

**BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS
PROGRAMI**

**ÇEVİRİMİÇİ ÖĞRENME ORTAMLARINDA ÖĞRENME ANALİTİĞİ
VERİLERİ VE MAKİNE ÖĞRENMESİ KULLANARAK AKADEMİK
BAŞARININ DEĞERLENDİRİLMESİ**

**HAZIRLAYAN
RAMAZAN TEKİNARSLAN**

YÜKSEK LİSANS TEZİ

ANKARA - 2022

**BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS
PROGRAMI**

**ÇEVİRİMİÇİ ÖĞRENME ORTAMLARINDA ÖĞRENME ANALİTİĞİ
VERİLERİ VE MAKİNE ÖĞRENMESİ KULLANARAK AKADEMİK
BAŞARININ DEĞERLENDİRİLMESİ**

**HAZIRLAYAN
RAMAZAN TEKİNARSLAN**

YÜKSEK LİSANS TEZİ

**TEZ DANIŞMANI
DOÇ. DR. MUSTAFA SERT**

ANKARA - 2022

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı çerçevesinde Ramazan Tekinarslan tarafından hazırlanan bu çalışma, aşağıdaki jüri tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 24 / 01 / 2022

Tez Adı: Çevrimiçi Öğrenme Ortamlarında Öğrenme Analitiği Verileri Ve Makine Öğrenmesi Kullanarak Akademik Başarının Değerlendirilmesi

Tez Jüri Üyeleri (Unvanı, Adı - Soyadı, Kurumu)	İmza
--	-------------

Doç. Dr. , Mustafa SERT, Başkent Üniversitesi	
---	-------

Prof. Dr. , Hamit ERDEM, Başkent Üniversitesi	
---	-------

Doç. Dr. , Derya YILMAZ, Gazi Üniversitesi	
--	-------

ONAY

Prof. Dr. Faruk ELALDI

Fen Bilimleri Enstitüsü Müdürü

Tarih : ... / ... /

BAŞKENT ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
YÜKSEK LİSANS / DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

Tarih: 21 / 02 / 2022

Öğrencinin Adı, Soyadı: Ramazan Tekinarslan

Öğrencinin Numarası: 21920113

Anabilim Dalı: Bilgisayar Mühendisliği Anabilim Dalı

Programı: Bilgisayar Mühendisliği Tezli Yüksek Lisans

Danışmanın Unvanı/Adı, Soyadı: Doç. Dr. Mustafa Sert

Tez Başlığı: Çevrimiçi Öğrenme Ortamlarında Öğrenme Analitiği Verileri ve Makine Öğrenmesi Kullanarak Akademik Başarının Değerlendirilmesi

Yukarıda başlığı belirtilen Yüksek Lisans tez çalışmamın; Giriş, Ana Bölümler ve Sonuç Bölümünden oluşan, toplam 54 sayfalık kısmına ilişkin, 21/ 02/ 2022 tarihinde şahsım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 8' dir. Uygulanan filtrelemeler:

1. Kaynakça hariç
2. Alıntılar hariç
3. Beş (5) kelimeden daha az örtüşme içeren metin kısımları hariç

“Başkent Üniversitesi Enstitüleri Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Usul ve Esaslarını” inceledim ve bu uygulama esaslarında belirtilen azami benzerlik oranlarına tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Öğrenci İmzası:.....

ONAY

Tarih: 21 / 02 / 2022

Öğrenci Danışmanı Unvan, Adı, Soyadı, İmza:

Doç. Dr. Mustafa Sert

TEŞEKKÜR

Yazar, bu çalışmanın gerçekleşmesinde katkılarından dolayı, aşağıda adı geçen kişilere teşekkür eder.

Sayın Doç. Dr. Mustafa SERT'e (tez danışmanı), çalışmanın sonuca ulaştırılmasında ve karşılaşılan güçlüklerin aşılmasında her zaman yardımcı ve yol gösterici olduğu için...

Sayın Prof. Dr. Erkan TEKİNARSLAN'a (Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dekanı) tezle ilgili yaptığı değerlendirmeler ile katkıları için...

Sayın Doç. Dr. Levent ÖZBEK'e (Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü), tezle ilgili yaptığı değerlendirmeler ile katkıları için...

Sayın Dr. Öğr. Üyesi Nadide Gizem Akgülğil MUTLU'ya (Başkent Üniversitesi Güzel Sanatlar, Tasarım ve Mimarlık Fakültesi), tezle ilgili tartışmalardaki katkıları için...

Sayın Dr. Öğr. Üyesi Bora Mihrişah TÜRKKAN'a (Başkent Üniversitesi Güzel Sanatlar, Tasarım ve Mimarlık Fakültesi), tezle ilgili tartışmalardaki katkıları için...

Sayın Öğr. Gör. Öner Emre TARTAN'a (Başkent Üniversitesi Teknik Bilimler Meslek Yüksekokulu), model başarımlarının elde edilmesinde yaptığı değerlendirmeler ve destekleri için...

ÖZET

Ramazan TEKİNARSLAN

ÇEVİRİMİÇİ ÖĞRENME ORTAMLARINDA ÖĞRENME ANALİTİĞİ VERİLERİ VE MAKİNE ÖĞRENMESİ KULLANARAK AKADEMİK BAŞARININ DEĞERLENDİRİLMESİ

Başkent Üniversitesi Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

2022

Covid - 19 salgını döneminde daha önce de kullanılan ancak yaygın olmayan çevrimiçi öğrenme ortamlarının sayısı artmıştır. Bu ortamlarda oluşan öğrenme analitiği verileri üzerinde makine öğrenmesi yöntemleri ile öğrenci başarısının tahmini ve sınıflandırma çalışmaları son yıllarda önem kazanmaktadır. Çevrimiçi öğrenme ortamından elde edilen öğrenme analitiği verileriyle öğrenci başarısı arasındaki ilintinin anlaşılması amacıyla; öğrenci başarısının tahmini ve sınıflandırması yapılmıştır. Veri kümesi üzerinde *ilinti* (korelasyon), *özniteliğin önemi*, *fisher puanı*, *selectKbest* ve *bilgi kazancı* öznitelik seçim yöntemleri kullanılarak öğrenci başarısı ile ilintili olan özniteliklerin tespiti yapılmıştır. Seçimi yapılan özniteliklerin normalizasyonu ve verilerin one hot encoding (OHE) yöntemi ile temsili sağlanmıştır. Öğrenci başarısının tahmini ve sınıflandırılmasında geleneksel makine öğrenmesi algoritmaları Rastgele orman (Random Forest-RF), Çok katmanlı algılayıcı (Multilayer Perceptron-MLP) ve k-en yakın komşu (k-Nearest Neighbours, k-NN) probleme uygulanmıştır. Bu çalışmada, öğrenci başarısının tahmini ve sınıflandırılması için verilerin OHE temsiline, öznitelik seçimine ve Evrimsel Sinir Ağı (ESA)-(Convolutional Neural Network-CNN) mimarisine dayalı bir yöntem önerilmektedir. Belirtilen amaç doğrultusunda Başkent Üniversitesi'nin 2020-2021 akademik yılının çevrimiçi öğrenme ortamı olan Moodle verileri ile 2013-2014 yıllarına ait İngiltere'deki Open University çevrimiçi öğrenme veri kümesi kullanılmıştır. Başkent Üniversitesi veri kümesinde OHE temsili ve temsilsiz üçlü sınıflandırma çalışmasında önerilen ESA modeli %92 doğruluk başarımları ile geleneksel makine öğrenmesi yöntemlerinden yüksektir. Literatürde Open University veri kümesi üzerinde yapılan ikili, üçlü ve dördü sınıflandırma çalışma sonuçları ile önerilen ESA modeli sonuçları karşılaştırılmıştır. İkili sınıflandırmada %95,43 ile en yüksek başarımları bulunurken üçlü sınıflandırmada %88 ve dördü sınıflandırmada %73,32 değerleri diğer çalışmalara göre daha yüksek olarak

bulunmuştur. Öğrenci başarısını tahmininde hata değerlendirme ölçütü olarak belirlenen kök ortalama kare hata (RMSE) ve ortalama mutlak hata (MAE) değerleri önerilen ESA modelinde %1'in altında kalarak diğer modellere göre düşük bir hata oranı vermiştir. Farklı veri kümeleri üzerinde ayrı ayrı kullanılan öznetelik seçimi, verilerin OHE temsili ve ESA mimarisine dayalı yöntem bu çalışmada birlikte kullanılarak literatüre katkı sunmaktadır.

ANAHTAR KELİMELEER: Öğrenci başarısının sınıflandırılması, Evrişimsel Sinir Ağı, OHE temsili, Öğrenme analitiğı, Öznetelik seçimi, Covid-19



ABSTRACT

Ramazan TEKİNARSLAN

ASSESSMENT OF ACADEMIC SUCCESS IN ONLINE LEARNING ENVIRONMENTS USING LEARNING ANALYTICS DATA AND MACHINE LEARNING

Başkent University Institute of Science and Engineering

Department of Computer Engineering

2022

During the Covid-19 pandemic, the use of online learning environments is rapidly increasing. Estimation and classification studies of student success with machine learning methods on learning analytics data generated in these environments have gained importance in recent years. In order to understand the relationship between learning analytics data obtained from the online learning environment and student success; in this thesis, we deal with the estimation and classification of student success using the learning analytic data. With these aims, we propose a method based on One-Hot-Encoding (OHE) representation of data, feature selection, and Convolutional Neural Network (CNN) architecture for the estimation and classification of student success. We determine the features related to student success by using correlation, feature importance, fisher score, selectKbest, and knowledge gain feature selection methods on the data set. We also perform the normalization of the selected features and transform the representation of the data with OHE method. To demonstrate the efficacy of the proposed CNN-based architecture we also employ traditional machine learning algorithms such as Random Forest (RF), Multilayer Perceptron (MLP), and k-Nearest Neighbor, (k-NN) in the analyses. For the learning analytics data, we use the Moodle data, which is the online learning environment of Başkent University of the 2020-2021 academic year, and the Open University online learning dataset of years 2013-2014 in England. The results on the Başkent University dataset show that the proposed CNN model with- and without-OHE in three-class classification (fail, pass, distinction) score is higher than the traditional machine learning methods. We also compare the results of binary (fail, pass), three-class (withdrawn, fail, pass) and four-class (withdrawn, fail, pass, distinction) classification performance of our proposed CNN-based architecture on the Open University dataset. We achieved better results than the literature with the highest accuracy rates of 95.43% in two-class classification, 88% in three-class classification and 73.32% in four-class classification. For the estimation of student's grade, Root

mean square error (RMSE) and mean absolute error (MAE) values remained below 1% in the proposed CNN-based model, giving a low error rate compared to other models. As a result, the proposed method achieves promising and better results in the evaluations.

KEYWORDS: Student performance classification, Convolutional Neural Network, One-hot encoding (OHE), Learning analytics, Feature selection, Covid-19



İÇİNDEKİLER

Sayfa

TEŞEKKÜR.....	i
ÖZET.....	ii
ABSTRACT	iv
İÇİNDEKİLER.....	vi
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ.....	ix
SİMGELER VE KISALTMALAR LİSTESİ.....	xi
1. GİRİŞ	1
1.1. Problem Tanımı	3
1.2. Motivasyon	4
1.3. Tezin Amacı.....	4
1.4. Tezin Kapsamı	4
1.5. Tezin Katkıları.....	5
1.6. Tezin Organizasyonu.....	5
2. LİTERATÜR TARAMASI	6
2.1. Öğrenme Analitiği	6
2.2. Makine Öğrenmesi Literatürü	8
3. TEMEL BİLGİLER.....	12
3.1. Çevrimiçi Öğrenme Ortamı.....	12
3.2. Öznitelik Seçimi	13
3.2.1 İlinti (Korelasyon).....	13
3.2.2 Bilgi Kazancı (information gain)	14
3.2.3 Fisher Puanı.....	14
3.2.4 SelectKBest	15
3.2.5 Öznitelik Önemi (Feature Importance)	15
3.3. Regresyon ve Sınıflandırma.....	15
3.3.1 Regresyon.....	16
3.3.2 Sınıflandırma	16
3.3.3 Regresyon ve Sınıflandırma Algoritmaları.....	16

3.3.2.1.	Rastgele Orman.....	16
3.3.2.2.	Çok Katmanlı Algılayıcı	18
3.3.2.3.	Evrişimsel Sinir Ağı	19
3.3.2.4.	k-En Yakın Komşu.....	21
3.4.	Değerlendirme Ölçütleri	22
4.	ÖZNİTELİK ANALİZİ.....	24
4.1.	Veri Kümesi Ön Hazırlık	24
4.1.1.	Moodle Günlükleri.....	24
4.1.2.	Veri İşleme	25
4.1.3.	Normalizasyon.....	27
4.1.4.	İstatistikler.....	27
4.2.	Öznitelik Analizi ve Seçimi	29
4.2.1	Davranışsal ve Zihinsel Öznitelikler	29
4.2.2	Öznitelik Seçimi.....	30
4.3.	One Hot Encoding Yöntemi.....	32
5.	ÖĞRENCİ BAŞARIMININ DEĞERLENDİRMESİ.....	34
5.1.	Modellerin Tasarım Kriterleri	34
5.1.1.	Rastgele Orman.....	34
5.1.2.	k-En Yakın Komşu.....	36
5.1.3.	Çok Katmanlı Algılayıcı	36
5.1.4.	Evrişimsel Sinir Ağı	37
5.2.	DeneySEL Çalışmalar	38
5.2.1.	Veri Kümesi.....	38
5.2.2.	Sınıflandırma Çalışmaları.....	42
5.2.3.	Öğrenci Başarım Tahmini.....	49
5.3.	Değerlendirmeler	51
SONUÇ.....		52
KAYNAKLAR.....		55

TABLÖLAR LİSTESİ

	Sayfa
Tablo 4.1. Ders erişim günlüğü	25
Tablo 4.2. Veri kümesi alan açıklamaları	26
Tablo 4.3. Normalizasyon uygulanan alanlar örneği	27
Tablo 4.4. Davranışsal etkinlik veri kümesi.....	29
Tablo 4.5. Zihinsel etkinlik veri kümesi	29
Tablo 4.6. Öznitelik seçim yöntemlerinin normalize edilmiş sonuçları	30
Tablo 4.7. Veri kümesi.....	32
Tablo 4.8. Label Encoding temsili	32
Tablo 4.9. One Hot Encoding temsili.....	33
Tablo 4.10. Kayıt örneği	33
Tablo 4.11. ESA encoder karşılaştırmaları	33
Tablo 5.1. k-NN karşılaştırmaları	36
Tablo 5.2. ESA model özeti	37
Tablo 5.3. OHE yöntemli ESA model özeti.....	37
Tablo 5.4. Başkent Üniversitesi veri kümesi istatistikleri.....	39
Tablo 5.5. OLUAD veri kümesi.....	42
Tablo 5.6. Puan ölçeği.....	42
Tablo 5.7. BÜ veri kümesi üzerinde tasarlanan modellerin sınıflandırma (doğruluk, f1-ölçeği, kesinlik, duyarlılık) karşılaştırmaları	45
Tablo 5.8. Başkent Üniversitesi veri kümesi üzerinde tasarlanan modellerin sınıflandırma (AUC alan) karşılaştırmaları	48
Tablo 5.9. OULAD veri kümesi üzerinde önerilen yöntemin sınıflandırma başarımı ile literatür karşılaştırması.....	49
Tablo 5.10. ESA modeli tahmin özeti.....	50
Tablo 5.11. Başkent Üniversitesi veri kümesinde öğrenci başarı tahmini hata sonuçları.....	50
Tablo 5.12. Başkent Üniversitesi veri kümesi üzerinde başarımlar karşılaştırmaları	51

ŞEKİLLER LİSTESİ

	Sayfa
Şekil 3.1. RF sınıflandırma yapısı [29]	17
Şekil 3.2. RF regresyon yapısı	17
Şekil 3.3. MLP'nin gizli katmanlı yapısı	18
Şekil 3.4. MLP nöron çıkışının hesaplanması.....	18
Şekil 3.5. ESA mimarisi.....	20
Şekil 3.6. k-NN yeni örneğin sınıflandırması	21
Şekil 4.1. Veri kümesi ön hazırlık iş akışı.....	24
Şekil 4.2. Örnek günlük dosyası.....	24
Şekil 4.3. Veri kümesi örneği.....	26
Şekil 4.4. Dönem sonu geçme notu dağılım grafiği.....	28
Şekil 4.5. Başarı durumu sayıca dağılımı.....	28
Şekil 4.6. Öznitelik secim grafiği.....	31
Şekil 4.7. Karışıklık matrisi.....	31
Şekil 4.8. Kayıt örneği OHE temsili	33
Şekil 5.1. Öğrenci başarımının değerlendirilmesi.....	34
Şekil 5.2. Başkent Üniversitesi veri kümesindeki “Başarısız”, “Başarılı”, “Yüksek Başarılı” sınıflandırması için oluşturulan örnek RF ağaç yapısı	35
Şekil 5.3. MLP katman yapısı	36
Şekil 5.4. ESA modeli eğitim ve doğrulama kayıp grafiği	38
Şekil 5.5. OULAD ders yapısı	39
Şekil 5.6. OLUAD veri kümesi yapısı	40
Şekil 5.7. OULAD öznitelik seçimi	41
Şekil 5.8. Sınıflandırma yöntemleri	42
Şekil 5.9. RF OHE karışıklık matrisi	43
Şekil 5.10. RF karışıklık matrisi.....	43
Şekil 5.11. k-NN OHE karışıklık matrisi	43
Şekil 5.12. k-NN karışıklık matrisi	43
Şekil 5.13. MLP OHE karışıklık Matrisi.....	44
Şekil 5.14. MLP karışıklık Matrisi.....	44
Şekil 5.15. ESA OHE karışıklık matrisi.....	44
Şekil 5.16. ESA karışıklık matrisi.....	44

Şekil 5.17. RF OHE üçlü sınıf.....	46
Şekil 5.18. RF üçlü sınıf.....	46
Şekil 5.19. k-NN OHE üçlü sınıf	46
Şekil 5.20. k-NN üçlü sınıf.....	46
Şekil 5.21. MLP OHE üçlü sınıf	47
Şekil 5.22. MLP üçlü sınıf	47
Şekil 5.23. ESA OHE üçlü sınıf.....	47
Şekil 5.24. ESA üçlü sınıf	47
Şekil 5.25. Başarı tahmin yöntemleri	50



SİMGELER VE KISALTMALAR LİSTESİ

AİK	Alıcı İşletim Karakteristiği (Receiver Operating Characteristic-ROC)
BLSTM	Bidirectional Long Short - Term Memory
CRF	Conditional Random Fields
ÇKA	Çok Katmanlı Algılayıcı (Multilayer Perceptron-MLP)
DVM	Destek Vektör Makinesi (Support Vector Machine-SVM)
ESA	Evrişimsel Sinir Ağı (Convolutional Neural Network-CNN)
EK	Etiket Kodlama (Label Encoding)
GA	Gradyan Artırması (Gradient Boosting Machines-GBM)
k-EYK	k-En Yakın Komşu (k-Nearest Neighbors-k-NN)
KOKH	Kök Ortalama Kare Hata (Root Mean Square Error-RMSE)
LR	Lojistik Regresyon (Logistic Regression-LR)
MTK	Model Tasarım Kriterleri
OHE	One-Hot Encoding
OMH	Ortalama Mutlak Hata (Mean Absoulate Error-MAE)
ÖSA	Özyinelemeli Sinir Ağı (Recurrent Neural Network-RNN)
RO	Rastgele Orman (Random Forest-RF)
UKSB	Uzun Kısa Süreli Bellek Long Short –Term Memory-LSTM)

1. GİRİŞ

Çevrimiçi öğrenme ortamları daha önceden uzaktan eğitim adıyla kullanılmasına rağmen [1] Covid 19 salgını döneminde kullanımı yaygınlaşarak artmıştır. Özellikle 2000 yıllardan itibaren bilgi ve iletişim teknolojileri alanında gelişmeler ve internetin yaygın kullanımı bu ortamlarda büyük miktarda verilerin oluşmasını sağlamıştır. Oluşan bu veriler üzerinde; veri analitik yöntemlerinin öğrenme süreçlerine uyarlanması ‘öğrenme analitiği’ kavramını doğurmuştur.

Öğrenme analitiği, öğrenmeyi ve öğrenmenin olduğu ortamları anlamak ve iyileştirmek için öğrenciler ve içerikler hakkındaki verilerin ölçümü, toplanması, analizi ve raporlanması olarak görülebilir [2].

Çevrimiçi uygulamalar hayatımızın birçok alanında kullanıldığı gibi son yıllarda öğrenme süreçlerinde de yaygın olarak kullanılmaktadır. Çevrimiçi ortamlardaki tüm etkileşimler, paylaşımlar, güncellemeler arkasında dijital izler bırakır ve bu izler büyük verileri oluşturur [3].

Öğrencinin çevrimiçi öğrenme ortamlara katılımından elde edilen öğrenme analitiği verileri ile başarısı arasındaki ilintinin anlaşılması, öğrenci başarımlarının otomatik olarak tahminini ve sınıflandırmasını mümkün kılması açısından önemlidir.

Öğrenme analitiği ve makine öğrenmesi yöntemleri birlikte kullanılarak öğrenme ortamı ve içeriklerinin iyileştirilmesi, başarısız olabilecek risk altındaki öğrencilerin tespiti vb. süreçlerde geri bildirimler sunmaktadır [4] [5].

Çevrimiçi öğrenme ortamından elde edilen öğrenme analitiği verileri üzerinde öğrenci başarısının tahmini ve sınıflandırmasında öğrenme analitiği ve makine öğrenmesi yöntemlerinin beraber kullanıldığı çalışmalar bulunmaktadır [6] [7] [8]. Bu ve benzeri çalışmalarda kullanılan yöntemler ve sonuçları ile ilgili ayrıntılı inceleme literatür taraması bölümünde verilmiştir. Aşağıda incelenen araştırmalar, tez çalışmasında kullanılan Rastgele Orman (Random Forest-RF), Çok Katmanlı Algılayıcı (Multilayer Perceptron-MLP), K-En Yakın Komşu (k-Nearest Neighbours, k-NN) ve Evrimsel Sinir Ağı (ESA)-(Convolutional Neural Network-CNN) makine öğrenmesi algoritmalarının seçiminde motivasyon sağlamıştır.

Namoun vd. 2020 yılından itibaren son on yılda yayınlanan 62 makaleyi incelemiştir [8]. Çalışmalarında hem veri madenciliği hem de öğrenme analitiği yöntemlerini kullanarak değerlendirme yapılmış. Öğrenme çıktılarının sınıflandırılmasında başarı oranı en yüksek beş ve en kötü beş makine öğrenmesi algoritmasının sunulması, gelecekteki yapılacak çalışmalarda

kullanılacak makine öğrenmesi algoritmaları için bir fikir vermektedir. Rastgele Orman algoritması sınıflandırma başarımları en yüksekler içerisinde yer almaktadır [9].

Makine öğrenmesi yöntemleri ile yapılan çalışmaların bir kısmı çevrimiçi öğrenme ortamı verileri üzerinde bir kısmı da üniversitelerin veri tabanlarından elde edilen öğrenme sonuçları üzerinden yapılmıştır Öğrenci başarısının tahmini ve sınıflandırılmasında bu veri kümeleri üzerinde farklı makine öğrenmesi yöntemleri kullanılmıştır [10] [11]. Aşağıda incelenen araştırma sonuçlarında MLP ve k-NN yöntemlerinde elde edilen başarımları sonuçları tez çalışmasında bu yöntemlerin kullanılmasında motivasyon sağlamıştır.

Cagliero vd. çalışmalarında İtalya Torino Politeknik Üniversitesi'nde yeni kayıt olan 5000 öğrencinin öğrenme ortamında belli aralıklar ile her dersten girdiği sınavlar, eğitim videolarına erişim sayıları demografik bilgilerini sınıflandırıp başarısını incelemişler [11]. Çalışmalarında istatistiksel analizler ve MLP, RF, Karar Ağacı (Decision Tree-DT) ve k-NN yöntemlerini kullanmışlar. En başarılı sonuç %95 ile MLP ve %90 ile k-NN olarak bulunmuş. Gelecekte daha farklı derslerde ve öğrenci profilinde sınıflandırma yapılarak çalışmalar yapılabilceği söylenmiş.

Makine öğrenmesinin bir alt dalı olan derin öğrenme algoritmalarından Evrimsel Sinir Ağı (ESA) [12], çevrimiçi öğrenme ortamından elde edilen verilerin sınıflandırılmasında yaygın olarak kullanılmaktadır [13] [14] [15] [16].

Karimi vd. çalışmalarında İngiltere'deki Open University veri kümesi [17] üzerinde öğrencilerin derslere haftalık erişimleri üzerinden sınıflandırma yaparak başarımlarını incelemişler [16]. Çalışmalarında veri kümesinde öğrencinin demografik bilgilerinde OHE temsili ve ESA, doğrusal regresyon ve destek vektör makinesi yöntemlerini kullanmışlar. En başarılı sonuç %87 ile ESA yönteminde elde edilmiştir.

Veri kümeleri üzerindeki kategorik değişkenlerin sayısal forma dönüştürülmesinde genel olarak etiket kodlama (label encoding) ve kategorik değişkenlerin ikili temsili anlamına gelen One-Hot Encoding-OHE kullanılır. Etiket kodlama ağaç tabanlı yöntemlerde ve sıralı kategorik değişkenler için kullanılır. OHE temsili ise n sınıflı değişkenlerde n boyutlu ikili (0,1) temsili olarak ifade edilir. Bu temsil kukla (dummy) değişken diye tanımlanır [18]. Veri kümesinin karakteristiği sınıflandırmada OHE temsiline daha uygun olduğundan ikili temsil yöntemi tez çalışmasında tercih edilmiştir.

Bu çalışmada çevrimiçi öğrenme ortamından elde edilen öğrenme analitiği verileri kullanılmıştır. Öğrenme analitiği verileri üzerinde beş farklı öznelik seçim yöntemi uygulanarak öğrenci başarısı ile en çok ilintili olan öznelikler belirlenmiştir. Belirlenen öznelikler ile oluşan veri kümesi üzerinde normalizasyon ve OHE temsili yöntemi

uygulanmıştır. Öğrenci başarısının tahmini ve sınıflandırması için öznitelik seçimine, ders etkinliklerinin OHE temsili ve Evrişimsel Sinir Ağına (ESA) dayalı bir model bu çalışmada önerilmiştir.

1.1. Problem Tanımı

Çevrimiçi öğrenme ortamına öğrencinin katılımından elde edilen öğrenme analitiği verileri ile öğrenci başarısı arasındaki ilintinin anlaşılması, öğrenci başarısının sınıflandırılması ve tahmini bu çalışmanın problemi/görevleri olarak tanımlanabilir.

Öğrenme analitiği verileri, öğrencinin çevrimiçi ortamda kayıt olduğu derslerin tanımlı etkinliklerine (ders dokümanları, ders videoları, duyuru vb.) erişim sayıları ve girdiği sınavlardan aldığı puanlar olarak ifade edilir. Bu çalışmada, Öğrenci başarısının sınıflandırması ve tahmini görevlerinde çevrimiçi öğrenme ortamından elde edilen öğrenme analitiği verileri veri kümesini oluşturmuştur.

Sınıflandırma görev tanımlarında makine öğrenmesi algoritması k adet sınıf için çözümü aşağıdaki fonksiyon ile belirler:

$$f: \mathbb{R}^n \rightarrow \{1, \dots, k\} \quad (1.1)$$

Fonksiyon, x adet girdi özniteliğini kullanarak y olarak bir sayısal değer ile sınıfı tanımlar ve $y=f(x)$ fonksiyonu ile ifade edilir.

Öğrenci başarısının sınıflandırması görevinde veri kümesindeki 9 adet öznitelik, makine öğrenmesi sınıflandırma fonksiyonu ile belirlenen başarısız, başarılı ve yüksek başarılı sınıflarından birini tanımlamada kullanılmaktadır.

Öğrenci başarısının tahmini görevlerinde makine öğrenmesi algoritması girilen n adet girdi özniteliğini kullanarak aşağıdaki fonksiyon ile sayısal bir değer hesaplar:

$$f: \mathbb{R}^n \rightarrow \mathbb{R} \quad (1.2)$$

Başarının tahmini sonucu sınıflandırmaya benzese de formatları farklıdır.

Başarım tahmini fonksiyonu ile veri kümesindeki 9 adet öznitelik kullanılarak, öğrencinin dönem sonu geçme notu tahmin edilmektedir.

1.2. Motivasyon

Gelişen teknoloji ve içinde bulunulan Covid-19 salgını, çevrimiçi öğrenme ortamlarının yaygınlaşmasını ve önemini artırmıştır. Öğrencilerin çevrimiçi öğrenme ortamları ile etkileşimleri akademik başarılarının sınıflandırılması ve tahmininde etkin rol oynamaktadır. Bu ortamlardan elde edilen öğrenme analitiği verileriyle sınıflandırmanın otomatik yapılabilmesi, dönem sonu başarısız olabilecek öğrencilerin erken tespiti bu verilerin önemini bir kat daha artırmıştır. Öğrencilerin çevrimiçi öğrenme ortamında yürütülen derslere katılımından elde edilen veriler, öğrenme analitiği verileri olarak görülür [2].

1.3. Tezin Amacı

Çevrimiçi öğrenme ortamından elde edilen öğrenme analitiği verileri ile öğrenci başarısının arasındaki ilintinin incelenmesi bu tezin ana amacını oluşturmaktadır. Öğrencilerin çevrimiçi ortamda yürütülen derslerin etkinliklerine katılımı ve girdiği sınavlardan aldığı puanlar ile akademik başarısının sınıflandırması otomatik yapılabilir. Çevrimiçi öğrenme ortamında tanımlı bu etkinliklere öğrencinin erişim sayısının az olması dönem sonu olası başarısızlıkların erken tespiti açısından fayda sağlayabilir. Belirtilen ana amaç doğrultusunda elde edilen öğrenme analitiği verileri üzerinde makine öğrenmesi yöntemleri ile öğrenci başarısının tahmini ve sınıflandırılması yapılmıştır.

1.4. Tezin Kapsamı

Öğrenci başarısının çevrimiçi ortamlardan elde edilen öğrenme analitiği verileri ile ilintisinin incelenmesi amacıyla: Başkent Üniversitesi çevrimiçi öğrenme ortamı Moodle öğrenme analitiği verileri üzerinde öğrenci başarısının tahmini ve sınıflandırması yapılmıştır. Belirlenen veri kümesi Covid-19 salgını döneminde (2020-21 akademik yıl) çevrimiçi ortamda yürütülen 2 dersin öğrenme analitiği verilerini kapsamaktadır.

Öğrencinin başarısı ile öğrenme analitiği verilerinin ilintisi, öğrencinin girdiği sınavlar ve ders etkinliklerine erişim sayıları kapsamında değerlendirilmiştir. Bu çalışmada yanıtları aranan araştırma soruları şunlardır:

- 1) Öğrenme ortamındaki davranışsal etkinliklere (ders dokümanları, ders videoları, ders izlencesi vb.) öğrencinin erişiminden elde edilen özniteliklerin, öğrencinin başarısına etkisi var mıdır?
- 2) Öğrencinin çevrimiçi ortamda katıldığı zihinsel etkinliklere (sınavlar, ödev) katılımından elde edilen özniteliklerin, başarısındaki katkısı nedir?

- 3) Öznitelik seçiminin, öğrenci başarısındaki etkisi nedir?
- 4) Öğrencinin cinsiyeti ve dersin açıldığı dönemin başarıya etkisi var mıdır?

Tez kapsamında önerilen yöntemlerin etkinliğini sınamak amacıyla, araştırmacıların kullanımına açık 2013 ve 2014 yıllarına ait İngiltere’deki Open University veri kümesi üzerinde ise sınıflandırma çalışması yapılmıştır.

1.5. Tezin Katkıları

Öğrenci başarısının tahmini ve sınıflandırması görev tanımlarına uyan literatürde incelenen güncel çalışmalar ve değerlendirmeleri literatür taraması bölümünde verilmiştir. İncelenen çalışmalarda başarının tahmini ve sınıflandırması başarı oranlarında makine öğrenmesi algoritmalarından RF, k-NN, MLP ve ESA modelinin diğerlerine oranla öne çıktığı görülmüştür. Özellikle derin öğrenme algoritmalarından olan ESA modeli genelde görsel sınıflandırma alanında kullanırken [13], son yıllarda çevrimiçi ortamlardan elde edilen öğrenme analitiği verilerinden öğrenci başarısının tahmini ve sınıflandırması alanında kullanımı yaygınlaşmaktadır [16] [19] [20].

Veri kümelerindeki kategorik değişkenlerin sayısallaştırılmasında genel olarak etiket kodlama (label encoding) [18] ve one hot encoding [18] yöntemleri kullanılmaktadır. Kategorik değişken değerlerinin sıralı olduğu durumlarda label diğer durumlarda genel olarak one hot encoding yöntemi tercih edilir [18].

Beş farklı öznitelik seçme (*ilinti, bilgi kazancı, fisher puanı, selectkbest, özneliliğin önemi*) yöntemi sonuçları ve verinin OHE temsili yöntemleri beraber kullanılarak model başarımında daha az öznitelik ile de yüksek doğruluk başarı oranı elde edilebileceği görülmüştür.

Bu çalışmada öznitelik seçimi, verinin OHE temsili ve Evrimsel Sinir Ağı mimarisine dayalı bir model önerilerek literatüre katkı sunulmuştur.

1.6. Tezin Organizasyonu

Bu çalışmanın diğer bölümleri şu şekilde düzenlenmiştir: Bölüm 2’ de yapılan literatür taraması anlatılmış, Bölüm 3’ de temel bilgiler verilmiş, Bölüm 4’ de öznitelik analizi yapılmış, Bölüm 5’ de öğrenci başarımının değerlendirilmesi yapılmış, sonuçlar ve tartışmalar ise Bölüm 6’ da sunulmuştur.

2. LİTERATÜR TARAMASI

Öğrencinin akademik başarısının değerlendirilmesinde öğrenme analitiği ve makine öğrenmesi kullanılarak yapılmış çalışmalardan elde edilen sonuçlar, öğrenme analitiği ve makine öğrenmesi başlıkları altında incelenmiştir. Bu alanlarda incelenen çalışmalarda öğrenci başarısının tahmini ve sınıflandırma görev tanımı esas alınmıştır.

2.1. Öğrenme Analitiği

Çevrimiçi öğrenme ortamların yaygınlaşması ve öğrenmede önem kazanması ile birlikte bu ortamlarda öğrenenlerin bıraktığı izler öğrenme analitiğinin gelişmesine katkı vermiştir. Öğrencinin başarısının tahmini için öğrenme analitiği ve makine öğrenmesi yöntemleri beraber kullanılarak alanında birçok çalışma yapılmıştır [4] [5] [6] [7] [8]. Bu çalışmaların an kullandığı öğrenme analitiği teknikleri ve sonuçları hakkında bilgiler aşağıda verilmiştir.

Daud vd. çalışmalarında Pakistan'daki çeşitli üniversitelerdeki lisans ve ön lisans bölümlerinde okuyan 3000 öğrencinin, 2004-2011 yılları arasındaki verilerini incelemişler [4]. Çalışmalarında sınıflandırma modelleri yardımı ile öğrencilerin bilgileri, ailesinin varlık durumu, geliri ve harcama bilgilerinden oluşturulan veri kümesi kullanmıştır. Çalışmada öğrenci başarısında aile harcamalarının etkisi üzerine sınıflandırmada Karar ağacı (Decision Tree-DT), Destek vektör makinesi (Support Vector Machine-SVM), Naive Bayes (NB) yöntemleri kullanılmıştır. Öğrenci başarısında, öğrencinin özelliğini yanı sıra aile harcamalarının etkisi sınıflandırılmasında %86,7 ile SVM yöntemi en iyi başarımla sonuç vermiştir. Gelecekte çevrimiçi öğrenme ortam verilerinden farklı yöntemler ile sınıflandırma çalışmaları yapılabileceği vurgusu yapılmıştır.

Jha vd. çalışmalarında çevrimiçi öğrenme ortamı Moodle üzerinden yürütülen bilgisayar bilimleri dersi verileri üzerinden öğrenci başarısını incelemişler [5]. Çalışmalarında veri gruplandırması (k-means) ile sınıflandırma yöntemleri (karar ağacı) kullanılmıştır. Çalışmada başarısız olabilecek öğrencilerin erken tespiti ve grup çalışmasının başarıda önemli olduğu tespiti yapılmış. Bu tespiti ek olarak eğitim içeriklerinin öğrenme ortamına uygunluğunun da başarıda önemli olduğu söylenebilir.

Lahmire vd. çalışmalarında Portekiz'de iki farklı ortaöğretim okulunda 2005 ve 2006 yıllarına ait matematik dersi öğrenci verileri kullanılarak öğrenci başarısının tahminini incelemişler [6]. Çalışmalarında öğrenme analitiği verilerinde Bayes optimizasyonu ve Destek vektör makine ile En yakın komşu yöntemlerini kullanmış ve Destek vektör makinesinin

doğruluk skoru %64.52, en yakın komşu algoritmasının doğruluk skoru %57.51 olarak elde edilmiş. Çalışmada elde edilen başarımın artırılması için diğer optimizasyon ve veri madenciliği tekniklerinin ilerdeki çalışmalarda dikkate alınması gerektiği vurgulanmıştır.

Namoun vd. 2010-2020 yılları arasında elektronik dergilerde yayınlanan 62 makaleyi çalışmalarında incelemişler [8]. Çalışmalarında veri madenciliği ve öğrenme analitiği yöntemlerini kullanmış ve değerlendirme öğrenme çıktılarının tahmini, öğrencilerin öğrenmesini tahmin etmek için geliştirilen analitik modeller ve öğrencinin başarısını etkileyen baskın faktörler olarak üç başlıkta yapılmış. Çalışmada öğrenme çıktılarının tahmininde başarım oranı en yüksek beş ve en kötü beş makine öğrenmesi algoritması listelenmesi, gelecekteki yapılacak çalışmalar için bir fikir verecektir. Başarım oranında öne çıkan ilk beş algoritma içinde yer alan çalışmalar [9] [21] [22] aşağıda incelenmiştir.

Al-Shabandar vd. çalışmalarında Harvard University'nin çevrimiçi öğrenme ortamı günlüklerinden, öğrencinin başarısı ile ders etkinliklerine erişimi arasındaki ilişkiyi incelemişler [9]. Çalışmalarında istatistiksel yöntemler ile sınıflandırmada Rastgele orman, Destek vektör makinesi ve Sinir ağı yöntemlerini kullanmışlar. Çalışma sonuçlarında, öğrenme ortamı kaynaklarına erişimin (tıklama) başarı ile ilintisinin öznelitellere göre %57 ile %64 arasında olduğu görülmektedir. Sınıflandırma doğruluk başarım oranında ise RF %98 ile en başarılı yöntem olarak değerlendirilmiştir. Gelecekteki çalışmada, pasif katılım sağlayan öğrencilerin başarısının tahmin ve sınıflandırma çalışmaları yapılacağı belirtilmiştir.

Zaporozhko vd. çalışmalarında çevrimiçi öğrenme ortamından elde edilen öğrenme analitiği verilerinden başarısız olacak öğrencilerin erken tespitini incelemişler [21]. Öğrenme analitiği verilerini, sinir ağı yaklaşımına dayalı çok katmanlı algılayıcı yöntemi kullanarak başarısız olacak öğrencileri erken tahmin etmişler. Çalışmada başarısız olacakların erken tespitinin, öğrenciye özelleştirilmiş eğitim ile başarımın artırılabilceği vurgulanmış. Bu veriler doğrutulusunda eğitim ortamının dinamik olarak tasarlanmasının önemi üzerinde durulmuş. Gelecekte öğrenme ortam verileri üzerinde yapılacak çalışmaların karar destek sistemlerine destek olabileceği vurgusu yapılmış.

Sokkhey vd. çalışmalarında özel bir üniversitenin 56.000 öğrenci kaydını, öğrenci başarısının erken tahmini için incelemişler [22]. Çalışmalarında eğitimsel veri madenciliği teknikleri ve Rastgele orman ile k-En yakın komşu yöntemlerini kullanmışlar. Öğrenci başarısının erken tespitinde özel tasarlanmış RF yönteminde doğruluk başarım oranı %99,72 olarak bulunurken, k-NN yönteminde doğruluk başarım oranı %95,95 elde edilmiş. Başarısı düşük öğrencinin erken tespitinin, eğitimsel veri madenciliği ile entegre web tabanlı bir öğretim programının öğrenci başarısını artıracağı yönünde vurgu yapılmış.

2.2. Makine Öğrenmesi Literatürü

Üniversitelerde çevrimiçi öğrenme ortamlarda yürütülen derslerin etkinliklerine (ders notları, videolar, kaynaklar vb.) öğrencilerin erişimleri analizler için önemli veri oluşturmaktadır. Bu verilerin değerlendirilmesi bazen ders bazında bazen yarıyıl bazında olabiliyor [23] [24] [25]. Çevrimiçi öğrenme ortamlarından öğrenci ve kayıt olduğu dersler ile ilgili tutulan bilgilerden öğrencinin başarısının tahmini ve sınıflandırılması ile birçok çalışma yapılmıştır [26] [27] [28]. Yapılan çalışmalarından bazılarının kullandığı makine öğrenmesi algoritmaları ve sonuçları ile ilgili bilgiler aşağıda verilmiştir.

Kadoic vd. çalışmalarında Zagreb Üniversitesi'nde 73 öğrencinin ders bazında etkinliklere ne zaman katıldığı ile başarısı arasındaki ilişkiyi incelemişler [10]. Çalışmalarında Moodle günlüklerinin istatistiki (anova, korelasyon vb.) yöntemleri kullanarak analizi yapılmış ve öğrencinin etkinliklere sınavdan kısa zaman önce eriştiği bilgisi grafik olarak verilmiş. Çalışmada kız öğrencilerin etkinliklere katılımının daha fazla olduğu ve buna bağlı olarak başarılarının da aynı oranda yükseldiği görülmüş. Daha fazla ders ve etkinlik sayısı ile çalışmanın genişletilmesi farklı çıkarımlar hakkında fikir verebilir.

Bhutto vd. çalışmalarında Kalboard çevrimiçi öğrenme ortamı verilerinden 500 öğrencinin verilerini kullanarak öğrenci başarısının sınıflandırmasını incelemişler [23]. Çalışmalarında sınıflandırma Lojistik regresyon (Logistic Regression-LR) ve SVM yöntemleri ile yapılmıştır. Çalışmada üçlü sınıflandırma (0-69 (bad), 70-89 (average), 90-100 (good) yapılmış ve en iyi doğruluk başarımları oranı %79,3 ile SVM olurken LR %73,4 olarak elde edilmiştir. Öğrenci başarısının sınıflandırmasının yanı sıra başarıda etkin olan özelliklerin seçimi ile öğrenme ortamının etkili biçimde öğretilmesinin öğrenci motivasyonunu artıracak ve başarı oranını artıracak öngörülmektedir. Gelecekte çevrimiçi öğrenme ortamlarının daha da yaygınlaşacağı ve başarının tamamen öğrenciye bağlı olduğu vurgusu öne çıkmaktadır. Bu nedenle öğrenme ortamı ile öğrencinin etkileşimi ve ihtiyaca göre yeniden tasarlanması öğrenci başarısının tahmininde önemli rol oynayacaktır. Gelecekte farklı sınıflandırma yöntemleri ve özellik seçimi kullanılarak başarının artırılacağı ifade edilmektedir.

Taranan makalelerde çevrimiçi öğrenme ortam verileri üzerinde yapılan sınıflandırma çalışmalarının [29] [30] [31] [32] yanı sıra yüz yüze yapılan ders verileri ile üniversite veri tabanlarından alınan kayıtlar üzerinden de çalışma yapılmış [33] [34] [35] [36] [37]. Çevrimiçi öğrenme ortamı verileri üzerinde yapılan çalışmaların bazılarının sonuçları aşağıda sunulmuştur.

Bognar vd. çalışmalarında Macaristan'daki Dunaujvaros Üniversitesi'nde çevrimiçi öğrenme ortamı olan Moodle üzerinde yürütülen bir derse kayıtlı 57 öğrencinin başarılarının sınıflandırmasını incelemişler [27]. Çalışmalarında Moodle etkinlikleri olarak tanımlanan sınav, ders notları ve kitaplara erişimi ile Moodle ortamı dışından erişim sağlanan video, minitab, pdf vb. etkinliklerine erişimlerinden oluşan veri kümesini kullanmışlar. Çalışmada öğrenci başarısının sınıflandırmasında veri kümesi üzerinde eğitim, test ve doğrulama için 5 kat çapraz doğrulama LR yöntemi ile beraber kullanılmıştır. Moodle etkinliklerine erişimde başarının sınıflandırılmasında LR yöntemi %84 olur iken Moodle dışı etkinliklere erişimde aynı yöntem %91 doğruluk başarımları ile en yüksek sonucu elde etmiştir. Moodle dışı etkinliklerin (video, minitab vb) sistem içerisine bütünleşik olarak tanımlanabilmesi öğrenme ortamının kullanımını yaygınlaştıracaktır diye ifade edilmiş. Bu sayede şu an için Moodle siteminde belirlenebilen risk altındaki öğrencilerin erken tespiti daha etkin kullanabilecektir vurgusu yapılmış. Gelecekte benzer kurslarda başarısız olabilecek öğrencilerin erken tespiti çalışmalarının yapılacağı söylenmiş.

Purwoningsih vd. çalışmalarında Endonazya'daki Turbeka Üniversitesinde iki farklı dönemden 1710 öğrencinin moodle üzerinden eğitim içeriğine erişimlerini ve başarısını incelemişler [31]. Çalışmalarında sayısal alanların normalizasyonu (minmax) ve kategorik alanların temsili teknikleri ile beş farklı makine öğrenmesi yöntemi kullanılmış. Çalışmanın sonucunda değerlendirmede kullanılan 5 yöntemden, destek vektör makinesi doğruluk başarımında daha baskın görünmüş. Bu sonuçların farklı dersler ile doğrulanmasının ve eğitim içeriklerinin niteliği artırılarak öğrenci başarısının artırılacağı vurgulanmış.

İngiltere'deki Open University çevrimiçi kurslarının veri kümesini [17] araştırmacıların kullanımına açmıştır. Veri kümesi üzerinde yapılan çalışmaların bazıları kurs ve dönem bazında [32], bazıları da bütün veri kümesi üzerinde olmuş [36].

Open University'de dersler yılda iki kez açılıyor. Şubat ayında açılan dersler, açıldığı yılın sonuna 'j' eklenerek (2013j), ekim ayında açılanlar derslerde ise yılın sonuna 'b' eklenerek (2013b) gösterilirler. Veri kümesi üzerinde literatürde makine öğrenmesi algoritmaları ile yapılan çalışmalardan iki tanesi incelenmiştir.

Ruiz-Rodriguez vd. çalışmalarında 2013 ve 2014 yıllarında açılan derslerden bir tanesi üzerinden başarı sınıflandırmasını incelemişler [32]. Çalışmalarında ikili ve dördü sınıflandırma yaparak rastgele orman ve sinir ağı yöntemlerini kullanmışlardır. Dersin 2013 verilerini eğitim, 2014 verilerini test olarak kullanmışlar. İkili sınıflandırmada rastgele orman %91, sinir ağı %92 doğruluk başarımları elde etmiş. Dördü sınıflandırmada ise rastgele orman %70, sinir ağı %72 doğruluk başarımları elde etmiş. Çalışma sonucu olarak;

sınıflandırma sayısı arttıkça başarı oranının düşeceği bilgisi çıkarılabilir. Gelecekteki çalışmalarda sınıflandırma tipleri, etkinlik türleri vb. değiştirilerek çalışmalar yapılabileceği vurgulanmış.

Alshabandar vd. çalışmalarında açık üniversite veri kümesi üzerinde öğrenci başarısının tahmini için regresyon hata ve sınıflandırma metriklerini incelemişlerdir [36]. Çalışmalarında üçlü sınıflandırmayı k kat (fold) 10 alarak, 6 farklı yöntemi kullanmış ve en iyi başarı sonucunu %86.8 ile Gradyan artırması (GBM) yöntemi ile elde etmişler. Davranışsal öğrenme ortamı verileri 6 farklı zaman dilimine bölünmüş ve regresyon hata metriği olarak Kök ortalama kare hata (root mean square error-RMSE) yöntemi kullanılmış. En iyi sonucu üçüncü zaman diliminde rasgele orman yöntemiyle 8,131 olarak elde etmişler. Regresyonda 6 farklı zaman diliminde hata metriğinin hesaplanması iyi olmakla birlikte sınıflandırmada dersi bırakan öğrencilerin final notunun eksikliğinin etkisi üzerinde durulmamıştır.

Makine öğrenmesi yöntemlerinden olan derin öğrenme (deep learning) algoritmaları 2006 yılından itibaren uygulanmaya başlanmıştır [38] [39]. İletişim teknolojilerinin artması ve internetin yaygın kullanımı ile oluşan veri kümeleri (kelime, fotoğraf, video vb.) üzerinde gelişmiş derin öğrenme algoritmaları sınıflandırma, tıbbi görüntü analizi, tanıma, tahmin etme vb. işlemleri için yıllar içerisinde farklı alanlarda kullanılmaya başlamıştır [40] [41] [42]. Derin öğrenme algoritmaları ile farklı veri kümeleri üzerinde, farklı sınıflandırma [13] [14] [15] ve zaman serileri kullanılarak öğrenci başarısını tahminine yönelik birçok çalışma yapılmıştır [16] [19] [20] [43]. Bu çalışmalardan bazılarının kullandığı derin öğrenme algoritmalarının sonuçları aşağıda değerlendirilmiştir.

Malekian vd. çalışmalarında Melbourne Üniversitesi'nde çevrimiçi öğrenme ortamından 2013 (51306 öğrenci) ve 2014 (33975 öğrenci) dönemlerine ait öğrenci ve ders bilgileri üzerinde zaman aralıklarına bağlı olarak başarı tahmini incelemesi yapılmış [43]. Çalışmalarında öğrencinin başarısını tahmin etmek için veri kümesi zaman aralıklarında gruplandırılarak, geleneksel makine öğrenmesi algoritmalarından Lojistik regresyon (LR) ve MLP'nin yanı sıra derin öğrenme algoritmalarından Özyinelemeli sinir ağı (RNN) yapısının özel bir kullanımı olan Uzun-kısa vadeli bellek (LSTM) yöntemi kullanılmış. En iyi doğrulama skorunu %87.1 ile LSTM verirken LR %83.3 ve MLP %83.7 değerlerini almış. Gelecekteki çalışmada farklı zaman aralıkları ve farklı gruplandırmalar ile yeni derin öğrenme mimarileri ile başarımın artacağı söylenmiş.

Uliyan vd. çalışmalarında Suudi Arabistan'da bulunan Ha'il Üniversitesi'nin Bilgisayar Bilimleri Bölümü'nün 1149 öğrencisinin verilerinden başarımın tahmini ve sınıflandırmasını incelemişler [20]. Çalışmalarında öğrenci başarımının tahmini ve sınıflandırması için derin

öğrenme algoritmalarından çift yönlü uzun kısa zaman belleği (Bidirectional Long Short - Term Memory- BLSTM) yöntemini koşullu rastgele alanlar (Conditional Random Fields-CRF) ile beraber kullanmışlar. BLSTM yöntemi ile CRF %90 başarımlı oranı elde ederken, RF ve LR %93 başarımlı oranı elde etmişler. Öğrencinin başarı tahmini ve sınıflandırması yapılırken kullanılan niteliklerin önemine ve öğrenci demografik bilgilerine işaret edilmiş. Gelecekte öznelitlikler üzerinde farklı normalizasyon işlemleri ile çalışma yapılabileceği vurgulanmış.

Karimi vd. çalışmalarında İngiltere'deki Open University veri kümesi [17] üzerinde 2013 ve 2014 yıllarında açılan 22 derse kayıtlı 32,595 öğrencinin ikili ve dörütlü başarı sınıflandırmalarını incelemişler [44]. Çalışmalarında veri kümesinde her iki dönemde açılan dersin bir tanesini eğitim değeri test amaçlı öğrencinin etkinliklere haftalık katılımı üzerinden kullanmışlar. Çalışmada sınıflandırma da ESA, LR ve SVM yöntemi ikili ve dörütlü sınıflandırmada *f1* değerlendirme ölçütü ile kullanılmış. Haftalık etkinliklere erişim değerlendirmelerde ikili ve dörütlü sınıflandırmada ESA yönteminin öne çıktığı görülmüştür. Gelecekte sosyal medya verileri üzerinden sınıflandırma çalışmalarının yapılacağı ifade edilmiştir.

Taraması yapılan ve incelenen literatür çalışmalarında veri kümesi üzerinde normalizasyon, öznelitlik seçim yöntemi ile kategorik değişkenlerin sayısallaştırılmasında label encoding ve one hot encoding yönteminin kullanıldığı görülmüştür. Bu çalışmada veri kümesinde kategorik değişkenlerin yanı sıra öznelitlik seçimi yapılan bütün alanlarda one hot encoding yöntemi ile ESA mimarine dayalı yöntem beraber kullanılmıştır.

3. TEMEL BİLGİLER

Bu bölümde tezde kullanılan veri kümesinin elde edildiği çevrimiçi öğrenme ortamı, öznitelik seçimi, sınıflandırma ve regresyon başlıkları altında katalog bilgisi verilecektir.

3.1. Çevrimiçi Öğrenme Ortamı

Çevrimiçi öğrenme ortamları, öğrencinin internet üzerinden yürütülen ders etkinliklerine (ödev, sınavlar, ders dokümanları, anket, tartışma grupları vb.) katıldığı web tabanlı platformlar olarak görülür [45]. Bu ortamlar üzerinden yürütülen ders eğitimlerinin etkili olabilmesi için, ortam özellikleri ile beraber düşünülmelidir. Çevrimiçi öğrenme ortamları üzerinde yürütülen derse göre farklı özellikler sunsa da bu özellikleri; kayıt işleminden, eğitim içeriklerinin paylaşımı, eğitim sürecinin takibi, kişiselleştirme, ödev, sınavlar, canlı dersler, duyuru, raporlama ve analiz ana başlıkları olarak alabiliriz.

Çevrimiçi öğrenme ortamını, öğrenme yönetim sistemleri, içerik yönetim sistemleri ve sanal sınıf yazılım bileşenleri oluşturur. Öğrenme yönetim sistemlerini, ücretli ticari (Blackboard, Google Classrom vb) ve açık kaynak kodlu ücretsiz yazılımlar (Moodle, Chamilo, ATutor vb.) olarak iki kategoride incelenebilir. Ücretli yazılımlardan olan Blackboard; öğrencilerin eğitimde ilerlemesini raporlaması, mesaj ve faklı işletim sistemi desteği avantaj olarak görülürken, düz ara yüzü ve çok menü oluşturması dezavantaj olarak görülür. Ücretsiz yazılımlardan olan Moodle, öğretici bol eğitim ve materyali ortamı sunması, çoklu dil desteği ve toplu ders oluşturma özellikleri ile öne çıkarken, yerel video desteğinin olmaması ve ara yüzü konuları dezavantaj olarak değerlendirilebilir.

Başkent Üniversitesi'nde çevrimiçi öğrenme ortamı olarak Moodle sunulmaktadır. Haftalık olarak ders izlencesinde belirtilen konular sistem üzerinde tanımlamaya uygundur. Belirtilen konu başlıklarının eğitim süresine uyumlu, haftalık etkinlik olarak tanımlanabilmesi bir avantaj olarak görülmektedir. Sistem üzerinde tanımlanan bu etkinliklere öğrencilerin erişimlerinin, günlüklerden elde edilebiliyor olması Moodle sistemini öne çıkarmaktadır.

İncelenen literatürde, eğitim kurumlarında yaygın olarak ücretsiz açık kaynak kodlu olan Moodle kullanıldığı görülmüştür. Öne çıkan özelliklerinin yanı sıra, Başkent Üniversitesi'nde de kullanılması Moodle sisteminin tercih edilmesinde etkin olmuştur.

3.2. Öznitelik Seçimi

Öznitelik Seçimi (feature selection), orijinal veri kümesini temsil edebilecek en iyi altkümenin seçimi olarak tanımlanmaktadır. Öznitelik seçimi (diğer adıyla nitelik seçimi veya değişken seçimi), kullanılan algoritmaya göre özellikleri değerlendirerek veri kümesindeki n adet özellik arasından en iyi k adet özelliği seçme işlemidir [46].

Öğrencilerin başarısını etkileyen birçok faktör bulunmaktadır. Bunların bazıları kontrol altına alınabilir, ölçülebilir faktörlerdir. Bu faktörlerin, öğrenci başarısını hangi düzeyde etkilediğinin bilinmesi gerekmektedir. Bu yolla veri kümesindeki tüm öznitelikler yerine, veri kümesini daha iyi temsil edebilecek daha az sayıda öznitelik belirlenebilirse hem regresyon hem de sınıflandırma yöntemlerinin tahmin başarımları artabilir [47].

Öznitelik seçim yöntemlerini; filtreleme (*filter*), sarmal (*wrapper*) ve gömülü (*embedded*) başlıkları olmak üzere üç grupta incelemek mümkündür. Filtreleme yöntemlerinde, özniteliklerden istatistiki bilgiler alınabilirken, sarmal yöntemlerde öznitelikler üzerinde arama işlemleri gerçekleştirilir ve son olarak gömülü yöntemlerde ise en iyi bölen ölçütünü bulan bir yapı kullanılmaktadır [46].

Filtreleme yöntemleri istatistiksel sonuçları elde ederken, uzaklık, bağımlılık, tutarlılık ve bilgi ölçümlerini veren fonksiyonları kullanır. Bu fonksiyonlar aracılığı ile bir değer (puan) hesaplanır. Hesaplanan bu değerler içinden en yüksek değerlere sahip öznitelikler, veri kümesini temsil etmek için seçilir.

Bu çalışmanın amacı doğrultusunda çevrimiçi öğrenme ortam verileri ile öğrenci başarısının ilintisi incelenmiştir. Başarının belirlenmesinde, elde edilen çevrimiçi ortam verilerinden, filtreleme yöntemleri ile öznitelik seçimi yapılmıştır. Çalışmada kullanılan öznitelik seçim yöntemleri hakkında bilgiler aşağıda verilmiştir.

3.2.1 İlinti (Korelasyon)

İki sayısal ölçüm arasında doğrusal bir ilişki olup olmadığını, varsa bu ilişkinin yönünü ve şiddetinin ne olduğunu belirlemek için kullanılan istatistiksel bir yöntemdir. İlinti katsayısı iki değişken arasındaki doğrusal ilişkinin ölçüsü olup incelenen değişkenlerin birimlerinden bağımsızdır ve $-1 \leq r \leq 1$ arasındadır. İlinti katsayısı aşağıdaki formül ile hesaplanır:

$$r = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sqrt{(\sum X_i^2 - n \bar{X}^2)(\sum Y_i^2 - n \bar{Y}^2)}} \quad (3.1)$$

Burada X serisinin ortalaması $\bar{X}$, Y serisinin ortalaması $\bar{Y}$ ve n gözlem sayısıdır.

3.2.2 Bilgi Kazancı (information gain)

Bilgi kazancı, hedef değişkenin sınıfının tespitinde en ayırt edici özneliği belirlemek için her öznelik için ölçülen istatistiksel bir özelliktir. Her özneliğin bilgi kazancı ölçülürken entropi (E) modeline dayalı bir yapı kullanılır. Veri kümesindeki kayıtların (K) bir s sınıfına ilişkin entropisi [48] aşağıdaki şekilde hesaplanır:

$$Entropy(K) = E = \sum_{i=1}^s -p_i \log_2 p_i \quad (3.2)$$

Burada p_i , K veri kümesindeki i sınıfının oranıdır.

Veri kümesindeki her bir kaydın (K), özneliklerinin (N) bilgi kazancı (BK) aşağıdaki formül ile hesaplanır:

$$BK(K, N) = Entropy(K) - \sum_{v \in Values(N)} \frac{|K_v|}{|K|} Entropy(K_v) \quad (3.3)$$

Burada v , özneliklerin (N) bir elemanını ifade eder.

3.2.3 Fisher Puanı

Fisher Puanı yöntemi, her bir sınıf için özelliklere ait ortalama ve standart sapma değerlerini kullanarak bir ilişki puanı aşağıdaki formüle göre hesaplanır:

$$F(X_i) = \frac{|\mu_i^+ - \mu_i^-|}{\sigma_i^+ - \sigma_i^-} \quad (3.4)$$

Formülde bulunan + ve – işaretleri iki sınıflı bir problem için farklı sınıfları ve değerleri sınıfların aritmetik ortalamalarını ve değerleri sınıflara ait standart sapma değerlerini göstermektedir.

İki sınıfı birbirinden ayırmaya yardımcı olacak özellikler bulunmaya çalışan bu yöntem ile özellik seçim işlemi, özelliklerin hesaplanan skorlara göre büyükten küçüğe doğru sıralanmasıdır.

3.2.4 SelectKBest

Sınıflandırma problemlerinde, öznitelik sayısı (k değeri) arasından hedef değere uyumlu en yüksek puanı alanı belirlemede kullanılır.

Ki kare testi kategorik değişkenler arasındaki ilişkiyi ölçmek için kullanılır. Ki-kare testini kullanarak, kategorik özellik ile hedef değişken arasında ilişkiyi ölçtükten sonra en iyi k kadar (SelectKBest) özelliği seç diyebileceğimiz hazır fonksiyonlar vardır. Ki kare (X^2) aşağıdaki formüle göre hesaplanır:

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (3.5)$$

Burada O gözlemlenen sıklık sayısını, E ise beklenen sıklık sayısını ifade etmektedir.

3.2.5 Öznitelik Önemi (Feature Importance)

Denetimli öğrenmede filtreleme yöntemlerinden bir tanesi de özneliliğin önem derecesidir. Makine öğrenmesi metotlarında hedef değişkenin belirlenmesinde diğer özniteliklerin önem derecesini verir.

Öznitelik önemi dediğimiz kavram, verilerinizin her özneliliği için bir puan verir, puan verilen özneliliğe yakın veya ilgili olan özniteliklere hedef değişkeninize yönelik olarak puanlama ile devam eden bir yapıdır. Ne kadar ilgili ise öznitelik puanı o kadar yüksek olur. Öznitelik önemi (ÖÖ) aşağıdaki formüle göre hesaplanır:

$$\text{ÖÖ}_i = \frac{\sum_j \text{norm} \text{ÖÖ}_{ij}}{T} \quad (3.6)$$

Burada i indisi modeldeki tüm ağaçlardan hesaplanan öznitelik önemini, normÖÖ normalize edilmiş özellik önemini T ise modeldeki toplam ağaç sayısını ifade etmektedir.

3.3. Regresyon ve Sınıflandırma

Bu çalışmada kullanılacak modellerin başarımını değerlendirmek için regresyon ve sınıflandırma ölçütleri kullanılmıştır. Başarım ölçütlerinde, sınıflandırma karar ağacında, bağımlı değişkenler kategorik iken regresyonda bağımlı değişkenler sayısaldir. Regresyon ve sınıflandırma ölçütlerin tanımı ayrı ayrı başlıklarda verilmiştir. Bu başarım ölçütlerinde

kullanılan Rastgele orman, Çok katmanlı algılayıcı, k-En yakın komşu ve Evrişimsel sinir ağı yöntemleri hakkında bilgi ise regresyon ve sınıflandırma algoritmaları başlığında verilmiştir.

3.3.1 Regresyon

Regresyon problemlerinde, makine öğrenmesi algoritması n adet öznitelik girişine göre sayısal bir değeri tahmin eder [12]. Regresyon problemlerinin çözümünde, algoritma belirli sayıdaki özniteliği girdi olarak alır ve aşağıdaki fonksiyon ile bir sayısal değer elde eder:

$$f : \mathbb{R}^n \rightarrow \mathbb{R} \quad (3.7)$$

Burada n öznitelik sayısını ifade eder.

Makine öğrenmesi algoritmaları sayısal değer tahminini belirli bir hata oranı ile yukarıdaki fonksiyon aracılığı ile yapar.

3.3.2 Sınıflandırma

Sınıflandırma görevlerinde, makine öğrenmesi algoritması n tane özniteliğin girişine göre sınıfının, k tane sınıf içerisinde aşağıdaki fonksiyon ile tahmin edilmesidir [12].

$$f : \mathbb{R}^n \rightarrow \{1, \dots, k\} \quad (3.9)$$

Burada n öznitelik sayısını, k sınıf sayısını ifade eder.

Öğrenen algoritmada, x adet öznitelik için y sayısal değeri ile sınıfı $y=f(x)$ fonksiyonu ile tanımlanır.

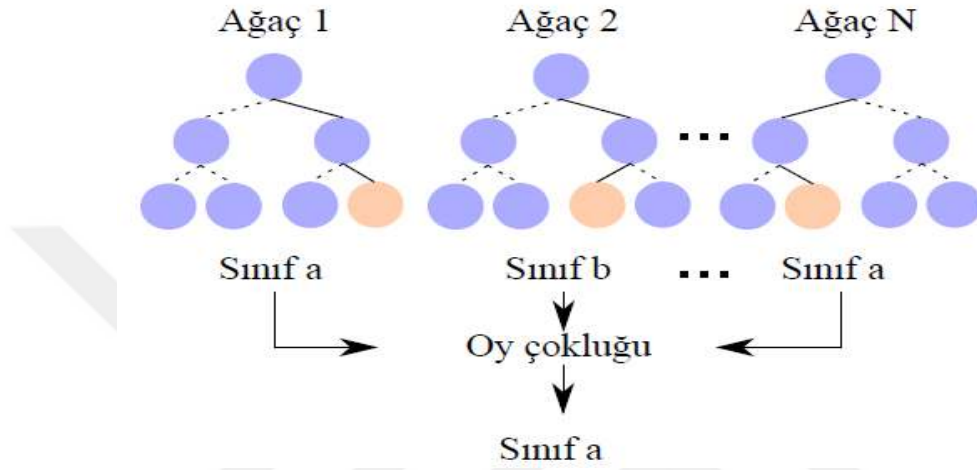
3.3.3 Regresyon ve Sınıflandırma Algoritmaları

Bu çalışmada regresyon ve sınıflandırma için Rastgele orman, Çok katmanlı algılayıcı, Evrişimsel sinir ağı ve k-En yakın komşu ve yöntemleri kullanılmıştır. Kullanılan yöntemler hakkında bilgilendirme aşağıda verilmiştir.

3.3.2.1. Rastgele Orman

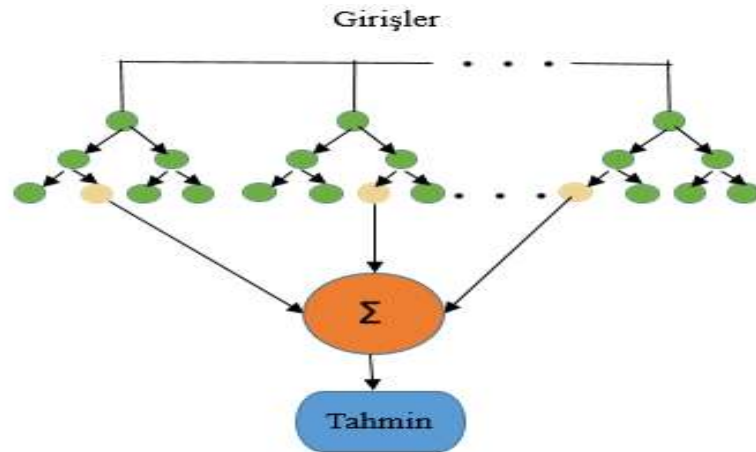
Sınıflandırma, regresyon ve diğer görevler için, eğitim aşamasında çok sayıda karar ağacı oluşturarak problemin tipine göre sınıf veya sayı tahmini yapan bir toplu öğrenme yöntemidir. Sınıflandırma yönteminde her ağaçta değişkenlerin farklı olmasından dolayı uyum sorunu pek oluşmaz.

RF yöntemi, birleştirilmiş karar ağaçları ile sınıflandırma yapan bir algoritmadır. Veri kümesindeki bir kayıt için algoritmada yer alan her bir ağaçta sınıflandırma işlemi yapılır. Her bir ağacın yaptığı oylama ile sınıf belirlenir. Diğer bir ifade ile her bir ağaç bir sınıf tahmini yapar ve en çok oyu alan sınıf, yöntemin sonucu haline gelir. RF yönteminde birden fazla karar ağacından elde edilen sonuçlar bir araya getirilerek güvenilir tek bir sonuç bulunur [32] [47]. RF yönteminin N tane ağaç ile yapısı Şekil 3.2’ de gösterilmiştir.



Şekil 3.1. RF sınıflandırma yapısı [32]

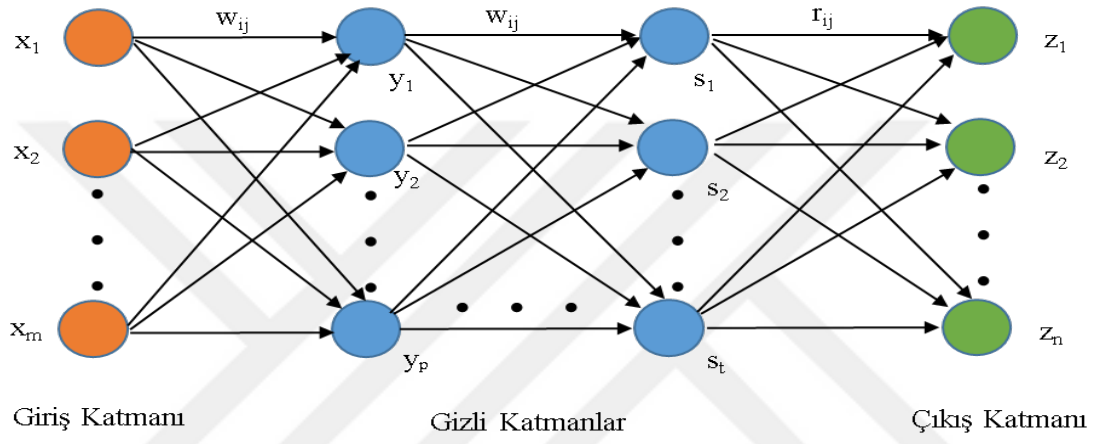
Regresyon probleminin sayısal tahmini için çok sayıda üretilen ağacın her birisinin tahmini toplanarak belirlenen en optimal tahmin sonucu Şekil 3.1’ de gösterilmiştir.



Şekil 3.2. RF regresyon yapısı

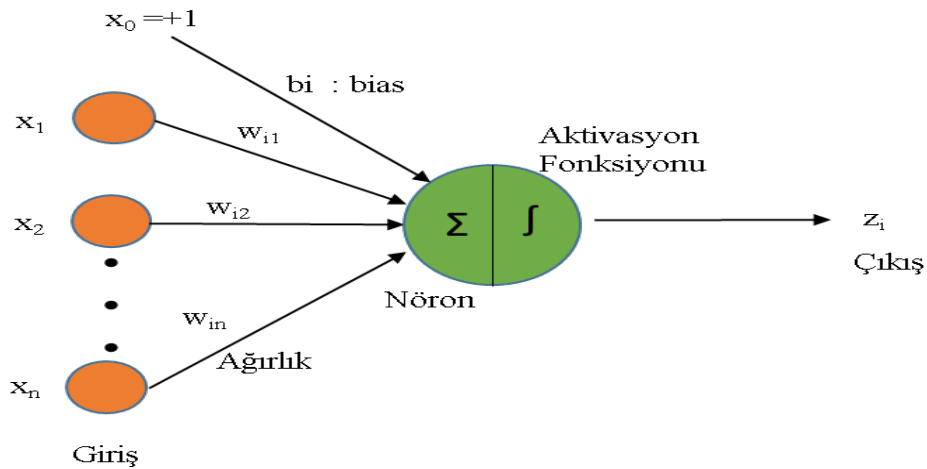
3.3.2.2. Çok Katmanlı Algılayıcı

Çok katmanlı algılayıcı, bir dizi girdiden bir dizi çıktı üreten bir ileri beslemeli yapay sinir ağıdır. MLP'nin yapısı giriş katmanı, çıkış katmanı ve gizli katman olmak üzere üç tip katmandan oluşur ve Şekil 3.3' de gösterilmiştir. İşlenecek öznitelikler giriş katmanından alınır, sınıflandırma ve tahmin gibi görevler çıktı katmanı tarafından gerçekleştirilir. Giriş ve çıkış katmanı arasına yerleştirilen rastgele sayıdaki gizli katman, MLP'nin hesaplama motorudur. Veriler girişten çıkış katmanına doğru ilerler ve nöronlar geri yayılım öğrenme algoritması ile eğitilir [49].



Şekil 3.3. MLP'nin gizli katmanlı yapısı

MLP'nin yapısında her bir nöronda girişler ağırlıkları ile çarpılıp toplanır ve aktivasyon fonksiyon değeri elde edilir. Bu değer sonraki katmana giriş olarak verilir. Her bir nöron için yapılan hesaplama Şekil 3.4' de verilmiştir.



Şekil 3.4. MLP nöron çıkışının hesaplanması

Nöronların (n) girdilerinin ağırlıklı (w) toplamı hesaplanarak, sapma değeri (b :bias) eklenir. Hesaplanan bu değer bir aktivasyon fonksiyonundan (f) geçirilir. Her katmandaki nöronlarda hesaplama işlemi aşağıdaki formül ile yapılır:

$$z_i = f(n_i) = f\left(\sum_{j=1}^n w_{ij}x_j + b_i\right) \quad (3.10)$$

Regresyon problemlerinde, giriş katmanı her biri bir öznitelığe karşılık gelen çok sayıda nöron içerir, her gizli katman en az iki adet nörondan oluşur ve çıktı katmanı ise tahmin edilen değere karşılık gelir. Bu yöntemde ileri doğru bir iletimle, hatayı en az tutmak için hesaplamaları geriye doğru algoritma ile yapmaktadır. Bu işlem gerçekleşirken, giriş katmanında her bir öznitelik (g) ile ağırlık değerleri (a) çarpılarak toplanır ve her katmanda ağırlık değerleri güncellenerek, hatanın en az olması amaçlanır [50]. Çarpım sonucu elde edilen her katmanın çıktısı, bir sonraki katmanın girdisi olarak kabul edilerek ağırlıklı toplam bir aktivasyon fonksiyonundan (f) geçirilerek çıkış değeri (CD) aşağıdaki şekilde hesaplanır:

$$CD = f\left(\sum_i a_{ip} g_i\right) \quad (3.11)$$

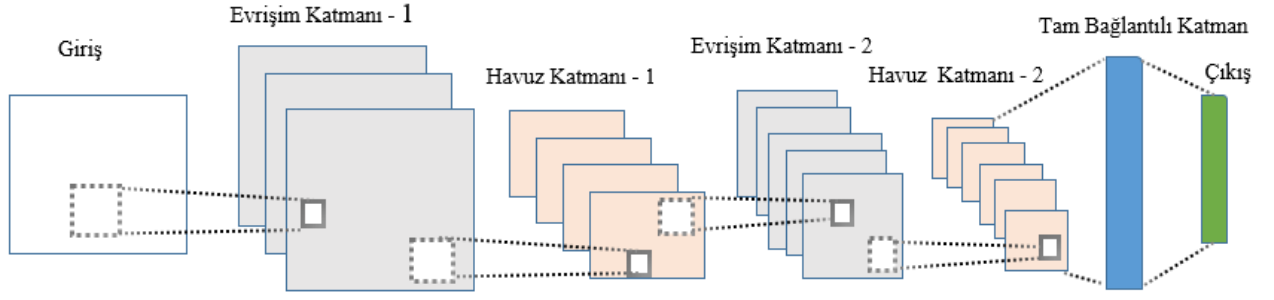
Buradan elde edilen çıkış değeri (CD) ile hedeflenen çıkış değeri (HCD) farkı bize yapılan hatayı verecektir.

3.3.2.3. Evrişimsel Sinir Ağı

Evrişimsel ağlar, ilk defa 1989 yılında LeCun tarafından tanımlanan ve Evrişimsel sinir ağı (ESA) olarak ifade edilen, tek boyutlu zaman serileri ve iki boyutlu görsel verilerini işlemek için kullanılan özel yapay bir sinir ağıdır. Bu ağda, *evrişim* adı verilen matematiksel bir işlem kullanılır. Katmanların en az birinde, matris çarpımı yerine bu *evrişim* kullanılır [12]. Geleneksel evrişimsel sinir ağı giriş, evrişim (*convolution*), havuzlama (*pooling*), tam bağlantı (*full connected*) ve çıkış katmanlarından oluşur. Genellikle görsel bilginin analiz edilmesinde kullanılır.

Giriş değerleri, evrişim katmanında, belirli adımlarla filtrelendikten sonra aktivasyon fonksiyonu ile öznitelik haritası oluşturulur. Her evrişim katmanından sonra hesaplamayı kolaylaştırmak için havuz katmanı kullanılır. Havuz katmanında, bir katmandaki nöronların hesaplanan değerlerini, sonraki katmandaki tek bir nörona birleştirerek verinin boyutunu azaltıp bir nevi örnekleme işlemi yapar. Havuzlamada maksimum ve ortalama olmak üzere iki

tür kullanılır. İlk yöntemde geçerli matrisin maksimum değerini seçerek algılanan özellikler korunur. İkinci yöntemde ise geçerli matris değerlerinin ortalaması alınır ve boyut azaltarak öznitelik haritası örneklenmiş olur. Tam bağlantılı katman genelde evrişim ağının sonunda ve hedeflenen sınıf değerlerinin optimize edilmesinde kullanılır. ESA mimarisi Şekil 3.5’ de verilmiştir.



Şekil 3.5. ESA mimarisi

Katmanlardaki (h) hesaplama işlemi, giriş değerleri (x) ve ağırlıkları (w) çarpılarak toplanır ve sapma değeri (b :bias) eklenerek aktivasyon fonksiyonundan (f) geçirilerek aşağıdaki biçimde hesaplanır:

$$h_i = f\left(\sum_i w_i x_i + b\right) \quad (3.12)$$

Algoritmanın ağırlık güncellemesi ve optimizasyonu, gradyanların geri yayılmasıyla yapılır. Tam bağlantılı katmanın *softmax* fonksiyonu ile çıktısı her bir sınıfın olasılığını taşır. En yüksek olasılığa (p) sahip sınıf çıktı olarak kabul edilir ve normalize üstel fonksiyon genellemesiyle aşağıdaki biçimde hesaplanır:

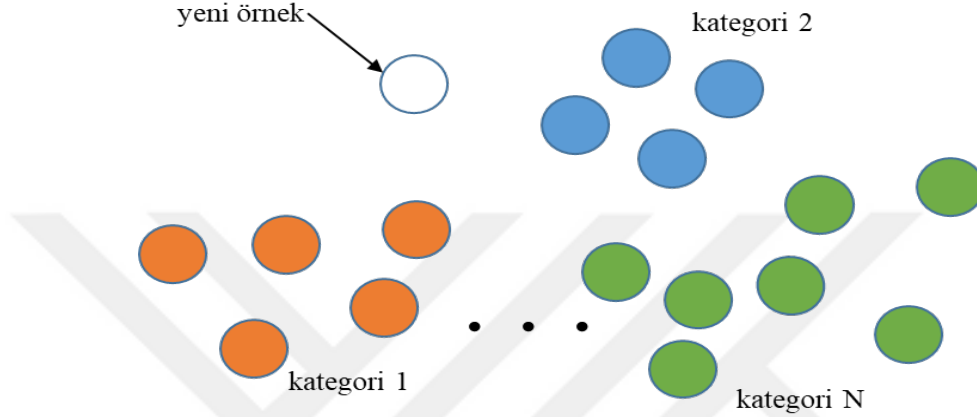
$$p_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \quad (3.13)$$

Fonksiyon, $x \in \mathbb{R}^n$ skorlarının bir vektörünü girdi olarak alır ve p çıkış olasılık vektörünü oluşturur.

ESA mimarisin genelde son katmanı olan tam bağlantılı katmanda hedeflenen değer optimize edilir. Bu katmanda sınıflandırma için *softmax* fonksiyonu kullanılırken, regresyon problemlerinde sürekli bir değeri tahmin etmek için *lineer* fonksiyon kullanılır. Model, sürekli bir değeri kayıp (*loss*) fonksiyonu aracılığıyla, belirli oranda bir hata ile tahmin eder

3.3.2.4. k-En Yakın Komşu

Makine öğrenmesi yöntemlerinden en basit örnek tabanlı sınıflandırma algoritmalarındandır. Sınıflandırmayı veri kümesindeki her bir kaydın diğer kayıtlara uzaklığı üzerinden yapar. Denetimli öğrenme algoritmaları gibi eğitim aşaması yoktur. İlk defa karşılaşılan bir örnek, veri kümesinde yer alan kayıtlara olan benzerliğine göre sınıflandırılmaktadır (kategori) ve Şekil 3.6’ da gösterilmektedir.



Şekil 3.6. k-NN yeni örneğin sınıflandırması

Veri kümesinde yer alan her bir kayıt n boyutlu sayısal alanlar olarak belirtilir. Sayısal olarak belirlenen bu uzayda kayıtlar bir noktayı temsil edecek biçimde tutulur. Bilinmeyen yeni bir kayıt ile karşılaşıldığında veri kümesinden bu yeni kayıta en yakın k adet komşunun sınıf etiketleri çoğunluk oylaması ile yeni kaydın sınıfı belirlenir. Diğer yandan k değerinin seçiminde, belirli bir yöntemi yoktur. Yapılan testlerle ya da sezgisel olarak belirlenir. Mesafe fonksiyonuyla (*euclidean*, *manhattan*, *minkowski*) ölçülen komşuları arasındaki en yakın mesafedeki sınıfa atanır [51]. Mesafe fonksiyonlarından *euclidean*, belirlenen k komşuluğunda, yeni bir örnek geldiğinde komşulara benzerliği aşağıdaki formül ile hesaplar:

$$d(x, y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (3.14)$$

Burada n boyutlu öklid uzayında, x ve y iki noktanın bileşenlerini ifade eder.

Regresyon problemlerinde, bulunan bu uzaklık ile en yakın k komşu ve öznitelik değerlerine göre çıktı değeri tahmin edilir. Çıktı değeri ile hedeflenen değer arasındaki fark regresyonda hata olarak kabul edilir.

3.4. Değerlendirme Ölçütleri

Başarı tahmininde bulunurken oluşan hata, Kök ortalama kare hata (Root Mean Square Error-RMSE) ve Ortalama mutlak hata (Mean Absoulate Error-MAE) ölçütleri ile değerlendirilmiştir. Bu ölçütler ile ilgili açıklama ve formülleri aşağıda verilmiştir.

RMSE: Bir makine öğrenmesi modelinin, tahmin ettiği değerler ile gerçek değerleri arasındaki uzaklığın bulunmasında sıklıkla kullanılan, hatanın büyüklüğünü veren bir ölçektir. RMSE tahmin hatalarının (kalıntıların) standart sapmasıdır.

$$RMSE = \sqrt{\frac{\sum_{j=1}^n e_j^2}{n}} \quad (3.15)$$

MAE: Ortalama mutlak hata iki sürekli değişken arasındaki farkın ölçüsüdür. MAE, her gerçek değer ile veriye en iyi uyan çizgi arasındaki ortalama dikey mesafedir. MAE aynı zamanda her veri noktası ile en iyi uyan çizgi arasındaki ortalama yatay mesafedir.

$$MAE = \frac{1}{n} \sum_{j=1}^n |e_j| \quad (3.16)$$

Sınıflandırma yöntemlerinin başarımını değerlendirirken *doğruluk* (accuracy), *kesinlik* (presicion), *duyarlılık* (recall) ve *f-ölçeği* temel kavramları kullanılmıştır. Doğruluk ölçüm sonuçlarının gerçek değere yakınlığını gösterir, kesinlik ise tekrarlı ölçümlerin birbirine olan yakınlığını gösterir. Duyarlılık ise pozitif olarak tahmin etmemiz gereken işlemlerin ne kadarını pozitif olarak tahmin ettiğimizi gösteren bir ölçüttür. Bu çalışmada üç başarı durumunu içeren bir sınıflandırma problemi üzerinde çalışılmıştır. Sınıflandırma yöntemlerinin başarımı karmaşıklık matrisinden elde edilen değerler (doğruluk, kesinlik vb.) ile ölçülür. Karmaşıklık matrisinde doğru tahmin edilen hedef (artı) sınıf TP, doğru tahmin edilen diğer sınıflar (eksi) TN, yanlış tahmin edilen artı sınıf FP ve yanlış tahmin edilen eksi sınıf FN olmak üzere; *doğruluk*, *kesinlik*, *duyarlılık* ve *f-ölçeği* aşağıdaki formüller ile hesaplanır:

$$Doğruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.17)$$

$$Kesinlik = \frac{TP}{TP + FP} \quad (3.18)$$

$$Duyarlılık = \frac{TP}{TP + FN} \quad (3.19)$$

$$F - \text{ölçeği} = \frac{2xDuyarlılıkxKesinlik}{Duyarlılık + Kesinlik} \quad (3.20)$$

Karmaşıklık matrisinde elde edilen değerler (TP, TN, FP, FN) her sınıf için farklı olsa da doğruluk değeri hep aynı olur.

4. ÖZNİTELİK ANALİZİ

Bu bölüm veri kümesi ön hazırlık ile öznitelik analizi ve seçimi başlıkları altında incelenecektir.

4.1. Veri Kümesi Ön Hazırlık

Bu bölümde işlenecek başlıklar Şekil 4.1’ de iş akışı ile verilmiştir



Şekil 4.1. Veri kümesi ön hazırlık iş akışı

4.1.1. Moodle Günlükleri

2020-2021 Akademik yılın güz ve bahar yarıyıllarından sırasıyla Bilgisayar Donanımı (Bilp103) ve Veri Tabanı (Bilp106) dersleri seçilmiştir. Bu derslerin etkinlikleri duyuru, yardımcı ders kaynakları, ders izlencesi, haftalık ders notları, haftalık çevrimiçi videolar, sınavlar (quiz, ara sınav, final), ödevler olarak Moodle sisteminde tanımlanmıştır.

Derslerin günlükleri (log) Moodle sisteminde rapor bölümünden excel dosyası yapısında alınmıştır. Günlük dosyasının bir örneği Şekil 4.2’ de verilmiştir.

	A	B	C	D	E	F	G	H	I	J
1	Zaman	Kullanıcı	Etkilenen	Etkinlik baş	Bileşen	Etkinlik adı	Açıklama	Köken	IP adresi	
2	2/09/21, 1		T-	Ders: B	Günlükler	Günlük rap	The user with id '20996' \ web		193.14	
3	2/09/21, 1			Ders: B	Sistem	Ders görün	The user with id '20996' \ web		193.1	
4	27/04/21, 1		T-	Ders: B	Sistem	Ders görün	The user with id '20996' \ web		193.14	
5	27/04/21, R			Ders: B	Etkinlik rap	Etkinlik rap	The user with id '20996' \ web		193.1	
6	27/04/21, R			Ders: B	Sistem	Ders görün	The user with id '20996' \ web		193.14	
7	25/04/21, R			Ders: B	Ders katılım	Görüntüler	The user with id '20996' \ web		31.2	
8	25/04/21, R		Un	Ders: B	İstatistikle	Görüntüler	The user with id '20996' \ web		31.	

Şekil 4.2. Örnek günlük dosyası

Günlük dosyasının içerdiği alanlar ve açıklamaları Tablo 4.1’ de verilmiştir.

Tablo 4.1. Ders erişim günlüğü

Alan Adı	Açıklama
Zaman	Hangi tarih ve saatte erişim sağlandığı bilgisi
Kullanıcının tam adı	Sisteme giren kullanıcı
Etkilenen kullanıcı	Yapılan işlemde hangi kullanıcının etkilendiği
Etkinlik bağlamı	Etkinliğin kapsamı
Bileşen	Etkinliğin bileşenleri
Etkinlik adı	Erişilen etkinliğin adı
Açıklama	Kullanıcıların hangi etkinliğe eriştiği bilgisi
Köken	Erişim hangi ortamdan
IP adresi	Erişilen ip numarası

4.1.2. Veri İşleme

Moodle sisteminde her biri farklı bir numara ile otomatik tanımlı ders etkinliklerine (ders notları, çevrimiçi ders videoları, ders izlencesi, duyuru, yardımcı kaynaklar vb.) öğrencilerin hangi gün erişim sağladığı bilgisi Moodle günlük kayıtlarında ‘Açıklama’ alanında yer almaktadır. Ders günlük kayıtları Moodle sisteminden raporlar bölümünden excel dosyası olarak alınmıştır.

Öğrenci sınıf listesi bilgileri (öğrenci numarası, adı soyadı, dönemi, ders kodu ve cinsiyet) ve çevrimiçi ortamda girdiği sınavlar (quiz, ara sınav ve final sınavı) ile ödevden aldığı notların ağırlıklarına göre hesaplanmış halini içeren bilgiler dönem sonu bir excel dosyasında tutulmaktadır.

Günlük dosyası ve dönem sonu oluşturulan dosya üzerinden alan eşleştirmeleri yapılarak elde edilen veri kümesinin örneği Şekil 4.3’ de verilmiştir. Veri kümesi örneği elde edilirken gerçekleştirilen işlem adımları aşağıda belirtilmiştir.

Adım 1: Günlük dosyası ilgili alanlar tiplerine göre tanımlanarak günlük tablosu olarak veri tabanına atılır. Tabloya dönem, ders kodu ve öğrenci numarası alanları eklenir. Dönem sonu oluşturulan dosya ilgili alanlar tanımlanarak dönem sonu tablosu olarak veri tabanında oluşturulur.

Adım 2: Moodle sisteminden etkinlik raporundan her bir etkinliğin numarası ve adı alınıp ilgili alanlar ile veri tabanında etkinlik tablosu oluşturulur.

Adım 3: Günlük tablosunda açıklama alanında yer alan etkinlik numarası ile etkinlik tablosundaki etkinlik numarası alanı eşleştirilir. Her bir eşleştirme sonucu saydırılıp öğrencinin o etkinliğe kaç kere erişim sağladığı etkinlik erişim sayıları tablosuna yazılır

Adım 4: Dönem sonu tablosu ve etkinlik erişim sayıları tablolarından öğrenci numarası, dönem ve ders kodu alanları eşleştirilerek gerekli diğer alanlar (sınav notları, etkinlik erişim sayısı vb) eklenerek veri kümesi tablosu oluşturulur.

derskodu	donem	adsoyad	ders_slb	duyuru	kaynak	hdtop	hvtop	quiz	arasinav	odev	final	notu
BİLP103	2020211	Abdullah K...	2	0	2	21	30	12,75	24,3	0	0	37,05
BİLP103	2020211	...	2	0	1	10	14	10,5	0	15	0	25,5
BİLP103	2020211	...	0	0	0	7	0	8,25	21,3	14,25	22	65,8
BİLP103	2020211	...	2	0	1	16	0	11,25	16,8	14,25	24	66,3
BİLP103	2020211	...	7	1	1	31	10	15	25,8	15	35,2	91
BİLP103	2020211	...	3	0	0	4	0	8,25	20,4	14,25	12,8	55,7
BİLP103	2020211	...	12	3	3	52	4	12	26,4	15	30	83,4
BİLP103	2020211	...	2	1	2	43	8	9	28,8	15	25,2	78

Şekil 4.3. Veri kümesi örneği

Veri kümesi örneğinde ‘notu’ alanı dönem sonu geçme notunu ifade eder. Dönem sonu geçme notunda her bir ders için ölçüt olarak çan eğrisi sistemi uygulanarak öğrencinin başarı durumu belirlendi. Oluşan veri kümesinin içerdiği alanlar ve tanımları ile beraber örneği de Tablo 4.2’de verilmiştir.

Tablo 4.2. Veri kümesi alan açıklamaları

Alan Adı	Açıklama
derskodu	Dersin hangi kod ile açıldığı
donem	Dersin açıldığı yarıyıl
ogrencino	Öğrenci numarası
adsoyad	Öğrencinin adı ve soyadı
cinsiyet	Öğrencinin cinsiyeti
ders_slb	Etkinlik olarak tanımlanan dersin izlencesine erişim sayısı
duyuru	Ders hakkında yapılan duyuruya erişim sayısı
kaynak	Ders için önerilen yardımcı kaynaklara erişim sayısı
hdtop	Haftalık ders dokümanlarına toplam erişim sayısı
hvtop	Haftalık çevrimiçi ders videolarına toplam erişim sayısı
quiz	Öğrencinin quiz sınavının ağırlığına göre puanı
odev	Öğrencinin ödevin ağırlığına göre puanı
arasinav	Öğrencinin ara sınav ağırlığına göre puanı
final	Öğrencinin final sınavının ağırlığına göre puanı
notu	Dönem sonu notu (100 üzerinden)
sonuc	Başarı durumu (başarısız, başarılı, yüksek başarılı)

Veri kümesi 4 kategorik 12 sayısal alan olmak üzere toplam 16 alandan oluşmaktadır. Veri kümesinde iki ders için toplam 83 kayıt vardır.

4.1.3. Normalizasyon

Öğrencilerin Moodle sitesinde tanımlı ders etkinliklerine (ders_slb, duyuru, kaynak, hktop, hvtop) erişim sayılarının sınırları birbirinden bağımsız olduğundan bu etkinlik alanlarında min-max normalizasyon (0-1 Aralığında sayı üretir) yöntemi aşağıdaki formül ile hesaplandı:

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4.1)$$

Min-max normalizasyonda x her bir alandaki değeri, min(x) x'in bulunduğu sütundaki en küçük değeri, max(x) x'in bulunduğu sütundaki en büyük değeri ifade etmektedir.

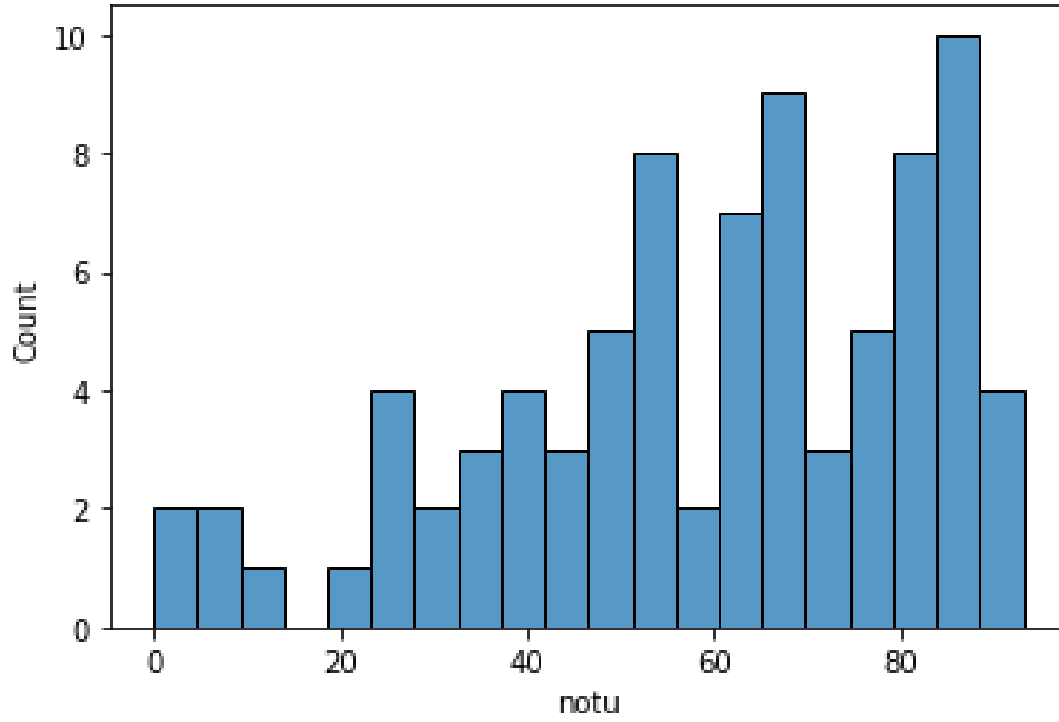
Moodle çevrimiçi öğrenme ortamından elde edilen ders etkinliklerine (ders_slb, duyuru, kaynak, hktop, hvtop) öğrencilerin erişim sayılarında min-max normalizasyon yöntemi uygulanarak alanlar 0 ile 1 Aralığında ifade edilmiştir. Öğrencinin çevrim içi ortamda girdiği sınavlardan aldığı notlar ağırlıklarına göre olduğu gibi alınmıştır. Bu işlemlerin sonunda oluşan veri kümesi alanlarının bir kısmından örnek görüntü Tablo 4.3' te verilmiştir.

Tablo 4.3. Normalizasyon uygulanan alanlar örneği

ders_slb	duyuru	kaynak	hktop	hvtop	quiz	arasinav	odev	final	notu	sonuc
0,154	0	0,25	0,16	0,33	13	24,3	0	0	37,1	Başarısız
0,154	0	0,13	0,07	0,15	11	0	15	0	25,5	Başarısız
0	0	0	0,05	0	8,3	21,3	14,3	22	65,8	Başarılı
0,154	0	0,13	0,12	0	11	16,8	14,3	24	66,3	Başarılı
0,538	0,17	0,13	0,23	0,11	15	25,8	15	35	91	Yüksek Başarılı
0,231	0	0	0,02	0	8,3	20,4	14,3	13	55,7	Başarılı

4.1.4. İstatistikler

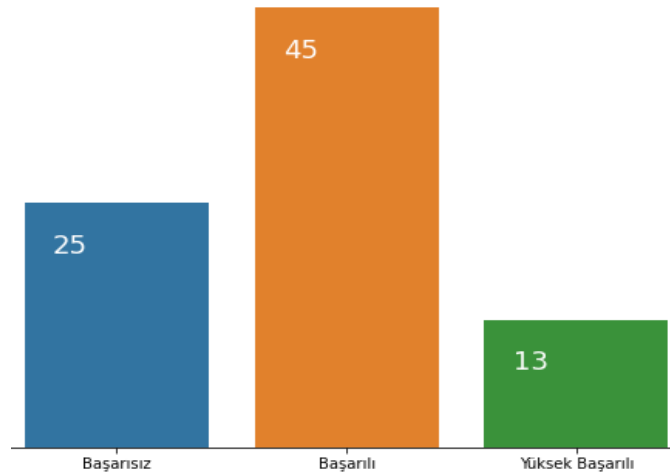
Öğrencinin dönem sonu geçme notu üzerinden bazı istatistiki sonuçlar elde edilerek değerlendirmeler yapılmıştır. Dönem sonu geçme notunun dağılım grafiği Şekil 4.4' de verilmiştir.



Şekil 4.4. Dönem sonu geçme notu dağılım grafiği

Dönem sonu geçme notunun dağılımı konusunda bir fikir elde edebilmek için istatistiksel yöntemlerden Kolmogorov-Smirnov ve Shapiro-Wilk testini uyguladık. Sırasıyla bu yöntemlerden elde edilen p değerleri 0,059 ve 0,072 dağılımın normal dağılımdan anlamlı bir fark sergilemediğini gösteriyor.

Başarı durumunun (başarılı, başarısız ve yüksek başarılı) sayıca dağılımını gösteren grafik Şekil 4.5’ de verilmiştir. Grafikte başarısız 25, başarılı 43 ve yüksek başarılı 13 olarak bulunmuştur.



Şekil 4.5. Başarı durumu sayıca dağılımı

4.2. Öznitelik Analizi ve Seçimi

Öznitelik analizi- işlemleri için Tablo 4.2’ de verilen veri kümesinden öğrencinin numarası ve adı soyadı alanları çalışmada kullanılmayacağı için çıkartılmıştır. Öznitelik seçimi için bölüm 3.2.’ de verilen *ilinti* (korelasyon), *bilgi kazancı*, *fisher puanı*, *selectKbest* ve *öznitelik önemi* yöntemleri 12 öznitelik için kullanılmıştır. Bunun yanı sıra, davranışsal ve zihinsel özniteliklerin başarıma olan etkileri de ayrı ayrı incelenmiştir.

4.2.1 Davranışsal ve Zihinsel Öznitelikler

Çevrimiçi öğrenme ortamında bir ders yürütülürken tanımlanan etkinlikler davranışsal ve zihinsel olarak ifade edilir. Ders için davranışsal olarak ifade edilen etkinlikler; ders dokümanları, ders videoları vb. olarak görülür. Öğrencinin bu etkinliklere erişiminden elde edilen bilgiler ile davranışsal öznitelikler veri kümesi Tablo 4.4’ de verilmiştir.

Tablo 4.4. Davranışsal etkinlik veri kümesi

Alan Adı	Açıklama
ders_slb	Etkinlik olarak tanımlanan dersin izlencesine erişim sayısı
duyuru	Ders hakkında yapılan duyuruya erişim sayısı
kaynak	Ders için önerilen yardımcı kaynaklara erişim sayısı
hdtop	Haftalık ders dokümanlarına toplam erişim sayısı
hvtop	Haftalık çevrimiçi ders videolarına toplam erişim sayısı

Bu tablodaki davranışsal özniteliklerin, öğrenci başarısının belirlenmesinde Rastgele orman yöntemi ile üçlü sınıflandırma doğruluk başarımları oranı %54 olarak bulunmuştur.

Öğrencinin çevrimiçi ortamda katıldığı zihinsel etkinlikler; öğrencinin girdiği sınavlar (quiz, ara sınav, final vb.) ile yaptığı ödevler olarak görülür. Öğrencinin girdiği sınavlardan aldığı puanlardan oluşan zihinsel etkinlikler veri kümesi Tablo 4.5’ de verilmiştir.

Tablo 4.5. Zihinsel etkinlik veri kümesi

Alan Adı	Açıklama
quiz	Öğrencinin quiz sınavının ağırlığına göre puanı
odev	Öğrencinin ödevin ağırlığına göre puanı
arasnav	Öğrencinin ara sınav ağırlığına göre puanı
final	Öğrencinin final sınavının ağırlığına göre puanı

Zihinsel etkinlik veri kümesindeki özniteliklerin, öğrenci başarısındaki etkisi Rastgele orman yöntemi ile üçlü sınıflandırmada %89 olarak bulunmuştur.

4.2.2 Öznitelik Seçimi

Bu yöntemlerden *İlinti* öznitelik seçiminde Pandas [52] ve Numpy [53] Kütüphaneleri kullanılmıştır. Özniteliklerin ikili olarak birbirleri arasındaki ilişki derecesini veren Pearson ilinti katsayısı hesaplanmıştır. Aşağıda verilen tüm öznitelik seçim yöntemlerinde Sklearn Kütüphanesi kullanılmıştır [54].

Bilgi Kazancı yöntemi ile sınıflandırmada Mutual (karşılıklı) fonksiyonu kullanılarak her öznitelik ile hedef değişken açıklayıcılığı ve özniteliklerin karşılıklı bağımlılığı ölçülmüştür.

Fisher Puanı yöntemi ile her bir sınıf için özelliklere ait ortalama ve standart sapma değerlerini kullanarak bir ilişki puanı Fisher Score fonksiyonu ile bulunmuştur.

SelectKbest öznitelik seçiminde sınıflandırmada 12 öznitelik (k değeri) arasından hedef değere uyumlu en yüksek puanı alanı belirlemede ki kare yöntemi kullanılmıştır.

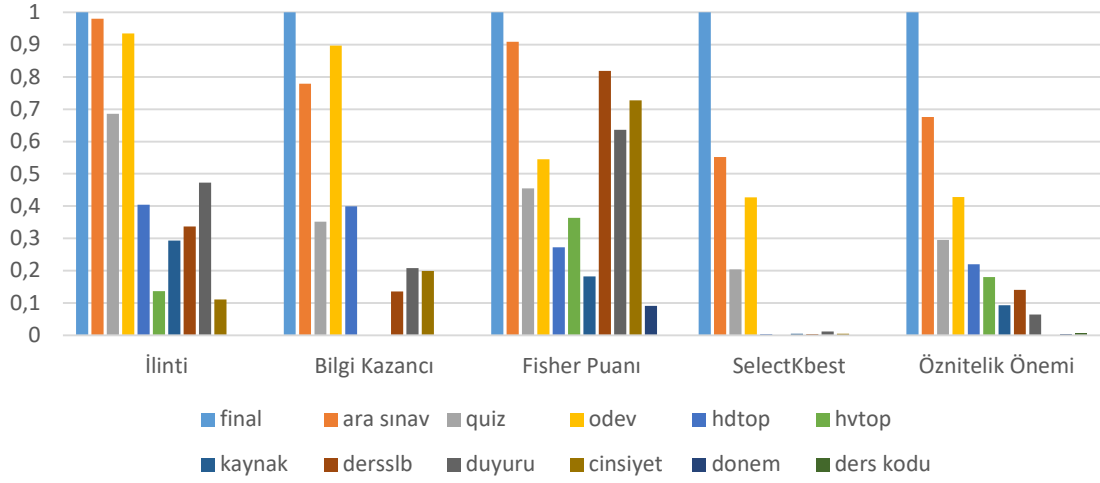
Öznitelik Önemi yöntemi ile sınıflandırmada Rastgele Orman yöntemi uygulanarak hedef değişken belirlenmesinde diğer özniteliklerin önem derecesi hesaplanmıştır.

Bu öznitelik seçim yöntemlerinden elde edilen sonuçlar farklı ölçeklerde olduğundan min-max formülü kullanılarak normalizasyon işlemi yapılmıştır. Normalizasyon sonrası yöntemlerin 12 alan için sonuçları Tablo 4.6' da verilmiştir.

Tablo 4.6. Öznitelik seçim yöntemlerinin normalize edilmiş sonuçları

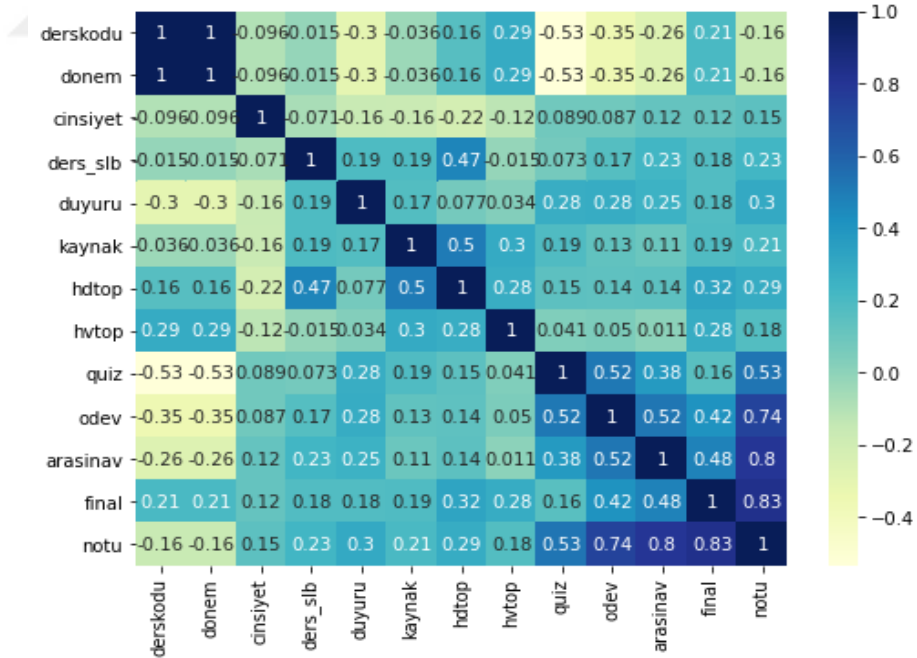
Öznitelikler / Özn. Seç. Yöntemi	final	ara sınav	quiz	odev	hktop	hvtop	kaynak	dersslb	duyuru	cinsiyet	donem	ders kodu
İlinti	1	0,98	0,69	0,9	0,4	0,137	0,293	0,34	0,472	0,111	0	0
Bilgi Kazancı	1	0,78	0,35	0,9	0,4	0	0	0,14	0,208	0,2	0	0
Fisher Puanı	1	0,91	0,45	0,5	0,27	0,364	0,182	0,82	0,636	0,727	0,09	0
SelectKbest	1	0,55	0,2	0,4	0	0,001	0,005	0	0,012	0,005	0	0
Öznitelik Önemi	1	0,68	0,29	0,4	0,22	0,18	0,094	0,14	0,065	0	0	0,01

Tablo 4.6' dan elden edilen değerlerin grafik üzerinde karşılaştırıldığı durum Şekil 4.6' da gösterilmiştir



Şekil 4.6. Öznitelik secim grafiği

Tüm öznitelik seçim yöntemlerinde final alanı en yüksek değere sahiptir. Cinsiyet, dönem ve ders kodu alanları öznitelik seçim yöntemlerinde hedef değişkenin belirlenmesinde en az değere sahiptir. Bütün alanların öğrencinin dönem sonu geçme notuna etkisini gösteren *karişıklık matrisi* Şekil 4.7’ de gösterilmiştir. Sağ taraftaki barda en koyudan en açığa ilişki derecesini (-1,1) göstermektedir. İlişki derecesinin en yüksek olduğu durum ‘1’ dir.



Şekil 4.7. Karişıklık matrisi

Şekil 4.7’ de de görüldüğü gibi finalin %83, arasınnavın %80, ödevin %74 ve quizin %53 değerleri ile dönem sonu geçme notu arasında ilişki derecesi yüksektir. Son üç sıradaki cinsiyet %15 ile düşük ders kodu ve dönem -%16 ile negatif bir ilişki derecesi vermektedir. İlk dokuz alan üzerinde çalışılacak veri kümesi olarak Tablo 4.7’ de verilmiştir.

Tablo 4.7. Veri kümesi

Alan Adı	Açıklama
ders_slb	Etkinlik olarak tanımlanan dersin izlencesine erişim sayısı
duyuru	Ders hakkında yapılan duyuruya erişim sayısı
kaynak	Ders için önerilen yardımcı kaynaklara erişim sayısı
hdtop	Haftalık ders dokümanlarına toplam erişim sayısı
hvtop	Haftalık çevrimiçi ders videolarına toplam erişim sayısı
quiz	Öğrencinin quiz sınavının ağırlığına göre puanı
odev	Öğrencinin ödevin ağırlığına göre puanı
arasinav	Öğrencinin ara sınav ağırlığına göre puanı
final	Öğrencinin final sınavının ağırlığına göre puanı
notu	Yarıyıl sonu notu (100 üzerinden)
sonuc	Başarı durumu (başarısız, başarılı, yüksek başarılı)

4.3. One Hot Encoding Yöntemi

Makine öğrenmesi algoritmalarında kategorik veriler (sayısal olmayanlar) doğrudan kullanılmazlar. Verilerin kullanılmadan önce sayısal biçimde ifade edilmeleri gerekir. Kategorik verilerin sayısal olarak kodlanmasında genel olarak Label Encoding ve One-Hot Encoding (OHE) yöntemleri kullanılır [18] [55].

Label Encoder yönteminde kategorik verinin aldığı değerın alfabetik sıralamasına göre benzersiz bir sayı ile temsili sağlanır. Örneğin renk özelliğinde kırmızı, mavi ve yeşil değerlerinin sayısallaştırılması Tablo 4.8’ deki gibi olur.

Tablo 4.8. Label Encoding temsili

Renk	Değeri
Kırmızı	0
Mavi	1
Yeşil	2

Bu yöntem genellikle sıralı kategorik verilerin sayısallaştırılmasında ve ağaç tabanlı yapılarda kullanılmaktadır.

Tablo 4.9. One Hot Encoding temsili

Renk		Kırmızı	Mavi	Yeşil
Kırmızı		1	0	0
Mavi		0	1	0
Yeşil		0	0	1

Renk		Kırmızı	Mavi	Yeşil
Kırmızı		1	0	0
Mavi		0	1	0
Yeşil		0	0	1

Tablo 4.10. Kayıt örneği									
ders_slb	duyuru	kaynak	hdtop	hvtop	quiz	arasinav	odev	final	notu
0,15385	0	0,25	0,15625	0,32609	12,75	24,3	0	0	37,5

ders_slb	duyuru	kaynak	hdtop	hvtop	quiz	arasinav	odev	final	notu
0,15385	0	0,25	0,15625	0,32609	12,75	24,3	0	0	37,5

[illegible]

Veri kümesinde her iki yöntemin ESA modeli ile karşılaştırması Tablo 4.11’ de verilmiştir. Veri kümesinin karakteristiğinden dolayı tez çalışmasında OHE temsili yöntemi uygulandı.

	Label Encoder	One Hot Encoder
Model Adı	Doğruluk (%)	Doğruluk (%)
ESA	91	92

5. ÖĞRENCİ BAŞARIMININ DEĞERLENDİRMESİ

Bu bölümde öğrenci başarımlarının sınıflandırılması ve tahmini için tasarlanan modeller, deneysel çalışmalar ve değerlendirmeleri Şekil 5.1’ deki iş akışı ile verilmiştir.



Şekil 5.1. Öğrenci başarımının değerlendirilmesi

Öğrenci başarısının sınıflandırılması ve tahmini için kullanılan RF, k-NN, MLP ve ESA modellerin tasarım kriterleri iş akışında ilk bölümde verilmiştir. İlk bölümde tasarlanan modellerin veri kümeleri üzerinde regresyon ve sınıflandırmaları deneysel çalışmalar başlığı ile sunulmuştur. Değerlendirmeler başlığında ise veri kümesinde öznitelik seçimi yapılmadan ve modellerin tasarım kriterleri değiştirilerek yapılan deneysel çalışmalar karşılaştırmalı olarak değerlendirilmiştir.

5.1. Modellerin Tasarım Kriterleri

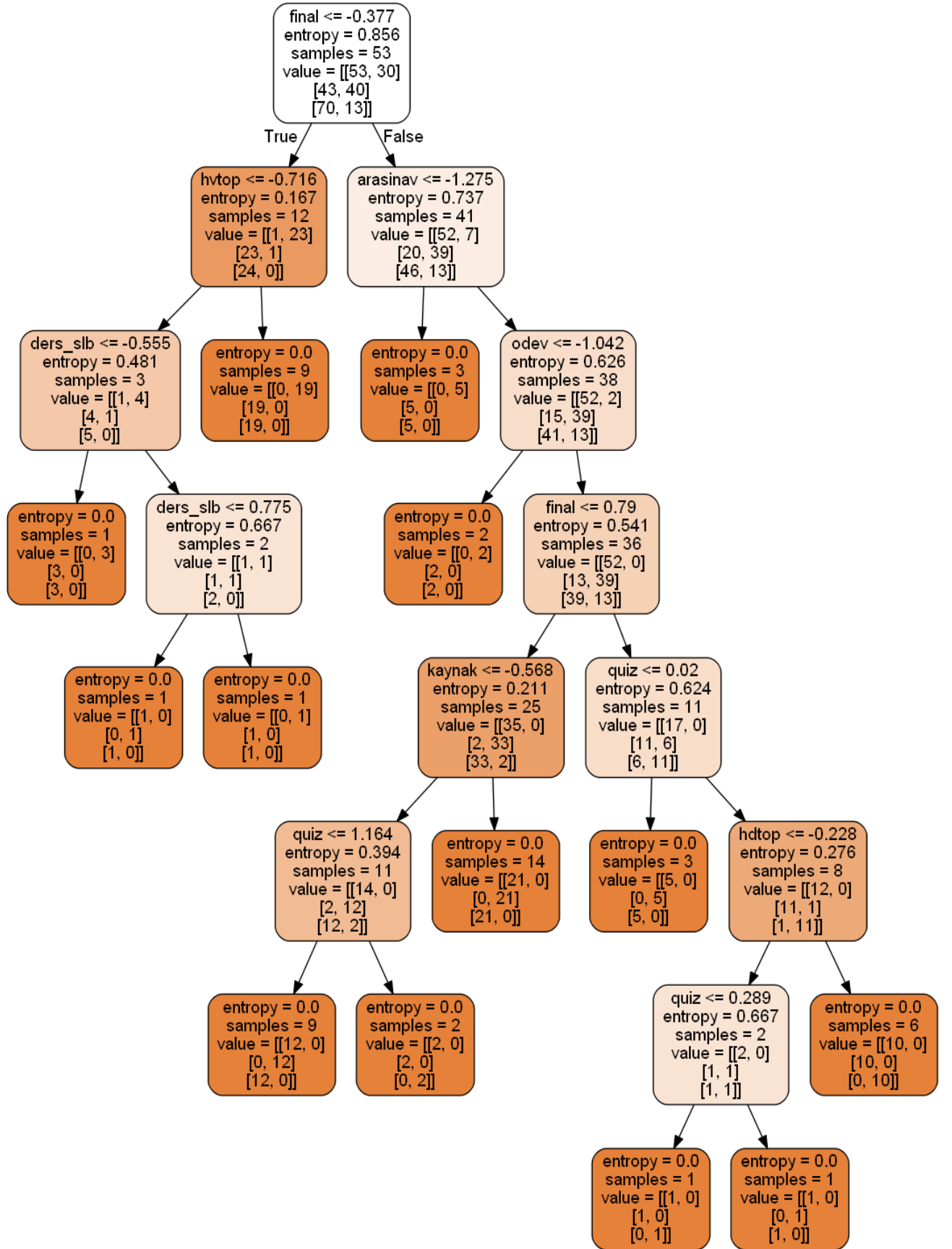
Genelde kategorik değişkenlerin temsilinde kullanılan OHE yöntemi bu çalışmada hedef değişkeni belirleyen bütün alanlarda kullanılmıştır. Modellerin tasarımında kullanılan parametreler optimize edilerek en iyi başarımlar sonucu elde edilmiştir. Modeller optimize edilmiş parametreler ile Tablo 4.8’ de verilen veri kümesi üzerinde OHE yöntemi uygulanmadan ve aynı veri kümesi üzerinde OHE yöntemi uygulanarak iki farklı sonuç elde edilmiştir.

Veri kümesi üzerinde modellerin sınıflandırma ve başarımlar tahmininde eğitim, test ve doğrulama işlemleri için 5 kat çapraz doğrulama yöntemi uygulanmıştır.

Tezde kullanılan RF, k-NN, MLP ve ESA modellerinin tasarımındaki kriterler aşağıda sunulmuştur. Sınıflandırma için RF, k-NN ve MLP modellerinde Sklearn Kütüphanesi [54] kullanılmıştır.

5.1.1. Rastgele Orman

Sınıflandırma modelinde parametreler ($n_estimators=300$, $criterion= 'entropy'$) optimize edilerek en iyi başarımlar sonucu elde edilmiştir. Modelde düzensizliğin ifadesi olan *entropy* sıfırlanıncaya kadar ağaç dallanmaya devam eder. RF sınıflandırma modeli için oluşturulan örnek bir ağacın yapısı Şekil 5.2’ de gösterilmiştir.



Şekil 5.2. Başkent Üniversitesi veri kümesindeki “Başarısız”, “Başarılı”, “Yüksek Başarılı” sınıflandırması için oluşturulan örnek RF ağaç yapısı

5.1.2. k-En Yakın Komşu

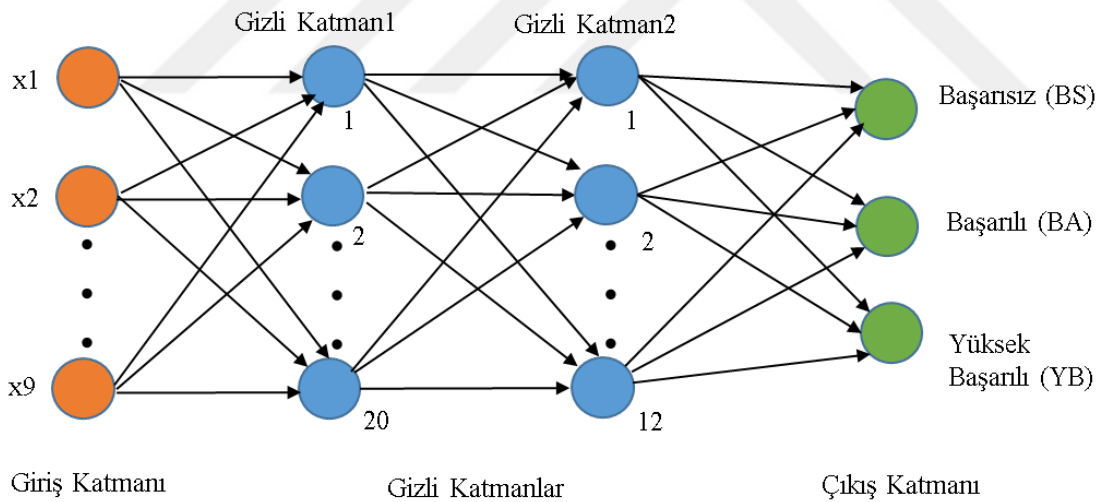
Sınıflandırma modelinde parametreler ($n_neighbors=5$, $metric= 'minkowski'$, $p=1$) optimize edilerek en iyi başarımlar sonucu elde edilmiştir. Farklı komşuluklar (3 ve 7) alınarak yapılan karşılaştırmalar Tablo 5.1’ de verilmiştir.

Tablo 5.1. k-NN karşılaştırmaları

Model Adı	5 komşuluğu	3 komşuluğu	7 komşuluğu
	Doğruluk (%)	Doğruluk (%)	Doğruluk (%)
k-NN OHE	72	69	66
k-NN	89	87	86

5.1.3. Çok Katmanlı Algılayıcı

Sınıflandırma modelinde parametreler optimize edilerek en iyi başarımlar sonucu elde edilmiştir. Giriş ve çıkış katmanı hariç iki gizli katman modelde kullanılmıştır. Modelin optimize edilmiş parametreler ile iki gizli katmanlı yapısı Şekil 5.3’ de gösterilmiştir.



Şekil 5.3. MLP katman yapısı

Deneysel olarak, giriş katmanında 9 öznelik ve çıkış katmanında aktivasyon fonksiyonu olarak relu seçilmiştir. Gizli katmanların ilkinde 20 nöron ikinci katmanda ise 12 nöron alınmıştır. Sınıflandırmada çıkış katmanı üç sınıf (Başarısız (BS), Başarılı (BA) ve Yüksek Başarılı (YB)) olarak belirlenmiştir.

5.1.4. Evrişimsel Sinir Ağı

Sınıflandırma için Keras Tensorflow Kütüphanesi kullanılmıştır [56]. Evrişimsel modelin bütün çıkış katmanında *activation* fonksiyonu için ‘softmax’ ve diğer bütün katmanlarda *activation fonksiyonu* için ‘relu’ değeri kullanılmıştır. İlk katmanda parametre olarak *filters=64, kernel_size=4* ve *input_shape=(9,1)* olarak alınmıştır. İkinci katmanda *filters=32, kernel_size=2* olarak alınmıştır. Modelin derlemesinde *loss* parametresi ‘sparse_categorical_crossentropy’ değerini, *optimizer* parametresi ‘adam’ ve *metrics* parametresi ‘accuracy’ değerini almıştır. Modelde *epochs* değeri olarak 135 alınmıştır. Modelin bu parametreler ile özeti Tablo 5.2’ de verilmiştir.

Tablo 5.2. ESA model özeti

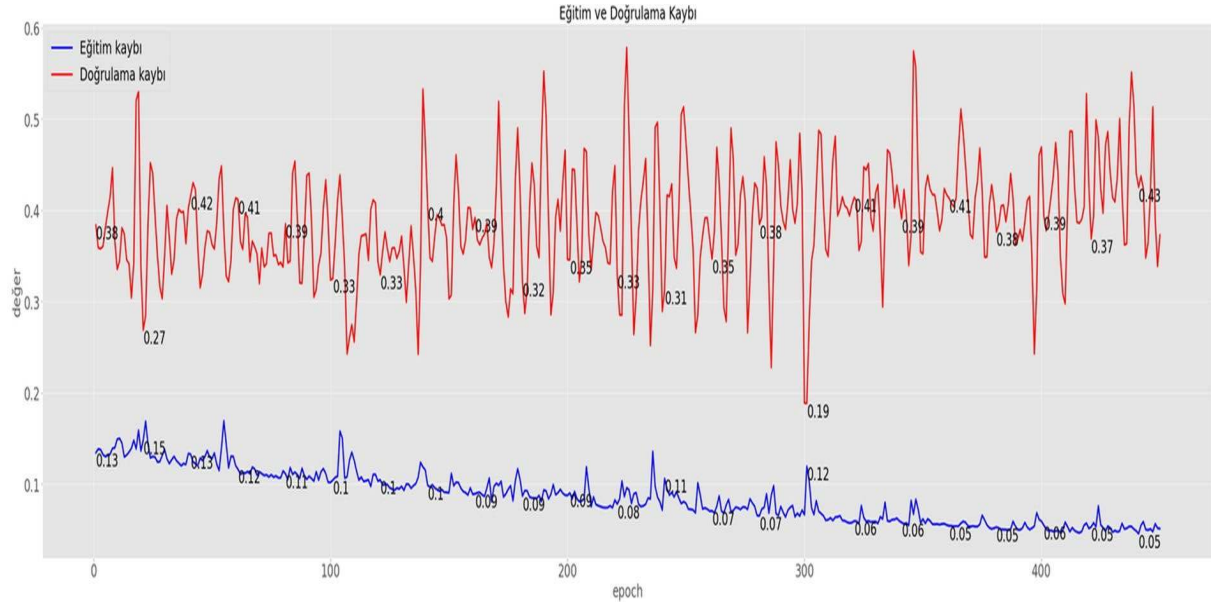
Layer (type)	Output Shape	Param #
conv1d (conv1D)	(None, 6, 64)	320
max_pooling1d (MaxPooling1D)	(None, 3, 64)	0
conv1d_1 (Conv1D)	(None, 2, 32)	4128
max_pooling1d_1 (MaxPooling1D)	(None, 1, 32)	0
dense (Dense)	(None, 1, 120)	3960
dense_1 (Dense)	(None, 1, 84)	10164
flatten (Flatten)	(None, 84)	0
dense_2 (Dense)	(None, 3)	255
Toplam Parametreler: 18,827		
Eğitilebilir Parametreler: 18,7827		

Veri kümesi üzerinde OHE yöntemi uygulanarak modelin aynı parametreler ile özeti Tablo 5.3’ de verilmiştir.

Tablo 5.3. OHE yöntemli ESA model özeti

Layer (type)	Output Shape	Param #
conv1d (conv1D)	(None, 256, 64)	320
max_pooling1d (MaxPooling1D)	(None, 128, 64)	0
conv1d_1 (Conv1D)	(None, 127, 32)	4128
max_pooling1d_1 (MaxPooling1D)	(None, 63, 32)	0
dense (Dense)	(None, 63, 120)	3960
dense_1 (Dense)	(None, 63, 84)	10164
flatten (Flatten)	(None, 5292)	0
dense_2 (Dense)	(None, 3)	15879
Toplam Parametreler: 34,451		
Eğitilebilir Parametreler: 34,451		

Modelin eğitim ve doğrulama kayıp grafikleri *epochs* değeri 450 alınarak ve belli aralıklar ile grafiğin aldığı değerler yazdırılmış ve sonuç Şekil 5.4’ de gösterilmiştir. Eğitimin 120-150 *epochs* arasında kesilebileceği görülüyor. Eğitim 135 *epochs* değerinde kesilmiştir.



Şekil 5.4. ESA modeli eğitim ve doğrulama kayıp grafiği

5.2. Deneysel Çalışmalar

Bu bölüm üzerinde deneysel çalışmalar yapılan veri kümeleri, sınıflandırma deneyleri ve regresyon başlıkları altında incelenecektir.

5.2.1. Veri Kümesi

Modellerin optimize edilmiş tasarım kriterleri doğrultusunda bu çalışmada iki farklı veri kümesi kullanılmıştır. İlkı Başkent Üniversitesi’nin çevrimiçi öğrenme ortamı olan Moodle’den elde edilen öğrenme analitiği verileri iken ikincisi İngiltere’de araştırmacıların kullanımına açılan Open University veri kümesi (OULAD) olmuştur [17].

a) Başkent Üniversitesi çevrimiçi öğrenme ortamı verileri

Başkent Üniversitesi çevrimiçi öğrenme ortamında dersler 14 hafta üzerinden dönemlik olarak tanımlanmaktadır. Her bir dersin etkinlikleri konu başlıklarına ayrılarak haftalık olarak Moodle sistemine girilir. Öğrencinin bu etkinliklere erişim sayıları, kayıt olduğu dersler, girdiği sınavlardan aldığı notlar veri kümesinin yapısını oluşturur. Veri kümesinin yapısı Tablo 5.4’ de verilmiştir.

Başkent Üniversitesi'nde 2020-2021 akademik yılında çevrimiçi öğrenme ortamı Moodle üzerinde yürütülen Bilgisayar Donanımı (Bilp103) ve Veri Tabanı (Bilp106) dersleri seçilmiştir. Moodle günlükleri işlenerek dersler için tanımlanan etkinliklere erişim sayıları, çevrimiçi girdiği sınavlar, öğrenci ve ders bilgilerinden öznitelik seçimi yapılarak 83 kayıttan oluşan Tablo 4.8' deki veri kümesi elde edilmiştir. Veri kümesinin istatistiksel sonuçları ve aldığı değerler sınırlarıyla Tablo 5.4' de verilmiştir.

Tablo 5.4. Başkent Üniversitesi veri kümesi istatistikleri

Alan Adı	Türü	Aralığı	Min	Max	Ortalama	Standart Sapma
derskodu	kategorik	-	-	-	-	-
donem	int64	-	-	-	-	-
ogrenci_no	int64	-	-	-	-	-
adsoyad	kategorik	-	-	-	-	-
cinsiyet	kategorik	-	-	-	-	-
ders_slb	float64	0-1	0	1	0,243744	0,232892
duyuru	float64	0-1	0	1	0,090361	0,180899
kaynak	float64	0-1	0	1	0,191261	0,227943
hdtop	float64	0-1	0	1	0,22741	0,193124
hvtop	float64	0-1	0	1	0,161472	0,219273
quiz	float64	0-15	0	15	0,809036	4,487585
arasinav	float64	0-30	0	29,1	17,56265	8,120476
odev	float64	0-15	0	15	10,73855	5,51693
final	float64	0-45	0	45	24,23735	12,692606
notu	float64	0-100	0	93,05	59,34759	23,505475
sonuc	kategorik	-	-	-	-	-

b) Open University veri kümesi

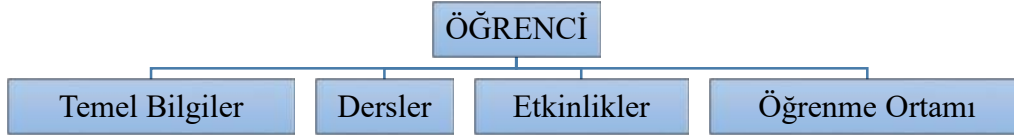
Open University [17] uzaktan eğitim ortamında bir ders genellikle 9 aylık zaman dilimlerinde yürütülür. Dersin öğrenimi süresinde öğrencinin katılacağı bir dizi etkinlik (sınav, ödev vb.) ve ders bitiminde gireceği bir final sınavı bulunmaktadır. Çevrimiçi öğrenme ortamında yürütülen bir dersin yapısı Şekil 5.5' de verilmiştir.



Şekil 5.5. OULAD ders yapısı

Veri kümesinde 2013 ve 2014 yıllarına ait öğrenci verileri birbirleriyle ilişkili tablolarda tutulur. Öğrencinin temel bilgileri, kayıt olduğu dersler, ders etkinlikleri ve çevrimiçi öğrenme

ortamındaki davranışları veri kümesinin genel yapısı oluşturur ve Şekil 5.6’ da verilmiştir. Veri kümesinde 22 derse kayıtlı 32,593 öğrencinin verileri vardır.



Şekil 5.6. OLUAD veri kümesi yapısı

Open University’nin uzaktan eğitim ortamında yürüttüğü derslerin verileri üç başlıkta incelenebilir [32].

- Demografik: Yaşadığı şehir, cinsiyet vb. temel öğrenci bilgileri.
- Başarı: Ders için tanımlanan etkinliklerden (tek başına, grup ve eğitmen eşliğinde) aldığı puanlar ve final puanı toplamı.
- Öğrenme davranışı: Çevrimiçi öğrenme ortamında etkinliklere erişim (gün, tıklama sayıları vb.) bilgileri.

Veri kümesindeki öğrencilerin başarı durumunu, ders etkinlikleri (sınav, ödev vb.) %50 ağırlıkta ve final sınavı ağırlığı da %50 olarak eşit belirlemektedirler. Başarı durumu *başarısız (fail)*, *başarılı (pass)*, *yüksek başarılı (distinction)* ve *dersten çekildi (withdrawn)* olarak tanımlıdır [17].

Veri kümesinde üçlü sınıflandırma (fail, pass, distinction) yapılırken dersten çekilenler (withdrawn) bu çalışma kapsamına alınmamıştır. Sınavlara girerek başarısız olan (puanları sıfır) öğrenci ile dersten çekilen öğrencinin sınav puanları sıfır olduğu veri kümesinde görülmüştür. Bu durum sınıflandırma başarımını etkileyeceği için dersten çekilenlerin bilgisi bu çalışmada kullanılmamıştır.

Her dersin etkinlik sayısı ve ağırlıkları farklıdır. Etkinliklerin puanı 0 ile 100 arasındadır ve 40 puan altı o etkinlik için başarısız olarak tanımlıdır [17]. Öğrencinin ağırlıklı etkinlik puanı (AEP) aşağıdaki şekilde hesaplanır:

$$AEP = \sum_{i=1}^{ES} \frac{EP_i * EA_i}{100} \quad (5.1)$$

Burada ES etkinlik sayısını, EP etkinlik puanını ve EA ise etkinlik ağırlığının ifade etmektedir.

Öğrencinin etkinlikten başarılı olması için 40'ın üzerinde puan alması gerektiği tanımlıdır [17]. Öğrencinin etkinliklerden geçme oranı (EGO) aşağıdaki şekilde hesaplanır:

$$EGO = \frac{BES}{ES} \quad (5.2)$$

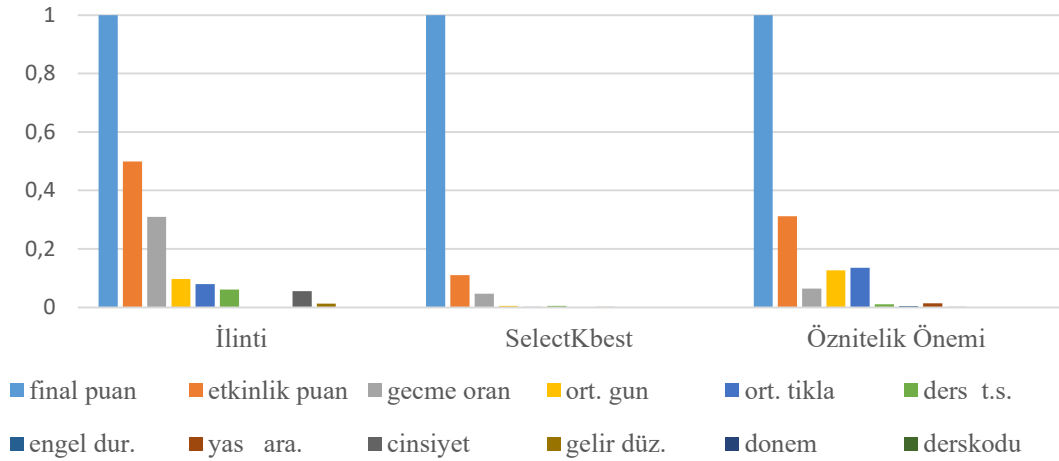
Burada ES etkinlik sayısını, BES başarılı etkinlik sayısını ifade etmektedir.

Öğrenme ortamı ve ders içeriklerine öğrencilerin eriştikleri gün ve tıklama sayıları tanımlıdır. Her ders için tanımlı etkinliğe öğrencinin eriştiği gün ve tıklama sayılarının ortalamaları alınmıştır.

Modelin performansını (eğitim süresi, doğruluk, karmaşıklık) önemli ölçüde etkileyen gereksiz ve alakasız özniteliklerin sayısı azaltılmalıdır [32]. Öznitelik seçiminde *ilinti* (korelasyon), *selectkbest* ve *öznitelik önemi* yöntemleri Sklearn kütüphanesi ile beraber kullanıldı.

İlinti, öznitelikler arasındaki ilişki derecesini (-1,1) belirlemede kullanılır. *SelectKbest*, öznitelik sayısı (k değeri) arasından hedef değere uyumlu en yüksek puan alanı belirlemede kullanılır. *Öznitelik önemi*, hedef değişkenin belirlenmesinde diğer özniteliklerinden en ilgili olanların önem derecesini verir.

Her bir seçme yöntemi farklı sayı aralıklarında puanlama (*selectkbest*, *ilinti*) ve derecelendirme (*öznitelik önemi*) yaptığından yöntem sonuçları (4.1) ile normalize edilerek Şekil 5.7' de verilmiştir.



Şekil 5.7. OULAD öznitelik seçimi

Öznitelik seçme yöntemleri sonucuna göre öğrencinin başarıyla en ilintili olan ilk 6 öznitelik seçilmiştir. Bu öznitelikler ile üzerinde çalışılacak veri kümesi ve açıklamaları Tablo 5.5' te gösterilmiştir.

Tablo 5.5. OLUAD veri kümesi

Öznitelik Adı	Açıklama
ders_tekrar_say	Öğrencinin dersi tekrar sayısı
basari_durumu	Dersteki başarı durumu
etkinlik_puani	Ağırlıklı etkinlik puanı
gecme_orani	Başarılı olduğu etkinliklerin oranı
final_puani	Final sınavından aldığı puan
ort_gun	Her bir etkinliğe eriştiği ortalama gün
ort_tıklama	Her bir etkinliğe ortalama tıklama

5.2.2. Sınıflandırma Çalışmaları

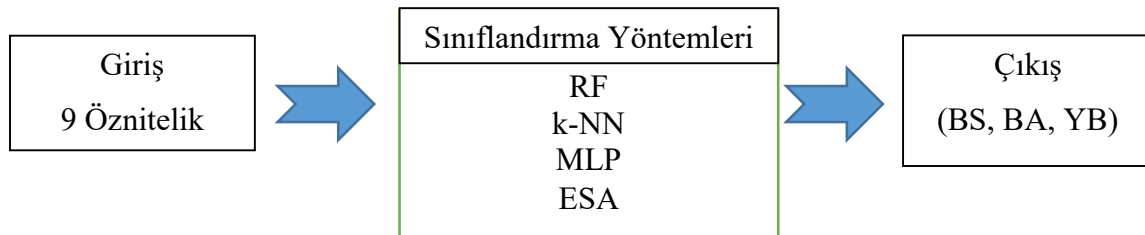
Başkent Üniversitesi (BÜ) çevrimiçi öğrenme ortamı verileri ve öğrenci bilgilerinden elde edilen Tablo 4.8’ de verilen veri kümesi üzerinde sınıflandırma çalışması eğitim, test ve doğrulama işlemleri 5 kat çapraz doğrulama ile yapılmıştır.

Başkent Üniversitesi çevrimiçi öğrenme ortamı verileri temel alınarak öğrencilerin dönem sonu geçme notuna göre başarıları Tablo 5.6’ da verilen puan ölçeğine göre sınıflandırılmıştır.

Tablo 5.6. Puan ölçeği

Puan	Başarı Durumu
0-49	Başarısız (BS)
50-89	Başarılı (BA)
90-100	Yüksek Başarılı (YB)

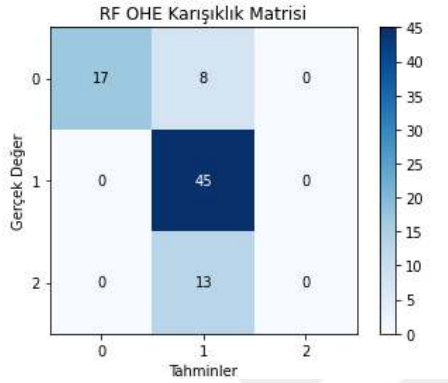
Belirlenen puan ölçeğine göre başarı durumunun sınıflandırılmasında bu çalışmada seçilen makine öğrenmesi yöntemleri (RF, k-NN, MLP, CNN) kullanılmıştır. Öznitelik seçim yöntemleri ile belirlenen 9 öznitelik sınıflandırma yöntemlerine girdi olarak verilip çıktı olarak Tablo 5.6’ da verilen başarı durumlarından biri elde edilmiştir. Her bir kayıt için sınıflandırma yönteminin işlem süreci Şekil 5.8’ de verilmiştir.



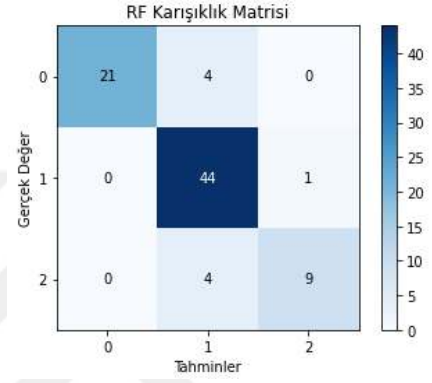
Şekil 5.8. Sınıflandırma yöntemleri

Modeller verilen tasarım kriterleri doğrultusunda Tablo 4.8’ de verilen veri kümesi üzerinde OHE yöntemi kullanılarak ve kullanılmadan sonuçlar elde edilmiştir. Sınıflandırmada değerlendirme ölçütü olarak karışıklık matrisinden (BS-0, BA-1, YB-2) elde edilen *doğruluk* (accuracy), *kesinlik* (presicion), *duyarlılık* (recall), *f1*-ölçeği kullanılmıştır.

RF modeli veri kümesinde OHE yöntemi kullanarak (RF OHE) ve OHE yöntemi kullanılmadan (RF) elde edilen karışıklık matrisi sırasıyla Şekil 5.9’ da ve Şekil 5.10’ da verilmiştir.

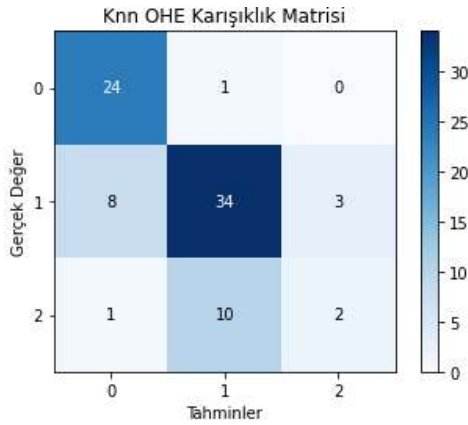


Şekil 5.9. RF OHE karışıklık matrisi

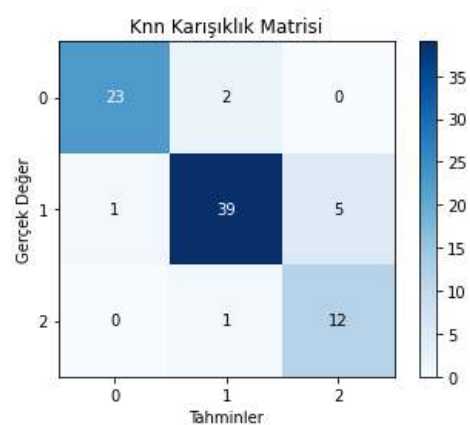


Şekil 5.10. RF karışıklık matrisi

k-NN modeli veri kümesinde OHE yöntemi kullanarak (k-NN OHE) ve OHE yöntemi kullanılmadan (k-NN) elde edilen karışıklık matrisi sırasıyla Şekil 5.11’ de ve Şekil 5.12’ de verilmiştir.

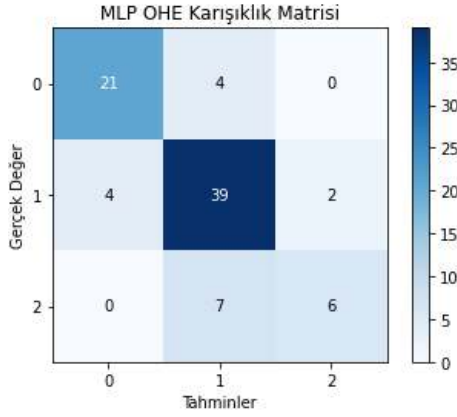


Şekil 5.11. k-NN OHE karışıklık matrisi

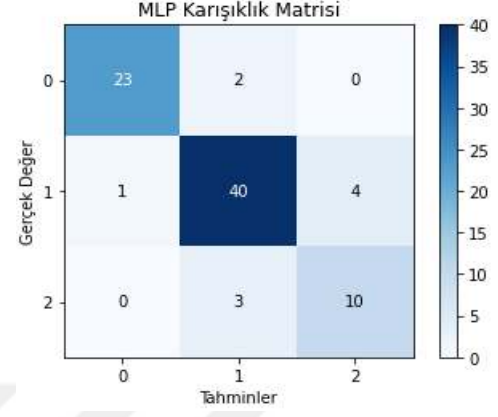


Şekil 5.12. k-NN karışıklık matrisi

MLP modeli veri kümesinde OHE yöntemi kullanarak (MLP OHE) ve OHE yöntemi kullanılmadan (MLP) elde edilen karışıklık matrisi sırasıyla Şekil 5.13’ de ve Şekil 5.14’ de verilmiştir.

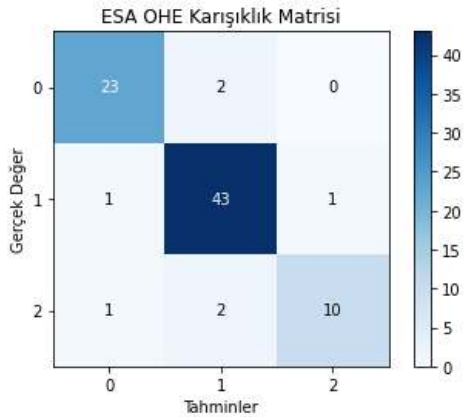


Şekil 5.13. MLP OHE karışıklık Matrisi

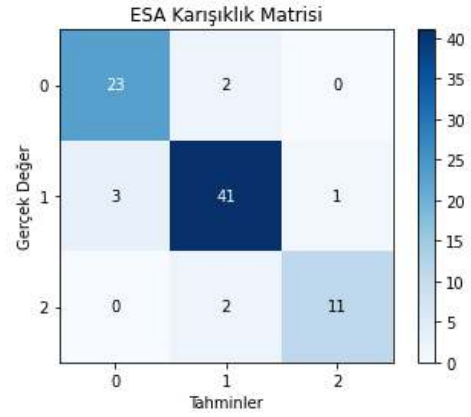


Şekil 5.14. MLP karışıklık Matrisi

ESA modeli veri kümesinde OHE yöntemiyle (ESA OHE) ve OHE yöntemi kullanılmadan (ESA) elde edilen karışıklık matrisi sırasıyla Şekil 5.15’ de ve Şekil 5.16’ de verilmiştir.



Şekil 5.15. ESA OHE karışıklık matrisi



Şekil 5.16. ESA karışıklık matrisi

Başkent Üniversitesi (BÜ) veri kümesi üzerinde, bütün modellerin karışıklık matrisinden elde edilen değerlendirme ölçütlerine göre karşılaştırmaları Tablo 5.7’ te verilmiştir.

Tablo 5.7. BÜ veri kümesi üzerinde tasarlanan modellerin sınıflandırma (doğruluk, F1-ölçeği, kesinlik, duyarlılık) karşılaştırmaları

Model Adı	Doğruluk (%)	F1-ölçeği (%)			Kesinlik (%)			Duyarlılık (%)		
		BS	BA	YB	BS	BA	YB	BS	BA	YB
ESA OHE	92	92	93	83	92	91	91	92	96	76
ESA	90	90	91	88	88	91	92	92	91	85
RF OHE	75	81	81	0	100	68	0	68	100	0
RF	89	91	91	78	100	85	90	84	98	69
MLP OHE	80	84	82	57	84	78	75	84	87	46
MLP	88	94	89	74	96	89	71	92	89	77
k-NN OHE	72	83	76	22	73	76	40	96	76	15
k-NN	89	96	90	76	100	91	69	92	89	85

Değerlendirme ölçütlerine bakıldığında doğruluk %92 ile ESA OHE en yüksek gözüküyor. Diğer modellerin OHE yöntemi ile sonuçları % 72 ile % 80 arasında değişmektedir. Katman kullanan modellerin (ESA, MLP) başarımlarının katman kullanmayan modellere (k-NN, RF) göre dah iyi olduğu görülüyor.

OHE yöntemi kullanmadan gerçekleştirilen başarımlar oranları % 88 ile % 90 arasında değişmekte ve aralarında çok da fark olmadığı izlenmektedir.

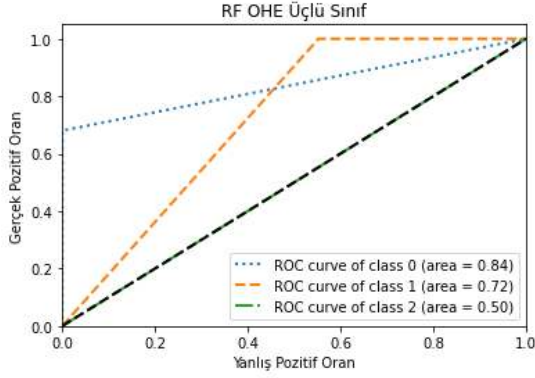
Sınıfların değerlendirmesinde F1-ölçeğinde k-NN modeli BS sınıfı %96 ile en yüksek değere sahip iken ESA OHE modeli BA sınıfında % 93 ve ESA modeli YB sınıfında % 88 ile en yüksek değerine ulaşmıştır.

Kesinlik değerlendirmesinin BA sınıfında 3 model (RF, RF OHE, k-NN) % 100 olur iken BA sınıfında da yine farklı 3 model (ESA OHE, ESA, k-NN) %91 ile en yüksek değerini almıştır. YB sınıfında ESA modeli %92 ile en yüksek değere sahiptir.

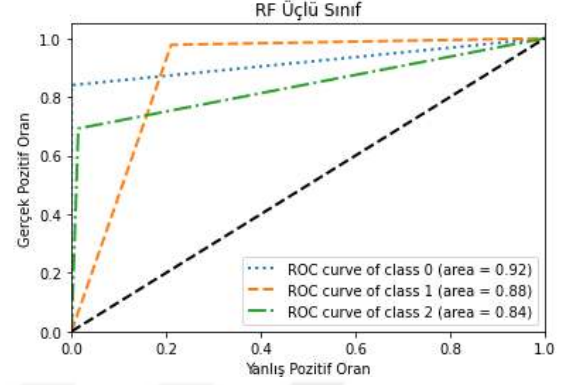
Duyarlılık değerlendirmesinde k-NN OHE modeli BS sınıfında %96 değerini ve RF OHE modeli BA sınıfında % 100 ile en yüksek değerini almıştır. YB sınıfında ise % 85 ile ESA ve Knn modelleri en yüksek değere ulaşmıştır.

Her bir model OHE yöntemi ve OHE yöntemi kullanılmadan çalıştırılarak her sınıf için Alıcı İşletim Karakteristiği (Receiver Operating Characteristic-ROC) eğrisi grafiği çizdirilmiştir. Her sınıfın ROC eğrisi altında kalan alanları (AUC) grafik üzerinde gösterilmiştir.

RF modeli veri kümesinde her sınıf için OHE yöntemi kullanarak ve kullanmadan ROC eğrisi çizdirilmiştir. Sonuçlar grafik olarak sırasıyla Şekil 5.17' de ve Şekil 5.18' de gösterilmiştir.

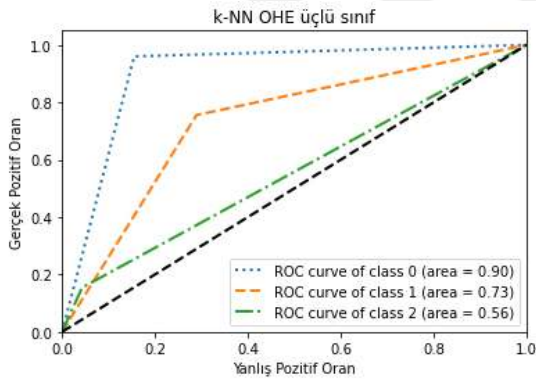


Şekil 5.17. RF OHE üçlü sınıf

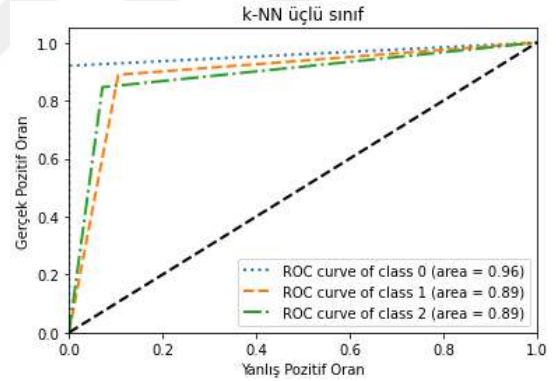


Şekil 5.18. RF üçlü sınıf

k-NN modeli veri kümesinde her sınıf için OHE yöntemi kullanarak ve kullanmadan ROC eğrisi çizdirilmiştir. Sonuçlar sırasıyla Şekil 5.19' da ve Şekil 5.20' de gösterilmiştir.

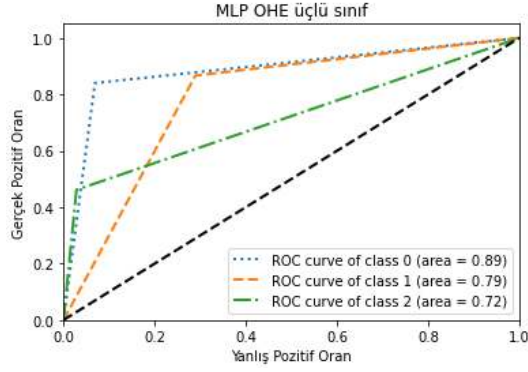


Şekil 5.19. k-NN OHE üçlü sınıf

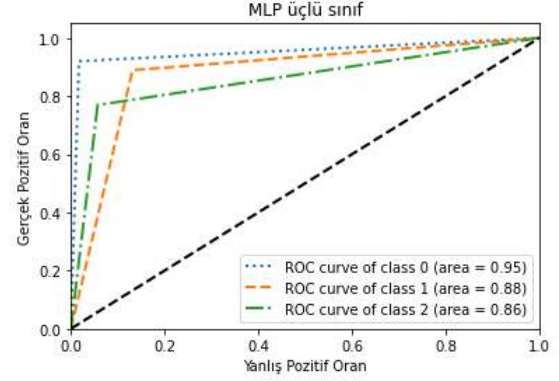


Şekil 5.20. k-NN üçlü sınıf

MLP modeli veri kümesinde her sınıf için OHE yöntemi kullanarak ve kullanmadan ROC eğrisi çizdirilmiştir. Sonuçlar grafik olarak sırasıyla Şekil 5.21' de ve Şekil 5.22' de gösterilmiştir.

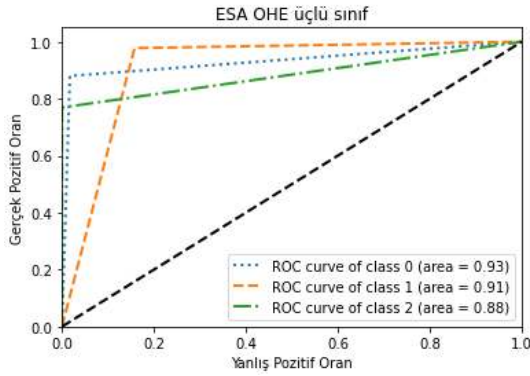


Şekil 5.21. MLP OHE üçlü sınıf

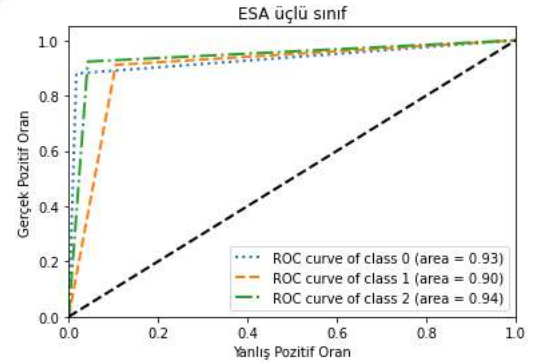


Şekil 5.22. MLP üçlü sınıf

ESA modeli veri kümesinde her sınıf için OHE yöntemi kullanarak ve kullanmadan ROC eğrisi çizdirilmiştir. Sonuçlar grafik olarak sırasıyla Şekil 5.23' de ve Şekil 5.24' de gösterilmiştir.



Şekil 5.23. ESA OHE üçlü sınıf



Şekil 5.24. ESA üçlü sınıf

Bütün modellerde her bir sınıf için çizdirilen ROC eğrisi altında kalan AUC alanı değerlendirme ölçütlerine göre karşılaştırmalar Tablo 5.8’ de verilmiştir.

Tablo 5.8. Başkent Üniversitesi veri kümesi üzerinde tasarlanan modellerin sınıflandırma (AUC alan) karşılaştırmaları

Model Adı	AUC (%)		
	BS	BA	YB
ESA OHE	93	91	88
ESA	93	90	94
RF OHE	84	72	50
RF	92	88	84
MLP OHE	89	79	72
MLP	95	88	86
k-NN OHE	90	73	56
k-NN	96	89	89

Modellerin ROC (Receiver Operating Characteristic) eğrisi altında kalan her bir sınıf için AUC değerlerinin geneline bakıldığında %96 ile k-NN modeli BS sınıfında en yüksek değere sahiptir.

BA sınıfında ESA OHE modeli %91 değeri alır iken YB sınıfında %95 ile ESA modeli en yüksek değeri almıştır.

Her bir sınıf için ROC eğrisi altında kalan alan değeri (AUC) ne kadar yüksek ise modellerin başarımı o denli yüksek olur. Burada bütün sınıflar için AUC değeri %50 ile %96 arasında değişim göstermektedir. Alanın ellinin üzerinde olması kabul edilebilir bir sonuç olarak yorumlanabilir.

Open University veri kümesi (OULAD) üzerinde önerilen ESA modelinin etkinliğini test edebilmek için geleneksel makine öğrenmesi yöntemleri (RF, k-NN, MLP) ve literatürdeki güncel çalışmalar ile kıyaslamalar sunulmuştur.

Tablo 5.4’te verilen veri kümesindeki toplam 32,593 kayıt üzerinde sınıflandırmada eğitim, test ve doğrulama işlemleri için 10 kat çapraz doğrulama yöntemi uygulanmıştır. Veri kümesi üzerinde OHE yöntemi ve önerilen ESA mimarisi (ÖY) ile sınıflandırma Geçti (G), Kaldı (K) ve Yüksek Başarılı (YB) olarak alınıp 10 kat çapraz doğrulama yöntemi ile sonuçlar elde edilmiştir. Sınıflandırma değerlendirme ölçütü olarak *doğruluk* (accuracy), *kesinlik* (precision), *duyarlılık* (recall) ve *f1* ölçüğü kullanılmıştır.

Önerilen yöntemin etkinliğini değerlendirmek amacıyla literatürdeki güncel çalışmalar [12] [13] ve literatürde yaygın olarak kullanılan makine öğrenmesi yöntemleri (RF, k-NN, MLP) ile karşılaştırma sonuçları Tablo 5.9’ da verilmiştir.

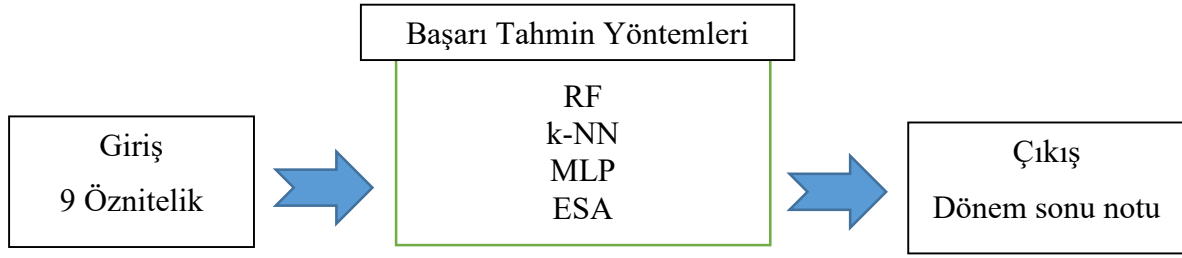
Tablo 5.9. OULAD veri kümesi üzerinde önerilen yöntemin sınıflandırma başarımı ile literatür karşılaştırması
(Önerilen Yöntem: ÖY, Yüksek Başarılı: YB, K: Kaldı, Geçti: G, Çekildi: Ç)

Model Adı	Sınıf	Doğruluk (%)	F1-skoru (%)				Kesinlik (%)				Duyarlılık(%)			
			YB	K	G	Ç	YB	K	G	Ç	YB	K	G	Ç
RF [12]	YB, K, G, Ç	70	-	-	-	-	73	48	78	67	60	46	89	57
NN [12]	YB, K, G, Ç	72	-	-	-	-	77	55	78	66	53	37	90	77
ESA (ÖY)	YB, K, G, Ç	73,32	78	34	85	67	76	39	81	74	81	30	89	61
RF [12]	K, G	91	-	-	-	-	-	93	90	-	-	84	96	-
NN [12]	K, G	92	-	-	-	-	-	92	92	-	-	86	95	-
ESA (ÖY)	K, G	95,43	-	96	95	-	-	98	93	-	-	30	89	-
GBM [13]	K, G, Ç	86.6	-	86	72	99	-	90	66	99	-	84	93	99
ESA (ÖY)	K, G, Ç	88	-	79	95	86	-	80	93	88	-	78	98	83
RF	YB, K, G	87	79	73	91	-	86	88	87	-	73	63	95	-
k-NN	YB, K, G	85	79	66	89	-	82	77	87	-	77	57	92	-
MLP	YB, K, G	88	83	80	92	-	80	83	92	-	86	77	91	-
ESA (ÖY)	YB, K, G	89	83	80	92	-	85	83	91	-	81	78	93	-

Literatürdeki güncel çalışmalar ile yapılan kıyaslamalar aynı sınıf ve yöntemler kullanılarak yapılmıştır. Önerilen model ESA ikili sınıflandırmada %95,43 başarımlı oranı elde ederken üçlü sınıflandırmada %88 ve dördümlü sınıflandırmada %73,22 başarımlı oranı elde etmiştir. Önerilen ESA modeli %89 ile diğer makine öğrenmesi modellerinden daha iyi başarımlı sonucu elde ettiği görülüyor. Buradan sınıf sayısı arttıkça başarımlı oranının düştüğü sonucunu çıkarabiliriz.

5.2.3. Öğrenci Başarımlı Tahmini

Öğrencinin çevrimiçi öğrenme ortam verileri ve girdiği sınavlar sonucunda oluşan dönem sonu geçme notu üzerinden başarımlı tahmini diğer bir ifade ile regresyon deneyleri yapılmıştır. Öznitelik seçim yöntemi sonuçlarına göre belirlenen Tablo 4.8’ deki ilk 9 tane öznitelikten öğrencinin dönem sonu geçme notunun tahmininde bu çalışma için seçilen makine öğrenmesi yöntemleri kullanılmıştır. Başarımlı tahmini için belirlenen yöntemlerin işleyişi Şekil 5.24’ de verilmiştir. Giriş olarak 9 öznitelik alınmış ve regresyon yöntemleri ile dönem sonu geçme notu tahmin edilmeye çalışılmıştır.



Şekil 5.25. Başarı tahmin yöntemleri

Regresyon yöntemleri ile dönem sonu notu tahmin edilirken oluşan hataların değerlendirmesinde ölçüt olarak kök ortalama kare hata (RMSE) ve mutlak ortalama hata (MAE) kullanıldı. ESA modelinin OHE yöntemi ile özeti Tablo 5.10’ da verilmiştir.

Tablo 5.10. ESA modeli tahmin özeti

Layer (type)	Output Shape	Param #
conv1d (conv1D)	(None, 256, 64)	320
max_pooling1d (MaxPooling1D)	(None, 128, 64)	0
conv1d_1 (Conv1D)	(None, 127, 32)	4128
max_pooling1d_1 (MaxPooling1D)	(None, 63, 32)	0
dense (Dense)	(None, 63, 120)	3960
dense_1 (Dense)	(None, 63, 84)	10164
flatten (Flatten)	(None, 5292)	0

Toplam Parametreler: 18,572

Eğitilebilir Parametreler: 18,572

Her bir model için elde edilen sonuçlar karşılaştırmalı olarak Tablo 5.11’ de verilmiştir.

Tablo 5.11. Başkent Üniversitesi veri kümesinde öğrenci başarı tahmini hata sonuçları

Model Adı	RMSE (%)	MAE(%)
ESA OHE	0,89	0,61
ESA	0,93	0,65
RF OHE	13,19	10,68
RF	5,98	4,54
MLP OHE	27,71	23,61
MLP	52,23	48,57
k-NN OHE	15,01	12,18
k-NN	5,29	3,8

Kök ortalama kare hata ölçütünde ESA modelinin her ikisi de %1’ in altında kalmıştır. Yine aynı şekilde ortalama mutlak hata ölçütünde ESA modelleri %1’in altında kalmıştır.

5.3. Değerlendirmeler

Öznitelik seçim yöntemleri sonucu ile deneysel çalışmalarda kullanılan veri kümesinden çıkartılan dönem, ders kodu ve cinsiyet bilgileri eklenerek modellerin optimize edilmiş tasarım kriterleri için deneyler yapılmıştır. Deneysel çalışma sonuçları model temelinde aşağıda tartışılmıştır.

RF modeli için bütün özniteliklerin bulunduğu Tablo 4.2’ deki veri kümesi üzerinde optimize edilmiş tasarım kriterleri kullanılarak deneysel çalışma yapılmıştır. Modelde OHE yöntemi kullanılarak ve kullanılmadan elde edilen doğruluk başarımları sırası ile %72 ve %86 bulunmuştur.

MLP modeli için Tablo 4.2’ deki veri kümesinde aynı tasarım kriterleri ile deneysel çalışma yapılmıştır. Modelin OHE yöntemi ile başarımları %76 bulunurken OHE yöntemi olmadan bulunan sonuç %82 olmuştur.

k-NN modeli OHE yöntemi kullanarak Tablo 4.2’ deki veri kümesinde tasarım kriterleri değiştirilmeden elde edilen başarımları %70 olarak bulunmuştur. Modelde OHE yöntemi kullanmadan %87 elde edilmiştir.

ESA modeli, tasarım kriterleri başlığı altında verilen aynı parametreler ile Tablo 4.2’de verilen veri kümesi üzerinde deneysel çalışmalar yapılmıştır. Yapılan deneysel çalışmaların sonuçları OHE yöntemi kullanılarak ve kullanılmadan elde edilmiştir. Sırasıyla çalışmalardan elde edilen doğruluk başarımları oranları %90 ve %88 olarak bulunmuştur.

Yapılan bu deneysel çalışmalarda elde edilen sonuçların değerlendirmeleri iki başlıkta toplanmıştır. Hiç bir özneliğin çıkartılmadığı durum ‘Öznitelik Seçimsiz’ ve öznelik seçimi yapılarak elde edilen modellerin sonuçları ‘Öznitelik Seçimli’ başlığında Tablo 5.12’ de verilmiştir.

Tablo 5.12. Başkent Üniversitesi veri kümesi üzerinde başarımları karşılaştırmaları

Model Adı	Öznitelik Seçimli	Öznitelik Seçimsiz
	Doğruluk (%)	Doğruluk (%)
ESA OHE	92	90
ESA	90	88
RF OHE	75	72
RF	89	86
MLP OHE	80	76
MLP	88	82
k-NN OHE	72	70
k-NN	89	87

ESA modelinin her iki durumda da en iyi sonucu verdiği görülmüştür.

SONUÇ

Bu çalışmada öğrencinin çevrimiçi öğrenme ortamına katılımından elde edilen öğrenme analitiği verileri ile başarısı arasındaki ilinti; öğrencinin başarısının sınıflandırılması ve tahmini görev tanımı temelinde incelenmiştir.

Belirlenen görevler temelinde, öğrenme ortamı verileri ile öğrenci başarısı arasındaki ilintinin anlaşılması esas amaç olarak belirlenmiştir. Bu amaç doğrultusunda çevrimiçi öğrenme ortamlarından elde edilen veriler üzerinde makine öğrenmesi yöntemleri (RF, k-NN, MLP ve ESA) kullanılarak öğrenci başarısının sınıflandırması ve tahmini yapılmıştır.

Belirtilen amaç doğrultusunda bu çalışmada, veri kümesi olarak Başkent Üniversitesi çevrimiçi öğrenme ortamı olan Moodle verileri ve İngiltere'deki Open University'nin 2013-2014 yıllarına ait verileri kullanılmıştır. Başkent Üniversitesi veri kümesi üzerinde hem başarının sınıflandırması hem de başarının tahmini yapılırken, Open University veri kümesinde sadece başarı sınıflandırma çalışması yapılmıştır.

Bu çalışmanın giriş bölümündeki tezin kapsamı alt başlığında belirlenen araştırma soruları ilinti (korelasyon) ve öznitelik seçim yöntemleri ile incelenmiştir. Sadece davranışsal öznitelikler (ders videoları, ders dokümanları vb. erişim bilgileri) alındığında; davranışsal özniteliklerin %54 doğruluk değeri ile öğrenci başarımında etkili olduğu görülmüştür. Zihinsel özniteliklerin (sınav, ödev puanları), öğrenci başarısına etkisinde ise doğruluk başarıım oranı %89 olarak elde edilmiştir. Davranışsal ve zihinsel özniteliklere, öğrencinin cinsiyeti, dersin kodu ve açıldığı dönem bilgisi de eklenerek öznitelik seçimi yapılmadan oluşturulan veri kümesinde, doğruluk başarıım oranı %90 olarak bulunmuştur. Öznitelik seçimi yapılarak, davranışsal ve zihinsel özniteliklerin beraber kullanılmasında elde edilen doğruluk başarıım oranı %92 olarak elde edilmiştir. Öznitelik seçimi yapılmadan bulunan %90 başarıım oranı ile öznitelik seçimi yapılarak bulunan %92 başarıım oranı, öznitelik seçiminin önemi ele aldığımız problemde de doğrulanmıştır. Veri kümesinin karakteristiği yanı sıra ders kodu ve açıldığı dönemin başarıya negatif etkisi olduğu karışıklık matrisinde görülmektedir. Aynı matriste cinsiyetin öğrenci başarıya etkisi %15 olarak görülmektedir.

Öznitelik seçimi yapılmadan kullanılan sınıflandırma çalışmaları yöntemlerinde problemin karmaşıklığı artmakta ve başarıım oranı olumsuz yönde etkilenmesi bir dezavantaj olarak görülür. Bunun yanında veri kümesini daha az öznitelik ile temsil edilmesini sağlayan öznitelik seçimi, başarıım oranını artırması yöntemin avantajı olarak değerlendirilir.

Veri kümesinde öğrenci başarısının oluşmasında ilintisi olan özniteliklerin belirlenmesi kapsamında öznitelik seçim yöntemleri (*ilinti, bilgi kazancı, fisher puanı, özniteliğin önemi*)

kullanılmıştır. Öznitelik seçim yöntemlerinin her biri başarı ile ilintisi olan öznitelikleri seçerken farklı teknik ve sınırlarda puanlama yaparak kullanır. Çalışmada öznitelik seçim yöntemleri ile elde edilen bu değerler min max yöntemi kullanılarak normalize edilmiştir. Her iki veri kümesi içinde öznitelik seçim yöntemi uygulanarak sonuçlar elde edilmiştir. Bu sonuçlarda öğrenci başarısı ile en çok ilintisi olan özniteliklerin; zihinsel etkinlik diye tanımlanan öğrencinin girdiği sınavlar, sonrasında davranışsal etkinlik diye tanımlanan öğrenme ortamına erişim sayıları ve son olarak da öğrenci demografik bilgileri olduğu görülmüştür.

Öznitelik seçim yöntemleri ile belirlenen öznitelikler; makine öğrenmesi algoritmalarıyla öğrenci başarısının sınıflandırması ve tahmini için kullanmadan önce veri ön işleme sürecinden geçirilmiştir. Öznitelikler içinde farklı sınır değerlerinde olanlar (öğrenme ortamı erişim sayıları) min max yöntemi ile normalize edilmiştir. Kategorik değerler içeren özniteliklerin sayısallaştırılmasında literatürde yaygın olarak kullanılan label encoding ve one hot encoding (OHE) yöntemleri deneysel çalışmalar yapılmıştır. Yapılan deneysel çalışmalarda, veri kümesi üzerinde her iki yöntem ile sınıflandırma doğruluk başarımları elde edilmiştir. Başarım oranlarında benzer sonuçlar elde edilse de label encoding yönteminin genelde sıralı değerlerde kullanıldığı göz önüne alınarak, veri kümesinin karakteristiğinden dolayı çalışmada verilerin sayısal temsili OHE yöntemi tercih edilmiştir.

Öğrenci başarısının sınıflandırması ve tahmininde, literatür incelemelerinde öne çıkan geleneksel makine öğrenmesi yöntemleri (RF, k-NN, MLP) ve son yıllarda kullanımı yaygınlaşan güncel derin öğrenme algoritması Evrişimsel sinir ağı (ESA) kullanılmıştır. ESA yöntemi genelde görsel sınıflandırma alanında kullanılmasına rağmen son yıllarda çevrimiçi öğrenme ortamı verileri üzerinden de sınıflandırma çalışmalarında kullanılmaktadır.

Çevrimiçi öğrenme ortamı öğrenme analitiği verileri kullanarak, öğrenci başarısının tahmini ve sınıflandırması için ders etkinliklerinin OHE temsiline, öznitelik seçimine ve Evrişimsel Sinir Ağına (ESA) dayalı bir yöntem önerilmiştir.

Çalışmada önerilen ESA mimarisine dayalı yöntemin, verilerin ikili sayısal temsili OHE ile beraber kullanımı sonuç almada, veri kümesi boyutuna bağlı olarak zaman kaybına neden olması bir dezavantaj olarak görülebilir. Ancak sınıflandırmada geleneksel makine öğrenmesi yöntemlerine göre başarımlarında iyi olması avantaj olarak kabul edilebilir.

Veri kümesi üzerinde yapılan çalışmalar verilerin OHE temsili ile ve bu temsil olmadan yapılmıştır. Önerilen model, hem geleneksel makine öğrenmesi yöntemleri ile hem de literatür ile kıyaslamaları yapılmıştır.

Geleneksel makine öğrenmesi yöntemlerinin üçlü sınıflandırmada OHE temsili ile doğruluk başarımları oranı %72 - %80 arasında değişirken önerilen ESA modelinde %92 gibi bir başarımlar

oranı elde edilmiştir. Verilerin temsili olmadan elde edilen sonuçlarda ise yine önerilen model ESA %90 başarı oranı vardır.

Literatür ile kıyaslamalarda ikili, üçlü ve dördü sınıflandırmalar ile kıyaslamalar yapılmıştır. İkili kıyaslamalarda literatürdeki çalışma %91 - %92 başarı oranında sonuç alırken, ESA modelinde %95,43 sonucu bulunmuştur. Üçlü sınıflandırma başarı oranlarında literatür %86,6 olur iken ESA modeli %88 değerini elde etmiştir. Benzer şekilde dördü sınıflandırmada ESA modeli %73,32 başarı oranında literatür çalışmaları ise %70 - %72 arasında değerler elde etmişlerdir.

Öğrenci başarısının tahmininde aynı makine öğrenmesi yöntemleri kullanılmıştır. Değerlendirme kriteri olarak ise kök ortalama kare hata (RMSE) ve ortalama mutlak hata (MAE) ölçütleri kullanılmıştır. Önerilen ESA modelinin OHE temsili ile başarı tahmini hata oranında RMSE %0,89 olurken MAE %0,61 değerleriyle %1'in altında kalmışlardır.

Çalışmada önerilen yöntem olan ESA modeli, sınıflandırma için kullanıldığı katmanlarda (filtreleme, havuzlama) yaptığı otomatik düzeltmeler sonucunda elde edilen küçük hata oranı ve sınıflandırma başarımında diğer yöntemlere göre daha öne çıkmaktadır.

Çalışma sonuçları; OHE yöntemi ile kodlanan özneliklerin kodlanmadığı duruma kıyasla daha başarılı olduğunu göstermektedir.

Öznelik seçim yöntemi sonuçlarına göre hedef değişken ile en çok ilintisi olan öznelikler kullanıldığında model başarımının, bütün alanların kullanımına kıyasla daha başarılı olduğunu göstermektedir.

Çevrimiçi öğrenme ortamlarından elde edilen öğrenme analitiği verileri ile öğrenci başarısının sınıflandırması ve tahmini için öznelik seçimi, OHE temsili ve ESA yöntemi literatürde ayrı ayrı kullanımı vardır. Bu çalışmada üçü bir arada kullanılarak sınıflandırma tahmininde başarı oranını olumlu anlamda artırdığı görülmüştür.

Çalışmada elde edilen bütün sonuçlar makine öğrenmesi yöntemleri ile çevrimiçi öğrenme ortamlarında öğrenci başarımının tahmininin yapılabileceğini göstermektedir.

Bu çalışma, öğrencinin çevrimiçi öğrenme ortamında kaldığı süre, öğrencinin ortamı kullanma becerisi ve eğitim materyallerinin (video, ders dokümanları, uygulama vb.) kalitesi bilgileri ölçülerek genişletilebilir. Gelecekteki çalışma planında, verilerin OHE dışında farklı temsili ile farklı veri kümelerinde çapraz doğrulama yapılarak öğrenci başarısının tahmini ve sınıflandırmasında doğruluk başarı oranı artırılması yer almaktadır.

KAYNAKLAR

- [1] A. H. Almala, “A Constructivist Conceptual Framework for a Quality e-Learning Environment,” *Distance Learn.*, vol. 2, no. 5, 2005.
- [2] A. Bozkurt, “Öğrenme analitiği: e-öğrenme, büyük veri ve bireyselleştirilmiş öğrenme,” *Açıköğretim Uygulamaları ve Araştırmaları Derg.*, vol. 2, no. 4, pp. 55–81, Oct. 2016.
- [3] A. H. Kıcıman, E. A. Tot, E. Eren, G. Çetintav, G. Karataş, and T. Guler, “2016-2020 yılları arasında Öğrenme Analitiği ile ilgili Yapılmış SSCI İndeksli Makalelerin Sistemantik Olarak İncelenmesi,” *Öğretim Teknol. ve Hayat Boyu Öğrenme Derg. - Instr. Technol. Lifelong Learn.*, vol. 2, no. 1, pp. 135–152, Jun. 2021.
- [4] A. Daud, M. D. Lytras, N. R. Aljohani, F. Abbas, R. A. Abbasi, and J. S. Alowibdi, “Predicting student performance using advanced learning analytics,” *26th Int. World Wide Web Conf. 2017, WWW 2017 Companion*, pp. 415–421, 2017.
- [5] S. Jha, M. Jha, and L. O’Brien, “Analysing Computer Science Course Using Learning Analytics Techniques,” in *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering, CSDE 2019*, 2019.
- [6] S. Lahmiri, R. G. Saade, D. Morin, and F. Nebebe, “Learning Analytics based on Bayesian Optimization of Support Vector Machines with Application to Student Success Prediction in Mathematics Course,” *Proc. 2020 5th Int. Conf. Cloud Comput. Artif. Intell. Technol. Appl. CloudTech 2020*, pp. 1–5, 2020.
- [7] O. H. T. Lu, A. Y. Q. Huang, and S. J. H. Yang, “Impact of teachers’ grading policy on the identification of at-risk students in learning analytics,” *Comput. Educ.*, vol. 163, no. December 2020, p. 104109, 2021.
- [8] A. Namoun and A. Alshanqiti, “Predicting student performance using data mining and learning analytics techniques: A systematic literature review,” *Appl. Sci.*, vol. 11, no. 1, pp. 1–28, 2021.

- [9] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, "Machine learning approaches to predict learning outcomes in Massive open online courses," in *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017-May.
- [10] N. Kadoic and D. Oreski, "Analysis of student behavior and success based on logs in Moodle," *2018 41st Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2018 - Proc.*, pp. 654–659, 2018.
- [11] L. Cagliero, L. Canale, L. Farinetti, E. Baralis, and E. Venuto, "Predicting student academic performance by means of associative classification," *Appl. Sci.*, vol. 11, no. 4, pp. 1–22, 2021.
- [12] Y. B. and A. C. Ian Goodfellow, *Deep Learning*. MIT Press, 2016.
- [13] J. Wang, Y. Qian, Q. Ye, and B. Wang, "Image retrieval method based on metric learning for convolutional neural network," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 231, no. 1, 2017.
- [14] Y. Sun, B. Xue, M. Zhang, and G. G. Yen, "Evolving Deep Convolutional Neural Networks for Image Classification," *IEEE Trans. Evol. Comput.*, vol. 24, no. 2, 2020.
- [15] S. Lyu and J. Liu, "Convolutional recurrent neural networks for text classification," *J. Database Manag.*, vol. 32, no. 4, 2021.
- [16] H. Karimi, J. Huang, and T. Derr, "A Deep Model for Predicting Online Course Performance," *Assoc. Adv. Artif. Intell.*, 2020.
- [17] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Data Descriptor: Open University Learning Analytics dataset," *Sci. Data*, vol. 4, 2017.
- [18] Y. Ma and Z. Zhang, "Travel Mode Choice Prediction Using Deep Neural Networks with Entity Embeddings," *IEEE Access*, vol. 8, pp. 64959–64970, 2020.
- [19] F. D. Pereira *et al.*, "Explaining individual and collective programming students' 2019; behaviour by interpreting a black-box predictive model," *IEEE Access*, pp. 117097–117119, 2021.

- [20] D. Uliyan, A. S. Aljaloud, A. Alkhalil, H. S. Al Amer, M. A. E. A. Mohamed, and A. F. M. Alogali, "Deep Learning Model to Predict Students Retention Using BLSTM and CRF," *IEEE Access*, vol. 9, pp. 135550–135558, 2021.
- [21] V. V. Zaporozhko, D. I. Parfenov, and V. M. Shardakov, "Development Approach of Formation of Individual Educational Trajectories Based on Neural Network Prediction of Student Learning Outcomes," in *Advances in Intelligent Systems and Computing*, 2020, vol. 1126 AISC, 2020.
- [22] P. Sökkhey and T. Okazaki, "Developing web-based support systems for predicting poor-performing students using educational data mining techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, 2020.
- [23] D. I. F. S. Engr. Sana Bhutto, Dr. Qasim Ali Arain, Maleeha Anwar, "Through Supervised Machine Learning," *Predict. Students' Acad. Perform. Through Supervised Mach. Learn.*, 2020.
- [24] C. Felix and S. R. Sobral, "Predicting students' performance using survey data," *IEEE Glob. Eng. Educ. Conf. EDUCON*, vol. 2020-April, pp. 1017–1023, 2020.
- [25] C. Y. Ko and F. Y. Leu, "Examining Successful Attributes for Undergraduate Students by Applying Machine Learning Techniques," *IEEE Trans. Educ.*, vol. 64, no. 1, pp. 50–57, 2021.
- [26] M. Glavas, M. Brkic Bakaric, and M. Matetic, "Applying advanced linear models in the task of predicting student success," *2018 41st Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2018 - Proc.*, pp. 744–748, 2018.
- [27] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," *Comput. Educ.*, vol. 131, no. July 2018, pp. 22–32, 2019.
- [28] R. Gampa and A. Baynes, "Visual analytic workflow to understand students' performance in computer science courses," *Proc. - Front. Educ. Conf. FIE*, vol. 2018-Octob, pp. 1–5, 2019.

- [29] A. M. Derocchis, A. Michalenko, L. E. Boucheron, and S. J. Stochaj, "Extending academic analytics to engineering education," *Proc. - Front. Educ. Conf. FIE*, vol. 2018-Octob, pp. 1–5, 2019.
- [30] L. Bognar and T. Fauszt, "Different learning predictors and their effects for moodle machine learning models," *11th IEEE Int. Conf. Cogn. Infocommunications, CogInfoCom 2020 - Proc.*, pp. 405–410, 2020.
- [31] T. Purwoningsih, H. B. Santoso, and Z. A. Hasibuan, "Data analytics of students' profiles and activities in a full online learning context," *2020 5th Int. Conf. Informatics Comput. ICIC 2020*, pp. 1–8, 2020.
- [32] M. L. Ruiz-Rodriguez, J. Andres Sandoval-Bringas, and M. A. Carreno-Leon, "Classification of student success using Random Forest and Neural Networks," *Proc. - 2020 3rd Int. Conf. Incl. Technol. Educ. CONTIE 2020*, pp. 98–103, 2020.
- [33] R. Ghorbani and R. Ghousi, "Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques," *IEEE Access*, vol. 8, pp. 67899–67911, 2020.
- [34] S. Jayaprakash, S. Krishnan, and J. Jaiganesh, "Predicting Students Academic Performance using an Improved Random Forest Classifier," *2020 Int. Conf. Emerg. Smart Comput. Informatics, ESCI 2020*, pp. 238–243, 2020.
- [35] J. M. Gallego-Romero, C. Alario-Hoyos, I. Estévez-Ayres, and C. Delgado Kloos, "Analyzing learners' engagement and behavior in MOOCs on programming with the Codeboard IDE," *Educ. Technol. Res. Dev.*, vol. 68, no. 5, 2020.
- [36] R. Alshabandar, A. Hussain, R. Keight, and W. Khan, "Students Performance Prediction in Online Courses Using Machine Learning Algorithms," *Proc. Int. Jt. Conf. Neural Networks*, 2020.
- [37] L. Bognár, T. Fauszt, and G. Z. Nagy, "Analysis of Conditions for Reliable Predictions by Moodle Machine Learning Models," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 6, pp. 106–121, 2021.

- [38] G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Comput.*, vol. 18, no. 7, 2006.
- [39] Y. Bengio, “Learning deep architectures for AI,” *Found. Trends Mach. Learn.*, vol. 2, no. 1, 2009.
- [40] L. Deng and D. Yu, “Deep learning: Methods and applications,” *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, 2013.
- [41] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8689 LNCS, no. PART 1.
- [42] J. Donahue *et al.*, “DeCAF: A deep convolutional activation feature for generic visual recognition,” in *31st International Conference on Machine Learning*, 2014, vol. 2.
- [43] D. Malekian, J. Bailey, and G. Kennedy, “Prediction of students’ assessment readiness in online learning environments: The sequence matters,” *PervasiveHealth Pervasive Comput. Technol. Healthc.*, pp. 382–391, 2020.
- [44] H. Karimi, T. Derr, J. Huang, and J. Tang, “Online Academic Course Performance Prediction using Relational Graph Convolutional Neural Network,” *Proc. 13th Int. Conf. Educ. Data Mining, EDM 2020*, no. Edm, pp. 444–450, 2020.
- [45] Ş. Çalışkan, “Çevrimiçi Öğrenme Ortamının Kullanılabilirlik Analizi Ve Etkililiği : Ahmet Yesevi Üniversitesi Örneği,” *Yüksek Lisans Tezi*, pp. 1–158, 2019.
- [46] H. Budak, “Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım,” *Süleyman Demirel Üniversitesi Fen Bilim. Enstitüsü Derg.*, vol. 22, no. Özel, p. 21, 2018.
- [47] M. Gök, “Makine Öğrenmesi Yöntemleri ile Akademik Başarının Tahmin Edilmesi,” *Gazi Üniversitesi Fen Bilimleri Dergisi Part C Tasarım ve Teknoloji*, vol. 5, no. 3, pp. 139–148, 2017.
- [48] T. M. Mitchel, *Machine Learning*, 1 ed. New York, NY, USA: McGraw-Hill, Inc., 1997.

- [49] S. Abirami and P. Chitra, “Energy-efficient edge based real-time healthcare support system,” in *Advances in Computers*, vol. 117, no. 1, Aca. Press Inc., 2020, pp. 339–368.
- [50] M. Kayakuş and M. Terzioğlu, “Yapay Sinir Ağları ve Çoklu Doğrusal Regresyon Kullanarak Emeklilik Fonu Net Varlık Değerlerinin Tahmin Edilmesi,” *Bilişim Teknoloji Dergisi*, pp. 95–103, 2021.
- [51] E. Taşcı and A. Onan, “K- En Yakın Komşu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi,” *Akademik Bilişim*, 2016.
- [52] Pandas, “Pandas Community.” [Online]. Available: <https://pandas.pydata.org/>. [Accessed: 02-Feb-2022].
- [53] Numpy, “Numpy Community.” [Online]. Available: <https://numpy.org/>. [Accessed: 02-Feb-2022].
- [54] Scikit, “Scikit-Learn Machine Learning in Python.” [Online]. Available: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning. [Accessed: 02-Feb-2022].
- [55] S. Bagui, D. Nandi, S. Bagui, and R. J. White, “Machine Learning and Deep Learning for Phishing Email Classification using One-Hot Encoding,” *J. Comput. Sci.*, vol. 17, no. 7, 2021.
- [56] Keras, “Introduction to Keras for researcher.” [Online]. Available: https://keras.io/api/layers/convolution_layers/. [Accessed: 02-Feb-2022].