T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ VE EĞİTİM VERİSİ ÜZERİNE BİR UYGULAMA: ULUSLARARASI MATEMATİK VE FEN EĞİLİMLERİ ARAŞTIRMASI 2015 TÜRKİYE ÖRNEĞİ

Enes FİLİZ

DOKTORA TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Doç. Dr. Ersoy ÖZ

Temmuz, 2019

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESI YÖNTEMLERİ VE EĞİTİM VERİSİ ÜZERİNE BİR UYGULAMA: ULUSLARARASI MATEMATİK VE FEN EĞİLİMLERİ ARAŞTIRMASI 2015 TÜRKİYE ÖRNEĞİ

Enes FİLİZ tarafından hazırlanan tez çalışması 29/07/2019 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstatistik Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Doç. Dr. Ersoy ÖZ

Yıldız Teknik Üniversitesi

Danışman

Jüri Üyeleri

Doç. Dr. Ersoy ÖZ, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Filiz KARAMAN, Üye
Yıldız Teknik Üniversitesi _____

Prof. Dr. Semra ERPOLAT TAŞABAT, Üye
Mimar Sinan Güzel Sanatlar Üniversitesi _____

Dr. Öğr. Üyesi Öyküm Esra AŞKIN, Üye
Yıldız Teknik Üniversitesi _____

Dr. Öğr. Üyesi Seda BAĞDATLI KALKAN, Üye
İstanbul Ticaret Üniversitesi _____

Danışmanım Doç. Dr. Ersoy ÖZ sorumluluğunda tarafımca hazırlanan “Makine Öğrenmesi Yöntemleri ve Eğitim Verisi Üzerine Bir Uygulama: Uluslararası Matematik ve Fen Eğilimleri Araştırması 2015 Türkiye Örneği” başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Enes FİLİZ

İmza

TEŞEKKÜR

Doktora eğitimim süresince bana her türlü ilmi ve manevi desteğini esirgemeyen, çalışmalarımda hep yanımda olan, bu konuda çalışmaya yönlendiren, çalışmalarımın her aşamasında zamanını ve bilgisini esirgemeyen çok kıymetli hocam sayın Doç. Dr. Ersoy ÖZ'e ve yine çalışmalarım esnasında her türlü yardımı esirgemeyen Dr. Öğr. Üyesi Öyküm Esra AŞKIN'a ve görevlendirme ile geldiğim Yıldız Teknik Üniversitesi İstatistik Bölümü'nün tüm değerli hocalarına en içten teşekkürlerimi sunarım.

Ayrıca ev ve sosyal hayatımda bana hep destek olan, yardımını esirgemeyen, moral motivasyonumu artıran sevgili eşim Ümmü Gülsüm FİLİZ'e; dualarını ve desteklerini hep hissettiğim annem, babam ve kardeşlerime şükranlarımı sunarım.

Enes FİLİZ

İÇİNDEKİLER

SİMGE LİSTESİ vii

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ ix

TABLO LİSTESİ x

ÖZET xii

ABSTRACT xiv

1 Giriş 1
1.1 Literatür Özeti 3
1.2 Tezin Amacı 8
1.3 Hipotez 9

2 Makine Öğrenmesi 11
2.1 Makine Öğrenmesi Algoritmaları 11
2.1.1 K-En Yakın Komşu Algoritması 12
2.1.2 Naive Bayes Algoritması 13
2.1.3 Yapay Sinir Ağları 13
2.1.4 Karar Ağacı Algoritmaları 14
2.1.5 Destek Vektör Makineleri 16
2.1.6 Lojistik Regresyon Algoritması 22
2.2 Öznitelik Seçim Algoritmaları 22
2.2.1 Cfs Subset Öznitelik Seçim Algoritması 24
2.2.2 Correlation Attribute Öznitelik Seçim Algoritması 24
2.2.3 One-R Öznitelik Seçim Algoritması 24
2.2.4 Gain Ratio Öznitelik Seçim Algoritması 25
2.2.5 Info Gain Öznitelik Seçim Algoritması 25
2.2.6 ReliefF Öznitelik Seçim Algoritması 26
2.3 Sınıflandırma Algoritmaları İçin Performans Ölçüm Kriterleri 26
2.3.1 Doğru Pozitif Oranı (Duyarlılık) 27
2.3.2 Yanlış Pozitif Oranı 27
2.3.3 Precision (Hassasiyet) 28
2.3.4 F-Ölçütü 28
2.3.5 Alıcı İşlem Karakteristiği Eğrisi (ROC) 28
2.3.6 Kappa (κ) İstatistiği 28
2.3.7 Ortalama Mutlak Hata (MAE) 29
2.3.8 Kök Ortalama Kare Hata (RMSE) 29
2.3.9 Matthews Korelasyon Katsayısı (MCC) 30
2.4 k-katlı Çapraz Doğrulama 30

3 Uygulama 32
3.1 Deneysel Kurulum ve Uygulama 32

3.2 Weka ... 33
3.3 Fen Bilimleri için Veri Seti ... 33
3.3.1 Fen Bilimleri Veri Seti için Ölçekler ... 35
3.3.2 Fen Bilimleri Veri Seti için Uygulama ... 40
3.4 Matematik için Veri Seti ... 41
3.4.1 Matematik Veri Seti için Ölçekler ... 43
3.4.2 Matematik Veri Seti için Uygulama ... 49

4 Sonuçlar **51**
4.1 Fen Bilimleri Veri Seti için Sonuçlar ... 51
4.2 Matematik Veri Seti için Sonuçlar ... 67

5 Sonuç ve Öneriler **86**
5.1 Değerlendirme ve Tartışma 86

Kaynakça **91**

A Öznitelik ve Ölçek Bilgileri **99**

Tezden Üretilmiş Yayınlar **111**

SİMGE LİSTESİ

α_i	Destek vektör makinelerinde i. lagrange çarpanı
∂	Kısmi türev
ξ	Destek vektör makinelerinde soft hata
μ_i	Destek vektör makinelerinde güncellenmiş lagrange parametreleri
$K(x_i^T x_j)$	Destek vektör makinelerinde çekirdek fonksiyonu
$\bar{\gamma}$	Öznitelik alt kümesinin korelasyonu
κ	Kappa istatistiği

KISALTMA LİSTESİ

CFS	Correlation-Based Öznitelik Seçim Algoritması
ÇKA	Çok Katmanlı Algılayıcı
DN	Doğru Negatif
DP	Doğru Pozitif
DVM	Destek Vektör Makineleri
DVM-POLY	Destek Vektör Makinesi Poly Çekirdek Fonksiyonu (DVM-POLY)
DVM-PUK	Pearson VII Fonksiyon Tabanlı Çekirdek Fonksiyonu Evrensel
DVM-RBF	Destek Vektör Makinesi Radyal Tabanlı Çekirdek Fonksiyonu
EVM	Eğitim Veri Madenciliği
IEA	Uluslararası Eğitim Başarılarını Değerlendirme Birliği
IRT	Item Response Theory
KA-C4.5	C4.5 Karar Ağacı
KA-RepTree	RepTree Karar Ağacı
KA-RF	Random Forest Karar Ağacı
KKT	Karush-Kuhn-Tucker
k-NN	k-En Yakın Komşu Algoritması
LR	Lojistik Regresyon
MAE	Ortalama Mutlak Hata
MCC	Matthews Korelasyon Katsayısı
mRMR	Minimum Redundancy Maximum Relevance
NB	Naive Bayes
RMSE	Kök Ortalama Kare Hata
ROC	Alıcı İşlem Karakteristiği Eğrisi
TIMSS	Uluslararası Matematik ve Fen Eğilimleri Araştırması
YN	Yanlış Negatif
YP	Yanlış Pozitif
YSA	Yapay Sinir Ağları

ŞEKİL LİSTESİ

Şekil 2.1 Çok Katmanlı Algılayıcı Örneği 14
Şekil 2.2 DVM'de doğrusal olarak ayrılabilir durum için optimal ayırıcı hiper düzlem 17
Şekil 2.3 DVM'de doğrusal olarak ayrılamayan durum için optimal ayırıcı hiper düzlem 20
Şekil 2.4 Bir çekirdek fonksiyonu ile yüksek boyutlu bir özellik uzayına haritalama 21
Şekil 4.1 Fen Bilimleri veri seti adım 1 için başarılı algoritmaların ROC eğrileri 53
Şekil 4.2 Fen Bilimleri veri seti adım 2 için başarılı algoritmaların ROC eğrileri 55
Şekil 4.3 Fen Bilimleri veri seti adım 3.1 için başarılı algoritmaların ROC eğrileri 57
Şekil 4.4 Fen Bilimleri veri seti adım 3.2 için başarılı algoritmaların ROC eğrileri 59
Şekil 4.5 Fen Bilimleri veri seti adım 3.3 için başarılı algoritmaların ROC eğrileri 61
Şekil 4.6 Fen Bilimleri veri seti adım 3.4 için başarılı algoritmaların ROC eğrileri 63
Şekil 4.7 Fen Bilimleri veri seti adım 3.5 için başarılı algoritmaların ROC eğrileri 65
Şekil 4.8 Fen Bilimleri veri seti için sınıflandırma sonuçları 66
Şekil 4.9 Matematik veri seti adım 1 için başarılı algoritmaların ROC eğrileri 69
Şekil 4.10 Matematik veri seti adım 2 için başarılı algoritmaların ROC eğrileri 71
Şekil 4.11 Matematik veri seti adım 3.1 için başarılı algoritmaların ROC eğrileri 73
Şekil 4.12 Matematik veri seti adım 3.2 için başarılı algoritmaların ROC eğrileri 75
Şekil 4.13 Matematik veri seti adım 3.3 için başarılı algoritmaların ROC eğrileri 77
Şekil 4.14 Matematik veri seti adım 3.4 için başarılı algoritmaların ROC eğrileri 79
Şekil 4.15 Matematik veri seti adım 3.5 için başarılı algoritmaların ROC eğrileri 80
Şekil 4.16 Matematik veri seti adım 3.6 için başarılı algoritmaların ROC eğrileri 82
Şekil 4.17 Matematik veri seti için sınıflandırma sonuçları 84

TABLO LİSTESİ

Tablo 2.1 Çekirdek fonksiyonları ve formülleri 22
Tablo 2.2 Karşılaştırma tablosu 27
Tablo 3.1 Öğrencilerin Fen Bilimleri başarısının sınıflandırılmasında kullanılan tüm değişkenler 34
Tablo 3.2 Fen Bilimleri veri seti için ev eğitim kaynakları ölçeğinin içerdiği sorular 35
Tablo 3.3 Fen Bilimleri veri seti için öğrenci zorbalığı ölçeğinin içerdiği sorular 36
Tablo 3.4 Fen Bilimleri veri seti için öğrencilerin okula ait olma hissi ölçeğinin içerdiği sorular 37
Tablo 3.5 Fen Bilimleri veri seti için fen bilimlerindeki öğrenmeye olan sempatisi ölçeğinin içerdiği sorular 37
Tablo 3.6 Fen Bilimleri veri seti için öğrencinin fen bilimlerine verdiği değer ölçeğinin içerdiği sorular 38
Tablo 3.7 Fen Bilimleri veri seti için fen bilimlerinde öğrencinin güveni ölçeğinin içerdiği sorular 39
Tablo 3.8 Fen Bilimleri veri seti için tahmin için fen bilimleri başarısı çok düşük ölçeğinin içerdiği soru 39
Tablo 3.9 Öğrencinin fen bilimlerini öğrenmeyi sevmesi ölçeğinin içerdiği sorular 40
Tablo 3.10 Fen Bilimleri veri seti için fen bilimleri ödevinde haftalık harcanan zaman ölçeğinin içerdiği soru 40
Tablo 3.11 Öğrencilerin Matematik başarısının sınıflandırılmasında kullanılan tüm değişkenler 42
Tablo 3.12 Matematik veri seti için ev eğitim kaynakları ölçeğinin içerdiği sorular 44
Tablo 3.13 Matematik veri seti için öğrencilerin okula ait olma hissi ölçeğinin içerdiği sorular 44
Tablo 3.14 Matematik veri seti için öğrenci zorbalığı ölçeğinin içerdiği sorular 45
Tablo 3.15 Matematik veri seti için öğrencinin matematiği öğrenmeyi sevmesi ölçeğinin içerdiği sorular 45
Tablo 3.16 Matematik veri seti için matematik öğrenmeye olan sempatisi ölçeğinin içerdiği sorular 46
Tablo 3.17 Matematik veri seti için matematikte öğrencinin güveni ölçeğinin içerdiği sorular 47
Tablo 3.18 Matematik veri seti için öğrencinin matematiğe verdiği değer ölçeğinin içerdiği sorular 48
Tablo 3.19 Matematik veri seti için tahmin için matematik başarısı çok düşük ölçeğinin içerdiği soru 48
Tablo 3.20 Matematik veri seti için matematik ödevinde haftalık harcanan zaman ölçeğinin içerdiği soru 49
Tablo 4.1 Fen Bilimleri veri seti adım 1 için karşılaştırma matrisi 52
Tablo 4.2 Fen Bilimleri veri seti adım 1'e göre sınıflandırma sonuçları 52
Tablo 4.3 Fen Bilimleri veri seti adım 2 için karşılaştırma matrisi 54
Tablo 4.4 Fen Bilimleri veri seti adım 2'ye göre sınıflandırma sonuçları 54
Tablo 4.5 Fen Bilimleri veri seti adım 3.1 için karşılaştırma matrisi 56
Tablo 4.6 Fen Bilimleri veri seti adım 3.1'e göre sınıflandırma sonuçları 56

Tablo 4.7 Fen Bilimleri veri seti adım 3.2 için karşılaştırma matrisi 58
Tablo 4.8 Fen Bilimleri veri seti adım 3.2'ye göre sınıflandırma sonuçları 58
Tablo 4.9 Fen Bilimleri veri seti adım 3.3 için karşılaştırma matrisi 60
Tablo 4.10 Fen Bilimleri veri seti adım 3.3'e göre sınıflandırma sonuçları..... 60
Tablo 4.11 Fen Bilimleri veri seti adım 3.4 için karşılaştırma matrisi 62
Tablo 4.12 Fen Bilimleri veri seti adım 3.4'e göre sınıflandırma sonuçları..... 62
Tablo 4.13 Fen Bilimleri veri seti adım 3.5 için karşılaştırma matrisi 64
Tablo 4.14 Fen Bilimleri veri seti adım 3.5'e göre sınıflandırma sonuçları..... 64
Tablo 4.15 Fen Bilimleri veri seti adım 1 - adım 3 arasında genel sonuç tablosu..... 66
Tablo 4.16 Matematik veri seti adım 1 için karşılaştırma matrisi 68
Tablo 4.17 Matematik veri seti adım 1'e göre sınıflandırma sonuçları..... 68
Tablo 4.18 Matematik veri seti adım 2 için karşılaştırma matrisi 70
Tablo 4.19 Matematik veri seti adım 2'ye göre sınıflandırma sonuçları 70
Tablo 4.20 Matematik veri seti adım 3.1 için karşılaştırma matrisi..... 72
Tablo 4.21 Matematik veri seti adım 3.1'e göre sınıflandırma sonuçları 72
Tablo 4.22 Matematik veri seti adım 3.2 için karşılaştırma matrisi..... 74
Tablo 4.23 Matematik veri seti adım 3.2'ye göre sınıflandırma sonuçları..... 74
Tablo 4.24 Matematik veri seti adım 3.3 için karşılaştırma matrisi..... 76
Tablo 4.25 Matematik veri seti adım 3.3'e göre sınıflandırma sonuçları 76
Tablo 4.26 Matematik veri seti adım 3.4 için karşılaştırma matrisi..... 78
Tablo 4.27 Matematik veri seti adım 3.4'e göre sınıflandırma sonuçları 78
Tablo 4.28 Matematik veri seti adım 3.5 için karşılaştırma matrisi..... 79
Tablo 4.29 Matematik veri seti adım 3.5'e göre sınıflandırma sonuçları 80
Tablo 4.30 Matematik veri seti adım 3.6 için karşılaştırma matrisi..... 81
Tablo 4.31 Matematik veri seti adım 3.6'ya göre sınıflandırma sonuçları..... 82
Tablo 4.32 Matematik veri seti adım 1 - adım 3 arasında genel sonuç tablosu 83

ÖZET

Makine Öğrenmesi Yöntemleri ve Eğitim Verisi Üzerine Bir Uygulama: Uluslararası Matematik ve Fen Eğilimleri Araştırması 2015 TÜRKİYE Örneği

Enes FİLİZ

İstatistik Anabilim Dalı

Doktora Tezi

Danışman: Doç. Dr. Ersoy ÖZ

Eğitim alanında makine öğrenmesi, eğitim verilerini analiz etme ve modelleme yeteneğine sahip önemli bir araştırma alanıdır. Elde edilen çıktılar araştırmacıların ve eğitim planlamacılarının mevcut eğitim stratejilerinin sistematik sorunlarını anlamalarına ve gözden geçirmelerine yardımcı olmaktadır. Bu çalışma Uluslararası Matematik ve Fen Bilimleri Eğilimleri Araştırması (TIMSS) ile ilgilidir. TIMSS-2015 raporunda yer alan 8. sınıf Türk öğrencilerinin fen bilimleri ve matematik veri setine makine öğrenmesi algoritmaları uygulanmaktadır. Tez çalışması temel olarak iki noktaya odaklanmaktadır: ilk olarak öğrencilerin fen bilimleri ve matematik başarısını sınıflandırmada en başarılı algoritmaları bulmak; ikinci olarak ta fen bilimleri ve matematik başarısını sınıflandırmada en önemli öznitelikleri ortaya çıkarmaktır. Fen bilimleri ve matematik veri seti için lojistik regresyon, destek vektör makinesi polynomial çekirdeği en iyi algoritmalar olarak belirlenmiştir. Bu iki veri setindeki tüm öznitelikler arasında sınıflandırmadaki en etkili özellikleri belirlemek için farklı öznitelik seçim yöntemleri kullanılmıştır. Fen bilimleri veri seti için “Paylaşımlı tablet bilgisayar” , “Eğitimde ilerlemek istediğiniz düzey”, “Kaç ay ek ders alındığı” ve “Fen bilimlerinde öğrencinin güveni” öznitelikleri; matematik veri seti için ise “Ev eğitim kaynakları” ve “Matematikte öğrencinin güveni” öznitelikleri sınıflandırmada en etkili öznitelikler olarak belirlenmiştir.

Anahtar Kelimeler: Sınıflandırma algoritmaları, makine öğrenmesi, fen bilimleri başarısı, matematik başarısı, TIMSS-2015.

ABSTRACT

Machine Learning Methods and an Application on Educational Data: The Trends in International Mathematics and Science Study 2015 TURKEY Case

Enes FİLİZ

Department of Statistics

Doctoral Thesis

Advisor: Assoc. Prof. Dr. Ersoy ÖZ

Machine learning in the field of education is an important research area which has an ability of analyzing and modeling educational data. Obtained outputs help researchers and education planners understand and revise the systematic problems of current educational strategies. This thesis study deals with an important international study, namely Trends International Mathematics and Science Study (TIMSS). Machine learning methods are applied to TIMSS-2015 8th grade Turkish students' data. The study has mainly twofold: to find best performer algorithm(s) for classifying students' science and mathematics success and to extract important features on success. The most appropriate algorithm is found as logistic regression and support vector machines with polynomial kernel for the science and mathematics datasets. Different feature selection methods are used in order to extract the most effective features in classification among all features in the original datasets. "Computer Tablet Shared", "How Far in Education Do You Expect to Go", "Extra Lessons How Many Month" and "Student Confident in Science" are found the most important features in all feature selection methods for science dataset. "Home Educational Resources" and "Student Confident in Mathematics" are found the most important features in all feature selection methods for mathematics dataset.

Keywords: Classification algorithms, machine learning, science success, mathematic success, TIMSS-2015.

1

Giriş

Teknolojik ve bilimsel gelişmelerin artması ile birlikte veri ifadesinin popülerliği ileriye taşınmıştır. Veri ifadesinin bu denli popüler olması veriden elde edilen bilginin önemini ortaya koymaktadır. Verilerin toplanması ardından analiz edilmesi ve sonucunda yararlı bilginin ortaya çıkarılması veri ile bilgi arasındaki kuvvetli ilişkiyi göstermektedir. Veriyi işlemek ve bilgiye ulaşmak için kullanılan en önemli yöntemlerden olan veri madenciliği ve makine öğrenmesi veri sayısının giderek artmasıyla sıklıkla kullanılmaya başlanmıştır. Toplanan verinin miktarı büyüdükçe, boyutu ve karmaşıklığının da artması ile birlikte, veri madenciliği ve makine öğrenmesi teknolojik ilerleme için gerekli bir bileşen olarak ortaya çıkmaktadır. Günümüzde veri madenciliği ve makine öğrenmesi, bilgi teknolojisinin temel dayanak noktalarından biri haline gelmiş ve bununla birlikte yaşamımızın bir parçası olmuştur. Fazla sayıda verinin işlenmesi, analiz edilmesi önemli sonuçlar ortaya çıkarmaktadır. Makine öğrenmesi, birçok farklı istatistiksel prosedürü içeren disiplinlerarası bir çalışma alanıdır. Makine öğrenmesi yöntemlerinin standart istatistik yöntemlerdeki gibi varsayımlarının olmaması yöntemlerinin tercih nedenlerinden birisidir. Örneğin, doğrusal regresyon gibi geleneksel istatistiksel yöntemlerden farklı olarak, makine öğrenmesi yöntemleri doğrusallık, homojenlik veya normallik varsayımları gerektirmez. Ayrıca veri madenciliği ve makine öğrenmesi, eldeki verileri bilgiye dönüştürmek; kümelemek, sınıflandırma yapmak, tahmin etmek için büyük veri setlerine uygulamakla ilgilidir [1-3]. Kullanıcılar veya tüketiciler artık daha akıllı cihazlar talep ederken, akıllı sistemler makine öğrenmesini kullanarak gelişimlerine devam etmektedir. Akıllı sistemlerin bir parçası olarak makine öğrenmesi, arama motorları, eğitimin gelişimi, sağlık alanındaki gelişmeler, ekonomik analizler gibi araçların en kritik bileşenlerinden birisi olarak ortaya çıkmaktadır. Eksiklikleri gidermek, yanlışları düzeltebilmek için makineler eğitilebilir, böylece makineler çeşitli sorunları otomatik olarak algılayabilir, tanılayabilir ve çözüm üretebilir.

Akıllı sistemler, akıllı ve derin algılara dayalı makine öğrenmesinde son teknolojiyi geliştirmede hızlı bir ilerleme kaydetmiştir [4].

Makine öğrenmesi eğitim, tıp, astronomi, finans gibi farklı alanlarda sıklıkla uygulanmaktadır. Bu alanlardan eğitim alanında kullanılabilen makine öğrenmesi algoritmaları Eğitim Veri Madenciliği (EVM) olarak adlandırılır. EVM, eğitimcilere ve eğitim planlamacılarına büyük ve karmaşık eğitim veri kümelerini daha iyi anlamalarını ve bu veri kümelerinden çıkardıkları sonuçlara göre durum tespiti yapmalarına ve gelecek için planlamalar yapmalarına olanak sağlar. Temel istatistiksel analizlerle, veri kümelerinin genel davranışları yorumlanabilir. Bununla birlikte, bilinmeyen veya erişilemeyen yararlı bilgilerin keşfi EVM yöntemleri kullanılarak yapılabilir. Çıkarılan faydalı bilgiler, eğitimciler ve eğitim alanındaki karar vericiler tarafından kullanılabilir. Ayrıca, bu bilgiler eğitim sisteminin en önemli bileşeni olan öğrencilerin başarılarının mevcut durumunu izler ve bu nedenle hatalı eğitim stratejilerine çözümler önerilebilir.

EVM'yi etkin bir şekilde kullanmak için, güvenilir veri kümeleriyle çalışmak hayati bir rol oynar. Uluslararası Eğitim Başarılarını Değerlendirme Birliği (IEA: International Association for the Evaluation of Educational Achievement), katılımcı ülkelerin eğitim sistemlerinin politikalarının ve uygulamalarının etkilerini izlemek amacıyla güvenilir veri setlerini elde etmek ve raporlamak için birçok karşılaştırmalı çalışma yürütmektedir. Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS: Trends in International Mathematics and Science Study), öğrenci başarısını uluslararası düzeyde değerlendiren en büyük IEA projesidir ve 60'tan fazla ülke bu projeye katılmaktadır. Proje her dört yılda bir dördüncü ve sekizinci sınıf matematik ve fen öğrencileriyle gerçekleştirilmektedir. TIMSS, yalnızca eğitim politikalarının etkileri ve uygulamaları ile ilgili güvenilir bilgi sağlamakla kalmaz, aynı zamanda katılımcı ülkelerin öğrenci başarısı açısından, katılımcı ülkelerin sonuçları arasında karşılaştırma yapmalarını sağlar [5].

TIMSS verileri kullanılarak önemli çalışmalar yapılmış olmasına rağmen, bu çalışmaların birçoğunda regresyon ve faktör analizi gibi farklı klasik istatistiksel yöntemler kullanmıştır. Ancak, bu yöntemlerin, özellikle büyük ve karmaşık veri kümeleri için bazı sınırlamaları vardır. Ayrıca, normallik, değişkenlerin

homojenliği ve doğrusallık gibi klasik istatistiksel varsayımların yokluğu EVM'yi daha kullanılabilir hale getirmiştir [6].

1.1 Literatür Özeti

Makine öğrenmesinin geçmişi çok eski yıllara dayanmamakla beraber bilgisayar sistemlerinin gelişmesine paralel olarak ve teorisinde yaşanan gelişmelerden dolayı günümüzde yaygın olarak kullanılmaktadır. Makine öğrenmesi yönteminin eğitim, astronomi, biyoloji, finans, pazarlama, sigorta ve tıp gibi birçok alanda uygulaması mevcuttur. Veri madenciliği ile makine öğrenmesinin iç içe olduğunu ortaya koyan Mitchell, 1997 yılında makine öğrenmesinin literatürde görülmüş en iyi tanımlarından birini yapmıştır. Bir başka çalışmada metin sınıflandırmasına Naive Bayes (NB) varsayımını kullanan iki farklı sınıflandırma modeli kullanılmıştır [7, 8]. Farklı bir çalışmada ensemble yönteminin, tek bir sınıflandırıcıdan daha iyi performans gösterdiğini açıklanmıştır [9]. Önemli bir diğer çalışma ise danışmanlı öğrenme yöntemlerine dayalıdır. Pang, Lee ve Vaithyanathan, destek vektör makineleri (DVM), NB ile maksimum entropi sınıflandırıcılarını kullanarak film değerlendirmeleri için yaptıkları çalışmada "Movie Review" veri setine uygulamıştır [10]. Kavzaoğlu ve Çölkesen, çalışmalarında DVM ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerini araştırmışlardır. En yüksek performansı gösteren radyal tabanlı ve Pearson VII (>%94) kernel fonksiyonları olarak belirlemişlerdir [11]. Başka bir çalışmada makine öğrenmesinde kullanılan NB, Random Forest karar ağacı (KA-RF) ve DVM algoritmaları yardımıyla twitter üzerine duygu analizi yapılmış ve çalışmanın sonuçları karşılaştırılmalı olarak verilmiştir [12].

Makine öğrenmesi yöntemleri ile ilgili çalışmalar arttıkça literatür taramasına yönelik çalışmalarda yapılmaya başlamıştır. Blum ve Langley, yaptıkları çalışmada önemsiz bilgi içeren veri setlerinin kullanımı hakkında makine öğrenmesi yöntemleri ile yapılan çalışmaları incelemişlerdir. Makine öğrenmesinde hem ampirik hem de teorik çalışmalarda bu konulardaki ilerlemelerden bahsedilmiş ve farklı yöntemleri karşılaştırmak için genel bir çerçeve göstermişlerdir [13]. Sebastiani, çalışmasında makine öğrenmesi içinde yer alan metin kategorizasyonuna yönelik temel yaklaşımları incelemiştir [14]. Liao, çalışmasında

bilgi yönetimi teknolojilerinin ve uygulamalarının 1995-2002 yılları arasında nasıl geliştiğini keşfetmek için bir literatür taraması yaparak gelişimini araştırmıştır. 234 bilgi yönetimi uygulaması makalesine dayanarak, bilgi yönetimi teknolojilerini incelemiş ve sınıflandırmıştır [15]. 1991-2010 yılları arasında makine öğrenmesi modelinde ampirik çalışmaların sistematik bir literatür taramasını yapmışlarıdır [16]. Ayrıca makine öğrenmesi yöntemlerinin görüş madenciliğinde kullanılması üzerine bir literatür çalışması yapmıştır. Onan ve Korukoğlu, görüş madenciliğinde makine öğrenmesinin güçlü ve zayıf tarafları ortaya konmuştur [17].

Makine öğrenmesi kavramı, tüm dünyada eğitim araştırmaları alanında kullanılmıştır. Mevcut eğitim yöntemlerini incelemek, analiz etmek; gelecek için yeni ve daha iyi teknikler geliştirmek için kullanılan en önemli yaklaşımlardan biri haline gelmiştir. Araştırmacılar yaptıkları çalışmalar ile öğrencilerin öğrenme süreçleri ile ilgili daha iyi bir fikre sahip olur ve mevcut veri setlerini daha iyi anlayabilirler. Makine öğrenmesi sadece araştırmacılara yardımcı olmakla kalmaz, aynı zamanda pratik yaklaşımlarla faydalı bilgileri yayarak öğrencileri de destekler [18]. Eğitim konusuyla ilgili mevcut literatür ve verilerin kapsamlı bir incelemesi, 2007 yılında Romero ve Ventura tarafından yapılmış ve makine öğrenmesinin test etme, hipotez oluşturma ve öğrenci performansında gelişim için tekrarlı bir araç olarak hareket ettiği sonucuna varmışlardır. Eğitimciler de makine öğrenmesi tarafından oldukça fazla değer kazanabilecek öğrencilerin değerlendirilmesiyle ilgili bilgileri ayırt ettiklerinden çok fazla kazanç sağlarlar. Makine öğrenmesinin tanımları, EVM'nin kullanımı ve gelecekteki makine öğrenmesi kullanımı hakkındaki araştırma yazılarında tartışılmıştır [18, 19]. EVM hakkındaki güncel literatürde, EVM'nin tarihçesi ve kuruluşundan bu yana meydana gelen değişiklikleri, EVM alanındaki son gelişmeler ve yıllar içinde nasıl büyüdüğü, veri madenciliği ve analitik yöntemlerin eşzamanlı uygulaması yer almaktadır [20-22]. Peña-Ayala tarafından yapılan bir başka önemli çalışmada EVM için iki yönlü bir yaklaşım sunulmuştur. Bunlardan birincisi, yıllar boyunca EVM'nin gelişim tarihine dikkat çekerken, çalışmanın diğer kısmı analizlerini ve kullandıkları veri madenciliği yönteminin sonuçlarını içermektedir [23].

Daha önce belirtildiği gibi, makine öğrenmesinde veri kümelerini kategorize etmek, sonuçlarını analiz etmek ve tahmin yapmak için kullanılan çeşitli

algoritmalar mevcuttur. Makine öğrenmesinde kullanılan algoritmalar ile ilgili yapılan çeşitli çalışmalar bulunmaktadır. Kotsiantis, Pierrakeas ve Pintelas çalışmalarında çeşitli makine öğrenmesi tekniklerinin karşılaştırmalı bir değerlendirmesini yapmış ve NB algoritmasının bir yazılım destek aracı geliştirmek için en uygun olduğunu belirtmişlerdir [24]. Bir diğer çalışmada, 2002'den 2015'e kadar en başarılı öğrenciler için performans tahmini yöntemlerini bulmak için literatür incelemesi yapılmıştır. En sık kullanılan makine öğrenmesi yöntemlerinin sırasıyla Yapay Sinir Ağları (YSA), Karar ağacı (KA), DVM, k-en yakın komşu algoritması (k-NN) ve NB olduğunu göstermişlerdir [25]. Ramaswami ve Bhaskaran, 2010 yılında akademik performansı incelemek için bir anket yapılmış ve hem öğrencilere hem de okul müdürlerine uygulanmıştır. Ki-kare tahmin modellerini kullanarak, anket tahmin kurallarını ortaya koymuşlardır [26]. Baradwaj ve Pal tarafından yapılan bir başka araştırmada, öğrenci performansını değerlendirmek için makine öğrenmesi sınıflandırma yöntemlerinin yeterli olduğunu göstermeyi amaçlamışlardır. Çalışmada karar ağacı yöntemi kullanılmıştır [27]. Rajni ve Malaya, özellikle yükseköğrenim kategorisine yönelik bir başka çalışmada, makine öğrenmesini eğitim bölümü için uygun bir model olarak ortaya koymuşlardır. Bu çalışmada, mühendislik giriş sınavı verilerini tahmin etmek ve mühendislik eğitimi planlama problemini çözmek için YSA ve KA modelleri kullanılmıştır [28]. Bazı çalışmaların farklı sonuçlara yol açtığı gözlemlenmiştir. Örneğin, KA'nın, diğer yöntemlerden daha iyi sonuçlar verdiğini ve yorumlandığını göstermiştir. Martinez Abad ve Chaparro Caso López tarafından yapılan çalışmada, KA tekniklerine dayanarak, akademik başarı faktörlerinin makine öğrenmesi yöntemleri kullanılarak yapılan istatistiksel analizlerle elde edilebileceğini ve kişisel faktörlerin akademik performansta önemli bir rol oynadığını öne sürmüştür [29]. Portekiz'e yönelik bir çalışmada, ortaokul öğrencilerinin performanslarına yönelik model geliştirmek için YSA, KA-RF ve DVM kullanmış; bir öğrencinin geçmiş performansının ve akademik başarısının yakın bir ilişki içinde olduğu sonucunu göstermiştir; aynı zamanda, en iyi tahmin modelinin böyle bir durumda ne olacağını göstermiştir [30]. Horáková, Houška ve Dömeová, YSA'nın 120 metin örneği için sınıflandırma ve regresyon ağaçlarına kıyasla daha doğru bir tahmin yöntemi olduğunu bulmuştur [31]. Osmanbegović ve Suljić, Tuzla Üniversitesi öğrencilerinin başarı seviyelerini tahmin etmek için

veri madenciliği yöntemlerini kullanmıştır. Web tabanlı öğrenme sisteminin, öğrencinin yaratıcılığının gelişiminin önemli bir parçası olduğu kanıtlamışlardır [32]. Delen, öğrenciler arasında yıpranmayı tahmin etmede analitik modeller geliştirmek için YSA, DVM, C5 karar ağacı algoritması ve LR kullanmış; DVM'nin bu alanda en iyi sonuçları verdiğini ortaya koymuştur [33]. Gorostiaga ve Rojo-Álvarez, PISA 2009 veri seti yardımıyla İspanya'da ki öğrencilerin matematik sonuçlarını incelemiştir. LR, Fisher ayırıcı analizi ve DVM algoritmaları yardımıyla matematik sonuçlarını sınıflandırmışlardır. DVM'nin diğer algoritmalardan daha iyi performans gösterdiğini belirlemişlerdir [34].

Dünyanın çeşitli bölgelerindeki birçok ülkeden gelen eğitim verilerini inceleyen TIMSS vb. çalışmalar, bu tür çalışmalara katılan ülkelerin eğitim politikalarının iyileştirilmesine ve diğer ülkeler tarafından takip edilen eğitim normlarına kıyasla öğrencilerin başarı oranlarının iyileştirilmesine yardımcı olur. TIMSS-2011 veri setini kullanarak eğitim başarısı için yapılan bir çalışmada, öğrencinin güveninin öğrenci başarısında belirleyici faktör olduğunu ortaya konmuştur. Tahmin ve sınıflandırma performansını hesaplamak için Lojistik Regresyon (LR) ve YSA teknikleri kullanılmıştır [35]. Schreiber, öğrencilerin akademik performansını belirlemek için çeşitli veri madenciliği yöntemleri arasında en iyi sınıflandırma algoritması olarak LR algoritmasını belirlemiştir [36].

TIMSS çalışmasını, çeşitli ülkeler, farklı yaş grupları ve farklı yıllar açısından incelemek için çok sayıda çalışma yapılmıştır. Bu araştırmalarda amaca yönelik olarak çeşitli teknikler kullanmıştır. Kiray, Gök ve Bozkır çalışmalarında, TIMSS-1999, PISA-2003 ve PISA-2006 veri kümelerinde fen ve matematik başarı ile ilgilenmişlerdir. Sonuca ulaşmak için KA ve kümelemeyi kullanmışlardır [37]. Bir diğer çalışmada TIMSS-2007 veri setinden Tayvan'daki sekizinci sınıf öğrencilerinin fen ve matematik başarıları incelenmiştir. Karar ağaçlarından CART, ID3, CHAID, C4.5 ile NB, YSA, k-NN ve parçacık sürüsü optimizasyonu kullanılmıştır [38]. Topçu, Erbilgin ve Arıkan tarafından yapılan çalışmada ise Türk ve Koreli öğrencilerin fen ve matematik sınavlarındaki başarı faktörlerini tespit etmek için TIMSS-2011 verileri ele alınmıştır. TIMSS-2011 verilerine NB, karar ağaçları, YSA ve LR uygulayarak Türkiye'den sekizinci sınıf öğrencilerinin başarılarının sınıflandırılmasına yardımcı olacak en iyi algoritmaları bulmak için bir analiz

yapılmıştır [39, 40]. Hammouri, çalışmasında öğrencilerin matematik başarısını etkileyen faktörleri bulmayı amaçlamıştır. TIMSS veri seti yardımıyla elde edilen sonuçlara göre tutum, başarı, matematiksel yeteneklere duyulan güven ve matematiğin önemine ilişkin algı, öğrencinin matematik başarısını belirlediği ortaya çıkarılmıştır [41]. Diğer bir çalışmada, TIMSS-2003 veri setindeki faktörler incelenmiş ve bu faktörleri kullanarak Doğu Asya ve Amerika'daki yüksek ve düşük başarılı öğrencilerin matematiksel farkındalık kavramları araştırılmıştır [42]. Yoo, TIMSS-2011 veri setinden yaralanarak, Kore'deki 4. sınıf öğrencilerinin matematik puanlarını hedef kitle olarak belirlemiş ve LR yönteminden yaralanarak sonuçları incelemiştir [43].

Makine öğrenmesi algoritmaları, makine tarafından okunabilen bilgilerden sonuç çıkarır. Bu sonuçlar eldeki verinin kalitesine bağlıdır. Veriler yetersiz, alakasız ya da gereksiz bilgiler içeriyorsa, makine öğrenmesi algoritmalarının performansları doğruluğu düşük sonuçlar üretebilir. Öznitelik seçim algoritmaları, olabildiğince fazla alakasız ve gereksiz bilgiyi tanımlamaya ve bu gereksiz bilgileri ortaya çıkarmaya çalışan algoritmalardır. Yapılan bir çalışmada öznitelik seçiminin "iyiliğini" belirlemek için korelasyona dayalı bir sezgisel tarama kullanan yeni bir öznitelik seçim algoritması olarak correlation based (CFS) öznitelik seçim algoritması açıklanmış ve etkinliği üç makine öğrenme algoritması ile değerlendirilmiştir. Öznitelik seçimi ile üç algoritma için önemli iyileştirmeler sağlanmıştır [44]. Jiang, Meng ve Meng, çalışmalarında sonsuz değer niteliğindeki seçicilik tahminine odaklanmışlardır. Correlation attribute öznitelik seçim algoritmasından yararlanarak web veritabanına en az bağıntılı öznitelikleri araştırmışlardır [45]. Novakovic, çalışmasında öznitelik seçim algoritmalarının NB sınıflandırıcısının doğruluğu üzerindeki etkisini araştırmıştır. Altı öznitelik seçim algoritması arasından One-R algoritmasının NB'yi en iyi duruma getirdiği belirlenmiştir [46]. Bir başka çalışmada yüksek boyutlu veri setleri üzerinde çalışılmıştır. Bu çalışmada gain ratio öznitelik seçim algoritması kullanılmış ve veri kümelerinin nitelikleri sıralanmıştır [47]. Diğer bir çalışmada öznitelik seçim algoritmalarının karşılaştırmaları yapılmıştır. Bu çalışmada CFS öznitelik seçim algoritması, fast correlation-based filter yöntemi, interact algoritması, information gain öznitelik seçim algoritması, reliefF öznitelik seçim algoritması ve minimum redundancy maximum relevance (mRMR) yöntemi kullanılmıştır [48]. Öznitelik

seçimi, makine öğrenmesi sürecinde etkin bir rol oynar. Genel olarak, veri kümesi fazla sayıda nitelik içerir. Ancak, etkili sınıflandırma sürecinde bazı veriler çok bilgi içerirken bazıları daha az bilgi içerir. Buradan yola çıkarak Gnanambal, Thangaraj, Meenatchi, & Gayathri tarafından yapılan çalışmada bilgi analizi için WEKA (Waikato Environment for Knowledge Analysis) programı kullanılarak öznitelik seçiminden önce ve sonra sınıflandırma algoritmalarının başarılarını karşılaştırmışlardır [49]. Takcı, çalışmasında kalp krizi tahmini için makine öğrenmesi yöntemlerini kullanan bilgisayar destekli tanı sistemlerini incelemiştir. Kalp krizi geçirmeyi öngören en iyi makine öğrenmesi yöntemini ve en iyi öznitelik seçim algoritmasını bulmayı amaçlamıştır. Araştırmanın sonucunda, en iyi makine öğrenmesi algoritması doğrusal çekirdekli destek vektör makinesi, en iyi öznitelik seçim algoritması da reliefF olarak belirlenmiştir [50]. Ogura, öğrencilerin başarılarını etkileyen öznitelikler (faktörler) ile ilgilenmiştir. Çalışmanın sonucunda okul dışında harcanan fazladan zamanın, öğrencilerin başarı düzeyleri üzerinde önemli bir öznitelik olduğu belirlenmiştir [51]. Anıl, PISA-2006 Türkiye için öğrencilerin fen bilimleri verilerini incelemiş ve "Bilgisayar Ortamı" özniteliğinin öğrencilerin başarı için önemli bir değişken olduğunu ortaya koymuştur [52]. Özer ve Anıl, PISA 2006'da ki Türkiye fen bilimleri çalışmasına odaklanmıştır. Okul dışında, fen bilimleri derslerine harcanan fazladan zamanın, PISA 2006'da ki Türkiye fen bilimleri çalışmasında önemli bir unsur olarak belirlenmiştir [53].

1.2 Tezin Amacı

Bu çalışma, EVM çerçevesinde TIMSS çalışmalarının güncel literatürüne katkıda bulunmayı amaçlamaktadır. Tez çalışmasında başlıca iki araştırma sorusu ele alınmaktadır: (1) TIMSS-2015 Türkiye sekizinci sınıf öğrencilerinin fen ve matematik başarılarının sınıflandırılması için hangi EVM yönteminin daha uygun olduğunu belirlemek ve (2) hangi özniteliklerin öğrencilerin başarılarını belirlemede etkili olduğunu ortaya çıkartmak.

İlk araştırma sorusu olan en uygun EVM yöntemini bulmak için, EVM çalışmalarında literatürde en sık kullanılan algoritmalar seçilmiştir. Bunlar; k-NN,

NB, YSA, KA-RepTree, KA-RF ve KA-C4.5 algoritmaları, DVM-POLY, DVM-RBF, DVM-PUK çekirdek fonksiyonları ve LR analizi şeklindedir.

Bu çalışmadaki ikinci araştırma sorusu, sınıflandırmada en etkili öznitelikleri bulmaktır. Eğitim çeşitli değişkenlerden etkilendiğinden, öğrencilerin performansları için EVM algoritmalarının uygulanması, öğrencilerin zayıf performanslarını tanımak ve etkili özniteliklerin eğitim üzerindeki etkilerini araştırmak için esastır [26]. Orijinal TIMSS veri seti, öğrencilerin başarısına yönelik birçok potansiyel etkileyici öznitelikleri içerir ve burada en etkili olanları ortaya koymak çok önemlidir. Doğru tanımlama yapmak, karar vericilere mevcut durumları iyileştirmelerine yardımcı olur. Bu çalışmada öğrencilerin başarısında etkili olan öznitelikleri ortaya çıkarmak için öznitelik seçim algpritması olarak reliefF, CFS, info gain, correlation attribute, gain ratio ve one-r kullanılmıştır. Ayrıca, sınıflandırma algoritmalarındaki değişkenlerin sayısını, maksimum sınıflandırma performansını korurken; mümkün olduğunca düşük tutmak esastır. Daha az öznitelik kullanılarak sınıflandırma performansında önemli bir performans kaybı olmamasını sağlamak öznitelik seçiminin gerekliliğini ortaya koymaktadır. Bunun sonucu olarak makine öğrenmesi algoritmalarının sınıflandırma yapması, veri setindeki özniteliklerin seçimine büyük özen gösterilmesini ve sadece etkili olanların dâhil edilmesini önerir.

1.3 Hipotez

TIMSS çalışmalarının güncel literatürüne katkıda bulunmayı amaçlayan bu çalışmada makine öğrenmesi yöntemlerinin öğrencilerin başarılarının sınıflandırılmasında kullanılabileceği gösterilmek istenmektedir. Ayrıca öznitelik seçimi yardımı ile elde edilecek daha az sayıdaki öznitelik ile yapılacak sınıflandırmalarda sınıflandırma algoritmalarının performansları için önemli düşüşler olmayacağı beklenmektedir.

TIMSS-2015 veri setinin son yayımlanan (tezin savunma tarihi itibarı ile) veri seti olması bu çalışmanın özgün noktalarından biridir. TIMSS-2015 veri setinde Türkiye çerçevesinde sekizinci sınıf öğrencilerinin başarılarının sınıflandırılmasında çok sayıda makine öğrenmesi ve öznitelik seçimi algoritması kullanılması bu çalışma açısından oldukça önemlidir.

Verilen iki araştırma sorusu detaylı bir şekilde incelenerek, TIMSS-2015 Türkiye sekizinci sınıf öğrencilerinin fen bilimleri ve matematik başarılarının sınıfandırılmasında etkili çözümler ortaya konulacağı savunulmaktadır.

2

Makine Öğrenmesi

2.1 Makine Öğrenmesi Algoritmaları

Teknolojinin ilerlemesi ve verilerin dijital ortamda saklanması ile birlikte, veri miktarı gün geçtikçe artmaktadır. Buna bağlı olarak veri tabanlarının sayısında da artış gerçekleşmiştir. Ayrıca, veri saklama hem daha basit bir hale gelmiş hem de verinin kendisi daha ulaşılabilir hale gelmiştir. Veri tabanı teknolojisi, istatistik, yapay zekâ, makine öğrenmesi, örüntü tanımlama ve veri görselleştirmesi gibi pek çok teknik alan arasında köprü görevi gören veri madenciliği, çok disiplinli bir alandır. Veri madenciliği bu tezin giriş bölümünde anlatıldığı üzere eğitim, astronomi, biyoloji, finans, pazarlama, sigorta ve tıp gibi birçok alanda uygulanmaktadır [54].

Veri madenciliğinin bir alt kolu olan makine öğrenmesi, veriden öğrenmeye dayalı yöntemler olarak düşünülebilir. Farklı bir bakış açısıyla bilgisayarın kendi kendine problemi çözmeyi öğrenmesi olarak tanımlanabilir. Makine öğrenmesi, elde edilen çıktıların sınıflandırılmasını, kümelenmesini ya da tahmin edilmesini sağlayacak algoritmaları barındırır. Makine öğrenmesinde farklı algoritmalar kullanılır. Eğer veri setinde çıktı biliniyorsa danışmanlı, bilinmiyorsa danışmansız algoritmalardan yararlanılır [55].

Veri madenciliği yöntemleri temel olarak tahmin, sınıflandırma, kümeleme ve birliktelik kuralları şeklinde 4 bölümden oluşur. Tahmin yöntemleri olarak Regresyon analizi, Bayes ağları, LR, KA ve YSA; sınıflandırma için k-NN, NB, YSA, KA, DVM ve LR; kümeleme için k-means, modele dayalı kümeleme, tam bağlantı kümeleme; birliktelik kuralları için Apriori, Carma, Sequence, GRI, Eclat, FP-Growth gibi yöntemler kullanılır.

Klasik istatistiksel yaklaşımlardaki varsayımlar aranmamasına rağmen makine öğrenmesi yöntemlerini etkileyen bazı faktörler mevcuttur. Bu faktörler, sırasıyla

veri seti, sonuca etki eden değişkenlerin bulunması, öğrenme stratejisinin belirlenmesi, algoritma ve algoritmaya ait parametrelerin belirlenmesidir.

Makine öğrenmesi hayatımızın her alanında karşımıza çıkmaktadır. Literatürde önerilen algoritmalar makine öğrenmesinin temelini oluşturmaktadır. Bu algoritmalar sınıflandırma, kümeleme ve tahmin etme işlemlerinde yol gösterici olmaktadır. Makine öğrenmesinde sık kullanılan temel algoritmalar k-NN, NB, YSA, KA, DVM ve LR şeklindedir. Sınıflandırma da kullanılan bu yöntemlerin bazıları tahmin sürecinde de kullanılabilir.

2.1.1 K-En Yakın Komşu Algoritması

İlk kez Fix ve Hodges tarafından 1951 yılında ortaya konulan k-NN, sınıflandırma yapabilmek için tüm eğitim verilerini kullanan örnek tabanlı bir öğrenme yöntemidir [56]. Başka bir deyişle, bir veri setinde sınıflandırma yapmak için değişkenler arasında en yakın komşuları bulur. Buradaki en önemli nokta veri noktaları arasındaki mesafedir. Genellikle Manhattan, Minkowski, Mahalanobis ve Öklid gibi uzaklık ölçüleri veriler arasındaki mesafenin hesaplanmasında kullanılır. Ayrıca, bu algoritmanın performansı; değişkenlere, yani uygun bir benzerlik fonksiyonunun değerine ve k parametresine bağlıdır [57, 58]. k-NN algoritmasında k, uzaklıkları belirlemek için kullanılan bir parametredir. Herhangi bir öğrenme bileşenini içermeyen çok basit, sezgisel bir yöntemdir. Çok geniş bir uygulama yelpazesinde göreceli olarak yüksek performansı nedeniyle karşılaştırmalar veya kıyaslama için referans sınıflandırıcılar arasında yapılan sınıflandırma çalışmalarında sıklıkla kullanılabilir [59, 60]. Literatürde k-NN algoritmasının k değeri için farklı formüller yer almaktadır. Genel formülü Denklem (2.1)'de verilmiştir:

$$Oy(x_i) = \begin{cases} \infty \quad , & eğer\ d(x_i, q) = 0\ \ ise \\ \dfrac{1}{d(x_i, q)} \quad , & aksi\ takdirde \end{cases} \tag{2.1}$$

Denklem (2.1)'de, x_i uzaydan alınan rasgele bir $x \in X$ örneğidir [7]. d aradaki uzaklığı göstermektedir. Birbirine yakın komşular daha yüksek oy almaktadır. q ise komşu ile sınıfı bulunmak istenen nokta arasındaki uzaklığın tersi olarak belirlenmesinde kullanılmaktadır [61].

2.1.2 Naive Bayes Algoritması

NB algoritması veri madenciliğinde araştırmacılar tarafından kullanılan en etkili öğrenme algoritmalarından biri olarak kabul edilmektedir [62]. Bir tür Bayes ağı olan bu algoritmanın en iyi şekilde çalışması için iki koşulun gerçekleşmesi gerekir. Birinci koşul, sınıfların belirli koşullar altında birbirlerinden bağımsız olmasıdır. İkincisi koşul ise sonuçları etkileyebilecek olan değişkenlerin belirli olmasıdır [63]. Bağımsızlık varsayımı yerine getirildiğinde, NB sınıflandırıcılarının öğrenme süreci daha basit hale gelir ve gözlemlenebilir faktörlerin vektörü kullanılarak en uygun şekilde atama belirlenir [60]. Ayrıca eldeki veri setinin birleşik olasılıklarının sınıflandırma başarılarının bulunmasında da NB sınıflandırma algoritmasından yararlanılmaktadır [64]. Genel formülü Denklem (2.2)'de verilmiştir:

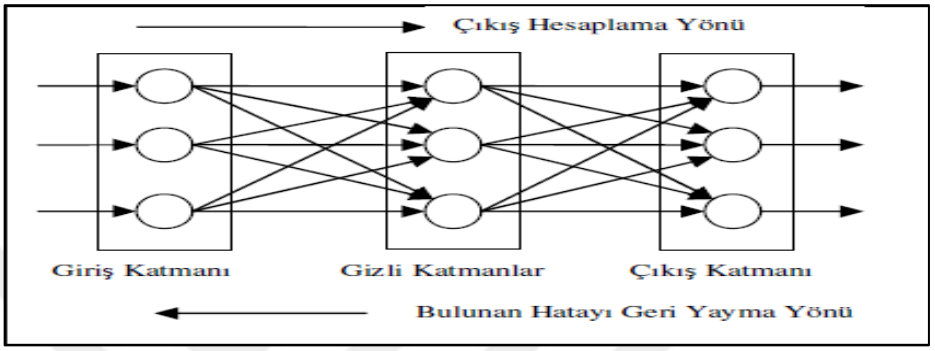
$$P(\vec{x}|c_j) = \prod_{i=1}^{n} P(x_{a_i}|c_j) \quad i = 1, 2, \dots, n\,; j = 1, 2, \dots, k) \tag{2.2}$$

Denklem (2.2)'de $X = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_m \end{bmatrix}$ m örneklemlerinin örnek uzay formudur. $\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \in \mathcal{R}^{m*n}$, n niteliklerin numarası ve verinin m sayıda gözlem matrisini gösterir. $c_1, c_2, \dots\ c_k$ örnek uzayda sınıf değerleridir [61].

2.1.3 Yapay Sinir Ağları

İnsan beyni süreçlerini taklit edecek şekilde çalışan YSA, geçmişteki deneyimlerin sonuçlarına dayanarak hareket eder. Bu nedenle, karmaşık doğrusal olmayan durumlarla çalışırken bile, YSA modellerinde standart istatistik yöntemlerinde olduğu gibi varsayımlara gerek duymaz. Çok katmanlı algılayıcı (ÇKA), eğitim sürecinde geri yayılım algoritması kullanan bir yapay sinir ağı modelidir. ÇKA bir giriş katmanı, gizli katmanlar ve bir çıkış katmanı olarak üç unsurdan oluşur. Eldeki bilgi bir nörondan diğerine ağırlık değeri ile taşınır. Bir ÇKA algoritmasının ilk adımı rasgele ağırlıklar atamaktır. İkinci adımda, girişler (bağımsız değişkenler) sigmoid veya lojistik işlevini kullanarak ileri doğru yayılır, böylece her gizli katman için çıktı değerleri (bağımlı değişkenler) üretilir. Bundan sonra, üçüncü adımda

ağırlıkları ve önyargıları güncelleyerek hata geriye doğru yayılır. Hatalar her çıktı ve gizli katman için hesaplanır. Son adımda, ağırlıklar ve önyargılar güncellenerek Adım 2'ye geri döndürülür. Genel hata en aza indirilene kadar adımlar tekrarlanır [6]. Yapay sinir ağları modelleri; ağın yapısına, ağdaki düğümlerin özelliklerine, kullanılmakta olan eşik değerine, ağın ileri ya da geri beslemeli olmasına, ağırlık matrisi değerlerinin sabit ya da değişken olmasına, ağırlık matrislerinin simetriklik veya asimetriklik durumuna ve eğitim ya da öğrenme kurallarına bağlıdır [65, 66].



Şekil 2.1 Çok Katmanlı Algılayıcı Örneği [67]

2.1.4 Karar Ağacı Algoritmaları

KA algoritmaları, hem sınıflandırma hem de tahminde kullanılabilecek en yaygın veri madenciliği yöntemlerinden biridir. KA algoritmaları, kolay yorumlanabildiği için sınıflandırma tekniklerinde sıklıkla tercih edilmektedir. Genelleme hatasını minimum seviyeye düşürerek veri setinden bir karar ağacı oluşturulması amaçlanmaktadır. KA algoritmaları C4.5, RF, Raptree gibi farklı algoritmalar kullanılarak çözülmektedir [68]. Karar ağacı akış çizelgesi bir ağaç yapısına benzemektedir. Üretilen sınıflandırma kuralları bu ağaçtan kolayca belirlenebilmektedir [69]. Bu kurallar sayesinde programcıların da daha kolay program yazmaları sağlanır [61].

2.1.4.1 C4.5 Karar Ağacı Algoritması

C4.5, J48 olarak bilinen açık kaynaklı bir Weka uygulamasıdır. Bu algoritma özellikle örüntü tanıma sorunları için kullanışlıdır. Bu algoritma, yalnızca kategorik değişkenlerle çalışan ve karar ağacı oluştururken her adımda bilgi kazancı hesaplayan ID3 algoritmasının eksikliklerini ve yetersizliklerini gidermek için 1993 yılında Quinlan tarafından geliştirilmiştir [70]. Eğitim verilerindeki eksik

değerleri gözlemlenebilir niteliklerden tahmin ederek belirleyebilir. Sürekli ve ayrık nitelikler içeren veri setlerinde de kullanılabilir. Bir ağaç oluştururken, ağacın boyutunu azaltmak için budama yöntemi kullanılır. Veriler, en iyi şekilde kategorize edilene kadar tekrarlı bir şekilde sınıflandırılır [71]. Bu algoritma, bilgi kazancı ve entropi kullanarak karar ağacı üretir. Bilgi kazancı verilen özelliğin eğitim setinden ne kadar iyi ayrıldığını gösteren bir değerdir. Her karar düğümünde, en yüksek bilgi kazancına sahip en yararlı özellik seçilir. Eğitim veri seti S ve bilgi kazancı $\text{Gain}_{(\text{S.A})}$ Denklem (2.3) ile verildiği biçimde tanımlanır:

$$Gain_{(S,A)} = Entropy_{(S)} - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} Entropy_{(S_v)} \tag{2.3}$$

A özelliğinin tüm olası değerleri $Value(A)$ ile gösterilir. S_V örnek uzayın bir alt kümesidir. Homojenlik, Denklem (2.4)'de verilen entropi ile ölçülür:

$$Entropy_{(S)} = -\sum_{i=1}^{c} P_i log_2 P_i \tag{2.4}$$

Burada P_i, S'nin i. sınıfına aittir. c ise sınıf numarasını gösterir [40, 72].

2.1.4.2 Random Forest Karar Ağacı Algoritması

KA-RF algoritması, karar ağaçlarını geliştirmek için eğitim verilerinin değişimlerinden yararlanır. Eğitim verilerinin değişimlerini belirlemek için orijinal eğitim verilerinin parçaları rastgele bir şekilde değiştirilir. Her ağaç budama yapılamadan önce olabildiğince ilerletilir. Daha sonra her ağacın kendi sınıflandırmasını yapmasına izin verilir ve böylece karar süreci uygulanır [73]. Bu yöntemin, karar ağacındaki diğer algoritmalardan performans bakımından çok daha üstün olduğunu belirlenmiştir [74].

2.1.4.3 RepTree Karar Ağacı Algoritması

RepTree algoritması, regresyon ağacı mantığını kullanarak çeşitli ağaçlar oluşturur ve yapılan karar ağaçları arasından en iyisini seçer. KA işlemlerini yapmak için, ortalama kare hata kriterinden yararlanır [75]. RepTree karar ağacı algoritması, hızlı bir karar ağacı öğrenmesidir. Bu algoritma karar ağacından bilgi edinir ve

varyansı düşürür. Başlangıçta, sayısal özelliklerin değerlerini kullanır ve sonra eksik gözlemler için C4.5'in kesirli örneklerini kullanır [76].

2.1.5 Destek Vektör Makineleri

DVM, ilk olarak 1995 yılında Cortes ve Vapnik tarafından geliştirilmiştir [77]. Güçlü matematiksel alt yapısı, yaygın olarak erişilebilen yazılım uygulamaları ve kullanım kolaylığı DVM'nin başarısının kilit faktörleridir [78].

DVM, verileri iki kategoriye ayırmak için n-boyutlu bir hiper düzlem oluşturur [79]. Parametrik olmayan kontrollü sınıflandırıcılardan biri olan DVM, sınıflandırmada kullanılan denetimli bir öğrenme yöntemidir [25]. Veriler doğrusal bir şekilde ayrılmışsa, doğrusal DVM kullanılır. Doğrusal olarak ayrılamazsa, doğrusal olmayan DVM uygulanır [80]. Çalışma prensibi aşağıda belirtilen kurallara göre devam eder:

1. Birimler doğrusal veya doğrusal olmayan DVM'leri olan iki (veya daha fazla) gruba ayrılır.

2. Doğrusal olmayan DVM'ler, bir çekirdek işlevi kullanarak verileri doğrusal DVM'lere dönüştürür.

3. Doğrusal olmayan DVM'ler girdi verilerini daha yüksek boyutlu bir alana eşler, bunun yanında daha yüksek sınıflandırma yapabilirler.

4. DVM operatörü, sınıflandırma problemini çözmek için kullanılan çekirdek tiplerini içerir.

Farklı çekirdek seçimi farklı sonuçlar verebilir [81]. Doğru çekirdek seçimi öğrenme kapasitesini önemli ölçüde etkileyebilir [82].

2.1.5.1 Doğrusal Destek Vektör Makineleri

Bir doğrusal DVM, verileri maksimum mesafe ile iki sınıfa ayıran bir optimum hiper düzlemin belirlemeyi amaçlamaktadır. Doğrusal DVM verilerin doğrusal olarak ayrılabilmesi veya ayrılamaması durumlarına göre iki şekilde incelenebilir.

Doğrusal olarak ayrılabilir durum, en temel DVM algoritmasıdır. Amaç bilinmeyen bir veri seti ile karşılaşıldığında sınıflama hatasını minimum yapacak doğruyu

belirlemektir. X, n örneklemlerin oluşturduğu eğitim setini göstermek üzere bu küme Denklem (2.5)'deki gibi ifade edilebilir:

$$X = \left\{(x_1, y_1), \dots, (x_n, y_n), x_i \in R^d, y_i \in \{-1, +1\}\right\} \tag{2.5}$$

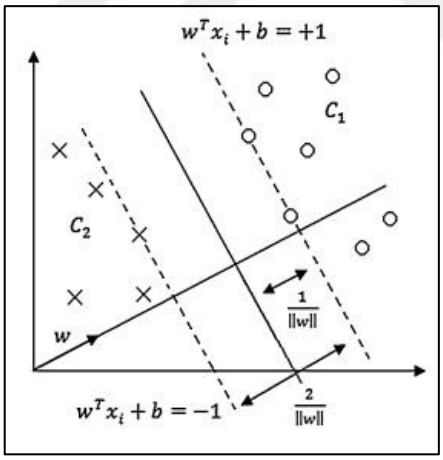
Buradaki x_i ve y_i sırasıyla giriş vektörlerini ve bu giriş vektörlerine karşılık gelen etiketleri göstermektedir [83, 84]. Karar fonksiyonu $signum(f(x))$ ile gösterilir ve burada $f\,(x)$, hiper düzlemle ilişkilendirilmiş ve Denklem (2.6)'deki biçimde tanımlanmıştır:

$$f(x) = w^T x_i + b, \;\; w \in R^d \;\; ve \;\; b \in R \tag{2.6}$$

Buradaki amaç, Şekil 2.2'de gösterilen optimum hiper düzlemi birlikte tanımlayan w vektörünü ve skaler b'yi belirlemektir. Hiper düzlemin aşağıda verilen Denklem (2.7) ve Denklem (2.8) kısıtlamalarını sağlaması gerekmektedir:

$$w^T x_i + b \geq +1, y_i = +1 \; için, \tag{2.7}$$

$$w^T x_i + b \leq -1, y_i = -1 \; için. \tag{2.8}$$



Şekil 2.2 DVM'de doğrusal olarak ayrılabilir durum için optimal ayırıcı hiper düzlem [80]

Denklem (2.7) ve Denklem (2.8) kısıtları tek bir formda aşağıda verilen Denklem (2.9) ile yazılabilir:

$$y_i(w^T x_i + b) \geq +1, \quad i = 1, \dots, n. \tag{2.9}$$

Sınır bandı, hiper düzlemler $w^T x_i + b \geq +1$ ve $w^T x_i + b \leq -1$ arasındaki bölge olarak tanımlanmaktadır. Sınır bandının genişliği $\frac{2}{\|w\|}$ olarak hesaplanabilir. Doğrusal ayrılabilir durum için, maksimum sınır bandı, doğrusal kısıtlı aşağıda verilen Denklem (2.10) ve Denklem (2.11) ile karesel optimizasyon probleminde olduğu gibi mesafenin tersinin en aza indirilmesiyle belirlenebilir:

$$\text{Amaç fonksiyonu: } min \frac{1}{2\|w\|^2}, \tag{2.10}$$

$$\text{Kısıtlar: } y_i(w^T x_i + b) \geq +1 \text{ , i} = 1, \dots, n \tag{2.11}$$

Sonuç olarak optimum bir hiper düzlemin var olduğu ortaya konulmuştur. Böylece negatif örnekler ve pozitif örnekler hiper düzlemin iki tarafına ayrılmış olur [84, 85]. Denklem (2.9)'deki optimizasyon problemi, Lagrange çarpanları ile sınırlandırılmamış bir optimizasyon problemi olarak Denkem (2.12)'deki gibi formüle edilebilir:

$$L_p = \frac{1}{2\|w\|^2} - \sum_{i=1}^{n} \alpha_i [y_i(w^T x_i + b) - 1] \tag{2.12}$$

Denklem (2.12)'de, α_i lagrange çarpanlarıdır. Lagrange fonksiyonu L_p, w ve b ile karşılaştırıldığında en aza indirilmeli ve $\alpha_i \geq 0$'a göre maksimize edilmelidir. Karush-Kuhn-Tucker (KKT) koşullarını kullanarak, Denklem (2.12)'deki w ve b tasarım değişkenleri α_i cinsinden ifade edilebilir, bu da problemi Lagrange çarpanları α_i'ye göre sadece maksimizasyon gerektiren dual bir problem haline dönüştürür:

$$\frac{\partial L_p}{\partial w} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i y_i x_i = w \tag{2.13}$$

$$\frac{\partial L_p}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{2.14}$$

Bunun sonucunda ikinci dereceden bir programlama probleminin dual formu Denklem (2.15)'deki gibi elde edilebilir:

$$L_d = \sum_{i=1}^{n} \alpha_i - 1/2 \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (2.15)$$

Ardından dual problem α_i açısından çözülür ve bu duruma kısıtlamalarda dahildir.

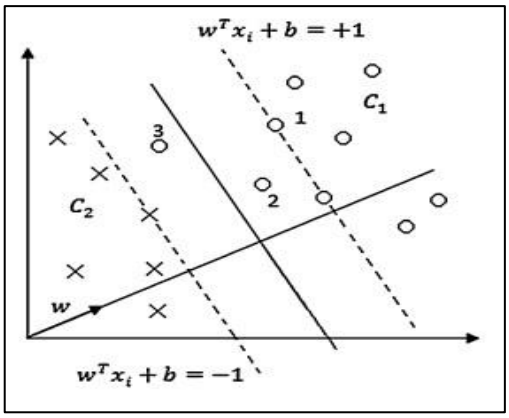
$$\sum_{i=1}^{n} \alpha_i y_i = 0 \quad ve \quad \alpha_i \geq 0 \quad (2.16)$$

Pozitif α_i değerlerine karşılık gelen x_i örneklerine destek vektörleri denir. w vektörü Denklem (2.13) kullanılarak belirlenir. Son bilinmeyen b parametresi tüm destek vektör makineleri için ortalama alınarak belirlenir ve Denklem (2.17)'deki gibi ifade edilebilir [80, 84].

$$b = y_i - w^T x_i \quad (2.17)$$

Verilerin doğrusal olarak ayrılabildiği durumda, hiper düzlemin sınıf etiketlerini ve veri örneklerini en doğru şekilde birbirinden ayırdığı kabul edilmektedir. Ancak bazı durumlarda veri setindeki örneklerin özel dağılımından dolayı böyle bir ayırıcı hiper düzlemin belirlenmesinde zorluklarla karşılaşılabilir. Karşılaşılan zorluk neticesinde doğrusal karar sınırları ile doğrusal ayrılabilirliği sağlamak mümkün olmaz. Veriler doğrusal olarak ayrılamazsa Denklem (2.9) kısıtlamaları geçerli olmaz. Bu durum, Şekil 2.3'de gösterildiği gibi, sınırın içine veya karar sınırının yanlış tarafına düşen veri örneğine karşılık gelir. Doğrusal olarak ayrılamayan verileri işlemek için, formüle her bir örnek için bir soft hata **ξ** ekleyerek Denklem (2.19)'deki gibi genelleştirilir:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \qquad i = 1, .., n\,, \;\; \xi_i > 0 \quad (2.18)$$



Şekil 2.3 DVM'de doğrusal olarak ayrılamayan durum için optimal ayırıcı hiper düzlem [80]

Yeni optimizasyon problemi iki amacın birleşimi olarak tanımlanmaktadır: marjı maksimize etmek ve hatayı minimize etmek. Eğer $\xi_i = 0$ ise, x_i örnekleri doğru şekilde sınıflandırılır (Şekil 2.3'deki 1 noktası). $0 > \xi_i > 1$ ise x_i örnekleri de doğru şekilde sınıflandırılır, ancak bunlar sınırın içindedir (Şekil 2.3'deki 2 noktası). $\xi_i \geq 0$ ise x_i örnekleri yanlış sınıflandırılmıştır (Şekil 2.3'deki 3 noktası). Yanlış sınıflandırmaların sayısı 1'den büyük olan ξ sayısına eşittir. Öte yandan, ayrılmayan noktaların sayısı pozitif ξ_i sayısına eşittir. Yani $\sum_{i=1}^{n} \xi_i$, sınıfı belirlenemeyen noktaların toplam sayısıdır. Bu hata terimini de kullanarak L_p, Denklem (2.19)'deki biçimde yazılır.

$$L_p = \frac{1}{2\|w\|^2} + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^{n} \mu_i \xi_i \tag{2.19}$$

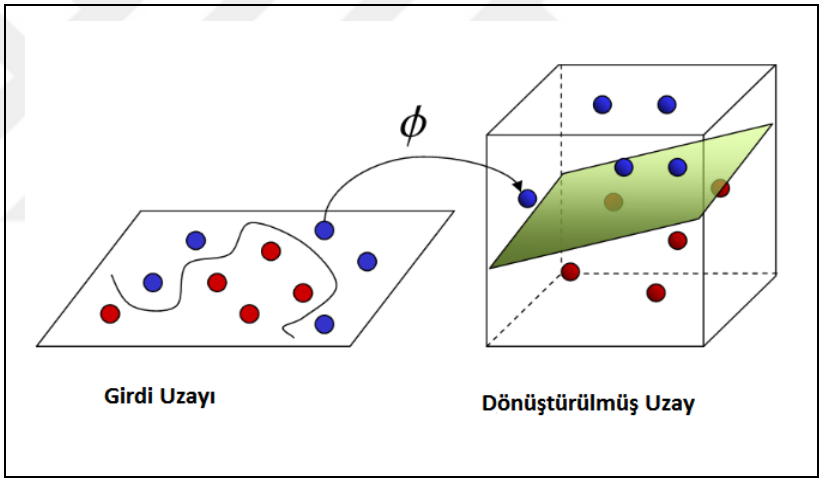
ξ_i değerlerinin, yeni Lagrange parametreleri μ_i ile pozitif olması gerekmektedir. C parametresi, destek vektörlerinin sayısı ile doğrusal olarak ayrılmayan noktaların sayısı arasındaki ağırlığı ayarlamak için kullanılır. Dual problem Denklem (2.20) ve Denklem (2.21) ile gösterilir [80, 84]:

$$L_d = \sum_{i=1}^{n} \alpha_i - 1/2 \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{2.20}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \quad ve\ 0 \leq \alpha_i \leq C \tag{2.21}$$

2.1.5.2 Doğrusal Olmayan Destek Vektör Makineleri

Veri doğrusal olmayan bir şekilde ayrılabilir olduğunda, doğrusal DVM kullanmak yerine doğrusal olmayan DVM kullanmak daha iyi bir yaklaşımdır. Doğrusal olmayan DVM girdileri dönüştürmek için ilk olarak doğrusal olmayan bir çekirdek belirler ve sonrasında doğrusal DVM kullanır. Doğrusal olmayan çekirdek fonksiyonu büyük olasılıkla büyük bir matristir. Dolayısıyla doğrusal olmayan DVM'nin yüksek boyutlu bir uzayda girdiyi eşlemesi çok uzun zaman alabilir. Burada amaç, verinin dönüştürüldüğü yeni boyutta en büyük sınırlı ayırıcı hiper düzlemi ortaya çıkarmaktır. Doğrusal olmayan ayrılabilir veriyi doğrusal ayrılabilir veriye dönüştürmek için veri, doğrusal olmayan bir çekirdek fonksiyonu kullanarak yüksek boyutlu bir özellik uzayı formunda haritalanır. Örnek olarak Şekil 2.4'te gösterildiği gibi bir çekirdek fonksiyonunu yüksek boyuta haritalanması gösterilmiştir.



Şekil 2.4 Bir çekirdek fonksiyonu ile yüksek boyutlu bir özellik uzayına haritalama [86]

Çekirdek fonksiyonlarının özellik uzayında karşılık gelen iç çarpımı olması gerekir ki böylece dönüşmüş olurlar. Çoğunlukla polinom (DVM-POLY), radyal tabanlı (DVM-RBF) ve Pearson VII fonksiyon tabanlı evrensel (DVM-PUK) çekirdek fonksiyonları kullanılır, ancak bunun dışında başka çekirdek fonksiyonları da kullanılmaktadır.

Çekirdek fonksiyonu, etkin bir benzerlik ölçüsüdür ve lineer bir modeli doğrusal olmayan bir modele dönüştürebilir. Bu çalışmada kullanılacak olan çekirdek fonksiyonlarının formülleri Tablo 2.1'de verilmiştir:

Tablo 2.1 Çekirdek fonksiyonları ve formülleri

Çekirdek	$K(x_i^T x_j)$
Radial basis function	$\exp(-\gamma\|x_i - x_j\|^2), \gamma > 0$
Polynomial	$(\gamma x_i^T X_j + r)^d, \gamma > 0$
Puk	$\frac{1}{\left[1 + \left(\frac{2 * \sqrt{\|x-y\|^2 \sqrt{2^{\left(\frac{1}{\omega}\right)} - 1}}}{\sigma}\right)^2\right]^{\omega}}$

2.1.6 Lojistik Regresyon Algoritması

LR, standart regresyon modellerinde olduğu gibi bağımlı ve bağımsız değişkenler arasındaki ilişkiyi belirlemek için kullanılır. Bağımlı değişken 0 veya 1 değerine sahipse, ikili LR kullanılır [87]. LR analizinde amaç, bağımlı ve bağımsız değişkenler arasındaki ilişki veya ilişkileri, en az değişken ile en iyi uyuma sahip olacak şekilde tanımlamaktır [88, 89]. Bu yöntemin doğrusal regresyona göre tercih edilmesinin temel nedeni doğrusal regresyonda geçerli olan varsayımların aranmaması ve doğrusal regresyona göre daha esnek olmasıdır [90]. Bir diğer ayrımda LR'de varsayımlar açısından incelendiğinde en küçük kareler yöntemine nazaran daha az kısıt mevcuttur [91]. Lojistik regresyonda parametreler, analitik olarak elde edilemediğinden dolayı iteratif bir yöntem olan maksimum olabilirlik tekniğinden yararlanılarak tahmin yapılmaktadır [92].

Lojistik fonksiyonun formülü Denklem (2.22)'deki biçimde yazılabilir:

$$g(z) = \frac{1}{1 + \mathrm{e}^{-\mathrm{z}}}, \mathcal{R} \rightarrow [0,1] \tag{2.22}$$

Denklem (2.22)'de, $g(z)$ değeri, 0 ile 1 arasında değer almaktadır. Eğer aldığı değer 0,5 değerinden büyük ise 1'e yaklaştığı, küçük ise 0'a yaklaştığı söylenir [61].

2.2 Öznitelik Seçim Algoritmaları

Öznitelik seçimi, M<N olmak üzere, N tane öznitelik sahibi olan bir veri seti için ölçüt fonksiyonu, M'nin bütün alt kümelerinde en iyi olacak şekilde M özniteliğini belirleme durumudur [93].

Makine öğrenmesi algoritmaları sınıflandırma yaparken en önemli noktalardan biri hangi değişkenlerin etkili olduğunu öğrenmektir. Algoritmaların sınıflandırma yapması, veri setindeki özniteliklerin seçimine büyük özen gösterilmesini ve sadece etkili olanların dahil edilmesini gerektirir. Sınıflandırma algoritmalarında kullanılan değişken sayısının olabildiğince azaltılması ve sınıflandırma sonuçlarında bunu yaparken önemli bir düşüş olmaması son derece önemlidir.

Öznitelik seçimi, veri setinde sınıflandırma başarısını etkileyen en etkili öznitelikleri seçme işlemidir. Veri setindeki bazı özniteliklerde önemli bilgi bulunmaz. Öznitelik uzayının yüksek boyutlu olmasının olumsuz taraflarından kurtulmak için boyut indirgeme yapılmalıdır. Boyut indirgeme işlemi ya öznitelik seçimi ya da öznitelik çıkartımı yapılır. Öznitelik seçimi, veri setindeki özniteliklerden yeni bir öznitelik alt kümesi oluşturarak düzenlenir. Asıl özniteliklerin yerine dönüşüm yapılarak elde edilen özniteliklerin kullanılması öznitelik çıkarımının öznitelik seçiminden farkıdır. Öznitelik sayısının durumu sınıflandırıcıların başarı performansını olumlu ya da olumsuz yönde etkileyebilir [94].

Öznitelik seçimi yapılmasındaki önemli noktalardan biri de, örnek sayısının düşük, nitelik sayısının ise fazla olduğu durumlarda makine öğrenmesi algoritmalarının işlem yapmasının zor olmasıdır. Öznitelik seçimi, farklı algoritmalarla yapılarak olumlu ve olumsuz sonuçlar oluşturabilir. Bunlar filtre modelli, sarmal modelli ve gömülü modelli algoritmalardır. Filtre modelli algoritmanın olumlu yönü sınıflandırma algoritmasından bağımsız ve hızlı olarak işlem yapmasıdır. Olumsuz yönü ise diğer yöntemlere göre başarı oranının daha düşük seviyede olmasıdır. Sarmal modelli algoritmanın olumlu yönü başarı oranını yüksek tutması ve uygulamasının kolay olmasıdır. İşlem yükünün fazlalığından dolayı yavaş çalışması ise olumsuz yönüdür. Gömülü modelli algoritmanın olumlu yönü, başarı oranını yüksek tutarak işlem yükünü düşürmesidir. Olumsuz yönü ise karar ağaçlarından bağımsız olarak kullanılamamalarıdır [95].

Farklı öznitelik seçim algoritmaları literatürde mevcuttur. Bunlar, Cfs subset, correlation attribute, one-R, gain ratio, info gain ve relifF öznitelik seçim algoritmalarıdır.

2.2.1 Cfs Subset Öznitelik Seçim Algoritması

CFS Subset öznitelik seçim algoritması, veri setindeki değişkenleri korelasyon yardımı ile değerlendirerek en iyi öznitelikleri bulmayı amaçlar. Bu algoritma aralarında düşük korelasyonlu, sınıf etiketleri arasında ise yüksek korelasyonlu öznitelikleri seçer [95, 96]. Bu algoritmada, yüksek korelasyonlu özellikler veri setinden çıkarılır çünkü ilişkisiz özellikler daha iyi sınıflandırma başarısı ortaya çıkarır [97]. Korelasyon tabanlı öznitelik seçim katsayısı Denklem (2.23)'deki gibi hesaplanır:

$$Skor_s = \frac{k\bar{\gamma}_{cf}}{\sqrt{k + k(k-1)\bar{\gamma}_{ff}}} \tag{2.23}$$

Denklem (2.23)'de, verilen s, k tane özelliğe sahip öznitelik alt kümesini göstermektedir. $\bar{\gamma}_{cf}$, öznitelik alt kümesinin sınıfı ile korelasyonunu belirtirken; $\bar{\gamma}_{ff}$, öznitelik alt kümesindeki özniteliklerin korelasyonunu ifade eder [95].

2.2.2 Correlation Attribute Öznitelik Seçim Algoritması

İkili olan özellikler ve sonuç arasındaki Pearson korelasyon değerini ölçer. Bu yöntemde değişkenler nominal ölçek ile ölçülürse, genel korelasyon için ağırlıklı ortalama hesaplanarak sonuçlar belirlenir [45]. Her bir öznitelik ile hedef sınıf özniteliği arasındaki ilişkiyi ölçmek için Pearson korelasyon yöntemini kullanmak uygundur. Nominal öznitelikleri değer bazında değerlendirmektedir. Her değer bir gösterge olarak hareket eder [49].

2.2.3 One-R Öznitelik Seçim Algoritması

Veri setinde test edilen öznitelikler üzerindeki sınıflandırma kuralları, tek bir öznitelik değerine dayanan One-R öznitelik seçim algoritması ile elde edilir. One-R algoritmasında, önemli nokta en düşük hata oranına sahip özniteliği seçmektir. Sonuç olarak, öznitelik değerinin çoğunluğuna ait olmayan örneklerin oranı hata oranına katkıda bulunacaktır [98]. One-R algoritması, tüm veriyi test ederek belirli kurallar doğrultusunda karar ağacını oluşturur. Bu teknik, veri yapısının uygun bir şekilde çalışılmasını sağlayan, son derece hassas ve yüksek doğruluk oranına sahip bir algoritmadır [99]. One-R öznitelik seçim algoritması eğitim setinde özellikleri

hata oranına göre sıralamaktadır. Sayısal olarak değer taşıyan özniteliklerin tamamını sürekli olarak ele alır ve değer aralığını birkaç farklı aralığa bölmek için basit bir yöntem kullanır. En basit öznitelik seçim uygulamalarından biridir. Yalnızca bir özelliğe dayalı basit kurallar üretir [46].

2.2.4 Gain Ratio Öznitelik Seçim Algoritması

Gain ratio, entropiye dayalı bir öznitelik seçim yöntemidir. Info gain yöntemi farklı değerlere sahip öznitelikler seçme eğilimindedir, bu nedenle bazen sonuçlar yanlı olarak elde edilir. Bu yanlılık durumunu azaltmak için gain ratio yöntemi kullanılır. Yanlılığı azaltmak için bölme bilgilerinden yararlanılır. Gain ratio, bilgi kazancı değerlerinin ve bölünmüş bilginin oranıyla elde edilen normalleşmedir [100]. Gain ratio, yanlılığını azaltan bilgi kazancının farklı bir versiyonudur. Gain ratio, öznitelik seçiminde değişken(branches) sayısı ve boyutunu dikkate alır. Bir bölünmenin kendi içindeki bilgisini hesaba katarak bilgi kazancını düzeltir. İçsel bilgi, örneklerin dallara dağılma entropisidir. Özgün bilgiler büyüdükçe özniteliğin değeri de düşer [47, 69]. Gain ratio için formül Denklem (2.24) ile verilmiştir:

$$Gain\ ratio(\text{Öznitelik}) = \frac{Gain(\text{Öznitelik})}{Intrinsic_info(\text{Öznitelik})} \tag{2.24}$$

2.2.5 Info Gain Öznitelik Seçim Algoritması

Info gain öznitelik seçim algoritmasının temeli entropi kavramı ile ilgilidir. Entropi, bir sistemdeki düzensizliğin ölçüsü olarak ifade edilebilir. Yüksek entropiye sahip verilerde daha fazla bilgi bulunmaktadır. Sınıflandırma işlemi, öznitelikleri kullanarak ne kadar bilgi kazanılabileceğini ortaya koymaktadır. Öznitelik seçimi sınıflardan ne kadar bağımsız olursa bilgi o kadar düşük olur [101, 102]. Bu algoritma ile öznitelik seçimi yapmak için her bir özniteliğin bilgi kazancı değerinin hesaplanması gerekmektedir. Eğer yapılan işlemler sonucunda elde edilen kazanç verileri önceden belirlenmiş olan eşik değerden düşük çıkarsa, bu özniteliklerin sınıflandırma sürecinde kullanılması uygun değildir [94].

Info gain formülü Denklem (2.25) ile verilebilir:

$$IG(t) = -\sum_{i=1}^{m} P(C_i) log P(C_i) + P(t) \sum_{i=1}^{m} P(C_i \mid t) log P(C_i \mid t) + P(\bar{t}) \sum_{i=1}^{m} P(C_i \mid \bar{t}) log P(C_i \mid \bar{t}) \quad (2.25)$$

Verilen Denklem (2.25)'de, m toplam sınıf sayısını, P(t) ve P($\bar{t}$) t teriminin kümeye ait olup olmamasının olasılığını, P(C_i) C_i sınıfının olasılığını, P($C_i \mid t$) ve P($C_i \mid \bar{t}$) t'nin kümede olduğu ya da olmadığı durumda C_i'nin koşullu olasılığı göstermektedir [103].

2.2.6 ReliefF Öznitelik Seçim Algoritması

ReliefF öznitelik seçim algoritması, özniteliklerin değerlerini ve aralarında bağımlılık olup olmadığını belirlemeye çalışır. Bu sonucu elde etmek için, örneklerin ait olduğu ve ağırlıklandırılmadığı sınıflardaki en yakın örnekleri karşılaştırır. Başlangıçta ikili sınıf problemleri için geliştirilmiş ve ardından çoklu sınıf problemlerine de uyarlanmıştır [104].

ReliefF öznitelik seçim algoritması, Relief istatistiksel modelinin geliştirilmiş halidir. Bu yöntem, ilk olarak veri setinden bir örnek seçer. İkinci olarak ilgili örneğin, kendi sınıflarındaki diğer örneklerle yakınlığı ile farklı sınıflarla olan uzaklığına göre bir model oluşturur ve bu işlemlerin ardından öznitelik seçme işlemini gerçekleştirir [48]. ReliefF formülü Denklem (2.26) ile verilebilir:

$$S_i = \frac{\sum_{j=1}^{m} -fark(x_{ij,} en\ yakın\ aynı_{ij}) + fark(x_{ij,} en\ yakın\ farklı_{ij})}{m} \quad (2.26)$$

Burada m, verideki tüm örnek sayısını gösterir. *j.* örnekteki *i.* değerin farklı sınıfa sahip en yakın örneğe olan uzaklığını $fark(x_{ij,} en\ yakın\ farklı_{ij})$ değeri; *j.* örnekteki *i.* niteliğin aynı sınıfa sahip en yakın örneğe olan uzaklığını $fark(x_{ij,} en\ yakın\ aynı_{ij})$ değeri ifade eder [95].

2.3 Sınıflandırma Algoritmaları İçin Performans Ölçüm Kriterleri

Sınıflandırma algoritmalarının başarı oranının belirlenmesi sürecinde farklı performans kriterlerinin sonuçları hesaplanır ve karşılaştırılır. Hangi algoritmanın en etkili olduğunu belirlemek için birçok kriter kullanılmaktadır. Bu kriterler; doğru pozitif (DP) oranı, yanlış pozitif (YP) oranı, doğru negatif (DN) oranı, yanlış

negatif (YN) oranı, hassasiyet (precision), f-ölçütü (f-measure), alıcı işlem karakteristiği eğrisi (ROC: Receiver Operating Characteristic), kappa (κ) istatistiği, ortalama mutlak hata (MAE: Mean Absolute Error), kök ortalama kare hata (RMSE: Root Mean Square Error), Matthews korelasyon katsayısı (MCC: Matthews Correlation Coefficient) olarak söylenebilir. Burada bazı değerlerin hesaplanabilmesi için karşılaştırma tablosunu bilmek gerekir. Karşılaştırma tablosunun bileşenleri Tablo 2.2'de verilmiştir;

Tablo 2.2 Karşılaştırma tablosu

		Tahmin	
		a	**b**
Gerçek Değer	**a**	DP (Doğru Pozitif)	YN (Yanlış Negatif)
	b	YP (Yanlış Pozitif)	DN (Doğru Negatif)

Tablo 2.2'de doğru pozitif (DP): doğru pozitif tahmin sayısını, yanlış pozitif (YP): yanlış pozitif tahmin sayısını, doğru negatif (DN): doğru negatif tahmin sayısını ve yanlış negatif (YN): yanlış negatif tahmin sayısını göstermektedir.

2.3.1 Doğru Pozitif Oranı (Duyarlılık)

Doğru sınıflandırılmış pozitif örneklerin, modeldeki toplam pozitif örnek sayısına oranıyla elde edilir ve Denklem (2.27)'de gösterildiği biçimde ifade edilir.

$$DP\ Oranı = \frac{DP}{DP + YN} \tag{2.27}$$

2.3.2 Yanlış Pozitif Oranı

Gerçekte negatif olan fakat pozitif sınıflanmış olan örneklerin toplam negatif örnek sayısına oranı ile bulunur ve Denklem (2.28)'de gösterildiği biçimde ifade edilir.

$$YP\ Oranı = \frac{YP}{YP + DN} \tag{2.28}$$

2.3.3 Precision (Hassasiyet)

Hassasiyet değeri, doğru sınıflandırılmış pozitif örneklerin toplam pozitif örneklerin sayısına oranı ile elde edilir ve Denklem (2.29)'de gösterildiği biçimde ifade edilir..

$$Hassasiyet = \frac{DP}{DP + YP} \tag{2.29}$$

2.3.4 F-Ölçütü

F-ölçütü, hassasiyet ve DP oranının harmonik ortalaması ile belirlenir [61]. F-ölçütünün hesaplanması için Denklem (2.30) kullanılabilir.

$$F - \text{ölçütü} = \frac{2 * Hassasiyet * DP\ Oranı}{Hassasiyet + DP\ Oranı} \tag{2.30}$$

2.3.5 Alıcı İşlem Karakteristiği Eğrisi (ROC: Receiver Operating Characteristic)

ROC eğrisi genellikle sınıflandırma algoritmalarının performansını ölçmek için kullanılır, burada eğri altındaki alan sınıflandırıcının nasıl çalıştığını gösterir [105]. Bu eğri, Y ekseninde DP değerine ve X ekseninde (1-DN) değerine sahiptir. ROC eğrisinin altında kalan alanın değeri ne kadar yüksek olursa, algoritmanın yaptığı sınıflandırmanın da o kadar başarılı olduğu söylenebilir. Değerler sayısal olarak hesaplanabilir. ROC eğrisinin altındaki alan, algoritmaların performansı ile ilgili görsel sonuçların yanı sıra sayısal sonuçlar da sunmaktadır. Böylece, farklı algoritmaların performanslarının karşılaştırılması görsel olarak kolayca yorumlanabilir [106].

2.3.6 Kappa (κ) İstatistiği

κ istatistiği, bir sınıflandırma algoritmasının performans başarısı ile ilgilenir. Kategorik değişkenler için yapılan analizlerin anlaşılmasında kullanılan uygun bir

istatistiki veridir. Aynı zamanda ki-kare tablosuna dayanan bir değerdir [107]. κ istatistiği 1'e yaklaştıkça, sınıflandırıcılar arasındaki anlaşma daha yüksektir. p_0 ve p_e'nin iki kategorik değişken arasındaki gözlemlenen ve beklenen değerleri göstermek üzere κ istatistiği Denklem (2.31) ile hesaplanır.

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{2.31}$$

p_0 ve p_e ise sırası ile Denklem (2.32) ve Denklem (2.33) ile hesaplanır.

$$p_0 = \frac{DP + DN}{DP + YN + YP + DN} \tag{2.32}$$

$$p_e = \frac{(DP + YP) * (DP + YN) + (YN + DN) * (YP + DN)}{(DP + YN + YP + DN)^2} \tag{2.33}$$

2.3.7 Ortalama Mutlak Hata (MAE: Mean Absolute Error)

Ortalama mutlak hata istatistiği, bir modelin tahmin edilen ve gözlenen değerleri arasındaki farkları göstermeye yardımcı olur [108]. Bu değer, tahmin edilen ve gözlemlenen değerler arasındaki mutlak farkların ortalamasını hesaplar ve Denklem (2.34)'de gösterildiği biçimde ifade edilir.

$$MAE = n^{-1} \sum_{i=1}^{n} \left| P_i - O_i \right| \tag{2.34}$$

$P_i - O_i$ modelin tahmin hatasını gösterir.

2.3.8 Kök Ortalama Kare Hata (RMSE: Root Mean Square Error)

MAE'ye benzer şekilde bir modelin tahmin edilen ve gözlenen değerleri arasındaki farkları göstermeye yardımcı olur. Kök ortalama kare hatası, tahmin edilen ve gözlemlenen değerler arasındaki kare farklarının ortalama karekköküne eşittir ve Denklem (2.35)'de gösterildiği biçimde ifade edilir.

$$RMSE = \sqrt{n^{-1} \sum_{i=1}^{n} \left| P_i - O_i \right|^2} \tag{2.35}$$

$P_i - O_i$ modelin tahmin hatasını gösterir.

2.3.9 Matthews Korelasyon Katsayısı (MCC: Matthews Correlation Coefficient)

Matthews korelasyon katsayısı (MCC) -1 ile 1 arasında değerler alır ve karşılaştırma matrisindeki bileşenler kullanılarak elde edilir. Pozitif değer elde eden MCC'lerin doğru tahminler ürettiği sonucuna varılabilir. MCC = 1 değeri mükemmel tahminler yağıldığı anlamına gelir. MCC, Denklem (2.36)'de verildiği şekilde hesaplanır [40].

$$MCC = \frac{(DP * DN) - (YP * YN)}{\sqrt{(DP + YP) * (DP + YN) * (DN + YP) * (DN + YN)}} \quad (2.36)$$

2.4 k-katlı Çapraz Doğrulama

Makine öğrenmesi yöntemlerini uygularken kritik noktalardan biri eğitim ve test verilerinin (kümelerinin) belirlenmesidir. Veri seti temel olarak eğitim ve test veri seti olarak iki gruba ayrılır. Bölünme, % 50 - % 50 ya da % 60 - % 40 gibi değerlerle belirlenebilir. k-katlı çapraz doğrulama yöntemi veri setini k eşit parçaya böler. Bunların k-1 tanesi eğitim veri seti, kalan kısmı ise test verisidir. Her bir parça, test kümesi olarak alınarak işlem k kez tekrar edilir. Tüm sonuçların ortalaması hesaplandığında, sınıflandırma değerleri belirlenmiş olur [109].

Literatürde k değerinin 2, 5, 10 olarak kullanıldığı görülmektedir [110]. Breiman, Friedman, Stone ve Olshen, k = 10 seçmenin en iyi seçim olduğunu söylemişlerdir [111]. Bu çalışmada da, k değeri için sıklıkla kullanılan 10 değeri alınmıştır. Fen bilimleri veri seti için k=10 alınarak 4481 öğrencinin verileri ile 10-katlı çapraz doğrulama yapılmıştır. 10 parçanın 9 tanesi eğitim seti, 1 tanesi test grubu olarak seçilmiştir. 4033 öğrencinin verisi eğitim setinde, 448 öğrencinin verisi de test grubunda yer almaktadır. Bu işlem her defasında test grubunu değiştirmek suretiyle 10 kez tekrar edilir. Matematik veri seti için yine k=10 alınarak 4577 öğrencinin verileri ile 10-katlı çapraz doğrulama yapılmıştır. 10 parçanın 9 tanesi eğitim seti, 1 tanesi test grubu olarak seçilmiştir. 4119 öğrencinin verisi eğitim

setinde, 458 öğrencinin verisi de test grubunda yer almaktadır. Bu işlem her defasında test grubunu değiştirmek suretiyle 10 kez tekrar edilir.

3

Uygulama

Bu çalışmada 2016 yılında yayınlanan TIMSS-2015 veri setinden yararlanılmıştır. TIMSS, dördüncü ve sekizinci sınıflardaki öğrencilerin matematik ve fen bilimleri alanındaki seviyelerini belirlemek için yapılan bir çalışmadır. Her 4 yılda bir gerçekleştirilen bu çalışmaya, 2015 yılında, sekizinci sınıf öğrencilerinin başarılarının değerlendirilmesi için 39 ülkenin yanı sıra 7 katılımcı ülke iştirak etmiştir. TIMSS'in ilk aşamasında, bu ülkelerden bir dizi okul orantılı bir şekilde belirlenmiştir. İkinci aşamada ise seçilen sınıflardan bazıları rastgele seçilmiştir. Çalışmada kullanılan anket, öğrencilerin verdiği cevaplara göre işlenmiştir [112].

Bu çalışmanın uygulaması iki bölümde incelenmiştir. Birincisi TIMSS veri setinde Türkiye'deki sekizinci sınıf öğrencilerinin fen bilimleri başarılarının sınıflandırılması, ikinci olarak ta yine sekizinci sınıf öğrencilerinin matematik başarılarının sınıflandırılmasıdır. Bu sınıflandırma başarılarının bulunması için makine öğrenmesi algoritmalarından yararlanılacaktır. Ayrıca bu başarı sınıflandırmalarını etkileyen değişkenler farklı öznitelik seçim algoritmaları ile belirlenecektir.

3.1 Deneysel Kurulum ve Uygulama

Bu çalışmanın deneysel kurulumu aşağıdaki şekilde verilebilir:

- TIMSS-2015 Türkiye 8. sınıf öğrencilerinin sonuçları odak grubu olarak ele alınır. Veri setinden mevcut olmayan ve eksik gözlemler temizlenir.
- Eğitim ve test veri kümeleri, 10-katlı çapraz doğrulama kullanılarak elde edilir.
- Sonuçlara etki edebilecek tüm potansiyel değişkenlerin kullanılmasıyla en iyi performans gösteren algoritma bulunur.

- Sınıflandırma başarısı için önemli olan ölçekler yardımıyla algoritmaların sınıflandırma başarıları incelenir.
- Başarılı sınıflandırmadaki en önemli özellikler 6 farklı öznitelik seçim algoritması kullanılarak elde edilir ve en iyi sınıflandırıcılar bu etkili değişkenler kullanılarak bulunur.
- 8. Sınıf fen bilimleri ve matematik öğrencilerinin başarısının sınıflandırılması, Tablo 3.1 ve Tablo 3.11'deki veri seti kullanılarak k-NN, NB, DVM-POLY, DVM-RBF, DVM-PUK, YSA, KA-C4.5, KA-RepTree, KA-RF ve LR ile yapılır.
- 8. Sınıf fen bilimleri ve matematik öğrencilerinin başarısının sınıflandırılmasında etkili olan öznitelikleri belirlemek için Cfs subset, correlation attribute, one-R, gain ratio, info gain ve relifF öznitelik seçim algoritmaları kullanılır.

3.2 Weka

Bu çalışmada kullanılacak verinin bağımlı ve bağımsız değişken değerleri ARFF dosyası olarak hazırlanmış ve analizleri yapmak amacıyla üzere WEKA programına aktarılmıştır. WEKA, veri madenciliğinde kullanılan makine öğrenmesi algoritmalarını ve öznitelik seçim algoritmalarını bünyesinde barındıran Java programlama dili tabanlı bir yazılımdır. Waikato üniversitesi tarafından geliştirilmiş ve Waikato Environment for Knowledge Analysis kelimelerinin baş harflerinden programın adı oluşturulmuştur [113].

3.3 Fen Bilimleri için Veri Seti

Birinci uygulama bölümünde, Türkiye'deki sekizinci sınıf öğrencilerinin fen bilimleri başarılarının sınıflandırılması ile ilgilenilmiştir. Fen bilimleri için veri seti 2943 kız, 3136 erkek olmak üzere toplam 6079 öğrenciden oluşmaktadır. Bununla birlikte, veri kümesindeki yanlış ve eksik bilgiler veri setinden çıkarılmıştır. Sonuç olarak 2273 kız, 2208 erkek olmak üzere toplamda 4481 öğrenci uygulamaya dahil edilmiştir. Veri setinin modellenmesinde tüm değişkenler kullanılmaktadır. Ancak, "yaş" gibi bazı değişkenlerde değişkenlik olmadığı için veri kümesinden

çıkarılmıştır. Tablo 3.1'de 35 bağımsız ve 1 bağımlı değişken olmak üzere veri setinde kullanılan değişkenler gösterilmiştir. 35 bağımsız değişken öğrencilerin fen bilimleri başarısının sınıflandırılması için önemli faktörlerdir ve bu faktörler makine öğrenmesi yöntemleri için öznitelik olarak ifade edilmektedir. Tablo 3.1'de görüldüğü gibi, "Fen bilimleri birinci değerlendirme sonucu" (BSSSCI01) bağımlı değişkendir ve öğrencilerin fen başarısını ifade etmektedir. TIMSS-2015'in ortalama fen bilimleri puanı, 100 standart sapma ile 500 olarak belirlenmiştir [5]. Bu bilgiler doğrultusunda bir öğrencinin fen bilimleri puanı eğer 500'den yüksek ise 1 (başarılı); 500'den düşük ise 0 (başarısız) olarak kodlanmıştır. Türkiye için hesaplanan ortalama başarı değeri ise 493 olarak bulunmuştur.

Tablo 3.1 Öğrencilerin Fen Bilimleri başarısının sınıflandırılmasında kullanılan tüm değişkenler

Faktör adı	Tanım	Cevap
ITSEX	Cinsiyet	1= Kadın, 2=Erkek
BSBG03	Aile ile iletişim	1=Her zaman, 2=Zaman zaman, 3=Bazen, 4= Asla
BSBG05	Dijital bilgi cihaz sayısı	1=Hiç, 2=1-3 cihaz, 3=4-6 cihaz, 4=7-10 cihaz, 5= 10 cihazdan daha fazla
BSBG06A	Kendi tablet bilgisayarı	1=Evet, 2=Hayır
BSBG06B	Paylaşımlı tablet bilgisayar	1=Evet, 2=Hayır
BSBG06C	Çalışma masası	1=Evet, 2=Hayır
BSBG06F	Cep telefonu	1=Evet, 2=Hayır
BSBG06G	Oyun sistemi	1=Evet, 2=Hayır
BSBG06H	Isınma sistemi	1=Evet, 2=Hayır
BSBG06I	Soğutma sistemi	1=Evet, 2=Hayır
BSBG06J	Çamaşır makinesi	1=Evet, 2=Hayır
BSBG06K	Bulaşık makinesi	1=Evet, 2=Hayır
BSBG08	Eğitimde ilerlemek istediğiniz düzey	1=Orta okul, 2= Lise, 3= Lise sonrası (yüksek öğrenim olmayan), 4=Üniversite Önlisans, 5=Üniversite lisans, 6=Lisansüstü
BSBG11	Hangi sıklıkta okula gidilmediği	1=Haftada bir ya da daha fazla, 2=İki haftada bir, 3=Ayda bir, 4=Çoğunlukla giderim
BSBG12	Okul günlerinde hangi sıklıkla kahvaltı yapıldığı	1=Hergün, 2=Çoğu günler, 3=Bazen, 4=Nerdeyse hiç
BSBG13A	Hangi sıklıkta tablet bilgisayar kullanıldığı (evde)	1=Neredeyse hergün, 2=Haftada bir ya da iki, 3=Ayda bir ya da iki, 4=Neredeyse hiç
BSBG13B	Hangi sıklıkta tablet bilgisayar kullanıldığı (okulda)	1=Neredeyse hergün, 2=Haftada bir ya da iki, 3=Ayda bir ya da iki, 4=Neredeyse hiç
BSBG13C	Hangi sıklıkta tablet bilgisayar kullanıldığı (diğer yerlerde)	1=Neredeyse hergün, 2=Haftada bir ya da iki, 3=Ayda bir ya da iki, 4=Neredeyse hiç
BSBG14A	Ders kitabına erişim	1=Evet, 2=Hayır
BSBG14B	Ödevlere erişim	1=Evet, 2=Hayır
BSBG14C	Sınıf arkadaşlarıyla işbirliği	1=Evet, 2=Hayır
BSBG14D	Öğretmen ile iletişim	1=Evet, 2=Hayır
BSBG14E	Fen bilimlerinde yardımcı bilgiye erişim	1=Evet, 2=Hayır
BSBS25AB	Öğretmenin hangi sıklıkta ödev verdiği (Fen bilimleri)	1=Her gün, 2=Haftada 3 ya da 4 kez, 3= Haftada 1 ya da 2 kez, 4=Haftada bir ya da hiç, 5=Asla
BSBS26AB	Son bir yılda alınan ek ders (Fen bilimleri)	1=Evet, sınıfta başarılı iken, 2=Evet, sınıfta başarısız iken, 3=Hayır
BSBS26BB	Kaç ay ek ders alındığı (Fen bilimleri)	1=Hiç, 2=4 aydan az

		, 3=4 ya da 8 ay, 4=8 aydan fazla
BSBGHER	Ev eğitim kaynakları	
BSBGSSB	Öğrencilerin okula ait olma hissi	
BSBGSB	Öğrenci zorbalığı	
BSBGSLS	Öğrencinin fen bilimlerini öğrenmeyi sevmesi	
BSBGESL	Fen bilimlerindeki öğrenmeye olan sempatisi	
BSBGSCS	Fen bilimlerinde öğrencinin güveni	
BSBGSVS	Öğrencinin fen bilimlerine verdiği değer	
BSDSLOWP	Tahmin için fen bilimleri başarısı çok düşük	1=Evet, 2=Hayır
BSDSWKHS	Fen bilimleri ödevinde haftalık harcanan zaman	1=3 saat ya da daha fazla, 2=45 dakikadan fazla 3 saatten az, 3=45 dakikadan az
BSSSCI01	Fen bilimleri birinci değerlendirme sonucu	0=Başarısız, 1= Başarılı

*Veri setinin orijinal hali Ek A.1'de verilmiştir.

3.3.1 Fen Bilimleri Veri Seti için Ölçekler

TIMSS-2015 çalışması ölçek adı verilen tek bir gizli yapıyı ölçmek için geliştirilmiştir. Raporlamada, Item Response Theory (IRT) ölçeklendirme yöntemlerinden biri olan Rasch kısmi kredi modeli kullanılmıştır [114]. Çalışmada 9 adet ölçek mevcuttur. Bu ölçekler sırasıyla; "Ev eğitim kaynakları" (BSBGHER), "Öğrenci zorbalığı" (BSBGSB), "Öğrencilerin okula ait olma hissi" (BSBGSSB), "Fen bilimlerindeki öğrenmeye olan sempatisi" (BSBGESL), "Öğrencinin fen bilimlerine verdiği değer" (BSBGSVS), "Fen bilimlerinde öğrencinin güveni" (BSBGSCS), "Tahmin için fen bilimleri başarısı çok düşük" (BSDSLOWP), "Öğrencinin fen bilimlerini öğrenmeyi sevmesi" (BSBGSLS) ve "Fen bilimleri ödevinde haftalık harcanan zaman" (BSDSWKHS) şeklindedir.

Fen Bilimleri veri setinde ev eğitim kaynakları (BSBGHER) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.2'de verilmiştir.

Tablo 3.2 Fen Bilimleri veri seti için ev eğitim kaynakları ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBG04	Evdeki kitap sayısı	1) 0-10 2) 11-25 3) 26-100 4) 101-200 5) 200'den fazla
BSDG06S	Ev çalışmalarındaki desteklerin sayısı	1) Yok 2) İnternet bağlantısı ya da kendi odası olma 3) İkiside
BSDGEDUP	Ebeveylerinin eğitim düzeyi	1) İlkokul, tamamlanmamış orta okul ya da hiç okula gitmemiş 2) Orta okul 3) Lise 4) Lise sonrası (yüksek öğrenim olmayan) 5) Üniversite ya da üzeri

*Veri setinin orijinal hali Ek A.2'de verilmiştir.

Fen Bilimleri veri setinde öğrenci zorbalığı (BSBGSB) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.3'te verilmiştir.

Tablo 3.3 Fen Bilimleri veri seti için öğrenci zorbalığı ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBG16A	Benimle, ad ya da soyadımla dalga geçilmesi	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16B	Beni oyun ya da aktivitelerine katmamaları	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16C	Hakkımda yalan yayılması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16D	Benden birşeyler çalınması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16E	Beni itmeleri ya da incitmeleri	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16F	Yapmak istemediğim şeylerin yaptırılması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16G	Hakkımda utanç verici bilgi yayılması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16H	Web üzerinden hakkımda utanç verici şeylerin yayınlanması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16I	Tehdit edilmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.3'te verilmiştir.

Fen Bilimleri veri setinde öğrencilerin okula ait olma hissi (BSBGSSB) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.4'te verilmiştir.

Tablo 3.4 Fen Bilimleri veri seti için öğrencilerin okula ait olma hissi ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBG15A	Okulda olmayı sevmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15B	Okuldayken güvende hissetmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15C	Okuldayken kendini okula ait hissetmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15D	Okulda arkadaşlarını görmekten hoşlanmak	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15E	Okuldaki öğretmenlerin adil olması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15F	Bu okula gitmekten gurur duymak	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15G	Okulda çok şey öğrenmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.4'te verilmiştir.

Fen Bilimleri veri setinde fen bilimlerindeki öğrenmeye olan sempatisi (BSBGESL) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.5'te verilmiştir.

Tablo 3.5 Fen Bilimleri veri seti için fen bilimlerindeki öğrenmeye olan sempatisi ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBS22A	Öğretmenimin benden ne yapmamı beklediğini biliyorum	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS22B	Öğretmenimi anlamak kolaydır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS22C	Öğretmenimin ne dediği ile ilgileniyorum	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS22D	Öğretmenim benim ilginç şeyler yapmamı sağlıyor	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS22E	Öğretmenim sorularıma açık cevap verir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS22F	Öğretmenim bana fen bilimlerini iyi açıklar	1) Kesinlikle katılıyorum

		2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS22G	Öğretmenim öğrendiklerimi göstermeme izin verir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS22H	Öğretmenim öğrenmemize yardımcı olacak şeyler yapar	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS22I	Öğretmenim hata yaptığımda nasıl daha iyisini yapmam gerektiğini söyler	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

***Veri setinin orijinal hali Ek A.5'te verilmiştir.

Fen Bilimleri veri setinde öğrencinin fen bilimlerine verdiği değer (BSBGSVS) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.6'da verilmiştir.

Tablo 3.6 Fen Bilimleri veri seti için öğrencinin fen bilimlerine verdiği değer ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBS24A	Fen bilimleri öğrenmek günlük hayatta bana yardımcı olacaktır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS24B	Diğer okul konularını öğrenmek için fen bilimlerine ihtiyaç vardır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS24C	Gitmek istediğim üniversite için fen bilimlerinde başarılı olmak gerekir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS24D	İstediğim iş için fen bilimlerinde başarılı olmak gerekir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS24E	Fen bilimleri kullanmayı içeren bir iş istiyorum	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS24F	Dünyada ilerlemek için fen bilimlerini öğrenmek önemlidir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS24G	Yetişkin olduğumda fen bilimleri bana daha fazla iş imkanı sağlayacak	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS24H	Ailem fen bilimlerinde iyi düzeyde olmamın önemli olduğunu düşünür	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS24I	Fen bilimlerinde başarılı olmak önemlidir.	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.6'da verilmiştir.

Fen Bilimleri veri setinde fen bilimlerinde öğrencinin güveni (BSBGSCS) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.7'de verilmiştir.

Tablo 3.7 Fen Bilimleri veri seti için fen bilimlerinde öğrencinin güveni ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBS23A	Fen bilimlerinde genellikle başarılıyımdır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS23B	Bir çok sınıf arkadaşıma göre fen bilimlerinde daha çok zorlanırım	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS23C	Fen bilimleri güçlü yanlarımdan biri değildir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS23D	Fen bilimlerindeki şeyleri kolaylıkla öğrenirim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS23E	Fen bilimlerinin zor problemlerini çözmekte başarılıyım	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS23F	Öğretmenim fen bilimlerinde iyi olduğumu söyler	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS23G	Fen bilimleri benim için diğer konulardan daha zordur	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS23H	Fen bilimleri beni şaşırtır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.7'de verilmiştir.

Fen Bilimleri veri setinde tahmin için fen bilimleri başarısı çok düşük (BSDSLOWP) ölçeğini oluşturan soru ve cevap seçeneği Tablo 3.8'de verilmiştir.

Tablo 3.8 Fen Bilimleri veri seti için tahmin için fen bilimleri başarısı çok düşük ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSDSLOWP	Tahmin için fen bilimleri başarısı çok düşük	1) Evet 2) Hayır

*Veri setinin orijinal hali Ek A.8'de verilmiştir.

Fen Bilimleri veri setinde öğrencinin fen bilimlerini öğrenmeyi sevmesi (BSBGSLS) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.9'da verilmiştir.

Tablo 3.9 Öğrencinin fen bilimlerini öğrenmeyi sevmesi ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBS21A	Fen bilimlerini öğrenirken eğlenirim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS21B	Keşke fen bilimleri çalışmak zorunda kalmasam	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS21C	Fen bilimleri sıkıcıdır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS21D	Fen bilimlerinde bir çok ilginç şey öğrenirim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS21E	Fen bilimlerini severim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS21F	Okulda fen bilimleri öğrenmeyi heyecanla bekliyorum	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS21G	Fen bilimleri dünyadaki sistemin nasıl işlediğini öğretir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS21H	Fen bilimleri deneyleri yapmayı severim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBS21I	Fen bilimleri en sevdiğim konulardan biridir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.9'da verilmiştir.

Fen Bilimleri veri setinde fen bilimleri ödevinde haftalık harcanan zaman (BSDSWKHS) ölçeğini oluşturan soru ve cevap seçeneği Tablo 3.10'da verilmiştir.

Tablo 3.10 Fen Bilimleri veri seti için fen bilimleri ödevinde haftalık harcanan zaman ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSDSWKHS	Fen bilimleri ödevinde haftalık harcanan zaman	1) 3 saat ya da daha fazla 2) 45 dakikadan fazla 3 saatten az 3) 45 dakikadan az

*Veri setinin orijinal hali Ek A.10'da verilmiştir.

3.3.2 Fen Bilimleri Veri Seti için Uygulama

Tez çalışmasının bu bölümünde, Türkiye'deki sekizinci sınıf fen bilimleri öğrencilerinin TIMSS-2015 sonuçları kullanılmıştır. Yanlış ve eksik değerler veri

setinden çıkarılmıştır. Eğitim ve test veri setlerini türetmek için 10-katlı çapraz doğrulama kullanılmıştır. Fen Bilimleri veri seti uygulaması için izlenen süreç aşağıda verilmiştir:

Adım 1: Algoritmaların sınıflandırma performanslarını sınıflandırma kriterleri ile belirlemek için Tablo 3.1'de gösterilen 35 değişkenin tamamı kullanılmıştır. Bu adım, algoritmaların performansını gösterir. Ayrıca ölçeklerin ve öznitelik seçiminin önemini ortaya koymak için diğer tüm aşamalarla (Adım 2 ve Adım 3) karşılaştırma yapmak için kullanılmaktadır.

Adım 2: Bu adımda veri setinde belirtilen ölçekler kullanılmıştır. Ölçekler kullanılarak algoritmalar çalıştırılmış ve sınıflandırma performansları Tablo 4.4'te verilmiştir. Kullanılan ölçekler; "Ev eğitim kaynakları" (BSBGHER), "Öğrenci zorbalığı" (BSBGSB), "Öğrencilerin okula ait olma hissi" (BSBGSSB), "Fen bilimlerindeki öğrenmeye olan sempatisi" (BSBGESL), "Öğrencinin fen bilimlerine verdiği değer" (BSBGSVS), "Fen bilimlerinde öğrencinin güveni" (BSBGSCS), "Tahmin için fen bilimleri başarısı çok düşük" (BSDSLOWP), "Öğrencinin fen bilimlerini öğrenmeyi sevmesi" (BSBGSLS) ve "Fen bilimleri ödevinde haftalık harcanan zaman" (BSDSWKHS)'dır. Bu adımın amacı, öğrencilerin fen bilimleri başarısının sınıflandırılmasında ölçeklerin etkinliğini belirlemektir.

Adım 3: Öznitelik seçim algoritmaları kullanılarak sınıflandırma başarısında en etkili olan öznitelikler ortaya çıkarılmıştır. Kullanılan öznitelik seçim yöntemleri Cfs subset, correlation attribute, one-R, gain ratio, info gain ve relifF'dir. Öznitelik seçim algoritmaları, 35 değişkene uygulanarak en etkili öznitelikler elde edilmiştir. Seçilen özniteliklerin üzerine analizler tekrar yapılmıştır. Yeni öznitelik kümeleri kullanılarak, tüm sınıflandırma algoritmaları için sınıflandırma başarı değerleri elde edilmiş ve Tablo 4.5 - Tablo 4.14'te verilmiştir.

3.4 Matematik için Veri Seti

Fen bilimleri veri setinde olduğu gibi, yine TIMSS-2015 çalışmasının sekizinci sınıftaki Türk öğrencilerinin matematik başarıları ele alınmıştır. Toplam 6079 öğrencinin katıldığı çalışmada 2943 kız ve 3136 erkek öğrenci bulunmaktadır. Bazı eksik ve yanlış değerlerin göz ardı edilmesi ile birlikte 2303 kız, 2274 erkek ve

toplamda 4577 öğrenci çalışmaya dahil edilmiştir. Veri setinin modellenmesinde tüm değişkenler kullanılmaktadır. Ancak, "yaş" gibi bazı değişkenlerde değişkenlik olmadığı için veri kümesinden çıkarılmıştır. Toplamda 36 değişken kullanılmış ve değişkenler Tablo 3.11'de gösterilmiştir. Bu değişkenlerden "Matematik birinci değerlendirme sonucu" (BSMMAT01) değişkeni öğrencilerin matematik başarısını gösterir ve çalışmada bağımlı değişken olarak seçilmiştir. TIMSS-2015'in ortalama matematik puanı, 100 standart sapma ile 500 olarak belirlenmiştir [5]. Bu nedenle, TIMSS çalışmasının ortalama değeri merkez nokta olarak kabul edilmiş ve öğrencilerin "Matematik birinci değerlendirme sonucu" puanı eğer 500'den büyük ise 1 (başarılı), değilse 0 (başarısız) olarak kodlamıştır. TIMSS tarafından belirlenen sekizinci sınıftaki Türk öğrencilerin ortalama matematik puanları 458'dir. Diğer 35 değişken ise bağımsız değişkenler olarak tanımlanmış, bir başka ifadeyle bu değişkenler öğrencilerin matematik başarısının sınıflandırılmasında önemli değişkenler olarak ele alınmıştır.

Tablo 3.11 Öğrencilerin matematik başarısının sınıflandırılmasında kullanılan tüm değişkenler

Faktör adı	Tanım	Cevap
ITSEX	Cinsiyet	1= Kadın, 2=Erkek
BSBG03	Aile ile iletişim	1=Her zaman, 2=Zaman zaman, 3=Bazen, 4= Asla
BSBG05	Dijital bilgi cihaz sayısı	1=Hiç, 2=1-3 cihaz, 3=4-6 cihaz, 4=7-10 cihaz, 5= 10 cihazdan daha fazla
BSBG06A	Kendi tablet bilgisayarı	1=Evet, 2=Hayır
BSBG06B	Paylaşımlı tablet bilgisayar	1=Evet, 2=Hayır
BSBG06C	Çalışma masası	1=Evet, 2=Hayır
BSBG06F	Cep telefonu	1=Evet, 2=Hayır
BSBG06G	Oyun sistemi	1=Evet, 2=Hayır
BSBG06H	Isınma sistemi	1=Evet, 2=Hayır
BSBG06I	Soğutma sistemi	1=Evet, 2=Hayır
BSBG06J	Çamaşır makinesi	1=Evet, 2=Hayır
BSBG06K	Bulaşık makinesi	1=Evet, 2=Hayır
BSBG08	Eğitimde ilerlemek istediğiniz düzey	1=Orta okul, 2= Lise, 3= Lise sonrası (yüksek öğrenim olmayan), 4=Üniversite Önlisans, 5=Üniversite lisans, 6=Lisansüstü
BSBG11	Hangi sıklıkta okula gidilmediği	1=Haftada bir ya da daha fazla, 2=İki haftada bir, 3=Ayda bir, 4=Çoğunlukla giderim
BSBG12	Okul günlerinde hangi sıklıkla kahvaltı yapıldığı	1=Hergün, 2=Çoğu günler, 3=Bazen, 4=Nerdeyse hiç
BSBG13A	Hangi sıklıkta tablet bilgisayar kullanıldığı (evde)	1=Neredeyse hergün, 2=Haftada bir ya da iki, 3=Ayda bir ya da iki, 4=Neredeyse hiç
BSBG13B	Hangi sıklıkta tablet bilgisayar kullanıldığı (okulda)	1=Neredeyse hergün, 2=Haftada bir ya da iki, 3=Ayda bir ya da iki, 4=Neredeyse hiç
BSBG13C	Hangi sıklıkta tablet bilgisayar kullanıldığı (diğer yerlerde)	1=Neredeyse hergün, 2=Haftada bir ya da iki, 3=Ayda bir ya da iki, 4=Neredeyse hiç
BSBG14A	Ders kitabına erişim	1=Evet, 2=Hayır
BSBG14B	Ödevlere erişim	1=Evet, 2=Hayır
BSBG14C	Sınıf arkadaşlarıyla işbirliği	1=Evet, 2=Hayır

BSBG14D	Öğretmen ile iletişim	1=Evet, 2=Hayır
BSBG14E	Matematikte yardımcı bilgiye erişim	1=Evet, 2=Hayır
BSBM38AA	Öğretmenin hangi sıklıkta ödev verdiği (Matematik)	1=Her gün, 2=Haftada 3 ya da 4 kez, 3= Haftada 1 ya da 2 kez, 4=Haftada bir ya da hiç, 5=Asla
BSBM39AA	Son bir yılda alınan ek ders (Matematik)	1=Evet, sınıfta başarılı iken, 2=Evet, sınıfta başarısız iken, 3=Hayır
BSBM39BA	Kaç ay ek ders alındığı (Matematik)	1=Hiç, 2=4 aydan az, 3=4 ya da 8 ay, 4=8 aydan fazla
BSBGHER	Ev eğitim kaynakları	
BSBGSSB	Öğrencilerin okula ait olma hissi	
BSBGSB	Öğrenci zorbalığı	
BSBGSLM	Öğrencinin matematik öğrenmeyi sevmesi	
BSBGEML	Matematikteki öğrenmeye olan sempatisi	
BSBGSCM	Matematikte öğrencinin güveni	
BSBGSVM	Öğrencinin matematiğe verdiği değer	
BSDMLOWP	Tahmin için matematik başarısı çok düşük	1=Evet, 2=Hayır
BSDMWKHW	Matematik ödevinde haftalık harcanan zaman	1=3 saat ya da daha fazla, 2=45 dakikadan fazla 3 saatten az, 3=45 dakikadan az
BSMMAT01	Matematik birinci değerlendirme sonucu	0=Başarısız, 1= Başarılı

*Veri setinin orijinal hali Ek A.11'de verilmiştir.

3.4.1 Matematik Veri Seti için Ölçekler

TIMSS-2015 çalışması ölçek adı verilen tek bir gizli yapıyı ölçmek için geliştirilmiştir. Raporlamada, Item Response Theory (IRT) ölçeklendirme yöntemlerinden biri olan Rasch kısmi kredi modeli kullanılmıştır [114]. Çalışmada 9 adet ölçek mevcuttur. Bunlar sırasıyla; "Ev eğitim kaynakları" (BSBGHER), "Öğrencilerin okula ait olma hissi" (BSBGSSB), "Öğrenci zorbalığı" (BSBGSB), "Öğrencinin matematik öğrenmeyi sevmesi" (BSBGSLM), "Matematikteki öğrenmeye olan sempatisi" (BSBGEML), "Matematikte öğrencinin güveni" (BSBGSCM), "Öğrencinin matematiğe verdiği değer" (BSBGSVM), "Tahmin için matematik başarısı çok düşük" (BSDMLOWP) ve "Matematik ödevinde haftalık harcanan zaman" (BSDMWKHW) olarak verilebilir.

Matematik veri setinde ev eğitim kaynakları (BSBGHER) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.12'de verilmiştir.

Tablo 3.12 Matematik veri seti için ev eğitim kaynakları ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBG04	Evdeki kitap sayısı	1) 0-10 2) 11-25 3) 26-100 4) 101-200 5) 200'den fazla
BSDG06S	Ev çalışmalarındaki desteklerin sayısı	1) Yok 2) İnternet bağlantısı ya da kendi odası olma 3) İkiside
BSDGEDUP	Ebeveylerinin eğitim düzeyi	1) İlkokul, tamamlanmamış orta okul ya da hiç okula gitmemiş 2) Orta okul 3) Lise 4) Lise sonrası (yüksek öğrenim olmayan) 5) Üniversite ya da üzeri

*Veri setinin orijinal hali Ek A.12'de verilmiştir.

Matematik veri setinde öğrencilerin okula ait olma hissi (BSBGSSB) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.13'te verilmiştir.

Tablo 3.13 Matematik veri seti için öğrencilerin okula ait olma hissi ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBG15A	Okulda olmayı sevmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15B	Okuldayken güvende hissetmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15C	Okuldayken kendini okula ait hissetmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15D	Okulda arkadaşlarını görmekten hoşlanmak	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15E	Okuldaki öğretmenlerin adil olması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15F	Bu okula gitmekten gurur duymak	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG15G	Okulda çok şey öğrenmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.13'te verilmiştir.

Matematik veri setinde öğrenci zorbalığı (BSBGSB) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.14'te verilmiştir.

Tablo 3.14 Matematik veri seti için öğrenci zorbalığı ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBG16A	Benimle, ad ya da soyadımla dalga geçilmesi	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16B	Beni oyun ya da aktivitelerine katmamaları	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16C	Hakkımda yalan yayılması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16D	Benden birşeyler çalınması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16E	Beni itmeleri ya da incitmeleri	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16F	Yapmak istemediğim şeylerin yaptırılması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16G	Hakkımda utanç verici bilgi yayılması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16H	Web üzerinden hakkımda utanç verici şeylerin yayınlanması	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBG16I	Tehdit edilmek	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.14'te verilmiştir.

Matematik veri setinde öğrencinin matematiği öğrenmeyi sevmesi (BSBGSLM) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.15'te verilmiştir.

Tablo 3.15 Matematik veri seti için öğrencinin matematiği öğrenmeyi sevmesi ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBM17A	Matematik öğrenirken eğlenirim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM17B	Keşke matematik çalışmak zorunda kalmasam	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM17C	Matematik sıkıcıdır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM17D	Matematikte birçok ilginç şey öğrenirim	1) Kesinlikle katılıyorum 2) Katılıyorum

		3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM17E	Matematiği severim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM17F	Rakam içeren herhangi bir okul çalışmasını severim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM17G	Matematk problemlerini çözmeyi severim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM17H	Okulda matematik öğrenmeyi heyecanla bekliyorum	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM17I	Matematik en sevdiğim konulardan biridir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.15'te verilmiştir.

Matematik veri setinde matematik öğrenmeye olan sempatisi (BSBGEML) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.16'da verilmiştir.

Tablo 3.16 Matematik veri seti için matematik öğrenmeye olan sempatisi ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBM18A	Öğretmenimin benden ne yapmamı beklediğini biliyorum	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM18B	Öğretmenimi anlamak kolaydır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM18C	Öğretmenimin ne dediği ile ilgileniyorum	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM18D	Öğretmenim benim ilginç şeyler yapmamı sağlıyor	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM18E	Öğretmenim sorularıma açık cevap verir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM18F	Öğretmenim bana matematiği iyi açıklar	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM18G	Öğretmenim öğrendiklerimi göstermeme izin verir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM18H	Öğretmenim öğrenmemize yardımcı olacak şeyler yapar	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum

		4) Kesinlikle katılmıyorum
BSBM18I	Öğretmenim hata yaptığımda nasıl daha iyisini yapmam gerektiğini söyler	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM18J	Öğretmenim söylediklerimi dinler	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.16'da verilmiştir.

Matematik veri setinde matematikte öğrencinin güveni (BSBGSCM) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.17'de verilmiştir.

Tablo 3.17 Matematik veri seti için matematikte öğrencinin güveni ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBM19A	Matematikte genellikle başarılıyımdır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM19B	Birçok sınıf arkadaşıma göre matematikte daha çok zorlanırım	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM19C	Matematik güçlü yanlarımdan biri değildir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM19D	Matematikteki şeyleri kolaylıkla öğrenirim	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM19E	Matematik beni tedirgin eder	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM19F	Matematik zor problemlerini çözmekte başarılıyım	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM19G	Öğretmenim matematikte iyi olduğumu söyler	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM19H	Matematik benim için diğer konulardan daha zordur	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM19I	Matematik beni şaşırtır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.17'de verilmiştir.

Matematik veri setinde öğrencinin matematiğe verdiği değer (BSBGSVM) ölçeğini oluşturan sorular ve cevap seçenekleri Tablo 3.18'de verilmiştir.

Tablo 3.18 Matematik veri seti için öğrencinin matematiğe verdiği değer ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSBM20A	Matematik öğrenmek günlük hayatta bana yardımcı olacaktır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM20B	Diğer okul konularını öğrenmek için matematiğe ihtiyaç vardır	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM20C	Gitmek istediğim üniversite için matematikte başarılı olmak gerekir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM20D	İstediğim iş için matematikte başarılı olmak gerekir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM20E	Matematik kullanmayı içeren bir iş istiyorum	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM20F	Dünyada ilerlemek için matematiği öğrenmek önemlidir	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM20G	Yetişkin olduğumda matematik bana daha fazla iş imkanı sağlayacak	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM20H	Ailem matematikte iyi düzeyde olmamın önemli olduğunu düşünür	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum
BSBM20I	Matematikte başarılı olmak önemlidir.	1) Kesinlikle katılıyorum 2) Katılıyorum 3) Katılmıyorum 4) Kesinlikle katılmıyorum

*Veri setinin orijinal hali Ek A.18'de verilmiştir.

Matematik veri setinde tahmin için matematik başarısı çok düşük (BSDMLOWP) ölçeğini oluşturan soru ve cevap seçeneği Tablo 3.19'da verilmiştir.

Tablo 3.19 Matematik veri seti için tahmin için matematik başarısı çok düşük ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSDMLOWP	Tahmin için matematik başarısı çok düşük	1) Evet 2) Hayır

*Veri setinin orijinal hali Ek A.19'da verilmiştir.

Matematik veri setinde matematik ödevinde haftalık harcanan zaman (BSDMWKHW) ölçeğini oluşturan soru ve cevap seçeneği Tablo 3.20'de verilmiştir.

Tablo 3.20 Matematik veri seti için matematik ödevinde haftalık harcanan zaman ölçeğinin içeriği

Soru kodu	Soru	Cevap
BSDMWKHW	Matematik ödevinde haftalık harcanan zaman	1) 3 saat ya da daha fazla 2) 45 dakikadan fazla 3 saatten az 3) 45 dakikadan az

*Veri setinin orijinal hali Ek A.20'de verilmiştir.

3.4.2 Matematik Veri Seti için Uygulama

Tez çalışmasının bu bölümünde, Türkiye'deki sekizinci sınıf matematik öğrencilerinin TIMSS-2015 sonuçları kullanılmıştır. Yanlış ve eksik değerler veri setinden çıkarılmıştır. Eğitim ve test veri setlerini türetmek için 10-katlı çapraz doğrulama kullanılmıştır. Matematik veri seti uygulaması için izlenen süreç aşağıda listelenmiştir:

Adım 1. Algoritmaların sınıflandırma performanslarını, sınıflandırma kriterleri ile belirlemek için Tablo 3.11'de gösterilen 35 değişkenin tamamı kullanılmıştır. Bu adım, algoritmaların performansını gösterir. Ayrıca ölçeklerin ve öznitelik seçiminin önemini ortaya koymak için diğer tüm aşamalarla (Adım 2 ve Adım 3) karşılaştırma yapmak için kullanılmaktadır.

Adım 2. Veri seti bölümünde açıklanan ölçekler bu adımda sınıflandırma başarısını etkileyen en önemli öznitelikler olarak kullanılmaktadır. Ölçekler kullanılarak algoritmalar çalıştırılmış ve sınıflandırma performansları Tablo 4.19'da verilmiştir. Kullanılan ölçekler; "Ev eğitim kaynakları" (BSBGHER), "Öğrencilerin okula ait olma hissi" (BSBGSSB), "Öğrenci zorbalığı" (BSBGSB), "Öğrencinin matematik öğrenmeyi sevmesi" (BSBGSLM), "Matematikteki öğrenmeye olan sempatisi" (BSBGEML), "Matematikte öğrencinin güveni" (BSBGSCM), "Öğrencinin matematiğe verdiği değer" (BSBGSVM), "Tahmin için matematik başarısı çok düşük" (BSDMLOWP) ve "Matematik ödevinde haftalık harcanan zaman" (BSDMWKHW)'dır. Burada amaç, öğrencilerin matematik başarısının sınıflandırılmasında ölçeklerin etkinliğini belirlemektir.

Adım 3. Öznitelik seçim algoritmaları kullanılarak öğrencilerin matematik başarısının sınıflandırılmasında en etkili olan öznitelikler ortaya çıkarılmıştır. Kullanılan öznitelik seçim yöntemleri Cfs subset, correlation attribute, one-R, gain ratio, info gain ve relifF'dir. Öznitelik seçim algoritmaları, 35 değişkene

uygulanarak en etkili öznitelikler elde edilmiştir. Seçilen öznitelikler kullanılarak analizler tekrar yapılmıştır. Yeni öznitelik kümeleri kullanılarak, tüm sınıflandırma algoritmaları için sınıflandırma başarı değerleri elde edilmiş ve Tablo 4.20 - Tablo 4.31'de verilmiştir.

4

Sonuçlar

TIMSS-2015 Türkiye 8. sınıf öğrencilerinin fen bilimleri ve matematik başarılarını sınıflandırmak ve etkili öznitelikler belirlemek amacıyla gerçekleştirilen analizlerin sonuçları ve yorumlarıbu kısımda verilecektir. Weka programı ile yapılan analiz sonuçları fen bilimleri veri seti için Tablo 4.1 – Tablo 4.14'te, matematik veri seti için Tablo 4.16 – Tablo 4.31'de gösterilmiştir. Daha önce belirtilmiş olan sınıflandırma kriterlerine göre (DP oranı, YP oranı, Precision, κ istatistiği, F-ölçütü, MAE, RMSE, ROC ve MCC) algoritmaların sınıflandırma performansları incelenmiştir. Uygulamanın 3.4.2 ve 3.5.2 bölümlerinde bahsedildiği üzere analizler 3 adımda gerçekleşmiştir. Adım 1'de tüm veri seti, adım 2'de ölçekler ve adım 3'te de öznitelik seçim algoritmaları kullanılmıştır. Her adımda en iyi algoritma, DP oranı ile belirlenmiştir. Ayrıca en iyi seçilen algoritmanın DP oranı dışında diğer sınıflandırma kriterlerine göre değerlendirmesi de yapılmıştır. Aynı zamanda, en iyi olarak seçilen algoritmanın ROC alanlarının değerleri ile sınıflandırma performansları görsel olarak sunulmuştur. Fen bilimleri veri seti için Tablo 4.15'te; matematik veri seti için Tablo 4.32'de elde edilen sonuçların genel gösterimi mevcuttur.

4.1 Fen Bilimleri Veri Seti için Sonuçlar

Adım 1; Tüm değişkenler kullanılarak elde edilen karşılaştırma matrisi değerleri Tablo 4.1'de ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.2'de verilmiştir. Tablo 4.2'de görüldüğü gibi, DP oranına göre LR (0,738) en iyi algoritma olarak belirlenmiştir. LR için YP Oranı (0,272), Precision (0,738), F-Ölçütü (0,737), κ istatistiği (0,469), RMSE (0,415), ROC alanı (0,820) ve MCC (0,470) değerleri bu sonucu desteklemektedir. Ayrıca, KA-RF ve DVM-POLY, LR ile karşılaştırıldığında benzer performanslara sahip olduğu söylenebilir.

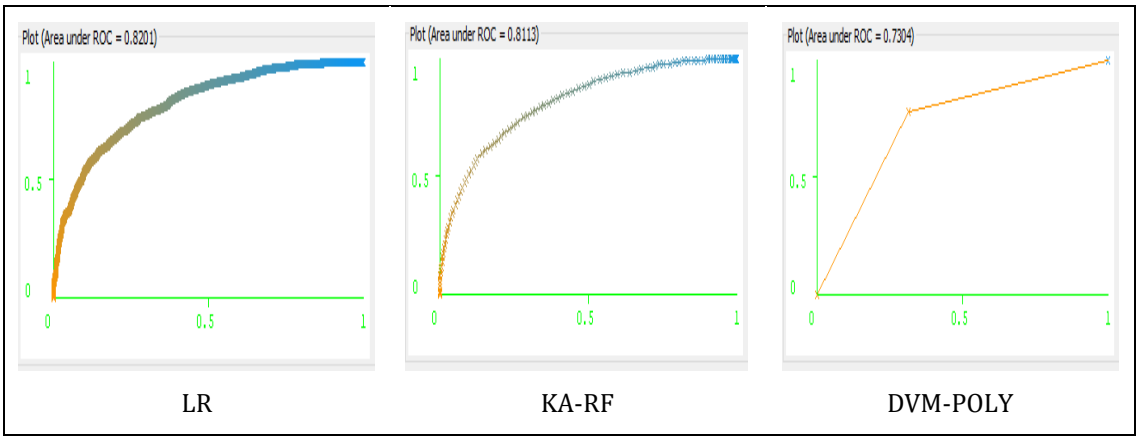
Tablo 4.1 Fen Bilimleri veri seti adım 1 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	1071	967	1261	777	1276	762	1323	715	1378	660
b=1	779	1664	514	1929	728	1715	647	1796	542	1901
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	1254	784	1397	641	1325	1254	784	1397	641	1325
b=1	642	1801	549	1894	522	642	1801	549	1894	522

Tablo 4.1'de verilen karşılaştırma matrisine göre LR, 4481 öğrenci verisinin 3308 (1384 + 1924) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 2735 (1071 + 1664), NB 3190 (1261 + 1929), KA-C4.5 2991 (1276 + 1715), KA-RepTree 3119 (1323 + 1796), KA-RF 3279 (1378 + 1901), YSA 3055 (1254 + 1801), DVM-POLY 3291 (1397 + 1894), DVM-RBF 3246 (1325 + 1921) ve DVM-PUK 3237 (1342 + 1895) tane doğru sınıflandırma yapmıştır.

Tablo 4.2 Fen Bilimleri veri seti adım 1'e göre sınıflandırma sonuçları (35 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,610	0,712	0,668	0,696	0,732	0,682	0,734	0,724	0,722	**0,738**
YP Oranı	0,404	0,304	0,339	0,312	0,277	0,329	0,274	0,288	0,288	**0,272**
Precision	0,608	0,712	0,667	0,695	0,731	0,681	0,734	0,724	0,722	**0,738**
F-Ölçütü	0,608	0,709	0,667	0,696	0,731	0,681	0,734	0,723	0,721	**0,737**
κ statistic	0,208	0,413	0,329	0,385	0,457	0,355	0,463	0,440	0,437	**0,469**
MAE	0,390	0,306	0,354	0,373	0,371	0,322	**0,266**	0,276	0,278	0,342
RMSE	0,624	0,469	0,542	0,463	0,422	0,530	0,515	0,525	0,527	**0,415**
ROC Area	0,605	0,789	0,661	0,739	0,811	0,749	0,730	0,718	0,717	**0,820**
MCC	0,209	0,416	0,329	0,386	0,457	0,355	0,463	0,442	0,438	**0,470**



Şekil 4.1 Fen Bilimleri veri seti adım 1 için başarılı algoritmaların ROC eğrileri

Şekil 4.1'de, fen bilimleri veri seti ve adım 1 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,605'in üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,820).

Adım 2; Tek bir gizli yapıyı ortaya çıkarmak için geliştirilen ölçekler kullanılarak elde edilen karşılaştırma matrisi değerleri Tablo 4.3'te ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.4'te verilmiştir. Kullanılan 9 adet ölçek şunlardır: "Ev eğitim kaynakları", "Öğrenci zorbalığı", "Öğrencilerin okula ait olma hissi", "Fen bilimlerindeki öğrenmeye olan sempatisi", "Öğrencinin fen bilimlerine verdiği değer", "Fen bilimlerinde öğrencinin güveni", "Tahmin için fen bilimleri başarısı çok düşük", "Öğrencinin fen bilimlerini öğrenmeyi sevmesi" ve "Fen bilimleri ödevinde haftalık harcanan zaman". Tablo 4.4 incelendiğinde, DP oranına göre DVM-POLY (0,703) en iyi algoritma olarak belirlenmiştir. DVM-POLY'nin YP Oranı (0,303), Precision (0,702), F-Ölçütü (0,702), κ istatistiği (0,400), MAE (0,298) ve MCC (0,400) değerleri bu sonucu desteklemektedir. Ayrıca, LR ve DVM-PUK'un, DVM-POLY ile karşılaştırıldığında benzer performanslara sahip olduğu söylenebilir.

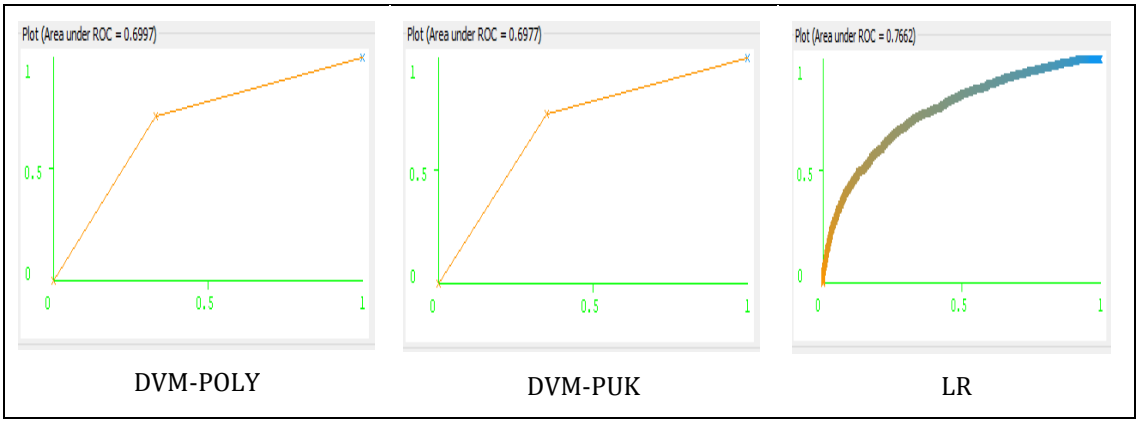
Tablo 4.3 Fen Bilimleri veri seti adım 2 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	1198	840	1372	666	1301	737	1318	720	1344	694
b=1	866	1577	791	1652	711	1732	746	1697	707	1736
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	1384	654	1362	676	1156	882	1324	714	1309	729
b=1	708	1735	657	1786	522	1921	621	1822	612	1831

Tablo 4.3'de verilen karşılaştırma matrisine göre DVM-POLY, 4481 öğrenci verisinin 3148 (1362 + 1786) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 2775 (1198 + 1577), NB 3024 (1372 + 1652), KA-C4.5 3033 (1301 + 1732), KA-RepTree 3015 (1318 + 1697), KA-RF 3080 (1344 + 1736), YSA 3199 (1384 + 1735), DVM-RBF 3077 (1156 + 1921), DVM-PUK 3146 (1324 + 1822) ve LR 3140 (1309 + 1831) tane doğru sınıflandırma yapmıştır.

Tablo 4.4 Fen Bilimleri veri seti adım 2'ye göre sınıflandırma sonuçları (9 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,619	0,675	0,677	0,673	0,687	0,696	**0,703**	0,687	0,702	0,701
YP Oranı	0,386	0,325	0,330	0,331	0,317	0,307	**0,303**	0,333	0,307	0,309
Precision	0,620	0,677	0,677	0,673	0,688	0,697	**0,702**	0,687	0,701	0,700
F-Ölçütü	0,619	0,675	0,677	0,673	0,687	0,696	**0,702**	0,682	0,701	0,700
κ statistic	0,233	0,348	0,348	0,341	0,370	0,388	**0,400**	0,359	0,397	0,394
MAE	0,381	0,383	0,389	0,394	0,388	0,381	**0,298**	0,313	**0,298**	0,391
RMSE	0,616	0,460	0,487	0,468	0,451	0,447	0,545	0,560	0,546	**0,442**
ROC Area	0,614	0,737	0,696	0,721	0,749	0,758	0,700	0,677	0,698	**0,766**
MCC	0,233	0,348	0,348	0,341	0,370	0,389	**0,400**	0,364	0,397	0,394



Şekil 4.2 Fen Bilimleri veri seti adım 2 için başarılı algoritmaların ROC eğrileri

Şekil 4.2'de, fen bilimleri veri seti ve adım 2 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,614'ün üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,766).

Adım 3; Bu adımda, altı farklı öznitelik seçim algoritmasına (Cfs subset, correlation attribute, one-R, gain ratio, info gain ve relifF) göre, sınıflandırma başarısında en etkili öznitelikler belirlenmiştir. Ardından algoritmaların sınıflandırma performansları hesaplanmıştır. Bu işlemler adım 3.1 - adım 3.5 ile aşağıda verilmiştir:

Adım3.1; Cfs-subset öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Aile ile iletişim", "Paylaşımlı tablet bilgisayar", "Çalışma masası", "Eğitimde ilerlemek istediğiniz düzey", "Hangi sıklıkta okula gidilmediği", "Okul günlerinde hangi sıklıkla kahvaltı yapıldığı", "Öğretmen ile iletişim", "Öğretmenin hangi sıklıkta ödev verdiği (Fen bilimleri)", "Son bir yılda alınan ek ders (Fen bilimleri)", "Kaç ay ek ders alındığı (Fen bilimleri)", "Ev eğitim kaynakları", "Fen bilimlerinde öğrencinin güveni" ve "Tahmin için fen bilimleri başarısı çok düşük" olarak belirlenmiştir. Bu 13 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.5'te ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.6'da verilmiştir. DP oranına göre DVM-POLY (0,734) ve LR (0,733) en iyi algoritmalar olarak belirlenmiştir. DVM-POLY için YP oranı (0,277), Precision (0,733), F-Ölçütü (0,733), κ istatistiği (0,460), MAE (0,266) ve MCC (0,461) değerleri bu sonucu desteklemektedir. LR için ise RMSE (0,421) ve ROC alanı (0,809) bu sonucu desteklemiştir.

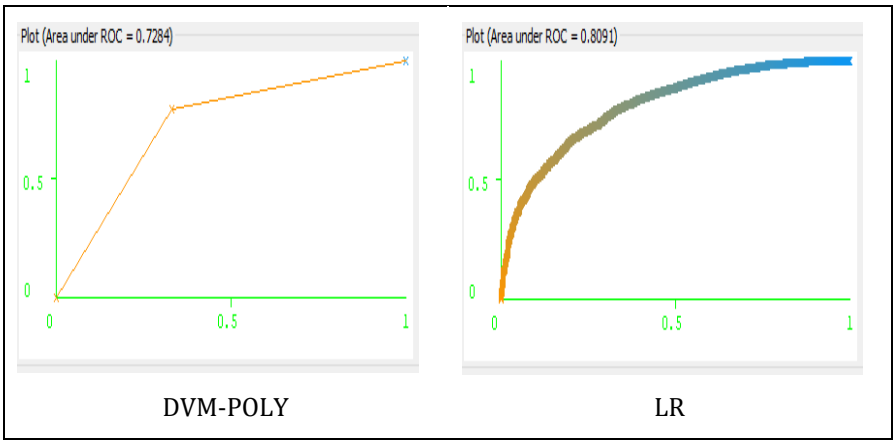
Tablo 4.5 Fen Bilimleri veri seti adım 3.1 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	1217	821	1214	824	1288	750	1319	719	1360	678
b=1	815	1628	448	1995	669	1774	630	1813	608	1835
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	1377	661	1359	679	1271	767	1343	695	1348	690
b=1	610	1833	513	1930	488	1955	590	1853	508	1935

Tablo 4.5'te verilen karşılaştırma matrisine göre DVM-POLY, 4481 öğrenci verisinin 3289 (1359 + 1930) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 2845 (1217 + 1628), NB 3209 (1214 + 1995), KA-C4.5 3062 (1288 + 1774), KA-RepTree 3132 (1319 + 1813), KA-RF 3195 (1360 + 1835), YSA 3210 (1377 + 1833), DVM-RBF 3226 (1271 + 1955), DVM-PUK 3196 (1343 + 1853) ve LR 3283 (1348 + 1935) tane doğru sınıflandırma yapmıştır.

Tablo 4.6 Fen Bilimleri veri seti adım 3.1'e göre sınıflandırma sonuçları (13 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,635	0,716	0,683	0,699	0,713	0,716	**0,734**	0,720	0,713	0,733
YP Oranı	0,371	0,304	0,325	0,310	0,295	0,290	**0,277**	0,296	0,296	0,279
Precision	0,635	0,718	0,683	0,698	0,712	0,716	**0,733**	0,720	0,712	0,732
F-Ölçütü	0,635	0,712	0,683	0,698	0,713	0,716	**0,733**	0,717	0,712	0,731
κ statistic	0,264	0,419	0,359	0,391	0,420	0,427	**0,460**	0,429	0,419	0,457
MAE	0,366	0,313	0,364	0,370	0,361	0,346	**0,266**	0,280	0,287	0,354
RMSE	0,604	0,454	0,493	0,453	0,435	0,434	0,516	0,529	0,536	**0,421**
ROC Area	0,636	0,794	0,707	0,754	0,783	0,791	0,728	0,712	0,709	**0,809**
MCC	0,264	0,425	0,360	0,391	0,420	0,427	**0,461**	0,432	0,420	0,458



Şekil 4.3 Fen Bilimleri veri seti adım 3.1 için başarılı algoritmaların ROC eğrileri

Şekil 4.3'te, fen bilimleri veri seti ve adım 3.1 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,636'nın üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,809).

Adım 3.2; Correlation attribute ve one R öznitelik seçim algoritmaları kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Paylaşımlı tablet bilgisayar", "Eğitimde ilerlemek istediğiniz düzey", "Hangi sıklıkta okula gidilmediği", "Son bir yılda alınan ek ders (Fen bilimleri)", "Kaç ay ek ders alındığı (Fen bilimleri)", "Ev eğitim kaynakları" ve "Fen bilimlerinde öğrencinin güveni" olarak belirlenmiştir. Bu iki öznitelik seçim algoritması aynı öznitelikleri seçmiştir. Seçilen 7 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.7'de ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.8'de verilmiştir. DVM-POLY (0,714), DVM-PUK (0,714) ve LR (0,713) en iyi algoritmalar olarak belirlenmiştir. DVM-POLY ve DVM-PUK için YP oranı (0,294), Precision (0,713), F-Ölçütü (0,713), κ istatistiği (0,421) ve MAE (0,286) değerleri bu sonucu desteklemektedir. Ayrıca DVM-PUK için MCC (0,422) değeri sonucu desteklemektedir. LR için ise RMSE (0,431) ve ROC alanı (0,791) bu sonucu desteklemiştir.

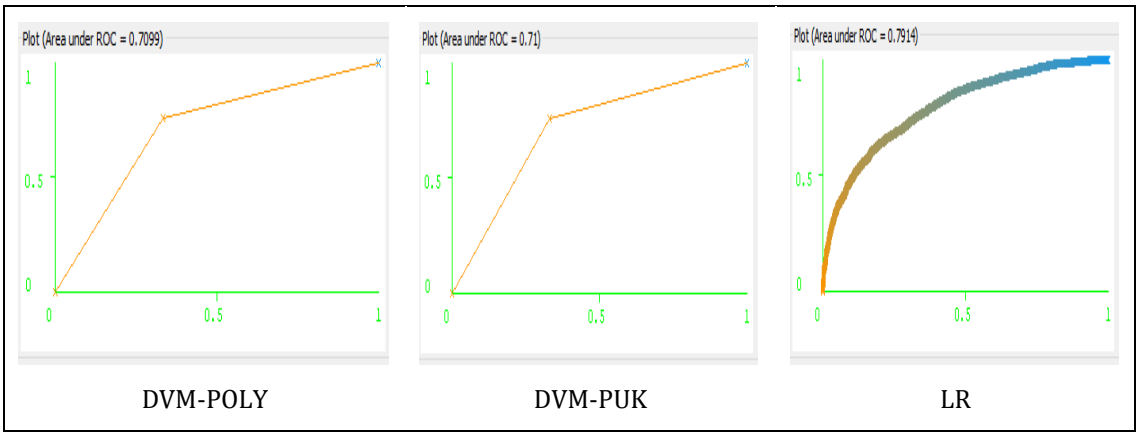
Tablo 4.7 Fen Bilimleri veri seti adım 3.2 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	1245	793	1268	770	1336	702	1346	692	1280	758
b=1	870	1573	539	1904	642	1801	639	1804	708	1735
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	1402	636	1357	681	1300	738	1359	679	1322	716
b=1	667	1776	601	1842	563	1880	603	1840	572	1871

Tablo 4.7'de verilen karşılaştırma matrisine göre 4481 öğrenci verisinin, DVM-POLY 3199 (1357 + 1842) ve DVM- PUK 3199 (1359 + 1840) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 2818 (1245 + 1573), NB 3172 (1268 + 1904), KA-C4.5 3137 (1336 + 1801), KA-RepTree 3150 (1346 + 1804), KA-RF 3015 (1280 + 1735), YSA 3178 (1402 + 1776), DVM-RBF 3180 (1300 + 1880) ve LR 3193 (1322 + 1871) tane doğru sınıflandırma yapmıştır.

Tablo 4.8 Fen Bilimleri veri seti adım 3.2'ye göre sınıflandırma sonuçları (7 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,629	0,708	0,700	0,703	0,673	0,709	**0,714**	0,710	**0,714**	0,713
YP O ranı	0,374	0,306	0,307	0,304	0,335	0,294	**0,294**	0,302	**0,294**	0,298
Precision	0,630	0,707	0,699	0,702	0,672	0,710	**0,713**	0,709	**0,713**	0,712
F-Ölçütü	0,629	0,706	0,700	0,703	0,672	0,709	**0,713**	0,708	**0,713**	0,711
κ statistic	0,254	0,405	0,394	0,400	0,339	0,414	**0,421**	0,410	**0,421**	0,417
MAE	0,371	0,330	0,377	0,377	0,370	0,366	**0,286**	0,290	**0,286**	0,371
RMSE	0,589	0,452	0,462	0,450	0,472	0,435	0,535	0,539	0,535	**0,431**
ROC Area	0,653	0,782	0,733	0,756	0,733	0,784	0,710	0,704	0,710	**0,791**
MCC	0,254	0,408	0,394	0,400	0,339	0,414	0,421	0,412	**0,422**	0,418



Şekil 4.4 Fen Bilimleri veri seti adım 3.2 için başarılı algoritmaların ROC eğrileri

Şekil 4.4'te, fen bilimleri veri seti ve adım 3.2 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,653'ün üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,791).

Adım 3.3; Gain ratio öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Aile ile iletişim" , "Paylaşımlı tablet bilgisayar", "Çamaşır makinesi", "Eğitimde ilerlemek istediğiniz düzey", "Son bir yılda alınan ek ders (Fen bilimleri)", "Kaç ay ek ders alındığı (Fen bilimleri)", "Ev eğitim kaynakları", "Fen bilimlerinde öğrencinin güveni" ve "Tahmin için fen bilimleri başarısı çok düşük" olarak belirlenmiştir. Seçilen 9 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.9'da ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.10'da verilmiştir. DP oranına göre DVM-PUK (0,719) en iyi algoritma olarak belirlenmiştir. DVM-PUK için YP oranı (0,291), Precision (0,719), F-Ölçütü (0,718), κ istatistiği (0,431), MAE (0,281) ve MCC (0,432) değerleri bu sonucu desteklemektedir. LR, DVM-POLY ve YSA algoritmaları da DVM-PUK'a yakın sonuçlar vermiştir.

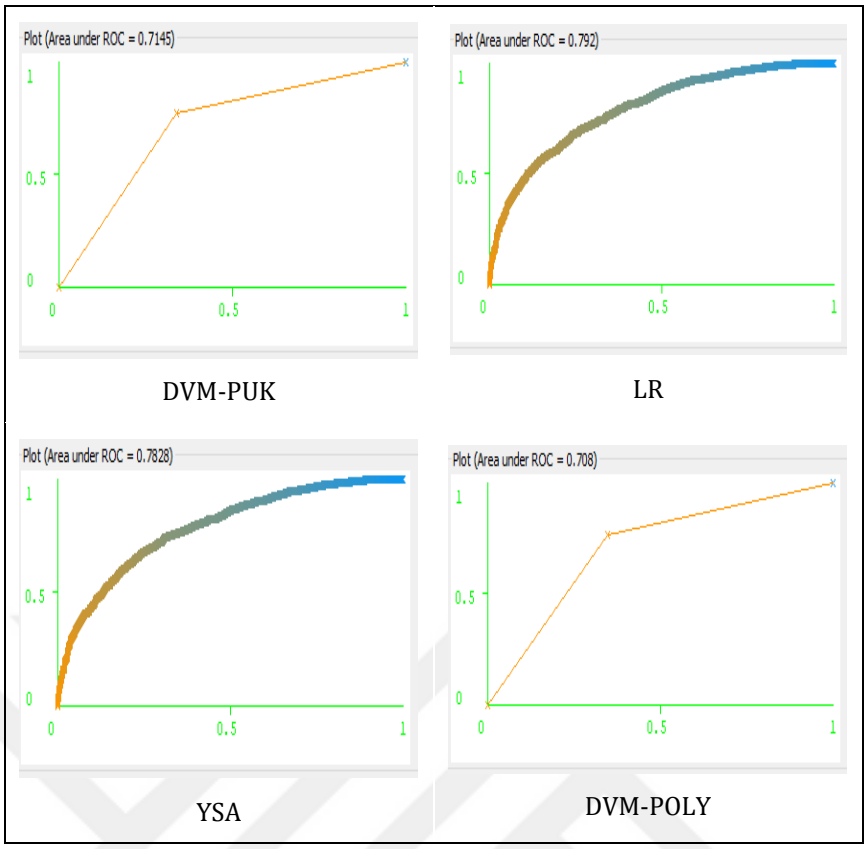
Tablo 4.9 Fen Bilimleri veri seti adım 3.3 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	1287	751	1220	818	1329	709	1294	744	1265	773
b=1	838	1605	515	1928	655	1788	625	1818	692	1751
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	1409	629	1325	713	1264	774	1343	695	1296	742
b=1	658	1785	572	1871	550	1893	562	1881	546	1897

Tablo 4.9'da verilen karşılaştırma matrisine göre DVM- PUK 4481 öğrenci verisinin 3224 (1343 + 1881) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 2892 (1287 + 1605), NB 3148 (1220 + 1928), KA-C4.5 3117 (1329 + 1788), KA-RepTree 3112 (1294 + 1818), KA-RF 3016 (1265 + 1751), YSA 3194 (1409 + 1785), DVM-POLY 3196 (1325 + 1871), DVM-RBF 3157 (1264 + 1893) ve LR 3193 (1296 + 1897) tane doğru sınıflandırma yapmıştır.

Tablo 4.10 Fen Bilimleri veri seti adım 3.3'e göre sınıflandırma sonuçları (9 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,645	0,703	0,696	0,694	0,673	0,713	0,713	0,705	**0,719**	0,713
YP O ranı	0,357	0,315	0,312	0,315	0,336	**0,291**	0,297	0,309	**0,291**	0,300
Precision	0,647	0,703	0,695	0,694	0,672	0,713	0,712	0,704	**0,719**	0,712
F-Ölçütü	0,646	0,699	0,695	0,694	0,672	0,713	0,712	0,702	**0,718**	0,711
κ statistic	0,288	0,393	0,385	0,381	0,339	0,422	0,418	0,399	**0,431**	0,416
MAE	0,361	0,330	0,374	0,378	0,366	0,364	0,287	0,296	**0,281**	0,369
RMSE	0,570	0,453	0,461	0,450	0,473	0,435	0,536	0,544	0,530	**0,430**
ROC Area	0,660	0,782	0,739	0,755	0,733	0,783	0,708	0,698	0,714	**0,792**
MCC	0,288	0,396	0,385	0,381	0,339	0,422	0,419	0,401	**0,432**	0,417



Şekil 4.5 Fen Bilimleri veri seti adım 3.3 için başarılı algoritmaların ROC eğrileri

Şekil 4.5'te, fen bilimleri veri seti ve adım 3.3 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,660'ın üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,792).

Adım 3.4; Info gain öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Dijital bilgi cihaz sayısı", "Paylaşımlı tablet bilgisayar", "Eğitimde ilerlemek istediğiniz düzey", "Hangi sıklıkta okula gidilmediği", "Son bir yılda alınan ek ders (Fen bilimleri)", "Kaç ay ek ders alındığı (Fen bilimleri)", "Ev eğitim kaynakları" ve "Fen bilimlerinde öğrencinin güveni" olarak belirlenmiştir. Bu 8 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.11'de ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.12'de verilmiştir. DP oranına göre DVM-PUK (0,714) ve LR (0,712) en iyi algoritmalar olarak belirlenmiştir. DVM-PUK için YP oranı (0,293), Precision (0,713), F-Ölçütü (0,714), κ istatistiği (0,422), MAE (0,286) ve MCC (0,422) değerleri bu sonucu desteklemektedir. LR için ise RMSE (0,431) ve ROC alanı

(0,791) bu sonucu desteklemiştir. DVM-POLY ise bu iki algoritmaya yakın sonuçlar vermiştir.

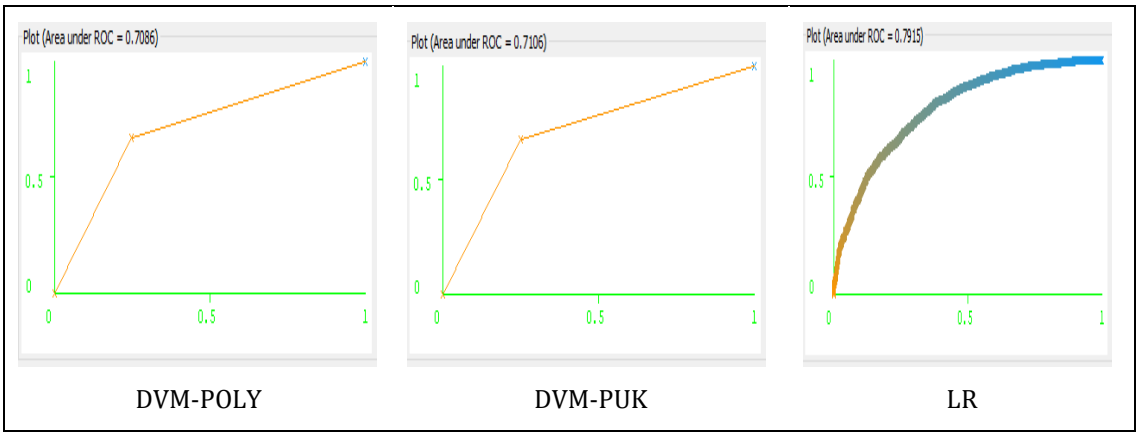
Tablo 4.11 Fen Bilimleri veri seti adım 3.4 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	1218	820	1292	746	1347	691	1324	714	1317	721
b=1	866	1577	559	1884	696	1747	628	1815	717	1726
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	1388	650	1355	683	1310	728	1373	665	1323	715
b=1	643	1800	605	1838	570	1873	617	1826	577	1866

Tablo 4.11'de verilen karşılaştırma matrisine göre DVM-PUK 4481 öğrenci verisinin 3199 (1373 + 1826) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 2795 (1218 + 1577), NB 3176 (1292 + 1884), KA-C4.5 3094 (1347 + 1747), KA-RepTree 3139 (1324 + 1815), KA-RF 3043 (1317 + 1726), YSA 3188 (1388 + 1800), DVM-POLY 3193 (1355 + 1838), DVM-RBF 3183 (1310 + 1873) ve LR 3189 (1323 + 1866) tane doğru sınıflandırma yapmıştır.

Tablo 4.12 Fen Bilimleri veri seti adım 3.4'e göre sınıflandırma sonuçları (8 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,624	0,709	0,690	0,701	0,679	0,711	0,713	0,710	**0,714**	0,712
YP O ranı	0,381	0,304	0,314	0,308	0,326	0,394	0,295	0,301	**0,293**	0,299
Precision	0,624	0,708	0,691	0,700	0,679	0,711	0,712	0,710	**0,713**	0,711
F-Ölçütü	0,624	0,707	0,691	0,700	0,679	0,711	0,712	0,709	**0,714**	0,711
κ statistic	0,243	0,408	0,376	0,394	0,353	0,418	0,419	0,412	**0,422**	0,415
MAE	0,375	0,326	0,374	0,379	0,372	0,364	0,287	0,290	**0,286**	0,371
RMSE	0,605	0,453	0,472	0,453	0,462	0,437	0,536	0,538	0,535	**0,431**
ROC Area	0,640	0,783	0,725	0,750	0,741	0,780	0,709	0,705	0,711	**0,791**
MCC	0,243	0,410	0,376	0,394	0,353	0,418	0,419	0,413	**0,422**	0,416



Şekil 4.6 Fen Bilimleri veri seti adım 3.4 için başarılı algoritmaların ROC eğrileri

Şekil 4.6'da, fen bilimleri veri seti ve adım 3.4 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,640'ın üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,791).

Adım 3.5; ReliefF öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Paylaşımlı tablet bilgisayar", "Eğitimde ilerlemek istediğiniz düzey", "Hangi sıklıkta okula gidilmediği", "Kaç ay ek ders alındığı (Fen bilimleri)" ve "Fen bilimlerinde öğrencinin güveni" olarak belirlenmiştir. Bu 5 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.13'te ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.14'te verilmiştir. DP oranına göre LR (0,704) en iyi algoritma olarak belirlenmiştir. Precision (0,703), F-Ölçütü (0,702), κ istatistiği (0,398), RMSE (0,440), ROC alanı (0,771) ve MCC (0,399) değerleri bu sonucu desteklemektedir. DVM-POLY, YSA ve DVM-PUK algoritmaları da LR'ye yakın sonuçlar vermiştir.

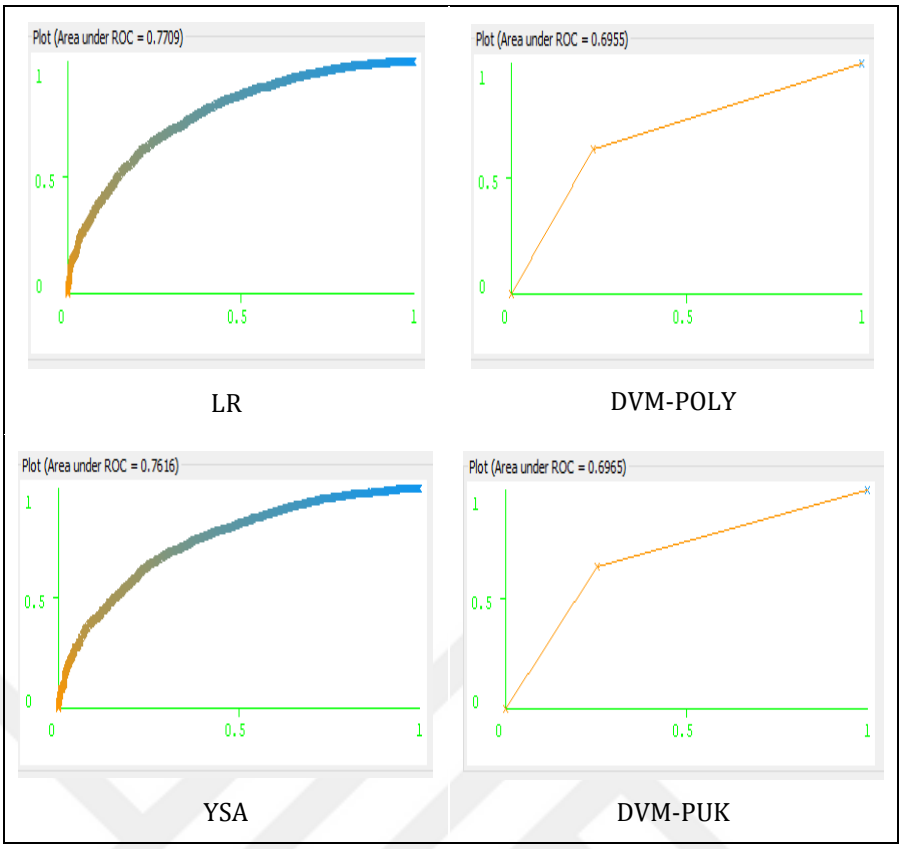
Tablo 4.13 Fen Bilimleri veri seti adım 3.5 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	1270	768	1127	911	1293	745	1281	757	1210	828
b=1	774	1669	451	1992	622	1821	631	1812	685	1758
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	1354	684	1274	764	1169	869	1319	719	1282	756
b=1	662	1781	572	1871	494	1949	621	1822	572	1871

Tablo 4.13'te verilen karşılaştırma matrisine göre LR 4481 öğrenci verisinin 3153 (1282 + 1871) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 2939 (1270 + 1669), NB 3119 (1127 + 1992), KA-C4.5 3114 (1293 + 1821), KA-RepTree 3093 (1281 + 1812), KA-RF 2968 (1210 + 1758), YSA 3135 (1354 + 1781), DVM-POLY 3145 (1274 + 1871), DVM-PUK 3141 (1319 + 1822) ve DVM-RBF 3118 (1169 + 1949) tane doğru sınıflandırma yapmıştır.

Tablo 4.14 Fen Bilimleri veri seti adım 3.5'e göre sınıflandırma sonuçları (5 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,656	0,696	0,695	0,690	0,662	0,700	0,702	0,696	0,701	**0,704**
YP O ranı	0,350	0,328	0,315	0,320	0,349	**0,306**	0,311	0,324	0,308	0,309
Precision	0,656	0,699	0,694	0,689	0,661	0,699	0,701	0,697	0,700	**0,703**
F-Ölçütü	0,656	0,690	0,694	0,689	0,661	0,699	0,700	0,691	0,700	**0,702**
κ statistic	0,306	0,375	0,382	0,372	0,315	0,394	0,394	0,377	0,395	**0,398**
MAE	0,386	0,355	0,395	0,390	0,384	0,387	**0,298**	0,304	0,299	0,387
RMSE	0,518	0,453	0,455	0,452	0,482	0,444	0,546	0,552	0,547	**0,440**
ROC Area	0,670	0,765	0,731	0,746	0,708	0,762	0,695	0,686	0,697	**0,771**
MCC	0,306	0,384	0,382	0,373	0,316	0,394	0,396	0,383	0,395	**0,399**



Şekil 4.7 Fen Bilimleri veri seti adım 3.5 için başarılı algoritmaların ROC eğrileri

Şekil 4.7'de, fen bilimleri veri seti ve adım 3.5 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,670'in üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,771).

Adım 1 ile adım 3 arasında elde edilen sonuçlar Tablo 4.15'te özetlenmiştir. Tablo 4.15 incelendiğinde tüm adımlarda, LR ve DVM-POLY, sınıflandırma kriterlerine göre en başarılı sınıflandırma sonuçlarını vermiştir. Adım 1 ve adım 3.1 dışında DVM-PUK, LR ve DVM-POLY algoritmalarına benzer sonuçlar göstermiştir. Adım 1 için KA-RF; Adım 3.3 ve adım 3.5 için YSA algoritması, LR ve DVM-POLY algoritmalarına yakın sonuçlar vermiştir. Ayrıca Tablo 4.15'te adım 3 için etkili öznitelikler gösterilmiştir. "Paylaşımlı tablet bilgisayar" (BSBG06B), "Eğitimde ilerlemek istediğiniz düzey" (BSBG08), "Kaç ay ek ders alındığı (Fen bilimleri" (BSBS26BB) ve "Fen bilimlerinde öğrencinin güveni" (BSBGSCS) adım 3.1 – adım 3.5 için belirlenen ortak özniteliklerdir. Öğrencilerin fen başarısını sınıflandırmada

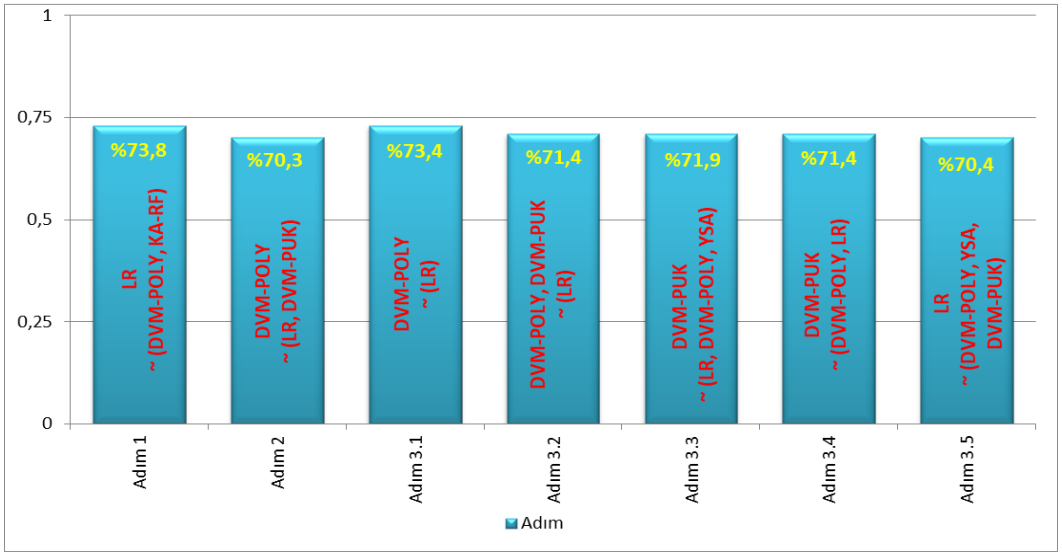
bu özniteliklerin en etkili öznitelikler olduğu söylenebilir. Tüm adımlardaki ROC eğrileri incelendiğinde LR'nin en başarılı sonuçlar verdiği gözlemlenmiştir.

Tablo 4.15 Fen Bilimleri veri seti adım 1 - adım 3 arasında genel sonuç tablosu

Adım	Algoritmalar	Değişkenler
Adım 1	LR, DVM-POLY, KA-RF	Tüm değişkenler
Adım 2	DVM-POLY, LR, DVM-PUK	BSBGHER, BSBGSSB, BSBGSB, BSBGSLS, BSBGESL, BSBGSCS, BSBGSVS, BSDSLOWP, BSDSWKHW
Adım 3.1	DVM-POLY, LR	BSBG03, BSBG06B, BSBG06C, BSBG08, BSBG11, BSBG12, BSBG14D, BSBS25AB, BSBS26AB, BSBS26BB, BSBGHER, BSBGSCS, BSDSLOWP
Adım 3.2	DVM-POLY, DVM-PUK, LR	BSBG06B, BSBG08, BSBG11, BSBS26AB, BSBS26BB, BSBGHER, BSBGSCS
Adım 3.3	DVM-PUK, LR, DVM-POLY, YSA	BSBG03, BSBG06B, BSBG06J, BSBG08, BSBS26AB, BSBS26BB, BSBGHER, BSBGSCS, BSDSLOWP
Adım 3.4	DVM-POLY, DVM-PUK, LR	BSBG05, BSBG06B, BSBG08, BSBG11, BSBS26AB, BSBS26BB, BSBGHER, BSBGSCS
Adım 3.5	LR, DVM-POLY, YSA, DVM-PUK	BSBG06B, BSBG08, BSBG11, BSBS26BB, BSBGSCS

Sınıflandırma algoritmalarının performansları arasındaki değişkenliği araştırmak için sınıflandırma kriterlerinden yararlanılır. Bu çalışmada DP oranı değişkenlikleri göstermek için seçilmiştir.

Fen Bilimleri veri setine yönelik olarak adım 1 – adım 3.5 arasında gerçekleştirilen tüm sınıflandırmalar için sınıflandırma kriterlerinden DP oranı baz alınarak belirlenen en iyi algoritmaları gösteren grafik Şekil 4.8'de verilmiştir. Şekil 4.8'e göre en iyi algoritmaya çok yakın başarı gösteren algoritmalar "~(...)" şeklinde gösterilmiştir.



Şekil 4.8 Fen Bilimleri veri seti için sınıflandırma sonuçları

Adım 1'de Tablo 3.1'de yer alan 35 özniteliğin tümü kullanılarak sınıflandırma algoritmalarının performansları belirlenmiştir. Bu adımda en yüksek DP oranı, LR (% 73,8) algoritmasına aittir. Ayrıca DVM-POLY (%73,4) ve KA-RF (%73,2) algoritmalarının DP oranları LR'ye yakın sonuçlar vermiştir. LR algoritmasının adım 1'de elde edilen başarı oranı adım 1 ile adım 3 arasındaki en yüksek sınıflandırma başarı değeri olarak bulunmuştur. Adım 1'e göre elde edilen sınıflandırma başarı değerleri diğer adımlar ile karşılaştırılmıştır. Adım 2 için 9 ölçek kullanılarak elde edilen sonuçlar incelendiğinde DP oranına göre DVM-POLY (% 70,3), DVM-PUK (% 70,2) ve LR (% 70,1) algoritmalarının en başarılı sınıflandırma performansı gösterdiği belirlenmiştir. Adım 3.1'de 13 etkili öznitelik kullanılarak elde edilen sonuçlar göz önüne alındığında DP oranına göre DVM-POLY (% 73,4) ve LR (% 73,3) algoritmaları sınıflandırmada en başarılı sonuçları vermiştir. Adım 3.2 için 7 etkili öznitelik kullanılarak elde edilen sonuçlar dikkate alındığında DP oranına göre DVM-POLY (% 71,4), DVM-PUK (% 71,4) ve LR (% 71,3) sınıflandırmada en başarılı algoritmalar olarak bulunmuştur. Adım 3.3'te 9 etkili öznitelik kullanılarak elde edilen sonuçlar incelendiğinde DP oranına göre DVM-PUK (% 71,9), LR (% 71,3), DVM-POLY (% 71,3) ve YSA (% 71,3) algoritmalarının en başarılı sınıflandırma performansı gösterdiği belirlenmiştir. Adım 3.4 için 8 öznitelik kullanılarak elde edilen sonuçlar göz önüne alındığında DP oranına göre DVM-PUK (%71,4), DVM-POLY (%71,3) ve LR (%71,2) algoritmaları sınıflandırmada en başarılı sonuçları vermiştir. Adım 3.5'te 5 öznitelik kullanılarak elde edilen sonuçlar dikkate alındığında DP oranına göre LR (%70,4), DVM-POLY(%70,2), DVM-PUK (%70,1) ve YSA (%70,0) sınıflandırmada en başarılı algoritmalar olarak bulunmuştur. Özellikle adım 3.1'de elde edilen algoritmaların sınıflandırma performanslarının, tüm adımlar incelendiğinde adım 1'e çok yakın sonuçlar vermesi öznitelik seçiminin ne denli önemli olduğunu vurgulamaktadır.

4.2 Matematik Veri Seti için Sonuçlar

Adım 1; Tüm değişkenler kullanılarak elde edilen karşılaştırma matrisi değerleri Tablo 4.16'da ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.17'de verilmiştir. Tablo 4.17 incelendiğinde, DP oranına göre KA-RF (0,805) ve LR (0,803) en iyi algoritma olarak belirlenmiştir. KA-RF için

Precision (0,804) ve MCC (0,584) değerleri; LR için YP Oranı (0,231), F-Ölçütü (0,802), κ istatistiği (0,581), RMSE (0,368) ve ROC alanı (0,880) değerleri bu sonucu desteklemektedir. Ayrıca DVM-POLY'nin, KA-RF ve LR ile karşılaştırıldığında benzer performanslara sahip olduğu söylenebilir.

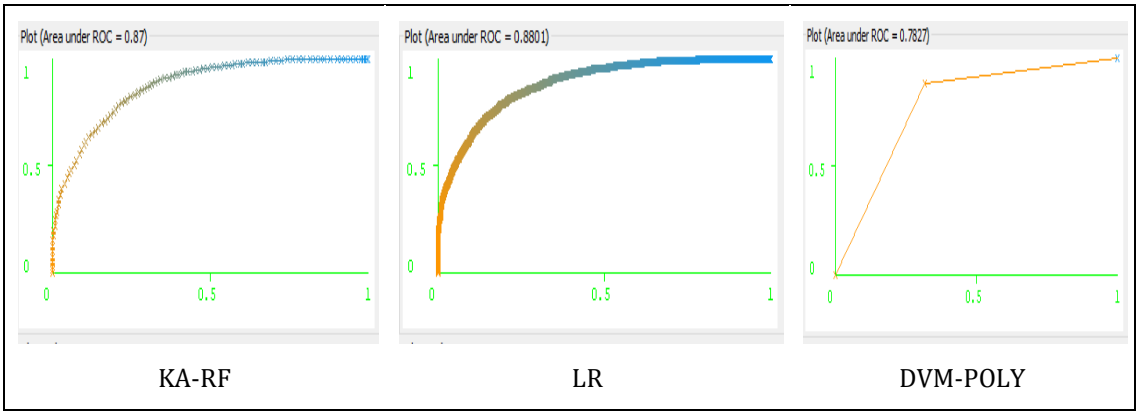
Tablo 4.16 Matematik veri seti adım 1 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	1860	925	2064	721	2225	560	2372	413	2478	307
b=1	786	1006	410	1382	588	1204	648	1144	586	1206
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	2241	544	2459	326	2463	322	2487	298	2411	374
b=1	585	1207	569	1223	639	1153	710	1082	526	1266

Tablo 4.16'da verilen karşılaştırma matrisine göre KA-RF 4577 öğrenci verisinin 3684 (2478 + 1206) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 2866 (1860 + 1006), NB 3446 (2064 + 1382), KA-C4.5 3429 (2225 + 1204), KA-RepTree 3516 (2372 + 1144), YSA 3448 (2241 + 1207), DVM-POLY 3682 (2459 + 1223), DVM-RBF 3616 (2463 + 1153), DVM-PUK 3569 (2487 + 1082) ve LR 3677 (2411 + 1266) tane doğru sınıflandırma yapmıştır.

Tablo 4.17 Matematik veri seti adım 1'e göre sınıflandırma sonuçları (35 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,626	0,753	0,749	0,768	**0,805**	0,753	0,804	0,790	0,780	0,803
YP Oranı	0,397	0,241	0,278	0,278	0,242	0,275	0,239	0,262	0,283	**0,231**
Precision	0,632	0,765	0,749	0,766	**0,804**	0,752	0,803	0,789	0,780	0,802
F-Ölçütü	0,628	0,755	0,749	0,765	0,801	0,753	0,801	0,786	0,773	**0,802**
κ statistic	0,226	0,497	0,472	0,502	0,579	0,480	0,579	0,545	0,518	**0,581**
MAE	0,374	0,261	0,278	0,300	0,303	0,250	**0,196**	0,210	0,220	0,270
RMSE	0,611	0,432	0,471	0,413	0,376	0,479	0,442	0,458	0,469	**0,368**
ROC Area	0,614	0,845	0,744	0,815	0,870	0,827	0,783	0,764	0,748	**0,880**
MCC	0,227	0,502	0,472	0,505	**0,584**	0,480	0,583	0,551	0,528	0,582



Şekil 4.9 Matematik veri seti adım 1 için başarılı algoritmaların ROC eğrileri

Şekil 4.9'da, matematik veri seti ve adım 1 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,614'ün üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,880).

Adım 2; Tek bir gizli yapıyı ortaya çıkarmak için geliştirilen ölçekler kullanılarak elde edilen karşılaştırma matrisi değerleri Tablo 4.18'de ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.19'da verilmiştir. Kullanılan 9 adet ölçek şunlardır: "Ev eğitim kaynakları", "Öğrencilerin okula ait olma hissi", "Öğrenci zorbalığı", "Öğrencinin matematik öğrenmeyi sevmesi", "Matematikteki öğrenmeye olan sempatisi", "Matematikte öğrencinin güveni", "Öğrencinin matematiğe verdiği değer", "Tahmin için matematk başarısı çok düşük" ve "Matematik ödevinde haftalık harcanan zaman". Tablo 4.19 incelendiğinde, DP oranına göre DVM-PUK (0,792) ve LR (0,790) en iyi algoritmalar olarak belirlenmiştir. DVM-PUK için Precision (0,792), F-Ölçütü (0,786), κ istatistiği (0,546), MAE (0,208) ve MCC (0,555); LR için F-Ölçütü (0,786), κ istatistiği (0,546), RMSE (0,385) VE ROC alanı (0,854) değerleri bu sonucu desteklemektedir. Ayrıca, DVM-POLY'nin, LR ve DVM-PUK ile karşılaştırıldığında benzer performanslara sahip olduğu söylenebilir.

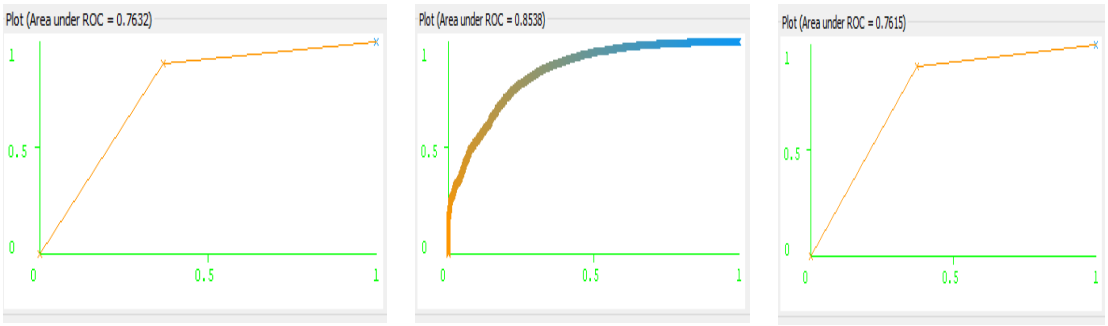
Tablo 4.18 Matematik veri seti adım 2 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	2131	654	2389	396	2399	386	2374	411	2424	361
b=1	712	1080	675	1117	634	1158	636	1156	628	1164
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	2385	400	2498	287	2625	160	2492	293	2441	344
b=1	588	1204	670	1122	941	851	660	1132	619	1173

Tablo 4.18'de verilen karşılaştırma matrisine göre DVM-PUK 4577 öğrenci verisinin 3624 (2492 + 1132) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 3211 (2131 + 1080), NB 3506 (2389 + 1117), KA-C4.5 3557 (2399 + 1158), KA-RF 3588 (2424 + 1164) KA-RepTree 3530 (2374 + 1156), YSA 3589 (2385 + 1204), DVM-POLY 3620 (2498 + 1122), DVM-RBF 3476 (2625 + 851) ve LR 3614 (2441 + 1173) tane doğru sınıflandırma yapmıştır.

Tablo 4.19 Matematik veri seti adım 2'ye göre sınıflandırma sonuçları (9 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,702	0,766	0,777	0,771	0,784	0,784	0,791	0,759	**0,792**	0,790
YP Oranı	0,334	0,285	0,270	0,274	0,264	**0,256**	0,268	0,342	0,265	0,259
Precision	0,700	0,763	0,775	0,769	0,782	0,782	**0,792**	0,777	**0,792**	0,788
F-Ölçütü	0,701	0,762	0,774	0,768	0,780	0,782	0,785	0,741	**0,786**	**0,786**
κ statistic	0,370	0,495	0,520	0,509	0,534	0,538	0,544	0,453	**0,546**	**0,546**
MAE	0,299	0,293	0,292	0,300	0,297	0,291	0,209	0,241	**0,208**	0,299
RMSE	0,546	0,411	0,420	0,410	0,390	0,391	0,457	0,491	0,456	**0,385**
ROC Area	0,685	0,822	0,791	0,818	0,845	0,844	0,762	0,709	0,763	**0,854**
MCC	0,370	0,499	0,524	0,512	0,538	0,540	0,553	0,491	**0,555**	0,551



Şekil 4.10 Matematik veri seti adım 2 için başarılı algoritmaların ROC eğrileri

Şekil 4.10'da, matematik veri seti ve adım 2 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,685'in üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,854).

Adım 3; Bu adımda, altı farklı öznitelik seçim algoritmasına (Cfs subset, correlation attribute, one-R, gain ratio, info gain ve relifF) göre, sınıflandırma başarısında en etkili öznitelikler belirlenmiştir. Ardından algoritmaların sınıflandırma performansları hesaplanmıştır. Bu işlemler adım 3.1 - adım 3.6 ile aşağıda verilmiştir.

Adım 3.1; Cfs-subset öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Aile ile iletişim", "Eğitimde ilerlemek istediğiniz düzey", "Hangi sıklıkta okula gidilmediği", "Okul günlerinde hangi sıklıkla kahvaltı yapıldığı", "Hangi sıklıkta tablet bilgisayar kullanıldığı (okulda)", "Öğretmenin hangi sıklıkta ödev verdiği (Matematik)", "Kaç ay ek ders alındığı (Matematik)", "Ev eğitim kaynakları", "Matematikte öğrencinin güveni" ve "Tahmin için matematk başarısı çok düşük" olarak belirlenmiştir. Bu 10 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.20'de ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.21'de verilmiştir. DP oranına göre LR (0,795) ve DVM-POLY (0,795) en iyi algoritmalar olarak belirlenmiştir. LR için Precision (0,794), F-Ölçütü (0,793), κ istatistiği (0,563), RMSE (0,378), ROC alanı (0,867) ve MCC (0,565); DVM-POLY için Precision (0,794) ve MAE (0,205) değerleri bu sonucu desteklemektedir. Ayrıca, DVM-PUK, LR ve DVM-POLY algoritmalarına benzer sonuçlar göstermiştir.

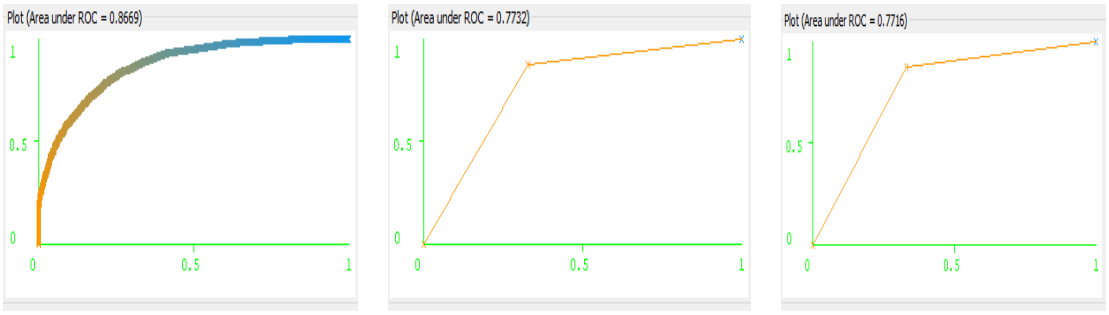
Tablo 4.20 Matematik veri seti adım 3.1 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	2100	685	2100	685	2339	446	2350	435	2376	409
b=1	695	1097	388	1404	553	1239	575	1217	565	1227
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	2375	410	2437	348	2512	273	2436	349	2410	375
b=1	549	1243	589	1203	718	1074	594	1198	561	1231

Tablo 4.20'de verilen karşılaştırma matrisine göre LR 4577 öğrenci verisinin 3641 (2410 + 1231) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 3197 (2100 + 1097), NB 3504 (2100 + 1404), KA-RepTree 3567 (2350 + 1217), KA-C4.5 3578 (2339 + 1239), KA-RF 3603 (2376 + 1227), YSA 3618 (2375 + 1243), DVM-POLY 3640 (2437 + 1203), DVM-RBF 3586 (2512 + 1074) ve DVM-PUK 3634 (2436 + 1198) tane doğru sınıflandırma yapmıştır.

Tablo 4.21 Matematik veri seti adım 3.1'e göre sınıflandırma sonuçları (10 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,698	0,766	0,782	0,779	0,787	0,790	**0,795**	0,783	0,794	**0,795**
YP Oranı	0,332	0,228	0,250	0,256	0,249	0,244	0,249	0,282	0,251	0,243
Precision	0,698	0,777	0,780	0,777	0,785	0,789	**0,794**	0,785	0,792	**0,794**
F-Ölçütü	0,698	0,768	0,780	0,778	0,785	0,789	0,792	0,776	0,791	**0,793**
κ statistic	0,367	0,522	0,537	0,530	0,546	0,554	0,560	0,525	0,557	**0,563**
MAE	0,301	0,274	0,284	0,293	0,285	0,278	**0,205**	0,217	0,206	0,287
RMSE	0,547	0,405	0,414	0,403	0,388	0,386	0,453	0,465	0,454	**0,378**
ROC Area	0,684	0,853	0,811	0,830	0,853	0,856	0,773	0,751	0,772	**0,867**
MCC	0,367	0,527	0,538	0,531	0,548	0,555	0,563	0,537	0,561	**0,565**



Şekil 4.11 Matematik veri seti adım 3.1 için başarılı algoritmaların ROC eğrileri

Şekil 4.11'de, matematik veri seti ve adım 3.1 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,684'ün üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,867).

Adım 3.2; Correlation attribute öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Eğitimde ilerlemek istediğiniz düzey", "Son bir yılda alınan ek ders (Matematik)", "Kaç ay ek ders alındığı (Matematik)", "Ev eğitim kaynakları", "Öğrencinin matematik öğrenmeyi sevmesi", "Matematikte öğrencinin güveni" ve "Tahmin için matematk başarısı çok düşük" olarak belirlenmiştir. Bu iki öznitelik seçim algoritması aynı öznitelikleri seçmiştir. Seçilen 7 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.22'de ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.23'te verilmiştir. DP oranına göre DVM-POLY (0,791), LR (0,790) ve YSA (0,789) en iyi algoritmalar olarak belirlenmiştir. DVM-POLY için Precision (0,790), F-Ölçütü (0,787), MAE (0,209) ve MCC (0,554); LR için F-Ölçütü (0,787), κ istatistiği (0,550), RMSE (0,381) ve ROC alanı (0,863); YSA için YP oranı (0,247), F-Ölçütü (0,787) ve κ istatistiği (0,550) değerleri bu sonucu desteklemektedir.

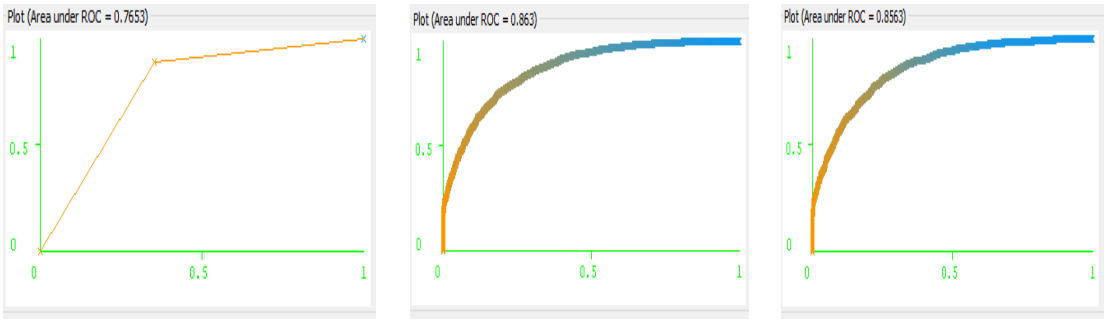
Tablo 4.22 Matematik veri seti adım 3.2 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	2182	603	2241	544	2375	410	2357	428	2325	460
b=1	662	1130	522	1270	633	1159	608	1184	589	1203
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	2374	411	2463	322	2511	274	2457	328	2420	365
b=1	556	1236	634	1158	772	1020	651	1141	594	1198

Tablo 4.20'de verilen karşılaştırma matrisine göre DVM-POLY 4577 öğrenci verisinin 3621 (2463 + 1158) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 3312 (2182 + 1130), NB 3511 (2241 + 1270), KA-C4.5 3534 (2375 + 1159) KA-RepTree 3541 (2357 + 1184), KA-RF 3528 (2325 + 1203), YSA 3610 (2374 + 1236), DVM-RBF 3531 (2511 + 1020), DVM-PUK 3598 (2457 + 1141) ve LR 3618 (2420 + 1198) tane doğru sınıflandırma yapmıştır.

Tablo 4.23 Matematik veri seti adım 3.2'ye göre sınıflandırma sonuçları (7 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,724	0,767	0,772	0,774	0,771	0,789	**0,791**	0,771	0,786	0,790
YP O ranı	0,310	0,254	0,273	0,267	0,265	**0,247**	0,261	0,301	0,267	0,253
Precision	0,722	0,768	0,770	0,771	0,769	0,787	**0,790**	0,774	0,785	0,789
F-Ölçütü	0,723	0,767	0,769	0,771	0,769	**0,787**	**0,787**	0,762	0,781	**0,787**
κ statistic	0,417	0,512	0,511	0,516	0,513	**0,550**	0,548	0,495	0,536	**0,550**
MAE	0,276	0,274	0,300	0,300	0,285	0,287	**0,209**	0,229	0,214	0,291
RMSE	0,522	0,409	0,408	0,403	0,399	0,385	0,457	0,478	0,463	**0,381**
ROC Area	0,717	0,840	0,819	0,830	0,842	0,856	0,765	0,735	0,759	**0,863**
MCC	0,417	0,512	0,514	0,518	0,514	0,551	**0,554**	0,510	0,543	0,553



Şekil 4.12 Matematik veri seti adım 3.2 için başarılı algoritmaların ROC eğrileri

Şekil 4.12'de, matematik veri seti ve adım 3.2 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,717'nin üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,863).

Adım 3.3; One-R öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Dijital bilgi cihaz sayısı", "Eğitimde ilerlemek istediğiniz düzey", "Kaç ay ek ders alındığı (Matematik)", "Ev eğitim kaynakları", "Öğrencinin matematik öğrenmeyi sevmesi" ve "Matematikte öğrencinin güveni" olarak belirlenmiştir. Bu iki öznitelik seçim algoritması aynı öznitelikleri seçmiştir. Seçilen 6 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.24'te ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.25'te verilmiştir. DP oranına göre DVM-POLY (0,784) ve LR (0,784) en iyi algoritmalar olarak belirlenmiştir. DVM-POLY için Precision (0,784), MAE (0,216) ve MCC (0,538); LR için F-Ölçütü (0,779), κ istatistiği (0,532), RMSE (0,390) ve ROC alanı (0,844) değerleri bu sonucu desteklemektedir. Bunlara ek olarak YSA, DVM-POLY ve LR algoritmalarına benzer sonuçlar göstermiştir.

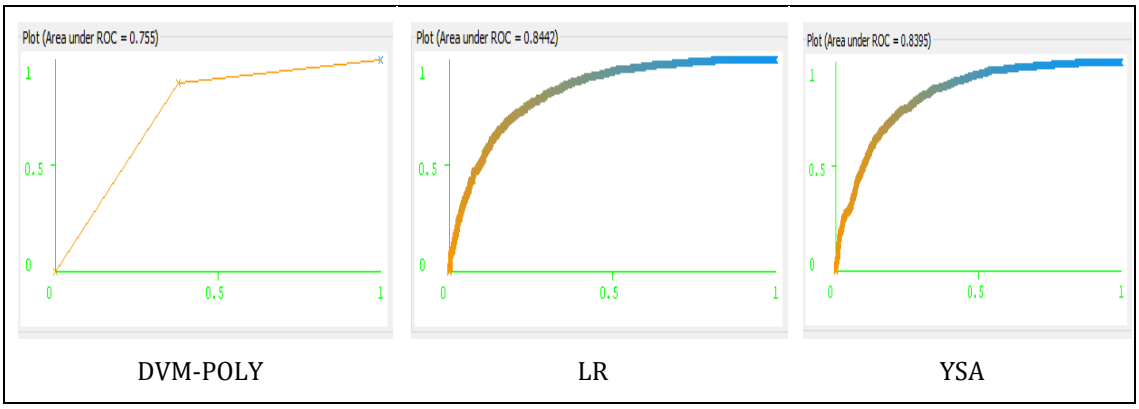
Tablo 4.24 Matematik veri seti adım 3.3 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	2118	667	2325	460	2390	395	2403	382	2340	445
b=1	694	1098	596	1196	652	1140	673	1119	652	1140
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	2350	435	2477	308	2545	240	2461	324	2443	342
b=1	573	1219	680	1112	856	936	673	1119	648	1144

Tablo 4.24'te verilen karşılaştırma matrisine göre DVM-POLY 4577 öğrenci verisinin 3589 (2477 + 1112) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 3216 (2118 + 1098), NB 3521 (2325 + 1196), KA-C4.5 3530 (2390 + 1140) KA-RepTree 3522 (2403 + 1119), KA-RF 3480 (2340 + 1140), YSA 3569 (2350 + 1219), DVM-RBF 3481 (2545 + 936), DVM-PUK 3580 (2461 + 1119) ve LR 3587 (2443 + 1144) tane doğru sınıflandırma yapmıştır.

Tablo 4.25 Matematik veri seti adım 3.3'e göre sınıflandırma sonuçları (6 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,703	0,769	0,771	0,769	0,760	0,780	**0,784**	0,761	0,782	**0,784**
YP O ranı	0,329	0,267	0,277	0,282	0,284	**0,256**	0,274	0,324	0,274	0,268
Precision	0,702	0,767	0,769	0,767	0,757	0,778	**0,784**	0,767	0,781	0,782
F-Ölçütü	0,702	0,767	0,767	0,765	0,757	0,778	0,778	0,748	0,777	**0,779**
κ statistic	0,374	0,509	0,507	0,502	0,486	0,531	0,530	0,465	0,526	**0,532**
MAE	0,298	0,284	0,309	0,311	0,301	0,299	**0,216**	0,240	0,218	0,307
RMSE	0,544	0,409	0,417	0,409	0,406	0,393	0,465	0,489	0,467	**0,390**
ROC Area	0,692	0,826	0,790	0,812	0,824	0,839	0,755	0,718	0,754	**0,844**
MCC	0,374	0,510	0,511	0,507	0,489	0,532	**0,538**	0,487	0,534	0,537



Şekil 4.13 Matematik veri seti adım 3.3 için başarılı algoritmaların ROC eğrileri

Şekil 4.13'te, matematik veri seti ve adım 3.3 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,692'nin üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,844).

Adım 3.4; Gain ratio öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Aile ile iletişim", "Eğitimde ilerlemek istediğiniz düzey", "Ev eğitim kaynakları", "Matematikte öğrencinin güveni" ve "Tahmin için matematk başarısı çok düşük" olarak belirlenmiştir. Seçilen 5 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.26'da ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.27'de verilmiştir. DP oranına göre DVM-PUK (0,793) ve LR (0,791) en iyi algoritmalar olarak belirlenmiştir. DVM-PUK için Precision (0,793), MAE (0,207) ve MCC (0,557); LR için F-Ölçütü (0,791), RMSE (0,384) ve ROC alanı (0,858) değerleri bu sonucu desteklemektedir. Ayrıca YSA, DVM-PUK ve LR algoritmalarına göre yakın sonuçlar vermiştir.

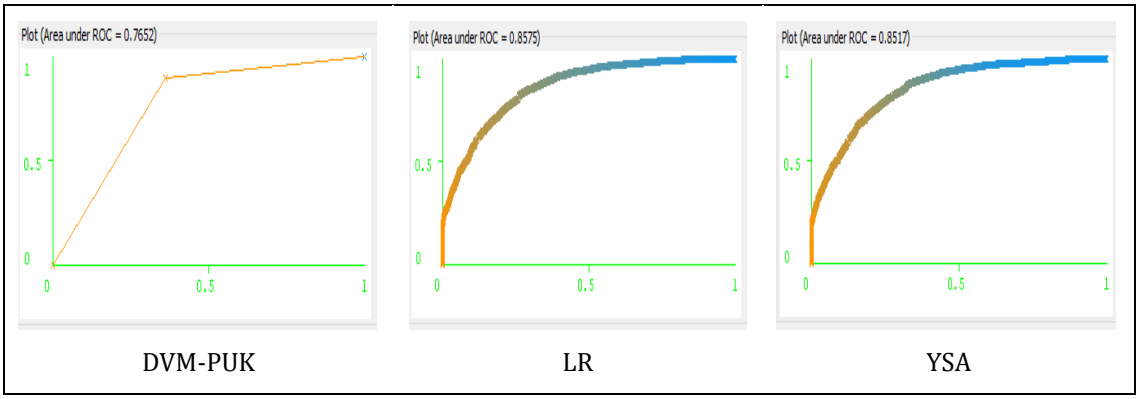
Tablo 4.26 Matematik veri seti adım 3.4 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	2335	450	2227	558	2385	400	2374	411	2311	474
b=1	697	1095	457	1335	590	1202	602	1190	632	1160
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	2423	362	2459	326	2637	148	2486	299	2419	366
b=1	589	1203	641	1151	903	889	649	1143	592	1200

Tablo 4.26'da verilen karşılaştırma matrisine göre DVM-PUK 4577 öğrenci verisinin 3629 (2486 + 1143) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 3430 (2335 + 1095), NB 3562 (2227 + 1335), KA-C4.5 3587 (2385 + 1202) KA-RepTree 3564 (2374 + 1190), KA-RF 3471 (2311 + 1160), YSA 3626 (2423 + 1203), DVM-POLY 3610 (2459 + 1151), DVM-RBF 3536 (2637 + 889) ve LR 3619 (2419 + 1200) tane doğru sınıflandırma yapmıştır.

Tablo 4.27 Matematik veri seti adım 3.4'e göre sınıflandırma sonuçları (5 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,749	0,778	0,784	0,779	0,758	0,792	0,789	0,770	**0,793**	0,791
YP O ranı	0,300	0,234	0,257	0,262	0,281	0,251	0,263	0,327	0,262	0,252
Precision	0,746	0,781	0,782	0,776	0,756	0,790	0,788	0,789	**0,793**	0,789
F-Ölçütü	0,745	0,779	0,781	0,776	0,756	0,789	0,784	0,753	0,788	**0,791**
κ statistic	0,461	0,539	0,537	0,527	0,485	**0,554**	0,542	0,479	0,550	0,551
MAE	0,291	0,290	0,300	0,301	0,290	0,295	0,211	0,230	**0,207**	0,297
RMSE	0,448	0,397	0,395	0,396	0,417	0,386	0,460	0,479	0,455	**0,384**
ROC Area	0,781	0,849	0,833	0,836	0,816	0,852	0,763	0,721	0,765	**0,858**
MCC	0,464	0,540	0,539	0,529	0,486	**0,557**	0,548	0,516	**0,557**	0,554



Şekil 4.14 Matematik veri seti adım 3.4 için başarılı algoritmaların ROC eğrileri

Şekil 4.14'te, matematik veri seti ve adım 3.4 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,721'in üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,858).

Adım 3.5; Info gain öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Eğitimde ilerlemek istediğiniz düzey", "Ev eğitim kaynakları", "Matematikte öğrencinin güveni" ve "Tahmin için matematk başarısı çok düşük" olarak belirlenmiştir. Bu 4 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.28'de ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.29'da verilmiştir. DP oranına göre DVM-POLY (0,791), LR (0,790) ve KA-C4.5 (0,789) en iyi algoritmalar olarak belirlenmiştir. DVM-POLY için Precision (0,791), MAE (0,209) ve MCC (0,554); LR için RMSE (0,386) ve ROC alanı (0,854); KA-C4.5 için ise YP oranı (0,248), F-Ölçütü (0,787), κ istatistiği (0,550) değerleri bu sonucu desteklemektedir.

Tablo 4.28 Matematik veri seti adım 3.5 için karşılaştırma matrisi

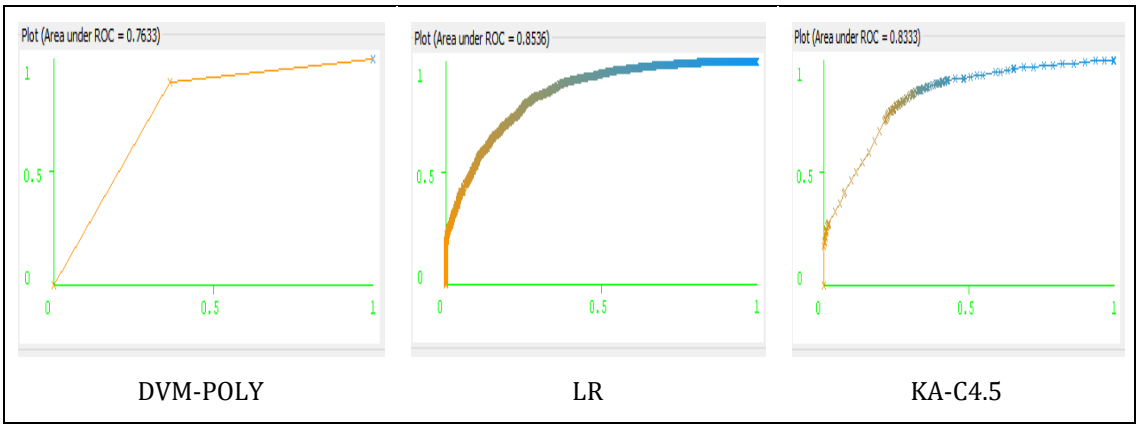
	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	2389	396	2370	415	2385	400	2391	394	2336	449
b=1	693	1099	568	1224	566	1226	597	1195	625	1167
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	2412	373	2483	302	2628	157	2480	305	2435	350

b=1	599	1193	654	1138	899	893	661	1131	610	1182

Tablo 4.28'de verilen karşılaştırma matrisine göre DVM-POLY 4577 öğrenci verisinin 3621 (2483 + 1138) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 3488 (2389 + 1099), NB 3594 (2370 + 1224), KA-C4.5 3611 (2385 + 1226) KA-RepTree 3586 (2391 + 1195), KA-RF 3503 (2336 + 1167), YSA 3605 (2412 + 1193), DVM-RBF 3521 (2628 + 893), DVM-PUK 3611 (2480 + 1131) ve LR 3617 (2435 + 1182) tane doğru sınıflandırma yapmıştır.

Tablo 4.29 Matematik veri seti adım 3.5'e göre sınıflandırma sonuçları (4 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,762	0,785	0,789	0,783	0,765	0,788	**0,791**	0,769	0,789	0,790
YP Oranı	0,291	0,251	**0,248**	0,258	0,275	0,256	0,265	0,327	0,267	0,256
Precision	0,759	0,783	0,787	0,781	0,763	0,786	**0,791**	0,786	0,789	0,789
F-Ölçütü	0,757	0,783	**0,787**	0,781	0,763	0,785	0,786	0,753	0,784	**0,787**
κ statistic	0,485	0,542	**0,550**	0,536	0,499	0,544	0,546	0,477	0,541	0,548
MAE	0,293	0,291	0,303	0,300	0,293	0,297	**0,209**	0,231	0,211	0,300
RMSE	0,431	0,391	0,393	0,393	0,415	0,388	0,457	0,480	0,459	**0,386**
ROC Area	0,796	0,848	0,833	0,839	0,815	0,848	0,763	0,721	0,761	**0,854**
MCC	0,490	0,544	0,551	0,539	0,500	0,547	**0,554**	0,513	0,549	0,552



Şekil 4.15 Matematik veri seti adım 3.5 için başarılı algoritmaların ROC eğrileri

Şekil 4.15'te, matematik veri seti ve adım 3.5 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,721'in üzerinde olduğunu söylemek mümkündür. Bu durum

algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,854).

Adım 3.6; ReliefF öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre en etkili öznitelikler "Paylaşımlı tablet bilgisayar", "Çalışma masası", "Hangi sıklıkta okula gidilmediği", "Matematikte yardımcı bilgiye erişim", "Son bir yılda alınan ek ders (Matematik)", "Kaç ay ek ders alındığı (Matematik)", "Ev eğitim kaynakları", "Öğrencinin matematik öğrenmeyi sevmesi", "Matematikte öğrencinin güveni" ve "Tahmin için matematk başarısı çok düşük" olarak belirlenmiştir. Bu 10 öznitelik kullanılarak elde edilen karşılaştırma tablosu değerleri Tablo 4.30'da ve algoritmaların sınıflandırma kriterlerine göre sınıflandırma başarıları Tablo 4.31'de verilmiştir. DP oranına göre DVM-POLY (0,799) ve LR (0,797) en iyi algoritmalar olarak belirlenmiştir. DVM-POLY için Precision (0,798), F-Ölçütü (0,796), κ istatistiği (0,568), MAE (0,201) ve MCC (0,572); LR için YP oranı (0,243), RMSE (0,378) ve ROC alanı (0,865) değerleri bu sonucu desteklemektedir.

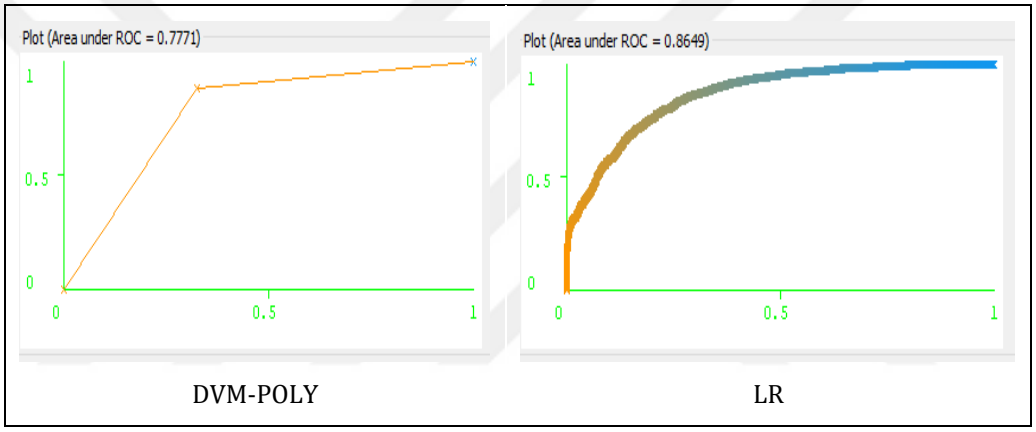
Tablo 4.30 Matematik veri seti adım 3.6 için karşılaştırma matrisi

	k-NN		**NB**		**KA-C4.5**		**KA-RepTree**		**KA-RF**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**	**a**	**b**
a=0	2130	655	2207	578	2372	413	2398	387	2376	409
b=1	692	1100	497	1295	595	1197	620	1172	609	1183
	YSA		**DVM-POLY**		**DVM-RBF**		**DVM-PUK**		**LR**	
	a	**b**	**a**	**b**	**a**	**a**	**b**	**a**	**b**	**a**
a=0	2383	402	2451	334	2492	293	2439	346	2424	361
b=1	571	1221	584	1208	700	1092	638	1154	567	1225

Tablo 4.30'da verilen karşılaştırma matrisine göre DVM-POLY 4577 öğrenci verisinin 3659 (2451 + 1208) tanesini doğru sınıflandırmıştır. Ayrıca k-NN 3230 (2130 + 1100), NB 3502 (2207 + 1295), KA-C4.5 3569 (2372 + 1197) KA-RepTree 3570 (2398 + 1172), KA-RF 3559 (2376 + 1183), YSA 3604 (2383 + 1221), DVM-RBF 3584 (2492 + 1092), DVM-PUK 3593 (2439 + 1154) ve LR 3649 (2424 + 1225) tane doğru sınıflandırma yapmıştır.

Tablo 4.31 Matematik veri seti adım 3.6'ya göre sınıflandırma sonuçları (10 değişken)

	Sınıflandırma Algoritmaları									
Kriterler	*k-NN*	*NB*	*KA-C4.5*	*KA-RepTree*	*KA-RF*	*YSA*	*DVM-POLY*	*DVM-RBF*	*DVM-PUK*	*LR*
DP Oranı	0,706	0,765	0,780	0,780	0,778	0,787	**0,799**	0,783	0,785	0,797
YP O ranı	0,327	0,250	0,260	0,265	0,264	0,250	0,245	0,279	0,265	**0,243**
Precision	0,705	0,767	0,778	0,778	0,775	0,785	**0,798**	0,784	0,784	0,796
F-Ölçütü	0,705	0,766	0,777	0,777	0,775	0,785	**0,796**	0,777	0,781	0,795
κ statistic	0,380	0,511	0,529	0,527	0,524	0,546	**0,568**	0,525	0,535	0,566
MAE	0,293	0,271	0,292	0,299	0,289	0,277	**0,201**	0,217	0,215	0,288
RMSE	0,540	0,411	0,416	0,403	0,394	0,388	0,448	0,466	0,464	**0,378**
ROC Area	0,694	0,840	0,797	0,824	0,845	0,854	0,777	0,752	0,760	**0,865**
MCC	0,380	0,511	0,531	0,530	0,526	0,548	**0,572**	0,536	0,540	0,568



Şekil 4.16 Matematik veri seti adım 3.6 için başarılı algoritmaların ROC eğrileri

Şekil 4.16'da, matematik veri seti ve adım 3.6 için sınıflandırmada başarılı algoritmaların ROC alanları gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,694'ün üzerinde olduğunu söylemek mümkündür. Bu durum algoritmaların iyi sınıflandırma yaptığı gösterir. LR ise en başarılı ROC değerini vermiştir (0,865).

Adım 1 ile adım 3 arasında elde edilen sonuçlar Tablo 4.32'de özetlenmiştir. Tablo 4.32 incelendiğinde tüm adımlarda LR algoritması sınıflandırma kriterlerine göre en başarılı sınıflandırma sonuçlarını vermiştir. Adım 3.4 dışında DVM-POLY, LR sınıflandırma algoritmasına en yakın sonuçları verdiği görülmüştür. Adım 1 için KA-RF; Adım 2, adım 3.1, adım 3.4 için DVM-PUK; Adım 3.2, adım 3.3, adım 3.4 için YSA ve adım 3.5 için KA-C4.5 algoritması, LR ve DVM-POLY algoritmalarına yakın sonuçlar vermiştir. Ayrıca Tablo 4.32'de adım 3.1 – adım 3.6 için etkili öznitelikler gösterilmiştir. "Ev eğitim kaynakları" (BSBGHER) ve "Matematikte öğrencinin

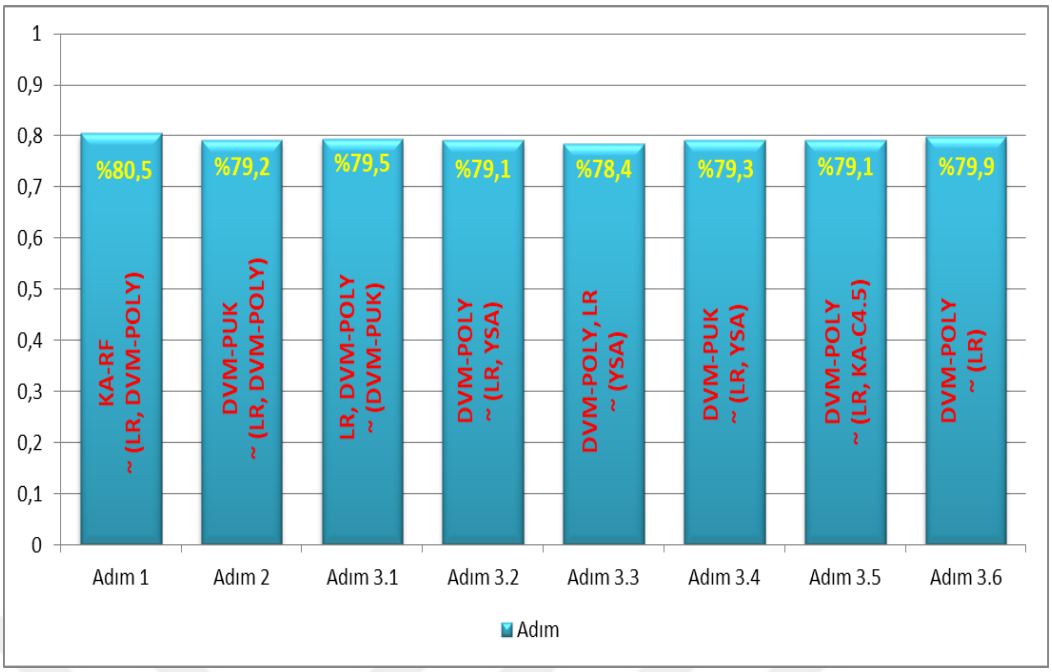
güveni" (BSBGSCM) adım 3.1 - adım 3.6 için belirlenen ortak özniteliklerdir. Öğrencilerin matematik başarısını sınıflandırmada bu özniteliklerin en etkili öznitelikler olduğu söylenebilir. ROC eğrileri incelendiğinde LR'nin en başarılı sonuçlar verdiği gözlemlenmiştir.

Tablo 4.32 Matematik veri seti adım 1 - adım 3 arasında genel sonuç tablosu

Adım	Algoritmalar	Değişkenler
Adım 1	KA-RF, LR, DVM-POLY	All features
Adım 2	DVM-PUK, LR, DVM-POLY	BSBGHER, BSBGSSB, BSBGSB, BSBGSLM, BSBGEML, BSBGSCM, BSBGSVM, BSDMLOWP, BSDMWKHW
Adım 3.1	LR, DVM-POLY, DVM-PUK	BSBG03, BSBG08, BSBG11, BSBG12, BSBG13B, BSBM38AA, BSBM39BA, BSBGHER, BSBGSCM, BSDMLOWP
Adım 3.2	DVM-POLY, LR, YSA	BSBG08, BSBM39AA, BSBM39BA, BSBGHER, BSBGSLM, BSBGSCM, BSDMLOWP
Adım 3.3	DVM-POLY, LR, YSA	BSBG05, BSBG08, BSBM39BA, BSBGHER, BSBGSLM, BSBGSCM
Adım 3.4	DVM-PUK, LR, YSA	BSBG03, BSBG08, BSBGHER, BSBGSCM, BSDMLOWP
Adım 3.5	DVM-POLY, LR, KA-C4.5	BSBG08, BSBGHER, BSBGSCM, BSDMLOWP
Adım 3.6	DVM-POLY, LR	BSBG06B, BSBG06C, BSBG11, BSBG14E, BSBM39AA,BSBM39BA, BSBGHER, BSBGSLM, BSBGSCM, BSDMLOWP

Sınıflandırma algoritmalarının performansları arasındaki değişkenliği araştırmak için sınıflandırma kriterlerinden yararlanılır. Bu çalışmada DP oranı değişkenlikleri göstermek için seçilmiştir.

Matematik veri setine yönelik olarak adım 1 – adım 3.6 arasında gerçekleştirilen tüm sınıflandırmalar için sınıflandırma kriterlerinden DP oranı baz alınarak belirlenen en iyi algoritmaları gösteren grafik Şekil 4.17'de verilmiştir. Şekil 4.17'ye göre en iyi algoritmaya çok yakın başarı gösteren algoritmalar "~(…)" şeklinde gösterilmiştir.



Şekil 4.17 Matematik veri seti için sınıflandırma sonuçları

Adım 1'de, Tablo 3.11'de verilen 35 özniteliğin tümü kullanılarak sınıflandırma algoritmalarının performansları belirlenmiştir. Bu adımda en yüksek DP oranına sahip algoritma KA-RF (% 80,5) olarak belirlenmiştir. Ayrıca DVM-POLY (% 80,4) ve LR (% 80,3) algoritmalarının DP oranları KA-RF'ye benzer sonuçlar göstermiştir. KA-RF algoritmasının adım 1'de elde edilen başarı oranı adım 1 ile adım 3 arasındaki en yüksek sınıflandırma başarı değerleri olarak bulunmuştur. Adım 1'e göre elde edilen sınıflandırma başarı değerleri diğer adımlar ile karşılaştırılmıştır. Adım 2 için 9 ölçek kullanılarak elde edilen sonuçlar incelendiğinde DP oranına göre DVM-PUK (% 79,2), LR (% 79,0) ve DVM-POLY (% 79,1) algoritmalarının en başarılı sınıflandırma performansına sahip oldukları belirlenmiştir. Adım 3.1'de 10 etkili öznitelik kullanılarak elde edilen sonuçlar göz önüne alındığında DP oranına göre LR (% 79,5), DVM-POLY (% 79,5) ve DVM-PUK (%79,4) algoritmaları sınıflandırmada en başarılı sonuçları vermiştir. Adım 3.2 için 7 etkili öznitelik kullanılarak elde edilen sonuçlar dikkate alındığında DP oranına göre DVM-POLY (% 79,1), LR (% 79,0) ve YSA (% 78,9) sınıflandırmada en başarılı algoritmalar olarak bulunmuştur. Adım 3.3'de 6 etkili öznitelik kullanılarak elde edilen sonuçlar incelendiğinde DP oranına göre DVM-POLY (% 78,4), LR (% 78,4) ve YSA (% 78,0) sınıflandırmada en başarılı algoritmalar olarak seçilmiştir. Adım 3.4 için 5 öznitelik kullanılarak elde edilen sonuçlar göz önüne alındığında DP

oranına göre DVM-PUK (% 79,3), LR (% 79,1) ve YSA (% 79,2) algoritmaları en başarılı sınıflandırma performansına sahip oldukları belirlenmiştir. Adım 3.5'te 4 öznitelik kullanılarak elde edilen sonuçlar dikkate alındığında DP oranına göre DVM-POLY (% 79,1), LR (% 79,0) ve KA-C4.5 (% 78,9) sınıflandırmada en başarılı algoritmaları olarak bulunmuştur. Adım 3.6'da 10 etkili öznitelik kullanılarak elde edilen sonuçlar incelendiğinde DP oranına göre DVM-POLY (% 79,9) ve LR (% 79,7) sınıflandırmada en başarılı algoritmalar olarak seçilmiştir. Özellikle adım 3.6'da elde edilen algoritmaların sınıflandırma performanslarının, tüm adımlar incelendiğinde adım 1'e çok yakın sonuçlar vermesi öznitelik seçiminin ne denli önemli olduğunu vurgulamaktadır.

5

Sonuç ve Öneriler

5.1 Değerlendirme ve Tartışma

Eğitimsel karar alma süreçlerinde öğrencilerin başarısını etkileyen öznitelikleri belirlemek çok önemlidir. Öğrencilerin başarısını etkileyen öznitelikler bilindiğinde, öğrencinin başarısının bu öznitelikler için önlemler alınarak veya mevcut koşulları değiştirerek arttırılacağı varsayılabilir. Bu fikre dayanarak, bu çalışmanın giriş bölümünde de belirtildiği gibi çalışma iki araştırma sorusuna odaklanmaktadır: ilk olarak, öğrencilerin fen bilimleri ve matematik başarısını sınıflandırmak için hangi algoritmaların uygun olduğunu ve ikinci olarak ise öznitelik seçim algoritmaları yardımıyla sınıflandırmadaki en önemli özniteliklerin hangileri olduğunu ortaya çıkartmaktır. Bu araştırma soruları TIMSS-2015 8. sınıf Türk öğrencilerinin fen bilimleri ve matematik veri setleri ele alınmıştır.

Fen bilimleri ve matematik veri seti üzerine yapılan analizlere göre birinci araştırma sorusu için LR ve DVM-POLY en uygun algoritmalar olarak bulunmuştur. Literatürde yapılan incelemelere göre mevcut pek çok çalışmada, LR ve DVM-POLY'nin öğrencilerin fen bilimleri ve matematik başarısını sınıflandırmada en uygun olduğu belirlenmiştir. LR, öğrencilerin akademik performansını belirlemek için çeşitli veri madenciliği yöntemleri arasında en iyi sınıflandırma algoritması olarak birçok çalışmada ön plana çıkmıştır [36, 40]. Yoo çalışmasında, TIMSS-2011 Kore 4. sınıf öğrencilerinin matematik puanları hedef kitle olarak almıştır. Veri setine tahmin metodu olarak LR uygulamış ve başarılı sonuçlar elde etmiştir [43]. Delen, tarafından yapılan bir başka araştırma, öğrenciler arasında yıpranmayı tahmin etmek için analitik modeller geliştirmek için YSA, DVM, C5 karar ağacı algoritması ve LR kullanmış; DVM'nin bu alanda en iyi sonuçlar verdiğini ortaya koymuştur [33]. Gorostiaga ve Rojo-Álvarez, LR, Fisher ayırıcı analizi ve DVM (hem doğrusal hem de doğrusal olmayan) algoritmaları yardımıyla, İspanyol öğrencilerin akademik puanlarının PISA-2009 matematik sonuçlarını

sınıflandırmak için kullanmışlardır. DVM'nin diğer algoritmalardan daha iyi performans gösterdiğini belirtmişlerdir [34]. Ayrıca fen bilimleri seti için DVM-PUK algoritmasının; matematik veri seti için ise DVM-PUK ve YSA algoritmalarının DVM-POLY ve LR algoritmalarına yakın performanslar gösterdiği belirlenmiştir.

DVM sınıflandırma performanslarını belirlemede çekirdek seçimi önemli bir süreçtir. Farklı çekirdek fonksiyonlarının seçimi sonucunda farklı sınıflandırma performansları elde edilir [81]. Buna bağlı olarak, fen bilimleri ve matematik veri seti için DVM-POLY ve DVM-PUK algoritmalarının çoğu adımda en iyi sınıflandırma performanslarına sahip olduğu görülmüştür. DVM-RBF algoritmasının ise hiçbir adımda en iyi sınıflandırma algoritması olarak bulunmadığı belirlenmiştir. Bunun yanında fen bilimleri veri seti için tüm adımlardaki sınıflandırma performansları incelendiğinde, k-NN, NB, KA-RepTree ve KA-C4.5 algoritmaları hiçbir adımda en iyi performans algoritması olarak bulunamamıştır. Ayrıca, üç KA algoritmasının sınıflandırma performansları tüm adımlarda incelendiğinde, sadece KA-RF algoritması her iki veri seti içinde adım 1'de başarılı sınıflandırma performansı göstermiştir. Matematik veri seti için k-NN, NB ve KA-RepTree algoritmaları hiçbir adımda en iyi performansı göstermemiştir. Üç KA algoritmasının sınıflandırma performansları tüm adımlarda incelendiğinde, KA-RF algoritması adım 1'de ve KA-C4.5 algoritması adım 3.5'te başarılı sınıflandırma performansı göstermiştir.

Bu çalışmanın ikinci araştırma sorusu için, öğrencilerin fen bilimleri ve matematikteki başarılarıyla ilgili en önemli öznitelikleri belirlemek için farklı öznitelik seçim yöntemleri kullanılmaktadır. Farklı öznitelik seçim yöntemlerini kullanmanın amacı, sınıflandırma başarısından ödün vermeden daha az öznitelik kullanmaktır. Sınıflandırma algoritmalarında kullanılan öznitelik sayısının olabildiğince azaltılması ve sınıflandırma sonuçlarında bunu yaparken önemli bir düşüş olmaması son derece önemlidir. Fen bilimleri veri seti için öznitelik seçim algoritmaları kullanılarak elde edilen sonuçlara göre "Paylaşımlı tablet bilgisayar", "Eğitimde ilerlemek istediğiniz düzey", "Kaç ay ek ders alındığı (Fen bilimleri)" ve "Fen bilimlerinde öğrencinin güveni" öznitelikleri ortak öznitelikler olarak belirlenmiştir. Matematik veri seti için öznitelik seçim algoritmaları kullanılarak elde edilen sonuçlara göre "Ev eğitim kaynakları" ve "Matematikte öğrencinin güveni" öznitelikleri ortak öznitelikler olarak belirlenmiştir. Bu öznitelikler, önceki

pek çok çalışmada da önemli öznitelikler olarak bulunmuştur. “Öğrencinin güveni” özniteliğinin çeşitli çalışmalarda öğrencilerin fen bilimleri başarısı için önemli bir unsur olduğu kanıtlanmıştır [35, 40-42]. PISA-2006 Türkiye için öğrencilerin fen bilimleri verilerini inceleyen Anıl tarafından “Bilgisayar ortamı” bir başka önemli öznitelik olarak bulunmuştur [52]. Okul dışında, fen bilimleri derslerine harcanan fazladan zaman, PISA 2006'daki Türkiye fen bilimleri çalışmasında önemli bir unsur olarak görülmüştür [53]. Ayrıca, Ogura'nın araştırmasında, okul dışında harcanan fazladan zaman, öğrencilerin başarı düzeyleri üzerinde önemli bir öznitelik olarak belirlenmiştir [51]. Bu çalışmada matematik veri seti için önemli bir öznitelik olarak bulunan “Ev eğitim kaynakları”, Topçu, Erbilgin ve Arıkan çalışmasında, öğrenci başarısında önemli bir öznitelik olarak bulmuş ve eğitim kaynaklarına kolayca ulaşabilen öğrencilerin başarılı olduklarına dikkat çekmişlerdir [39].

Çalışmanın uygulama kısmında Türkiye’deki 8. sınıf öğrencilerinin fen bilimleri ve matematik başarılarını sınıflandırmda kullanılabilecek en etkili sınıflandırma algoritmalarının LR ve DVM-POLY olduğu ortaya çıkmıştır. Bu noktada fen bilimleri başarılarını sınıflandırma problemi ele alındığında LR’nin DP oranının 0,738 olduğu ve bu değerin tüm değişkenleri kullanarak elde edildiği görülmektedir. Bunun anlamı, eğer bağımsız değikenlerin yani özniteliklerin değerlerinin bilinmesi durumunda herhangi bir öğrencinin başarılı olup olamayacağının belirlenmesi %73,8 oranında doğru bir şekilde yapılabilecektir. Ancak daha az sayıda öznitelik kullanılarak da sınıflandırma başarısında tatminkar sonuçlar elde edilmiştir. Fen bilimleri başarısı için 13 öznitelik ile DP oranı 0,734 olan DVM-POLY algoritması en iyi sınıflandırıcı olmakla beraber, sadece burada elde edilen 13 özniteliğin değerlerinin bilinmesi durumunda herhangi bir öğrencinin başarılı olup olamayacağının belirlenmesi %73,4 oranında doğru bir şekilde yapılabilecektir. Benzer şekilde matematik başarılarını sınıflandırma problemi dikkate alındığında KA-RF’nin DP oranının 0,805 olduğu ve bu değerin tüm değişkenleri kullanarak elde edildiği görülmektedir. Bunun anlamı, eğer özniteliklerin değerlerinin bilinmesi durumunda herhangi bir öğrencinin başarılı olup olamayacağının belirlenmesi %80,5 oranında doğru bir şekilde yapılabilecektir. Matematik başarısı için 10 öznitelik ile DP oranı 0,799 olan DVM-POLY algoritması en iyi sınıflandırıcı olmakla beraber, sadece burada elde edilen

10 özniteliğin değerlerinin bilinmesi durumunda herhangi bir öğrencinin başarılı olup olamayacağının belirlenmesi %79,9 oranında doğru bir şekilde yapılabilecektir.

Bu çalışmada literatürde yer alan pek çok çalışmadan farklı bir istatistiksel bakış açısı ile eğitim alanında katkı sağlayabilecek sonuçlar elde edilmiştir. TIMSS-2015 fen bilimleri ve matematik alanlarındaki veri setleri incelemiştir. Bu veri seti, tezin savunma tarihi itibarı ile yayımlanan en son TIMSS veri setidir. TIMSS-2015 veri setinde Türkiye çerçevesinde sekizinci sınıf öğrencilerinin başarılarının sınıflandırılmasında çok sayıda makine öğrenmesi ve öznitelik seçimi algoritması kullanılması bu çalışma açısından oldukça önemlidir. Ayrıca, TIMSS literatürü ile ilgili daha önceki araştırmalar incelendiğinde, kümeleme, tahmin etme ve regresyonda, standart istatistiksel tekniklerin (örneğin faktör analizi, ANOVA) kullanıldığı görülmektedir. Belirtilen bu teknikler kesin varsayımlarından dolayı bir miktar önem kaybetmektedirler. Bundan dolayı makine öğrenmesi algoritmaları tercih edilen bir teknik haline gelmektedir. Tez çalışmasında genellikle makine öğrenmesi literatüründe en sık tercih edilen algoritmalar kullanılmıştır. Bu çalışmaya özgü diğer bir özellik, öğrencilerin başarısının sınıflandırılması incelenirken veri azaltmaya verilen önemdir. Sınıflandırma sırasında sadece en alakalı özniteliklerin dikkate alınması gerektiğini ve tüm özniteliklerin kullanılmasına gerek olmadığını kanıtlamaktadır. Bu katkılara ek olarak, eğitimciler ve eğitim politikası belirleyiciler bu çalışmada çıkarılan en önemli öznitelikleri kullanabilirler. Yararlı eğitim stratejileri geliştirmek ve böylece öğrencilerin akademik başarısını arttırmak için, önemli öznitelikleri bilmek hayati bir rol oynamaktadır. Ayrıca, bu araştırmanın bulguları daha önceki araştırmalarla desteklenmiştir.

Bu çalışmanın bazı sınırlamaları mevcuttur. Öncelikle, TIMSS çalışması fen bilimleri ve matematik gibi birçok alanın başarısını değerlendirir, ancak bu çalışmada sonuçlar sadece TIMSS-2015 fen bilimleri ve matematik veri setine dayanmaktadır. Ayrıca, bu çalışmanın bulguları, IEA tarafından geliştirilen ve kişisel olarak raporlanan anketlerin sonuçlarına bağlıdır. Diğer bir sınırlama, öğrencilerin fen bilimleri ve matematik başarısının sınıflandırılmasını etkileyen öznitelikler, sadece sekizinci sınıf öğrencileri için çıkarılmıştır.

Gelecekte TIMSS veri setlerinde birden fazla ülke kullanılarak ve ülkelerin başarı durumlarının incelenmesi üzerine makine öğrenmesi yöntemlerinin uygulanması ile kapsamlı çalışmalar yapılabilir. Bu tez çalışması gelecekte yapılacak bu tür çalışmalara ışık tutacaktır.

Kaynakça

[1] S. Sinharay, "An NCME instructional module on data mining methods for classification and regression," Educational Measurement: Issues and Practice, cilt 35, sayı 3, s. 38-54, 2016.

[2] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, ... ve M. Kudlur, "Tensorflow: A system for large-scale machine learning," in 12th USENIX Symposium on Operating Systems Design and Implementation OSDI-16, 2016, s. 265-283.

[3] H. Bunke, "Graph-based tools for data mining and machine learning," in International Workshop on Machine Learning and Data Mining in Pattern Recognition, 2016, s. 7-19.

[4] S. Y. Rhee, J. Park ve A. Inoue, "Soft Computing in Machine Learning," 2014.

[5] I. V. Mullis, M. O. Martin, P. Foy ve A. Arora, "TIMSS 2011 international results in mathematics," 2012.

[6] J. Han, M. Kamber ve J. Pei, "Data mining: Concept and techniques (3rd ed.)," 2012.

[7] T. M. Mitchell, (1997), "Machine Learning," Burr Ridge, IL: McGraw Hill, cilt 45, sayı 37, s. 870-877, 1997.

[8] A. McCallum ve K. Nigam, "A Comparison of Event Models For Naive Text Classification," AAAI-98 Workshop On Learning For Text Categorization, 1998, s. 41-48.

[9] T. G. Dietterich, "Ensemble methods in machine learning," in International workshop on multiple classifier systems, 2000, s. 1-15.

[10] B. Pang, L. Lee ve S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing, cilt 10, 2002, s. 79-86.

[11] T. Kavzaoğlu ve İ. Çölkesen, "Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi," Harita Dergisi, sayı 144, 2010, s. 73-82.

[12] M. Meral ve B. Diri, "Twitter Üzerinde Duygu Analizi," IEEE 22nd Signal Processing and Communicatios Applications Conference (SIU 2014), 2014, s. 690-693.

[13] A. L. Blum ve P. Langley, "Selection of relevant features and examples in machine learning," Artificial intelligence, cilt 97, sayı 1-2, s. 245-271. 1997.

[14] F. Sebastiani, "Machine learning in automated text Categorization," ACM computing surveys, cilt 34, sayı 1, s. 1-47, 2002.

[15] S. H. Liao, "Knowledge management technologies and applications—literature review from 1995 to 2002," Expert systems with applications, cilt 25, sayı 2, s. 155-164, 2003.

[16]J. Wen, S. Li, Z. Lin, Y. Hu ve C. Huang, “Systematic literature review of machine learning based software development effort estimation models,” Information and Software Technology, cilt 54, sayı 1, s. 41-59, 2012.

[17]A. Onan ve S. Korukoğlu, “Makine Öğrenmesi yöntemlerinin görüş madenciliğinde kullanılması üzerine bir literatür araştırması,” Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, cilt 22, sayı 2, s. 111-122, 2016.

[18]C. Romero ve S. Ventura, “Educational data mining: a review of the state of the art,” The Institute of Electrical and Electronics Engineers (IEEE) Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), cilt 40, sayı 6, s. 601-618, 2010.

[19]C. Romero ve S. Ventura, S. “Educational data mining: A survey from 1995 to 2005,” Expert systems with applications, cilt 33, sayı 1, s. 135-146, 2007.

[20]R. S. Baker, ve K. Yacef, “The state of educational data mining in 2009: A review and future visions,” Journal of Educational Data Mining, cilt 1, sayı 1, s. 3-17, 2009.

[21]S. K. Mohamad ve Z. Tasir, “Educational data mining: A review,” Procedia-Social and Behavioral Sciences, cilt 97, s. 320-324, 2013.

[22]G. Siemens ve R. S. Baker, “Learning analytics and educational data mining: towards communication and collaboration,” in: Proceedings of the 2nd international conference on learning analytics and knowledge, 2012, s. 252-254.

[23]A. Peña-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works,” Expert systems with applications, cilt 41, sayı 4, s. 1432-1462, 2014.

[24]S. Kotsiantis, C. Pierrakeas ve P. Pintelas, “Predicting Students' performance In Distance Learning Using Machine Learning Techniques,” Applied Artificial Intelligence, cilt 18, sayı 5, s. 411-426, 2004.

[25]A. M. Shahiri, W. Husain ve N. A. Rashid, “A review on predicting student's performance using data mining techniques,” Procedia Computer Science, cilt 72, s. 414-422, 2015.

[26]M. Ramaswami ve R. Bhaskaran, “A CHAID based performance prediction model in educational data mining,” International Journal of Computer Science Issues, cilt 7, sayı 1, s. 10-18, 2010.

[27]B. K. Baradwaj ve S. Pal, “Mining educational data to analyze students' performance,” International Journal of Advanced Computer Science and Applications, cilt 2, sayı 6, s. 63-69, 2011.

[28]J. Rajni ve D. B. Malaya, “Predictive analytics in a higher education context,” IT Professional, cilt 17, sayı 4, s. 24-33, 2015.

[29]F. Martínez Abad ve A. A. Chaparro Caso López, “Data-mining techniques in detecting factors linked to academic achievement,” School Effectiveness and School Improvement, cilt 28, sayı 1, s. 39-55, 2017.

[30] P. Cortez ve A. M. G. Silva, “Using data mining to predict secondary school student performance,” in: A. Brito & J. Teixeira (Eds.), Proceedings of 5th Annual Future Business Technology Conference, 2008, s. 5–12.

[31] T. Horáková, M. Houška ve L. Dömeová, “Classification of the educational texts styles with the methods of artificial intelligence,” Journal of Baltic Science Education, cilt 16, sayı 3, s. 324-336, 2017.

[32] E. Osmanbegović ve M. Suljić, “Data mining approach for predicting student performance,” Economic Review, cilt 10, sayı 1, s. 3-12, 2012.

[33] D. Delen, “A comparative analysis of machine learning techniques for student retention management,” Decision Support Systems, cilt 49, sayı 4, s. 498-506, 2010.

[34] A. Gorostiaga ve J. L. Rojo-Álvarez, “On the use of conventional and statistical-learning techniques for the analysis of PISA results in Spain,” Neurocomputing, cilt 171, s. 625-637, 2016.

[35] O. E. Askin ve F. Gokalp, “Comparing the predictive and classification performances of logistic regression and neural networks: a case study on timss 2011,” Procedia-Social and Behavioral Sciences, cilt 106, s. 667-676, 2013.

[36] J. B. Schreiber, “Scoring Above the International Average: A Logistic Regression Model of the TIMSS Advanced Mathematics Exam,” Multiple Linear Regression Viewpoints, cilt 28, sayı 1, s. 22-30, 2002.

[37] S. A. Kiray, B. Gok ve A. S. Bozkir, “Identifying the factors affecting science and mathematics achievement using data mining methods,” Journal of Education in Science, Environment and Health, cilt 1, sayı 1, s. 28-48, 2015.

[38] P. F. Pai, C. T. Chen, Y. M. Hung, W. Z. Hung ve Y. C. Chang, “A group decision classifier with particle swarm optimization and decision tree for analyzing achievements in mathematics and science,” Neural Computing and Applications, cilt 25, sayı 7-8, s. 2011-2023, 2014.

[39] M. S. Topçu, E. Erbilgin ve S. Arıkan, “Factors Predicting Turkish and Korean Students' Science and Mathematics Achievement in TIMSS 2011,” Eurasia Journal of Mathematics, Science & Technology Education, cilt 12, sayı 7, s. 1711-1737, 2016.

[40] S. Kılıç-Depren, Ö. E. Askin ve E. Öz, “Identifying the Classification Performances of Educational Data Mining Methods: A Case Study for TIMSS,” Educational Sciences: Theory & Practice, cilt 17, sayı 5, s. 1605-1623, 2017.

[41] H. Hammouri, “Attitudinal and motivational variables related to mathematics achievement in Jordan: Findings from the Third International Mathematics and Science Study (TIMSS),” Educational Research, cilt 46, sayı 3, s. 241-257, 2010.

[42] S. Liu ve L. Meng, “Re-examining factor structure of the attitudinal items from TIMSS 2003 in cross-cultural study of mathematics self-concept,” Educational Psychology, cilt 30, sayı 6, s. 699-712, 2010.

[43]J. E. Yoo, "TIMSS 2011 Student and Teacher Predictors for Mathematics Achievement Explored and Identified via Elastic Net. Frontiers in psychology," cilt 9, makale sayısı 317, 2018.

[44]M. A. Hall ve L. A. Smith, "Practical feature subset selection for machine learning," 1998.

[45]F. Jiang, W. Meng ve X. Meng, X. "Selectivity estimation for exclusive query translation in deep web data integration," in: International Conference on Database Systems for Advanced Applications, 2009, s. 595-600.

[46]J. Novakovic, "The impact of feature selection on the accuracy of naïve bayes Classifier," in 18th Telecommunications forum TELFOR, cilt 2, 2010, s. 1113-1116.

[47]R. P. Priyadarsini, M. L. Valarmathi ve S. Sivakumari, "Gain ratio based feature selection method for privacy preservation," ICTACT Journal on soft computing, cilt 1, sayı 4, s. 201-205, 2011.

[48]V. Bolón-Canedo, N. Sánchez-Marono, A. Alonso-Betanzos, J. M. Benítez ve F. Herrera, "A review of microarray datasets and applied feature selection methods," Information Sciences, cilt 282, s. 111-135, 2014.

[49]S. Gnanambal, M. Thangaraj, V.T. Meenatchi ve V. Gayathri, "Classification Algorithms with Attribute Selection: an evaluation study using WEKA," International Journal of Advanced Networking and Applications, cilt 9, sayı 6, s. 3640-3644, 2018.

[50]H. Takci, "Improvement of heart attack prediction by the feature selection methods," Turkish Journal of Electrical Engineering & Computer Sciences, cilt 26, sayı 1, s. 1-10, 2018.

[51]Y. Ogura, "Background to Japanese student achievement in science and mathematics," in S. J. Howie ve T. Plomp (Eds.), "Content of learning mathematics and science," Lessons learned from TIMSS, s. 313–331, 2006.

[52]D. Anil, "Factors effecting science achievement of science students in programme for international students' achievement (PISA) in Turkey," Education and Science, cilt 34, sayı 152, s. 87-100, 2009.

[53]Y. Ozer ve D. Anil, "Öğrencilerin Fen ve Matematik Başarılarını Etkileyen Faktörlerin Yapısal Eşitlik Modeli İle İncelenmesi," Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, cilt 41, s. 313-324, 2011.

[54]S. Özekes, "Veri Madenciliği Modelleri Ve Uygulama Alanları," İstanbul Ticaret Üniversitesi Dergisi, sayı 3, s. 65-82, 2003.

[55]H. Nizam ve S. S. Akın, "Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması," XIX. Türkiye'de İnternet Konferansı, 2014, s. 129-136.

[56]S. Jiang, G. Pang, M. Wu ve L. Kuang, "An improved K-nearest-neighbor algorithm for text categorization," Expert Systems with Applications, cilt 39, sayı 1, s. 1503-1509, 2012.

[57]B. Li, S. Yu ve Q. Lu, "An improved k-nearest neighbor algorithm for text categorization," Proceedings of the 20th International Conference on

Computer Processing of Oriental Languages, 2003, https://arxiv.org/ftp/cs/papers/0306/0306099.pdf.

[58] B. Daş ve İ. Türkoğlu, "DNA Dizilimlerindeki Nükleotit Çiftlerinin Frekans Değerlerine Göre Farklı Sınıflandırma Yöntemleri ile Karşılaştırılması," Tıp Teknolojileri Ulusal Kongresi (TIPTEKNO'2014), 2014, s. 191-194.

[59] P. Horton ve K. Nakai, "Better prediction of protein cellular localization sites with the k nearest neighbors classifier," in: ISMB-97 Proceedings, 1997, s. 147–152.

[60] E. Öz, S. Kurt, M. Asyalı, H. Kaya, ve Y. Yücel, "Feature based quality assessment of DNA sequencing chromatograms," Applied Soft Computing, cilt 41, s. 420–427, 2016.

[61] M. E. Balaban ve E. Kartal, "Veri Madenciliği Ve Makine Öğrenmesi," 2015.

[62] H. Zhang, "The optimality of naive Bayes," in: FLAIRS Conference, 2004, s. 562-567.

[63] G. H. John ve P. Langley, "Estimating continuous distributions in Bayesian classifiers," in: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, 1995, s. 338-345.

[64] M. F. Amasyalı, B. Diri ve F. Türkoğlu, "Farklı Özellik Vektörleri İle Türkçe Dökümanların Yazarlarının Belirlenmesi," Türkiye Yapay Sinir Ağları Sempozyumu (TAINN), 2006.

[65] Z. Şen, "Yapay Sinir Ağları İlkeleri," 2004.

[66] E. Yakut, B. Elmas ve S. Yavuz, "Yapay Sinir Ağları ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini," Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, cilt 19, sayı 1, s.139-157, 2014.

[67] K. K. Çevik ve E. Dandıl, "Yapay sinir ağları için net platformunda görsel bir eğitim yazılımının geliştirilmesi," Bilişim Teknolojileri Dergisi, cilt 5, sayı 1, s. 19-28, 2012.

[68] L. Rokach ve O. Maimon, "Decision Trees," Data Mining and Knowledge Discovery Handbook, Springer, s. 165-192, 2005.

[69] J. Han ve M. Kamber, "Data Mining Concepts and Techniques," 2001.

[70] J. R. Quinlan, "C4.5: programs for machine learning," 2014.

[71] C. S. Dangare ve S. S. Apte, "Improved study of heart disease prediction system using data mining classification Techniques," International Journal of Computer Applications, cilt 47, sayı 10, s. 44–48, 2012.

[72] V. Sugumaran, V. Muralidharan ve K. I. Ramachandran, "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing," Mechanical Systems and Signal Processing, cilt 21, sayı 2, s. 930–942, 2007.

[73] X. W. Chen ve M. Liu, "Prediction of protein–protein interactions using random decision forest framework," Bioinformatics, cilt 21, sayı 24, s. 4394-4400, 2005.

[74] L. Breiman, "Random forests," Machine learning, cilt 45, sayı 1, s. 5-32, 2001.

[75]S. Kalmegh, “Analysis of WEKA data mining algorithm REPTree, Simple CART and RandomTree for classification of Indian news,” International Journal of Innovative Science, Engineering & Technology, cilt 2, sayı 2, s. 438-446, 2015.

[76]D. B. Srinivasan ve P. Mekala, “Mining Social Networking Data for Classification Using REPTree,” International Journal of Advance Research in Computer Science and Management Studies, cilt 2, sayı 10, s. 155-160, 2014.

[77]C. Cortes ve V. Vapnik, “Support-vector networks,” Machine learning, cilt 20, sayı 3, s. 273-297, 1995.

[78]M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt ve B. Scholkopf, “Support vector machines,” IEEE Intelligent Systems and their applications, cilt 13, sayı 4, s. 18-28, 1998.

[79]S. Haykin, “Neural Networks: A comprehensive Foundation,” 1999.

[80]E. Alpaydın, “Introduction to machine learning,” 2004.

[81]J. Shawe-Taylor, P. L. Bartlett, R. C. Williamson ve M. Anthony, “Structural risk minimization over data-dependent hierarchies,” The Institute of Electrical and Electronics Engineers (IEEE) Transactions on Information Theory, cilt 44, sayı 5, s. 1926-1940, 1998.

[82]P. K. Varshney ve M. K. Arora, “Advanced image processing techniques for remotely sensed hyperspectral data,” 2004.

[83]C. Campbell ve Y. Ying, “Learning with support vector machines,” Synthesis lectures on artificial intelligence and machine learning, cilt 5, sayı 1, s. 1-95, 2011.

[84]E. Öz ve H. Kaya, “Support vector machines for quality control of DNA sequencing,” Journal of Inequalities and Applications, cilt 85, sayı 1, 2013.

[85]V. Vapnik, “Statistical learning Theory,” Wiley, s.156-160, 1998.

[86]C. Moreira, “Learning to rank academic experts,” Doktora tezi, Technical University of Lisbon, 2011.

[87]D. W. Hosmer ve S. Lemeshow, “Applied Logistic Regression (2nd ed.),” 2000.

[88]Ö. Çokluk, “Lojistik regresyon Analizi: Kavram ve Uygulama,” Kuram ve Uygulamada Eğitim Bilimleri, cilt 10, sayı 3, s. 1357-1407, 2010.

[89]D. Atasoy, “Lojistik regresyon analizinin incelenmesi ve bir uygulaması,” Yüksek lisans tezi, Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Sivas, 2001.

[90]K. Sümbüloğlu, “Lojistik Regresyon Analizi,” 2015, http://78.189.53.61/-/bs/ess/k_sumbuloğlu.pdf.

[91]A. K. Özdemir, S. Tolun ve E. Demirci, “Endeks Getirisi Yönünün İkili Sınıflandırma Yöntemiyle Tahmin Edilmesi: İMKB-100 Endeksi Örneği,” Niğde Üniversitesi İİBF Dergisi, cilt 4, sayı 2, s. 45-59, 2011.

[92]A. S. Albayrak, “Uygulamalı Çok Değişkenli İstatistik Teknikler,” 2006.

[93]P. M. Narendra ve K. Fukunaga, “A branch and bound algorithm for feature subset selection,” IEEE Transactions on computers, cilt 26, sayı 9, s. 917-922, 1977.

[94]H. Gündüz, "Borsa İstanbul (BIST) 100 Endeksi Yönünün Ekonomi Haberleri İle Tahmin Edilmesi," Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 2013.

[95]A. Gümüşçü, İ. B. Aydilek ve R. Taşaltın, "Mikro-dizilim Veri Sınıflandırmasında Öznitelik Seçme Algoritmalarının Karşılaştırılması," Harran Üniversitesi Mühendislik Dergisi, cilt 1, sayı 1, s. 1-7, 2016.

[96]J. H. Gennari, P. Langley ve D. Fisher, "Models of incremental concept formation," Artificial intelligence, cilt 40, sayı 1-3, s. 11-61, 1989.

[97]M. A. Hall, "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning," 2000.

[98]Z. Muda, W. Yassin, M. N. Sulaiman ve N. I. Udzir, "Intrusion detection based on k-means clustering and OneR classification,"in: 7th International Conference on Information Assurance and Security (IAS), 2011, s. 192-197.

[99]D. Kabakchieva, "Predicting student performance by using data mining methods for classification," Cybernetics and information technologies, cilt 13, sayı 1, s. 61-72, 2013.

[100]A. G. Karegowda, A. S. Manjunath ve M. A. Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," International Journal of Information Technology and Knowledge Management, cilt 2, sayı 2, s. 271-277, 2010.

[101]T. M. Cover ve J. A. Thomas, "Elements of information theory," 2012.

[102]M. S. Aktas ve O. Kalıpsız, "Veri Madenciliğinde Özellik Seçim Tekniklerinin Bankacılık Verisine Uygulanması Üzerine Araştırma ve Karşılaştırmalı Uygulama," 9. Ulusal Yazılım Mühendisliği Sempozyumu (UYMS 2015), 2015, s. 9-11.

[103]A. K. Uysal, "New approaches to enhancing the performance of text classification," Doktora Tezi, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, 2013.

[104]I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF," in European conference on machine learning, 1997, s. 171-182.

[105]A. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," Pattern Recognition, cilt 30, sayı 7, s. 1145–1159, 1997.

[106]E. Filiz ve E. Oz, E. "Finding the Best Algorithms and Effective Factors in Classification of Turkish Science Student Success," Journal Of Baltic Science Education, cilt 18, sayı 2, s. 239-253, 2019.

[107]A. Donner ve N. Klar, "The statistical analysis of kappa statistics in multiple samples," Journal of clinical epidemiology, cilt 49, sayı 9, s. 1053-1058, 1996.

[108]C. J. Willmott ve K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Climate research, cilt 30, sayı 1, s. 79-82, 2005.

[109]E. Filiz ve E. Öz, “Classification of BIST-100 Index'Changes via Machine Learning Methods,” Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi, cilt 39, sayı 1, s. 117-129, 2017.

[110]S. Erpolat ve E. Öz, “Kanser Verilerinin Sınıflandırılmasında Yapay Sinir Ağları ile Destek Vektör Makinelerinin Karşılaştırılması,” İstanbul Aydın Üniversitesi Fen Bilimleri Dergisi, cilt 2, sayı 5, s. 71-83, 2010.

[111]L. Breiman, J. Friedman, C. J. Stone ve R. A. Olshen, “Classification algorithms and regression trees,” Classification and regression trees, s. 246-279, 1984.

[112]S. LaRoche, M. Joncas ve P. Foy, “Sample Design in TIMSS 2015,” M. O. Martin, I. V. S. Mullis ve M. Hooper, (Eds.) “Methods and Procedures in TIMSS 2015,” 2016.

[113]E. Frank, M. A. Hall ve I. H. Witten, “The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques,” (Dördüncü baskı), 2016.

[114]G. N. Masters ve B. D. Wright, B.D. “The partial credit model,” in M.J. van de Linden ve R.K. Hambleton (Eds.), “Handbook of modern item response theory,” 1997.

A

Öznitelik ve Ölçek Bilgileri

A.1 Öğrencilerin Fen Bilimleri başarısının sınıflandırılmasında kullanılan tüm değişkenler (Orijinal tanımlamalar)

Factor Name	Description	Domain
ITSEX	Sex of Students	1= Female, 2=Male
BSBG03	OFTEN SPEAK AT HOME	1=Always, 2=Almost Always, 3=Sometimes, 4= Never
BSBG05	DIGITAL INFORMATION DEVICES	1=None, 2=1-3 devices, 3=4-6 devices, 4=7-10 devices, 5= More than 10 device
BSBG06A	COMPUTER TABLET OWN	1=Yes, 2=No
BSBG06B	COMPUTER TABLET SHARED	1=Yes, 2=No
BSBG06C	STUDY DESK	1=Yes, 2=No
BSBG06F	OWN MOBILE PHONE	1=Yes, 2=No
BSBG06G	GAMING SYSTEM	1=Yes, 2=No
BSBG06H	HEATING SYSTEMS	1=Yes, 2=No
BSBG06I	COOLING SYSTEMS	1=Yes, 2=No
BSBG06J	WASHING MACHINE	1=Yes, 2=No
BSBG06K	DISHWASHER	1=Yes, 2=No
BSBG08	HOW FAR IN EDUCATION DO YOU EXPECT TO GO	1=Finish lower secondary, 2= Finish upper secondary, 3= Finish post-secondary, non-tertiary, 4=Finish short-cycle tertiary, 5=Finish bachelor's or equivalent, 6=Finish post graduate degree
BSBG11	ABOUT HOW OFTEN ABSENT FROM SCHOOL	1=Once a week or more, 2=Once every two weeks, 3=Once a month, 4=Never or almost never
BSBG12	HOW OFTEN BREAKFAST ON SCHOOL DAYS	1=Every day, 2=Most days, 3=Sometimes, 4=Never or almost never
BSBG13A	HOW OFTEN USE COMPUTER TABLET\HOME	1=Every day or almost every day, 2=Once or twice a week, 3=Once or twice a month, 4=Never or almost never
BSBG13B	HOW OFTEN USE COMPUTER TABLET\SCHOOL	1=Every day or almost every day, 2=Once or twice a week, 3=Once or twice a month, 4=Never or almost never
BSBG13C	HOW OFTEN USE COMPUTER TABLET\OTHER	1=Every day or almost every day, 2=Once or twice a week, 3=Once or twice a month, 4=Never or almost never
BSBG14A	ACCESS TEXTBOOKS	1=Yes, 2=No
BSBG14B	ACCESS ASSIGNMENTS	1=Yes, 2=No
BSBG14C	COLLABORATE WITH CLASSMATES	1=Yes, 2=No
BSBG14D	COMMUNICATE WITH TEACHER	1=Yes, 2=No
BSBG14E	FIND INFO TO AID IN MATH	1=Yes, 2=No
BSBs25AB	HOW OFTEN TEACHER GIVE YOU HOMEWORK/SCIENCE	1=Every day, 2=3 or 4 times a week, 3=1 or 2 times a week, 4=Less than once a week, 5=Never
BSBS26AB	EXTRA LESSONS LAST 12 MONTH\SCIENCE	1=Yes, to excel in class, 2=Yes, to keep up in class, 3=No, 9= Omitted or Invalid
BSBS26BB	EXTRA LESSONS HOW MANY MONTH\SCIENCE	1=Did not attend, 2=Less than 4 months, 3=4-8 months, 4=More than 8 months
BSBGHER	Home Educational Resources	
BSBGSSB	Students Sense of School Belonging	
BSBGSB	Student Bullying	
BSBGSLS	Students Like Learning Science	
BSBGESL	Engaging Teaching in Science Lessons	

BSBGSCS	Student Confident in Science	
BSBGSVS	Students Value Science	
BSDSLOWP	Science Achievement Too Low for Estimation	1=Yes, 2=No
BSDSWKHS	Weekly Time Spent on Science Homework	1=3 Hours or more, 2=More than 45 minutes but less than 3 hours, 3=45 minutes or less
BSSSCI01	1ST PLAUSIBLE VALUE SCIENCE	0=Not successful, 1= Successful

A.2 Fen Bilimleri veri seti için ev eğitim kaynakları ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBG04	Number of books in the home	1) 0-10 2) 11-25 3) 26-100 4) 101-200 5) More than 200
BSDG06S	Number of home study supports	1) None 2) Internet connection or own room 3) Both
BSDGEDUP	Highest level of education of either parent	1) Finished some primary or lower secondary or did not go to school 2) Finished lower secondary 3) Finished upper secondary 4) Finished post-secondary education 5) Finished university or higher

A.3 Fen Bilimleri veri seti için öğrenci zorbalığı ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBG16A	Made fun of me or called me names	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16B	Left me out of their games or activities	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16C	Spread lies about me	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16D	Stole something from me	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16E	Hit or hurt me (e.g., shoving, hitting, kicking)	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16F	Made me do things I didn't want to do	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16G	Shared embarrassing information about me	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

BSBG16H	Posted embarrassing things about me online	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16I	Threatened me	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.4 Fen Bilimleri veri seti için öğrencilerin okula ait olma hissi ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBG15A	I like being in school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15B	I feel safe when I am at school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15C	I feel like I belong at this school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15D	I like to see my classmates at school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15E	Teachers at my school are fair to me	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15F	I am proud to go to this school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15G	I learn a lot in school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.5 Fen Bilimleri veri seti için fen bilimlerindeki öğrenmeye olan sempatisi ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBS22A	I know what my teacher expects me to do	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS22B	My teacher is easy to understand	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS22C	I am interested in what my teacher says	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS22D	My teacher gives me interesting things to do	1) Agree a lot 2) Agree a little 3) Disagree a little

		4) Disagree a lot
BSBS22E	My teacher has clear answers to my questions	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS22F	My teacher is good at explaining science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS22G	My teacher lets me show what I have learned	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS22H	My teacher does a variety of things to help us learn	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS22I	My teacher tells me how to do better when I make a mistake	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.6 Fen Bilimleri veri seti için öğrencinin fen bilimlerine verdiği değer ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	**Soru**	**Cevap**
BSBS24A	I think learning science will help me in my daily life	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS24B	I need science to learn other school subjects	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS24C	I need to do well in science to get into the university of my choice	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS24D	I need to do well in science to get the job I want	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS24E	I would like a job that involves using science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS24F	It is important to learn about science to get ahead in the world	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS24G	Learning science will give me more job opportunities when I am an adult	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS24H	My parents think that it is important that I do well in science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS24I	It is important to do well in science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.7 Fen Bilimleri veri seti için fen bilimlerinde öğrencinin güveni ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBS23A	I usually do well in science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS23B	Science is more difficult for me than for many of my classmates	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS23C	Science is not one of my strengths	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS23D	I learn things quickly in science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS23E	I am good at working out difficult science problems	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS23F	My teacher tells me I am good at science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS23G	Science is harder for me than any other subject	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS23H	Science makes me confused	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.8 Fen Bilimleri veri seti için tahmin için fen bilimleri başarısı çok düşük ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSDSLOWP	Science Achievement Too Low for Estimation	1) Yes 2) No

A.9 Öğrencinin fen bilimlerini öğrenmeyi sevmesi ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBS21A	I enjoy learning science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS21B	I wish I did not have to study science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS21C	Science is boring	1) Agree a lot

		2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS21D	I learn many interesting things in science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS21E	I like science	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS21F	I look forward to learning Science in school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS21G	Science teaches me how things in the world work	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS21H	I like to conduct Science experiments	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBS21I	Science is one of my favorite subjects	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.10 Fen Bilimleri veri seti için fen bilimleri ödevinde haftalık harcanan zaman ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSDSWKHS	Weekly Time Spent on Science Homework	1) 3 Hours or more 2) More than 45 minutes but less than 3 hours 3) 45 minutes or less

A.11 Öğrencilerin Matematik başarısının sınıflandırılmasında kullanılan tüm değişkenler (Orijinal tanımlamalar)

Factor Name	Description	Domain
ITSEX	Sex of Students	1= Female, 2=Male
BSBG03	OFTEN SPEAK AT HOME	1=Always, 2=Almost Always, 3=Sometimes, 4= Never
BSBG05	DIGITAL INFORMATION DEVICES	1=None, 2=1-3 devices, 3=4-6 devices, 4=7-10 devices, 5= More than 10 device
BSBG06A	COMPUTER TABLET OWN	1=Yes, 2=No
BSBG06B	COMPUTER TABLET SHARED	1=Yes, 2=No
BSBG06C	STUDY DESK	1=Yes, 2=No
BSBG06F	OWN MOBILE PHONE	1=Yes, 2=No
BSBG06G	GAMING SYSTEM	1=Yes, 2=No
BSBG06H	HEATING SYSTEMS	1=Yes, 2=No
BSBG06I	COOLING SYSTEMS	1=Yes, 2=No
BSBG06J	WASHING MACHINE	1=Yes, 2=No
BSBG06K	DISHWASHER	1=Yes, 2=No
BSBG08	HOW FAR IN EDUCATION DO YOU EXPECT TO GO	1=Finish lower secondary, 2= Finish upper secondary, 3= Finis secondary, non-tertiary, 4=Finish short-cycle tertiary, 5=Finisl

		bachelor's or equivalent, 6=Finish post graduate degree
BSBG11	ABOUT HOW OFTEN ABSENT FROM SCHOOL	1=Once a week or more, 2=Once every two weeks, 3=Once a month, 4=Never or almost never
BSBG12	HOW OFTEN BREAKFAST ON SCHOOL DAYS	1=Every day, 2=Most days, 3=Sometimes, 4=Never or almost never
BSBG13A	HOW OFTEN USE COMPUTER TABLET\HOME	1=Every day or almost every day, 2=Once or twice a week, 3=Once or twice a month, 4=Never or almost never
BSBG13B	HOW OFTEN USE COMPUTER TABLET\SCHOOL	1=Every day or almost every day, 2=Once or twice a week, 3=Once or twice a month, 4=Never or almost never
BSBG13C	HOW OFTEN USE COMPUTER TABLET\OTHER	1=Every day or almost every day, 2=Once or twice a week, 3=Once or twice a month, 4=Never or almost never
BSBG14A	ACCESS TEXTBOOKS	1=Yes, 2=No
BSBG14B	ACCESS ASSIGNMENTS	1=Yes, 2=No
BSBG14C	COLLABORATE WITH CLASSMATES	1=Yes, 2=No
BSBG14D	COMMUNICATE WITH TEACHER	1=Yes, 2=No
BSBG14E	FIND INFO TO AID IN MATH	1=Yes, 2=No
BSBM38AA	HOW OFTEN TEACHER GIVE YOU HOMEWORK/MATHEMATICS	1=Every day, 2=3 or 4 times a week, 3=1 or 2 times a week, 4=Less than once a week, 5=Never
BSBM39AA	EXTRA LESSONS LAST 12 MONTH\MATHEMATICS	1=Yes, to excel in class, 2=Yes, to keep up in class, 3=No, 9= Omitted or Invalid
BSBM39BA	EXTRA LESSONS HOW MANY MONTH\MATHEMATICS	1=Did not attend, 2=Less than 4 months, 3=4-8 months, 4=More than 8 months
BSBGHER	Home Educational Resources	
BSBGSSB	Students Sense of School Belonging	
BSBGSB	Student Bullying	
BSBGSLM	Students Like Learning Mathematics	
BSBGEML	Engaging Teaching in Math Lessons	
BSBGSCM	Student Confident in Mathematics	
BSBGSVM	Students Value Mathematics	
BSDMLOWP	Mathematics Achievement Too Low for Estimation	1=Yes, 2=No
BSDMWKHW	Weekly Time Spent on Math Homework	1=3 Hours or more, 2=More than 45 minutes but less than 3 hours, 3=45 minutes or less
BSMMAT01	1ST PLAUSIBLE VALUE MATHEMATICS	0=Not successful, 1= Successful

A.12 Matematik veri seti için ev eğitim kaynakları ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	**Soru**	**Cevap**
BSBG04	Number of books in the home	1) 0-10 2) 11-25 3) 26-100 4) 101-200 5) More than 200
BSDG06S	Number of home study supports	1) None 2) Internet connection or own room 3) Both
BSDGEDUP	Highest level of education of either parent	1) Finished some primary or lower secondary or did not go to school 2) Finished lower secondary 3) Finished upper secondary 4) Finished post-secondary education 5) Finished university or higher

A.13 Matematik veri seti için öğrencilerin okula ait olma hissi ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBG15A	I like being in school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15B	I feel safe when I am at school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15C	I feel like I belong at this school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15D	I like to see my classmates at school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15E	Teachers at my school are fair to me	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15F	I am proud to go to this school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG15G	I learn a lot in school	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.14 Matematik veri seti için öğrenci zorbalığı ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBG16A	Made fun of me or called me names	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16B	Left me out of their games or activities	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16C	Spread lies about me	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16D	Stole something from me	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16E	Hit or hurt me (e.g., shoving, hitting, kicking)	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16F	Made me do things I didn't want to do	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16G	Shared embarrassing information about me	1) Agree a lot

		2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16H	Posted embarrassing things about me online	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBG16I	Threatened me	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.15 Matematik veri seti için öğrencinin matematiği öğrenmeyi sevmesi ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBM17A	I enjoy learning mathematics	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM17B	I wish I did not have to study mathematics	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM17C	Mathematics is boring	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM17D	I learn many interesting things in mathematics	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM17E	I like mathematics	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM17F	I like any schoolwork that involves numbers	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM17G	I like to solve Mathematics problems	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM17H	I look forward to Mathematics class	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM17I	Mathematics is one of my favorite subjects	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.16 Matematik veri seti için matematik öğrenmeye olan sempatisi ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBM18A	I know what my teacher expects me to do	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM18B	My teacher is easy to understand	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM18C	I am interested in what my teacher says	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM18D	My teacher gives me interesting things to do	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM18E	My teacher has clear answers to my questions	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM18F	My teacher is good at explaining mathematics	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM18G	My teacher lets me show what I have learned	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM18H	My teacher does a variety of things to help us learn	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM18I	My teacher tells me how to do better when I make a mistake	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM18J	My teacher listens to what I have say	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.17 Matematik veri seti için matematikte öğrencinin güveni ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSBM19A	I usually do well in mathematics	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM19B	Mathematics is more difficult for me than for many of my classmates	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM19C	Mathematics is not one of my strengths	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM19D	I learn things quickly in mathematics	1) Agree a lot

		2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM19E	Mathematics makes me nervous	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM19F	I am good at working out difficult Mathematics problems	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM19G	My teacher tells me I am good at mathematics	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM19H	Mathematics is harder for me than any other subject	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM19I	Mathematics makes me confused	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot

A.18 Matematik veri seti için öğrencinin matematiğe verdiği değer ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	**Soru**	**Cevap**
BSBM20A	I think learning mathematics will help me in my daily life	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM20B	I need mathematics to learn other school subjects	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM20C	I need to do well in mathematics to get into the university of my choice	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM20D	I need to do well in mathematics to get the job I want	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM20E	I would like a job that involves using mathematics	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM20F	It is important to learn about mathematics to get ahead in the world	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM20G	Learning mathematics will give me more job opportunities when I am an adult	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM20H	My parents think that it is important that I do well in mathematics	1) Agree a lot 2) Agree a little 3) Disagree a little 4) Disagree a lot
BSBM20I	It is important to do well in mathematics	1) Agree a lot 2) Agree a little

	3) Disagree a little 4) Disagree a lot

A.19 Matematik veri seti için tahmin için matematik başarısı çok düşük ölçeğinin içeriği soru (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSDMLOWP	Mathematics Achievement Too Low for Estimation	1) Yes 2) No

A.20 Matematik veri seti için matematik ödevinde haftalık harcanan zaman ölçeğinin içeriği (Orijinal tanımlamalar)

Soru kodu	Soru	Cevap
BSDMWKHW	Weekly Time Spent on Math Homework	1) 3 Hours or more 2) More than 45 minutes but less than 3 hours 3) 45 minutes or less

Tezden Üretilmiş Yayınlar

İletişim Bilgisi: enesf@yildiz.edu.tr

Makaleler

1. E. Filiz ve E. Oz, E. “Finding the Best Algorithms and Effective Factors in Classification of Turkish Science Student Success,” Journal Of Baltic Science Education, cilt 18, sayı 2, s. 239-253, 2019. (SSCI)

2. E. Filiz ve E. Öz, “Classification of BIST-100 Index'Changes via Machine Learning Methods,” Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi, cilt 39, sayı 1, s. 117-129, 2017. (EBSCO-Host, ECONLIT, TÜBİTAK - TR Dizin Ulakbim)

Konferans Bildirileri

1. E. Filiz, Ö.E. Aşkin ve E. Öz, "The Classification with Support Vector Machines and Logistic Regression: The case study of TIMSS 2015 Turkey Results," 19. Uluslararası Ekonometri Yöneylem Araştırması ve İstatistik Sempozyumu, ANTALYA, TÜRKiYE, 2018, s. 1325-1331.

2. E. Filiz, Ö.E. Aşkin ve E. Öz, "Finding The Best Machine Learning Algorithms in Predicting The Successes of 8th Grade Turkish Students," 19. Uluslararası Ekonometri Yöneylem Araştırması ve İstatistik Sempozyumu, ANTALYA, TÜRKiYE, 2018, s. 1118-1124.

3. E. Filiz, T. Karadağ, Ö.E. Aşkin ve E. Öz, "Comparison of Classification Achievements of Logistic Regression and Decision Trees Algorithms on TIMSS 2015 Data", 4th International Researchers, Statisticians And Young Statisticians Congress, İZMİR, TÜRKIYE, 2018, s. 86-86.

4. S. Akoğul, E. Filiz, H.A. Karaboğa ve E. Öz, "Büyük Dünya Endeksleri ve Makroekonomik Göstergeler Kullanılarak BIST-100 Endeksinin Yönünün Makine Öğrenmesi Yöntemleri ile Sınıflandırılması", XVIII. Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu 2017, TRABZON, TÜRKIYE, 2017, s. 412-412.