T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Doktora Tezi

YAPAY ZEKÂ YÖNTEMLERİ İLE ÜNİVERSİTE ÖĞRENCİLERİNİN DERS MOTİVASYONLARININ TAHMİN EDİLMESİ

MİTHAT YAVUZARSLAN

Enformatik Anabilim Dalı

Enformatik Programı

DANIŞMAN
Doç. Dr. ÇİĞDEM SELÇUKCAN EROL

Nisan, 2023

İSTANBUL

Bu çalışma, 26.04.2023 tarihinde aşağıdaki jüri tarafından Enformatik Anabilim Dalı, Enformatik Programında Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Doç. Dr. ÇİĞDEM SELÇUKCAN EROL(Danışman)
İstanbul Üniversitesi
Enformatik Bölümü

Prof. Dr. SEVİNÇ GÜLSEÇEN
İstanbul Üniversitesi
Enformatik Bölümü

Prof. Dr. GONCA KIZILKAYA CUMAOĞLU
Yeditepe Üniversitesi
Eğitim Fakültesi

Dr. Öğr. Üyesi YALÇIN ÖZKAN
İstinye Üniversitesi
İktisadi, İdari ve Sosyal Bilimler Fakültesi

Doç. Dr. MURAT GEZER
İstanbul Üniversitesi
Enformatik Bölümü

İntihal Programı Beyanı

20.04.2016 tarihli Resmi Gazete'de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi'nin abonesi olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü'nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

Proje Destekleri

Tezden Üretilmiş Yayınların Künye Bilgileri

Yavuzarslan, M. & Erol, Ç, 2022. Öğrenme Yönetim Sistemi Log Kayıtlarının Akademik Başarı Tahmininde Kullanılması. *Bilişim Teknolojileri Dergisi*, 15(2), s.199-207, DOI: 10.17671/gazibtd.837884

ÖNSÖZ

Öncelikle, bu tez çalışmasının gerçekleştirilmesinde ilgi ve desteğini paylaşmakta hiçbir zaman tereddüt etmeyen danışmanım Doç. Dr. Çiğdem Selçukcan Erol'a ne kadar teşekkür etsem azdır.

Enformatik bölümündeki öğrenciliğim sırasında yol göstericiliğinden faydalandığım ve tez sürecindeki çok değerli öneri, yorum ve katkıları dolayısıyla Prof. Dr. Sevinç Gülseçen'e teşekkür ederim.

Doktora süreci de dâhil olmak üzere 2010'dan bu yana uzun süren öğrencilik dönemimde desteğiyle her aşamada yanımda olan Prof. Dr. Gonca Kızılkaya Cumaoğlu'na sonsuz teşekkürlerimi sunarım.

Son olarak, motivasyon ile ilgili bu tezi tamamlamak için gerekli motivasyonu sağlayan değerli eşim Gökçe Altın Yavuzarslan'a ve desteklerini her zaman yanımda hissettiğim aileme teşekkür ederim.

Nisan 2023 Mithat Yavuzarslan

İÇİNDEKİLER

Sayfa No

ÖNSÖZ ..... **iv**
İÇİNDEKİLER ..... **v**
ŞEKİL LİSTESİ ..... **viii**
TABLO LİSTESİ ..... **x**
SİMGE VE KISALTMA LİSTESİ ..... **xiii**
ÖZET ..... **xiv**
SUMMARY ..... **xvi**
1. GİRİŞ **1**
1.1. ARAŞTIRMANIN AMACI 2
1.2. ARAŞTIRMA SORULARI 2
1.3. SINIRLILIKLAR VE VARSAYIMLAR 2
2. GENEL KISIMLAR **4**
2.1. YAPAY ZEKÂ 4
2.2. VERİ TABANLARINDA BİLGİ KEŞFİ 6
2.2.1. Veri Seçimi 9
2.2.2. Veri Önişleme 9
2.2.3. Veri Dönüştürme 10
2.2.4. Modelleme 11
2.2.5. Modelin Performansının Ölçülmesi 11
2.2.5.1. Hold-out Yöntemi *12*
2.2.5.2. Tekrarlı Hold-out Yöntemi *12*
2.2.5.3. Çapraz Geçerleme *12*
2.2.5.4. Karmaşıklık Matrisi *13*
2.2.6. Değerlendirme 15
2.3. MAKİNE ÖĞRENMESİ 16
2.3.1. Danışmanlı Öğrenme 16
2.3.1.1. k-En Yakın Komşu Algoritması *18*
2.3.1.2. Naive Bayes Algoritması *19*
2.3.1.3. Destek Vektör Makineleri Algoritması *19*
2.3.1.4. C5.0 Algoritması *20*

2.3.1.5. CART Algoritması *21*
2.3.1.6. Rastgele Orman Algoritması *22*
2.3.1.7. Lojistik Regresyon Algoritması *22*
2.3.2. Danışmansız Öğrenme 23
2.3.3. Pekiştirmeli Öğrenme 24
2.4. EĞİTİMDE YAPAY ZEKÂ KULLANIMI 25
2.5. MOTİVASYON 29
2.5.1. Öz-Yeterlik Teorisi 30
2.5.2. Öz-Belirlenim Teorisi 31
2.5.3. Hedef Yönelimi Teorisi 31
2.5.4. Keller'in ARCS Motivasyon Teorisi 32
2.6. ÖĞRENME YÖNETİM SİSTEMLERİ VE LOG DOSYALARI 36
2.7. LİTERATÜR TARAMASI 38
2.7.1. ÖYS Log Dosyalarının Kullanıldığı Makine Öğrenmesi Çalışmaları 39
2.7.2. ÖYS Log Dosyalarının Kullanıldığı Motivasyon Odaklı Makine Öğrenmesi Çalışmaları 42
3. MALZEME VE YÖNTEM 45
3.1. ARAŞTIRMANIN ÖRNEKLEMİ 45
3.2. VERİ TOPLAMA ARAÇLARI 46
3.2.1. Derse İlgi Ölçeği (DİÖ) 46
3.2.2. Temel Bilgisayar Uygulamaları Dersi ve Öğrenme Yönetim Sistemi (ÖYS) Log Dosyaları 47
3.3. VERİ TOPLAMA 52
3.4. VERİ ANALİZİ 53
3.4.1. Veri Setinin Ön İşlenmesi 56
3.4.2. Modelin Eğitilmesi 61
3.4.3. Modelin Değerlendirilmesi 61
4. BULGULAR 63
4.1. DERSE İLGİ ÖLÇEĞİ (DİÖ) GEÇERLİLİK VE GÜVENİRLİK ANALİZİ BULGULARI 63
4.2. RASTGELE ALT ÖRNEKLEME (TEKRARLI HOLD-OUT) YÖNTEMİYLE ELDE EDİLEN BULGULAR 70
4.3. ÇAPRAZ GEÇERLEME YÖNTEMİYLE ELDE EDİLEN BULGULAR 77
4.4. HOLD-OUT YÖNTEMİYLE ELDE EDİLEN BULGULAR 84
4.4.1. Yukarı Örnekleme İle Elde Edilen Bulgular 85
4.4.2. SMOTE Yöntemi İle Elde Edilen Bulgular 88

4.4.3. MWMOTE Yöntemi İle Elde Edilen Bulgular....91
5. TARTIŞMA VE SONUÇ 95
KAYNAKLAR....100
EKLER....121
EK 1. TBA Sonuçlarına Ait Döndürülmüş Faktör Matrisi, Faktörlere Ait Güvenirlik Katsayıları, Madde Ortalama ve Standart Sapma Puanları (Madde Sayısı=31)....121
EK 2. DİÖ – 31 Maddeli Yapısal Eşitlik Modeli Diyagramı....122
EK 3. DİÖ – 25 Maddeli Yapısal Eşitlik Modeli Diyagramı....123
EK 4. Etik Kurulu İzni....124
EK 5. Bilgilendirilmiş Gönüllü Onam Formu....125
EK 6. Derse İlgi Ölçeği....126
ÖZGEÇMİŞ....128

ŞEKİL LİSTESİ

Şekil 2.1: Veri Tabanlarında Bilgi Keşfi Adımlarına Genel Bakış (Fayyad ve diğ., 1996, s.41) 8

Şekil 2.2: K-Katlamalı Çapraz Geçerleme Adımları 13

Şekil 2.3: Danışmanlı Öğrenme Diyagramı (Liu & Wu, 2012, s.3244) 17

Şekil 2.4: Destek Vektör Makineleri ile Doğrusal Sınıflandırma 20

Şekil 2.5: Standart Lojistik Fonksiyon 23

Şekil 2.6: Danışmansız Öğrenme Diyagramı (Kantardzic, 2020, s.110) 23

Şekil 2.7: Keller'in Motivasyon ve Performans Makro Modeli (Keller, 2010) 33

Şekil 3.1: Öğrenme Yönetim Sistemi Ekran Görüntüsü 48

Şekil 3.2: ÖYS Log Kayıtlarına Erişim ve Log Verisinin İşlenmesi Süreci 50

Şekil 3.3: Araştırmanın Veri Önişleme, Modelleme ve Performans Değerlendirme Süreci 55

Şekil 3.4: I. ve II. Veri Setlerinin "Faculty" Özniteliğine Ait Frekans Değerleri Grafikleri 57

Şekil 3.5: I. ve II. Veri Setlerinin "Motivasyon" Hedef Niteliğine Ait Frekans Değerlerini Gösteren Histogramlar 59

Şekil 4.1: I. SET (N=207) Özniteliklerine Ait Pearson Korelasyon Testi Sonuçları 71

Şekil 4.2: Rastgele Örnekleme Uygulanan I. SET (N=207)'in "Motivasyon" Hedef Niteliğine Ait İki Sınıflı Frekans Değerleri (Pozitif Sınıf ≥ 100.0) 72

Şekil 4.3: II. SET (N=200)'in "Motivasyon" Hedef Niteliğine Ait İkili Sınıf Frekans Değerleri 73

Şekil 4.4: I. SET (N=207) ve II. SET (N=200)'in "Motivasyon" Hedef Niteliğine Ait Üç Sınıflı Frekans Değerleri (Yüksek ≥ 110.0, 110.0>Orta ≥ 90.0, 90.0>Düşük) 75

Şekil 4.5: Çapraz Geçerleme Uygulanan I. SET (N=207)'in "Motivasyon" Hedef Niteliğine Ait İki Sınıflı Frekans Değerleri (Pozitif Sınıf ≥ 100.0) 77

Şekil 4.6: ROC-AUC Kriteriyle 10-Katlı Çapraz Geçerleme Uygulanan İki Sınıflı I.SET (N=207) ile Eğitilen CART Karar Ağacı Görselleştirmesi 82

Şekil 4.7: Yukarı Örnekleme Uygulanan I. SET (N=286)'in "Motivasyon" Hedef Niteliğine Ait Frekans Değerleri (Pozitif Sınıf ≥ 4,0) 85

Şekil 4.8: Yukarı Örnekleme Uygulanan I. SET (N=286) Özniteliklerine Ait Pearson Korelasyon Testi Sonuçları.....86

Şekil 4.9: SMOTE Uygulanan I. SET (N=286)'in "Motivasyon" Hedef Niteliğine Ait Frekans Değerleri (Pozitif Sınıf ≥ 4,0).....88

Şekil 4.10: SMOTE Uygulanan I. SET (N=286) Özniteliklerine Ait Pearson Korelasyon Testi Sonuçları 89

Şekil 4.11: MWMOTE Uygulanan I. SET (N=286)'in "Motivasyon" Hedef Niteliğine Ait Frekans Değerleri 91

Şekil 4.12: MWMOTE Uygulanan I. SET (N=286) Özniteliklerine Ait Pearson Korelasyon Testi Sonuçları 92

TABLO LİSTESİ

Sayfa No

Tablo 2.1: İkili Sınıflandırma Problemleri için Karmaşıklık Matrisi 14

Tablo 2.2: Eğitsel Veri Madenciliği ve Öğrenme Analitiği Arasındaki Farklar (Siemens & Baker, 2012; Romero & Ventura, 2013) 28

Tablo 2.3: Moodle Örnek Log Kaydı 38

Tablo 3.1: Araştırmanın Katılımcılarının Bağlı Bulundukları Fakültelere Göre Dağılımı 45

Tablo 3.2: Araştırmanın Katılımcılarının Sınıf Düzeylerine ve Cinsiyetlerine Göre Dağılımı 46

Tablo 3.3: Derse İlgi Ölçeği (DİÖ) Madde ve Alt Boyutları.....47

Tablo 3.4: Veri Seti Öznitelikleri, Veri kaynakları ve Özniteliklerin Açıklamaları.....51

Tablo 3.5: Veri Analizinde Kullanılan R Paket ve Fonksiyonları 54

Tablo 3.6: İkiye Ayrılan Veri Setinin Özelliklerinin Karşılaştırılması.....56

Tablo 3.7: Araştırmada Kullanılan Sınıflandırma Algoritmaları Fonksiyonlarına Ait Parametreler.....61

Tablo 4.1: DİÖ Faktör Analizi Örneklemine Ait Tanımlayıcı Demografik Bilgiler (N=305) 63

Tablo 4.2: DİÖ Faktör Analizi Örnekleminin Sınıflara Göre Ortalama, Mod, Medyan, En Düşük ve En Yüksek Değerleri.....64

Tablo 4.3: Derse İlgi Ölçeği (DİÖ) Güvenilirlik Analizi Sonuçlarının Literatürdeki Sonuçlarla Karşılaştırılması.....64

Tablo 4.4: KMO ve Bartlett Küresellik Testi Sonuçları 66

Tablo 4.5: YEM Analizi Sonucunda Elde Edilen CMIN/DF, NFI, CFI, GFI ve RMSEA Model Uyum İndeksi Değerleri ve Önerilen Değerlerin Karşılaştırılması.....68

Tablo 4.6: 25 Maddeli ve 4 Faktörlü DİÖ Alt Boyutlarına Ait Korelasyon Analizi Sonuçları.....68

Tablo 4.7: Araştırma Örnekleminin (N=407) Sınıflara Göre DİÖ Ortalama, Mod, Medyan, En Düşük ve En Yüksek Değerleri.....69

Tablo 4.8: Araştırma Örnekleminin (N=407) Fakültelere Göre DİÖ Ortalama, En Düşük ve En Yüksek Değerleri 70

Tablo 4.9: Rastgele Alt Örnekleme Uygulanan İki Sınıflı I. SET(N=207) ile Eğitilen Modellerin Performansları..... 72

Tablo 4.10: Rastgele Alt Örnekleme Uygulanan İki Sınıflı I. SET (N=207) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları 73

Tablo 4.11: Rastgele Alt Örnekleme Uygulanan İki Sınıflı I.SET (N=207) ile Eğitilen Naive Bayes Modeline Ait Öznitelik Önem Sıralaması 74

Tablo 4.12: Rastgele Alt Örnekleme Uygulanan Üç Sınıflı I. SET(N=207) ile Eğitilen Modellerin Performansları 76

Tablo 4.13: Rastgele Alt Örnekleme Uygulanan Üç Sınıflı I. SET(N=207) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları 76

Tablo 4.14: k-Katlamalı Çapraz Geçerleme Uygulanan İki Sınıflı I.SET (N=207) ile Test Edilen Algoritmalara Ait Doğruluk Değerleri 78

Tablo 4.15: k-Katlamalı Çapraz Geçerleme Uygulanan İki Sınıflı I.SET (N=207) ile Test Edilen Algoritmalara Ait ROC-AUC, Duyarlılık ve Özgüllük Değerleri 79

Tablo 4.16: k-Katlı Çapraz Geçerleme Uygulanan İki Sınıflı I. SET(N=207) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları 80

Tablo 4.17: ROC-AUC Kriteriyle 10-Katlı Çapraz Geçerleme Uygulanan İki Sınıflı I.SET (N=207) ile Eğitilen C5.0 Modeline Ait Öznitelik Önem Sıralaması..... 81

Tablo 4.18: k-Katlamalı Çapraz Geçerleme Uygulanan Üç Sınıflı I.SET (N=207) ile Test Edilen Algoritmalara Ait Doğruluk Değerleri 83

Tablo 4.19: k-Katlı Çapraz Geçerleme Uygulanan Üç Sınıflı I. SET(N=207) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları 84

Tablo 4.20: Yukarı Örnekleme Uygulanmış I. SET (N=286) ile Eğitilen Modellerin Hold-out ile Test Edilen (%70-%30) Performansları 87

Tablo 4.21: Yukarı Örnekleme Uygulanmış I.SET (N=286) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları 87

Tablo 4.22: SMOTE Uygulanmış I. SET (N=286) ile Eğitilen Modellerin Hold-out ile Test Edilen (%70-%30) Performansları 90

Tablo 4.23: SMOTE Uygulanmış I.SET (N=286) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları..... 90

Tablo 4.24: MWMOTE Uygulanmış I. SET (N=286) ile Eğitilen Modellerin Hold-out ile Test Edilen (%70-%30) Performansları 93

Tablo 4.25: MWMOTE Uygulanmış I.SET (N=286) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları.....93

SİMGE VE KISALTMA LİSTESİ

Kısaltmalar	Açıklama
AFA	: Açımlayıcı Faktör Analizi
AMOS	: Yapısak Eşitlik Modeli Analizi Programı *(Analysis of Moment Structures)*
ARCS	: Dikkat, İlişki, Güven, Tatmin *(Attention, Relevance, Confidence, Satisfaction)*
AUC	: Eğimin Altındaki Alan *(Area Under the Curve)*
CART	: Sınıflandırma ve Regreston Ağacı *(Classification and Regression Tree)*
CFI	: Karşılaştırmalı Uyum İndeksi *(Comparative Fit Index)*
CMIN/DF	: Ki-karenin Serbestlik Derecesine Oranı *(Minimum Discrepancy Function by Degrees of Freedom divided)*
DFA	: Doğrulayıcı Faktör Analizi
DİÖ	: Derse İlgi Ölçeği
GFI	: Uyum İyiliği İndeksi *(Goodness of Fit Index)*
KMO	: Kaiser-Meyer-Olkin Testi
MWMOTE	: Çoğunluk Ağırlıklı Azınlık Yukarı Örnekleme Tekniği *(Majority Weighted Minority Oversampling Technique)*
NFI	: Normlaştırılmış Uyum İndeksi *(Normed Fit Index)*
ÖYS	: Öğrenme Yönetim Sistemi
RMSEA	: Ortalama Karekök Sapması *(Root Mean Square Error of Approximation)*
ROC	: Alıcı İşletim Karakteristiği *(Receiver Operating Characteristic)*
SMOTE	: Sentetik Azınlık Yukarı Örnekleme Tekniği *(Synthetic Minority Over-Sampling Technique)*
SPSS	: Sosyal Bilimler İçin İstatistik Paketi Yazılımı *(Statistical Package for Social Sciences)*
TBA	: Temel Bileşenler Analizi
YEM	: Yapısal Eşitlik Modellemesi

ÖZET

DOKTORA TEZİ

YAPAY ZEKÂ YÖNTEMLERİ İLE ÜNİVERSİTE ÖĞRENCİLERİNİN DERS MOTİVASYONLARININ TAHMİN EDİLMESİ

MİTHAT YAVUZARSLAN

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Enformatik Anabilim Dalı

Danışman : Doç. Dr. ÇİĞDEM SELÇUKCAN EROL

Çevrimiçi öğrenme ortamlarında öğrencilerin derse olan ilgilerini saptamak yüz yüze öğrenme ortamlarına kıyasla daha zordur. Diğer yandan Öğrenme Yönetim Sistemleri (ÖYS), öğrencilerin çevrimiçi olarak sunulan derslerin ortam ve materyalleriyle olan etkileşimlerinden log adı verilen kayıt verisi oluşturmakta ve depolamaktadır. Söz konusu verinin yapay zekâ yöntemleriyle işlenerek öğrencilerin derse yönelik motivasyon düzeyleri ve motivasyon düzeylerini etkileyen değişkenleri ortaya çıkarmak mümkün olmaktadır. Bu çalışmada üniversite öğrencilerine ait ÖYS log verisi kullanılarak bir derse yönelik motivasyon düzeylerinin tahmin edilmesi amaçlanmıştır. Araştırma bir üniversitede 2020-2021 Güz, 2020-2021 Bahar ve 2021-2022 Güz dönemlerinde açılan Temel Bilgisayar Uygulamaları dersine kayıtlı 407 öğrenciyle yürütülmüştür. Öğrencilerden ders dönemini kapsayan 14 hafta boyunca kayıtlı oldukları dersin yürütüldüğü Moodle tabanlı ÖYS sunucusunda depolanan log kayıtları elde edilmiştir. Ayrıca öğrencilerin derse yönelik motivasyon düzeylerini belirlemek amacıyla John M. Keller tarafından geliştirilen ARCS motivasyon teorisini temel alan Derse İlgi Ölçeği (DİÖ) çalışma kapsamında Türkçe'ye uyarlama çalışmaları yapılarak uygulanmıştır.

Çalışmada iki veri seti kullanılmıştır ve birinci veri seti ile modelleme ve test işlemleri yapılmış olup ikinci veri seti ise birinci veri setiyle geliştirilen modeller için geçerleme seti olarak kullanılmıştır. Motivasyon hedef niteliği DİÖ ile elde edilen toplam puanların ikili ve üçlü sınıflara indirgenmesiyle elde edilmiştir. Çalışmada kullanılan yapay zekâ modelleri k-En Yakın Komşu, Naive Bayes, Destek Vektör Makineleri, C5.0, CART, Rastgele Orman ve Lojistik Regresyon algoritmaları ile eğitilmiştir. Performans değerlendirme amacıyla rastgele alt örnekleme, k-katlı çapraz geçerleme ve hold-out yöntemleri kullanılmış olup hold-out ile

yürütülen analizlerde veri setine çeşitli yukarı örnekleme yöntemleri uygulanmıştır. Değerlendirme için ise özellikle doğruluk ve ROC değerleri performans metrikleri olarak kullanılmıştır.

Araştırmanın sonucunda ÖYS log dosyalarından elde edilen veriyle öğrencilerin derse yönelik motivasyon düzeylerini tahmin etmenin mümkün olduğu görülmüştür. Bu doğrultuda birinci ve ikinci aşama bulgularında rastgele alt örnekleme ve k-katlı çapraz geçerleme ile %70 civarında genel sınıflandırma başarısı görülürken yukarı örnekleme uygulanmış veri seti kullanılarak hold-out ile test edilen modellerin performansları %80 seviyesine yakın ölçülmüştür. Diğer yandan özellikle birinci aşamada test edilen modeller arasında en yüksek performanslar karar ağacı algoritmalarıyla (C5.0, CART ve Rastgele Orman) elde edilmiştir. Motivasyon hedef niteliğine etki eden öznitelikleri belirleyebilmek adına yüksek performans gösteren algoritmaların öznitelik önemi sıralamalarına bakılmış ve ÖYS log dosyaları ile oluşturulan özniteliklerden "ToplamOturumSuresi", "SinavOdakliCalisma" ve "SayfadaKalmaOrtalama" gibi süre ifade eden özniteliklerin derse yönelik motivasyon düzeyi üzerinde diğer özniteliklere kıyasla daha belirleyici oldukları görülmüştür. Araştırmanın sonuçları ilgili literatürdeki bulgular göz önüne alınarak tartışılmış ve gelecekte yürütülecek benzer çalışmalara dair öneriler sunulmuştur.

Nisan 2023, 145 sayfa.

Anahtar kelimeler: öğrenme yönetim sistemleri, motivasyon, sınıflandırma, yapay zekâ, log dosyası

SUMMARY

Ph.D. THESIS

PREDICTION OF UNDERGRADUATE STUDENTS' COURSE MOTIVATION USING ARTIFICIAL INTELLIGENCE METHODS

MİTHAT YAVUZARSLAN

İstanbul University

Institute of Graduate Studies in Sciences

Department of Informatics

Supervisor : Assoc. Prof. Dr. ÇİĞDEM SELÇUKCAN EROL

In online learning environments, determining students' interest in the course is more difficult compared to face-to-face learning environments. On the other hand, Learning Management Systems (LMS) create and store record data called logs of students' interactions with the environment and materials offered online. It is possible to reveal the motivation levels of students towards the course and the variables that affect their motivation levels by processing this data with artificial intelligence methods. This study aims to predict the motivational levels of university students towards a course using LMS log data. The research was conducted with 407 students registered for the Basic Computer Applications course opened in the Fall 2020-2021, Spring 2020-2021, and Fall 2021-2022 semesters at a university. Log records stored in the Moodle-based LMS server where students were registered for the course during the 14 weeks covering the course period were obtained from the students. Additionally, the Course Interest Scale (CIS), based on the ARCS motivation theory developed by John M. Keller, was adapted to Turkish and applied within the scope of the study to determine the students' motivation levels towards the course.

In the study, two datasets were used considering students' LMS usage intensity and the branches they are registered for. Modeling and testing were carried out with the first data set, and the second data set was used as a validation set for the models developed with the first data set. The motivation target was obtained by reducing the total scores obtained from CIS to binary and

triple classes. Artificial intelligence models used in the study were trained with k-Nearest Neighbor, Naive Bayes, Support Vector Machines, C5.0, CART, Random Forest, and Logistic Regression algorithms. Random subsampling, k-fold cross-validation, and hold-out methods were used for performance evaluation, and oversampling methods were applied to the dataset in the analyses carried out with hold-out. Accuracy and ROC values were specifically used as performance metrics.

As a result of the research, it was seen that it is possible to predict students' motivation levels towards the course using the data obtained from the LMS log files. In this direction, while around 70% overall classification success was observed with random subsampling and k-fold cross-validation in the first and second stage findings, the performances of the models tested with hold-out using the oversampled datasets were measured close to 80%. On the other hand, the highest performances among the models tested, especially in the first stage, were obtained with decision tree algorithms (C5.0, CART, and Random Forest). To determine the features that affect the motivation target property, the variable importance rankings of the algorithms with high performance were examined, and it was observed that features expressing duration such as "TotalSessionTime," "ExamFocusedStudy," and "AverageTimeOnPage" from the features created with LMS log files were effective on students' motivation towards the course. The results of the research were discussed considering the findings in the relevant literature and suggestions for similar studies to be carried out in the future were presented.

April 2023, 145 pages.

Keywords: learning management systems, motivation, classification, artificial intelligence, log file

1. GİRİŞ

Büyük veri kavramının ortaya çıkışı ve teknolojik ilerlemenin önemli bir belirleyeni haline gelmesiyle birlikte yapay zekâ uygulamaları kısa zamanda küresel bir yaygınlık kazanmıştır. Yapay zekânın da parçası olduğu bu dönüşümü tanımlamak için kullanılan Endüstri 4.0 kavramı eğitim alanında karşılığını bireyselleştirilmiş öğrenme, öğrenme analitiği, akıllı eğitsel ajanlar, sanal/arttırılmış gerçeklik ve e-değerlendirme gibi yapay zekâ temelli uygulamaları kapsayan Eğitim 4.0 kavramında bulmuştur (Ciolacu ve diğ., 2017). Bu doğrultuda yapay zekânın günümüzde öğrenme ve öğretim faaliyetlerini yenilikçi hale getirme ve eğitim ile ilgili süregelen problemlere çözüm sunabilme potansiyeli taşıdığı kabul edilmektedir (UNESCO, 2019).

Bu tez çalışmasının GENEL KISIMLAR bölümünde Yapay Zekâ, Veri Madenciliği ve Veri Tabanlarında Bilgi Keşfi kavramları açıklanarak Eğitimde Yapay Zekâ Kullanımı uygulamaları hakkında bilgi verilmiş, Makine Öğrenmesi kavramı açıklanmış ve Eğitsel Veri Madenciliği ve Öğrenme Analitiği çalışmaları sunulmuştur. Bu tez çalışması dâhilinde yürütülen araştırmanın bir değişkeni olarak Motivasyon kavramı ve ARCS Motivasyon Teorisi hakkındaki bölümlere ek olarak araştırmanın gerçekleştirildiği ve tez verisinin elde edildiği Öğrenme Yönetim Sistemi ve Log Dosyaları başlığı açıklanmıştır. Son olarak Literatür Taraması bölümü içerisinde literatürde yer alan ve motivasyon dahil olmak üzere eğitim ile ilgili değişkenleri irdeleyen çalışmalar bulgularıyla birlikte sunulmuştur.

MALZEME VE YÖNTEM bölümünde araştırmanın katılımcıları hakkında demografik bilgiler özetlenmiş ve çalışmanın veri toplama amacıyla kullanılan Derse İlgi Ölçeği (DİÖ) ve demografik bilgi toplama formu hakkında bilgi verilmiştir. Ayrıca DİÖ'ye ait geçerlik ve güvenirlik analizleri sonuçları da bu bölümde yer almaktadır. Ardından araştırmanın yürütüldüğü Öğrenme Yönetim Sistemi ve verinin toplandığı ders ortam ve içeriği ile ilgili açıklamalar sunulmuş olup toplanan verinin analizi için kullanılan yapay zekâ algoritmaları açıklanmıştır.

BULGULAR bölümünde araştırma sürecinde elde edilen veri seti üzerinde yürütülen yapay zekâ algoritmalarından elde edilen sonuç değerler tablolar ile gösterilmiştir ve elde edilen bulgular ilgili başlıklar altında karşılaştırmalı olarak açıklanmıştır.

TARTIŞMA VE SONUÇ bölümünde ise elde edilen bulgular araştırmanın amacı doğrultusunda ilgili literatür göz önüne alınarak tartışılmış ve bu doğrultuda araştırmanın sonuçlarının mevcut literatüre katkısı açıklanmıştır.

1.1. ARAŞTIRMANIN AMACI

Çevrimiçi öğrenme ortamlarında eğitmenin öğrencilerle olan etkileşimi yüz yüze öğrenme ortamlarına kıyasla kısıtlı durumdadır ve öğrencilerin derse olan ilgileri hakkında bilgi sahibi olmak da aynı ölçüde zorlaşmaktadır. Diğer yandan çevrimiçi öğrenme ortamları kullanıcılardan elde ettikleri veriyi depolayarak sistem yöneticilerinin erişimine sunmaktadırlar. Söz konusu veri doğru yöntemlerle işlendiğinde öğrencinin öğrenme ortamı içerisindeki davranışlarını anlamak mümkün olmaktadır. Bu doğrultuda yürütülen tez çalışmasında bir Öğrenme Yönetim Sistemi'nden (ÖYS) elde edilen veri seti üzerinde yapay zekâ teknikleri kullanılarak öğrencilerin derse yönelik motivasyonlarının tahmin edilmesi amaçlanmaktadır.

1.2. ARAŞTIRMA SORULARI

Bu tez çalışmasında aşağıdaki araştırma sorularına cevap aranmaktadır:

1. Katılımcıların ÖYS kullanım verisi ile derse yönelik motivasyonları arasında ilişki var mıdır?
2. Katılımcıların ÖYS kullanım verisi ile elde edilen veri setine ait hangi öznitelikler derse yönelik motivasyonlarını etkilemektedir?

1.3. SINIRLILIKLAR VE VARSAYIMLAR

Bu tez çalışması kapsamında yürütülen araştırmanın sınırlılıkları aşağıdaki gibidir:

1. Bu araştırma Yeditepe Üniversitesi'nde 2020-2021 Güz, 2020-2021 Bahar ve 2021-2022 Güz dönemlerinde açılan Temel Bilgisayar Uygulamaları dersine kayıtlı, çeşitli bölümlerde okumakta olan ve araştırmaya gönüllü olarak katılan 407 öğrenciyle sınırlıdır.
2. Bu araştırma katılımcıların kullanmış oldukları Moodle tabanlı Öğrenme Yönetim Sistemi'nden (ÖYS) elde edilen ve Keller (2010) tarafından geliştirilen olan Derse İlgi Ölçeği (DİO) kullanılarak toplanan veriyle sınırlıdır.

Tez çalışmasının varsayımları aşağıda sıralanmıştır:

1. Araştırmanın katılımcıları Derse İlgi Ölçeği (DİÖ)'nin maddelerini içtenlikle cevaplamışlardır.
2. Araştırma süresince katılımcıların kullanmış oldukları Öğrenme Yönetim Sistemi (ÖYS)'nden elde edilen verinin etkileşim zaman ve miktarını temsil ettiği varsayılmaktadır.

2. GENEL KISIMLAR

Tezin Genel Kısımlar bölümünde Yapay Zekâ'nın tanımı ve yaygınlaşmasındaki etmenlere değinilmiş olup bu çalışmanın yöntemiyle ilişkili analiz sürecini açıklayabilmek adına Veri Tabanlarında Bilgi Keşfi başlığı bu süreci oluşturan alt başlıklarıyla birlikte aktarılmıştır. Eğitimde Yapay Zekâ kullanımı başlığı ilgili araştırmaları içerecek şekilde açıklandıktan sonra motivasyon kavramına ve literatürdeki çeşitli motivasyon teorilerine değinilmiştir. Öğrenme Yönetim Sistemleri ve Log dosyaları hakkında bilgi verilen başlıklardan sonra ise çalışmanın Literatür Taraması bölümü benzer yöntemlerle yürütülen çalışmalar ile doğrudan motivasyon değişkenini irdeleyen çalışmaları kapsayacak şekilde iki başlıkta incelenmiştir.

2.1.YAPAY ZEKÂ

Yapay zekâ teknolojileri özellikle son yıllarda giderek yaygınlaşmakta ve kişisel amaçlarla kullanılan mobil cihazlardan sağlık ve finans gibi toplumsal yaşam açısından önem arz eden sektörlerdeki uygulamalara kadar geniş bir kullanım alanını kapsamaktadır. Yapay zekâ çalışmalarının artışındaki en önemli sebeplerin başında dünyada üretilen veri miktarındaki devasa artış ve bu duruma eşlik eden işlemci hızı ve depolama kapasitesi gibi donanım teknolojilerindeki gelişmelerin geldiği bilinmektedir (Natale & Ballatore, 2020). Diğer bir deyişle büyük miktarda veri ile beslenerek geliştirilebilen yapay zekâ uygulamaları hem geçmişe kıyasla daha fazla miktarda ve türde verinin üretildiği hem de bu verinin işlenmesinde kullanılan donanımların geliştiği bir dönemde bu koşullara yapay zekâ odaklı çalışmalar için özelleşen programlama dilleri ve geliştirme paketlerinin öne çıkmalarının da eşlik etmesiyle birlikte bir yaygınlaşma olanağına erişmiştir. Kreutzer ve Sirrenberg (2020) yapay zekânın özelikle son yıllarda kazandığı ivmenin arkasındaki etmenleri (1) ürün, hizmet ve süreçlerin dijitalizasyonu ile birlikte fiziksel ortamdaki işlerin veriye dönüştürülebilmeleri, (2) bilişim sistem ve teknolojilerinin performanslarındaki katlanarak hızlanan artış ve (3) nesneler, insanlar ve süreçler arasındaki internet tabanlı bağlantı sıklıklarının ve çeşitlerinin çoğalması olmak üzere üç başlık altında toplamışlardır. Bu gelişmelere ek olarak IBM, Microsoft ve Google ve gibi önde gelen teknoloji şirketlerinin yapay zekâ alanına dönük yatırımlarını arttırması ve aynı şirketler tarafından "Yapay Zekânın Demokratizasyonu" hedefi doğrultusunda kod kütüphaneleri ve geliştirme araçlarının ücretsiz olarak sunulması sayesinde ise çok daha fazla kişinin yapay zekâ uygulamalarının geliştirme süreçlerine katılabilmeleri sağlanmıştır (Garvey, 2018).

Rich (1983) en genel tanımıyla yapay zekâyı mevcut durumda insanlar tarafından daha iyi yapılan işlerin bilgisayarlar tarafından yapılabilmesine dönük çalışmaların bütünü olarak ifade etmiştir. Bir işin bilgisayarlar tarafından insan gibi yapılması haricinde bilgisayarların insan gibi düşünebilmesini vurgulayan bir tanıma göre ise yapay zekâ kavramı insana özgü zihinsel yetilerin hesaplamalı modeller aracılığıyla incelenmesine dönük çalışmaları nitelemektedir (Charniak & McDermott, 1985). Bu iki tanımın kapsamıyla paralel olarak Russel ve Norvig (2010) yapay zekâ üzerine yapılan tanımların insan gibi düşünmek, insan gibi davranmak, rasyonel düşünmek ve rasyonel davranmak olarak dört kategoride toplanabileceğini ileri sürmüşlerdir. Bu yaklaşıma göre Rich (1983) tarafından yapılan tanım insan gibi davranmayı; Charniak ve McDermott (1985)'in tanımı ise rasyonel düşünmeyi yapay zekâ açıklamalarının merkezine yerleştirmektedir. Nilsson (1998) ise yapay zekâyı rasyonel davranma çerçevesinden irdeleyen bir yaklaşımla algılama, sorgulama, öğrenme ve iletişim kurma gibi davranışları aklın yol göstericiliğinde sergilemesi beklenen insan yapımı makinelerin geliştirilmesi olarak ele almıştır. Ayrıca insan zihninin işleyişi ile bilgisayarların çalışma mekanizmasını bağdaştıran teorilerden yola çıkarak yapay zekânın bilgisayarların düşünebilir hale getirilmesi ve bu doğrultuda akıllı makineleri yaratma çabası olarak tanımlandığı da görülmektedir (Haugeland,1985; Winograd, 1991). Yapılan tanımlar ortaklaştırıldığında yapay zekânın insan davranışlarını ve düşüncelerini rasyonel bir çerçevede taklit ve icra eden teknolojiler ve bu teknolojilerin geliştirilmesini amaçlayan çalışmaları kapsadığı söylenebilir.

Yapay zekânın son yıllardaki yükselişini açıklamak için kullanılan kavramların başında "Büyük Veri (Big Data)" gelmektedir. Büyük veri yalnızca devasa miktardaki veriyi ifade eden hacim (*volume*) ile değil aynı zamanda verinin üretim ve değişim hızını tanımlamak için kullanılan yoğunluk (*velocity*) ve işlenecek veriye ait kaynakların ve veri türlerinin kalabalıklaşmasını niteleyen çeşitlilik (*variety*) kavramlarıyla da açıklanmaktadır (Kitchin & McArdle, 2016). Diğer yandan Boyd ve Crawford (2012) ise Büyük Verinin kültürel, teknolojik ve bilimsel bir olgu olduğunu öne sürerek hesaplama gücü ve algoritmaların etkililiğinin büyük veri setlerinin işlenmesi doğrultusunda maksimize edilmesi, bu veri setlerinin sosyal, ekonomik ve teknik kararlara yön gösterecek örüntülerin ortaya çıkarılması için analiz edilmesi ve büyük verinin önceden erişilmesi imkânsız olan yüksek düzeydeki bilgilere sahip olmanın aracı olduğuna dair genel kabul ve mitlerin yaygınlaşması olmak üzere üç sürecin etkileşimine dayandığını belirtmişlerdir. Kısacası Büyük Veri, hem veri üretimi ve analizindeki teknolojik ilerlemeleri ve bu sayede verinin niceliksel ve niteliksel açıdan geçirdiği dönüşümü ifade eden hem de bu

dönüşümün bilgi üretiminde doğurduğu yeni ihtiyaçları vurgulayan geniş bir kavramdır (De Mauro ve diğ., 2015).

Büyük Veri olgusu ile paralel olarak yapay zekâ çalışmalarına hız kazandıran bir diğer gelişme ise derin öğrenme başlığı altında kullanılan tekniklerin başarı oranı yüksek yapay zekâ modellerinin geliştirilmesinde kullanılmalarıdır (Deng, 2014). Akpınar (2017)'a göre yapay zekâ çalışmalarının birinci döneminde bilgisayarların daha zeki hale getirilmesi için nasıl programlanması gerektiğine odaklanılıyorken bugün ise bilgisayarların öğrenebilir şekilde nasıl programlanacağı sorusu gündemdedir. Bir makine öğrenmesi alt dalı olan derin öğrenme disiplini içerisinde kullanılan yöntemler büyük veri ile geliştirilen yapay zekâ modellerinde başarılı sonuçlar göstermiştir ve özellikle bilgisayarlı görme, ses işleme ve doğal dil işleme gibi günümüzün önde gelen yapay zekâ uygulama alanlarında sıklıkla kullanılmaktadır (Jordan & Mitchell, 2015).

Özet olarak, yapay zekâ araştırma ve uygulamaları son yıllarda bir dizi gelişmeye bağlı olarak yükselişini sürdürmekte ve etkilediği disiplinlerin sayısını artırmaktadır. Yapay zekâ teknolojilerinin bir işi insan gibi yaparak veya rasyonel düşünerek insan tarafından icra edilen iş ve süreçlerin otomatizasyonuna katkı sağlamalarının yanında insanlara kurumsal çerçevede belirli kararların alınmasında yol gösterici olmaları da mümkündür. Büyük veri yığınlarından anlamlı bilgilerin elde edilmesi ve karar alma süreçlerine yol göstermesi amacıyla kullanılan yapay zekâ yöntemleri veri madenciliği adı verilen süreç içerisinde incelenmektedir.

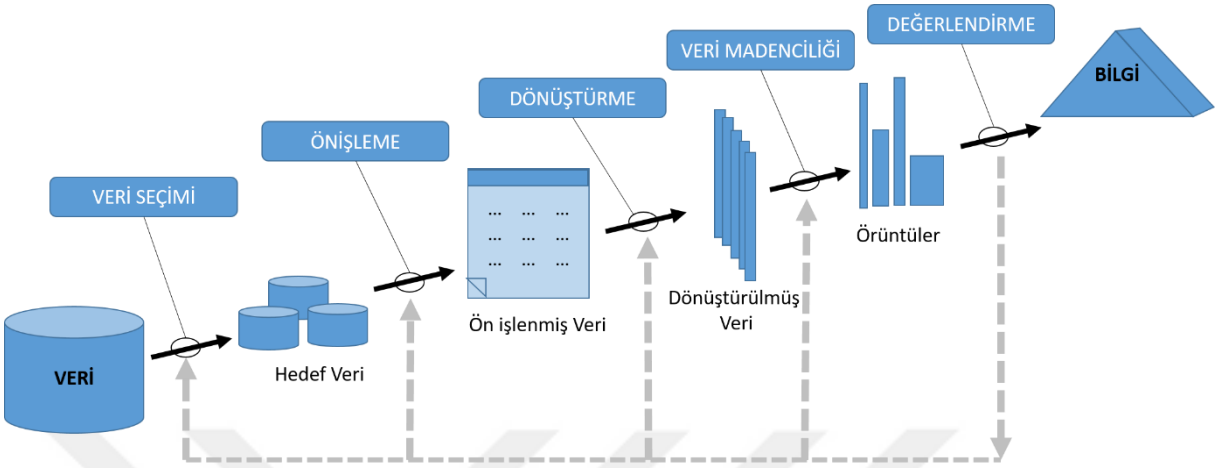
2.2.VERİ TABANLARINDA BİLGİ KEŞFİ

Bilgi teknolojilerinin ve özellikle veri tabanı endüstrisinin evrimi 1960'larda veri toplama/veri tabanı yaratımı evresinden, 1970'lerde gelişim gösteren veri tabanı yönetim sistemlerine ve 1980'lerden sonra ise veri madenciliği ve veri tabanlarında bilgi keşfi evresine ulaşılmıştır (Han & Kamber, 2006). Fayyad ve diğerleri (1996) tarafından yapılan ve genel kabul gören bir tanıma göre "veri tabanlarında bilgi keşfi veri içerisindeki geçerli, özgün, potansiyel olarak kullanışlı ve sonuçları anlaşılabilir olan örüntülerin saptandığı karmaşık bir süreçtir" (s. 40). Yazarlara göre veri tabanlarında bilgi keşfi veri yığını içerisindeki kullanışlı bilgilerin keşfedildiği bütünsel bir süreci ifade ederken veri madenciliği ise bu süreç içerisinde örüntülerin ortaya çıkarılması amacıyla ilgili probleme dönük algoritmaların kullanıldığı adımdır. Aynı şekilde Kreutzer ve Sirrenberg (2020) de yapay zekâ uygulamalarının kapsamlı hedefinin veri tabanlarında bilgi keşfi olduğunu vurgulamışlar ve veri madenciliğinin aksine

veri tabanlarında bilgi keşfinin yalnızca veri ile ilgili işlemleri değil aynı zamanda ulaşılan sonuçların değerlendirilmesini de içerdiğine dikkat çekmişlerdir. Öte yandan bazı tanımlamalarda veri madenciliği ve veri tabanlarında bilgi keşfi arasındaki ayrıma çekinceyle yaklaşıldığı da görülmektedir. Örneğin, Han & Kamber (2006) veri madenciliğinin veri tabanlarında bilgi keşfinin bir adımı olduğunu kabul etmekle beraber endüstride, medyada ve araştırmacılar arasında veri madenciliği tanımının giderek popülerlik kazandığını göz öne almak gerektiğini savunmaktadırlar.

Kimi zaman makine öğrenmesi ile de yakın anlama gelecek şekilde kullanılmakta olan veri madenciliğinde veri yığınları içerisinden başka yöntemlerle elde edilmesi mümkün olmayan, anlamlı ve önceden bilinmeyen bilgi veya bilgilerin ortaya çıkarılması amaçlanmakta iken makine öğrenmesi çalışmalarının nihai amacı ise bilgisayarların insanın öğrenme eylemini taklit ederek eğitilmesi, bu sayede yeni bilgi ve becerilerin elde edilmesi ve öğrenme performansının çeşitli senaryolarla sınanarak etkililiğin artırılmasıdır (Hua ve diğ., 2009; Silahtaroğlu, 2016). Bu bağlamda veri madenciliği anlamlı makine öğrenmesi modellerinin yaratılması sürecini kapsamaktadır ve makine öğrenmesi algoritmaları hakkındaki bilgilerin geliştirilmesiyle değil bu algoritmaları veri setlerine uygulayarak ortaya çıkarılan bilgiler sayesinde bir probleme dönük olarak yeni bakış açılarının kazanılmasıyla ilgilenmektedir (Kühl ve diğ., 2019). Benzer şekilde Fayyad ve diğerleri (1996) de veri tabanlarında bilgi keşfi sürecinin veri madenciliği adımının veri içerisindeki örüntüleri ortaya çıkarmak adına makine öğrenmesi, örüntü tanıma ve istatistik yöntemlerini temel aldığını belirtmişlerdir.

Şekil 2.1'de görüldüğü gibi veri tabanlarında bilgi keşfi süreci döngüsel ve kendi içinde etkileşime açık bir dizi adımdan oluşmaktadır.



Şekil 2.1: Veri Tabanlarında Bilgi Keşfi Adımlarına Genel Bakış (Fayyad ve diğ., 1996, s.41)

Veri tabanlarında bilgi keşfi sürecinin temel adımları şu şekilde özetlenmiştir (Brachman & Anand, 1996): Uygulama alanının anlaşılması ve bilgi keşfinin amacının belirlenmesi; bilgi keşfinin odaklandığı probleme dönük olarak uygun verinin seçilerek hedef veri setinin oluşturulması; kirli ve kayıp veri için ön işleme ve temizleme işlemlerinin yapılması, veri setinin problemi temsil eden özniteliklere indirgenmesi, bilgi keşif sürecinin amaçları ile veri madenciliği yönteminin eşleştirilmesi, uygun veri madenciliği algoritmasının seçilmesi, veri madenciliği ile modelleme işleminin gerçekleştirilmesi, ortaya çıkarılan örüntülerin yorumlanması, elde edilen bilgi veya bilgilerin değerlendirilmesi ve kullanılması.

Han & Kamber (2006) ise veri tabanlarında bilgi keşfi süreci adımlarını şu şekilde listelemişlerdir:

1- Veri Temizleme: gürültülü ve hatalı verinin temizlenmesi,
2- Veri Bütünleştirme: birden fazla veri kaynağının bütünleştirilmesi,
3- Veri Seçimi: analizin amacı ile ilgili verinin veri tabanından elde edilmesi,
4- Veri dönüştürme: verinin madencilik işlemlerine uygun şekilde bütünleştirme veya özetlemeye tabi tutularak dönüştürülmesi,
5- Örüntünün değerlendirmesi: bilgiyi temsil eden örüntülerin belirli ölçütlere dayanarak değerlendirilmesi,

6- Bilginin sunumu: görselleştirme ve bilgi temsil tekniklerini kullanarak veri madenciliği ile elde edilen bilginin sunulması.

Veri tabanlarında bilgi keşif süreci adımlarının sayısı farklı yaklaşımlara göre değişiklik gösterse de yaklaşımların tümü veri yığınlarından anlamlı ve kullanışlı bilginin elde edilmesi için yapılması gerekenleri aşamalı ve döngüsel bir sistematik içerisinde sunmaktadırlar.

2.2.1. Veri Seçimi

Veri tabanlarında bilgi keşif sürecinin amacıyla uyumlu olarak bir dönemi kapsayan bütünleşik veri tabanlarına veri ambarı denmektedir (Silahtaroğlu, 2016). Veri ambarları üzerinden gerçekleştirilecek veri seçiminde çözülmeye çalışılan problem veya ortaya çıkarılması beklenen bilgi ve örüntüler ile potansiyel olarak ilişkili olması beklenen verinin elde edilmesi gerekmektedir. Veri madenciliği çalışmalarında genellikle bir problem için kullanılacak veri bütünleşik halde tek bir kaynakta bulunmadığından dolayı farklı veri tabanları ve kaynaklarından elde edilen veri de analiz süreçlerine dâhil edilebilir (Feelders ve diğ., 2000).

2.2.2. Veri Önişleme

Veri madenciliği algoritmaları veri seti içerisinde kayıp, kirli, gürültülü veya yanlış girilmiş değerlerin olmadığı varsayımıyla uygulanmaktadır. Bundan dolayı bir veri madenciliği uygulamasında modelin düşük performans göstermesini, veri içindeki örüntülerin gizli kalmasını ve anlamsız çıktıların görülmesini önleme amacıyla çabanın büyük bir kısmı temizleme, doldurma ve normalizasyon gibi veri ön işleme adımları için harcanmaktadır (Zhang ve diğ., 2003).

Veri temizleme, özellikle kayıp veya eksik veriye sahip kayıtların veri setinden çıkarılması işlemini içerdiği gibi tüm kayıp alanlara ortalama değerin girilmesi ve ayrıca çeşitli yöntemlerle kayıp değerlerin tahmin edilmesi ile de yapılabilmektedir (Silahtaroğlu, 2016). Doldurulan verinin amaçlanan madencilik çalışması için doğru bir temsile sahip olmaması ihtimalinden dolayı bu yöntemlere kıyasla kayıp verinin regresyon, Bayesyen veya bir karar ağacı yöntemiyle tahmin edilmesi yöntemi sıklıkla tercih edilmektedir (Han & Kamber, 2006). Bu sayede kayıp veriye rastlanan kayıtlarda el ile doldurma ve ortalama değerin atanması yöntemlerine kıyasla özgün değerlerin yer alması sağlanabilmektedir.

Veri önişleme aşamasında giderilmesi gereken bir diğer sorun ise gürültülü veridir. Veri girişinde yapılan hatalar ve yanlış kodlamalar gürültülü değerlere sebep olabilmekte ve veri bir ortam tarafından üretiliyorsa depolama ve dönüştürme ile ilgili işlemler de veri setinde gürültülü değerlerin gözükmesine yol açabilmektedir (Pyle, 1999). Gürültülü veri yanlış bir değerde ise veri setinden çıkarılmasına veya uç bir değerde ise dönüştürülmesine karar verilebilir.

2.2.3. Veri Dönüştürme

Çeşitli kaynaklardan elde edilen ve bir veri setinde bütünleştirilen veri tutarsızlık, gürültü ve aykırı (uç) değerlerden kaynaklanan sorunların giderilmesi amacıyla dönüştürme işlemlerine tabi tutulmalıdır. Özellikle gürültülü verinin dönüştürülmesi doğrultusunda ilgili özniteliğe ait değerlerin gruplara ayırarak grup ortalamasının elemanlara atandığı kutulama yöntemi, gruplar içindeki alt ve üst sınır değerleri yardımıyla gürültünün giderilmesi yöntemi ve kümeleme ile uç değerlerin tespit edilerek yeni değerlerin atanması yöntemi kullanılabilmektedir (Silahtaroğlu, 2016).

Veri setinde yer alan değerlerin ölçü birimleri hem analiz sürecini etkileyebilmekte hem de uygulanacak algoritmaların çalıştıkları veri tipleriyle uyumsuzluk gösterebilmektedir. Bundan dolayı özellikle farklı büyüklükte değerlerin bulunduğu veri setlerinde tüm değerlerin 0 ve 1 veya -1,+1 aralığında değerlere sahip olmasını ifade eden normalizasyon işlemi gerçekleştirilmelidir (Akpınar, 2017).

Büyük miktarda veri her ne kadar veri madenciliği uygulamaları için gerekli olsa da veri setindeki özniteliklerin veya boyutların miktarındaki fazlalık algoritmaların performansını olumsuz yönde etkileyebilmektedir. Akpınar (2017)'a göre

> "Artan boyut sayısının yol açtığı ortak sorun, boyut sayısı arttıkça, uzay hacminin artması, bunun paralelinde mevcut verinin giderek uzayda seyrekleşmesi ve istatistik anlamda elde edilen sonuçların güvenilirliğinin azalmasıdır" (s.144).

Boyut indirgeme, bir veri setindeki özniteliklerin sayısının orijinal (indirgenmemiş) veri setinin temsiliyetini kaybetmeden azaltılması işlemidir (García ve diğ., 2015). Bu sayede sonuca etkisi bulunmayan özniteliklerin çıkarılması veya aynı değişkeni temsil eden birden fazla özniteliğin tek boyuta indirgenmesi ile birlikte orijinal veri setinden daha etkili ve güvenilir sonuçlar elde

etmek mümkün olmaktadır. Ayrıca sayısal indirgeme ve ayrıştırma gibi veri setindeki değerlerin aralıklı veya kategorik değerlerle yer değiştirerek kullanıldığı yöntemler de bir boyut indirgeme ve veri dönüştürme yöntemi olarak tercih edilmektedir (Han & Kamber, 2006).

2.2.4. Modelleme

Chung ve Gray (1999), uygun algoritmanın seçilmesi ve modellemenin gerçekleştirilmesi aşamalarının veri tabanlarında bilgi keşfi sürecinin en önemli noktaları olduğunu vurgulamışlardır. Modelleme aşamasında makine öğrenmesi ve istatistik temelli yöntemler problemin ihtiyacı göz önünde bulundurularak seçilmekte ve önişleme ve dönüştürme aşamalarından geçirilen veri seti üzerinde yürütülmektedir.

Veri madenciliği modelleri kullanılarak yürütülen bilgi keşfi süreci Fayyad ve diğerleri (1996) tarafından iki temel kategoride incelenmektedir. (1) Tahmin (*Prediction*), bir veri tabanındaki değişkenlere ait örüntülerin bulunarak gelecekteki girdilerin tahmininde kullanılmasını ifade ederken; (2) Betimleme (*Description*) ise veri tabanı içindeki örüntülerin anlaşılabilir ve kişiler tarafından yorumlanabilir biçimlerde ortaya çıkarılmasıdır. Yazarlar bu noktada bahsi geçen iki kategori arasındaki ayrımın tahmin modellerinin de betimleme içerebilmesinden dolayı keskin olamayabileceğini ancak veri madenciliğinin yürütülme amacının anlaşılması açısından bu ayrıma ihtiyaç olduğunu belirtmişlerdir.

Modelleme aşamasında birden fazla yöntem veri tabanlarında bilgi keşfi sürecinin döngüsel yapısına uygun olarak veri seti içerisindeki örüntülerin farklı biçimlerde ortaya çıkarılmasını sağlamak adına uygulanmaktadır. Ancak kullanılacak yöntemlere veri madenciliğinin amacı ve veri setinin yapısı göz önüne alınarak karar verilmesi gerekmektedir (Kantardzic, 2020). Veri madenciliği sürecindeki modelleme aşamasında makine öğrenmesi algoritmaları kullanılmaktadır.

2.2.5. Modelin Performansının Ölçülmesi

Kullanılan algoritma ile birlikte modelin performansına etki eden bir diğer unsur ise eğitim ve test verisini ayırma yöntemidir. Danışmanlı öğrenmede ve özel olarak sınıflandırma problemlerinde amaç veriyi önceden tanımlanmış ayrı sınıflara etkili şekilde atamaktır. Bu amacın gerçekleştirilmesi için mevcut verinin bir kısmı eğitim seti olarak ayrılır ve modelin eğitilmesinde kullanılırken; veri setinin geri kalan kısmı test verisi olarak ayrılır ve öğrenen

modelin performansının test edilmesi amacıyla kullanılır (Korjus ve diğ., 2016). Eğitim ve test verisinin ayrılması ve sınıflandırma performansının test edilmesi işlemini gerçekleştirmek amacıyla sıklıkla kullanılan yöntemlerden bazıları hold-out,, tekrarlı hold-out ve çapraz geçerleme (*cross validation)* olarak sıralanmaktadır.

2.2.5.1. Hold-out Yöntemi

Hold-out yönteminde veri seti içerisinden bir defaya mahsus olarak eğitim setini oluşturacak örnekler rasgele seçilmekte ve geri kalan örnekler ise test verisi için kullanılmaktadır. Veri madenciliği uygulamalarında genellikle test verisi için tüm veri setinin %10 ila %30'u ayrılırken, eğitim seti için tüm veri setinin %90 ila %70'i ayrılmaktadır (Berrar, 2018). Diğer yandan veri setinin büyüklüğü olanak tanıdığı ölçüde eğitim ve test verisini eşit miktarda (%50-%50) örneklere ayırmak da mümkündür. Bu durumda ikiye ayrılan veri setinin her bir bölmesi hem eğitim hem test verisi olacak şekilde iki defa etkililik ölçümü gerçekleştirmek mümkün olmaktadır (Tempola ve diğ., 2021).

2.2.5.2.Tekrarlı Hold-out Yöntemi

Tahminleyici modelin eğitilmesi için seçilecek olan örneklemin hedef niteliğe ait sınıfları yeterince temsil edememesi veya belirli sınıfları aşırı derecede temsil etmesi sapma (*bias)* durumuna sebep olmaktadır (Tancev & Pascale, 2020). Veri setinin bir defa eğitim ve test verisine ayrılması yöntemi olan standart hold-out yerine bu işlemin birden fazla seçimle gerçekleştirildiği tekrarlı hold-out yöntemi sapma durumunun azaltılması için kullanılmaktadır. Tekrarlı hold-out yönteminde, birden fazla ayırma işlemi gerçekleştirilerek her döngüde elde edilen performans değerlendirme ölçütlerinin ortalaması alınarak sınıflandırma performansı belirlenmektedir (Torunoğlu ve diğ., 2011).

2.2.5.3.Çapraz Geçerleme

Çapraz geçerleme, verinin farklı setler halinde, tekrarlanan döngülerle ve her tekrarlamada geçerleme verisi kullanılacak şekilde eğitilerek test edildiği yöntemdir (Molinaro ve diğ., 2005). Çapraz geçerleme yöntemi örneklem seçiminin birden fazla ve rastgele yapılması dolayısıyla tekrarlı hold-out yöntemine benzemekte ancak çapraz geçerlemedeki tekrarlarda seçilen test setlerinin birbiriyle çakışmaması yönünden ise farklılaşmaktadır (Berrar, 2018). Bir çapraz geçerleme türü olan k-katlı çapraz geçerlemede veri seti belirlenen k sayıda parçaya

ayrılmakta ve her bir parça sırasıyla sınıflandırıcı modelin test edilmesi amacıyla kullanılmaktadır. Modelin eğitimi ise her bir tekrarda geriye kalan k-1 parça kullanılarak gerçekleştirilmektedir.



Şekil 2.2: K-Katlamalı Çapraz Geçerleme Adımları

Şekil 2.2'de k değerinin ve dolayısıyla veri setinin ayrılacağı alt set sayısının 10 olarak belirlendiği bir k-katlamalı çapraz geçerleme örneğine ait adımlar görülmektedir. İlk adımda birinci set geçerleme seti olarak ayrılmakta ve geri kalan setlerle ise modelin eğitimi gerçekleştirilmektedir. Bir sonraki adımda ise ikinci set geçerleme seti olarak kullanılmakta ve k sayıda adım sonunda elde edilen performans ölçümlerinin ortalaması modelin değerlendirilmesi için kullanılmaktadır (Berrar, 2018).

2.2.5.4.Karmaşıklık Matrisi

Makine öğrenmesinde sınıflandırma problemlerinin performans ölçümleri karmaşıklık matrisi *(confusion matrix)* adı verilen kare matris biçiminde görselleştirilen bir tablo aracılığı ile yapılmaktadır (Caelen, 2017). Karmaşıklık matrisi genel olarak belirli bir davranışın hangi sıklıkla doğru ve yanlış algılandığı hakkında bilgi vermektedir. Sınıflandırma problemleri söz konusu olduğunda modelin doğru ve yanlış tahmin ettiği örnek miktarları karmaşıklık matrisinde yer almakta ve bu değerler kullanılarak hesaplanan etkililik, duyarlılık, özgüllük ve F-skoru gibi değerler performans değerlendirmesi amacıyla kullanılmaktadır (Ruuska ve diğ., 2018).

Tablo 2.1: İkili Sınıflandırma Problemleri için Karmaşıklık Matrisi

		TAHMİN	
		Doğru	**Yanlış**
GERÇEK	**Doğru**	DP (Doğru Pozitif)	YN (Yanlış Negatif)
	Yanlış	YP (Yanlış Pozitif)	DN (Doğru Negatif)

Tablo 2.1 'de gösterilmekte olan karmaşıklık matrisinde dikey eksende modelin doğru ve yanlış olarak sınıflandırdığı örnek miktarları yatay eksende yer alan gerçek sınıf değerinin doğru veya yanlış olması göz önüne alınarak yazılmaktadır. Doğru Pozitif (DP) ve Doğru Negatif (DN) modelin pozitif ve negatif sınıflara yönelik doğru tahmin ettiği örneklerin sayısını göstermektedir. Yanlış Pozitif (YP) ve Yanlış Negatif (YN) değerleri ise modelin pozitif ve negatif sınıflara yönelik yanlış tahmin ettiği örnekleri ifade etmektedir.

Karmaşıklık matrisi kullanılarak hesaplanan doğruluk değeri modelin genel performansının belirlenmesi amacıyla toplamda doğru tahmin edilen örnek miktarının tüm örneklem miktarına bölünmesiyle elde edilmektedir (Sokolova ve diğ., 2006).

$$doğruluk = \frac{DP + DN}{DP + YP + YN + DN} \tag{2.1}$$

Sınıflara ait örneklerin eşit veya birbirine yakın dağılım göstermediği veri setlerinde veya modelin tek bir sınıfın örneklerini doğru tahmin etmesinin beklendiği durumlarda genel sınıflandırma başarısını doğruluk ölçütüyle belirlemek yeterli ve açıklayıcı olmamaktadır (Ye, 2003). Sınıflandırma performansını ayrı sınıfları temel alarak ölçülmesinde duyarlılık ve özgüllük değerleri kullanılmaktadır ve bu değerler modelin bir sınıfa ait doğru tahmin ettiği örnek sayısının ilgili sınıfın toplam gerçek örnek sayısına bölünmesiyle elde edilmektedir.

$$duyarlılık = \frac{DP}{DP + YN} \tag{2.2}$$

$$\text{özgüllük} = \frac{DN}{YP + DN} \tag{2.3}$$

Dengesiz veri setlerinde bir sınıflandırma modelinin pozitif sınıfa ait başarısı (duyarlılık) negatif sınıfa ait başarısından (özgüllük) daha yüksek olabilmektedir. Bundan dolayı duyarlılık ve kesinlik değerlerinin ağırlıklı ortalaması alınarak hesaplanan F-skoru kullanılmaktadır.

$$kesinlik = \frac{DP}{DP + YP} \tag{2.4}$$

$$F - Skoru = \frac{2 \times kesinlik \times duyarlılık}{kesinlik + duyarlılık} \tag{2.5}$$

Kesinlik ölçütü pozitif sınıfa ait doğru tahminlerin tüm pozitif tahminlerin sayısına bölünmesiyle elde edilmekte ve pozitif sınıf tahminlerinin ne kadarının doğru olduğunu belirlemek için hesaplanmaktadır. Aritmetik ortalama yerine harmonik ortalamanın kullanılmasının sebebi aralarında büyük fark olabilen duyarlılık ve kesinlik değerlerinden düşük değerde olana daha yakın bir değer elde edilerek veri setindeki dengesizliğe karşı bir denge durumunun gözetilmesidir (Wong, 2020).

Özellikle ikili sınıflandırma problemlerinde kullanılan bir performans değerlendirme yöntemi olan ROC *(Receiver Operating Characteristic)* duyarlılık değeri ile "$1 - \text{Özgüllük}$" şeklinde hesaplanan yanlış pozitif oranının x ve y eksenlerinde oluşturulan eğirişidir. AUC (*Area Under Curve)* ise bu eğrinin altında kalan alanın 0 ve 1 arasında bir değerle ölçülerek pozitif ve negatif sınıfları birbirinden ayırma derecesini ifade etmektedir (Sokolova ve diğ., 2006).

2.2.6. Değerlendirme

Veri tabanlarında bilgi keşfi sürecinde veri madenciliği ve modelleme aşamalarının ardından ortaya çıkarılan örüntülerin yorumlanması ve işlevlendirilmesi gerekmektedir. Yorumlama aşamasındaki uzman görüşlerin de değerlendirmesi sonucunda keşif sürecinin herhangi bir adımına daha doğru sonuçlar elde etme amacıyla geri dönülebilmektedir. Öte yandan yorumlama aşamasında ortaya çıkarılan örüntülerin açıklanması veya kullanılan modelin ve

çıktılarının incelenmesi amacıyla görselleştirme yöntemi tercih edilebilmektedir (Fayyad ve diğ., 1996).

Fayyad (2001), veri madenciliği aşaması sonrasında mevcut bilgilerin geçmiş bilgilerle eşleştirilmesinin ve elde edilen bilgilerin doğru yorumlanmasının bilgilerin kullanışlılığını saptamak açısından gerekli olduğuna dikkat çekmiştir. Yazara göre veri madenciliği ile ortaya çıkarılan örüntü ve bilgilerin kullanışlı sayılmalarının ölçütü probleme özgü olarak belirlenen ilginçlik düzeyini aşıyor olması, diğer bir deyişle önceden bilinmeyen ve mevcut bilgilerle kıyaslandığında ilginç bulunan bilgilerin gerçek anlamda keşfedilmiş bilgi sayılabileceğidir. Kısacası veri tabanlarında bilgi keşfi sürecinin sonunda yapılan değerlendirme veri madenciliğinin çıktısı olan bilgi ve örüntülerin yorumlanarak yeni bilgilerin tanımlı hale getirilmesini amaçlamaktadır.

2.3.MAKİNE ÖĞRENMESİ

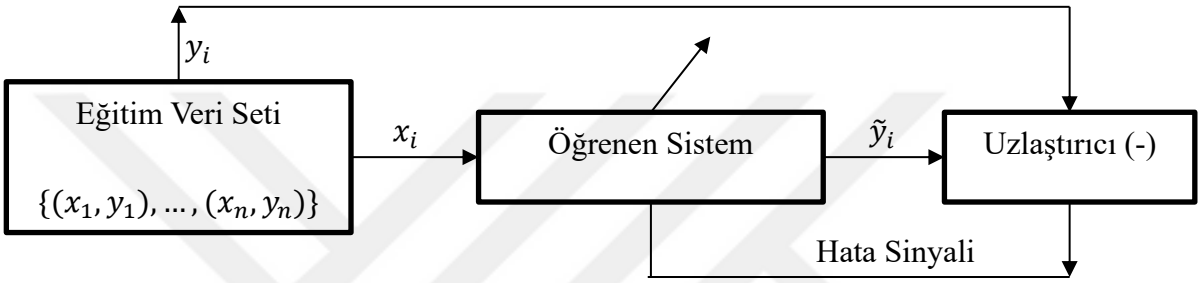
Yapay zekâ ve makine öğrenmesi kavramları son yıllarda artan uygulamalar ve tanımları itibariyle ortak yönlere sahip olmalarından dolayı birbirlerinin yerine sıklıkla kullanılmaktadır. Ancak bu kavramların tanımlama açısından birbirlerinden farklılaştığı bilinmektedir. Yapay zekâ, insanlar tarafından yürütülen faaliyetlerin bilgisayarlar tarafından yürütülmesi doğrultusunda yapılan araştırma ve uygulama alanını niteleyen genel bir kavramdır. Yapay zekâ alanının özel bir alt boyutu olarak ise makine öğrenmesi kavramı çeşitli problemlerin hesaplamalı yöntem ve kaynaklar ile çözülmesi amacıyla etkili algoritmaların tasarlanması ve geliştirilmesi olarak tanımlanmaktadır (Mohri ve diğ., 2021). Örneğin makine öğrenmesi yöntemlerden bir tanesi olan derin öğrenme, özellikle son yıllarda görüntü işleme ve konuşma tanıma uygulamalarında sıklıkla kullanılmaktadır (Kühl ve diğ., 2019).

Makine öğrenmesi algoritmaları, bir öğrenme problemine dönük performans ölçütlerini iyileştirme ve problemle ilgili bir görevin gerçekleştirilmesinde deneyimlere dayanarak herhangi bir faaliyette bulunacak olan modellerin geliştirilmesinde kullanılmaktadırlar. Bu doğrultuda kullanılan yöntemler danışmanlı öğrenme, danışmansız öğrenme ve pekiştirmeli öğrenme başlıkları altında incelenmektedir.

2.3.1. Danışmanlı Öğrenme

Danışmanlı öğrenme (*supervised learning*) ile önceden etiketlenmiş örneklere sahip bir eğitim verisi kullanılarak modelin eğitilmesi ve eğitim sırasında modele gösterilmeyen örnekler

üzerinde tahmin yürütülmesi amaçlanmaktadır (Kantardzic, 2020). Diğer bir deyişle makineler insan öğrenmesine benzer şekilde geçmiş deneyimleri gerçek dünya problemleri üzerinde uygulamak üzere çıktı değerleri belli olan eğitim verisi ile eğitilmekte ve bu sayede yeni örnekleri tahmin edebilmektedirler (Liu, 2007). Danışmanlı öğrenmede amaç eğitim verisine ait etiketlenmiş örneklerin temel alınarak yapay bir sistemin kurulması ve bu sisteme ait girdiler ve çıktılar arasındaki ilişkinin bilgisinin bir modele dönüştürülerek yeni girdiler üzerinde tahmin yürütülmesidir.



Şekil 2.3: Danışmanlı Öğrenme Diyagramı (Liu & Wu, 2012, s.3244)

Şekil 2.3'te gösterilen Danışmanlı Öğrenme diyagramında 'x' sistem girdilerini, 'y' sistem çıktılarını, 'i' ise eğitim veri seti indekslerini nitelemektedir. Danışmanlı öğrenme süreci içerisinde Öğrenen Sistem x_i olarak verilen sistem girdileri ile eğitilmekte ve $\tilde{y}_i$ çıktılarını üretmektedir. Öğrenen Sistem'in çıktıları y_i olarak gösterilen veri seti içerisindeki gerçek sınıf etiketleriyle karşılaştırılmakta ve bu iki sonuç arasındaki fark Hata Sinyali olarak hesaplanarak Öğrenen Sistem'e iletilmektedir. Bu süreç sonunda Öğrenen Sistem'e ait parametreler yeniden düzenlenerek gerçek sonuçlarla tahmin edilen sonuçlar arasındaki fark azaltılmaya ve optimal bir model elde edilmeye çalışılır (Liu & Wu, 2012).

Danışmanlı öğrenme, eğer kategorik değerlerin veya sınıf etiketlerinin tahmin edilmesi amacıyla uygulanacaksa sınıflandırma, sürekli değerlerin tahmin edilmesi için kullanılacaksa regresyon yöntemleriyle gerçekleştirilmektedir. Veri madenciliği ve makine öğrenmesi çalışmalarında sıklıkla kullanılmakta olan sınıflandırma yöntemleri, öğrenen sistemin birden fazla sınıfa ait girdi ve çıktı örneklerini (eğitim seti) temel alarak bu örnekler içerisindeki kuralları haritalandırılmasını ve bu kuralların yeni örnekler için genelleştirilmesini kapsadığı için tipik bir danışmanlı öğrenme problemi olarak değerlendirilmektedir (Osisanwo ve diğ., 2007).

Kotsiantis (2007) danışmanlı makine öğrenmesi kapsamında kullanılan sınıflandırma algoritmalarını mantık tabanlı algoritmalar (örn. karar ağaçları), algılayıcı tabanlı algoritmalar (örn. Yapay sinir ağları), istatistiksel öğrenme algoritmaları (örn. Naive Bayes), uzaklık tabanlı öğrenme algoritmaları (örn. k-En Yakın Komşu algoritması) ve Destek Vektör Makineleri olarak gruplandırmıştır. Veri madenciliği sürecinin modelleme aşamasında kullanılacak algoritmaya karar vermek için birden fazla algoritmanın performans değerlendirme ölçütleri kıyaslanmakta ve tahmin veya sınıflandırma başarısı en yüksek algoritma aranmaktadır.

Bu başlık altında öğrencilerin derse dönük motivasyonlarının tahmin edilmesi amacıyla geliştirilen sınıflandırma modelinin eğitilmesinde kullanılan k-en yakın komşu, Naive Bayes, Destek Vektör Makineleri, C5.0, CART, Rastgele Orman, ve Lojistik Regresyon algoritmaları hakkında bilgi verilmiştir.

2.3.1.1. k-En Yakın Komşu Algoritması

Cover ve Hart (1967) tarafından geliştirilen ve en basit sınıflandırıcı olarak bilinen k-en yakın komşu yöntemi, sınıflandırılmak istenen bir örneğin hangi sınıfa atanacağına en yakın komşularının ait olduğu sınıfı göz önüne alarak karar vermektedir. Diğer bir deyişle, sınıflandırılmak istenen bir x örneğinin bulunduğu uzayda mesafesi en yakın k adet örnek bulunmakta ve bu k adet yakın örneklerin çoğunluğunun ait olduğu sınıf etiketi sınıflandırılmak istenen x örneğine atanmaktadır (Kataria & Singh, 2008). k-en yakın komşu algoritmasının adımları aşağıda açıklanmıştır:

- Sınıflandırma yapılırken kullanılacak komşu sayısının (k değeri) belirlenmesi,
- Test setindeki her bir örnek için eğitim setindeki örneklere olan uzaklığın belirlenen bir ölçütle hesaplanması,
- k değerine göre en yakın komşu sayısı ve sınıflarının sıralanması,
- Test setindeki örneğin sınıf çoğunluğuna sahip olan en yakın komşuların sınıfına atanması (Kantardzic, 2020).

Bu çalışmada k-en yakın komşu algoritması adımlarında örnekler arası uzaklığın belirlenmesi amacıyla Öklid uzaklığı esas alınmıştır.

$$d(x,y) = \left(\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \right) \quad (2.6)$$

Eşitlik 2.6'da x_i ve y_i bir uzayda bulunan iki örneği ve $d(x,y)$ ise bu örnekler arasındaki mesafeyi göstermektedir.

2.3.1.2. Naive Bayes Algoritması

Naive Bayes algoritması temelini Bayes teoreminden alan olasılık temelli bir sınıflandırma yöntemidir ve amaç, belirli bir problem alanındaki bir dizi yeni veri ile bir dizi sınıflandırma alanı arasındaki en iyi eşlemeyi bulmaktır (Rish, 2001). Bu eşleşme bulunurken tüm sınıflar için bir olasılık değeri hesaplanır ve sınıflandırma en yüksek olasılığa sahip olan sınıf için yapılır (Lewis, 1998).

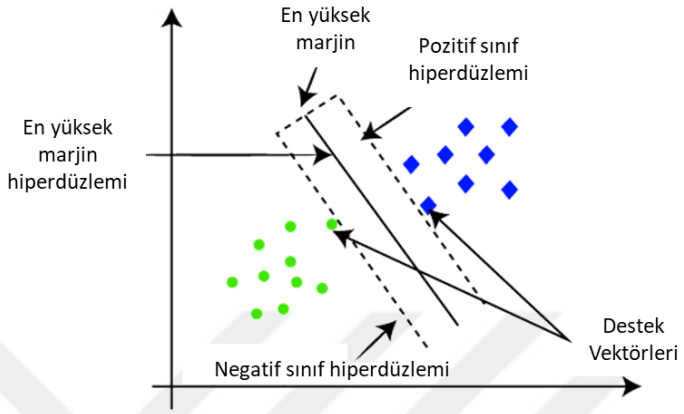
$$P(C_k|x) = P(C_k) \times \frac{P(x|C_k)}{P(x)} \quad (2.7)$$

Eşitlik 2.7'de gösterilen Naive Bayes formülünde C_k , hedef niteliğe ait sınıfları içeren C kümesinin rastgele bir elemanıdır ve x ise veri setindeki bir vektöre veya özniteliğe ait rastgele seçilmiş bir örnektir. Dolayısıyla $P(C_k|x)$ x örneğinin C_k sınıfından olma olasılığını, $P(x|C_k)$ ise x örneğinin C_k sınıfında görülme sıklığını göstermektedir. $P(C_k)$ ve $P(x)$ ifadeleri ise C_k sınıfının veri setindeki genel olasılığını ve x örneğinin eğitim seti içerisindeki görülme sıklığını göstermektedir.

2.3.1.3. Destek Vektör Makineleri Algoritması

Cortes ve Vapnik (1995) tarafından ikili sınıflandırma problemleri için geliştirilen Destek Vektör Makineleri kavramsal olarak girdi vektörlerinin yüksek boyutlu bir öznitelik uzayına doğrusal olmayan bir şekilde eşlenmesi fikrine dayanmaktadır. Bu yöntemdeki amaç ise bu uzayda hiperdüzlem adı verilen doğrusal bir karar düzlemi oluşturarak örnekleri sınıflara ayırmaktır ve bu hiperdüzlem sınıfların en yakın noktaları arasında en yüksek uzaklığı bularak oluşturulmaktır. Ayrıca doğrusal bir ayırıcı oluşturulamadığında örnekler daha yüksek boyutlu bir uzayda izdüşümleri alınarak doğrusal olarak ayrılabilir hale getirilmekte ve ayırıcı düzlemin yanlış tarafında kalan örneklerin ağırlık değerleri düşürülerek modelin performansına olan

etkileri azaltılmaktadır. Destek Vektör Makinelerinin kullanıldığı sınıflandırma problemleri bilinen yöntemlerle çözülebilecek ikinci dereceden bir optimizasyon problemine dönüştürülebilmektedir (Meyer, 2015).



Şekil 2.4: Destek Vektör Makineleri ile Doğrusal Sınıflandırma

Şekil 2.4'te görüldüğü gibi Destek Vektör Makineleri ile iki sınıfı birbirinden ayıran en yüksek marjin hiperdüzlemi veya optimal hiperdüzlem, destek vektörleri olarak adlandırılan en yakın negatif ve pozitif örneğe eşit mesafede konumlanmaktadır.

2.3.1.4. C5.0 Algoritması

C5.0 algoritması karar ağaçları temelli bir sınıflandırma yöntemidir. Karar ağaçları, eğitim veri setinde yer alan öznitelik değerlerinin sınıflandırıcı ayrım noktalarını kurallar dizisi olarak belirlenerek sınıflandırıcı modelin oluşturulduğu teknikleri kapsamaktadır. C5.0 algoritması bir önceki versiyonu C4.5 e kıyasla gürültülü ve kayıp verilere daha duyarlı olarak geliştirilmiş bir sınıflandırıcıdır ve aşırı uyum ve hatalı budama gibi sorunları giderilmiştir. Ayrıca kurallar setini oluştururken sınıflandırmaya etkisi olan ve olmayan öznitelikleri ayırmada önceki algoritmalara göre daha başarılı olduğu bilinmektedir (Pandya & Pandya, 2015).

C5.0 algoritması karar ağacındaki ayırma potansiyeli en yüksek özniteliği saptarken örnekleri ayırmada homojenlik gösteren değerleri entropi veya bilgi kazanımı hesaplayarak belirlemeye çalışır ve en yüksek entropi değerine sahip olan öznitelik ayrım için kullanılmaktadır. Bilgi kazancına sahip olmayan özniteliklerin varlığı durumunda ise karar düğümüne beklenen değer atanmaktadır (Akpınar, 2017).

$$\text{Entropi(S)} = \sum_{i=1}^{c} -p_i \log_2(p_i) \tag{2.8}$$

Eşitlik 2.8'de, verili bir S data setinde c sayıda farklı sınıf etiketi gösterilmekte ve p_i değeri örneklerin i sınıfına ait olma oranlarını ifade etmektedir.

2.3.1.5. CART Algoritması

C5.0 gibi bir karar ağacı algoritması olan CART *(Classification and Regression Trees)*, ayırma ölçütü fonksiyonu olarak entropi yerine Gini indeksi yöntemini kullanmaktadır ve her bir düğüm için ikili dallar oluşturmaktadır (Breiman ve diğ., 2017). Dolayısıyla öznitelikler ve değerleri kök düğümden başlamak üzere her aşamada iki alt düzeye ayrılmakta ve sonraki düzeylerde örneklerin sınıflandırılabilmesi adına düğümlerin aşamalı olarak homojenleştirilmesi amaçlanmaktadır (Akpınar, 2017). Kategorik değişkenlerini içeren sınıflandırma problemleri haricinde sürekli değişkenlerin varlığında regresyon ağacı olarak da kullanılmaktadır (Chang ve diğ., 2006)

$$Gini_{sol} = 1 - \sum_{i=1}^{k} \left(\frac{L_i}{T_{sol}}\right)^2 \tag{2.9}$$

$$Gini_{sağ} = 1 - \sum_{i=1}^{k} \left(\frac{R_i}{T_{sağ}}\right)^2 \tag{2.10}$$

$$Gini_j = \left(\frac{1}{n}|T_{sol}|Gini_{sol} + \left|T_{sağ}\right|Gini_{sağ}\right) \tag{2.11}$$

Eşitlik 2.9, Eşitlik 2.10 ve Eşitlik 2.11'de Gini indeksinin hesaplama adımları gösterilmektedir ve k hedef niteliğe ait sınıf sayısını, $T_{sağ}$ ve T_{sol} sırasıyla ağacın sağ ve sol düğümlerinde yer alan toplam örnek miktarını ifade etmekteyken R_i ve L_i ise sağ ve sol düğümlerde i kategorisine ait örnek miktarını göstermektedir. Ardından $Gini_j$ değeri en küçük olan değişken bölünme için kullanılır ve düğüm olarak seçilmektedir. Ağacın yapılandırılması tamamlanana kadar diğer öznitelikler için de aynı işlemler sürdürülmektedir (Rutkowski ve diğ., 2014).

2.3.1.6. Rastgele Orman Algoritması

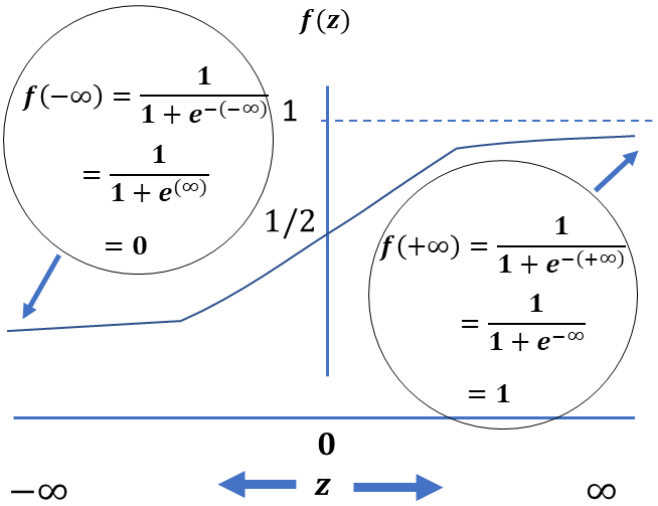
Rastgele Orman algoritması diğer karar ağacı algoritmaları gibi örnekleri sınıflandıran bir ağaç yapısı oluşturmakta ancak önyükleme olarak anılan eğitim setini birbirinden farklı kümelere ayırma işleminden ve ağacın hangi öznitelik ve değerlerle dallara ayrılacağının belirlenmesinde rastgelelik özelliğinden faydalanmaktadır. Bu noktada karar ağacının düğümlerini ayıracak en uygun dalı belirlemek yerine rastgele seçilen değişkenler içerisinden en iyi değişkeni kullanmaktadır. Rastgele seçilen değişkenler arasından yapılacak seçim için ise CART algoritmasında olduğu gibi Gini indeksi kullanılarak ağacın yapılandırılması sürdürülmektedir (Breiman, 2001).

2.3.1.7. Lojistik Regresyon Algoritması

Lojistik regresyon, ikili sınıflandırma problemleri için kullanılan bir makine öğrenmesi algoritmasıdır. Bir olayın meydana gelme olasılığını bir dizi bağımsız değişkene dayalı olarak tahmin etmek için kullanmakta ve bağımsız değişkenler ile olayın meydana gelme olasılığı veya aynı anlama gelmek üzere bir örneğin bir sınıfa ait olma olasılığı arasındaki ilişkiyi bir lojistik fonksiyon kullanarak tahmin ederek çalışır. Lojistik fonksiyonun çıktısı, olayın meydana gelme olasılığı olarak yorumlanabilecek 0 ile 1 arasında bir değerdir (Kleinbaum & Klein, 2002).

$$f(z) = P(y = 1|x) = \frac{1}{1 + e^{-z}} \tag{2.12}$$

Eşitlik 2.12'de lojistik fonksiyonun tanımı görülmektedir. $P(y = 1|x)$ ifadesi x öznitelik girdileri verildiğinde olayın gerçekleşme olasılığını, e euler sabitini, z ise girdi özniteliklerinin doğrusal kombinasyonu ve karşılıklı ağırlıklarını göstermektedir.

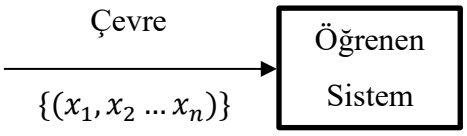


Şekil 2.5: Standart Lojistik Fonksiyon

Şekil 2.5 incelendiğinde lojistik fonksiyon z değeri $-\infty$ değer aldığında 0'a ve ∞ değer aldığında ise 1'e eşitlenecektir. Lojistik fonksiyonun z nin değerinin yüksekliğinden veya düşüklüğünden bağımsız olarak 1 ve 0 aralığında değer üretiyor olması bu fonksiyonu olasılık temelli yöntemlerde kullanışlı hale getirmiştir.

2.3.2. Danışmansız Öğrenme

Danışmanlı öğrenmeden farklı olarak danışmansız öğrenme (*Unsupervised Learning*) modelinde hedef çıktılar önceden belli değildir ve amaç veri içerisindeki doğal yapıyı ortaya çıkarmaktır (Atalay & Çelik, 2017).



Şekil 2.6: Danışmansız Öğrenme Diyagramı (Kantardzic, 2020, s.110)

Dayan (1999)'a göre danışmansız öğrenme ilk aşamada saf, yapılandırılmamış ve gürültü olarak nitelendirilebilecek veri içerisindeki desenlerin bulunması işlemidir.

Şekil 2.6'da görüldüğü gibi danışmansız öğrenme süreci yalnızca problem durumu ile ilgili çevreden edinilen ve girdi değerlerine sahip olan fakat sınıf etiketleri bilinmeyen veriyi kullanarak gerçekleştirilmektedir. Kantardzic (2020)'e göre danışmansız öğrenme modelinin özellikle vurgulanması gereken yönü görev-bağımsız oluşudur. Diğer bir deyişle öğrenen

sistemin başarı ölçütü danışmanlı öğrenmede olduğu gibi bir görevin (örn. sınıflandırma) ne kadar iyi yerine getirildiği değil girdi değerleri sunulan verinin öğrenen sistem tarafından ne kadar iyi temsil edildiğidir.

Kümeleme analizi tipik bir danışmansız öğrenme modeli örneği olarak görülmektedir ve bu yöntemler ile veri içerisinde birbirine benzeyen nesnelerin gruplandırılarak kümelere ayrıştırılması amaçlanmaktadır (Akpınar, 2017). Kümeleme yöntemlerinin genelinde benzerlik, iki veya daha fazla boyuta sahip bir uzaydaki örneklerin birbirlerine olan uzaklıkları hesaplanarak elde edilmekte ve yakın örnekler aynı küme içerisine atanmaktadır. Bu işlemin ardından yeni örnekler oluşan kümeler ile olan benzerlikleri ölçüsünde gruplandırılırlar (Han & Kamber, 2006).

2.3.3. Pekiştirmeli Öğrenme

Pekiştirmeli veya takviyeli öğrenmede (*Reinforcement Learning*) sistem bir danışmandan yoksun olarak yalnızca yürüttüğü bir takım eylemlerin sonucuna dair aldığı ödül veya ceza biçimindeki geribildirimler sayesinde eğitilmektedir (Russell & Norvig, 2010). Pekiştirmeli öğrenmede yürütülecek eylemler sisteme önceden tanımlanmamakta fakat bunun yerine sistemden en fazla ödül getiren eylemleri keşfetmesi beklenmektedir.

Sutton & Barto (1998) pekiştirmeli öğrenmedeki deneme-yanılma ile arama ve gecikmeli ödül mekanizmalarının bu yöntemi diğer makine öğrenmesi yöntemlerinden ayıran iki temel özellik olduğunu belirtmişlerdir. Deneme-yanılma ile arama doğrultusunda sistem daha fazla ödül elde etmek için hem geçmişte denediği eylemler arasından en fazla ödül getirenleri seçmekte hem de deneme-yanılma yoluyla gelecekte daha iyi seçimler yapabilmesini sağlayacak eylemleri keşfetmeyi sürdürmektedir. Ödülün veya geribildirimin gecikmeli yapısı ise pekiştirmeli öğrenmeyi özellikle danışmanlı öğrenmeden ayırmaktadır. Danışmanlı öğrenmede öğrenme etiketlenmiş örnekler üzerinden gerçekleştirilirken pekiştirmeli öğrenmede ise sistem çevreyle etkileşime girerek sonuçları eyleme geçtikten sonra ortaya çıkacak davranışlarda bulunmakta ve bu sonuçlara göre gelecekteki eylemlerini düzenlemektedir.

2.4. EĞİTİMDE YAPAY ZEKÂ KULLANIMI

Eğitimde yapay zekâ kullanımı ve bu doğrultudaki araştırmalar yapay zekâ destekli öğretim sistemleri başta olmak üzere akıllı chatbotlar, eğitsel ajanlar ve yazı analizi gibi öğrenme sürecini destekleyici teknolojileri kapsamaktadır (Holmes ve diğ., 2019). Eğitimde yapay zekâ kullanımının başlıca amacı öğrenenin davranış, yeti ve zihinsel durumunu modelleyebilen ve öğrenme sürecinin iyileştirilmesi adına kişiye özgü geribildirimler verebilen akıllı eğitsel ajanların ve sistemlerin geliştirilmesidir (Woolf ve diğ., 2013).

Chen ve diğerleri (2020), yapay zekânın eğitimde kullanımının eğitim yönetimi, öğretim ve öğrenme alanlarına olan katkılarını incelemişlerdir. Yazarlara göre Zeki Öğretim Sistemleri (*Intelligent Tutoring Systems*) sayesinde öğrencilerin durumuna özgü geribildirim sağlayarak rehberlik ve danışmanlık faaliyetlerinin yükünün azaltılması ve öğrenme stillerine göre kişiselleştirilmiş öğrenme planları hazırlanabilmesi mümkündür. Web tabanlı eğitsel ajanlardan robotik teknolojisiyle birleştirilmiş yapay zekâ uygulamalarına kadar geniş bir çerçevede rastlanan çalışmalarda sınavların değerlendirilmesi, sınıf dışında destekleyici öğretim faaliyetlerinin sürdürülmesi ve öğrenciye özgü öğretim yönteminin seçilmesi gibi normal şartlarda öğretmenlerce üstlenilen görevlerin gerçekleştirilebildiği görülmüştür. Ancak eğitimde yapay zekâ kullanımının henüz başlangıç evresinde olması ve eğitim ile yapay zekâ arasındaki ilişkinin en sade haliyle insan ve insanlarla işbirliği halindeki makinelerin bir entegrasyonu olması nedeniyle özellikle öğretmenlerin eğitim sürecindeki rollerinin tümüyle yapay zekâ teknolojilerine devredilemeyeceği vurgulanmaktadır (Cope ve diğ., 2020). Sijing ve Lan (2018), yapay zekânın ödev okuma, not verme, öğrenme materyallerinin ve kaynakların hazırlanması, eğitim içeriğinin aktarımı gibi tekrarlayan işleri yapabilir hale gelmelerinin öğretmenlerin iş yükünü azaltacağını ancak eğitim ortamlarındaki bu değişimin öğretmenlere dönük beklentileri tam aksine arttıracağını belirtmişlerdir. Bu doğrultuda yapay zekâ, tekrarlayan işleri yapma zorunluluğundan kurtardığı öğretmenleri bilgi aktarıcısı rolünden amaçlı rehber rolüne doğru bir değişikliğe zorlamaktadır. Yapay zekâ teknolojilerinin yardımıyla iş yükü azalan öğretmenlerin yine yapay zekâ desteğiyle öğrenme sürecindeki sorunları keşfetmeleri ve öğrenme hedefleri doğrultusunda rehberlik yapma ve kendine özgü rehberlik yöntemleri geliştirme becerilerini geçmişe kıyasla daha etkili bir biçimde sergilemeleri olanaklı hale gelmektedir.

Luckin ve diğerleri (2016), eğitimde kullanılan yapay zekâ modellerini (1) öğretim yöntemleri ve bilgisi ile eğitilerek geribildirim, ipucu ve pekiştireçler ile öğrenme sürecine yardımcı olan pedagojik model; (2) öğretilecek konuların bilgisi ile eğitilen alan modeli; (3) öğrenenin geçmiş deneyimleri, öğrenme hızı, duygu durumu ve benzeri kişisel bilgileri ile eğitilen öğrenen modeli olmak üzere üç başlıkta incelemişlerdir. Özellikle öğrenen modeli yapay zekâ yöntemlerinin yapısına uygun olarak öğrenenin sistemle kurduğu etkileşimlerle gelişmekte ve bu sayede pedagojik model ve alan modeli öğrenciye en uygun geribildirim ve içerikleri sunabilmektedir. Ayrıca öğrenen davranışının modellenmesi ile öğrenmenin nasıl gerçekleştiği, bir problemin çözümü için öğrencilerin hangi adımları izlediği, başarıya etki eden etmenlerin neler olduğu gibi eğitim bilimleri araştırmalarınca irdelenen soruların da cevaplanması mümkündür. Bu doğrultuda Baker (2000), eğitimle ilgili bir durumun anlaşılması ve tahmin edilmesi amacıyla öğrenenin bilişsel süreçlerinin modellendiği yapay zekâ modellerini öğrencilere geribildirimler veren akıllı eğitsel modellerden ve doğrudan öğrencilerle etkileşimde olmayıp yalnızca öğretim süreçlerinin veya ortamlarının tasarlanması için bir temel olarak kullanılan modellerden ayrı olarak sınıflandırmıştır. Bahsi geçen üç modelin yanında eğitim yöneticilerinin karar verme ve politika geliştirme süreçlerini desteklemek amacıyla geliştirilen politika-yapıcı modeller de eğitimde kullanılan yapay zekâ modelleri arasında gösterilmektedir (Hwang ve diğ., 2020).

Son yıllarda eğitimde yapay zekâ kullanımı kapsamında öne çıkan iki disiplin eğitsel veri madenciliği ve öğrenme analitiğidir. Her iki disiplinde de eğitim ortamlarından elde edilen yüksek miktarda veri analiz edilerek yürütülen uygulama ve araştırmaların sayesinde öğrenme ve öğretim deneyimlerinin iyileştirilmesi amaçlandığından bu terimler birbirleriyle aynı anlama gelecek şekilde kullanılabildikleri gibi (Aldowah ve diğ., 2019), "eğitimde yapay zekâ kullanımı" üst başlığında birleştirilerek kullanıldıkları da görülmektedir (Berendt ve diğ., 2020).

Eğitsel veri madenciliği eğitim ile ilgili ortamlardan elde edilen verinin açıklanmasına dönük yöntemlerin geliştirilmesi ve bu yöntemleri öğrenenlerin ve öğrenme ortamlarının daha iyi anlaşılması için kullanılması ile ilgilenmektedir (Siemens & Baker, 2012). Daha detaylı bir tanıma göre ise eğitsel veri madenciliği, herhangi bir başka yöntemle analiz edilmesi zor ya da imkânsız olan büyük miktardaki eğitim verisi içindeki örüntülerin makine öğrenmesi ve istatistik temelli hesaplamalı yöntemlerle saptanması ve bu yöntemlerin araştırılması, geliştirilmesi ve uygulanması ile ilgilidir (Romero & Ventura, 2013). Baker & Yacef (2009),

eğitsel veri madenciliğinin yalnızca bilinen veri madenciliği yöntemlerinin eğitim verisi üzerinde uygulanması olmadığını ve makine öğrenmesi, psikometri, istatistik, bilgi görselleştirme ve hesaplamalı modelleme gibi birden fazla disiplinin eğitim bilimleri araştırmalarının konularıyla olan kesişimi içerisinde ve bu alana özgün yöntemlerle yürütülebileceğini öne sürmüşlerdir. Paralel olarak Romero ve Ventura (2007) eğitsel veri madenciliğinde kullanılan yöntemleri istatistik, görselleştirme, kümeleme, sınıflama, uç değer tespiti, birliktelik kuralları, örüntü madenciliği ve metin madenciliği olarak sıralamakla beraber bu alandaki uygulamaların farklı kullanıcı grupları, öğrenme ortamlarının türleri ve ortamların sağladıkları veri türleri gibi etmenleri göz önüne alarak gerçekleştirilmesi gerektiğini belirtmişlerdir (Romero & Ventura, 2010).

Öğrenme analitiğinin ortaya çıkışı ise ticari işletmelerde verimliliğin artırılması amacıyla kullanılan bilişim teknolojilerindeki gelişmelerin sunduğu veri işleme, analiz ve raporlama olanaklarının benzer şekilde eğitim ortamlarının ve kurumlarının verimliliğini artırmak amacıyla da kullanılmasıyla gerçekleşmiştir (Bozkurt, 2016). İş analitiği ile yakından ilişkili olan bir uygulama alanı olarak gelişen öğrenme analitiği aynı zamanda öğrenmeyi ve öğrenme ortamlarını anlama ve iyileştirme amaçları doğrultusunda öğrenenlerden ve öğrenenlerle ilişkili ortamlardan elde edilen verinin toplanması, ölçülmesi, analiz edilmesi ve raporlanmasını da kapsadığı için bir araştırma sahası olarak da değerlendirilmektedir (Papamitsiou & Economides, 2014).

Eğitsel veri madenciliği ve öğrenme analitiği disiplinlerinin her ikisi de veri temelli bir yaklaşımı benimsemekte ve veri analizi sayesinde eğitim ile ilgili problemlerin saptanarak planlama, organizasyon ve karar verme gibi süreçlere dönük iyileştirmeler yapılması amacında ortaklaşmaktadırlar. Ancak söz konusu disiplinlerin her ne kadar ortak yanları bulunsa da kullanılan teknikler, keşif biçimleri, çıkış ve odak noktaları açısından birtakım farklılıklar görülebilmektedir.

Tablo 2.2: Eğitsel Veri Madenciliği ve Öğrenme Analitiği Arasındaki Farklar (Siemens & Baker, 2012; Romero & Ventura, 2013)

	Öğrenme Analitiği	**Eğitsel Veri Madenciliği**
Çıkış Noktası	Semantik web, akıllı müfredat, öğrenme çıktılarının tahmin edilmesi, sistematik müdahaleler	Eğitim yazılımları, öğrenen modeli, ders çıktılarının tahmin edilmesi
Odak Noktası	Verinin ve veri analizinin sonuçlarının açıklanması	Kullanılan veri madenciliği yöntemlerinin karşılaştırması ve açıklaması
İndirgeme ve Bütünsellik	Sistemlerin tüm karmaşıklığıyla birlikte bütünsel olarak anlaşılması	Sistemlerin parçalara indirgenerek analiz edilmesi ve parçalar arası ilişkilerin incelenmesi
Yöntem	Sosyal ağ analizi, söylem analizi, etki analizi, kavram analizi	Sınıflandırma, kümeleme, bayesyen modelleme, birliktelik kuralları, modelleme, görselleştirme
Keşfetme Biçimi	Karar alma sürecinin desteklenmesi önemlidir, otomatik keşif bu hedefin gerçekleştirilmesi için bir araçtır.	Otomatik keşfetme önemlidir, karar alma sürecinin desteklenmesi bu hedefin gerçekleştirilmesi için bir araçtır.

Tablo 2.2'de yapılan karşılaştırmaya göre eğitsel veri madenciliği modelleme sayesinde otomatik hale getirilmiş uygulamaların eğitim ortamına entegre edilmesini ve modelin geliştirilmesi için kullanılan yöntemlerin teknik analizini önemsemekte ve eğitsel bağlamın açıklanması ve iyileştirilmesi amacına bu açıdan yaklaşmaktayken; öğrenme analitiği ise aynı yöntemlerle veri yığını içerisinden elde edilen anlamlı bilginin öğretmen ve öğrenciler üzerindeki çıktılarına odaklanmaktadır. Diğer bir deyişle eğitsel veri madenciliği öğrenme ile ilgili büyük miktardaki verinin anlamlandırılmasına dönük teknik zorluklara odaklanırken, öğrenme analitiği ise elde edilen anlamlı bilginin eğitsel bağlamı iyileştirecek şekilde nasıl kullanılabileceği sorusuna cevap aramaktadır (Ferguson, 2013).

Kimi yazarlar eğitsel veri madenciliği ve öğrenme analitiği arasındaki benzerlik ve farklılıkları kabul etmekle birlikte bu iki disiplin arasında bir hiyerarşik ilişkinin de geçerli olduğunu ileri

sürmüşlerdir. Chatti ve diğerleri (2012)'ne göre öğrenme analitiği, eğitsel veri madenciliğinde temel olarak kullanılan sınıflandırma, kümeleme ve birliktelik kuralları gibi yöntemlerin yanı sıra istatistiksel yöntemler, görselleştirme araçları ve sosyal ağ analizi gibi veri madenciliği kapsamı dışındaki yöntemleri de barındıran ve ayrıca bu yöntemlerin etkililiğini öğretim ve öğrenmenin geliştirilmesi doğrultusunda inceleyen daha geniş bir alandır. Benzer şekilde Steiner ve diğerleri (2014), söz konusu iki kavramı ayrı veya birbirinin yerine kullanmaktansa öğrenme analitiği adlandırmasını eğitsel veri madenciliği ve öğrenme analitiği ile ilgili tüm araştırma ve uygulamaları kapsayacak şekilde kullanmayı önermişlerdir.

Siemens & Baker (2012), eğitsel veri madenciliği ve öğrenme analitiği arasındaki ayrımların iki araştırma grubu arasındaki genel eğilimleri yansıttığını ancak yürütülen çalışmalar göz önüne alındığında Tablo 1'de belirtilen ayrım noktaları arasında geçişkenlik olabileceğini ve bundan dolayı bir alanda yürütülen araştırmanın diğer alanın yaklaşımından öğeler barındırabileceğini belirterek iki alanın araştırma grupları arasında birliğin sağlanmasını önermişlerdir. Öte yandan eğitim ile ilgili büyük miktardaki veriyi anlamlı ve işlevli hale getirecek uygulamaların etkililiğini yalnızca kullanılan yapay zekâ yöntemlerinin teknik boyutu açısından değerlendirilmemesi ve tüm bu süreci eğitim bilimleri, psikoloji, pedagoji ve bilgisayar bilimlerinin bir kesişim noktası olarak ele alınması bir araştırma problemi olarak öğrenme sürecini daha iyi anlama olanaklarını da artırma potansiyeline sahiptir (Papamitsiou & Economides, 2014)

2.5. MOTİVASYON

Motivasyon, eğitim bilimleri araştırmalarında öğrenme süreci ile ilgili etmenlerin ve özellikle akademik başarının açıklanması için önemli bir kavram olarak görülmektedir (Mega ve diğ., 2014). Motivasyon temel olarak insan davranışının uyarılma, güç ve yönünü etkileyen süreçleri nitelemektedir (Barto, 2012) ve motive olmak bir işi yapmak için harekete geçmeyi ifade etmektedir (Ryan & Deci, 2000a). Öğrenme açısından motivasyon ise öğrencinin kendisi ve çevresi hakkında sahip olduğu algılardan kaynaklanan ve sunulan bir eğitsel faaliyeti tamamlamayı seçmesine, bu faaliyetten sebat ederek öğrenme amacıyla meşgul olmasına neden olan bir olgu olarak tanımlanmıştır (Viau, 2015).

Schunk ve diğerleri (2014), motivasyonun içsel kaynakları olduğunu ve bundan dolayı bir öğrenenin nasıl motive olduğunu anlamak için zihnin bilgiyi işleme sürecini incelemek gerektiğini öne süren bilişsel yaklaşımın ve motivasyonu not ve takdir gibi ödüllerden

kaynaklanan, öğrenenin bu ödülleri elde etmek amacıyla öğrenme ortamındaki dışsal uyaranlara artan ve devamlı şekilde verdiği tepkiler olarak tanımlayan davranışçı yaklaşımın temel noktalarını içeren genelleştirilmiş bir motivasyon tanımı yapmışlardır. Yazarlara göre "Motivasyon, hedef-yönelimli faaliyetlerin tetiklendiği ve sürdürüldüğü bir süreçtir"(s.5). Bu tanıma göre motivasyon süreci bilişsel yaklaşımda da vurgulandığı şekilde öğreneni eğitsel faaliyete yöneltecek hedef veya hedefleri içermektedir. Hedefler planlı veya plansız olabilir ve tecrübe yoluyla değişebilir veya değişmeyebilir ancak önemli olan öğrenenlerin gerçekleştirmeye istekli veya isteksiz olacakları bir hedefin varlığının bilincinde olmalarıdır. Ayrıca hedeflerle kurulan ilişkinin davranış yoluyla ortaya çıktığı fiziksel veya zihinsel faaliyetler de motivasyon tanımının önemli bir bileşenidir. Fiziksel faaliyetler çaba, ısrar ve diğer tüm görünür eylemleri ifade ederken; zihinsel faaliyetler ise planlama, tekrar yapma, problem çözme, bilgilerin düzenlenmesi ve ilerlemenin değerlendirilmesi gibi eylemleri kapsamaktadır. Tetikleme, öğrenme hedeflerinin bilincinde olan öğrenenin hedefini gerçekleştirmek için girişkenlik göstermesidir. Son olarak, motivasyon bir süreç olarak tanımlandığı ve izlenebildiği için özellikle öğrenenin zor durumlara, problemlere, başarısızlıklara ve gerilemelere nasıl tepki verdiğini niteleyen sürdürme eylemi ile görünür hale gelmektedir.

Araştırmacılar arasında motivasyonun tanımı ve işleyişi üzerine bir ortaklık bulunmadığından dolayı bugüne dek motivasyonu açıklayan birden fazla teori öne sürülmüştür (Pelaccia & Viau, 2017). Bu doğrultuda özellikle öz-yeterlik, öz-belirlenim ve hedef teorileri öne çıkmaktadır.

2.5.1. Öz-Yeterlik Teorisi

Öz yeterlik, bireyin bir işi başarabileceğine veya başaramayacağına dair kendi hakkındaki inancıdır (Bandura, 1977). Bu yaklaşıma göre bir işi başarmak için gereken yetenekleri hakkında öz-yeterliliğe sahip bir öğrenci öğrenme için çaba harcama ve zorlayıcı durumlarla başa çıkmada daha istekli olmaktadır (Schunk & Pajares, 2009). Öz-yeterlilik için belirleyici faktör tek başına bireylerin yetenekleri değildir ve aynı yeteneklere sahip iki farklı kişinin akademik performansları birbirine kıyasla değişkenlik gösterebilir veya farklı ortamlardaki aynı kişi öz-yeterlik inancının farklılaşmasından dolayı zayıf performans sergileyebilir (Bandura, 1997). Öğrenciler geçmiş başarı ve başarısızlıklarından edindikleri deneyimlerle ilişkili olarak (Lackaye & Margalit, 2006) mevcut yeteneklerinin bir işi gerçekleştirmek için yeterli olup

olmadığına dair bir inanca sahip olmaktadırlar ve bu öz-yeterlik inancı çaba, azim ve davranış olarak izlenebilen motivasyonu belirlemektedir (Schunk, 1991).

2.5.2. Öz-Belirlenim Teorisi

Bu teoriye göre bireyin iradesini kullanma süreci olarak tanımlanan öz-belirlenim, güç ve sınırlarının farkında olan insanın en temel psikolojik ihtiyaçları olan yetkinlik, özerklik ve aidiyet ihtiyaçlarını karşılamak için belirlediği yollar ve aldığı kararları ifade etmektedir (Schunk, 1991). Öz-belirlenim teorisi, bireylerin davranışını benlikleri tarafından onaylanan ve isteyerek eyleme geçtiklerinde öz-belirlenimli; çevreden kaynaklanan bir sebepten ötürü zorunlu olarak sergiledikleri eylemleri ise kontrollü eylemler olmak üzere ikiye ayırmaktadır (Deci ve diğerleri, 1991). İnsan davranışındaki bu ayrım motivasyon söz konusu olduğunda içsel motivasyon ve dışsal motivasyon ayrımına yol açmaktadır. İçsel motivasyon, öğrenenin bir davranışı özü itibariyle tatmin edici bulduğu için, dışsal motivasyon ise temel olarak bir sonuç veya ödülün elde edilmesi isteğiyle gerçekleştirmesi olarak tanımlanmaktadır (Legault, 2016). Diğer bir deyişle, bir öğrenci bir konuyu ilginç veya eğlenceli bulduğu için öğrenmeye istekli olabilirken başka bir öğrenci dönem sonunda yüksek not almak veya öğretmenden takdir görmek amacıyla öğrenme için çaba harcayabilmektedir. Ryan ve Deci (2000a), özellikle bireyin bir işe dönük ilgisinden kaynaklanan ve yetkinlik ve özerklik gibi psikolojik ihtiyaçların tatmin edilmesini sağlayan içsel motivasyonun öz-belirlenimli davranışı tanımladığını vurgulamışlardır. İçsel sebeplerden bağımsız olarak bir sonuca erişmek için yürütülen dışsal motivasyon davranışları ise farklı öz-belirlenim seviyelerine sahiptirler ve davranışın sonuçları öğrenme süreci içerisinde ne düzeyde içselleştirilirler ise öz-belirlenim seviyesi o denli artmaktadır. Özetle, öz-belirlenim teorisi insan motivasyonunun kaynağını öncelikle içsel psikolojik ihtiyaçlara dayandırmakta ve motivasyon sürecini psikolojik ihtiyaçların giderilmesi amacıyla insanın çevreyle kurduğu ilişki ekseninde açıklamaktadır (Ryan & Deci, 2000b).

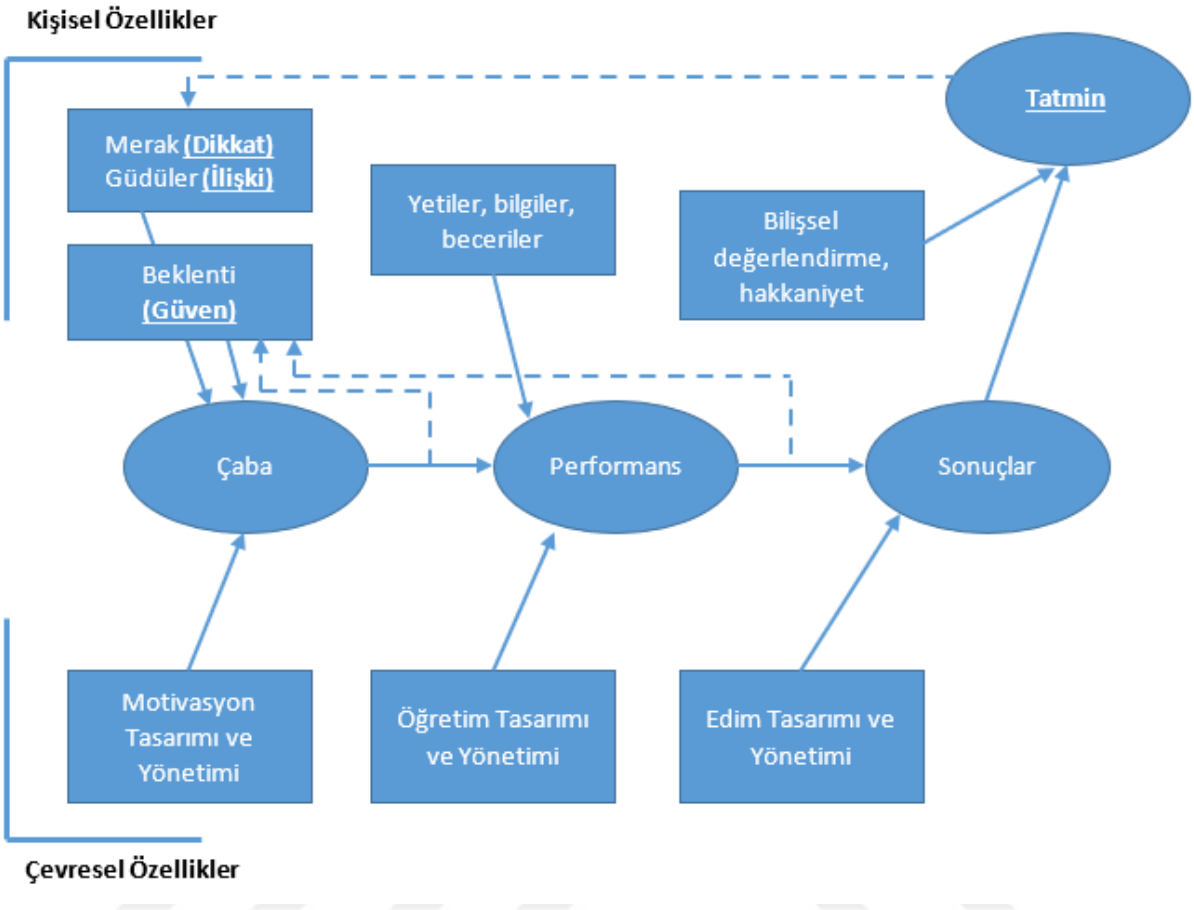
2.5.3. Hedef Yönelimi Teorisi

Hedef yönelimleri öğrenenin başarı için gereken görevlere bağlanma amacı ve nedenlerini nitelemektedir (Pintrich, 2003). Bu yaklaşıma göre öğrencilerin bir işi başarmak için gereken hedeflere bağlanma davranışları istek ve motivasyonlarının işlevsel ifadeleri anlamına gelmektedir (Seifert, 2004). Hedef yönelimi teorisi, öğrencilerin bir hedefe neden ulaşmak istedikleri ve öğrenme görevlerine nasıl yaklaştıkları ve nasıl bağlandıkları ile ilgilenmektedir

ve bu doğrultuda hedef yönelimlerini ustalık hedef yönelimi ve performans hedef yönelimi olmak üzere iki kategoride incelemektedir (Schunk ve diğ., 2014). Ames (1992)'e göre ustalık hedef yönelimli öğrenciler yeni beceriler geliştiren, yaptıkları işi anlamaya çalışan, öğrenme sürecinde tatmin düzeylerini yükselten ve kendi belirledikleri standartlara göre bir işte ustalaştıkları hissiyatını taşıyan kişilerdir. Öğrencilerin ustalık hedef yönelimli davranışları onları öğrenme görevi kapsamında elde ettikleri bilgi ve beceriler aracılığıyla yetkinlik düzeylerini artırmaya yöneltmektedir (Mega ve diğ., 2014). Performans hedef yönelimli öğrenciler ise yeteneklerini sergileyerek diğer öğrencilerden farklı olduklarını göstermek ve bu performansları sayesine sosyal görünürlüklerini artırma eğilimindedirler (Pintrich, 2000).

2.5.4. Keller'in ARCS Motivasyon Teorisi

Keller (2010), öne sürülmüş olan motivasyon teorileri arasındaki ayrımların farklı araştırma sahalarını tanımlamada etkili olduğunu ancak tek bir teoriyi temel alarak yürütülen çalışmaların motivasyonu açıklamada eksik noktalar bıraktığını ve bundan dolayı bütünsel bir motivasyon teorisine ihtiyaç olduğunu savunmuştur. ARCS (*Attention, Relevance, Confidence, Satisfaction* – Dikkat, İlişki, Güven, Tatmin) modeli bu ihtiyacı karşılamak üzere bilişsel ve davranışçı teorilerin dayanaklarını da içerdiği için açıklayıcı gücü yüksek ve aynı zamanda öğretim tasarımı ve öğrenme stratejileri sayesinde öğrencilerde motivasyonun sağlanması ve sürdürülmesine dönük standartları içerdiği için eğitim ortamlarına uyarlanabilir bir teori olarak geliştirilmiştir (Song & Keller, 2001).



Şekil 2.7: Keller'in Motivasyon ve Performans Makro Modeli (Keller, 2010)

Şekil 2.7'de görülen model motivasyonel süreçleri girdi, çıktı ve ilişkileri bulunan bir sistematik süreç olarak görselleştirilmektedir. En üst satırda motivasyon, performans ve öğrenmeye etki eden psikolojik özellikler gösterilirken, orta satırda ise motivasyonun ölçülebilir çıktıları sıralanmıştır. En alt satırda ise bu çıktılara etki eden çevresel özellikler gösterilmektedir (Keller, 2010). ARCS modelinin temel aldığı dört kategori (Dikkat, İlişki, Güven ve Tatmin) alt kategorileri olmakla birlikte temel olarak bir insanın tamamen motive olması için gereken durumları kapsamaktadır (Keller, 2000):

Dikkat: Bir öğrencinin derse dönük ilgisinin uyandırılması için öncelikle dikkatlerinin çekilmesi gerekmektedir (Keller & Suzuki, 2004). Bireylerin bir nesne veya duruma dönük ilk izlenimlerinin tutumlarıyla ilişkili olduğu bilindiğinden özellikle öğrencilerin bir ders veya konunun ilk dakikalarında ilgilerinin uyandırılması gelecekteki motivasyonel düzeylerinin önemli bir belirleyeni olarak görülmektedir (Keller & Burkman, 1993). Öğrencinin öğrenme hedefleriyle ilgili uyaranlara veya uyaranların bir bölümüne odaklanması öğrenmeye dönük

meraklarını tetiklemekte ve bir derse dönük daha fazla çaba harcamalarını sağlamaktadır (Keller, 2010). Dikkatin elde edildikten sonra sürdürülmesi ise öğretim stratejilerinin çeşitlendirilmesi ile mümkündür. Öğrenciler ne kadar etkili olursa olsun sürekli aynı öğretim stratejilerinin kullanıldığı ortamlarda ilgilerini zamanla kaybetme eğilimindedirler (Keller, 2008).

İlişki: Öğrencinin merak ve ilgisi uyandırılsa bile öğrenme içeriğinin öğrencilerin önem verdikleri hedeflerle, geçmişten getirdikleri ilgi alanlarıyla veya öğrenme stilleriyle bağlantısı kurulamadığı durumda motivasyonun sürdürülmesi zorlaşmaktadır (Keller, 2000). Diğer bir deyişle, öğrencilerin bir derse motive olmaları ve çaba göstermeleri için ilgili dersin hedef ve kazanımlarının öğrencinin bireysel hedefleriyle ilişkili olduğuna dair bir inanç ve güdülenme gereklidir (Keller, 1987). Bununla birlikte öğrenme hedefleri açıkça tanımlandığı takdirde öğrencilerin yalnızca dersten geçmek amacıyla değil büyük oranda kişisel ilgileri ve seçimleri dolayısıyla çaba göstermeleri sağlanabilmektedir (Keller & Suzuki, 2004). Yüksek bir motivasyon düzeyi için gerekli olan içsel motivasyonu sağlamak amacıyla en sık başvurulan yöntemler ders kazanımlarını öğrencilerin gelecekteki iş yaşamlarıyla veya yakın gelecekteki akademik gereksinimleriyle ilişkilendirmek, ayrıca vaka çalışmalarından, benzetimlerden, öğrencilerin mevcut ilgi alanları ve deneyimleriyle ilişkili örneklerden faydalanmaktır (Keller, 2000).

Güven: Öğrencilerin kendi içsel hedefleriyle tutarlı, öğrenme stilleriyle uyumlu ve geçmişte edindikleri bilgi ve becerileriyle ilişkili konulara dönük bir güdülenmeyi sürekli kılabilmeleri için öğrenme hedeflerini gerçekleştirme doğrultusunda karşılaşacakları kavramsal karmaşa, anlaşılmazlık ve benzeri zor durumların üstesinden gelebileceklerine dair bir güvene sahip olmaları gerekmektedir (Arnone & Small, 1995). Güven değişkeni öğrencinin kişisel kontrol algısını ve başarı beklentisini ifade etmekle birlikte öğrenme hedeflerini gerçekleştireceklerine ve başarılarını kontrol edebileceklerine dair inançlarının geliştirilmesine dönük tüm stratejileri kapsamaktadır (Keller, 2010). Öğrenme süreci ve görevleri üzerindeki odaklanmalarını kontrol edilebilir bir çabayla sürdüren öğrenciler çabalarının sonuçlarına dair daha az kaygı duymakta ve bu sonuçların getireceği potansiyel başarılara odaklandıkları oranda da daha üretken olabilmektedirler (Keller, 2008).

Tatmin: Öğrenmeye dönük motivasyona sahip olmakla birlikte elde edilen motivasyonun sürdürülmesi için öğrencilerin başarıları ve öğrenme deneyimleri hakkında olumlu izlenimlere

sahip olmaları gerekmektedir. Ayrıca öğrencilerin başarılarının ardından ilgi görmeleri, öğrendikleri bilgi ve becerileri uygulama olanağı bulmaları, öğrenme için gereken iş yükünün yeterli olduğuna inanmaları, konuların içeriği ve ölçme-değerlendirme araçlarının öğrenme hedefleriyle tutarlı olduğunu gözlemlemeleri ve notlandırmada adil davranıldığına dönük bir hakkaniyet duygusuna sahip olmaları tatmin düzeyini yükseltmektedir (Keller, 2000). Şekil 4'te gösterilen modelde tatmin kategorisi içsel ve dışsal değişkenlerin çıktısı olan performansın sonuçları ile öğrenenin bu sonuçlar hakkındaki bilişsel değerlendirmelerinin birleşiminden oluşmaktadır. Bilişsel değerlendirme sürecinde ise öğrenciler performanslarının sonuçlarını beklentileriyle olduğu kadar diğer öğrencilerin elde ettikleriyle de kıyaslamakta ve bunun sonucunda olumlu veya olumsuz bir yargıya sahip olmaktadırlar (Keller, 2010).

Keller'in ARCS modeli motivasyonun işleyişini açıklamak için birden fazla yaklaşımın temel önermelerinin birleştirildiği bir makro model olarak yapılandırılmıştır. Örneğin, güven kategorisi öğrencinin bir işi yapabileceğine olan inancını içermekte ve bu açıdan öz-yeterlik teorisine (Bandura, 1977) yaklaşmaktadır. Öte yandan tatmin kategorisi modelde gösterilen içsel ve dışsal değişkenlerin çıktılarına bağlı olarak belirlenmekte ve bu ölçüde motivasyonu içsel ve dışsal etmenlere bağlı olarak açıklayan ve sınıflandıran öz-belirlenim teorisinden (Deci & Ryan, 1985) faydalanmaktadır.

Şekil 2.7'de görselleştirilen modelin son satırı motivasyon, öğrenme ve performansın belirlenmesinde çevrenin rolünü içermektedir. Keller (2010)'e göre "Motivasyon, öğretmenin ve öğretim materyallerinin merak uyandırma derecesinden, öğrencinin kişisel özellikleriyle ilişkili uyaranlar setiyle birlikte güven duygusunu pekiştiren zorluk seviyelerinden, ve çaba göstermeyi önleyecek stres faktörlerinin yokluğundan etkilenir" (s.6). Bundan dolayı öğretim materyallerinin özellikleri ve kullanımı, öğretmenin sergilediği öğretim stratejileri ve tutumu, müfredatın içeriği ve ders içi etkinlikler gibi etmenler öğrenci motivasyonunun çevresel belirleyenleri arasında yer almaktadırlar. Son olarak, modelin çıktılar satırında yer alan geribildirimler öğrencinin harcadığı çaba ve sonuçlarının gelecekte elde edilecek başarılar için güven duygusunu beslediğini, tatmin kategorisinden dikkat ve ilişki kategorilerine yönelen geribildirim ise kişinin öğrenme hedeflerine yönelik bir dizi davranışın sonuçlarıyla ilgili gerçek deneyimlerinin gelecekte bu ve benzer hedeflere atfedilen değeri etkileyeceğini göstermektedir (Keller, 2010).

ARCS Motivasyon Modeli, öne çıkan motivasyon teorileri ile birlikte motivasyonel gelişim için gerekli öğretim tasarımı ve stratejileri uygulamalarını da barındırması sebebiyle e-öğrenme, bilgisayar tabanlı öğrenme ve uzaktan eğitim ortamlarında yürütülen araştırmalarda gerek öğrencilerin motivasyon düzeylerini ortaya çıkarmak gerekse de öğrencilerde motivasyonun artırılmasına dönük öğrenme materyalleri ve öğretim programları geliştirmek için kullanılmaktadır (Song & Keller, 2001; Chang & Lehman, 2002; Huang ve diğ., 2004; Huett ve diğ., 2008; Chen, 2012; Chang & Chen, 2015; Ma & Lee, 2021).

2.6. ÖĞRENME YÖNETİM SİSTEMLERİ VE LOG DOSYALARI

Öğrenme Yönetim Sistemleri (ÖYS), öğretim materyallerinin, ödevlerin, duyuruların paylaşılması ve ayrıca öğrencilerin kendi aralarında veya öğretmenlerle iletişimini olanaklı kılan web tabanlı bir sistemdir (Lonn & Teasley, 2009). ÖYS'ler ayrıca notlandırma, devamsızlık ve kayıt takibi gibi ders yönetimiyle ilgili süreçlerin yürütülmesi için de kullanılmaktadır (Coates ve diğ., 2005). Falvo & Johnson (2007), ÖYS'leri öğrencilere dönük etkinliklerin ve öğrenme materyallerinin sağlanmasının yanı sıra derse katılım ve öğrenme sürecindeki ilerleyişin veri sistemleri ve değerlendirme araçları sayesinde izlenmesini mümkün kılarak eğitimcilere yardımcı olan sistemler olarak tanımlamışlardır. Günümüzde ÖYS kullanımı yüz yüze eğitimi destekleyici bir unsur olarak görülse de özellikle COVID-19 pandemisinin çıkışıyla birlikte üniversiteler öğretim faaliyetlerini yüz yüze eğitimin yapılamadığı bu dönemde başta ÖYS'ler olmak üzere çeşitli e-öğrenme teknolojileri ile uzaktan yürütmüşlerdir (Raza ve diğ., 2021).

Coates ve diğerleri (2005) ÖYS'lerin işlev ve özelliklerini şu şekilde sıralamışlardır:

- Asenkron ve senkron iletişim (Duyurular, tartışma forumları, mesajlaşma)
- İçerik geliştirme ve iletimi (Öğrenme içeriğinin geliştirilmesi ve kaynakların depolanması, link paylaşımı)
- Biçimlendirici ve özetleyici değerlendirme (Sınav yapma, ödev verme, grup çalışması ve geribildirim)
- Sınıf ve kullanıcı yönetimi (derse kaydetme ve kaydolma, zaman çizelgeleri, öğrenci etkinliklerini yönetme)

ÖYS modelleri, sistem içerisindeki özelliklerin seçimi ve kullanıcılara aktarılmasındaki farklılıklara göre ayrılmaktadır. Bir organizasyonun kendi girişimiyle geliştirilen ve yalnızca

organizasyonun eğitsel ihtiyaçlarına cevap verecek ÖYS'lerin kullanımına rastlanmakla birlikte genel kullanıma açık ve özelleştirilebilen tescilli ÖYS hizmetleri de bulunmaktadır. Ancak yıllar içerisinde giderek popülerlik kazanan bir diğer model ise açık kaynak kodlu ÖYS'ler olmuştur (Babo & Azevedo, 2012). En fazla bilinen temsilcisi Moodle platformu olan açık kaynak kodlu ÖYS'ler kullanıcılar tarafından geliştirilmeye açık ve ücretsiz erişilebilen sistemler olmakla birlikte sosyal-yapılandırmacı yaklaşıma ait ilkeleri temel alarak geliştirilmesi yönünden diğer modellerden ayrılmaktadır (Al-Aljan & Zedan, 2008).

Moodle, geliştiricileri tarafından "eğitimcilere, yöneticilere ve öğrencilere kişiselleştirilmiş öğrenme ortamları oluşturmalarını sağlamak amacıyla tasarlanmış sağlam, güvenli ve entegre bir sistem" olarak tanımlanmıştır (Moodle, 2020, s.1). Moodle tabanlı sistemler yapısı ve miktarı değiştirilebilen belirli kategorilere ait modüllerden oluşmaktadır (Kumar ve diğ., 2011):

- *İletişim Modülü*; dosya yükleme ve indirme, tartışma forumları, e-posta gönderimi, gerçek zamanlı mesajlaşma gibi tüm iç ve dış iletişim özelliklerini barındırmaktadır.
- *Üretkenlik Modülü*; arama, takvim, yardım, ilerleme takibi ve gözden geçirme modüllerini içermektedir.
- *Öğrenci Katılımı Modülü*; öz-değerlendirme ve grup çalışması modülünü içermektedir.
- *Yönetim Modülü*; kullanıcı yetkilendirme, kimlik doğrulama, ders kayıt entegrasyonu gibi modülleri barındırmaktadır.
- *Ders Aktarımı Modülü*; ders yönetimi, çevrimiçi notlandırma, öğrenci takibi ve sınav yapma özelliklerini kapsamaktadır.
- *Müfredat Geliştirme Modülü;* ders programının oluşturulması doğrultusunda kullanılacak şablonları ve özelleştirme modüllerini içermektedir.

Moodle tabanlı ÖYS'ler ayrıca site yönetimi paneli aracılığıyla erişilen raporlama özelliği sayesinde öğretmen, öğrenci veya ziyaretçilerin etkinliklerinin log kayıtlarıyla izlenebilmesine olanak sunmaktadır. Kullanıcıların web sayfalarında gezinirken bıraktıkları izlerin çeşitli bilgilerle kayıt altına alındığı log dosyaları, bir web sayfasının gösterilmesi için gerekli olan dosyaları içeren web sunucularında depolanmaktadır. Web sunucuları arasında değişiklik gösterse de kullanıcı adı (IP adresi), ziyaret yolu, zaman bilgisi ve ziyaret edilen sayfa gibi bilgiler çoğunlukla log kayıtları tarafından gösterilmektedir (Grace ve diğ., 2011).

Tablo 2.3: Moodle Örnek Log Kaydı

Zaman	31/05/20 03:12:30 PM
Kullanıcı adı	student1
Etkilenen kullanıcı	-
Etkinlik bağlamı	Assignment: Upload Excel Assignment 3 Here
Bileşen	Assignment
Etkinlik adı	Course module viewed
Açıklama	The user with id '87' viewed the 'assign' activity with course module id '131'
Köken	web
IP adresi	1xx.xx.1xx.1xx

Tablo 2.3'te gösterilen Moodle sunucusuna ait örnek log kaydında zaman, kullanıcı adı, ziyaret edilen sayfa (etkinlik bağlamı) gibi log kayıtlarının genelinde rastlanan bilgiler yer aldığı gibi bileşen, etkinlik adı ve açıklama bilgileri ise Moodle'a özgü kullanıcı etkinliklerinin takip edilebilmesini sağlamaktadır. Log dosyaları özellikle web kullanım madenciliği başlığı altında yürütülen araştırmalarda kullanıcıların web sitesiyle olan etkileşimlerinin analiz edilmesiyle birlikte doğrudan görülmesi mümkün olmayan davranış örüntülerinin ortaya çıkarılması amacıyla kullanılmaktadır (Mughal, 2018).

Eğitimde yapay zekâ çalışmalarının genelinde rastlanmakta olan son kullanıcıdan veri toplama yönteminin aksine kullanıcıların belirli bir teknolojiyi kullanma sürecinde üretmekte oldukları verinin kullanılması, web kullanım madenciliği uygulamalarının sağlamış olduğu olanaklarla diğer veri türleriyle birlikte analiz edilerek farklı örüntülerin ortaya çıkarılmasına katkı sunabilmektedir. Ayrıca log dosyalarının kullanımı madencilik uygulamalarının da kullanım süreciyle bütünleştirilmesi sağlanabilmekte ve kullanıcıların sistemi kullanmakta oldukları süre içerisinde tahmin sürecinin otomatize edilerek süreklileştirilmesi ve gerçek zamanlı hale getirilmesi mümkün olmaktadır (Tsiakmaki ve diğ., 2019).

2.7. LİTERATÜR TARAMASI

Öğrenme Yönetim Sistemi (ÖYS) log dosyaları kullanılarak yürütülen çalışmalarda yapay zekâ yöntemleri kullanılarak en fazla tahmin edilmeye çalışılan değişkenin akademik başarı veya akademik performans olduğu görülmektedir (Aldowah ve diğ., 2019). Motivasyon veya motivasyon ile ilgili değişkenlerin yapay zekâ yöntemleriyle tahmin edilmesine dönük çalışmalarda ise ÖYS log dosyalarının kullanıldığı çalışmalara rastlanmakla beraber (Cocea &

Weibelzahl, 2007; Hershkovitz & Nachmias, 2009, Kularbphettong ve diğ., 2012) öğrencilerin derse dönük motivasyonsuzluğunun bir göstergesi olarak dersi veya okulu bırakma davranışının önceden tahmin edilmesine dönük çalışmalar ağırlıktadır (Cohen, 2017; Kondo ve diğ., 2017; Sukhbaatar ve diğ., 2018; De Oliveira ve diğ., 2019; Ortigosa ve diğ., 2019; Figueroa-Cañas & Sancho-Vinuesa, 2020; Tamada ve diğ., 2022). Bu doğrultuda ilgili literatür taraması bölümü veri setini ÖYS log dosyalarını temel alarak oluşturan makine öğrenmesi çalışmaları ve ÖYS log dosyalarının kullanıldığı motivasyon odaklı makine öğrenmesi çalışmaları olmak üzere iki bölümde incelenmektedir.

2.7.1. ÖYS Log Dosyalarının Kullanıldığı Makine Öğrenmesi Çalışmaları

Araştırmacılar ÖYS log dosyalarını kullanılarak elde dilen veri setleri üzerinde başta akademik başarı ve performans olmak üzere öğrenme stilleri ve etkileşim gibi değişkenleri de incelemişlerdir. Örneğin, Romero ve diğerleri (2010) tarafından yürütülen çalışmada mühendislik bölümlerinde okuyan 438 öğrencinin Moodle log dosyaları ile oluşturulan ödev yükleme, geçilen/başarısız olunan sınav, materyal okuma miktarları ve ödevlere, sınavlara ve forumlara harcanan toplam süre özniteliklerini içeren veri seti kullanılarak akademik başarı seviyelerinin tahmin edilmesi amaçlanmıştır. Akademik başarı hedef niteliği ise öğrencilerin sınav notlarının denk geldiği aralığa göre başarısız, geçer, iyi ve mükemmel olmak üzere dört sınıfa atanmasıyla oluşturulmuştur. Birden fazla sınıflandırma algoritmasının kullanıldığı çalışmada başarı oranının ortalama %65 olduğu ve karar ağaçları, kural indirgeme ve bulanık mantık algoritmalarının etkililiği ve sonuçlarının anlaşılırlığı açısından öne çıktıkları görülmektedir.

Benzer çalışmaların ÖYS'den elde edilen log kayıtlarının öğrencilerin sistem içerisindeki aktivitelerini niteleyen miktar veya yüzdelik değerlere dönüştürüldüğü öznitelikleri içerdiği görülebilmekte (Calvo-Flores ve diğ., 2006; Bravo & Ortigosa, 2009; Ademi ve diğ., 2019; Huang ve diğ., 2020) ve ayrıca bir takım çalışmaların ise öğrencilerin sistemle çeşitli yönlerden etkileşimlerinin zaman cinsine dönüştürüldüğü özniteliklere sahip veri setlerini kullandıkları gözlenmektedir (Romero ve diğ., 2008a; Romero ve diğ., 2008b; Akçapınar, 2014). Araştırmalarda kullanılan sınıflandırma algoritmaları farklılık gösteriyor olsa da ÖYS log dosyaları kullanılarak yürütülecek eğitsel veri madenciliği çalışmalarının verinin toplanması, veri önişleme, veri madenciliği ve sonuçların yorumlanması adımlarını içeren standart bir süreç olarak tarif edilebileceği de vurgulanmaktadır (Romero ve diğ. 2008b).

Çalışmalarda ÖYS'den elde edilen log dosyaları kullanılarak oluşturulan öznitelikler sistemin kullanımındaki ortak noktalar dolayısıyla (sisteme giriş yapma sayısı, sayfa görüntüleme sayısı vb.) benzerlik gösterebilmekte ancak derslerin müfredat, etkinlik veya ölçme-değerlendirme yöntemlerine göre farklılaşabilmektedir. Örneğin Ai & Laffey (2007) çalışmalarında yükseköğretim seviyesinde Moodle kullanılarak yürütülen derslerde ortak olarak oluşturulabilecek olan oturum başına sayfa görüntüleme sayısı, anasayfa görüntüleme sayısı, sisteme giriş sayısı ve dosya indirme sayısı gibi öznitelikleri kullanmışlardır. Diğer yandan Figueira (2017) tarafından yürütülen çalışmada, forum görüntüleme ve forum mesajı gönderme sayısı gibi dersin işleyişine özel özniteliklerin yanında Moodle log dosyalarında olay tanımlamaları içerisinde yer alan oluşturma, görüntüleme, güncelleme ve silme gibi ifadelerin her bir öğrenci için sıklığının hesaplandığı bir veri seti oluşturulmuştur. Huang ve diğerleri (2020) ise araştırma yürüttükleri dersin forum, quiz, video ve tüm moodle aktivitelerinin temel alınarak oluşturulduğu öznitelikleri içeren geniş kapsamlı bir veri seti kullanmışlardır. Model geliştirmede veri setindeki örnek sayısı kadar öznitelik sayısının da fazla olması performansı artıran bir etmen olarak görülmektedir. Romero ve diğerleri (2010), eğitimciler için geliştirdikleri tahmin yazılımında kullanıcıların Moodle log dosyalarını yüklemeleri ardından ne gibi öznitelikleri oluşturacaklarını seçmeleri sağlanmaktadır. Bundan dolayı Moodle içerisindeki ortak olayların özniteliklere dönüştürüldüğü veri setlerinde oluşturulan modeli genellemek mümkün olsa da gerek akademik başarının tahmin edilmesi doğrultusunda geliştirilmiş bir yazılım yoluyla gerekse de Moodle log dosyalarının kullanıldığı tüm eğitsel veri madenciliği çalışmalarında özniteliklerin dersin müfredatı ve etkinliklerini göz önüne alarak oluşturulduğu gözlenmektedir.

Eğitsel veri madenciliği çalışmalarında akademik başarının yanı sıra öğrenme stilleri ve etkileşim gibi değişkenlerin de ÖYS log dosyaları kullanılarak tahmin edildiği görülmektedir. Efrati ve diğerleri (2014) ise ÖYS log dosyaları ile kaynak görüntüleme ve forum ziyareti gibi öznitelikleri içeren veri seti üzerinde yürüttükleri beklenti-maksimizasyon kümeleme analizi sonucunda elde edilen kümeleri bağımsız, rekabetçi ve öznel-içi öğrenme stilleriyle eşleştirmişlerdir. Özetle, öğrencilerin ÖYS içerisindeki aktivitelerini içeren log dosyalarını kullanarak sınıflandırma veya kümeleme gibi farklı yöntemlerle öğrenme stillerinin tahmin edilmesi veya keşfedilmesinin mümkün olduğu görülmüştür. Öte yandan Liyanage ve diğerleri (2016), öğrencilerin bireysel öğrenme stillerini tahmin edecek bir model geliştirmek amacıyla ÖYS log verisini kullanarak sayfa ziyareti ve harcanan zaman ortalamalarını hesaplamışlar ve

her bir öğrenci için sistemin ortalama kullanılma oranını ifade eden bir öznitelik oluşturmuşlardır. Anket yoluyla elde edilen öğrenme stilleri güçlü aktif, ortalama aktif, dengeli, ortalama yansıtıcı ve güçlü yansıtıcı olmak üzere beş hedef niteliği sınıfını oluşturmuştur. J48, Bayesyen ağlar, Naive Bayes ve rasgele orman algoritmalarının kullanıldığı çalışmada en yüksek başarı yüzdelerine %65 ile %90 aralığında J48 algoritması ile ulaşılmıştır.

Moodle log dosyaları aynı zamanda sistem içerisindeki öğretmen, öğrenci, materyal vb. öznelerin aralarındaki etkileşimin bilgisini de barındırmaktadır. Özdağoğlu ve diğerleri (2019), Moodle log dosyaları içerisindeki kullanıcılara ait aktivite verisi üzerinde süreç madenciliği, örüntü analizi ve kümeleme yöntemlerini uygulayarak öğrencilerin aktivite sıraları, birbirleri arasındaki aktivite ağı ve sistemdeki genel süreç akışını ortaya çıkarmışlardır. Bu çalışmanın benzer çalışmalara kıyasla önemli bir noktası öğrenciler arasındaki işbirliği ağının ortaya çıkarılmasında log satırlarındaki IP adresi bilgisinden yararlanarak yakın veya aynı lokasyondan sisteme bağlanan öğrencilerin saptanabilmesidir. Ayrıca bu alanda yürütülmüş ve öğrenci-öğrenci, öğrenci-içerik, öğrenci-öğretmen, öğrenci-sistem etkileşimlerinin, ÖYS içerisindeki aktivitelerin, özelliklerin ve işlevlerin kullanım sıklığı ve katılım yönteminin sınıflandırılması amacıyla regresyon ve açınsayıcı analiz yöntemlerinden faydalanıldığı çalışmalara da rastlanmaktadır (Agudo-Peregrina ve diğ., 2014; Iglesias-Pradas ve diğ., 2015).

Türkiye'de yürütülen tez çalışmalarına bakıldığında ise Moodle log dosyalarını öğrenci hareketliliğini belirlemek ve ders çalışma etkinlikleri ve bağlılık gibi değişkenlerle olan ilişkilerini makine öğrenmesi yöntemlerinin kullanılmadan analitik düzeyde incelendiği çalışmalara rastlanmaktadır (Çiftçi, 2006; Topal, 2020). Diğer yandan Özbay (2020), Moodle logları ile öğrenci hareketliliğini veri madenciliği kullanarak analiz ettiği çalışmasında log dosyaları ile elde ettiği çeşitli eylemlere ait sayısal öznitelikler oluşturarak kümeleme analizi uygulamış ve ardından kümelere atadığı sınıf etiketlerini kullanarak oluşturduğu veri setini karar ağacı algoritmalarıyla test etmiştir. Çalışma sonucunda %85 doğruluk değerine sahip CART algoritması ile karar ağacı oluşturulmuştur. Yıldırım (2018) tarafından yürütülen çalışmada da log dosyalarından elde edilen veriye uygulanan kümeleme ile etkileşim sınıfları belirlenmiş ve farklı etkileşim gruplarına özgün gezinme hareketliliği sıralı örüntü ve sınıflandırma analizleriyle incelenmiştir. Akçapınar (2014) ise araştırma kapsamında geliştirilen bir çevrimiçi öğrenme ortamına ait log dosyalarını kullanarak oluşturduğu 28 öznitelikli veri setiyle üniversite öğrencilerinin akademik başarılarını tahmin etmeyi

amaçlamıştır. Araştırma sonucunda bir dersten geçen ve kalan öğrencileri sınıflandırmada k-en yakın komşu ve CN2 algoritmalarıyla %86 başarı elde edildiği belirtilmiştir.

2.7.2. ÖYS Log Dosyalarının Kullanıldığı Motivasyon Odaklı Makine Öğrenmesi Çalışmaları

Veri setinin ÖYS log dosyaları kullanılarak oluşturulduğu çalışmalarda akademik başarı değişkenine kıyasla sayıları az olsa da öğrencileri motivasyon düzeylerine göre sınıflandırma veya kümelemeyi amaçlayan araştırmalara da rastlanmaktadır. Cocea & Weibelzahl (2007) tarafından yürütülen çalışmada HTML Tutor isimli bir çevrimiçi interaktif öğrenme ortamına ait log dosyalarından elde ettikleri sayfada harcanan zaman, sayfa görüntüleme sayısı, olunan test sayısı ve sayfa okuma süresi gibi özniteliklerden oluşan ve 48 öğrenciden elde edilen 943 örnekli bir veri seti kullanılmıştır. Araştırmacılar özellikle sayfa ve testler için harcanan zamanı temel alarak ve uzman görüşünden faydalanarak bağlanma, bağlantısızlık ve nötr olmak üzere veri setindeki örnekleri hedef niteliğe ait üç sınıf değeriyle etiketlemişlerdir. Yazarların çok daha düşük bir örneklemle gerçekleştirdikleri pilot çalışmada bir öğrencinin bağlanmaya sahip olup olmadığını anlamanın sistemde geçirilen 45 dakikadan sonra mümkün olduğu ve bu doğrultuda motivasyonel düzeylerin ve özellikle bağlantısız öğrencilerin erken tespiti için daha az zamanda yürütülen çalışmalara ihtiyaç duyulduğu öne sürülmüştür (Cocea & Weibelzahl; 2006). Yürütülen çalışmada ise öğrencilere sistemde materyal okuma ve test için 10 dakikalık bir zaman tanınmıştır. Analiz sonucunda regresyon yöntemiyle %87 başarı elde edilmiş olup bağlantısızlık sınıfının diğer sınıflara göre başarılı olarak sınıflandırma oranı ise Bayesyen Ağlar algoritması ile %93 olarak kaydedilmiştir.

Hershkovitz ve Nachmias (2009) tarafından gerçekleştirilen çalışmada öğrencilerin çevrimiçi öğrenme ortamındaki davranışlarının log dosyalarını temel alarak incelenmesi ve bir motivasyon ölçme aracının geliştirilmesi hedeflenmiştir. Üniversite öğrencilerinin kelime haznesini geliştirmeleri için tasarlanan ve ezberleme, uygulama, oyun, kendini test etme ve arama modlarına sahip olan öğrenme ortamı katılımcıların kullanımına 3 hafta ile 3 ay süresince açılmaktadır. 1444 örnekli veri setinin kullanıldığı çalışmada log dosyalarından göreve harcanan süre, ortalama oturum süresi, ortalama aktivite adımı, oturumlar arası ortalama süre, kelime işaretleme, sınav aktivite yüzdesi ve oyun aktivite yüzdesi öznitelikleri elde edilmiştir. Hiyerarşik küme analizi ile oluşan 3 adet küme motivasyon boyutlarını ifade eden bağlanma, yönelim ve kaynak değişkenlerine atanmıştır. Atama işlemi sonrasında bağlanma kümesi

göreve harcanan süre, ortalama oturum süresi değişkenlerini, yönelim kümesi ortalama aktivite adımı, oturumlar arası ortalama süre, sınav aktivite yüzdesi ve oyun aktivite yüzdesi değişkenlerini, kaynak kümesi ise kelime işaretleme değişkenini içermektedir.

Kularbphettong ve diğerleri (2012) tarafından yürütülen çalışmada e-öğrenme ortamındaki öğrencilerin motivasyonel davranışlarının analizi için bir model oluşturulması amaçlanmıştır. 3002 öğrencinin kişisel kayıtlarıyla birlikte Moodle tabanlı ÖYS'den elde edilen log dosyaları kullanılarak test ve ödev ile ilgili aktiviteleri ifade eden öznitelikler oluşturulmuştur. Apriori algoritması ile yürütülen birliktelik kuralları analizi sonucunda öğretmenlerin ilgilenecekleri düşünülen kurallar listelenmiş ve aralarında özellikle not özniteliğinin motivasyonu ifade edecek biçimde değişkenliğe sahip olduğu düşünülerek çok düşük, düşük, orta, yüksek ve çok yüksek etiketleri bu öznitelik temel alınarak örneklere atanmıştır. Ardından 901 örnekle gerçekleştirilen analizde ise tüm etiketler için sınıflandırma başarısının %66 ile %89 arasında değiştiği gözlenmiştir.

Motivasyon ile ilgili eğitsel veri madenciliği çalışmalarında amaç büyük oranda motivasyonu düşük veya motivasyonsuz öğrencilerin erkenden saptanarak gerekli müdahale veya değişikliklerin yapılmasını kolaylaştırmaktır. Bununla bağlantılı olarak motivasyonsuzluğun davranışsal bir sonucu olarak değerlendirilebilecek olan dersi veya okulu bırakma eğilimi konusunda ÖYS log dosyaları üzerinde yapay zekâ yöntemleri kullanılarak yapılan çalışmalara da rastlanmaktadır. De Oliveira ve diğerleri (2019), iki farklı lisansüstü ders kapsamında yürüttükleri çalışmada dersi bırakma eğilimi olan öğrencilerin erken tespit edilebilmesi için bir öğrenme analitiği sistemi geliştirmeyi hedeflemişlerdir. Moodle platformundan elde edilen ve 166 öğrenciye ait log kayıtlarında 30 günden fazla sisteme giriş yapmayan öğrenciler dersi bırakmış olarak etiketlenmiştir. Forum ve ders sayfasında yapılan etkinlikler temel alınarak oluşturulan özniteliklere öğrencilere ait farklı kaynaklardan elde edilen veri de eklenmiştir. Birden fazla sınıflandırma algoritmasının kullanıldığı analiz sonucunda en yüksek başarı ekstra rasgele orman algoritması ile elde edilmiştir.

Benzer bir çalışmada ise Figueroa-Cañas ve Sancho-Vinuesa (2020) dersi bırakma riski taşıyan öğrencilerin önceden saptanabilmesi için geliştirilecek modelde 197 üniversite öğrencisinin Moodle ÖYS içerisindeki log kayıtlarını kullanmışlardır. Araştırmacılar diğer çalışmalardan farklı olarak veri setini ödevlerin son teslim tarihlerine denk gelecek üç farklı zamana ayırmışlar ve ayrıca oluşturulan öznitelikleri de bağlanma ve başarı olmak üzere iki sete bölmüşlerdir.

Dersi bırakma hedef niteliği ise araştırmacılar tarafından final notu tamamlanmamış olarak gözüken veya zorunlu final sınavına girmeyen öğrencilerin etiketlenmesiyle oluşturulmuştur. Bağlanma ile ilgili özniteliklerinin yer aldığı veri setinde başarı seviyeleri %60'lar dolayındayken Başarı ile ilgili özniteliklerin aldığı veri setinde %70'li seviyelere ulaşılmıştır. Yazarlar benzer problemler için sınıflandırma başarıları düşük olsa da ağaç temelli sınıflandırma algoritmalarının kullanımını anlaşılır ve yorumlanabilir olmalarından dolayı önermektedirler. Dersi bırakma riski taşıyan öğrencilerin durumunu yorumlanabilir bir biçimde sunma doğrultusunda özellikle öğretmenlere dönük görselleştirme araçlarının geliştirildiği çalışmalar da bulunmaktadır (Brito ve diğ., 2019; Dobashi ve diğ., 2019; Ortigosa ve diğ., 2019; Distante ve diğ, 2020).

Özet olarak, ÖYS log kayıtları ile gerçekleştirilen ve motivasyon ile ilgili eğitsel veri madenciliği çalışmalarında hedef niteliğe ait etiketler araştırmacıların veri setinde ilişkili olduğu düşünülen bir değişkenin dönüştürülmesiyle veya uzmanlar tarafından veri setindeki değişkenlerin incelenmesiyle oluşturulmaktadır. Motivasyon haricindeki değişkenleri irdeleyen kimi araştırmalarda ise hedef niteliğin belirlenmesi adına öğrencilerden anket yoluyla veri toplanması tercih edilmiştir (Arroyo & Woolf, 2005; Liyanage ve diğ., 2016). Bu çalışmada kullanılacak veri setinin motivasyon hedef niteliği katılımcılardan anket yoluyla toplanarak oluşturulmuştur ve bu sayede literatürdeki çalışmalarda benimsendiği görülen motivasyon ile ilgili olduğu düşünülen değişkenleri hedef niteliğe dönüştürme yaklaşımının yerine doğrudan derse yönelik motivasyon değişkeni incelenmiştir. Ayrıca literatürde danışmanlı öğrenme yöntemi ile yürütülmüş olan çalışmalardaki gibi birden fazla sınıflandırma algoritmasının performansları ile birlikte ilgili çalışmalardan farklı olarak motivasyon değişkeninin en fazla etkileyen değişkenler de incelenmiştir.

3. MALZEME VE YÖNTEM

Tezin Malzeme ve Yöntem bölümünde araştırmanın örneklemi, veri toplama süreci, veri toplama araçları, veri analizi yöntemleri doğrultusunda kullanılan algoritmalar ve veri analizi süreci veri tabanlarında bilgi keşfi sürecinin aşamalarıyla paralel olarak açıklanmıştır.

3.1. ARAŞTIRMANIN ÖRNEKLEMİ

Araştırmaya 2020-2021 Güz, 2020-2021 Bahar ve 2021-2022 Güz akademik dönemlerinde Yeditepe Üniversitesi'ndeki farklı bölümlerde okumakta olan ve Temel Bilgisayar Uygulamaları dersine kayıtlı 407 öğrenci katılmıştır. Bu öğrencilerden 207 kişi araştırmacı tarafından yürütülen şubelere, geri kalan 200 kişi ise aynı dersin farklı eğitmenler tarafından yürütülen şubelerine kayıtlıdır. Katılımcıların bağlı bulundukları fakültelere göre dağılımı Tablo 3.1'de gösterilmektedir.

Tablo 3.1: Araştırmanın Katılımcılarının Bağlı Bulundukları Fakültelere Göre Dağılımı

Fakülte	Kişi Sayısı	Yüzdelik Dağılım
Mühendislik Fakültesi	211	51,84%
Fen-Edebiyat Fakültesi	79	19,41%
İktisat Fakültesi	36	8,85%
Sağlık Bilimleri	25	6,14%
Ticari Bilimler Fakültesi	20	4,91%
İletişim Fakültesi	19	4,67%
Mimarlık Fakültesi	8	1,97%
Eczacılık Fakültesi	6	1,47%
Eğitim Fakültesi	2	0,49%
Güzel Sanatlar Fakültesi	1	0,25%
TOPLAM	407	100%

Tablo 3.1'e bakıldığında Mühendislik Fakültesi öğrencilerinin araştırmaya en fazla sayıda katılım gösteren grup olduğu görülmektedir. Katılımcıların araştırmaya katıldıkları sıradaki sınıf düzeyleri ve cinsiyete göre dağılımları Tablo 3.2'de sunulmuştur.

Tablo 3.2: Araştırmanın Katılımcılarının Sınıf Düzeylerine ve Cinsiyetlerine Göre Dağılımı

Sınıf	Erkek	Kadın	TOPLAM ve YÜZDE (Sınıf)
1	18	37	55 (%13,51)
2	27	41	68 (%16,71)
3	47	31	78 (%19,16)
4	99	89	188 (%46,19)
5	12	6	18 (%4,42)
TOPLAM VE YÜZDE (Cinsiyet)	203 (%49,88)	204 (%50,12)	

Tablo 3.2 incelendiğinde araştırmaya en fazla 4. sınıfta okumakta olan öğrencilerin katılım gösterdiği görülmektedir. Cinsiyet dağılımına bakıldığında ise araştırmaya katılan kadın ve erkek öğrenci sayısının birbirine yakın olduğu gözlenmektedir.

3.2. VERİ TOPLAMA ARAÇLARI

Bu bölümde araştırma verisinin toplanmasında kullanılan araçlar açıklanmıştır.

3.2.1. Derse İlgi Ölçeği (DİÖ)

Derse İlgi Ölçeği (DİÖ), Keller (2010) tarafından ARCS Motivasyon Teorisi'ni temel alarak geliştirilmiş olan ve öğrencilerin bir eğitmen tarafından yürütülen derse dönük öğrenme motivasyonlarını ve motivasyonel tepkilerini ölçmek amacıyla geliştirilmiştir. Yazar, DİÖ'nün yüz yüze eğitim ortamlarının yanında senkron veya asenkron çevrimiçi öğrenme ortamları gibi bilgi aktarımının gerçekleştiği her bağlamda motivasyonel tutumları ölçmek için kullanılabileceğini belirtmiştir. Ancak DİÖ, duruma özel bir ölçek olduğundan dolayı öğrencilerin öğrenmeye dönük genel motivasyon düzeylerini ölçmemektedir ve ancak öğrencilerin bir derse dönük olarak nasıl motive olduklarını ölçmek için kullanılmaktadır. Tablo 3.3'te DİÖ'ye ait maddelerin hangi ARCS Motivasyon Teorisi alt boyutlarına ait oldukları gösterilmektedir.

Tablo 3.3: Derse İlgi Ölçeği (DİÖ) Madde ve Alt Boyutları

Dikkat	İlişki	Güven	Tatmin
1	2	3	7*
4*	5	6*	12
10	8*	9	14
15	13	11*	16
21	20	17*	18
24	22	27	19
26*	23	30	31*
29	25*	34	32
	28		33

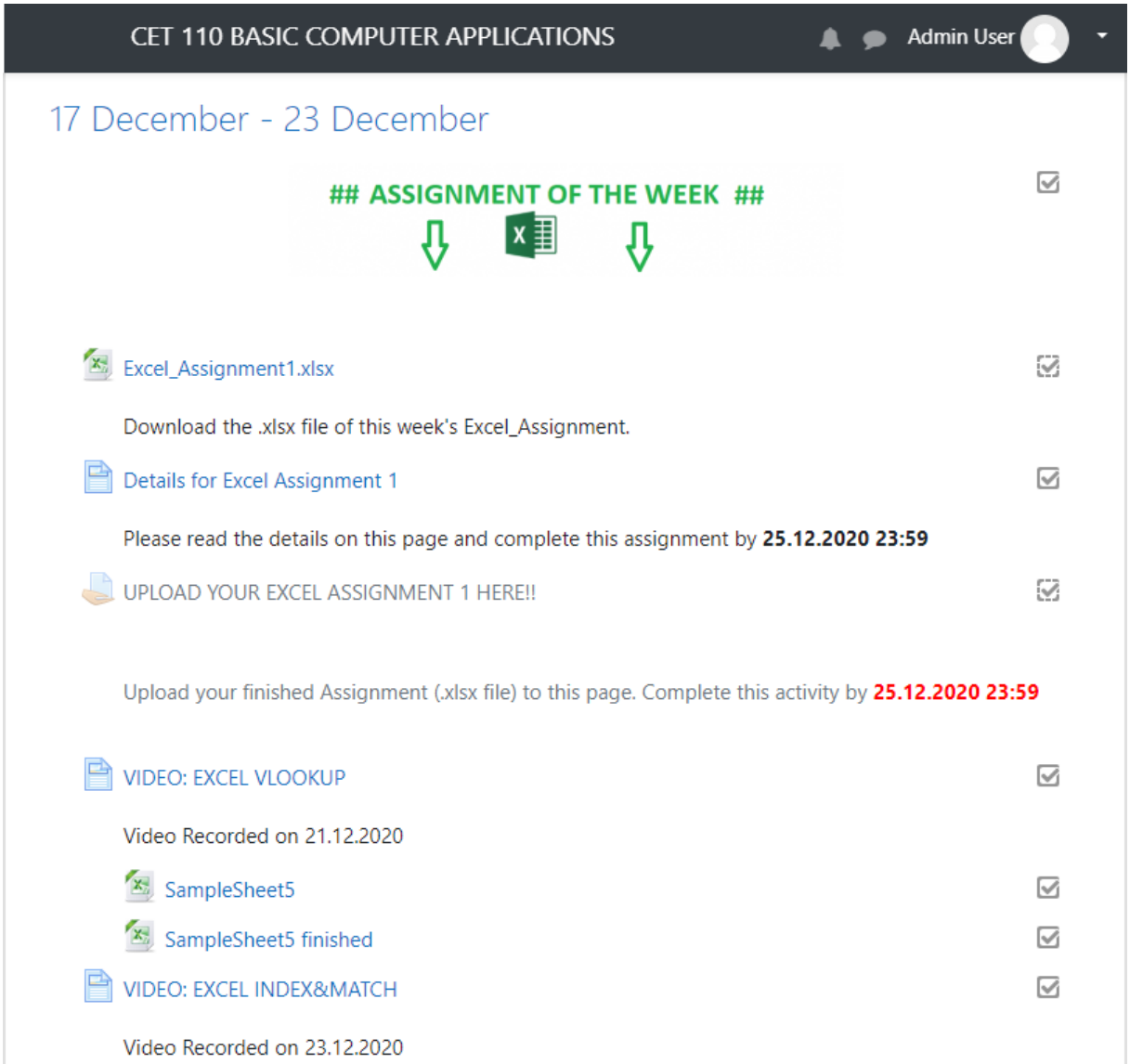
*: Ters puanlanan maddeler

Dikkat, İlişki, Güven ve Tatmin olmak üzere ARCS Motivasyon Teorisi'ni temel alan dört alt boyuta ait 34 maddeden oluşan DİÖ, 5'li likert tipinde geliştirilmiştir. Ölçekteki olumsuz ifade içeren 4, 6, 7, 8, 11, 17, 25, 26 ve 31. maddelerin ters puanlanması gerekliliği belirtilmiştir. Bu ölçeğin Türkçe'ye adaptasyonu amacıyla yapılan en son çalışma Dinçer (2015) tarafından yürütülmüştür ancak söz konusu araştırmanın katılımcıları üniversite öğrencilerinden oluşmamaktadır. Acar (2009) tarafından üniversite öğrencileriyle yürütülen adaptasyon çalışması kapsamında Türkçe çevirisi yapılan DİÖ'deki maddelerin ise yeniden çevrilmesinin yapılan DİÖ çevirilerinin çeşitliliğine katkı sağlayacağı düşünülmüştür. Ayrıca araştırmaya katılan öğrencilerin anket maddelerini anadillerinde eksiksiz bir şekilde anlayarak cevaplamalarını sağlamak amacıyla DİÖ ölçme aracı bu araştırma kapsamında Türkçe'ye çevrilmiştir (EK-6) ve geçerlik-güvenirlik analizi çalışmaları IBM SPSS Statistics 24 *(Statistical Package for Social Sciences)* ve AMOS Graphics 24 *(Analysis of Moment Structures)* programları kullanılarak yürütülmüştür.

3.2.2. Temel Bilgisayar Uygulamaları Dersi ve Öğrenme Yönetim Sistemi (ÖYS) Log Dosyaları

Bu tez araştırması 2020-2021 Güz, 2020-2021 Bahar ve 2021-2022 Güz akademik dönemlerinde Yeditepe Üniversitesi bünyesinde açılan ve 2020-2021 Güz döneminde uzaktan eğitim yöntemiyle, geri kalan dönemlerde ise hibrit eğitim yöntemiyle işlenen "Temel Bilgisayar Uygulamaları" dersi kapsamında yürütülmüştür. Söz konusu ders çeşitli bölümden öğrencilere akademik ve iş hayatlarında yardımcı olacak bilgisayar kullanım becerilerini

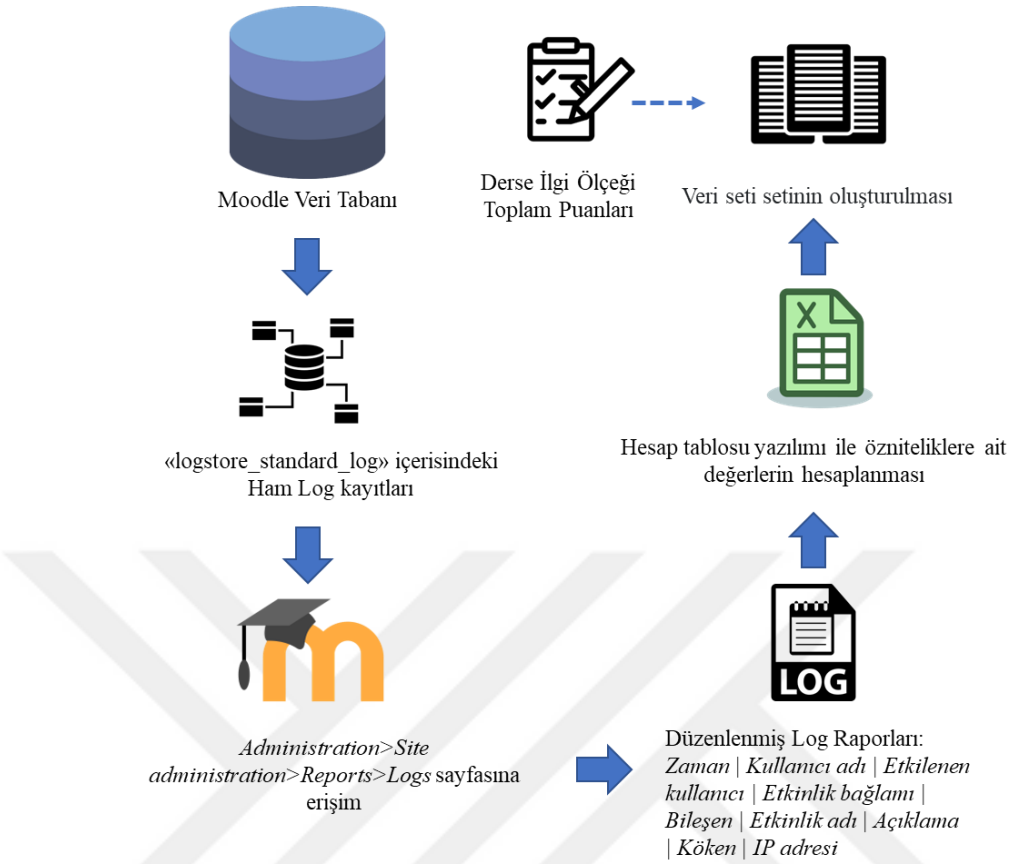
kazandırmayı amaçlamaktadır ve her bir dönem için öğretim süresi 14 haftadır. Eğitim dili İngilizce olan dersin öğretim programı genel hatlarıyla günümüz bilgi teknolojileri becerilerini kapsayacak şekilde e-posta ve bulut hizmetlerinin kullanımı, kelime işlemci programının ödev, tez ve rapor yazımına dönük doküman oluşturma ve düzenlemeyi sağlayan araç ve özelliklerinin kullanımı, sunum hazırlama programının kullanımı ve hesap tablosu programı ile veri analizi ve görselleştirme başlıklarından oluşmaktadır. Bu derse ait öğrenme materyalleri ve ilgili tüm içerikler öğrencilerin erişimine Moodle tabanlı bir ÖYS kullanılarak sunulmuştur. Şekil 3.1'de dersin bir haftasına ait örnek ÖYS ekran görüntüsü görülmektedir.



Şekil 3.1: Öğrenme Yönetim Sistemi Ekran Görüntüsü

Şekil 3.1'de bir örneği sunulan haftalık içerikte eğer varsa ilgili haftada verilen ödeve ait materyaller, ödev soruları ve açıklamaları içeren okuma materyali sayfası ve ödev yükleme sayfasının yanı sıra tüm haftalarda paylaşılan ve o haftanın konu anlatımını içeren video kaydı ve videoda üzerinde çalışılan dosyaların paylaşıldığı görülmektedir. Videolar araştırmacı tarafından hazırlanmış olup 2020-2021 Güz dönemi süresince her bir konunun işleneceği haftanın başında kaydedilip konular işlendikten sonra öğrencilerin erişimine açılmıştır. Geri kalan dönemlerde ise video içerikleri gerek duyulduğuna güncellenmiştir. Videoların paylaşılması özellikle uzaktan ve hibrit eğitimin sürdürüldüğü dönemin şartları göz önüne alındığında konu tekrarı yapmak isteyen veya ödevlerin çözümü için geçmiş konuların gösterildiği videolardan yararlanmak isteyen öğrencileri daha fazla ÖYS kullanmaya yöneltmiştir. Bahsi geçen içeriklere ek olarak öğrenciler ÖYS üzerinden güncel devamsızlık durumlarını takip edebilmekte ve güncel duyurulara erişebilmektedirler.

Derse kayıtlı öğrencilerin sistemde geçirdikleri 14 haftalık süre dâhilindeki aktiviteleri sistem yöneticisinin erişimine açık olan "Log Raporları" sayfasında ve Moodle tabanlı ÖYS içerisindeki veritabanında tutulmaktadır. Log verisinin elde edilmesi ve işlenmesi sürecini açıklayan diyagram Şekil 3.2'de gösterilmiştir.



Şekil 3.2: ÖYS Log Kayıtlarına Erişim ve Log Verisinin İşlenmesi Süreci

Şekil 3.2'de araştırmanın yürütüldüğü ÖYS üzerindeki derse kayıtlı öğrencilerin ders sayfaları ve içerikleriyle olan etkileşim verisinin tutulduğu log kayıtlarına erişim ve bu kayıtların işlenerek araştırmada kullanılacak olan veri setine dönüştürülmesi süreci özetlenmektedir. Ham log kayıtları Moodle web aplikasyonunun kurulu olduğu sunucuya ait veri tabanında tutulmakta ve Moodle yönetici paneli bir arayüz vasıtasıyla log raporlarını düzenlenmiş olarak sistem yöneticisinin erişimine sunmaktadır. Veri setinin oluşturulması amacıyla erişilen log kayıtları üzerinde Microsoft Excel 365 hesap tablosu programının işlevleri kullanılarak her bir öğrenci için özniteliklerin hesaplamaları gerçekleştirilmiştir. Özniteliklerin hesaplanmasında özellikle kullanıcının sistem içerisinde sergilediği etkileşimlerin tarih ve saat verisinin tutulduğu "Zaman" kayıtlarından, bu etkileşime bağlı aktivitenin kategorisi ilgili "*Course, Assignment, Page, File"* gibi anahtar kelimeler barındıran "Etkinlik Bağlamı" kayıtlarından ve etkileşimin içeriğiyle ilgili açıklamalar içeren "Etkinlik adı" ve "Açıklama" kayıtlarından yararlanılmıştır. Bu sayede bir öğrencinin sistemdeki etkileşimi süresince hangi aktiviteyi ne sıklıkta ve ne kadar

sürede yaptığını zaman ve miktar cinsinden değerleri tutan öznitelikler elde etmek mümkün olmuştur.

Tablo 3.4: Veri Seti Öznitelikleri, Veri kaynakları ve Özniteliklerin Açıklamaları

Öznitelik	kaynak	Açıklama
OturumAcmaSayisi	Log Kayıtları	Öğrencinin ÖYS'ye yaptığı giriş sayısı
GecmisKonularaBakma	Log kayıtları	Öğrencinin mevcut hafta içerisinde geçmiş haftalara ait sayfa ve içeriklerle girdiği etkileşim sayısı
Gezinme Sayısı_Toplam	Log kayıtları	Öğrencinin ziyaret ettiği toplam sayfa sayısı
GezinmeSayisi_Ortalama	Log kayıtları	Öğrencinin ziyaret ettiği toplam sayfa sayısının oturum açma sayısına bölünmesiyle elde edilen bir girişte ziyaret edilen ortalama sayfa sayısı
ToplamOturumSuresi	Log kayıtları	Öğrencinin ders dönemi boyunca ÖYS'de harcadığı toplam süre (saniye).
OrtalamaOturumSuresi	Log kayıtları	Öğrencinin toplam oturum süresinin oturum açma sayısına bölünmesiyle elde edilen bir ziyaretinde ortalama harcadığı süre (Saniye)
MateryalIndirme	Log kayıtları	Öğrencinin bir konu veya ödev ile ilgili paylaşılan dosyaları indirme sayısı
OdevDenemeSayisi	Log kayıtları	Öğrencinin ödev ile ilgili sayfa ve içeriklerle girdiği etkileşim sayısı
OdeveHarcananZaman	Log kayıtları	Öğrencinin ödev sayfasında geçirdiği süre (saniye)
SayfadaKalma_Ortalama	Log kayıtları	Öğrencinin toplam oturum süresinin toplam gezinme sayısına bölünmesiyle elde edilen bir sayfada harcadığı ortalama süre (Saniye)
SınavOdakliCalisma	Log kayıtları	Öğrencinin sınav haftaları içerisinde ÖYS'de harcadığı süre (saniye)
VideoIzlemeSuresi	Log kayıtları	Öğrencinin video içeriklerinin bulunduğu sayfalarda harcadığı süre (saniye)
YuklenenOdev	Log kayıtları	Öğrencinin sisteme yüklediği ödev sayısı
SınavNotu	ÖYS	Öğrencinin dönem sonu not ortalaması
Devamsizlik	ÖYS	Öğrencinin dönem içerisinde devamsızlık yapığı saatler
Faculty	Anket	Öğrencinin kayıtlı olduğu fakülte
Class	Anket	Öğrencinin kaçıncı sınıf olduğu bilgisi
Gender	Anket	Öğrencinin cinsiyeti
OgretimTuru	Araştırmacı	İlgili dönemde dersin işlenme biçimi (uzaktan veya hibrit)
Motivasyon (HEDEF NİTELİK)	DİÖ	Öğrencinin DİÖ üzerinde cevapladığı maddelerin toplam puanı

Tablo 3.4'te oluşturulan veri setinin öznitelikleri, veri kaynakları ve özniteliklerin açıklamaları listelenmektedir. 13 adet öznitelik ÖYS log kayırları kullanılarak türetilmiş olup geri kalan özniteliklerin oluşturulmasında ders kapsamında hazırlanan listeler ve anketler kullanılmıştır. Hedef nitelik için öğrencilerin faktör analizi sonucunda elde edilen DİÖ'nün 25 maddesine verdikleri cevapların ortalamaları hesaplanmıştır. Veri setinde ToplamOturumSuresi, OrtalamaOturumSuresi, OdeveHarcananZaman, SayfadaKalma_Ortalama, SınavOdakliCalisma ve VideoIzlemeSuresi adıyla geçen süre cinsinden özniteliklerin oluşturulmasında benimsenen yöntem genel hatlarıyla ve ilgili bağlam göz önünde bulundurularak etkileşimin başladığı ve bittiği log kayıtları arasındaki sürenin hesaplanması biçimindedir. Genel olarak web kullanım madenciliği uygulamalarında karşılaşıldığı üzere Zhou ve diğerleri (2017), kullanıcıların bir sayfayı ziyaretleri esnasında bilgisayardaki işlerini yarım bırakmaları, başka bir program veya web sayfası açmaları ve benzeri log raporlarına yansımayan durumların sayfada geçirilen süreleri hesaplamada kısıtlılık doğurabileceğini belirtmekle birlikte sistemdeki oturum süreleri *(session time)* kullanılarak yapılacak hesaplamaların sayfada geçirilen gerçek süreye varsayımsal olarak yaklaşabileceğini vurgulamışlardır. Bundan dolayı Log kayıtlarında etkileşimin başlangıcının görülebildiği ancak bitişiyle ilgili kaydın bulunmadığı durumlarda öğrencinin söz konusu etkinliğinin 30 dakika sürdüğü varsayılarak ilgili öğrencinin süre hesaplamalarına eklenmiştir.

3.3.VERİ TOPLAMA

Bu araştırmanın verisi araştırmanın yürütüldüğü 2020-2021 Güz, 2020-2021 Bahar ve 2021-2022 Güz akademik dönemlerinde aynı zamanda dersin eğitmeni de olan araştırmacı tarafından etik kurulu izni alınarak toplanmıştır (EK-4). Ayrıca tüm katılımcılardan araştırmaya gönüllü katılımlarını beyan eden onam formu onayı alınmıştır (EK-5).

Dersin yürütüldüğü ÖYS'nin kurulum, test ve içerik geliştirme aşamaları veri toplama süreci öncesinde gerçekleştirilmiştir. ÖYS'nin kurulu olduğu sunucunun veri tabanında depolanan log kayıtları ders dönemi başlangıcından final sınavı gününden bir gün önceye kadarki tarihleri kapsayacak şekilde ve öğrencilerin isim-soyisim bilgileri anonimleştirilerek raporlanmıştır. Bu bağlamda log kaydı verisinin toplanma süreci ilgili dersin işlendiği yarıyıl boyunca sürdürülmüştür.

DİÖ'de yer alan "Bu dersten yüksek bir not almak şans işidir." ifadesine sahip 6. madde, "Kendi notlarımla arkadaşlarımınkini karşılaştırdığımda bana adil davranıldığını düşünüyorum."

İfadesine sahip 14. madde, "Sınavlardan almayı düşündüğüm notla aldığım notu karşılaştırdığımda dersin hocasının değerlendirmelerinden memnunum." ifadesine sahip 18. madde ve "Bu derste çalışmalarımın karşılığını not ve diğer geribildirimlerle yeterince aldığımı düşünüyorum." ifadesine sahip 32. maddenin cevaplanabilmesi için söz konusu ölçek katılımcılara dersin vize sınavı sonrası notların duyurulmasını takiben uygulanmıştır. DİÖ'nün bu araştırma kapsamında Türkçe'ye çevrilen versiyonundaki ölçek maddelerinin yer aldığı Google Forms sayfası araştırmacı tarafından hazırlanarak dijital ortamda öğrencilerle paylaşılmıştır. Ders döneminin 8. haftasına denk gelen DİÖ uygulaması esnasında öğrencilere veri toplama amacı, araştırmanın kapsamı, gönüllü katılım esası ve kişisel veri gizliliği hakkında gerekli açıklamalar sözlü ve yazılı olarak yapılmış olup cevaplar dijital ortamda 15 dakikalık süre içerisinde toplanmıştır. Öğrencilere ait demografik bilgiler ise DİÖ uygulaması esnasında aynı dijital form içerisinde yer alan anket sorularından elde edilmiştir.

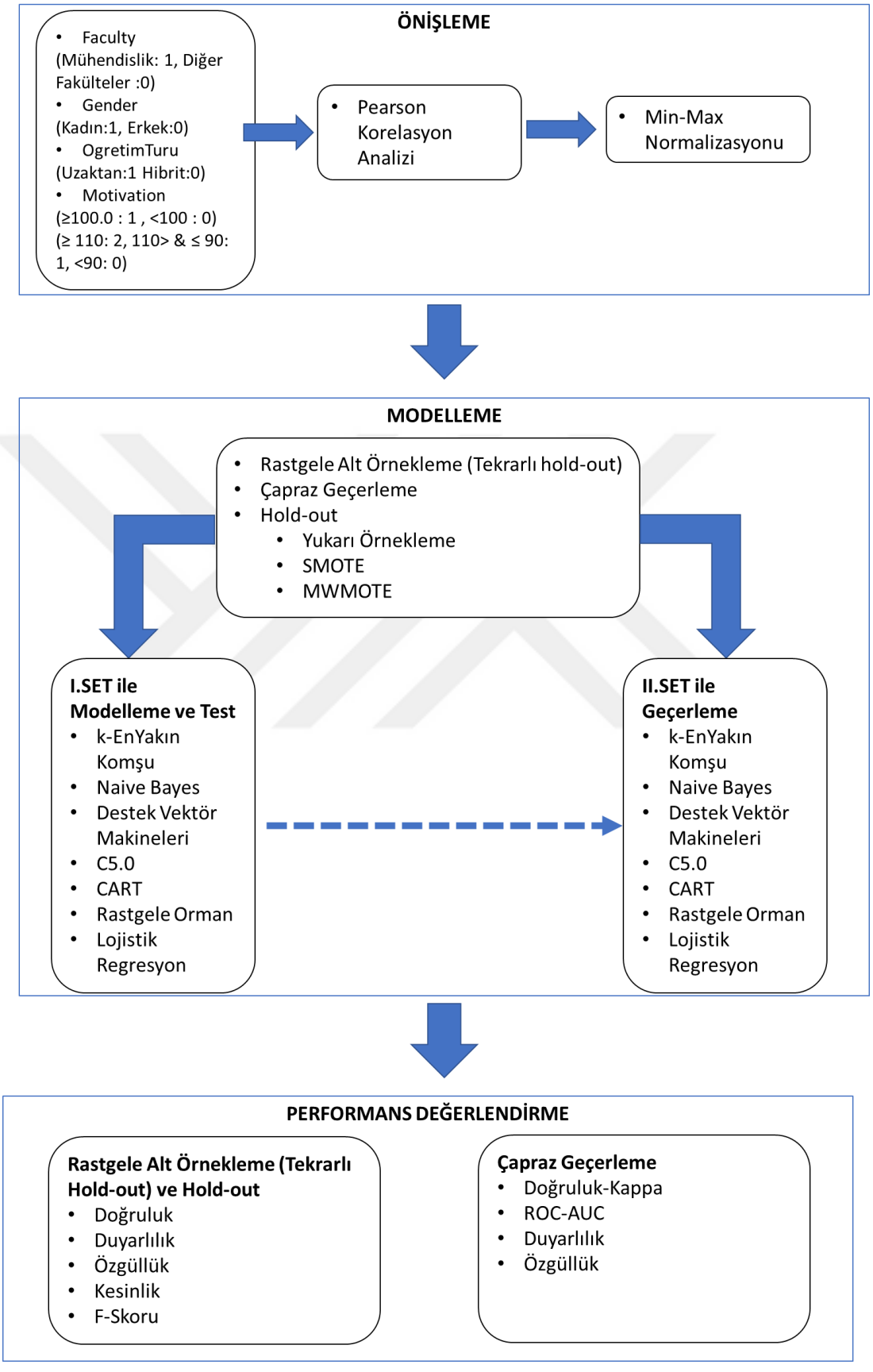
3.4. VERİ ANALİZİ

Veri analiz sürecini kapsayan işlemler istatistiksel programlama ve veri bilimi alanlarında sıklıkla kullanılan R Studio Entegre Geliştirme Ortamı 2022.12.0 Build 353 sürümü ile R programlama dilinin 4.2.2 versiyonu kullanılarak yürütülmüştür. R dili, istatistiksel programlama alanında çalışan araştırmacılar tarafından 1990'lı yıllarda geliştirilmiştir ve zamanla dışarıdan geliştiricilerin katkılarıyla genişleyen paket kütüphaneleri sayesinde doğrusal ve doğrusal olmayan modelleme, temel istatistiksel analiz, zaman serileri analizi, sınıflandırma, kümeleme ve benzeri analizleri yürütmek için kullanışlı hale getirilmiştir (Jones ve diğ., 2009). Veri analizinde kullanılan R paket ve fonksiyonları Tablo 3.5'te sunulmuştur.

Tablo 3.5: Veri Analizinde Kullanılan R Paket ve Fonksiyonları

Kullanım	Paket	Fonksiyon	Geliştirici
C5.0 Algoritması	C50	C5.0	Kuhn ve Quinlan (2023)
Veri Ayırma	caret	createDataPartition	Kuhn (2022)
Model Değerlerndirme ve Karmaşıklık Matrisi	caret	predict,confusionMatrix	Kuhn (2022)
Model Eğitimi	caret	train	Kuhn (2022)
Çapraz Geçerleme	caret	trainControl	Kuhn (2022)
k-En Yakın Komşu Algoritması	class	knn	Venables ve Ripley (2002)
Normalizasyon	clusterSim	data.Normalization	Walesiak ve Dudek (2020)
Naive Bayes Algoritması	e1071	naiveBayes	Meyer ve diğerleri (2023)
Destek Vektör Makineleri Algoritması	e1071	svm	Meyer ve diğerleri (2023)
MWMOTE Yöntemi	imbalance	mwmote	Cordón ve diğerleri (2018)
Rastgele Orman Algoritması	randomForest	randomForest	Liaw ve Wiener (2002)
Yukarı Örnekleme	ROSE	ovun.sample	Lunardon ve diğerleri (2014)
CART Algoritması	rpart	rpart	Therneau ve Atkinson (2022)
CART Karar Ağacı Görselleştirmesi	rpart	rpart.plot	Therneau ve Atkinson (2022)
Korelasyon Analizi	stats	cor	R Core Team (2022)
Lojistik Regresyon Algoritması	stats	glm	R Core Team (2022)
SMOTE Yöntemi	themis	smote	Hvitfeldt (2022)

Bu başlık altında öğrencilerin ders motivasyonlarının tahmin edilmesi amacıyla yürütülen veri analizi işlemleri veri tabanlarında bilgi keşfi adımlarıyla paralel olarak veri önişleme, modelleme ve değerlendirme başlıklarıyla açıklanmıştır. Önişleme ve modelleme sürecine dair süreç Şekil 3.3'te gösterilmektedir.



Şekil 3.3: Araştırmanın Veri Önişleme, Modelleme ve Performans Değerlendirme Süreci

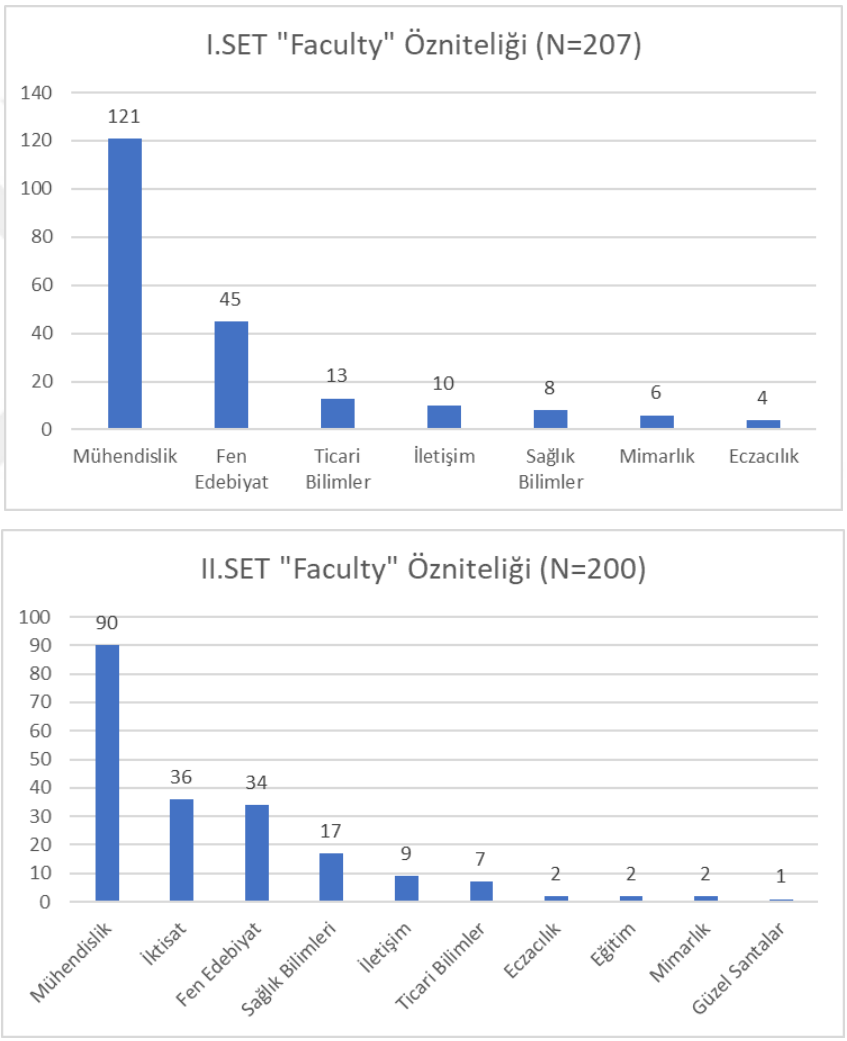
3.4.1. Veri Setinin Ön İşlenmesi

Veri önişleme sürecinin ilk adımında toplanan 407 katılımcının yer aldığı veri setinden 207 örneklik kayıt bu öğrencilerin araştırmacının yürütmüş olduğu şubelerde bulunmaları ve dolayısıyla aynı ders içerik ve materyallerine sahip ÖYS'lere kayıtlı olmaları sebebiyle modelleme seti olarak ayrılmıştır. Geri kalan öğrenciler ise aynı dersin araştırmacıdan farklı eğitmenlerce yürütülen şubelerine kayıtlı olup dersin ÖYS'yi ders içi etkinlik, materyal paylaşımı ve duyurulara erişim amacıyla daha kısıtlı bir çerçevede kullanmışlardır. 200 kişilik bu kayıtlar ise araştırmada geliştirilen modelin yeni veriyle test edilmesi amacıyla kullanılmıştır. Veri setinin bu farklılık gözetilerek işlevlerine göre ayrılması Tablo 3.6'da özetlenmiştir.

Tablo 3.6: İkiye Ayrılan Veri Setinin Özelliklerinin Karşılaştırılması

Veri Seti	Kayıt Sayısı	İşlevi	Moodle Kullanımı	DİÖ Uygulaması	Kullanılan Öznitelikler
I. SET	207	Modelleme	Yoğun	Evet	OgretimTuru, OturumAcmaSayisi, GecmisKonularaBakma, Gezinme Sayısı_Toplam, GezinmeSayisi_Ortalama, ToplamOturumSuresi, OrtalamaOturumSuresi, MateryalIndirme, OdevDenemeSayisi, OdeveHarcananZaman, SayfadaKalma_Ortalama, SınavOdakliCalisma, VideoIzlemeSuresi, YuklenenOdev, SınavNotu, Devamsizlik, Faculty, Class, Gender, Motivasyon (HEDEF NİTELİK)
II. SET	200	Validasyon	Kısıtlı	Evet	OgretimTuru, OturumAcmaSayisi, GecmisKonularaBakma, Gezinme Sayısı_Toplam, GezinmeSayisi_Ortalama, ToplamOturumSuresi, OrtalamaOturumSuresi, MateryalIndirme, SayfadaKalma_Ortalama, SınavOdakliCalisma, Faculty, Class, Gender, Motivasyon (HEDEF NİTELİK)

Şekil 3.4'te I. ve II. veri setinde öğrencinin kayıtlı olduğu fakülte bilgisini içeren "Faculty" özniteliğine ait frekans değerleri incelendiğinde I. veri setinde Mühendislik bölümü öğrencilerinin diğer fakültelere kayıtlı öğrencilere kıyasla daha fazla sayıda oldukları ve II. Veri setinde ise yarıya yakın sayıda öğrencinin Mühendislik fakültesine kayıtlı olduğu görülmektedir. Ayrıca iki veri setinden yalnızca birinde yer alan fakültelerin varlığı veri setlerindeki "Faculty" özniteliğine ait değerler arasında tutarsızlık doğurmaktadır.



Şekil 3.4: I. ve II. Veri Setlerinin "Faculty" Özniteliğine Ait Frekans Değerleri Grafikleri

Bu durumun modelleme ve test aşamalarında tutarsızlık ve hatalara yol açmaması adına Faculty özniteliğinin örnek çoğunluğuna sahip Mühendislik ve diğer fakülteler biçiminde Eşitlik 3.1'de gösterilen koşul çerçevesinde ikili kategorik değerlere dönüştürülmüştür.

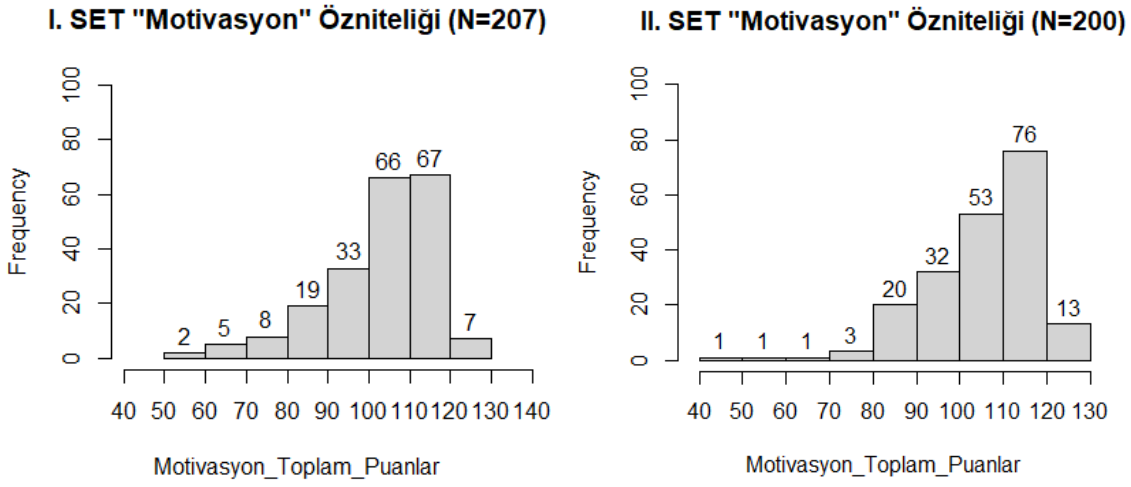
$$Faculty\begin{cases}1\ eğer\ Faculty = "Mühendislik" \\ 0\ eğer\ Faculty \neq "Mühendislik"\end{cases} \quad (3.1)$$

R dili kütüphanelerinde yer alan paketler ve bu paketler sayesinde işletilen algoritmalar sayısal değerlerle çalıştıklarından dolayı veri setinde sayısal olmayan değerlerin dönüştürülmesi gerekmektedir. Bu sebeple veri setinde "Gender" adındaki özniteliğe ait Kadın ve Erkek kategorik metin değerleri ve "OgretimTuru" özniteliğine ait Uzaktan ve Hibrit metin değerleri Eşitlik 3.2 ve 3.3'te gösterildiği gibi 1 ve 0 olarak sayısal hale getirilmiştir.

$$Gender\begin{cases}1\ eğer\ Gender = "Kadın" \\ 0\ eğer\ Gender = "Erkek"\end{cases} \quad (3.2)$$

$$OgretimTuru\begin{cases}1\ eğer\ OgretimTuru = "Uzaktan" \\ 0\ eğer\ OgretimTuru = "Hibrit"\end{cases} \quad (3.3)$$

I. ve II. veri setlerinde "Motivasyon" başlığıyla yer alan ve öğrencilerin DİÖ toplam puan değerlerini içeren hedef niteliğe ait frekans değerleri Şekil 3.5'te sunulmuştur.



Şekil 3.5: I. ve II. Veri Setlerinin "Motivasyon" Hedef Niteliğine Ait Frekans Değerlerini Gösteren Histogramlar

Şekil 3.5'te gösterilen frekans değerleri göz önüne alınarak "Motivasyon" hedef niteliğine ait değerleri düşük ve yüksek motivasyonu niteleyecek şekilde ikili sınıfa dönüştürmek için 100.0 değeri ayırıcı değer olarak belirlenmiştir. "Motivasyon" hedef niteliğini üçlü sınıfa dönüştürmek için ise 110.0 değeri yüksek ve orta motivasyonu ayıracak değer olarak, 90.0 değeri ise orta ve düşük motivasyon sınıflarını ayıracak değer olarak belirlenmiştir. Değerlerin ikili sınıfa dönüştürülme işlemi Eşitlik 3.4'te, üçlü sınıfa dönüştürülme işlemi ise Eşitlik 3.5'te gösterilmiştir.

$$Motivasyon_{\,2}\begin{cases}1\ eğer\ Motivasyon_TPuan \geq 100.0\\0\ eğer\ Motivasyon_TPuan < 100.0\end{cases} \tag{3.4}$$

$$Motivasyon_{\,3}\begin{cases}2\ eğer\ Motivasyon_TPuan \geq 110.0\\1\ eğer\ Motivasyon_TPuan \geq 90.0\ ve\ Motivasyon_TPuan < 110.0\\0\ eğer\ Motivasyon_TPuan \geq 0.0\ ve\ Motivasyon_TPuan < 90.0\end{cases} \tag{3.5}$$

Analiz aşaması öncesinde ve kategorik değer dönüşümlerinin ardından tüm örnekleme yöntemleri için Pearson Korelasyon Analizi uygulanarak yüksek korelasyon gösteren özniteliklerin belirlenmesi amaçlanmıştır.

Veri önişleme adımlarının bir sonraki aşamasında veri setindeki tüm numerik değerler normalize edilerek 0 ve 1 arası değerlere dönüştürülmüşlerdir. Makine öğrenmesinde sıklıkla kullanılan bir veri önişleme tekniği olan normalizasyon, veri setindeki değerleri ortak bir ölçeklendirmeye tabi tutarak değer aralıkları arasındaki farkların doğurabileceği sorunları önlemek amacıyla uygulanmaktadır. Normalizasyon işleminin özellikle sınıflandırma algoritmalarının eğitim sürecini hızlandırdığı ve daha iyi performans sağladığı belirtilmektedir (Al Shalabi, 2006). Bu çalışmada kullanılan veri setinin normalizasyonu için min-max normalizasyonu yöntemi kullanılmıştır.

$$x_i' = \frac{(x_i - \min(x))}{(\max(x) - \min(x))} \tag{3.6}$$

Eşitlik 3.5'te x_i' 0 ve 1 arasında normalize edilmiş değerleri, x_i değeri ise gerçek değerleri nitelemektedir. *min* değeri normalize edilmek istenen özniteliğe ait örnekler arasında görülen en düşük değeri ve *max* değeri aynı özniteliğe ait örneklerin en yüksek değerini ifade etmektedir.

Hedef niteliğin ikili sınıfa indirgendiği tüm durumlarda pozitif ve negatif sınıf dağılımlarının dengesiz halde oldukları görülmüştür. Dengesiz veri seti, hedef niteliğin sınıflarına ait örneklerin eşit veya birbirlerine yakın olarak dağılmadıkları durumu ifade etmektedir (Chawla, 2009). Bu durumun düzeltilmesi amacıyla veri setindeki örnekler arasındaki dengesizlik yukarı örnekleme, SMOTE *(Synthetic Minority Oversampling Technique)* ve MWMOTE *(Majority Weighted Minority Oversampling Technique)* teknikleri kullanılmıştır. Yukarı örnekleme, azınlıkta olan sınıf örneklerinin kopyalanması yoluyla yeni örnekler türetilerek veri setindeki dengesizliğin giderilmesi yöntemidir (Mohammed ve diğ., 2020). SMOTE tekniği ise iki sınıf arasındaki dengesizliği azınlık sınıfa ait örnekleri aynı örneklerle çoğaltılması yerine bu örneklerin sentetik kopyalarını oluşturarak gidermektedir (Gosain & Sardana, 2017). SMOTE algoritmasının işleyişi Chawla ve diğerleri (2002) tarafından aşağıdaki adımlarla açıklanmıştır:

1- Azınlık sınıftan bir x_0 örneği seçilir.

2- Seçilen x_0 örneğine azınlık sınıftan en yakın K adet komşu belirlenir.

3- x_0 örneğinin K adet komşusundan bir örnek rastgele olarak seçilir ve k değeri seçilen komşunun sırasını ifade edecek şekilde x_k olarak gösterilir.

4- x_0 ile x_k arasında aşağıdaki formül doğrusal interpolasyon hesaplanır ve yeni z sentetik örneği oluşturulur. Formüldeki w değeri ise 0 ve 1 arasında rasgele bir değişkeni ifade etmektedir:

$$z = x_0 + w(x_k - x_0) \tag{3.7}$$

5- Bu işlemler M sayıda sentetik örnek türetilene kadar tekrar edilir.

MWMOTE ise SMOTE tekniğinin yol açabileceği düşünülen gürültülü verinin sentetik olarak çoğaltılması problemini ve yanlış sentetik örneklerin oluşturulmasını önlemek amacıyla geliştirilmiş bir örnekleme yöntemidir. MWMOTE tekniği en zorlu azınlık sınıf örneklerini analiz ederek her bir örneğe çoğunluk sınıfına olan uzaklıklarını gözeterek bir ağırlık değeri

atamaktadır. Sentetik örnekler daha sonra bir kümeleme yaklaşımı kullanılarak ağırlıklı değeri atanmış olan azınlık sınıfı örneği bir azınlık sınıfı kümesi içinde yer alması gözetilerek türetilmektedir (Barua ve diğ., 2014). Yukarı örnekleme tekniklerinin yanı sıra bu çalışma dâhilinde kullanılan veri setinin mevcut örnek sayısı korunarak gerçekleştirilen çapraz geçerleme ve rastgele alt örnekleme veya tezin ikinci bölümünde açıklandığı haliyle tekrarlı hold-out tekniğiyle elde edilen sonuçlar da paylaşılmıştır.

3.4.2. Modelin Eğitilmesi

Öğrencilerin derse dönük motivasyonlarının tahmin edilmesi amacıyla geliştirilen sınıflandırma modellerinin eğitilmesinde k-en yakın komşu, Naive Bayes, Destek Vektör Makineleri, C5.0, CART, Rastgele Orman, ve Lojistik Regresyon algoritmaları kullanılmıştır ve bu algoritmaların uygulanması için kullanılan R fonksiyonlarına gönderilen parametreler Tablo 3.12'de sunulmuştur.

Tablo 3.7: Araştırmada Kullanılan Sınıflandırma Algoritmaları Fonksiyonlarına Ait Parametreler

Algoritma	Fonksiyon	Parametreler
C5.0 Algoritması	C5.0	C5.0(train_Attributes, train_Target)
k-En Yakın Komşu Algoritması	knn	knn(train = train_Attributes,test = test_Attributes,cl = train_Target,k = 5)
Naive Bayes Algoritması	naiveBayes	naiveBayes(motivation ~.,data=train_data)
Destek Vektör Makineleri Algoritması	svm	svm(formula= motivation ~.,data=train_data,type="C-classification",kernel="linear")
Rastgele Orman Algoritması	randomForest	randomForest(motivation~.,data=train_data,proximity=TRUE)
CART Algoritması	rpart	rpart(motivation~., train_data,method="class")
Lojistik Regresyon Algoritması	glm	glm(motivation~.,family= "binomial", data=train_data)

3.4.3. Modelin Değerlendirilmesi

Bu araştırmada eğitilen yapay zekâ modellerinin performanslarının değerlendirilmesi amacıyla tanımlama ve açıklamaları Genel Kısımlar bölümünün Danışmanlı Öğrenme başlığı altında yapılan yöntemlerden rastgele alt örnekleme (tekrarlı hold-out), çapraz geçerleme ve hold-out kullanılmıştır. Hold-out yöntemi sırasıyla yukarı örnekleme, SMOTE ve MWMOTE yöntemleriyle örnek sayısı artırılan veri setine uygulanmıştır. Sonuçlar karmaşıklık matrisi ile

incelenmiştir ve modellerin performanslarını karşılaştırmak adına karmaşıklık matrisi ile hesaplanan doğruluk, duyarlılık, özgüllük ve F-skoru değerleri Bulgular bölümünde sunulmuştur.

4. BULGULAR

Tezin Bulgular bölümünde öğrencilerin ders motivasyonlarını tahmin etmek amacıyla farklı yöntemlerle eğitilen sınıflandırıcı yapay zekâ modellerine ait sonuçlar önce kullanılan örnekleme yöntemi, analiz edilen veri seti ve ardından hedef niteliğin ikili sınıfa indirgendiği değerler esas alınarak yapılan bir sıralamayla sunulmuştur.

4.1. DERSE İLGİ ÖLÇEĞİ (DİÖ) GEÇERLİLİK VE GÜVENİRLİK ANALİZİ BULGULARI

Bu başlık altında DİÖ'nün mevcut araştırmada kullanılmak üzere yürütülen Türkçe'ye adaptasyon çalışmaları aktarılmaktadır. Birinci aşamada araştırmanın katılımcıları dışındaki Temel Bilgisayar Uygulamaları ve benzer içeriğe sahip derslere kayıtlı 305 üniversite öğrencisine ait tanımlayıcı demografik bilgiler Tablo 4.1'de sunulmuştur.

Tablo 4.1: DİÖ Faktör Analizi Örneklemine Ait Tanımlayıcı Demografik Bilgiler (N=305)

Sınıf	**Cinsiyet**	Temel Bilgisayar Uygulamaları	Öğretim Teknolojileri	Bilişim Teknolojileri	Toplam
1	Erkek	12	-	12	107
	Kadın	15		68	
2	Erkek	16	4	-	52
	Kadın	18	7	7	
3	Erkek	19	-	-	32
	Kadın	11	2	-	
4	Erkek	63	-	-	104
	Kadın	41	-	-	
5	Erkek	9	-	-	10
	Kadın	1	-	-	
Toplam		205	13	87	305 (E: 135, K:170)

Tablo 4.1 incelendiğinde DİÖ faktör analizi için erişilen katılımcıların çoğunluğunu Kadın öğrencilerin oluşturduğu ve tüm katılımcılar içerisinde 1. ve 4. Sınıf öğrencilerinin sayısı diğer

sınıflara ait katılımcılardan fazla olduğu görülmektedir. Tablo 4.2'de mevcut örneklemin sınıflara göre ortalama, mod, medyan, en düşük ve en yüksek değerleri gösterilmiştir.

Tablo 4.2: DİÖ Faktör Analizi Örnekleminin Sınıflara Göre Ortalama, Mod, Medyan, En Düşük ve En Yüksek Değerleri

	1. Sınıf (N=107)	**2. Sınıf (N=52)**	**3. Sınıf (N=32)**	**4. Sınıf (N=104)**	**5. Sınıf (N=10)**	**Tüm Örneklem (N=305)**
Ortalama	100,93	104,52	105,84	106,36	102,1	103,95
Mod	106	112	115	117	-	113
Medyan	104	106	113	109	100,5	106
En Düşük	48	52	49	68	86	48
En Yüksek	123	124	125	124	122	125

Bu örneklemden elde edilen veri üzerinde yürütülen güvenirlik analizi ile literatürde rastlanan DİÖ güvenirlik analizi sonuçlarına ait Cronbach Güvenirlik Katsayısı puanları Tablo 4.3'te karşılaştırmalı olarak sunulmuştur.

Tablo 4.3: Derse İlgi Ölçeği (DİÖ) Güvenilirlik Analizi Sonuçlarının Literatürdeki Sonuçlarla Karşılaştırılması

Çalışmalar	**Cronbach Alpha Katsayısı**
Keller, 2010 (Ölçeğin geliştiricisi)	*0,95*
Huett ve diğerleri, 2008	0,93
Acar, 2009	0,93
Metraglia ve Villa, 2014	0,92
Ucar ve Kumtepe, 2019	0,91
Peak ve diğerleri, 2008	0,91
Wah, 2015	0,88
Bu çalışma	**0,87**
Al Qatawneh, 2014	0,87
Naveed & Bhatti, 2017	0,86
Lee, 2018	0,82
Gomes ve diğ., 2018	0,81
Metraglia ve diğerleri, 2015	0,76

Güvenilirlik analizi sonucu elde edilen Cronbach Alpha katsayısı bir ölçme aracının içsel tutarlılığını ölçmek amacıyla kullanılmakta ve 0,70 ile 0,95 arası değerler uygun olarak kabul

edilmektedir (Tavakol & Dennick, 2011). Tablo 4.3'teki değerler incelendiğinde bu çalışmada elde edilen Cronbach Alpha katsayısının (α=0,87) kabul edilebilir aralıkta ve aynı ölçek için yürütülen güvenilirlik analizleri sonucunda elde edilen katsayılara kıyasla ortalama değerde olduğu görülmüştür.

İkinci aşamada Geçerlik analizinin yürütülmesi amacıyla ölçek maddelerine Açımlayıcı Faktör Analizi (AFA) ve Doğrulayıcı Faktör Analizi (DFA) uygulanmıştır. AFA, bir ölçme aracına ait maddelerin boyutlara dayalı yapısını ortaya çıkarmak ve bu boyutlarla olan ilişkisini incelemek amacıyla yürütülmektedir. Diğer bir deyişle AFA ile amaçlanan bir ölçme aracına ait maddelerin hangi faktörlere nasıl dağıldıklarını incelemektir ve genellikle şu adımlardan oluşmaktadır: (1) Değişkenlerin seçilmesi, (2) Faktör sayısının belirlenmesi, (3) Faktörlerin çıkarılması, (4) Faktörlerin döndürülmesi, (5) Faktörlerin yorumlanması (Fabrigar & Wegener, 2011). DFA ise bir ölçme aracının AFA ile belirlenen faktör yapısını doğrulamak için kullanılmaktadır. AFA ile ölçme aracına ait maddelerin belirli sayıdaki faktöre dağılımı ortaya çıkarılırken; DFA ile faktörler arasındaki ilişkilerin düzeyi ve maddeler ve faktörler arasındaki aidiyet ilişkileri Yapısal Eşitlik Modeli (YEM) kullanılarak incelenmektedir (Erkorkmaz ve diğ., 2013). YEM, bir modelin kuramsal ve istatistik temelli yönünün o modelin yapısı arasındaki belirleyici ilişkinin incelenmesin sağlayan yöntemdir (Baştürk ve diğ., 2013).

AFA ve DFA öncesinde DİÖ'ye ait maddeler arası korelasyon katsayıları incelenmiş olup diğer maddelerle yüksek negatif korelasyon gösteren 10. madde ile pozitif veya negatif korelasyon göstermeyen 30. maddenin veri setinden çıkarılmasına karar verilmiştir. Ayrıca araştırma kapsamında veri toplanan ders bir dönem uzaktan eğitim yöntemiyle geri kalan dönemlerde ise hibrit eğitim yöntemiyle yürütüldüğünden dolayı katılımcıların belirgin bir cevap vermekte zorlanabilecekleri "Sınıftaki öğrenciler bu derse aktif olarak katılıyorlar." ifadesine sahip 22. maddenin de veri setinden çıkarılması kararlaştırılmıştır.

31 maddeli DİÖ için AFA kapsamında yürütülecek analizlerin öncesinde bir ölçme aracının faktör analizine uygunluğunu belirlemek amacıyla KMO (Kaiser-Meyer-Olkin) ve Bartlett's testleri yürütülmüştür. KMO testi, örneklem büyüklüğünün yeterliliğini modeldeki her bir değişken için ölçmektedir ve 0,8 ila 1,0 arası değerlerin ideal olarak değerlendirilmektedir. Barttlet Küresellik Testi ise ölçek maddeleri arası ilişki olmadığına dönük boş hipotezi test ederek bir anlamlılık değeri üretmektedir ve 0,05'ten düşük değerler ilgili veri seti üzerinde faktör analizi yürütülebileceğini göstermektedir (Shrestha, 2021).

Tablo 4.4: KMO ve Bartlett Küresellik Testi Sonuçları

Kaiser-Meyer-Olkin (KMO) Testi		0,901
Bartlett Küresellik Testi	Yaklaşık Ki-Kare	4659,569
	Serbestlik derecesi (df)	465
	Anlamlılık (p)	0,000*

*<,05

Tablo 4.4 incelendiğinde bu araştırma kapsamında elde edilen DİÖ veri setinin faktör analizi için uygun değerlere sahip olduğu görülmektedir (KMO=0,901; Bartlett's p<,05). Faktör analizinin bir sonraki adımında belirli sayıdaki değişkeni bileşenler adıyla anılan daha az sayıda değişkene indirgemek amacıyla kullanılan Temel Bileşenler Analizi (TBA) uygulanmıştır (Richardson, 2009). TBA, bir yandan verideki boyutluluğu doğrusal fonksiyonlar ile azaltmakta diğer yandan ise verinin istatistiksel çeşitliliğini varyansı maksimize ederek korumaktadır (Jolliffe & Cadima, 2016). TBA sonucunda 4 faktörlü yapının toplam varyansın %53,64'ünü açıkladığı görülmüştür ve bu değer sosyal bilimler alanında yürütülen TBA sonuçlarında ideal toplam varyans açıklama değerinin %50 ve üzeri olması gerektiğini belirten çalışmalarla uyumludur (Tinsley & Tinsley, 1987; Merenda, 1997). Ayrıca TBA çıkarım metodu olarak Eigenvalue uygulandığında 6 faktörlü bir yapı elde edilmekte ve böylelikle toplam varyansın %61,02 açıklansa da ölçeğin orjinalinin 4 faktörlü yapıda olması ve çalışmada faktörlere ait puanların yerine ortalama ölçek puanı kullanılacağından dolayı ideal değerin üzerinde toplam varyans açıklama değerine sahip olan 4 faktörlü yapının kullanılmasına karar verilmiştir.

EK-1'de sunulan TBA ile elde edilen Döndürülmüş Faktör Matrisi sonuçları hangi maddelerin kaçıncı faktör altında gruplandırıldığını göstermektedir. Bu sonuçlara göre ölçek maddelerinin Tablo 3.3'te sunulan boyutlarla aynı biçimde dağılmadıkları gözlenmekte ise de Keller (2010) tarafından DİÖ'nün geliştirildiği çalışmada yalnızca güvenirlik analizi sonuçlarının yürütülmüş olması ve benzer şekilde DİÖ'nün Türkçe uyarlamasını yapan Dinçer (2015)'in bulguları kapsamında da ölçek maddeleri dağılımının orijinal çalışmada yer alan madde-boyut dağılımıyla aynı biçimde olmadığının görülmesi sebebiyle ölçek maddelerinin dağılımı ile ilgili bu durumun DFA uygulaması için bir engel teşkil etmediği düşünülmüştür.

Ayrıca EK-1'de verili maddelerin boyutlara dağılımı göz önüne alınarak hesaplanan Cronbach Güvenirlik Katsayılarının da uygun değerlerde oldukları görülmektedir. Yine ilgili tablodaki matris üzerinde Faktör 1'e ait 14, Faktör 2'ye ait 9, Faktör 3'e ait 6 ve Faktör 4'e ait 2 maddenin bulunduğu saptanmıştır ve tüm ölçek için madde puan ortalamaları 2,22 ile 4,72 arasındadır. EK-1'de yer alan döndürülmüş bileşen matrisindeki 31 maddenin 4 faktöre dağılımı esas alınarak DFA uygulanmıştır. YEM yöntemiyle yürütülen analiz sonucunda elde edilen ilk modele ait faktör yükleri ve kovaryans değerlerini gösteren diyagram EK-2'de görülmektedir.

31 maddeli DİÖ için uygulanan ilk YEM analizi sonuçlarına bakıldığında model uyumunun ölçümü için kullanılan RMSEA değerinin kabul edilebilir düzeyde olmadığı görülmektedir. RMSEA *(Root Mean Square Error of Approximation- yaklaşık hataların ortalama karekökü)*, yapısal eşitlik modellemesi analizlerinde genel kabul gören mutlak uyum indeksidir ve modele ait hipotezin mükemmelliğe veya mutlak uyuma ne kadar yakın veya uzak olduğunu ölçmektedir (Xia & Yang, 2019). Browne ve Cudeck (1993), 0,05 veya daha düşük RMSEA değerinin modelin serbestlik derecesine yakın bir uyumu gösterdiğini; 0,08-0,05 arasında bir RMSEA değerinin ise model açısından kabul edilebilir bir yaklaşım hatası değerini nitelediğini ancak 0.1 ve üstü değerlerin model kurmak için yetersiz olduğunu belirtmişlerdir. Ayrıca MacCallum ve diğerleri (1996) de benzer şekilde RMSEA için 0,1'den yüksek değerlerin zayıf, 0,08 ile 0,1 arasındaki değerlerin sınırda, 0,05 ile 0,08 arasındaki değerlerin kabul edilebilir ve 0,05'ten düşük değerlerin mükemmel uyum olarak kabul edilmesi gerektiğini vurgulamışlardır.

Bu doğrultuda modelin iyileştirilmesi amacıyla öncelikle düşük faktör yüklerine sahip olan "Bu derste başarılı olacağıma inancım tam." ifadesine sahip 3. madde, "Bu derste başarılı olmak için çok çalışmam gerekiyor." ifadesine sahip 7. madde, "Bu dersi geçip geçemeyeceğim bana bağlıdır" ifadesine sahip 9. madde," Bu derste yüksek hedefler belirleyerek mükemmele ulaşmaya çalışıyorum." ifadesine sahip 13. madde, "Bu dersteki öğrenciler dersin konularına ilgili görünüyorlar." ifadesine sahip 15. madde, ve "Bu derse sıkı çalıştığım takdirde başarılı olacağıma inanıyorum." ifadesine sahip 27. madde veri setinden çıkarılmıştır. Ayrıca YEM analizi sonucunda elde edilen düzeltme indisleri *(modification indices)* tablosu kullanılarak aynı faktöre ait maddeler arasındaki kovaryanslar oluşturulmuştur. Modelin iyileştirilmesi için yapılan değişiklikler sonucunda EK-3'teki diyagram elde edilmiştir.

Model iyileştirme adımlarının uygulanması ile birlikte RMSEA değeri 0,066 olarak kabul edilebilir seviyede ölçülmüştür. RMSEA değeri ile birlikte YEM analizi sonucunda elde edilen

modelin uyumunun değerlendirilmesi için kullanılan diğer ölçütler ve literatürde önerilen değerlerle karşılaştırılması Tablo 4.5‘te sunulmuştur.

Tablo 4.5: YEM Analizi Sonucunda Elde Edilen CMIN/DF, NFI, CFI, GFI ve RMSEA Model Uyum İndeksi Değerleri ve Önerilen Değerlerin Karşılaştırılması

Ölçüt	Elde Edilen Değer	Önerilen Değer
CMIN/DF	2,309	≤3 veya ≤5 (Kline, 2015; Marsh & Hocevar, 1985)
NFI	0,856	0,9'a yakın değerler (Bentler & Bonett, 1980)
CFI	0,912	≥0,9 (Fan & diğ., 1999)
GFI	0,867	≥0,9 (Hu & Bentler, 1998)
RMSEA	0,066	≤0,08 (Browne & Cudeck, 1993)

Tablo 4.5’teki sonuçlar incelendiğinde elde edilen model uyum indeksi değerlerinin literatürde önerilen değer aralıkları içerisinde veya bu aralıklara yakın değerde oldukları görülmüştür (CMIN/DF=2,309; NFI=0,856; CFI=0,912; GFI=0,867; RMSEA=0,066). Bir yapısal eşitlik modeli uyumunun değerlendirilmesi açısından farklı istatistiksel hesaplamaları temel alan birden fazla uyum indeksinin literatürde kabul gören değerlere sahip olması sebebiyle ölçeğin bu haliyle kullanılmasına karar verilmiştir. Faktör analizi süreci sonunda oluşan 25 maddeli ve 4 faktörlü DİÖ için korelasyon analizi sonuçları Tablo 4.6’da sunulmuştur.

Tablo 4.6: 25 Maddeli ve 4 Faktörlü DİÖ Alt Boyutlarına Ait Korelasyon Analizi Sonuçları

	Faktör 1	**Faktör 2**	**Faktör 3**	**Faktör 4**	**DİÖ Ortalaması**
Faktör 1	1	,258**	,642**	,223**	,851**
Faktör 2		1	,264**	-,317**	,660**
Faktör 3			1	,133*	,752**
Faktör 4				1	,161**
DİÖ Ortalaması					1
**. 0,01’den düşük anlamlı korelasyon					
*. 0,05’ten düşük anlamlı korelasyon					

Tablo 4.6'da gösterilen korelasyon analizi sonuçlarına bakıldığında faktör analizi ile elde edilen alt boyutlar arasındaki korelasyon katsayılarınn -0,317 ile 0,642 arasında değişkenlik gösterdiği ve DİÖ ortalaması da dahil olmak üzere tüm alt boyutlar arasında istatistiksel olarak anlamlı ilişkilerin varlığı saptanmıştır. Sonuç olarak, 5'li likert tipindeki DİÖ'nün 25 maddeli ve 4 faktörlü mevcut şekliyle hesaplanabilecek en yüksek toplam puan 125, en düşük toplam puan 25 ve ortalama puan 75'tir. Bu tez araştırması kapsamında faktör analizi sonucunda elde edilen 25 maddeli DİÖ'nün belirtilen aralıklardaki toplam puanı kullanılmıştır. Tablo 4.7'de araştırmanın veri setinin elde edildiği katılımcılara ait DİÖ toplam puanlarınn ortalama, mod, medyan, en düşük ve en yüksek değerleri bağlı bulundukları sınıfa göre gösterilmektedir.

Tablo 4.7: Araştırma Örnekleminin (N=407) Sınıflara Göre DİÖ Ortalama, Mod, Medyan, En Düşük ve En Yüksek Değerleri

	1. Sınıf (N=55)	**2. Sınıf (N=68)**	**3. Sınıf (N=78)**	**4. Sınıf (N=188)**	**5. Sınıf (N=18)**	**Tüm Örneklem (N=407)**
Ortalama	96,945	100,897	106,667	106,782	102,444	104,256
Mod	104	113	114	114	97	102
Medyan	103	103	111	109	100,5	106
En Düşük	48	50	71	61	88	48
En Yüksek	123	124	125	124	122	125

Tablo 4.7 incelendiğinde 1.sınıf öğrencilerinin DİÖ toplam puanları ortalamasının diğer sınıflardan düşük ve tüm örneklem ortalamasının da altında olduğu görülmektedir. En yüksek DİÖ puan ortalaması 106,782 ile 4. sınıf öğrencilerine aittir. Tablo 4.8'de araştırmanın veri setinin elde edildiği katılımcılar ait DİÖ toplam puanlarının ortalama, en düşük ve en yüksek değerleri gösterilmiştir.

Tablo 4.8: Araştırma Örnekleminin (N=407) Fakültelere Göre DİÖ Ortalama, En Düşük ve En Yüksek Değerleri

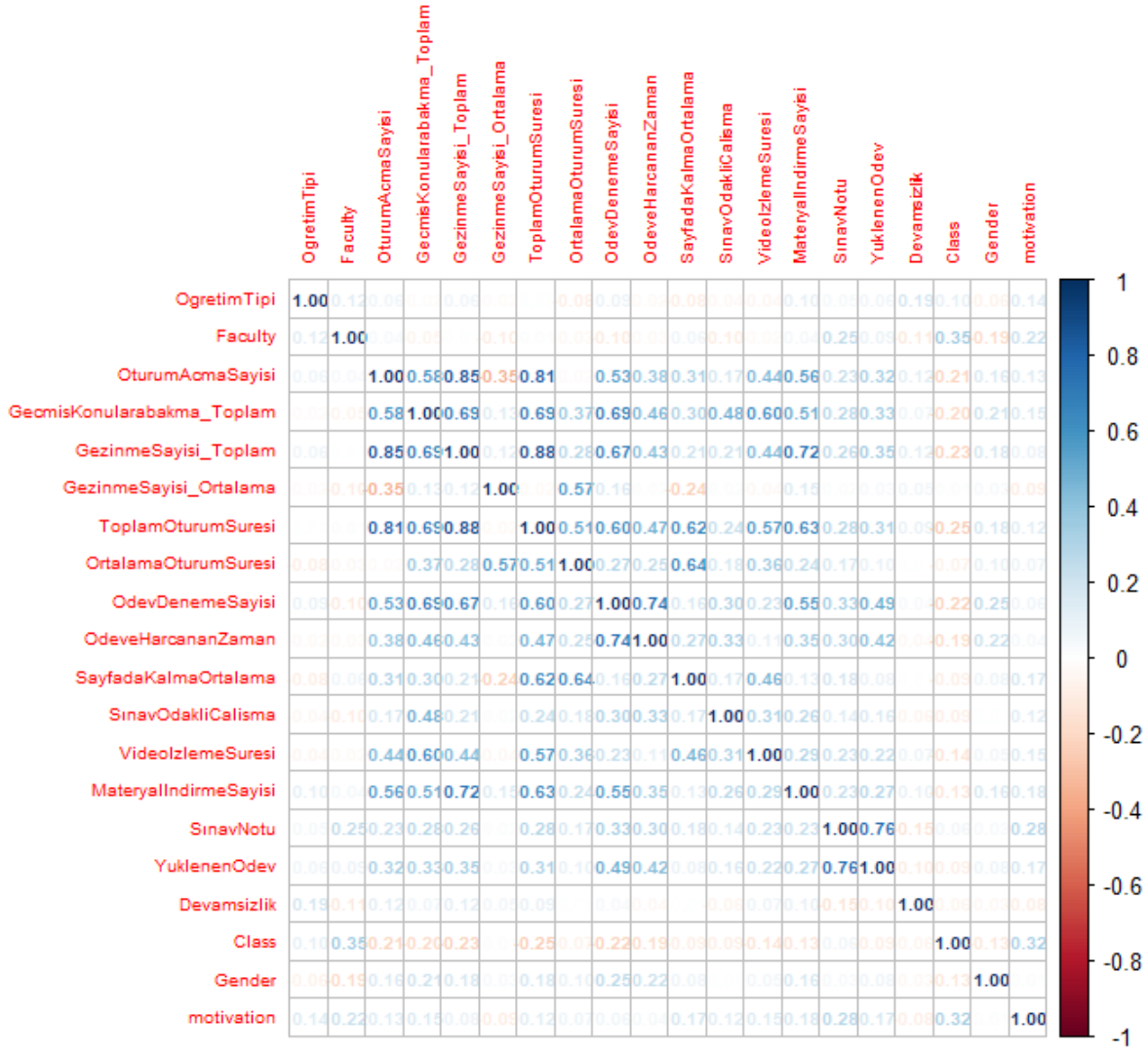
Fakülte	Kişi Sayısı	Ortalama	En Düşük	En Yüksek
Eczacılık	6	90,83	74	106
Eğitim	2	89,50	86	93
Fen-Edebiyat	79	98,48	48	121
Güzel Sanatlar	1	107	107	107
İktisat	36	106,44	87	124
İletişim	19	107,05	85	117
Mimarlık	8	98,63	76	113
Mühendislik	211	106,00	54	125
Sağlık Bilimleri	25	106,24	73	121
Ticari Bilimler	20	107,25	89	123

Tablo 4.8'e bakıldığında DİÖ toplam puanının ortalamasına göre en yüksek üç grubun Ticari Bilimler, İletişim ve İktisat Fakültesi öğrencileri oldukları görülmektedir.

4.2. RASTGELE ALT ÖRNEKLEME (TEKRARLI HOLD-OUT) YÖNTEMİYLE ELDE EDİLEN BULGULAR

Bu bölümün ilk aşamasında sadece I.SET (N=207) rastgele alt örnekleme yöntemiyle modellenmiş ve kendi içinde %70-%30 ayrımı esas alınarak test edilmiş olup ikinci aşamada ise I.SET'in tüm örnekleri kullanılarak geliştirilen modeller II.SET (N=200)'in örnekleriyle test edilmiştir. Son aşamada ise 3 sınıfa indirgenen I. SET (N=207) ile birinci ve ikinci aşamadaki işlemler tekrarlanmıştır.

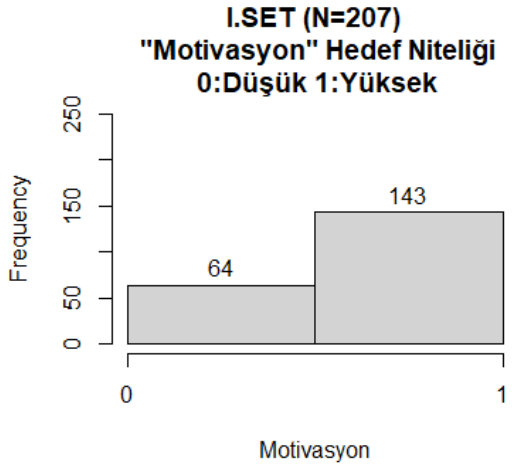
Modelleme işlemlerinden önce I.SET'in öznitelikleri arasındaki korelasyon katsayılarını görebilmek amacıyla Pearson Korelasyon testi uygulanmıştır ve elde edilen sonuçlar Şekil 4.1'de gösterilmiştir.



Şekil 4.1: I. SET (N=207) Özniteliklerine Ait Pearson Korelasyon Testi Sonuçları

Şekil 4.1'de 0,80'den yüksek korelasyon değerlerinin "OturumAcmaSayisi" ve "GezinmeSayisi_Toplam" öznitelikleri arasında (r=0,85; p<0,01), "ToplamOturumSuresi" ve "OturumAcmaSayisi" öznitelikleri arasında (r=0,81; p<0,01) ve "ToplamOturumSuresi" ve "GezinmeSayisi_Toplam" öznitelikleri arasında (r=0,88; p<0,01) olduğu görülmektedir. Süre cinsinden bir özniteliğin veri setinde kalmasını sağlamak amacıyla "OturumAcmaSayisi" ve "GezinmeSayisi_Toplam" özniteliklerinin veri setinden çıkarılmasına karar verilmiştir.

"Motivasyon" hedef niteliğinin ikili sınıfa "4.0" değerine göre indirgendiğinde oluşan pozitif ve negatif sınıfların frekans değerleri Şekil 4.2'de gösterilmiştir.



Şekil 4.2: Rastgele Örnekleme Uygulanan I. SET (N=207)'in "Motivasyon" Hedef Niteliğine Ait İki Sınıflı Frekans Değerleri (Pozitif Sınıf ≥ 100.0)

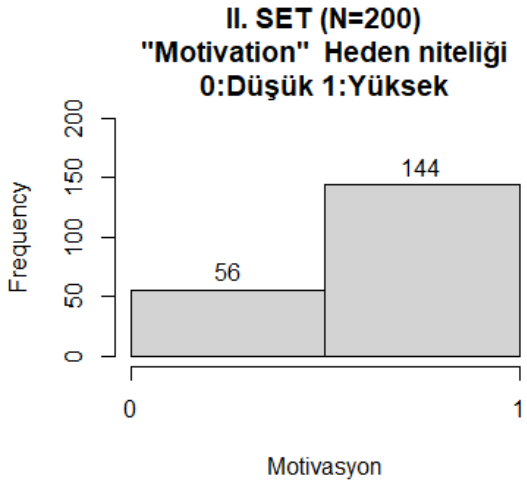
Şekil 4.2'de "Motivasyon" hedef niteliğinde tutulan DİO toplam puanları 100,0 değerine göre iki sınıfa ayrıldığında negatif sınıfa ait 64 ve pozitif sınıfa ait 143 örnek bulunmaktadır.

I.SET (N=207) kullanılarak 200 tekrarlı rastgele alt örnekleme (tekrarlı hold-out) yöntemiyle elde edilen ve her tekrarda eğitim-test verisi ayrımı %70'a %30 olarak belirlenen modellerin performans bulguları Tablo 4.9'da sunulmuştur.

Tablo 4.9: Rastgele Alt Örnekleme Uygulanan İki Sınıflı I. SET(N=207) ile Eğitilen Modellerin Performansları

		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	0,697	0,861	0,335	0,742	0,796
	k-EYK k=5	0,702	0,889	0,289	0,735	0,804
	k-EYK k=7	0,694	0,898	0,243	0,724	0,801
	k-EYK k=9	0,690	**0,911**	0,200	0,716	0,801
	Naive Bayes	0,700	0,767	**0,551**	**0,792**	0,777
	DVM	0,717	0,866	0,387	0,759	0,808
	C5.0	0,672	0,790	0,409	0,748	0,767
	CART	0,684	0,806	0,414	0,754	0,777
	Rast.Orm.	**0,727**	0,902	0,341	0,752	**0,820**
	Loj. Reg.	0,716	0,865	0,386	0,758	0,807

Tablo 4.9'daki performans değerlerine bakıldığında doğruluk değerlerinin 0,67 ile 0,72 aralığında olduğu ve en yüksek performans gösteren algoritmanın Rastgele Orman olduğu görülmektedir. Negatif sınıfa dair örnek sayısının azlığı dolayısıyla negatif sınıf örneklerinin doğru sınıflandırılma oranını ifade eden özgüllük değerleri düşük seviyelerdedir. Özgüllük değeri en yüksek algoritmanın ise 0,55 ile Naive Bayes algoritması olduğu gözlenmiştir.



Şekil 4.3: II. SET (N=200)'in "Motivasyon" Hedef Niteliğine Ait İkili Sınıf Frekans Değerleri

Şekil 4.3'te II.SET (N=200)'e ait 56 negatif ve 144 pozitif örnek bulunduğu görülmektedir. I.SET (N=207)'in tüm örnekleri kullanılarak oluşturulan modeller II.SET (N=200) üzerinde test edilmiştir ve bu analizin sonuçları Tablo 4.10'da sunulmuştur.

Tablo 4.10: Rastgele Alt Örnekleme Uygulanan İki Sınıflı I. SET (N=207) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları

		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	0,565	0,563	0,571	0,771	0,651
	k-EYK k=5	0,623	0,740	0,321	0,737	0,738
	k-EYK k=7	0,660	0,757	0,411	0,768	0,762
	k-EYK k=9	0,635	0,715	0,429	0,763	0,738
	Naive Bayes	0,425	0,243	**0,893**	**0,854**	0,378
	DVM	**0,725**	0,813	0,500	0,807	**0,810**
	C5.0	0,690	**0,819**	0,357	0,766	0,792
	CART	0,435	0,313	0,750	0,763	0,443
	Rast.Orm.	0,365	0,161	0,890	0,789	0,266
	Loj. Reg.	0,705	0,778	0,518	0,806	0,792

Tablo 4.10 incelendiğinde doğruluk değeri en yüksek algoritmanın 0,725 değerine sahip Destek Vektör Makineleri algoritması olduğu görülmektedir. Ayrıca Destek Vektör Makineleri algoritmasının performansına yakın doğruluk ve özgüllük değerlerine sahip olan Lojistik Regresyon algoritması da göze çarpmaktadır.

Rastgele alt örnekleme uygulanan I.SET (N=207) ile geliştirilen modeller arasında diğer algoritmalara kıyasla ve tüm performans değerleri göz önüne alındığında yüksek performans sergileyen Naive Bayes algoritmasıyla oluşturulan modelde hangi özniteliklerin "Motivasyon" değişkenini en fazla oranda yordadığını görebilmek amacıyla öznitelik önemi değerleri hesaplanarak Tablo 4.11'de sunulmuştur.

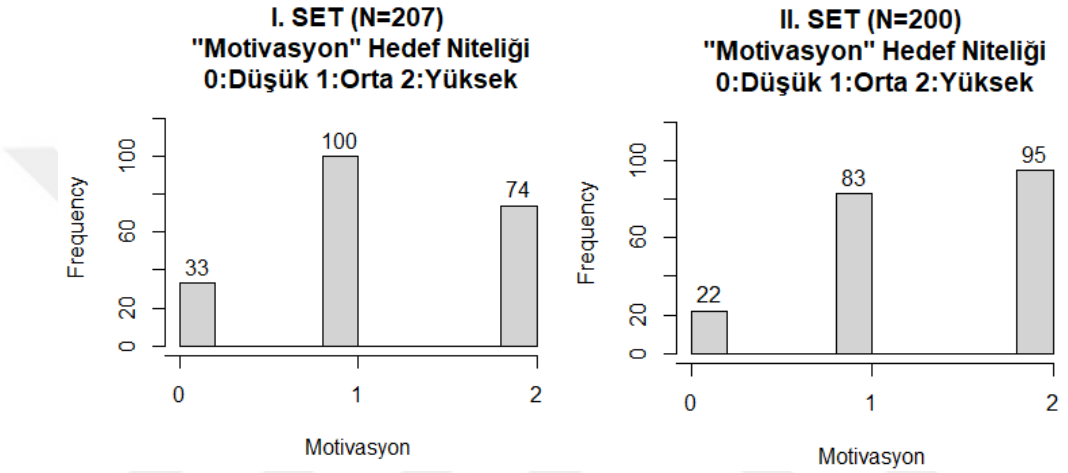
Tablo 4.11: Rastgele Alt Örnekleme Uygulanan İki Sınıflı I.SET (N=207) ile Eğitilen Naive Bayes Modeline Ait Öznitelik Önem Sıralaması

Öznitelik	Öznitelik Önemi
SınavNotu	100
YuklenenOdev	67,92
GecmisKonularabakma_Toplam	66,98
MateryalIndirmeSayisi	55,77
ToplamOturumSuresi	51,44
SınavOdakliCalisma	50,44
Faculty	49,78
VideoIzlemeSuresi	43,78
SayfadaKalmaOrtalama	42,9
OdevDenemeSayisi	35,79
OdeveHarcananZaman	33,35
OrtalamaOturumSuresi	32,8
Devamsizlik	30,19
Class	26,53
OgretimTuru	19,7
Gender	12,38
GezinmeSayisi_Ortalama	0

Tablo 4.11'de gösterilen sıralamaya göre "SınavNotu" özniteliği hedef niteliğin pozitif ve negatif olarak sınıflandırılmasında diğer özniteliklere göre yüksek önem derecesine sahiptir. ÖYS logları kullanılarak oluşturulan öznitelikler incelendiğinde ise "YuklenenOdev", "GecmisKonularaBakma_Toplam", "MateryalIndirmeSayisi", "ToplamOturumSuresi" ve

"SinavOdakliCalisma" özniteliklerinin önem değerlerinin 0,50 ve üstünde oldukları görülmüştür.

Son aşamada "Motivasyon" hedef niteliğinin üçlü sınıfa ayrıldığı I.SET (N=207) ile gerçekleştirilen modelleme işlemleri gerçekleştirilmiş ve bu modeller II.SET (N=200) ile test edilmiştir. Şekil 4.4'te üçlü sınıfa ayrılan "Motivasyon" özniteliğine ait frekans dağılımları görülmektedir.



Şekil 4.4: I. SET (N=207) ve II. SET (N=200)'in "Motivasyon" Hedef Niteliğine Ait Üç Sınıflı Frekans Değerleri (Yüksek ≥ 110.0, 110.0>Orta ≥ 90.0, 90.0>Düşük)

Hedef niteliği üçlü sınıfa ayrılan I. SET (N=207) ile yürütülen modelleme ve test bulguları Tablo 4.12'de; I. SET (N=200) ile elde edilen modellerin II.SET (N=207) ile test edilmesi işlemlerine ait bulgular ise Tablo 4.13'te sunulmuştur.

Tablo 4.12: Rastgele Alt Örnekleme Uygulanan Üç Sınıflı I. SET(N=207) ile Eğitilen Modellerin Performansları

		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	0,482	0,316	0,549	0,527	0,405
	k-EYK k=5	0,498	0,265	0,582	0,526	0,435
	k-EYK k=7	0,508	0,242	0,615	0,503	-
	k-EYK k=9	0,514	0,236	0,631	0,490	-
	Naive Bayes	0,441	0,355	0,371	0,648	0,322
	DVM	0,463	0,287	0,552	0,464	-
	C5,0	0,471	0,334	0,524	0,537	0,347
	CART	0,486	0,371	0,546	0,516	0,463
	Rast,Orm,	0,508	0,221	0,673	0,400	-
	Loj, Reg,	0,482	0,242	0,907	0,102	0,392

Tablo 4.13: Rastgele Alt Örnekleme Uygulanan Üç Sınıflı I. SET(N=207) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları

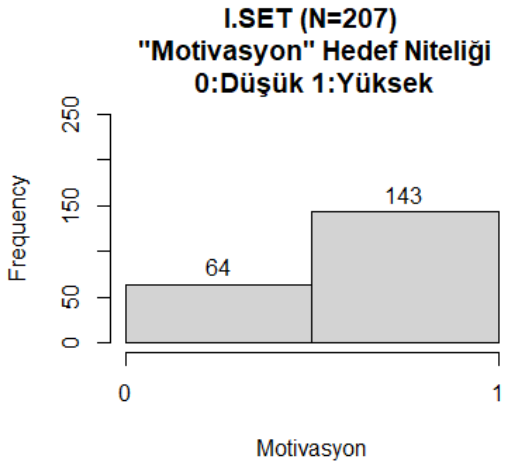
		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	0,330	0,456	0,462	0,508	0,141
	k-EYK k=5	0,373	0,340	0,553	0,472	0,127
	k-EYK k=7	0,380	0,338	0,657	0,344	0,161
	k-EYK k=9	0,399	0,263	0,730	0,243	0,191
	Naive Bayes	0,270	0,545	0,470	0,402	0,145
	DVM	0,430	0,273	0,675	0,359	0,176
	C5.0	0,335	0,409	0,651	0,325	0,150
	CART	0,355	0,318	0,771	0,171	0,184
	Rast.Orm.	0,222	0,732	0,284	0,795	0,111
	Loj. Reg.	0,350	0,273	0,771	0,137	0,171

Tablo 4.12'deki değerler incelendiğinde I.SET (N=207) ile yürütülen modelleme ve test işlemlerinde 0,44 ile 0,55 arası doğruluk değerleri görülmektedir Tablo 4.13'teki II.SET (N=200) ile yürütülen testin sonuçlarına göre ise doğruluk değeri aralığı 0,22 ile 0,39 arasındadır. Her iki tabloda da hedef niteliği üçlü sınıfa ayrılan veri setlerinin performans bulgularının düşük seviyede oldukları görülmüştür.

4.3.ÇAPRAZ GEÇERLEME YÖNTEMİYLE ELDE EDİLEN BULGULAR

Bu bölümdeki bulgular ilk aşamada I.SET (N=207)'e sırasıyla 5 ve 10 katlamalı çapraz geçerleme yöntemi doğruluk ve ROC-AUC değerleri temel alınarak modellenmiş ve test edilmiştir. Ardından ilk aşamada çapraz geçerleme işlemi sonucu oluşturulan en uygun modeller II.SET(N=200) örnekleriyle test edilmişlerdir. Bu aşamalarda kullanılacak olan I. SET (N=207) rastgele alt örnekleme için kullanılan veri setiyle aynı örnek ve öznitelik sayısına sahip olduğu için Pearson korelasyon analizi sonucunda "OturumAcmaSayisi" ve "GezinmeSayisi_Toplam" özniteliklerinin veri setinden çıkarılması işlemi aynı şekilde uygulanmıştır.

Şekil 4.5'te çapraz geçerleme işlemi için kullanılacak I.SET (N=207)'e ait pozitif ve negatif sınıfların dağılımı görülmektedir.



Şekil 4.5: Çapraz Geçerleme Uygulanan I. SET (N=207)'in "Motivasyon" Hedef Niteliğine Ait İki Sınıflı Frekans Değerleri (Pozitif Sınıf ≥ 100.0)

Şekil 4.5 incelendiğinde "Motivasyon" hedef niteliğinin 100 toplam puan değerine göre ayrılan negatif sınıf örnekleri 64 ve pozitif sınıf örnekleri ise 143 olarak sayılmaktadır.

I.SET (N=207) üzerinde gerçekleştirilen k-katlamalı çapraz geçerleme işlemleri kapsamında tüm algoritmalar için sırasıyla 5 ve 10 katlama değerleri kullanılmış olup en uygun değerin seçilmesi için Doğruluk değeri kullanılmıştır. Bu işlem için "caret" kütüphanesi içindeki "trainControl" fonksiyonunda "method= 'cv'" ve "number=5" veya "number=10" parametreleri kullanılmıştır.

Tablo 4.14: k-Katlamalı Çapraz Geçerleme Uygulanan İki Sınıflı I.SET (N=207) ile Test Edilen Algoritmalara Ait Doğruluk Değerleri

K-En Yakın Komşu Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı (k-değeri=9)	0,730
Doğruluk 10-Katlamalı (k-değeri=5)	0,711
Naive Bayes Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,701
Doğruluk 10-Katlamalı	0,724
Destek Vektör Makineleri Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,691
Doğruluk 10-Katlamalı	0,720
C5.0 Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,754
Doğruluk 10-Katlamalı	**0,739**
CART Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,681
Doğruluk 10-Katlamalı	0,710
Rastgele Orman Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,744
Doğruluk 10-Katlamalı	0,734
Lojistik Regresyon Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,744
Doğruluk 10-Katlamalı	0,738

Tablo 4.14 incelendiğinde algoritmalara ait doğruluk değerlerinin 0,69-0,75 aralığında olduğu ve en yüksek doğruluk değerinin 5-katlamalı çapraz geçerleme ile C5.0 algoritmasına ait olduğu görülmektedir.

Tablo 4.15: k-Katlamalı Çapraz Geçerleme Uygulanan İki Sınıflı I.SET (N=207) ile Test Edilen Algoritmalara Ait ROC-AUC, Duyarlılık ve Özgüllük Değerleri

K-En Yakın Komşu Algoritması	**ROC-AUC**	**Duyarlılık**	**Özgüllük**
ROC 5-Katlamalı (k değeri=7)	0,643	0,201	**0,937**
ROC 10-Katlamalı (k değeri=9)	0,646	0,183	0,909
Naive Bayes Algoritması	**ROC-AUC**	**Duyarlılık**	**Özgüllük**
ROC 5-Katlamalı	0,722	0,327	0,846
ROC 10-Katlamalı	0,732	0,326	0,854
Destek Vektör Makineleri Algoritması	**ROC-AUC**	**Duyarlılık**	**Özgüllük**
ROC 5-Katlamalı	0,731	0,391	0,861
ROC 10-Katlamalı	**0,743**	0,302	0,888
C5.0 Algoritması	**ROC-AUC**	**Duyarlılık**	**Özgüllük**
ROC 5-Katlamalı	0,679	0,365	0,853
ROC 10-Katlamalı	0,737	**0,507**	0,900
CART Algoritması	**ROC-AUC**	**Duyarlılık**	**Özgüllük**
ROC 5-Katlamalı	0,618	0,373	0,761
ROC 10-Katlamalı	0,711	**0,507**	0,818
Rastgele Orman Algoritması	**ROC-AUC**	**Duyarlılık**	**Özgüllük**
ROC 5-Katlamalı	0,725	0,422	0,895
ROC 10-Katlamalı	0,733	0,319	0,902
Lojistik Regresyon Algoritması	**ROC-AUC**	**Duyarlılık**	**Özgüllük**
ROC 5-Katlamalı	0,672	0,360	0,867
ROC 10-Katlamalı	0,703	0,436	0,883

Tablo 4.15'te I.SET (N=207)'e uygulanan 5 ve 10 katlamalı çapraz geçerleme sonrası en uygun modelin seçilmesi için ROC-AUC değerlerinin kullanıldığı modellere ait ROC-AUC, duyarlılık ve özgüllük değerleri listelenmiştir. ROC-AUC sonuçları incelendiğinde tüm algoritmalar için değer aralığının 0,61 ve 0,74 aralığında olduğu görülmektedir. Özgüllük değerlerinin tümü 0,76 ve üzerinde ölçülmüştür. Algoritmaların performansları negatif sınıf örneklerinin doğru tahmin edilme performansını belirten duyarlılık değerleri ile birlikte değerlendirildiğinde sırasıyla 0,737 ve 0,711 ROC-AUC değerlerine sahip 10-katlamalı C5.0 ve CART algoritmaları öne çıkmaktadır.

Tablo 4.16'da I.SET (N=207)'e k-katlı çapraz geçerleme uygulanarak geliştirilen en uygun modellerin II.SET (N=200) ile test edilmesiyle elde edilen bulgular yer almaktadır.

Tablo 4.16: k-Katlı Çapraz Geçerleme Uygulanan İki Sınıflı I. SET(N=207) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları

	DOĞRULUK ile 5-KATLAMALI				
	DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
k-EYK k=9	0,635	0,715	0,429	0,763	0,738
Naive Bayes	0,330	0,090	**0,946**	0,813	0,163
DVM	**0,725**	0,813	0,500	0,807	0,810
C5.0	0,715	**0,861**	0,339	0,770	**0,813**
CART	0,435	0,313	0,750	0,763	0,443
Rast.Orm.	0,405	0,222	0,875	**0,821**	0,350
Loj. Reg.	0,705	0,778	0,518	0,806	0,792
	DOĞRULUK ile 10-KATLAMALI				
	DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
k-EYK k=9	0,635	0,715	0,429	0,763	0,738
Naive Bayes	0,330	0,090	**0,946**	**0,813**	0,163
DVM	**0,725**	**0,813**	0,500	0,807	**0,810**
C5.0	0,510	0,417	0,750	0,811	0,550
CART	0,435	0,313	0,750	0,763	0,443
Rast.Orm.	0,365	0,167	0,875	0,774	0,274
Loj. Reg.	0,705	0,778	0,518	0,806	0,792
	ROC-AUC ile 5-KATLAMALI				
	DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
k-EYK k=9	0,635	0,715	0,429	0,763	0,738
Naive Bayes	0,425	0,243	**0,893**	**0,854**	0,378
DVM	0,425	0,243	**0,893**	**0,854**	0,378
C5.0	0,510	0,417	0,750	0,811	**0,550**
CART	0,435	0,313	0,750	0,763	0,443
Rast.Orm.	0,365	0,160	**0,893**	0,793	0,266
Loj. Reg.	**0,705**	**0,778**	0,518	0,806	0,792
	ROC-AUC ile 10-KATLAMALI				
	DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
k-EYK k=9	0,635	0,715	0,429	0,763	**0,738**
Naive Bayes	0,330	0,090	**0,946**	**0,813**	0,163
DVM	0,330	0,090	**0,946**	**0,813**	0,163
C5.0	0,405	0,250	0,804	0,766	0,377
CART	0,435	0,313	0,750	0,763	0,443
Rast.Orm.	0,340	0,118	0,911	0,773	0,205
Loj. Reg.	**0,705**	**0,778**	0,518	0,806	0,792

Tablo 4.16 incelendiğinde uygun modelin seçiminde doğruluk değerinin kullanıldığı 5-katlı çapraz geçerleme bulgularında Destek Vektör Makineleri, C5.0 ve Lojistik Regresyon

algoritmalarının 0,7 ve üzerinde doğruluk değerlerine sahip oldukları görülmüştür. Destek Vektör Makineleri algoritmasının performansının ROC-AUC kriteri kullanıldığında düştüğü ve C5.0 algoritmasının ise doğruluk kriteri ile 5-katlamalı çapraz geçerleme sonuçları dışında elde edilen performansının benzer şekilde düşük olduğu gözlenmiştir. Doğruluk, duyarlılık ve özgüllük değerleri göz önüne alındığında Lojistik Regresyon ve Destek Vektör Makineleri algoritmalarının diğer algoritmalara kıyasla düşük ve yüksek motivasyon sınıflarını birbirinden ayırmada daha başarılı oldukları söylenebilir.

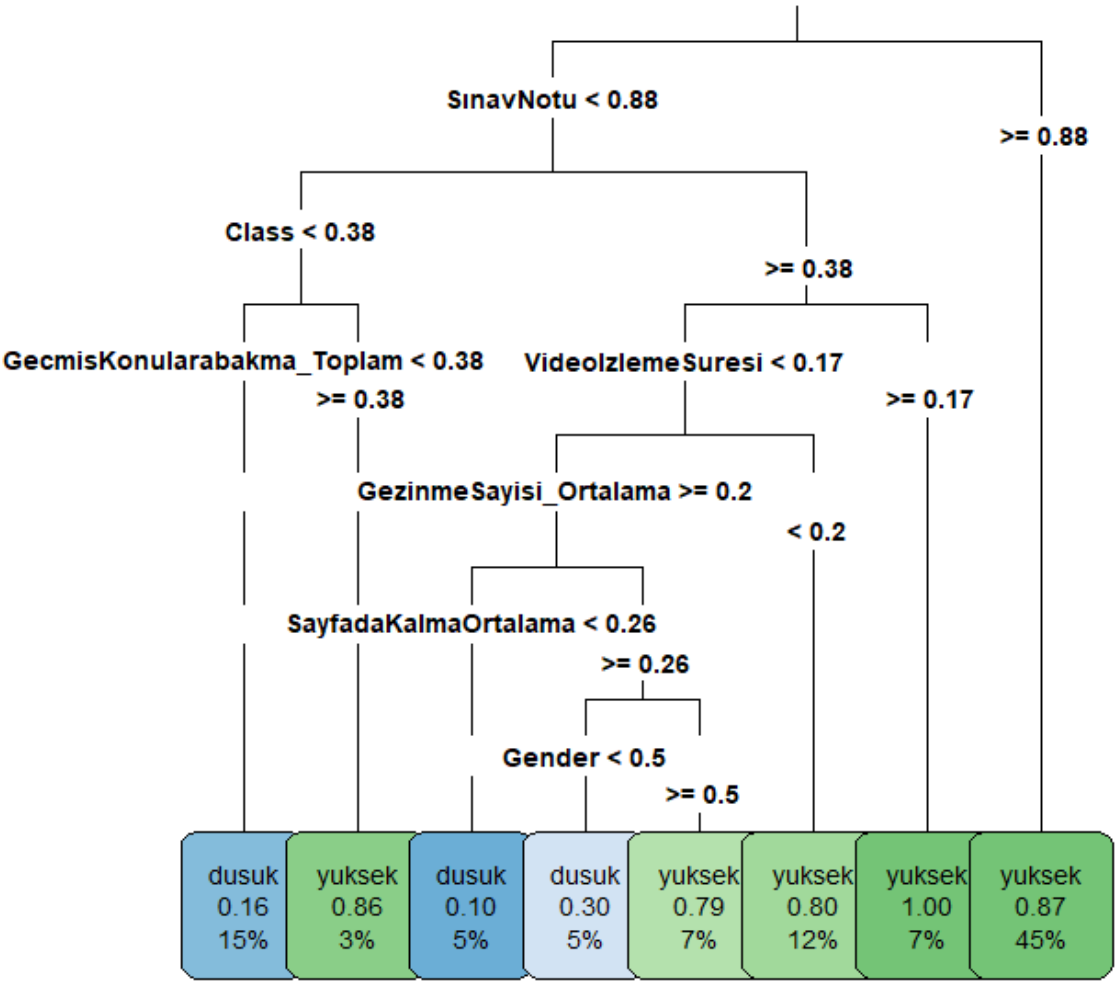
Hedef niteliğin sınıflandırılmasına etki eden özniteliklerin belirlenebilmesi amacıyla I. SET (N=207)'e ROC-AUC kriteriyle uygulanan 10-katlı çapraz geçerleme sonrasında performansları birbirlerine yakın olarak ölçülen C5.0 algoritmasına ait öznitelik önemi değerleri ve CART algoritmasına ait karar ağacı görselleştirmesi Tablo 4.17 ve Şekil 4.6'da sunulmuştur.

Tablo 4.17: ROC-AUC Kriteriyle 10-Katlı Çapraz Geçerleme Uygulanan İki Sınıflı I.SET (N=207) ile Eğitilen C5.0 Modeline Ait Öznitelik Önem Sıralaması

Öznitelik	Öznitelik Önem Değeri
GezinmeSayisi_Ortalama	100
OgretimTuru	100
ToplamOturumSuresi	100
SayfadaKalmaOrtalama	100
SınavNotu	100
Faculty	100
SınavOdakliCalisma	100
Class	100
GecmisKonularabakma_Toplam	99,52
MateryalIndirmeSayisi	98,55
YuklenenOdev	97,58
Gender	96,62
OrtalamaOturumSuresi	92,75
OdevDenemeSayisi	92,75
OdeveHarcananZaman	88,89
VideoIzlemeSuresi	77,29
Devamsizlik	74,88

Tablo 4.17 incelendiğinde C5.0 ile oluşturulan modelde "OgretimTuru", "SinavNotu" ve "Faculty" gibi özniteliklerin yanı sıra ÖYS log kayıtlarıyla oluşturulmuş olan

"ToplamOturumSuresi", "SayfadaKalmaOrtalama" ve "SinavOdakliCalisma" gibi zaman temelli özniteliklerin yüksek önem değerleri dikkat çekmektedir.



Şekil 4.6: ROC-AUC Kriteriyle 10-Katlı Çapraz Geçerleme Uygulanan İki Sınıflı I.SET (N=207) ile Eğitilen CART Karar Ağacı Görselleştirmesi

Şekil 4.6'daki CART karar ağacı görselleştirmesinde ilk düğümlerin "SinavNotu" ve "Class" özniteliklerinden ayrıştığı ve ardından karar ağacının ÖYS log kayıtları kullanılarak oluşturulmuş olan "GecmisKonularabakma_Toplam", "VideoIzlemeSuresi", "GezinmeSayisi_Ortalama", "SayfadaKalmaOrtalama" öznitelikleriyle yapılandırıldığı görülmektedir.

"Motivasyon" hedef nitelikleri üçlü sınıfa ayrılan I.SET (N=207) ve II.SET (N=200) ile yürütülen modelleme ve test işlemlerine ait sonuçlar Tablo 4.18 ve 4.19'da sunulmuştur.

Tablo 4.18: k-Katlamalı Çapraz Geçerleme Uygulanan Üç Sınıflı I.SET (N=207) ile Test Edilen Algoritmalara Ait Doğruluk Değerleri

K-En Yakın Komşu Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı (k-değeri=9)	0,541
Doğruluk 10-Katlamalı (k-değeri=5)	0,551
Naive Bayes Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,464
Doğruluk 10-Katlamalı	0,459
Destek Vektör Makineleri Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,439
Doğruluk 10-Katlamalı	0,449
C5.0 Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,512
Doğruluk 10-Katlamalı	0,513
CART Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,517
Doğruluk 10-Katlamalı	0,474
Rastgele Orman Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,541
Doğruluk 10-Katlamalı	0,498
Lojistik Regresyon Algoritması	**Doğruluk**
Doğruluk 5-Katlamalı	0,482
Doğruluk 10-Katlamalı	0,491

Tablo 4.18'deki hedef niteliği üç sınıfa ayrılan I.SET(N=207) ile yürütülen 5 ve 10 katlı çapraz geçerleme bulgularına bakıldığında doğruluk değerlerinin 0,43 ile 0,54 arasında olduğu ve tüm algoritmaların düşük performans sergilediği görülmektedir.

Tablo 4.19: k-Katlı Çapraz Geçerleme Uygulanan Üç Sınıflı I. SET(N=207) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları

	DOĞRULUK ile 5-KATLAMALI				
	DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
k-EYK k=7	0,385	0,364	0,651	0,359	0,167
Naive Bayes	0,240	0,727	0,373	0,538	0,142
DVM	0,415	0,273	0,675	0,359	0,154
C5.0	0,265	0,636	0,470	0,538	0,131
CART	0,355	0,318	0,771	0,171	0,184
Rast.Orm.	0,240	0,818	0,265	0,855	0,122
Loj. Reg.	0,395	0,273	0,723	0,256	0,171
	DOĞRULUK ile 10-KATLAMALI				
	DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
k-EYK k=7	0,365	0,318	0,639	0,333	0,146
Naive Bayes	0,270	0,545	0,470	0,402	0,145
DVM	0,415	0,273	0,675	0,359	0,154
C5.0	0,290	0,636	0,301	0,718	0,131
CART	0,355	0,318	0,771	0,171	0,184
Rast.Orm.	0,220	0,727	0,265	0,795	0,110
Loj. Reg.	0,395	0,273	0,723	0,256	0,171

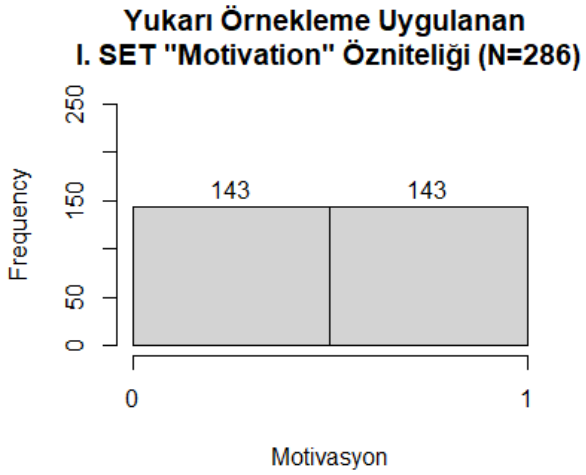
Tablo 4.19 incelendiğinde hedef niteliği üç sınıfa ayrılan I. SET (N=207) ile eğitilen modellerin II.SET (N=200) örnekleri ile yürütülen teste ait sonuçların rastgele alt örnekleme yöntemiyle elde edilen sonuçlara paralel olarak doğruluk değerleri 0,22 ile 0,41 arasında olacak şekilde düşük seviyede olduğu görülmüştür.

4.4.HOLD-OUT YÖNTEMİYLE ELDE EDİLEN BULGULAR

Bu bölüm içerisinde I.SET(N=207) toplam örnek sayısı farklı yukarı örnekleme yöntemleriyle arttırılarak modellenmiş ve önce hold-out yöntemiyle ardından ise II.SET (N=200) kullanılarak test edilmiştir.

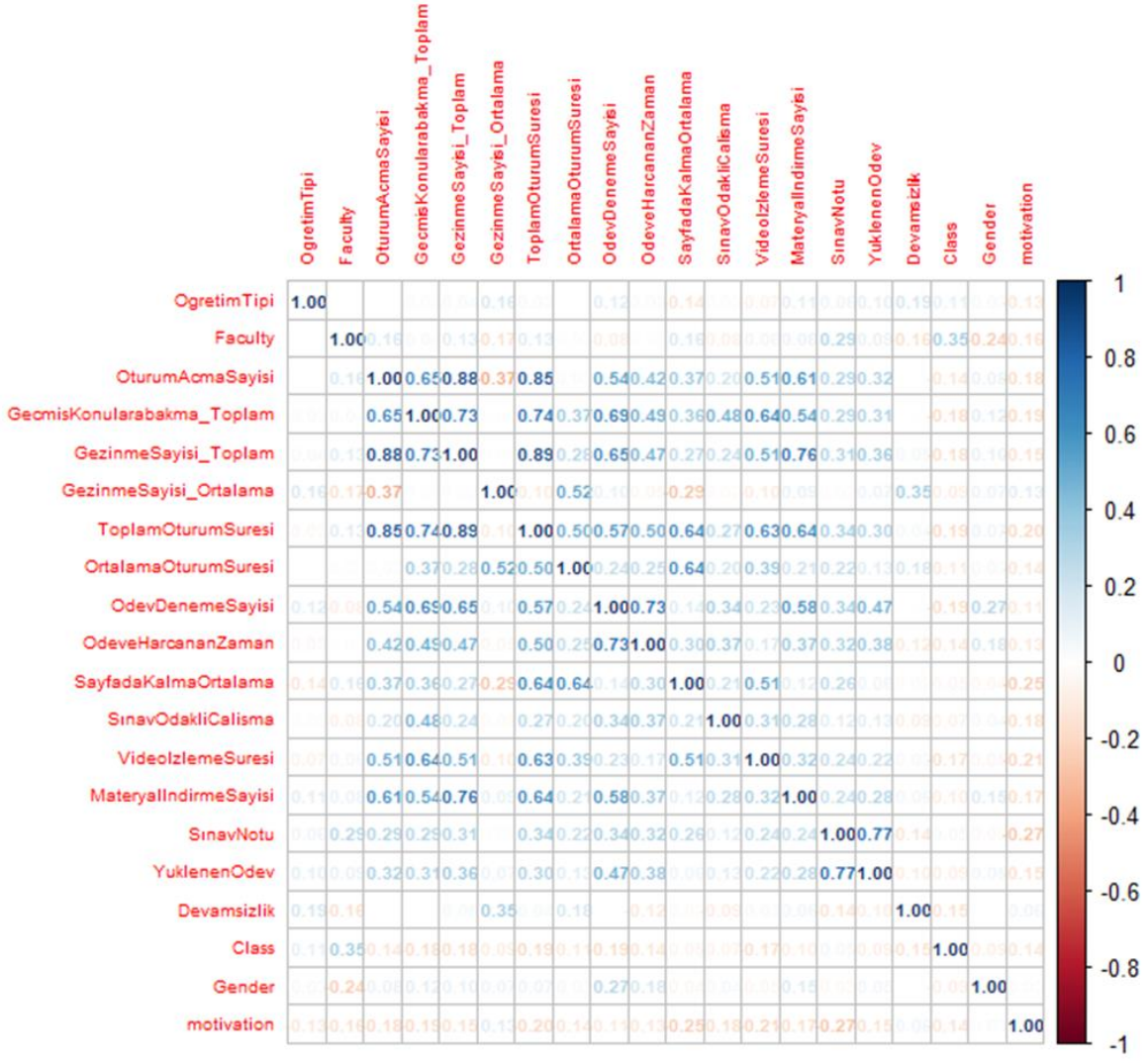
4.4.1. Yukarı Örnekleme İle Elde Edilen Bulgular

Bu bölümde 207 örnekli I.SET'te dengesiz dağılan sınıf etiketleri yukarı örnekleme yöntemiyle dengeli hale getirilmiştir ve hold-out yöntemi ile eğitilen ve test edilen modellere ait bulgular sunulmuştur. Yukarı örnekleme işlemi R kütüphanesi içerisinde bulunan ve "ROSE" paketinde yer alan "ovun.sample" fonksiyonu ile gerçekleştirilmiştir. Yukarı örnekleme işlemi sonrası oluşan pozitif ve negatif sınıfa ait frekans değerleri Şekil 4.7'de görülmektedir.



Şekil 4.7: Yukarı Örnekleme Uygulanan I. SET (N=286)'in "Motivasyon" Hedef Niteliğine Ait Frekans Değerleri (Pozitif Sınıf ≥ 4,0)

Modelleme işlemlerinden önce öznitelikler arası korelasyon katsayılarını incelemek amacıyla yürütülen Pearson Korelasyon Testi sonuçları Şekil 4.8'de sunulmuştur.



Şekil 4.8: Yukarı Örnekleme Uygulanan I. SET (N=286) Özniteliklerine Ait Pearson Korelasyon Testi Sonuçları

Şekil 4.8'deki korelasyon analizi bulguları göz önüne alınarak 0,80 ve üzeri pozitif yönlü ilişki gösteren "OturumAcmaSayisi" ve "GezinmeSayisi_Toplam" özniteliklerinin veri setinden çıkarılmasına ve veri setinin 17 öznitelik ve 1 hedef nitelik ile kullanılmasına karar verilmiştir.

Yukarı örnekleme işlemiyle dengeli hale getirilen I.SET(N=286)'in %70'lik bölümü modelin eğitimi amacıyla, %30'luk bölümü ise test amacıyla rastgele olarak ayrılmıştır. Bu analize dair sonuçlar Tablo 4.20'de gösterilmektedir.

Tablo 4.20: Yukarı Örnekleme Uygulanmış I. SET (N=286) ile Eğitilen Modellerin Hold-out ile Test Edilen (%70-%30) Performansları

		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	0,607	0,476	0,738	0,645	0,548
	k-EYK k=5	0,548	0,429	0,667	0,563	0,486
	k-EYK k=7	0,607	0,500	0,714	0,636	0,560
	k-EYK k=9	0,560	0,500	0,619	0,568	0,532
	Naive Bayes	0,548	0,476	0,619	0,556	0,513
	DVM	0,560	0,619	0,500	0,553	0,584
	C5.0	0,738	**0,643**	0,833	0,794	0,711
	CART	0,619	0,476	0,762	0,667	0,556
	Rast.Orm.	**0,762**	0,643	**0,881**	**0,844**	**0,730**
	Loj. Reg.	0,488	0,476	0,500	0,488	0,482

Tablo 4.20'de gösterilmekte olan bulgulara bakıldığında iki karar ağacı algoritması olan C5.0 ve Rastgele Orman algoritmalarının 0,70 ve üzeri doğruluk değerine sahip oldukları gözlenmektedir. Geri kalan algoritmaların performans değerlerinin düşük olduğu söylenebilir.

Bu aşamanın ardından I.SET (N=286)'in tüm örnekleri kullanılarak eğitilen modeller II.SET (N=200) ile test edilmiştir. Bu analize dair sonuçlar Tablo 4.21'de sunulmaktadır.

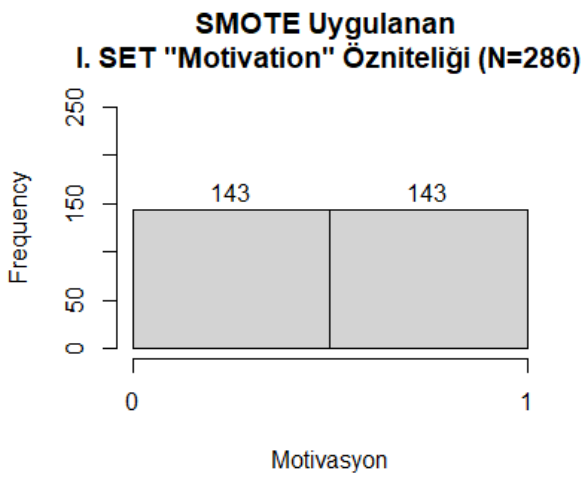
Tablo 4.21: Yukarı Örnekleme Uygulanmış I.SET (N=286) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları

		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	0,400	0,285	0,696	0,707	0,406
	k-EYK k=5	0,365	0,229	0,714	0,673	0,342
	k-EYK k=7	0,395	0,236	0,804	0,756	0,360
	k-EYK k=9	0,440	0,313	0,768	0,776	0,446
	Naive Bayes	0,340	0,104	**0,946**	0,833	0,185
	DVM	0,455	0,292	0,875	**0,857**	0,435
	C5,0	0,360	0,167	0,857	0,750	0,273
	CART	0,435	0,271	0,857	0,830	0,408
	Rast,Orm,	0,345	0,125	0,911	0,783	0,216
	Loj, Reg,	**0,610**	**0,813**	0,089	0,696	**0,750**

Tablo 4.21 incelendiğinde tüm algoritmaların performanslarının düşük seviyede olduğu görülmüştür.

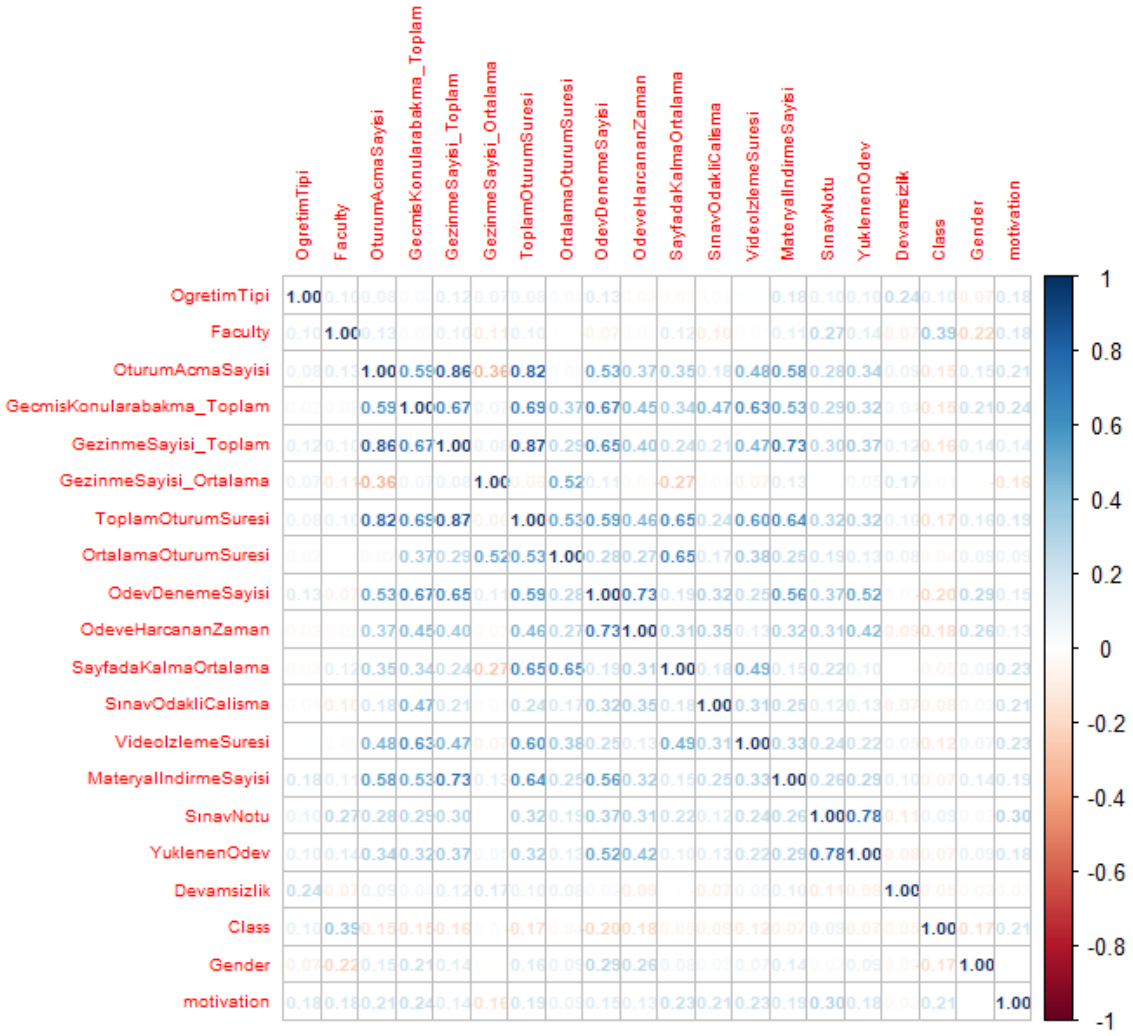
4.4.2. SMOTE Yöntemi İle Elde Edilen Bulgular

Bu bölümdeki modelleme çalışmaları I.SET (N=207)'in pozitif ve negatif sınıflar arasındaki dengesizliği SMOTE yöntemiyle giderilerek oluşturulan veri seti üzerinde yürütülmüştür. SMOTE yönteminin uygulanması için R paketleri içerisinde yer alan "themis" kütüphanesinin "smote" fonksiyonu kullanılmıştır. Şekil 4.9'da SMOTE uygulaması sonrasında oluşan pozitif ve negatif sınıfa ait frekans değerleri görülmektedir.



Şekil 4.9: SMOTE Uygulanan I. SET (N=286)'in "Motivasyon" Hedef Niteliğine Ait Frekans Değerleri (Pozitif Sınıf ≥ 4,0)

SMOTE uygulaması sonrası oluşan 286 örnekli veri setinin öznitelikleri arasındaki korelasyon katsayılarının incelenmesi amacıyla Pearson Korelasyon Testi sonuçları Şekil 4.9'da sunulmuştur.



Şekil 4.10: SMOTE Uygulanan I. SET (N=286) Özniteliklerine Ait Pearson Korelasyon Testi Sonuçları

Şekil 4.10'daki korelasyon analizi bulguları göz önüne alınarak önceki bölümlerde elde edilen sonuçlara benzer şekilde 0,80 ve üzeri pozitif yönlü ilişki gösteren "OturumAcmaSayisi" ve "GezinmeSayisi_Toplam" özniteliklerinin veri setinden çıkarılmasına ve böylelikle veri setinin 17 öznitelik ve 1 hedef nitelik ile kullanılmasına karar verilmiştir.

SMOTE uygulanmış I.SET (N=286) kullanılarak hold-out yöntemiyle eğitilen ve test edilen modellere ait performans bulguları Tablo 4.22'de sunulmuştur.

Tablo 4.22: SMOTE Uygulanmış I. SET (N=286) ile Eğitilen Modellerin Hold-out ile Test Edilen (%70-%30) Performansları

		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	**0,786**	0,714	**0,857**	**0,833**	0,769
	k-EYK k=5	0,738	0,690	0,786	0,763	0,725
	k-EYK k=7	0,726	0,690	0,762	0,744	0,716
	k-EYK k=9	0,714	0,667	0,762	0,737	0,700
	Naive Bayes	0,738	0,690	0,786	0,763	0,725
	DVM	0,714	**0,762**	0,667	0,696	0,727
	C5.0	0,667	0,738	0,595	0,646	0,689
	CART	0,702	0,643	0,762	0,730	0,684
	Rast.Orm.	0,774	0,762	0,786	0,780	**0,771**
	Loj. Reg.	0,714	0,738	0,690	0,705	0,721

Tablo 4.22'deki bulgulara bakıldığında C5.0 algoritması hariç tüm algoritmaların doğruluk değerlerinin 0,7 ve üzerinde olduğu görülmektedir. En yüksek doğruluk değeri k değerinin 3 olarak alındığı k-en yakın komşu algoritmasıyla 0,786 olarak ölçülmüştür. Bu algoritmaya en yakın yüksek performans ise 0,774 doğruluk değeri ile Rastgele Orman algoritmasına aittir.

SMOTE uygulanmış I.SET (N=286)'in tüm örnekleriyle geliştirilen modeller II.SET (N=200) kullanılarak test edilmiş ve bu işleme dair bulgular Tablo 4.23'te gösterilmiştir.

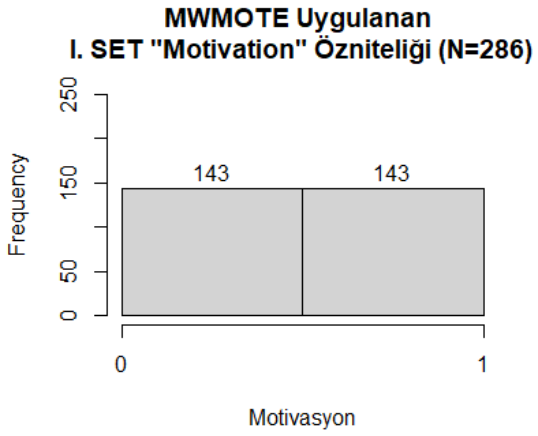
Tablo 4.23: SMOTE Uygulanmış I.SET (N=286) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları

		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	0,480	0,354	0,804	0,823	0,495
	k-EYK k=5	0,460	0,340	0,768	0,790	0,476
	k-EYK k=7	0,445	0,292	0,839	0,824	0,431
	k-EYK k=9	0,435	0,271	0,857	0,830	0,408
	Naive Bayes	0,425	0,250	**0,875**	0,837	0,385
	DVM	0,530	0,417	0,821	**0,857**	0,561
	C5,0	0,430	0,299	0,768	0,768	0,430
	CART	**0,605**	**0,688**	0,393	0,744	**0,715**
	Rast,Orm,	0,415	0,236	**0,875**	0,829	0,368
	Loj, Reg,	0,530	0,438	0,768	0,829	0,573

Tablo 4.23 incelendiğinde bir önceki aşamada modellenen algoritmaların II. SET(N=200) örnekleri üzerinde 0,415 ve 0,605 aralığında doğruluk değerleriyle düşük sınıflandırma performansına sahip oldukları görülmüştür.

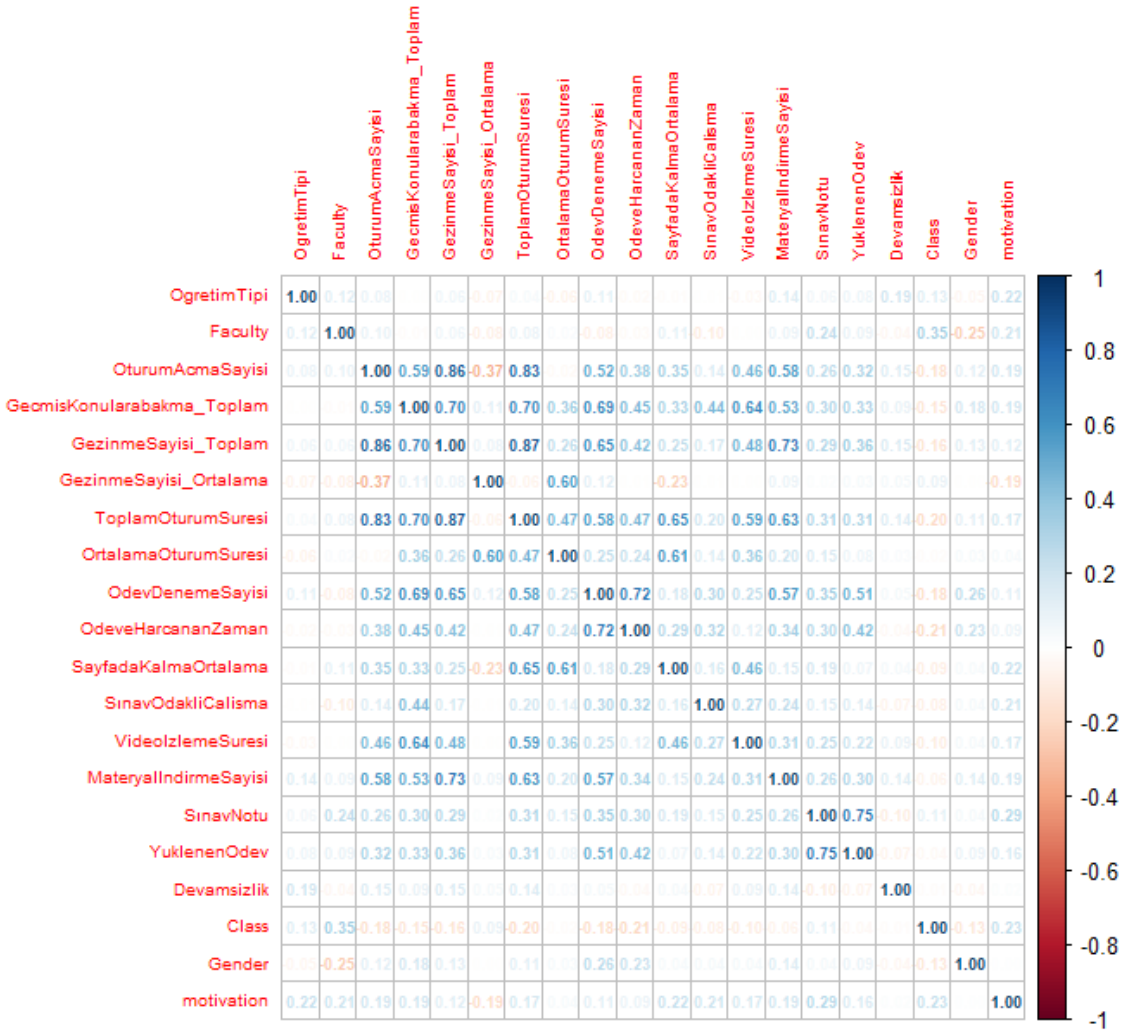
4.4.3. MWMOTE Yöntemi İle Elde Edilen Bulgular

Bu bölümde I.SET (N=207)'in hedef niteliğine ait pozitif ve negatif sınıf örnekleri arasındaki dengesizlik MWMOTE yöntemiyle giderilerek veri setinin örnek sayısı 286'ya çıkarılmıştır ve modelleme ve test işlemleri gerçekleştirilmiştir. MWMOTE uygulaması için R paketleri içerisindeki "imbalance" kütüphanesi ile "mwmote" fonksiyonu kullanılmıştır. Bu uygulamanın ardından dengeli hale getirilen I.SET (N=286)'in pozitif ve negatif sınıflarına ait frekans değerleri Şekil 4.11'de görülmektedir.



Şekil 4.11: MWMOTE Uygulanan I. SET (N=286)'in "Motivasyon" Hedef Niteliğine Ait Frekans Değerleri

MWMOTE yöntemiyle pozitif ve negatif sınıf örnekleri eşitlenen I.SET (N=286) ile yürütülen Pearson Korelasyon analizi sonuçları Şekil 4.12'de görülmektedir.



Şekil 4.12: MWMOTE Uygulanan I. SET (N=286) Özniteliklerine Ait Pearson Korelasyon Testi Sonuçları

Şekil 4.12'deki korelasyon katsayıları incelendiğinde "OturumAcmaSayisi", "GezinmeSayisi_Toplam" ve "ToplamOturumSuresi" özniteliklerini birbirleriyle 0,80 ve üzerinde pozitif yönlü korelasyon gösterdikleri görülmektedir. Bu sebeple "OturumAcmaSayisi" ve "GezinmeSayisi_Toplam" özniteliklerinin veri setinden çıkarılması kararlaştırılmıştır.

MWMOTE ile elde edilen I.SET (N=286) üzerinde eğitim ve test kümeleri %70'e %30 olarak bölünerek uygulanan hold-out yöntemine ait bulgular Tablo 4.24'te yer almaktadır.

Tablo 4.24: MWMOTE Uygulanmış I. SET (N=286) ile Eğitilen Modellerin Hold-out ile Test Edilen (%70-%30) Performansları

		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	0,762	**0,786**	0,738	0,750	0,767
	k-EYK k=5	0,762	0,738	0,786	0,775	0,756
	k-EYK k=7	**0,798**	0,762	**0,833**	**0,821**	**0,790**
	k-EYK k=9	0,750	0,714	0,786	0,769	0,741
	Naive Bayes	**0,798**	**0,786**	0,810	0,805	0,795
	DVM	0,762	0,762	0,762	0,762	0,762
	C5.0	0,667	0,643	0,690	0,675	0,659
	CART	0,667	0,643	0,690	0,675	0,659
	Rast.Orm.	**0,798**	0,762	**0,833**	**0,821**	**0,790**
	Loj. Reg.	0,750	0,738	0,762	0,756	0,747

Tablo 4.24 incelendiğinde tüm algoritmaların doğruluk değerlerinin 0,66 ile 0,79 aralığında olduğu görülmektedir. 0,798 olarak ölçülen en yüksek doğruluk değerini Naive Bayes, Rastgele Orman ve k değeri 7 olarak belirlenen k-en yakın komşu algoritmaları paylaşmaktadır. Ayrıca doğruluk, özgüllük ve duyarlılık değerleri birlikte düşünüldüğünde bu algoritmaların birbirleriyle yakın performans sergiledikleri söylenebilir.

Bu aşamanın ardından MWMOTE uygulanmış I.SET (N=286)'in örnekleri kullanılarak geliştirilen modeller 200 örnekli II.SET ile test edilmiştir. Bu analize dair sonuçlar Tablo 4.25'te sunulmuştur.

Tablo 4.25: MWMOTE Uygulanmış I.SET (N=286) ile Eğitilen Modellerin II.SET (N=200) ile Test Performansları

		Performans Değerlendirme Ölçütleri				
		DOĞRULUK	**DUYARLILIK**	**ÖZGÜLLÜK**	**KESİNLİK**	**F-SKORU**
Algoritmalar	**k-EYK k=3**	0,495	0,410	0,714	0,787	0,539
	k-EYK k=5	0,440	0,347	0,679	0,735	0,472
	k-EYK k=7	0,460	0,361	0,714	0,765	0,491
	k-EYK k=9	0,475	0,403	0,661	0,753	0,525
	Naive Bayes	0,425	0,229	0,929	0,892	0,365
	DVM	0,510	0,361	0,893	0,897	0,515
	C5,0	0,460	0,333	0,786	0,800	0,471
	CART	0,430	0,438	0,411	0,656	0,525
	Rast,Orm,	0,410	0,250	0,821	0,783	0,379
	Loj, Reg,	0,485	0,326	0,893	0,887	0,477

SMOTE ile elde edilen sonuçlara benzer şekilde Tablo 4.25'te görülen 0,51 ve altındaki doğruluk değerleri tüm sınıflandırma algoritmalarının düşük seviyeli performans sergilediklerini göstermektedir.

5. TARTIŞMA VE SONUÇ

2020 yılından itibaren yaklaşık 2 yıl boyunca tüm dünyayı etkileyen COVID-19 pandemisi sebebiyle yükseköğretim seviyesi dâhil olmak üzere tüm öğretim faaliyetleri uzaktan ve hibrit yöntemlerle yürütülmüştür. Pandemi sırasında eğitimin sürdürülmesi için zorunlu bir yöntem olan uzaktan eğitimin gelecekte büyük nüfusları etkileyebilecek benzer kriz dönemlerinin ortaya çıkması durumunda da uygulanma ihtimalinin olduğu öngörülebilir. Ayrıca bu süre içerisindeki çevrimiçi öğrenme deneyimlerinin pandemi sonrası eğitim-öğretim faaliyetlerinde kalıcı izler bıraktığı ve gelecekte çevrimiçi öğrenme platformlarının yüz yüze eğitim sürecinin destekçisi olarak kullanılmasının giderek önem kazandığı vurgulanmaktadır (Rapanta ve diğ., 2021). Bundan dolayı günümüzde çevrimiçi öğrenme ortamlarındaki öğrenci davranışlarının başta yapay zekâ olmak üzere hesaplamalı yöntemler kullanılarak çözümlenmesi ve bu yöntemler sayesinde yüz yüze eğitime kıyasla uzaktan eğitim ortamlarında gözlemlenmesi zor olan faktörlerin açıklanması da aynı derecede önem kazanmıştır.

Bu tez çalışmasında üniversite öğrencilerinin bir çevrimiçi öğrenme ortamı olan Moodle Öğrenme Yönetim Sistemi (ÖYS) içerisindeki log adı verilen etkileşim verisi kullanılarak derse dönük motivasyon düzeylerinin tahmin edilmesi amaçlanmıştır. Çalışma kapsamında katılımcıların ÖYS kullanım verisi ile derse yönelik motivasyon düzeyleri arasında yapay zekâ yöntemleriyle ortaya çıkarılması amaçlanan ilişki incelenmiş olup katılımcıların ÖYS kullanım verisi ile elde edilen veri seti özniteliklerinden hangilerinin derse yönelik motivasyonu etkilediği sorusuna da cevap aranmıştır.

Çalışmanın bulgularında genel olarak ÖYS log dosyalarıyla oluşturulan ve öğrencilerin sistem içerisinde sergiledikleri davranışları niteleyen özniteliklerden oluşan veri ile kayıtlı oldukları derse yönelik motivasyon düzeylerinin tahmin edilmesinin mümkün olduğu görülmektedir. Birbirine yakın performans bulguları gözlenen rastgele alt örnekleme ve çapraz geçerleme yöntemleriyle ikili sınıflandırma için %70 civarında genel sınıflandırma başarısı ve %50 civarında ise düşük motivasyon düzeyindeki öğrencileri tahmin etme başarısı görülmekte iken Yukarı Örnekleme, SMOTE ve MWMOTE yöntemleriyle elde edilen genel sınıflandırma başarısının %80'lere yaklaştığı ve düşük motivasyonlu öğrencilerin tahmin edilme başarısının ise %80'in üzerine çıkabildiği gözlenmiştir. Ancak çalışma kapsamında sentetik veri kullanılmadan çapraz geçerleme ile oluşturulan ve test edilen modellerin görece daha düşük performanslarına rağmen daha dengeli ve doğru sonuçlar verdiği düşünülmektedir. Figueroa-

Cañas ve Sancho-Vinuesa (2020), %70 civarında ölçtükleri doğruluk değerlerini düşük olarak değerlendirmelerine rağmen eğitmenlerin dersi bırakma riski mevcut olarak sınıflanan öğrencilere müdahale etme uyarısını vermek için yeterli olduğunu belirtmişlerdir. Benzer olarak Baker ve diğerleri (2015) de bir çevrimiçi öğrenme ortamından elde ettikleri aktivite verisi ile dersi bırakma davranışını tahmin etmeyi amaçladıkları çalışmalarında %65-%70 arası sınıflandırma başarısı ile bir teşhis aracı geliştirilemeyeceğini ancak bu düzeydeki bulguların bir uyarı sistemi için uygun olduğunu vurgulamışlardır. Diğer yandan De Oliveira ve diğerleri (2019), dersi bırakma eğilimini genel sınıflandırma başarısını %84 olarak ölçtükleri çalışmalarında risk altındaki öğrencilerin doğru tahmin edilme düzeylerindeki göreli düşüklüğün dengesiz veri setinden kaynaklı olduğunu ve yukarı veya aşağı örnekleme yöntemleriyle bu sorunun giderilebileceği önerisinde bulunmuşlardır. Literatürdeki benzer çalışmalara ait bulgular göz önüne alındığında bu araştırmanın bulgularının literatürdeki bulgulara kıyasla ortalama sınıflandırma başarısına sahip olduğu ve özellikle düşük motivasyon sınıfının tahmin edilme başarısının önerildiği gibi yukarı örnekleme yöntemleriyle arttırılabildiği görülmüştür.

Veri analizlerinin ikinci aşamasını oluşturan validasyon seti ile elde edilen sonuçlar göz önüne alındığında rastgele alt örnekleme ve çapraz geçerleme bulgularında birinci aşamadaki performans değerlerine benzer olarak ikili sınıflandırmada genel başarının %70'ler seviyesinde ve düşük motivasyon düzeyindeki öğrencileri doğru tahmin etme başarısının ise %50'ler seviyesinde olduğu saptanmıştır. Ancak yukarı örnekleme, SMOTE ve MWMOTE yöntemleriyle elde edilen ikinci aşama sonuçlarının birinci aşama sonuçlarına kıyasla neredeyse yarı yarıya düştükleri gözlenmiştir. Bu durumun özellikle negatif sınıfa ait türetilmiş veya sentetik örneklerle eğitilen modellerin gerçek veriye ait örüntüleri tanımada uyuşmazlık göstermesinden kaynaklandığı düşünülmektedir. Öte yandan validasyon için kullanılan veri setinin Moodle ÖYS ile geçirilen zaman ve etkileşim yoğunluğu açısından kısıtlı bir çerçevede kullanan öğrencilerden elde edildiği not edilmelidir. Bu sebeple birinci aşamada eğitilen modellerin aktivite çeşidi ve yoğunluğu fazla veri ile test edilmesi durumunda performanslarının artacağı öngörülmektedir. Son olarak, "Motivasyon" hedef niteliğinin üçlü sınıfa ayrılarak test edildiği bulgularda görülen düşük performansın ise toplam örnek sayısının az olmasıyla ilişkili olduğu düşünülmektedir.

Çalışmanın bir diğer öne çıkan sonucu ise birinci aşamada test edilen algoritmalar arasında en yüksek performansın çoğunlukla karar ağacı algoritmalarına ait olmasıdır. Rastgele alt örnekleme bulgularında Rastgele Orman algoritmasının ikinci en yüksek performans gösteren algoritma olduğu görülmüş olup çapraz geçerleme bulgularında ise C5.0 ve CART algoritmalarının en yüksek performansa sahip algoritmalar oldukları gözlenmiştir. Ayrıca yukarı örnekleme, SMOTE ve MWMOTE bulgularının tümünde Rastgele Orman algoritması birinci veya ikinci en yüksek performansa sahip algoritma olarak gözükmektedir. ÖYS veya benzeri çevrimiçi öğrenme ortamlarından elde edilen log verisinin eğitimle ilgili değişkenlerin tahmin edilmesi amacıyla kullanıldığı çalışmalarda karar ağaçları algoritmalarının yüksek performans sergilemesi rastlanılır bir bulgudur. De Oliveira ve diğerleri (2019) tarafından yürütülen çalışmada öğrencilerin dersi bırakma eğilimleri Moodle ÖYS log kayıtlarını kullanarak tahmin edilmeye çalışılmış ve *Extra Trees* algoritmasının en yüksek performansı sergilediği görülmüştür. Benzer şekilde Liyanage ve diğerleri (2016), Moodle ÖYS logları ile oluşturulan veri seti ile öğrencilerin öğrenme stillerini sınıflandırmak amacıyla yürüttükleri çalışmada en yüksek performans J48 karar ağacına aittir ve bu algoritmaya en yakın performans yine bir karar ağacı algoritması olan Rastgele Orman tarafından sergilenmiştir. Diğer yandan Romero ve diğerleri (2010), ÖYS log dosyalarının kullanıldığı makine öğrenmesi çalışmalarında sonuçların eğitimciler tarafından anlaşılabilir olması ve hedef niteliğe etki eden özniteliklerin görselleştirilerek incelebilmesinin mümkün olması sebebiyle karar ağaçları ile oluşturulan modellerin uygulamalarda kullanılmasını önermişlerdir. Benzeri çalışmalarda da görüldüğü üzere karar ağaçları algoritmalarının ÖYS log verisiyle gösterdiği yüksek performansın yanında karar verme sürecine katkı sağlayacak şekilde yorumlanabilmeleri açısından avantajlı oldukları görülmektedir.

Çalışma kapsamında rastgele alt örnekleme yöntemiyle elde edilen Naive Bayes ve çapraz geçerleme ile elde edilen C5.0 algoritmalarıyla elde edilen modellere ait öznitelik önemleri hangi özniteliklerin motivasyon sınıflandırmasında belirleyici olduğunu görmek amacıyla raporlanmıştır. Her iki bulguda da beklenen bir sonuç olarak "SinavNotu" özniteliğinin sınıflandırmada yüksek belirleyiciliğe sahip olduğu görülmüştür. Bir işi başarma yönelimi olarak tanımlanan motivasyon, öğrenme süreci ve öğrenme ortamı konusu olduğunda akademik başarıyla ilişkili ve onu açıklayan bir faktör olarak değerlendirilmektedir (Pintrich & De Groot, 1990; Mega ve diğ., 2014). Bu doğrultuda örneğin Kularbphettong ve diğerleri (2012) tarafından yürütülen çalışmada, bu tez araştırmasının amacıyla benzer şekilde Moodle ÖYS log

dosyalarını kullanarak öğrencilerin motivasyon düzeylerinin tahmin edilmesi amaçlanmış olup motivasyon hedef niteliği sınıflarının belirlenmesi için akademik başarı ve motivasyon arasındaki ilişkiye dayanarak doğrudan ilgili dersin sınav puanları kullanılmıştır. Ancak Figueroa-Cañas ve Sancho-Vinuesa (2020), elde ettikleri sınıflandırma modellerinde dersi bırakma eğilimine en fazla katkısı olan etmenin devamlı öğrenme davranışı olduğunu ve derslerin zorluk ve kolaylık seviyelerinin değişkenlik göstermesi sebebiyle benzer analizlerde sınav notunun belirleyiciliğine güvenmemek gerektiğini belirtmişlerdir. "SinavNotu" haricinde öznitelik önem sıralamalarında yüksek belirleyicilik gösteren ve ÖYS log dosyaları kullanılarak oluşturulmuş özniteliklere bakıldığında ise hem Naive Bayes hem de C5.0 modellerinde "ToplamOturumSuresi" ve "SinavOdakliCalisma" özniteliklerinin yer aldığı görülmektedir. Ayrıca Şekil 4.6'da gösterilen CART karar ağacı görselinde Naive Bayes modeline ait öznitelik önem tablosunda da yüksek değere sahip olduğu görülen "GecmisKonularabakma_Toplam" ve C5.0 modelinde de yüksek öznitelik önem değerine sahip "SayfadaKalmaOrtalama" öznitelikleri yer almaktadır. Bu bağlamda zaman veya süre temelli değişkenlerin sıklık veya miktar belirten değişkenlere göre birden fazla algoritmanın öznitelik önemi sıralamalarında yer almalarının sebebinin süre değerlerinin miktarlara kıyasla öğrencinin ders ile ilgili bir aktiviteye olan ilgisini daha iyi açıklaması olduğu düşünülmektedir. Diğer bir deyişle dersin etkinlik rutini gereği ÖYS içerisinde gerçekleştirdiği sayfa ziyareti sayısı yüksek olan öğrenci ile belirli bir hedef yönelimi çerçevesinde bu hedefi gerçekleştirme amacıyla ihtiyaç duyduğu konu ve materyalleri içeren daha az sayıda sayfa ziyaret eden fakat her bir sayfada belirli bir zaman harcayan öğrencinin derse olan ilgilerinin birbirinden farklı olması beklenir. Bu çalışmanın bulgularında öznitelik önem sıralamalarındaki süre türünden özniteliklerin görülmesi bu türdeki değişkenlerin öğrencilerin derse yönelik motivasyon düzeylerini sınıflandırmada daha nitelikli bir ayrım olanağı sunduğunu göstermektedir. Ek olarak, Romero ve diğerleri (2010), düşük performans gösteren modellerin iyileştirilmesi için "çevrimiçi" verinin yanında öğrencilere ait "çevrimdışı" verinin de toplanarak veri setindeki özniteliklerin artırılması çözümünü önermiştir. Diğer modellere kıyasla yüksek performans gösteren CART modeline ait karar ağacı görseli ve C5.0 modeline ait öznitelik önem değeri sıralamasında ÖYS log dosyaları kullanılarak oluşturulan özniteliklerin yanında öğrencilere ait sınıf, cinsiyet ve fakülte gibi özniteliklerin de görülmüş olması bu öneriyi doğrular niteliktedir.

Özetle, bu çalışma üniversite öğrencilerinin Moodle ÖYS ile etkileşim verisini kullanarak derse yönelik motivasyon düzeylerinin yapay zekâ yöntemleri kullanılarak tahmin edilebileceğini ve

ÖYS kullanım verisi ile derse dönük motivasyon arasındaki ilişkinin yapay zekâ yöntemleri ile modellenebileceğini göstermiştir. Çalışmanın sonuçlarının uzaktan veya hibrit olarak ÖYS üzerinden yürütülen derslerin eğitmenlerine ve ÖYS kullanılan eğitim kurumlarına derse yönelik yüksek ve düşük motivasyona sahip öğrenci gruplarını saptamak ve bu düşüklük veya yüksekliğin kaynaklarını inceleyerek öğrenme ortamı, öğretim materyalleri ve öğretim programları hakkında verilecek kararlara yol gösterici olabileceği düşünülmektedir.

Gelecekte aynı doğrultuda yapılacak çalışmaların daha fazla katılımcıyla yürütülmesi önerilmektedir. Katılımcı sayısı arttıkça ÖYS kullanım verisinin çeşitlilik ve yoğunluk kazanacağı ve eğitilen modellerin derse yönelik motivasyon düzeyini saptamada daha yüksek performans sergileyebileceği ve hatta ikili sınıflandırma yerine çok sınıflı analizler gerçekleştirilebileceği öngörülmektedir. Ayrıca öğrencilerin ÖYS kullanım sıklığını arttıracak içerik ve araçların sunulması da kullanım verisine çeşitlilik kazandırarak veri setinin daha fazla öznitelikle oluşturulmasını sağlayabilir. Bu noktada Moodle ÖYS içerisindeki forumların bahsi geçen konuya katkısı biliniyor olsa da (Örn. Romero ve diğ., 2010; Romero ve diğ., 2013; De Oliveira ve diğ., 2019) mevcut çalışmada katılımcıların forum kullanımına teşvik edilmelerine rağmen son yıllarda sosyal medyanın yaygınlaşması ve buna bağlı olarak forum kullanımının azalmasından kaynaklanan alışkanlıklar sebebiyle bu girişimin gerçekleştirilemediği not edilmelidir. Ancak katılımcıların ders ile ilgili sorularına cevap aramak amacıyla arama motorlarını ve video içerik sitelerinin yanında mobil mesajlaşma uygulamasında kurdukları grupları da benzer amaçla kullandıkları görülmüştür. Gelecekteki çalışmalarda katılımcıların bu ve benzeri uygulamalara eğitmen öncülüğünde yönlendirilmesiyle önceki çalışmalarda rastlanılan forum ile ilgili özniteliklere ait verinin ikame edilmesi ve toplanması mümkün olabilir. Son olarak gelecekteki çalışmalar için tek bir ders yerine farklı içerikteki derslerden elde edilen sonuçların karşılaştırılması, kullanılan öğretim materyallerinin ve öğretim tasarımının motivasyonu artırmaya yönelik bir bakış açısıyla ele alınması ve geliştirilen modeller arasından yüksek performans sergileyen model ile derse yönelik motivasyon düzeyinin tahmin edilmesi amacıyla bir araç ve ara yüz geliştirilmesi önerilmektedir.

KAYNAKLAR

Acar, S., 2009, Web Destekli Performans Tabanlı Öğrenmede ARCS Motivasyon Stratejilerinin Öğrencilerin Akademik Başarılarına, Öğrenmenin Kalıcılığına, Motivasyonlarına ve Tutumlarına Etkisi, Doktora tezi, Gazi Üniversitesi.

Ademi, N., Loshkovska, S. & Kalajdziski, S., 2019, Prediction of student success through analysis of moodle logs: Case study. In *ICT Innovations 2019. Big Data Processing and Mining: 11th International Conference, ICT Innovations 2019, Ohrid, North Macedonia, October 17–19, 2019, Proceedings 11* (pp. 27-40). Springer International Publishing.

Agudo-Peregrina, Á.F., Iglesias-Pradas, S., Conde-González, M.Á. & Hernández-García, Á., 2014, Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in human behavior*, *31*, pp.542-550.

Ai, J. & Laffey, J., 2007, Web mining as a tool for understanding online learning. *MERLOT Journal of Online Learning and Teaching*, *3*(2), pp.160-169.

Akçapınar, G., 2014, Çevrimiçi Öğrenme Ortamındaki Etkileşim Verilerine Göre Öğrencilerin Akademik Performanslarının Veri Madenciliği Yaklaşımı ile Modellenmesi, Doktora Tezi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü.

Akpınar, H., 2017, *DATA: Veri Madenciliği Veri Analizi*, Papatya Yayıncılık Eğitim A.Ş. Istanbul, ISBN: 97-605-4220-81-6.

Al Qatawneh, K.S., 2014, Correlation between curriculum's conceptual base and students' motivation towards English language learning. *Dirasat: Educational Sciences*, *41*(2), pp.997-1008.

Al Shalabi, L., Shaaban, Z. & Kasasbeh, B., 2006, Data mining: A preprocessing engine. *Journal of Computer Science*, *2*(9), pp.735-739.

Al-Ajlan, A. & Zedan, H., 2008, Why moodle. In *2008 12th IEEE International Workshop on Future Trends of Distributed Computing Systems* (pp. 58-64). IEEE.

Aldowah, H., Al-Samarraie, H. & Fauzy, W.M., 2019, Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, *37*, pp.13-49.

Ames, C., 1992, Classrooms: Goals, Structures, and Student Motivation. *Journal of Educational Psychology*, *84*(3), p.261.

Arnone, M.P. & Small, R.V., 1995, Arousing and Sustaining Curiosity: Lessons from the ARCS model. In *Proceedings of the Annual National Conference of the Association of Educational Communications and Technology (AECT), Anaheim, CA.*

Arroyo, I., & Woolf, B. P. (2005). Inferring learning and attitudes from a Bayesian network of log file data. In *Proceedings of the 2005 Conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology* (pp. 868-870). IOS Press.

Atalay, M. and Çelik, E., 2017. Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları-Artificial Intelligence and Machine Learning Applications in Big Data Analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *9*(22), pp.155-172.

Babo, R. & Azevedo, A., 2012, *Higher education institutions and learning management systems: adoption and standardization*. United States of America: Information Science Reference. ISBN: 978-1-60960-884-2.

Baker, M.J., 2000. The roles of models in Artificial Intelligence and Education research: A Pospective View. *Journal of Artificial Intelligence and Education*, *11*, pp.122-143.

Baker, R.S. & Yacef, K., 2009, The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, *1*(1), pp.3-17.

Baker, R.S., Lindrum, D., Lindrum, M.J. & Perkowski, D., 2015, Analyzing Early At-Risk Factors in Higher Education E-Learning Courses. *International Educational Data Mining Society*.

Bandura, A., 1977, Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, *84*(2), p.191.

Bandura, A., 1997, *Self-Efficacy: The Exercise of Control*, W. H. Freeman & Company, NY, ISBN: 0-7167-2626-2.

Barto, A.G., 2012. Intrinsic Motivation and Reinforcement Learning. *Intrinsically Motivated Learning in Natural and Artificial Systems*, pp.17-47.

Barua, S., Islam, M. M., Yao, X., & Murase, K., 2014, 'MWMOTE--Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning', *IEEE Transactions on Knowledge and Data Engineering*, 26(2), pp. 405-425. doi: 10.1109/tkde.2012.232.

Baştürk, S., Dönmez, G., & Dicle, A. N., 2013, Geçerlik ve güvenirlik. İçinde: S. Baştürk (Ed.), *Bilimsel Araştırma Yöntemleri* (pp. 162-194). Ankara: Vize Yayıncılık.

Berendt, B., Littlejohn, A. & Blakemore, M., 2020. AI in Education: Learner Choice and Fundamental Rights. *Learning, Media and Technology*, *45*(3), pp.312-324.

Berrar, D., 2018, Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*, Volume 1,pp. 542.

Boyd, D., & Crawford, K., 2012, Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679.

Bozkurt, A., 2016, Öğrenme analitiği: e-öğrenme, büyük veri ve bireyselleştirilmiş öğrenme. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, *2*(4), pp.55-81.

Brachman, R. J. & Anand, T., 1996. The process of knowledge discovery in databases. In Fayyad, U. M. et al. (Eds.), *Advances in knowledge discovery and data mining*. AAAI Press / The MIT Press.

Bravo, J. & Ortigosa, A., 2009, Detecting Symptoms of Low Performance Using Production Rules. In: *2nd International Conference on Educational Data Mining*, Spain, pp. 31-40.

Breiman, L., 2001. Random forests. *Machine learning*, *45*, pp.5-32.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 2015, *Classification and Regression Trees*. CRC press; 1984, Florida US, ISBN: 0-412-04841-8.

Brito, M., Medeiros, F. & Bezerra, E., 2019, A report-type plugin to indicate dropout risk in the virtual learning environment moodle. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)* (Vol. 2161, pp. 127-128). IEEE.

Browne, M. W. & Cudeck, R., 1993, Alternative Ways of Assessing Model Fit, In Bollen, K. and Long, J. (eds.) *Testing Structural Equation Models*, Sage, pp. 136-162.

Caelen, O., 2017, A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, *81*(3-4), pp.429-450.

Calvo-Flores, M.D., Galindo, E.G., Jiménez, M.P. & Pineiro, O.P., 2006, Predicting students' marks from Moodle logs using neural network models. *Current Developments in Technology-Assisted Education*, *1*(2), pp.586-590.

Chang, L. Y., Wang, H. W., 2006, Analysis of Traffic Injury: An Application of Nonparametric Classification Tree Techniques, *Accident Analysis Prevention*, 38: 1019-1027.

Chang, M.M. & Lehman, J.D., 2002, Learning Foreign Language Through an Interactive Multimedia Program: An Experimental Study on The Effects of The Relevance Component of the ARCS Model, *CALICO Journal*, pp.81-98.

Chang, N.C. & Chen, H.H., 2015, A motivational analysis of the ARCS model for information literacy courses in a blended learning environment. *Libri*, 65(2), pp.129-142.

Charniak, E. and Mcdermott, D.,1985, *Introduction to Artificial Intelligence*. Reading, Mass.: Addison-Wesley.

Chatti, M.A., Dyckhoff, A.L., Schroeder, U. & Thüs, H., 2012, A reference model for learning analytics, *International Journal of Technology Enhanced Learning*, *4*(5-6), pp.318-331.

Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P., 2002, SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, pp.321-357.

Chen, L., Chen, P. and Lin, Z., 2020. Artificial Intelligence in Education: A review. *IEEE Access*, *8*, pp.75264-75278.

Chen, Y.T., 2012, The effect of thematic video-based instruction on learning and motivation in e-learning. *International Journal of Physical Sciences*, *7*(6), pp.957-965.

Chung, H.M. & Gray, P., 1999, *Critical Issues in Data Mining*, Journal of Management Information Systems (16:1), pp 11-16.

Ciolacu, M., Tehrani, A.F., Beer, R. & Popp, H., 2017, October. Education 4.0—Fostering student's performance with machine learning methods. In *2017 IEEE 23rd international symposium for design and technology in electronic packaging (SIITME)* (pp. 438-443). IEEE.

Ciolacu, M., Tehrani, A.F., Beer, R. & Popp, H., 2017, October. Education 4.0—Fostering student's performance with machine learning methods. In *2017 IEEE 23rd international symposium for design and technology in electronic packaging (SIITME)* (pp. 438-443). IEEE.

Coates, H., James, R. and Baldwin, G., 2005, A Critical Examination of the Effects of Learning Management Systems on University Teaching and Learning. *Tertiary education and Management*, *11*(1), pp.19-36.

Cocea, M. & Weibelzahl, S., 2007, Eliciting motivation knowledge from log files towards motivation diagnosis for Adaptive Systems. In *User Modeling 2007: 11th International Conference, UM 2007, Corfu, Greece, July 25-29, 2007. Proceedings 11* (pp. 197-206). Springer Berlin Heidelberg.

Cocea, M. and Weibelzahl, S., 2006, Can log files analysis estimate learners' level of motivation?, in *Workshop week Lernen—Wissensentdeckung—Adaptivitat*, pp. 32-35. Hildesheim, Germany.

Cohen, A., 2017, Analysis of Student Activity in Web-Supported Courses as a Tool for Predicting Dropout. *Educational Technology Research and Development*, *65*, pp.1285-1304.

Cordón, I., García, S., Fernández, A., & Herrera, F., 2018, Imbalance: Oversampling algorithms for imbalanced classification in R, *Knowledge-Based Systems*, 161, 329-341.

Cortes, C., & Vapnik, V., 1995, Support-Vector Networks. *Machine learning*, *20*, 273-297.

Cover, T., & Hart, P., 1967, Nearest neighbor pattern classification. *IEEE transactions on information theory*, *13*(1), 21-27.

Çiftçi, S., 2006, Uzaktan Eğitimde Öğrencilerin Ders Çalışma Etkinliklerinin Log Verileri Analiz Edilerek İncelenmesi, Yüksek Lisans Tezi, Gazi Üniversitesi.

Dayan, P., Sahani, M. & Deback, G., 1999. Unsupervised learning. *The MIT encyclopedia of the cognitive sciences*, pp.857-859.

De Mauro, A., Greco, M., & Grimaldi, M., 2015, What is Big Data? A Consensual Definition and a Review of Key Research Topics. *American Institute of Physics (AIP) Conference Proceedings*, 1644(1):97-104.

De Oliveira, M.M., Barwaldt, R., Pias, M.R. & Espíndola, D.B., 2019, Understanding the student dropout in distance learning. In *2019 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE.

Deci, E.L. & Ryan, R.M., 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer Science & Business Media NY, ISBN: 978-1-4899-2273-1.

Deci, E.L., Vallerand, R.J., Pelletier, L.G. & Ryan, R.M., 1991, Motivation and education: The self-determination perspective. *Educational psychologist*, *26*(3-4), pp.325-346.

Deng, L., 2014, Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing*, 7(3-4), pp.197–387.

Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F., 2023, e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, R package version 1.7-13, https://CRAN.R-project.org/package=e1071.

Dinçer, S., 2015. Farklı Eğitsel Arayüzler Kullanılarak Hazırlanan Bilgisayar Destekli Öğretim Yazılımlarının Öğrencilerin Akademik Başarılarına, Motivasyonlarına, Derse İlgilerine, Bilgisayar Destekli Öğretimi Değerlendirmelerine ve Bilişsel Yüklerine Etkisi, Doktora Tezi, Çukurova Üniversitesi.

Distante, D., Villa, M., Sansone, N. & Faralli, S., 2020, MILA: A SCORM-compliant interactive learning analytics tool for moodle. In *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)* (pp. 169-171). IEEE.

Dobashi, K., Fulford, C.P. & Lin, M.F.G., 2019, A heat map generation to visualize engagement in classes using moodle learning logs. In *2019 4th international conference on information technology (InCIT)* (pp. 138-143). IEEE.

Efrati, V., Limongelli, C. & Sciarrone, F., 2014, A data mining approach to the analysis of students' learning styles in an e-learning community: A case study. In *Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II 8* (pp. 289-300). Springer International Publishing.

Erkorkmaz, Ü., Etikan, İ., Demir, O., Özdamar, K. & Sanisoğlu, S.Y., 2013, Doğrulayıcı faktör analizi ve uyum indeksleri. *Turkiye Klinikleri Journal of Medical Sciences*, *33*(1), pp.210-223.

Fabrigar, L.R. & Wegener, D.T., 2011. *Exploratory factor analysis*. Oxford University Press.

Falvo, D.A. & Johnson, B.F., 2007, The use of learning management systems in the United States. *TechTrends*, *51*(2), p.40.

Fayyad, U., 2001, The digital physics of data mining. *Communications of the ACM*, *44*(3), pp.62-65.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P., 1996, From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37-37.

Feelders, A., Daniels, H., & Holsheimer, M., 2000, Methodological and practical aspects of data mining. *Information & Management*, *37*(5), 271-281.

Ferguson, R., 2012, Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5-6), pp.304-317.

Figueira, Á., 2017, Mining Moodle logs for grade prediction: a methodology walk-through. In *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 1-8).

Figueroa-Cañas, J. & Sancho-Vinuesa, T., 2020, Early prediction of dropout and final exam performance in an online statistics course. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, *15*(2), pp.86-94.

García, S., Luengo, J., & Herrera, F., 2015. *Data Preprocessing Data Mining*, 1st ed. Cham, Switzerland: Springer,

Garvey, Colin, 2018, A framework for evaluating barriers to the democratization of artificial intelligence. *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. No. 1.

Gomes, A., Ke, W., Lam, C.T., Marcelino, M.J. & Mendes, A., 2018, October. Student motivation towards learning to program. In *2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1-8). IEEE.

Gosain, A. & Sardana, S., 2017, Handling class imbalance problem using oversampling techniques: A review. In *2017 international conference on advances in computing, communications and informatics (ICACCI)* (pp. 79-85). IEEE.

Grace, L. K. J., Maheswari, V., & Nagamalai, D., 2011, Analysis of Web Logs and Web User in Web Mining. *International Journal of Network Security & Its Applications (IJNSA)*, 3(1), 35-45.

Han, J., & Kamber, M., 2006, *Data Mining: Concepts and Techniques*, 2nd edition, Morgan Kaufmann Publishers. San Francisco, CA, USA.

Haugeland, J.,1985, *Artificial Intelligence: The Very Idea.* Massachusetts Institute of Technology, MIT PRESS.

Hershkovitz, A. & Nachmias, R., 2009, Learning about online learning processes and students' motivation through web usage mining. *Interdisciplinary Journal of E-Learning and Learning Objects*, *5*(1), pp.197-214.

Holmes, W., Bialik, M. & Fadel, C., 2019. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. *The Center for Curriculum Redesign, Boston, MA.*

Hua, W., Ma C. & Zhou L., 2009, A brief review of machine learning and its application. *Information Engineering and Computer Science, ICIECS 2009. International Conference on. IEEE*, pp. 1–4.

Huang, A.Y., Lu, O.H., Huang, J.C., Yin, C.J. & Yang, S.J., 2020, Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs. *Interactive Learning Environments*, *28*(2), pp.206-230.

Huang, D.W., Diefes-Dux, H., Imbrie, P.K., Daku, B. and Kallimani, J.G., 2004. Learning motivation evaluation for a computer-based instructional tutorial using ARCS model of motivational design. In *34th Annual Frontiers in Education, 2004. FIE 2004.* (pp. T1E-30). IEEE.

Huett, J.B., Kalinowski, K.E., Moller, L. and Huett, K.C., 2008, Improving the motivation and retention of online students through the use of ARCS-based e-mails. *The American Journal of Distance Education*, *22*(3), pp.159-176.

Hvitfeldt, E., 2022, themis: Extra Recipes Steps for Dealing with Unbalanced Data, *R package version 1.0.0*, https://CRAN.R-project.org/package=themis.

Hwang, G.J., Xie, H., Wah, B.W. & Gašević, D., 2020. Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, p.100001.

Iglesias-Pradas, S., Ruiz-de-Azcárate, C. & Agudo-Peregrina, Á.F., 2015, Assessing the suitability of student interactions from Moodle data logs as predictors of cross-curricular competencies. *Computers in Human Behavior*, *47*, pp.81-89.

Jolliffe, I.T. & Cadima, J., 2016, Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), p.20150202.

Jones, O., Maillardet, R. & Robinson, A., 2014. *Introduction to scientific programming and simulation using R*. CRC Press.

Jordan, M. I., & Mitchell, T. M., 2015, Machine learning: Trends, perspectives, and prospects, *Science*, *349*(6245), 255-260.

Kantardzic, M., 2020, *Data Mining: Concepts, Models, Methods, and Algorithms*, 3rd Edition, John Wiley & Sons, New Jersey, US.

Kataria, A., & Singh, M. D., 2013, A review of data classification using k-nearest neighbour algorithm, *International Journal of Emerging Technology and Advanced Engineering*, *3*(6), 354-360.

Keller, J. & Suzuki, K., 2004, Learner Motivation and E-Learning Design: A Multinationally Validated Process. *Journal of Educational Media*, *29*(3), pp.229-239.

Keller, J. M., & Burkman, E., 1993, Motivation Principles. In M. Fleming & W. H. Levie (Eds.), *Instructional Message Design: Principles From the Behavioral and Cognitive Sciences*. Englewood Cliffs, NJ: Educational Technology Press.

Keller, J. M., 2010, *Motivational Design for Learning and Performance*. New York, NY: Springer.

Keller, J.M., 1987. Strategies for stimulating the motivation to learn. *Performance and instruction*, *26*(8), pp.1-7.

Keller, J.M., 2000, How to Integrate Learner Motivation Planning Into Lesson Planning: The ARCS Model Approach. *VII Semanario, Santiago, Cuba*, *1*, p.13.

Keller, J.M., 2008, First Principles of Motivation to Learn and E-learning. *Distance Education*, *29*(2), pp.337-338.

Kitchin, R. & McArdle, G., 2016, What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets, *Big Data & Society*, 3(1).

Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M. and Klein, M., 2002, *Logistic regression* (p. 536), New York: Springer-Verlag.

Kondo, N., Okubo, M. & Hatanaka, T., 2017, July. Early detection of at-risk students using machine learning based on LMS log data. In *2017 6th IIAI international congress on advanced applied informatics (IIAI-AAI)* (pp. 198-201). IEEE.

Korjus, K., Hebart, M.N. & Vicente, R., 2016. An efficient data partitioning to improve classification performance while keeping parameters interpretable. *PloS one*, *11*(8), p.e0161788.

Kotsiantis, S. B., 2007, Supervised machine learning: A Review of Classification Techniques. *Informatica: An International Journal of Computing and Informatics, 31, 249-268.*.

Kreutzer, Ralf T., & Marie Sirrenberg, 2020, *Understanding artificial intelligence*. Berlin, Germany: Springer International Publishing,

Kuhn, M., & Quinlan, R., 2023, C50: C5.0 Decision Trees and Rule-Based Models, *R package version 0.1.8*, https://CRAN.R-project.org/package=C50.

Kuhn, M., 2022, caret: Classification and Regression Training, *R package version 6.0-93*, https://CRAN.R-project.org/package=caret.

Kularbphettong, K., Waraporn, P. & Tongsiri, C., 2012, Analysis of student motivation behavior on e-learning based on association rule mining. *International Journal of Information and Communication Engineering*, *6*(6), pp.794-797.

Kumar, S., Gankotiya, A.K. & Dutta, K., 2011, A comparative study of moodle with other e-learning systems. In *2011 3rd International Conference on Electronics Computer Technology* (Vol. 5, pp. 414-418). IEEE.

Kühl N., Goutier M., Hirt R., & Satzger G., 2019 ,Machine Learning in Artificial Intelligence: Towards a Common Understanding, in *Proceedings of International Conference on DSystem Sciences* (HICSS-52), pp. 1–11.

Lackaye, T.D. & Margalit, M., 2006, Comparisons of achievement, effort, and self-perceptions among students with learning disabilities and their peers from different achievement groups. *Journal of learning disabilities*, *39*(5), pp.432-446.

Lee, J., 2018, Visual Stimuli of Foods Increase Academic Performance and Create High Beta and Gamma EEG Oscillations, Master's Thesis, Seoul National University.

Legault, L., 2020. Intrinsic and Extrinsic Motivation. *Encyclopedia of Personality and Individual Differences*, pp.2416-2419.

Lewis, D. D., 1998, Naive (Bayes) at Forty: The independence assumption in information retrieval, In *Machine Learning: ECML-98: 10th European Conference on Machine Learning Chemnitz, Germany, April 21–23, 1998 Proceedings 10* (pp. 4-15), Springer Berlin Heidelberg.

Liaw, A., & Wiener, M., 2002, Classification and Regression by randomForest, *R News*, 2(3), 18-22.

Liu, B., 2007, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, Springer Berlin Heidelberg New York, ISBN: 78-3-540-37881-5.

Liu, Q., & Wu Y., 2012, *Supervised Learning, Encyclopedia of the Sciences of Learning*, In: Seel, N. M., (ed.), Springer New York, ISBN: 978-1-4419-1427-9, Pp. 3243-3245.

Liyanage, M.P.P., KS, L.G. & Hirakawa, M., 2016. Detecting learning styles in learning management systems using data mining. *Journal of Information Processing*, *24*(4), pp.740-749.

Lonn, S. & Teasley, S.D., 2009, Saving time or innovating practice: Investigating perceptions and uses of Learning Management Systems. *Computers & education*, *53*(3), pp.686-694.

Luckin, R.; Holmes, W., Griffiths, M., & Forcier, L. B., 2016, *Intelligence Unleashed: an Argument for AI in Education*. Pearson Education, London.

Lunardon, N., Menardi, G., & Torelli, N., 2014, ROSE: a Package for Binary Imbalanced Learning, *R Journal*, 6(1), 82-92.

Ma, L. & Lee, C.S., 2021, Evaluating the effectiveness of blended learning using the ARCS model. *Journal of computer assisted learning*, *37*(5), pp.1397-1408.

MacCallum, R.C., Browne, M.W. & Sugawara, H.M., 1996, Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, *1*(2), p.130.

Mega, C., Ronconi, L. & De Beni, R., 2014, What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement. *Journal of educational psychology*, *106*(1), p.121.

Merenda, P.F., 1997, A guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. *Measurement and Evaluation in counseling and Development*, *30*(3), pp.156-164.

Metraglia, R. & Villa, V., 2014, Engineering graphics education: webcomics as a tool to improve weaker students' motivation. *Research Journal of Applied Sciences, Engineering and Technology*, *7*(19), pp.4106-4114.

Metraglia, R., Villa, V., Baronio, G. & Adamini, R., 2015, High school graphics experience influencing the self-efficacy of first-year engineering students in an introductory engineering graphics course. *The Engineering Design Graphics Journal*, *79*(3).

Meyer, D., 2015, Support vector machines, The interface to LIBSVM in *package e1071*, Austria, Available: https://cran.rproject.org/web/packages/e1071/vignettes/svmdoc.pdf

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F., 2023, e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, *R package version 1.7-13*, https://CRAN.R-project.org/package=e1071.

Mohammed, R., Rawashdeh, J. & Abdullah, M., 2020, April. Machine learning with oversampling and undersampling techniques: overview study and experimental results. In *2020 11th international conference on information and communication systems (ICICS)* (pp. 243-248). IEEE.

Mohri, M., Rostamizadeh, A., & Talwalkar, A., 2012, Foundations of Machine Learning, MIT Press, Cambridge, Massaschusetts, ISBN : 978-0-262-01825-8.

Molinaro, A.M., Simon, R. & Pfeiffer, R.M., 2005. Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, *21*(15), pp.3301-3307.

Moodle, 2020, About Moodle – Moodle Docs, https://docs.moodle.org/401/en/About_Moodle, [Ziyaret tarihi: 2 Mayıs 2021]

Mughal, M.J.H., 2018, Data mining: Web data mining techniques, tools and algorithms: An overview. *International Journal of Advanced Computer Science and Applications*, *9*(6).

Natale, S. & Ballatore, A. (2017). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence: The International Journal of Research into New Media Technologies*, 26(1), pp.3–18.

Naveed, T. and Bhatti, N.M., 2017. Validation of course interest survey tool in MBBS students of a private medical college of Pakistan. *Journal of Islamic International Medical College (JIIMC)*, *12*(4), pp.201-207.

Nilsson, N.J.,2003, *Artificial Intelligence : A New Synthesis*. San Francisco, California: Kaufmann.

Ortigosa, A., Carro, R.M., Bravo-Agapito, J., Lizcano, D., Alcolea, J.J. & Blanco, O., 2019, From lab to production: Lessons learnt and real-life challenges of an early student-dropout prevention system. *IEEE Transactions on Learning Technologies*, *12*(2), pp.264-277.

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J., 2017, Supervised machine learning algorithms: classification and

comparison, *International Journal of Computer Trends and Technology (IJCTT)*, *48*(3), 128-138.

Özdağoğlu, G., Öztaş, G.Z. & Çağliyangil, M., 2019, An application framework for mining online learning processes through event-logs. *Business Process Management Journal*, *25*(5), pp.860-886.

Özbay, Ö., 2015, Öğretim Yönetim Sistemi Üzerinde Üniversite (Lisans) Düzeyindeki Öğrenci Hareketliliğinin Veri Madenciliği Yöntemleriyle Analizi, Yüksek Lisans Tezi, Başkent Üniversitesi.

Pandya, R., & Pandya, J., 2015, C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *International Journal of Computer Applications*, *117*(16), 18-21.

Papamitsiou, Z. & Economides, A.A., 2014, Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, *17*(4), pp.49-64.

Peak, P., Kalinowski, K., Tyler-Wood, T. & Huett, J., 2008, Indicators Impacting Student Success in Distance Education. In *Society for Information Technology & Teacher Education International Conference* (pp. 637-641). Association for the Advancement of Computing in Education (AACE).

Pelaccia, T. & Viau, R., 2017. Motivation in medical education. *Medical teacher*, *39*(2), pp.136-140.

Pintrich, P.R. & De Groot, E.V., 1990, Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, *82*(1), p.33.

Pintrich, P.R., 2000. Multiple Goals, Multiple Pathways: The Role of Goal Orientation in Learning and Achievement. *Journal of Educational Psychology*, *92*(3), p.544.

Pintrich, P.R., 2003. A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. *Journal of Educational Psychology*, *95*(4), p.667.

Pyle, D., 1999, *Data Prepataion for Data Mining*, Morgan Kaufmann Publishers, Inc. San Francisco, California.

R Core Team, 2022, R: A language and environment for statistical computing, *R Foundation for Statistical Computing*, Vienna, Austria, https://www.R-project.org/.

Rapanta, C., Botturi, L., Goodyear, P., Guàrdia, L. & Koole, M., 2021, Balancing technology, pedagogy and the new normal: Post-pandemic challenges for higher education. *Postdigital Science and Education*, *3*(3), pp.715-742.

Raza, S.A., Qazi, W., Khan, K.A. & Salam, J., 2021, Social isolation and acceptance of the learning management system (LMS) in the time of COVID-19 pandemic: an expansion of the UTAUT model. *Journal of Educational Computing Research*, *59*(2), pp.183-208.

Rich, E., Knight, K. & Nair, S.B.,2009,. *Artificial Intelligence.* 3rd ed. New Delhi, India: Tata McGraw Hill Education Private Limited.

Richardson, M., 2009, Principal Component Analysis. University of Oxford (Special Topic). online: http://aurora.troja.mff.cuni.cz/nemec/idl/09bonus/pca.pdf, [Ziyaret Tarihi: 8 Nisan 2022].

Rish, I., 2001, An Empirical Study of the Naive Bayes Classifier. In: *IJCAI 2001 Workshop on Empirical Methods in Artificial İntelligence*, 3 (22), pp. 41-46.

Romero, C. & Ventura, S., 2007, Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, *33*(1), pp.135-146.

Romero, C. & Ventura, S., 2010, Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)*, *40*(6), pp.601-618.

Romero, C. & Ventura, S., 2013, Data mining in education. *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, *3*(1), pp.12-27.

Romero, C., Espejo, P.G., Zafra, A., Romero, J.R. & Ventura, S., 2010, Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), pp.135-146.

Romero, C., Ventura, S. & García, E., 2008b. Data mining in course management systems: Moodle case study and tutorial. *Computers & education*, *51*(1), pp.368-384.

Romero, C., Ventura, S., Espejo, P.G. & Hervás, C., 2008a, Data mining algorithms to classify students. In *Educational data mining 2008*.

Russell, S. J. & Norvig P., 2010, *Artificial Intelligence : A Modern Approach*, Upper Saddle River, N.J. : Prentice Hall.

Rutkowski, L., Jaworski, M., Pietruczuk, L. & Duda, P., 2014. The CART decision tree for mining data streams. *Information Sciences*, *266*, pp.1-15.

Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P. and Mononen, J., 2018, Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behavioural processes*, *148*, pp.56-62.

Ryan, R.M. & Deci, E.L., 2000b, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, *55*(1), p.68.

Ryan, R.M. and Deci, E.L., 2000a. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), pp.54-67.

Schunk, D. H., & Pajares, F., 2009, Self-Efficacy Theory. In K. R. Wentzel & A. Wigfield (Eds.), Handbook of Motivation at School, pp. 35-53, New York: Routledge.

Schunk, D.H., 1991, Self-efficacy and academic motivation. *Educational psychologist*, *26*(3-4), pp.207-231.

Schunk, D.H., Meece, J.R. & Pintrich, P.R., 2014. *Motivation in education: Theory, research, and applications*. Pearson Higher Ed. ISBN: 1-292-04147-1.

Seifert, T., 2004. Understanding student motivation. *Educational research*, *46*(2), pp.137-149.

Shrestha, N., 2021, Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, *9*(1), pp.4-11.

Siemens, G. & Baker, R.S.D., 2012, Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* ,pp. 252-254.

Sijing, L. & Lan, W., 2018, August. Artificial intelligence education ethical problems and solutions. In *2018 13th International Conference on Computer Science & Education (ICCSE)* (pp. 1-5). IEEE.

Silahtaroğlu, G., 2016, *Veri Madenciliği: Kavram ve Algoritmaları*, Papatya Bilim Akademik Yayınevi, ISBN: 978-975-6797-81-5.

Sokolova, M., Japkowicz, N. & Szpakowicz, S., 2006, Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In *AI 2006: Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, December 4-8, 2006. Proceedings 19* (pp. 1015-1021). Springer Berlin Heidelberg.

Song, S.H. & Keller, J.M., 2001, Effectiveness of motivationally adaptive computer-assisted instruction on the dynamic aspects of motivation. *Educational technology research and development*, *49*(2), pp.5-22.

Steiner, C.M., Kickmeier-Rust, M.D. & Albert, D., 2014, Learning analytics and educational data mining: An overview of recent techniques. *Learning analytics for and in serious games*, *6*, pp.61-75.

Sukhbaatar, O., Ogata, K. & Usagawa, T., 2018, October. Mining educational data to predict academic dropouts: a case study in blended learning course. In *TENCON 2018-2018 IEEE region 10 conference* (pp. 2205-2208). IEEE.

Sutton, R.S. & Barto, A.G., 1998, *Reinforcement Learning: An Introduction*, MIT press, Cambridge, MA.

Tamada, M.M., Giusti, R. & Netto, J.F.D.M., 2022, Predicting students at risk of dropout in technical course using LMS logs. *Electronics*, *11*(3), p.468.

Tancev, G. & Pascale, C., 2020. The relocation problem of field calibrated low-cost sensor systems in air quality monitoring: a sampling bias. *Sensors*, *20*(21), p.6198.

Tavakol, M. & Dennick, R., 2011, Making sense of Cronbach's alpha. *International journal of medical education*, *2*, p.53.

Tempola, F., Rosihan, R. & Adawiyah, R., 2021, Holdout validation for comparison classfication Naïve bayes and KNN of recipient Kartu Indonesia pintar. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1125, No. 1, p. 012041). IOP Publishing.

Therneau, T., & Atkinson, B., 2022, rpart: Recursive Partitioning and Regression Trees, *R package version 4.1.19*, https://CRAN.R-project.org/package=rpart.

Tinsley, H.E. & Tinsley, D.J., 1987, Uses of factor analysis in counseling psychology research. *Journal of counseling psychology*, *34*(4), p.414.

Topal, M., 2020, Oyunlaştırma ile Zenginleştirilmiş Çevrimiçi Öğrenmenin Başarı, Çevrimiçi Bağlılık ve Öğrenme Motivasyonu Üzerindeki Etkisi, Doktora Tezi, Sakarya Üniversitesi.

Torunoğlu, D., Çakirman, E., Ganiz, M.C., Akyokuş, S. & Gürbüz, M.Z., 2011, Analysis of preprocessing methods on classification of Turkish texts. In *2011 International Symposium on Innovations in Intelligent Systems and Applications* (pp. 112-117). IEEE.

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S. & Ragos, O., 2019, Implementing AutoML in educational data mining for prediction tasks. *Applied Sciences*, *10*(1), p.90.

Ucar, H. & Kumtepe, A.T., 2019, Effects of the ARCS-V-based motivational strategies on online learners' academic performance, motivation, volition, and course interest. *Journal of Computer Assisted Learning*, *36*(3), pp.335-349.

UNESCO, 2019, Planning Education in the AI Era: Lead the Leap, *International Conference on Artificial Intelligence and Education, 16-18 May 2019 Beijing*, United Nations Educational, Scientific and Cultural Organization, 1-11.

Venables, W. N., & Ripley, B. D., 2002, *Modern Applied Statistics with S*, Fourth Edition, Springer, New York, ISBN 0-387-95457-0.

Viau, R., 2015, *Okulda Motivasyon: Okulda Güdülenme ve Güdülenmeyi Öğrenme*, Anı Yayıncılık (Çev.: Y. Budak), Ankara, ISBN: 978-605-170-012-0.

Wah, L.K., 2015, The effects of instruction using the arcs model and Geogebra on upper secondary students' motivation and achievement in learning combined transformation. *Asia Pacific Journal of Educators and Education*, *30*(5), pp.141-158.

Walesiak, M., & Dudek, A., 2020, The Choice of Variable Normalization Method in Cluster Analysis, In Soliman KS (Ed.), *Education Excellence and Innovation Management: A 2025 Vision to Sustain Economic Development During Global Challenges*, 325-340, ISBN 978-0-9998551-4-1.

Winograd, T., 1991, *The Boundaries of Humanity: Humans, Animals, Machines,* James Sheehan and Morton Sosna, eds., Berkeley: University of California Press, pp. 198-223.

Wong, T.T., 2020, Linear approximation of F-measure for the performance evaluation of classification algorithms on imbalanced data sets. *IEEE Transactions on Knowledge and Data Engineering*, *34*(2), pp.753-763.

Woolf, B.P., Lane, H.C., Chaudhri, V.K. & Kolodner, J.L., 2013. AI grand challenges for education. *AI magazine*, *34*(4), pp.66-84.

Xia, Y. & Yang, Y., 2019, RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior research methods*, *51*, pp.409-428.

Ye, N., 2003. *The Handbook of Data Mining*. Lawrence Erlbaum Associates, Inc., Publishers, ISBN: 0-8058-4081-8.

Yıldırım, D., 2018, Uzaktan Eğitim Öğrencilerinin Akademik Başarıları, Etkileşim ve Gezinme Örüntülerinin İlişkisel Çözümlenmesi, Doktora Tezi, Hacettepe Üniversitesi.

Zhang, G., Cheng, Z., He, A., & Huang, T., 2003, A WWW-based learner's learning motivation detecting system. In *Proceedings of International Workshop on" Research Directions and Challenge Problems in Advanced Information Systems Engineering*, Honjo City, Japan.

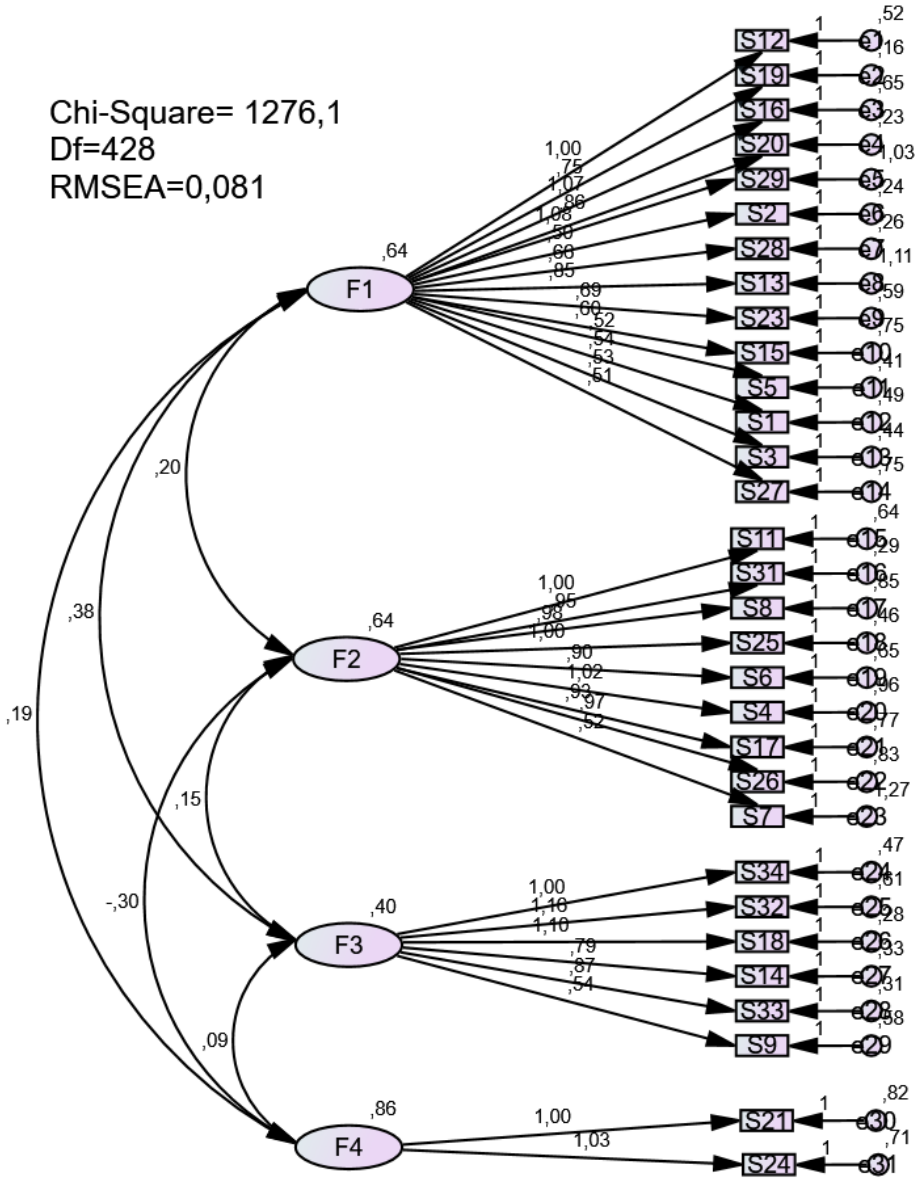
Zhou, Q., Quan, W., Zhong, Y., Xiao, W., Mou, C. & Wang, Y., 2017, Predicting high-risk students using Internet access logs. *Knowledge and Information Systems*, *55*, pp.393-413.

EKLER

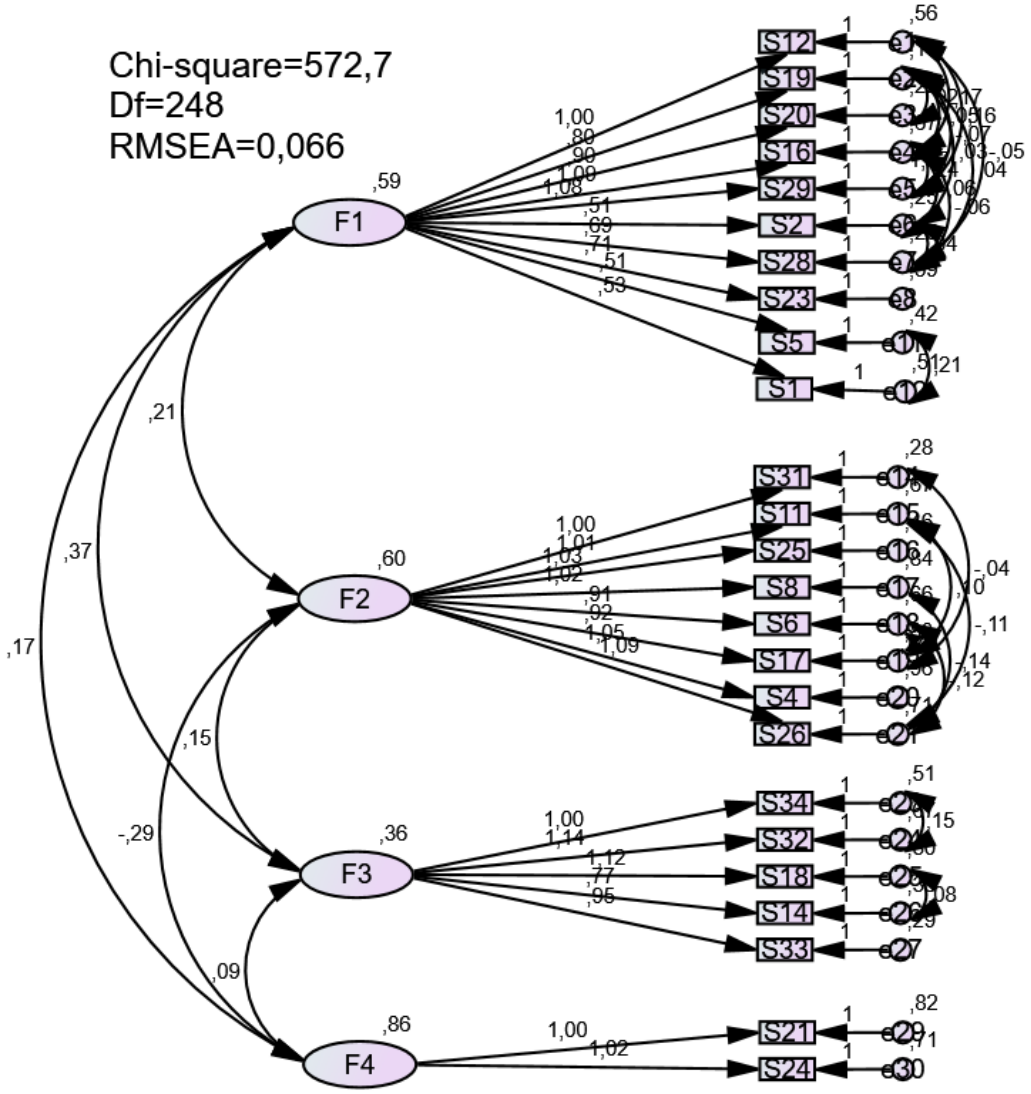
EK 1. TBA Sonuçlarına Ait Döndürülmüş Faktör Matrisi, Faktörlere Ait Güvenirlik Katsayıları, Madde Ortalama ve Standart Sapma Puanları (Madde Sayısı=31)

	Faktör 1	**Faktör 2**	**Faktör 3**	**Faktör 4**		
Cronbach Alpha Değeri	*0,89*	*0,87*	*0,83*	*0,70*	Ortalama	Standart Sapma
S12	0,786				3,98	1,08
S19	0,735				4,61	0,73
S16	0,727				3,96	1,17
S20	0,727				4,47	0,83
S29	0,694				3,52	1,33
S2	0,659				4,72	0,64
S28	0,63				4,60	0,73
S13	0,613				3,77	1,26
S23	0,592				4,20	0,94
S15	0,497				3,91	0,99
S5	0,494				4,58	0,76
S1	0,479				4,36	0,82
S3	0,475				4,57	0,79
S27	0,444				4,37	0,96
S11		0,83			4,13	1,13
S31		0,747			4,59	0,93
S8		0,727			4,08	1,21
S25		0,699			4,43	1,05
S6		0,658			4,35	1,08
S4		0,655			3,93	1,28
S17		0,649			4,02	1,15
S26		0,576			4,21	1,20
S7		0,572			3,26	1,20
S34			0,73		4,31	0,94
S32			0,725		4,28	1,07
S18			0,678		4,47	0,87
S14			0,656		4,62	0,76
S33			0,573		4,48	0,79
S9			0,507		4,44	0,83
S21				0,63	2,22	1,30
S24				0,601	2,80	1,28
Çıkarım Yöntemi: Temel Bileşenler Analizi						
Döndürme Yöntemi: Kaiser Normalizasyonu ile Varimax						

EK 2. DİÖ – 31 Maddeli Yapısal Eşitlik Modeli Diyagramı



EK 3. DİÖ – 25 Maddeli Yapısal Eşitlik Modeli Diyagramı



EK 4. Etik Kurulu İzni

Tarih ve Sayı: 06/08/2020-103971

BEL9RFR5S

T.C.
İSTANBUL ÜNİVERSİTESİ REKTÖRLÜĞÜ
Sosyal ve Beşeri Bilimler Araştırmaları Etik Kurulu Başkanlığı

Sayı :35980450-663.05-
Konu :Mithat YAVUZARSLAN

Sayın Mithat YAVUZARSLAN

İlgi :23.06.2020 tarih ve 28166 sayılı başvurunuz.

Sorumlu araştırıcılığını üstlendiğiniz 2020/80 dosya numaralı "Yapay Zeka Yöntemleri ile Üniversite Öğrencilerinin Ders Motivasyonlarının Tahmin Edilmesi" başlıklı çalışma, Kurulumuzun 10.07.2020 tarih ve 08 sayılı toplantısında görüşülerek etik yönden uygun bulunmuş olup, karar ekte sunulmuştur.

Bilgilerinizi rica ederim.

e-İmzalı
Prof. Dr. N. Tolga SARUÇ
Başkan

EK :
Sosyal ve Beşeri Bilimler Araştırmaları Etik Kurulu Kararı

06/08/2020 B.İşl. V. : G.AKALAN

Doğrulamak İçin:http://194.27.128.66/envision.Sorgula/belgedogrulama.aspx?V=BEL9RFR5S
Ayrıntılı bilgi için irtibat : Gülsüm AKALAN (Süleyman ARIK Vekaletiyle) Dahili : 11896
İstanbul Üniversitesi Merkez Kampüsü
34452 Beyazıt/Fatih-İstanbul
Tel : 0212 440 20 89 Faks : 0212 440 20 88
e-posta : sosyalbilimleretikkurul@istanbul.edu.tr Elektronik Ağ : www.istanbul.edu.tr

K-Q TSE-ISO-EN 9000

Bu belge 5070 sayılı Elektronik İmza Kanununun 5. Maddesi gereğince güvenli elektronik imza ile imzalanmıştır.

EK 5. Bilgilendirilmiş Gönüllü Onam Formu

Sizi Mithat Yavuzarslan tarafından Doç. Dr. Çiğdem Selçukcan Erol danışmanlığında İstanbul Üniversitesi Enformatik bölümü bünyesinde yürütülen "Yapay Zekâ Yöntemleri ile Üniversite Öğrencilerinin Ders Motivasyonlarının Tahmin Edilmesi" başlıklı araştırmaya davet ediyoruz. Bu araştırmanın amacı üniversite öğrencilerinin derse dönük motivasyonlarının elde edilen çeşitli veriler üzerinde uygulanacak yapay zekâ algoritmaları ile tahmin edilmesidir.

Araştırmada sizden tahminen 15 dk. ayırmanız istenmektedir. Araştırmaya sizin dışınızda tahminen 400 kişi katılacaktır. Bu çalışmaya katılmak tamamen gönüllülük esasına dayanmaktadır. Çalışmanın amacına ulaşması için sizden beklenen, kimsenin baskısı veya telkini altında olmadan, size en uygun gelen cevapları içtenlikle verecek şekilde cevaplamanızdır. Bu formu okuyup onaylamanız, araştırmaya katılmayı kabul ettiğiniz anlamına gelecektir. Ancak, çalışmaya katılmama veya katıldıktan sonra herhangi bir anda çalışmayı bırakma hakkına da sahipsiniz. Bu çalışmadan elde edilecek bilgiler tamamen araştırma amacı ile kullanılacak olup kişisel bilgileriniz gizli tutulacaktır; ancak verileriniz yayın amacı ile kullanılabilir. İletişim bilgileriniz ise sadece izninize bağlı olarak ve farklı araştırmacıların sizinle iletişime geçebilmesi için "ortak katılımcı havuzuna" aktarılabilir.

Eğer araştırmanın amacı ile ilgili verilen bu bilgiler dışında şimdi veya sonra daha fazla bilgiye ihtiyaç duyarsanız araştırmacıya şimdi sorabilir veya mithat.yavuzarslan@yeditepe.edu.tr e-posta adresinden ulaşabilirsiniz. Araştırma tamamlandığında genel/size özel sonuçların sizinle paylaşılmasını istiyorsanız lütfen araştırmacıya iletiniz.

Yukarıda yer alan ve araştırmadan önce katılımcıya verilmesi gereken bilgileri okudum ve katılmam istenen çalışmanın kapsamını ve amacını, gönüllü olarak üzerime düşen sorumlulukları anladım. Çalışma hakkında yazılı ve sözlü açıklama aşağıda adı belirtilen araştırmacı/araştırmacılar tarafından yapıldı. Bana, çalışmanın muhtemel riskleri ve faydaları sözlü olarak da anlatıldı. Kişisel bilgilerimin özenle korunacağı konusunda yeterli güven verildi.

Bu koşullarda söz konusu araştırmaya kendi isteğimle, hiçbir baskı ve telkin olmaksızın katılmayı kabul ediyorum.

İSİM SOYİSİM

EK 6. Derse İlgi Ölçeği

Bu ankette 34 soru bulunmaktadır. Lütfen tüm soruları almakta olduğunuz dersi düşünerek cevaplayınız. Soruları cevaplarken kendi düşüncelerinizi başkalarının etkisi altında kalmadan ve içtenlikle ifade etmeniz beklenmektedir.

1 = Doğru Değil

2 = Biraz Doğru

3 = Orta Derecede Doğru

4 = Genellikle Doğru

5 = Çok Doğru

1- Dersin hocası dersin konularına karşı bizi nasıl heveslendireceğini bilir.	1	2	3	4	5
2- Bu derste öğrendiğim şeylerin bana faydalı olacağını düşünüyorum.	1	2	3	4	5
3- Bu derste başarılı olacağıma inancım tam.	1	2	3	4	5
4- Bu derste ilgimi çeken çok az şey var.	1	2	3	4	5
5- Dersin hocası ders konularının önemini kavramımızı sağlıyor.	1	2	3	4	5
6- Bu dersten yüksek bir not almak şans işidir.	1	2	3	4	5
7- Bu derste başarılı olmak için çok çalışmam gerekiyor.	1	2	3	4	5
8- Bu dersin içeriğiyle daha önceki bilgilerim arasında bir bağlantı göremiyorum.	1	2	3	4	5
9- Bu dersi geçip geçemeyeceğim bana bağlıdır.	1	2	3	4	5
10- Dersin hocası konuyu anlatırken belirsiz noktalar bırakıyor.	1	2	3	4	5
11- Bu dersin içeriği bana çok zor geliyor.	1	2	3	4	5
12- Bu dersin beni fazlasıyla tatmin edeceğini hissediyorum.	1	2	3	4	5
13- Bu derste yüksek hedefler belirleyerek mükemmele ulaşmaya çalışıyorum.	1	2	3	4	5
14- Kendi notlarımla arkadaşlarımınkini karşılaştırdığımda bana adil davranıldığını düşünüyorum.	1	2	3	4	5

15- Bu dersteki öğrenciler dersin konularına ilgili görünüyorlar.	1	2	3	4	5
16- Bu derse çalışmak eğlenceli.	1	2	3	4	5
17- Dersin hocasının ödevlerime vereceği notu tahmin etmem çok zor.	1	2	3	4	5
18- Sınavlardan almayı düşündüğüm notla aldığım notu karşılaştırdığımda dersin hocasının değerlendirmelerinden memnunum.	1	2	3	4	5
19- Bu dersin bana kattıklarından memnunum.	1	2	3	4	5
20- Bu dersin içeriği beklenti ve hedeflerimle uyumludur.	1	2	3	4	5
21- Dersin hocası ilginç ve beklenmedik şekilde ders işliyor.	1	2	3	4	5
22- Sınıftaki öğrenciler bu derse aktif olarak katılıyorlar.	1	2	3	4	5
23- Bu derste hedeflerime ulaşmam için derste iyi olmam önemlidir.	1	2	3	4	5
24- Dersin hocası derste ilginç öğretim teknikleri kullanıyor.	1	2	3	4	5
25- Bu dersin bana pek fayda sağlayacağını düşünmüyorum.	1	2	3	4	5
26- Bu dersi dinlerken sıklıkla hayallere dalarım.	1	2	3	4	5
27- Bu derse sıkı çalıştığım takdirde başarılı olacağıma inanıyorum.	1	2	3	4	5
28- Bu dersin bana faydaları açık ve nettir.	1	2	3	4	5
29- Bu dersin soruları ve problemleri sık sık merakımı artırıyor.	1	2	3	4	5
30- Bu dersin zorluk seviyesini ne çok zor ne de çok kolay buluyorum	1	2	3	4	5
31- Bu ders beni fazlasıyla hayal kırıklığına uğrattı.	1	2	3	4	5
32- Bu derste çalışmalarımın karşılığını not ve diğer geribildirimlerle yeterince aldığımı düşünüyorum.	1	2	3	4	5
33- Bu derste yapmam gereken çalışmaların yükü dersin yapısına uygun.	1	2	3	4	5
34- Bu dersteki performansıma dair yeterli geribildirim alıyorum.	1	2	3	4	5

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Mithat Yavuzarslan
Doğum Yeri	
Doğum Tarihi	Tarih girmek için tıklayın veya dokunun.
Uyruğu	☑ T.C. ❒ Diğer:
Telefon	
E-Posta Adresi	
Web Adresi	

Eğitim Bilgileri	
Lisans	
Üniversite	Yeditepe Üniversitesi
Fakülte	Eğitim Fakültesi
Bölümü	Bilgisayar ve Öğretim Teknolojileri Öğretmenliği
Mezuniyet Yılı	17.06.2013

Yüksek Lisans	
Üniversite	Yeditepe Üniversitesi
Enstitü Adı	Eğitim Bilimleri Enstitüsü
Anabilim Dalı	Eğitim Programları ve Öğretim Ana Bilim Dalı
Programı	Eğitim Programları ve Öğretim

Doktora	
Üniversite	İstanbul Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Enformatik Anabilim Dalı
Programı	Enformatik Programı

Makale ve Bildiriler

Yavuzarslan, M., Olgun, H., & Yazıcı, S., 2019, A Pilot Study on the Comparison Between Blended and F2F Learning Methods in a SQL Course. *The Turkish Online Journal of Educational Technology, Special Issue for IETC-ITEC-INTE 2019*, 2, 17-25.

Yavuzarslan, M., Demirkol, D., & Gülseçen, S., 2020, The role of virtual reality implementations in rehabilitation of patients with stroke: A literature review. *Journal of Information Systems and Management Research*, 2(1), 42-49.

Yavuzarslan, M., 2020, Hesaplama/ Tablo/ Grafik Programları. İçinde: S. Bayram, G. Kızılkaya Cumaoğlu, & İ. Kuşbeyzi Aybar (Ed.) *Bilişim Teknolojileri: Kuramsal Yaklaşım ve Uygulamalar.* (pp. 173-202). Ankara: Nobel Akademik Yayıncılık.