



**ÖĞRENCİLERİN AKADEMİK BAŞARISININ
DEĞERLENDİRMESİNDE MAKİNE
ÖĞRENME ALGORİTMALARININ
KULLANILMASI**

Yüksek Lisans Tezi

Derya ÇINAR

Eskişehir 2023

**ÖĞRENCİLERİN AKADEMİK BAŞARISININ DEĞERLENDİRMESİNDE
MAKİNE ÖĞRENME ALGORİTMALARININ KULLANILMASI**

Derya ÇINAR

Yüksek Lisans Tezi

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Donanımı Bilim Dalı

Danışman: Dr. Öğr. Üyesi Sevcan YILMAZ GÜNDÜZ

Eskişehir

Eskişehir Teknik Üniversitesi

Lisansüstü Eğitim Enstitüsü

Haziran 2023

Bu tez çalışması BAP Komisyonu tarafından kabul edilen 21LÖT098 no.lu proje kapsamında desteklenmiştir.

JÜRİ VE ENSTİTÜ ONAYI

Derya ÇINAR'ın ÖĞRENCİLERİN AKADEMİK BAŞARISININ DEĞERLENDİRMESİNDE MAKİNE ÖĞRENME ALGORİTMALARININ KULLANILMASI başlıklı çalışması 19/06/2023 tarihinde aşağıdaki jüri tarafından değerlendirilerek “Eskişehir Teknik Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği”nin ilgili maddeleri uyarınca, Bilgisayar Mühendisliği Anabilim dalında Yüksek Lisans Tezi olarak kabul edilmiştir.

Jüri Üyeleri

Unvan Adı Soyadı

İmza

Üye	: Dr. Öğr. Üyesi Sevcan YILMAZ GÜNDÜZ	
Üye	: Dr. Öğr. Üyesi Mehmet KILIÇARSLAN	
Üye	: Dr. Öğr. Üyesi Sinem BOZKURT KESER	

Prof. Dr. Semra KURAMA

Lisansüstü Eğitim Enstitüsü Müdürü

19/06/2023

DANIřMAN ONAYI

Daniřmanlıęını yrttęm Yksek Lisans ęrencisi Derya INAR, ęRENCİLERİN AKADEMİK BAřARISININ DEęERLENDİRMEĐİNDE MAKİNE ęRENME ALGORİTMALARININ KULLANILMASI bařlıklı tez alıřmasını tamamlamıřtır. Hazırlamıř olduęu tez tarafımca incelenmiř ve ęrencinin tez savunma sınavına alınması bilimsel ve etik aıdan uygun grlmřtr.

Tez Daniřmanı

Dr. ęr. yesi Sevcan YILMAZ GNDZ

ÖZET

ÖĞRENCİLERİN AKADEMİK BAŞARISININ DEĞERLENDİRMESİNDE MAKİNE ÖĞRENME ALGORİTMALARININ KULLANILMASI

Derya ÇINAR

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Donanımı Bilim Dalı

Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Haziran 2023

Danışman: Dr. Öğr. Üyesi Sevcan YILMAZ GÜNDÜZ

Günümüzde yapay zekâ birçok alanda kullanılmaktadır. Bu alanlardan biri de eğitimidir. Eğitim faaliyetlerinin yanı sıra başarıyı sınıflandırmak, başarıyı etkileyen faktörleri belirlemek, başarısızlık nedenlerini tespit etmek ve öğrencinin akademik başarı düzeyini önceden tahmin etmek için makine öğrenme yöntemleri kullanılabilmektedir. Başarısızlık durumunda tedbir alınıp, bu konuda çözüm üretilebilmesi için bu durumun öngörülebilmesi çok önemlidir. Başarısızlık kadar başarının da öngörülebilmesi teşvik ve destekle başarının arttırılmasını ve istikrarı sağlanabilir. Bu çalışmada ortaöğretim öğrencileri ile hazırlanmış 2 farklı veri seti ve bu verilerin farklı türevlerinin kullanıldığı 12 farklı çalışma sunulmuştur. Her bir çalışmada sekiz adet makine öğrenme algoritması: derin öğrenme algoritması, yapay sinir ağları, basit lojistik regresyon, yinelemeli sınıflandırıcı optimize edici, karar tablosu, bulanık mantık tabanlı algoritma, tek kural, lojistik model ağacı algoritmaları, klasik, torbalama ve yükseltme yöntemleri önce tüm öznitelikler kullanarak daha sonra da belirli öznitelikler seçilerek denenmiştir. Öznitelik seçilerek yapılan çalışmalarda öznitelikler farklı algoritmalar ile belirlenmiştir. 10 kat çaprazlama yöntemi kullanılarak başarı sınıfları tahmin edilmeye çalışılmış ve doğruluk oranı, kesinlik, duyarlılık ve f-değeri hesaplanmıştır. Çıkan sonuçlar benzeri çalışmalar ile karşılaştırılmıştır. Yapılan çalışmalar sonucunda veri setinin özellikleri ile doğruluk oranı arasında ilişki olduğu gözlemlenmiştir.

Anahtar Sözcükler: Makine Öğrenmesi, Sınıflandırma, Öğrenci Başarı Tahmini

ABSTRACT

PREDICTION OF SECONDARY SCHOOL STUDENTS' SUCCESS WITH MACHINE LEARNING

Derya ÇINAR

Department of Computer Engineering

Programme in Computer Hardware

Eskişehir Technical University, Institute of Graduate Programs, June 2023

Supervisor: Asst. Prof. Dr. Sevcan YILMAZ GÜNDÜZ

In today's world, artificial intelligence is utilized in numerous fields, and education is one of them. Alongside educational activities, machine learning methods can be employed to classify success, determine influential factors, identify causes of failure, and predict students' academic achievement levels in advance. It is crucial to anticipate this situation in order to take precautions and develop solutions in case of failure. If success situations can be predicted as accurately as failure, encouragement, and support can be provided to enhance and stabilize success. Twelve different research studies that use two distinct datasets prepared with secondary school students, along with various derivatives of these datasets were presented. In each study, eight machine learning algorithms were employed, namely: Deep Learning Algorithm, Artificial Neural Networks, Simple Logistic Regression, Iterative Classifier Optimizer, Decision Table, Fuzzy Rule Induction Algorithm, One-Rule and Logistic Model Tree. Initially, all attributes were employed to test classical, bagging, and boosting methods. Subsequently, a selection process was conducted to determine specific attributes for further analysis. In the studies conducted with feature selection, different algorithms were used to determine the features. The 10-fold crossover method was employed to predict success classes, and various performance metrics such as accuracy rate, precision, recall, and F-measure were calculated. Finally, the obtained results were compared with similar studies. As a result of the analysis, it was observed that there is a relationship between the characteristics of the dataset and the accuracy rate.

Keywords: Machine Learning, Classification, Student Achievement Prediction.

TEŞEKKÜR

Tez çalışmam süresince, bilgi ve deneyimlerini benden esirgemeyen tez danışmanım Sayın Dr. Öğr. Üyesi Sevcan YILMAZ GÜNDÜZ'e katkılarından dolayı içtenlikle teşekkürler ederim.

Lisansüstü eğitime başlamamdan bitirme sürecine kadar tecrübelerini ve yardımlarını her an paylaşan kuzenlerim Gamze ve Yasin Doğan'a, hayatımın her evresinde sürekli yanımda olup beni destekleyen aileme sonsuz teşekkür ederim.

Derya ÇINAR



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Eskişehir Teknik Üniversitesi tarafından kullanılan “bilimsel intihal tespit programı”yla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Derya ÇINAR

İÇİNDEKİLER

Sayfa

BAŞLIK SAYFASI	I
JÜRİ VE ENSTİTÜ ONAYI.....	II
DANIŞMAN ONAYI	III
ÖZET	IV
ABSTRACT.....	V
TEŞEKKÜR	VI
ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....	VII
İÇİNDEKİLER	VIII
TABLolar DİZİNİ	XI
GÖRSELLER DİZİNİ	XII
SİMGELER VE KISALTMALAR DİZİNİ.....	XIII
1. GİRİŞ	1
2. LİTERATÜR.....	3
2.1. Geliştirilen Çalışma.....	7
3. YÖNTEM VE ALGORİTMALAR.....	8
3.1. Makine Öğrenmesi ve Sınıflandırma	8
3.1.1. Klasik makine öğrenme yöntemleri	8
3.1.2. Torbalama yöntemi	9
3.1.3. Yükseltme yöntemi	9
3.2. Sınıflandırma Algoritmaları	10
3.2.1. Derin öğrenme	10
3.2.2. Yapay sinir ağları	11
3.2.3. Basit lojistik regresyon.....	11
3.2.4. Yinelemeli sınıflandırıcı optimize edici	12
3.2.5. Karar tablosu	12
3.2.7. Tek kural	13

3.2.8. Lojistik model ağacı	14
3.3. Öznitelik Seçme Algoritmaları	14
3.3.1. Bilgi kazanç özelliği değerlendirmesi	14
3.3.2. Korelasyon tabanlı öznitelik altkütmesi değerlendiricisi.....	14
3.3.3. Sınıflandırıcı altküme değerlendirmesi	14
3.3.4. Korelasyon öznitelik değeri	15
3.3.5. Kazanç oranı özelliği değerlendirmesi	15
3.3.6. Tek kural özniteliği değerlendirmesi.....	15
3.3.7. F özniteliğini atarak değerlendirme	15
3.3.8. Simetrik belirsiz öznitelik değerlendirmesi	15
3.4. Öznitelik Sıralama Metotları	15
3.4.1. En iyi en önce sıralama metodu	15
3.4.2. Açgözlü adım sıralama metodu.....	16
3.4.3. Sıralı metodu.....	16
3.5. Sınıflandırmada Kullanılan Metrikler.....	16
3.6. Sınıflandırmada Kullanılan Program	18
3.7. Sınıflandırmada Kullanılan Yöntemlerin Seçenekleri	18
4. VERİ SETLERİ VE ÖN İŞLEMLERİ.....	20
4.1. Veri Seti Tanıtımı.....	20
4.2. Veri Seti Ön İşleme	21
4.3. Veri Setlerinde Özniteliklerin Belirlenmesi.....	22
5. DENEYSEL ÇALIŞMALAR.....	24
5.1. İki Sınıflı Sınıflandırma Sonuçları	24
5.1.1. Matematik dersi 2 sınav notu içeren deney sonuçları.....	24
5.1.2. Matematik dersi 1 sınav notu içeren deney sonuçları.....	26
5.1.3. Matematik dersi sınav notu içermeyen deney sonuçları.....	27
5.1.4. Portekiz Dili dersi 2 sınav notu içeren deney sonuçları.....	29
5.1.5. Portekiz Dili dersi 1 sınav notu içeren deney sonuçları.....	31
5.1.6. Portekiz Dili dersi sınav notu içermeyen deney sonuçları.....	32
5.2. Beş Sınıflı Sınıflandırma Sonuçları	33
5.2.1. Matematik dersi 2 sınav notu içeren deney sonuçları.....	34
5.2.2. Matematik dersi 1 sınav notu içeren deney sonuçları.....	35

5.2.3. Matematik dersi sınav notu içermeyen deney sonuçları.....	37
5.2.4. Portekiz Dili dersi 2 sınav notu içeren deney sonuçları.....	38
5.2.5. Portekiz Dili dersi 1 sınav notu içeren deney sonuçları.....	40
5.2.6. Portekiz Dili dersi sınav notu içermeyen deney sonuçları.....	41
5.2.7. Algoritmaların başarısını genel bir değerlendirme.....	43
5.3. Veri Seti İle Yapılmış Çalışmalar	48
6. SONUÇ VE TARTISMA	54
KAYNAKÇA.....	57



TABLÖLAR DİZİNİ

Sayfa

Tablo 3.1. Metriklerde Kullanılan Örüntü	16
Tablo 3.2. Weka programında torbalama işlem ayarları	18
Tablo 3.3. Weka programında yükseltme işlem ayarları	19
Tablo 4.1. Veri setlerindeki öznitelikler, açıklaması ve alacağı değer aralıkları	20
Tablo 4.2. Notların sınıflandırılması / not baremi 2’li sınıflandırma sistemi	21
Tablo 4.3. Notların sınıflandırılması / not baremi 5’li sınıflandırma sistemi	22
Tablo 4.4. Weka programında öznitelik seçme algoritmaları	22
Tablo 5.1. Matematik dersi final notu çıkartılmış bütün sonuçlar	25
Tablo 5.2. Matematik dersi final ve 2.sınav notları çıkartılmış 2 sınıflı sonuçlar	26
Tablo 5.3. Matematik dersi bütün sınav notları çıkartılmış 2 sınıflı sonuçlar	28
Tablo 5.4. Portekiz Dili dersi final notu çıkartılmış 2 sınıflı sonuçlar.....	30
Tablo 5.5. Portekiz Dili dersi final ve 2.sınav notları çıkartılmış 2 sınıflı sonuçlar	31
Tablo 5.6. Portekiz Dili dersi bütün sınav notları çıkartılmış 2 sınıflı sonuçlar	33
Tablo 5.7. Matematik dersi final notu çıkartılmış 5 sınıflı sonuçlar.....	34
Tablo 5.8. Matematik dersi final ve 2.sınav notları çıkartılmış 5 sınıflı sonuçlar	36
Tablo 5.9. Matematik dersi bütün sınav notları çıkartılmış 5 sınıflı sonuçlar	37
Tablo 5.10. Portekiz Dili dersi final notu çıkartılmış 5 sınıflı sonuçlar.....	39
Tablo 5.11. Portekiz Dili dersi final ve 2.sınav notu çıkartılmış 5 sınıflı sonuçlar	40
Tablo 5.12. Portekiz Dili dersi bütün sınav notları çıkartılmış 5 sınıflı sonuçlar	42
Tablo 5.13. Benzer çalışmalardan 2 sınıflı çalışmalar	49
Tablo 5.14. Benzer çalışmalardan 3 ve 5 sınıflı çalışmalar	51

GÖRSELLER DİZİNİ

Sayfa

Görsel 5.1. Derin Öğrenme algoritma grafiği	44
Görsel 5.2. Çok Katmanlı Algılayıcı algoritma grafiği.....	44
Görsel 5.3. Basit Lojistik Algoritma Grafiği	45
Görsel 5.4. Yinelemeli Sınıflandırıcı Optimize Edici Algoritma Grafiği	46
Görsel 5.5. Karar Tablosu Algoritma Grafiği	46
Görsel 5.6. FURIA Algoritma Grafiği	47
Görsel 5.7. Tek-Kural Algoritma Grafiği	47
Görsel 5.8. Lojistik Model Ağacı Algoritma Grafiği.....	48

SİMGELER VE KISALTMALAR DİZİNİ

Adaboost	: Adaptive Boosting
ANOVA	: ANalysis Of VAriance (Varyans Analizi)
ANN	: Artificial Neural Network(Yapay Sinir Ağları)
BGA	: Binary Genetic Algorithm(İkili Genetik Algoritma)
CNN	: Convolutional Neural Network(Evreşimsel Sinir Ağları)
DL	: Deep Learning (Derin Öğrenme)
DL4MLP	: Deep Learning For Multilayer Perceptron(Çok Katmanlı Algılayıcı İçin Derin Öğrenme)
DNN	: Deep Neural Network(Derin Sinir Ağları)
DO	: Doğruluk Oranı
DT	: Decision Table(Karar Tablosu)
DTree	: Decision Tree (karar ağacı)
Duy	: Duyarlılık
EBGA	: Enhanced Gelişmiş Binary Genetic Algorithm(Gelişmiş İkili Genetik Algoritma)
F-1	: F1- Score(F1-Değeri)
FURIA	: Fuzzy Unordered Rule Algorithm(Bulanık Sırasız Kural Tümevarım Algoritması)
ICO	: Iteratif Classifier Optimizer(Yinelemeli Sınıflandırıcı Optimize Edici)
Kes	: Kesinlik
k-NN	: k-Nearest Neighbors (k en yakın komşu)
LMT	: Logistic Model Tree (Lojistik Model Ağacı)
LR	: Logistic Regression (Lojistik Regresyon)

MLP	: Multi Layer Perceptron (Çok Katmanlı Algılayıcı)
NB	: Naive Bayes (Naif Bayes)
NN	: Neural Network (Sinir Ağları)
NV	: Naive Predictor (Naif Tahminci)
One-R	: Tek Kural (One Rule)
RF	: Random Forest (Rassal Orman)
SL	: Simple Logistic(Basit Lojistik Regresyon)
SMOTE	: Sentetik Azınlık Örneklem Arttırma Yöntemi(Synthetic Minority Over-Sampling Technique)
SVM	: Destek Vektör Makinesi(Support Vector Machine)
WEKA	: Waikato Environment for Knowledge Analysis

1. GİRİŞ

Ülkelerin geleceği gençlerinin başarısı ve geleceği için atacağı adımlar ile bağlantılıdır. Başarı adımlarından biri de öğrencilerin eğitim hayatında geçtikleri süreçlerdir. Öğrencilerin akademik başarıları hem kendi geleceklerini hem de toplumu şekillendiren bir olgudur. Doğru yönlendirilemeyen veya erken müdahale edilemeyen başarısızlıklar büyüyerek sorun haline gelir (Rao et al., 2016).

2022 eğitim öğretim yılında ülkemizde ortaöğretim düzeyinde öğrenim gören öğrenci sayısı yaklaşık 6,5 milyondur. Ortaöğretim, meslek seçimi için önemli bir basamaktır. Bu yıllarda elde edilen başarılar öğrencilerin hem yükseköğretimde üniversite seçimini etkiler, hem de onların meslek hayatında başarılı ve özgüvenli bireyler olmalarını sağlar. Aksi takdirde ortaöğretim dönemindeki başarısızlıkları telafi etmek zordur (Keskin & Yapıcı, 2008).

Öğrencilerin notlarının önceden tahmin edilebilmesi, öğrencileri doğru yönlendirmek için önemlidir (Rao et al., 2016). Öğrencilerin akademik başarılarını arttırmada başarıyı düşüren faktörlerin belirlenmesi, bu faktörleri ortadan kaldırmak, bu mümkün değilse bile etkisini azaltmak etkilidir. Fakat bu faktörlerin belirlenmesi kolay değildir. Sebebi de öğrencilerin akademik başarı performansını etkileyen çok fazla faktör olmasıdır (Al-Radaideh et al., 2006). Ayrıca bu işlemi zorlaştıran bir diğer sebep de ölçülemeyen faktörlerin varlığıdır.

Öğrenci başarısı sosyo-ekonomik ve fiziksel şartlarına bağlı olabilir. Yaşadığı toplum, aile yaşantısı, okul şartları, maddi imkânları gibi ölçülebilir; yetenek, ilgi, zekâ gibi kolay ölçülemeyen bireysel farklılıklarda başarıyı etkileyen faktörlerdir (Şama & Tarım, 2007). Ekonomik şartlar, kardeş sayısı, aile bireylerinin eğitim düzeyleri ve meslekleri de başarıyı etkili faktörlerdendir (Sezer, 2007).

Öğrenci başarısını tahmin etmek için kullanılan veri setlerinde yer alan anket cevapları ve okul sistemlerinden alınan notlar ile başarıyı tahmin etmek ve bu verilerden hangilerinin etkili olduğunu bulmak yapay zekâ yöntemleri ile mümkündür (Iqbal et al., 2017).

Yapılan anketler ve tutulan kayıtlarla öğrenciler ile ilgili çok fazla ve çok farklı veriler tutulabilmektedir. Bu verilerin türleri ve boyutu değişebilmektedir (Yadav & Pal, 2012). Böyle büyük verilerle çalışmak çıkarım yapmak ve tahmin yapmak için daha elverişlidir. Öğrenci başarısının tahmin edilmesinde son zamanlarda birçok alanda popüler olan makine öğrenme algoritmaları kullanılabilmektedir. Makine öğrenme

algoritmaları ile öğrenci verileri kullanılarak öğrencilerin olası başarı durumları tahmin edilmeye çalışılmaktadır (Gadhavi & Patel, 2017).

Makine öğrenme teknikleri kullanıldıkları veri seti türüne ve işlevine göre değişmektedir. Bu tekniklerden bazıları kümeleme, sınıflandırma, ilişki analizi gibi tekniklerdir (Han et al., 2001). Sınıflandırma sonuçlarını daha verimli hale getirmek için basit yöntemler dışında birkaç yöntemin birlikte kullanıldığı bileşik makine yöntemleri de kullanılabilir (Erdoğan et al., 2017). Her algoritmanın verimli çalıştığı veri seti, setin içeriğine göre farklılık göstermektedir.

Büyük verileri sınıflandırarak genellikle risk analizi, görüntü işleme, veri tahmin etme işlemlerinde kullanılır (Han et al., 2001). Öğrenci performans tahmini için kullanılan teknik sınıflandırmadır (Soofi & Awan, 2017). Yapılan literatür çalışmalarından elde edilen bilgilere göre en çok tercih edilen sınıflandırma algoritmalarından bazıları k en yakın komşu(k-NN), destek vektör makineleri(SVM), lojistik regresyon(LR), rastgele orman(RF), karar ağaçları(DTree), Naive Bayes(NB) algoritmalarıdır (Abdualgalil & Abraham, 2020).

Sınıflandırma işlemi var olan verilerin niteliklerine göre kategorize etme işlemidir. Bu işlem eğitim verileri ile sistemi eğiterek sınıf etiketlerini tahmin etme işlemidir. Bu işlemde, sınıfı belli olan verilerle sınıfı tahmin edilecek verilerin benzerliğine göre grupta yapılır (Abdualgalil & Abraham, 2020).

Bu çalışmada öğrencilerin akademik başarıları tahmin edilirken basit makine öğrenme teknikleri ve bileşik makine öğrenme tekniklerinden torbalama(bagging) ve yükseltme(boosting) kullanılmıştır.

Çalışmada açık kaynak erişimli olan Portekiz’de ortaöğretim öğrenci verileri ile hazırlanmış, sosyo-ekonomik şartlarının yanı sıra matematik ve Portekiz Dili dersine ait 3 not bilgisi içeren iki veri seti kullanılmış ve bu setlerdeki not bilgileri kademeli olarak çıkartılarak başarı sınıfları tahmin edilmeye çalışılmıştır. En verimli yöntemi bulmak için açık kaynak kodlu Weka(Waikato Environment for Knowledge Analysis) programında birçok algoritma farklı şekillerde denenmiştir. Testler önce tüm öznitelikler kullanılarak, daha sonra öznitelikler azaltılarak klasik, torbalama ve yükseltme yöntemleri ile 8 farklı algoritma ile yapılmıştır. Her sete 48 adet test 3 farklı şekilde uygulandıktan sonra, 2 kademeli sınıflandırma ile aynı işlemler uygulanmıştır. Çalışmada toplam 376 test sonucu sunulmuştur.

2. LİTERATÜR

Eğitim bilimleri için öğrenci başarı sınıflarını öngörebilmek, eğitimin kalitesini arttırmak için önemli ve ihtiyaç duyulan bir konudur. Bu alanda ülkemizde ve dünyada yapılan çeşitli çalışmalar mevcuttur. Bu bölümde bu çalışmalardan bir kısmına değinilecektir.

Cortez ve Silva (2008) yaptıkları çalışmada Portekiz’de iki farklı lisede öğrenim gören öğrencilerin 33 farklı öznelik ve 3 sınav notunu barındıran gerçek dünya verilerini içeren 2 farklı veri seti ile makine öğrenme tekniklerini kıyaslamışlardır. Çalışmayı derinleştirmek amacıyla sınav notlarını kademeli olarak çıkartarak öğrenci akademik başarısını tahmin eden sistemler oluşturmuşlardır. Cortez ve Silva çalışmalarında bütün deneylerini R programlama ortamında gerçekleştirmişlerdir. Açık kaynak bir kütüphane olan RMiner Kütüphanesi ile derin öğrenme yöntemlerinden rastgele orman, sinir ağları(NN), destek vektör makineleri, karar ağaçları ve naif tahminci(NV) algoritmaları kullanmışlardır. Çalışmalarında önceki yılların başarı durumlarının yüksek oranda etkili olduğu ancak not haricindeki diğer ilgili özelliklerinde etkili olduğunu görmüşlerdir. Çalışmalarında doğru tahminleri elde etmek için her türevi için 20 kere 10 kat çaprazlama yaparak 200 değer elde etmişlerdir. Verilerin %10’u ile 10 kerede bütün verileri test etmişlerdir. Sistemlerin tahmin başarısı en yüksek olan veri seti 2 sınav notunu da içeren çalışma, en başarılı algoritma ise naif tahminci olarak tespit edilmiştir.

Bentaleb ve Abouchabaka (2022) 3 sınıflı matematik ve Portekiz dili derslerinin notları ile hazırlanan veri setleri ile Python programında yaptıkları çalışmada karar ağaçları, rastgele orman ve destek vektör makineleri algoritmaları arasında %86 doğruluk oranıyla en başarılı algoritmanın rastgele orman algoritması olduğunu belirtmişlerdir. Bozkurt Keser ve Aghalarova (2022) çalışmalarında yeni bir hibrit topluluk öğrenme algoritması (HELA) kullanmışlardır. Yaptıkları bu çalışmada hibrit topluluk öğrenme algoritmasına girdi olarak 3 farklı yükseltme algoritması ve bu algoritmaların farklı kombinasyonlarını kullanmışlardır. Çalışmalarında veri seti olarak Portekiz’de ortaöğretim öğrencilerin matematik ve Portekiz Dili dersine ait notları ve diğer bilgilerini içeren ikili ve beşli sınıflandırma bulunan veri setlerini kullanmışlardır. Yaptıkları çalışma sonuçlarına göre hibrit topluluk öğrenme algoritmasının yalın yükseltme algoritmalarından daha başarılı olduğunu belirtmişlerdir.

Garcia (2022) 2 kademeli(geçti-kaldı) matematik ve İspanyolca derslerine ait not bilgileriyle hazırlanan veri setiyle yaptıkları sınıflandırma çalışmasında karar ağaçları algoritmasını Python’da deneyerek %91 doğruluk oranında bir başarı elde etmişlerdir. Quy ve arkadaşları (2022) yaptıkları çalışmada Portekiz Dili dersi notları ile 2 kademeli(geçti-kaldı) bir sınıflandırma üzerinde çalışmışlar ve klasik yöntemlerin yanı sıra Adaboost yükseltme yöntemini denemişlerdir. Çalışma sonucunda %95 oranında bir başarı elde etmişlerdir. Özkan ve arkadaşları (2018) sınıflandırma ve en etkili öz niteliği bulmak amacıyla yaptıkları çalışmada matematik veri seti ile 2 kademeli(geçti-kaldı) sınıflandırma yapmışlardır. R programında yaptıkları çalışmalarının sonucunda farklı algoritmalar içinde %0.86 oranıyla C5.0 algoritmasının başarılı olduğunu belirtmişlerdir. Başer ve arkadaşları (2020) 2 kademeli matematik dersi veri seti ile yaptıkları çalışmada WEKA programında çeşitli algoritmalar ile sınıflandırma çalışması yapmışlar ve tek kural(One-R) algoritması ile %92.15 doğruluk oranına ulaşmışlardır.

Yadav ve Pal (2012) yaptıkları çalışmada mühendislik öğrencilerinin bir yıl önceki sınav notlarını kullanarak o yıl alacakları notları tahmin etmeye çalışmışlardır. Çalışmalarında karar ağacı algoritmalarını kullanmışlar, en başarılı karar ağacı algoritmasının da C4.5(karar ağacı algoritmalarından biri) algoritması olduğunu belirtmişlerdir. Benzer bir çalışmayı Fırat Üniversitesi’nin öğrenci bilgilerini içeren veri seti ile yapan Şengür (2013), karar ağaçları ve yapay sinir ağları(ANN) yöntemlerini kullanmıştır. Çalışmasında mezuniyet notlarını en iyi tahmin eden algoritmanın yapay sinir ağları olduğunu gözlemlemiştir. Yine bir üniversite öğrencilerinin bilgileriyle hazırlanan bir veri setiyle çalışan Hakyemez (2015) birinci sınıf öğrencilerinin başarısını farklı algoritmalar ile tahmin etmeye çalışmış ve en başarılı sonucu rastgele orman algoritması ile ele ettiğini belirtmiştir. Iqbal ve arkadaşları (2017) ise farklı yöntemler deneyerek 225 üniversite öğrencisinin akademik performanslarını tahmin eden bir çalışma yayınlamışlardır. Veri setinde öğrencilerin geçmiş akademik bilgileri, mülakat sonuçları, ders notları gibi bilgileri içermektedir. Çalışma sonucunda en iyi tahmin sonucunu veren yöntemin Kısıtlı Boltzmann Makineleri yöntemi olduğu belirtilmiştir. Gazi Üniversitesi öğrencilerine ait bilgilerle oluşturulan veri seti ile Erdoğan ve arkadaşları (2017) öğrencilerin geçme-kalma durumunu tahmin etmek için çalışma yapmışlardır. Bu çalışma da basit ve bileşik makine öğrenme yöntemlerinden rastgele orman, lojistik model ağacı(LMT), torbalama ve hızlandırma gibi algoritmaları

denemişler ve torbalama yöntemi ve rastgele orman alt sınıflandırıcısı kullanarak 5 kat çapraz doğrulama ile en başarılı sonuca ulaşıldığını belirtmişlerdir.

Aydoğan ve Karcı (2018) meslek yüksekokulu öğrencileri verileri ile yaptıkları çalışmada denetimli ve denetimsiz öğrenme algoritmalarıyla öğrenci başarısını etkileyen en önemli faktörlerden birinin meslek liselerinden sınavsız geçişin, diğer faktörün ise öğrencilerin gelir durumu olduğunu belirtmişlerdir. Elde ettikleri bulgulara göre meslek lisesinden mezun olarak meslek yüksekokullarına sınavsız geçiş yapan öğrencilerin başarısının, sınav puanı ile yerleşen öğrencilere göre daha düşük olduğunu; ayrıca öğrenci gelir durumu yüksek olan öğrencilerin gelir durumu düşük olan öğrencilerden daha başarılı olduğunu tespit etmişlerdir. Adak ve Duralioğlu (2023) yaptıkları çalışmada 10. ve 11. sınıflarda okuyan 87 bilişim teknolojileri öğrencisinin alan dal derslerindeki sınav notları ile yaptıkları çalışmada 3 sınav notunu değişimli olarak kullanarak diğer sınav notlarını tahmin eden deneyler yapmışlardır. Deneylerde 87 öğrenciye ait veri yetersiz olduğu için sentetik veriler oluşturularak; doğrusal regresyon, k en yakın komşu ve karar ağaçları algoritmalarını kullanmışlardır. K en yakın komşu ve karar ağaçları algoritmalarında en etkili öznelilikler kullanılmış ve k en yakın komşu algoritmasının en başarılı algoritma olduğu belirtilmiştir.

Yapılan çalışmaların hepsi başarılı olmamıştır. Al-Radaideh vd (2006) üniversite öğrencilerinin programlama dersi verilerini kullanarak ders notlarını harf notu olarak tahmin etme üzerine çalışmıştır. Çalışmasında yeterli veri olmadığı için doğru tahmin oranının yeterince yüksek olmadığını belirtmiştir.

Başarılı olan çalışmalardan birinde ise Khan ve Khilya (2015) ortaokul öğrencilerinin geçme notunu harf sistemiyle sınıflandırmış ve % 85 başarı oranı elde etmiştir. Agrawal ve Mavani (2015) öğrenci performanslarının tahmininde sınıflandırma için makine öğrenmesi yöntemlerinden yapay sinir ağları kullanmışlardır. Deneyler aynı veri setinin farklı kayıt sayıları ile yapılmıştır. En yüksek kayıt sayısı en verimli tahmin sonucunu vermiştir. Rao ve arkadaşları (2016) öğrenci performansını tahmin etmek için makine öğrenme yöntemleri olan karar tabloları(DT), Naive Bayes sınıflandırma ve rastgele orman algoritmalarını kullandığı çalışmasında en iyi tahmin sonucunu veren rastgele orman algoritması olduğunu tespit etmiştir. V. Vijayalakshmi ve K. Venkatachalapathy (2019) öğrenci performans tahmini için farklı algoritmalar kullanarak R programı ile modeli eğitmiş ve en başarılı sonuç veren algoritmanın %84 doğruluk oranı ile derin sinir ağı(DNN) olduğunu belirtmiştir. Çalışmalarında karar

ağaçları, destek vektör makineleri, Naive Bayes, rastgele orman, k en yakın komşu, derin sinir ağı gibi birçok algoritma kullanmışlardır.

Ortaöğretim öğrencilerinin bilgileriyle oluşturulan veri seti ile çalışan Gök (2017), çalışmasında Türkçe ve matematik ders notlarını tahmin etmeye çalışmış ve matematik dersi için doğrusal regresyon, Türkçe dersi için rastgele orman algoritmalarının en başarılı algoritmalar olduğunu gözlemlemiştir.

Kim ve arkadaşları (2018) Udacity kurs sitesi için derin öğrenme yöntemleri kullanarak ders performansını tahmin etmeye çalışmışlar ve GritNet derin öğrenme algoritmasıyla öğrenci performansının en iyi tahmin sonucuna ulaşmışlardır.

Tosunoğlu ve arkadaşları (2021) makine öğrenmesi yöntemleriyle eğitim bilimleri alanında yazılmış makaleleri incelemişlerdir. İncelemelerini 2015-2020 yılları arasında yayınlanmış 201 makale ile gerçekleştirmişlerdir. Makalelerde en fazla kullanılan algoritmanın %22,5 oranı ile karar ağaçları algoritması olduğunu tespit etmişlerdir.

Güvenç ve arkadaşları (2022) Bilişim Sistemleri Mühendisliğine Giriş dersini alan öğrencilerin akademik başarısını tahmin etmek için farklı makine öğrenme yöntemleri denemişler, fakat 71 öğrenciye ait veri seti örneklem yetersizliğinden dolayı başarılı sonuç elde edememişlerdir. Bunun üzerine örneklem arttırmak amacıyla Sentetik Azınlık Örneklem Arttırma Yöntemi (SMOTE) ile veri setini 300 öğrenci kaydına çıkartmış ve en başarılı algoritmanın doğru tahmin başarısını %60 dan %97,87 oranına k en yakın komşu algoritmasıyla yükseltmişlerdir. Koca ve Adem (2022) yaptıkları blok tabanlı kodlama araçlarının kullanılması ile ilgili çalışmalarında özyeterlik inançlarını derin öğrenme ve makine öğrenmesi yöntemleri ile sınıflandırmışlardır. Kullandıkları veri setinde 190 bilişim öğretmenine ait 39 öznitelik bulunduğu ve bu veriler yetersiz olduğu için sentetik azınlık örneklem arttırma yöntemi ile 262 örnek oluşturularak WEKA programında karar ağaçları, rastgele orman, K-Star, çok katmanlı algılayıcı(MLP), Naive Bayes, destek vektör makineleri, k en yakın komşu ve evreşimsel sinir ağıları(CNN) algoritmaları ile sınıflandırmışlar ve en yüksek doğruluk oranının % 99.30 ile evreşimsel sinir ağıları olduğunu belirtmişlerdir. Yavuzarslan ve Erol (2022) 93 öğrenciye ait 10 haftalık log kayıtları ile yaptıkları çalışmada farklı algoritmaları kullanarak öğrenci başarısını tahmin etmeye çalışmışlar fakat veri setinin verimli çalışmaması sebebiyle sentetik azınlık örneklem arttırma yöntemi, yukarı örnekleme ve sınıf örnekleme yöntemleri ile yeni veri seti türevleri oluşturarak

çalışmayı yenilemişlerdir. Yeni oluşturdukları bu veri seti türevleri ile yaptıkları çalışmalarda %80 üzeri başarıya ulaşmışlardır.

Salameh ve ark. (2022) yaptıkları çalışmada 2 kademeli(geçti-kaldı) sınıflandırmayı MATLAB(MATrix LaBoratory) programında gerçekleştirmiş ve öznitelik seçme algoritmalarından gelişmiş ikili genetik algoritmasını(EBGA) ve ikili genetik algoritmasını(BGA) kullanmışlardır. Çalışmalarında k en yakın komşu, karar ağaçları, Naive Bayes, destek vektör makineleri ve gizli Dirichlet modelleme Latent Dirichlet allocation) yöntemini, Naive Bayes algoritmasının ve gelişmiş ikili genetik algoritma öznitelik seçme yönteminin, %0,87 oranında en başarılı yöntemler olduğunu belirtmişler.

Kayalı ve Buyrukoğlu (2022) öğrencilerin akademik başarılarını tahmin etmek için yaptıkları çalışmada ileri seçim, varyans analizi(Anova), gömülü(embedded) öznitelik seçim yöntemlerini ve karar ağaçları, rastgele orman, destek vektör makineleri algoritmalarını uygulayarak 2 veri seti ile toplam 24 farklı deney yapmışlardır. Deneylerin sonucunda 1.veri setinde rastgele orman algoritmasında tüm öznitelikler kullanılarak elde edilen %84 olan doğruluk oranını ileri seçim teknikleriyle öznitelik seçilerek %86'ya yükseltildiğini; 2. veri setinde ise rastgele orman algoritması ve destek vektör makineleri algoritmalarıyla bütün öznitelikler kullanılarak %82 olan doğruluk oranının %90'a çıkarıldığını belirtmişlerdir.

2.1. Geliştirilen Çalışma

Yapılan bu çalışmada daha önceki çalışmalardan farklı olarak benzer iki veri seti ile 48 algoritma denenmiş ve en verimli sonuç alınan 8 algoritma seçilmiştir. Ayrıca aynı veri seti ve aynı algoritmalar ile 2 farklı yöntem kullanılarak deneyler tekrarlanmış ve 8 algoritma 3 farklı şekilde çalıştırılmıştır.

Öznitelik seçerek yani öznitelikleri azaltarak çalışmadaki deneylerin verimliliğini artırıp arttırmayacağını test etmek için yapılan işlemlerin hepsi öznitelik seçilerek tekrarlanmıştır.

Ayrıca öğrenci akademik başarısını sınıflandırmada not verilerinin etkinliğini test etmek amacıyla veri setlerinden kademeli olarak not bilgileri çıkartılarak veri setleri türetilmiştir. Türetilen bu veri setlerine tüm deneyler uygulanmış ve sonuçlar kıyaslanmıştır.

Birçok çalışmanın aksine aynı anda hem öznitelik seçerek hem de topluluk öğrenme algoritmalarını kullanarak deneyler derinleştirilmiştir.

3. YÖNTEM VE ALGORİTMALAR

Makine öğrenme yöntemleri tek başına veya birleştirilerek farklı şekillerde uygulanabilir. Bir algoritma yalın halde kullanılabildiği gibi birkaç algoritma aynı anda veya sıralı şekilde de çalıştırılabilir. Daha verimli sonuçlar elde etmek için farklı yöntemler geliştirilmiştir.

3.1. Makine Öğrenmesi ve Sınıflandırma

Makine öğrenmesi insan beynini taklit eder. İnsanlar öğrendiklerinden ve tecrübelerinden yola çıkarak karşılaştıkları durumlar ile ilgili çıkarımlarda bulunması gibi makine öğrenmesi de daha önce kendisine sunulan verilerden istenilen durumlar için tahminlerde bulunabilir. İnsanlar durumları değerlendirirken duyguları öngörülerini etkilerken makineler daha rasyonel sonuçlar üretebilir. Ayrıca makineler çok daha büyük veriler ile çalıştığı için daha doğru tahminler üretebilmektedir (İnik & Ülker, 2017). Makine öğrenmesi insan zekâsına benzetilerek tasarlanmış, algoritmalarla tahmin sonuçlarını hesaplayan bir veri bilimi dalıdır (El Naqa & Murphy, 2015). Makine öğrenme algoritmaları büyük verileri hızlı bir şekilde analiz eden, modelleyen ve sonuç olarak bir tahmin üreten algoritmalarlardır. Verilerin belirli bir kısmı eğitim için algoritmaya sonucuyla beraber sunulur ve algoritmanın verinin geri kalanına dair tahmin üretmesi sağlanır.

Bu çalışmada basit sınıflandırma yöntemlerinin yanı sıra birleşik yöntemler de test edilerek, sınıflandırma tahminini iyileştirilmek için çalışılmıştır. Farklı sınıflandırma algoritmaları önce klasik yöntemler ile daha sonra torbalama ve yükseltme yöntemleri ile test edilmiştir. Kullanılan yöntemler ile ilgili genel bilgi aşağıda verilmiştir.

3.1.1. Klasik makine öğrenme yöntemleri

Öğrenme, önceden sınıfı belli olan veri kümelerini kullanarak veriler arasındaki örüntüyü keşfetmektir (Alpaydın, 2009). Her algoritmanın öğrenme ve sınıflama yöntemi farklıdır. Klasik makine öğrenme yöntemlerinde algoritmalar tekil kullanılmaktadır. Yapı olarak tek bir algoritma ile öğrenme işlemi yapılır. Bazı algoritmalar verileri tek bir düzende işlerken, bazı algoritmalar çıktılarını tekrar girdi olarak alır ve süreci baştan başlatarak ve bu işlemi birkaç defa tekrarlayarak en iyi sonuçlara ulaşmaya çalışır.

Bu çalışmanın ilk aşamasında birçok makine öğrenme algoritması denenmiş ve en verimli 8 algoritma seçilmiştir.

3.1.2. Torbalama yöntemi

Klasik makine öğrenme yöntemlerinde tek bir algoritma ile işlemler yapılırken torbalama yönteminde algoritmalar bir arada veya sıralı bir şekilde uygulanabilir. Torbalama(Bagging) algoritması, Breiman tarafından önerilen topluluk sınıflandırıcı için bir yöntemdir (Breiman, 1996). Torbalama terimi, önyükleme (bootstrap) ve toplama/birleştirme (aggregating) terimlerinin oluşturduğu bir kısaltmadır (Uyanık & Kasapbaşı, 2021). Bu yöntem tek bir eğitim setinden farklı eğitim setleri oluşturarak başarıyı arttırmayı hedefler. Eğitim setinden %63 e yakın bir değerde gerçek örnek alınır, eğitilen bu setler diğer verilere uygulanır. Sonuçlar birleştirilir (Kökcü et al., 2014).

Eğitim-test şeklinde ayrılan verilerden eğitim verilerini, tek algorithmada farklı parametreler seçilerek veya farklı öğrenme algoritmaları kullanarak, heterojen oluşan topluluk öğrenme modellerinden biri torbalamadır (Wang et al., 2018).Torbalama algoritması bir topluluk öğrenme metodudur. Farklı örneklemeler ile eğitim setini eğitirken farklı algoritmaları birleştirir. Bu şekilde birçok sonuç elde edilir. Elde edilen bu sınıflandırma sonuçlarından çoğunluk çıktı olarak alınır (Onan, 2018). Öğrenme setleri çoğunluk oyu alan, eğitilmiş sistem ile tekrar çalıştırılır; sonuçlar yenilenir, yeni ve son çıktılar sonuç olarak sunulur (Breiman, 1996).

3.1.3. Yükseltme yöntemi

Yükseltme işlemi zayıf eğitilmiş sistemleri birleştirerek güçlü bir eğitilmiş sistem oluşturur. Yükseltme işlemlerinin temel amacı sınıflandırıcıları eğitmektir (Güzel, 2020). Genel olarak karar ağaçları kullanılarak, sınıflandırmayı tahmin eden modellerdir. Çok gürültülü verilerde verimli bir şekilde kullanılır (Aydın & Aslan, 2017). Eğitim hatalarını azalttığı için genellemeyi sağlıklı yapar. Eksik verilerin olduğu durumlarda da uygun çalışabilir (Song et al., 2011).

Yükseltme yöntemi ile çalışan algoritmalarından biri de Adaboost(Adaptive Boosting) algoritmasıdır. Adaboost bir topluluk öğrenme algoritmasıdır. Farklı ve düşük sonuç veren algoritmaları birleştirerek karma ve daha yüksek sonuç veren eğitim modeli oluşturur. Hangi algoritmaların daha iyi sınıflandırma yapacağını; karşılaştırarak veya algoritmaları birleştirerek belirler. En iyi sonucu eğitim modeli olarak seçer (Freud & Schapire, 1997).

Bu algoritmanın diğer topluluk öğrenme yöntemlerinden daha fazla tercih edilme sebebi daha hızlı olması ve hafızada daha az yer kaplamasıdır. Daha kullanılabilir bir

algoritmadır. Bu yöntemde her olasılık için örnekleri seçerken yanlış çıkan seçimlere önem verilir. Her tekrar için örneklerin seçimi güncellenir ve yanlış seçimlere odaklanılır. Bu yöntem ağaç sınıflandırıcılarına göre daha başarılı uygulanabilir (Estevez et al., 2009). Bu seçimlerden ve sınıflandırıcılardan bir topluluk oluşturulur. Bu zayıf sınıflandırıcılardan hata oranı en az olanlar seçilerek güçlü sınıflandırıcılar oluşturulur. Bu güçlü sınıflandırıcılara dâhil olamayan seçimler silinir. Güçlü sınıflandırıcılar eğitim olarak kullanılır. Adaboost algoritması aynı anda sınıflandırma ve öznitelik seçme işlemlerini yapar (Gao et al., 2010). Kolektif sınıflandırıcıların geneli tekil yöntemlere göre daha başarılıdır. Adaboost çok kullanılan bir yükseltme algoritmasıdır (Şimşek et al., 2012).

3.2. Sınıflandırma Algoritmaları

Bu çalışmada farklı algoritmalar denenerek farklı çalışma alanları için en iyi sonuç veren 8 algoritma ele alınmıştır.

3.2.1. Derin öğrenme

Derin öğrenme, biz insanların bilgileri öğrenme şeklini taklit ederek öğrenen bir yapay zekâ ve makine öğrenme sistemidir. İstatistik kullanılarak ve tahminde bulunarak modelleme yapan bir veri bilimidir (Krizhevsky et al., 2012). Derin öğrenme sisteminde çalışılan verilerin çoğu yapılandırılmış, etiketlenilmiş ve tablo haline getirilmiş olur. Bunun nedeni, sistemi eğitmek için girdilerle beraber çıktılara da ihtiyaç duymasıdır (Onar, 2020).

Derin öğrenme yapısında en az 3 veya daha fazla katman bulunur. Bu katmanlarda hesaplanan modeller soyutlanarak veri temsilleri öğrenilir. Bu süreçte her katmandan çıkan sonuç yani geri dönüşümle nasıl bir algoritma oluştuğu bulunur. Büyük veri kümelerinde bu işlemler her katman için uygulanarak karmaşık yapılar çözülmüş olur. Eğitilen sistemlerden elde edilen algoritmalar kullanılarak yeni veriler için öngörüler oluşturulur (IBM, 2021). Derin öğrenme algoritmaları görüntü işleme, metin ve konuşma işleme gibi alanlar dışında sınıflandırma işlemlerinde de kullanılır (Doğan & Türkoğlu, 2019).

DeepLearning4J, çok katmanlı algılayıcılarla sınıflandırma ve regresyon yapmak için kullanılan bir derin öğrenme algoritma kütüphanesidir. Çeşitli “arff” dosyaları ile uyumlu çalıştığı için bu çalışmada Weka programında “DeepLearning4JMLp” kullanılmıştır (WEKA, 2022).

3.2.2. Yapay sinir ağıları

Bir yapay sinir ağı modeli olan çok katmanlı algılayıcı yapay sinir ağlarının önemli bir temelidir (Işıkhah, 2020). Denetimli öğrenme algoritmalarından biridir. Sisteme hem giriş hem de çıkış verileri verilir ve algoritmanın öğrenmesi beklenir (Orhan, 2023). Sınıflandırma için eşik değeri bulunur ve sınıflandırma bu değere göre oluşturulur. Bu değer probleme göre belirlenir. Tekrar sayısı artırılarak öğrenme güçlendirilir (Dede, 2020). Çok katmanlı algılayıcı, “XOR” problemlerini çözmek amacıyla geliştirilmiştir. Hata yayma modeli de denilir. Delta öğrenme metodunu kullanır (Azgınoğlu, 2018). Girişler çok olduğu zaman yeterli nöron olmadığında katmanla işleme devam edilir. Giriş ve çıkış katmanları arasında da gizli katmanlar bulunur. Giriş katmanına gelen verileri ara katmanlara iletir. Probleme göre ara katman sayısı değişir. Bir katmandan çıkan veriler bir sonraki ara katmana iletilir. Sistemdeki çıkış sayısı, çıkış katmanındaki nesne sayısıdır (Bayır, 2006).

Bu algoritmada matematiksel fonksiyonlardan biri kullanılabilir. En çok tang, sigmoid, threshold, lineer ve hard limiter kullanılır. Ayrıca çok katmanlı algılayıcıda delta öğrenme kuralı kullanılır. Ara eğitim seti giriş(input) ve çıkış(output) olarak verilir (Işıkhah, 2020).

Bu modelin geri yayımlı yapay sinir ağı modeli de vardır. Giriş değerlerinden çıkış değerlerini elde edecek fonksiyonu kullanarak sınıflandırma yapar. Önce ileri doğru hesaplanır sonra aynı işlem geri doğru yapılarak hesaplanır. İleri doğru hesaplamada veriler istenilen fonksiyona göre hesaplanarak çıkış elde edilir. Eğitim verilerinde elde edilen çıkış verileri ile olması gereken veriler arasında farklılık varsa hata olduğu anlaşılır. İkinci aşama geriye doğru hatayı dağıtmaya çalışmaktır. İlk olarak rastgele verilen ağırlık değerleri her tekrar da değiştirilerek yok edilmeye çalışılır (Azgınoğlu, 2018). Çok katmanlı algılayıcıların çok geniş kullanım alanları vardır. Örneğin siber güvenlik, büyük verilerin analizi gibi alanlarda yaygın kullanılmaktadır.

3.2.3. Basit lojistik regresyon

Basit lojistik regresyon(SL) veri setinde ikili bir sonucu tahmin etmek için kullanılan istatistiksel bir yöntemdir. Sonuçlar veri setindeki bağımsız değişkenler arasındaki bağlantıları analiz ederek bağımlı değişken için bir değer bulmayı hedefler. Bulunan bu değer ile kesin bir yargı oluşturulur (McDonald, 2022).

Basit lojistik regresyonda birden fazla giriş değişkeni kullanılabilir. Bu değişkenlerden okunan değerler ile daha önceki sonuçlar kıyaslanarak yeni girdiler için

olasılıklar hesaplanarak bir sonuç üretilir (Terzi, 2023). Basit lojistik regresyon, etiketli bir değişken ve bir ölçüm değişkeni varsa ve ölçüm değişkeninde yapılacak değişimlerin etiketli değişken üzerinde değişikliklere neden olup olmayacağını öğrenmek istediğimiz durumlarda kullanılır. Örneğin iki değer alan bir değişken varsa (E/H, var/yok vb.) bu etiketli değişkendir. Ölçüm değişkeni bağımsız, etiketli değişken bağımlı değişkendir. Bu durumda bağımlı değişkendeki değeri tahmin etmek için bağımsız değişkendeki değer aralıkları önemlidir (Statstates, 2022). Birden fazla bağımsız değişken kullanılabilir. Ölçüm değerleri birbirinden bağımsız olur, yani bir ölçüm değeri diğerini etkilemez. Her bir bağımsız değişken için üretilen bağımlı değişken değeri, birbirinden bağımsız ele alınarak tek bir değer elde edilir (Terzi, 2023).

Basit lojistik regresyonun doğrusal regresyondan farkı bağımlı değişkenin etiketli bir değer olmasıdır. Kullanım amaçlarından biri bağımsız değişkenin bağımlı değişkeni etkileyebilmek için alacağı değer aralığını görmektir. Diğer amacı da etiketli değişkenin (bağımlı değişken) alacağı değeri tahmin etme olasılığıdır (Terzi, 2023).

Basit lojistik fonksiyonda sonuç 2 değer arasında kesin bir kaniya varır. Değişken ya da değişkenler birbirini etkilemez. Tek bir etiketli değer üretilir (Statstates, 2022).

3.2.4. Yinelemeli sınıflandırıcı optimize edici

Yinelemeli algoritmalar her bir adım için komşu düğümlerden elde edilen tahminleri kullanır (Manikandan et al., 2018). Bu algoritmada veri setindeki ilk kaydı sınıflandırıldıktan sonra bulunan hatalar geri besleme olarak kullanılmakta ve bu şekilde değiştirilmektedir. Bu şekilde bütün çıktılar bir sonraki işlemin iyileştirilmiş girdileri olur. Bu yapı beyin sinir ağı gibi tanımlanabilmektedir. Kayıtlar ağda tekrarlanarak sistem eğitilir (Fan & Huang, 2011). Bu sınıflandırıcı çapraz doğrulama yapan bir algoritma için en iyi tekrar sayısını seçer (Vural, 2017).

Algoritma ilk aşamada bir sınıflandırma yapar. Çıkan sonuçlardaki elde edilen hataları tekrar sisteme girdi olarak verir ve sınıflandırma tekrarlanır. Bu yinelemeli işlemde her adım bir önceki adımın sonuçlarıyla beslendiği için düğümler birbirini etkiler. Tekrarlanan işlemler sistemi eğitir (Fan & Huang, 2011).

3.2.5. Karar tablosu

Birden fazla koşulu olan durumlar için kullanılan karar algoritmasıdır. Tabloda satır ve sütunlarda koşullar, kesişim yerlerinde ise eylemler bulunur. Tablo şeklinde programlanan kural – mantık düzenidir. Bir satır ve ya sütun ekleyerek yeni kurallar eklenebilir. Bir eyleme birden fazla koşulla ulaşılabilir ya da tam tersi bir koşulla birden

fazla eylem yönlendirilebilir. Ortaya çıkan eylemlere göre eğitilen algoritma, sınıflandırma yapar (Vanthienen, 1992), (IBM, 2021). Karar tabloları basittir ve kolay yorumlanabilir. Geliştirmek kolaydır. Etkili türevleri oluşturulabilir ve test etmek için uygulama alanı geniştir. Çok karmaşık veriler bile karar ağacına kolay dönüşebilir (Hamilton, 2023).

3.2.6. Bulanık sırasız kural tümevarım algoritması

Bu algoritma bulanık kural ya da kural esnetme yaklaşımını kullanır. Keskin olmayan sınırlar içeren sınıflandırmalarda yani kesin sınırları olmayan iki sınıf arasında farklı sınıflar var ise kullanılır (Salem, 2018). Bir veri setinde iyi-orta-kötü, çok iyi-iyi-orta-geçer-kalır gibi durumların olduğu sınıflandırmalar için kullanılması uygundur. Keskin sınıflandırma kurallarından daha avantajlıdır. Sınıflar arasında keskin sınırlar olmadığı zamanlarda yani bulanık mantığın temeli olan yumuşak geçişli sınırları olan sınıflandırmalarda verimli çalışır. Sınıflandırma yaparken sonuçlar keskin sınırlara dönüştürülür (Salem, 2018).

Bu çalışmada bulanık sırasız kural tümevarım algoritması olarak WEKA programında FURIA(Bulanık Kural Tümevarım algoritması) kullanılmıştır. FURIA kural eğitimcisi olarak Ripper algoritmasının geliştirilmiş ve aynı zamanda basitleştirilmiş halidir. Farklı olarak bulanık ve sırasız kuralları içerir (Hühn & Hüllermeier, 2009).

3.2.7. Tek kural

Tek kural algoritması genellikle veri madenciliğinde etkilidir ve makine öğrenmesinde de kullanılır. Tek aşamalı bir karar ağacı gibi çalışır. Basit bir çalışma yapısı vardır. Hedef sınıfı en iyi tahmin eden özniteliği seçmeye çalışır. Hata verilerinden kural çıkarımı yapar. Yani ayırt edici özniteliklerden en iyisini alır ki bu öznitelik en az hatalı sonucu verendir; buna göre kurallar belirler (Atmaca, 2017).

Veri setleri düşünüldüğünde sonuç sınıflandırmasına en az hata veren özelliği alır ve sonuç ile özellik arasında bir kural oluşturur. Bütün kayıtları bu oluşan kurala göre sınıflandırır(devamsızlığı 10 günden az olanlar başarılıdır gibi) (Uzun, 2023). Tek seviyeli karar ağacı gibi değerlendirilebilir. Her durum için bir kural çıkartılır. Kurallar hata tabanlı belirlenir. Çıkarımlar hataları iyileştirmek yoluyla yapılır (Holte, 1993).

3.2.8. Lojistik model ağacı

Lojistik model ağacı algoritması karar ağacı indüksiyonu ve lojistik regresyon algoritmalarının birleştirilerek oluşturulduğu bir algoritmadır (Karadağ, 2014). Denetimli sınıflandırma modelidir. Bu yöntem de daha önce model olarak bulunan ağaç yapısına dayanır. Yapraklarında sabitleri olan karar ağaçlarından, parçalı doğrusal bir regresyon modeli sağlanır (Landwehr, 2003). Lojistik kısmında logitboost algoritması kullanılarak her düğümde bir model oluşturulur. Düğüm belirli yöntemlerle parçalanır. Her kademenin başlangıcı bir üst düğümdeki sonuca göredir. Son aşamada ağaç budanır (Landwehr et al., 2005). Lojistik model ağacı çapraz doğrulama kullanarak eğitim verileriyle uyuşmayan tekrarlarını bulur. Bunun için farklı hızlarda çalışan çeşitleri bulunur (Sumner et al., 2005).

Ağaç yapısı her alt düğüme geçerken logitboost algoritmasıyla lojistik regresyon oluşturur. Bu işlemde düğümlerdeki bilgiler alt yapraklara da taşınarak parçalanır. Parçalama yöntemi C4.5 kıstasıdır (Aksu & Karaman, 2017). Ağaç budanırken çapraz doğrulama yaparak sadeleşir ve sınıflandırmaya uygun hale gelir. Bu veri setinde olduğu gibi hem sayısal hem sayısal olmayan değerlerin sınıflandırılmasında lineer ve ağaç tümevarım modelleri başarılıdır (Aksu & Karaman, 2017).

3.3. Öznitelik Seçme Algoritmaları

Weka programında 8 algoritma kullanılarak öznitelik seçme işlemi yapılmıştır.

3.3.1. Bilgi kazanç özelliği değerlendirme

Bilgi kazançlarını ölçerken sınıfa göre ölçerek öznitelikleri değerlendirir. Eksik veri varsa bu verileri farklı bir değer olarak görür veya değerlere dağıtırken orantılı bir şekilde dağıtır. Her özniteliği bulduğu değere göre sıralama yapar (Witten et al., 2011).

3.3.2. Korelasyon tabanlı öznitelik altkümesi değerlendiricisi

Tahmin edeceği sınıf ile arasında korelasyon değeri yüksek olduğu halde birbirleriyle yüksek değerde ilişki olmayan özniteliklerin altkümesini belirler. Belirlenen bu alt kümedeki öznitelikleri sıralar (Witten et al., 2011).

3.3.3. Sınıflandırıcı altküme değerlendirme

Bir sınıflandırıcı yardımıyla öznitelikleri değerlendirir. Parametre olarak nesne düzenleyicide sınıflandırıcı belirlenir ve ayrı bir veri kümesinde veya eğitim verileri üzerinde öznitelik kümeleri değerlendirilir (Witten et al., 2011).

3.3.4. Korelasyon öznitelik değeri

Sınıf ile bir öznitelik arasındaki korelasyon değerini Pearson korelasyonunu kullanarak ölçer ve özniteliğe değer olarak alır. Bu değerler sıralanır. Etiketli özniteliklerin aldığı değerler birer gösterge olarak değerlendirilir. Bu etiketli özniteliklerin korelasyon değeri bulunurken ağırlıklı ortalama kullanılır.

3.3.5. Kazanç oranı özelliği değerlendirmesi

Özniteliğin değeri bulunurken kazanç oranını sınıfa göre ölçer. Bu kazanç oranına göre sıralanır (Witten et al., 2011).

3.3.6. Tek kural özniteliği değerlendirmesi

Tek kural sınıflandırıcısının kullandığı basit doğruluk ölçütünü kullanır. Parametre olarak katlama sayısı alınır. Tek kural algoritması gibi basit kesikli hale getirme yöntemini kullanır (Witten et al., 2011).

3.3.7. F özniteliğini atarak değerlendirme

Örnekleme tabanlı bir yaklaşımla öznitelikleri sıralar. Önce rastgele örneklemeler yaparak farklı ve aynı sınıflara sahip komşu örneklemeleri kontrol eder. Parametreleri bakılacak komşu sayısı, örnekleme sayısı, komşulara ağırlık ile mesafe verilip verilemeyeceği gibi parametrelerdir. Bu parametrelere göre sıralama yapar (Witten et al., 2011).

3.3.8. Simetrik belirsiz öznitelik değerlendirmesi

Sınıfa göre simetrik belirsizlik ölçütünü kullanarak bir özniteliğin değerini belirler. Bu değere göre sıralama yapar (Witten et al., 2011).

3.4. Öznitelik Sıralama Metotları

Öznitelik seçme algoritmalarında seçilen öznitelikler belirli metotlarla sıralanır. Öznitelik seçme yöntemine göre sıralama metodu Weka programında seçilir.

3.4.1. En iyi en önce sıralama metodu

Bütün alt kümeleri, tam kümeden geriye doğru çıkararak, boş kümeden ileriye doğru tepe tırmanışı yaparak veya orta bir noktadan başlayıp her iki yöne doğru arayarak belirlenir. Belirlenen altkümeler önbellege alınır. Verimlilik açısından önbelletteki boyutu bir parametredir. Genellikle öznitelik altküme yöntemi en iyi ilk prensibine göre sıralanır. Nedeni tek bir yönde arama yapmadığı için aynı alt kümeye bağlı kalmamasıdır. Geriye dönüş yaparak öznitelik ekleyip çıkartabilir (Witten et al., 2011).

3.4.2. Ağgözlü adım sıralama metodu

Öznitelik altkümelerini kapsamlı tarar. En iyi en önce sıralama gibi boş alt kümelerden ileriye doğru veya tam kümeden geriye doğru arama yaparak gidebilir. Fakat en iyi en önce sıralama metodu gibi geriye dönerek arama yapamaz. Kalan özniteliklerden en iyi özniteliği ekleyerek veya silerek değerlendirme metriğini azaltır azaltmaz sonlandırır (Witten et al., 2011).

3.4.3. Sıralı metodu

Bütün öznitelikleri bireysel olarak değerlendirmelerine göre sıralar. Bir eşik değeri belirleyerek atılması veya korunması gereken öznitelik sayısını belirler. Daha düşük sırada olanları kaldırarak öznitelik seçer. Aldığı değere göre kalan öznitelikleri sıralar (Witten et al., 2011).

3.5. Sınıflandırmada Kullanılan Metrikler

Algoritmaların sınıflandırma başarısını ölçerken bazı metriklere bakılarak karar verilir. Bu çalışmada kullanılan metrikler doğruluk oranı(DO), kesinlik(Kes), duyarlılık(Duy), F1-değeri (F1-Skor)'dir.

Bu metrikler hesaplanırken algoritmaların ürettikleri tahminlerin doğru-yanlış sayıları oranlanarak bulunur.

Bir tahmin işleminde gerçekte olan değerler ve tahmin edilen değerler bulunur. Basit gösterimi Tablo 3.1'deki gibidir.

Bu değerler doğru ve yanlış tahminler olabilir. Sonuçlar doğru tahmin edilmiş pozitif-gerçek pozitif(GP), doğru tahmin edilmiş negatif-gerçek negatif(GN), yanlış tahmin edilerek doğru işaretlenmiş negatif – yanlış negatif(YN) ve yanlış işaretlenmiş pozitif- yanlış pozitif(YP) olabilir.

Tablo 3.1. Metriklerde Kullanılan Örüntü

	Tahmin Edilen Değerler		
		Pozitif	Negatif
Gerçek Değerler	Pozitif	Gerçek Pozitif(GP)	Yanlış Pozitif(YP)
	Negatif	Yanlış Negatif(YN)	Gerçek Negatif(GN)

Sınıflandırmada en fazla kullanılan metrik doğruluk oranıdır. Doğruluk oranı doğru tahmin edilen sınıfların bütün tahminlere oranıdır. Doğruluk oranı bütün tahminlerde bizim doğru tahmin ettiğimiz sınıf yüzdesini verir.

$$DO = \frac{GP+GN}{GP+GN+YN+YP} \quad (3.1)$$

Denklemden de görüldüğü üzere doğru tahmin edilmiş sınıfların oranı olduğu için sınıflandırmada önemli bir başarı ölçütüdür. Negatif veya pozitif doğru tahminleri hesaplayarak algoritmaların verimliliğini gösterir.

Kullandığımız diğer bir metrik ise kesinlik değeridir. Kesinlik ise gerçek pozitiflerin bütün pozitiflere oranını verir. Algoritma verilerden bir sınıf belirlediğinde bu sınıfta pozitif bulunduğu değerlere bakar. Doğru tahmin ettiği olumlu sınıf sayısının, toplam olumlu sınıf sayısına oranına denir. Bu oran yukarıdaki Tablo 3.1'deki parametreler kullanılarak şu şekilde hesaplanabilir.

$$Kes = \frac{GP}{GP+YP} \quad (3.2)$$

Denklemden görüldüğü üzere bulunmak istenen sınıfa dahil olmayan veriler denkleme alınmamakta, sadece pozitif sınıf verileri ele alınmaktadır.

Bir diğer metrik ise duyarlılıktır. Duyarlılık bir sınıflandırma yaparken o sınıfa dâhil olan verilerden ne kadarını doğru tahmin ettiğimizi gösterir. Bu da doğru tahmin edilen sınıf verilerinin sayısının gerçekte o sınıfa ait olan veri sayısına oranıdır. Yani doğru sınıfı ne kadar tahmin edebildiğimizi orantısal olarak gösterir. Denklemi aşağıdaki gibidir.

$$Duy = \frac{GP}{GP+YN} \quad (3.3)$$

F-1 değeri ise bu çalışmada kullanılan son metriktir. Bu değer bulunan kesinlik ve duyarlılık değerlerinin harmonik ortalamasını verir. Yanlış pozitif ve yanlış negatifinde etkili olması için harmonik ortalama alınır.

$$F - 1 = 2 * \frac{kesinlik*duyarlılık}{kesinlik+duyarlılık} \quad (3.4)$$

F-1 değeri sınıflandırmadaki hataların etkisini ölçmek üzere hesaplanan bir değerdir.

3.6. Sınıflandırmada Kullanılan Program

Bu çalışmada deneyler WEKA programıyla gerçekleştirilmiştir. WEKA programı Waikato Üniversitesi'nin geliştirmiş olduğu bir programdır ve java tabanlıdır. Ayrıca açık kaynak kodlu bir programdır.

İçerisinde farklı sınıflandırma, kümeleme, regresyon, öznelik seçme gibi farklı algoritmalar bulunur. Farklı dosya formatlarını desteklemektedir. Bu çalışmada arff formatı kullanılmıştır (Coşkun & Abdullah, 2011), (WEKA, 2022).

Bu çalışmada WEKA programında sınıflandırma ve öznelik seçme işlemleri yapılmıştır.

3.7. Sınıflandırmada Kullanılan Yöntemlerin Seçenekleri

Klasik yöntemlerin sonuçları alındıktan sonra torbalama yöntemi uygulanmıştır. Weka programında torbalama işlemleri daha önce seçilen 8 algoritma ile gerçekleştirilmiştir. Weka programında torbalama işlemleri aşağıdaki Tablo 3.2'deki seçimler yapılarak uygulanmıştır.

Tablo 3.2. *Weka programında torbalama işlem ayarları*

Seçenekler	Seçimler
Torba boyutu yüzdesi	100
Parti boyutu	100
Sınıflandırıcı	Kullanılan bütün algoritmalar
Hata ayıklamak	Hayır
Kapasiteyi kontrol etme	Hayır
Ondalık Basamak Sayısı	2
Yineleyici Sayısı	10
Ağırlıkları kullanarak kopyaları temsil etme	Hayır
Düğüm	1
Torba dışı tahminleri sakla	Hayır

Torbalama işlemlerinden sonra aynı deneyler yükseltme yöntemi ile gerçekleştirilmiştir. Yükseltme işlemleri için her algoritmanın verimli çalışmadığı gözlemlenmiş ve WEKA programında yükseltme deneyleri Adaboost algoritması ile yapılmıştır. Tablo 3.3'da Adaboost ile yükseltme yapılırken seçilen özellikler gösterilmektedir.

Tablo 3.3. *Weka programında yükseltme işlem ayarları*

Seenekler	Seimler
Torba boyutu yüzdesi	100
Parti boyutu	100
Sınıflandırıcı	Kullanılan bütün algoritmalar
Hata ayıklamak	Hayır
Kapasiteyi kontrol etme	Hayır
Ondalık Basamak Sayısı	2
Yineleyici Sayısı	10
Ağırlıkları kullanarak kopyaları temsil etme	Hayır
Düğüm	1
Torba dışı tahminleri sakla	Hayır

4. VERİ SETLERİ VE ÖN İŞLEMLERİ

Çalışmada kullanılan veri setleri ham ve gerçek bilgileri içerir. Deneyleri gerçekleştirmek için bazı işlemler yapılması gerekir.

4.1. Veri Seti Tanıtımı

Bu çalışmada 2 farklı veri seti ve türevleriyle çalışılmıştır. Veri setleri açık kaynak kodlu sunulmaktadır. (Cortez & Silva, 2008). Bu setler benzer veriler içermektedir ve Portekiz’de ortaöğretim öğrencilerinin matematik ve Portekiz dili derslerini içeren veri setleridir. Bu veri setlerinde öğrencilerin başarısını etkileyen kısıtlara uygun anket sorularına verdikleri cevaplar, notlar ve bir başarı sınıfı bulunmaktadır. Birinci veri setinde(Set 1) öğrenci bilgileri, matematik dersi notları ve başarı sınıfı, ikinci veri setinde(Set 2) ise yine öğrenci bilgileri, Portekiz Dili dersi notları ve başarı sınıfı bulunmaktadır.

Set 1 matematik dersi için 208 kadın, 187 erkek toplam 395 öğrenci, Set 2 ise Portekiz Dili dersi için 383 kadın, 266 erkek toplam 649 öğrenci kaydı içermektedir. Veri setlerinde 33 öznitelik bulunmaktadır.

Setlerde genel olarak fiziksel şartlarını, aile sosyal-ekonomik düzeylerini, aile yaşantılarını, öğrenci çalışma disiplinlerini içeren alanlardan oluşan bir anket niteliğindedir. Çalışmada bu alanlara verdikleri cevaplar ile elde ettikleri başarı oranları doğru tahmin edilmeye çalışılmıştır. Mahremiyeti sağlamak ve öğrencilerin anket sorularına baskı altında hissetmeden cevap verebilmeleri için öğrencilerin kimlik veya herhangi bir ayırt edici bilgileri kullanılmamaktadır. Veri setlerinin öznitelikleri aşağıda bulunan Tablo 4.1’de verilmiştir. Bu tablo içeriğinde özniteliklerin adı, açıklaması ve alabileceği değerler gösterilmiştir.

Tablo 4.1. Veri setlerindeki öznitelikler, açıklaması ve alacağı değer aralıkları

Öznitelik adı	Açıklaması	Alabileceği değer aralıkları
Cinsiyet	öğrencinin cinsiyeti	ikili: kadın veya erkek
Yaş	öğrencinin yaşı	sayısal: 15'ten 22'ye kadar
Okul	öğrencinin okulu	ikili: Gabriel Pereira veya Mousinho da Silveira
Adres	öğrencinin ev adresi türü	ikili: kentsel veya kırsal
Aile durumu	ebeveynin birlikte yaşama durumu	ikili: birlikte veya ayrı yaşayan
Anne eğitim	annenin eğitimi	sayısal: 0'dan 4'a'ya kadar
Anne mesleği	annenin işi	etiketli: at_home, health, services, other, teacher

Tablo 4.1. (Devam) *Veri setlerindeki öznitelikler, açıklaması ve alacağı değer aralıkları*

Öznitelik adı	Açıklaması	Alabileceği değer aralıkları
Baba mesleği	babasının işi	etiketli: at_home, health, services, other, teacher
Veli	öğrencinin velisi	etiketli: anne, baba veya diğer
Aile boyutu	ailede birey sayısı	ikili: 3 veya > 3
Destek	ekstra eğitim okul desteği	ikili: evet veya hayır
Aile desteği	aile eğitim desteği	ikili: evet veya hayır
Etkinlikler	müfredat dışı etkinlikler	ikili: evet veya hayır
Dışarı çıkma	arkadaşlarla dışarı çıkmak	sayısal: 1 - çok düşükten 5 - çok yüksek
Hafta sonu alkol	hafta sonu alkol tüketimi	sayısal: 1 - çok düşük ila 5 - çok yüksek
Hafta içi alkol	hafta içi alkol tüketimi	(sayısal: 1 - çok düşük ila 5 - çok yüksek)
Sağlık	mevcut sağlık durumu	sayısal: 1 - çok kötü ila 5 - çok iyi
Devamsızlık	okulda devamsızlık sayısı	sayısal: 0'dan 93'e kadar
Not1	birinci dönem notu	sayısal: 0'dan 20'ye kadar
Not2	ikinci dönem notu	sayısal: 0'dan 20'ye kadar
Final notu	final notu	sayısal: 0'dan 20'ye kadar

4.2. Veri Seti Ön İşleme

Veri setlerinde aile bilgileri, öğrenci zaman kullanımı, önceki yıllarda başarı ya da başarısızlıkları gibi bilgilerin yanı sıra 0 ile 20 arasında değerlendirme puanı olan iki sınav notuyla final sınav notu bulunmaktadır. Öğrenci başarısını tahmin etmek için final sınavı kullanılmıştır. Portekiz’de sınıf başarısı final sınav notu ile belirlenir. İlk aşamada sınıflar sadece iki seviyeli(geçti-kaldı) başarı sınıfları oluşturulmuştur. Bu sınıflar başarı kademelerini belirtmektedir. İkinci aşamada ise beş seviyeli(çok iyi, iyi, orta, geçer, kalır şeklinde) başarı sınıfı oluşturulmuştur (Cortez & Silva, 2008).

Veri setindeki final notuna göre not baremi ve sınıflar aşağıdaki tablolarda gösterilmiştir. Bu sınıflandırma bu veri setlerini kullanan Cortez ve Silva (2008)’nın makalesinden elde edilmiştir. Tablo 4.2’de iki kademeli geçme kalma sınırları gösterilmiştir.

Tablo 4.2. *Notların sınıflandırılması / not baremi 2’li sınıflandırma sistemi*

Ülke Düzey	Geçer	Kalır
Portekiz	10-20	0-9

Aşağıdaki Tablo 4.3’te veri setinin hazırlandığı Portekiz’de 5 kademeli not sisteminin sınırları gösterilmiştir.

Tablo 4.3. *Notların sınıflandırılması / not baremi 5'li sınıflandırma sistemi*

Ülke	Çok iyi	İyi	Orta	Geçer	Kalır
Portekiz	16-20	14-15	12-13	10-11	0-9

Veri setlerine final sınav notuna göre başarı sınıfı eklendikten sonra final notu çıkartılarak “excel” formatındaki veriler “arff” dosya formatına dönüştürülerek deneylere hazır hale getirilmiştir.

4.3. Veri Setlerinde Özniteliklerin Belirlenmesi

Deneylerin ikinci aşaması en etkili öznitelikler belirlenerek yapılmıştır. Öznitelik seçimi WEKA programında öznitelik seçme algoritması kullanılarak yapılmıştır. Bu algorithmada setin tamamı eğitim amaçlı kullanılmıştır. Weka programında öznitelik seçimi yapan 11 algoritma ve 3 metot bulunmaktadır. Bazı metotlar her algoritma ile çalışırken bazıları sadece belirli algoritmalarla çalışmaktadır. Uygun algoritmaları bulmak için hepsi denenerek öznitelik seçimi yapılmıştır.

Tablo 4.4. *Weka programında öznitelik seçme algoritmaları*

Attribute Evaluator
InfoGainAttributeEval
CfsSubsetEval
ClassifierSubsetEval
CorrelationAttributeEval
GainRatioAttributeEval
OneRAttributeEval
ReliefFAttributeEval
SymmetricalUnsetAttributeEval

Aşağıdaki tablolarda torbalama ve yükseltme yöntemleri için WEKA programında sunulan seçimler ve deneyleri gerçekleştirirken uygulanan seçimler aşağıdaki tablolarda verilmiştir.

Aşağıdaki Tablo 4.5'te ise öznitelikler seçilirken kullanılan arama metotları listelenmiştir. Her algoritma belirli bir metot ile çalışmaktadır. En verimli algoritmayı tespit edebilmek için, bu metotlar her algorithmada denendikten sonra en verimli sonuç veren metotlar kullanılmıştır.

Tablo 4.5. *Weka Programında Öznitelik Seçmek için Arama Metotları*

Search Method
BestFirst
GreedyStepwise
Ranker

Yukarıdaki tablolarda verilen seçimler ile öznitelik seçimi iki veri setine, 3 türevi için de uygulanmış, veri setlerinde farklı öznitelikler verimli olarak belirlenmiştir. Deneyler bu öznitelikler ile yapılmıştır. Belirtilen öznitelik seçme yöntemleri denenerek her tablo için farklı öznitelikler belirlenmiş ve bu öznitelikler çalışma adımlarında listelenmiştir.

5. DENEYSEL ÇALIŞMALAR

Bu çalışmada bu 2 veri seti için 8 farklı algoritma kullanılarak sonuçlar kıyaslanmış ve bu tür veriler içeren veri setlerinde hangi algoritmanın daha verimli olacağını belirlemek ve öğrenci başarı düzeylerini en yüksek oranda tahmin etmek hedeflenmiştir. Deneylerde kullanılan algoritmalar bu 2 veri setine de uygulanmış, sonuçları iyileştirmek için en verimli öznitelikler seçilerek tekrar denenmiştir.

Yaptığımız çalışmalarda 3 farklı yöntem ile 8 farklı algoritma 2 deney setine 2 farklı sınıflandırma ile uygulanmıştır. İlk uygulamada bütün öznitelikler dâhil edilmiş, ikinci uygulamada ise özniteliklerin bazıları çıkartılarak deney tekrarlanmıştır. Öznitelik seçimi WEKA programı aracılığıyla yapılmış ve en etkili öznitelikler seçilmeye çalışılmıştır. Seçilen öznitelikler çalışma adımlarında verilmiştir.

5.1. İki Sınıflı Sınıflandırma Sonuçları

Çalışmanın bu kısmında sınıflandırma geçti-kaldı şeklinde iki sınıflı yapılmaktadır. Önce matematik dersi, sonra Portekiz Dili dersi veri setleriyle sınıflandırma yapılmıştır.

5.1.1. Matematik dersi 2 sınav notu içeren deney sonuçları

Bu deneyde matematik veri setinden sadece final notu çıkartılarak 2 seçenekli sınıf(geçti-kaldı) tahmin edilmeye çalışılmıştır. Deneyler 3 yöntem(klasik-torbalama-yükseltme) ve 8 algoritmayla tüm öznitelikler ve seçilen öznitelikler kullanılarak uygulanmış ve yapılan deneylerin sonuçları Tablo 5.1’de verilmiştir.

Daha önce bahsedilen öznitelik seçme yöntemleri ile belirlenen öznitelikler aşağıdaki gibidir.

- Not2(G2)
- Not1(G1)
- Başarısızlıklar(failures)
- Dışarı çıkma(goout)
- Yüksek öğrenim(higher)

Bu deneyde seçilen özniteliklere bakıldığında ara sınav notlarının ve geçmiş dönem başarısızlıkları gibi okul başarısı; gelecekte yükseköğrenim hedefleri ve dışarı çıkma süresi gibi ders çalışmayı etkileyen özniteliklerin sınıflandırmada etkili olduğunu göstermektedir. Genel olarak değerlendirildiğinde matematik dersi başarısının sosyal ve ekonomik çevreden ziyade öğrencinin ders çalışması ile bağlantılı bir durum olduğu düşünülebilir.

Tablo 5.1. Matematik dersi final notu çıkartılmış bütün sonuçlar

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL	87,595	0,899	0,876	0 8 9	89,114	0,906	0,891	0,893
	MLP	87,342	0,873	,87	0,873	88,101	0,880	0,881	0,880
	SL	92,152	0,922	0,922	0,922	91,899	0,924	0,919	0,920
	ICO	91,646	0,920	0,916	0,917	91,139	0,914	0,911	0,912
	DT	90,886	0,913	0,909	0,910	91,392	0,919	0,914	0,915
	FURIA	87,848	0,878	0,878	0,878	91,392	0,916	0,914	0,915
	One-R	91,899	0,924	0,919	0,920	91,899	0,924	0,919	0,920
	LMT	92,658	0,927	0,927	0,927	91,646	0,921	0,916	0,918
Torbalama Yöntemleri	DL	86,835	0,897	0,868	0,872	88,101	0,897	0,881	0,884
	MLP	88,354	0,883	0,884	0,883	89,620	0,896	0,896	0,896
	SL	92,152	0,921	0,922	0,921	90,633	0,908	0,906	0,907
	ICO	91,392	0,918	0,914	0,915	91,392	0,918	0,914	0,915
	DT	91,899	0,922	0,919	0,920	91,646	0,921	0,916	0,918
	FURIA	91,899	0,920	0,919	0,919	91,899	0,920	0,919	0,919
	One-R	91,899	0,924	0,919	0,920	91,899	0,924	0,919	0,920
	LMT	91,646	0,916	0,916	0,916	90,127	0,901	0,901	0,901
Yükseltme Yöntemleri	DL	-	-	-	-	-	-	-	-
	MLP	87,342	0,873	0,873	0,873	89,367	0,895	0,894	0,894
	SL	90,633	0,907	0,906	0,906	91,899	0,923	0,919	0,920
	ICO	91,899	0,919	0,919	0,919	90,886	0,909	0,909	0,909
	DT	89,367	0,893	0,894	0,893	88,861	0,888	0,889	0,888
	FURIA	87,848	0,878	0,878	0,878	91,392	0,916	0,914	0,915
	One-R	91,139	0,912	0,911	0,911	90,633	0,910	0,906	0,907
	LMT	90,127	0,901	0,901	0,901	88,861	0,889	0,889	0,889

Bu veri setini oluşturarak ve kullanarak benzer çalışmaları yapan Cortez ve Silva (2008), yaptıkları çalışmada bir naif tahminci algoritması ile % 91,9 doğruluk oranında bir başarıya ulaşmışlardır. Yine aynı veri setinin bu parçası ile çalışmalar yapan Başer ve arkadaşları (2020) çalışmalarında % 92,41 doğruluk oranı ile en başarılı algoritma olarak yinelemeli sınıflandırma iyileştiriciyi belirlemişlerdir.

Bu çalışmada en verimli sonuç, 10 kat çaprazlama tekniğiyle bir ağaç algoritmasıyla bulunmuştur. Çalışmada yüksek doğruluk oranına ulaşılması veri setinde 2 adet sınav bilgisinin bulunması ile mümkün olmuştur.

5.1.2. Matematik dersi 1 sınav notu içeren deney sonuçları

Çalışmanın bu bölümünde matematik dersi veri setinden 2.sınav ve final notları çıkartılarak deneyler tekrarlanmış ve sınıf geçme durumuna göre başarı sınıfı tahmin edilmiştir. Yine bu çalışmada da bütün deneyler tekrarlanarak Tablo 5.2'deki sonuçlara göre en iyi yöntem ve en iyi algoritma tespit edilmeye çalışılmıştır.

Tablo 5.2. Matematik dersi final ve 2.sınav notları çıkartılmış 2 sınıflı sonuçlar

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL	82,785	0,840	0,828	0,831	83,03	0,838	0,830	0,833
	MLP	79,494	0,795	0,795	0,79	4,810	0,846	0,848	0,847
	SL	84,304	0,842	0,843	0,842	84,051	0,840	0,841	0,840
	ICO	83,798	0,836	0,838	0,833	82,532	0,830	0,825	0,827
	DT	80,253	0,799	0,803	0,799	84,557	0,845	0,846	0,845
	FURIA	82,532	0,822	0,825	0,823	84,304	0,840	0,843	0,840
	One-R	83,798	0,843	0,838	0,840	83,798	0,843	0,838	0,840
	LMT	84,304	0,842	0,843	0,842	84,051	0,840	0,841	0,840
Toralama Yöntemleri	DL	82,785	0,842	0,828	0,831	84,304	0,848	0,843	0,845
	MLP	82,279	0,822	0,823	0,822	85,570	0,855	0,856	0,855
	SL	84,304	0,842	0,843	0,842	83,544	0,835	0,835	0,835
	ICO	83,798	0,837	0,838	0,838	83,544	0,835	0,835	0,835
	DT	83,291	0,832	0,833	0,833	84,051	0,840	0,841	0,840
	FURIA	83,038	0,829	0,830	0,830	84,051	0,838	0,841	0,839
	One-R	82,025	0,830	0,820	0,823	82,025	0,830	0,820	0,823
	LMT	84,304	0,844	0,843	0,844	84,304	0,841	0,843	0,841
Yükseltme Yöntemleri	DL	-	-	-	-	46,835	0,560	0,468	0,479
	MLP	79,241	0,792	0,792	0,792	84,557	0,844	0,846	0,844
	SL	85,317	0,852	0,853	0,852	84,051	0,840	0,841	0,840
	ICO	80,760	0,807	0,808	0,807	82,532	0,824	0,825	0,825
	DT	79,494	0,795	0,795	0,795	84,810	0,848	0,848	0,848
	FURIA	82,532	0,822	0,825	0,823	84,304	0,840	0,843	0,840
	One-R	83,544	0,833	0,835	0,834	83,038	0,828	0,830	0,829
	LMT	82,532	0,825	0,825	0,825	83,291	0,834	0,833	0,833

Bu aşamada önceki deneye göre sınıf tahmininin başarı oranı düşmüştür. Sınav bilgilerinin azaltılması, başarı sınıfını bulmanın not bilgilerine bağlı olduğu makine

öğrenmesini zayıflatmıştır. %85,570 olan en yüksek başarı oranını, torbalama yöntemiyle öznitelik seçilerek çok katmanlı algılayıcı algoritması vermiştir.

Sınav notunun çıkartılmasıyla farklı öznitelikler sınıflandırmada etkili olmuştur. Veri setinin bu bölümünde öznitelikler aynı yöntemleri kullanarak aşağıdaki gibi elde edilmiştir.

- not1(G1)
- başarısızlıklar(failures)
- yüksek öğrenim(higher)

Benzer çalışmaları yapan Cortez ve Silva (2008) çalışmalarının bu bölümünde en yüksek doğruluk oranına % 83,8 ile yine naif tahminci algoritmasıyla elde etmişlerdir.

Seçili özniteliklere bakıldığında yine birinci deneydeki gibi önceki yıllardaki başarı durumu ve gelecek hedefleri etkilidir. Bu da ders çalışma durumunun bu derste etkili olduğunu göstermektedir.

Torbalama algoritması klasik yöntemle göre tek başına doğru sınıflandırma yüzdesini arttıramazken, öznitelikleri azaltarak torbalama yapıldığında daha verimli sonuç bir yapay sinir ağı algoritması olan çok katmanlı algılayıcı algoritması ile elde edilmiştir. Aynı algoritma klasik yöntemle çalıştırıldığında diğer algoritmalarından çok daha düşük bir performans sergilerken, öznitelik seçimi ile diğer algoritmalara yakın sonuç üretmiştir.

5.1.3. Matematik dersi sınav notu içermeyen deney sonuçları

Çalışmanın bu aşamasında not bilgilerinin tamamı çıkartılarak sistemin tahmini daha da zorlaştırılmıştır. Herhangi bir sınav notu içermeyen özniteliklerle doğru sınıf tahmininin doğruluk oranı çok düşüktür. Sonuçlar Tablo 5.3'te gösterilmiştir.

Sistemin karmaşıktığı çalışmalarda, basit bir yapıya sahip olan klasik yöntemler daha verimli çalışabilir. Çalışmanın bu aşaması bu duruma bir örnektir.

Seçilen öznitelikler şöyledir:

- başarısızlıklar(failures)
- sebep(reason)
- destek(schoolsup)
- dışarı çıkma(goout)
- yüksek öğrenim(higher)
- anne mesleği(mjob)

Bu çalışmada ise farklı algoritmalar ile bu oran % 2,03 oranında arttırılmıştır. Fakat not bilgisi içermediği için tahmin oranı düşmüş; seçili öznitelik sayısı artmıştır. Not bilgisinin bulunmadığı durumlar, özniteliklerin en etkili olduğu durumlardır. Okul seçme nedeni de yine okul başarısı ile bağlantılı bir durumdur.

Tablo 5.3. Matematik dersi bütün sınav notları çıkartılmış 2 sınıflı sonuçlar

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL	69,114	0,675	0,691	0,678	71,392	0,698	0,714	0,682
	MLP	66,582	0,658	0,666	0,661	65,823	0,633	0,658	0,639
	SL	70,633	0,692	0,706	0,658	71,899	0,711	0,719	0,676
	ICO	72,911	0,716	0,729	0,707	72,152	0,706	0,722	0,699
	DT	69,367	0,671	0,694	0,667	71,139	0,693	0,711	0,684
	FURIA	68,101	0,651	0,681	0,644	72,152	0,706	0,722	0,700
	One-R	70,886	0,690	0,709	0,681	70,886	0,690	0,709	0,681
	LMT	70,633	0,692	0,706	0,658	71,899	0,711	0,719	0,676
Torbalama Yöntemleri	DL	68,608	0,672	0,686	0,676	71,139	0,693	0,711	0,687
	MLP	65,823	0,638	0,658	0,643	67,089	0,647	0,671	0,650
	SL	69,114	0,666	0,691	0,661	71,899	0,708	0,719	0,681
	ICO	71,899	0,703	0,719	0,695	72,405	0,710	0,724	0,697
	DT	68,608	0,657	0,686	0,641	71,139	0,695	0,711	0,677
	FURIA	68,608	0,660	0,686	0,656	71,392	0,697	0,714	0,689
	One-R	71,392	0,697	0,714	0,683	71,646	0,701	0,716	0,687
	LMT	70,886	0,691	0,709	0,690	68,101	0,655	0,681	0,654
Yükseltme Yöntemleri	DL	60,000	0,556	0,600	0,570	50,127	0,559	0,501	0,516
	MLP	65,570	0,642	0,656	0,647	67,848	0,660	0,678	0,663
	SL	68,608	0,660	0,686	0,656	71,646	0,707	0,716	0,674
	ICO	67,342	0,649	0,673	0,652	71,139	0,694	0,711	0,680
	DT	66,076	0,647	0,661	0,652	70,127	0,680	0,701	0,672
	FURIA	68,101	0,651	0,681	0,644	72,152	0,706	0,722	0,700
	One-R	69,367	0,674	0,694	0,675	70,886	0,694	0,709	0,666
	LMT	64,557	0,633	0,646	0,638	62,279	0,607	0,623	0,613

Tabloda da görüldüğü üzere önceki deney sonuçlarına göre düşük olan doğruluk oranı %72,911 ile yinelemeli sınıflandırıcı optimize edici algoritma ile sağlanmıştır. Öznitelik seçimi veya farklı yöntemlere başvurulması bu oranı arttırmada yeterli olmamıştır.

Diğer deneylerden farklı olarak bu deneyde anne mesleği etkili öznitelik olarak görülmektedir. Anne mesleği öğrencinin çalışma düzenini, alışkanlıklarını, kısacası başarı durumunu etkileyen bir özniteliktir.

Çalışmalarımıza referans aldığımız Cortez ve Silva (2008) deneylerinde en başarılı algoritmanın destek vektör algoritması olduğunu ve % 70,6 doğruluk oranına sahip bir sınıflandırma yaptıklarını belirtmişlerdir.

5.1.4. Portekiz Dili dersi 2 sınav notu içeren deney sonuçları

Çalışmanın bu aşamasında veri seti değiştirilerek Portekiz Dili dersi notlarını içeren ikinci veri seti kullanılmıştır. Aynı özniteliklere sahip olan fakat farklı veriler içeren bu veri seti ile yapılan çalışmada sınav notları dâhil olduğu için başarı tahmin oranı yüksektir. Aşağıdaki Tablo 5.4’de bu veri seti ile elde edilen sonuçlar gösterilmiştir. Bu çalışmada kullanılan seçilmiş öznitelikler şöyledir:

- not2(G1)
- not2(G2)
- başarısızlıklar(failures)
- yüksek öğrenim(higher)
- okul(school)

Seçili özniteliklere bakıldığında ders çalışmasını etkileyen faktörlerden(geçmiş başarısızlıklar ve yükseköğrenim hedefleri) farklı olarak okul özneliği bulunmaktadır. Okulların sosyal yapısı, bulunduğu yerin kırsal veya merkezi olması, öğrencilerin aile yapıları, okuldaki öğrencilerin başarı düzeyleri okul kültürünü oluşturmaktadır. Okul kültürü ise okuldaki öğrencilerin günlük hayatta kullandıkları konuşma dilini, hatta bu dildeki eğitimi, dolaylı olarak da dil derslerinin başarısını etkilemektedir. Çok kitap okuyan, günlük hayatta düzgün bir konuşma dili kullanan öğrenciler, dil derslerinde daha başarılı olabilmektedir.

Tabloda görüldüğü üzere bu çalışmada aynı doğruluk oranına sahip birkaç algoritma tespit edilmiştir. Bu algoritmalarından tek kural algoritması farklı yöntemlerle kullanıldığında da aynı sonucu vermiştir. Hem öznitelik seçildiğinde hem de torbalama yöntemi ile aynı değeri göstermektedir. Ayrıca lojistik model ağacı ve yinelemeli sınıflandırıcı optimize edici algoritmaları da aynı doğruluk oranına ulaşmıştır. Ayrıca matematik dersine oranla Portekiz Dil dersinin başarı sınıfını tahmin etmek daha verimli

sonuçlanmıştır. Bu durum, ülkede konuşulan dile ait ders notlarının çevresel faktörler gibi veri setinde bulunan özniteliklere daha bağımlı olmasından kaynaklanmaktadır.

Tablo 5.4. Portekiz Dili dersi final notu çıkartılmış 2 sınıflı sonuçlar

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL	90,909	0,910	0,909	0,910	91,834	0,919	0,918	0,919
	MLP	89,676	0,894	0,897	0,895	93,374	0,932	0,934	0,932
	SL	93,220	0,929	0,932	0,929	93,220	0,929	0,932	0,930
	ICO	93,374	0,932	0,934	0,932	92,912	0,927	0,929	0,927
	DT	92,604	0,923	0,926	0,922	93,220	0,930	0,932	0,928
	FURIA	93,220	0,929	0,932	0,929	92,296	0,919	0,923	0,919
	One-R	93,683	0,935	0,937	0,933	93,683	0,935	0,937	0,933
	LMT	93,683	0,935	0,937	0,933	93,220	0,929	0,932	0,930
Torbalama Yöntemleri	DL	91,063	0,909	0,911	0,910	92,142	0,920	0,921	0,921
	MLP	91,063	0,905	0,911	0,906	92,912	0,927	0,929	0,927
	SL	92,296	0,920	0,923	0,921	93,529	0,933	0,935	0,933
	ICO	93,683	0,935	0,937	0,935	93,374	0,932	0,934	0,933
	DT	93,220	0,930	0,932	0,928	92,604	0,922	0,926	0,922
	FURIA	93,220	0,930	0,932	0,930	93,220	0,930	0,932	0,931
	One-R	93,683	0,935	0,937	0,933	93,683	0,935	0,937	0,933
	LMT	92,758	0,925	0,928	0,926	92,296	0,920	0,923	0,921
Yükseltme Yöntemleri	DL	-	-	-	-	63,790	0,739	0,638	0,679
	MLP	89,676	0,894	0,897	0,895	93,220	0,930	0,932	0,930
	SL	93,374	0,931	0,934	0,932	93,220	0,929	0,932	0,930
	ICO	91,680	0,916	0,917	0,916	93,683	0,935	0,937	0,935
	DT	91,680	0,913	0,917	0,915	92,296	0,922	0,923	0,922
	FURIA	93,220	0,929	0,932	0,929	92,604	0,922	0,926	0,923
	One-R	93,066	0,929	0,931	0,929	93,066	0,928	0,931	0,929
	LMT	92,142	0,920	0,921	0,921	90,909	0,905	0,909	0,906

Doğruluk oranını %93.683 olarak sonuç alınan algoritmalar tek kural, lojistik model ağacı ve yinelemeli sınıflandırma iyileştirici algoritmaları olmuştur. Cortez ve Silva (2008) yaptıkları çalışmalarında yine bir ağaç algoritması olan karar ağacı algoritması ile % 93 oranına ulaşmışlardır.

5.1.5. Portekiz Dili dersi 1 sınav notu içeren deney sonuçları

Bu çalışmanın ikinci aşamasında Portekiz Dili dersi veri setinden ikinci sınav notu da çıkartılarak deneyler tekrarlanmıştır. Yapılan deneylerden aşağıdaki Tablo 5.5'teki sonuçlar elde edilmiştir.

Tablo 5.5. Portekiz Dili dersi final ve 2.sınav notları çıkartılmış 2 sınıflı sonuçlar

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL4MLP	88,598	0,882	0,886	0,884	89,368	0,890	0,894	0,892
	MLP	87,211	0,866	0,872	0,868	89,368	0,886	0,894	0,888
	SL	91,217	0,907	0,912	0,908	90,755	0,901	0,908	0,902
	ICO	91,988	0,918	0,920	0,918	91,988	0,918	0,920	0,918
	DT	91,988	0,918	0,920	0,918	92,142	0,919	0,921	0,920
	FURIA	90,909	0,905	0,909	0,906	91,834	0,916	0,918	0,917
	OneR	91,988	0,918	0,920	0,918	91,988	0,918	0,920	0,918
	LMT	91,217	0,907	0,912	0,908	90,755	0,901	0,908	0,902
Torbalama Yöntemleri	DL4MLP	89,985	0,896	0,900	0,897	89,831	0,894	0,898	0,896
	MLP	88,444	0,876	0,884	0,879	89,831	0,893	0,898	0,895
	SL	89,676	0,890	0,897	0,892	90,447	0,898	0,904	0,898
	ICO	91,988	0,918	0,920	0,918	91,834	0,916	0,918	0,917
	DT	90,601	0,900	0,906	0,902	91,525	0,912	0,915	0,913
	FURIA	90,755	0,903	0,908	0,905	91,834	0,916	0,918	0,917
	OneR	91,988	0,918	0,920	0,918	91,988	0,918	0,920	0,918
	LMT	89,368	0,890	0,894	0,892	89,676	0,893	0,897	0,895
Yükseltme Yöntemleri	DL4MLP	-	-	-	-	-	-	-	-
	MLP	87,057	0,865	0,871	0,867	89,214	0,886	0,892	0,888
	SL	90,909	0,903	0,909	0,904	90,755	0,901	0,908	0,902
	ICO	87,673	0,874	0,877	0,875	90,755	0,907	0,908	0,907
	DT	89,060	0,888	0,891	0,889	91,217	0,913	0,912	0,913
	FURIA	90,909	0,905	0,909	0,906	91,834	0,916	0,918	0,917
	OneR	88,290	0,874	0,883	0,877	91,525	0,911	0,915	0,912
	LMT	87,827	0,876	0,878	0,877	88,598	0,879	0,886	0,882

Önceki çalışmanın aksine aynı değeri verebilen başka algoritma bulunmamaktadır.

Bu çalışmada derin öğrenme algoritması ile sınıflandırma yapılamamıştır. Tabloda görüldüğü üzere veri setinden sınav notu çıkartılınca sınıflandırmada ki başarı oranı düşmüştür. En yüksek doğruluk oranına sahip olan algoritma öznitelikler

azaltılarak uygulanan ve % 92,142 doğruluk oranı ile karar tablosu olmuştur. Öznitelik seçme yöntemleri ile elde edilen öznitelikler şunlardır:

- not1(G1)
- başarısızlıklar(failures)
- yüksek öğrenim(higher)
- okul(school)
- anne meslek(mjob)

Sınav notlarından birinin çıkartılmasıyla etkili öznitelikler çoğalmasına rağmen, sınıflandırma başarısı düşmüştür. Yine bu deneyde de okul kültürünün ve anne mesleğinin etkisi görülmektedir.

Çalışmalarını referans aldığımız Cortez ve Silva (2008) yaptıkları çalışmada bu deneyde %90,1 doğruluk oranı ile rastgele orman algoritmasını en başarılı sınıflandırıcı olarak belirtmişlerdir.

5.1.6. Portekiz Dili dersi sınav notu içermeyen deney sonuçları

Çalışmanın bu bölümünde bütün not bilgileri çıkartılarak deneyler tekrarlanmış ve not içermediği için düşen doğruluk oranı ile elde edilen sonuçlar aşağıdaki Tablo 5.6'da verilmiştir.

Seçilen öznitelikler şunlardır.

- başarısızlıklar(failures)
- okul(school)
- anne eğitim(medu)
- yüksek öğrenim(higher)

Not bilgisi içermeyen dil dersi sınıflandırma tahmininde anne eğitimi etkili öznitelik seçilmiştir. Bu durum okul kültürü kadar aile kültürünün de dil dersi başarısında etkili olduğunu göstermektedir. Öğrenciler alışkanlıklarını önce aileden, sonra sınıf ve okul ortamından almaktadırlar. Bu durumda ilk öğretmenleri aileleri; bebeklik ve çocukluk dönemlerinde en fazla zaman geçirdikleri için en çok da anneleri olmaktadır. Annelerin eğitimi çocukların yetişmesinde en önemli faktörlerdendir. Zamanla bu etki azalıp, öğrenciler arkadaş ve sosyal çevrelerinden etkilense de çocukluk döneminde iyi eğitim almış gençler olumsuz etkenlerden en az etkilenenlerdir. Bu durum okul başarısını da etkilemektedir.

Tablo 5.6. *Portekiz Dili dersi bütün sınav notları çıkartılmış 2 sınıflı sonuçlar*

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL	85,516	0,837	0,855	0,843	85,054	0,825	0,851	0,831
	MLP	82,435	0,817	0,824	0,821	84,746	0,823	0,847	0,829
	SL	85,824	0,833	0,858	0,826	85,824	0,833	0,858	0,826
	ICO	85,054	0,819	0,851	0,822	84,900	0,814	0,849	0,815
	DT	85,054	0,825	0,851	0,831	86,133	0,841	0,861	0,844
	FURIA	86,133	0,840	0,861	0,841	86,441	0,845	0,864	0,848
	One-R	84,284	0,784	0,843	0,787	84,746	0,806	0,847	0,799
	LMT	85,670	0,834	0,857	0,838	85,516	0,828	0,855	0,828
Torbalama Yöntemleri	DL	85,824	0,838	0,858	0,842	85,362	0,826	0,854	0,829
	MLP	84,592	0,824	0,846	0,830	85,208	0,830	0,852	0,836
	SL	85,362	0,827	0,854	0,831	85,824	0,832	0,858	0,829
	ICO	84,592	0,811	0,846	0,815	85,978	0,836	0,860	0,835
	DT	85,516	0,833	0,855	0,808	84,746	0,819	0,847	0,825
	FURIA	85,670	0,830	0,857	0,830	85,208	0,833	0,852	0,839
	One-R	84,592	0,796	0,846	0,786	84,438	0,788	0,844	0,785
	LMT	84,438	0,826	0,844	0,832	83,975	0,814	0,840	0,823
Yükseltme Yöntemleri	DL	-	-	-	-	-	-	-	-
	MLP	82,435	0,811	0,824	0,817	84,438	0,819	0,844	0,826
	SL	85,824	0,833	0,858	0,826	85,824	0,833	0,858	0,826
	ICO	84,746	0,837	0,847	0,842	86,133	0,840	0,861	0,842
	DT	85,208	0,835	0,852	0,841	85,978	0,839	0,860	0,843
	FURIA	86,133	0,840	0,861	0,841	86,441	0,845	0,864	0,848
	One-R	85,516	0,830	0,855	0,833	85,670	0,831	0,857	0,833
	LMT	83,513	0,821	0,835	0,827	83,359	0,811	0,834	0,819

Aynı veri seti ile benzer çalışmaları yapan Cortez ve Silva (2008) çalışmalarının bu bölümünde % 85 doğruluk oranı ile yine rastgele orman algoritmasının en başarılı algoritma olduğunu belirtmişlerdir.

5.2. Beş Sınıflı Sınıflandırma Sonuçları

Çalışmanın bu aşamasında sınıflandırma 5 kademeli yapılır. Başarı durumuna göre yapılan sınıflandırmada başarısız öğrenci kaldı, başarılı öğrencilerin başarı kademeleri gösterilir.

5.2.1. Matematik dersi 2 sınav notu içeren deney sonuçları

Çalışmanın ikinci aşamasında sınıflandırma sistemi değiştirilerek final notuna göre beş kademeli sınıflandırmaya geçilmiştir. Beş kademeli sınıflandırmada sınıflar arasında geçişler daha hassas olduğu için doğruluk oranları bir miktar düşük olacaktır.

İlk deneyimiz matematik dersi veri seti ile sadece final notu çıkartılarak yapılmış ve deney sonuçları aşağıdaki Tablo 5.7'deki gibidir.

Tablo 5.7. Matematik dersi final notu çıkartılmış 5 sınıflı sonuçlar

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL	52,658	0,517	0,527	0,505	53,165	0,513	0,532	0,474
	MLP	53,165	0,523	0,532	0,526	71,392	0,714	0,714	0,714
	SL	71,646	0,716	0,716	0,715	72,658	0,727	0,727	0,725
	ICO	77,975	0,785	0,780	0,777	76,962	0,773	0,770	0,768
	DT	77,722	0,782	0,777	0,775	78,481	0,791	0,785	0,782
	FURIA	75,190	0,753	0,752	0,752	77,468	0,778	0,775	0,772
	One-R	78,481	0,791	0,785	0,782	78,481	0,791	0,785	0,782
	LMT	71,646	0,716	0,716	0,715	71,899	0,721	0,719	0,719
Torbalama Yöntemleri	DL	50,380	0,772	0,772	0,771	55,443	0,548	0,554	0,503
	MLP	55,696	0,541	0,557	0,547	73,418	0,736	0,734	0,735
	SL	71,899	0,721	0,719	0,718	73,924	0,743	0,739	0,739
	ICO	77,215	0,772	0,772	0,771	76,203	0,764	0,762	0,760
	DT	77,975	0,784	0,780	0,777	78,228	0,788	0,782	0,780
	FURIA	77,215	0,772	0,772	0,772	77,468	0,777	0,775	0,774
	One-R	78,481	0,791	0,785	0,782	78,481	0,791	0,785	0,782
	LMT	74,177	0,746	0,742	0,741	70,127	0,698	0,701	0,699
Yükseltme Yöntemleri	DL	30,127	?	0,301	?	31,899	?	0,319	?
	MLP	56,962	0,560	0,570	0,564	71,646	0,715	0,716	0,715
	SL	69,873	0,701	0,699	0,699	72,658	0,727	0,727	0,725
	ICO	70,127	0,703	0,701	0,701	72,405	0,720	0,724	0,721
	DT	75,443	0,757	0,754	0,755	77,215	0,773	0,772	0,770
	FURIA	75,190	0,753	0,752	0,752	77,468	0,778	0,775	0,772
	One-R	78,481	0,791	0,785	0,782	78,481	0,791	0,785	0,782
	LMT	74,430	0,747	0,744	0,744	68,608	0,684	0,686	0,685

Öznitelik azaltılarak yapılan deneylerde kullanılan öznitelikler ise şöyledir:

- not2(G2)
- not1(G1)

- başarısızlıklar(failures)
- anne mesleği(mjob)
- dışarı çıkma(goout)

5 kademeli sınıflandırma deneyinde en başarılı algoritmalarda seçilen özniteliklerle yapılan deneyler ile öznitelik seçmeden yapılan deneyler genellikle aynı sonucu vermiştir. Bu durum seçilmeyen özniteliklerden sınıflandırmayı olumsuz etkileyen bir öznitelik bulunmadığını göstermektedir. Matematik dersi başarısı için yine ders çalışmasını etkileyen faktörler ile ara sınav notlarının etkisi görülmektedir. Yine anne mesleği, ders çalışma alışkanlıklarında etkili olduğu için ders başarısında da etkilidir.

Tabloda görüldüğü üzere öznitelik seçimi algoritmaların genelinde sonuçları etkilememiş, sınıflandırma yöntemlerine bakılmaksızın tek kural algoritması veri setinin bu türevi için %78,481 doğruluk oranıyla en iyi sınıflandırıcı algoritma olmuştur. Aynı orana karar tablosunun da öznitelik seçilerek klasik yöntemle ulaşıldığı görülmektedir.

Yine tabloda gözlemlenebilen, sınav notları içeren veri seti türevlerinde hem daha yüksek doğruluk oranına erişilmiş hem de farklı algoritma ve yöntemler aynı oranı vermiştir.

Cortez ve Silva (2008) yaptıkları çalışmanın bu aşamasında %78,5 oranla en başarılı algoritmanın naif tahminci algoritması olduğunu belirtmişlerdir. Bu çalışmada üç farklı yöntem(klasik-torbalama-yükseltme), iki farklı uygulama(öznitelik seçerek-hepsini kullanarak) ve birçok algoritma kullanarak bu oran arttırılamamış, aynı oran bulunmuştur.

5.2.2. Matematik dersi 1 sınav notu içeren deney sonuçları

Son çalışılan setten ikinci sınav notu çıkartılarak çalışıldığında aşağıdaki Tablo 5.8.'de elde edilmiştir.

Bu çalışmada ki % 61,013 doğruluk oranı torbalama yöntemi kullanılarak ve öznitelik seçilerek karar tablosu algoritması ile elde edilmiştir.

Seçili öznitelikler ise şu şekildedir.

- not1(G1)
- başarısızlıklar(failures)
- dışarı çıkma(goout)
- romantik ilişki(romantic)
- özel ders(paidclass)

Seçili öz niteliklere bakıldığında matematik dersi için sınıf kademelerini belirlemede öğrencinin ders ilgisini azaltabilecek romantik ilişki ile ders eksiklerini tamamlayabilecek özel ders öz niteliklerinin etkili olduğu görülmektedir.

Tablo 5.8. Matematik dersi final ve 2.sınav notları çıkartılmış 5 sınıflı sonuçlar

		Tüm Öz nitelikler				Seçili Öz nitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL	45,317	0,434	0,453	0,435	50,380	0,470	0,504	0,453
	MLP	46,582	0,453	0,466	0,457	53,671	0,526	0,537	0,528
	SL	55,190	0,533	0,552	0,537	58,734	0,580	0,587	0,578
	ICO	57,215	0,567	0,572	0,568	57,722	0,578	0,577	0,566
	DT	53,671	0,553	0,537	0,541	58,228	0,590	0,582	0,578
	FURIA	54,177	0,513	0,542	0,510	51,899	0,496	0,519	0,470
	OneR	57,975	0,571	0,580	0,558	57,975	0,571	0,580	0,558
	LMT	53,671	0,527	0,537	0,531	58,734	0,580	0,587	0,578
Torbalama Yöntemleri	DL	45,570	0,425	0,456	0,432	49,620	0,465	0,496	0,448
	MLP	48,354	0,458	0,484	0,468	59,494	0,590	0,595	0,591
	SL	54,937	0,535	0,549	0,541	59,241	0,589	0,592	0,586
	ICO	57,975	0,567	0,580	0,572	58,228	0,569	0,582	0,571
	DT	58,734	0,581	0,587	0,579	61,013	0,611	0,610	0,606
	FURIA	55,949	0,544	0,559	0,550	57,468	0,558	0,575	0,557
	OneR	59,241	0,580	0,592	0,564	59,241	0,580	0,592	0,564
	LMT	56,962	0,562	0,570	0,565	55,190	0,553	0,552	0,552
Yükseltme Yöntemleri	DL	27,848	?	0,278	?	27,848	?	0,278	?
	MLP	48,354	0,472	0,484	0,476	56,203	0,553	0,562	0,556
	SL	56,709	0,556	0,567	0,560	58,734	0,580	0,587	0,578
	ICO	55,443	0,543	0,554	0,548	56,709	0,562	0,567	0,563
	DT	53,671	0,539	0,537	0,537	58,987	0,596	0,590	0,590
	FURIA	54,177	0,513	0,542	0,510	51,899	0,496	0,519	0,470
	OneR	57,215	0,555	0,572	0,557	57,215	0,555	0,572	0,557
	LMT	51,392	0,512	0,514	0,513	53,671	0,531	0,537	0,533

Genel olarak bakıldığında öğrenciler kendi çabaları ile çoğu zaman dersi geçebilecek kadar başarıyı elde edebilmektedir. Fakat matematik dersi gibi sayısal yetenek ve mantık gerektiren derslerde başarı düzeylerini yüksek tutmak için desteğe ihtiyaç duyabilmektedir. Bu deneyde özel ders öz niteliğinin etkili olması başarı kademelerinin ayırt edilmesi için ihtiyaç duyulmuştur.

Çalışmada birbirine yakın sonuçlar elde edilse de en başarılı algoritma torbalama yöntemi ile öznitelikler seçilerek karar tablosu algoritması olmuştur. Sınıflandırmada ki doğruluk oranı % 61,013'tür. Bir önceki çalışmaya göre düşük olsa da Cortez ve Silva (2008)'nin çalışmaları sonucu naif tahminci algoritması ile buldukları % 60,5 doğruluk oranından biraz fazladır.

5.2.3. Matematik dersi sınav notu içermeyen deney sonuçları

Bu veri seti ile olan son çalışmamızda bütün not bilgileri kaldırılarak beş kademeli bir sınıflandırma sonuçları aşağıdaki Tablo 5.9'da gösterildiği gibi elde edilmiştir.

Tablo 5.9. Matematik dersi bütün sınav notları çıkartılmış 5 sınıflı sonuçlar

		Tüm Öznitelikler				Seçili Öznitelikler			
		D.O. %	Kes.	Duy.	F-1	D.O. %	Kes.	Duy.	F-1
Klasik Yöntemler	DL4MLP	31,899	0,318	0,319	0,317	31,646	0,252	0,316	0,278
	MLP	25,570	0,250	0,256	0,251	30,633	0,278	0,306	0,288
	SL	34,684	0,324	0,347	0,320	35,949	?	0,359	?
	ICO	36,962	0,420	0,370	0,298	34,177	?	0,342	?
	DT	27,342	0,232	0,273	0,239	31,139	?	0,311	?
	FURIA	32,658	0,276	0,327	0,285	32,911	?	0,329	?
	OneR	34,684	?	0,347	?	32,911	?	0,329	?
	LMT	32,152	0,302	0,322	0,303	35,190	?	0,352	?
Torbalama Yöntemleri	DL4MLP	32,152	0,311	0,322	0,315	35,190	0,326	0,352	0,332
	MLP	31,392	0,299	0,314	0,304	32,152	0,310	0,322	0,314
	SL	32,911	0,310	0,329	0,316	35,190	0,301	0,352	0,311
	ICO	34,937	0,322	0,349	0,324	34,684	0,311	0,347	0,311
	DT	34,177	0,313	0,342	0,269	33,671	0,289	0,337	0,300
	FURIA	33,924	0,314	0,339	0,319	31,392	0,286	0,314	0,284
	OneR	34,684	?	0,347	?	31,646	?	0,316	?
	LMT	32,911	0,319	0,329	0,323	29,367	0,293	0,294	0,293
Yükseltme Yöntemleri	DL4MLP	20,760	0,236	0,208	0,206	26,076	?	0,261	?
	MLP	29,873	0,287	0,299	0,291	33,418	0,322	0,334	0,326
	SL	34,684	0,324	0,347	0,320	35,443	0,302	0,354	0,305
	ICO	32,911	0,304	0,329	0,311	32,152	0,301	0,322	0,303
	DT	27,342	0,232	0,273	0,239	29,873	?	0,299	?
	FURIA	32,658	0,276	0,327	0,285	32,405	?	0,324	?
	OneR	32,658	?	0,327	?	31,899	0,270	0,319	0,267
	LMT	33,165	0,326	0,332	0,328	29,114	0,290	0,291	0,290

En zor çalışılan ve en düşük oranda tahmin elde edilen deney olmuştur. Not içermeyen bir veri setinde sınıflandırmayı tahmin etmenin zorluğunun yanı sıra dersin zorluğu başarı tahminini elde etmeyi daha da zorlaştırmıştır.

Bazı deneylerde kesinlik ve F-1 değerleri hesaplanamamıştır. Doğruluk oranı da en yüksek % 36,962 oranı ile yinelemeli sınıflandırıcı optimize edici algoritmayla elde edilmiştir. Bu orana sadece klasik yöntem ve bütün öznitelikler kullanılarak ulaşılmıştır.

Bu öznitelikler ise şöyledir:

- başarısızlıklar(failures)
- destek(schoolsup)
- dışarı çıkma(goout)
- yüksek öğrenim(higher)
- anne meslek(mjob)

Not bilgisi içermeyen bu deneyde seçilen etkili özniteliklerden biri de okul desteğidir. Okulda öğrencilere ihtiyaç duydukları derslerde verilen destek eğitimleri öğrencilerin başarı kademelerini etkilemektedir. Sınıflandırma tahmini yaparken de bu öznitelik belirleyici etkiye sahiptir. Ayrıca not içermeyen bu deney için 2 kademeli sınıflandırmada da etkili olan anne mesleği, dışarı çıkma sıklığı, geçmiş başarısızlıklar, yükseköğrenim hedefleri gibi öznitelikler; matematik dersi başarısını genel olarak etkilediği için bu deneyde de etkili özniteliklerdendir.

Benzer çalışmayı yapan Cortez ve Silva (2008) yaptıkları deneyler sonucu rastgele orman algoritması ile % 33,5 doğruluk oranında bir sınıflandırma yapmışlardır.

5.2.4. Portekiz Dili dersi 2 sınav notu içeren deney sonuçları

Daha önce ikili sınıflandırma yapılan, Portekiz Dili dersine ait notlar içeren bu veri seti ile 5 kademeli bir sınıflandırma çalışılmış ve sonuçlar Tablo 5.10'da verilmiştir.

Seçilen öznitelikler ise aşağıdaki gibidir.

- not1(G1)
- not2(G2)
- başarısızlıklar(failures)
- çalışma süresi(studytime)
- özel ders(Paidclass)

2 kademeli sınıflandırmada benzer deney yapılırken sosyal-kültürel çevre ile ilgili öznitelikler etkili olurken, bu deneyde ise başarı kademelerini belirlemek için öğrencinin ders çalışma durumunu belirten öznitelikler etkili olmuştur. Dil bilgisi birikimleri geçme- kalma durumunu ortaya koyabiliyorken, 5 kademeli sınıflandırmada sınıf ayrımlarını daha fazla öznitelik ile belirlemek mümkün olmaktadır. Bu deneyde de öğrencinin ders çalışma süresi ve özel ders alma durumu, yine diğer çalışmalarda da etkili olan notları ve geçmiş başarısızlıkları etkili özniteliklerdir.

Tablo 5.10. *Portekiz Dili dersi final notu çıkartılmış 5 sınıflı sonuçlar*

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL	52,851	0,528	0,529	0,526	60,247	0,580	0,602	0,577
	MLP	57,165	0,573	0,572	0,572	73,190	0,731	0,732	0,729
	SL	71,341	0,715	0,713	0,713	72,265	0,724	0,723	0,720
	ICO	74,114	0,740	0,741	0,733	75,655	0,758	0,757	0,752
	DT	74,114	0,740	0,741	0,733	76,425	0,777	0,764	0,760
	FURIA	71,341	0,711	0,713	0,708	74,422	0,745	0,744	0,737
	One-R	76,733	0,782	0,767	0,764	76,733	0,782	0,767	0,764
	LMT	71,649	0,718	0,716	0,716	72,265	0,724	0,723	0,720
Torbalama Yöntemleri	DL	52,697	0,525	0,527	0,522	60,555	0,593	0,606	0,584
	MLP	58,089	0,579	0,581	0,578	72,727	0,730	0,727	0,725
	SL	72,111	0,722	0,721	0,720	72,265	0,723	0,723	0,721
	ICO	75,963	0,760	0,760	0,756	74,576	0,745	0,746	0,741
	DT	76,425	0,776	0,764	0,760	76,425	0,777	0,764	0,760
	FURIA	71,341	0,715	0,713	0,710	75,193	0,760	0,752	0,748
	One-R	76,733	0,782	0,767	0,764	76,733	0,782	0,767	0,764
	LMT	71,957	0,720	0,720	0,719	68,875	0,689	0,689	0,688
Yükseltme Yöntemleri	DL	23,267	?	0,233	?	67,334	0,671	0,673	0,672
	MLP	54,391	0,542	0,544	0,543	71,957	0,720	0,720	0,719
	SL	71,186	0,713	0,712	0,711	72,265	0,724	0,723	0,720
	ICO	68,875	0,688	0,689	0,688	73,498	0,737	0,735	0,734
	DT	73,498	0,737	0,735	0,731	74,884	0,750	0,749	0,744
	FURIA	71,341	0,711	0,713	0,708	74,422	0,745	0,744	0,737
	One-R	76,117	0,771	0,761	0,756	76,117	0,771	0,761	0,756
	LMT	71,803	0,718	0,718	0,718	67,334	0,671	0,673	0,672

Tabloda görüldüğü üzere ikili sınıflandırmaya göre daha düşük bir performans elde edilse de matematik dersi veri setine göre daha yüksek başarı elde edilmiştir. Yine

aynı çalışmayla benzer algoritma verimli sonuç vermiştir. Bu algoritma da tek kural algoritması olmuştur. Yükseltme yönteminde istenilen oran elde edilememiş, klasik ve torbalama yöntemlerinde tek kural algoritması hem bütün özniteliklerle hem de seçili özniteliklerle en yüksek doğruluk oranı olan % 76,733'e ulaşmıştır.

Ayrıca benzer çalışmayı yapan Cortez ve Silva (2008) veri setinin bu türevinde % 76,1 doğruluk oranına karar ağaçları algoritması ile ulaşmıştır.

5.2.5. Portekiz Dili dersi 1 sınav notu içeren deney sonuçları

Bu çalışmada da son kullanılan veri seti türevinden ikinci sınav notunu çıkartarak elde edilen sonuçlar Tablo 5.11.'deki gibidir.

Tablo 5.11. Portekiz Dili dersi final ve 2.sınav notu çıkartılmış 5 sınıflı sonuçlar

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL	44,992	0,449	0,450	0,449	52,388	0,511	0,524	0,498
	MLP	42,835	0,428	0,428	0,428	59,014	0,593	0,590	0,588
	SL	58,398	0,592	0,584	0,584	61,633	0,616	0,616	0,612
	ICO	62,096	0,617	0,621	0,615	62,250	0,617	0,622	0,614
	DT	62,866	0,645	0,629	0,627	62,404	0,641	0,624	0,623
	FURIA	57,319	0,565	0,573	0,565	61,633	0,616	0,616	0,611
	One-R	61,787	0,618	0,618	0,613	61,787	0,618	0,618	0,613
	LMT	58,398	0,592	0,584	0,584	61,633	0,616	0,616	0,612
Torbalama Yöntemleri	DL	44,838	0,448	0,448	0,447	52,388	0,511	0,524	0,498
	MLP	46,995	0,471	0,470	0,469	59,168	0,589	0,592	0,587
	SL	56,395	0,565	0,564	0,563	61,171	0,609	0,612	0,606
	ICO	61,787	0,617	0,618	0,615	60,555	0,601	0,606	0,600
	DT	61,325	0,619	0,613	0,611	62,250	0,627	0,622	0,620
	FURIA	58,552	0,581	0,586	0,581	61,017	0,604	0,610	0,603
	One-R	62,404	0,625	0,624	0,619	62,404	0,625	0,624	0,619
	LMT	55,162	0,552	0,552	0,551	51,464	0,507	0,515	0,510
Yükseltme Yöntemleri	DL	26,348	?	0,263	?	27,119	?	0,271	?
	MLP	45,917	0,465	0,459	0,460	59,014	0,593	0,590	0,588
	SL	56,857	0,575	0,569	0,567	61,633	0,616	0,616	0,612
	ICO	58,398	0,585	0,584	0,583	59,938	0,596	0,599	0,596
	DT	62,558	0,642	0,626	0,624	62,404	0,641	0,624	0,623
	FURIA	57,319	0,565	0,573	0,565	61,633	0,616	0,616	0,611
	One-R	57,319	0,565	0,573	0,565	62,096	0,622	0,621	0,617
	LMT	53,775	0,535	0,538	0,536	52,542	0,526	0,525	0,526

Bir önceki çalışmaya göre doğruluk oranı oldukça düşüktür. Benzer sonuçlar elde edilse de en yüksek başarı oranı olan % 62,866 oranına sadece karar tablosu ulaşabilmiştir. Torbalama yöntemi ile yapılan derin öğrenme algoritma deneylerinde kesinlik ve F1-değeri hesaplanamamıştır.

Deneylerin ikinci aşamasında bu veri seti için aşağıdaki öznitelikler seçilmiştir.

- not1(G1)
- başarısızlıklar(failures)
- çalışma süresi(studytime)
- özel ders(paidclass)
- anne eğitimi(Medu)
- baba eğitimi(Fedu)

Bu deneyde ki öznitelik seçimleri de diğer deneylere benzer şekilde aile bireylerinin eğitimi, öğrencilerin ders çalışması ve mevcut başarı durumlarını gösteren öznitelikler şeklinde belirlenmiştir.

Aynı çalışma matematik dersi için yapıldığında seçili öznitelikler kullanılarak yapılan sınıflandırma tahmini en iyi sonucu verirken, bu çalışmada bütün öznitelikler kullanılarak bir miktar daha yüksek başarı elde edilmiştir. Bu durum özniteliklerin dersler üzerinde etkisini göstermektedir.

Cortez ve Silva (2008) yaptıkları çalışmanın bu aşamasında karar ağaçları ile % 62,9 doğruluk oranına ulaştıklarını makalelerinde göstermişlerdir. Bu çalışma algoritma olarak yakın algoritmalar ile aynı sonuca ulaşıldığını göstermektedir. Bu durum bu veri seti türevi ile iyileştirme çalışmalarının bir sonuç vermediğini göstermektedir.

5.2.6. Portekiz Dili dersi sınav notu içermeyen deney sonuçları

Son çalışma olan Portekiz Dili dersi veri setinde bütün sınav notları çıkartılmıştır. Oldukça düşük doğruluk oranına sahip sonuçlar Tablo 5.12’de görülmektedir.

Bu son çalışma da basit lojistik ve lojistik model ağacı algoritmalarının en yüksek oran olan % 38,521 oranında kaldığı gösterilmiştir. Oldukça düşük olan bu değer not bilgisi içermemekten kaynaklanmaktadır. Ayrıca torbalama ve yükseltme yöntemleri ile sonuçlar iyileştirilememiştir. Öznitelik azaltarak yapılan deneylerde de doğruluk oranı artmamıştır.

Aynı veri seti türevi ile çalışan Cortez ve Silva (2008) % 36,7 doğruluk oranı ile rastgele orman algoritmasının en verimli sonucu verdiğini göstermişlerdir.

Kullanılan öznitelikler ise aşağıdaki gibi belirlenmiştir.

- başarısızlıklar(failures)
- çalışma süresi(studytime)
- özel ders(paidclass)
- baba eğitimi(Fedu)
- anne eğitimi(Medu)

Tablo 5.12. Portekiz Dili dersi bütün sınav notları çıkartılmış 5 sınıflı sonuçlar

		Tüm Öznitelikler				Seçili Öznitelikler			
		DO %	Kes	Duy	F-1	DO %	Kes	Duy	F-1
Klasik Yöntemler	DL4	37,134	0,373	0,371	0,372	33,898	0,305	0,339	0,294
	MLP	31,433	0,315	0,314	0,314	31,895	0,308	0,319	0,307
	SL	38,521	0,385	0,385	0,376	35,439	0,341	0,354	0,307
	ICO	35,593	0,342	0,356	0,327	32,049	0,276	0,320	0,272
	DT	31,895	0,310	0,319	0,286	29,430	0,263	0,294	0,260
	FURIA	31,895	0,317	0,319	0,298	30,046	?	0,300	?
	One-R	31,125	?	0,311	?	31,125	?	0,311	?
	LMT	38,521	0,385	0,385	0,376	35,439	0,341	0,354	0,307
Torbalama Yöntemleri	DL	35,901	0,361	0,359	0,359	34,515	0,333	0,345	0,313
	MLP	35,439	0,352	0,354	0,352	33,128	0,327	0,331	0,315
	SL	35,439	0,356	0,354	0,352	34,977	0,347	0,350	0,306
	ICO	37,905	0,382	0,379	0,368	32,203	0,304	0,322	0,292
	DT	34,977	0,362	0,350	0,294	31,895	0,304	0,319	0,280
	FURIA	33,282	0,322	0,333	0,322	31,125	0,293	0,311	0,282
	One-R	31,587	?	0,316	?	31,587	?	0,316	?
	LMT	34,977	0,347	0,350	0,346	30,509	0,303	0,305	0,297
Yükseltme Yöntemleri	DL	19,106	0,221	0,191	0,182	24,191	?	0,242	?
	MLP	34,207	0,338	0,342	0,338	31,895	0,308	0,319	0,307
	SL	38,059	0,380	0,381	0,373	35,439	0,341	0,354	0,307
	ICO	34,207	0,339	0,342	0,337	28,814	0,279	0,288	0,273
	DT	31,895	0,310	0,319	0,286	29,430	0,263	0,294	0,260
	FURIA	31,895	0,317	0,319	0,298	30,046	?	0,300	?
	One-R	29,430	?	0,294	?	29,738	?	0,297	?
	LMT	34,669	0,348	0,347	0,345	26,965	0,269	0,270	0,267

Not bilgileri çıkartılarak bir önceki deneye göre sınıflandırma tahmini zorlaşmıştır. Bu durumda etkili öznitelik sayısı artmış ve etkili özniteliklerden bazıları da aile eğitimi olmuştur. Dil derslerinde aile bireylerinin eğitimleri ve meslekleri

öğrenci üzerinde etkili olmuştur. Ayrıca çok kademeli sınıflandırmalarda öğrencilerin ders çalışmaları ile ilgili öznitelikler etkilidir. Burada da ders çalışma süresi ve özel ders alma durumu başarı kademesini tahmin etmede etkili olmuştur.

Bu tez çalışmasında benzer özniteliklere sahip 2 veri setinden 12 farklı veri seti türetilerek deneyler yapılmış, her bir veri seti türevi için bir tablo oluşturularak seçilen bütün deney sonuçları gösterilmiştir, benzer çalışmalar ile karşılaştırılarak sonuçlar yorumlanmıştır.

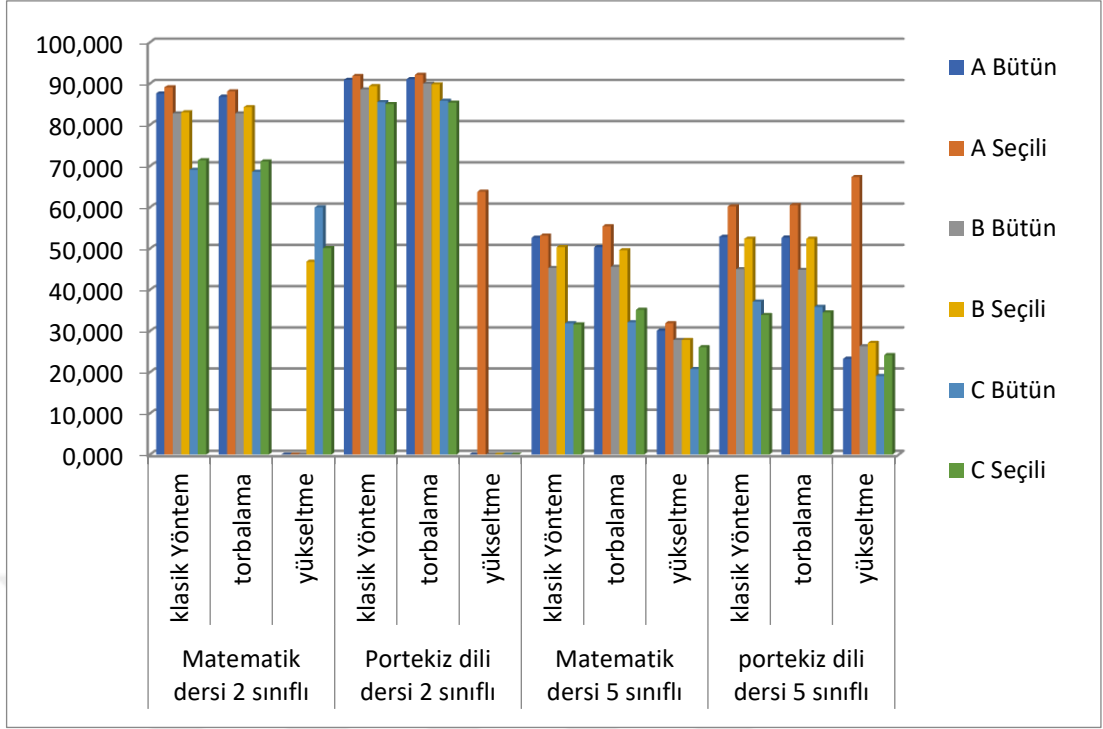
5.2.7. Algoritmaların başarısını genel bir değerlendirme

Bu çalışmada bazı algoritmalar kısmı olarak bazı veri setlerinde en yüksek doğruluk oranını verirken, bazı veri setlerinde birbirine yakın değerler göstermiştir. Her algoritma 12 farklı veri türeviyle denenmiş, bu veriler özniteliklerin tümü ve seçili olanları ile ayrı ayrı gerçekleştirilmiştir. Ayrıca 3 yöntem için bütün uygulamalar tekrarlanarak bir algoritmaya ile 72 adet deney yapılmıştır.

Grafiklerde A: 2 sınav notu içeren veri setini, B: 1 sınav notu içeren veri setini, C: sınav notu içermeyen veri setini ifade etmektedir.

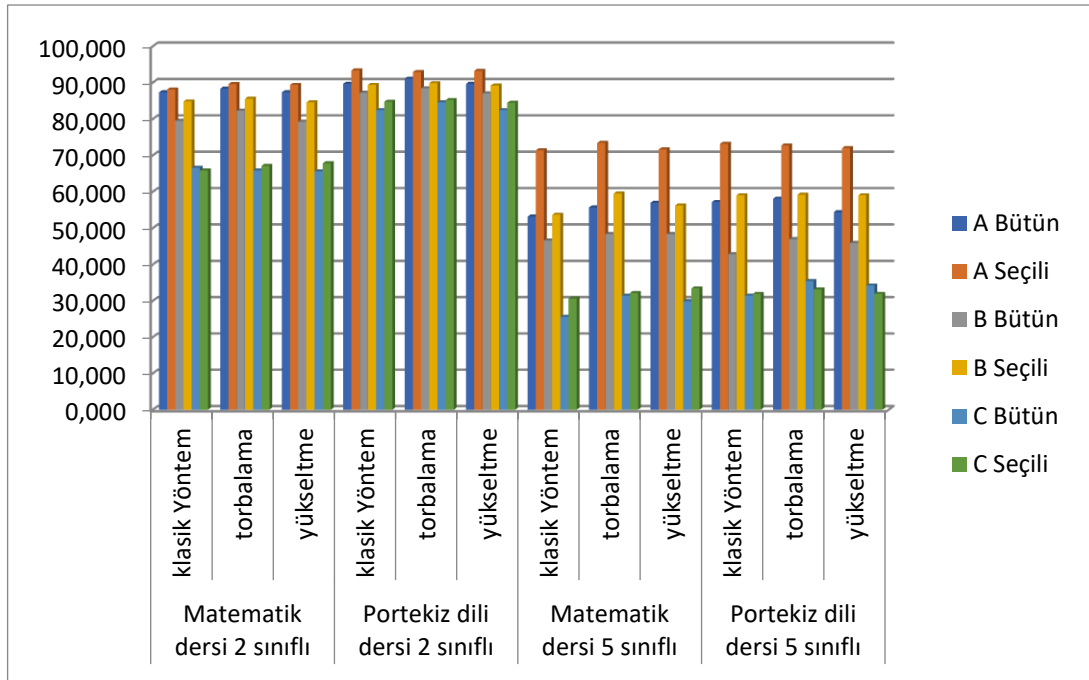
Bütün: özniteliklerin tümünü kullanarak, Seçili: öznitelikler seçilerek ve eksiltiyle yapılan deneyleri ifade etmektedir. Grafiklerde algoritmaların veri setlerinde gösterdiği değerler verilmiştir.

Aşağıdaki Görsel 5.1’de derin öğrenme yöntemi olan DL4MLP algoritması ile yapılan deney sonuçlarını göstermektedir. Tablodan da anlaşılacağı üzere 2 sınıflı veri setlerini tahmin etmek daha kolay iken yükseltme yöntemi bu algoritma ile verimli sonuç vermemektedir. Ayrıca Portekiz dili dersinin 2 sınıflı tahmin başarısı diğerlerinden daha yüksektir.



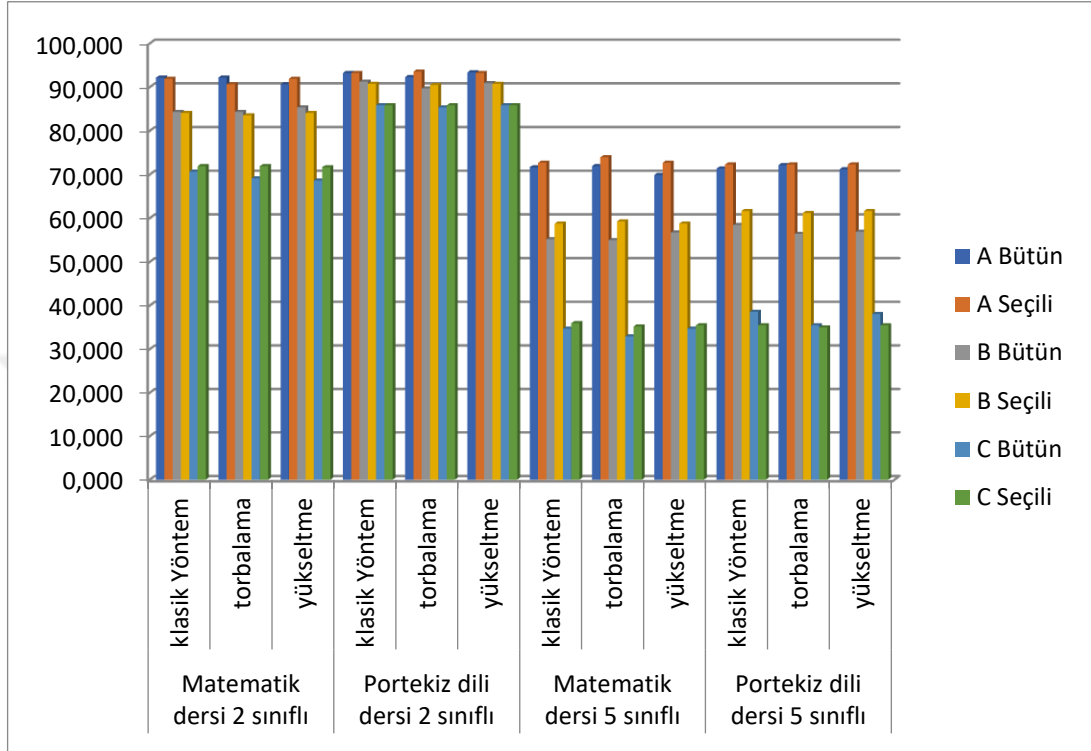
Görsel 5.1. Derin Öğrenme algoritma grafiği

Görsel 5.2’de çok katmanlı algılayıcı algoritmasının yapılan bütün deneylerdeki sonuçları verilmiştir. Bu algıtmada öznelilik seçme işlemi ikili sınıflandırmada bütün özneliliklere göre çok az iyileştirme yaparken, beş kademeli sınıflandırma da bütün özneliliklere göre çok daha etkilidir. Yine bu algıtmada da Portekiz Dili dersi veri setinde daha başarılı bir sınıflandırma tahmini olmuştur.



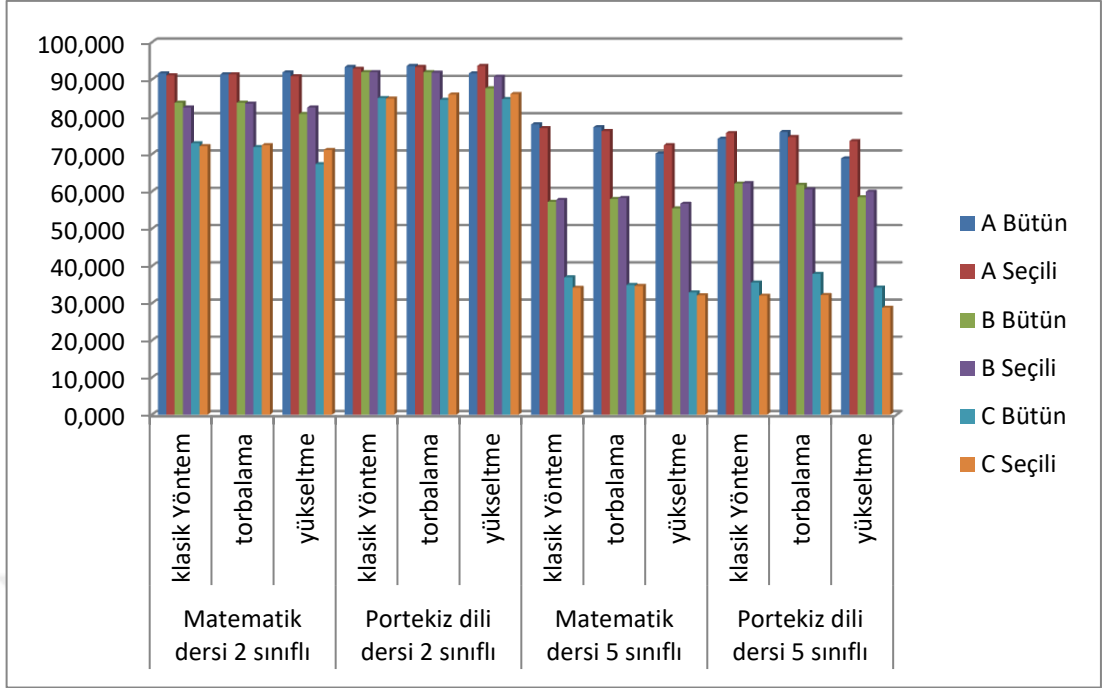
Görsel 5.2. Çok Katmanlı Algılayıcı algoritma grafiği

Görsel 5.3'te basit lojistik algoritmasının deney sonuçları görülmektedir. Algoritma 2 sınıflı Portekiz dili dersinde bütün veri setleri ile yüksek doğruluk oranında sınıflandırma tahmini yaparken, 5 kademeli sınıflandırma da özellikle not bilgisi olmayan veri setlerinde sınıflandırmada tahmin oranı düşüktür.



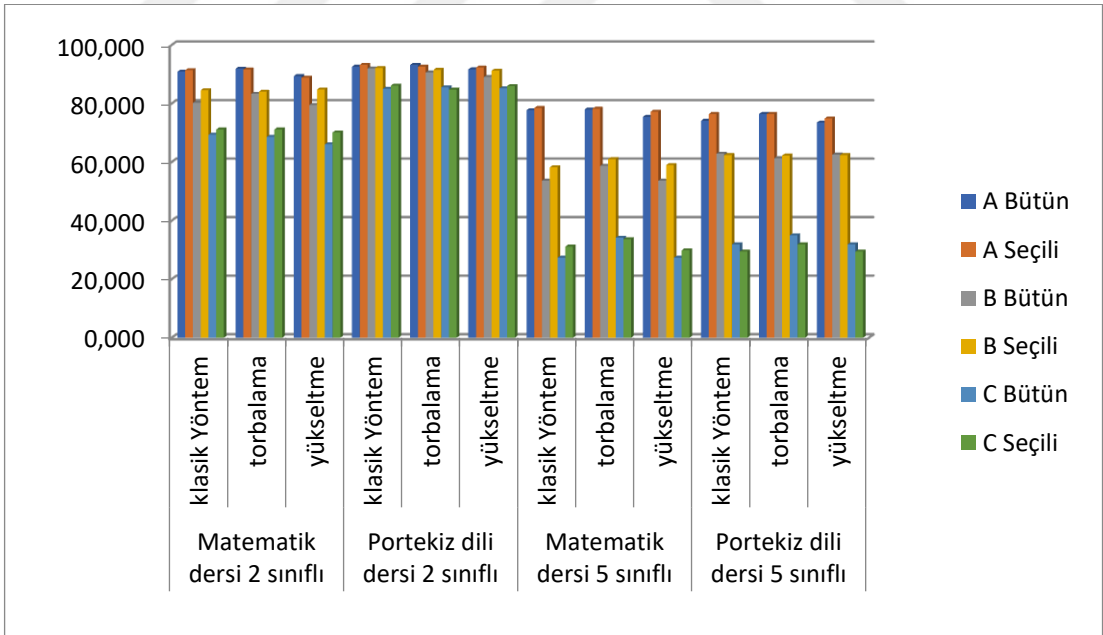
Görsel 5.3. Basit Lojistik Algoritma Grafiği

Görsel 5.4'te yinelemeli sınıflandırma optimize edici algoritmasının deney sonuçları görülmektedir. Bu algoritmada 2'li sınıflandırmada daha başarılıdır. Ayrıca Portekiz dili dersi veri setini sınıflandırılmasında daha başarılıdır.



Görsel 5.4. Yinelemeli Sınıflandırıcı Optimize Edici Algoritma Grafiği

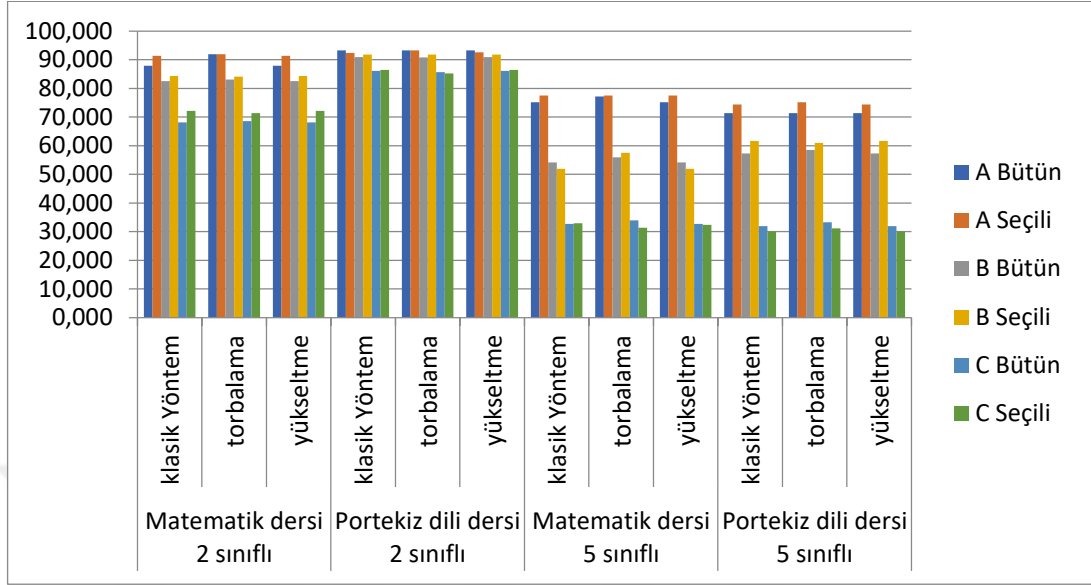
Karar tablosu algoritması ile yapılan deney sonuçları Görsel 5.5'te verilmiştir. Bu algoritmada seçili öz niteliklerin daha başarılı olduğu görülmektedir.



Görsel 5.5. Karar Tablosu Algoritma Grafiği

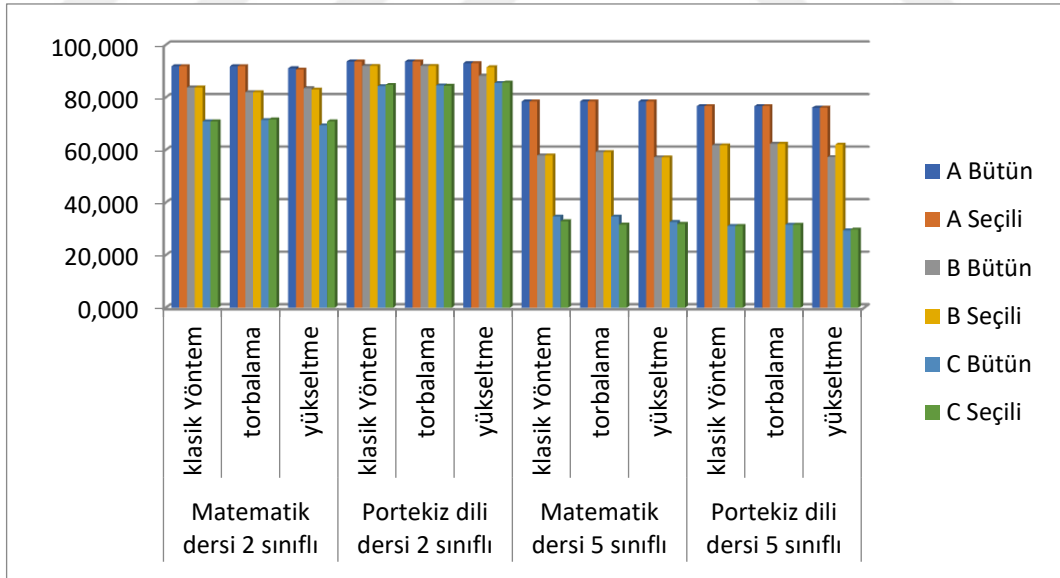
Bulanık sırasız kural tümevarım algoritmasıyla yapılan deney sonuçları Görsel 5.6'da verilmiştir. Algoritmanın en başarılı olduğu veri seti Portekiz dili dersi ile 2'li sınıflandırma olmuştur. 3 yöntemde de aynı sonuçlara ulaşmıştır. 2'li sınıflandırmada

yöntemlerin değişmesi ve not bilgilerinin çıkartılması sonuçları önemli derecede etkilemezken, 5 kademeli sınıflandırma da sınav notlarının etkisi görülmektedir.



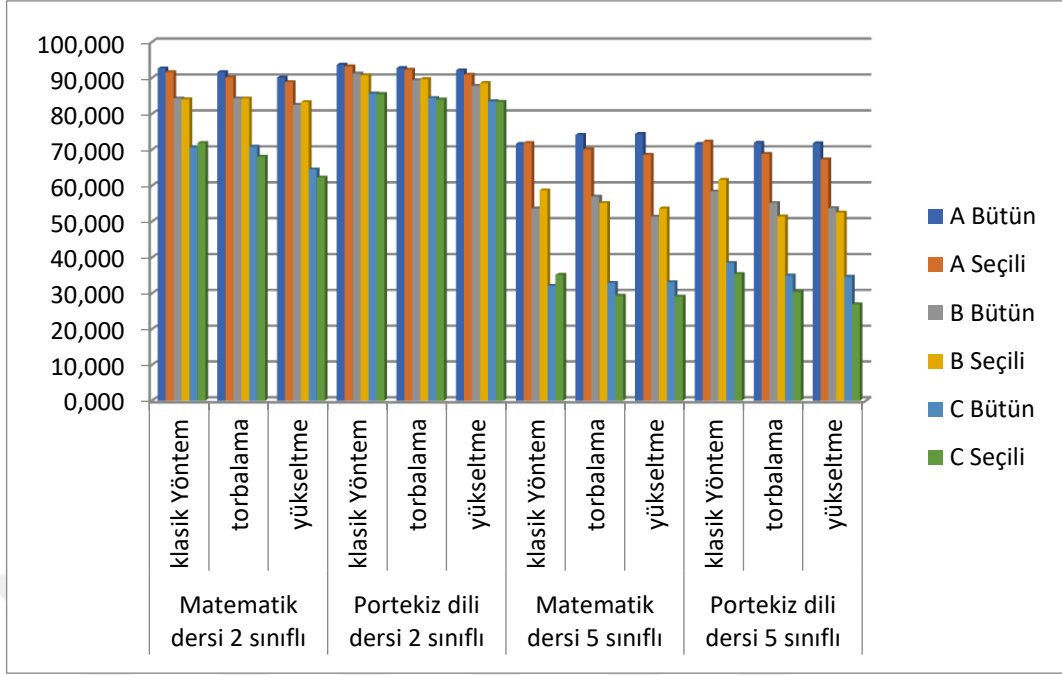
Görsel 5.6. FURIA Algoritma Grafiği

Tek-Kural algoritması bu çalışmada en yüksek doğruluk oranına sahip olan algoritmadır. Görsel 5.7’de görüldüğü üzere benzer setlerdeki değerler birbirine çok yakın, hatta bir çoğu aynıdır.



Görsel 5.7. Tek-Kural Algoritma Grafiği

Son algoritma olan lojistik model ağacı diğer algoritmalar ile benzer sonuçlar göstermektedir. Bu algoritmada klasik yöntemler daha başarılı olmuştur.



Görsel 5.8. Lojistik Model Ağacı Algoritma Grafiği

5.3. Veri Seti İle Yapılmış Çalışmalar

Bu çalışmada kullanılan veri setleri Paula Cortez (2008) tarafından oluşturularak açık kaynak kodlu olarak yayınlanmıştır. Gerçek verilerden oluşması, yeterince büyük ve anlaşılır verilerden oluşması ve açık kaynak kodlu olması nedeniyle birçok araştırmada da kullanım sebebi olmuştur. Bu veri setleri ile yapılan birçok çalışma mevcuttur. Aşağıdaki Tablo 5.13'te bu çalışmalardan yakın zamanda yapılmış ve konu olarak benzer fakat farklı yöntem ve algoritmalar kullanan bazı çalışmalar gösterilmiştir. Tabloda yapılan çalışmaların kullandığı veri seti türü, çalışmayı yapan yazarlar, çalışmanın motivasyon nedeni, çalışmanın yapıldığı programlama ortamı, çalışmada kullanılan algoritmalar ve çalışma sonucunda elde edilen bulgular sunulmuştur.

Bu veri setleri ile yapılan çalışmalar genel olarak değerlendirildiğinde % 90 üzeri bir başarı ile sınıflandırmada kullanılmaya elverişli bir veri seti olduğu sonucuna varılabilir.

Çalışmaların çoğu % 93 doğruluk oranına sahip iken, matematik dersi veri setinde HELA algoritmasının, Portekiz Dili veri setinde ise tek kural algoritmasının en yüksek doğruluk oranına sahip olduğu görülmektedir.

İkili sınıflandırma çalışmalarında ağaç tabanlı algoritmalar daha başarılıdır. Bunun sebebinin sınıflandırmayı en çok etkileyen özneliliğin ön sınav notları olduğu

düşünülmektedir. Sınıf sayısının az olduğu durumlarda ağaç algoritmalarında kurallar daha kolay ve net oluşabilmektedir. Bu veri setlerinde de 2.sınav notu algoritma kurallarını en çok etkilediği için diğer özniteliklerin etkisi daha azdır ve etiketli değerler ile verimli çalışan algoritmaların başarısı daha düşüktür.

Tablo 5.13. Benzer çalışmalardan 2 sınıflı çalışmalar

Kullanılan Veri Seti	Yazarlar	Motivasyon	Kullanılan Program	Kullanılan Algoritmalar	Bulgular (DO)
Matematik veri seti 2'li sınıf	(Özkan et al., 2018)	Sınıflandırma ve en etkili özniteliği bulma	R	C5.0, BoostedC5, RT,SVM, LR,RF, DL	86 % C5.0
Matematik veri seti 2'li sınıf	(Başer et al., 2020)	Sınıflandırma	WEKA	ICO, OneR, LogitBoost, YSA	92.2 % One-R
Matematik veri seti 2'li sınıf	(Ünal, 2020)	Sınıflandırma	WEKA	NB, RF, DT	93.67 % RF
Matematik veri seti 2'li sınıf	(Bozkurt Keser & Aghalarova, 2022)	Sınıflandırma	Python	HELA, GB, XGboost, LightGBM	96,6 % HELA
Matematik veri seti 2'li sınıf	(Cortez & Silva, 2008)	Sınıflandırma	R	NV,NN, SVM, DT,RF	91,9 % NV
Matematik veri seti 2'li sınıf	Bu çalışma	Sınıflandırma	WEKA	DL4MLP, MLP,SL, ICO, DT, FURIA, ONE-R, LMT	92.7 % LMT

Tablo 5.13. (Devam) Benzer çalışmalardan 2 sınıflı çalışmalar

Kullanılan Veri Seti	Yazarlar	Motivasyon	Kullanılan Program	Kullanılan Algoritmalar	Bulgular (DO)
Portekiz Dili veri seti 2'li sınıf	(Ünal, 2020)	Sınıflandırma	WEKA	NB, RF, DT	93.2 % RF
Portekiz Dili veri seti 2'li sınıf	(Bozkurt Keser & Aghalarova, 2022)	Sınıflandırma	Python	Hela, GB, XGboost LightGBM	91,3 % HELA
Portekiz Dili veri seti 2'li sınıf	(Cortez & Silva, 2008)	Sınıflandırma	R	NV, NN, SVM, DT RF	93,0 % DT
Portekiz Dili veri seti 2'li sınıf	Bu çalışma	Sınıflandırma	WEKA	DL4MLP MLP SL ICO DT FURIA ONE-R LMT	93.7 % One-R, ICO, LMT

Matematik dersi veri seti hemen her ülkede okutulan bir ders olduğu için daha çok tercih edilmiştir. Matematik veri seti ile yapılan çalışmalarda en iyi doğruluk oranına hibrit bir yöntem olan HELA algoritması ile ulaşmışlardır.

Portekiz Dili dersi veri seti Portekiz’de kullanılan dile ait bir ders olmasına rağmen algoritmaların verimliliğini test etmek veya karşılaştırmak amacıyla çalışmalara konu olmuştur. Bu veri seti ile yapılan deneylerde en başarılı sonuç ise mevcut çalışmada elde edilen % 93,7 doğruluk oranı olmuştur. Bu oran tek kural, yinelemeli sınıflandırıcı optimize edici ve lojistik model ağacı algoritmaları ile elde edilmiştir.

Veri setlerinin bir diğer sınıflandırma şekli ise 3 ve 5 kademeli sınıflandırma olmuştur. 3 kademeli sınıflandırmada araştırmacılar kendi belirledikleri not baremine

göre sınıflandırma yaparken, 5 kademeli sınıflandırmada ise araştırmacılar Portekiz’de okullarda kullanılan not baremi kullanmışlardır. Yapılan çalışmalardan bazı örnekler Tablo 5.14’te verilmiştir.

Tablo 5.14. Benzer çalışmalardan 3 ve 5 sınıflı çalışmalar

Kullanılan Veri Seti	Yazarlar	Motivasyon	Kullanılan Program	Kullanılan Algoritmalar	Bulgular (DO)
Veri setleri birleştirilmiş 3'lü sınıf	(Kayalı & Buyrukoğlu, 2022)	Öznitelik seçme yöntemlerini kıyaslama	R	DTree RF, SVM	86 % RF
Veri setleri birleştirilmiş 3'lü sınıf	(Bentaleb & Abouchabaka, 2022)	Sınıflandırma	Python	SVM, RF, NB, LR, XB	79 % LR
Veri setleri birleştirilmiş 5 sınıflı	(Peker, 2016)	En etkili özniteliği bulma	-	C4.5, FFANN, SVM,KELM	KELM
Matematik veri seti 5'li sınıf	(Kızılkaya & Oğuzlar, 2018)	Cinsiyetleri ve başarı durumları arasındaki ilişki	-	C4.5, FFANN, SVM,KELM	KELM
Matematik veri seti 5'li sınıf	(Felicia & Ferren, 2022)	Sınıflandırma	RapidMiner	GLM, RF, NB	48.6 % NB
Matematik veri seti 5'li sınıf	(Bozkurt Keser & Aghalarova, 2022)	Sınıflandırma	Python	Hela,GB, XGboost LightGBM	78.2 % HELA
Matematik veri seti 5'li Sınıf	(Ünal, 2020)	Sınıflandırma	WEKA	NB, RF, DT	79.5 % J48
Matematik veri seti 5'li sınıf	(Cortez & Silva, 2008)	Sınıflandırma	R	NV, NN, SVM, DT, RF	78,5 % NV

Tablo 5.14. (Devam) Benzer çalışmalardan 3 ve 5 sınıflı çalışmalar

Kullanılan Veri Seti	Yazarlar	Motivasyon	Kullanılan Program	Kullanılan Algoritmalar	Bulgular (DO)
Matematik veri seti 5'li sınıf	Bu çalışma	Sınıflandırma	WEKA	DL4MLP, MLP, SL, ICO, DT FURIA ONE-R LMT	78,5 % One-R, DT
Portekiz Dili veri seti 5'li sınıf	(Ünal, 2020)	Sınıflandırma	WEKA	NB, RF, DT	77.2 % RF
Portekiz Dili veri seti 5'li sınıf	(Bozkurt Keser & Aghalarova, 2022)	Sınıflandırma	Python	Hela, GB, XGboost LightGBM	78.5% HELA
Portekiz Dili veri seti 5'li sınıf	(Cortez & Silva, 2008)	Sınıflandırma	R	NV, NN, SVM, DT RF	76,1 % DT
Portekiz Dili veri seti 5'li sınıf	Bu çalışma	Sınıflandırma	WEKA	DL4MLP, MLP, SL, ICO,DT, FURIA, ONE-R, LMT,	76.5 % One-R

3-5 kademeli sınıflandırma için kullanılan veri setlerinin değişik versiyonları ile farklı çalışmalar yapılmıştır. Çalışmaların bir kısmında sınıflandırmayı farklı bir not baremi kullanarak, iki seti birleştirerek veya iki değer sonucunu ortak alarak yorumlamışlardır. Yine veri setlerini birleştirerek 5 kademeli sınıflandırma yapılan çalışmalarda mevcuttur.

Veri setlerinin sınıflandırmayı etkileyen etiketli özniteliklerinin farklı olmasına rağmen en etkili öznitelik yine ön sınav notlarıdır. Bu nedenle en çok başarı gösteren algoritmaların başında kural tabanlı ağaç yapıları gelmektedir.

5 kademeli sınıflandırmalarda 2 kademeli sınıflandırmaya göre doğru sınıf tahminlerini yapmak daha zordur. Hem sınıf sayısının fazla olması, hem de sınıflar arasında geçişlerin birbirine yakın olmasından kaynaklı doğruluk oranı 2'li sınıflandırmaya göre daha düşüktür.

Bu veri setleriyle yapılan çalışmaların bir kısmı sınıflandırmada etkili öznitelikleri bulma veya etkili öznitelikleri bulan algoritmaları tespit etme üzerine yapılmıştır.

Bazı çalışmalarda sınıflandırma sonuçlarını iyileştirmek için veri çoğaltma veya bazı verileri çıkartma yöntemleri uygulanmıştır. Verilen benzer çalışmalar gerçek dünya verileri ile çalışan araştırmacıların verileri olmuştur. Mevcut çalışma da aynı şekilde var olan veriler ile yapılmış herhangi bir artırma veya azaltma yöntemi uygulanmamıştır.

Sonuçlara baktığımızda matematik veri seti ile çalışan Ünal (2020) en yüksek doğruluk oranını J48 algoritması ile öznitelik seçerek %79,5 oranında bulmuştur.

Portekiz Dili veri setinde ise hibrit bir yöntem olan HELA algoritmasını kullanarak Bozkurt Keser ve Aghalarova (2022) %78,5 oranında doğruluk oranına ulaşmışlardır. Yalın algoritmaları kullanarak yapılan çalışmalarda bu orana ulaşamamıştır.

6. SONUÇ VE TARTISMA

Bu çalışmada iki derse ait notlar içeren iki veri seti ile notları kademeli olarak çıkartarak 12 farklı türevi ile deneyler yapılmıştır. Deneylerde farklı algoritmalar denendikten sonra 8 farklı algoritma seçilmiş, seçilen bu 8 algoritma 3 farklı yöntem ile tekrarlanmıştır. Bu yöntemler klasik yöntem, torbalama yöntemi ve yükseltme yöntemi şeklindedir. Daha sonra bütün deneyler öznitelik seçme yöntemleriyle öznitelikler azaltılarak tekrarlanmıştır. Bu aşamada yapılan 12 veri seti, 3 farklı yöntem ve 2 farklı öznitelik durumu için 8 algoritma denenerek toplam 576 deney sonucu tablolar halinde sunulmuştur.

Deneyler önce klasik yöntem ile 8 algoritma test edilerek başlanmış, sonrasında aynı algoritmalar diğer iki yöntem için de uygulanmıştır. Dördüncü aşamada sırası ile aynı algoritma ve yöntemler öznitelik azaltılarak denenmiştir. Toplam 6 aşamadan oluşan bir deney seti, 12 veri seti türevine de uygulanmıştır.

Çalışmalarda farklı algoritmalar denenmiş ve en yüksek doğruluk oranını veren algoritmalar 8 algoritma belirlenerek bütün deneylere eklenmiştir. Genel olarak bakıldığında farklı algoritmalar farklı setlerde başarılı olsa da en fazla başarı gösteren algoritma tek kural algoritması olmuştur. Bu da öğrenci akademik başarısını etkileyen faktörlerin benzer sebeplerden kaynaklandığını göstermektedir.

Bu 12 veri seti türevli çalışmaya genel olarak bakıldığında veri setlerinden etkili öznitelikler çıkartıldığında sınıflandırmada ki başarı düzeyi düşmektedir. Bu çalışmada ki en etkin öznitelik ikinci sınav notu(G2) olmuştur. İkinci sınav notları final notlarına yakın oldukları için etkisi diğerlerine göre daha fazladır.

Ayrıca not bilgisi fazla olan çalışmalarda birden fazla algoritma verimli çalışırken, özellikle not bilgisi içermeyen deneylerde tek bir algoritma en yüksek doğruluğa ulaşmıştır.

Özellikle not bilgilerini içeren özniteliklerin çıkartıldığı deneylerde bazı algoritmalar sonuç üretememiştir. Sınıflandırma yapabilen algoritmaların doğruluk oranı sonuçları diğer veri seti çalışmalarına göre daha düşüktür. Fakat bu çalışma türevi daha önce not bilgileri kaydı olmayan öğrenciler için, okula yeni başlayan veya ilk defa veri kaydı tutulacak durumlar için yapılabilecek bir çalışmadır.

Öznitelik seçimlerinde görüldüğü gibi öğrencilerin başarısını etkileyen en önemli faktörler önceki dönem-yıllardaki başarıları ve aile yaşantılarıdır. Genel olarak öznitelikler bir bire benzer çıksa da bazı setlerde farklı özniteliklere de rastlanmaktadır.

Veri setlerine bakıldığında farklı kademeli sınıflandırmalarda Portekiz dili dersi veri setinin daha yüksek doğruluk oranında sınıflandırıldığı görülmektedir. Bunun nedeninin de o ülkede konuşulan dil ile ilgili bir ders olduğu için özniteliklerin sınıflandırmada etkili olduğu düşünülmektedir.

Ayrıca Portekiz Dili dersi için anne-baba eğitimi ve meslekleri özellikle başarıyı etkileyen önemli faktörlerdendir. Bunun nedeni öğrencilerin bu derste başarı düzeylerinin; aile bireylerinin eğitimi, aile kültürü, öğrenci ile geçirilen zaman, ilgi ve aile ortamındaki konuşma şekli(ağız-lehçe) gibi durumlardan etkilendiği düşünülmektedir.

Matematik dersi için ebeveynlerin meslekleri dışında öğrencinin geçmiş dönem başarı durumu ve yükseköğrenime olan isteği ile ilgili öznitelikler en etkili özniteliklerdir. Ülkemizde olduğu gibi yükseköğrenime devam etmek isteyen öğrenciler matematik dersine daha çok önem vermektedir. Matematik dersi başarısı; sayısal zekâ düzeyi gibi veri setlerinde olmayan bir özneliğe ve ders çalışma alışkanlıklarına bağlıdır. Bu veri setinde ki dışarı çıkma, ders çalışma süresi, özel ders gibi özniteliklerin bu nedenlerle etkili olduğu düşünülmektedir.

Geçmiş dönem başarısızlık durumu öğrenci başarısını etkileyen en önemli faktörlerdendir. Eğer öğrencinin geçmiş dönem başarısızlığı o dönemlerde yaşadığı sağlık sorunu, psikolojik olumsuzluklar, maddi-manevi sıkıntılar gibi faktörlere bağlı ise ve bu durumlar ortadan kalkmamış ise bu öznelik başarı sınıfını bulmada etkili olacaktır. Makine öğrenimi bu özneliğe bakarak öğrencinin gelecek dönem başarısının düşük olacağını öngörecektir. Bu durum makine öğrenimi başarısı için verimli, öğrencilerin başarısı için olumsuz bir durumdur.

2 kademeli sınıflandırmanın beş kademeli sınıflandırmaya göre daha başarılı olduğu görülmektedir. Bunun nedeninin sınıflar arasında geçişin iki kademelide daha kesin olması ve az sınıf olduğu için sınıflandırmanın daha kolay ve başarılı olmasıdır. 5 kademeli sınıflandırmada ise not baremi geçer notlarda her 2 notta bir sınıf değişmektedir. Ayrıca kalır not alan öğrencilerin profilleri birbirine çok yakın iken, geçer not alan fakat ara kademelerde bulunan öğrencilerin profillerini ayırt etmek daha zordur. Kısacası gerçek hayatta da öğrencilerin geçme- kalma durumunu mevcut verilerine bakarak tahmin etmek daha kolay iken, geçeceği düşünülen öğrencilerin hangi kademede olacağını tahmin etmek zordur. Makine öğrenmesi de insanların

öğrenme mantığını taklit eden bir sistem olduğu için gerçek dünya verilerinde dar kalıplarda sınıflandırma yapması daha zordur.

Çalışmanın amaçlarından biri klasik yöntemlerin yanı sıra farklı yöntemler deneyerek ve öznitelikleri azaltarak veya etkili öznitelikleri bularak sınıflandırmada en başarılı algoritma ve yöntemleri tespit etmektir. Bu çalışmanın farklı aşamalarında hem farklı yöntemler hem de öznitelik seçilerek yapılan deneyler en yüksek doğruluk oranına sahip olmuştur. Özellikle bir topluluk öğrenme yöntemi olan torbalama yöntemi bazı deneylerde başarıyı arttırmıştır.

Çalışma genel olarak hedefine ulaşarak veri setleri ile yapılan sınıflandırmalarda doğruluk oranları belirli oranlarda yükseltilerek iyileştirmeler elde edilmiştir. İki çalışmada ise aynı oranda kalmıştır.

İlerleyen çalışmalarda farklı yöntemler denenerek sınıflandırma doğruluk oranları artırılabilir. Çalışmada veri setinin makine öğrenmesine zorluk çıkaran öznitelikleri belirlenerek sonuçlar iyileştirilebilir.

Ayrıca benzer bir veri seti ülkemizde ortaöğretim öğrencileri için oluşturulabilirse, ortaöğretim başarı düzeylerini arttırmak için gerekli tedbirlerin alınması konusunda rehberlik edebilir. Bu çalışmadan yola çıkılarak okullarda kullanılabilecek uygulamalar geliştirilip gelecek dönemler için öngörülebilir. Bu uygulamalar hem rehberlik öğretmenlerinin hem de sınıf rehber öğretmenlerinin öğrenciyi tanıma ve risk durumları için gerekli tedbirleri alma konusunda erken uyarı sistemi olarak değerlendirilebilir.

Ülkemizde ortaöğretimde özellikle lise dönemlerinde öğrenciler sınıfı en fazla 2 kez tekrar edebilirler. Herhangi bir sınıf düzeyinde 2. defa sınıf tekrarına kalırsa örgün eğitimden ayrılmak zorunda kalmaktadır. Bu durumu engellemek için öğrenci başarı durumu öngörülebilir ve etkileyen faktörler bulunabilirse, sorunların çözümü için öğrenci ile birebir çalışmalar yapılabilir veya özel destek alması sağlanabilir. Başarı potansiyeli olan fakat olumsuz şartlardan dolayı yeterli düzeye çıkamayan öğrenciler belirlenebilirse, maddi-manevi destekler ile öğrencilerin daha iyi şartlarda eğitime devam etmeleri sağlanabilir.

Bu çalışma ile öğrenci başarısını arttırmak, başarısını desteklemek, olumsuzlukları önlemek adına adımlar atılabilir.

KAYNAKÇA

- Hakyemez, T. (2015). *İlk yıl öğrencilerinin akademik performansına etki eden faktörlerin araştırılması ve bu faktörlere bağlı olarak başarılarının tahminine yönelik bir karar destek sistemi tasarımı*. Sosyal Bilimler Enstitüsü, Yönetim Bilişim Sistemleri. Sakarya: Sakarya Üniversitesi.
- Abdualgalil, B., & Abraham, S. (2020). Applications of Machine Learning Algorithms and Performance Comparison: A Review. *International Conference on Emerging Trends in Information Technology and Engineering*, 1-6.
- Adak, M. F., & Duralioğlu, Ö. (2023). Makine Öğrenmesi Yöntemleri Kullanılarak Öğrencilerin Kazanım Bilgileri ile Sınavlardaki Başarı Durumunun Tahmini. *Zeki Sistemler Teori ve Uygulamaları Dergisi*, 6(1), 43-51.
- Agrawal, H., & Mavani, H. (2015). Student Performance Prediction using Machine. *International Journal of Engineering Research and Technology*, 111-113.
- Aksu, M., & Karaman, E. (2017). Karar Ağaçları ile Bir Web Sitesinde Link Analizi ve Tespiti. *Acta Infologica*, 1 (2), 84-91.
- Alpaydın, E. (2009). *Introduction to machine learning*. MIT press.
- Al-Radaideh, Q. A., Al-Shawakfa, E., & Al-Najjar, M. (2006). Mining Student Data Using Decision Trees. *International Arab Conference on Information Technology*, Yarmouk University, Jordan.
- Atmaca, K. (2017, 03 07). <https://kenanatmaca.com>. 01 2023 tarihinde <https://kenanatmaca.com/oner-algoritmasi/> adresinden alındı
- Aydın, F., & Aslan, Z. (2017). Yapay öğrenme yöntemleri ve dalgacık dönüşümü kullanılarak nöro dejeneratif hastalıkların teşhisi. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 32(3), 749-766.
- Aydoğan, M., & Karcı, A. (2018). Meslek Yüksekokulu Öğrencilerinin Başarı Performanslarının Makine. *International Symposium on Multidisciplinary Studies and Innovative Turkey*.
- Azgınoğlu, N. (2018, Mayıs 15). www.nuhazginoglu.com. <http://www.nuhazginoglu.com/2018/05/15/cok-katmanli-algilayicilar-multi-layer-perceptron/> adresinden alınmıştır
- Başer, S. h., Hökelekli, O., & Adem, K. (2020). Ortaöğretimde Öğrenim Gören Öğrenci Performanslarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi. *Bilgisayar Bilimleri ve Teknolojileri Dergisi*, 1(1), 22-27.
- Başer, S., Hökelekli, O., & Adem, K. (2020). Ortaöğretimde Öğrenim Gören Öğrenci Performanslarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi. *Bilgisayar Bilimleri ve Teknolojileri Dergisi*, 22-27.
- Bayır, F. (2006). *Yapay Sinir Ağları ve Tahmin Modellemesi Üzerine Bir Uygulama*. İstanbul Üniversitesi, Sayısal Yöntemler Bilim Dalı. İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

- Bentaleb, A., & Abouchabaka, J. (2022). ENSEMBLE LEARNING FOR MINING EDUCATIONAL DATA. *Journal of Theoretical and Applied Information Technology*, 2715-2722.
- Bozkurt Keser, S., & Aghalarova, S. (2022). HELA: A novel hybrid ensemble learning algorithm for predicting academic performance of students. *Education and Information Technologies*, 27, 4521–4552.
- Breiman, L. (1996). Bagging Predictors. *Kluwer Academic Publishers*, 24, 123-140.
- Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. *Proceedings of 5th Annual Future Business Technology Conference* (s. 5-12). Porto: EUROSIS, ISBN 978-9077381-39-7.
- Coşkun, C., & Abdullah, B. (2011). Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması. *Akademik Bilişim Konferansı Bildirileri*. Malatya: İnönü Üniversitesi.
- Dede, M. (2020, Şubat 16). <https://muhtalipdede.medium.com/https://muhtalipdede.medium.com/yapay-sinir-a%C4%9Flar%C4%B1-d811e802a381-adresinden-alınmıştır>
- Doğan, F., & Türkoğlu, İ. (2019). Derin Öğrenme Modelleri ve Uygulama Alanlarına İlişkin Bir Derleme. *DÜMF Mühendislik Dergisi*, 10(2), 409-445.
- El Naqa, I., & Murphy, M. J. (2015). *Machine Learning in Radiation Oncology*. Springer, Cham.
- Erdoğan, Z., Namlı, Ö., & Akarsu, C. (2017). Öğrenci başarısının bileşik makine öğrenme teknikleri kullanılarak tahmini. *The Proceedings of International Symposium on Industry 4.0 and Applications* (s. 150-155). Karabük, Turkey: İstanbul Üniversitesi-Cerrahpaşa Akademik Veri Yönetim Sistemi.
- Estevez, P. A., Tesmer, M., Perez, C. A., & Zurada, J. M. (2009). Normalized Mutual Information Feature Selection. *IEEE Neural Networks*, 189-201.
- Fan, S., & Huang, B. (2011). Training Iterative Collective Classifiers with. *Journal of Machine Learning Research*, 119(2), 2122-2159.
- Felicia, & Ferren. (2022). Exploring Secondary School Performance by Using Machine Learning Algorithms. *Journal of Educational Analytics (JEDA)*, 41-60.
- Freud, Y., & Schapire, R. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 119-139.
- Gadhavi, M., & Patel, D. (2017). Student final grade prediction based on linear regression. *Indian Journal of Computer Science and Engineering (IJCSE)*, 274-279.
- Gaoa, C., Sang, N., & Tanga, Q. (2010). On Selection and Combination of Weak Learners in AdaBoost. *Pattern Recognition Letters*, 31(9), 991-1001.
- GÖK, M. (2017). Makine Öğrenmesi Yöntemleri İle Akademik Başarının Tahmin Edilmesi. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, 5(3), 139 - 148.

- Güldal, H. (2018). Karar ağacı algoritmalarının eğitsel veriler üzerindeki performanslarının incelenmesi. *13th International Balkan Education and Science Congress* (s. 6-10). Edirne: Trakya University.
- Güvenç, E., Sakal, M., Çetin, G., & Özkaraca, O. (2022). Öğrencilerin Dersteki Niteliklerinin Makine Öğrenmesi Teknikleri Kullanılarak Sınıflandırılması. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 10, 1359-1371.
- Güzel, K. (2020, 10 28). <https://kadirguzel.medium.com/boosting-nedir-ad%C4%B1m-ad%C4%B1m-adaboost-algoritmas%C4%B1-439cce20ab9a> adresinden alınmıştır
- Hamilton, T. (2023, Nisan 8). <https://www.guru99.com/decision-table-testing.html> adresinden alınmıştır
- Han, J., Kamber, M., & Pei, J. (2001). *Data mining: concepts and techniques*. San Francisco: ELSEVIER.
- Holte, R. (1993). Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. *Kluwer Academic Publishers, Boston. Manufactured in The Netherlands*.
- Hühn, J., & Hüllermeier, E. (2009). FURIA: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, 293–319.
- IBM. (2021, Mart 04). https://www.ibm.com/docs/en/bpmoc?topic=SS964W/com.ibm.wbpm.admin.doc/topics/cbre_busiru_decisiontable.html adresinden alınmıştır
- İnik, Ö., & Ülker, E. (2017). Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri. *GAZİOSMANPAŞA BİLİMSEL ARAŞTIRMA DERGİSİ (GBAD)*, 6(3), 85-104.
- Iqbal, Z., Qadir, J., Mian, A., & Kamiran, F. (2017). Machine Learning Based Student Grade. *arXiv*, arXiv:1708.08744v1.
- Işıkhani, E. (2020, Ocak 1). <https://isikhanelif.medium.com/multi-layer-perceptron-mlp-nedir-4758285a7f15> adresinden alınmıştır
- Karadağ, M. (2014). *Karar ağaçları ile lojistik regresyon analizinin performanslarının simülasyon çalışması ile karşılaştırılması*. Trakya Üniversitesi, Biyoistatistik Anabilim Dalı. EDİRNE: Trakya Üniversitesi, Sosyal Bilimler Enstitüsü.
- Kayalı, S., & Buyrukoğlu, S. (2022). Makine Öğrenmesi Yöntemleriyle Öğrencilerin Akademik Performanslarının Sınıflandırılması. *2nd International Conference on Educational Technology and Online Learning*. Balıkesir: ICETOL.
- Keskin, H., & Yapıcı, Ş. (2008). Başarılı Ve Başarısız Öğrencilerin Kişilik Özellikleri İle İlgili Öğretmen Ve Veli Görüşleri. *Kocatepe Üniversitesi Kuramsal Eğitim Bilim Dergisi*, Cilt 1, Sayı 1, 20 - 32, 01.06.2008.
- Khan, B., & Khiyal, M. (2015). Final Grade Prediction of Secondary School Student Using Decision Tree. *International Journal of Computer Applications*, 32-36.

- Kim, B.-H., Vizitei, E., & Ganapathi, V. (2018). GritNet: Student Performance Prediction. *Udacity*.
- Kızılkaya, Y. M., & Oğuzlar, A. (2018). BAZI DENETİMLİ ÖĞRENME ALGORİTMALARININ R PROGRAMLAMA DİLİ İLE KIYASLANMASI. *Karadeniz Uluslararası Bilimsel Dergi*(37), 90-98.
- Koca, B., & Adem, K. (2022). Bilişim Teknolojileri Öğretmenlerinin Blok Tabanlı Kodlama Araçlarına İlişkin Öz Yeterlik İnançlarınınEvrışimsel Sinir Ağı ile Sınıflandırılması. *Avrupa Bilim ve Teknoloji Dergisi*(36), 50-54.
- Kökcü, B. N., Köse, R. D., Bulut, F., & Amasyalı, M. (2014). Kolektif Öğrenme Algoritmalarıyla Çocuklarda Obezite Hastalığına. *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu*, 200-205.
- Krizhevsky, A., Sutskever, I., & E. Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 1090–1098.
- Landwehr, N. (2003). *Logistic Model Trees*. Freiburg University, Department of Computer Science. Freiburg: Springer.
- Landwehr, N., Hall, M., & Frank, E. (2005). Logistic Model Trees. *2005 Springer Science + Business Media, Inc. Manufactured in The Netherlands*, 59, 161–205.
- López García, L. (2022). Tree to predict if students want to continue with their high school studies by implementing machine learning. *Journal Educational Theory*, 6-9.
- Manikandan, G., Vasudev, A., & Balasubramanian, A. (2018). A Survey to Identify an Efficient Classification Algorithm forHeart Disease Prediction. *International Journal of Pure and Applied Mathematics*, 119(12), 13337-13345.
- McDonald, J. H. (2022, Nisan 24). [https://stats.libretexts.org/Bookshelves/Applied_Statistics/Book%3A_Biological_Statistics_\(McDonald\)/05%3A_Tests_for_Multiple_Measurement_Variables/5.06%3A_Simple_Logistic_Regression](https://stats.libretexts.org/Bookshelves/Applied_Statistics/Book%3A_Biological_Statistics_(McDonald)/05%3A_Tests_for_Multiple_Measurement_Variables/5.06%3A_Simple_Logistic_Regression) adresinden alınmıştır
- Onan, A. (2018). Firma Başarısızlığının Tahmin Edilmesi İçin Kümelemeye Dayalı Bir Sınıflandırıcı. *alphanumeric journal*, 6(2).
- Onar, A. (2020, Eylül 21). [www.alitunacanonar.medium.com. https://alitunacanonar.medium.com/derin-%C3%B6%C4%9Frenme-y%C3%B6ntemleri-ve-uygulamalar%C4%B1-1ce215de40e8](https://alitunacanonar.medium.com/derin-%C3%B6%C4%9Frenme-y%C3%B6ntemleri-ve-uygulamalar%C4%B1-1ce215de40e8) adresinden alınmıştır
- Orhan, U. (2023, Ocak 28). <https://ceng.cu.edu.tr>. Ocak 28, 2023 tarihinde Çukurova Üniversitesi/Mühendislik Fakültesi/Bilgisayar Mühendisliği: <https://ceng.cu.edu.tr/uorhan/DersNotu/Ders07.pdf> adresinden alındı
- Özkan, Y., Önay Koçoğlu, F., & Selçukcan Erol, Ç. (2018). Öğrenci Performansının Yapay Zeka Derin Öğrenme Algoritması Ile Öngörülmesi. *INNOVATIONS IN LEARNING FOR THE FUTURE* (s. 136-145). İstanbul: İstanbul Üniversitesi.

- Peker, M. (2016). Predicting and Analyzing Students' Performance: An Educational Data Mining Approach . *2016 Sixth International Conference on Advanced Computing & Communication Technologies* , 36-40.
- Quy, T. L., Nguyen, T. H., Friege, G., & Ntoutsis, E. (2022). Evaluation of group fairness measures in student performance prediction problems. *Institute for Didactics of Mathematics and Physics*.
- Ragab, M., Abdel Aal, A. M., Jifri, A. O., & Omran, N. F. (2021). Enhancement of Predicting Students Performance Model Using Ensemble Approaches and Educational Data Mining Techniques. *Wireless Communications and Mobile Computing*.
- Rao, K., Rao, M., & Ramesh, B. (2016). Predicting Learning Behavior of Students using Classification Techniques. *International Journal of Computer Applications*, 15-19.
- Salameh Shreem, S., Turabieh, H., Al Azwar, S., & Baothman, F. (2022). Enhanced binary genetic algorithm as a feature selection to predict student performance. *Soft Computing*, 1811-1823.
- Salem, O. (2018). Event Detection in Wireless Body Area Networks Using Kalman filter and power divergence. *IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT*, 15(3), 1018-1034.
- Sezer, Ö. (2007). SINIF TEKRAR EDEN ÖĞRENCİLERİN BAZI DEMOGRAFİK ÖZELLİKLERİ İLE BU ÖĞRENCİLER VE ÖĞRETMENLERİNİN SINIF TEKRAR ETME HAKKINDAKİ GÖRÜŞLERİ. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, güz, 8(14), 31-48.
- Song, J., Lu, X., Liu, M., & Wu, X. (2011). A new LogitBoost algorithm for multiclass unbalanced data classification. *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 974-977.
- Soofi, A. A., & Awan, A. (2017). Classification Techniques in Machine Learning: Applications and Issues. *Journal of Basic & Applied Sciences*, 459-465.
- Statstates. (2022, Ocak 21). <https://www.statstest.com>. Ocak 21, 2023 tarihinde StatsTest: <https://www.statstest.com/simple-logistic-regression/> adresinden alındı
- Sumner, M., Frank, E., & Hall, M. (2005). Speeding up logistic model tree induction. *PKDD. Springer*, 675-683.
- Şama, E., & Tarım, K. (2007). ÖĞRETMENLERİN BAŞARISIZ OLARAK ALGILADIKLARI. *Türk Eğitim Bilimleri Dergisi*, Kış 2007, 5(1), 135-154.
- Şengür, D. (2013). Öğrencilerin akademik başarılarının veri madenciliği metotları ile tahmini. Eğitim Bilimleri Enstitüsü, Bilgisayar ve Öğretim Teknolojileri Eğitimi. Elazığ: Fırat Üniversitesi.
- Şimşek, H., Demiral, Y., Aslan, Ö., & Ünal Toğrul, B. (2012). Bir Üniversite Hastanesinde Koroner Kalp Hastalarına Uygulanan Tedavi Oranları,. *Dokuz Eylül Üniversitesi Tıp Fakültesi Dergisi*, 26(22), 111-117.

- Terzi, Y. (2023, Nisan 17). <https://avys.omu.edu.tr/>. Nisan 17, 2023 tarihinde Ondokuz Mayıs Üniversitesi Akademik Veri Yönetim Sistemi: <https://avys.omu.edu.tr/storage/app/public/yukselt/118305/Lojistik%20regresyon%20analizi.pdf> adresinden alındı
- Tosunoğlu, E., Yılmaz, R., Özeren, E., & Sağlam, Z. (2021). Eğitimde Makine Öğrenmesi: Araştırmalardaki Güncel Eğilimler Üzerine İnceleme. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, 3(2), 178-199.
- Uyanık, F., & Kasapbaşı, M. C. (2021). Telekomünikasyon Sektörü için Veri Madenciliği ve Makine Öğrenmesi Teknikleri ile Ayrılan Müşteri Analizi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9(3), 172-191.
- Uzun, E. (2023, Nisan 18). <https://erdincuzun.com>. Nisan 18, 2023 tarihinde https://erdincuzun.com/makine_ogrenmesi/oner/ adresinden alındı
- Ünal, F. (2020). *Data mining for student performance prediction in education*.
- Vanthienen, J. (1992). A more general comparison of the decision table and tree: A response. *DTEW Research Report 9224*.
- Vijayalakshmi, V., & Venkatachalapathy, K. (2019). Comparison of Predicting Student's Performance. *I.J. Intelligent Systems and Applications*, 34-45.
- Vural, T. (2017). *Veri Madenciliği Kullanılarak Mikrodalga Verileri İle Beyindeki Kitlesel Lezyonların Teşhisi*. Yıldız Teknik Üniversitesi, Matematik Mühendisliği Bilim Dalı. İstanbul: Yıldız Teknik Üniversitesi / Fen Bilimleri Enstitüsü / Matematik Mühendisliği Ana Bilim Dalı.
- Wang, Z., Wang, Y., & Srinivasan, R. (2018). A Novel Ensemble Learning Approach to Support Building Energy Use Prediction. *Energy and Buildings*, 159:109-122.
- WEKA. (2022, Aralık 25). *deeplearning4j:Deep Learning using Weka*. (WEKA, Düzenleyen) Aralık 25, 2022 tarihinde <https://deeplearning.cms.waikato.ac.nz/user-guide/data/> adresinden alındı
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (3. baskı b.). Morgan Kaufmann .
- Yadav, S. K., & Pal, S. (2012). Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification. *World of Computer Science and Information Technology Journal (WCSIT)*, 51-56.
- Yavuzarslan, M., & Erol, Ç. (2022). Öğrenme Yönetim Sistemi Log Kayıtlarının Akademik Başarı Tahmininde Kullanılması. *BİLİŞİM TEKNOLOJİLERİ DERGİSİ*, 15(2), 199-207.
- Yürekli, A. (2019). Lisansüstü Eğitim Enstitüsü Tez Yazım Kılavuzu. *Eskişehir Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü*, 1-12.

ÖZGEÇMİŞ

ORCID NO: 0009000547292172

Ad Soyad : Derya ÇINAR

Yabancı Dil : İngilizce

Eğitim ve Mesleki Geçmişi:

- 2005, Süleyman Demirel Üniversitesi, Teknik Eğitim Fakültesi, Elektronik ve Bilgisayar Eğitimi, Bilgisayar sistemleri Öğretmenliği
- 2016, Pamukkale Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği
- 2005-halen, Bilişim Teknolojileri Öğretmenliği, MEB

Yayınları ve/veya Bilimsel/Sanatsal Faaliyetleri:

- Derya Çınar, Sevcan Yılmaz Gündüz, “Prediction of Secondary School Students’ Mathematics Success with Machine Learning” 13th International Istanbul Scientific Research Congress on Life, Engineering and Applied Sciences Proceedings Book, S:94-104, ISBN: 978-605-72563-8-6, 9-30 Nisan 2023, İstanbul