.46%, respectively.
- Observation of the classification accuracies of Model 4 and Model 5 suggests that gender plays a significant role for prediction of admission decision. Similarly, the observed decrease in classification accuracy from Model 8 to Model 9 reveals that NSPE markedly increases the classification accuracy for Relief-F.
- Observation of the classification accuracies of Model 14 and Model 15 suggests that NSPE plays a significant role for prediction of admission decision for F-Score.
- Compared to the classification accuracies obtained by KMC, the machine learning methods SVM, MLP, RBF and TB achieve on the average 19.12%, 17.29%, 12.98% and 11.36% more accurate classifications, respectively.

4.3. Discussion for Classification Accuracy on the 2007 Dataset

- The prediction models including the attributes,gender, NSSE, NSPE, GPA, specialization area, vertical jump test score, coordination and skill test score, 30-meter dash test score, 20-meter shuttle run test score gives the highest accuracy for both Relief-F and F-score.
- For Relief-F based prediction models, the ones including the 30-meter dash test score give the lowest accuracies whereas the models including the 20-meter shuttle run test score give the lowest accuracies for F-Score based prediction models.
- The smallest and highestaccuraciesvary in the range of 66.10%-94.29%, 59.32%-91.43%, 55.93%-88.57%, 55.93%-80.85% and 46.81%-73.50% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied Relief-F.
- The smallest and highestaccuracies vary in the range of 71.33%-97.67%, 70.63%-95.35%, 52.78%-90.21%, 71.93%-82.46% and 51.16%-75.86% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied F-Score.
- For the 2007 dataset, the overall average accuracies for Relief-F based, F-score based and CFS based prediction models are 72,82%, 70,31% and 66,46%, respectively.
- Observation of the classification accuracies of Model 35 and Model 36 suggests that vertical jump test plays a significant role for prediction of admission decision for Relief-F and F-Score.
- Compared to the classification accuracies obtained by KMC, the machine learning methods SVM, MLP, RBF and TB achieve on the average 23.01%, 17.66%, 15.06% and 10.87% more accurate classifications, respectively.

4.4. Discussion on Other Performance Criteria

- The average sensitivity values obtained by RBF on both datasets are higher than those obtained by other classifiers. In other words, RBF performs the best for correctly predicting the admission decision of students who are actually admitted to the School.
- The lowest sensitivity values for both datasets have been obtained by MLP, independent of which validation technique has been used.
- The results show that TB achieves the highest specificity values compared to those obtained by other classifiers for both datasets. In other words, TB performs the best for correctly predicting the admission decision of students who are not actually admitted to the School.
- The good performance of TB for correctly predicting the admission decision of students who are not actually admitted to the School compansates its relatively poor performance for correctly predicting the admission decision of students who are actually admitted to the School.
- According to the results obtained, the average PPV values obtained by TB on both datasets are higher than those obtained by other classifiers.
- The lowest PPV values for both datasets have been obtained by K-Means Clustering, independent of which validation technique has been used.
- The ranking of the classifiers in terms of their PPV values is TB, SVM, RBF Network, MLP and KMC, independent of the dataset used and the model validation technique applied.
- The average NPV's values obtained by RBF on both datasets are higher than those obtained by other classifiers.
- Investigation of average NPV's's obtained by each classifier reveals that the performance of TB and RBF is comparable for both dataset.

- For the 2006 dataset, the smallest and highest sensitivities vary in the range of 15.5%-86.66%, 0%-84.25%, 26.33%-89.66%, 15%-51.50% and 37.33%-77.16% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied Relief-F and 0%-87.33, 0%-88%, 46.83%-99.50%, 19%-50.41 and 60.58%-72.07 for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied F-Score.
- For the 2006 dataset, the smallest and highest specifities vary in the range of 88.12%-99.22%, 70.57%-97.17%, 72.40%-91.31%, 85.88%-98.84% and 59.58%-80.01% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied Relief-F and 93.27%-99.61%, 80.75%-93.04%, 58.78%-71.12%, 87.09%-98.74% and 46.12%-81.24%for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied F-Score.
- For the 2006 dataset, the smallest and highest PPV's vary in the range of 20%-97.82%, 0%-94.69%, 48%-78.70%, 25.72%-82.14% and 32.81%-53.58% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied Relief-F and 0%-94.77%, 0%-95.14%, 32.06%-57.16%, 50.86%-93.18% and 33.16%-56.81% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied F-Score.
- For the 2006 dataset, the smallest and highest NPV's vary in the range of 77.44%-95.22%, 71.53%-95.67%, 73.99%-95.45%, 73.66%-82.88% and 73.66%-87.29% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied Relief-F and 72.03%-95.27%, 72.05%-95.43%, 74.73%-99.73%, 75.06%-83.41% and 72.42%-

87.21% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied F-Score.

- For the 2007 dataset, the smallest and highest sensitivities vary in the range of 40.50%-82.66%, 21.25%-73.25%, 19.50%-92.50%, 47.75%-71.25% and 50.33%-79.50% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied Relief-F and 14.50%-83.66%, 27.50%-79.08%, 26.50%-94%, 47%-62.83% and 48.75%-70.50% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied F-Score.
- For the 2007 dataset, the smallest and highest specificities vary in the range of 84.57%-93.76%, 76.02%-95.48%, 79.61%-85.76%, 91.62%-96.41% and 42.45%-75.16% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied Relief-F and 89.74%-94.66%, 85.05%-93.45%, 77.45%-87.65%, 91.09%-94.02% and 56.23%-73.68% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied F-Score.
- For the 2007 dataset, the smallest and highest PPV's vary in the range of 65.60%-87.95%, 32.22%-88.02%, 14.14%-77.78%, 80.21%-86.50% and 42.32%-53.23% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied Relief-F and 40.88%-89.57%, 43.75%-86.84%, 41.40%-78.88%, 76.34%-85.33% and 45.22%-52.57% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied F-Score.

- For the 2007 dataset, the smallest and highest NPV's vary in the range of 74.01%-91.28%, 65.57%-84.99%, 66.33%-96.10%, 78.09%-86.59% and 71.32%-78.90% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied Relief-F and 68.22%-92%, 71.25%-90.34%, 69.57%-97.07%, 77.26%-83.41% and 73.20%-81.26% for the SVM-based, MLP-based, RBF-based, TreeBoost-based and K-means-based prediction models, respectively among models that applied F-Score.
- For the 2006 dataset, the overall average sensitivities for Relief-F based, F-score based and CFS based prediction models are 53.87%, 49.74% and 49.31%, respectively.
- For the 2006 dataset, the overall average specificities for Relief-F based, F-score based and CFS based prediction models are 84.68%, 82.48% and 80.73%, respectively.
- For the 2006 dataset, the overall average PPV's for Relief-F based, F-score based and CFS based prediction models are 57.88%, 53.43% and 61.60%, respectively.
- For the 2006 dataset, the overall average NPV's for Relief-F based, F-score based and CFS based prediction models are 83.78%, 82.09% and 76.07%, respectively.
- For the 2007 dataset, the overall average sensitivities for Relief-F based, F-score based and CFS based prediction models are 59.66%, 55.65% and 46.93%, respectively.
- For the 2007 dataset, the overall average specificities for Relief-F based, F-score based and CFS based prediction models are 83.70%, 84.65% and 80.72%, respectively.
- For the 2007 dataset, the overall average PPV's for Relief-F based, F-score based and CFS based prediction models are 66.34%, 63.76% and 56.10%, respectively.

- For the 2007 dataset, the overall average NPV's for Relief-F based, F-score based and CFS based prediction models are 80.17%, 79.67% and 75.29%, respectively.

5. CONCLUSION

In this thesis, new hybrid admission decision prediction models have been developed by using different machine learning methods including SVM, MLP, RBF Network, TB and KMC combined with feature selection algorithms to investigate the effect of the predictor variables on the admission decision of a candidate to the School of of Physical Education and Sports at Cukurova University. Experiments have been conducted on two datasets, which are actual test results of candidates who applied to the school in 2006 and 2007. By using the Relief-F and F-score feature selection algorithms, ranking of the attributes has been calculated. Then, based on these ranking scores, several models have been developed by removing the attribute with the lowest score at a time. In contrast to Relief-F and F-score algorithms, the CFS algorithm gave a set of selected variables to develop a single model. 10-fold cross validation and percentage splits of data have been used. The performance of SVM, MLP, RBF Network, TB and KMC classifiers on both datasets are evaluated using classificaiton accuracy, sensitivity, specificity, PPV and NPV.

Considering the results obtained, one can reach several conclusions. First of all, SVM-based prediction models show better performance than the models developed by other classification methods. The order of the classification methods for prediction of admission decision, from the best to the worst, is SVM, MLP and RBF Network, TB, KMC. Secondly,Relief-F-based models in general exhibit higher performance than the CFS-based and F-Score-based models.Thirdly, observation of the resultsof 2006 dataset suggests that gender and NSPE play a significant role for prediction of admission decision for Relief-F andNSPE plays a significant role for prediction of admission decision for F-Score.Observation of the resultsof 2007 dataset suggestsvertical jump test plays a significant role of prediction of admission decision for Relief-F and F-Score. According to CFS algorithm, NSPE, GPA, special area, vertical jump test score, 30-m dash test score adn 20-meter shuttle run test score are more important features of admission decision for 2006 dataset and NSPE, special area and 20-meter shuttle run test score are more important features of admission decision for 2007 dataset.

Although this is a detailed study,future work can be performed in a number of different areas. Different feature selection algorithms can be combined with classification methods to identify the relevant and irrelevant features for prediction of admission decision. Additional research on new data sets can be carried out to investigate the effect of predictor variables on the admission decision. The datasets can be merged into a single one in order to have a larger dataset, which in turn can be used to develop new admission decision models with the help of machine learning methods combined with feature selection.

REFERENCES

AÇIKKAR, M., and AKAY, M.F., 2008. Predicting the admission decision of a candidate to the School of Physical Education and Sports at Cukurova University by using artificial neural networks. EECS'08, 5th International Symposium on Electrical and Computer Systems, Gemikonagi, Cyprus.

AÇIKKAR, M., AKAY, M.F., ABUT, F., and ISOGLU, O., 2014. Predicting the Admission Decision of a Candidate to the School of Physical Education and Sports at Cukurova University by Using Multilayer Perceptron Combined with Feature Selection, Second International Symposium on Engineering, Artificial Intelligence & Applications (ISEAIA2014), North Cyprus, Nov, 5-7, pp. 15.

AKAY, M.F., and AÇIKKAR, M., 2009. Support Vector Machines For Predicting The Admission Decision Of A Participant To The School Of Physical Education And Sports At Cukurova University. Expert Systems with Applications, vol.36, pp.7228-7233.

APPEL, R., FUCHS, T., DOLL'AR, P. , PERONA, P., 2013. Quickly Boosting Decision Trees Pruning Underachieving Features Early. Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA.

BOSER, B.E., 1992. A Training Algorithm for Optimal Margin Classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory 5, 144-152, Pittsburgh.

BOSWELL, D., 2002. Introduction to Support Vector Machines. August 6.

CHEN, Y. W., LIN, C. J., 2003. Combining SVMs with various feature selection strategies. NIPS 2003 feature selection challenge, 1-10.

ÇUKUROVA ÜNİVERSİTESİ BEDEN EĞİTİMİ VE SPOR YÜKSEKOKULU, " Özel Yetenek Sınavı ", http://besyo.cu.edu.tr/besyo/TR/OzelYetenekSinavi.aspx (Erişim tarihi: 11 Şubat 2015)

FLETCHER, R.H., SUZANNE, W., 2005. Clinical epidemiology : the essentials, Fourth edition, Baltimore, Md.: Lippincott Williams & Wilkins. p. 45.

ISBN 0-7817-5215-9

FRIEDRICHS, F., and IGEL, C., 2005. Evolutionary tuning of multiple SVM parameters. Neurocomputing, vol. 64, pp. 107-117.

GHISELLI, E. E., 1964. Theory of Pyschological Measurement. McGraw-Hill Book Company.

GHOSH-DASTİDAR, S., ADELI, H., DADHMER, N., 2008. Principal Component Analysis Enhanced Cosine Radial Basis Function Neural Network for Robust Epilepsy and Seizure Detection. Biomedical Engineering, Vol. 55, No. 2, pp. 512 – 518.

GOMM, J.B., YU, D.L, 2000. Selecting Radial Basis Function Network Centers with Recursive Orthogonal Least Squares Training. IEEE Transactions on Neural Networks, 2: 306-314.

GUO, X.C., LIANG, Y.C., WU, C.G., and WANG, C.Y., 2006. PSO-based hyper- parameters selection for LS-SVM classifiers. Proceedings of Neural Information Processing, Belin, pp. 1138-1147.

HALL, M. A., and SMITH, L. A., 1997. Feature Subset Selection: a correlation based filter approach. In 1997 International Conference on Neural Information Processing and Intelligent Information Systems. Springer, Berlin, pp. 855-858.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H., 2009. 10. Boosting and Additive Trees. The Elements of Statictical Learning. Second Edition, Springer, New York, pp. 337–384. ISBN 0-387-84857-6

HSU, C.W., CHANG, C.C., and LIN, C.J., 2003. A practical guide to support vector classification.

IMRIE, C.E., DURUCAN, S., 2000. A River flow prediction using artificial neural networks: generalization beyond the calibration range. Journal of Hydrology, Vol. 233, No. 1-4, pp. 138-153.

JAGANATHAN, P., RAJKUMAR, N., and KUPPUCHAMY, R., 2012. A Comparative Study of Improved F-Score with Support Vector Machine and RBF Network for Breast Cancer Classification. International Journal of Machine Learning and Computing, Vol. 2, No. 6.

JI, L., and WANG, B., 2007. Parameters selection for SVR based on the SCEM-

UA algorithm and its application on monthly runoff prediction. Proceedings of the 2007 International Conference on Computational Intelligence and Security, pp. 48-51.

KAREGOWDA, A., JAYARAM, M.A., 2011. Feature Subset Selection using Cascaded GA & CFS: A Filter Approach in Supervised Learning. International Journal of Computer Applications (0975 – 8887) Volume 23– No.2.

LEE, J.Y., Yeh, R.Y., and Hsing-Kuo Pao, H. K., 2004. An Introduction to Support Vector Machines. National Taiwan University of Science and Technology.

NICOLAOS, B., KARAYIANNIS, M.M., 2003. The Construction and Training of Reformulated Radial Basis Function Neural Networks. IEEE Transactions on Neural Networks, 4: 835-844.

PARK, J., SANDBERG, I.W., 1991. Universal approximation using radial basis function networks Neural Comput., pp. 246–257.

SU, T., 2003. Another look at non-random methods for initializing K-means clustering. Master Project, Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA.

SUN, Y., LOU, X., and BAO, B., 2011. A Novel Relief Feature Selection Algorithm Based on Mean-Variance Model. Journal of Information & Computational Science 8, vol. 16, no. 2011, pp. 3921–3929.

TOPEND SPORTS, http://www.topendsports.com/testing/tests/sprint-30meters.htm (Erişim tarihi: 11 Şubat 2015)

VAPNİK, V., 1999. The Nature of Statistical Learning Theory. 2nd edition, Springer, 1999.

WRIGHT, G. B., FLYER, N., FORNBERG, B., 2013. Radial Basis Function-generated Finite Differences: A Mesh-free Method for Computational Geosciences, October 17, pp1-30

XIE, J., JIANG, S., XIE, W., GAO, X., 2011. An Efficient Global K-means Clustering Algorithm. Journal of Computers, Vol 6, No 2, pp. 271-279.

WITTEN, I. H., FRANK, E., and HALL, M. A., 2011. Data mining: practical machine learning tools and techniques. Morgan Kaufmann, Amsterdam.

CURRICULUM VITAE

Gözde ÖZSERT YİĞİT was born in Gaziantep-Turkey, in 1988. She completed the elementary education at Dayı Ahmet Ağa, Gaziantep. She went to Fitnat-Nuri Tekerekoğlu high school. She graduated from the department of Computer Engineering, Çukurova University, Adana, at 2012.She started MSc program at the department of Computer Engineering since 2014 with Assoc. Prof. Dr. Fatih Akay.