

T.C.
İSTANBUL BEYKENT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI

**MOODLE SİSTEMİNDEKİ VERİLER İLE MAKİNE
ÖĞRENMESİ ALGORİTMALARI KULLANILARAK
ÖĞRENCİLERİN DÖNEM SONU AKADEMİK
PERFORMANSLARININ TAHMİN EDİLMESİ**

Yüksek Lisans Tezi

Tezi Hazırlayan
Buket DÖNMEZ

İstanbul, 2023

T.C.
İSTANBUL BEYKENT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI

**MOODLE SİSTEMİNDEKİ VERİLER İLE MAKİNE
ÖĞRENMESİ ALGORİTMALARI KULLANILARAK
ÖĞRENCİLERİN DÖNEM SONU AKADEMİK
PERFORMANSLARININ TAHMİN EDİLMESİ**
Yüksek Lisans Tezi

Tezi Hazırlayan
Buket DÖNMEZ

Öğrenci No
2220003097

ORCID ID
0000-0002-7510-2832

Danışman
Dr. Öğr. Üyesi Ege KİPMAN

İstanbul, 2023

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum **“Moodle Sistemindeki Veriler İle Makine Öğrenmesi Algoritmaları Kullanılarak Sonu Akademik Performanslarının Tahmin Edilmesi”** başlıklı bu çalışmanın, bilimsel ahlak ve geleneklere uygun şekilde tarafımdan yazıldığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, yararlandığım eserlerin tamamının kaynaklarda gösterildiğini ve çalışmamın içinde kullanıldıkları her yerde bunlara atıf yapıldığını, patent ve telif haklarını ihlal edici bir davranışımın olmadığını belirtir ve bunu onurumla doğrularım. 21/11/2023

Buket DÖNMEZ

T.C.
İSTANBUL BEYKENT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜ
TEZLİ YÜKSEK LİSANS SINAV TUTANAĞI

21/11/2023

Enstitümüz **Bilgisayar Mühendisliği** Anabilim Dalı **Bilgisayar Mühendisliği** Programı yüksek lisans öğrencilerinden 2220003097 numaralı **Buket DÖNMEZ**'in “*İstanbul Beykent Üniversitesi Lisansüstü Eğitim – Öğretim Yönetmeliği*”nin ilgili maddesine göre hazırlayarak, Enstitümüze teslim ettiği “**Moodle Sistemindeki Veriler ile Makine Öğrenmesi Algoritmaları Kullanılarak Öğrencilerin Dönem Sonu Akademik Performanslarının Tahmin Edilmesi**” konulu tezini, Yönetim Kurulumuzun 21/11/2023 tarih ve 2023/46 sayılı toplantısında seçilen ve On-Line toplanan biz jüri üyeleri huzurunda, İstanbul Beykent Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 29. maddesinin 3. fıkrası gereğince 45 dakika süre ile Zoom programı aracılığıyla on-line olarak aday tarafından savunulmuş ve sonuçta adayın tezi hakkında “**OYBİRLİĞİ**” ile “**KABUL**” kararı verilmiştir.

İşbu tutanak, 2 nüsha olarak hazırlanmış ve Enstitü Müdürlüğü'ne sunulmak üzere tarafımızdan düzenlenmiştir.

DANIŞMAN
Dr. Öğr. Üyesi Eg*** KI***
(İstanbul Beykent Üniversitesi)

ÜYE
Doç. Dr. At*** YI***
(İstanbul Beykent Üniversitesi)

ÜYE
Dr. Öğr. Üyesi Şe*** KO*** AY***
(İstanbul Gedik Üniversitesi)

Adı ve Soyadı : Buket DÖNMEZ
Danışmanı : Dr. Öğr. Üyesi Ege KİPMAN
Derecesi ve Tarihi : Yüksek Lisans (Tezli), 2023
Alanı : Bilgisayar Mühendisliği
Anahtar Kelimeler : Makine Öğrenmesi , Makine Öğrenmesi Algoritmaları, Yapay Zeka, MOODLE, Akademik Performans

ÖZ

MOODLE SİSTEMİNDEKİ VERİLER İLE MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANILARAK ÖĞRENCİLERİN DÖNEM SONU AKADEMİK PERFORMANSLARININ TAHMİN EDİLMESİ

Teknolojinin eğitim sektöründeki etkileri incelendiğinde, akademik performansın ölçümü ve değerlendirilmesi üzerindeki dönüştürücü gücü özellikle dikkat çekicidir. Bu alanda, yapay zeka ve makine öğrenimi algoritmaları, öğrencilerin gelecekteki akademik başarılarını tahmin etme kapasitesi ile ön plana çıkmaktadır. Öğrencilerin geçmiş başarıları, öğrenme süreçleri ve çeşitli faktörleri büyük veri analizine tabi tutarak, kişiselleştirilmiş eğitim yaklaşımlarına imkan tanımakta ve eğitim kurumlarının stratejik planlama becerilerini geliştirmektedir.

Bu araştırmada, MOODLE eğitim yönetim sistemi üzerinden elde edilen verilerin, beş farklı makine öğrenme algoritması (Random Forest, Karar Ağaçları, K-Nearest Neighbors, Support Vector Machines ve Naive Bayes) kullanılarak analiz edilmesi ele alınmıştır. İkili ve çoklu sınıflandırma teknikleriyle yürütülen bu analizler, algoritmaların öğrenci başarısını tahmin etmede gösterdikleri yüksek başarıyı ortaya koymuştur. Çalışma, makine öğrenme algoritmalarının eğitimde nasıl bir potansiyel taşıdığını ve performans tahmini gibi kritik süreçlerde ne derecede etkili olabileceğini göstermektedir. Literatür taramasıyla zenginleştirilen bu tez, benzer çalışmaların detaylı bir analizini sunmakta ve mevcut algoritmaların eğitim verileri üzerindeki uygulanabilirliğini değerlendirmektedir. MOODLE veri setinin işlenmesi, veri temizleme, özellik seçimi ve normalizasyon gibi ön işleme adımlarını kapsamaktadır.

Name and Surname : Buket DÖNMEZ
Supervisor : Asst. Prof. Dr. Ege KIPMAN
Degree and Date : Master's (Thesis), 2023
Major : Computer Engineering
Key Words : Machine Learning, Machine Learning, Algorithms,
Artificial Intelligence, MOODLE, Academic Performance

ABSTRACT

PREDICTING STUDENTS' ACADEMIC PERFORMANCE AT THE END OF THE SEMESTER BY USING MACHINE LEARNING ALGORITHMS WITH THE DATA IN MOODLE SYSTEM

The transformative power of technology in measuring and assessing academic performance is particularly striking when analysing the impact of technology in education. In this area, artificial intelligence and machine learning algorithms stand out for their ability to predict students' future academic success. By subjecting students' past performance, learning processes and various factors to big data analysis, it enables personalised educational approaches and improves the strategic planning capabilities of educational institutions.

In this study, the data obtained from the MOODLE learning management system were analysed using five different machine learning algorithms (Random Forest, Decision Trees, K-Nearest Neighbors, Support Vector Machines and Naive Bayes). These analyses, conducted with binary and multiple classification techniques, revealed the high success of the algorithms in predicting student achievement. The study shows the potential of machine learning algorithms in education and how effective they can be in critical processes such as performance prediction. Enriched with a literature review, this thesis provides a detailed analysis of similar studies and evaluates the applicability of existing algorithms on training data. The processing of the MOODLE dataset includes pre-processing steps such as data cleaning, feature selection and normalisation.

İÇİNDEKİLER

Sayfa No.

ÖZ

ABSTRACT

ŞEKİLLER LİSTESİ	v
TABLolar LİSTESİ	vi
KISALTMALAR	viii
SÖZLÜK.....	x
GİRİŞ	1

1. LİTERATÜR TARAMASI

2. YAPAY ZEKA, MAKİNE ÖĞRENMESİ, UZAKTAN EĞİTİM

2.1. Yapay Zeka.....	8
2.1.1. Yapay Zeka'nın Kısa Tarihi ve Gelişimi	9
2.1.2. Yapay Zeka ve Eğitim.....	10
2.1.3. Yapay Zeka'nın Eğitimde Uygulanması.....	11
2.1.4. Yapay Zeka'nın Eğitim de Kullanımının Avantaj ve Dezavantajları	12
2.2. Makine Öğrenimi	13
2.2.1. Çalışma Teknikleri	16
2.2.1.1. Danışmanlı (Denetimli) Öğrenme.....	17
2.2.1.2. Danışmansız (Denetimsiz) Öğrenme	17
2.2.1.3. Yarı Danışmanlı Öğrenme	18
2.2.1.4. Pekiştirmeli (Takviyeli) Öğrenme	19
2.2.2. Makine Öğrenmesi Algoritmaları	19
2.2.2.1. Karar Ağaçları (DT).....	20
2.2.2.2. Random Forest (RF).....	21
2.2.2.3. Naive Bayes	22
2.2.2.4. Komşu (K-Nearest Neighbors, KNN) Algoritması.....	23
2.2.2.5. Destek Vektör Makineleri (SVM).....	24
2.2.2.6. Doğrusal Olmayan Destek Vektör Makinesi	24
2.2.2.7. Doğrusal Destek Vektör Makineleri	26
2.2.2.8. Lojistik Regresyon	27

2.2.2.9. Extreme Gradient Boosting.....	28
2.2.2.10. Bagging	29
2.3. Uzaktan Eğitim	29
2.3.1. Uzaktan Eğitim Teknolojileri.....	31
2.3.2. Uzaktan Eğitim Modelleri.....	32
2.3.2.1. Eş Zamanlı (Senkron)	32
2.3.2.2. Eş Zamanlı Olmayan (Asenkron)	32
2.3.3. MOODLE Öğrenim Yönetim Sistemi	34
2.3.4. MOODLE Sistemi Üzerinden Toplanabilen Verilerin Örnekleri	36
2.3.5. MOODLE Sisteminin Öğrenci Performansının Analizi İçin Kullanılabilirliği.....	37
2.3.5.1. MOODLE İle Veri Toplamanın Avantajları.....	38
2.3.5.2. MOODLE İle Veri Toplamanın Dezavantajları.....	38
3. YÖNTEM	
3.1. Araştırma Modeli.....	40
3.2. Çalışma Kümesi	41
3.3. Veri Seti	41
3.3.1. Veri Setinin Hazırlanması	41
3.3.2. Öznitelik Bilgileri.....	42
3.3.3. Veri İşleme Araçları.....	42
3.3.3.1. Python	43
3.3.3.2. Pandas Kütüphanesi	43
3.3.3.3. Numpy Kütüphanesi	44
3.3.3.4. Scikit-Learn Kütüphanesi	44
3.3.3.5. Matplotlib Kütüphanesi.....	45
3.3.3.6. Anaconda	45
3.3.3.7. Jupyter Notebook	45
3.4. Veri Kaynakları ve Veri Toplama Araçları	46
3.4.1. Veri toplama yöntemleri ve teknikleri	46
3.4.2. CRISP-DM Yöntemi Süreçleri	47
3.4.2.1. İşin Anlaşılması ve Problemin Tanımlanması	47
3.4.2.2. Verinin Anlaşılması	48
3.4.2.3. Verinin Hazırlanması	49

3.4.2.4. Modelin Oluşturulması	50
3.4.2.5. Modelin Değerlendirilmesi	50
3.4.2.6. Canlıya Alma	51
3.4.3. Öğrenci Performansı Tahmini.....	52

4. BULGULAR

4.1. Veri Seti Nitelik Analizi.....	53
4.2. İkili Sınıflandırma.....	54
4.3. Çoklu Sınıflandırma	55
4.4. Kategorik Değişkenlerin Tip Dönüşümü	55
4.5. İkili Sınıflandırma Yapılarak Elde Edilen Sonuçlar	56
4.5.1. Random Forest Algoritması Sonuçları.....	56
4.5.1.1. Random Forest ile Önemli Özellikler Analizi	56
4.5.1.2. Random Forest En Önemli Beş Özellik ve Katsayıları.....	57
4.5.1.3. Random Forest Algoritma Sonuçları	58
4.5.1.4. Random Forest Tahmini İçin Çapraz Doğrulama	58
4.5.2. Karar Ağaçları İkili Sınıflandırma	59
4.5.2.1. Karar Ağaçları Algoritma Sonuçları	59
4.5.2.2. Karar Ağaçları Tahmini için Çapraz Doğrulama	59
4.5.3. Naive Bayes İkili Sınıflandırma.....	60
4.5.3.1. Naive Bayes Algoritma Sonuçları.....	60
4.5.3.2. Naive Bayes Tahmini için Çapraz Doğrulama	61
4.5.4. KNN İkili Sınıflandırma	61
4.5.4.1. KNN Algoritma Sonuçları	61
4.5.4.2. KNN Tahmini için Çapraz Doğrulama	62
4.5.5. SVM İkili Sınıflandırma	63
4.5.5.1. SVM Algoritma Sonuçları	63
4.5.5.2. SVM Tahmini için Çapraz Doğrulama	63
4.6. Çoklu Sınıflandırma Yapılarak Elde Edilen Sonuçlar.....	64
4.6.2. Çoklu Sınıflandırma Karar Ağaçları Algoritma Sonuçları	65
4.6.2.1. Çoklu Sınıflandırma Karar Ağaçları Cross Validation Sonuçları.....	67
4.6.2.2. Çoklu Sınıflandırma Karar Ağaçları GridSearch Sonuçları	67
4.6.3. Çoklu Sınıflandırma ile KNN Algoritma Sonuçları.....	68
4.6.3.1. Çoklu Sınıflandırma KNN Cross Validation Sonuçları	69

4.6.3.2. Çoklu Sınıflandırma KNN GridSearch Sonuçları.....	70
4.6.1. Random Forest ile Önemli Özellikler Analizi	71
4.6.1.1. Random Forest En Önemli Bes Özellik ve Katsayıları.....	71
Kaynak: Yazar tarafından oluşturulmuştur.	72
4.6.1.2. Çoklu Sınıflandırma ile Random Forest Algoritma Sonuçları.....	72
4.6.1.3. Çoklu Sınıflandırma Random Forest Çapraz Doğrulama Sonuçları.	72
4.6.1.4. Çoklu Sınıflandırma Random Forest GridSearch Sonuçları.....	73
4.6.4. Çoklu Sınıflandırma ile SVM Algoritma Sonuçları	74
4.6.4.1. Çoklu Sınıflandırma SVM Cross Validation Sonuçları.....	75
4.6.4.2. Çoklu Sınıflandırma SVN GridSearch Sonuçları	75
4.6.5. Çoklu Sınıflandırma ile Naive Bayes Algoritma Sonuçları	76
4.6.5.1. Çoklu Sınıflandırma Naive Bayes Cross Validation Sonuçları	77
4.6.5.2. Çoklu Sınıflandırma Naive Bayes GridSearch Sonuçları	77
SONUÇ	79
KAYNAKÇA.....	84
EKLER	
Ek-1: MOODLE Verileri Kullanım İzni	100

ŞEKİLLER LİSTESİ

	Sayfa No.
Şekil 1. Yapay Zeka, Makine Öğrenmesi ve Derin Öğrenme	15
Şekil 2. Makine Öğrenimi İçerisindeki Öğrenme Türlerinin Hiyerarşik Yapısı	16
Şekil 3. Karar Ağaçları Şematik Gösterimi	20
Şekil 4. Doğrusal Olmayan Regresyon.....	25
Şekil 5. Doğrusal Destek Vektör Makineleri.....	26
Şekil 6. Bütünleme Notuna Etki Eden Özellik Sıralaması	57
Şekil 7. Bütünleme Notuna Etki Eden Özellik Sıralaması	71
Şekil 8. İlk Beş Önemli Özellik.....	66

TABLÖLAR LİSTESİ

	Sayfa No.
Tablo 1. Veri Seti Nitelik Adları ve Türü	42
Tablo 2. Veri Seti Nitelik Adları ve Türü	53
Tablo 3. Veri Özelliklerinin Sayısal Değerleri	54
Tablo 4. Bütünleme Notu İkili Sınıflandırma	54
Tablo 5. Çoklu Sınıflandırma Not Dağılımı	55
Tablo 6. Algoritma Başarı Oranları ve Geçen Süre	56
Tablo 7. Etki Eden Özellikler ve Katsayıları	57
Tablo 8. Random Forest Tahmin Başarı Oranı	58
Tablo 9. Random Forest İçin Çapraz Doğrulama	58
Tablo 10. Karar Ağaçları Tahmin Başarı Oranı	59
Tablo 11. Karar Ağaçları İçin Çapraz Doğrulama Değerleri	60
Tablo 12. Naive Bayes Tahmin Başarı Oranı	60
Tablo 13. Naive Bayes İçin Çapraz Doğrulama Değerleri	61
Tablo 14. KNN Tahmin Başarı Oranı	62
Tablo 15. KNN İçin Çapraz Doğrulama Değerleri	62
Tablo 16. SVM Tahmin Başarı Oranı	63
Tablo 17. SVM İçin Çapraz Doğrulama Değerleri	64
Tablo 18. Tahmin Başarı Oranları ve Tahmin Süreleri.....	64
Tablo 19. Etki Eden Özellikler ve Katsayıları	72
Tablo 20. Random Forest Çoklu Sınıflandırma Başarı Oranı.....	72
Tablo 21. RF Çapraz Doğrulama Uygulanarak Elde Edilen Sonuçlar.....	73
Tablo 22. RF GridSearch Sonuçları	74
Tablo 23. Veri Seti Önem Katsayıları.....	65
Tablo 24. Karar Ağaçları Çoklu Sınıflandırma Başarı Oranı	66
Tablo 25. KA Cross Validation Uygulanarak Elde Edilen Sonuçlar	67
Tablo 26. Karar Ağaçları GridSearch Sonuçları	68
Tablo 27. KNN Çoklu Sınıflandırma Başarı Oranı.....	69
Tablo 28. KNN Cross Validation Uygulanarak Elde Edilen Sonuçlar	69
Tablo 29. KNN GridSearch Sonuçları	70

Tablo 30. SVM Çoklu Sınıflandırma Başarı Oranı.....	74
Tablo 31. SVM Cross Validation Uygulanarak Elde Edilen Sonuçlar	75
Tablo 32. SVM GridSearch Sonuçları	76
Tablo 33. Naive Bayes Çoklu Sınıflandırma Başarı Oranı	76
Tablo 34. NB Cross Validation Uygulanarak Elde Edilen Sonuçlar	77
Tablo 35. Naive Bayes GridSearch Sonuçları.....	78



KISALTMALAR

Akt.	: Aktaran
ANO	: Ağırlıklı Not Ortalaması
BDÖ	: Bilgisayar destekli öğretimin
Bkz	: Bakınız
CSV	: Virgülle Ayrılmış Olarak Veriler
Çev.	: Ceviren
DM	: Data Mining (Veri Madenciliği)
DTÖS	: Diyalog tabanlı öğretici sistemler
DVM	: Doğrusal Destek Vektör Makineleri
EVM	: Eğitimde Veri Madenciliği
GANÖ	: Genel Ağırlıklı Not Ortalaması
IBM	: Bilişim Teknoloji Şirketi
IOS	: Iphone İşletim Sistemi
IQ	: Zeka Seviyesi
KA	: Karar Ağaçları
KNN	: K-Nearest Neighbors
LMS	: Öğrenme Yönetim Sistemi (Learning Management System)
LR	: Logistic Regression
LRA	: Lojistik regresyon analizi
MIT	: Massachusetts Teknoloji Enstitüsü
ML	: Makine Öğrenmesi
MOODLE	: Modular-Object-Oriented-Dynamic-Learning-Environment yani Esnek Nesne Yönelimli Dinamik Öğrenme Ortamı
NB	: Naive Bayes
OBS	: Öğrenci Bilgi Sistemi
ÖYS	: Öğrenme Yönetim Sistemi
PCA	: Temel Bileşen Analizi
RF	: Random Forest
RF	: Rastgele Ormanlar RO (Random Forest
SVM	: Destek Vektör Makineleri

URL : Tekduzen Kaynak Bulucu
vd. : Ve diğlerleri
YSA : Yapay Sinir Ağları
YZ : Yapay Zeka (Artificial Intelligence - AI),



SÖZLÜK

CSV: Değerleri ayırmak için virgül kullanan bir metin dosyası biçimidir.

Ensemble: Bir öğrenci ile modeli eğitmek yerine, birden fazla öğrenci ile modeli oluşturmayı sağlayan öğrenmedir

F1-Score (F1 Skoru): F1 skoru, kesinlik ve duyarlılık arasında bir denge sağlayan bir metriktir. Özellikle dengeli sınıflandırma problemlerinde kullanılır.

MOODLE: Özgür ve açık kaynak kodlu bir uzaktan eğitim sistemidir.

Precision (Kesinlik): Kesinlik, doğru pozitif tahminlerin toplam pozitif tahminlere oranını ifade eder. Kesinlik, "kesin olarak pozitif tahminlerin ne kadarının doğru olduğunu" gösterir.

Recall (Duyarlılık): Duyarlılık, gerçek pozitif tahminlerin toplam pozitif verilere oranını ifade eder. Duyarlılık, "gerçek pozitiflerin ne kadarının tespit edildiğini" gösterir.

Support (Destek): Destek, her bir sınıf için tahmin edilen örnek sayısını ifade eder. Bu metrik, her sınıfın model tarafından ne kadar desteklendiğini gösterir.

GİRİŞ

Bu platform, öğrenci performansının değerlendirilmesi ve öğretim sürecinin iyileştirilmesi için değerli veriler sağlamaktadır.

Tez çalışmasının amacı, MOODLE sistemi üzerindeki verilerin kullanımıyla öğrencilerin dönem sonu akademik performanslarının tahmin edilmesidir. Bu çalışma, öğrenci performansının daha iyi anlaşılması, müfredatın ve öğretim sürecinin iyileştirilmesi için bir temel sağlamayı hedeflemektedir. Veri analitiği ve makine öğrenmesi tekniklerinin eğitim alanında kullanımının artmasıyla, bu alanda yapılan araştırmaların önemi ve etkisi de artmaktadır.

Çalışmada, öncelikle literatür taraması yapılmış ve MOODLE sistemi üzerindeki verilerin kullanılmasıyla öğrenci performansını tahmin etme konusunda yapılan çalışmalar incelenmiştir. Bu çalışmaların sonuçları ve kullanılan yöntemler, tezin temelini oluşturmaktadır

Ardından, veri toplama ve ön işleme adımları gerçekleştirilmiştir. MOODLE sistemi üzerindeki öğrenci etkileşimleri, sınav sonuçları, ödev teslimleri ve diğer ilgili veriler toplanmış ve bu verilerin analizi için ön işleme adımları uygulanmıştır

Daha sonra, farklı makine öğrenmesi algoritmaları uygulanmıştır. Karar ağaçları, yapay sinir ağları, destek vektör makineleri, random forest gibi algoritmalar, öğrenci performansını tahmin etmek için kullanılmış ve bu algoritmaların performansları karşılaştırılmıştır

Sonuçlar ve tartışma bölümünde, uygulanan algoritmaların performansı ve elde edilen sonuçlar ayrıntılı olarak incelenmiştir. Ayrıca, elde edilen bulguların öğretim sürecine etkisi ve bu çalışmanın sınırlamaları da tartışılmıştır.

Sonuç olarak, bu tez çalışması, MOODLE sistemindeki verilerin kullanılmasıyla öğrencilerin dönem sonu akademik performanslarının tahmin edilmesi konusunda bir katkı sağlamayı amaçlamaktadır. Bu çalışma, öğretmenlere ve yöneticilere öğrenci performansını takip etmek, müdahale etmek ve öğrenme sürecini iyileştirmek için değerli bir araç sunmayı hedeflemektedir.

Yapılan araştırma kapsamında MOODLE üzerinde yer alan gerçek veriler ile makine öğrenmesi algoritmaları kullanılarak öğrencilerin dönem sonu akademik performanslarının tahminini yapabilecek algoritma için 5 farklı makine öğrenmesi

algoritması kullanılarak en etkili model geliştirilmiştir. Elde edilen bilgiler, öğrencilerin dönem sonu akademik başarılarının tahmini yapılarak başarılarını arttırabilecek yöntemler geliştirmek ve öğrencilerin başarısız olmalarının önüne geçmeyi sağlamak. Araştırmada önerilecek modelin diğer eğitim kurumlarında kullanılabilir. Bu çalışma, makine öğrenmesi gibi yapay zekanın bir bileşenini kullanarak, öğrencilerin akademik performanslarının izlenmesi, rehberlik sağlanma konusunda önemli bir katkı sağlayabilir.

Araştırmada kapsamında öğrenci verisi üzerine veri madenciliği teknikleri uygulanarak elde edilen modeller, rastlantısal bağlantılara dayanmamaktadır.



1. LİTERATÜR TARAMASI

Günümüzde eğitim teknolojilerinin gelişmesiyle birlikte, eğitim süreçleri ve öğrenci performanslarının değerlendirilmesi için veri madenciliği teknikleri kullanılmaktadır. Çevrimiçi öğrenme yönetim sistemlerinden biri olan MOODLE (Learning Management System - LMS) sistemi birçok eğitim kurumu tarafından tercih edilmektedir. Öğrencilerin dönem sonu akademik performanslarını tahmin etmek, eğitimcilerin öğrencilerin gereksinimlerini belirlemeleri ve müdahale etmeleri için önemli bir adımdır.

Makine öğrenmesi algoritmaları tahmin yapmak için ideal algoritmalar olduğu bilinmektedir. Bu algoritmalar, büyük veri kümelerinden öğrenir ve öğrencilerin gelecekteki performansını tahmin etmek için bu bilgileri kullanır. MOODLE veri seti, makine öğrenmesi algoritmalarının öğrenmesi ve uygulanması için çok uygun olduğu görülmektedir. Bunun nedeni genellikle çeşitli öğrenci bilgilerini içerir: notlar, katılım, forum etkinliği, ders materyalleri etkileşimi vb. Bu kapsamda, eğitim yönetim sistemleri üzerinde yapılan çalışmalar da oldukça önem kazanmıştır. Bu tez çalışması, MOODLE eğitim yönetim sistemi üzerindeki verilerin makine öğrenmesi algoritmaları kullanılarak analiz edilmesiyle öğrencilerin dönem sonu akademik performanslarının tahmin edilmesini amaçlamaktadır.

Literatür taraması sonucunda, öğrenci performanslarının tahmini için özellikle makine öğrenmesi algoritmalarının kullanıldığı birçok çalışmanın bulunduğu görülmüştür. Bu çalışmalar arasında, verilerin toplanması, öznelitliklerin seçimi, model oluşturma ve değerlendirme aşamaları gibi konulara yer veren birçok yöntem bulunmaktadır. Ayrıca, öğrencilerin eğitim yönetim sistemlerindeki etkileşimlerinin de performans tahmininde önemli bir rol oynadığı belirtilmektedir.

Aşağıdaki literatür çalışması, seçilen bazı araştırmaların sonuçlarına dayanarak, yapay zeka destekli makine öğrenmesi algoritmalarının kullanımı ile öğrencilerin akademik performanslarının tahmin edilebileceği konusunda benzer sonuçlara ulaşmıştır.

Amrieh vd. (2016) tarafından yapılan bu araştırma, Moodle sistemi üzerinden toplanan verileri ensemble öğrenme yöntemleriyle analiz ederek öğrencilerin dönem sonundaki başarılarını önceden tahmin etmeyi hedeflemektedir. Araştırmada, öğrenci

başarısını tahmin etmek için hangi değişkenlerin kritik olduğunu belirlemek amacıyla Random Forest, Gradient Boosting ve AdaBoost gibi ensemble öğrenme algoritmaları kullanılmıştır.

Shahiri vd. 2020 tarihli çalışmasında, öğrenci performansının tahminlenmesi için kullanılabilecek farklı veri madenciliği ve makine öğrenmesi yöntemleri ele alınmaktadır. Ayrıca, bu tahminlemede kullanılabilecek verilerin kaynaklarından biri olarak MOODLE'den alınan bilgileri öne sürmektedirler.

Yağcı'nın 2022 yılında gerçekleştirdiği "Educational data mining: prediction of students' academic performance using machine learning algorithms" başlıklı çalışmasında, makine öğrenimi teknikleri kullanarak üniversite öğrencilerinin final sınav sonuçlarını tahmin etmek amacıyla bir model önerilmektedir. Bu model, ara sınav sonuçları ile fakülte ve bölüm bilgileri gibi değişkenleri içermektedir. Seçilen algoritmalar Random Forests, SVM, Lojistik Regresyon, Naïve Bayes ve k-NN ile model, %70-75 aralığında bir doğruluk oranıyla sınıflandırma yapabilmektedir. Bu sonuçlar, bu tür makine öğrenimi modellerinin öğrenci sınav performans tahminlerinde etkili bir şekilde kullanılabileceğini ve bu tahminlerin öğrencilerin başarısını artırmada yardımcı bir araç olarak değerlendirilebileceğini göstermektedir. Ayrıca, çalışma bu alanda yapılacak gelecekteki araştırmalara yönelik olarak modelin ve kullanılan parametrelerin genişletilmesi konusunda literatüre katkıda bulunmaktadır.

Zhang vd. 2020 tarihli çalışması "Using learning analytics to predict students performance in MOODLE LMS" başlığı altında, MOODLE Öğrenim Yönetim Sistemi (LMS) içerisinde öğrenci başarısının tahminlenmesinde Öğrenme Analitiği'nin nasıl kullanılabileceğini ele almaktadır. Bu analizler, MOODLE LMS platformunda yer alan öğrenci etkileşimleri, online etkinlikler, sınavlar ve ödevler gibi verilere dayanarak gerçekleştirilmektedir. Çalışmada, Öğrenme Analitiği yaklaşımları ve Makine Öğrenmesi tekniklerinin birleştirilerek kullanılmasının sonuçları incelenmiş ve bu yöntemlerin öğrenci performansının tahminindeki etkinlikleri üzerine değerlendirmeler yapılmıştır.

Karadayı ve Küçük'ün 2019 tarihli çalışmasında, öğrenci performansının tahminlenmesi sürecinde karar ağacı algoritmasının nasıl kullanılabileceği konusu üzerinde durulmaktadır. Bu araştırmada, öğrenci başarısını doğrudan veya dolaylı

şekilde etkileyebilecek olan faktörlerin neler olduğu tespit edilmeye çalışılmıştır. Tespit edilen bu faktörler temel alınarak karar ağacı algoritmasıyla öğrenci performansının tahmin edilme yöntemi incelenmiştir. Çalışmanın sonuç kısmında, gerçekte yapılan bir uygulamanın sonuçları ile karar ağacı algoritmasının öğrenci başarısını tahminlemedeki başarısı değerlendirilip ele alınmıştır.

Hatice Nazlı Bastem'in 2021 yılında yaptığı çalışma, öğrenci akademik performansının tahminlenmesi konusunda yapay zeka ile desteklenen makine öğrenmesi algoritmalarını mercek altına almaktadır. Bu çalışma, literatür incelemesi ve veri analizi gibi yöntemlerle gerçekleştirilmiş olup, sonuçlarına göre makine öğrenmesi yöntemleriyle desteklenen yapay zeka uygulamaları, öğrencilerin akademik başarısını tahmin etmek için oldukça etkili bir yaklaşım sunmaktadır. Araştırmada, öğrenci performansının tahminlenmesinde sıkça başvuru alan algoritmalarından bazıları - karar ağaçları, destek vektör makineleri, yapay sinir ağları, m ve rasgele ormanlar gibi - detaylı olarak ele alınmıştır.

Yıldız ve Çelik (2021) çalışmasında, modern yapay zeka teknolojilerinden biri olan derin öğrenme algoritmalarının, öğrenci başarısını tahmin etmede nasıl bir etkiye sahip olduğu üzerine detaylı bir analiz sunmaktadırlar. Bu çalışma, öğrencilere ait sınav sonuçları, ödevler gibi çeşitli akademik verileri kullanarak, bu verilerin derin öğrenme teknikleriyle nasıl modellediğini ve bu modellemenin sonucunda ne kadar başarılı tahminler yapılabildiğini incelemiştir. Çalışmanın kritik bir yönü, bu algoritmaların uygulanabilirliği ve etkinliği hakkında detaylı bilgi veren sonuçlarıdır. Bu, eğitim alanında, özellikle eğitim teknolojilerinin ve yapay zeka uygulamalarının giderek daha fazla benimsendiği bu dönemde, pedagojik yaklaşımların ve öğrenci değerlendirme yöntemlerinin daha da iyileştirilmesine katkıda bulunabilir.

Yıldız ve Kaya'nın 2020 tarihli çalışması, makine öğrenmesi tekniklerinin öğrenci başarısının tahmininde ne kadar etkili olduğunu araştırmıştır. Araştırma kapsamında, öğrencilerin Türkçe, matematik ve fen bilimleri derslerindeki performansları bu tekniklerle ölçülmüştür. Araştırmanın bulguları, makine öğrenmesi teknolojisinin öğrenci performans tahminlerinde etkin bir araç olduğunu ortaya koymuştur.

Dalkılıç ve Kurt, 2019 yılında, öğrenci performansını tahmin etmek amacıyla yapay sinir ağları metodolojisiyle bir araştırma yaptılar. Matematik dersi başarısını,

öğrenciye özgü verilere (yaş, cinsiyet, anne-baba eğitim seviyesi gibi) dayanarak tahmin etmeyi amaçladılar. Bu çalışmanın sonuçları ve metodolojisi makalede detaylandırılmıştır.

2023'te Bingöl, eğitsel veri madenciliği ve yapay zekâ tekniklerini kullanarak öğrenci başarısının nasıl tahmin edilebileceğine dair bir çalışma gerçekleştirdi. Yapılan literatür taraması, yapay zekâ ve makine öğrenmesi tekniklerinin, öğrenci performansının tahmininde etkili olduğunu gösterdi. Bingöl'ün çalışması, eğitim alanındaki çeşitli makine öğrenimi algoritmalarını vurgulayarak, eğitimcilere öğrenci performans değerlendirmesinde ve strateji geliştirmede yardımcı olmayı hedefliyor.

Paramita'nın 2021 tarihli çalışması, e-öğrenme ortamında öğrenci performansının tahmin edilmesine odaklanıyor. Çalışma, Open University'e ait büyük bir öğrenci veri seti üzerinde, birçok farklı algoritma kullanarak yapıldı. Özellikle k-NN ve ANN algoritmaları, bu çalışmada dikkate değer sonuçlar elde etti. Araştırma, makine öğreniminin öğrenci başarısını tahminlemedeki potansiyelini vurgulayarak, gelecekteki çalışmalara zemin hazırlıyor.

2006 yılında Zhang, Zhou, Briggs ve Nunamaker, çevrimiçi öğrenme ortamlarında öğrenci performansının nasıl tahmin edilebileceğine dair geniş çaplı bir meta-analiz yayınladılar. Bu analizde, farklı öğrenci veri kaynaklarına dayanarak yapılan araştırmaların sonuçları değerlendirildi. Bu genel inceleme, çevrimiçi öğrenme ortamında öğrenci başarısını tahmin etmek için kullanılan yöntemlerin genelde başarılı olduğunu belirtiyor.

Tüm bu çalışmalar, yapay zeka destekli makine öğrenmesi algoritmalarının öğrenci performansının tahmin edilmesi için etkili bir yöntem olduğunu göstermektedir. Bu alanda yapılacak daha kapsamlı çalışmalar, öğrencilerin akademik performansının tahmin edilmesinde yapay zeka destekli makine öğrenmesi algoritmalarının daha da geliştirilmesine katkı sağlayabilir.

Öğrencilerin akademik performansının MOODLE üzerindeki verilerle tahmin edilmesi üzerine birçok araştırma yapılmıştır. Bu araştırmalar, öğrencilerin MOODLE platformundaki etkileşimleri, sınav sonuçları, ödev teslimleri ve diğer verileri kullanarak öğrenci performansını tahmin etmeyi amaçlamaktadır. Bu tür araştırmalar, eğitimde veri analitiği ve öğrenci değerlendirme süreçlerinde yapay zeka tekniklerinin kullanılması konularında önemli bir rol oynamaktadır.

Bu literatür taraması, MOODLE eğitim yönetim sistemi ve diğer çevrimiçi öğrenme ortamları üzerindeki verilerin makine öğrenmesi algoritmaları kullanılarak analiz edilerek öğrencilerin dönem sonu akademik performanslarının tahmin edilmesi konusunda yapılan araştırmaların genel bir incelemesini sunmaktadır.

Sonuç olarak, yapay zeka destekli makine öğrenmesi algoritmaları, öğrencilerin akademik performanslarının tahmini için etkili bir yöntem olarak kabul edilmektedir. Bu alanda yapılacak daha kapsamlı çalışmalar, öğrencilerin akademik başarılarını artırmak için önemli bir katkı sağlayabilir.



2. YAPAY ZEKA, MAKİNE ÖĞRENMESİ, UZAKTAN EĞİTİM

Yapay zeka kavramı ve kısa tarihi ve gelişimi kavramları incelenmiş ve yapay zekanın eğitim ile bağlantısı araştırılmıştır. Yapay zekanın eğitimde uygulanması ile ilgili kavramlar ve uygulamalara değinilmiş, avantaj ve dezavantajları araştırılmıştır. Makine öğrenmesi kavramı, makine öğrenimi algoritmaları incelenmiştir.

2.1. Yapay Zeka

Yapay Zeka, günümüzde hızla gelişen teknolojik alanlardan biri olup, Yılmaz'a (2019) göre insan düşünce yeteneklerini bilgisayar sistemlerinde modellemeyi hedefler. Bu alandaki kökenleri 1950'li yıllara dayanan AI, Chollet (2019) tarafından bilgisayar bilimcilerinin "bilgisayarlar düşünebilir mi?" sorusu etrafında tanımlanmıştır. Matematik, psikoloji, dil bilimi ve bilgisayar bilimleri gibi farklı bilim dallarından beslenen Yapay Zeka, 1956'da Dartmouth Koleji'nde düzenlenen konferansta ilk kez resmi bir isimle anıldı. Toosi, Bottino, Saboury, Siegel ve Rahmim'e (2021) göre, John McCarthy bu disiplini "zeki makineler yapma bilimi ve mühendisliği" olarak tanımlamıştır. Bu tanımlamaların yanı sıra, Nabiye (2012) yapay zeka çalışmalarının amacının, yüksek düzeydeki bilişsel süreçleri taklit etmek olduğunu belirtir.

Akpınar'a (2019) göre, Yapay Zeka, makinelerin insana benzer düşünce ve özellikler kazanmasını amaçlar. Bu amacı destekleyen Chassigno vd. (2018), makinelerin insana benzer performans sergilemesi gerektiğini vurgular. Böylece, günümüzde birçok alanda, özellikle yüz tanıma, otonom sürüş sistemleri ve akıllı asistanlar gibi uygulamalarda AI'nin aktif olarak kullanıldığına tanık olmaktadır. Riedel vd. (2017), (Akt., Kır ve Şenocak, 2022), eğitimde AI'nin önemine dikkat çekerken, AI'nin alt dallarının makine öğrenimi ve doğal dil işleme gibi konuları kapsadığını belirtir. Ayrıca Charniak vd. (2018) AI'nin temel hedefinin, insanın yaptığı işleri bilgisayarlar aracılığıyla gerçekleştirmek olduğunu ifade eder, bu da bu alandaki çalışmaların yoğun bir şekilde sürdüğüne işaret eder.

Yapay Zeka'nın eğitim sektöründe de büyük bir potansiyeli olduğu aşikardır. Zawacki-Richter ve ekibinin 2019'da yürüttüğü çalışmada, eğitimde yapay zeka uygulamaları dört temel boyutta incelenmiştir. İlk olarak, "Profil Oluşturma ve

Tahmin" boyutu, öğrenci verilerini derinlemesine analiz ederek eğitim planlamasına ve öğrenci performansı tahminlerine yönlendirmede bulunmayı amaçlar. "Akıllı Öğretim Sistemleri" boyutu ise, öğrencilerin bireysel öğrenme gereksinimlerini tanımlayarak onlara özelleştirilmiş geri bildirimler sunar. "Ölçme ve Değerlendirme" boyutu, eğitimin kalitesini değerlendirme ve öğrenci katılımını ölçme kapasitesi sunarken; "Uyarlanabilir Sistemler ve Kişiselleştirme" boyutu, öğrencilere bireysel öğrenme deneyimleri sunmayı hedefler.

Eğitimde bu boyutların tanımlanması, yapay zeka uygulamalarının eğitimdeki geniş yelpazesini ve potansiyelini ortaya koymaktadır. Yapay Zeka, sadece eğitim alanında değil, sağlık, otomasyon, sanat ve eğlence gibi birçok sektörde de teknolojik devrimin öncüsü olmuştur. Günümüzde, bu teknolojik gelişmeler sayesinde, insan zekasının benzeri özellikleri makinelerde görmekteyiz. Bu kapsamlı entegrasyon, Yapay Zeka'nın geleceğin teknolojik evrimindeki rolünün ne denli kritik olduğunu vurgular. Bu nedenle, Yapay Zeka'nın daha fazla araştırılması ve potansiyel uygulama alanlarının keşfedilmesi, modern dünyanın en önemli gündem maddelerinden biridir.

2.1.1. Yapay Zeka'nın Kısa Tarihi ve Gelişimi

Yapay Zeka kavramı, Adipat ve diğerleri (2022) ile Dennehy (2020) tarafından da vurgulandığı gibi, modern teknolojinin hızla ilerlemesi ve dijitalleşme süreciyle daha belirgin bir şekilde hayatımızda yer almaya başladı. Fakat, Yapay Zeka fikri aslında çok daha eski zamanlara, hatta antik Yunan mitolojisine kadar uzanmaktadır. Mayor (2018) (Akt. McTear, 2020) bu mitolojik öykülerde günümüz Yapay Zeka teknolojisiyle benzerlik gösteren karakterlere işaret ederken, Skalfist (2019) (Akt., Okur, 2023) bu karakterleri, Yapay Zeka'nın potansiyelini simgeleyen formlar olarak tanımlamaktadır.

Tarihsel olarak bakıldığında, 17. yüzyılda Yapay Zeka'nın temellerinin atıldığını görebiliriz. Sayar (2021), bu dönemde Gottfried Leibniz'in "Adım Hesaplayıcı" adlı aleti icat etmesiyle bilgisayarın temelini atıldığını belirtirken, George Boole'un mantık ilkesi ve Bool Cebiri'ni tanıtarak bilgisayar bilimine önemli katkılarda bulunduğunu ifade eder. Ayrıca, Frege'nin mantıksal önermeleri grafiklerle ifade edebilen bir sistem oluşturduğuna dikkat çeker. 18. yüzyılda ise, Thomas

Bayes'in Bayes Teoremi, Yapay Zeka çalışmalarına kritik bir katkı sağlamıştır. Bu teoremin, Nilsson (2019) tarafından da belirtildiği gibi, mühendislikte temel bir rol oynamıştır. Teknolojik ilerlemelerle paralel olarak, Joseph Marie Jacquard'ın 1801'de programlanabilir kartlarla dokuma tezgahlarını otomatikleştirmesi, programlamanın önemli bir aşamasını temsil eder. Charles Babbage'in tasarladığı "Fark Motoru" ve "Analitik Motor" Yapay Zeka'nın gelişimine önemli katkılar sağlamıştır. Ada Lovelace'in bu motorlar için yazdığı programlar, Yapay Zeka'nın evriminde kritik bir role sahiptir. Bu tarihsel dönüm noktaları, Yapay Zeka'nın bugünkü karmaşık yapı ve uygulamalarının temelini atmıştır.

2.1.2. Yapay Zeka ve Eğitim

Teknolojik ilerlemeler sayesinde eğitimin sınırları genişlemiş, geleneksel sınıf ortamlarının dışına taşarak her yerde erişilebilir bir hâl almıştır. Bu evrimde, yapay zekanın eğitimdeki rolü giderek artmaktadır. Yapay zeka, eğitimde bireyler için kişiselleştirilmiş öğrenme deneyimleri, proje tabanlı öğrenme ve esnek servis sunmaktadır, bu da öğrencilere daha etkin ve bireysel öğrenme fırsatları sağlamaktadır (Smith, 2021).

Massachusetts Teknoloji Enstitüsü (MIT) Yapay Zeka Laboratuvarı yöneticisi Edward Fredkin'in tarihsel perspektiften bakarak yapay zekanın önemini vurgulaması, bu teknolojinin evrenin oluşumu ve canlılığın başlangıcı kadar kritik bir aşamada olduğunu göstermektedir (Copeland, 1998). Ancak yapay zekanın eğitimdeki asıl etkisi, insanoğlunun bilgiye erişim ve kullanım şeklini değiştirerek, öğrenme yöntemlerini ve süreçlerini dönüştürmüş olmasıdır (Tuğluk ve Gök-Çolak, 2019).

Yapay zeka, eğitimde bilişsel yetenekleri geliştirmeye odaklanırken, aynı zamanda öğrencilere yapay zeka sistemleriyle uyumlu çalışabilecek yetenekler kazandırmayı da hedeflemektedir (Günay ve Şişman, 2019). Ancak bu, eğitim sistemlerinin sadece teknik becerilere odaklanması gerektiği anlamına gelmez. Yaratıcı ve eleştirel düşünme gibi insana özgü yeteneklerin yapay zeka tarafından tam anlamıyla benimsenememesi, bu yeteneklerin eğitimdeki önemini daha da vurgulamaktadır (Sariel, 2017; Akt., Coşkun, F. ve Gülleroğlu, H. D., 2021). Bu nedenle, modern eğitim sistemleri, teknik becerilerin yanı sıra yaratıcılığı ve eleştirel düşünmeyi de teşvik etmektedir.

Devletlerin eğitim stratejileri, yapay zeka sistemlerini de dikkate alarak şekilleniyor. Özellikle büyük güçler olan ABD, Rusya, Çin ve Avrupa'nın bu alanda yaptığı yatırımlar, uluslararası eğitim standartlarını ve değerlendirme yöntemlerini etkiliyor. OECD'nin, eğitim değerlendirme programlarına dijital okuryazarlık ve yaratıcılık gibi becerileri eklemesi, bu değişikliğin bir yansımasıdır (Günay ve Şişman, 2019).

Sonuç olarak, yapay zeka, eğitim sistemlerini hem içerik hem de yöntem açısından yeniden şekillendiriyor. Teknolojik ilerlemenin yanı sıra, insana özgü becerilere verilen değerin korunması, dengeli bir eğitim sisteminin oluşturulmasında kritik bir rol oynamaktadır (Karaca ve Telli, 2019).

2.1.3. Yapay Zeka'nın Eğitimde Uygulanması

Yapay zekayı eğitim sürecinde kullanılan birçok sistemi bulunmaktadır. Bunlar uzman sistemler, akıllı öğretici sistemler ve diyalog tabanlı öğretici sistemler olarak incelenmiştir.

- **Uzman Sistemler**

Son yıllarda yapay zeka alanındaki ilerlemeler, uzman sistemlerin gelişimine katkıda bulunmuştur. Uzman sistemler, belirli bir alanda uzmanlaşmış görevleri yapay zeka algoritmaları ile yerine getiren bilgisayar programlarıdır. Genellikle dört ana bileşeni vardır: bilgi güncelleme, bilgi tabanı, çıkarım/karar mekanizması ve kullanıcı arayüzü (Önder, 2003).

Yapay zeka genel problemleri çözmeye yönelikken, uzman sistemler özelleşmiş alanlara odaklanır. Örneğin, bir öğretmen sınıf yönetimi sorunlarına deneyimleriyle yanıt verebilirken, uzman sistemler çok daha geniş bir veri tabanını analiz ederek çözüm sunabilir. Uzman sistemler, kullanıcının karşısındaki problemleri çözmek için insan bilgisini temsil eden bilgi tabanlı sistemlerdir ve diğer bilgi sistemleri arasında en karmaşık yapıya sahiptir (Öz ve Baykoç, 2004, s. 276).

Eğitimde, uzman sistemlerin uygulama alanlarından biri uzaktan eğitimidir. Tıp eğitimi için geliştirilen MYCIN gibi sistemler, doktorlara teşhis ve tedavi konularında yardımcı olabilir (Holmes vd., 2019).

- **Akıllı Öğretici Sistemler**

Yapay zekanın eğitimdeki rolü evrilmeye devam ediyor. 1960'lar ve 1970'lerde bilgisayar destekli öğretimin (BDÖ) yükselmesi, eğitimde bir dönüm noktasıydı. Bu BDÖ sistemlerinin sınırlılıklarını aşmak adına akıllı öğretici sistemler (AÖS) ortaya çıktı. Bu sistemler, öğrencilere kişiselleştirilmiş öğrenme deneyimleri sunar. SCHOLAR gibi erken dönem AÖS'ler, Sokratik diyalog ilkesini benimseyerek çalışır. Akıllı öğretici sistemler, öğrencilere kişiselleştirilmiş ve verimli öğrenme deneyimleri sunma potansiyeline sahip olan yapay zeka uygulamalarıdır.

- **Diyalog Tabanlı Öğretici Sistemler**

Akıllı öğretici sistemlerin gelişimi, diyalog tabanlı öğretici sistemlerin (DTÖS) doğmasına zemin hazırlamıştır. CIRCSIM gibi sistemler, tıp eğitimi için geliştirilmiş olup, öğrencilere birebir öğretici diyaloglar oluşturarak dönüt sağlar (Rovick ve Michael, 1986; Akt. Arslan, K., 2020). Birebir öğretim, öğrencilere en etkili öğrenme deneyimlerinden birini sunar (Bloom, 1984). AutoTutor gibi diğer sistemler, öğrencilere derinlemesine bilgi kazandırmak için diyalog tabanlı yaklaşımı benimser. Bu sistemler, öğrencinin cevaplarını hızla değerlendirip geri dönüş sağlar, bu da öğrenme deneyiminin geliştirilmesine yardımcı olur.

2.1.4. Yapay Zeka'nın Eğitim de Kullanımının Avantaj ve Dezavantajları

Yapay zeka teknolojisinin eğitimde kullanımı, bir dizi önemli avantaj sunmaktadır. Bu avantajlar, öğrencilere kişiselleştirilmiş öğrenme deneyimleri sunma, öğretmenlerin etkili öğretim teknikleri uygulama, öğrenci ilerlemesini izleme ve hızlı geri bildirim sağlama, öğrencilere özel destek ve rehberlik sunma, farklı dilleri konuşan veya işitme-görme engelli bireyler için eğitimde fırsat eşitliği sağlama ve modern teknolojilerle etkileşimli dersler oluşturma şeklinde sıralanabilir.

YZ teknolojisi, öğrencilere kişiselleştirilmiş öğrenme deneyimleri sunarak daha yüksek katılımı ve eğitim başarılarını teşvik edebilir (Shen vd., 2021). Ayrıca, öğretmenler, öğrencilerin akademik geçmişi ve öğrenme ortamını dikkate alarak etkili öğretim tekniklerini uygulamak için YZ'yı kullanabilirler. Bu teknoloji, öğrencilerin ilerlemesini otomatik olarak izleyip değerlendirebilir ve öğretmenlere öğretim yöntemlerini buna göre uyarlamada destek sağlayabilir (Akgun ve Greenhow, 2022). YZ destekli chatbotlar, öğrencilere sık sorulan soruları

yanıtlayabilir, ödevler hakkında geri bildirim sağlayabilir ve öğrencilerin performansına dayalı olarak daha fazla çalışma materyali önerebilir (Akinwalere ve Ivanov, 2022).

YZ teknolojilerinin eğitimde kullanılması, farklı dilleri konuşan veya işitme-görme engelli bireyler için eğitimde fırsat eşitliği sağlamak açısından da önemlidir. Özellikle Presentation Translator gibi YZ tabanlı sistem uygulamaları, gerçek zamanlı altyazı sağlayarak farklı dilleri konuşan kişilere yardımcı olabilir. Bu sayede öğrenciler kendi dillerinde okuyabilir ve dinleyebilirler. Ayrıca, sanal gerçeklik, artırılmış gerçeklik ve oyunlaştırma gibi modern teknolojiler, derslerle entegre edilerek daha etkileşimli dersler oluşturmak için kullanılabilir (Kengam, 2020).

Ancak, YZ'nin eğitimde kullanılması geleneksel öğretim yöntemlerinin yerini almak yerine onları tamamlamalıdır (Kengam, 2020; Shen vd., 2021). Ayrıca, teknolojinin aşırı kullanımı ve insanlar arası etkileşimin azalması gibi risklere dikkat edilmelidir.

UNESCO Eğitimde Bilgi Teknolojileri Enstitüsü'nün 2020 raporuna göre, YZ'nin eğitimde kullanımıyla ilgili zorluklar da göz önünde bulundurulmalıdır. Bu zorluklar arasında bilgi teknolojilerine eşit erişim eksikliği, veri gizliliği, etik sorunlar ve teknoloji bağımlılığı gibi konular yer almaktadır. Bu nedenle, YZ teknolojisinin eğitimde dengeli bir şekilde kullanılması ve öğrencilere en iyi şekilde yardımcı olunması için çaba harcanmalıdır.

2.2. Makine Öğrenimi

Makine öğrenimi, bilgisayar sistemlerinin verileri analiz etmek, sonuçlar çıkarmak ve bilgileri öğrenmek amacıyla algoritmaları ve istatistiksel teknikleri kullanarak öğrenme yeteneğini geliştirdiği bir alandır (Keş, 2023; Bishop, 2006: 738). Bu disiplin, yapay zekanın alt dalı olarak kabul edilir ve eğitim verilerini algoritmalarına ileterek kararlar alır. Bu kararlar, matematiksel modellerle desteklenerek, bilgisayarların verileri anlamalarına ve tahminlerde bulunmalarına yardımcı olur.

Makine öğrenmesi, yapay zeka çalışmalarının önemli bir bileşeni olarak insan zekasının bilgisayarlarla entegre edilmesini hedefler (Irmak, 2023; Goodfellow ve

diğerleri, 2018). Bu alandaki çalışmalar, bilgisayar sistemlerinin insan benzeri düşünme, öğrenme, karar verme yeteneği kazanmasını, doğal dil işleme becerileri geliştirmesini ve benzer karmaşık görevleri yerine getirebilmesini amaçlar.

Makine öğrenmesi, kökleri 1950'lerdeki öncüler olan Alan Turing, John McCarthy, Marvin Minsky ve diğerlerine kadar uzanan bir alan olarak ortaya çıkmıştır. Makine öğrenmesi terimi ise ilk kez 1952'de Arthur Samuel tarafından kullanılmış ve bilgisayarların öğrenme yeteneğini açıklamak için "makine öğrenmesi" terimi ortaya çıkmıştır (Al Mayahi ve Al-Bahri, 2020).

Bu geniş alan, çeşitli disiplinlerin etkili bir şekilde kullanıldığı bir disiplin ve uygulama yelpazesine ilişkilidir. İstatistik, matematik, bilgisayar bilimi, yapay zeka, veri madenciliği, sinyal işleme, nöroloji ve psikoloji gibi disiplinler, makine öğrenmesinin gelişimine katkı sağlar. Ayrıca, finans, tıp, eğitim, e-ticaret, medya, spor gibi birçok uygulama alanında da kullanılmaktadır (Ersöz ve Çınar, 2021; Mitchell, 1997).

Makine öğrenmesi, ham verilerden ve deneyimlerden öğrenme yeteneği üzerine kurulmuştur ve bir makinenin, sahip olduğu girdi verilerinden örüntüler çıkartarak bilinmeyen verilere doğru tahminlerde bulunabilmesini amaçlar (Kurbanoğlu, 1992). Bu, geleneksel programlama yaklaşımlarının aksine, sistemlerin programlanmadığı, bunun yerine bir modelin belirli bir veri kümesi üzerinde eğitildiği ve bu eğitim sonucunda verilerdeki örüntüleri tanıyarak görevleri otomatikleştiren kuralların oluşturulduğu bir alandır (Chollet, 2019).

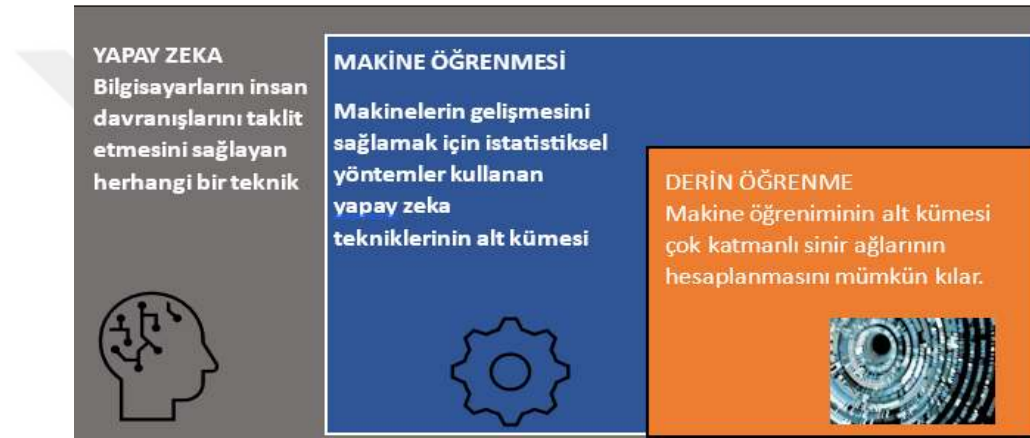
Mitchell (1997), makine öğrenmesini deneyim D ile artan performans ölçüsü P 'ye göre bilgisayar programlarının öğrenmesi olarak tanımlar. Bu tanım, makine öğrenmesinin temel fikrini özetlerken, derin öğrenme gibi alt disiplinlerin önemini de vurgular. Derin öğrenme, büyük veri setlerini işleme yeteneği ile öne çıkar ve genellikle büyük veri analizi ve modelleme için kullanılır.

Makine öğrenimi görevlerini başarıyla gerçekleştirmek için izlenmesi gereken temel adımlar, makine öğrenimi projelerinin temel taşlarıdır. Bu adımların başında geçmiş verilerin toplanması gelir. Bu verilerin çeşitliliği ve miktarı, makine öğrenme sonuçlarının güvenilirliğini etkileyen kritik faktörlerden biridir (Jain, 2015).

Veri kalitesi, herhangi bir analitik işlemin başarısını belirler. Verilerin kalitesini değerlendirmek, eksik veri veya aykırı değer gibi sorunları çözmek ve

sonuç olarak verileri analiz etmek için doğru adımları atmak gerekmektedir. Bu süreç içinde keşif analizi, verinin incelenmesi ve içerik kalitesinin artırılması için önemli bir yöntem olarak kullanılır (Jain, 2015; Sunil, 2016).

Model eğitimi ve seçimi, makine öğrenme sürecinin kritik bir aşamasını oluşturur. Bu aşamada, uygun bir algoritma seçimi ve verilerin model olarak temsil edilmesi önemlidir. Temizlenen veriler genellikle eğitim ve test verileri olarak ayrılır, bu sayede modelin geliştirilmesi ve doğruluğunun test edilmesi sağlanır (Jain, 2015).



Şekil 1. Yapay Zeka, Makine Öğrenmesi ve Derin Öğrenme

Kaynak: Usoft. (2023). *The significant difference between AI, ML and deep learning*. 20 Temmuz 2023 tarihinde <https://www.usoft.com/blog/differencebetween-ai-ml-and-deep-learning> adresinden edinilmiştir.

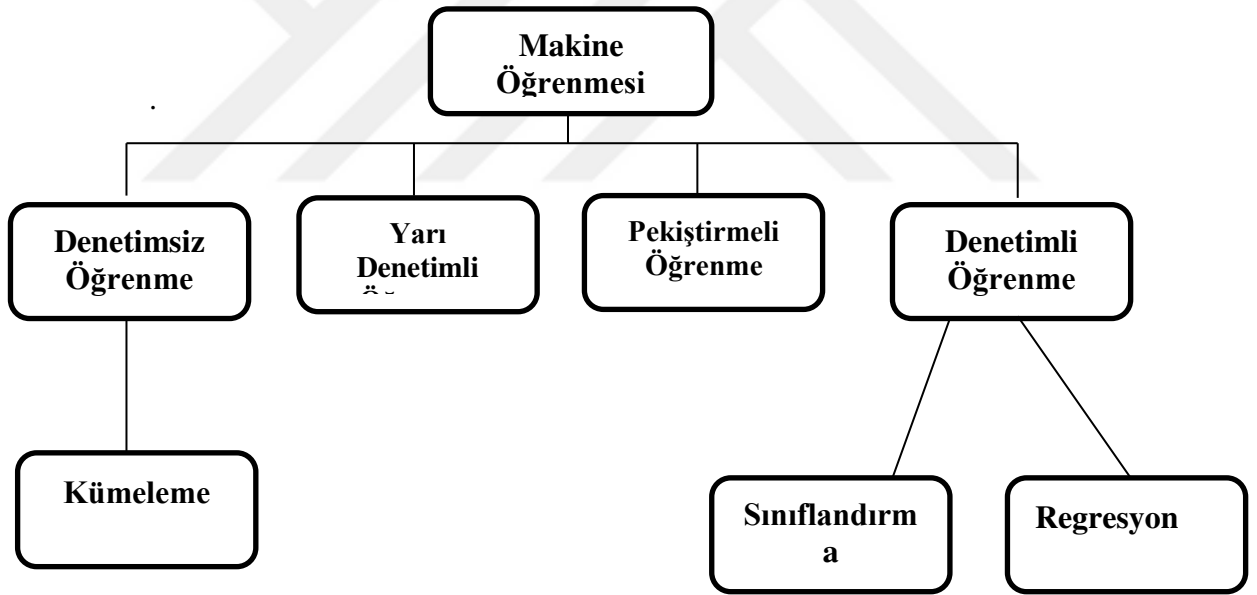
Model değerlendirmesi, modelin performansının test verileri kullanılarak kesinliğini belirler. Bu aşama, kullanılan algoritmanın başarı seviyesini değerlendirmenin yanı sıra, modelin iyileştirilmesi ve daha iyi sonuçlar elde etmek için gereken düzeltmelerin yapılmasını içerir (Jain, 2015).

Son olarak, elde edilen sonuçlara dayanarak performansın geliştirilmesi ve sürekli iyileştirme adımı gelir. Bu, verimliliği artırmak veya farklı bir modeli seçmek gibi adımları içerebilir. Bu nedenle, veri toplama ve hazırlama işlemlerine büyük bir özen göstermek önemlidir, çünkü bu adımlar makine öğrenme projelerinin başarısını büyük ölçüde etkileyebilir (Jain, 2015).

Bu beş temel adım, makine öğrenimi projelerinin başarılı bir şekilde tamamlanmasını sağlar ve veri odaklı kararların alınmasına yardımcı olur. Bu süreçler, verilerin doğru ve güvenilir bir şekilde kullanılmasını sağlayarak yapay zeka ve makine öğrenme uygulamalarının başarısını artırır.

2.2.1. Çalışma Teknikleri

Makine öğrenmesi, genellikle veri madenciliği ve derin öğrenme olarak adlandırılan iki temel alt kategoriye ayrılmaktadır. Her iki alt kategori, farklı öğrenme yaklaşımlarını ve teknikleri içinde barındırır. Makine öğrenmesi yöntemleri, denetimli, denetimsiz, pekiştirmeli ve yarı gözetimli algoritmaları içerir. Şekil 2'de, makine öğrenimi içindeki öğrenme türlerinin hiyerarşik yapısı açıkça gösterilmektedir



Şekil 2. Makine Öğrenimi İçerisindeki Öğrenme Türlerinin Hiyerarşik Yapısı

Kaynak: Akdoğan, F. (2021). *Makine öğrenmesi nedir? Ne işe yarar?* 25 Temmuz 2023 tarihinde <https://technogezgin.com/makine-ogrenmesi-nedir-ne-ise-yarar/> adresinden edinilmiştir.

Bu yapı, farklı öğrenme yaklaşımlarının nasıl birbirleriyle ilişkili olduğunu ve bu alanın çeşitliliğini vurgular.

2.2.1.1. Danışmanlı (Denetimli) Öğrenme

Denetimli öğrenme, etiketlenmiş veri örneklerine dayanarak bilinmeyen sonuçları tahmin eden bir makine öğrenme yöntemidir. Bu yaklaşım, eğitim veri setinden elde edilen örüntülerle test veri setindeki sonuçları tahmin etmeye çalışır. Başlangıçta, veri kümesi eğitim ve test veri setleri olarak bölünür ve bu eğitim verisi kullanılarak bir model geliştirilir. Bu model, test verisi üzerinde kullanılarak tahminlerde bulunur (Naqa ve Murphy, 2015, s. 7; Mahesh, 2020, s. 381).

Bu tahminleme süreci, sınıflandırma ve regresyon olmak üzere iki ana kategoriye ayrılır. Sınıflandırma, farklı kategorilere ayrılması gereken veri örneklerini tahmin etmeye çalışır. Örnek olarak "evet-hayır" ya da "kadın-erkek" gibi sınıflandırmalar düşünülebilir. Bu tür algoritmalar, girdi verilerini en uygun sınıfa yerleştirmeyi hedefler. Sınıflandırma işlemlerinde Yapay Sinir Ağları, Karar Ağaçları, Destek Vektör Makineleri ve Lojistik Regresyon gibi yöntemler sıkça tercih edilir. Bunun yanı sıra, regresyon algoritmaları, girdi ile çıktı arasındaki ilişkiyi tanımlamak ve bu ilişkiyi en iyi yansıtan fonksiyonu bulmak için kullanılır.

Denetimli öğrenme, geniş bir uygulama yelpazesine sahiptir. Görüntü bölme, sağlıkta tanı koyma, hile algılama, istenmeyen e-postaları engelleme ve sesli komut algılama gibi farklı alanlarda etkili bir şekilde tercih edilmektedir (Nassif vd, 2019: 19148).

2.2.1.2. Danışmansız (Denetimsiz) Öğrenme

Denetimsiz öğrenme, gözetimsiz öğrenme olarak da adlandırılır ve bu yöntemde herhangi bir insan müdahalesine ya da talimata ihtiyaç duyulmaz. Makineler bu yaklaşımda, veri kümelerini kendi başlarına analiz eder ve aralarındaki bağlantıları veya korelasyonları ortaya çıkarırlar. Sonuçların yorumu tamamen makine öğrenme algoritmalarının inisiyatifindedir. Özellikle büyük miktarda verinin bu algoritmalara sunulması, karar mekanizmasının daha etkili ve güçlü hale gelmesini sağlar. Temel olarak, denetimsiz öğrenme önceden etiketlenmemiş verilerdeki gizli örüntüleri ve yapıları belirlemeye yöneliktir. Bu algoritmalar, etiketlenmemiş veriyle öğrenir ve dış bir denetim olmaksızın tahminlerde bulunurlar. Bu tahminlerin ana amacı, veriyi benzerliklerine ve farklılıklarına göre

gruplandırmak ve sınıflandırmaktır. Bu öğrenilen bilgiler, yeni eklenen verinin hangi kategoriye ya da sınıfa ait olduğunu belirlemede kullanılır (Mahesh, 2020, s. 383; Janiesch, 2021, s. 687).

Denetimsiz öğrenme yöntemleri üç ana kategoriye ayrılır: kümeleme, boyut indirgeme ve anomali tespiti. Kümeleme, benzer veri öğelerini bir araya getirirken, farklı olanları ayırır ve bu kategoride K-Ortalama Kümeleme ile DBSCAN gibi yöntemler bulunmaktadır. Boyut indirgeme, veri setinin karmaşıklığını azaltarak onu daha anlaşılır kılar. Bu kategoride, Temel Bileşen Analizi (PCA) yaygın bir yöntem olarak bilinir. Anomali tespiti, veri setinde beklenmedik veya aykırı veri noktalarını belirlemek için kullanılır.

Bu öğrenme yönteminin kullanıldığı alanlar geniştir. Özellikle müşteri segmentasyonu, pazar sepeti analizi, web madenciliği ve öneri sistemleri gibi konularda denetimsiz öğrenme teknikleri başarılı bir şekilde kullanılmaktadır (Nassif vd., 2019, s. 19149; Moubayed vd., 2018, s. 39128-39129).

2.2.1.3. Yarı Danışmanlı Öğrenme

Yarı denetimli öğrenme, makine öğrenme alanında hem etiketli hem de etiketsiz verilerin birlikte kullanıldığı bir stratejidir. Bu strateji özellikle sınırlı sayıda etiketli veri bulunurken büyük miktarda etiketsiz veriyle karşılaşıldığında tercih edilir. Bu durum, özellikle veri etiketleme sürecinin maliyetli veya zor olabileceği büyük veri setleri için geçerlidir. Yarı denetimli öğrenmenin ana amacı, kullanılan etiketli veri ile sınıflandırma performansını maksimize etmektir (Naqa ve Murphy, 2015, s. 7).

Bu çalışma paradigması, Kendi Kendine Eğitim ve Düşük Yoğunluklu Ayırma gibi spesifik teknikler üzerine kurulmuştur. Bu teknikler, eksik etiketlenmiş geniş veri kümesinin analizine katkıda bulunarak sınıflandırma ile ilgili problemlere yanıt teşkil eder. Neticede, yarı denetimli öğrenme yaklaşımı, veri kümesinin daha derinlemesine ve etkin bir analizini mümkün kılarak doğru inferansların gerçekleştirilmesine imkan tanır.

2.2.1.4. Pekiştirmeli (Takviyeli) Öğrenme

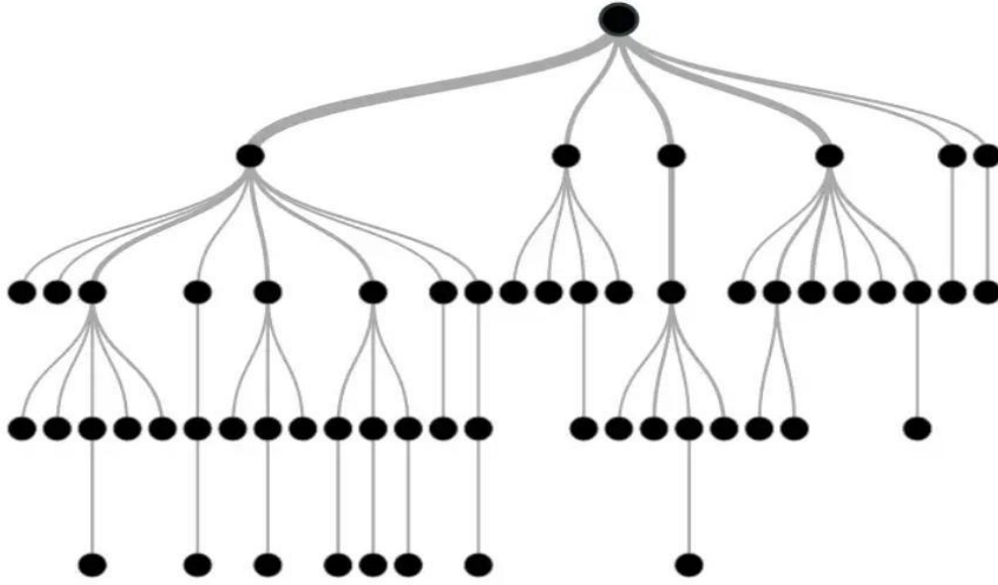
Pekiştirmeli öğrenme, yapay zeka ajanlarının çevre ile interaksyon kurup deneyim elde ettiği ve bu deneyimler aracılığıyla performansını artırdığı bir öğrenme paradigmasıdır. Ajanların temel amacı, eylemlerinden kaynaklanan olumlu sonuçları ödülle maksimize etmektir. Optimal ödül değerlerine ulaşabilmek için, deneme-yanılma mekanizması sıkça kullanılır. Pekiştirmeli öğrenme metodolojisinin kritik niteliklerinden biri, deneme-yanılma yaklaşımının gecikmeli ödül prensibi ile bütünleşmesidir (Moubayed vd., 2018, s. 39132).

Bu öğrenme stratejisi, önceden belirlenmiş etiketli veri setlerini temel almaz; bunun yerine ajanlar kendi yaşantılarından edindiği bilgilerle kendini geliştirir. Çeşitli disiplinlerde, özellikle oyun teorisinden bilgi teorisine, çoklu ajan sistemlerine kadar olan geniş spektrumda, pekiştirmeli öğrenme önemli bir yer tutmaktadır. Bu yaklaşım, Dinamik Programlama ve Monte Carlo Teknikleri gibi sektördeki prestijli algoritmalarından yararlanarak uygulanır (Kaelbling et al., 1996, s. 239).

2.2.2. Makine Öğrenmesi Algoritmaları

Makine öğrenmesi algoritmaları, otomatik sürüşten sesli asistanlara, kişiselleştirilmiş öneri sistemlerine kadar birçok modern teknolojinin geliştirilmesinde kritik bir role sahiptir (LeCun vd., 2015). Bu algoritmalarından elde edilen yenilikler, geniş bir uygulama yelpazesinde karmaşık sorunları çözmekte ve sürekli olarak teknolojik evrimi teşvik etmektedir.

Eğitim sektörü de bu algoritmaların sunduğu potansiyelin farkında. Özellikle veri analitiği ve makine öğrenme tekniklerini benimseyerek, öğrenci performansını değerlendirme ve öğrenme süreçlerini optimize etme eğilimindedirler. Çevrimiçi öğrenme yönetim sistemlerinin sağladığı zengin veri setleri, öğrencilerin ders etkileşimleri, test sonuçları ve diğer aktivitelerine dair değerli bilgiler sunmaktadır. Bu bilgiler, öğrencilerin akademik başarılarının tahmin edilmesi için bir temel oluşturabilir (Márquez-Vera vd., 2013). Ancak, bu tahminlerin getirdiği potansiyel faydaların yanı sıra, öğrenci verilerinin gizliliği ve etik sorunları da bulunmaktadır. Bu nedenle, bu konularda dikkatli ve bilinçli bir yaklaşım benimsemek esastır (Slade ve Prinsloo, 2013).



Şekil 3. Karar Ağaçları Şematik Gösterimi

Kaynak: Ulgen E.(2017). *Karar ağaçları*. 1 Ağustos 2023 tarihinde <https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-5-karar-agac-lari-c90bd7593010> adresinden edinilmiştir.

Önemli olan bir diğer husus ise, makine öğrenmesi algoritmalarının her birinin belirli sorun tiplerine yönelik özelleştirilmiş olmasıdır. Seçilecek algoritmanın belirlenmesi sürecinde, veri setinin özellikleri, karşılaşılan problemin doğası ve modelleme ihtiyaçları dikkate alınmalıdır.

2.2.2.1. Karar Ağaçları (DT)

Karar ağaçları, sınıflandırma ve regresyon meselelerine çözüm sunan prestijli bir algoritmadır (Quinlan, 1986). Bu yöntem, veri kümelerinde yer alan yapıları ve ilkelere dayalı olarak doğru projeksiyonlarda bulunabilme özelliğine sahiptir. Yorumlama kolaylığı, veri tabanı sistemleri ile entegrasyonundaki verimliliği ve güvenilirlik özellikleri sebebiyle sınıflandırma yaklaşımları içerisinde yoğun bir ilgi görmektedir (Ayık, 2007, s. 441-454).

Karar ağaçları, veriyi daha yönetilebilir alt kümelerine bölmek için belirli özelliklere dayalı karar kuralları kullanır. Bu ağaç yapısı, hem kategorik (sınıflandırma) hem de sürekli (regresyon) çıktıları elde etmek için uygundur.

Ağacın oluşturulması sırasında, kök düğümünden başlanarak çeşitli yöntemlerle ilerlenir. Bu yöntemler arasında Gini, Entropi ve Cart algoritmaları bulunur. Gini, her düğüm için optimum bölünmeyi belirlemek için kullanılırken, Entropi daha dengeli ağaçlar oluşturmayı amaçlar. Cart algoritması, en optimal ağaç yapısını elde etmek için ağacı sürekli böler.

Veri madenciliği ve öğrenci performans tahminleri gibi alanlarda karar ağaçları sıklıkla tercih edilir. Bu ağaçlar, veriyi parçalayarak karmaşık problemleri basit karar kurallarıyla çözer. Seçilen kök düğümü ve bölünme yöntemi, modelin genel performansını doğrudan etkileyebilir.

2.2.2.2. Random Forest (RF)

Random Forest algoritması, makine öğrenimi paradigmaları içerisinde hem sınıflandırma hem de regresyon meselelerine yönelik bir yaklaşım sunmaktadır. Bu algoritma, veri kümelerini aleatoryum bir biçimde parçalayarak çoklu karar ağaçlarını inşa eder. Prognostik süreçlerde, konstrüe edilen karar ağaçlarının çıktılarının aritmetik ortalaması temel alınır. RF, modelin aşırı öğrenme riskini significant derecede minimize edebilme kapasitesine sahiptir (Akşehir ve Kılıç, 2019). Random Forests algoritması, Breiman (1996) ve Ho (1998) tarafından tanıtılan Bagging ve Random Subspace metodolojilerinin entegrasyonu neticesinde evrilmiştir. Karar ağaçlarının her biri, orijinal veri kümesinden bootstrap metodolojisi kullanılarak çekilmiş heterojen örneklem grupları üzerinde eğitim görmektedir. Sonuçta, bu karar ağaçlarının çıktıları amalgam edilerek nihai sınıflandırma gerçekleştirilir. RF modelinde ayrıca değişkenlerin önem derecesi de hesaplanabilir. Bu özellik, çok sayıda değişkenin olduğu durumlarda modeli sadeleştirmek için kullanışlıdır (Akman, 2010). RF yönteminin genel amacı, birçok karar ağacının oluşturduğu topluluğun gücüyle daha doğru tahminler yapabilmektir (Bodea, 2020; Breiman, 2001). Rüzgar endüstrisinde de RF algoritması, rüzgar hızı ve gücü tahmini gibi konularda popüler bir yöntemdir. RO'nun, gürültülü ve ilgisiz girdilere karşı dayanıklı olduğu ve rüzgar tahmini için umut vaat ettiği belirtilmiştir (Lahouar ve Ben Hadj Slama, 2017: 1040-1051; Akt., Bhat vd., 2021).

2.2.2.3. Naive Bayes

Naive Bayes, olasılık temelli bir sınıflandırma metodolojisi olarak literatürde yer almaktadır (Duda ve Hart, 1973; Langley ve Sage, 1994; Langley vd., 1992) ve özellikle makine öğrenimi ile veri bilimi disiplinlerinde frekansla tercih edilmektedir. Bu sınıflandırıcı, özelliklerin belirli bir sınıfa olan etkisinin diğer özellik değerlerinden mutlak bağımsızlık gösterdiği varsayımını kabul eder, ki bu konsept sınıf koşullu bağımsızlık adı verilmiştir (Han vd., 2022). Bu algoritma, bir özellik vektörü tarafından belirlenen bir örnek için en muhtemel sınıfın tespitini gerçekleştirir (Rish, 2001). Eğitim sürecinde, NB algoritması, veri seti üzerinden her bir özelliğin koşullu olasılığını infer eder ve bu olasılıklar doğrultusunda bir sınıflandırma kararı alır (Friedman vd., 1997). Sınıflandırma süreci esnasında, a priori olasılıkların belirlenmesi akabinde sınıf koşullu olasılıklar hesaplanır ve bu bilgiler Bayes Teoremi'nin rehberliğinde kullanılarak sınıflandırma tahminleri gerçekleştirilir (Kotu ve Deshpande, 2014).

Naive Bayes'in adı, olasılık teorisinin öncülerinden olan Thomas Bayes'ten gelmektedir. Bayes Teoremi, $P(A | B) = (P(B|A).P(A)) / P(B)$ formülüyle ifade edilir. Naive Bayes algoritması, bu teoremi kullanan popüler bir sınıflandırıcıdır (Kör, 2017). Naive Bayes, belirli bir sınıf için özelliklerin birbirinden bağımsız olduğunu varsayar, bu da onun hesaplamaları basit ve etkili kılar. Bu basitliğine rağmen, birçok gerçek dünya uygulamasında etkili sonuçlar verdiği gözlemlenmiştir (Çelik ve Altunaydın, 2018).

Bayes sınıflandırıcıları, belirli bir sınıfa ait olma olasılıklarını tahmin edebilirler. Sınıflandırma algoritmalarını karşılaştıran araştırmalar, Naive Bayes sınıflandırıcısının, karar verme ağaçları ve bazı sinir ağlarıyla benzer performans gösterdiğini bulmuştur (Han ve Kamber, 2006).

Bu araştırmada tanımlanan tahmin modelinin oluşturulmasında, eğitim ve test veri setlerinin strüktürlenmesi sırasında, her kategoriden homojen dağılımı temsil edecek şekilde veri seçimi için stratifiye örnekleme yöntemi kullanılmıştır. Ayrıca, modelin optimizasyonunda Laplace düzeltmesi uygulanmıştır.

2.2.2.4. Komşu (K-Nearest Neighbors, KNN) Algoritması

K-En Yakın Komşu (K-Nearest Neighbors) algoritması, gözetimli öğrenme çerçevesinde sınıflandırma ve regresyon sorunlarında yaygın olarak kullanılan bir yöntemdir. Bu algoritma, bir veri ögesinin sınıfını belirlerken ya da değerini tahmin ederken, veri kümesindeki 'k' en yakın komşusunu temel alır. 'k' değeri, kullanıcıya özgü bir hiperparametre olarak belirlenir ve modelin performansı üzerinde belirleyici bir role sahiptir (Aha ve ark., 1991; Akt. Koyuncu ve Gelbal, 2020).

KNN, "örneklem tabanlı öğrenme" ya da "bellek tabanlı öğrenme" yaklaşımını benimser. Temel amacı, daha önce etiketlenmiş verilere dayanarak yeni veri noktalarını sınıflandırmaktır. Bu algoritma, veri noktalarının birbirlerine olan uzaklıklarını hesaplar ve bu uzaklıklara göre tahminde bulunur. Bu hesaplamada genellikle Öklid uzaklık formülü kullanılır, ancak Manhattan, Minkowski ve Chebyshev gibi farklı mesafe ölçüm yöntemleri de tercih edilebilir (Hu ve ark., 2016).

1950'lerin başında Evelyn Fix ve Joseph L. Hodges Jr. tarafından temelleri atılan KNN algoritması, sonraları T.M. Cover ve P.E. Heart gibi bilim insanları tarafından genişletilmiştir. Tembel bir algoritma olarak kabul edilen KNN, eğitim aşamasında modeli eğitmez, bunun yerine eğitim verilerini ezberler ve tahmin yapmak için her defasında tüm veri setini tarar.

Özellikle büyük veri setlerinde bu algoritmanın hesaplama maliyeti yüksek olabilir. Bunun yanında, KNN algoritması ölçeklendirmeye duyarlıdır. Farklı özellikler arasında büyük ölçek farkları, bazı özelliklerin diğerlerinden daha baskın olmasına sebep olabilir. Bu sebeple, modelin performansını optimize etmek için verilerin ölçeklendirilmesi ve normalize edilmesi önerilir (James et al., 2013).

Doğru k değerini seçmek bu algoritmanın performansını büyük ölçüde etkiler. Çok büyük bir k değeri modelin duyarlılığını azaltırken, çok küçük bir k değeri aşırı uyarlamaya neden olabilir (Kılınç ve ark., 2016).

Kısacası, KNN algoritması sınıflandırma ve regresyon problemlerinin çözümünde kullanılan basit ve etkili bir yöntemdir. Ancak, en uygun k değerinin seçimi, veri ölçeklendirmesi ve hesaplama maliyeti gibi konularda dikkatli olunmalıdır.

2.2.2.5. Destek Vektör Makineleri (SVM)

Destek Vektör Makineleri (SVM), makine öğrenimi ve veri bilimi alanlarında kritik bir teknolojidir. N-boyutlu bir uzayda veri noktalarını en optimal şekilde ayırabilen bir hiper düzlem belirlemekle görevlidir. Bu yaklaşım, yapay sinir ağlarının teorileriyle paralellikler gösterir (Haykin, 1999; Akt. Yakut, 2014). SVM'nin esas görevi, veri noktaları arasında en geniş ayrımı sağlayan bir karar sınırını tanımlamaktır (Cortes ve Vapnik, 1995). Bu ayrım, destek vektörleri adını alan spesifik veri öğeleri temel alınarak gerçekleştirilir.

SVM, çok boyutlu veri uzaylarında lineer olmayan karar sınırlarını etkili bir şekilde belirlemekte ve hem sınıflandırma hem de regresyon görevleri için uygundur (Hsu, 2010). Optimal hiper düzlemi tanımlama süreci, ikinci dereceden bir optimizasyon problemi çözümüne dayanır ve bu süreçte kritik destek vektörlerini belirler.

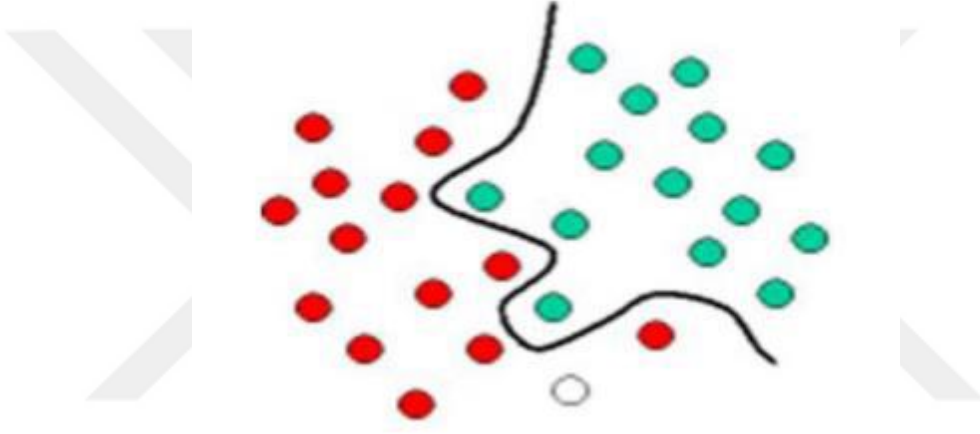
1963 yılında Alexey Chervonenkis ve Vladimir Vapnik tarafından ilk defa tanımlanan Destek Vektör Makineleri, 1995'te Vapnik ve ekibi tarafından daha ileri bir seviyeye taşınmıştır (Canbay, 2013). Bu metodoloji, öğrenci başarısının tahmini gibi makine öğrenimi görevlerinde etkili sonuçlar üretmektedir.

Yüz, ses ve el yazısı tanıma gibi çok sayıda uygulamada sıkça başvurulanan SVM, istatistiksel öğrenme teorilerine dayanan bir yaklaşımdır. Gözetim altında gerçekleştirilen öğrenme sürecinde, SVM algoritması, mevcut veri ve etiket kombinasyonlarını öğrenir, bu bilgileri temel alarak destek vektörlerini oluşturur (Dağdeviren, 2013).

2.2.2.6. Doğrusal Olmayan Destek Vektör Makinesi

Doğrusal olmayan Destek Vektör Makineleri karmaşık ve doğrusal olarak ayrılamaz veri kümeleri için tasarlanmış bir öğrenme algoritmasıdır. Gerçek dünya verileri genellikle doğrusal bir şekilde sınıflandırmak için çok karmaşıktır. Bu zorluğu aşmak için DVM, ayırma eğrileri adı verilen sınıflandırma özelliklerini kullanır. Bu ayırma eğrileri, temel işlevler olarak adlandırılan işlevlerin verilere uygulanmasıyla elde edilir. Bu işlevler, ayrılamaz bilgileri izole ederek verileri daha etkili bir şekilde temsil eder (Canbay, 2013).

Doğrusal olmayan DVM, doğrusal bir hiper düzlemle veriyi ayırlamadığında, çekirdek fonksiyonları veya 'kernel trick' olarak bilinen teknikleri kullanır. Bu fonksiyonlar, veriyi yüksek boyutlu bir özellik uzayına dönüştürür, böylece doğrusal ayrımın gerçekleştirilebileceği bir uzayda çalışılabilir (Busuttil, 2003; Suykens, 2002). Bu yaklaşım, gerçek hayatta doğrusal olarak ayırlamaz birçok veri kümesi olduğundan, destek vektör makinelerinin önemini artırmaktadır. Ayrıca, çekirdek fonksiyonları kullanılarak, yüksek boyutlu özellik uzayındaki tüm hesaplamalar girdi uzayında gerçekleştirilebilir, bu da işlem süresinde büyük bir kolaylık sağlar (Hamel, 2011).



Şekil 4. Doğrusal Olmayan Regresyon

Kaynak: Akca, M. (2020). Nedir Bu Destek Vektör Makineleri? (Makine Öğrenmesi Serisi-2). 5 Ağustos 2023 tarihinde <https://medium.com/deep-learning-turkiye/nedir-bu-destek-vekt%C3%B6r-makineleri-makine-%C3%B6%C4%9Frenmesi-serisi-2-94e576e4223e> adresinden edinilmiştir.

Destek Vektör Makineleri (DVM) yalnızca sınıflandırma amaçlı değil, aynı zamanda regresyon problemlerinde de etkili bir araç olarak kullanılmaktadır. Regresyon analizinde, lineer olmayan senaryolarla mücadele edebilmek adına çekirdek fonksiyonlarından yararlanılmaktadır (Çomak, 2008). Eğitim sürecinde, bir eğitim örneği temelinde fonksiyon oluşturulduğunda, diğer veri örnekleri için pre-kalıplı değerlerin kullanılmasıyla hesaplama süreleri optimize edilebilir (Kecman, 2001).

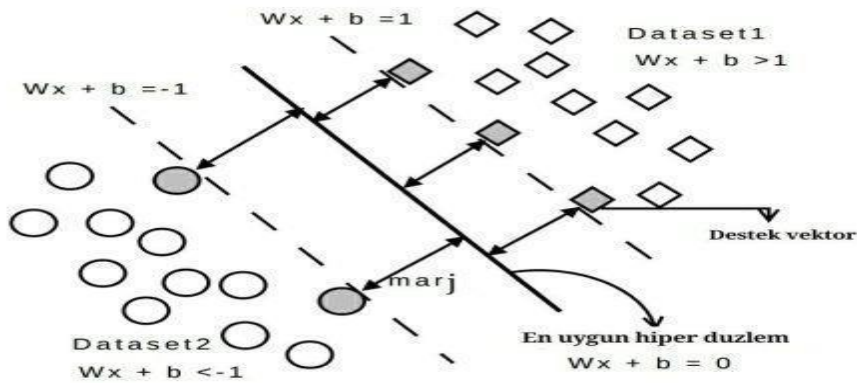
2.2.2.7. Doğrusal Destek Vektör Makineleri

Doğrusal Destek Vektör Makineleri (DVM), veri kümesinin iki sınıfın doğrusal bir sınır düzleminde ayrılabilir olduğu durumlarda kullanılan bir makine öğrenme algoritmasıdır. Bu algoritma, veri noktalarını iki sınıf arasında net bir şekilde ayıran doğrusal bir karar sınırı bulmayı amaçlar. Bu nedenle, veri noktalarının sınıflandırılması için doğrusal bir sınır kullanır (Canbay, 2013).

DVM'nin temel amacı, veri kümesindeki herhangi bir noktayı maksimum derecede uzakta olan iki sınıf arasında bir karar sınırı belirlemektir. Bu sırada sınıflar arasındaki en geniş marjı bulmaya çalışır. Ayrıca, bu marjın belirlenmesinde kritik bir rol oynayan destek vektörleri tanımlar. Destek vektörleri, sınıflandırma sürecinde belirleyici olan ve hiper düzlemi en iyi şekilde tanımlayan veri noktalarıdır (Canbay, 2013).

Doğrusal DVM, veri kümesinin doğrusal olarak ayrılabilir olduğu durumlar için uygundur. Ancak gerçek dünya verileri genellikle doğrusal olarak ayrılamayacak kadar karmaşıktır. Bu nedenle, doğrusal olmayan DVM veya diğer karmaşıklıkları ele alan SVM varyasyonları da kullanılır.

Doğrusal Destek Vektör Makineleri (DVM), veri kümesinin iki sınıfını doğrusal bir sınır düzleminde ayırabilen ve bu sınıfları en iyi şekilde ayrıştırabilen güçlü bir makine öğrenme algoritmasıdır (Canbay, 2013). DVM, sınıflar arasındaki marjı maksimize ederek ve destek vektörlerini belirleyerek bu başarıyı elde eder.



Şekil 5. Doğrusal Destek Vektör Makineleri

Kaynak: Şanlı, O. (2018). *Veri madenciliği giriş*. 12 Ağustos 2023 tarihinde <https://oyasanli.com/oyasblog/2018/09/06/veri-madenciliği-giris/> adresinden edinilmiştir.

2.2.2.8. Lojistik Regresyon

Lojistik regresyon analizi, makine öğrenimi ve veri analitiği alanlarında yaygın olarak kullanılan istatistiksel bir yöntemdir. Özellikle veri analitiği ve makine öğreniminin karmaşık ve büyük veri setleriyle çalışırken etkili olduğunu biliyoruz. Ancak makine öğrenimi modelleri bazen karmaşık olabiliyor ve eğitilmesi için önemli miktarda hesaplama kaynağına ihtiyaç duyabiliyor.

Lojistik regresyon, sınıflandırma problemleri için kullanılır ve özellikle ikili sonuçların elde edilmesini istediğimiz durumlarda tercih edilen bir yöntemdir (Hosmer Jr et al., 2013). Lojistik regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi belirlemekte kullanılır (Yıldırım, 2022). Bu yöntem, bağımlı değişkenin iki veya daha fazla kategorisine sahip olduğu durumlarda da kullanılabilir. Örneğin, ikili, sıralı veya çok kategorili lojistik regresyon olarak (Alpar, 2011; Kartal, 2015).

Bu yöntemde, bağımsız değişkenlerin sürekli veya kategorik olabilmesine rağmen, bağımlı değişkenin sadece kategorik olabileceği belirtilmektedir. Lojistik regresyon, 0 veya 1 değerlerini üretebilir ve bu değerler genellikle bir şeyin olup olmadığını (Özdemir, 2016) veya bir sınıfa ait olup olmadığını belirtmek için kullanılır. Bu, sigmoid fonksiyonu sayesinde $[0-1]$ aralığında bir y değeri üreten algoritma ile mümkündür (Ay Karakuş, 2018).

Lojistik regresyon analizi gerçekleştirilirken, tipik olarak veri kümesinin yaklaşık %70'i eğitim amaçlı, geriye kalan %30'luk kısmı ise test amaçlı ayrılır. Bu metodoloji, doğruluk oranının %78 olduğunu göstermiştir. Doğrusal regresyon modeline kıyasla, lojistik regresyon, belli başlı varsayımların sınırlamalarını taşımaksızın uygulanabilen bir modeldir. Özellikle, bağımlı değişkenin kategorik veya süreksiz olduğu durumlar için tercih edilir. Sürekli nitelikteki bir bağımlı değişkenin kategorik bir yapıya dönüştürülmesinde de lojistik regresyon analizi kullanılmaktadır (Tabachnick ve Fidell, 1996).

Bu analizin temel amacı, değişkenler arası ilişkileri tanımlamak ve bu ilişkileri en az değişkenle en iyi uyuma sahip olacak şekilde modellemektir (Atasoy, 2001). İlginç bir şekilde, lojistik modelin kullanılmasının tarihçesi 1845 yıllarına kadar uzanıyor. Bu model, ilk başta toplumsal nüfus artışının matematiksel olarak

açıklanmasında kullanılmış (Gürcan, 1998). Ancak son 20 yılda askeri, meteoroloji, iç göç hareketleri ve eğitim gibi farklı alanlarda kullanımı artmıştır.

Lojistik regresyon analizi, makine öğrenimi ve veri analitiği alanında önemli bir rol oynamaktadır. Bu yöntem, ikili ve kategorik sonuçların belirlenmesinde özellikle etkilidir ve çeşitli uygulama alanlarına sahiptir.

2.2.2.9. Extreme Gradient Boosting

XGBoost, Friedman (2001) tarafından ortaya konan gradyan artırma yaklaşımlarını esas alarak tasarlanmış GB tabanlı bir algoritmadır. XGBoost ile GB'nin her ikisi de gradyan artırma tekniklerine başvururken, XGBoost algoritması, model performansını optimize etmek adına daha düzenlenmiş bir yapı benimser (Carmonavd., 2019). XGBoost'un mimarisi, geçmiş gradyan artırma algoritmalarının kısıtlılıklarını aşma, kaynakları verimli bir biçimde kullanma ve üstün ölçeklenebilirlik sağlama prensiplerine dayanmaktadır. Bu yüzden, LightGBM ve CATBoost gibi alternatif algoritmalara kıyasla, modelin eğitim aşaması hem daha hızlıdır hem de özellikle aşırı uydurmaya karşı yeni bir düzenleme stratejisi benimsediği için daha robusttır (Al Daoud, 2019, s. 6-10).

XGBoost, spam tespiti, reklam eşleştirme, dolandırıcılık ve anormal olay tespit sistemleri gibi birçok alanda kullanılmaktadır (Chen ve Guestrin, 2016, s. 785-794). Bu yöntemin temel avantajı, tahminlerin veri odaklı olmasından kaynaklanmaktadır. Regresyon ve sınıflandırma ağacı genişletme algoritmaları kullanılarak elde edilen sonuçlar, oldukça etkilidir. XGBoost'un başarısı, yalnızca doğru tahminler yapabilmesinden değil, aynı zamanda çok çeşitli tahminlerde de bulunabilmesinden kaynaklanmaktadır.

XGBoost, gradyan artırma yöntemlerine dayanarak tasarlanmış ve optimize edilmiş bir kitaplıktır. Bu özellik, makine öğrenimi algoritmalarını hesaplamamanın ötesine taşıyarak, çok boyutlu büyük veri analizlerinde etkin bir şekilde kullanılmasını sağlar. XGBoost'un Kaggle gibi platformlarda popüler hale gelmesinin nedeni, kullanım kolaylığı ve tahmin etme yeteneğinin yüksek olmasıdır (Morde, 2019). Aşırı Gradyan Güçlendirme, sınıflandırma ve regresyon ağaçları algoritmasını temel alır ve temel gradyan artırma makinelerinin (GBM) çerçevesini geliştirir.

XGBoost, birçok yarışmada başarılı sonuçlar elde etti. Özellikle mağaza satış tahminleri, fiziksel olay sınıflandırması, web metni sınıflandırması ve müşteri davranışı tahmini gibi konularda etkili bir şekilde kullanılmıştır (Chen ve Guestrin, 2016). XGBoost'un Chen tarafından 2016 yılında önerildiğini ve sınıflandırma ve regresyon ağaçlarına (CART) dayandığını belirtmek de önemlidir. Bu yöntem, petrol fiyatları, güç yükleri ve kazalar gibi konularda tahmin yapma konusunda diğer modellere göre daha yüksek bir doğruluk sunmaktadır.

XGBoost, gradyan artırma tekniklerini temel alan güçlü ve ölçeklenebilir bir algoritmadır. Çeşitli uygulama alanlarında başarılı sonuçlar elde etmektedir ve makine öğrenimi alanında önemli bir konumdadır.

2.2.2.10. Bagging

Bagging tekniği, eğitim verilerini rastgele alt kümeler halinde seçerek çoklu ağaçların eğitilmesini sağlar. Her bir ağaç paralel ve bağımsız olarak öğrenir, ardından elde edilen sonuçlar bir araya getirilir. Bu yaklaşım, ağaçlar arasındaki bağımlılığı minimize eder ve overfitting sorununu azaltarak modelin genel başarımını yükseltir. Bagging yöntemi, Breiman (1996) tarafından ortaya konulmuştur ve uygulaması pratiktir.

Rastgele Alt-Uzay tekniği, özellik setindeki değişiklikleri kullanarak farklı eğitim kümelerinin oluşmasını teşvik eder. Mevcut eğitim verileri daha spesifik alt kümelerine ayrılır ve her temel öğrenme modeli bu spesifik kümeler üzerinden eğitim alır (Ho, 1998). Bu yaklaşım, yüksek boyutlu veri kümelerinde gereksiz özelliklerin varlığında özellikle avantajlıdır.

Hem Bagging hem de Rastgele Alt-Uzay teknikleri, topluluk öğrenme stratejileri arasında yer alır. Amacı, modelin genel performansını yükseltmek ve overfitting riskini azaltmaktır. Bagging, bağımsız ağaçları birleştirme prensibiyle hareket ederken, Rastgele Alt-Uzay özellik setindeki değişikliklere odaklanır. Bu iki yöntem, veri analizi ve makine öğrenimi alanlarında sıkça tercih edilir..

2.3. Uzaktan Eğitim

Eğitim, toplumun her kesimine ulaştırılarak bireylerin eğitim seviyelerini artırmak ve bu sayede toplumun genel eğitim seviyesini ileriye taşımak amacıyla yüz

yüze eğitimlerin ötesine geçmeye gereksinim duyar (Yekta ve Arıcı, 2005). Yüz yüze eğitim, bir dizi sorunu beraberinde getirebilir; eğitimi ulaşılabilir kılmada zorluklar, verimli kaynak kullanımının sağlanamaması, uygun dinlenme ortamlarının bulunmaması, sınırlı ulaşılabilirlik gibi sorunlarla karşı karşıya kalınabilir (Kaya vd., 2004). Bu sorunlar çözülmeli ve bir çözüm olarak Uzaktan Eğitim (UE) sistemi geliştirilmelidir. UE, 19. yüzyıldan bu yana geleneksel eğitimle birlikte teknolojik ilerlemelere ayak uyduran ve öğrencilere zamandan ve mekândan bağımsız olarak öğrenme fırsatı sunan bir eğitim modelidir (Schlosser ve Simonson, 2006; Akt. Omolade ve Samuel, 2022).

UE, öğrencilerin ve öğretmenlerin kendi hızlarında ve tercih ettikleri yerde öğrenmelerine olanak tanır (İşman, 2011; Kaya vd., 2004). UE, öğrencilere yaşam boyu eğitim fırsatı sunan, eğitimde bireysel ve toplumsal hedeflere katkıda bulunan ve kendi kendine öğrenmeye dayalı bir model sunar. Bozkurt'a (2017) göre UE, öğretim sürecindeki engelleri kaldırmak ve öğrencilerin, öğretmenlerin ve öğrenme materyallerinin tam potansiyellerine ulaşmalarını sağlamak için güncel teknolojiyi kullanır.

UE, teknoloji gelişmeleriyle birlikte evrim geçirir ve zamandan ve mekândan bağımsız olarak öğrenme etkinliklerine katılabilmek anlamına gelir (Morrison, 2003; Orakcıoğlu, 2019). Ayrıca UE, eğitim maliyetlerini düşürürken nitelikli bir eğitim ortamı sağlama potansiyeline sahiptir (Özbay, 2015).

UE, öğrenme sürecini daha esnek ve verimli hale getirir (İliç, 2020; Urdan ve Weggen, 2000). Bilgi teknolojisinin gelişimi, mobil cihazların ve internetin kolay erişilebilir olması sayesinde bilgiye erişimi ve internetten yararlanmayı kolaylaştırır (Yeşiltuna ve Tükel, 2015). UE, öğrenme ortamlarının oluşturulması, eğitim materyallerinin geliştirilmesi, yayın yapılması ve yönetilmesi konularında teknolojik cihazların avantajlarından yararlanır.

- Alkan (1996)(Akt. Kaya, 2002), UE'nin temel ilkelerini şöyle özetler:
- İnnovatif imkanlar oluşturma,
- İş hayatı ile eğitimi uyum içinde kılma,
- Eğitim yaklaşımlarını daha katılımcı bir hale getirme,
- Sürekli öğrenmeyi teşvik etme,
- Bireysel ihtiyaçları eğitime adapte etme,
- Var olan eğitim yapılarından optimum fayda sağlama,

- Teknolojik araçları eğitimde verimli bir biçimde entegre etme,
- Bireysel, toplumsal ve teknolojik ihtiyaçlara odaklanma,
- Geniş demografik kitlelere hitap etme,
- Bireysel öğrenimi ve toplu eğitimi uyumlu kılma,
- Eğitim talebi ve finansal kaynakları dengeli bir şekilde yönetme.

Deveci (2019), daha etkili bir UE sistemi için aşağıdaki önerileri sunar:

- İletişim, geri bildirim ve etkileşim öğelerini iyi planlama ve uygulama,
- Öğrenenler arasında etkili iletişim ve etkileşim ortamı yaratma,
- Öğrenci motivasyonunu artırmak için UE sistemi oluşturma,
- Motivasyonu artırmak için çoklu ortam uygulamalarına yer verme.

2.3.1. Uzaktan Eğitim Teknolojileri

UE süreçlerinin işleyişi ve sistem altyapısının oluşturulması açısından teknoloji büyük önem taşımaktadır. Okullarda, Zoom, Cisco Webex, Google Meet gibi çeşitli video konferans platformları ve MOODLE, Sakai, Perculus gibi öğrenme yönetim sistemleri kullanılmaktadır (Yamamoto ve Altun, 2020; Dikmen ve Bahçeci, 2020).

UE süreçlerinde uygun ders içeriklerinin geliştirilmesi, eğitim teknolojileri ve senaryolarına uygun eğitim senaryolarının hazırlanmasını gerektirir. Bu senaryoların oluşturulabilmesi için temel ofis yazılımları (ppt, doc vb.) uygulamalarına hakim olmanın yanı sıra, ihtiyaca göre Articulate, Captivate veya iSpring gibi zengin içerik geliştirme araçları hakkında da bilgi sahibi olunmalıdır (Altun, 2020). Bu araçlar kullanılarak, UE için zengin ders içerikleri hazırlamak mümkündür.

Örneğin, Articulate Storyline 360 yazılımı ile e-öğrenme içerikleri daha çok hikaye tarzında oluşturulabilir. Bu yazılımın tasarım arayüzü, Microsoft PowerPoint arayüzüne benzemekte ve aynı zamanda içerisinde bir zaman çizelgesi (timeline) bulundurmaktadır. Articulate Storyline 360 kullanılarak resimler, sesler, videolar, animasyonlar ve sorular eklenerek e-içerikler ve sunumlar oluşturulabilir. Bu sunumlar, hem mobil cihazlarda hem de bilgisayar ortamında sorunsuz bir şekilde çalıştırılabilir (İzmirli vd., 2018).

2.3.2. Uzaktan Eğitim Modelleri

Bir uzaktan eğitim faaliyetinin gerçekleştirilmesi için öncelikle hedef amaçlı bir eğitim modeli belirlenmeli ve geliştirilmelidir. Bununla birlikte bu modelin teknolojik altyapıyla da desteklenmesi gerekmektedir. Uzaktan eğitimde kullanılan modeller genelde eş zamanlı ve eş zamanlı olmayan (senkron ve asenkron) olmak üzere iki gruba ayrılmaktadır (Romiszowski, 2004). Uzaktan eğitimin eş zamanlı ve eş zamanlı olmayan şekilde gerçekleşmesi ile ilgili açıklamalar aşağıdaki tabloda verilmiştir.

2.3.2.1. Eş Zamanlı (Senkron)

- Canlı veya gerçek zamanlıdır.
- Öğrenciler aynı anda online olur ve derse katılır.

Örnek: Webinar iletişim, online chat.

2.3.2.2. Eş Zamanlı Olmayan (Asenkron)

- Canlı veya gerçek zamanlı değildir.
- Öğrenci kendine en uygun zamanda online olur veya derse katılır.

Örnek: Bireysel online, takım veya bütün grup çalışmaları.

Tabloda temelde iki farklı uzaktan eğitim modeli görülmektedir: Eş zamanlı ve eş zamanlı olmayan öğrenme modeli. Her bir modelin özelliğini bilmek, öğreticiye online veya karma öğretim sisteminin uzaktan eğitim sistemine nasıl entegre edilmesi konusunda karar verilmesine yardımcı olacaktır (Midkiff ve DaSilva, 2011; Akt. Demir, 2014). Uzaktan eğitimde kullanılan bu modellerin bazı avantajları ve dezavantajları bulunmaktadır (Midkiff ve DaSilva, 2011; Akt. Demir, 2014): Senkron modelin avantajları :

- Gerçek zamanlı tartışma ve beyin fırtınası ortamı oluşturur.
- Herhangi bir problemle karşılaşıldığında anında geri bildirim alınabilir.
- Öğrenci grup içinde olduğu için daha az izole olmuş olur.
- Mekâna bağlı engelleri ortadan kaldırır. Senkron modelin dezavantajları:

- Ders zamanının ayarlanması, ders zaman aralıklarının öğrenciye uymaması gibi problemler yaşanır.

- Sınav değerlendirmesinde gözetmenlik sorunu oluşur.

- Farklı grup çalışmalarında ihtiyaç duyulan teknolojik araçların temini sağlanamaz.

- Öğrenci tartışmalara istediği zaman katılamaz, kalabalık derslerde sadece dinleyen pozisyonuna düşer.

- Teknolojiyle barışık olmayan öğrenciler pasif kalır. Asenkron Modelin Avantajları:

- Yer ve zaman engeli ortadan kalkar.

- Herkesin eğitime katılım olanağı vardır..

- Eğitim uluslararası kimliğe bürünür.

- Öğrenci derse ve tartışmalara istediği şekilde katılabilir.

- Çekingen, iletişim yönü zayıf öğrencinin derse katılımı artar.

Asenkron Modelin Dezavantajları:

- Sanal ve dağıtık bir öğrenim topluluğu yaratır.

- Uygulamalı dersler için uygun değildir.

- Gözetmenli sınavlarda dışarıya bağımlıdır.

- Öğrencilerde izole edilmiş etkisi yaratabilir.

- Anında geri besleme alınmaz.

Tarihî süreç içerisinde eş zamanlı olmayan (asen kron) öğrenme modeli; mektupla öğretim, çoklu ortam modeli - hareketli görüntüler, resimler ve sesler gibi pek çok unsurdan oluşan bir sunumu anlatmaktadır.- (Uşun, 2004) ve esnek öğrenme -öğrenen; öğrenme zamanını başlatmada, istediği öğrenme düzeyini seçmede, yer ve zamanı kendine göre saptamada, öğrenme hızını belirlemede, öğrenme teknolojisini seçmede özgürdür.- (Race, 1998) modeli olarak incelenebilir. Eş zamanlı (senkron) modelde ise radyo, televizyon, telefon, uydu, İnternet gibi teknolojiler

kullanılmaktadır. Bu teknolojiler aracılığıyla iki yönlü ses, iki yönlü görüntü, sesli ve görüntülü konferans uygulamalarına olanak sağlanmaktadır (Uşun, 2004).

Eş zamanlı ve eş zamanlı olmayan öğrenmeyle birlikte teknoloji çağının getirmiş olduğu teknolojik avantajlar yeni bir öğrenme modelini de beraberinde getirmiştir. Gün geçtikçe yeni anlamlar ve nitelikler kazanmaya devam eden bu model karma eğitim modeli (blended learning, hybrid eğitim) olarak ifade edilir. Driscoll'a göre karma eğitim (harmanlanmış eğitim, hybrid eğitim) dört farklı kavramla anlatılır (Driscoll, 2002):

1. Web tabanlı teknoloji gereçlerinin farklı şekillerini karıştırarak ya da birleştirerek kullanmak (sanal sınıflar, kendi kendine öğrenme, birlikte öğrenme, ses, video ya da metin).

2. Eğitim teknolojisi kullanarak ya da kullanmayarak farklı eğitim bilimsel yaklaşımları birleştirerek en iyi eğitim çıktısını elde etmek (yapısalcı, davranışsal ve bilişsel yaklaşım vb.).

3. Farklı eğitim teknolojilerini yüz yüze eğitmen rehberliğindeki uygulamalarla birleştirmek (video teyp, CD-ROM, web tabanlı eğitim, film).

4. Eğitim teknolojisini güncel bilgilerle ve gelişmelerle birleştirerek veya karıştırarak öğrenme ve çalışma arasında uyumlu bir etki yaratmaya çalışmak. Harmanlanmış öğrenme geleneksel eğitimi desteklemek amacıyla teknolojik materyallerin kullanılması olarak düşünülmemelidir. Harmanlanmış öğrenmedeki denge çevrim içi ağırlıklı eğitim veren bir kurumun verimliliği arttırmak amacıyla çevrim içi eğitime ek olarak yüz yüze ders vermesi olarak da düşünülebilir (Picciano, 2008).

2.3.3. MOODLE Öğrenim Yönetim Sistemi

MOODLE, Modular Object Oriented Dynamic Learning Environment teriminin kısaltmasıdır ve Esnek Nesne Yönelimli Dinamik Öğrenme Ortamı olarak çevrilebilir. Açık kaynaklı bir uzaktan eğitim platformu olarak geliştirilmiştir. Bu platform, MySQL ve PostgreSQL gibi veri tabanı sistemleri üzerinde etkin bir şekilde çalışırken, PHP diline uyumlu olan Linux, Windows gibi işletim sistemlerinde de sorunsuzca faaliyet göstermektedir. MOODLE, çevrimiçi ders

yönetim sistemleri arasında yer alıp video konferans uygulamaları (Zoom, Cisco Webex, Google Meet gibi) ve diğer öğrenme yönetim sistemleri (Sakai, Perculus gibi) ile entegrasyon kabiliyetine sahiptir (Yamamoto ve Altun, 2020; Dikmen ve Bahçeci, 2020).

MOODLE, 1999 yılında Avustralya'nın Curtin Üniversitesi'nde Dr. Martin Dougiamas tarafından tasarlanmış ve hayata geçirilmiştir. Bu platformun evrimi sürekli olarak devam edip, düzenli olarak güncellenmekte ve yeni sürümleri yayınlanmaktadır (Höbarth, 2007; Akt. Arslan, 2009). 2012 yılı verilerine göre, MOODLE global bir ölçekte 216 ülkede 66,676 web sitesinde aktif olarak kullanılmakta ve 58,294,719 kayıtlı kullanıcıya sahiptir. MOODLE'in geniş bir kullanıcı kitlesi tarafından tercih edilmesinin temel sebepleri arasında; kullanıcıya odaklı arayüzü, maliyetsiz erişim imkanı, çoklu işletim sistemine uyumluluğu, ölçeklenebilir yapısı, aktif geliştirici topluluğu ve hızla gerçekleştirilen güncelleme süreçleri bulunmaktadır (Mısırlı, 2007; Aydın ve Biroğul, 2008).

MOODLE'in özellikleri şunlardır:

- MOODLE, tamamen ücretsiz bir sistemdir ve hem Windows hem de Linux işletim sistemleri altında çalışabilir.
- Ölçeklenebilir bir yapıya sahiptir ve büyük öğrenci ve ders sayılarına uyum sağlayabilir.
- Ticari rakipleriyle rekabet edebilecek düzeyde işlevselliğe sahiptir.
- Geniş bir geliştirici topluluğuna sahiptir ve yeni özellikler hızla eklenir.
- Öğrenci iletişimi için farklı araçlar sunar, öğrencilerin ders içeriği ile etkileşimde bulunmasını sağlar.
- Öğretmenlerin ders materyali sunma ve etkileşimli etkinlikler oluşturma olanağı vardır (Çevik, 2008).
- Ayrıca MOODLE, ders materyali sunma ve iletişim etkinlikleri oluşturma konusunda çeşitli araçlar sunar.

Ders materyali sunumu için kullanılabilecek araçlar şunlardır:

Düz Metin Dokümanları: Bu format, yalnızca metin bazlı içeriklerin oluşturulmasına imkan tanır.

Web Dokümanları: HTML kodlaması veya belirli bir editör yardımıyla etkileşimli web sayfalarının tasarımı gerçekleştirilebilir.

İnternet bağlantıları: MOODLE içinde dışarıdaki sayfalara bağlantılar ekleyebilirsiniz.

Dosya bağlantıları: Önceden yüklenmiş dosyalara bağlantılar oluşturabilirsiniz.

Etiketler: Başlıklarda metin veya resim ekleyebilirsiniz (Çevik, 2008).

MOODLE ayrıca öğrencilerin etkileşimli ders etkinlikleri yapmasını sağlayan araçlar sunar. Bu etkinlikler arasında öğrencilere ödevler verme, tek soruluk anketler oluşturma, bültenler oluşturma, dersler oluşturma, quizler oluşturma ve anketler düzenleme gibi seçenekler bulunur (Çevik, 2008).

Ayrıca, MOODLE ile öğrencilerin iletişimini artırmak için sohbet odaları, forumlar, sözlükler, wikiler ve işbirlikçi çalışmalar gibi iletişim etkinlikleri de sunulmaktadır (Çevik, 2008). Bu iletişim araçları, öğrencilerin birbirleriyle etkileşimde bulunmasını ve öğrenmeyi daha etkili hale getirmeyi amaçlamaktadır.

2.3.4. MOODLE Sistemi Üzerinden Toplanabilen Verilerin Örnekleri

MOODLE, bir Öğrenme Yönetim Sistemi (LMS) olarak geniş bir yelpazede eğitim verilerini toplama ve izleme yeteneği sunar (Romero ve Ventura, 2008; Akt. Yavuzarslan ve Erol, 2022). Öğrenci performans verileri, öğrenci başarısını ölçmek, öğretim stratejilerini ayarlamak ve öğrenme deneyimlerini kişiselleştirmek için kullanılabilir.

MOODLE üzerinden öğrenci performans verileri kolayca toplanabilir. Bu, notlar, ödev sonuçları ve quiz sonuçları gibi doğrudan öğrenci başarısını yansıtan verileri içerir. Öğretmenler, öğrencilerin dersler boyunca aldıkları notların özetini görüntüleyebilir veya belirli bir ödev veya quizin sonuçlarını analiz edebilir (Dougiamas, 2003).

Ayrıca, MOODLE, öğrencilerin çevrimiçi etkinliklere katılımını izleme olanağı sunar. Bu, forum gönderileri, çevrimiçi tartışmalara yapılan yorumlar ve sohbet odalarındaki etkinlikler gibi verileri içerir. Bu veriler, öğrencilerin çevrimiçi ortamda ne kadar aktif olduklarını ve belirli konuları veya etkinlikleri ne kadar anladıklarını göstermek için kullanılabilir (MOODLE, 2022).

MOODLE ayrıca, öğrenci etkinliklerini log dosyaları aracılığıyla izler. Log dosyaları, öğrencilerin derslere ne zaman eriştiklerini, hangi kaynakları ve etkinlikleri kullandıklarını ve MOODLE'da ne kadar zaman harcadıklarını ayrıntılı bir şekilde belgeleyen bilgileri içerir. Bu tür veriler, öğrenci katılımını ve etkileşimini analiz etmek için değerli bir kaynaktır (Zhang vd., 2023).

Ancak, bu verilerin toplanması ve kullanılması sırasında gizlilik ve veri koruma konularına dikkat edilmesi önemlidir. Öğrenci verileri, belirli etik ve yasal düzenlemelere tabidir ve bu düzenlemelerin titizlikle takip edilmesi gerekmektedir (Slade ve Prinsloo, 2013).

2.3.5. MOODLE Sisteminin Öğrenci Performansının Analizi İçin Kullanılabilirliği

MOODLE, çok yönlü bir Öğrenme Yönetim Sistemi (LMS) olarak geniş bir eğitim verilerini toplama ve analiz etme yeteneği sunar (Zhang vd., 2023).

MOODLE, öğrenci performansının analizinde kullanılacak çeşitli bilgileri sağlar. Veriler, öğrenci notlarını, ödev sonuçlarını ve sınav sonuçlarını içerir. Öğretmenler, bu bilgileri kullanarak öğrencinin genel akademik performansını değerlendirebilirler. Ayrıca, öğrencinin belirli konularda veya etkinliklerde ne kadar başarılı olduğunu anlamak için öğrenci notlarını ve ödev sonuçlarını analiz edebilirler (Dougiamas, 2003).

MOODLE aynı zamanda öğrencilerin çevrimiçi etkinliklere katılımını izler. Örneğin, bir öğrencinin forum gönderilerini veya çevrimiçi tartışmalara yaptığı yorumları analiz ederek, öğrencinin belirli bir konuyu ne kadar iyi anladığını veya konuyu ne kadar derinlemesine keşfettiğini belirlemeye yardımcı olabilir (MOODLE, 2021).

MOODLE, log dosyaları aracılığıyla öğrenci etkinliklerini de izler. Log dosyaları, öğrencilerin derslere ne zaman eriştiklerini, hangi kaynakları ve etkinlikleri kullandıklarını ve MOODLE'da ne kadar zaman harcadıklarını detaylı olarak belgeleyen bilgiler içerir. Bu tür veriler, öğrenci katılımını ve etkileşimini analiz etmek için değerli bir kaynak olabilir (Romero ve Ventura, 2008; Akt. Yavuzarslan ve Erol, 2022).

MOODLE, öğretmenlere ve eğitimcılara öğrenci performansını daha iyi anlama ve öğretim stratejilerini öğrenci ihtiyaçlarına göre ayarlama konusunda güçlü bir araç sunar. Ancak, bu verilerin kullanılması ve analiz edilmesi sırasında gizlilik ve veri koruma konularına özen gösterilmesi önemlidir (Slade ve Prinsloo, 2013).

2.3.5.1. MOODLE İle Veri Toplamanın Avantajları

MOODLE, öğrenci etkinliklerinin, ders materyallerine erişimin ve öğrenme sonuçlarının çeşitli yönlerini kapsayan geniş bir veri seti sağlar (Dougiamas, 2003). Ayrıca, MOODLE sistemi öğrencilerin etkinliklerini otomatik olarak izler ve kaydeder, bu da manuel veri toplama sürecinden tasarruf sağlar (Romero ve Ventura, 2008; Akt. Yavuzarslan ve Erol, 2022). Bu otomatik veri toplama süreci, eğitimcilerin daha fazla zamanlarını öğrenci etkileşimini analiz etmeye harcamalarına yardımcı olur.

MOODLE ayrıca, öğrencilerin öğrenme sürecine gerçek zamanlı olarak bakma olanağı sağlar, bu da eğitimcilerin öğrencilerin anlayışını ve etkileşimini hemen değerlendirebilecekleri geri bildirim sağlar (Zhang vd., 2023). Bu gerçek zamanlı analizler, öğrencilerin gerektiğinde desteklenmelerine veya özelleştirilmiş öğrenme planlarına erişmelerine olanak tanır.

2.3.5.2. MOODLE İle Veri Toplamanın Dezavantajları

MOODLE üzerinden veri toplama, öğrencinin gizliliği ve veri güvenliği ile ilgili konuları gündeme getirir (Slade ve Prinsloo, 2013). Özellikle hassas bilgilerin toplandığı durumlarda, bu konuların dikkate alınması ve uygun güvenlik önlemlerinin alınması önemlidir.

MOODLE'dan toplanan verinin doğru bir şekilde yorumlanması ve anlaşılması karmaşık olabilir. Verilerin anlamını çözümlemek için genellikle daha fazla analitik yetenek ve uzmanlık gereklidir (Zhang vd., 2023). Bu, veri analizi sürecinin karmaşıklığını artırabilir ve doğru kararlar almayı zorlaştırabilir.

MOODLE sadece çevrimiçi etkinlikleri kaydeder, bu da öğrencilerin ders dışı etkinlikleri veya MOODLE dışındaki etkileşimlerinin verilerini toplamanın zorluğunu oluşturabilir (Romero ve Ventura, 2008; Akt. Yavuzarslan ve Erol, 2022).

Bu durum, tam bir öğrenci deneyimi veri seti oluşturmanın zorluklarını ortaya çıkarabilir.

MOODLE kullanmanın avantajları ve dezavantajları vardır. Geniş veri toplama yetenekleri sunar, ancak veri güvenliği ve yorumlama karmaşıklığı gibi bazı zorluklarla karşılaşabilirsiniz. Bu nedenle, MOODLE kullanırken dikkatli bir şekilde planlama ve yönetim gereklidir.



3. YÖNTEM

Araştırma yöntemi belirlenirken araştırma modeli oluşturulmuş, çalışma kümesi belirlenmiş, veri seti oluşturulmuş ve veri seti hazırlanmıştır. Öznelik bilgileri, veri işleme araçları, Veri kaynakları ve veri toplama araçları incelenmiştir.

3.1. Araştırma Modeli

Bu çalışma, Gedik Üniversitesi öğrencilerinin öğrenme davranışlarını ve performanslarını daha iyi anlamak ve öğretim stratejilerini bu verilere dayalı olarak optimize etmek amacıyla bu çalışma kümesini kullanacaktır. Elde edilen bulgular, öğrencilerin öğrenme deneyimlerini iyileştirmeye yönelik öneriler sunmak ve eğitim kalitesini artırmak için kullanılacaktır.

Çalışma kümesi seçimi, araştırmanın geçerliliği ve sonuçlarının genelleme yapılabilirliği açısından büyük önem taşımaktadır. Bu nedenle, çalışma kümesinin dikkatli bir şekilde seçilmesi ve analiz edilmesi, bu araştırmanın başarıyla tamamlanmasının temelidir.

Bu tez, MOODLE eğitim yönetim sistemi üzerindeki verilerin makine öğrenmesi algoritmaları kullanılarak analiz edilerek öğrencilerin dönem sonu akademik performanslarının tahmin edilmesini amaçlamaktadır. Bu çalışmada, öğrencilerin MOODLE üzerindeki etkileşimleri, notları, quiz sonuçları, ödev teslimleri gibi verilerin analiz edilmesi ile bir özellik vektörü oluşturulmuştur.

Araştırmada, öncelikle veriler toplanmış ve işlenmiştir. Daha sonra, 5 farklı makine öğrenmesi algoritması kullanılarak iki sınıf şeklinde sınıflandırma yapılmıştır. Bu sınıflandırmada, notlar 1-50 arası başarısız, 50-100 arası başarılı olarak iki sınıfa ayrılmıştır. İkinci olarak, aynı verileri kullanarak çoklu sınıflandırma yapılmıştır. Bu sınıflandırmada notlar 0-20 , 20-40 , 40-60, 60-80, 80-100 olarak ayrılmıştır.

Sonuçlar, her iki sınıflandırma yöntemi için de iyi sonuçlar vermiştir. Çalışmanın sonuçları, MOODLE eğitim yönetim sistemi üzerinde makine öğrenmesi algoritmalarının kullanılarak öğrenci performansının tahmin edilmesinin mümkün olduğunu göstermiştir. Bu çalışma, öğrenci performansının daha doğru bir şekilde tahmin edilmesine ve öğrencilerin daha iyi bir şekilde yönlendirilmesine yardımcı

olabilir. Ayrıca, eğitim yönetim sistemlerindeki öğrenci etkileşimlerinin değerlendirilmesi ve öğrenci performansının öngörülmesi için makine öğrenmesi algoritmalarının kullanımının yaygınlaştırılmasına da katkı sağlayabilir.

Çalışmada araştırmanın gerçekleştirilmesi, İstanbul Gedik Üniversitesi MOODLE Sistemi veritabanlarındaki verilerin kullanılmıştır. Kullanılan veriler içerisinde kişisel veriler yer almamaktadır. Verilerin kullanılması kurum izni alınarak gerçekleştirilmiştir (EK-1).

3.2. Çalışma Kümesi

Bu çalışma, Gedik Üniversitesi öğrencilerinin MOODLE öğrenme yönetim sistemi ve OBS üzerinde bulunan öğrenci verilerine dayanmaktadır. MOODLE ve OBS üzerinde toplanan bu veriler, öğrencilerin akademik performansı, çevrimiçi etkinliklere katılımı ve öğrenme deneyimleri hakkında değerli bilgiler içermektedir.

Çalışma kümesi, Gedik Üniversitesi öğrencilerinin belirli bir döneme veya öğrenme sürecine ait verileri içermektedir. Bu veriler, öğrencilerin ders notları, ödev sonuçları, sınav skorları gibi çeşitli ölçümleri içermektedir.

3.3. Veri Seti

Veri seti için toplam 6 özellik ve 1952 örnek bulunmaktadır.

3.3.1. Veri Setinin Hazırlanması

Veri seti sınıflandırma ve regresyon modelleri altında oluşturulmuştur. Veri seti, bu tez çalışmasına da konu olan öğrencilerin akademik not başarılarını tahmini üzerine kurulmuş bir yapıdadır. Veri setindeki bağımlı değişken bütünleme notu olarak ele alınacaktır. Veri setinde bulunan özellikler Çizelge 1.'de gösterilmiştir.

Tablo 1. Veri Seti Nitelik Adları ve Türü

Sıra No	Nitelik Adı
1	Yaş
2	Fakülte
3	Ara Sınav Notu
4	Cinsiyet
5	Final Notu
6	Bütünleme Notu

Kaynak: Yazar tarafından oluşturulmuştur.

Bu araştırmada kullanılan veri seti "CSV" formatında sunulmuştur. Python dilinde CSV formatını okuyabilmek için "Pandas" kütüphanesi kullanılmıştır. Veri üzerinde hesaplamalar ve dönüşümler için "NumPy", veri seti analizi için ise "Pandas" kullanılmıştır.

3.3.2. Öznitelik Bilgileri

Bu çalışmada 1952 kişinin çeşitli özellikleri içeren MOODLE ve OBS den alınan veriler kullanılmıştır. Uygulama Python dilinde yazılmıştır. Uygulamada veri görselleştirme Matplotlib ve Seaborn kütüphaneleri kullanılarak oluşturulmuştur.

Çalışmada makine öğrenimi algoritmaları kullanılarak bağımsız olarak işlenmiştir. Toplamda 1952 veri içerisinde %70'lik kısmı eğitim amaçlı, %30'luk kısmı ise test amaçlı olarak ayrılmıştır. Bu araştırmanın odak noktası, kullanılan çeşitli algoritmalarla kaynaklı sonuçların kalitesini yükseltmek ve doğruluk oranını optimize etmektir.

3.3.3. Veri İşleme Araçları

Veri işleme aracı olarak python, pandas kütüphanesi, numpy kütüphanesi, scikit-learn kütüphanesi, matplotlib kütüphanesi kullanılmıştır. Ayrıca anaconda platformu ve Jupyter Notebook kullanılmıştır.

3.3.3.1. Python

Python, 1990'larda Guido Van Rossum tarafından tasarlanmış ve geliştirilmeye başlanmış bir programlama dilidir (Özgül, 2013). Python, yazım biçimi bakımından diğer programlama dillerine göre daha kolay bir yapıya sahiptir ve herhangi bir derleyiciye ihtiyaç duyulmadan kullanılabilir.

Python, ücretsiz ve açık kaynak kodlu bir programlama dilidir ve bu özelliği sayesinde geniş bir kullanıcı kitlesine hitap etmektedir. Python'ın en belirgin özelliklerinden biri, zengin bir kütüphane ekosistemine sahip olmasıdır. Bu kütüphaneler, Python'ı diğer programlama dillerinden ayıran önemli bir özelliktir. Özellikle veri bilimi ve makine öğrenmesi alanlarında kullanılan TensorFlow, Keras ve Scikit-learn gibi kütüphaneler Python'ın popülerliğini artırmaktadır (Arslan, 2019).

Ayrıca, veri analizi ve görselleştirme konularında sıklıkla kullanılan pandas, numpy ve matplotlib gibi kütüphaneler de Python'ın gücünü artıran araçlardır. Bu kütüphaneler sayesinde Python, veri analizi ve işleme süreçlerinde etkili bir şekilde kullanılabilmektedir.

Python'ın bu geniş kütüphane desteği, çeşitli endüstrilerde ve bilimsel alanlarda kullanılmasını kolaylaştırmaktadır. Ayrıca, Python topluluğunun sürekli olarak yeni kütüphaneler ve araçlar geliştirmesi, dilin gelişmeye ve güncel kalmaya devam etmesini sağlamaktadır. Bu nedenle Python, programcılar ve araştırmacılar için vazgeçilmez bir araç haline gelmiştir.

3.3.3.2. Pandas Kütüphanesi

Python için geliştirilmiş olan pandas kütüphanesi, birçok işlemi daha hızlı ve kolay bir şekilde gerçekleştirmeyi sağlamaktadır. Pandas kütüphanesi, Python yazılım dili için özel olarak geliştirilmiş bir veri işleme kütüphanesidir ve özellikle veri analizi ve manipülasyonu alanında oldukça güçlüdür (Chen, 2017).

Pandas, zaman serileri, sayısal dizi işlemleri ve veri yapıları oluşturmak ve bunlar üzerinde işlem yapmak için geniş bir araç yelpazesi sunar. Veri çerçeveleri ve seriler gibi veri yapıları, verileri düzenlemek ve analiz etmek için kullanılabilir. Bu sayede büyük veri kümeleri üzerinde hızlı ve etkili işlemler gerçekleştirilebilir.

Kütüphaneyi kullanmaya başlamak için sadece "import pandas" komutunu kullanmak yeterlidir (McKinney W., 2022). Pandas, veri bilimcileri ve veri analistleri tarafından yaygın olarak tercih edilen bir araçtır ve Python programcılarına veri işleme konusunda önemli avantajlar sunar. Bu nedenle, pandas kütüphanesi Python ekosisteminin vazgeçilmez bir parçası haline gelmiştir.

3.3.3.3. Numpy Kütüphanesi

Çok boyutlu diziler üzerinde matematiksel dizi işlemlerini hızlı ve verimli bir şekilde gerçekleştirmek amacıyla geliştirilen bir Python kütüphanesi, Harris ve diğerleri tarafından 2020 yılında oluşturulmuştur. Bu kütüphane, veri analitiği ve bilimsel hesaplamalar alanında geniş bir uygulama yelpazesine sahiptir.

Numpy kütüphanesini kullanmaya başlamak için sadece "import numpy" komutunu kullanmanız yeterlidir. Bu komut, kütüphanenin ilgili modülünü projenize ekler ve böylece çok boyutlu dizilerle çalışmaya başlayabilirsiniz (NumPy Developers, 2022). Numpy, Python programcıları için güçlü bir araçtır ve matematiksel işlemlerin yanı sıra veri manipülasyonu için de geniş olanaklar sunar.

Özetle, Numpy, çok boyutlu dizilerle çalışırken performansı artırmak ve işlemleri daha kolay hale getirmek için oldukça faydalı bir araçtır. Bu nedenle, veri bilimi, bilimsel hesaplamalar ve mühendislik gibi birçok alanda sıklıkla tercih edilmektedir.

3.3.3.4. Scikit-Learn Kütüphanesi

Python içerisinde bulunan birçok ücretsiz kütüphane arasında, Scikit-learn kütüphanesi de yer almaktadır. Scikit-learn, ücretsiz bir Python kütüphanesidir ve özellikle makine öğrenimi ve veri bilimi alanında yaygın olarak kullanılmaktadır. Bu kütüphane, içinde birçok farklı algoritma barındırması nedeniyle birden fazla işlemin kolayca gerçekleştirilmesine olanak tanır (Karabay, 2020).

Scikit-learn kütüphanesi, Python programcıları için büyük bir avantaj sunar. Makine öğrenimi projeleri geliştirirken kullanabileceğiniz bir dizi araç ve algoritma içerir. Ayrıca, veri madenciliği, veri analizi ve model oluşturma gibi birçok farklı görevde işlerinizi kolaylaştırabilir.

Özetlemek gerekirse, Scikit-learn kütüphanesi, Python ile makine öğrenimi ve veri bilimi uygulamaları geliştirmek isteyenler için vazgeçilmez bir kaynaktır. Ücretsiz olması ve geniş özelliklere sahip olması, bu kütüphanenin popülerliğini artırmaktadır.

3.3.3.5. Matplotlib Kütüphanesi

Matplotlib kütüphanesi, verileri belirli formatlarda ve farklı grafik türleriyle durağan veya etkileşimli olarak görselleştirmek için kullanılan bir Python kütüphanesidir. Bu kütüphane, verileri düzenleyebilir, yakınlaştırabilir veya uzaklaştırabilir ve yüksek kaliteli çıktılar üretebilir (Matplotlib Geliştirme Ekibi, 2012-2023).

Matplotlib, özellikle veri analizi, veri görselleştirme ve bilimsel hesaplama gibi alanlarda yaygın olarak kullanılmaktadır. Grafiklerin özelleştirilebilir olması ve farklı grafik türlerini desteklemesi, araştırmacıların ve veri bilimcilerinin çeşitli veri görselleştirme ihtiyaçlarını karşılamalarına olanak tanır.

Bu nedenle, Matplotlib kütüphanesi Python programcıları ve veri bilimcileri için önemli bir araçtır ve veri tabanlı görselleştirmelerde kullanılmak üzere tasarlanmıştır.

3.3.3.6. Anaconda

Anaconda, makine öğrenimi, veri analizi ve veri bilimi gibi uygulamalar için özel olarak tasarlanmış bir platformdur. Bu platform, birçok uygulama yazılımı ve kütüphane içeren birleşik bir Python ekosistemine ev sahipliği yapar (Kaplan, 2019). Anaconda platformu, bu tez çalışmasında kullanacağımız Jupyter Notebook dahil olmak üzere çeşitli araçları içerir.

3.3.3.7. Jupyter Notebook

Jupyter Notebook, Python programlama dilinin yanı sıra diğer programlama dillerine de destek sağlayabilen ve programlama sürecini kolaylaştıran bir yazılımdır (Tuzcu, 2020). Bu yazılımın arayüzü son derece basit ve kullanıcı dostudur. Basitliği sayesinde yazılan kodların kolayca okunmasına ve derlenmesine olanak tanır. Bu

nedenle, özellikle Python programlama dilini öğrenmek isteyenler için önemli bir yardımcı araçtır.

3.4. Veri Kaynakları ve Veri Toplama Araçları

Veriler veri toplama yöntemleri ve teknikleri belirlenerek düzenlenmiştir. CRISP-DM yöntemi ve süreçleri kullanılmıştır. Önce problem tanımlanmış, karşılaşılabilecek problemlere karşı verilerin anlaşılması adımı tamamlanmıştır. Problem tanımlandıktan sonra model oluşturulmuştur. Model oluşturulduktan sonra değerlendirilmesi yapılmış ve uygulama yapılmıştır.

3.4.1. Veri toplama yöntemleri ve teknikleri

Veri toplama, bilimsel araştırma, iş analitiği, toplum hizmetleri ve çeşitli diğer uygulamalar için temel bir işlevidir (Bryman ve Bell, 2015). Farklı bağlamlarda ve amaçlar doğrultusunda çeşitli veri toplama yöntemleri ve teknikleri bulunmaktadır. Bu yöntemler şunlardır:

Anketler: Anketler, farklı alanlarda yaygın olarak kullanılan bir veri toplama tekniğidir. İnternet, telefon veya yüz yüze görüşmeler aracılığıyla uygulanabilirler. Genellikle geniş bir hedef kitlenin görüşlerini, davranışlarını veya özelliklerini hızlı ve ekonomik bir şekilde toplamak için kullanılırlar (Fowler, 2013; Akt. Yirci vd., 2018).

Mülakatlar: Mülakatlar, daha derinlemesine bilgi elde etmek amacıyla kullanılır ve genellikle daha az katılımcıyla gerçekleştirilir. Ancak her bir katılımcıdan daha ayrıntılı bilgi alınır. Mülakatlar, genellikle bireysel deneyimler, anılar veya görüşler hakkında daha zengin ve karmaşık veriler elde etmek için kullanılır (Brinkmann ve Kvale, 2015).

Gözlemler: Gözlemler, bir olayı veya durumu doğrudan ve genellikle gerçek zamanlı olarak inceleme fırsatı sunar. İnsanların davranışlarını, etkileşimlerini ve çevresel bağlamı incelemek için özellikle yararlıdır (Angrosino, 2007; Akt. Kadriye ve Aydın, 2012). Doküman Analizi: Doküman analizi, mevcut belgeleri veya kayıtları inceleyerek veri toplama sürecini içerir. Bu tür dokümanlar, hükümet

raporları, şirket bilançoları, gazete makaleleri, bloglar, forumlar, sosyal medya gönderileri gibi birçok farklı kaynağı kapsayabilir (Bowen, 2009).

Büyük Veri ve Veri Madenciliği Perspektifi: Dijital teknolojinin ve internetin evrimiyle birlikte, büyük veri ve veri madenciliği modern veri analiz yöntemleri arasında önemli bir konuma sahip olmuştur. Bu teknikler, genellikle otomatik olarak toplanan büyük veri setlerinden karmaşık desenler, eğilimler veya ilişkiler çıkarmayı içerir (Witten vd., 2016).

Bu farklı veri toplama yöntemleri, belirli bir araştırma veya analiz amacına ve gereksinimlerine göre seçilmelidir.

3.4.2. CRISP-DM Yöntemi Süreçleri

Veri madenciliği çalışmalarında, problem tanımından başlayarak sonuca ulaşma süreci boyunca tüm adımların planlanmasında, genellikle CRISP DM döngüsü, yani Çapraz Endüstri Veri Madenciliği Süreci sıkça kullanılmaktadır (Şeker, 2018). Bu tez çalışmasında izlenecek temel yöntem, CRISP-DM adımları ve akış şemasına dayalı olarak şekillendirilecektir. CRISP-DM adımları ve akışı şu şekildedir:

- İş Probleminin Tanımlanması ve Kavramsallaştırılması
- Veri Setinin İncelenmesi ve Kavranması
- Veri Ön İşleme ve Düzenlenmesi
- Algoritma Uygulaması ve Model Oluşturma
- Modelin Performansının Ölçülmesi
- Üretim Ortamına Entegrasyonu

Bu adımlar, veri madenciliği sürecinin başından sonuna kadar olan sürecin temelini oluşturur ve bu çalışmanın yol haritasını belirler.

3.4.2.1. İşin Anlaşılması ve Problemin Tanımlanması

Eğitim ve öğretim süreçleri, toplumun gelişiminde temel bir rol oynamaktadır. Bu süreçlerin daha etkili ve verimli bir şekilde yönetilmesi,

öğrencilerin akademik başarılarını artırmak ve eğitim kurumlarının daha iyi sonuçlar elde etmesine katkı sağlamak için kritik öneme sahiptir. Bu durum göz önünde bulundurulduğunda, öğrenci başarısını öngörmek ve bu başarıyı etkileyen unsurları analiz etmek, eğitimsel çalışmalarda kritik bir inceleme alanı olarak öne çıkmıştır. Bu çalışma, öğrencilerin akademik performanslarını tahmin etmek için makine öğrenimi algoritmalarının kullanılmasını ve performansı etkileyen değişkenlerin belirlenmesini amaçlamaktadır. Bu amaçla, çalışmanın temel problemi ve hedefleri aşağıda daha ayrıntılı bir şekilde tanımlanmıştır.

Bu tez çalışmasının işi ve problemi, öğrencilerin eğitim ve öğretim süreçlerinde akademik performanslarının makine öğrenimi algoritmaları kullanılarak tahmin edilmesi amacını taşımaktadır. Ayrıca, bu çalışma öğrencilerin akademik başarılarını etkileyen özellikleri, nitelikleri ve değişkenleri tespit etmeyi amaçlamaktadır. Eğitim ve öğretim süreçlerinin daha hızlı, etkili ve kaliteli bir şekilde ilerlemesine katkı sağlayarak, öğrenciler hakkında daha kapsamlı ve nitelikli bilgilerin elde edilmesini hedeflemektedir. Bu bilgiler, öğrencilere, eğitmenlere, eğitim kurumlarına ve ilgili sorumlulara sağlam bir temel sunarak, daha bilimsel ve etkili kararlar alınmasına yardımcı olacaktır. Bu tez çalışması, akademik performans tahmini ve eğitim süreçlerinin iyileştirilmesi konularında önemli bir katkı sağlamayı amaçlamaktadır.

3.4.2.2. Verinin Anlaşılması

Modelin tasarlanması ve kurulması aşamasında, karşılaşılabilecek sorunları önlemek için verinin iyi bir şekilde anlaşılması çok önemlidir (Ataseven, 2008). Bu aşama, modelleme sürecinin başarılı bir şekilde ilerlemesi için kritik bir rol oynamaktadır. Verinin anlaşılması ve analizi, modelleme sürecinin temelini oluşturur. Verinin ne kadar iyi anlaşılıp analiz edilirse, modelleme aşaması da o kadar iyi bir şekilde planlanmış olur. Bu nedenle, bu aşamanın titizlikle yapılması gerekmektedir.

Bu aşamada, öğrenci akademik performansı ile ilgili veri setini ele alıyoruz. Veri seti, yaş, fakülte, ara sınav notu, cinsiyet, final notu ve bütünleme notu gibi altı önemli başlıktan oluşmaktadır. Bu başlıklar, öğrencilerin akademik başarılarını etkileyen faktörleri incelemek için önemli bir temel oluşturmaktadır. Veri setimiz,

toplam kaç öğrenciyi içeriyor ve bu öğrenciler hangi fakültelerden geliyorlar gibi temel istatistiksel bilgilere sahibiz. Ayrıca, yaş ve cinsiyet gibi demografik verilerle ilgili önemli özet istatistikleri de elde edebiliyoruz. Ana analizimiz, öğrenci başarısını belirlemek için bütünleme notunu incelemek olacaktır. Ayrıca, yaş, fakülte ve cinsiyet gibi faktörlerin akademik performans üzerindeki etkilerini de göz önünde bulunduracağız.

Bu aşamada, veri setimizi daha da ayrıntılı bir şekilde inceleyerek hangi istatistiksel yöntemleri kullanacağımızı belirlemeye çalışıyoruz. Veri madenciliği sürecinde daha ileri adımlar atmadan önce verileri anlamak ve özellikle analiz için uygun olanları seçmek önemlidir. Bu adım, model oluşturma ve sonuçların yorumlanması aşamalarına sağlam bir temel oluşturacaktır.

3.4.2.3. Verinin Hazırlanması

Bu aşamada, ham verilerin düzenlemesi işlemi için Python programlama dilinin önemli kütüphaneleri olan Pandas ve NumPy büyük bir öneme sahiptir. Pandas ve NumPy kütüphaneleri, veri hazırlama aşamasında kullanılan güçlü araçlardır. İşte bu aşamada takip edilen adımlar:

Veri Temizliği: Veri setindeki eksik veya hatalı verileri tanımlayarak düzeltilmelerini sağlıyoruz. Pandas kütüphanesi sayesinde bu verileri tespit edebilir ve düzeltebiliriz.

Veri Dönüşümü: Veri setindeki sayısal olmayan verileri sayısal forma dönüştürüyoruz. Örneğin, cinsiyet gibi kategorik verileri sayısal kodlamalarla değiştiriyoruz.

Özellik Mühendisliği: Veri setini analiz etmek için yeni özellikler oluşturabiliriz. Örneğin, ara sınav notu ile final notu arasındaki farkı hesaplayarak öğrencilerin performans trendini yakalayabiliriz.

Veri Bölme: Veri setini eğitim ve test verileri olarak ayırıyoruz. Bu, modelin eğitimi ve sonuçların değerlendirilmesi için kullanılır.

Özellik Seçimi: Hangi özelliklerin analize dahil edileceğini belirliyoruz. Araştırmamızın odaklandığı özellikler olan yaş, fakülte, ara sınav notu, cinsiyet, final notu ve bütünleme notu bu aşamada öne çıkan özelliklerdir.

Pandas ve NumPy kütüphaneleri, veri setinin işlenmesi ve düzenlenmesi için güçlü bir temel sağlar. Bu adımların tamamlanması, veri setinin model oluşturmak için uygun hale gelmesini sağlar. Verilerin düzenlenmesi ve hazırlanması, analiz ve tahmin yapma aşamalarında daha kesin sonuçlar elde etmemize yardımcı olur. Bu kütüphaneler, veri madenciliği sürecinin bu kritik aşamasında büyük önem taşır.

3.4.2.4. Modelin Oluşturulması

Bu bölümde, çalışmanın temelini oluşturan modelin oluşturulma sürecini detaylı bir şekilde ele alıyoruz. İlk adım olarak, uygulamada kullanılacak verilerin nasıl işleneceğini ve modelleme için nasıl hazırlanacağını anlatıyoruz. Veri setinin bölünmesi, temizlenmesi ve gerekli dönüşümlerin yapılması bu aşamanın önemli bir parçasıdır.

Ardından, modelin kendisi oluşturulur. Bu çalışmada, öğrenci akademik performansını tahmin etmek için SVM (Destek Vektör Makineleri), RF (Rastgele Orman), KA (Karar Ağaçları), KNN (En Yakın Komşu) ve NB (Naif Bayes) algoritmaları kullanılarak farklı modellemeler gerçekleştirilir. Hem ikili sınıflandırma hem de çoklu sınıflandırma senaryoları göz önünde bulundurularak her bir algoritma ayrı ayrı değerlendirilir.

Bu aşama, çalışmanın en önemli bölümlerinden biridir çünkü model, öğrenci akademik performansının tahminlenmesi ve sınıflandırılması için kullanılacaktır. İlgili veri madenciliği yöntemleri ve makine öğrenimi algoritmaları, öğrenci başarılarını anlamamıza ve daha iyi eğitim destekleri sunmamıza yardımcı olacaktır. Modelin sonuçları, çalışmanın ilerleyen aşamalarında daha ayrıntılı olarak ele alınacak ve sonuçların yorumlanmasına katkı sağlayacaktır.

3.4.2.5. Modelin Değerlendirilmesi

Araştırmanın evresi, projenin kapsamını, problem tanımını ve veri ön işleme ile model oluşturma süreçlerinin kapsamlı bir analizini kapsamaktadır. Bu evrede, karar ağaçları, k-en yakın komşu, random forest, destek vektör makineleri ve naive bayes gibi çeşitli makine öğrenimi algoritmalarının bağımsız modellenmeleri ve performans metrikleri temelinde karşılaştırmaları esastır. Her bir modelin

performansı ayrı ayrı incelenir ve elde edilen sonuçlar dikkatle değerlendirilir. İlgili metrikler kullanılarak her bir modelin ikili ve çoklu sınıflandırma senaryolarında nasıl performans gösterdiği belirlenir. Bu değerlendirme, istenen sonuçlara ulaşıp ulaşılamadığını anlamamıza yardımcı olur.

Ayrıca, bu değerlendirme süreci, modelleme aşamasının gözden geçirilmesi gerektiği durumlarda erken uyarı sistemini sağlar. Yani, eğer belirlenen başarı kriterleri karşılanmazsa, modelleme teknikleri ve yöntemi değerlendirme sonuçlarına bağlı olarak yeniden şekillendirilebilir. Bu esnek yaklaşım, araştırmanın başından sonuna kadar veri anlayışından başlayarak sonuçlara yönelik bir yaklaşımın temelini oluşturur. Model değerlendirmesi bu çalışmanın kritik bir aşamasıdır ve elde edilen sonuçlar, öğrenci akademik performansını tahmin etme amacını daha iyi anlamamıza ve geliştirmemize olanak tanır.

3.4.2.6. Canlıya Alma

Bu aşama, tüm önceki adımların başarıyla tamamlandığı ve modelin oluşturulduğu noktada gerçekleşir. Canlıya alma, uygulama sonuçlarını son kullanıcılara anlaşılır bir şekilde sunmayı içerir. Bu aşamada ayrıca, çalışma sonucunda elde edilen verilerin görselleştirilmesi ve bu görselleştirmelerin detaylı bir şekilde yorumlanarak analiz edilmesi de gerçekleştirilir.

Uygulamanın sonuçları, grafikler, tablolar ve açıklayıcı metinler aracılığıyla etkili bir şekilde görselleştirilir. Bu görselleştirmeler, elde edilen sonuçların daha kolay anlaşılmasına ve yorumlanmasına yardımcı olur. Ayrıca, uygulama sonuçları, modelin başarıları ve performansı ile ilgili önemli bilgileri içerir.

Canlıya alma aşaması, çalışmanın sonunda elde edilen sonuçların paydaşlara sunulmasını ve kullanılmasını sağlar. Aynı zamanda, modelleme sürecinin başlangıcından sonuna kadar olan veri anlayışını ve sonuçlara yönelik bir yaklaşımın temelini oluşturur. Bu aşama, çalışmanın tamamlanması ve elde edilen bulguların kullanıma hazır hale getirilmesi için kritik bir adımdır.

3.4.3. Öğrenci Performansı Tahmini

Öğrenci performansının tahmini, eğitimciler ve eğitim politika yapıcılarının son derece kritik bir rol oynamaktadır. Bu tahminler, öğrenci başarısını değerlendirmek ve daha etkili öğretim stratejileri geliştirmek amacıyla kullanılmaktadır (Kotsiantis, 2012). Geleneksel tahmin yöntemleri, öğrencinin demografik bilgileri, önceki akademik başarıları ve kişisel özellikleri gibi faktörleri içermektedir. Ancak, bu yöntemler, büyük ve karmaşık veri setlerini işlemekte yetersiz kalmakta ve bazen yanıltıcı sonuçlara neden olabilmektedir (Romero ve Ventura, 2008; Akt. Yavuzarslan ve Erol, 2022). Son yıllarda, öğrenci performans tahmininde makine öğrenmesi ve yapay zeka teknolojileri önemli bir rol oynamaya başlamıştır. Bu teknolojiler, geniş ve karmaşık veri setlerini daha etkili bir şekilde işleyebilir ve daha doğru tahminler yapabilir (Márquez-Vera et al., 2013). Örneğin, sınıflandırma algoritmaları, öğrencilerin belirli bir dersi geçme olasılığını tahmin etmek için kullanılabilir. Ayrıca, regresyon algoritmaları, bir öğrencinin belirli bir sınavda alacağı notu tahmin etmek için uygulanabilir. Yapay sinir ağları ise öğrenci başarısını etkileyen karmaşık faktörler arasındaki ilişkileri modelleyebilir ve gizli faktörleri tespit edebilir (Huang ve Fang, 2013).

Bu teknolojilerin kullanımı, öğrencilere daha kişiselleştirilmiş ve etkili öğrenme deneyimleri sunma potansiyeline sahiptir. Ancak, bu teknolojilerin etik ve gizlilik endişeleri göz önünde bulundurularak kullanılması gerekmektedir. Özellikle, öğrenci verilerinin korunması ve bu verilerin kullanımının şeffaf bir şekilde yönetilmesi büyük bir önem taşımaktadır (Slade ve Prinsloo, 2013). Bu nedenle, öğrenci performans tahmininde teknoloji kullanımı, dikkatle denetlenmeli ve bu hususlara özen gösterilmelidir.

4. BULGULAR

Yapılan analizler sonucunda ikili ve çoklu sınıflandırma yapılmıştır. Hem ikili sınıflandırmada hem de çoklu sınıflandırmada RF, KNN, KA, SVM, NB algoritmalarına göre analizler yapılmıştır. Yapılan analizlerin sonucunda bulguları yazılmıştır.

4.1. Veri Seti Nitelik Analizi

Veri seti 6 öznitelikten oluşmaktadır. 6 özneliğin 5 tanesi sayısal kalan 1 tanesi ise kategorik değişkendir. Özniteliklerin isimleri ve tipleri Tablo 2’ de gösterilmektedir.

Tablo 2. Veri Seti Nitelik Adları ve Türü

Sıra No	Nitelik Adı	Tip
1	Yaş	İnt64
2	Fakülte	Object
3	Ara Sınav Notu	İnt64
4	Cinsiyet	İnt64
5	Final Notu	İnt64
6	Bütünleme Notu	İnt64

Kaynak: Yazar tarafından oluşturulmuştur.

Veri seti incelendiğinde 1952 öğrenci yer almaktadır. Bu öğrencilerin yaş ortalaması 23.7’ dir. Öğrencilerin sınav ortalamalarına baktığımızda Tablo 3’ de görüldüğü gibi en düşük ortalama bütünleme sınavındadır.

Tablo 3. Veri Özelliklerinin Sayısal Değerleri

	Miktar	Ortalama	Standart Sapma	Min.	25%	50%	75%	Max.
Yaş	1952	23.7	4.7	18.0	21.0	23.0	24.0	59.0
Ara Sınav Notu	1952	32.5	29.7	0.0	0.0	29.0	55.0	100.0
Final Notu	1952	16.9	20.0	0.0	0.0	5.0	34.0	88.0
Bütünleme Notu	1952	1.3	0.4	1.0	1.0	1.0	2.0	2.0

Kaynak: Yazar tarafından oluşturulmuştur.

4.2. İkili Sınıflandırma

Yöntem bölümünün altında, CRISP-DM yöntemlerinin 3. aşaması olan verinin hazırlanması süreci detaylı olarak açıklanmaktadır. Bu aşamada, veri setinin modelleme aşamasına taşınabilmesi için "Bütünleme notu" özelliğini ikili sınıflandırma ile düzenleme işlemi gerçekleştirilmektedir.

Tablo 4. Bütünleme Notu İkili Sınıflandırma

İkili Sınıflama	Bütünleme Notu
Yüksek	$100 \geq \text{Bütünleme} \geq 51$
Düşük	$50 \geq \text{Bütünleme} \geq 0$

Kaynak: Yazar tarafından oluşturulmuştur.

Çizelge 4 incelendiğinde, bütünleme notları, 0 ile 100 arasındaki notlardan oluşmaktadır. Modelleme öncesinde, bu notlar 1 – 50 aralığı düşük notlar ve 51 – 100 aralığı yüksek notlar olarak sınıflandırılmıştır. İkili sınıflandırma işlemi, Tablo 4’ de görsel olarak sunulmuştur.

4.3. Çoklu Sınıflandırma

İlk tahmin modelimiz ikili sınıflandırma yöntemi kullanarak oluşturulacakken, ikinci tahmin modelimiz çoklu sınıflandırma yöntemi ile geliştirilecektir.

Tablo 5. Çoklu Sınıflandırma Not Dağılımı

Bütünleme Notu	Seviye
0-20	Çok düşük
21-40	Düşük
41-60	Orta
61-80	İyi
81-100	Çok iyi

Kaynak: Yazar tarafından oluşturulmuştur.

Bütünleme notları 1'den 20'ye kadar olan aralıkta bulunmaktadır ve bu notlar beş farklı kategoriye ayrılmıştır: çok iyi, iyi, orta, düşük ve çok düşük. Bu kategorizasyon, not dağılımını ve seviyelerini açıkça göstermektedir ve Tablo 5' te sunulmuştur.

4.4. Kategorik Değişkenlerin Tip Dönüşümü

Veri setinde 1 tane kategorik yani object tipinde özellik bulunmaktadır. Bu kategorik değişkenleri tahmin algoritmasına uygulamak için sayısal veri tipine dönüştürmek gerekmektedir. Bu bağlamda kategorik değişkenler sayısal veri tipine dönüştürülmüştür. Dönüştürülen tip uint8 değişken veri tipidir. Bu tip dönüşümünden sonra 5 özellik sayımız 6 özelliğe çıkmıştır bunun sebebi bir özelliğin bir alt özelliği olmasından kaynaklanan bir durumdur.

4.5. İkili Sınıflandırma Yapılarak Elde Edilen Sonuçlar

İkili sınıflandırma için kullanılan SVM, RF, KA, KNN ve NB makine öğrenimi algoritmaları, başarı oranlarına göre değerlendirildi.

Tablo 6. Algoritma Başarı Oranları ve Geçen Süre

Algoritma Adı	SVM	RF	KA	KNN	NB
Başarı Oranı	0,74	0,71	0,7	0, 68	0,66
Geçen Süre (Sn)	0.06	0.303	0.007	0.032	0.058

Kaynak: Yazar tarafından oluşturulmuştur.

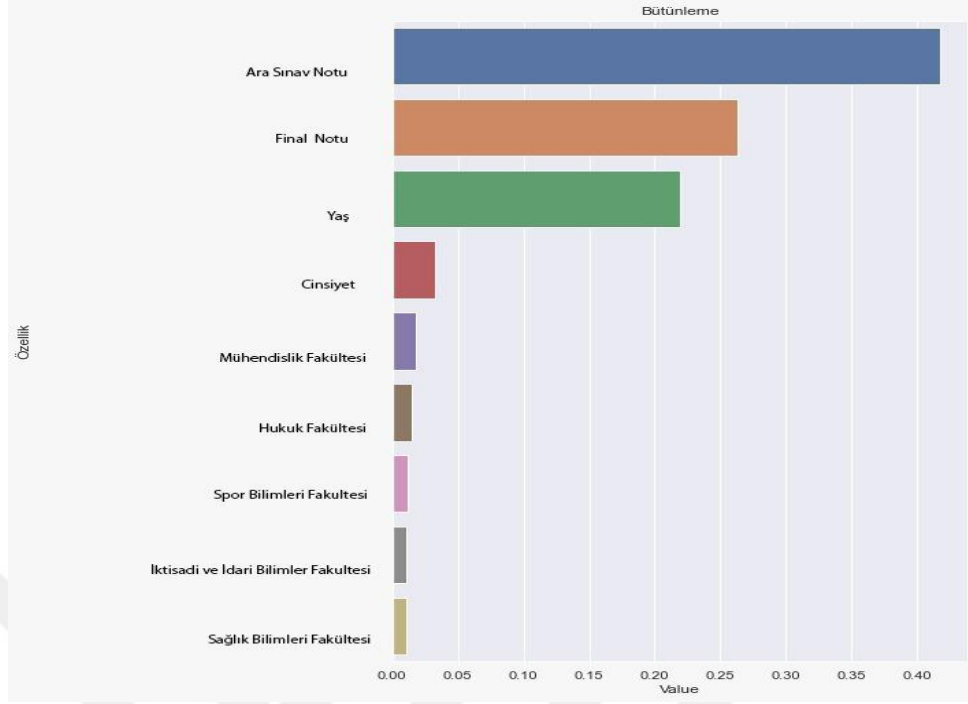
Değerlendirme sonucunda, en yüksek başarı oranı %74 ile SVM algoritması tarafından elde edildi. SVM'yi sırasıyla %71 ile RF, %70 ile KA, %68 ile KNN ve son olarak %66 ile NB algoritması takip etti. Başarı oranları ve tahmin süreleri Tablo 6'da ayrıntılı olarak sunulmuştur.

4.5.1. Random Forest Algoritması Sonuçları

Random Forest algoritmasının önemli özellik analizi ve katsayısı, algoritma sonuçlarının bulguları bulunmuştur. Doğrulamayı yapmak için çapraz doğrulama yapılmıştır. Her birinin bulgularının sonuçları ayrı ayrı başlıklar halinde gösterilmiştir.

4.5.1.1. Random Forest ile Önemli Özellikler Analizi

Bu tez çalışmasının odak noktası, öğrenci performanslarını tahmin etme başarısını değerlendirmek amacıyla kullanılan referans özelliğinin öğrencilerin bütünleme notu olduğudur. Bütünleme notu, diğer özelliklerden etkilenen bağımlı bir değişkendir.



Şekil 6. Bütünleme Notuna Etki Eden Özellik Sıralaması

Kaynak: Yazar tarafından oluşturulmuştur.

Random forest algoritması kullanılarak, bütünleme notuna en fazla etki eden özellikten en az etki eden özelliğe kadar olan ilişkiyi gösteren grafik Şekil 6' de sunulmuştur.

4.5.1.2. Random Forest En Önemli Beş Özellik ve Katsayıları

Bütünleme notu, bağımlı bir değişken olarak kabul edilir ve tüm özellikler üzerinde belirli bir etki katsayısına sahip olan bir değişkendir.

Tablo 7. Etki Eden Özellikler ve Katsayıları

Etki Sırası	Özellik Adı	Etki Katsayısı
1	Ara Sınav	0.42
2	Final Sınavı	0.26
3	Yaş	0.22
4	Cinsiyet	0.18
5	Mühendislik Fak.	0.017

Kaynak: Yazar tarafından oluşturulmuştur.

Bu etki katsayıları ve bağımsız değişkenlerin isimleri, ayrıntılı olarak Tablo 7' de sunulmuştur.

4.5.1.3. Random Forest Algoritma Sonuçları

Random Forest algoritmaları, çapraz doğrulama sonuçlarına göre 0,71 ile ikinci en yüksek başarı oranını elde ederken, aynı zamanda 0,30 ile ikinci en uzun süreyi gerektiren algoritma olarak öne çıkıyor.

Tablo 8. Random Forest Tahmin Başarı Oranı

Model	Random Forest Algoritması
Başarı Oranı	0,71
Geçen Süre	0,30

Kaynak: Yazar tarafından oluşturulmuştur.

Çapraz doğrulama sonuçları, Random Forest algoritmalarının ikili sınıflandırma görevlerinde yüksek doğruluk sağladığını göstermekle birlikte, işlem süresinin diğer algoritmalarla kıyaslandığında daha fazla zaman aldığını göstermektedir. Bu nedenle, algoritma seçerken hem başarı oranı hem de işlem süresini göz önünde bulundurmak önemlidir.

4.5.1.4. Random Forest Tahmini İçin Çapraz Doğrulama

Modelin doğruluğunu ve tahmin başarısını ölçmek amacıyla 5 farklı çapraz doğrulama işlemi uygulanmıştır.

Tablo 9. Random Forest İçin Çapraz Doğrulama

Sıra	Katsayı Değeri	Ortalama Katsayısı
1	0.69	
2	0.71	
3	0.65	0.68
4	0.66	
5	0.67	

Kaynak: Yazar tarafından oluşturulmuştur.

Çapraz doğrulama işlemlerinin sonuçlarına göre, 2. çapraz doğrulama işlemi %0,71 başarı oranıyla önemli bir başarı elde etmiştir. Çapraz doğrulama işlemlerinin ortalaması %0,68 olarak hesaplanmıştır. Ayrıca, çapraz doğrulama katsayıları ve ortalaması ayrıntılı olarak Tablo 9'da sunulmuştur.

4.5.2. Karar Ağaçları İkili Sınıflandırma

Karar Ağaçları algoritma sonuçları bulunmuş ve bu sonuçların doğruluğunun kontroölü için karar ağaçları algoritmaları çapraz doğrulama işlemi yapılmıştır.

4.5.2.1. Karar Ağaçları Algoritma Sonuçları

Karar Ağacı (KA) algoritması, çapraz doğrulama sonuçlarına göre başarı oranı açısından üçüncü sırada yer almaktadır ve %0,70 başarı oranına ulaşmıştır. Bu, Karar Ağacı' nın bu özel veri seti veya problem bağlamı için oldukça iyi bir performans sergilediğini göstermektedir. Ayrıca, belirtilen geçen süre (0,007) diğer algoritmaların işlem sürelerinden çok daha hızlı olduğunu gösterir. Bu hızlı işlem süresi, Karar Ağacı algoritmasının veri analizi veya tahminleme görevlerinde pratik ve verimli bir seçenek olduğunu göstermektedir.

Tablo 10. Karar Ağaçları Tahmin Başarı Oranı

Model	Karar Ağaçları
Başarı Oranı	0,70
Geçen Süre	0,007

Kaynak: Yazar tarafından oluşturulmuştur.

Karar Ağacı algoritması bu çalışma bağlamında iyi bir başarı oranı ve hızlı işlem süresi sunarak dikkat çekmektedir. Bu sonuçlar, Karar Ağacı'nın bu tür ikili sınıflandırma görevlerinde kullanılabilirliğini ve etkinliğini vurgulamaktadır.

4.5.2.2. Karar Ağaçları Tahmini için Çapraz Doğrulama

Modelin doğruluğunu ve tahmin başarısını ölçmek amacıyla 5 farklı çapraz doğrulama işlemi uygulanmıştır. Bu çapraz doğrulama işlemlerinin sonuçlarına göre, 1. çapraz doğrulama işleminde %0,68 oranında bir başarı elde edilmiştir.

Tablo 11. Karar Ağaçları İçin Çapraz Doğrulama Değerleri

Sıra	Katsayı Değeri	Ortalama Katsayısı
1	0.68	0.65
2	0.64	
3	0.63	
4	0.67	
5	0.65	

Kaynak: Yazar tarafından oluşturulmuştur.

Çapraz doğrulama işlemlerinin ortalaması ise %0,65 olarak hesaplanmıştır. Çapraz doğrulama katsayıları ve ortalaması, Tablo 11'de detaylı bir şekilde sunulmuştur.

4.5.3. Naive Bayes İkili Sınıflandırma

Naive Bayes algoritma sonuçları bulunmuştur. Bu sonuçların doğruluğunu kontrol edebilmek için naive bayes tahmini için çapraz doğrulama sonuçlarına ulaşılmıştır.

4.5.3.1. Naive Bayes Algoritma Sonuçları

Çapraz doğrulama sonuçlarına göre, Naive Bayes algoritmaları beş farklı algoritma içerisinde en düşük başarıya sahip olan sonuncusu olmuştur.

Tablo 12. Naive Bayes Tahmin Başarı Oranı

Model	Naive Bayes Algoritması
Başarı Oranı	0,66
Geçen Süre	0,05

Kaynak: Yazar tarafından oluşturulmuştur.

Naive Bayes tahmin başarı sonucu, bu özel algoritmanın bu belirli veri seti veya problem bağlamı için diğer alternatiflere göre daha az uygun olduğunu göstermektedir. Ancak, her algoritmanın güçlü ve zayıf yanları olabilir, bu nedenle sonuçların daha derin bir analizi gerekebilir.

4.5.3.2. Naive Bayes Tahmini için Çapraz Doğrulama

Çalışmamızda, modelin doğruluğunu ve tahmin başarısını değerlendirmek amacıyla 5 kez çapraz doğrulama yöntemi uygulanmıştır. Bu çapraz doğrulama işleminin sonuçlarına göre, 4. çapraz doğrulama aşamasında %0,70 başarı oranına ulaşılmıştır.

Tablo 13. Naive Bayes İçin Çapraz Doğrulama Değerleri

Sıra	Katsayı Değeri	Ortalama Katsayısı
1	0.67	
2	0.68	
3	0.56	0.66
4	0.70	
5	0.67	

Kaynak: Yazar tarafından oluşturulmuştur.

Karşılaştırma için, çapraz doğrulama yapılmadan elde edilen sonuç %0,66 olarak kaydedilmiştir. Ayrıca, çapraz doğrulama katsayıları ve ortalaması, Tablo 12'de ayrıntılı olarak sunulmuştur.

4.5.4. KNN İkili Sınıflandırma

KNN ikili sınıflandırma algoritma sonuçlarına ulaşılmıştır. Elde edilen sonuçların doğruluğunun kontrolü için KNN çapraz doğrulama analizi de yapılmış ve sonuçları bulunmuştur.

4.5.4.1. KNN Algoritma Sonuçları

K-En Yakın Komşu (KNN) algoritması, çapraz doğrulama sonuçlarına göre başarı oranı açısından dördüncü sırada yer almaktadır ve %0,68 başarı oranına sahiptir. Bu, KNN algoritmasının bu veri seti için uygun bir performans gösterdiğini göstermektedir. Ayrıca, belirtilen işlem süresi (0,03), diğer algoritmaların süreleri ile karşılaştırıldığında oldukça hızlı olduğunu gösterir.

Tablo 14. KNN Tahmin Başarı Oranı

Model	KNN Algoritması
Başarı Oranı	0,68
Geçen Süre	0,03

Kaynak: Yazar tarafından oluşturulmuştur.

KNN, bu çalışma bağlamında düşük işlem süresi ile dikkat çekiyor ve bu, KNN algoritmasının hızlı sonuçlara ihtiyaç duyulan durumlarda etkili bir seçenek olabileceğini göstermektedir. Ancak, başarı oranı açısından diğer bazı SVM, RF, KA algoritmaların gerisinde kalmış olabilir. Bu nedenle, algoritma seçerken hem başarı oranı hem de işlem süresi gibi faktörleri dikkate almak önemlidir.

4.5.4.2. KNN Tahmini için Çapraz Doğrulama

Çalışmamız, modelin doğruluğunu ve tahmin başarısını değerlendirmek için 5 kez çapraz doğrulama yöntemini kullanmıştır. Çapraz doğrulama sonuçlarına göre, 1. çapraz doğrulama aşamasında %0,68 başarı oranına ulaşılmıştır. Bu sonuç, modelin bu özel veri seti veya problem bağlamı için iyi bir performans sergilediğini göstermektedir. Ayrıca, çapraz doğrulama yapılmadan elde edilen sonuç %0,66 olarak kaydedilmiştir, bu da çapraz doğrulama yönteminin modelin performansını artırdığını göstermektedir.

Tablo 15. KNN İçin Çapraz Doğrulama Değerleri

Sıra	Katsayı Değeri	Ortalama Katsayısı
1	0.68	0.66
2	0.66	
3	0.63	
4	0.65	
5	0.67	

Kaynak: Yazar tarafından oluşturulmuştur.

Çapraz doğrulama katsayıları ve ortalaması, ayrıntılı olarak Tablo 15'de sunulmuştur, bu da sonuçların daha fazla analiz ve yorumlamaya açık olduğunu göstermektedir.

4.5.5. SVM İkili Sınıflandırma

SVM ikili sınıflandırma algoritma sonuçlarına ulaşılmıştır. Algoritma sonuçlarının doğruluğunun kontrolü için çapraz doğrulama yapılmış ve sonuçları bulunmuştur.

4.5.5.1. SVM Algoritma Sonuçları

SVM (Destek Vektör Makineleri) algoritması, ikili sınıflandırma tahminlemelerinde en yüksek başarı oranına sahip olan algoritma olarak öne çıkıyor; başarı oranı %0,74 olarak kaydedilmiştir. Bu, SVM'in bu özel veri seti veya problem bağlamında çok iyi bir performans sergilediğini göstermektedir.

Tablo 16. SVM Tahmin Başarı Oranı

Model	SVM Algoritması
Başarı Oranı	0,74
Geçen Süre	0,06

Kaynak: Yazar tarafından oluşturulmuştur.

Başarı oranı açısından SVM en yüksek performansa sahip olsa da, işlem süresi açısından bazı diğer algoritmaların gerisinde kalıyor. Bu nedenle, algoritma seçerken hem başarı oranı hem de işlem süresi gibi faktörleri dikkate almak önemlidir, çünkü her iki faktör de karar verme sürecinde etkili olabilir.

4.5.5.2. SVM Tahmini için Çapraz Doğrulama

Çalışmamız, modelin doğruluğunu ve tahmin başarısını değerlendirmek amacıyla 5 kez çapraz doğrulama yöntemini kullanmıştır. Çapraz doğrulama sonuçlarına göre, 2. çapraz doğrulama aşamasında %0,74 başarı oranına ulaşılmıştır. Bu sonuç, modelin bu özel veri seti veya problem bağlamı için oldukça yüksek bir performans sergilediğini göstermektedir. Ayrıca, çapraz doğrulama yapılmadan elde edilen sonuç %0,70 olarak kaydedilmiştir; bu da çapraz doğrulama yönteminin modelin performansını önemli ölçüde artırdığını göstermektedir.

Tablo 17. SVM İçin Çapraz Doğrulama Değerleri

Sıra	Katsayı Değeri	Ortalama Katsayısı
1	0.72	0.70
2	0.74	
3	0.72	
4	0.64	
5	0.70	

Kaynak: Yazar tarafından oluşturulmuştur.

Detaylı çapraz doğrulama katsayıları ve ortalamaları Tablo 17'de ayrıntılı olarak sunulmuştur, bu da sonuçların daha fazla analiz ve yorumlamaya açık olduğunu göstermektedir.

4.6. Çoklu Sınıflandırma Yapılarak Elde Edilen Sonuçlar

Çalışmanın birinci bölümünde, beş farklı algoritma kullanılarak ikili sınıflandırma işlemi gerçekleştirilmiş ve her algoritmanın başarı oranları ölçülmüş, ardından bu başarı oranları karşılaştırılmıştır. İkinci bölümde ise çoklu sınıflandırma işlemi uygulanarak aynı beş algoritmanın başarı oranları bir kez daha değerlendirilmiştir.

Tablo 18. Tahmin Başarı Oranları ve Tahmin Süreleri

Model	KNN	SVM	NB	RF	KA
Başarı Oranı	0,540	0,539	0,535	0,523	0,518
Tahmin Süre(Sn)	0.117	0.188	0.208	0.446	0,046

Kaynak: Yazar tarafından oluşturulmuştur.

Her bir algoritma için ayrı ayrı çapraz doğrulama işlemi uygulanmış ve her algoritmanın performansı bu şekilde incelenmiştir. Çoklu sınıflandırma işlemi ile elde edilen sonuçlar incelendiğinde, en yüksek tahmin başarı oranına %0,540 ile K-En Yakın Komşu (KNN) algoritması ulaşmıştır. KNN algoritmasını %0,539 başarı oranıyla SVM takip etmiş, ve %0,535 ile Naive Bayes algoritması da benzer bir başarı göstermiştir. Diğer algoritmaların başarı oranlarına göre sıralaması ise %0,523

başarı oranıyla Random Forest algoritması, ve %0,518 ile en düşük başarıya sahip olan Karar Ağacı algoritması şeklinde sıralanmıştır. Algoritmalarından elde edilen tahmin başarı oranları, Tablo 18'de ayrıntılı olarak sunulmuştur.

4.6.2. Çoklu Sınıflandırma Karar Ağaçları Algoritma Sonuçları

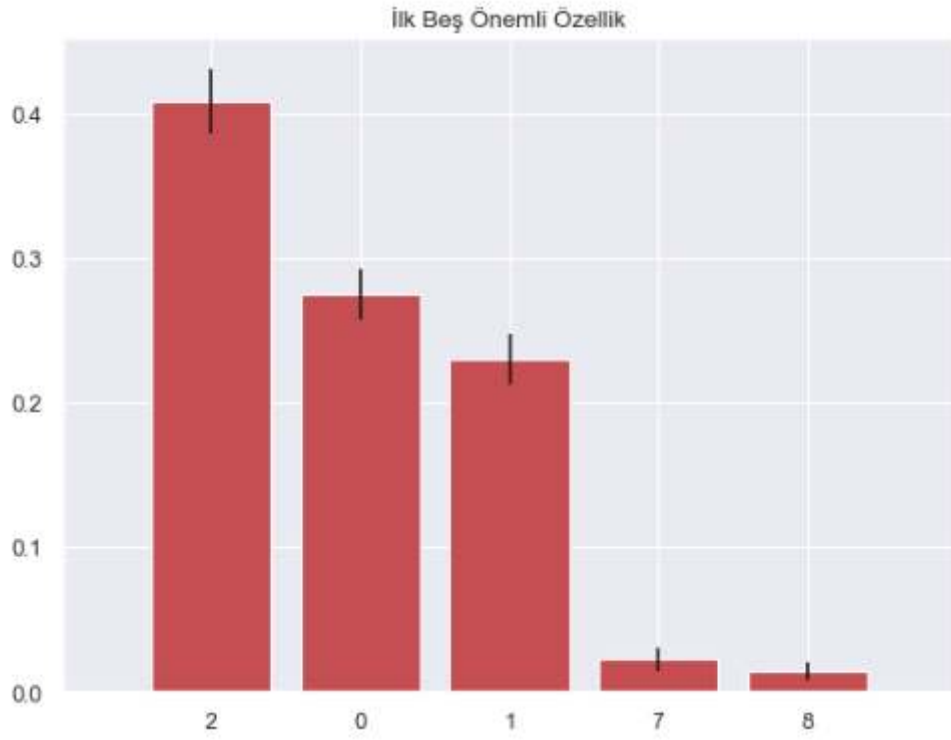
Karar ağaçları algoritması kullanılarak, veri setindeki verilere bir önem sırası atanmış ve bu verilerin önem dereceleri belirlenmiştir. İlgili özelliklerin ve onların önem katsayılarının listesi Tablo 23'te sunulmuştur.

Tablo 19. Veri Seti Önem Katsayıları

Sıra	Özellik İndeksi	Özellik Adı	Önem Katsayısı
1	1	AraSınavNotu	0.41
2	2	Final Notu	0.27
3	0	Yaş	0.23
4	8	İktisadi ve İdari Bilimler Fakültesi	0.023
5	3	BütünlemeNotu	0.014
6	4	Hukuk Fakültesi	0.013
7	6	Sağlık Bilimleri Fakültesi	0.013
8	5	Mühendislik Fakültesi	0.012
9	7	Spor Fakültesi	0.011

Kaynak: Yazar tarafından oluşturulmuştur.

Tablo 23'de görüldüğü üzere her özelliğin bir önem derecesi ve katsayısı vardır. Tahmin sonucunu etki eden ilk beş özellik ise aşağıda Şekil 8'de gösterilmiştir.



Şekil 7. İlk Beş Önemli Özellik

Kaynak: Yazar tarafından oluşturulmuştur.

Şekil 8’de görüldüğü üzere ilk beş önemli özellik sırasıyla final notu, yaş, ara sınav notu, spor bilimleri ve iktisadi idari fakültesi yer almaktadır.

Karar Ağaçları algoritmaları, çoklu sınıflandırma sonuçlarına göre başarı sırasında sonuncu sıradadır, başarı oranı yalnızca 0,518'dir. Ancak, işlem süresi açısından diğer algoritmaların önündedir, sadece 0,046 saniye süresiyle en hızlı işlem süresine sahiptir.

Tablo 20. Karar Ağaçları Çoklu Sınıflandırma Başarı Oranı

Model	KA
Başarı Oranı	0,518
Tahmini Süresi(sn)	0,046

Kaynak: Yazar tarafından oluşturulmuştur.

Karar Ağaçları algoritmalarının çoklu sınıflandırma görevlerinde düşük bir başarı gösterdiğini, ancak işlem süresi açısından avantajlı olduğunu göstermektedir. Bu nedenle, algoritma seçiminde hem başarı oranını hem de işlem süresini dikkate almak önemlidir.

4.6.2.1. Çoklu Sınıflandırma Karar Ağaçları Cross Validation Sonuçları

Modelin doğruluğunu ve tahmin başarısını ölçmek amacıyla 5 farklı çapraz doğrulama işlemi uygulanmıştır. Bu çapraz doğrulama işlemlerinin sonuçlarına göre, 1. çapraz doğrulama işlemi %0,55 başarı oranıyla önemli bir başarı elde etmiştir.

Tablo 21. KA Cross Validation Uygulanarak Elde Edilen Sonuçlar

	1	2	3	4	5	Ort
Çapraz Doğrulama	0,55	0,50	0,45	0,51	0,45	0,49

Çapraz doğrulama işlemlerinin ortalaması %0,49 olarak hesaplanmıştır. Ayrıca, çapraz doğrulama katsayıları ve ortalaması ayrıntılı olarak Tablo 24' de sunulmuştur.

4.6.2.2. Çoklu Sınıflandırma Karar Ağaçları GridSearch Sonuçları

Karar Ağaçları algoritmaları, çoklu sınıflandırma görevlerinde performanslarını değerlendirmek için kullanıldığında, sonuçlar farklı başarı düzeylerinde değişmektedir. İyi performans gösteren bir düzeyde, model pozitif sınıf örneklerini başarıyla tanıırken, kötü performans gösteren bir düzeyde model yanlış pozitif tahminler yapmaktadır. Ortalama başarı düzeyi, dengeli bir performansı yansıtırken, ölçütler arasındaki dengesizlik makro ortalama ile görülebilir. Bu sonuçlar, Karar Ağaçları algoritmalarının çoklu sınıflandırma görevlerinde performanslarının farklı düzeylerde olduğunu göstermektedir. Daha yüksek performans elde etmek için modelin iyileştirilmesi gerekebilir, bu da özellik mühendisliği veya hiperparametre ayarlamaları gibi stratejileri içerebilir.

Tablo 22. Karar Ağaçları GridSearch Sonuçları

Düzy	Duyarlılık	Kesinlik	Skor Başarısı
Çok İyi	0,29	0,24	0,26
İyi	0,68	0,69	0,68
Orta	0,49	0,3	0,38
Düşük	0,2	0,27	0,23
Çok Düşük	0,05	0,4	0,09
Ağırlıklı Ortalama	0,51	0,51	0,5
Makro Ortalama	0,34	0,38	0,33

Kaynak: Yazar tarafından oluşturulmuştur.

"Çok İyi" düzeyinde duyarlılık (recall) oldukça düşüktür (0,29), bu da modelin pozitif sınıf örneklerini tanıma yeteneğinin zayıf olduğunu gösterir. Aynı zamanda kesinlik (precision) da düşüktür (0,24), yani model pozitif olarak sınıflandırdığı örneklerin büyük bir kısmı gerçekten pozitif değildir. Bu, modelin yanlış pozitif tahminler yaptığını gösterir. Skor başarısı (F1-score) ise düşük bir değere sahiptir (0,26). "İyi" düzeyinde duyarlılık oldukça yüksektir (0,68), yani pozitif sınıf örneklerini başarılı bir şekilde tanır. Kesinlik de yüksektir (0,69), bu da modelin pozitif olarak sınıflandırdığı örneklerin çoğunun gerçekten pozitif olduğunu gösterir. Bu düzeyde skor başarısı yüksektir (0,68), bu da modelin iyi bir denge yakaladığını gösterir. "Orta" düzeyde duyarlılık (0,49) ve kesinlik (0,30) değerleri daha düşüktür, bu da modelin iyileştirilebileceğini gösterir. Skor başarısı ise orta bir değere sahiptir (0,38). "Düşük" düzeyinde hem duyarlılık (0,20) hem de kesinlik (0,27) düşüktür, bu da modelin performansının zayıf olduğunu gösterir. Skor başarısı da düşüktür (0,23). "Çok Düşük" düzeyinde duyarlılık oldukça düşüktür (0,05) ve kesinlik (0,40) daha yüksektir, ancak skor başarısı hala çok düşüktür (0,09). Bu düzeyde modelin pozitif sınıf örneklerini tanıma yeteneği çok zayıftır.

Ağırlıklı Ortalama başarı (0,50) genel olarak dengeli bir performansı gösterirken, Makro Ortalama (0,33) ölçütler arasındaki dengesizliği yansıtır.

4.6.3. Çoklu Sınıflandırma ile KNN Algoritma Sonuçları

KNN algoritması çoklu sınıflandırma sonuçlarına göre başarı sırasında birinci sıradadır, %0,540 başarı oranıyla en yüksek başarıya sahiptir ve işlem süresi 0,263

saniyedir. KNN algoritması en yüksek başarı oranına sahipken işlem süresi diğer algoritmalar arasında orta düzeydedir.

Tablo 23. KNN Çoklu Sınıflandırma Başarı Oranı

Model	KNN
Başarı Oranı	0,540
Tahmini Süresi(sn)	0,263

Kaynak: Yazar tarafından oluşturulmuştur.

Diğer algoritmaların başarı oranları ve işlem süreleri KNN'e göre sıralandığında, SVM ikinci en yüksek başarı oranına sahiptir ancak en hızlı işlem süresine sahip değildir. Bu nedenle, hangi algoritmanın kullanılacağına karar verirken hem başarı oranını hem de işlem süresini göz önünde bulundurmak önemlidir.

4.6.3.1. Çoklu Sınıflandırma KNN Cross Validation Sonuçları

Modelin doğruluğunu ve tahmin başarısını ölçmek amacıyla 5 farklı çapraz doğrulama işlemi uygulanmıştır.

Tablo 24. KNN Cross Validation Uygulanarak Elde Edilen Sonuçlar

	1	2	3	4	5	Ort
Çapraz Doğrulama	0,54	0,53	0,5	0,49	0,50	0,51

Kaynak: Yazar tarafından oluşturulmuştur.

Çapraz doğrulama işlemlerinin sonuçlarına göre, 1. çapraz doğrulama işlemi %0,54 başarı oranıyla önemli bir başarı elde etmiştir. Çapraz doğrulama işlemlerinin ortalaması %0,51 olarak hesaplanmıştır. Ayrıca, çapraz doğrulama katsayıları ve ortalaması ayrıntılı olarak Tablo 27' de sunulmuştur.

4.6.3.2. Çoklu Sınıflandırma KNN GridSearch Sonuçları

Tabloda KNN algoritmasının performansı farklı düzeylerde değerlendirilmiştir. Duyarlılık, kesinlik ve skor başarıları ölçütleri kullanılarak elde edilen sonuçlar aşağıdaki gibidir:

"Çok İyi" düzeyinde KNN, duyarlılık, kesinlik ve skor başarıları açısından düşük bir performans sergilemektedir. Bu durum, modelin doğru tahminler yapma yeteneğinin zayıf olduğunu göstermektedir.

"İyi" düzeyde KNN algoritması daha yüksek bir performans sergilemektedir. Model, bu düzeyde daha yüksek duyarlılık, kesinlik ve skor başarıları değerlerine sahiptir, bu da tahminlerin daha doğru olduğunu göstermektedir.

"Orta" düzeyde KNN, ortalama bir performans sergilemektedir.

"Düşük" düzeyde KNN algoritması, düşük bir performans göstermektedir.

"Çok Düşük" düzeyde KNN, yine düşük bir performans sergilemektedir.

Tablo 25. KNN GridSearch Sonuçları

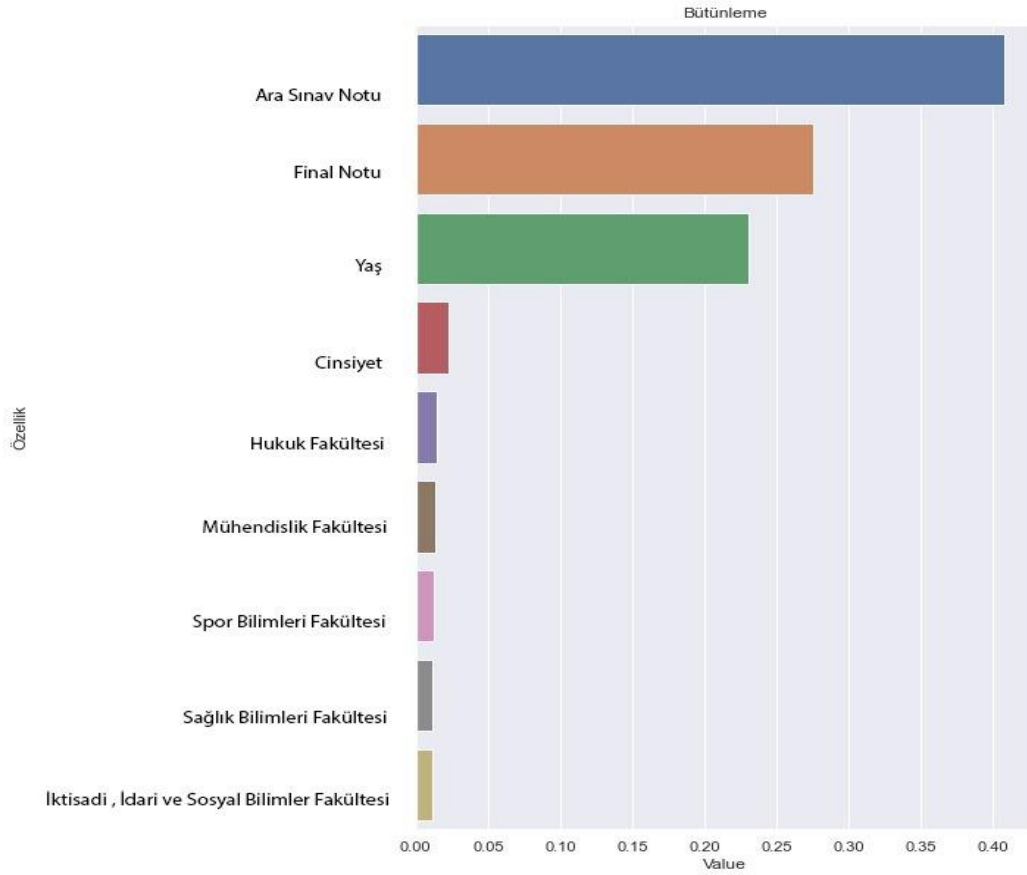
Düzye	Duyarlılık	Kesinlik	Skor Başarıları
Çok İyi	0,1	0,11	0,1
İyi	0,78	0,66	0,71
Orta	0,35	0,32	0,33
Düşük	0,18	0,29	0,22
Çok Düşük	0,1	0,36	0,16
Ağırlıklı Ortalama	0,53	0,53	0,51
Makro Ortalama	0,3	0,3	0,31

Kaynak: Yazar tarafından oluşturulmuştur.

Ağırlıklı ortalama sonuçlar, KNN algoritmasının genel başarı oranını değerlendirmek için kullanılmıştır. Ancak, makro ortalama sonuçlar, algoritmanın bazı sınıflarda daha zayıf bir performans gösterebileceğini göstermektedir. KNN algoritması farklı düzeylerde performans göstermektedir ve modelin hangi sınıf veya problem bağlamında daha uygun olduğunu belirlemek için daha fazla analiz gerekebilir.

4.6.1. Random Forest ile Önemli Özellikler Analizi

Bu tez çalışması, öğrenci performanslarının tahmin edilmesi üzerine odaklanmaktadır. Temel referans olarak alınan kriter, öğrencilerin bütünleme notlarıdır. Bütünleme notu, diğer faktörlerden etkilenen bağımlı bir değişken olarak ele alınmıştır. Random forest algoritması kullanılarak, bütünleme notunu en fazla etkileyen özellikten en az etkileyene kadar sıralama yapılmıştır ve bu sıralama Şekil 7'de sunulmuştur.



Şekil 8. Bütünleme Notuna Etki Eden Özellik Sıralaması

Kaynak: Yazar tarafından oluşturulmuştur.

4.6.1.1. Random Forest En Önemli Bes Özellik ve Katsayıları

Bütünleme notu, bağımlı bir değişken olarak tanımlanmıştır ve diğer özelliklere belirli bir oranda etki eder. Bu etkiyi yaratan bağımsız değişkenlerin isimleri ve onların bütünleme notu üzerindeki etkileme katsayıları Tablo 19'da listelenmiştir.

Tablo 26. Etki Eden Özellikler ve Katsayıları

Sıra	Özellik İndeksi	Özellik Adı	Önem Katsayısı
1	1	Ara Sınav Notu	0.41
2	2	Final Notu	0.27
3	0	Yaş	0.23
4	8	Cinsiyet	0.023
5	4	Hukuk Fakültesi	0.014

Kaynak: Yazar tarafından oluşturulmuştur.

4.6.1.2. Çoklu Sınıflandırma ile Random Forest Algoritma Sonuçları

Random Forest algoritmaları, çoklu sınıflandırma sonuçlarına göre başarı sırasında 0,523 ile dördüncü sıradadır ve tahmini süre açısından 0,446 ile beş algoritma içerisinde en yavaş işlem süresine sahiptir.

Tablo 27. Random Forest Çoklu Sınıflandırma Başarı Oranı

Model	RF
Başarı Oranı	0,523
Tahmini Süresi(sn)	0,446

Kaynak: Yazar tarafından oluşturulmuştur.

Ortaya çıkan sonuçlar, Random Forest algoritmalarının çoklu sınıflandırma görevlerinde orta düzeyde bir başarı gösterdiğini ancak işlem süresi açısından diğer algoritmaların gerisinde kaldığını göstermektedir. Algoritma seçiminde başarı oranı ve işlem süresi gibi faktörleri göz önünde bulundurmak önemlidir.

4.6.1.3. Çoklu Sınıflandırma Random Forest Çapraz Doğrulama Sonuçları

Modelin doğruluğunu ve tahmin başarısını ölçmek amacıyla 5 farklı çapraz doğrulama işlemi uygulanmıştır. Bu çapraz doğrulama işlemlerinin sonuçlarına göre, 1. çapraz doğrulama işlemi %0,53 başarı oranıyla önemli bir başarı elde etmiştir.

Tablo 28. RF Çapraz Doğrulama Uygulanarak Elde Edilen Sonuçlar

	1	2	3	4	5	Ort
Çapraz Doğrulama	0,53	0,50	0,48	0,49	0,50	0,50

Kaynak: Yazar tarafından oluşturulmuştur.

Çapraz doğrulama işlemlerinin ortalaması %0,50 olarak hesaplanmıştır. Ayrıca, çapraz doğrulama katsayıları ve ortalaması ayrıntılı olarak Tablo 20' de sunulmuştur.

4.6.1.4. Çoklu Sınıflandırma Random Forest GridSearch Sonuçları

Çok İyi seviyesinde modelin duyarlılığı (0,05) oldukça düşüktür, yani pozitif sınıfı doğru tahmin etme yeteneği zayıftır. Bununla birlikte, kesinlik (0,5) daha yüksektir, yani pozitif tahminlerin doğruluğu daha yüksektir. Bu sonuçlar, modelin negatif sınıfı daha iyi tahmin ettiğini, ancak pozitif sınıfta zayıf performans gösterdiğini gösterir. Skor Başarısı (0,09) çok düşüktür, bu da genel model performansının düşük olduğunu yansıtır. İyi seviyesinde modelin duyarlılığı (0,85) oldukça yüksektir, yani pozitif sınıfı doğru tahmin etme yeteneği iyidir. Kesinlik (0,67) daha düşüktür, yani pozitif tahminlerin doğruluğu daha düşüktür. Bu sonuçlar, modelin pozitif sınıfı iyi tahmin ettiğini ancak negatif sınıfta hatalar yaptığını gösterir. Skor Başarısı (0,75) iyi bir performansı yansıtır. Orta seviyede modelin duyarlılığı (0,28) ve kesinliği (0,32) düşüktür, yani hem pozitif sınıfı hem de negatif sınıfı doğru tahmin etme yeteneği düşüktür. Skor Başarısı (0,30) orta düzeyde bir performansı yansıtır. Düşük seviyede modelin duyarlılığı (0,30) ve kesinliği (0,40) yine düşüktür, bu da her iki sınıfı da sınırlı bir başarıyla tahmin ettiğini gösterir. Skor Başarısı (0,34) düşük bir performansı yansıtır. Çok Düşük seviyede modelin duyarlılığı (0,12) çok düşüktür, yani pozitif sınıfı tahmin etme yeteneği çok zayıftır. Bununla birlikte, kesinlik (0,88) çok yüksektir, yani pozitif tahminlerin doğruluğu çok yüksektir. Skor Başarısı (0,30) bu dengesiz performansı yansıtır. Ağırlıklı Ortalama (0,59) ve Makro Ortalama (0,33) sonuçları, bu sonuçların genel bir özetini sunar. Ağırlıklı Ortalama, farklı sınıfların veri dağılımlarını dikkate alarak ortalama performansı hesaplar. Makro Ortalama ise her sınıfın performansını eşit olarak değerlendirir.

Tablo 29. RF GridSearch Sonuçları

Düzye	Duyarlılık	Kesinlik	Skor Başarısı
Çok İyi	0,05	0,5	0,09
İyi	0,85	0,67	0,75
Orta	0,28	0,32	0,30
Düşük	0,30	0,40	0,34
Çok Düşük	0,12	0,88	0,30
Ağırlıklı Ortalama	0,59	0,57	0,55
Makro Ortalama	0,33	0,55	0,36

Kaynak: Yazar tarafından oluşturulmuştur.

Random Forest algoritması için Grid Search sonuçları, modelin performansının verilere göre değiştiğini gösterir. İyi performans gösterdiği durumlar olduğu gibi, düşük performans gösterdiği durumlar da bulunmaktadır. Modelin iyileştirilmesi gereken alanlar belirlenmelidir.

4.6.4. Çoklu Sınıflandırma ile SVM Algoritma Sonuçları

Modelin doğruluğunu ve tahmin başarısını ölçmek amacıyla 5 farklı çapraz doğrulama işlemi uygulanmıştır. Bu çapraz doğrulama sonuçlarına göre, SVM modelinin başarı oranı %0,539 olarak belirlenmiştir. Ayrıca, bu modelin tahmini işlem süresi 0,17 saniyedir. Çapraz doğrulama işlemleri sonucunda, SVM algoritmasının bu veri setinde yüksek bir başarı gösterdiği ve aynı zamanda hızlı tahminler gerçekleştirdiği görülmektedir.

Tablo 30. SVM Çoklu Sınıflandırma Başarı Oranı

Model	SVM
Başarı Oranı	0,539
Tahmini Süresi(sn)	0,17

Kaynak: Yazar tarafından oluşturulmuştur.

Diğer algoritmalarla karşılaştırıldığında, SVM modeli en yüksek başarı oranına sahip olan ikinci algoritma olarak dikkat çekmektedir. Bu sonuçlar, SVM algoritmasının bu özel veri seti için etkili bir seçenek olabileceğini göstermektedir.

4.6.4.1. Çoklu Sınıflandırma SVM Cross Validation Sonuçları

Modelin doğruluğunu ve tahmin başarısını ölçmek amacıyla 5 farklı çapraz doğrulama işlemi uygulanmıştır. Bu çapraz doğrulama işlemlerinin sonuçlarına göre, 1. çapraz doğrulama işlemi %0,53 başarı oranıyla önemli bir başarı elde etmiştir.

Tablo 31. SVM Cross Validation Uygulanarak Elde Edilen Sonuçlar

	1	2	3	4	5	Ort
Çapraz Doğrulama	0,54	0,53	0,50	0,52	0,55	0,53

Kaynak: Yazar tarafından oluşturulmuştur.

Çapraz doğrulama işlemlerinin ortalaması %0,50 olarak hesaplanmıştır. Ayrıca, çapraz doğrulama katsayıları ve ortalaması ayrıntılı olarak Çizelge 30' de sunulmuştur.

4.6.4.2. Çoklu Sınıflandırma SVN GridSearch Sonuçları

Tabloya göre, SVM (Destek Vektör Makineleri) algoritması için gerçekleştirilen Grid Search sonuçları incelendiğinde, farklı performans ölçütleri ve düzeylere göre çeşitlilik gösteriyor. Genel olarak, bu sonuçlar, modelin bazı düzeylerde iyi bir performans gösterdiğini, ancak diğerlerinde başarısız olduğunu göstermektedir. Özellikle "İyi" düzeyde, modelin duyarlılık (0,79) açısından yüksek bir başarı elde ettiği görülmektedir. Bu, modelin pozitif sınıfı doğru bir şekilde tanımlama yeteneğini vurgulamaktadır. Ancak, aynı düzeydeki kesinlik (0,68) ve başarı skoru (0,73) daha ılımlıdır, yani model bazen yanlış pozitif tahminlerde bulunabilir. Diğer yandan, "Orta" ve "Düşük" düzeylerde, modelin duyarlılık, kesinlik ve başarı skorlarında daha düşük performans gösterdiği görülmektedir. Bu, modelin bu düzeylerde bazı örnekleri yanlış sınıflandırdığını ve daha fazla

iyileştirmeye ihtiyaç duyduğunu göstermektedir. Sonuçlar ayrıca, "Çok İyi" ve "Çok Düşük" düzeylerdeki sonuçların genellikle başarısız olduğunu göstermektedir.

Tablo 32. SVM GridSearch Sonuçları

Düzye	Duyarlılık	Kesinlik	Skor Başarısı
Çok İyi	0	0	0
İyi	0,79	0,68	0,73
Orta	0,34	0,28	0,31
Düşük	0,26	0,3	0,28
Çok Düşük	0	0	0
Ağırlıklı Ortalama	0,54	0,48	0,51
Makro Ortalama	0,28	0,25	0,26

Kaynak: Yazar tarafından oluşturulmuştur.

Ağırlıklı Ortalama başarı skoru 0,51 ve Makro Ortalama başarı skoru 0,26'dır. Ağırlıklı Ortalama, farklı sınıfların dengesizliği göz önüne alırken, Makro Ortalama tüm sınıfları eşit olarak değerlendirir. Bu nedenle, SVM algoritması için Grid Search sonuçları, modelin belirli düzeylerde iyi bir performans sergilediğini, ancak genel olarak iyileştirilmesi gerektiğini göstermektedir. Modelin daha iyi bir performans elde etmesi için hiperparametre ayarları ve özellik mühendisliği teknikleri üzerinde çalışılabilir.

4.6.5. Çoklu Sınıflandırma ile Naive Bayes Algoritma Sonuçları

Naive Bayes algoritmasının başarı oranı açısından diğer algoritmalara göre rekabetçi bir performans sergilediğini gösteriyor.

Tablo 33. Naive Bayes Çoklu Sınıflandırma Başarı Oranı

Model	Naive Bayes
Başarı Oranı	0,535
Tahmini Süresi(sn)	0,24

Kaynak: Yazar tarafından oluşturulmuştur.

Başarı oranı 0.535 olarak ölçüldü ve diğer algoritmaların arasında yer alıyor. Ancak, işlem süresi 0.24 saniye olarak kaydedildi, bu da bazı diğer algoritmaların işlem sürelerinden daha hızlı olduğunu gösteriyor. Algoritma seçerken hem başarı oranını hem de işlem süresini dikkate almak önemlidir.

4.6.5.1. Çoklu Sınıflandırma Naive Bayes Cross Validation Sonuçları

Modelin doğruluğunu ve tahmin başarısını ölçmek amacıyla 5 farklı çapraz doğrulama işlemi uygulanmıştır. Bu çapraz doğrulama işlemlerinin sonuçlarına göre, 1. çapraz doğrulama işlemi %0,53 başarı oranıyla önemli bir başarı elde etmiştir.

Tablo 34. NB Cross Validation Uygulanarak Elde Edilen Sonuçlar

	1	2	3	4	5	Ort
Çapraz Doğrulama	0,53	0,49	0,39	0,56	0,51	0,50

Kaynak: Yazar tarafından oluşturulmuştur.

Çapraz doğrulama işlemlerinin ortalaması %0,50 olarak hesaplanmıştır. Ayrıca, çapraz doğrulama katsayıları ve ortalaması ayrıntılı olarak Tablo 33' de sunulmuştur.

4.6.5.2. Çoklu Sınıflandırma Naive Bayes GridSearch Sonuçları

Tablodaki sonuçlara dayalı olarak, Naive Bayes algoritması ile gerçekleştirilen Grid Search sonuçları değerlendirildi. Grid Search, farklı hiperparametre ayarları kullanarak algoritmanın performansını optimize etmeye çalışırken, farklı başarı düzeylerini ve ölçütlerini değerlendirildi. Sonuçlar aşağıdaki gibi elde edildi:

"Çok İyi" düzeyde, model yüksek bir duyarlılık (0,71) elde etti, yani gerçek pozitiflerin çoğunu doğru bir şekilde sınıflandırabiliyor. Ancak, düşük bir kesinlik (0,1) ve düşük bir başarı skoru (0,18) ile pozitif tahminlerde yanlış pozitiflerin yüksek olduğunu gösteriyor.

"İyi" düzeyde, model daha dengeli bir performans sergiliyor. Duyarlılık (0,45) ve kesinlik (0,71) daha dengeli bir şekilde dağılmıştır ve başarı skoru (0,55) bu düzeyde daha yüksektir.

Diğer düzeylerde ("Orta", "Düşük" ve "Çok Düşük"), modelin performansı daha düşüktür. Özellikle "Düşük" düzeydeki başarı skoru (0,2) oldukça düşüktür.

Ağırlıklı Ortalama başarı skoru 0,43 ve Makro Ortalama başarı skoru 0,31'dir. Ağırlıklı Ortalama, farklı sınıfların dengesizliğini göz önüne alırken, Makro Ortalama tüm sınıfları eşit olarak değerlendirir.

Tablo 35. Naive Bayes GridSearch Sonuçları

Düzyey	Duyarlılık	Kesinlik	Skor Başarısı
Çok İyi	0,71	0,1	0,18
İyi	0,45	0,71	0,55
Orta	0,43	0,3	0,36
Düşük	0,15	0,28	0,2
Çok Düşük	0,28	0,28	0,28
Ağırlıklı Ortalama	39	0,52	0,43
Makro Ortalama	0,41	0,34	0,31

Kaynak: Yazar tarafından oluşturulmuştur.

Bu sonuçlar, Naive Bayes algoritmasının farklı başarı düzeylerinde farklı performanslar gösterdiğini ve iyileştirme alanları olduğunu göstermektedir. Modelin daha iyi bir performans elde etmesi için hiperparametre ayarları ve özellik mühendisliği teknikleri üzerinde çalışılabilir.

SONUÇ

Bu tez çalışması, eğitim ve öğretim sürecine etki eden faktörlerin günümüz teknolojileri kullanılarak modern değerlendirme yöntemlerine tabi tutulup analiz edilmesini amaçlamıştır. Bu analiz, öğretmenlerden öğrencilere ve diğer paydaşlara daha ayrıntılı ve nitelikli bilgiler sunmayı hedeflemiştir. Ayrıca, eğitsel veri madenciliği ve makine öğrenme teknikleri kullanılarak öğrencilerin akademik başarılarını etkileyen faktörleri incelemiştir. CRISP-DM yöntemi kullanılarak yapılan bu çalışma, beş farklı makine öğrenmesi algoritması olan RF, KA, KNN, SVG ve NB ile analiz ve tahminler gerçekleştirmiştir.

Bu analizlerde, öğrencilerin bütünleme notlarını yüksek ve düşük notlar olarak ikili ve çoklu sınıflandırmada kullanılmıştır. Veri setinde bulunan kategorik nitelikler sayısal özelliklere dönüştürülmüş ve eğitim verisi ile test verisi olarak ayrılmıştır. Her bir algoritma için bağımsız değişken olarak bütünleme notu kullanılarak ayrı ayrı modeller oluşturulmuştur. Bu çalışma eğitim-öğretim sürecindeki faktörleri modern teknolojilerle analiz ederek paydaşlara daha fazla bilgi sunma amacı taşımaktadır. Ayrıca, beş farklı sınıflandırma algoritmasının performansı da ayrıntılı bir şekilde incelenmiş ve bu çalışmanın sonucunda elde edilen bulgular eğitim ve öğretim süreçlerinin daha iyi anlaşılmasına ve yönetilmesine katkı sağlayabilir. Çalışmada öncelikle, veri setindeki kategorik değişkenlerin sayısal forma dönüştürülmesi gerekti. Dönüşüm sonucunda, özellik sayısı artmıştır, çünkü bazı özellikler alt özelliklere ayrılmıştır.

Çalışmanın ikili sınıflandırma görevi için kullanılan beş farklı algoritmanın performans sonuçları incelendiğinde Karar ağaçları, çeşitli düzeylerde performans sergilemektedir. Özellikle "İyi" ve "Orta" düzeylerde kabul edilebilir sonuçlar elde edilirken, "Çok İyi" ve "Çok Düşük" düzeylerinde başarısızlık yaşanmıştır. Karar ağaçlarının belirli problem bağlamlarında etkili olabileceği, ancak bazı problemlerde daha fazla iyileştirmeye ihtiyaç duyabileceği göz önüne alınmalıdır.

Bu çalışma, öğrenci başarılarını tahmin etmek için farklı ikili sınıflandırma algoritmalarını değerlendirmek amacıyla gerçekleştirilmiştir. Öncelikle, veri setindeki kategorik değişkenlerin sayısal forma dönüştürülmesi gerekti. Dönüşüm sonucunda, özellik sayısı artmıştır, çünkü bazı özellikler alt özelliklere ayrılmıştır.

İkili sınıflandırma sonuçlarına göre, SVM, RF, KA, KNN ve NB makine öğrenimi algoritmaları başarı oranlarına göre değerlendirilmiştir. Bu değerlendirme sonucunda, en yüksek başarı oranı %74 ile SVM algoritması tarafından elde edilmiştir. SVM'yi sırasıyla %71 ile RF, %70 ile KA, %68 ile KNN ve son olarak %66 ile NB algoritması takip etmiştir. Random Forest algoritması özelinde, bütünleme notuna en fazla etki eden özellikten en az etki eden özelliğe kadar olan ilişkiyi gösteren bir grafik oluşturulmuştur. Buna göre, bütünleme notunu en fazla etkileyen özellikler sırasıyla ara sınav, final sınavı, yaş, cinsiyet ve mühendislik fakültesi olarak sıralanmıştır. K-En Yakın Komşu (KNN) algoritması, başarı oranlarına göre dördüncü sırada yer almıştır ve %0,68 başarı oranına sahiptir. Ayrıca, KNN'nin hızlı işlem süresi ile dikkat çektiği gözlemlenmiştir. Karar Ağacı (KA) algoritması, çapraz doğrulama sonuçlarına göre %0,70 başarı oranı ile üçüncü sırada yer almıştır. Ayrıca, belirtilen geçen süresi (0,007) diğer algoritmaların işlem sürelerinden oldukça hızlı olduğunu göstermektedir. Bu sonuçlar, Karar Ağacı algoritmasının bu tür ikili sınıflandırma görevlerinde başarılı ve veri analizi veya tahminleme görevlerinde pratik bir seçenek olduğunu vurgulamaktadır. Son olarak, Naive Bayes algoritması çapraz doğrulama sonuçlarına göre en düşük başarı oranına sahip olan algoritma olmuştur. Bu sonuç, bu özel algoritmanın bu belirli veri seti veya problem bağlamı için diğer alternatiflere göre daha az uygun olduğunu göstermektedir.

Çalışmanın çoklu sınıflandırma görevi için kullanılan beş farklı algoritmanın performans sonuçları incelendiğinde, K-En Yakın Komşu (KNN) algoritmasının %0,540 başarı oranıyla en yüksek performansa sahip olduğu görülmüştür. Bu, bu özel veri kümesi ve problem bağlamı için KNN'nin en iyi performansı gösterdiğini göstermektedir.

Sırasıyla ikinci ve üçüncü sırada, Destek Vektör Makineleri (SVM) ve Naive Bayes algoritmaları gelmektedir. SVM %0,539 başarı oranıyla KNN'ye oldukça yakın bir performans sergilemiştir, bu da SVM'in bu veri kümesinde etkili bir sınıflandırma yapabildiğini göstermektedir. Naive Bayes ise %0,535 başarı oranıyla dikkate değer bir sonuç elde etmiştir. Random Forest ve Karar Ağaçları algoritmaları ise daha düşük başarı oranlarına sahiptir (%0,523 ve %0,518). Bu sonuçlar, bu iki algoritmanın bu özel veri kümesi için diğer üç algoritma kadar etkili olmadığını

göstermektedir. Ayrıca, işlem süreleri incelendiğinde, KNN'nin en hızlı işlem süresine sahip olduğu görülmüştür. Bu, KNN'nin hem iyi performans gösterdiği hem de hızlı çalıştığı anlamına gelir.

Bu çalışma, öğrenci başarılarını tahmin etmek için farklı ikili ve çoklu sınıflandırma algoritmalarının performanslarını inceledi. Ancak, her algoritmanın güçlü ve zayıf yönleri olabilir ve sonuçlar daha derin bir analize ihtiyaç duyabilir. Algoritma seçimi yapılırken hem başarı oranı hem de işlem süresi gibi faktörlerin göz önünde bulundurulması önemlidir. Bu çalışma, öğrenci performanslarının tahmin edilmesi amacıyla farklı ikili sınıflandırma algoritmalarının kullanılabileceğini göstermiştir ve bu algoritmaların avantajlarını ve sınırlamalarını açığa çıkarmıştır. Gelecekteki çalışmalar, bu algoritmaların daha fazla iyileştirilmesi ve belirli veri setlerine uyarlanması açısından potansiyel sunmaktadır. Çoklu sınıflandırma için görevinde ise en iyi algoritma olarak K-En Yakın Komşu (KNN) önerilmektedir. Ancak, bu sonuçlar sadece bu özel veri kümesi ve problem bağlamı için geçerlidir. Farklı veri kümeleri ve problemler için farklı algoritmaların değerlendirilmesi gerekebilir.

Bu tez, ikili ve çoklu sınıflandırma algoritmalarının performansını inceledi ve bazı önemli bulguları ortaya koydu. Elde edilen sonuçlar, gelecekteki çalışmalara ve uygulamalara rehberlik edebilecek bazı önerilere yol açmaktadır:

Çalışmanın sonuçlarına dayalı olarak, gelecekteki araştırmalar için bir dizi öneri sunulmaktadır. İlk olarak, farklı veri setleri ve problem bağlamlarında, kullanılan algoritmaların genel geçerliliğini değerlendiren daha geniş kapsamlı çalışmaların yapılması gerekmektedir. Ayrıca, daha büyük veri setleri ile çalışarak algoritmaların daha kesin ve güvenilir tahminlerde bulunup bulunamayacağını incelemek önemlidir. Özellikle, öğrenci başarılarını tahmin etmek için daha fazla veri toplama ve temizleme işlemi yapılabilir, bu da model performansını artırabilir. Algoritmik performansın optimizasyonu ve üstün sonuçların elde edilmesi amacıyla, etkili özellik mühendisliği tekniklerinin derinlemesine incelenmesi gerekmektedir. Algoritma kombinasyonları üzerinde çalışmak, özellikle karmaşık veri setleri için daha iyi sonuçlar elde etmede yardımcı olabilir. Ayrıca, daha fazla veri analizi yapmak ve farklı performans metriklerini dikkate almak, algoritmaların güçlü ve zayıf yönlerini daha iyi anlamamıza yardımcı olabilir. Son olarak, işlem süreleri ve hesaplama maliyetleri de göz önünde bulundurularak, pratikte uygulanabilir

çözümler üzerine odaklanmak önemlidir. Bu öneriler, öğrenci başarılarını tahmin etmek için daha etkili yöntemler geliştirme yolunda gelecekteki araştırmalara rehberlik edebilir.

Bu çalışma, araştırma verilerinin kaynağının Moodle öğrenim yönetim sistemi olduğunu belirtmek önemlidir. Veriler, öğrenci başarılarının tahmin edilmesi amacıyla Moodle üzerinden toplanmıştır. Moodle, öğrencilerin kurslarla etkileşimlerini ve öğrenme süreçlerini takip etmek için yaygın olarak kullanılan bir çevrimiçi öğrenim platformudur. Bu kaynak, araştırma sürecinin şeffaflığını artırmak ve verilere erişim hakkında okuyucuları bilgilendirmek amacıyla bu çalışmada önemli bir rol oynamıştır.

Bu çalışma, öğrenci başarılarını tahmin etmek için farklı ikili sınıflandırma ve çoklu sınıflandırma algoritmalarını karşılaştırmış ve analiz etmiştir. İkili sınıflandırma sonuçlarına göre SVM algoritması %74 başarı oranı ile en yüksek performansı gösterirken, çoklu sınıflandırma görevinde ise K-En Yakın Komşu (KNN) %54 başarı oranı ile öne çıkmıştır. Bu sonuçlar, algoritmaların görev bağlamına ve veri kümesine göre farklı performanslar sergileyebileceğini vurgulamaktadır. Ayrıca, bu çalışma sonucunda özellik mühendisliği stratejilerinin önemi ortaya çıkmıştır. Özellikle, bütünleme notunu en fazla etkileyen özelliklerin tespit edilmesi, öğrenci başarılarını tahmin etmede daha iyi bir anlayış sağlayabilir. Bununla birlikte, daha fazla veri toplama ve temizleme süreçleri, model performansını artırabilir ve algoritmaların daha güvenilir tahminlerde bulunmasına yardımcı olabilir. Çalışmanın öneriler bölümünde belirtildiği gibi, gelecekteki araştırmalar, farklı veri setleri ve problem bağlamları için daha geniş kapsamlı testler içermelidir. Ayrıca, daha büyük veri kümeleri üzerinde çalışmak ve algoritmaların performansını iyileştirmek için özellik mühendisliği stratejilerini geliştirmek gereklidir. İşlem süreleri ve hesaplama maliyetleri de göz önünde bulundurularak, gerçek dünya uygulamaları için daha pratik çözümler üzerinde çalışılmalıdır.

Bu tez çalışması, ikili ve çoklu sınıflandırma algoritmalarının performansını inceledi ve bu algoritmaların farklı düzeylerdeki başarılarını ortaya koydu. Elde edilen sonuçlar, gelecekteki araştırmalar ve uygulamalar için bazı önemli önerilere yol açmaktadır. İlk olarak, veri kümesinin daha iyi anlaşılması ve özellik mühendisliği tekniklerinin kullanılması, algoritmaların performansını artırmak için önemlidir. Özellik seçimi yöntemleri, veri kümesindeki önemli özellikleri

tanımlamak ve gereksiz olanları elemek için kullanılabilir. Ayrıca, algoritmaların hiperparametrelerinin dikkatle ayarlanması gerekmektedir. Grid arama veya rastgele arama gibi hiperparametre ayarlama stratejileri, algoritmaların daha iyi sonuçlar vermesine yardımcı olabilir. Sınıf dengesizliği, çoklu sınıflandırma görevlerinde sıkça karşılaşılan bir sorundur. Bu zorluğu ele almak için özel sınıf dengesizliği azaltma veya artırma teknikleri kullanılabilir. Özellikle nadir sınıfların daha iyi tanımlanması için çözümler araştırılmalıdır. Ayrıca, bu çalışma belirli bir veri kümesi üzerinde gerçekleştirilmiştir ve gelecekte farklı veri kümeleri ve uygulama alanlarında benzer bir analizin yapılması, algoritmaların genel uygulanabilirliğini daha iyi anlamamıza yardımcı olabilir. Bununla birlikte, makine öğrenimi alanı sürekli olarak yeni algoritmalar geliştirmektedir. Bu nedenle, bu tezde incelenmeyen yeni nesil sınıflandırma algoritmalarının incelenmesi ve performanslarının değerlendirilmesi ilginç bir araştırma konusu olabilir. Son olarak, algoritmaların performansı, kullanılan eğitim verisinin miktarına bağlı olarak değişebilir. Gelecekteki çalışmalarda, farklı eğitim verisi miktarlarının algoritmaların performansına etkisi daha ayrıntılı bir şekilde incelenebilir. Bu öneriler, makine öğrenimi ve veri madenciliği alanındaki gelecekteki çalışmalar için yol gösterici olabilir ve algoritmaların daha etkili hale getirilmesine katkıda bulunabilir.

KAYNAKÇA

- Adipat, S., Laksana, K., Busayanon, K., Piatanom, P., Mahamarn, Y., Pakapol, P. ve Ausawasowan, A. (2022). The world of technology: Artificial intelligence in education. *Specialusis Ugdyimas*, 2(43), 2142-2146.
- Akgun, S. ve Greenhow, C. (2021). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2(3), 431-440.
- Akinwalere, S. N. ve Ivanov, V. (2022). Artificial intelligence in higher education: Challenges and opportunities. *Border Crossing*, 12(1), 1-15.
- Akman, M. (2010). *Veri madenciliğine genel bakış ve random forests yönteminin incelenmesi: Sağlık alanında bir uygulama* [Yüksek lisans tezi]. Ankara Üniversitesi.
- Akpınar, B. (2019). *Görüntü sınıflandırma için derin öğrenme ile bayesçi derin öğrenme yöntemlerinin karşılaştırılması* [Yüksek lisans tezi]. Afyon Kocatepe Üniversitesi.
- Akşehir, Z. D. ve Kılıç, E. (2019). Makine öğrenmesi teknikleri ile banka hisse senetlerinin fiyat tahmini. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 12(2), 30-39.
- Al Daoud, E. (2019). Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *International Journal of Computer and Information Engineering*, 13(1), 6-10.
- Al Mayahi, K. ve Al-Bahri, M. (2020, 14 October). *Machine learning based predicting student academic success* [Bildiri sunumu]. 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops. Brno.
- Alpar, R., (2011). *Uygulamalı çok değişkenli istatistiksel yöntemler*. Detay Yayıncılık.
- Altun, D. (2019). *Sanal gerçeklik ve yapay zeka*. Doğu Kitapevi.

- Amrieh, E. A., Hamtini, T. ve Aljarah, I. (2016). Mining educational data to predict student's academic performance using ensemble methods. *International Journal of Database Theory and Application*, 9(8), 119-136.
- Arslan, K. (2020). Eğitimde yapay zekâ ve uygulamaları. *Batı Anadolu Eğitim Bilimleri Dergisi*, 11(1), 71-88.
- Arslan, T. (2009). *Yabancı dil olarak Almanca öğretiminde web destekli öğrenme modeli moodle'in kullanımı ve öğrenme sürecine etkisi - yazma becerisi bağlamında görgül bir çalışma* [Doktora tezi]. Mersin Üniversitesi.
- Ataseven, S. (2008). *Üniversitelerin adaylar tarafından tercih edilme desenlerini veri madenciliği yöntemleri ile belirleyen bir model önerisi* [Doktora tezi]. İstanbul Kültür Üniversitesi.
- Atasoy, D. (2001). *Lojistik regresyon analizinin incelenmesi ve bir uygulaması*. [Yüksek lisans tezi]. Cumhuriyet Üniversitesi.
- Aydın, A. (2019). *Yapay Zekâ ve Gelecek*. Doğu Kitapevi.
- Aydın, C. ve Biroğul, S. (2008). E-öğrenmede açık kaynak kodlu öğretim yönetim sistemleri ve Moodle. *Bilişim Teknolojileri Dergisi*, 1(2).
- Ayık, Y. Z., Özdemir, A. ve Yavuz, U. (2010). Lise türü ve lise mezuniyet başarısının, kazanılan fakülte ile ilişkisinin veri madenciliği tekniği ile analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 441-454.
- Bahçeci, F. ve Gürol, M. (2010). Eğitimde akıllı öğretim sistemleri uygulamalarına yönelik bir model önerisi. *Engineering Sciences*, 5(2), 121-128.
- Balaban, M. E. ve Kartal, E. (2015). *Veri madenciliği ve makine öğrenmesi temel algoritmaları ve r dili ile uygulamaları*. Çağlayan Kitabevi.
- Bastem, H. N. (2021). *Student academic performance prediction via artificial intelligence using machine learning algorithms* [Yüksek lisans tezi]. Çankaya Üniversitesi.
- Bell, E., Bryman, A. ve Harley, B. (2022). *Business research methods*. Oxford University Press.

- Bhat, I. A., Tomar, K. ve Saini, S. (2021). Novel design of power generation with windmill using national grid. *International Journal of Innovative Research in Engineering ve Management (IJIREM)*, 8(6), 66. <https://doi.org/10.55524/ijirem.2021.8.6.12>.
- Bingöl, F. (2023). *Yapay zekâ destekli makine öğrenmesi algoritmaları ile eğitsel veri madenciliği alanında öğrencilerin akademik performanslarının tahmin edilmesi* [Yüksek lisans tezi]. İstanbul Gedik Üniversitesi.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4-16.
- BODEA, M. G. I. (2020). *Makine öğrenmesi teknikleri ile sosyal medya kullanımı üzerine bir duygu analizi çalışması* [Doktora tezi]. Kastamonu Üniversitesi.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40.
- Bozkurt, A. (2017). Türkiye’de uzaktan eğitimin dünü, bugünü ve yarını. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 3(2), 85-124.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123-140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- Brinkmann, S. ve Kvale, S. (2015). *Learning the craft of qualitative research interviewing*. Sage Publications.
- Busuttil, S. (2003). Support vector machines. University of Malta.
- Campbell, M., Hoane Jr, A. J. ve Hsu, F. H. (2002). Deep blue. *Artificial intelligence*, 134(1-2), 57-83.
- Canbay, Y. (2013). *Diyabet verilerinin destek vektör makineleri kullanılarak sınıflandırılması* [Yüksek lisans tezi]. Erciyes Üniversitesi.
- Carmona, P., Climent, F. ve Momparler, A. (2019). Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics ve Finance*, 61, 304-323.
- Charniak, E., McDermott, D. ve McDermott, D. V. (1985). *Introduction to artifical intelligence*. Addison-Wesley Company.

- Chassignol, M., Khoroshavin, A., Klimova, A. ve Bilyatdinova, A. (2018). Artificial intelligence trends in education: a narrative overview. *Procedia Computer Science*, 136, 16-24.
- Chen, D. Y. (2017). *Pandas for everyone: Python data analysis*. Addison-Wesley Professional.
- Chen, T. ve Guestrin, C. (2016, 16 August). *Xgboost: A scalable tree boosting system* [Bildiri Sunumu]. In Proceedings Of The 22nd Acm Sigkdd International Conference On Knowledge Discovery And Data Mining, San Francisco.
- Chollet, F., (2019). *Python ile derin öğrenme*. Buzdağı Yayınevi.
- Copeland, J. (1993). *Artificial intelligence: A philosophical introduction*. John Wiley and Sons.
- Cortes, C. ve Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>
- Coşkun, F. ve Gülleroğlu, H. D. (2021). Yapay zekanın tarih içindeki gelişimi ve eğitimde kullanılması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 54(3), 947-966. <https://doi.org/10.30964/auebfd.916220>
- Çelik, M. (2009). *Veri madenciliğinde kullanılan sınıflandırma yöntemleri ve bir uygulama*. [Yüksek lisans tezi]. İstanbul Üniversitesi.
- Çelik, Ö. ve Altunaydın, S. S. (2018). Makine öğrenimi yöntemleri ve uygulamaları üzerine bir araştırma. *Journal of Educational Technology ve Online Learning*, 1(3), 25-40.
- Çevik, A. (2008, May). Moodle öğrenme yönetim sistemi yönetimindeki karşılaşılabilecek olası sorunlar ve çözüm önerileri. In *8th International Educational Technology Conference* (pp. 31-34).
- Çolakoğlu, A. (2020). *Makine öğrenmesi algoritmaları ile avrupa havalimanları analizi* [Doktora tezi]. Pamukkale Üniversitesi
- Çomak, E. (2008). *Destek vektör makinelerinin etkin eğitimi için yeni yaklaşımlar* [Doktora tezi]. Selçuk Üniversitesi.

- Dagdeviren, E. (2013). *El yazisi rakam tanıma için destek vektor makinelerinin ve yapay sinir ağlarının karşılaştırması* [Yüksek lisans tezi]. İstanbul Üniversitesi.
- Dalkılıç, N. ve Kurt, A. A. (2019). Öğrenci başarısının yapay sinir ağları ile tahmin edilmesi. *Bilişim Teknolojileri Dergisi*, 12(3), 107-116.
- Demir, E. (2014). Uzaktan eğitime genel bir bakış. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 39, 203.
- Deveci, M. (2019). *Türkiye’de yeni iletişim teknolojilerinin yükseköğretim kurumlarında uzaktan eğitim alanında kullanımı* [Doktora tezi]. Marmara Üniversitesi.
- Dikmen, S. ve Bahçeci, F. (2020). Covid-19 pandemisi sürecinde yükseköğretim kurumlarının uzaktan eğitime yönelik stratejileri: Fırat Üniversitesi örneği. *Turkish Journal of Educational Studies*, 7(2), 78-98.
- Dougiamas, M. ve Taylor, P. C. (2003). Moodle: Using learning communities to create an open source course management system. *Association for the Advancement of Computing in Education (AACE)*, 171-178.
- Driscoll, M. (2002). Blended learning: Let’s get beyond the hype. *E-learning*, 1(4), 1-4.
- Duda, R. O. and Hart, P. E. (1973). *Pattern classification and scene analysis*. John Wiley & Sons.
- El Naqa, I. ve Murphy, M. J. (2015). *What is machine learning?* Springer International Publishing. https://doi.org/10.1007/978-3-319-18305-3_1
- Erçetin, Ö. Z. ve Baykoç, Ö. (2004). Tedarikçi seçimi problemine karar teorisi destekli uzman sistem yaklaşımı. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 19(3).
- Ersöz, F. ve Çınar, Y. (2021). Veri madenciliği ve makine öğrenimi yaklaşımlarının karşılaştırılması: Tekstil sektöründe bir uygulama. *Avrupa Bilim ve Teknoloji Dergisi*, (29), 397-414.
- Feng, Y., Cui, N., Gong, D., Zhang, Q. ve Zhao, L. (2017). Evaluation of random forests and generalized regression neural networks for daily reference

- evapotranspiration modelling. *Agricultural Water Management*, 193, 163-173.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Friedman, N., Geiger, D. ve Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29, 131-163.
- Günay, D. ve Şişman, B. (2019). Bilgi ve eğitim teknolojileri okuryazarlığı. AD Öğretim-Özçelik ve MN Tuğluk. *Eğitimde Ve Endüstride 21. Yüzyıl Becerileri*, 257-275.
- Gürçan, M. (1998). *Lojistik regresyon analizi ve bir uygulama*. [Yüksek lisans tezi]. Ondokuz Mayıs Üniversitesi.
- Hamel, L. H. (2011). *Knowledge discovery with support vector machines*. John Wiley and Sons.
- Han, J., Pei, J., and Tong, H. (2022). *Data mining: concepts and techniques*. Morgan Kaufmann.
- Hastie, T., Tibshirani, R., Friedman, J. H. ve Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer.
- Haykin, S. (2009). *Neural networks and learning machines*, 3/E. Pearson Education.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 20(8), 832-844.
- Holmes, W., Bialik, M. ve Fadel, C. (2023). *How digital technologies can serve humanity*. Globethics Publications.
- Hosmer Jr, D. W., Lemeshow, S. ve Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley and Sons.
- Hu, L. Y., Huang, M. W., Ke, S. W. ve Tsai, C. F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus*, 5(1), 1-9.

- Irmak, H. (2023). *Yapay zeka yöntemleri ile uzaktan eğitimdeki sorunların tespiti ve öğrencilerin akademik performanslarının tahmin edilmesi* [Doktora Tezi]. İstanbul Üniversitesi.
- İşler, B. ve Kılıç, M. (2021). Eğitimde yapay zekâ kullanımı ve gelişimi. *Yeni Medya Elektronik Dergisi*, 5(1), 1-11.
- İşman, A., (2011). *Uzaktan eğitim*. Pegem Yayınevi.
- İzmirli, S., Özdil, A., Yaşar, C. ve Hacıömeroğlu, G. (2018). E-öğrenme materyali geliştirilmesinde kullanılan yazarlık araçlarının karşılaştırılması. *International Journal of Computers in Education*, 1(1), 14-23.
- James, G., Witten, D., Hastie, T. ve Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.
- James, J. T. (2013). A new, evidence-based estimate of patient harms associated with hospital care. *Journal of patient safety*, 9(3), 122-128.
- Kadriye, Uzun ve Aydın, C. H. (2012). The use of virtual ethnography in distance education research. *Turkish Online Journal of Distance Education*, 13(2), 212-225.
- Kaelbling, L. P., Littman, M. L. ve Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4, 237-285.
- Karaca, B. ve Telli, G. (2019). Yapay zekânın çeşitli süreçlerdeki rolü ve tahminleme fonksiyonu. *Yapay Zeka ve Gelecek*, 172, 185.
- Karadayı, M. A. ve Küçük, Z. (2019). Öğrenci başarısının karar ağacı algoritması ile tahmin edilmesi: Bir uygulama. *Kastamonu Eğitim Dergisi*, 27(4), 1595-1604.
- Karakuş, B. A. (2018). *Derin öğrenme ve büyük veri yaklaşımları ile metin analizi* [Doktora tezi]. Fırat Üniversitesi
- Kartal, E. (2015). *Sınıflandırmaya dayalı makine öğrenmesi teknikleri ve kardiyolojik risk değerlendirmesine ilişkin bir uygulama* [Doktora tezi]. İstanbul Üniversitesi.
- Kaya, Z. (2002). *Uzaktan eğitim*. Pegem Yayınları.

- Kaya, Z., Erden, O., Çakır, H. ve Bağrsakçı, B. (2004). Uzaktan eğitimin temelleri dersindeki uzaktan eğitim ihtiyacı ünitesinin web tabanlı sunumunun hazırlanması. *Turkish Online Journal Of Educational Technology*, 3(3).
- Kecman, V. (2001). *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. MIT press.
- Kengam, J. (2020). Artificial intelligence in education. *Research Gate*, 18, 1-4. <https://doi.org/10.13140/RG.2.2.16375.65445> dergi
- Keş, T. (2023). *Makine öğrenmesi yöntemleri ile diyabet hastalığı tahminleme* [Yüksek lisans tezi]. Beykent Üniversitesi.
- Kır, Ş. ve Şenocak, D. (2022). Açık ve uzaktan öğrenme sistemlerinde yapay zekânın öğrenen destek hizmeti bağlamında kullanımı. *Dijital Teknolojiler ve Eğitim Dergisi*, 1(1), 39-56. <https://doi.org/10.5281/zenodo.6647642>
- Kilinc, D., Borandag, E., Yucalar, F., Tunali, V., Simsek, M. ve Ozcift, A. (2016). Classification of scientific articles using text mining with KNN algorithm and R language. *Marmara Journal of Pure and Applied Sciences*, 3, 89-94.
- Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artificial Intelligence Review*, 37, 331-344.
- Kotu, V. ve Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann.
- Koyuncu, İ. ve Gelbal, S. (2020). Comparison of data mining classification algorithms on educational data under different conditions. *Journal of Measurement and Evaluation in Education and Psychology*, 11(4), 325-345. <https://doi.org/10.21031/epod.696664>.
- Kör, H. (2017). *Bulut tabanlı çevrimiçi öğrenme ortamında etkinlik öneri sistemi tasarımı: eğitimsel veri madenciliği uygulaması* [Doktora Tezi]. Kırıkkale Üniversitesi.
- Kurbanoğlu, S. (1992). Uzman sistemler. *Türk Kütüphaneciliği*, 6(4), 189-193.
- Langley, P. ve Sage, S. (1994). *Induction of selective bayesian classifiers*. Morgan Kaufmann.

- Langley, P., Iba, W. ve Thompson, K. (1992, July). An analysis of bayesian classifiers. In *Aaai*, s. 223-228.
- LeCun, Y., Bengio, Y. ve Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*, 9(1), 381-386.
- Márquez-Vera, C., Cano, A., Romero, C. ve Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied intelligence*, 38, 315-330.
- McCarthy, J. (2007). From here to human-level AI. *Artificial Intelligence*, 171(18), 1174-1182.
- McTear, M. (2022). *Conversational ai: dialogue systems, conversational agents, and chatbots*. Springer Nature.
- Mısırlı, Z. A. (2007). *Web tabanlı öğrenme yönetim sistemine ilişkin öğrenci ve öğretmen görüşleri* [Yüksek lisans tezi]. Balıkesir Üniversitesi.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill Education.
- Morrison, D. (2003). Using activity theory to design constructivist online learning environments for higher order thinking: A retrospective analysis. *Canadian Journal of Learning and Technology*, 29(3).
- Moubayed, A., Injadat, M., Nassif, A. B., Lutfiyya, H. ve Shami, A. (2018). E-learning: Challenges and research opportunities using machine learning ve data analytics. *IEEE Access*, 6, 39117-39138.
- Nabiyev, V. V., (2012). *Yapay zeka: insan-bilgisayar etkileşimi*. Seçkin Yayıncılık.
- Nassif, A. B., Shahin, I., Attili, I., Azzeh, M. ve Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE access*, 7, 19143-19165.
- Nilsson, N. J., (2019). *Yapay zekâ geçmişi ve geleceği*. Boğaziçi Üniversitesi Yayınevi.

- Okur, E.,(2023). *Mobil robotlarda genişletilmiş kalman filtresi kullanılarak kapalı alan haritalandırma ve navigasyonda iyileştirme*. [Yüksek lisans tezi]. KTO Karatay Üniversitesi.
- ÖNDER, H. H. (2003). Uzaktan eğitimde bilgisayar kullanımı ve uzman sistemler. *TOJET: The Turkish Online Journal of Educational Technology*, 2(3).
- Özbay, Ö. (2015). Dünyada ve Türkiye’de uzaktan eğitimin güncel durumu. *Uluslararası Eğitim Bilimleri Dergisi*, (5), 376-394.
- Özdemir, Ş. (2016). *Eğitimde veri madenciliği ve öğrenci akademik başarı öngörüsüne ilişkin bir uygulama* [Doktora tezi.]. İstanbul Üniversitesi.
- Paramita, A. S. ve Tjahjono, L. M. (2021). Implementing machine learning techniques for predicting student performance in an e-learning environment. *International Journal of Informatics and Information Systems*, 4(2), 149-156.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1, 81-106.
- Race, P., (1998). *500 tips for open and flexible learning*. Psychology Press.
- Raza, K. (2020). *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*. Springer.
- Rish, I., Hellerstein, J. ve Thathachar, J. (2001). An analysis of data characteristics that affect naive Bayes performance. *IBM TJ Watson Research Center*, 30, 1-8.
- Romiszowski, A. J. (2004). How's the e-learning baby? Factors leading to success or failure of an educational technology innovation. *Educational Technology*, 44(1), 5-27.
- Sayar, B. (2021). *Yapay zekâ ve robot teknolojisinin gazetecilik pratikleri üzerindeki etkileri* [Doktoral tezi]. Marmara Üniversitesi.
- Shah, F. (1997). *Recognizing and responding to student plans in an intelligent tutoring system: CIRCSIM-Tutor*. Illinois Institute of Technology.
- Shahiri, A. M. ve Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414-422.

- Shen, J., Zhang, X. ve Lian, Z. (2021). Gender differences in human psychological responses to wooden indoor environment. *European Journal of Wood and Wood Products*, 79(1), 217-226.
- Slade, S. ve Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510-1529.
- Suykens, J. A., De Brabanter, J., Lukas, L. ve Vandewalle, J. (2002). Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*, 48(1-4), 85-105.
- Şeker, Ş. E. (2018). *CRISP-DM: Endüstriler arası standart işleme-veri madenciliği için*. YBS Ansiklopedi.
- Tabachnick, B. G., Fidell, L. S. ve Ullman, J. B. (2013). *Using multivariate statistics*. Pearson.
- Tan, A. H. ve Yu, P. S. (2003). Guest editorial: Text and web mining. *Applied Intelligence*, 18(3), 239.
- Taylor, R. W. (2002). Pros and cons of online learning—a faculty perspective. *Journal of European Industrial Training*, 26(1), 24-37.
- Telli, G. (2019). *Yapay zeka ve geleceğin meslekleri*. Doğu Kitapevi.
- Timms, M. J. (2016). Letting artificial intelligence in education out of the box: educational cobots and smart classrooms. *International Journal of Artificial Intelligence in Education*, 26, 701-712.
- Toosi, A., Bottino, A. G., Saboury, B., Siegel, E. ve Rahmim, A. (2021). A brief history of AI: how to prevent another winter (a critical review). *PET Clinics*, 16(4), 449-469.
- Tuğluk, M. N. ve Gök-Çolak, F. (2019). *Sanayi toplumu ve eğitimi*. Pegem Akademi.
- Uçar, A. ve Uludağ, M. H. (2018). Nesnelerin interneti (IoT) ile akıllı sınıf ve öğrenci takip sistemi tasarımı. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 9(2), 591-600.
- Uşun, S. (2006). *Bilgisayar destekli öğretimin temelleri*. Nobel Yayın Dağıtım.
- Ünal, A. ve Kılınç, İ. (2020). Yapay zeka işletme yönetimi ilişkisi üzerine bir değerlendirme. *Yönetim Bilişim Sistemleri Dergisi*, 6(1), 51-78.

- Vapnik, V. (1999). *The nature of statistical learning theory*. Springer Science and Business Media.
- Vassallo, D., Krishnamurthy, R. ve Fernando, H. J. (2020). Utilizing physics-based input features within a machine learning model to predict wind speed forecasting error. *Wind Energy Science Discussions*, 2020, 1-18.
- Weber, R., Schek, H. J. ve Blott, S. (1998, August). A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In *VLDB (Vol. 98, pp. 194-205)*.
- Witten, I. H., Frank, E., Hall, M. A., Pal, C. J. ve Mining, D. (2016). *Practical machine learning tools and techniques*. In *Data Mining (Vol. 2, p. 4)*. Morgan Kaufmann.
- Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11.
- Yakut, Y., Yakut, E. ve Yavuz, S. (2014). Yapay sinir ağıları ve destek vektör makineleri yöntemleriyle borsa endeksi tahmini. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19(1), 139-157.
- Yamamoto, G. T. ve Altun, D. (2020). Coronavirüs ve çevrimiçi (online) eğitimin önlenemeyen yükselişi. *Üniversite Araştırmaları Dergisi*, 3 (1), 25-34.
- Yavuzarslan, M. ve Erol, Ç. S. (2022). Öğrenme yönetim sistemi log kayıtlarının akademik başarı tahmininde kullanılması. *Bilişim Teknolojileri Dergisi*, 15(2), 199-207. <https://doi.org/10.17671/gazibtd.837884>
- Yekta, M. ve Arıcı, N. (2005). Mesleki ve teknik eğitimde çoklu ortam araçları kullanılmış web tabanlı öğretimin öğrenci başarısına etkisi. *Gazi Üniversitesi Ticaret ve Turizm Eğitim Fakültesi Dergisi*, (1), 144-153.
- Yeşiltuna, D. Ç. ve Tükel, İ. (2015). *Yeni medyanın yeni dünya düzeni*. Nobel yayıncılık
- Yıldırım, M. (2022). Film yorumları kullanılarak önerilen yapay zekâ tabanlı yöntemle duygu analizinin gerçekleştirilmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 34(2), 751-760.

- Yıldız, B. Ç. ve Çelik, E. (2021). Öğrenci başarısının derin öğrenme algoritmaları ile tahmini. *International Journal of Computer Science and Engineering*, 9(2), 39-47.
- Yıldız, N. D. ve Kaya, Ş. (2020). Makine öğrenmesi yöntemleri ile öğrenci performansının tahmini. *Eğitimde Nitel Araştırmalar Dergisi*, 8(3), 450-472.
- Yılmaz, A., (2019). *Yapay Zeka*. Kodlab.
- Yin, R. K. ve Moore, G. B. (1987). The use of advanced technologies in special education: Prospects from robotics, artificial intelligence, and computer simulation. *Journal of Learning Disabilities*, 20(1), 60-63.
- Yirci, R., Özalp, U. ve Kocabaş, İ. (2018). Relationship between mentor teachers' mentoring functions and beginning teachers' subjective happiness. *Kuram ve Uygulamada Eğitim Yönetimi*, 24(2), 387-435. <https://doi.org/10.14527/kuey.2018.010>
- Zawacki-Richter, O., Marín, V. I., Bond, M. ve Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhang, D., Zhou, L., Briggs, R. O. ve Nunamaker Jr, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information ve Management*, 43(1), 15-27.
- Zhang, K., Yılmaz, R., Ustun, A. B. ve Karaoğlu-Yılmaz, F. G. (2023). Learning analytics in formative assessment: A systematic literature review. *Journal of Measurement and Evaluation in Education and Psychology*, 14(Special Issue), 382-395. <https://doi.org/10.21031/epod.1272054>
- Zhang, Y., Ghandour, A. ve Shestak, V. (2020). Using learning analytics to predict students performance in moodle LMS. *International Journal of Emerging Technologies in Learning (iJET)*, 15(20). <https://doi.org/10.3991/ijet.v15i20.15915>

İnternet Kaynağı

- Hsu, C. W., Chang, C. C. ve Lin, C. J. (2016, 19 Mayıs). *A Practical Guide to Support Vector Classification* 12 Temmuz 2023 tarihinde <http://www.csie.ntu.edu.tw/~cjlin> adresinden edinilmiştir.
- Sunil, R. (2016, 15 Mayıs). *A complete tutorial which teaches data exploration in detail. Analytics Vidhya.* 21 Temmuz 2023 tarihinde <https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/> adresinden edinilmiştir.
- Karabay A. (2020, 12 Şubat). *Scikit-learn nedir?* 1 Temmuz 2023 tarihinde <https://www.karabayyazilim.com/blog/python/scikit-learn-nedir-2020-02-12-062241> adresinden edinilmiştir.
- Şanlı, O. (2018, 6 Eylül). *Veri madenciliği–Giris.* 20 Temmuz 2023 tarihinde <https://oyasanli.com/oyasblog/2018/09/06/veri-madenciligi-giris/> adresinden temin edilmiştir.
- MOODLE. (2022, 30 Nisan). *Assignment activity.* 10 Ağustos 2023 tarihinde https://docs.moodle.org/403/en/Assignment_activity adresinden edinilmiştir.
- Akdoğan F. (2020, 28 Ocak). *Makine öğrenmesi nedir? Ne işe yarar?* 2 Temmuz 2023 tarihinde <https://technogezgin.com/makine-ogrenmesi-nedir-ne-ise-yarar/> adresinden edinilmiştir.
- Morde, V. ve Setty, V. A. (2019, 8 Nisan). *XGBoost algorithm: long may she reign! Towards Data Science.* 2 Eylül 2023 tarihinde <https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d> adresinden edinilmiştir.
- Microsoft. (2021, 11 Nisan). *Multiclass logistic regression component.* 12.09.2023 tarihinde <https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/multiclass-logistic-regression?view=azureml-api-2> adresinden edinilmiştir.
- Matplotlib E. (2021, 07 Mayıs). *Matplotlib Kütüphanesi.* 3 Temmuz 2023 tarihinde <https://matplotlib.org/> adresinden edinilmiştir.

- Kaplan K. (2019, 13 Şubat). *Python ve Anaconda Kurulumu*. 9 Temmuz 2023 tarihinde <https://medium.com/kodcular/python-ve-anaconda-kurulumu-b8931bd80e64> adresinden edinilmiştir.
- NumPy D., (2022). *NumPy: the absolute basics for beginners*. 9 Temmuz 2023 tarihinde https://numpy.org/doc/stable/user/absolute_beginners.html adresinden temin edilmiştir.
- Akca, M. (2020, 25 Ağustos). *Nedir bu destek vektör makineleri?* 4 Ağustos 2023 tarihinde <https://medium.com/deep-learning-turkiye/nedir-bu-destek-vekt%C3%B6r-makineleri-makine-%C3%B6%C4%9Frenmesi-serisi-2-94e576e4223e> adresinden edinilmiştir.
- MyKinney, W., (2022, 20 Eylül). *Python for data analysis*. 7 Temmuz 2022 tarihinde <https://wesmckinney.com/book/> adresinden temin edilmiştir.
- Jain, K. (2015, 31 Ekim). *Machine Learning Basics for a new bie*. 2 Kasım 2023 tarihinde <https://www.analyticsvidhya.com/blog/2015/06/machine-learning-basics/> adresinden edinilmiştir.
- Ulgen E. (2017, 12 Kasım). *Makine öğrenimi bölüm-5 (karar ağaçları)*. 23 Temmuz 2023 tarihinde <https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-5-karar-agac-lari-c90bd7593010> adresinden temin edilmiştir.

