



**T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



Doktora Tezi

**TERS YÜZ SINIFLARDAKİ VIDEO İZLEME DAVRANIŞLARI
İNCELENEREK VERİ MADENCİLİĞİ ile BAŞARININ TAHMİN
EDİLMESİ**

Kadir Burak OLGUN

Enformatik Anabilim Dalı

Enformatik Programı

DANIŞMAN

Prof. Dr. Sevinç GÜLSEÇEN

II. DANIŞMAN

Dr. Alper BAYAZIT

Şubat, 2021

İSTANBUL

Bu çalışma, 12.02.2021 tarihinde aşağıdaki jüri tarafından Enformatik Anabilim Dalı, Enformatik Programında Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Prof. Dr. Sevil Güllü SEÇEN (Danışman)
İstanbul Üniversitesi
Fakülte

Doç. Dr. Gonca KIZILKAYA
CUMAOĞLU
Yeditepe Üniversitesi
Eğitim Fakültesi

Doç. Dr. Çiğdem ÇUKCAN EROL
İstanbul Üniversitesi
Fakülte

Doç. Dr. Tülgün TOSUN
Trakya Üniversitesi
Eğitim Fakültesi

Doç. Dr. Melih ENGİN
Bursa Uludağ Üniversitesi
İnegöl İşletme Fakültesi

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 aboneli olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.



ÖNSÖZ

Lisansüstü öğrenimim boyunca çok şey öğrendiğim ve tez çalışmamda desteğini hissettiğim değerli danışman hocam Sayın Prof. Dr. Sevinç GÜLSEÇEN'e çok teşekkür eder, sevgi ve saygılarımı sunarım.

Görev yaptığım üniversitede tanıştığım ve iş hayatımda olduğu gibi tez çalışmamda da bana yeni fikirler veren II. danışmanım Sayın Dr. Alper BAYAZIT'a teşekkür ederim.

Üniversite hayatım boyunca birçok çalışmam da olduğu gibi tez çalışmamda da yanımda olan ve her konuda değerli fikirlerini benimle paylaşan Sayın Doç. Dr. Gonca KIZILKAYA CUMAOĞLU'na çok teşekkür eder, sevgi ve saygılarımı sunarım.

Tez çalışmam boyunca özellikle analiz aşamasında değerli fikirlerini benimle paylaşan tez izleme komitesi üyesi Sayın Doç. Dr. Seda TOLUN TAYALI'ya teşekkür eder, sevgi ve saygılarımı sunarım.

Akademik hayatlarındaki her türlü tecrübelerini benimle paylaşan ve tez çalışmamda ilgi ve desteklerini esirgemeyen akademisyen arkadaşlarım Araş. Gör. Dr. Serra ÇELİK ve Araş. Gör. Dr. Emre AKADAL'a en içten teşekkürlerimi sunarım.

Son olarak tez çalışmam boyunca gösterdikleri her türlü destek için sevgili ailemin her bir ferdine ayrı ayrı teşekkür ederim.

Şubat 2021

[Kadir Burak OLGUN]

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	iv
İÇİNDEKİLER.....	v
ŞEKİL LİSTESİ	vii
TABLO LİSTESİ.....	ix
KISALTMA LİSTESİ.....	xii
ÖZET	xiii
SUMMARY	xv
1. GİRİŞ	1
1.1. ARAŞTIRMA PROBLEMİ	1
1.1.1. Alt Problemler	1
1.2. ARAŞTIRMANIN AMACI VE ÖNEMİ	1
1.3. VARSAYIMLAR	3
1.4. SINIRLILIKLAR	3
2. GENEL KISIMLAR.....	4
2.1. VERİ MADENCİLİĞİ.....	4
2.1.1. Veri, Enformasyon ve Bilgi.....	4
2.1.2. Veri Tabanı, Veri Ambarı ve Büyük Veri	7
2.1.3. Veri Madenciliği Teknikleri	9
2.1.4. Veri Madenciliği Uygulama Alanları	11
2.2. EĞİTSEL VERİ MADENCİLİĞİ.....	13
2.3. ÖĞRENME ANALİTİKLERİ	23
2.4. TERS YÜZ EĞİTİM.....	28
2.5. BENZER ÇALIŞMALAR	31
3. MALZEME VE YÖNTEM.....	37
3.1.ÇALIŞMA GRUBU VE GENEL ÖZELLİKLERİ.....	37
3.2.VERİ TOPLAMA ARAÇLARI.....	37
3.3.UYGULAMA SÜRECİ	40
3.4.VERİLERİN ANALİZ EDİLMESİ	43
4. BULGULAR.....	45
4.1. TANIMLAYICI İSTATİSTİKSEL BULGULAR.....	45

4.2. ARAŞTIRMANIN ANA PROBLEMİNE AİT BULGULAR	50
4.2.1. Rastgele Orman Yöntemi Kullanılarak Gerçekleştirilen Analizlere Ait Bulgular	52
4.2.2. Naive Bayes Yöntemi Kullanılarak Gerçekleştirilen Analizlere Ait Bulgular	58
4.2.3. Destek Vektör Makineleri Yöntemi Kullanılarak Gerçekleştirilen Analizlere Ait Bulgular	65
4.3. ARAŞTIRMANIN ALT PROBLEMLERİNE AİT BULGULAR.....	71
4.3.1. Açılan Farklı Oturum Sayısı ile Akademik Başarı.....	71
4.3.2. Toplam Aksiyon Sayısı ile Akademik Başarı	73
4.3.3. Video İzlerken Harcanan Zaman ile Akademik Başarı.....	74
4.3.4. Görüntülenen Farklı Video Sayısı ile Akademik Başarı	76
4.3.5. Video Yükleme Sayısı ile Akademik Başarı	78
4.3.6. Video Oynatma Sayısı ile Akademik Başarı	79
4.3.7. Video Duraklatma Sayısı ile Akademik Başarı.....	81
4.3.8. Videonun Sonuna Gelme Sayısı ile Akademik Başarı	83
4.3.9. Toplam Atlama Yapma Sayısı ile Akademik Başarı.....	84
4.3.10. İleri Atlama Yapma Sayısı ile Akademik Başarı	86
4.3.11. Geri Atlama Yapma Sayısı ile Akademik Başarı	88
5. TARTIŞMA VE SONUÇ	90
6. KAYNAKLAR	94
7. EKLER	104
8. ÖZGEÇMİŞ	109

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1: Bilgi hiyerarşisi.	4
Şekil 2.2: En popüler veri tabanı yönetim sistemleri sıralaması.	8
Şekil 2.3: Eğitsel veri madenciliği alanında yapılan çalışmaların yıllara göre dağılımı.	16
Şekil 2.4: Eğitsel veri madenciliği alanındaki çalışmaların yıllara göre yüzdelik dağılımı.	16
Şekil 2.5: Eğitsel veri madenciliği süreci 1.	17
Şekil 2.6: Eğitsel veri madenciliği süreci 2.	18
Şekil 2.7: Eğitsel veri madenciliği alanındaki çalışmaların kullanılan modellere göre yüzdelik dağılımı.	21
Şekil 2.8: Eğitsel veri madenciliği alanındaki çalışmaların araştırma konularına göre yüzdelik dağılımı.	22
Şekil 2.9: Eğitsel veri madenciliği alanındaki çalışmaların analiz yöntemlerine göre yüzdelik dağılımı.	23
Şekil 2.10: Öğrenme analitiği süreci.	26
Şekil 2.11: Öğrenme analitiği döngüsü.	27
Şekil 2.12: Eğitsel veri madenciliği alanındaki çalışmaların çalışma grubu sayılarına göre yüzdelik dağılımı.	32
Şekil 2.13: Eğitsel veri madenciliği alanındaki çalışmaların veri toplama araçlarına göre yüzdelik dağılımı.	32
Şekil 3.1: Ham verilerin görünümü.	38
Şekil 4.1: Öğrencilerin sınıflarına göre yüzdelik dağılımları.	46
Şekil 4.2: Üniversite yerleştirme puan türlerine göre yüzdelik dağılımı.	48
Şekil 4.3: Öğrenci yerleştirme puanlarına göre başarı durumu grafiği.	50
Şekil 4.4: Rastgele orman yönteminde 1-30 kayıt aralığı tahminleri.	54
Şekil 4.5: Rastgele orman yönteminde 31-60 kayıt aralığı tahminleri.	55
Şekil 4.6: Rastgele orman yönteminde 61-90 kayıt aralığı tahminleri.	56

Şekil 4.7: Rastgele orman yönteminde 91-110 kayıt aralığı tahminleri.	57
Şekil 4.8: Rastgele orman yönteminde 111-130 kayıt aralığı tahminleri.	57
Şekil 4.9: Rastgele orman yönteminde 131-149 kayıt aralığı tahminleri.	58
Şekil 4.10: Naive bayes yönteminde 1-30 kayıt aralığı tahminleri.	61
Şekil 4.11: Naive bayes yönteminde 31-60 kayıt aralığı tahminleri.	62
Şekil 4.12: Naive bayes yönteminde 61-90 kayıt aralığı tahminleri.	62
Şekil 4.13: Naive bayes yönteminde 91-110 kayıt aralığı tahminleri.	63
Şekil 4.14: Naive bayes yönteminde 111-130 kayıt aralığı tahminleri.	64
Şekil 4.15: Naive bayes yönteminde 131-148 kayıt aralığı tahminleri.	64
Şekil 4.16: Destek vektör makineleri yönteminde 1-30 kayıt aralığı tahminleri.....	67
Şekil 4.17: Destek vektör makineleri yönteminde 31-60 kayıt aralığı tahminleri.....	68
Şekil 4.18: Destek vektör makineleri yönteminde 61-90 kayıt aralığı tahminleri.....	69
Şekil 4.19: Destek vektör makineleri yönteminde 91-110 kayıt aralığı tahminleri.	69
Şekil 4.20: Destek vektör makineleri yönteminde 111-130 kayıt aralığı tahminleri.	70
Şekil 4.21: Destek vektör makineleri yönteminde 131-150 kayıt aralığı tahminleri.	71

TABLO LİSTESİ

Sayfa No

Tablo 2.1: Veri madenciliği model, teknik ve yöntem örnekleri.....	10
Tablo 2.2: Veri madenciliği uygulama alanları.	13
Tablo 3.1: Veri setindeki nitelikler ve açıklamaları.	39
Tablo 3.2: Temel bilgisayar uygulamaları haftalık ders akışı.	40
Tablo 3.3: Ders videolarının içerik ve süre bilgileri.....	41
Tablo 4.1: Öğrencilerin sınıflarına göre frekans dağılımı.	45
Tablo 4.2: Öğrencilerin cinsiyetlerine göre frekans dağılımları.....	46
Tablo 4.3: Öğrencilerin üniversite yerleştirme puan türlerine göre frekans dağılımları.	47
Tablo 4.4: Öğrencilerin dönem sonu başarı düzeyine göre frekans dağılımları.....	48
Tablo 4.5: Öğrencilerin yerleştikleri puan türlerine göre başarı durumlarının frekans dağılımları.....	49
Tablo 4.6: Örnek hata matrisi.	51
Tablo 4.7: Rastgele orman yönteminde veri dengesizliğine ait sayılar.	52
Tablo 4.8: Rastgele orman yönteminde veri setlerinin başarı düzeyine göre dağılımı.	52
Tablo 4.9: Rastgele orman yönteminde hata matrisi görünümü.....	53
Tablo 4.10: Rastgele orman yöntemine ait sonuç değerleri.	53
Tablo 4.11: Naive bayes yönteminde veri dengesizliğine ait sayılar.	58
Tablo 4.12: Naive bayes yönteminde veri setlerinin başarı düzeyine göre dağılımı.....	59
Tablo 4.13: Naive bayes yönteminde hata matrisi görünümü.	59
Tablo 4.14: Naive bayes yöntemine ait sonuç değerleri.....	60
Tablo 4.15: Destek vektör makineleri yönteminde veri dengesizliğine ait sayılar.	65
Tablo 4.16: Destek vektör makineleri yönteminde veri setlerinin başarı düzeyine göre dağılımı.....	65
Tablo 4.17: Destek vektör makineleri yönteminde hata matrisi görünümü.	66

Tablo 4.18: Destek vektör makineleri yöntemine ait sonuç değerleri.	66
Tablo 4.19: Açılan farklı oturum sayısına ait açıklayıcı analiz sonuçları.	71
Tablo 4.20: Açılan farklı oturum sayısına ait Mann-Whitney u testi sonuçları.	72
Tablo 4.21: Açılan farklı oturum sayısına ait p değeri.	72
Tablo 4.22: Toplam etkileşim sayısına ait açıklayıcı analiz sonuçları.	73
Tablo 4.23: Toplam etkileşim sayısına ait Mann-Whitney testi sonuçları.	73
Tablo 4.24: Toplam etkileşim sayısına ait p değeri.	74
Tablo 4.25: Video izlerken harcanan zamana ait açıklayıcı analiz sonuçları.	75
Tablo 4.26: Video izlerken harcanan zamana ait Mann-Whitney testi sonuçları.	75
Tablo 4.27: Video izlerken harcanan zamana ait p değeri.	76
Tablo 4.28: Görüntülenen farklı video sayısına ait açıklayıcı analiz sonuçları.	76
Tablo 4.29: Görüntülenen farklı video sayısına ait Mann-Whitney testi sonuçları.	77
Tablo 4.30: Görüntülenen farklı video sayısına ait p değeri.	77
Tablo 4.31: Video yükleme sayısına ait açıklayıcı analiz sonuçları.	78
Tablo 4.32: Video yükleme sayısına ait Mann-Whitney testi sonuçları.	78
Tablo 4.33: Video yükleme sayısına ait p değeri.	79
Tablo 4.34: Video oynatma sayısına ait açıklayıcı analiz sonuçları.	79
Tablo 4.35: Video oynatma sayısına ait Mann-Whitney testi sonuçları.	80
Tablo 4.36: Video oynatma sayısına ait p değeri.	80
Tablo 4.37: Video duraklatma sayısına ait açıklayıcı analiz sonuçları.	81
Tablo 4.38: Video duraklatma sayısına ait Mann-Whitney testi sonuçları.	81
Tablo 4.39: Video duraklatma sayısına ait p değeri.	82
Tablo 4.40: Videonun sonuna gelme sayısına ait açıklayıcı analiz sonuçları.	83
Tablo 4.41: Videonun sonuna gelme sayısına ait Mann-Whitney testi sonuçları.	83
Tablo 4.42: Videonun sonuna gelme sayısına ait p değeri.	84
Tablo 4.43: Toplam atlama yapma sayısına ait açıklayıcı analiz sonuçları.	84
Tablo 4.44: Toplam atlama yapma sayısına ait Mann-Whitney testi sonuçları.	85

Tablo 4.45: Toplam atlama yapma sayısına ait p değeri.	85
Tablo 4.46: İleri atlama yapma sayısına ait açıklayıcı analiz sonuçları.	86
Tablo 4.47: İleri atlama yapma sayısına ait Mann-Whitney testi sonuçları.	87
Tablo 4.48: İleri atlama yapma sayısına ait p değeri.	87
Tablo 4.49: Geri atlama yapma sayısına ait açıklayıcı analiz sonuçları.	88
Tablo 4.50: Geri atlama yapma sayısına ait Mann-Whitney testi sonuçları.	88
Tablo 4.51: Geri atlama yapma sayısına ait p değeri.	89



KISALTMA LİSTESİ

Kısaltmalar	Açıklama
DN	: Doğru Tahmin Edilmiş Negatif Örnekler
DP	: Doğru Tahmin Edilmiş Pozitif Örnekler
KÖSE	: Kolb Öğrenme Stilleri Envanteri
MCC	: Matthews Korelasyon Katsayısı
MVA	: Moodle Video Analitik
SMOTE	: Synthetic Minority Over-sampling Technique
STEM	: Fen, Teknoloji, Mühendislik ve Matematik
YN	: Yanlış Tahmin Edilmiş Negatif Örnekler
YP	: Yanlış Tahmin Edilmiş Pozitif Örnekler

ÖZET

DOKTORA TEZİ

TERS YÜZ SINIFLARDAKİ VIDEO İZLEME DAVRANIŞLARI İNCELENEREK VERİ MADENCİLİĞİ İLE BAŞARININ TAHMİN EDİLMESİ

Kadir Burak OLGUN

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Enformatik Anabilim Dalı

Danışman : Prof. Dr. Sevinç GÜLSEÇEN

II. Danışman : Dr. Alper BAYAZIT

Bu tez çalışmasında ters yüz sınıflardaki video izleme davranışlarına ait veriler kullanılarak öğrencilerin başarı performansları veri madenciliği yöntemlerinden olan sınıflandırma yöntemleri ile modellenmesi amaçlanmıştır. Başarı tahmini öğrenme ortamının daha verimli tasarlanabilmesi ve bireysel dönütlerin verilebilmesi açısından önem arz etmektedir. Çevrimiçi ortamlardan gelen büyük miktardaki veriler kullanılarak gerçekleştirilen tahminler ise eğitsel veri madenciliği alanında oldukça popüler hale gelmiştir.

Çalışmada 4 dönem boyunca Temel Bilgisayar Uygulamaları dersine kayıtlı olan farklı bölümlerdeki toplam 404 üniversite öğrencisine ait video izleme verileri kullanılmıştır. Öğrenciler 14 haftalık dönem boyunca yüz yüze eğitim öncesinde araştırmacı tarafından hazırlanan ortalama 15’şer dakikalık ders içeriği ile uyumlu 11 videoyu Moodle öğrenme yönetim sistemi üzerinden izlemişlerdir. Öğrencilerin atlama yapma sayısı, oturum açma sayısı, farklı video görüntüleme sayısı vb. verilerinin kaydedilmesi için öğrenme yönetim sistemi üzerine kurulan bir araçtan yararlanılmıştır. Dijital ortamdan toplanan 11 adet nitelik dışında 6 adet demografik bilgi de anket yöntemi ile toplanmış, veri setinde 17 adet nitelik elde edilmiştir. Öğrencilerin başarı performansı ise başarılı ve başarısız olarak belirlenmiştir.

Verilerdeki gizli örüntüleri açığa çıkarmak için kullanılan ve alandaki başarı tahmini çalışmalarında en çok tercih edilen sınıflandırma algoritmalarından olan Rastgele Orman,

Destek Vektör Makineleri ve Naive Bayes algoritmaları analiz aşamasında kullanılarak araştırmanın ana problemine yanıt aranmıştır. Veri setinin dengesiz olması nedeniyle SMOTE tekniği kullanılarak sentetik veriler oluşturulmuş ayrıca çapraz geçerlilik yöntemi de kullanılmıştır.

Elde edilen sonuçlar incelendiğinde ders başarı performanslarının yüksek doğruluk oranları ile tahmin edilebileceği ortaya konmuştur. En başarılı sınıflandırma oranına sahip algoritma toplamda %83,11 doğruluk oranıyla Rastgele Orman algoritması olmuştur. Destek Vektör Makineleri, Naive Bayes ve Rastgele Orman algoritmalarına oranla oldukça düşük doğruluk oranlarıyla tahmin yapabilmektedir. Başarısız öğrencilerin dönem sonu sınavına girmeden önce doğru tahmin edilebilmesinin, başarı düzeyini arttırabileceği düşünüldüğünde başarılı öğrencileri tahmin etmeye kıyasla daha önemli olduğu görülmektedir. Rastgele Orman algoritmasının başarısız öğrencilerin tahmin edilmesinde %87,67 ile kullanılan diğer algoritmalarından oldukça yüksek bir doğru tahmin oranı yakaladığı tespit edilmiştir. |

Şubat 2021, |126| sayfa.

Anahtar kelimeler: | Akademik başarı tahmini, eğitsel veri madenciliği, ters yüz sınıf, çevrimiçi öğrenme ortamı, video izleme davranışları |

SUMMARY

[Ph.D. THESIS]

**[PREDICTING SUCCESS BY EXAMINING THE VIDEO WATCHING
BEHAVIORS IN FLIPPED CLASSROOMS USING DATA MINING]**

[Kadir Burak OLGUN]

İstanbul University

Institute of Graduate Studies in Sciences

[Department of Informatics]

Supervisor : [Prof. Dr.] [Sevinç GÜLSEÇEN]

[Co-Supervisor : Dr. Alper BAYAZIT]

[In this thesis, it was aimed to model students' success levels via the classification methods, which are included in data mining methods, by using the data of video watching behaviors in the flipped classrooms. Prediction of success is important to design the learning environment more efficiently and to give individual feedback. Estimates made by using large amounts of data from online environments have become very popular in educational data mining.

In the present study, video watching data were obtained from 404 university students in different departments enrolled in Basic Computer Applications course for 4 semesters. Before the face-to-face education, the researcher prepared 11 videos compatible with course content with the average of 15 minutes. During the 14-week course period, students watched these videos through the Moodle learning management system. To record students' data related to number of jumping, number of logins, number of different video views etc., a tool built on the learning management system was used. 11 qualities collected from digital environment and 6 demographic information collected via the survey method, totally, 17 qualities were obtained in data set. Success performance of the students were determined as successful and unsuccessful.

Random Forest, Support Vector Machines and Naive Bayes algorithms, which are among the most preferred classification algorithms in success prediction studies in the field to reveal hidden patterns in the data, were used in the analysis phase to answer the main problem of the study. Since the data set was unstable, synthetic data were created via the smote technique, and cross validation method was also used. When the obtained results were examined, it was revealed that the course success could be predicted with a high accuracy rate.

The algorithm with the most successful classification rate was the Random Forest algorithm with a total accuracy rate of 83.11%. Support Vector Machines were able to predict with a very low accuracy rate compared to Naive Bayes and Random Forest algorithms. Considering the fact that success levels of unsuccessful students can be increased before final exams, prediction of unsuccessful students is thought to be more important than prediction of successful students. It was determined that the Random Forest algorithm with a accuracy rate of 87,67% in predicting unsuccessful students achieved a much higher correct prediction than other algorithms used. |

Feburary 2021, |126| pages.

Keywords: | Academic success prediction, educational data mining, flipped classroom, online learning environment, video watching behaviors |

1. GİRİŞ

Çalışmanın bu başlığı altında araştırma problemi açıklanacak, araştırmanın amacı ve önemi incelenecek, araştırmadaki varsayımlar ve sınırlılıklara değinilecektir.

1.1. ARAŞTIRMA PROBLEMİ

Bu araştırmada problem; “Öğrencilerin akademik başarıları video izleme davranışlarından elde edilen veriler ile tahmin edilebilir mi?” sorusudur. Ana problemten uzaklaşılarak cinsiyet farklılığı ya da tek başına başka bir nitelikten elde edilen veriler ile öğrenci başarısının tahmin edilmesi üzerinde durulmamış fakat her bir niteliğin akademik başarı ile ilişkisi alt problemler olarak incelenmiştir.

1.1.1. Alt Problemler

- Açılan farklı oturum sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- Toplam etkileşim sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- Video izlerken harcanan zaman ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- Görüntülenen farklı video sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- Video yükleme sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- Video oynatma sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- Video duraklatma sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- Videonun sonuna gelme sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- Toplam atlama yapma sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- İleri atlama yapma sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?
- Geri atlama yapma sayısı ile akademik başarı arasında anlamlı bir ilişki var mıdır?

1.2. ARAŞTIRMANIN AMACI VE ÖNEMİ

İnsanoğlu uzun yıllardır yer yüzünde ekolojik ayak izi bırakmakta ve bunların etkilerini ileriki yıllarda fazlasıyla yaşamaktadır. Teknolojinin ortaya çıkışı ile beraber artık dijital ayak izi kavramı da konuşulmaktadır. Gelişmeler bu ayak izlerinin de etkilerinin olumsuz olaylarla

insanların karşısına çıkabileceği gibi neden bu ayak izlerinden faydalı bilgiler kazanılmasını ki sorusunu bilim insanlarına sordurmuştur. Bu kapsamda insanların dijital ayak izlerinin kayıtları araştırılmaya başlanmış ve verinin önemi büyük ölçüde artmıştır.

Özellikle teknoloji şirketleri verinin önemini kavramada hızlı davranarak insanların dijital ortamlardaki davranış ve etkileşimlerini kayıt altına almakta, elde edilen veriler ile veri madenciliği yöntemlerini kullanarak müşterilerinin veya kullanıcılarının ilgi ve meraklarını tespit etmeye çalışmaktadırlar. Ayrıca sosyal medya üzerinden elde edilen veriler ile ilgili birçok çalışma yapılmakta ve hatta bu verilerin siyaset, alışveriş ve reklam sektörü gibi birçok alanda kullanıcıları yönlendirme amacıyla kullanıldığı bazı gazete ve blog yazarları tarafından iddia edilmektedir (BBC, 2018; Demirci, 2020; Kulaklı, 2020; TRT, 2019).

Şirketlerin reklam ve pazarlama stratejileri içerisine dahil ettiği dijital ayak izlerinin depolandığı veri tabanı ve veri ambarları eğitim alanında da kullanılmak istenmiştir. Fakat buradaki en büyük zorluk öğrencilerin öğrenim süreçlerinin çoğunlukla dijital ortamda yürütülmemesinden dolayı yeterli veriye ulaşamamasıdır. Bu yüzden yapılan çalışmalarda (Akçapınar, et al., 2019; Akçapınar & Bayazıt, 2019; Asif, et al., 2017; Flanagan, et al., 2019; Kokoç, et al., 2018; Akçapınar, et al., 2015) araştırmacılar genellikle öğrenim sürecini taşıdıkları dijital ortamlardan elde ettikleri kısıtlı sayıdaki verilerden faydalanarak veri madenciliği yöntemlerini eğitimde kullanmaya başlamakta ve bu yöntemlerden elde ettikleri sonuçlar ile öğrenme süreçlerini iyileştirmeden, akademik başarının tahmin edilmesi ve artırılması gibi alanlarda gelişmeler ortaya koymaktadırlar. Öğrencilerin dijital ayak izleri yerine beyanlarından elde edilen anket sonuçlarının da veri madenciliği yöntemleriyle değerlendirilerek eğitim alanında sonuçlara ulaşmaya çalışıldığı araştırmalar da mevcuttur (Devasia, et al., 2016; Şengür & Tekin, 2013). Fakat Pena-Ayala (2018) öğrencilerin başarısını tahmin etmede sadece eğilimlerinin ve demografik özelliklerinin yeterli olmayacağını ayrıca öğrenme yönetim sistemlerinden elde edilen etkileşim verilerinin de dikkate alınması gerektiğini vurgulamaktadır.

Romero ve Ventura (2007) yaptıkları araştırmada 1995 ile 2005 yılları arasındaki eğitsel veri madenciliği (Educational Data Mining) alanında yapılan çalışmaları ve araçları incelemiş ayrıca eğitsel veri madenciliğinin yeterince olgunlaşmış bir kavram haline gelebilmesi için daha fazla çalışma yapılması gerektiğini vurgulamışlardır. Özellikle öğrencilerin akademik performanslarının modellenmesi veri madenciliğinin eğitim alanında en popüler

uygulamalarındandır (Bousbia & Belamri, 2014). Eğitsel veri madenciliği alanında yapılan çalışmalar incelendiğinde genellikle çok fazla sınırlılık içerisinde yapıldıkları gözlemlenmiştir. Ayrıca özellikle veri tabanlarındaki hazır veriler kullanılmayıp araştırmacılar tarafından düzenli veri toplanarak yapılan araştırmalarda düşük sayılardaki veri seti ve niteliklerin de sayısının az oluşu, verilerin toplanmasının uzun zaman alması zorluklar çıkarmakta, çalışmanın sonuçları için doğru analizlerin yapılmasını etkilemektedir.

Buradan yola çıkarak araştırmacı, uzun süreli veriler toplayarak öğrenciler hakkında daha detaylı bilgiler elde etmeyi, özellik (nitelik) sayısının önceki araştırmalara göre daha fazla olmasını ve dijital ortamdan elde edilen veriler ile sonraki araştırmalar için verilerin toplanabilirliğini kolaylaştırmayı hedeflemektedir. Alandaki çalışmalar incelendiğinde ve günümüzde uzaktan eğitimin de öneminin artması göz önünde bulundurulduğunda öğrencinin akademik başarısının tahmin edilmesine video izleme davranışlarının etkileri merak edilerek, bu araştırmada elde edilecek etkileşim verileri kullanılarak öğrencilerin başarı performanslarının tahmin edilip edilemeyeceğinin sonuç olarak ortaya konması amaçlanmıştır.

1.3. VARSAYIMLAR

Bu araştırmada aşağıdaki maddeler varsayım olarak kabul edilmiştir.

- Seçilen örneklem evreni temsil etmektedir.
- Araştırmaya katılan tüm öğrenciler uygulanan ölçeği dürüstlük ve içtenlikle cevaplamış, video izlerken doğal bir biçimde etkileşimlerde bulunmuşlardır.
- Elde edilen veriler öğrencilerin gerçek düzeylerini ve doğal etkileşimlerini temsil etmektedir.

1.4. SINIRLILIKLAR

Araştırmadaki sınırlılıklar aşağıda maddeler halinde listelenmiştir.

- Araştırma İstanbul İlindeki Yeditepe Üniversitesi'nde farklı bölümlere kayıtlı 404 öğrenci ile sınırlıdır.
- Araştırmada toplanan veriler Temel Bilgisayar Becerileri isimli dersin konuları hakkında çekilmiş videolardan elde edilen etkileşimler ile sınırlıdır.

2. GENEL KISIMLAR

Bu bölümde araştırmanın temel konularını oluşturan kavram ve açıklamalar ile araştırma konusu ile ilgili literatür taramalarına yer verilmiş olup her biri alt başlıklar halinde incelenmiştir.

2.1. VERİ MADENCİLİĞİ

2.1.1. Veri, Enformasyon ve Bilgi

İçinde bulunduğumuz çağa birçok isim verilmesinin yanında özellikle bilgi çağı olarak adlandırıldığı bilinmektedir. Hewitt(2019)'e göre bilgi çağı bilgisayarın yaygınlaşması sonucu bilgi ve verilere hızlı erişimin günlük yaşamda büyük değişimler yarattığı bir dönem olarak tanıtılmıştır. Bunun yanında bilgi toplumu kavramında bilginin çoğalması ve yayılmasının paylaşmak olduğu ve bu çoğalmanın bilginin yönetilmesini karmaşıktırmasından dolayı bilgi yönetimi konusu gündeme gelmektedir (Gülseçen, 2014). Bilgi çağında bilginin nasıl elde edildiği irdelendiğinde farklı sınıflandırmalar bulunmaktadır, fakat birçok araştırmacı (Kanehisa, et al., 2014; Ok, 2013; Bosancic, 2020) bilginin enformasyondan onunsa veriden geldiğini savunmaktadır. Şekil 2.1'de bilgi hiyerarşisi görülmektedir.



Şekil 2.1: Bilgi hiyerarşisi.

Hiyerarşiye göre ilk olarak elde olan veridir, veri bir takım işlemlerden geçirildikten sonra enformasyona ve sonrasında bilgiye dönüşmektedir. Bu hiyerarşik şekle ayrıca bilgi piramidi de denmektedir. Bilgiye ulaşmak için eldeki verilerin organize edilerek enformasyona dönüştürülmesi ve ardından bu enformasyonun anlamlandırılması gerekmektedir. Gülseçen (2014) bilginin insanlar tarafından yorumlanan veri ve enformasyon olduğunu söylerken veri ve enformasyonun ise bilgisayarlar tarafından insanlara sunulduğunu belirtmektedir. İşlenmemiş ve bu haliyle hiçbir anlam ifade etmeyen her bir bilgi veridir. Kelleher & Tierney (2018)'e göre veri gerçek dünyadaki bir varlığın özellikleri ile beraber soyutlanmasıdır. Duan ve Diğerleri (2017) verilerin makinelere özgü olmadığını, dağınık elementlerin ve kavramların toplanmasından oluştuğunu savunmaktadır. Veri gözlemleri, gerçekleri ve ham enformasyonları içinde barındırırken işlenmediği takdirde neredeyse hiçbir anlam ifade etmemektedir (Monino & Sedkaoui, 2016). Hewitt (2019) veriyi her hangi bir insanın bunları kullanmasıyla ya da başka enformasyonlar ile ilişkiye sokulmasıyla bilgiye sentezlenen yorumlanmaya açık potansiyel değere sahip olan hammaddelere benzetmektedir. Zins (2007) 57 bilim insanının katılımıyla yaptığı çalışmada her bir katılımcıdan veri, enformasyon ve bilgi için farklı tanımlar almış fakat genel olarak veriyi; ampirik algıları temsil eden işaret kümeleri, enformasyonu; ampirik bilgiyi temsil eden bir işaret kümesi ve bilgiyi; kişinin doğru olduğuna inandığı düşüncelerin anlam ya da içeriğini temsil eden bir işaret kümesi olarak tanımlamaktadır. Verinin tanımları incelendiğinde enformasyona dönüşebilmesi için organize edilmesi, bir araya getirilmesi ve bir anlam ifade etmesi gerekmektedir.

Enformasyon için işlenmiş ve organize edilmiş, anlamlı bir hale getirilmiş veri setleri denebilir. Hewitt (2019) enformasyonu bir araya getirilmiş telefon numaralarından oluşan ve adlandırılmış bir rehberi örnek göstererek, herhangi bir insan tarafından kullanılan veri olarak tanımlamaktadır. Badia (2014)'ya göre enformasyon genellikle anlamlandırılmış veri olarak tanımlanmaktadır. Enformasyonun veri tabanları gibi alanlarda düzenlenmiş ve ilişkili bir biçimde anlamlandırılmış olarak tutulan verilerden oluştuğu iddia edilmektedir (Duan, et al., 2019). Kettinger & Li (2010) enformasyonun veriden gelip bilgiye ulaştığı hiyerarşinin aksine bir hiyerarşik yapı olmadığını ve enformasyonun hem verinin hem de bilginin ortak bir işlevi olduğunu savunmaktadır.

Bilginin ne anlam ifade ettiği eskilerden günümüze kadar tartışılan felsefi bir soruyken varoluştan bu yana insanların kararlarında ve davranışlarında etkin olan, akıllarının erebildiği

olgu, gerçek ve ilkelerin tamamıdır denebilir (Gölseçen, 2014). Bilginin insanların algılayabileceği bir kavram olduğu kanısına varmak mümkündür fakat benimsenmiş bir bilginin doğruluğu ve sürekliliği tartışılabilir. Gölseçen (2014)'e göre bilgi var olduğu dönemde ve mekanda doğru olarak kabul görebilir fakat daha sonra elde edilen verilerden oluşan enformasyonlar ile sentezlenerek güncellenmeli ya da değiştirilmelidir. Yani bilgi değişkendir denebilir ve bu değişkenlik yeni bilgilerin oluşmasına katkı sağlamaktadır. Bilgi kelimesinin Yunan kökenli Latince bir sözcük olduğu ve eski dönemlerde “informo” olarak biyoloji, ahlak ve pedagoji alanlarında kullanılan “information” sözcüğünün anlamı zaman içerisinde değişime uğrayarak Orta Çağ döneminde şekil verme, daha sonra ise bir şeyin başka bir bireye iletilmesi olarak değişmiştir (Uçak, 2010).

Felsefi açıdan bakıldığında bilgi kesinlik gerektirir ve insanların ispatlanmış inançlarıdır (Denkel, 1998). Kettinger & Li (2010) verinin bir durumun açıklanmasına atıfta bulunduğunu ve bilginin bunların altında yatan kavramları anlaşılabilir bir biçimde ilişkisel olarak temsil ettiğini iddia etmektedirler. Bir başka araştırmacı ise bilgiyi çoğunlukla insan çabasının bir unsuru olarak ifade etse de aslen bilimin temel amacı olduğunu belirtmektedir (Hewitt, 2019). Szczerbicki & diğerleri (2020) bilginin deneyimlerden ve paylaşımdan çıkarılabildiğini ayrıca saklanabilir, geliştirilebilir ve yeniden kullanılabilir olduğunu aktarmaktadır. Buradan yola çıkılarak bilginin farklı biçimlerde kategorilere ayrıldığı kanısına varılabilmektedir. Bilginin kategorizasyonu incelendiğinde temelde 3 farklı biçimde kategorize edildiği görülmektedir. Bunlar:

- İşlemsel(Prosedürel-Prosedürel) Bilgi ve Asli(Kavramsal) Bilgi
- Örtük ve Açık Bilgi
- Genel ve Özel Bilgi

Kavramların anlaşılmasında ve problem çözme becerisinin geliştirilmesinde hem asli hem de işlemsel bilgiye ihtiyaç vardır (Wuryaningrum, et al., 2020). Canobi (2009)'ye göre işlemsel bilgi, problem çözmek için ihtiyaç duyulan becerilerle ilişkili ve bir dizi adım izleme yeteneği içermektedir. İşlemsel bilgi belirli bir hedefe ulaşmak için izlenen adımlar dizisi ya da eylemler olarak tanımlanabilir (Lenz, et al., 2019).

Zander ve Kogut (1995)'e göre asli bilgi teorik niteliklidir ve işlemsel bilgiye oranla daha güçlü bir bilgi türüdür. Asli bilgi, öğrenenin rapor edebildiği ve bilinçli olarak farkında olduğu bir

açık bilgi olarak tanımlanabilir (Sun & Li, 2019). Haye ve Torres-Sahli (2017)'ye göre asli bilgi bir şeyi bilmek, işlemsel bilgi ise bir şeyin nasıl yapılacağını bilmek şeklinde yorumlanmaktadır. Örnek ile açıklamak gerekirse, asli bilgi üretilen bir arabanın her bir parçasının performansını nasıl etkileyeceği bilgisi iken işlemsel bilgi bu parçaların her birinin nasıl bir araya getirileceğinin bilgisidir (Ok, 2013).

Kişinin bilincindeki bilgiler açık bilgi, bilinç dışı günlük rutinler gibi bilgi ve beceriler örtük bilgiyi temsil etmektedir (Yalçın, et al., 2015). İfade edilebilen, başka bireylere aktarılabilen ve kayıt altına alınabilen bilgi açık bilgidir (Uçak, 2010). Lim ve Diğerleri (1999)'ne göre örtük bilgi kişinin kendi içinde olan yetenekleri ve sezilerinden oluşurken kolaylıkla paylaşılamamaktadır. Örtük bilgi dönüşüm geçirebilir, birey tarafından başka biriyle paylaşılabilirdiği zaman açık bilgi halini alır. Örtük bilgi insan beyninde olduğundan ve özelliklerinden dolayı erişilmesi, paylaşılması ve yönetilmesi oldukça zordur (Yıldırım & Özdemirci, 2019). Bu tanımlardan yola çıkarak örtük bilginin bireylerin kişisel tecrübeleri sonucu elde edildiği söylenebilir.

Genel bilgi ve özel bilgiyi birbirinden ayıran en önemli farklılığın transfer sürecindeki zorluk seviyesi olduğu söylenebilir. Ok (2013)'a göre genel bilgi bireylerin büyük bir çoğunluğunun bildiği ve bireyler arasında transferinin oldukça kolay olduğu bir bilgi türüdür. Örneğin, bir teknoloji sınıfındaki neredeyse tüm öğrencilerin bilgisayarın nasıl açılacağını bilmesi genel bilgidir.

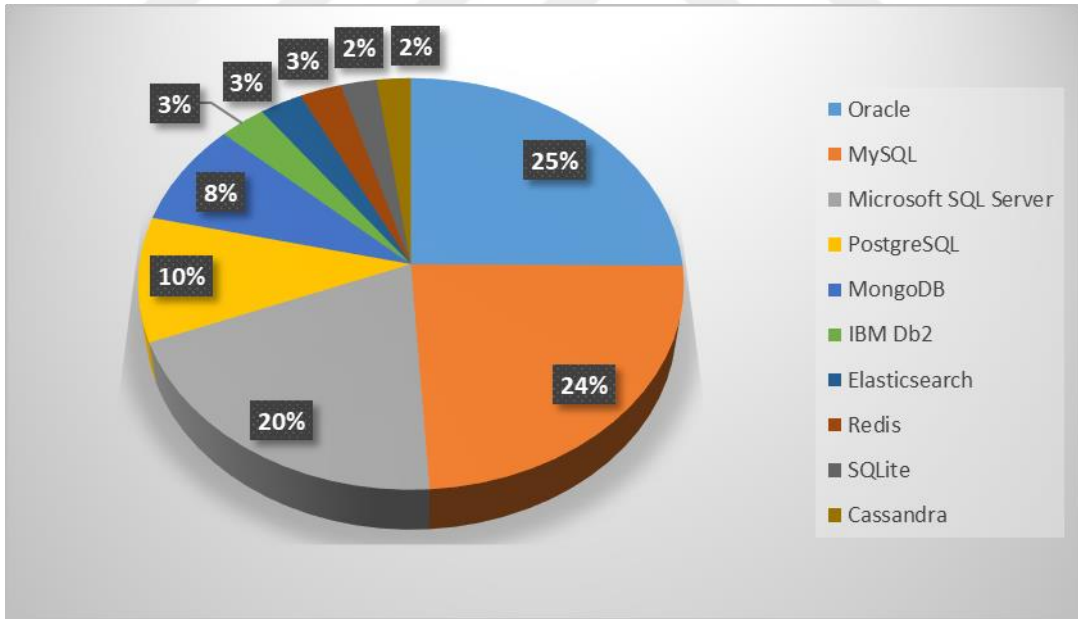
Özel bilgiyi Uçak (2010) alanında uzman bireylerin görüş, düşünce ve deneyimlerini birbirlerine aktarırken elde edilen ve değer ifade eden bilgi olarak belirtmektedir. Özel bilgi genel bilginin aksine az sayıda bireyin sahip olduğu ve bireyler arası aktarımın zor olduğu bilgi türüdür (Ok, 2013). Özel bilgiye örnek olarak alanında uzman bir bilişim teknolojileri öğretmenin anaokulu seviyesindeki öğrencilere bilgisayarın çalışma prensibini nasıl anlatması gerektiğini bilmesini gösterebiliriz.

2.1.2. Veri Tabanı, Veri Ambarı ve Büyük Veri

Teknolojinin gelişmesi ile beraber özellikle İnternet teknolojileri sosyal medya, blog siteleri, mobil uygulamalar gibi birçok ortamdan veri toplayan uygulamalar ile tek başlarına anlam ifade etmeseler de birçok veriyi toplar hale gelmiş ve bu verilerin birikmesi ve birleştirilmesi ile ortaya çıkan yığınların çeşitliliğinin, hacimsel büyüklüğünün artması ile beraber yeni çözümler

üretilmeye başlanmış, büyük veri kavramı bu şekilde ortaya atılmıştır (Doğan & Arslantekin, 2016; Şahin & Yurdugül, 2020). Büyük veriden önce verilerin depolanması ve saklanması amacıyla kullanılan veri ambarları ve veri tabanları incelenebilir.

Kalıpsız (2001)'a göre veri tabanı belirli bir amaç ya da birden çok amaç için gerekli olan, bir sistem halinde düzenli biçimde, kütük olarak bir bellekte saklanan veri toplumdur. Önceleri verilerin tek bir tablo içerisinde depolanıyor olmasının ardından veriler üzerindeki ilişkiler karmaşıklaştıkça birden fazla tablo barındıran ilişkisel veri tabanları ortaya çıkmıştır. İlişkisel veri tabanları veri tabanı yönetim sistemlerinin gelişimini tetiklemiştir (Ferrier & Stangret, 1982). Batory ve diğerleri (1988) veri tabanlarının büyük hacimli verileri düzenlemek ve saklamak için kullanıldığını, bunların içerisindeki bulunduğu veri tabanı yönetim sistemlerinin ise oldukça düşük maliyetli araçlar olduğunu belirtmektedirler. Günümüzde veri tabanı yönetim sistemleri oldukça gelişmiş ve büyük hacimli verilerin hızlı, güvenli biçimde depolanarak yönetilmesine imkan vermektedir. Şekil 2.2'de Haziran 2020 itibariyle bazı en popüler veri tabanı yönetim sistemlerinin sıralaması yüzdelik paylar halinde verilmektedir (Liu, 2020).



Şekil 2.2: En popüler veri tabanı yönetim sistemleri sıralaması.

Büyük miktardaki ham verilerin doğrudan kullanımının mümkün olmayışından dolayı veriler temizleme, çıkarma ve dönüştürme gibi bazı ön işlemlerden geçirilir ve farklı bir biçim alarak veri ambarlarında saklanabilir (Qu & Jiang, 2019). Veri ambarı birbirlerinden ayrık olan veri

tabanı uygulamalarının bütünleştirildiği ve bu sayede özellikle kurumlarda karar destek sistemleri için alt yapı oluşturan bir ortam olarak tanımlanabilir (Doğan & Arslantekin, 2016).

Monino ve Sedkaoui (2016)'ye göre büyük veri kavramı herhangi bir kuruluşun yönetebileceği veri işleme, kullanma ve saklama açısından yeni çözümler gerektiren kritik büyüklüğe ulaştığında kullanılır ve hız, hacim ve çeşitlilik gibi özellikleri veri tabanını artık büyük veri olarak nitelendirmek için kriter olarak belirlenebilir. Büyük verinin verimli bir şekilde kullanımının önemi ortada iken yüksek hacim ve yüksek hız özelliklerinin daha da büyümesi verilerin bakımı ve kullanımını oldukça zorlaştırmaktadır (Qu & Jiang, 2019). Büyük verilerden söz edildiğinde kurumların bu verileri kullanarak yeni bilgiler elde etmesi için veri madenciliği kullanılabilmektedir.

2.1.3. Veri Madenciliği Teknikleri

Veri madenciliği enformasyon olarak görünen ya da sadece anlamsız veriler olarak bulunan büyük verilerden bilginin çıkarılması işleminden dolayı aynı zamanda bilgi keşfi olarak da adlandırılmaktadır. Aksoy ve Diğerleri (2020) ham verilerin kullanılabilir bilgiye dönüştürülmesi amacıyla kullanılan veri madenciliği kavramının ilk olarak veri tabanından bilgi keşfi olarak ortaya çıktığını belirtmişlerdir. Veri madenciliği bilgisayar bilimi ya da yapay zeka gibi disiplinlerdeki algoritmaları kullanarak ve tüm teknolojileri bir araya getirerek veri içerisindeki faydalı bilgiler ile bunlar arasındaki önemli ilişkileri ortaya çıkarmaya odaklanmaktadır (Monino & Sedkaoui, 2016). Berry ve Linoff (2004)'a göre veri madenciliği iş faaliyetlerinin öğrenmeye dayandırıldığı, bilgi doğrultusunda verilmiş kararların daha doğru olduğu ve sonuçlarının ölçülebilir olduğu bir tutum, teknoloji, süreç ve metodolojidir. Han ve diğerleri (2006) ise veri madenciliğini çok büyük miktarda veri içerisinde bilginin çıkarılması süreci olarak tanımlamaktadır. Bir başka çalışmada veri madenciliği, büyük bir veri havuzundaki verilerin analiz edilmesi ve derlenmesi ile aralarındaki ilişki ve örüntülerin ortaya çıkarılarak sonrasında faydalı bilgiye dönüştürülmesi süreci olarak açıklanmaktadır. (Razak, et al., 2018). Hicham ve diğerleri (2020) ise veri madenciliğini, son yıllarda sağlık hizmetleri, sigorta, finansal veri analiz, telekomünikasyon ve eğitim gibi çeşitli alanlarda karar vermeye yardımcı olan bilgi, kurallar, ilişkiler ve eğilimler keşfetmeyi sağlayan bir yöntem koleksiyonu olarak tanımlamaktadır.

Veri madenciliği modelleri tahmin edici ve tanımlayıcı modeller olarak ikiye ayrılmaktadır (Özçakır, 2013). Bu modellere ait bazı teknikler ve bu tekniklere ait bazı yöntemler Tablo 2.1 'de verilmektedir (Hand, et al., 2001; Özekes, 2003).

Tablo 2.1: Veri madenciliği model, teknik ve yöntem örnekleri.

Tahmin Edici Modeller		Tanımlayıcı Modeller	
Sınıflandırma	Regresyon	Kümeleme	Birliktelik Kuralları
Karar Ağaçları		Bölme Yöntemleri	Apriori Algoritması
Yapay Sinir Ağları		Hiyerarşik Yöntemler	Carma Algoritması
Genetik Algoritmalar		Yoğunluk Tabanlı Yöntemler	Sequence Algoritması
K-En Yakın Komşu		Izgara Tabanlı Yöntemler	Eclat Algoritması
Naive-Bayes		Model Tabanlı Yöntemler	GRI Algoritması
Bellek Temelli Nedenleme		FP-Growth	

Tahmin etme, bir değişkenin sadece diğer ilgili değişkenlerin açıklamasına dayalı olarak tanımlanmasını sağlayan bir veri madenciliği tekniğidir (Hicham, et al., 2020). Tahmin edici teknikler olan sınıflandırma ve regresyon veri madenciliğinde elde olan verilerden yola çıkarak sonraki gelecek olan veriler hakkında öngöründe bulunma amacıyla kullanılır denebilir. Han ve Diğerleri (2006)'ne göre bu iki yöntem önceki verileri değerlendirerek gelecek verilerin eğilimlerini tahmin ederken sınıflandırma kategorik, regresyon ise süreklilik gösteren veriler için kullanılmaktadır. Bu modeller için birçok teknik ya da algoritma bulunması kullanılan veri setlerine göre seçim yapılabilmesine imkan sağlamaktadır. Bunlar arasından kolay kullanım ve kolay yorumlanmaları açısından karar ağaçları en yaygın kullanılan tekniklerdendir (Özekes, 2003). Karar ağaçları iki aşamalı olarak çalışmaktadır. Birinci aşamada var olan veri seti analiz

edilerek algoritmanın öğrenme adımı gerçekleştirilir. İkinci aşamada ise algoritma bu öğrenme ile model geliştirir ve gelecek veri seti hakkında çıkarımda bulunur. Model oluşturulurken kurallar kullanılır ve bu kurallara uygunluğa göre tahminler gerçekleştirilir. Örneğin bir veri setindeki bilgilere göre eğitilmiş modelin kasko poliçesi risk tahminleri incelendiğinde;

EĞER $yas < 25$ VEYA $yas > 50$ VE $ehliyetYili < 2$ İSE $riskDurumu = Yüksek$

Şeklinde bir karar cümlesi ya da kurallar oluşturulmaktadır. Karar ağaçları bunlar gibi birçok kuralın artarda sıralanması ile dallanarak büyür ve her bir dallanma mutlaka bir sonuca çıkmaktadır.

Rastgele Orman algoritması sınıflandırma modelleri içerisinde oldukça başarılı ve sınıflandırma yaparken birden fazla karar ağacı ile bu işlemi gerçekleştirmesinden kaynaklı olarak birçok disiplinde kullanılabilen en çok tercih edilen algoritmalarından biridir (Akar & Güngör, 2012). Breiman (2001) Rastgele Orman algoritmasının her bir ağacı bağımsız olarak örneklediğini, ormandaki tüm ağaçlar için aynı dağılıma sahip rastgele bir vektörün değerleri kombinasyonu olduğunu, ormandaki ağaç sayısının arttıkça genelleme hatasının düştüğünü, AdaBoost ile benzerlikleri olabileceğini fakat daha sağlam ve güvenli hata oranları verdiğini iddia etmektedir. Akar ve Güngör (2012) uydu görüntülerinden elde edilen verileri kullanarak Rastgele Orman, Gentle AdaBoost, En Çok Benzerlik ve Destek Vektör Makineleri algoritmalarından elde ettikleri tahmin performanslarını kıyasladıklarında Rastgele Orman algoritmasının diğer algoritmalarından daha yüksek oranda doğru tahminde bulunduğunu saptamışlardır.

2.1.4. Veri Madenciliği Uygulama Alanları

Veri madenciliği çalışmaları günümüzde tahmin çıkarımı yapılmak istenen ve karar verme tekniklerinin uygulandığı birçok alanda gerçekleştirilmektedir. Ayrıca büyük miktarda veri üretilen herhangi bir alanda veri madenciliğinin kullanılması ve uygulanması mümkündür (Savaş, et al., 2012). Padhy ve Diğerleri (2012)'ye göre insanlar günlük yaşamlarında oldukça fazla ve çeşitli veriler üretmekte ve bu verilerin değerli olduğu artık fark edilerek gizli örüntülerin ortaya çıkarılması ve uzmanların atladığı tahmin bilgilerinin bulunması gibi işler için veri madenciliği kullanılmaktadır.

Özekes(2003)'e göre veri madenciliği işlevlerine göre sınıflama, kümeleme ve birliktelik kuralları olarak üç ana başlık altında, bankacılık sektöründe risk hesaplama, satış – pazarlama alanında müşteri potansiyellerini tahmin etme, insan kaynaklarında işe alım sürecinde karar verme, tıp alanında hastaların verileri ile karar verme, üretimde verimliliğin artırılması, biyoloji alanında bitki ve hayvanların özelliklerine göre sınıflandırılmaları, şehir planlamacılığında gruplandırma, alışveriş sektöründe market sepeti uygulaması gibi işlerde kullanılmaktadır. Veri madenciliğinin en çok kullanıldığı alanların tıp, biyoloji ve genetik alanları olduğu iddia edilmektedir (Savaş, et al., 2012).

Baykal (2006) veri madenciliğinin uygulama alanlarını maddeler halinde aşağıdaki gibi incelemiştir:

- Karar verme süreçleri
- Risk analizi
- Pazar araştırması
- Verimli kaynak kullanımı
- Geleceğe yönelik tahminde bulunma
- Benzerlik çalışmaları

Veri madenciliği perakende satışlar, biyoenformatik, terörle mücadele alanlarında oldukça yoğun biçimde kullanılmaktadır (Baker, 2010). Ayrıca Baker (2010) aynı çalışmasında veri madenciliğinin son yıllarda eğitim alanında da çok sık kullanılmaya başlandığını ve eğitsel veri madenciliği kavramının ortaya atıldığını belirtmektedir. Bunlar dışında veri madenciliğinin son yıllarda gizli dinleme, vergi kaçakçılığının tespiti gibi çeşitli hukuki süreçlerde de kullanıldığı görülmektedir (Savaş, et al., 2012). Özbay (2015) ise yaptığı çalışmaya göre Tablo 2.2'deki alanların başlıca uygulama alanları olduğunu belirtmiştir.

Tablo 2.2: Veri madenciliği uygulama alanları.

Uygulama Alanı	Alandaki Faydaları
Sağlık	Hastalıkların erken teşhis ve tedavisi, ilaç geliştirme ve yan etki tespitleri.
Endüstri ve Mühendislik	Üretim süreçlerinin takibi, kalite kontrol takibi, performans etkenlerinin belirlenmesi.
Kamu	Güvenlik sorunlarının tahmini ve kaynak yönetimi.
Pazarlama	Kar oranının artırılması, müşteri analizlerinin yapılması, satış tahminleri.
Bankacılık, Finans ve Borsa	Risk analizi, yatırım modellemesi, borsa tahminleri.
İnternet	İnternet güvenliği ve kişiselleştirilmiş web sayfaları
Eğitim	Öğrenci başarılarının tahmin edilmesi ve artırılması, ortamdaki aksaklıkların tespiti, verimli ve etkili ortamların geliştirilmesi.

Veri madenciliğinin uygulama alanlarının sınırlandırılması yanlış bir yaklaşım olacaktır. Verinin ortaya çıktığı ve saklandığı tüm alanların veri madenciliği uygulama alanı olarak anılabilmesi gerekmektedir (Özbay, 2015). Tablo 2.2 incelendiğinde belirtilen uygulama alanlarının oldukça geniş bir yelpazeyi kapsadığı görülmekte olup bu alanların teknolojik gelişmeler sonucu veri elde edilebilir yeni alanlar ile daha da genişletilmesi söz konusudur.

2.2. EĞİTSEL VERİ MADENCİLİĞİ

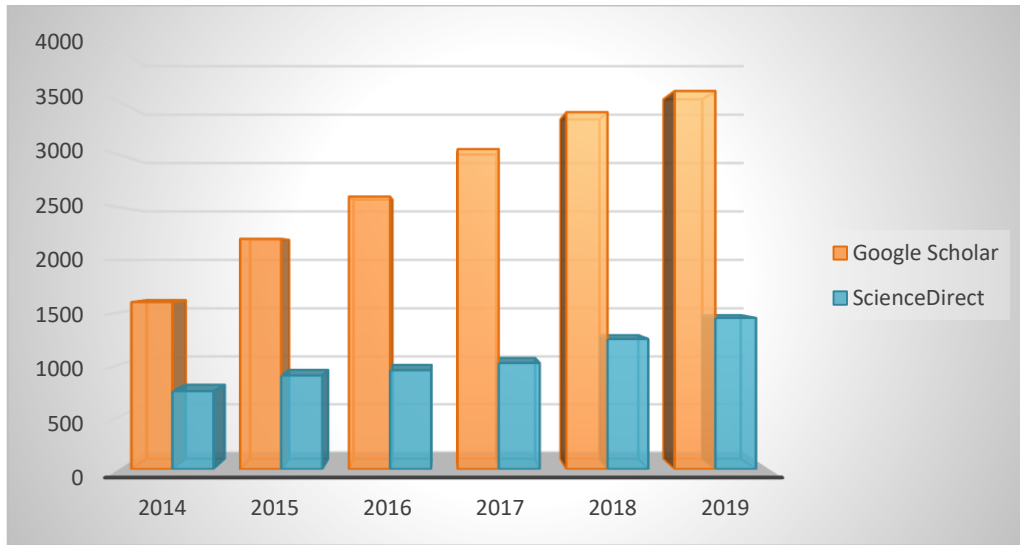
Son yıllarda özellikle açık eğitim kaynaklarının kullanımındaki artış öğrencilerin eğitimleri için en uygun materyallere ulaşması amacıyla bu kaynakların işlenmesi ve değerlendirilebilmesi için buradan gelen verilerin analizlerine yönelik talep oluşturmaktadır (Bulathwela, et al., 2020). Yu ve Diğerleri (2020), yükseköğretimde yaptıkları çalışma ile tahmin analizlerinin yöneticiler, öğretmenler ve öğrenciler gibi eğitim alanındaki farklı paydaşlar için uygulanabilir

bir takım iç görüler sağlayabileceğini ve bunun için öğrenme yönetim sistemlerinden elde edilen verilerin yeterli tahmin gücüne erişebileceğini belirtmişlerdir. Veri madenciliği, eğitim kalitesinin artırılmasına yardımcı olacak bulgular sağlayabilecek, eğitimde kullanılabilecek bilgisayar bilimlerinde bir alan olarak görülmektedir (Ktona, et al., 2014).

Veri madenciliği yöntemlerinin eğitim alanında kullanımı diğer alanlara kıyasla nispeten yeni sayılabilmektedir. Kurumların elindeki öğrenciler hakkındaki veriler artmaya başladıkça büyük veri havuzları oluşmuş ayrıca eğitimin İnternet ortamına taşınmaya başlaması ile dijital ortamlardan öğrenme-öğretme etkileşim verileri de elde edinmeye başlanmış, tüm bunların bir etkisi olarak da araştırmacıların ve eğitimcilerin arasında önemini günden güne arttıran eğitsel veri madenciliği kavramı kullanılmaya başlanmıştır (Güldal & Çakıcı, 2017; Romero & Ventura, 2007). Alandaki ilk uygulamanın üniversitelerindeki veri tabanından elde edilen öğrenci katılım verileri ile eğitimde kurumsal kararlar almayı hedefleyen Sanjeev ve Zytchow (1995) isimli araştırmacılar tarafından gerçekleştirildiği söylenebilir. Ardından yavaş bir ilerleme gösteren alanda, 2008 yılında Uluslararası Eğitsel Veri Madenciliği Çalışma Grubu (International Working Group in EDM) dünyanın birçok yerinden çok fazla araştırmacının dahil olduğu Uluslararası Eğitsel Veri Madenciliği Konferansı(International Conference on Educational Data Mining)'nı düzenlemiş ve günümüze kadar her yıl tekrarlanmasını sağlamış ayrıca hakemli bir dergi olan JEDM (Journal of Educational Data Mining)'yi kurarak alandaki araştırmacıların sayısının artmasına katkıda bulunmuştur (Olney, 2020; Romero & Ventura, 2010; Society, 2020). Merceron ve Yasef (2005), veri madenciliğinin eğitimde uygulanması işinin pedagojik bir bakış açısıyla yapılmasının önemini vurgulamış ve öğretme ve öğrenmeyi geliştirme amacıyla çok sayıda araştırmacının eğitsel veri madenciliği ile ilgilendiğini ancak analizlerin kolaylaştırılarak yaygınlaştırılması gerektiğini belirtmektedir. Eğitsel veri madenciliği öğrencilerin davranışlarını ve başarı performanslarını, alan bilgisi içerikleri, değerlendirme yöntemleri, eğitimsel fonksiyonlar ve uygulamalar gibi özellikleri karakterize etmek için kullanılan kalıplar oluşturmak ve tahminlerde bulunmak amacıyla kullanılmaktadır (Hicham, et al., 2020). Eğitim alanında özellikle performansın tahmin edilebilir olması oldukça önemlidir, son yıllarda araştırmacıların öğrencilerin öğrenmesini analiz etmek, eğitim sürecinin kalitesini iyileştirmek ve geniş bir bakış açısı elde etmek gibi amaçlar için veri madenciliği tekniklerini kullanmaları kaçınılmaz olmuştur (Razak, et al., 2018). Sınıflandırma, kümeleme, ilişkilendirme kuralları gibi veri madenciliği tekniklerinin birçok araştırmacı tarafından eğitim sürecinin iyileştirilmesi amacı güdülerek eğitim alanındaki verilere uygulandığı görülmektedir

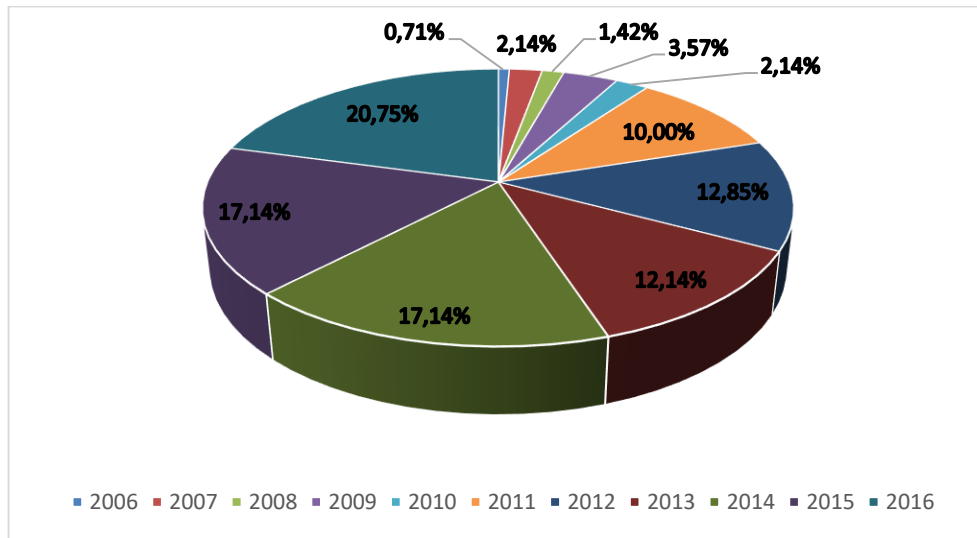
(Akçapınar, et al., 2015). Hicham ve diğerleri (2020) eğitsel veri madenciliğinin öğrencilerin nasıl öğrendiklerini anlamaya çalışan ve bunun için psikolojik yollar ile veri madenciliği tekniklerini kullanan bir araştırma ve analiz kavramı olarak yükselmekte olduğunu belirtmektedirler. Bu gelişmeler ele alındığında veri madenciliği yöntemlerinin eğitimde kullanılması eğitsel veri madenciliği alanının ortaya çıkmasına sebep olmuştur denebilir.

Alandaki birçok araştırmacı tarafından eğitsel veri madenciliği için ortak özellikleri olan tanımların yapıldığı görülmektedir. Baker (2010) eğitsel veri madenciliğini, eğitim ortamlarından toplanan eşsiz veri türlerini kullanarak öğrencilerin daha iyi öğrenme gerçekleştirmesi için yeni keşiflerin yapıldığı ve yöntemlerin geliştirildiği bir bilimsel araştırma alanı olarak tanımlamaktadır. Eğitsel veri madenciliği eğitimin her aşamasında oluşan büyük miktardaki verileri keşfederek, eğitim problemlerine çözüm aramak için analiz etme amacıyla yeni teknikler geliştirmeyi hedefleyen bir eğilimdir (Sachin & Vijay, 2012). Bir başka tanımda, eğitsel veri madenciliğinin eğitsel içeriklerden elde edilen benzersiz veri türlerinin keşfedilmesi için yöntemler geliştirilmesi ile ilgilenen bir disiplin olduğu vurgulanmaktadır (Romero, et al., 2008). Calders ve Pechenizkiy (2012) eğitsel veri madenciliğini, öğrencilerin nasıl öğrendiklerine ve öğrenme ortamlarına katkıda bulunan, eğitim verilerindeki artış ve kullanılabilirlik nedeniyle veri madenciliği için zengin bir uygulama alanı ve eğitim sistemlerinden elde edilen verileri araştırmak için yeni yöntem ve tekniklerin geliştirildiği çok disiplinli bir araştırma alanı olarak tanımlamaktadır. Eğitsel veri madenciliğinin amacı yalnızca verilerin bilgiye dönüştürülmesi değil, aynı zamanda öğrenmeyi iyileştirmek için eğitim ortamında nasıl değişiklikler yapılması gerektiğini ortaya çıkaran bilgileri filtrelemektir (Hicham, et al., 2020). Yapılan tanımlarda öğrenmenin daha verimli ve sorunsuz ortamlarda gerçekleşmesi amacı güdüldüğü ve eğitsel verilerin benzersizliği vurgulanmaktadır. Güldal ve Çakıcı(2017) eğitsel veri madenciliğinin ortaya çıkışının ve gelişme sürecinin henüz çok yeni olduğunu söylemektedir. Bu durum incelendiğinde yaklaşık 10 yıllık bir süreçte bile araştırmacıların birçoğu eğitsel veri madenciliğinin henüz yeni bir alan olduğunu ve gelişim sürecinde olduğunu belirtmektedir. Şekil 2.3 incelendiğinde yıllara göre eğitsel veri madenciliği alanındaki çalışma sayılarının grafiği görülmektedir (Hicham, et al., 2020).



Şekil 2.3: Eğitsel veri madenciliği alanında yapılan çalışmaların yıllara göre dağılımı.

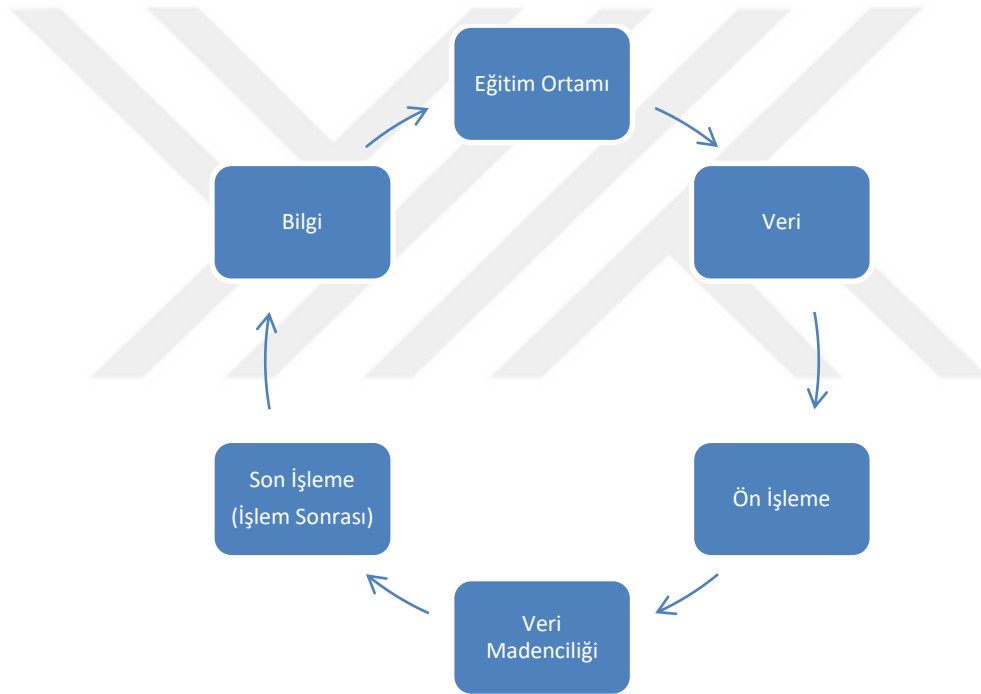
Tekin ve Öztekin (2018)'in yaptıkları çalışmada eğitsel veri madenciliği alanındaki çalışmaların %90'ının yabancı dergilerde %10'luk bir kısmının da yerli dergilerde yer aldığı, bu çalışmaların %34,3'nün nitel %65,7'sinin ise nicel çalışmalar olduğu ve %68,5 gibi büyük bir oranla çalışma grubunun üniversite öğrencilerinden oluştuğu ortaya çıkmıştır. Farklı veri tabanları taranarak yapılan araştırmada eğitsel veri madenciliği alanında yapılan çalışmaların yıllara göre yüzdelik dağılımı Şekil 2.4'te verilmiştir (Tekin & Öztekin, 2018).



Şekil 2.4: Eğitsel veri madenciliği alanındaki çalışmaların yıllara göre yüzdelik dağılımı.

Şekil 2.3 ve Şekil 2.4'e göre farklı veritabanları üzerinden yapılan taramalar sonucu eğitsel veri madenciliği ile ilgili yapılan çalışmaların sayısında artış görülmektedir. Bu artış eğitsel veri madenciliğinin araştırmacılar arasında popülaritesinin her geçen yıl yükseldiğini kanıtlamaktadır.

Romero ve Ventura (2010) eğitsel veri madenciliği sürecinin eğitim sistemlerinden elde edilen verilerin eğitim yazılımı geliştiriciler, öğretmenler ve eğitim araştırmacıları tarafından kullanılarak yararlı bilgelere dönüştürülmesi açısından diğer alanlarda uygulanan veri madenciliği süreçlerinden farklı olmadığını ve aynı adımlardan geçtiğini Şekil 2.5 ile açıklamaktadır.

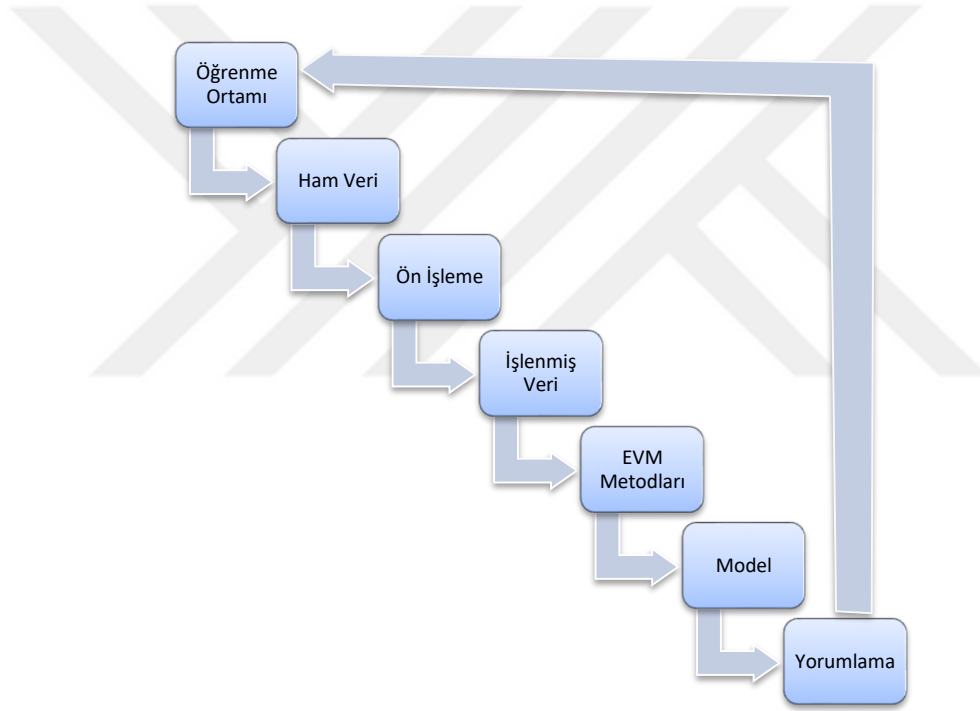


Şekil 2.5: Eğitsel veri madenciliği süreci 1.

Şekil 2.5'e göre eğitim ortamında elde edilen veriler kullanarak gerçekleştirilen eğitsel veri madenciliği sonucunda eğitim ortamında düzenlemeler yapılarak uygulama gerçekleştirilmekte ve bu yeni eğitim ortamından tekrar eğitsel veri madenciliğinde kullanılabilir yeni veriler elde edilebilmektedir. Şekil 2.5'teki veri madenciliği süreci aşağıda maddeler halinde açıklanmaktadır (Romero & Ventura, 2010).

- Ön işleme: Farklı eğitim ortamlarında elde edilen verilerin madencilik işlemleri uygun formatlara dönüştürülmesi, temizlenmesi ve entegre edilmesi işlemleridir.
- Veri madenciliği: Önceden işlemiş verilere amaca uygun olarak seçilen sınıflandırma, kümeleme vb. veri madenciliği tekniklerinin uygulanmasıdır.
- Son işleme (İşlem sonrası): Elde edilen sonuçların yorumlanarak eğitim ortamında düzenlemeler yapmak için kullanıldığı son adımdır.

Bousbia ve Belamri (2014) ile Linan ve Perez (2015) yaptıkları çalışmalarda, Romero ve Ventura (2010)'un araştırmasında sunduğu eğitsel veri madenciliği uygulama sürecini Şekil 2.6'daki gibi güncellemişlerdir.



Şekil 2.6: Eğitsel veri madenciliği süreci 2.

Romero ve Ventura (2010) müşterilerin ilgi alanlarını belirleyerek satış rakamlarını artırmak için veri madenciliği uygulanarak e-ticaret sistemlerinin başarıya ulaştığı gibi eğitim sistemlerinin değerlendirilerek eğitim kalitesinin artırılması ve etkili bir öğrenme sürecinin sağlanması amacıyla eğitim ortamlarında veri madenciliğinin uygulanması artan bir ilgi görmeye başlamış fakat henüz eğitimde beklenen ilerleme kaydedilememiştir.

Birçok alanda uygulanması ile veri madenciliğinin eğitimde uygulanması birbirlerine her ne kadar benzer görünseler de veri madenciliğinin eğitime uygulanmasını farklı kılan amaç,

kullanılan veriler ve uygulanan tekniklerdir (Romero & Ventura, 2007). Bu konuların önemi aşağıda maddeler halinde incelenmektedir (Romero & Ventura, 2010).

- Amaç: Eğitsel veri madenciliğinin amacı diğer alanlardaki uygulamalara göre daha öznel ve hassas ölçüm tekniklerine ihtiyaç duymaktadır.
- Veriler: Eğitim ortamlarından elde edilen verilerin türlerinde farklılıklar vardır ve bu verilerin eğitim alanına özgü veriler olması içsel anlamsal bilgilere, çok sayıda anlamlı hiyerarşik düzene sahip olduklarını gösterir. Dolayısıyla öğrencinin ve sistemin pedagojik yönlerinin de analize dahil edilmesi gerekmektedir.
- Teknikler: Eğitim verileri ve problemlerinin farklı özel niteliklere sahip oluşu geleneksel veri madenciliği tekniklerinin tümünün eğitim problemine uyarlanmasını engeller ve belirli eğitim sorunları için belirli tekniklerin kullanılması önerilmektedir.

Eğitsel veri madenciliği diğer alanlarda uygulanan veri madenciliğine göre farklı amaçlar güden, vizyon ve misyonları yönünden farklı açılardan bakan, yalnızca sınıflarını yönetmeye çalışan öğretmenler değil de kendi öğrenme süreçlerini anlamaya çalışan ya da geri bildirimler ile durum hakkındaki düşüncelerini desteklemeye çalışan öğrenciler gibi birçok grubu içermektedir (Hanna, 2004; Merceron & Yacef, 2005). Ayrıca sınıf ortamı, İnternet tabanlı eğitim sistemleri, öğrenme yönetim sistemleri, dijital anketler/testler/sınavlar/görsel-yazılı içerikler gibi ve artık eğitim ortamı olarak kullanılabilen sosyal ağlar, forumlar, sanal ortamlar, eğitici oyun ortamları gibi çeşitli eğitim sistemleri/ortamları birbirlerinden tamamen farklı veriler oluşturmakta, eğitsel veri madenciliğinde amaca yönelik çok farklı tekniklerin kullanılarak probleme çözüm aranmasını gerektirmektedir (Romero & Ventura, 2010).

Eğitsel veri madenciliği çalışmalarında kullanılan veriler aşağıdaki özelliklerine göre ayırt edilebilir (Bousbia & Belamri, 2014):

1. Veri Kullanılabilirliği:
 - a. Kurum veritabanları ya da öğrenme ortamları yazılımlarında kaydedilen veriler mevcuttur.
 - b. Araştırmalar sırasında üretilen veriler kullanılabilir.
 - c. Karşılaştırmalı veri havuzlarında veriler mevcuttur.
2. Kaynaklar:

- a. Öğrencilerin sınıf içi faaliyetleri hakkında notlar alan gözlemcilerin elde ettiği veriler.
- b. Öğrencilerin faaliyetlerini dijital ortamda kaydeden yazılımların elde ettiği veriler.
- c. Hem dijital ortamdan elde edilen hem de alınan notlardan elde edilen veriler.

3. Öğrenme Ortamı:

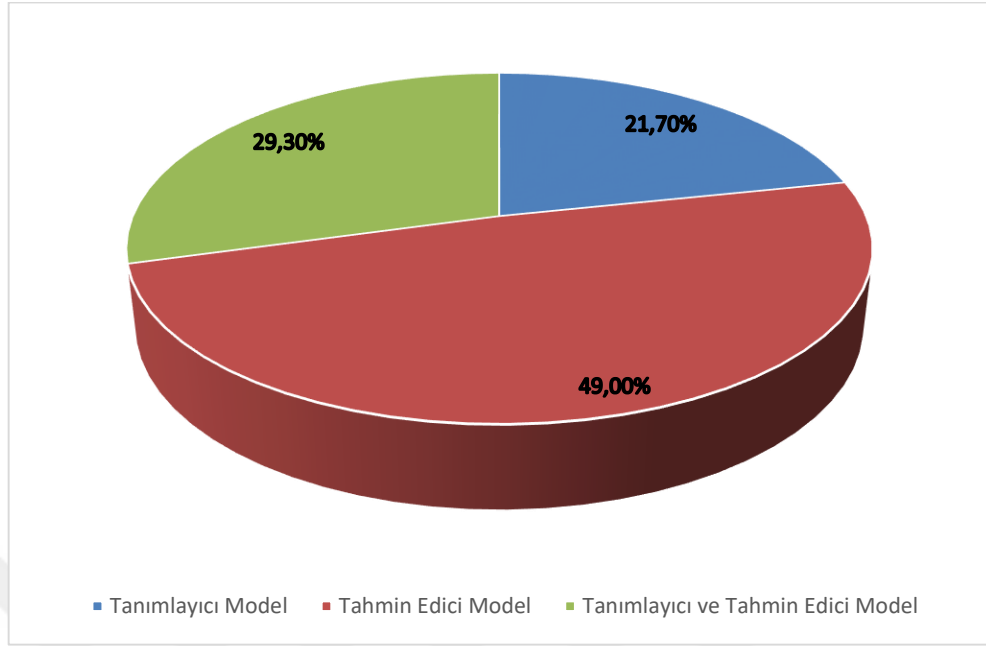
- a. Geleneksel eğitim.
- b. Dijital ortamlarda saklanabilir veri üreten bilgisayar tabanlı eğitim.

4. Veri Türleri:

- a. Nitel ya da nicel veriler.
- b. Kişisel ve / veya demografik veriler.
- c. Uygulanan anketlere verilen cevaplar.
- d. Eğitim ortamı içerisinde bilgisayar desteği ile elde edilen bireysel etkileşimler.
- e. Sosyal etkileşimler.
- f. Yüz ve vücut reaksiyonları.

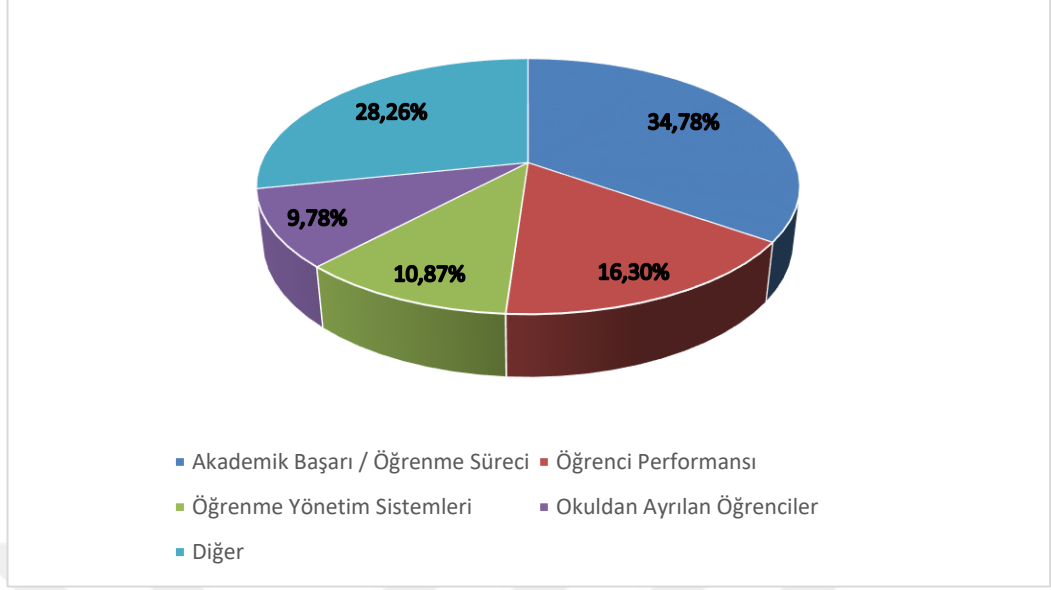
Tahmine dayalı modeller oluşturmak için öğrenme yönetim sistemlerinde gelen veriler ve klasik öğrenci yönetim sistemlerinden gelen demografik ve bilişsel bazı veriler kullanılır fakat kurumlardaki bazı öğrenci bilgilerinin gizliliği ve etik kaygılar başarılı tahminler elde etmek için ihtiyaç duyulabilecek bir takım verilerin elde edilmesini imkânsızlaştırmaktadır (Yu, et al., 2020). Eğitsel veri madenciliğinde verimli ve daha doğru tahminler yapabilmek için ihtiyaç duyulan tüm verilere ulaşmak etik kaygılar nedeniyle çoğu zaman mümkün olmasa da araştırmacılar tarafından toplanan demografik bilgiler ile bu eksiklikler giderilmeye çalışılmaktadır.

Eğitsel veri madenciliğinde de veri madenciliği yöntem ve teknikleri olan kümeleme, sınıflandırma ve birliktelik kuralları kullanılırken bunlar arasındaki en yaygın kullanım sınıflandırma olarak görülmektedir (Güldal & Çakıcı, 2017). Şekil 2.7’de eğitsel veri madenciliği alanında 2006 ile 2016 yılları arasında gerçekleştirilen çalışmalarda kullanılan veri madenciliği modelleme yöntemlerinin yüzdelik dağılımları verilmiştir (Tekin & Öztekin, 2018).



Şekil 2.7: Eğitsel veri madenciliği alanındaki çalışmaların kullanılan modellere göre yüzdelik dağılımı.

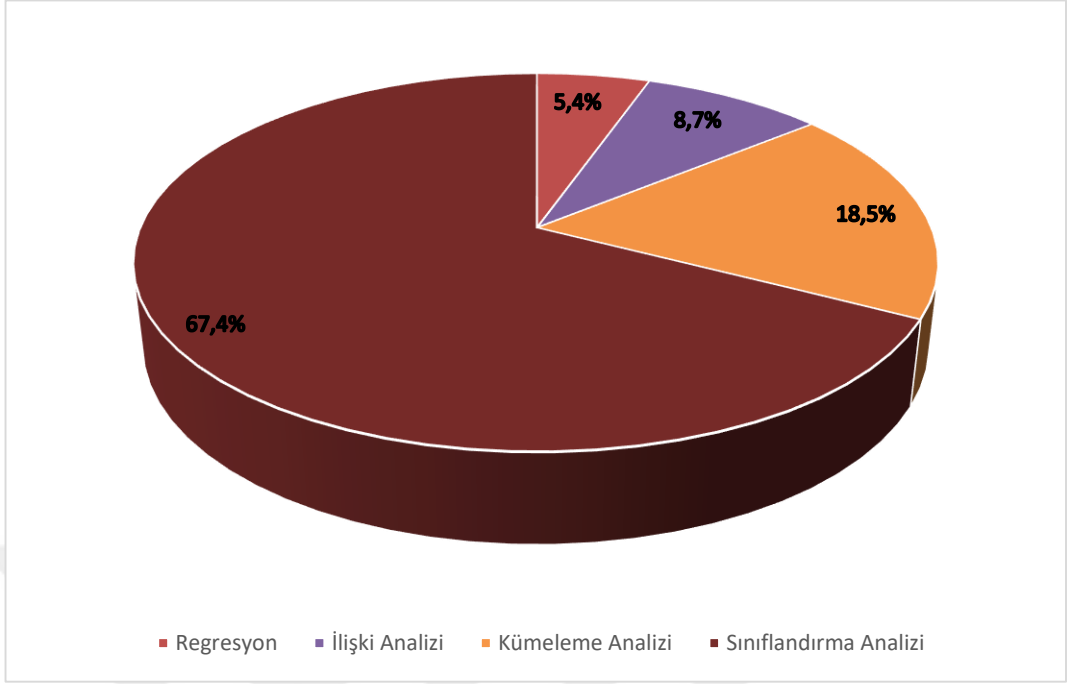
Şekil 2.7'ye göre 2006 ile 2016 yılları arasında eğitsel veri madenciliği alanında yapılan çalışmaların neredeyse yarısının tahmin edici model kullanılarak gerçekleştirildiği görülmektedir.



Şekil 2.8: Eğitsel veri madenciliği alanındaki çalışmaların araştırma konularına göre yüzdelik dağılımı.

Şekil 2.8'e göre 2006 ile 2016 yılları arasında eğitsel veri madenciliği alanında yapılan çalışmaların toplam %51,08'lik bölümünün akademik başarı ve öğrenci performansını araştırma konusu olarak belirlediği görülmektedir. Şekil 2.7 ve Şekil 2.8'deki veriler 2006 ile 2016 yılları arasında yapılan eğitsel veri madenciliği araştırmalarının yaklaşık olarak yarısında öğrencilerin akademik başarı ve öğrenci performanslarını konu alarak bunların tahmin edilmesi yönünde çalışmalar yapıldığını göstermektedir.

Öğrenci performansını tahmin etmeye yönelik araştırmalarda yaygın olarak kullanılan sınıflandırma yöntemlerinden ana öğrenme paradigmalarını temsil edenler Lojistik Regresyon (Logistic Regression), Karar Ağaçları (Decision Tree), k En Yakın Komşu (K-Nearest Neighbors), Destek Vektör Makineleri (Support Vector Machines) ve Rastgele Orman (Random Forest) algoritmalarıdır (Santamaria-Bonfil, et al., 2020).



Şekil 2.9: Eğitsel veri madenciliği alanındaki çalışmaların analiz yöntemlerine göre yüzdelik dağılımı.

Şekil 2.9'daki eğitsel veri madenciliğinde kullanılan veri analiz yöntemlerinin yüzdelik dağılımları incelendiğinde sınıflandırma tekniğinin bu alanda diğer tüm tekniklere göre çok daha fazla kullanıldığı görülmektedir (Tekin & Öztekin, 2018). Öğrencilerin performansını tahmin etmek için kullanılan eğitsel veri madenciliği tekniklerinden olan sınıflandırma, bilinmeyen verilerin sınıf etiketlerini oluşturabilen bir model oluşturmayı amaçlayan denetimli bir öğrenme tekniğidir (Hicham, et al., 2020).

Eğitsel veri madenciliğinin, öğrenme yönetim sistemlerindeki etkileşim verilerini kullanarak daha etkili öğrenme ortamları tasarlanmasında önemli rol oynayabileceği ve sensör teknolojilerinin eğitimde kullanılması ile beraber yakın gelecekte Nesnelerin İnterneti üzerinde yapılacak çalışmalara da rastlanabileceği öngörülmektedir (Şahin & Yurdugül, 2020).

2.3. ÖĞRENME ANALİTİKLERİ

Sınıf ortamında eğitimcilere destek olması amacıyla verilerin daha kolay görselleştirilmesi için araçlar geliştirmek ölçeklenebilirlik, kullanışlılık ve eğitim ekosisteminin kısıtlamaları açısından oldukça zorlu bir süreçtir (Munguia, et al., 2020). Romero ve Ventura (2020) çeşitli yeni eğitim ortamlarının öğrenciler hakkında topladığı yüksek eğitim değeri bulunan büyük

miktarda verinin manuel olarak analizinin imkansız olduğunu dolayısıyla bu tür verileri otomatik olarak analiz edecek öğrenme analitiği araçlarına ihtiyaç duyulduğunu belirtmektedir. Son yıllarda çevrimiçi öğrenme sistemlerinin popülaritesinin yükselişi hacim, hız ve çeşitlilik açısından önemli bir veri artışına neden olmuş, verilerin eğitimcilerin ihtiyaçlarına yönelik filtrelenmesi, anlaşılması ve analiz sonuçlarının ifade edilmesi önceki yıllara göre daha da zorlaşmıştır (Wang & Han, 2020). Ancak tüm bu zorluklara rağmen öğrenme süreçlerindeki karmaşıklıkları keşfetmek ve çözmek için gereken tüm ayrıntıların etkileşim verilerinde gizli olduğu söylenebilir (Russell, et al., 2020).

Öğrenme analitiği öğrencilerin performans tahmini, kavram yanlışlarının belirlenmesi gibi eğitim problemlerine diğer yaklaşımların yanı sıra makine öğrenmesi, doğal dil işleme ve istatistik uygulayan çok disiplinli bir alandır (Pena-Ayala, 2018). Siemens ve Gasevic (2012) öğrenme analitiğini öğrenenlerin bulundukları öğrenme ortamlarını anlayıp iyileştirmek amacıyla buradan elde edilen verilerin ölçülmesi, toplanması, analizi ve raporlanması olarak tanımlamaktadırlar. Öğrenme analitiği, eğitimde etkileşim sorunlarına yeni çözümler sunarak mevcut öğrenme ve öğretme modellerini geliştirme potansiyeline sahip, eğitim sürecini iyileştirmek için veri toplayabilen, işleyebilen, raporlayabilen, eğitim ortamlarında veri madenciliğinin uygulanmasını temsil etmek için kullanılan özel bir alandır (Aldowah, et al., 2019). Russel ve Diğerleri (2020) öğrenme analitiğinin birincil amacının öğrenci katılımı ile öğrenci performansı arasındaki ilişkiyi anlamak ve buna göre belirlenmiş müdahalelerden fayda sağlaması muhtemel olan öğrencileri tahmin etmek olduğunu vurgulamaktadırlar. Öğrenme analitiğinin yükseköğretim kademesindeki uygulamalarında en yaygın kullanımı, öğrenci başarısına katkıda bulunan kritik önemdeki faktörleri belirlemek ve müdahaleye ihtiyacı olan öğrencilerin erken tespit edilmesine yönelik tahmine dayalı modelleme kullanmaktır (Yu, et al., 2020). Eğitsel veri madenciliği sistemleri bileşenlerine göre indirgemeci bir şekilde incelerken, öğrenme analitiği makine öğrenmesi modelleri ile desteklenerek eğitim paydaşlarının karar vermelerinin de olduğu bütüncül bir bakış açısına odaklanmaktadır (Santamaria-Bonfil, et al., 2020). Şahin ve Yurdugül (2020) ise öğrenme analitiklerinin eğitsel veri madenciliği ile karıştırılmaması gerektiğini vurgulayarak eğitsel veri madenciliğini büyük verilere uygulanan veri madenciliği yöntemleri ile örüntüler keşfetmek olarak tanımlarken, öğrenme analitiklerini ise ortaya çıkarılan bu örüntülerden faydalanarak e-öğrenme ortamlarının iyileştirilmesi için çalışmak olarak belirtmektedirler. Eğitsel veriler üzerinde çok fazla öğrenme analitiği ve eğitsel veri madenciliği uygulamaları yapılmakta olup bu araştırma alanı literatüründe aşağıda

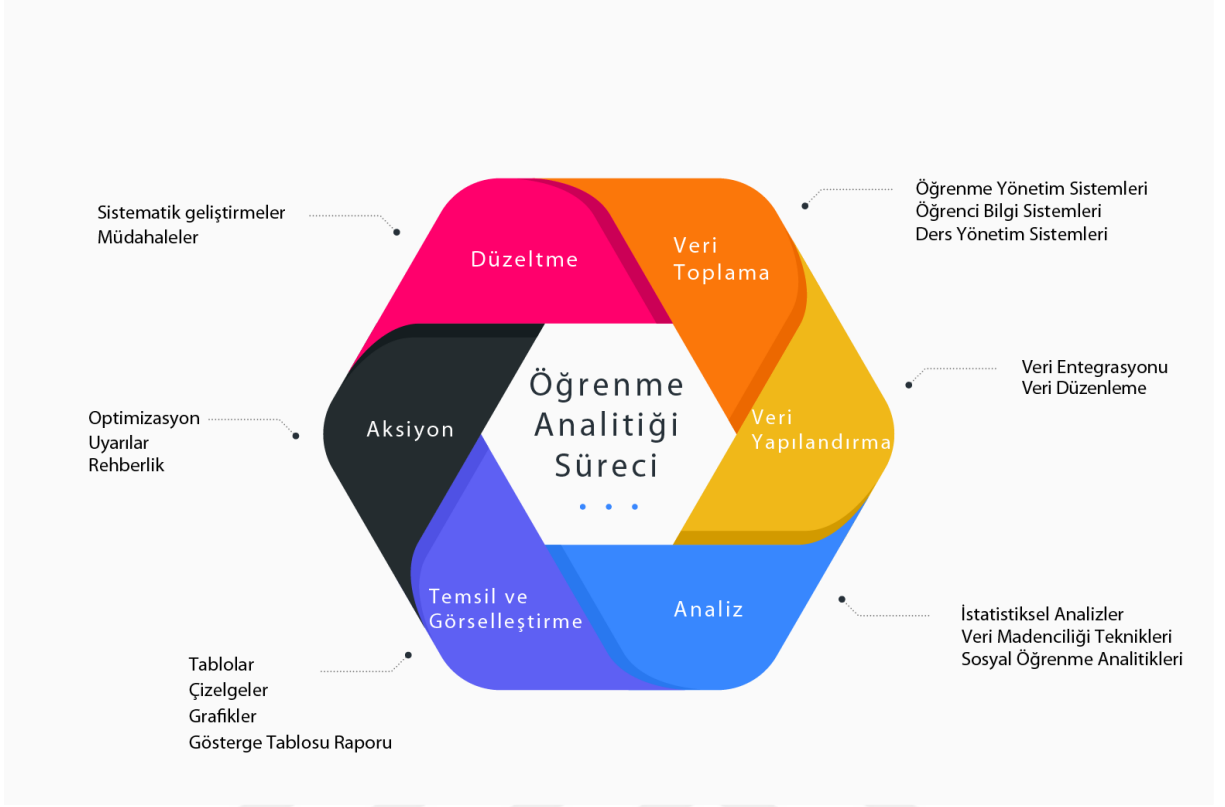
maddeler halinde tanımlanmış birbiri ile ilişkili çeşitli terimler kullanılmaktadır (Romero & Ventura, 2020).

- Akademi Analitiği ve Kurumsal Analitik: Dersler ve derece programları gibi akademik program etkileşimlerinin toplanması, analizi ve görselleştirilmesiyle ilgilidir.
- Öğretim Analitiği: Öğretim etkinliklerinin tasarımı, geliştirilmesi ve değerlendirilmesinin analiz edilmesini ifade eder.
- Veriye Dayalı Eğitim ve Eğitimde Veriye Dayalı Karar Verme: Öğrencilerin ve okulların başarısını iyileştirmeye yardımcı olacak kararlara rehberlik etmek için çeşitli eğitim verilerinin sistematik olarak toplanması ve analiz edilmesi anlamına gelir.
- Eğitimde Büyük Veri: Eğitim ortamlarından gelen verilere büyük veri tekniklerinin uygulanmasıdır.
- Eğitsel Veri Bilimi: Eğitim problemlerinin çözümü için eğitim ortamlarından toplanan verilerin kullanılması olarak tanımlanabilir.

Öğrenme analitiği hem öğrenme süreci hem de öğrenme ortamı üzerinde yararlı etkilere sahip olduğu inanılan öğrenci verilerinin analizine dayalı öğrenmeyi iyileştirmeyi sağlayan öğrenme sürecine müdahaleleri mümkün kılar ve bu alandaki araştırmaların öğrenmeyi nasıl etkilediği aşağıdaki üç operasyonel tanımla kategorize edilebilmektedir (Knobbout & Stappen, 2020).

- 1) Öğrenme ortamı
- 2) Öğrenme süreci
- 3) Öğrenme çıktısı

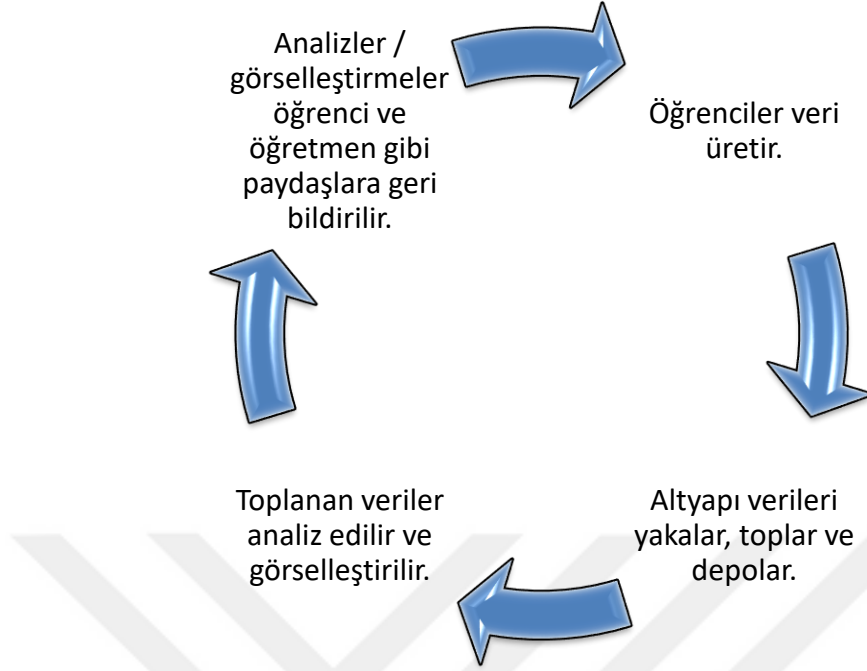
Öğrenme analitiği döngüsel bir sürecin içinde bulunmaktadır. Şekil 2.10'da Lal (2014) tarafından düzenlenmiş olan öğrenme analitiği süreci açıklamaları ile gösterilmektedir.



Şekil 2.10: Öğrenme analitiği süreci.

Şekil 2.10'daki aşamalarda ilk olarak öğrenme yönetim sistemleri, öğrenci bilgi sistemleri ve ders yönetim sistemleri gibi dijital öğretim ortamlarından veriler elde edilir. Elde edilen veriler veri madenciliği aşamalarından veri ön işlemeden geçirilerek düzenlenir ve öğrenme analitiğinde kullanılan analiz yöntemleri uygulanır. Ortaya çıkan örüntüler tablo, çizelge, grafik ve gösterge tablosu raporları kullanılarak görselleştirilir ve tüm eğitim paydaşları için anlaşılabilir hale getirilir. Daha sonra görselleştirilen bu örüntülerden uyarılar ve optimizasyon çalışması için notlar hazırlanır. Yapılan bu optimizasyon işlemi sonrası eğitim ortamlarında gerekli geliřtirilmeler tamamlanır ve döngü başa dönerek bu ortamlardan veri toplanmaya devam edilir.

Knobbout ve Stappen (2020)'e göre ise öğrenme süreci Şekil 2.11'deki gibi 4 temel aşamadan oluşan bir döngüdür.



Şekil 2.11: Öğrenme analitiği döngüsü.

Öğrenme döngüsündeki aşamalardan paydaşlara geri bildirim yapma işi öğrenme analitiğinin öğrenenler üzerinde etkili olabilmesi için gerçekleştirilen öğrenme analitiği müdahalesidir (Knobbout & Stappen, 2020).

Pena-Ayala (2018) öğrenme analitiği taksonomisinde uygulamalar alanını kaynaklar, öğrenci analizi ve işlevsellikler olarak üç başlık altında incelemekte ve bunların alt başlıklarını da aşağıdaki gibi açıklamaktadır.

❖ Kaynaklar

- Arayüzler: Öğrencilerin çok modlu etkileşimlerinin dijital ayak izlerini otomatik olarak yakalarlar.
- Araçlar: Farklı durumlarda kullanıcıların ihtiyaçlarını karşılamak için eğitim, pedagoji ve insan iletişimi teorilerine dayanan özel işlevler kullanabilmek için programlanmışlardır.
- Veri setleri: Eğitimdeki dijital teknolojiler vasıtasıyla kurumlar sosyo-demografik bilgiler, eğitim yeterlilikleri, geçme kalma oranları gibi öğrenciler hakkındaki çeşitli verileri üretir ve kullanırlar.

❖ Öğrenci Analizi

- İçselleştirme: Öğrenciler içinde bulundukları ruh haline göre öğrenmelerini içselleştirirler.
- Davranış: Öğrenme analitiklerinde en çok tercih edilen hedef öğrencilerin davranışlarıdır ve davranış sadece öğrencilerin tutumları ile niyetlerini göstermez, ayrıca performanslarını ve sonuçlarını açıklamaya da katkıda bulunmaktadır.
- Asiste etme: Öğrenciler bakış açıları farklılıkları nedeniyle farklı desteklere ihtiyaç duyarlar ve bireysel öğrenci başarısı için çeşitli işlevler ile yaklaşımlar gerekmektedir.
- Yansıtıcı yazım: Yansıtıcı yazma analizi gelişimi için gerekli tekniklerden yoksun öğrenciler ve eğitimciler için önemli bir stratejidir.
- Görselleştirme: Görselleştirmenin ana amacı büyük veri içerisinde gizli karmaşık kalıpların çeşitli eğitim paydaşları tarafından kolaylıkla yorumlanabilir ve etkileşimli bir şekilde görünmesini sağlamaktır.

❖ İşlevsellikler

- Riskli öğrenci tahmini: Özellikle öğrencinin okulu bırakma, başarısız olma gibi risklerinin incelenmesini temsil etmektedir.
- Öğrenci performansı tahmini: Öğrencilerin eğilimlerinin, demografik özelliklerinin yanı sıra öğrenme yönetim sistemlerinden elde edilen etkileşim bilgilerinin de dikkate alındığı başarı tahmini nesnesidir.
- Değerlendirme: Değerlendirme hesap verilebilirlik, zaman, geçerlilik ve güvenilirlik kavramlarının dikkate alınması gerektiği hem eğitimciler hem de eğitim sistemleri için zorlu bir görevdir.
- Geri Bildirim: Geri bildirim eğitimcilerin ve eğitim sistemlerinin zamanında ve bilge bir ölçü ile dikkatlice gerçekleştirilmesi gereken, öğrenciler için vazgeçilmez olan bir tavsiyedir.

2.4. TERS YÜZ EĞİTİM

İnternetin yaygınlaşması ile beraber ortaya çıkan uzaktan eğitim imkanları öğrenciler arasındaki coğrafi ve zamansal ayrımı ortadan kaldırılabilmekte ve bilgi bu imkanlar dahilinde çevrimiçi platformlardaki öğretim ortamları aracılığıyla dünyanın her köşesine

aktarılabilmektedir (Hung, et al., 2020). Eğitimde teknoloji kullanımının giderek artması ile birlikte ihtiyaçlardan doğan yeni öğretim modelleri de ortaya çıkmaktadır. Hem öğrencilerin multimedya ortamlarına erişim olanaklarının artması hem de eğiticilerin multimedya geliştirebilir hale gelmesi öğrenmenin, temas süresini fazlasıyla arttırmadan ya da ders kapsamından ödün vermeden entegre edilebileceği bir ortam yaratmıştır (Lage, et al., 2000). Ortaya çıkan bu ortam için ortak kanıya varılmış bir isim yerine dönüştürülmüş sınıf, ters yüz sınıf-eğitim ya da ters yüz edilmiş öğrenme isimleri Türkçe alanyazında kullanılmaktadır (Pehlivan & Arabacıoğlu, 2020).

Sınıfı tersine çevirmek kavramı olarak ortaya atılan ters yüz eğitim, sınıf içerisinde geleneksel olarak ortaya çıkan olayların artık sınıf dışında gerçekleşmesi anlamına gelmektedir ve bu uygulama ilk olarak 1996 yılında 10 ve 6 yıllık tecrübeleri olan iki eğitimci tarafından Miami Üniversitesi'nde microekonomi dersinde gerçekleştirilmiştir (Lage, et al., 2000). İleriki yıllarda özellikle üniversite seviyesinde popülerliğini arttıran ters yüz sınıflar günümüzde ilk ve orta öğretim düzeyinde de öğrencilerin öğrenimi için savunulan bir model haline gelmiştir (Strelan, et al., 2020).

Ters yüz sınıf modelinin İnternet ve bilgisayar ortamlarında gerçekleştirilebileceğinin düşünülmesinin aksine bilgisayar icat edilmeden önce bile uygulanan öğrencilere basılı bir makale verip, onlardan okumalarını istemek ve ardından sınıfa geldiklerinde makale üzerine tartışmak gibi teknikler de ters çevrilmiş sınıf olarak görülmektedir (Gren, 2020). Flipped Learning Network (2014) isimli organizasyon ters yüz eğitimin detaylı tanımını, doğrudan öğretimin grup öğrenme alanından bireysel öğrenme alanına taşındığı, ortaya çıkan grup alanının kavramları uygularken ve derslere yaratıcı bir şekilde katılırken öğrencilere rehberlik ettiği dinamik, etkileşimli bir öğrenme ortamının sağlandığı pedagojik bir yaklaşım olarak yapmaktadır. Strelan ve Diğerleri (2020)'ne göre ters yüz eğitim farklı öğretim alanlarında uygulanabilen, öğrenci başarısına olumlu etki sağlayan, önemli bir öğretim modeli türüdür. Ters yüz eğitim bazı sınırlılıklara sahip olsa da olumlu eğitim sonuçları sunan, öğrencilere bilgi aktarmak için sınıf zamanının kullanılmadığı ve buradan kalan zamanın öğretmen tarafından tartışma, öğrenciler tarafından yöneltilen soruları çözme, uygulamalar ve rehberlik gibi farklı öğrenme etkinlikleri ile değerlendirilebildiği bir öğretim modelidir (Akçayır & Akçayır, 2018). Zou ve Diğerleri (2020) ise ters yüz eğitimi öğretim içeriğinin çoğunlukla çevrimiçi olarak sınıftan önce sunulduğu geleneksel sınıf prosedürünün tersine çevrildiği, sınıflarda öğrencilerin

öğretmen rehberliğin problem çözme ve işbirliğine dayalı öğrenme görevlerini gerçekleştirdiği pedagojik bir yaklaşım olarak tanımlamaktadırlar. Bir başka çalışmada (Algayres & Triantafyllou, 2020) ise, geleneksel ders ve ev ödevi oturumlarının tersine çevrildiği, öğrencilere ders öncesi gerekli bilgileri edinmeleri için çevrimiçi materyal verilirken, ders saatinde bu bilgilerin uygulaması ve yansımalarının gerçekleştirildiği öğretim yöntemi ters yüz edilmiş sınıf olarak tanımlanmaktadır.

Flipped Learning Network eğitimcilerinin önerdiği FLIP kısaltmasından yola çıkarak hazırladıkları dört öge Cevikbas ve Kaiser (2020) tarafından aşağıdaki gibi listelenmektedir.

- F (Flexible environment)-Esnek ortam: Öğretmenler, bağımsız ya da grup çalışmasını teşvik etmek için öğretme-öğrenme ortamını amaca uygun olarak geliştirmelidir.
- L (Learning culture)-Öğrenme kültürü: Ters yüz öğrenmede sınıf dersi zamanı sorgulamaya, konuları ve kavramları derinlemesine öğrenmeye ve öğrenme fırsatları yaratmaya ayrılmalıdır.
- I (Intentional content)-Maksatlı içerik: Öğretmenlerin, derin anlama geliştirmede öğrencileri destekleme amacıyla ters yüz eğitimden nasıl yararlanabileceklerini sürekli olarak düşünmeleri, içerik ve öğrenciler tarafından keşfedilecek materyallere karar vermesi gerekmektedir.
- P (Professional educator)- Profesyonel eğitimci: Eğitimciler sürekli olarak öğrencileri gözlemlemeli, destek sağlamalı, kapsamlı geri bildirimler vermeli ve öğrencilerin çalışmalarını değerlendirmelidirler.

Öğrenme analitikleri, eğitsel veri madenciliği ve ters yüz eğitim oldukça aktif çalışma alanları olmalarına rağmen bu alanlar birleştirilerek yapılan çalışma sayısı hala oldukça azdır ve özellikle bu uygulamaların öğrenme yönetim sistemleri üzerinden elde edilen veriler ile sınırlı kalmasının yeterli olmayıp geliştirilmiş araçlar ile desteklenmiş öğrenme yönetim sistemlerinden faydalanılması gerektiği düşünülmektedir (Algayres & Triantafyllou, 2020).

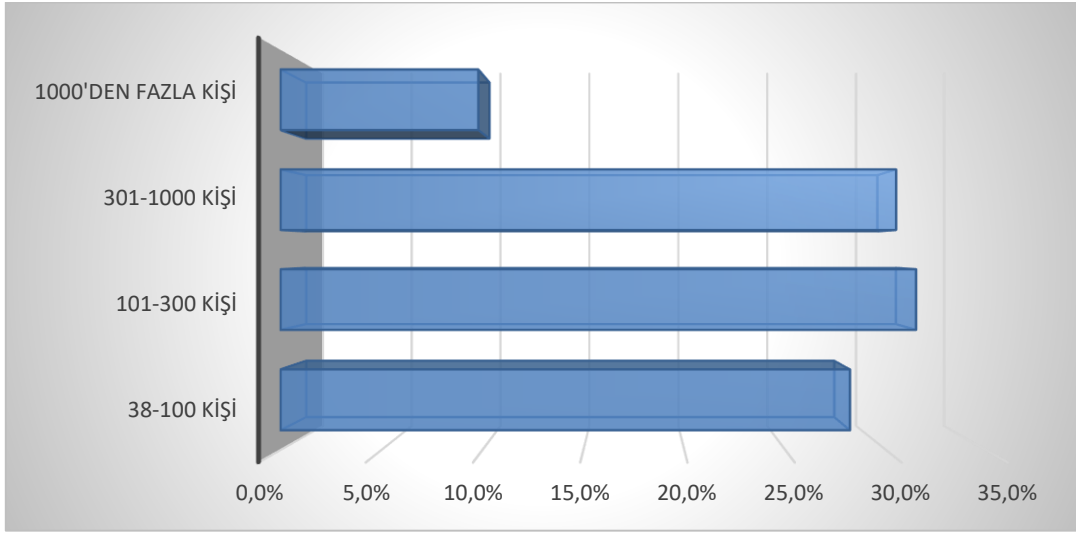
Algayres ve Triantafyllou (2020) yaptıkları çalışmada öğrenme analitiklerinin ve eğitsel veri madenciliği tekniklerinin ters yüz sınıflarda uygulanmasının aşağıda açıklanmış biçimde dört farklı şekilde gerçekleştirildiğini belirtmişlerdir.

- I. Tanılayıcı Analitikler: Eğitsel veri madenciliğine dayalı genel tanımlayıcı analitiklerdir. Burada öğrenme analitikleri öğrencilerin öğrenim platformuyla etkileşimlerini ve öğrenme performanslarının ölçüsünü izlemek için öğrenme yönetim sistemlerinin izlerinin kullanımına dayanmaktadır. “Ne oldu?” sorusuna cevap aranmaktadır.
- II. Teşhis Analitikleri: Öğrencilerin öğrenme kalıplarını gözlemleyerek, başarıları ve gelişmiş öğrenme deneyimlerini ters yüz sınıflardaki deneyim ve öğrenme analitiği kullanımı ile ilişkilendiren çalışmalarda kullanılmaktadır. “Neden oldu?” sorusuna cevap aranmaktadır.
- III. Tahmine Dayalı Analitikler: Öğrencilerin okulu bırakma riskinin ya da başarı performanslarının tahmin edilmesi ve müfredatın gidişatını düzenlemek için erken müdahale edilebilmesi için kullanılan, ters yüz sınıflarda kullanımı büyük potansiyele sahip analitiklerdir. “Ne olacak?” sorusuna cevap aranmaktadır.
- IV. Normatif Analitikler: Eğitimcilerin daha iyi bir öğrenme için en uygun araçları geliştirmesi, ortaya çıkarması ve kullanması amacıyla gerçekleştirilen bir çeşit sınıflandırma çalışmasıdır. “Nasıl gerçekleştirilebilir?” sorusuna cevap aranmaktadır.

Maddeler incelendiğinde ters yüz sınıflarda özellikle tahmine dayalı analitikler ve normatif analitiklerin eğitsel veri madenciliği teknikleriyle uygulamalarının henüz çok yeni ve gelişmekte olan bir alan olduğu, bu alanda çalışmaların yüksek potansiyele sahip olduğu söylenebilir.

2.5. BENZER ÇALIŞMALAR

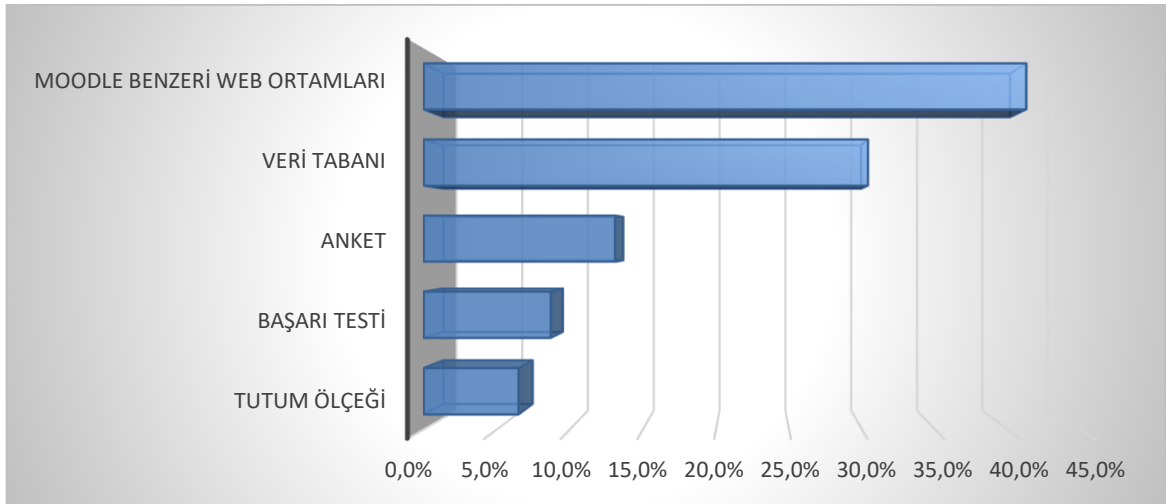
Eğitsel veri madenciliği alanında gerçekleştirilen 2006 ile 2016 yılları arasındaki araştırmaları inceledikleri çalışmalarında Tekin ve Öztekin (2018), yapılan araştırmalardaki çalışma gruplarının boyutlarını Şekil 2.12’de sunmuşlardır.



Şekil 2.12: Eğitsel veri madenciliği alanındaki çalışmaların çalışma grubu sayılarına göre yüzdelik dağılımı.

Alanda gerçekleştirilen çalışmaların %10'nundan azı 1000 kişiden fazla kişiden oluşan çalışma grubu ile çalışırken en yüksek yüzdeye %35'e yakın bir oranla sahip olan 101-300 kişiden oluşan çalışma grubuna sahip çalışmalardır.

Şekil 2.13'te eğitsel veri madenciliği çalışmalarında kullanılan veri toplama araçlarının yüzdelik dağılımları gösterilmektedir (Tekin & Öztekin, 2018).



Şekil 2.13: Eğitsel veri madenciliği alanındaki çalışmaların veri toplama araçlarına göre yüzdelik dağılımı.

Şekil 2.13 incelendiğinde, alandaki çalışmaların yaklaşık %5'i tutum ölçeklerinden elde edilen veriler ile çalışırken, neredeyse alandaki çalışmaların yarısını oluşturan bir kesim Moodle ya da benzeri web ortamlarını veri toplama aracı olarak kullanmaktadır.

Tekin (2014) yaptığı çalışmada üniversite düzeyindeki 127 öğrencinin sekiz dönem boyunca toplam 49 dersten aldığı ders notlarını kullanarak veri setini oluşturmuş ve daha sonra ilk iki yıllık ders notları ile algoritmasını eğiterek öğrencilerin mezuniyet notlarını tahmin etmeye çalışmıştır. Bu çalışmada sınıflandırma yöntemlerinden Yapay Sinir Ağları, Destek Vektör Makineleri ve Ekstrem Öğrenme Makinesi algoritmaları kullanılmış, bunlar arasından Destek Vektör Makineleri algoritmasını tahmin performansı %93,06 olarak hesaplanmıştır.

Razak ve Diğerleri (2018) araştırmalarında farklı bölümlerdeki 257 üniversite öğrencisinin cinsiyet, okudukları program, genel not ortalaması ve birinci dönem, ikinci dönem, üçüncü dönem, dördüncü dönem, beşinci dönem not ortalamalarını veri seti olarak kullanmışlardır. Bu verilerden yola çıkarak öğrenciler mezuniyet başarı performanslarını tahmin etmeye çalışan araştırmacılar Doğrusal Regresyon analizi ve J48 Karar Ağacı algoritması ile veri setini test etmişlerdir. Araştırmalarında genel not ortalaması ile 5 dönemlik not ortalamalarının öğrenci mezuniyet başarı performansı ile güçlü ilişkileri olduğunu tespit etmiştir. Doğrusal Regresyon analizi ile %96,2 ve J48 Karar Ağacı algoritması ile %82,5'lik bir tahmin başarısı elde etmişlerdir.

Asif ve Diğerleri (2017) gerçekleştirdikleri çalışmada 210 Bilgi Teknolojileri Bölümü öğrencisinin lise sertifika sınavındaki toplam puanları, aynı sınavdaki matematik, fizik ve kimya puanlarını, ayrıca üniversitedeki beş farklı dersten aldıkları puanları kullanarak mezuniyet performanslarını tahmin etmeyi hedeflemişlerdir. Rapid Miner uygulaması ile yapılan analizlerde Karar Ağacı, Naive Bayes, K-En Yakın Komşu, Yapay Sinir Ağları, Rastgele Orman algoritmaları kullanılmıştır. Bu algoritmalar arasından en yüksek tahmin başarısına %83,65 ile Naive Bayes algoritması erişirken, en düşük tahmin başarısına sahip algoritma %62.50 ile Yapay Sinir Ağları olmuştur.

Akçapınar ve Diğerleri (2019)'nin araştırmasında 90 üniversite öğrencisinin yaklaşık 65.000 satırlık e-kitap okuma verilerini kullanarak akademik performansı düşük olan öğrenciler tahmin edilmeye çalışılmıştır. Analiz esnasında 13 farklı tahmin algoritması kullanan araştırmacılar, ham verilere uygulanan Rastgele Orman algoritmasının diğer algoritmalarından daha başarılı

olduğunu tespit etmiştir. Veriler üzerinde dönüştürme işlemi uygulandığında J48 Karar Ağacı algoritmasının ve kategorik veriler kullanıldığında da Naive Bayes algoritmasının daha başarılı tahminler gerçekleştirdiğini ortaya koymuşlardır. Ayrıca kategorik veriler ile yapılan analizler ham veriler ile yapılan analizler ile benzer performans gösterirken, dönüştürülmüş veriler ile yapılan analizlerin ise daha düşük performans gösterdiği belirtilmektedir. Genel olarak tüm veri tiplerine uygulanan algoritmalar arasından en yüksek performansı %82,3'lük oranla ham verilere uygulanan Rastgele Orman algoritması sağlamaktadır.

Bir başka araştırmada (Chung & Sunbok, 2019) bölgelerindeki eğitim bakanlığının eğitim yönetimi sisteminden elde ettikleri 165.715 üniversite öğrencisinin derse katılım, notlar ve okul deneyimleri gibi verilerini kullanarak öğrencilerin okulu bırakma riskini tahmin etmişlerdir. Yapılan analizlerde Rastgele Orman algoritması kullanılmış ve tahmin performansı %95 olarak bulunmuştur. Ayrıca değişkenlerin önemi hesaplandığında ise en önemli değişken resmi olmayan devamsızlık olarak bulunmuştur.

Yu ve Diğerleri (2020) yaptıkları çalışmada 2.000 üniversite öğrencisinin kurumsal verilerini, öğrenme yönetim sistemlerinden elde edilen verilerini ve anketler ile toplanan verilerini inceleyerek yaptıkları karşılaştırmada kurumsal veriler ile öğrenme yönetim sistemleri verilerinin uzun vade ve kısa vade öğrenci başarısını doğru tahmin etmede yeterli güce sahip olduğunu fakat anket verilerinin çok az öngörücü güce sahip olduğunu ortaya koymuşlardır. Analiz için Lojistik Regresyon, Destek Vektör Makineleri ve Rastgele Orman sınıflandırma algoritmaları kullanılmıştır. Analiz sonuçlarına göre kısa vadede en iyi performansa sahip veri setinin kullanılan algoritmaların tümünde kurumsal veriler olduğu, uzun vadede ise öğrenme yönetim sistemleri verilerinin olduğu ortaya konmuştur.

Zhao ve Diğerleri (2020)'nin çalışmasında ise 132 lisansüstü öğrencisinin gelecekteki akademik başarısını tahmin etmek için standartlaştırılmış test puanları, lisans not ortalamaları ve okul sıralaması gibi lisansüstü eğitim başvurusundaki veriler kullanılmıştır. Analizlerde Destek Vektör Makineleri, Karar Ağacı, Rastgele Orman, Lojistik Regresyon, K-En Yakın Komşu, Naive Bayes, Yapay Sinir Ağları ve Toplu Öğrenme sınıflandırma algoritmaları uygulanmıştır. Ortalama %85'lik oranla Rastgele Orman ve Toplu Öğrenme algoritmaları en yüksek tahmin başarısına sahip algoritmalar olarak bulunmuştur.

Fernandes ve Diğerleri (2019) yaptıkları çalışmada Brezilya Federal Bölgesi'ndeki devlet okullarında lise öğrencilerinin akademik performanslarını tahmin etmek için öğrencilere ait notlar, devamsızlık ve demografik bazı değişkenlerden oluşan toplam 485.872 kayıtlık veri setini kullanmışlardır. Analizlerde Gradyan Artırma Makinesi sınıflandırma algoritması kullanılmıştır ve tüm veri setine uygulandığında başarı performansı %94 olarak hesaplanmıştır. Ayrıca akademik performansı tahmin etmek için en alakalı değişkenler notlar, devamsızlık, mahalle, okul ve yaş nitelikleri olarak ortaya konmuştur.

Akçapınar ve Coşgun (2019) ise ASSISTments adlı öğretim sistemindeki 514 orta öğretim öğrencisine ait etkileşim verilerini kullanarak öğrencilerin eğitim ve mesleki kariyerlerine STEM(Fen, Teknoloji, Mühendislik ve Matematik) ile ilgili herhangi bir alanda devam edip etmeyeceklerini tahmin ettikleri bir çalışma yapmışlardır. Çalışmada tahmin modelinin geliştirilmesi için Rastgele Orman, K-En Yakın Komşu, Destek Vektör Makineleri ve Güçlendirilmiş Genelleme Regresyon Modelleri algoritmalarını kullanmışlardır. Çalışmada en yüksek tahmin performansını %66 ile Destek Vektör Makineleri göstermiştir.

Baradwaj ve Pal (2011) tarafından yapılan çalışmada bir üniversitenin bilgi sisteminden elde edilen 50 öğrenciye ait farklı dönemlere ve derslere ait notlar ile derse devamlılık verileri kullanılarak öğrencilerin son dönem değerlendirme performansları tahmin edilmiştir. Analizlerde ID3 Karar Ağacı algoritması kullanılarak her bir nitelik ayrı ayrı incelenmiş ve önceki dönem notlarının öğrenci performansı tahmininde en etkili nitelik olduğu ortaya konmuştur.

Akçapınar ve Diğerlerinin (2015) çalışması, üniversite öğrencilerinin çevrimiçi öğrenme ortamlarından elde edilen etkileşim verileri kullanılarak öğrencilerin akademik performansını tahmin etmeyi amaçlamaktadır. 76 üniversite öğrencisinin verilerinden oluşan 10 nitelik barındıran veri setine Naive Bayes, Sınıflandırma Ağacı ve CN2 Kuralları sınıflandırma algoritmaları uygulanmış, üçlü sınıflandırma da Naive Bayes algoritması %75,4 tahmin başarıları gösterirken, ikili sınıflandırmada başarılı öğrencileri %91,8, başarısız öğrencileri ise %81,5 tahmin başarıları ile sınıflandırmıştır.

Özbay ve Ersoy (2017)'un benzer bir çalışmasında öğrenme yönetim sistemi üzerinden elde edilen 40 üniversite öğrencisinin etkileşim verileri kullanılarak, bu öğrencilerin akademik başarıları ile etkileşim verilerinin arasında anlamlı bir ilişkinin olduğu ortaya konmuştur.

Şengür ve Tekin (2013) çalışmalarında üniversite veri tabanından elde edilen 127 öğrenciye ait ders notlarını kullanarak mezuniyet ortalamalarını tahmin etmişlerdir. Analizde Yapay Sinir Ağları ve Karar Ağacı algoritmaları kullanılmış ve Yapay Sinir Ağları ile 25 tahminden 22'si oldukça yakın bir şekilde doğru tahmin edilirken, Karar Ağacı ile 25 tahminden 23'ü yakın değerlerde tahmin edilmiştir.

Devasia ve Diğerleri (2016) üniversite veri tabanından elde ettikleri başarı grafiği, öğrenci kabul detayları, öğrenci notları ve devamsızlık gibi verileri kullanarak öğrenciler dönem sonu başarı performansını tahmin etmişlerdir. Veri seti 700 öğrenciye ait 19 nitelikten oluşturulmuş, bu veri setine Naive Bayes, Regresyon, Karar Ağacı, Yapay Sinir Ağları algoritmaları uygulanmıştır. Elde edilen sonuçlarda en başarılı tahmin oranına sahip algoritmanın Naive Bayes olduğu ortaya konmuştur.

Hasnine ve Diğerleri (2018)'nin çalışmasında öğrencilerin e-kitap tıklama etkileşimleri kullanılarak Yapay Sinir Ağları, AdaBoost, Lojistik Regresyon, Naive Bayes, K-En Yakın Komşu, Destek Vektör Makineleri, Rastgele Orman ve CN2 Kuralları algoritmaları ile öğrencilerin ders sonu notları tahmin edilmiştir. En iyi ilişki gösteren 5 nitelik seçildiğinde CN2 Kuralları %88 tahmin başarıları gösterirken aynı oran tüm nitelikler ile çalıştırıldığında Yapay Sinir Ağları'nda da tespit edilmiştir.

Kokoç ve Diğerleri (2018) yaptıkları çalışmada öğrencilerin video izleme davranışlarının akademik başarı performanslarını nasıl tahmin ettiğini ortaya koymaya çalışmışlardır. Yapılan analizler sonucu akademik başarı ile video izleme davranışları arasında istatistiksel olarak anlamlı bir ilişki ortaya koymuş fakat güçlü bir korelasyon ortaya çıkmadığını savunmuşlardır. Araştırmada 101 öğrencinin 10 videodan elde edilen 5 nitelikten oluşan 5 haftalık verileri kullanılarak Çoklu Regresyon Modeli uygulanmıştır. Modelin öğrenme performansındaki varyansın %35,2'sini açıkladığı ortaya konmuştur. |

3. MALZEME VE YÖNTEM

Son yıllarda popülerliği gittikçe artan, verilerin içerisindeki gizli kalmış örüntüleri ortaya çıkararak yeni araştırmalara kaynak sağladığı gibi keşfedilmemiş enformasyonları gün yüzüne çıkaran veri madenciliği bu tezde bir araştırma modeli olarak kullanılmıştır.

3.1.ÇALIŞMA GRUBU VE GENEL ÖZELLİKLERİ

Bu araştırmada Yeditepe Üniversitesi'nde kitle dersi olarak okutulan ve birçok farklı bölümden öğrencinin kayıt olduğu Temel Bilgisayar Uygulamaları isimli dersin araştırmacı tarafından sürdürülen 4 ders yarıyılı boyunca verileri toplanmıştır. Veriler haftalık olarak Moodle sistemine yüklenmiş olan araştırmacı tarafından çekilmiş ders videolarından elde edilen etkileşimlerden oluşmaktadır. İlk elde edilen verilerden 457 öğrencinin verilerine ulaşılabilmişken, bunlardan derse devam etmeyen ya da dersi bırakan, aynı dersi daha önce almış olan öğrenciler çıkarıldığında toplam 404 öğrencinin etkileşim verileri kullanılmıştır. Öğrencilerin genel özelliklerine bakıldığında 18-22 yaş aralığında olmaları ve eğitim bilimleri, mühendislik, iktisadi idari bilimler, sağlık bilimleri gibi farklı alanlardaki bölümlerden geliyor olmaları sebebiyle araştırmanın ana problemini doğru bir biçimde analiz edebilmek için örneklemin evreni temsil ettiği söylenebilir.

3.2.VERİ TOPLAMA ARAÇLARI

Araştırmada kullanılacak veriler araştırmacı tarafından toplanmıştır. Veri madenciliğinde genellikle daha önce elde edilmiş binlerce satırlık veriler tercih edilir fakat eğitim alanında bir araştırma yapılması ve uzun süreli olarak toplanmış verilere ihtiyaç duyulmasından dolayı verilerin araştırmacı tarafından toplanması gerekmektedir. Bu amaçla verilerin elde edilmesi için ortam olarak MOODLE isimli öğrenme yönetim sistemi ve araç olarak da bir grup araştırmacı tarafından geliştirilen ve MOODLE sistemi üzerinde çalıştırılabilen Moodle Video Analytics Tool(MVA) kullanılmıştır (Bayazıt & Akçapınar, 2018). Bu araç öğrencilerin sistem üzerindeki videoları izlerken ki her bir etkileşimine id vererek kullanıcı bilgisi, etkileşim zamanı, etkileşimin türü, detay bilgisi, video bilgileri gibi verilerini kayıt altına almaktadır. MVA aracı kullanılarak 4 ders yarıyılı boyunca öğrencilerin derslere gelmeden önce izlediği o haftaki ders konusu ile ilgili ortalama 18 dakikalık videolarda bulundukları her bir etkileşim kayıt altına alınarak mümkün olduğunca çok kayıt elde edilmeye çalışılmıştır.

id	userid	sessionid	time	event	detail	videoid
29735	470	f2l7aml5kk3g1jq2m886obp450	1509887560	13	1159.6 saniyeye atlama yapıldı	1117
29736	470	f2l7aml5kk3g1jq2m886obp450	1509887561	8	1159.6 saniyede video oynatıldı	1117
29737	470	f2l7aml5kk3g1jq2m886obp450	1509887562	13	1187.7 saniyeye atlama yapıldı	1117
29738	470	f2l7aml5kk3g1jq2m886obp450	1509887562	8	1187.7 saniyede video oynatıldı	1117
29740	470	f2l7aml5kk3g1jq2m886obp450	1509887565	13	1190.9 saniyeye atlama yapıldı	1117
29742	470	f2l7aml5kk3g1jq2m886obp450	1509887566	13	1211.4 saniyeye atlama yapıldı	1117
29743	584	6dk8tiofbte220lk378utusdf7	1509887584	7	video sayfası yüklendi	1114
29744	584	6dk8tiofbte220lk378utusdf7	1509887589	8	0 saniyede video oynatıldı	1114
29745	470	f2l7aml5kk3g1jq2m886obp450	1509887593	10	1237.8 saniyede video sonuna gelindi	1117
29746	602	t3ak77u4hrs9v6tfh05anlb0	1509887602	7	video sayfası yüklendi	1116
29747	602	t3ak77u4hrs9v6tfh05anlb0	1509887612	8	0 saniyede video oynatıldı	1116
29748	602	t3ak77u4hrs9v6tfh05anlb0	1509887624	7	video sayfası yüklendi	1115
29750	584	6dk8tiofbte220lk378utusdf7	1509887626	13	41.6 saniyeye atlama yapıldı	1114
29751	602	t3ak77u4hrs9v6tfh05anlb0	1509887631	7	video sayfası yüklendi	1116
29753	584	6dk8tiofbte220lk378utusdf7	1509887661	13	85.5 saniyeye atlama yapıldı	1114
29755	584	6dk8tiofbte220lk378utusdf7	1509887662	13	93.8 saniyeye atlama yapıldı	1114
29757	584	6dk8tiofbte220lk378utusdf7	1509887668	13	112.5 saniyeye atlama yapıldı	1114
29759	584	6dk8tiofbte220lk378utusdf7	1509887670	13	133.2 saniyeye atlama yapıldı	1114
29760	584	6dk8tiofbte220lk378utusdf7	1509887671	9	151.7 saniyede video durduruldu.	1114
29761	584	6dk8tiofbte220lk378utusdf7	1509887672	8	151.7 saniyede video oynatıldı	1114
29763	584	6dk8tiofbte220lk378utusdf7	1509887673	13	165.2 saniyeye atlama yapıldı	1114
29765	584	6dk8tiofbte220lk378utusdf7	1509887713	13	197 saniyeye atlama yapıldı	1114

Şekil 3.1: Ham verilerin görünümü.

Şekil 3.1’de MVA aracının verdiği toplanan verilerin .csv formatındaki görünümü mevcuttur. Buradaki yapıda her bir öğrencinin anlık olarak hangi video üzerinde nasıl bir işlem gerçekleştirdiği detay olarak verilmektedir. Kullanıcı numaraları incelendiğinde birden fazla kullanıcının her bir etkileşiminin anlık olarak kayıt edildiği görünmekte fakat sayısal olarak bu verilerin pek bir anlam ifade etmemesinden dolayı veri madenciliğine hazır hale getirilmesi için bir takım aşamalardan geçmesi gerekmektedir. Bu aşamalardan önce her bir kullanıcının userid’sine göre etkileşim durumları bir yazılım kullanılarak sayısal olarak gruplandırılmış ve Şekil3.2’deki gibi bir görünüm elde edilmiştir.

userid	n_Session	n_TotalAction	d_Time	n_diffVideoCount	n_VideoLoad	n_VideoPlay	n_VideoPause	n_VideoFinish	n_VideoSeek	n_ForwardSeek	n_BackwardSeek
2	9	128	14	10	15	37	24	2	50	42	8
466	25	514	197	11	93	106	45	7	162	125	36
467	11	157	182	11	14	53	28	10	49	31	18
468	10	369	273	10	18	113	84	12	142	82	60
469	17	463	426	11	38	133	77	9	204	128	76
470	10	576	170	11	18	176	70	9	302	223	74
471	17	233	255	11	29	60	45	7	83	22	61
472	10	583	204	11	24	165	131	4	258	189	65
473	11	184	147	9	13	43	30	1	73	39	28
474	11	408	120	9	19	106	74	5	203	164	37
475	14	764	410	11	26	224	186	13	306	186	112

Şekil 3.2: userid’ye göre düzenlenmiş verinin görünümü.

Öğrencilerin video izleme davranışlarına ait her bir nitelik ve açıklaması Tablo3.1’de verilmiştir.

Tablo 3.1: Veri setindeki nitelikler ve açıklamaları.

Nitelik Adı	Açıklaması
n_Session	Açılan farklı oturum sayısı
n_TotalAction	Toplam etkileşim sayısı
d_Time	Video izlerken harcanan zaman
n_diffVideoCount	Görüntülenen farklı video sayısı
n_VideoLoad	Video yükleme sayısı
n_VideoPlay	Video oynatma sayısı
n_VideoPause	Video duraklatma sayısı
n_VideoFinish	Video sonuna gelme sayısı
n_VideoSeek	Toplam atlama yapma sayısı
n_ForwardSeek	İleri atlama yapma sayısı
n_BackwardSeek	Geri atlama yapma sayısı

Öğrencilerin öğrenme stillerinin de akademik başarıları ile bir ilişki içinde olabileceği literatürde ortaya çıkmış (Ekici, 2013) ve bu bakımdan tahmin algoritmalarına destek olabileceği düşüncesiyle öğrencilerin öğrenme stilleri ile ilgili veri elde edilmesi kararlaştırılmıştır. Bu amaç için kullanılacak envanterin arayışında literatürdeki araştırmalarda en çok tercih edilen KOLB Öğrenme Stili Envanteri seçilmiştir (Açıslı, 2016; Duman, 2008; Ekici, 2013; Güven & Kürüm, 2004). Kullanılacak envanterin Evin Gencel (2007) tarafından Türkçeleştirme çalışması gerçekleştirilmiş, geçerlik ve güvenirlik katsayıları çalışmanın geçerli ve güvenilir olduğunu göstermektedir.

Paquette ve diğerleri (2020) eğitsel veri madenciliği alanındaki son beş yıl içerisindeki araştırmaları inceleyerek sadece %15’lik bir kesimin analizlerinde demografik verilere yer verdiğini ortaya koymuştur. Yu ve diğerleri (2020) ise yaptıkları çalışmada kurumsal öğrenci verileri, öğrenme yönetim sistemlerinden elde edilen veriler ve anketlerden elde edilen verileri

kullanarak yapılan öğrenci başarı performansı tahminleri ile ilgili araştırmalardan yola çıkarak öğrenme yönetim sistemlerinden elde edilen veriler ile beraber kurumsal verilerin de kullanımı daha iyi performans gösterdiğini ortaya koymuşlardır. Bu araştırmalardan yola çıkarak benzer verilerin analiz sonuçlarına olumlu etkisi olacağı düşünülmüş ve öğrenciler hakkında toplanabilecek demografik veriler için bir form oluşturulmuştur.

3.3.UYGULAMA SÜRECİ

Temel Bilgisayar Uygulamaları Dersi içeriği göz önünde bulundurularak araştırmacı tarafından 11 ders videosu hazırlanmıştır. Videolar araştırmacının ders anlattığı ve ekran görüntüleri paylaştığı anlardan oluşmaktadır. Ters Yüz Sınıf Modeli kullanılarak dersi seçen tüm öğrencilere içerik olarak oluşturulan videolar sunulmuştur. Dersin haftalık içerikleri Tablo 3.2'deki gibidir.

Tablo 3.2: Temel bilgisayar uygulamaları haftalık ders akışı.

Hafta	Konu
1	Word'e Giriş ve Dosya Menüsü
2	Giriş Sekmesi
3	Ekle ve Düzen Menüsü
4	Şekiller, Grafikler, Resimler
5	Metin ve Paragraf Biçimlendirme
6	Başvurular Sekmesi, İçindekiler Tablosu Oluşturma
7	Denklemler, Görünüm Sekmesi ve Kaynakça Yönetimi
8	Kesmeler, Sayfa Numaralandırma ve Section Kavramı
9	Vize Sınavı
10	Excel'e Giriş, Hücre Biçimlendirme, Hücre Türleri, SUM,AVERAGE, PRODUCT Komutları
11	Hücre Renklendirme, Koşullar
12	Koşulların İç İç Kullanımı, AND ve OR ifadeleri, Koşullu Renklendirmeler

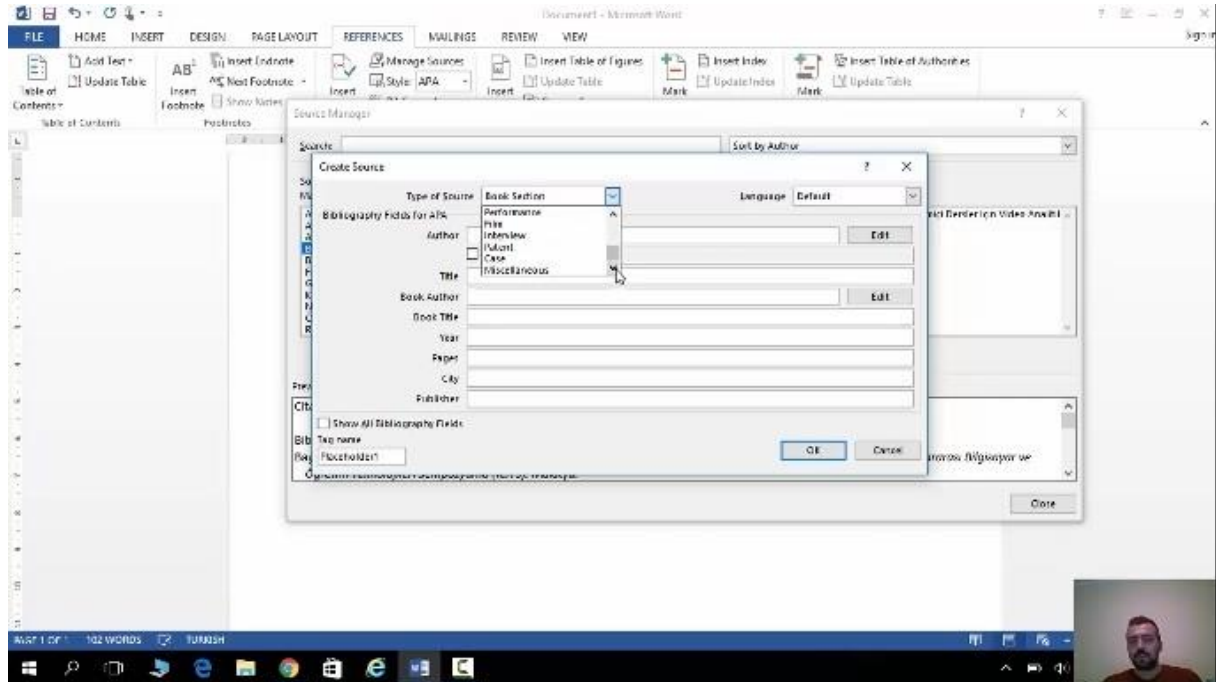
13	Koşullar ile İlgili Kapsamlı Örnekler
14	Tekrar

Ders videoları hazırlanırken haftalık ders akışına uygun içerikler gözetilmiş ve öğrencilerin anlamakta güçlük çekebileceği düşünülen konular üzerinde örnekler ile daha çok durulmuştur. Bu sebeple her bir videonun süresi farklılık göstermektedir. Tablo 3.3'te ders için hazırlanan videolar hakkında bilgiler bulunmaktadır.

Tablo 3.3: Ders videolarının içerik ve süre bilgileri.

Video No	Video İçeriği	Video Süresi
1	Word ile İlgili Genel Bilgiler	13d 32s
2	Giriş Sekmesindeki Tüm Detaylar	18d
3	Ekle ve Düzen Sekmelerinin İncelemesi	12d 15s
4	Şekiller, Resimler ve Grafiklerin Kullanımı	14d 32s
5	Metin ve Paragraflar ile İlgili Biçimlendirme Yöntemleri	21d 20s
6	Başvurular Sekmesinin İncelenmesi ve İçindekiler Tablosunun Oluşturulması	20d
7	Denklemlerin Kullanımı, Kaynak Yönetimi ve Kaynakça Oluşturulması	20d 37s
8	Kesmeler ile Section Oluşturma	8d 30s
9	Excel ile İlgili Genel Bilgiler, Hücre Kavramı, Hücrelerin Biçimlendirilmesi, Temel Fonksiyonlar	26d 41s
10	Hücrelerin Renklendirilmesi ve Koşullara Giriş	13d 45s
11	Karmaşık Koşul Yapıları ve Koşullu Renklendirme Menüsü	34d 40s

Videoların kaydedilmesinde Camtasia isimli uygulama kullanılmış olup hem bilgisayarın ekran görüntüsü hem de eğiticinin görüntüsü kayıt altına alınmıştır.



Şekil 3.2: Video Ders Örnek Ekran Görüntüsü.

Öğrencilerin dikkatinin dağılmaması için tüm videolarda eğiticinin arka plan görüntüsü Şekil3.3'te görüldüğü gibi düz bir renk olarak seçilmiştir. Öğrencilerin hazırlanan videoları izlemesini sağlamak amacıyla her hafta derslerden önce 1 saatlik süre verilmiş ve öğrencilerin videoları istedikleri etkileşimde bulunarak izlemeleri sağlanmıştır. Ayrıca her bir öğrencinin bağımsız olarak ve diğer video izleyen öğrencilerden etkilenmeden videoları izleyebilmeleri için öğrencilere kulaklıklar dağıtılmış ve öğrenci başına bir bilgisayar düşecek şekilde bir ortam düzenlenmiş bu sayede elde edilen verilerin tamamen kişisel olması mümkün kılınmıştır.

Ayrıca videolardan veri toplama işlemi 12. haftada son videonun izletilmesi ile sona ermiştir. Derslerin bitimi ile dersin final tarihi arasında 2 haftalık bir süre bulunmaktadır. Tüm bunlardan yola çıkarak verilerin analizi öğrenciler final sınavına girmeden 4 hafta önce gerçekleştirilebilir. Analiz sonuçlarına göre öğrencilerin başarılı olup olamayacağı 4 hafta öncesinden tespit edilebilmektedir. Dersler sonlanmadan 2 hafta önce başarısız olacağı tahmin edilen öğrenciler için önlemler alınabilmektedir.

3.4.VERİLERİN ANALİZ EDİLMESİ

Verilerin analizi için veri madenciliği kullanılmış ve ilk olarak veri madenciliği aşamaları sırasıyla gerçekleştirilmiştir. Veri madenciliğindeki aşamalardan olan veri önileme ve veri dönüştürme bu adımların en önemli ve karmaşık olanlarıdır (Berka & Bruha, 1998).

Veri madenciliği aşamaları (Özçakır, 2013) aşağıdaki şekilde izlenmiştir:

- Veri Seçimi
- Veri Önileme
- Veri İndirgeme
- Veri Madenciliği
- Değerlendirme

Veri seçimi aşamasında MVA aracından toplanan verilerdeki öğrenci id'leri ile Moodle sistemi üzerinden kullanıcı bilgilerinin çekilmesi ile öğrencilerin isimleri tespit edilmiştir. İsimleri tespit edilen öğrencilerden toplanan demografik bilgilerin bulunduğu formlar Excel uygulamasına girilmiş ve MVA aracından gelen veriler ile eşleştirme yapılmıştır. Bu sayede öğrencilerin kayıtlı oldukları bölüm, kaçınıcı sınıfta oldukları ve KÖSE sınıflandırma bilgileri veri kümesine eklenmiştir. Ayrıca dönem sonunda elde edilen not bilgileri de veri kümesine eklenerek veri bütünleştirme işlemi tamamlanmış ve tüm veriler tek bir ortamda toplanmıştır.

Veri Önileme aşamasında bazı verilere veri dönüştürme uygulanmıştır. Detay alanında sözel olarak bulunan öğrencilerin etkileşimleri her bir öğrenci için sayısal olarak dönüştürülmüştür.

Formlardan toplanan veriler içerisinde öğrencilerin okudukları bölümlerin çeşitliliği üniversite yerleşim puan türlerine göre EA, SAY, SÖZ, DİL olarak sınıflandırılmış ve yorumlama performansının artırılması amacıyla gruplama yapılmıştır.

MVA aracından elde edilen verilerde herhangi bir kayıp söz konusu değilken formlardan gelen bazı verilerde kayıplar söz konusudur. Formlar dersler başladığında toplandığından dolayı dönem içerisinde fark edilen bu kayıp verilerin tamamlanması için araştırmacı ilgili öğrenci ile görüşme yaparak kayıp değerleri el ile doldurmuştur. Kayıp verilerin tamamlanmasının ardından elde edilen veriler içerisinde dersi tekrar alan, derse devam etmeyen ya da dersi bırakan öğrencilerin verileri gürültülü veriye sebep olacağından veri kümesinden çıkarılmıştır.

Veri indirgeme aşamasında eldeki nitelik ve kayıt sayılarının mümkün olduğunca korunması sağlanmaya çalışılmıştır. Yapılan ilişki analizlerinde öğrencinin başarısı ile ilişkili görünmeyen birkaç nitelik bulunsa da algoritmalar bu nitelikler olmadan küçük bir miktar da olsa daha az başarılı tahminde bulunmuşlardır. Eğitsel veri madenciliğinde öğrencilerden sürekli veri toplandığında nitelik ve kayıt sayıları düşük olduğundan her bir kayıttın ve niteliğin önemli olduğu düşünülerek herhangi bir boyut indirgeme işlemi yapılmamıştır. Veri indirgeme aşaması içerisinde nitelik ya da kayıt sayısında bir kayıp oluşturmeyen ayırıklaştırma tekniği kullanılmıştır. Öğrencilerin dönem sonu notlarının 100'lük sisteme uygun biçimde girilmiş olması öğrenci başarısının tahmininde zorluk çıkaracağından üniversite yönetimince kararlaştırılan dersten başarılı olarak geçme notu temel alınarak öğrencilerin başarı durumu 0 (Başarısız) ve 1 (Başarılı) olarak kavram hiyerarşisinde bölümlenmiştir.

Bu çalışmada veri madenciliğinin tahmin edici modellerinde sınıflandırma kullanılmıştır. Sınıflandırma algoritmalarından literatür incelenerek en çok tercih edilenlerden olan Naive Bayes, Rastgele Orman ve Destek Vektör Makineleri veri setleri üzerinde denenmiş aralarından en iyi performansı sağlayan Rastgele Orman Algoritması tercih edilmiştir. Bu aşamada dengesiz veri setinin giderilmesi için tercih edilen yöntemlerden, sentetik veriler üreten SMOTE (Synthetic Minority Over-sampling Technique) tekniği kullanılarak verilerin başarılı ve başarısız olarak dağılımı dengelenmiştir.

Elde edilen sonuçların değerlendirilmesinde çalışmanın istenilene ulaşp ulaşmadığı araştırılmış ve alan yazındaki diğer araştırmalar ile karşılaştırılması yapılarak doğrulanması sağlanmaya çalışılmıştır.

Ayrıca öğrencilere elden dağıtılarak toplanan veriler de veri madenciliği sürecinde kullanılmıştır. Öğrencilerin uygulanan KOLB Öğrenme Stili envanterine verdikleri cevaplardan Aktif Deneyim, Yansıtıcı Gözlem, Soyut Kavramsallaştırma ve Somut Deneyim puanları hesaplanmış ve bu puanlarla yapılan hesaplamalar sonucu da öğrencinin öğrenme stili yerleştirme, değiştirme, ayırıştırma ya da özümseme olarak belirlenmiştir. |

4. BULGULAR

Araştırmada toplanan verilerden SPSS ve R Studio programları kullanılarak elde edilen bulgular bu bölümde başlıklar halinde sunulmaktadır.

4.1. TANIMLAYICI İSTATİSTİKSEL BULGULAR

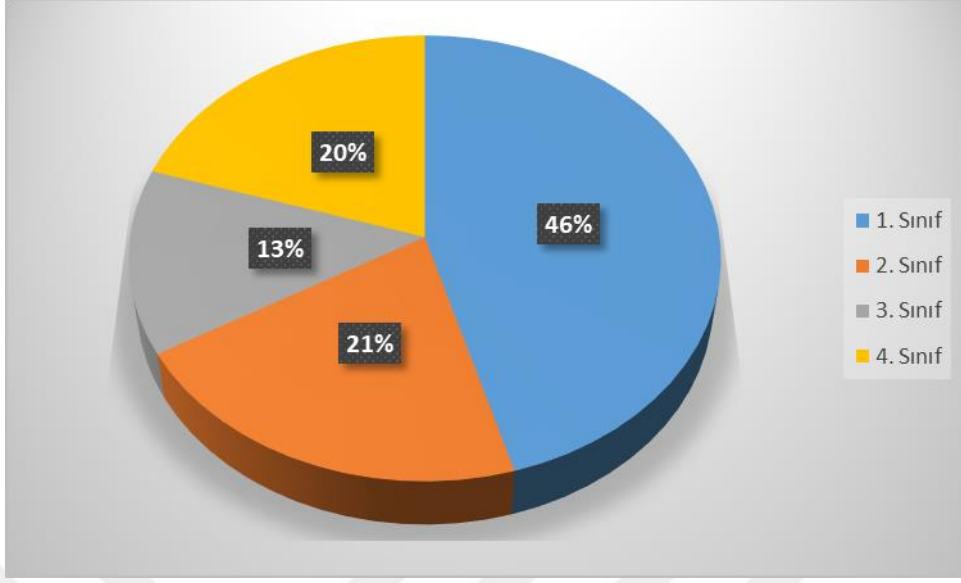
Toplanan verilerin ait olduğu öğrenciler hakkında bilgiler elde etme amacıyla tüm veri seti üzerinde tanımlayıcı istatistiksel hesaplamalar yapılmıştır. Tablo 4.1’de öğrencilerin kayıtlı oldukları lisans programlarında kaçınıcı sınıfta olduklarına dair frekans değerleri verilmiştir.

Tablo 4.1: Öğrencilerin sınıflarına göre frekans dağılımı.

Sınıf	Frekans	Yüzde
1	184	45,5
2	85	21,0
3	54	13,4
4	81	20,0
Toplam	404	100,0

Tablo 4.1’deki değerlere göre araştırmaya katılan öğrencilerden 1. sınıfta olanlar (184) en yüksek, 3. sınıfta olanlar (54) en düşük sayıya sahip olanlardır. Yüzdelik dağılımlar incelendiğinde ise sıralama 1. Sınıflar (%45,5), 2. Sınıflar (%21), 4. Sınıflar (%20) ve 3. Sınıflar (%13,4) şeklindedir. Araştırmaya katılan öğrencilerin yaklaşık yarısının 1. sınıf öğrencilerinden oluştuğu görülmektedir.

Şekil 4.1’de Tablo 4.1’deki verilerden elde edilen öğrencilerin sınıflarına göre yüzdelik dağılımlarına ait grafik gösterilmektedir.



Şekil 4.1: Öğrencilerin sınıflarına göre yüzdelik dağılımları.

Tablo 4.2’de araştırmaya katılan öğrencilerin cinsiyetlerine göre frekans değerleri ve yüzdelik dağılımları verilmektedir.

Tablo 4.2: Öğrencilerin cinsiyetlerine göre frekans dağılımları.

Cinsiyet	Frekans	Yüzde
Kız	290	71,8
Erkek	114	28,2
Toplam	404	100,0

Tablo 4.2 incelendiğinde araştırmadaki öğrencilerin cinsiyet dağılımında kız öğrenci sayısının (290) erkek öğrenci sayısına (114) göre daha fazla olduğu görülmektedir.

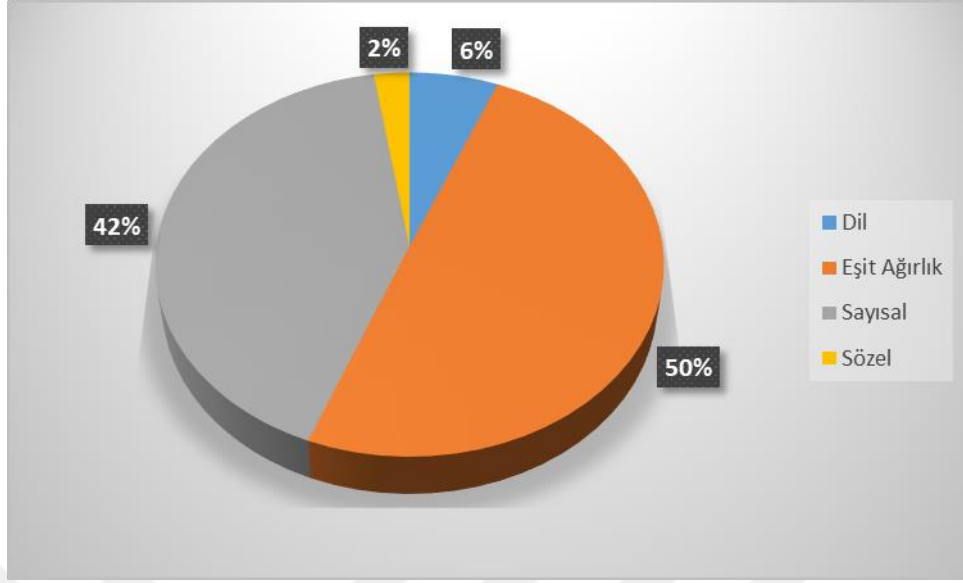
Tablo 4.3’de öğrencilerin okudukların bölümlerin üniversite yerleştirme puan türlerine göre sınıflandırılmış biçimde dağılımları verilmiştir.

Tablo 4.3: Öğrencilerin üniversite yerleştirme puan türlerine göre frekans dağılımları.

Puan Türü	Frekans	Yüzde
Dil	25	6,2
Eşit Ağırlık	201	49,8
Sayısal	168	41,6
Sözel	10	2,5
Toplam	404	100,0

Tablo 4.3 incelendiğinde araştırmadaki katılımcıların %6,2'sini Dil puan türü ile yerleşilen bölümlerde, %49,8'ini Eşit Ağırlık puan türü ile yerleşilen bölümlerde, %41,6'sını Sayısal puan türü ile yerleşilen bölümlerde ve %2,5'ini Sözel puan türü ile yerleşilen bölümlerde kayıtlı öğrenciler oluşturmaktadır. Araştırmada veri toplanan öğrencilerin büyük bir kısmının (toplam %91,3) Sayısal ve Eşit Ağırlık puan türü ile yerleşilen bölümlerde kayıtlı olduğu görünürken, Sözel ve Dil puan türü ile yerleşilen bölümlerde kayıtlı olan öğrencilerin toplam 35 (%8,7) kişilik bir azınlığı oluşturduğu görülmektedir.

Şekil 4.2'de Tablo 4.3'deki verilerden elde edilen öğrencilerin üniversite yerleştirme puan türlerine göre yüzdelik dağılımlarına ait grafik gösterilmektedir.



Şekil 4.2: Üniversite yerleştirme puan türlerine göre yüzdelik dağılım.

Şekil 4.2'deki pasta grafiğinde %50'lik bir payla katılımcı sayısının en yoğun olduğu alanın eşit ağırlık puan türünden olduğu görülmektedir.

Tablo 4.4'de araştırmaya katılan tüm öğrencilerin verilerin toplandığı derse ait her dönemin sonunda aldıkları notlarına göre sınıflandırılmış başarı düzeylerine ait frekans dağılımları verilmiştir.

Tablo 4.4: Öğrencilerin dönem sonu başarı düzeyine göre frekans dağılımları.

Başarı Durumu	Frekans	Yüzde
Başarısız	102	25,2
Başarılı	302	74,8
Toplam	404	100,0

Tablo 4.4 incelendiğinde toplam 404 katılımcıdan 102'sinin dönem sonlarında dersten başarısız olduğu, 302'sinin ise dersten başarılı sayılan bir not ile geçtiği görülmektedir. Başarısız ve

başarılı öğrenciler arasında yaklaşık 1/3'lük bir oran bulunmakta ve başarılı öğrencilerin sayısının daha fazla olduğu görülmektedir.

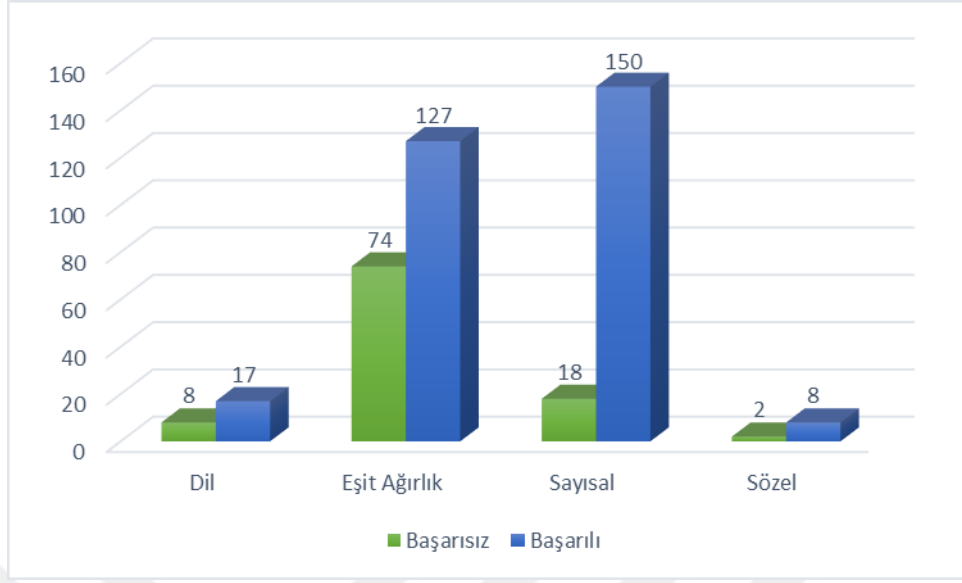
Tablo 4.5'te araştırmaya katılan öğrencilerin kayıtlı oldukları bölümlere yerleştikleri puan türlerine göre başarı durumlarının frekans dağılımları ve başarı yüzdeleri verilmiştir.

Tablo 4.5: Öğrencilerin yerleştikleri puan türlerine göre başarı durumlarının frekans dağılımları.

		Başarı Durumu			
		Başarısız	Başarılı	Başarı Yüzdesi	Toplam
Puan Türü	Dil	8	17	%68	25
	Eşit Ağırlık	74	127	%63,2	201
	Sayısal	18	150	%89,3	168
	Sözel	2	8	%80	10
Toplam		102	302		404

Tablo 4.5'e göre Eşit Ağırlık puan türü ile yerleşilen bölümlere kayıtlı öğrenciler Temel Bilgisayar Uygulamaları Dersinde en düşük başarı yüzdesine (%63,2) sahip öğrencilerdir. %68'lik oranla en düşük ikinci başarı yüzdesine sahip öğrenciler ise Dil puan türü ile yerleşilen bölümlere kayıtlı öğrenciler olmuştur. Sözel puan türüne göre yerleşilen bölümlere kayıtlı öğrencilerin %80'i başarılı olurken, en yüksek başarı yüzdesine sahip öğrencilerin %89,3 ile Sayısal puan türü ile yerleşilen bölümlere kayıtlı öğrencilerin olduğu görülmektedir.

Şekil 4.3'de Tablo 4.5'deki verilerden elde edilen öğrencilerin üniversite yerleştirme puan türlerine göre başarı durumlarına ait grafik gösterilmektedir.



Şekil 4.3: Öğrenci yerleştirme puanlarına göre başarı durumu grafiği.

4.2. ARAŞTIRMANIN ANA PROBLEMİNE AİT BULGULAR

Bu bölümde araştırmanın ana problemi olan “Öğrencilerin akademik başarıları video izleme davranışlarından elde edilen veriler ile tahmin edilebilir mi?” sorusuna ilişkin bulgular verilmiştir.

Araştırmanın analiz aşamasında başarılı öğrenci sayısı ile başarısız öğrenci sayısının arasında yaklaşık 3 kat gibi bir fark olduğundan verilerde kullanılan her algoritmadan önce oversampling işlemi SmoteClassif komutu ile gerçekleştirilmiştir.

Modellerin başarı performansı değerlendirilirken doğruluk oranı, duyarlılık, kesinlik ve kappa değerleri göz önünde bulundurulabilir. Hata matrisleri (Confusion Matrix) üzerinden bu değerlerden kappa hariç diğerleri hesaplanabilmektedir.

Tablo 4.6’da örnek bir hata matrisi verilmiştir.

Tablo 4.6: Örnek hata matrisi.

		Tahmin	
		1	0
Gerçek	1	DP	YP
	0	YN	DN

Tablo 4.6 üzerinden gidildiğinde doğruluk oranı doğru tahmin edilmiş örneklerin sayısının (DP+DN), toplam örnek sayısına (DP+YP+YN+DN) bölünmesi ile hesaplanmaktadır.

$$Doğruluk = \frac{DP+DN}{DP+YP+YN+DN} \quad (4.1)$$

Duyarlılık değeri doğru tahmin edilmiş pozitif örneklerin sayısının(DP), gerçekte pozitif olan tüm örneklerin sayısına(DP+YP) bölünmesi ile hesaplanmaktadır.

$$Duyarlılık = \frac{DP}{DP+YP} \quad (4.2)$$

Kesinlik değeri doğru tahmin edilmiş pozitif örneklerin sayısının(DP), pozitif olarak tahmin edilmiş toplam örneklerin sayısına(DP+YN) bölünmesi ile hesaplanmaktadır.

$$Kesinlik = \frac{DP}{DP+YN} \quad (4.3)$$

Kappa değeri bir tutarlılık ölçütü olarak tutarsız olan veriler için kalite ölçüm değeri olarak kullanılmasına rağmen, sınıflandırma performansının açıklanması için tek başına uygulanması zordur ve duyarlılık, doğruluk gibi metrikler ile beraber değerlendirilebilir (Wang, et al., 2019). Kappa değeri 0 ile 1 arasında olur. 0.00 ile 0.20 aralığındaki Kappa değeri hafif, 0.21 ile 0.40 aralığındaki Kappa değeri makul, 0.41 ile 0.60 aralığındaki Kappa değeri orta, 0.61 ile 0.80 aralığındaki Kappa değeri sağlam, 0,81 ile 1.00 aralığındaki Kappa değeri ise neredeyse mükemmel olarak ifade edilmektedir (Landis & Koch, 1977).

4.2.1. Rastgele Orman Yöntemi Kullanılarak Gerçekleştirilen Analizlere Ait Bulgular

Veri setinde Rastgele Orman yöntemi kullanılarak oluşturulacak model öncesi veriler başarılı ve başarısız olarak gruplandırıldığında ortaya çıkan dağılım ve sonrasında SmoteClassif komutu ile uygulanan Oversampling işleminin ardından oluşan dağılım Tablo 4.7'deki gibidir.

Tablo 4.7: Rastgele orman yönteminde veri dengesizliğine ait sayılar.

	Oversampling Öncesi	Oversampling Sonrası
Başarısız	102	295
Başarılı	302	302

Tablo 4.7'de görüldüğü üzere tüm veri setine Oversampling işlemi gerçekleştirilmeden önce başarısız öğrenci sayısı 102, başarılı öğrenci sayısı 302'dir. Veri setinin dengelenmesi için uygulanan SmoteClassif komutunda katsayı olarak 2.9 değeri kullanılmış ve veri setindeki yeni dağılım sayıları başarısız 295, başarılı 302 olarak belirlenmiştir.

Tablo 4.8'de eğitim veri seti ile test veri setlerinin başarılı ve başarısız öğrencilere göre dağılımı verilmektedir.

Tablo 4.8: Rastgele orman yönteminde veri setlerinin başarı düzeyine göre dağılımı.

	Eğitim Seti	Test Seti
Başarısız	222	73
Başarılı	227	75

Veri seti rpart kütüphanesi kullanılarak %25'i test veri seti, %75'i eğitim veri seti olarak ayrılmıştır. Tablo 4.8'de eğitim veri setinde 222 başarısız, 227 başarılı öğrenci ve test veri setinde 73 başarısız, 75 başarılı öğrenci olduğu görülmektedir. Veri setleri ayrılmadan önce set.seed() komutu uygulanarak rastgele veri seçimi sağlanmıştır.

Tablo 4.9'da ortaya çıkan modelin hata matrisi verilmektedir.

Tablo 4.9: Rastgele orman yönteminde hata matrisi görünümü.

	Tahmin		
	1	0	
Gerçek	1	59	16
	0	9	64

Model geliştirilirken Caret kütüphanesinden faydalanılmış ve train komutu içerisinde k katlamalı çapraz doğrulama (K-Fold Cross Validation) yöntemi kullanılarak k değeri 10 olarak girilmiştir. Tablo 4.9'a göre model 59 başarılı öğrenciyi ve 64 başarısız öğrenciyi doğru olarak tahmin etmektedir.

Tablo 4.10'da modele ait sonuç değerleri verilmiştir.

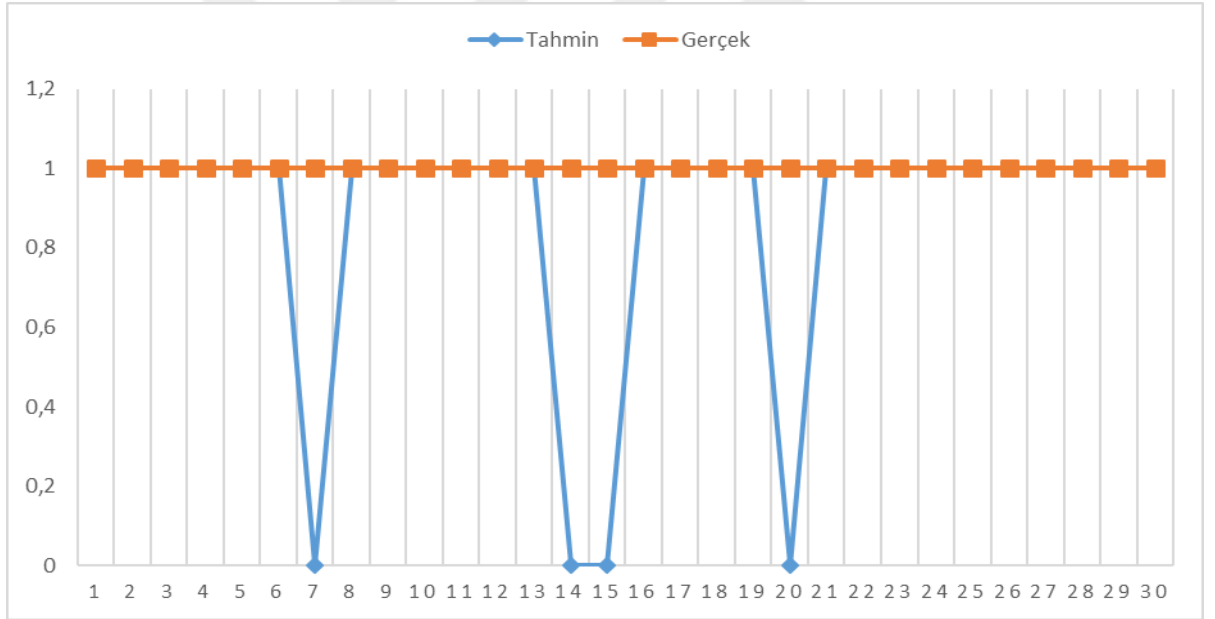
Tablo 4.10: Rastgele orman yöntemine ait sonuç değerleri.

Doğru Sınıflandırılmış Örnek Sayısı	123
Hatalı Sınıflandırılmış Örnek Sayısı	25
Toplam Örnek Sayısı	148
Doğruluk	0.8311
Duyarlılık	0.7867
Kesinlik	0.8676
Kappa Değeri	0.6625
MCC(Matthews korelasyon katsayısı)	0,6655
Pozitif Tahmin Oranı	0.7867
Negatif Tahmin Oranı	0.8767

Tablo 4.10 incelendiğinde modelin toplam 148 test verisinden 123'ünü doğru, 25'ini ise hatalı olarak sınıflandırdığı görülmektedir. Buna bağlı olarak modelin doğruluk oranı $(59+64) / (59+16+9+64) = 0.8311$ (%83,11) olarak hesaplanmıştır. Modelin duyarlılık değeri $59 / (59+16) = 0.7867$, kesinlik değeri ise $59 / (59+9) = 0.8676$ olarak hesaplanmıştır. Kappa değeri 0.6625, MCC değeri ise 0,6655 olarak bulunmuştur.

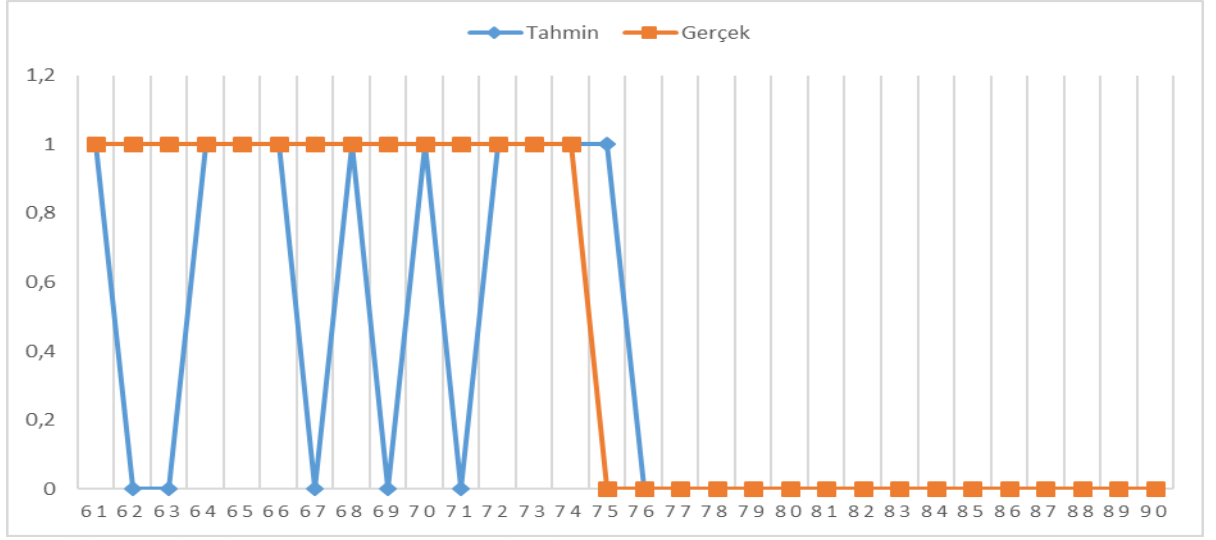
Modelin başarılı öğrencileri doğru tahmin edebilme oranı $59*100/75 = 0.7867$ (%78,67), başarısız öğrencileri doğru tahmin edebilme oranı ise $64*100/73=0.8767$ (%87,67) olarak hesaplanmıştır.

Şekil 4.4'de rastgele orman yöntemiyle oluşturulan modelin 1 ile 30. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.4: Rastgele orman yönteminde 1-30 kayıt aralığı tahminleri.

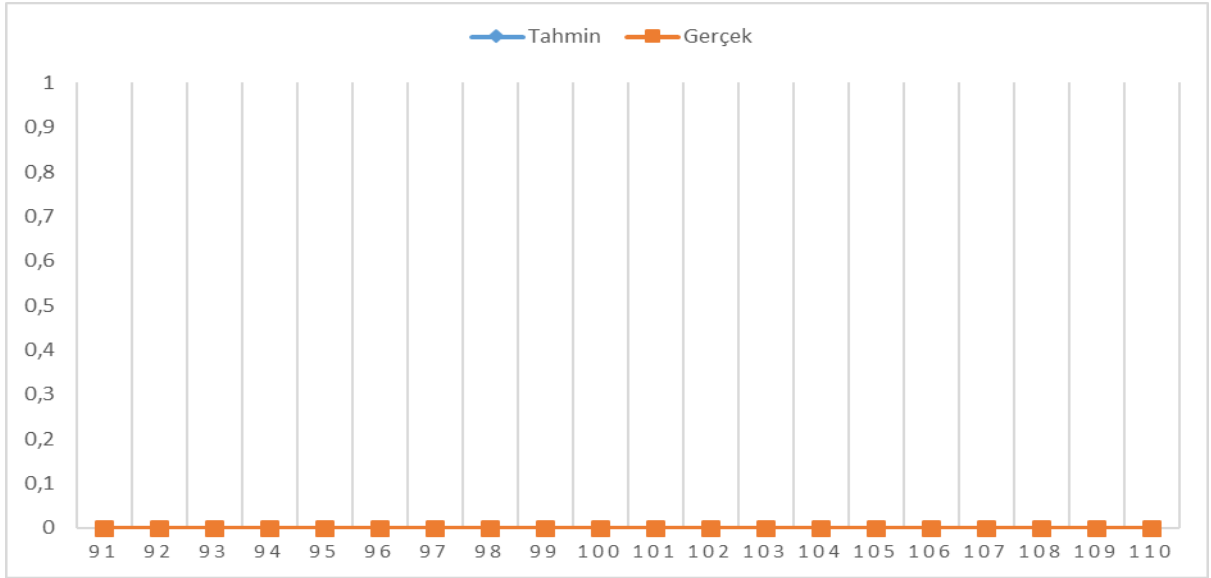
Şekil 4.4'e göre 7,14,15 ve 20 numaralı kayıttaki öğrenciler başarılı(1) olmalarına rağmen algoritma tarafından başarısız (0) olarak tahmin edilmiştir. Geri kalan 26 kayıt ise doğru bir şekilde başarılı (1) olarak tahmin edilmiştir.



Şekil 4.6: Rastgele orman yönteminde 61-90 kayıt aralığı tahminleri.

Şekil 4.6'ya göre 62,63,67,69 ve 71 numaralı kayıttaki öğrenciler başarılı (1) olmalarına rağmen algoritma tarafından başarısız (0) olarak tahmin edilmiş, 75 numaralı kayıttaki öğrenci ise başarısız (0) olmasına rağmen başarılı (1) olarak tahmin edilmiştir. Geri kalan 8 başarılı (1) öğrenci başarılı (1), 15 başarısız (0) öğrenci ise başarısız (0) olarak doğru bir biçimde sınıflandırılmıştır.

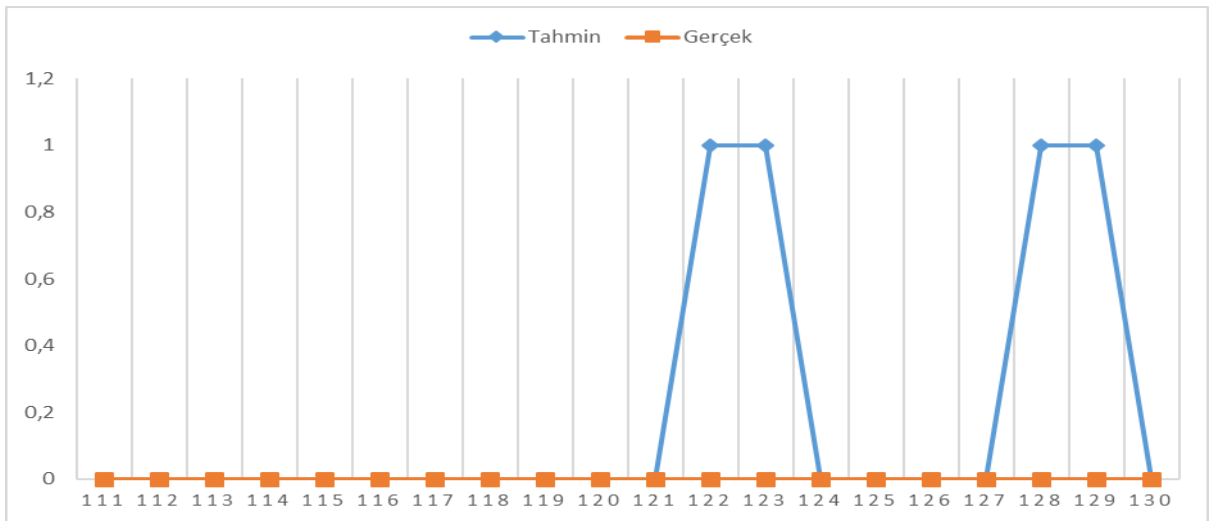
Şekil 4.7'da rastgele orman yöntemiyle oluşturulan modelin 91 ile 110. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.7: Rastgele orman yönteminde 91-110 kayıt aralığı tahminleri.

Şekil 4.7'ye göre toplam 30 öğrencinin tümü başarısızdır (0). Bu grafikte yer alan tüm kayıtlar algoritma tarafından başarısız (0) olarak doğru bir şekilde sınıflandırılmıştır.

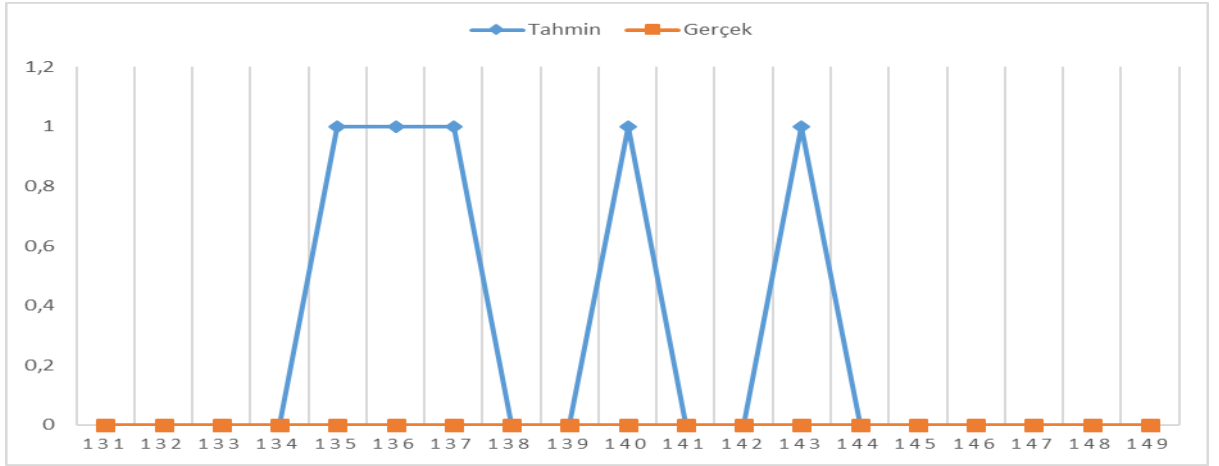
Şekil 4.8'de rastgele orman yöntemiyle oluşturulan modelin 111 ile 130. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.8: Rastgele orman yönteminde 111-130 kayıt aralığı tahminleri.

Şekil 4.8'e göre 122,123,128 ve 129 numaralı kayıttaki öğrenciler başarısız (0) olmalarına rağmen algoritma tarafından başarılı (1) olarak tahmin edilmiştir. Geri kalan 26 kayıt ise doğru bir şekilde başarısız (0) olarak sınıflandırılmıştır.

Şekil 4.9'da rastgele orman yöntemiyle oluşturulan modelin 131 ile 149. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.9: Rastgele orman yönteminde 131-149 kayıt aralığı tahminleri.

Şekil 4.9'a göre 135,136,137,140 ve 143 numaralı kayıttaki öğrenciler başarısız (0) olmalarına rağmen algoritma tarafından başarılı (1) olarak tahmin edilmiştir. Geri kalan 24 kayıt ise doğru bir şekilde başarısız (0) olarak sınıflandırılmıştır.

4.2.2. Naive Bayes Yöntemi Kullanılarak Gerçekleştirilen Analizlere Ait Bulgular

Rastgele Orman yönteminde olduğu gibi oluşturulacak model öncesi veriler başarılı ve başarısız olarak gruplandırıldığında ortaya çıkan dağılım ve sonrasında SmoteClassif komutu ile uygulanan Oversampling işleminin ardından oluşan dağılım Tablo 4.11'deki gibidir.

Tablo 4.11: Naive bayes yönteminde veri dengesizliğine ait sayılar.

	Oversampling Öncesi	Oversampling Sonrası
Başarısız	102	295
Başarılı	302	302

Tablo 4.11’de görüldüğü üzere tüm veri setine Oversampling işlemi gerçekleştirilmeden önce başarısız öğrenci sayısı 102, başarılı öğrenci sayısı 302’dir. Veri setinin dengelenmesi için uygulanan SmoteClassif komutunda katsayı olarak 2.9 değeri kullanılmış ve veri setindeki yeni dağılım sayıları başarısız 295, başarılı 302 olarak belirlenmiştir.

Eğitim veri seti ile test veri setlerinin başarılı ve başarısız öğrencilere göre dağılımı Tablo 4.12’de verilmektedir.

Tablo 4.12: Naive bayes yönteminde veri setlerinin başarı düzeyine göre dağılımı.

	Eğitim Seti	Test Seti
Başarısız	222	73
Başarılı	227	75

Veri seti Rastgele Orman yöntemindeki gibi rpart kütüphanesi kullanılarak %25’i test veri seti, %75’i eğitim veri seti olarak ayrılmıştır. Tablo 4.12’de eğitim veri setinde 222 başarısız, 227 başarılı öğrenci ve test veri setinde 73 başarısız, 75 başarılı öğrenci olarak ortaya çıkmıştır. Veri setleri ayrılmadan önce set.seed() komutu uygulanarak rastgele veri seçimi sağlanmıştır.

Tablo 4.13’te ortaya çıkan Naive Bayes yöntemine ait modelin hata matrisi verilmektedir.

Tablo 4.13: Naive bayes yönteminde hata matrisi görünümü.

		Tahmin		
		1	0	
Gerçek	1	32	43	
	0	16	57	

Model geliştirilirken Caret kütüphanesinden faydalanılmış ve train komutu içerisinde k katlamalı çapraz doğrulama yöntemi kullanılarak k değeri 10 olarak girilmiştir. Tablo 4.13’e

bakıldığında modelin 59 başarılı öğrenciyi ve 64 başarısız öğrenciyi doğru olarak tahmin ettiği görülmektedir.

Tablo 4.14'te modele ait sınıflandırma sayıları, doğruluk, duyarlılık ve kappa gibi değerler verilmiştir.

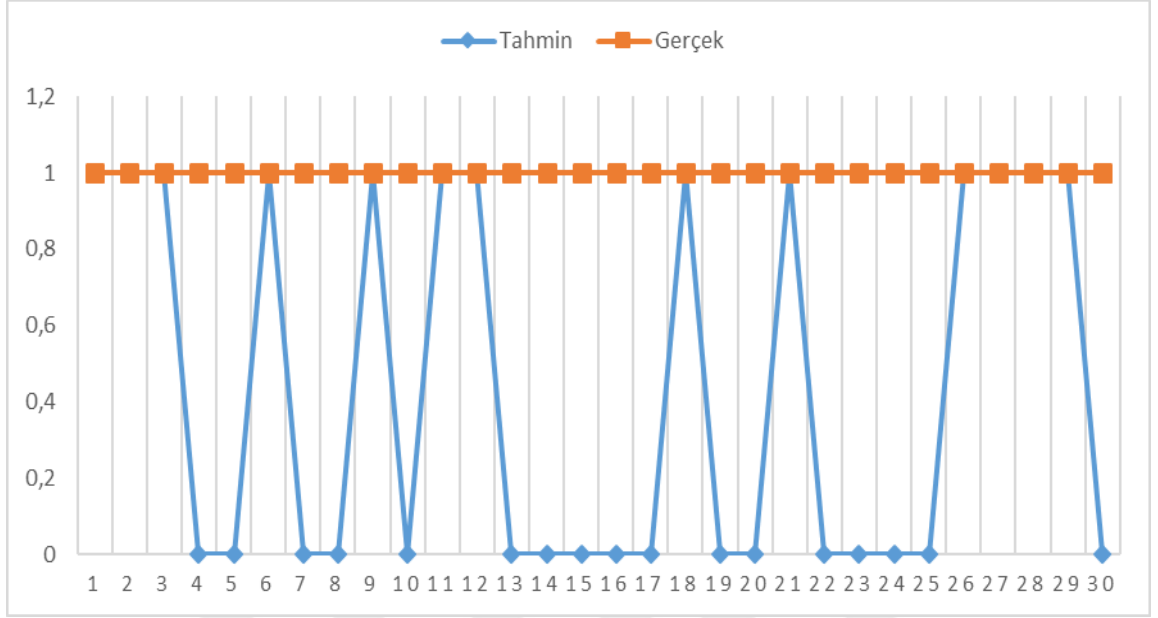
Tablo 4.14: Naive bayes yöntemine ait sonuç değerleri.

Doğru Sınıflandırılmış Örnek Sayısı	89
Hatalı Sınıflandırılmış Örnek Sayısı	59
Toplam Örnek Sayısı	148
Doğruluk	0.6014
Duyarlılık	0.4267
Kesinlik	0.6667
Kappa Değeri	0.2065
MCC	0.2216
Pozitif Tahmin Oranı	0.4267
Negatif Tahmin Oranı	0.7808

Tablo 4.14 incelendiğinde modelin toplam 148 test verisinden 89'unu doğru, 59'unu ise hatalı olarak sınıflandırdığı görülmektedir. Buna bağlı olarak modelin doğruluk oranı $(32+57) / (32+43+16+57) = 0.6014$ (%60,14) olarak hesaplanmıştır. Modelin duyarlılık değeri $32 / (32+43) = 0.4267$, kesinlik değeri ise $32 / (32+16) = 0.6667$ olarak hesaplanmıştır. Kappa değeri 0.2065, MCC değeri ise 0.2216 olarak bulunmuştur.

Modelin başarılı öğrencileri doğru tahmin edebilme oranı $32*100/75 = 0.4267$ (%42,67), başarısız öğrencileri doğru tahmin edebilme oranı ise $57*100/73=0.7808$ (%78,08) olarak hesaplanmıştır.

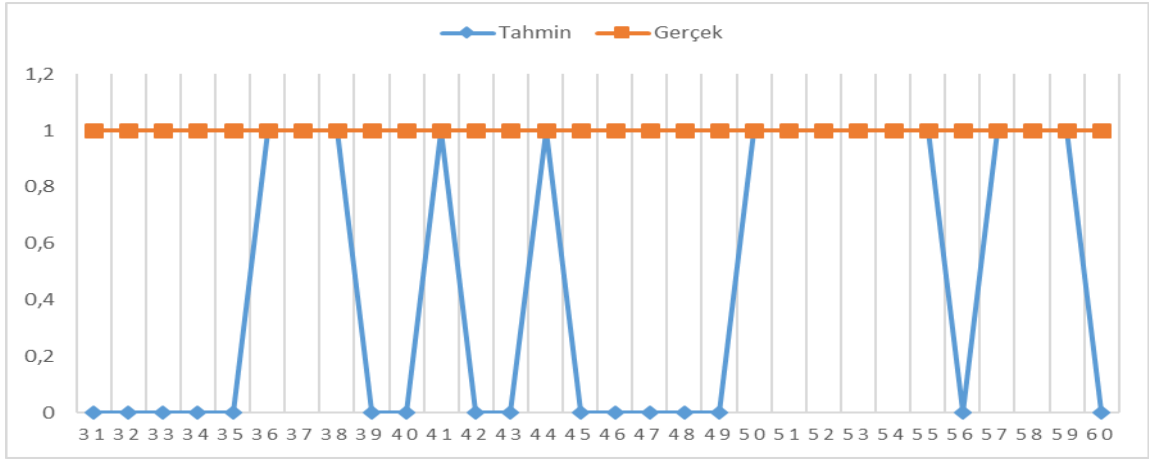
Şekil 4.10’da Naive Bayes yöntemiyle oluşturulan modelin 1 ile 30. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.10: Naive bayes yönteminde 1-30 kayıt aralığı tahminleri.

Şekil 4.10’a göre 4,5,7,8,10,13,14,15,16,17,19,20,22,23,24,25 ve 30 numaralı kayıttaki öğrenciler başarılı (1) olmalarına rağmen algoritma tarafından başarısız (0) olarak tahmin edilmiştir. Geri kalan 13 kayıt ise doğru bir şekilde başarılı (1) olarak sınıflandırılmıştır.

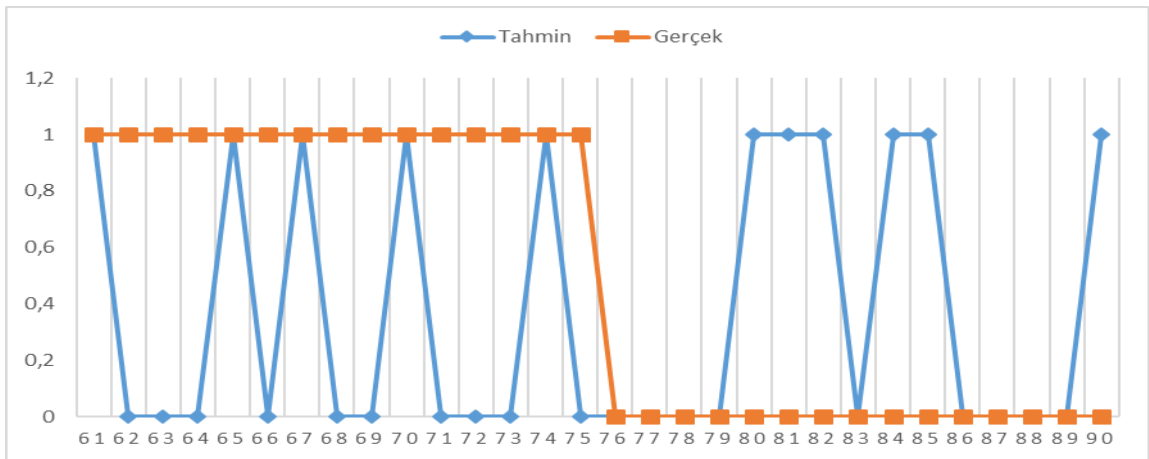
Şekil 4.11’de Naive Bayes yöntemiyle oluşturulan modelin 31 ile 60. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.11: Naive bayes yönteminde 31-60 kayıt aralığı tahminleri.

Şekil 4.11'e göre 31,32,33,34,35,39,40,42,43,45,46,47,48,49,56 ve 60 numaralı kayıttaki öğrenciler başarılı (1) olmalarına rağmen algoritma tarafından başarısız (0) olarak tahmin edilmiştir. Geri kalan 14 kayıt ise doğru bir şekilde başarılı (1) olarak sınıflandırılmıştır.

Şekil 4.12'de Naive Bayes yöntemiyle oluşturulan modelin 61 ile 90. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.

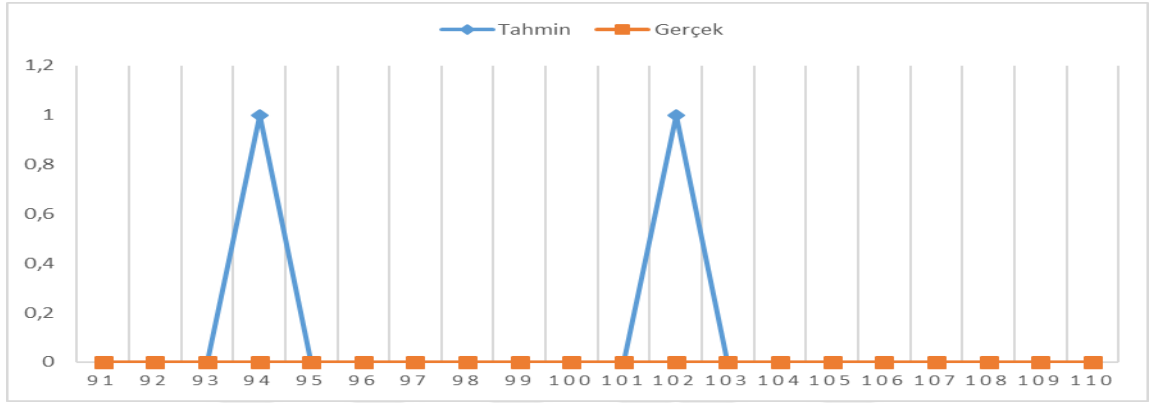


Şekil 4.12: Naive bayes yönteminde 61-90 kayıt aralığı tahminleri.

Şekil 4.12'ye göre 62,63,64,66,68,69,71,72,73 ve 75 numaralı kayıttaki öğrenciler başarılı (1) olmalarına rağmen algoritma tarafından başarısız (0) olarak tahmin edilmiş, 80,81,82,84,85 ve

90 numaralı kayıttaki öğrenciler ise başarısız (0) olmasına rağmen başarılı (1) olarak tahmin edilmiştir. Geri kalan 5 başarılı (1) öğrenci başarılı(1), 9 başarısız (0) öğrenci ise başarısız (0) olarak doğru bir biçimde sınıflandırılmıştır.

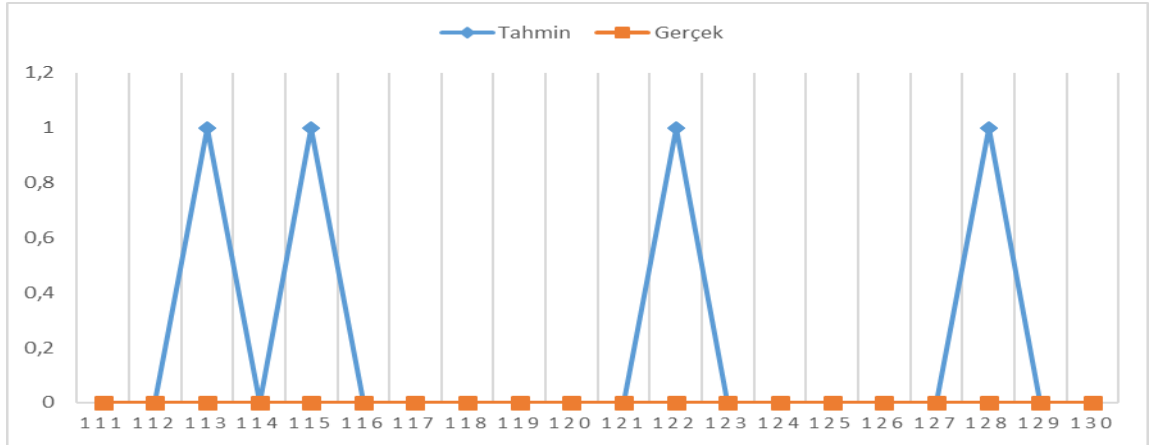
Şekil 4.13’de Naive Bayes yöntemiyle oluşturulan modelin 91 ile 110. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.13: Naive bayes yönteminde 91-110 kayıt aralığı tahminleri.

Şekil 4.13’e göre 94 ve 102 numaralı kayıttaki öğrenciler başarısız (0) olmalarına rağmen algoritma tarafından başarılı (1) olarak tahmin edilmiştir. Geri kalan 18 kayıt ise doğru bir şekilde başarısız (0) olarak sınıflandırılmıştır.

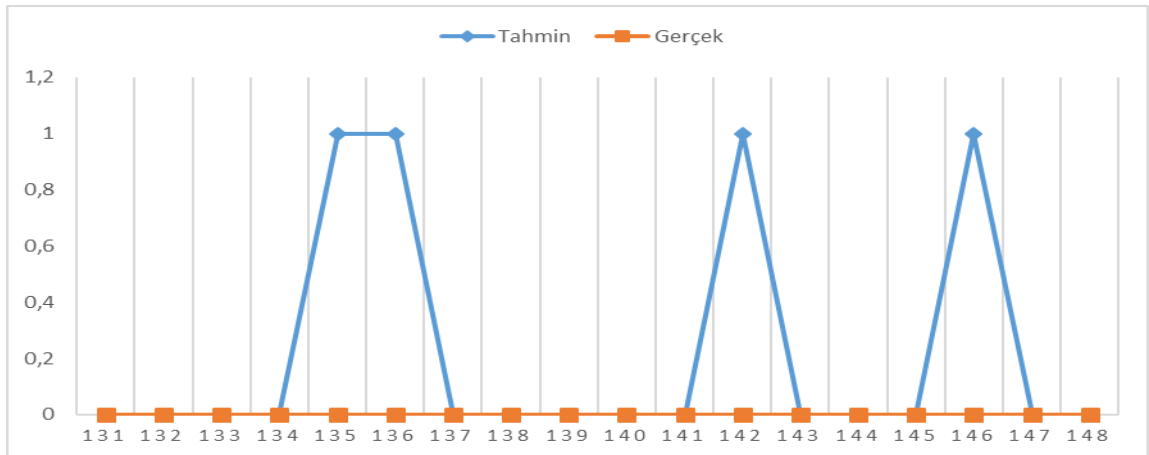
Şekil 4.14’de Naive Bayes yöntemiyle oluşturulan modelin 111 ile 130. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.14: Naive bayes yönteminde 111-130 kayıt aralığı tahminleri.

Şekil 4.14'e göre 113,115,122 ve 128 numaralı kayıttaki öğrenciler başarısız (0) olmalarına rağmen algoritma tarafından başarılı (1) olarak tahmin edilmiştir. Geri kalan 16 kayıt ise doğru bir şekilde başarısız (0) olarak sınıflandırılmıştır.

Şekil 4.15'de Naive Bayes yöntemiyle oluşturulan modelin 131 ile 148. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.15: Naive bayes yönteminde 131-148 kayıt aralığı tahminleri.

Şekil 4.15'e göre 135,136,142 ve 146 numaralı kayıttaki öğrenciler başarısız (0) olmalarına rağmen algoritma tarafından başarılı (1) olarak tahmin edilmiştir. Geri kalan 14 kayıt ise doğru bir şekilde başarısız (0) olarak sınıflandırılmıştır.

4.2.3. Destek Vektör Makineleri Yöntemi Kullanılarak Gerçekleştirilen Analizlere Ait Bulgular

Veri setinde Destek Vektör Makineleri yöntemi kullanılarak oluşturulacak model öncesi veriler başarılı ve başarısız olarak gruplandırıldığında ortaya çıkan dağılım ve sonrasında SmoteClassif komutu ile uygulanan Oversampling işleminin ardından oluşan dağılım Tablo 4.15'deki gibidir.

Tablo 4.15: Destek vektör makineleri yönteminde veri dengesizliğine ait sayılar.

	Oversampling Öncesi	Oversampling Sonrası
Başarısız	102	295
Başarılı	302	302

Tablo 4.15'de görüldüğü üzere tüm veri setine Oversampling işlemi gerçekleştirilmeden önce başarısız öğrenci sayısı 102, başarılı öğrenci sayısı 302'dir. Veri setinin dengelenmesi için uygulanan SmoteClassif komutunda katsayı olarak 2.9 değeri kullanılmış ve veri setindeki yeni dağılım sayıları başarısız 295, başarılı 302 olarak belirlenmiştir.

Tablo 4.16'da eğitim veri seti ile test veri setlerinin başarılı ve başarısız öğrencilere göre dağılımı verilmektedir.

Tablo 4.16: Destek vektör makineleri yönteminde veri setlerinin başarı düzeyine göre dağılımı.

	Eğitim Seti	Test Seti
Başarısız	221	74
Başarılı	226	76

Veri seti caTools kütüphanesi kullanılarak sample.split komutu ile %25'i test veri seti, %75'i eğitim veri seti olarak ayrılmıştır. Tablo 4.16'da eğitim veri setinde 221 başarısız, 226 başarılı öğrenci ve test veri setinde 74 başarısız, 76 başarılı öğrenci olduğu görülmektedir. Veri setleri ayrılmadan önce set.seed() komutu uygulanarak rastgele veri seçimi sağlanmıştır.

Tablo 4.17'de ortaya çıkan modelin hata matrisi verilmiştir.

Tablo 4.17: Destek vektör makineleri yönteminde hata matrisi görünümü.

	Tahmin		
	1	0	
Gerçek	1	50	26
	0	26	48

Model geliştirilirken e1071 kütüphanesinden faydalanılmış ve svm komutu içerisinde sınıflandırma makinesi olarak kullanılması için type değeri C-classification olarak girilmiştir. Ayrıca k katlamalı çapraz doğrulama uygulanması için cross değeri 10 olarak girilmiştir. Tablo 4.17'ye göre model 50 başarılı öğrenciyi ve 48 başarısız öğrenciyi doğru olarak tahmin etmektedir.

Tablo 4.18'de modele ait sonuç değerleri verilmiştir.

Tablo 4.18: Destek vektör makineleri yöntemine ait sonuç değerleri.

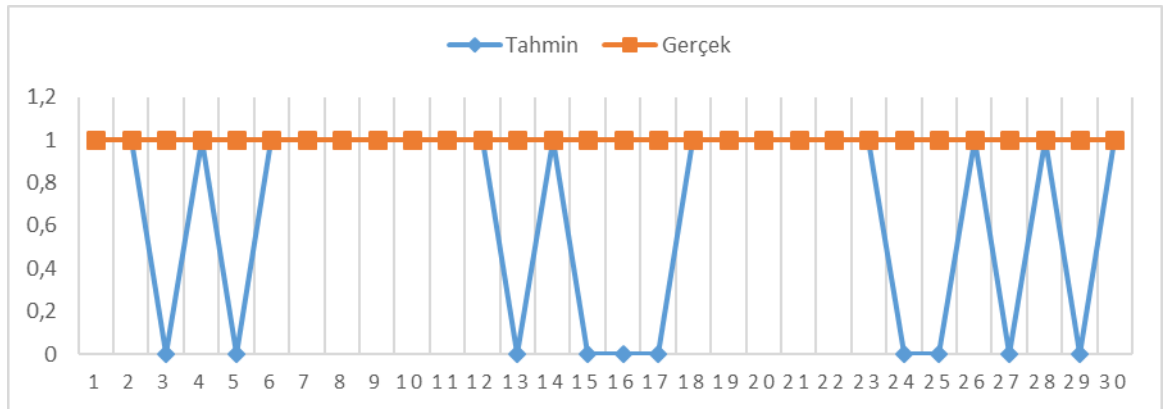
Doğru Sınıflandırılmış Örnek Sayısı	98
Hatalı Sınıflandırılmış Örnek Sayısı	52
Toplam Örnek Sayısı	150
Doğruluk	0.6533
Duyarlılık	0.6579

Kesinlik	0.6579
Kappa Değeri	0.3065
MCC	0.3065
Pozitif Tahmin Oranı	0.6579
Negatif Tahmin Oranı	0.6486

Tablo 4.18 incelendiğinde modelin toplam 150 test verisinden 98'ini doğru, 52'sini ise hatalı olarak sınıflandırdığı görülmektedir. Buna bağlı olarak modelin doğruluk oranı $(50+48) / (50+26+26+48) = 0.6533$ (%65,33) olarak hesaplanmıştır. Modelin duyarlılık değeri $50 / (50+26) = 0.6579$, kesinlik değeri ise $50 / (50+26) = 0.6579$ olarak hesaplanmıştır. Kappa değeri ve MCC değeri ise 0.3065 olarak bulunmuştur.

Modelin başarılı öğrencileri doğru tahmin edebilme oranı $50*100/76 = 0.6579$ (%65,79), başarısız öğrencileri doğru tahmin edebilme oranı ise $48*100/74=0.6486$ (%64,86) olarak hesaplanmıştır.

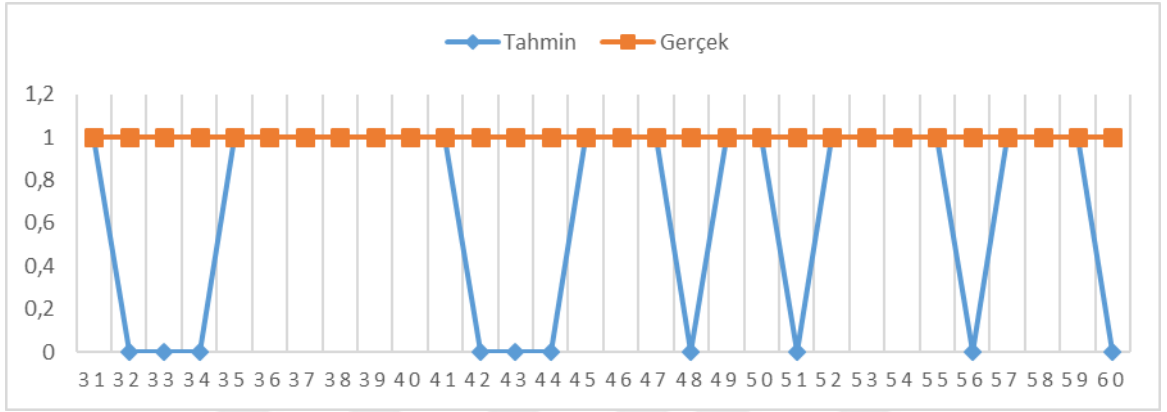
Şekil 4.16'da Destek Vektör Makineleri yöntemiyle oluşturulan modelin 1 ile 30. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.16: Destek vektör makineleri yönteminde 1-30 kayıt aralığı tahminleri.

Şekil 4.16'ya göre 3,5,13,15,16,17,24,25,27 ve 29 numaralı kayıttaki öğrenciler başarılı (1) olmalarına rağmen algoritma tarafından başarısız (0) olarak tahmin edilmiştir. Geri kalan 20 kayıt ise doğru bir şekilde başarılı (1) olarak sınıflandırılmıştır.

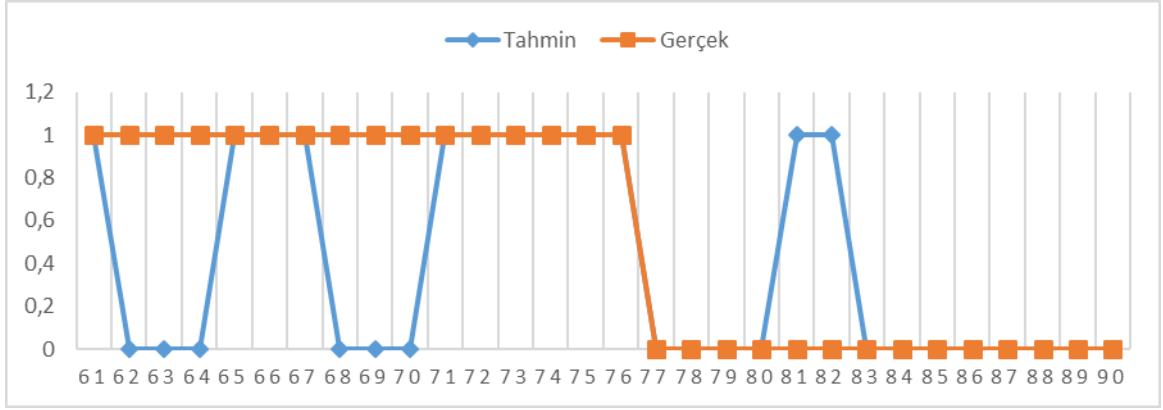
Şekil 4.17'de Destek Vektör Makineleri yöntemiyle oluşturulan modelin 31 ile 60. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.17: Destek vektör makineleri yönteminde 31-60 kayıt aralığı tahminleri.

Şekil 4.17'ye göre 32,33,34,42,43,44,48,51,56 ve 60 numaralı kayıttaki öğrenciler başarılı (1) olmalarına rağmen algoritma tarafından başarısız (0) olarak tahmin edilmiştir. Geri kalan 20 kayıt ise doğru bir şekilde başarılı (1) olarak sınıflandırılmıştır.

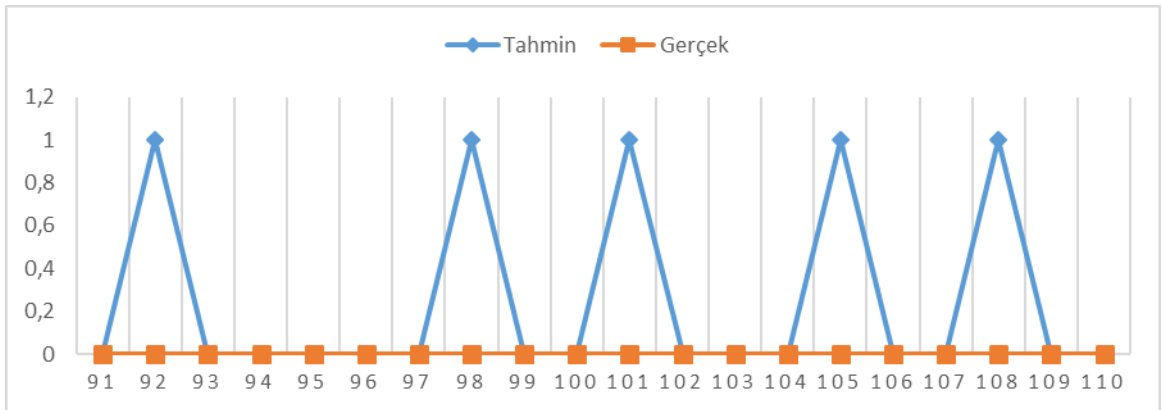
Şekil 4.18'de Destek Vektör Makineleri yöntemiyle oluşturulan modelin 61 ile 90. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.18: Destek vektör makineleri yönteminde 61-90 kayıt aralığı tahminleri.

Şekil 4.18'e göre 62,63,64,68,69 ve 70 numaralı kayıttaki öğrenciler başarılı (1) olmalarına rağmen algoritma tarafından başarısız (0) olarak tahmin edilmiş, 81 ve 82 numaralı kayıttaki öğrenciler ise başarısız (0) olmasına rağmen başarılı (1) olarak tahmin edilmiştir. Geri kalan 10 başarılı (1) öğrenci başarılı (1), 12 başarısız (0) öğrenci ise başarısız (0) olarak doğru bir biçimde sınıflandırılmıştır.

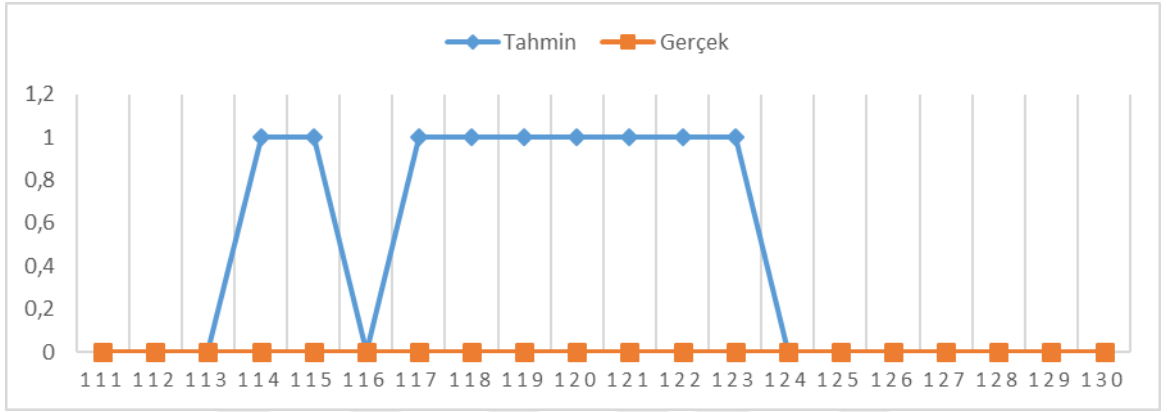
Şekil 4.19'da Destek Vektör Makineleri yöntemiyle oluşturulan modelin 91 ile 110. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.19: Destek vektör makineleri yönteminde 91-110 kayıt aralığı tahminleri.

Şekil 4.19'a göre 92,98,101,105 ve 108 numaralı kayıttaki öğrenciler başarısız (0) olmalarına rağmen algoritma tarafından başarılı (1) olarak tahmin edilmiştir. Geri kalan 15 kayıt ise doğru bir şekilde başarısız (0) olarak sınıflandırılmıştır.

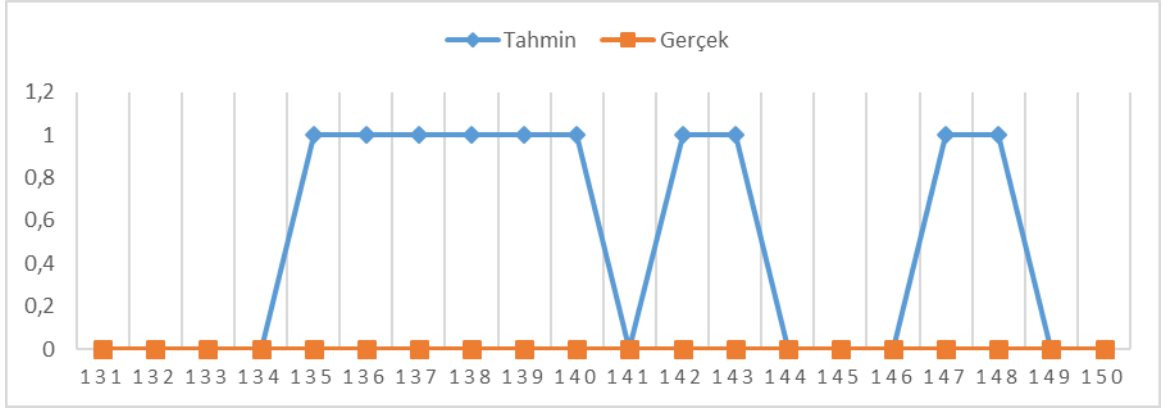
Şekil 4.20'de Destek Vektör Makineleri yöntemiyle oluşturulan modelin 111 ile 130. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.20: Destek vektör makineleri yönteminde 111-130 kayıt aralığı tahminleri.

Şekil 4.20'ye göre 114,115,117,118,119,120,121,122 ve 123 numaralı kayıttaki öğrenciler başarısız (0) olmalarına rağmen algoritma tarafından başarılı (1) olarak tahmin edilmiştir. Geri kalan 11 kayıt ise doğru bir şekilde başarısız (0) olarak sınıflandırılmıştır.

Şekil 4.21'de Destek Vektör Makineleri yöntemiyle oluşturulan modelin 131 ile 150. kayıtlar arasındaki öğrencileri başarılı ya da başarısız olarak tahmin etme düzeyleri grafik halinde gösterilmektedir.



Şekil 4.21: Destek vektör makineleri yönteminde 131-150 kayıt aralığı tahminleri.

Şekil 4.21'e göre 135,136,137,138,139,140,142,143,147 ve 148 numaralı kayıttaki öğrenciler başarısız (0) olmalarına rağmen algoritma tarafından başarılı (1) olarak tahmin edilmiştir. Geri kalan 10 kayıt ise doğru bir şekilde başarısız (0) olarak sınıflandırılmıştır.

4.3. ARAŞTIRMANIN ALT PROBLEMLERİNE AİT BULGULAR

Bu bölümde araştırmanın alt problemlerine ilişkin bulgular başlıklar halinde verilmiştir. Alt problemlerin analizlerinde SPSS programı kullanılarak verilerin normal dağılıma sahip olmadığı görülmüş ve parametrik olmayan Mann-Whitney u testi uygulanmış ve bazı açıklayıcı bulgulara yer verilmiştir.

4.3.1. Açılan Farklı Oturum Sayısı ile Akademik Başarı

Açılan farklı oturum sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.19'da bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.19: Açılan farklı oturum sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	13,91	9,788	4	185

Tablo 4.19'a göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklem farklı oturum açma ortalaması 13,91 ve standart sapması 9,788 olarak hesaplanmıştır. Örneklem içerisinde en az oturum açan ya da açanlar 4 kez oturum açmış, en çok oturum açan ya da açanlar 185 kez oturum açmışlardır.

Tablo 4.20'de uygulanan testin açılan farklı oturum sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.20: Açılan farklı oturum sayısına ait Mann-Whitney u testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
	Başarısız	102	186,06	18978,00
Değerler	Başarılı	302	208,05	62832,00
	Toplam	404		

Tablo 4.20'ye bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin açılan farklı oturum sayılarının sıralama ortalamaları (mean-rank) 186,06 ve başarılı öğrencilerin açılan farklı oturum sayılarının sıralama ortalamaları (mean-rank) ise 208,05 olarak bulunmuştur. Sıralama toplamı (sum of rank) 62832,00 değeri ile başarılı öğrencilerde daha yüksek iken, başarısız öğrenciler de bu değer 18978,00'dır.

Tablo 4.21'de Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.21: Açılan farklı oturum sayısına ait p değeri.

	Değer
Mann-Whitney U	13725,000
Wilcoxon W	18978,000
Z Skoru	-1,650

p değeri	,099
----------	------

Tablo 4.21'e göre yapılan testin Mann-Whitney U değeri 13725,000, Wilcoxon W değeri 18978,000, z skoru ise -1,650 olarak hesaplanmıştır. Açılan farklı oturum sayısına göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p=0,099$).

4.3.2. Toplam Etkileşim Sayısı ile Akademik Başarı

Toplam etkileşim sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.22'de bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.22: Toplam etkileşim sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	402,06	280,368	31	1712

Tablo 4.22'ye göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklem toplam etkileşim sayısı ortalaması 402,06 ve standart sapması 280,368 olarak hesaplanmıştır. Örneklem içerisinde en az etkileşim sayısına sahip olan öğrenci ya da öğrenciler 31 adet etkileşim gerçekleştirmiş, en çok etkileşim sayısına sahip olan öğrenci ya da öğrenciler 1712 adet etkileşim gerçekleştirmiştir.

Tablo 4.23'de uygulanan testin toplam etkileşim sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.23: Toplam etkileşim sayısına ait Mann-Whitney testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
Değerler	Başarısız	102	181,60	18523,00

Başarılı	302	209,56	63287,00
Toplam	404		

Tablo 4.23'e bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin toplam etkileşim sayılarının sıralama ortalamaları (mean-rank) 181,60 ve başarılı öğrencilerin toplam etkileşim sayılarının sıralama ortalamaları (mean-rank) ise 209,56 olarak bulunmuştur. Sıralama toplamı (sum of rank) 63287,00 değeri ile başarılı öğrencilerde daha yüksek iken, başarısız öğrenciler de bu değer 18523,00'dır.

Tablo 4.24'de Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.24: Toplam etkileşim sayısına ait p değeri.

	Değer
Mann-Whitney U	13270,000
Wilcoxon W	18523,000
Z Skoru	-2,091
p değeri	,037

Tablo 4.24'e göre yapılan testin Mann-Whitney U değeri 13270,000, Wilcoxon W değeri 18523,000, z skoru ise -2,091 olarak hesaplanmıştır. Toplam etkileşim sayısına göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık tespit edilmiştir ($p=0,037$).

4.3.3. Video İzlerken Harcanan Zaman ile Akademik Başarı

Video izlerken harcanan zamana göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.25'de bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.25: Video izlerken harcanan zamana ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	259,18	127,685	34	811

Tablo 4.25'e göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklem video izlerken harcanan zaman ortalaması 259,18 ve standart sapması 127,685 olarak hesaplanmıştır. Örneklem içerisinde video izlerken 34 dakika zaman harcayan öğrenci ya da öğrenciler minimum süreyi, en çok zaman harcayan öğrenci ya da öğrenciler ise 811 dakika ile maksimum süreyi temsil etmektedir.

Tablo 4.26'da uygulanan testin video izlerken harcanan zamana ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.26: Video izlerken harcanan zamana ait Mann-Whitney testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
Değerler	Başarısız	102	178,89	18247,00
	Başarılı	302	210,47	63563,00
	Toplam	404		

Tablo 4.26'ya bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin video izlerken harcanan zamana göre sıralama ortalamaları (mean-rank) 178,89 ve başarılı öğrencilerin video izlerken harcanan zamana göre sıralama ortalamaları (mean-rank) ise 210,47 olarak bulunmuştur. Sıralama toplamı (sum of rank) 63563,00 değeri ile başarılı öğrencilerde daha yüksek iken, başarısız öğrenciler de bu değer 18247,00'dır.

Tablo 4.27'de Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.27: Video izlerken harcanan zamana ait p değeri.

Değer	
Mann-Whitney U	12994,000
Wilcoxon W	18247,000
Z Skoru	-2,362
P değeri	,018

Tablo 4.27'ye göre yapılan testin Mann-Whitney U değeri 12994,000, Wilcoxon W değeri 18247,000, z skoru ise -2,362 olarak hesaplanmıştır. Video izlerken harcanana zamana göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık saptanmıştır ($p=0,018$).

4.3.4. Görüntülenen Farklı Video Sayısı ile Akademik Başarı

Görüntülenen farklı video sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.28'de bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.28: Görüntülenen farklı video sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	10,16	1,120	4	11

Tablo 4.28'e göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklemin görüntülenen farklı video sayısı ortalaması 10,16 ve standart sapması 1,120 olarak hesaplanmıştır. Örneklem içerisinde görüntülediği farklı video sayısı 4 olan öğrenci ya da öğrenciler minimum video sayısını, en çok farklı video görüntüleyen öğrenci ya da öğrenciler ise 11 video sayısı ile maksimum görüntülenen farklı video sayısını temsil etmektedir.

Tablo 4.29’da uygulanan testin görüntülenen farklı video sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.29: Görüntülenen farklı video sayısına ait Mann-Whitney testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
	Başarısız	102	173,55	17702,50
Değerler	Başarılı	302	212,28	64107,50
	Toplam	404		

Tablo 4.29’a bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin görüntülenen farklı video sayısına göre sıralama ortalamaları (mean-rank) 173,55 ve başarılı öğrencilerin görüntülenen farklı video sayısına göre sıralama ortalamaları (mean-rank) ise 212,28 olarak bulunmuştur. Başarılı öğrencilerin görüntülenen farklı video sayısına göre sıralama toplamı (sum of rank) 64107,50 değeri ile başarısız öğrencilerin görüntülenen farklı video sayısına göre sıralama toplamı (sum of rank) değerinden (17702,50) daha yüksektir.

Tablo 4.30’da Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.30: Görüntülenen farklı video sayısına ait p değeri.

	Değer
Mann-Whitney U	12449,500
Wilcoxon W	17702,500
Z Skoru	-3,146
P değeri	,002

Tablo 4.30'a göre yapılan testin Mann-Whitney U değeri 12449,500, Wilcoxon W değeri 17702,500, z skoru ise -3,146 olarak hesaplanmıştır. Görüntülenen farklı video sayısına göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık bulunmuştur ($p=0,002$).

4.3.5. Video Yükleme Sayısı ile Akademik Başarı

Video yükleme sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.31'de bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.31: Video yükleme sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	26,92	11,829	6	87

Tablo 4.31'e göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklem video yükleme sayısı ortalaması 26,92 ve standart sapması 11,829 olarak hesaplanmıştır. Örneklem içerisinde 6 kez video yükleme yapan öğrenci ya da öğrenciler minimumu temsil ederken, en çok video yükleme yapan öğrenci ya da öğrenciler 87 kez video yükleme yaparak maksimum değeri temsil etmektedir.

Tablo 4.32'de uygulanan testin video yükleme sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.32: Video yükleme sayısına ait Mann-Whitney testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
Değerler	Başarısız	102	189,71	19350,50
	Başarılı	302	206,82	62459,50
	Toplam	404		

Tablo 4.32'ye bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin video yükleme sayılarının sıralama ortalamaları (mean-rank) 189,71 ve başarılı öğrencilerin video yükleme sayılarının sıralama ortalamaları (mean-rank) ise 206,82 olarak bulunmuştur. Sıralama toplamı (sum of rank) 62459,50 değeri ile başarılı öğrencilerde daha yüksek iken, başarısız öğrenciler de bu değer 19350,50'dir.

Tablo 4.33'de Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.33: Video yükleme sayısına ait p değeri.

Değer	
Mann-Whitney U	14097,500
Wilcoxon W	19350,500
Z Skoru	-1,280
P değeri	,201

Tablo 4.33'e göre yapılan testin Mann-Whitney U değeri 14097,500, Wilcoxon W değeri 19350,500, z skoru ise -1,280 olarak hesaplanmıştır. Video yükleme sayısına göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p=0,201$).

4.3.6. Video Oynatma Sayısı ile Akademik Başarı

Video oynatma sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.34'de bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.34: Video oynatma sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	116,69	86,407	10	623

Tablo 4.34'e göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklemin video oynatma sayısı ortalaması 116,69 ve standart sapması 86,407 olarak hesaplanmıştır. Örneklem içerisinde video oynatma sayısı 10 olan öğrenci ya da öğrenciler minimum oynatma sayısını, en çok video oynatma gerçekleştiren öğrenci ya da öğrenciler ise 623 oynatma ile maksimum video oynatma sayısını temsil etmektedir.

Tablo 4.35'de uygulanan testin video oynatma sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.35: Video oynatma sayısına ait Mann-Whitney testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
	Başarısız	102	182,53	18618,50
Değerler	Başarılı	302	209,24	63191,50
	Toplam	404		

Tablo 4.35'e bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin video oynatma sayısına göre sıralama ortalamaları (mean-rank) 182,53 ve başarılı öğrencilerin video oynatma sayısına göre sıralama ortalamaları (mean-rank) ise 209,24 olarak bulunmuştur. Başarılı öğrencilerin video oynatma sayısına göre sıralama toplamı (sum of rank) 63191,50 değeri ile başarısız öğrencilerin video oynatma sayısına göre sıralama toplamı (sum of rank) değerinden (18618,50) daha yüksektir.

Tablo 4.36'da Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.36: Video oynatma sayısına ait p değeri.

	Değer
Mann-Whitney U	13365,500
Wilcoxon W	18618,500

Z Skoru	-1,997
P değeri	,046

Tablo 4.36'ya göre yapılan testin Mann-Whitney U değeri 13365,500, Wilcoxon W değeri 18618,500, z skoru ise -1,997 olarak hesaplanmıştır. Video oynatma sayısına göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık ortaya çıkarılmıştır ($p=0,046$).

4.3.7. Video Duraklatma Sayısı ile Akademik Başarı

Video duraklatma sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.37'de bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.37: Video duraklatma sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	80,03	63,378	5	393

Tablo 4.37'ye göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklem video duraklatma sayısı ortalaması 80,03 ve standart sapması 63,378 olarak hesaplanmıştır. Örneklem içerisinde 5 kez video duraklatma yapan öğrenci ya da öğrenciler minimumu temsil ederken, en çok video duraklatma yapan öğrenci ya da öğrenciler 393 kez video duraklatma yaparak maksimum değeri temsil etmektedir.

Tablo 4.38'de uygulanan testin video duraklatma sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.38: Video duraklatma sayısına ait Mann-Whitney testi sonuçları.

Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
---------------	---	------------------------	---------------------

	Başarısız	102	183,05	18671,00
Değerler	Başarılı	302	209,07	63139,00
	Toplam	404		

Tablo 4.38'e bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin video duraklatma sayılarının sıralama ortalamaları (mean-rank) 183,05 ve başarılı öğrencilerin video duraklatma sayılarının sıralama ortalamaları (mean-rank) ise 209,07 olarak bulunmuştur. Sıralama toplamı (sum of rank) 63139,00 değeri ile başarılı öğrencilerde daha yüksek iken, başarısız öğrenciler de bu değer 18671,00'dır.

Tablo 4.39'da Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.39: Video duraklatma sayısına ait p değeri.

	Değer
Mann-Whitney U	13418,000
Wilcoxon W	18671,000
Z Skoru	-1,945
P değeri	,052

Tablo 4.39'a göre yapılan testin Mann-Whitney U değeri 13418,000, Wilcoxon W değeri 18671,000, z skoru ise -1,945 olarak hesaplanmıştır. Video duraklatma sayısına göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık tespit edilememiştir ($p=0,052$).

4.3.8. Videonun Sonuna Gelme Sayısı ile Akademik Başarı

Videonun sonuna gelme sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.40'da bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.40: Videonun sonuna gelme sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	8,90	4,266	0	30

Tablo 4.40'a göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklem videonun sonuna gelme sayısı ortalaması 8,90 ve standart sapması 4,266 olarak hesaplanmıştır. Örneklem içerisinde hiçbir videonun sonuna gelmeyen (0) öğrenci ya da öğrenciler minimum video sonuna gelme sayısını, videoların sonuna en çok gelen öğrenci ya da öğrenciler ise 30 sonuna gelme sayısı ile maksimum video sonuna gelme sayısını temsil etmektedir.

Tablo 4.41'de uygulanan testin video sonuna gelme sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.41: Videonun sonuna gelme sayısına ait Mann-Whitney testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
	Başarısız	102	152,66	15571,00
Değerler	Başarılı	302	219,33	66239,00
	Toplam	404		

Tablo 4.41'e bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin video sonuna gelme sayısına göre sıralama ortalamaları (mean-rank) 152,66 ve başarılı öğrencilerin video sonuna gelme sayısına göre

sıralama ortalamaları (mean-rank) ise 219,33 olarak bulunmuştur. Başarılı öğrencilerin video sonuna gelme sayısına göre sıralama toplamı (sum of rank) 66239,00 değeri ile başarısız öğrencilerin video sonuna gelme sayısına göre sıralama toplamı (sum of rank) değerinden (15571,00) daha yüksektir.

Tablo 4.42’de Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.42: Videonun sonuna gelme sayısına ait p değeri.

	Değer
Mann-Whitney U	10318,000
Wilcoxon W	15571,000
Z Skoru	-5,004
P değeri	,000

Tablo 4.42’ye göre yapılan testin Mann-Whitney U değeri 10318,000, Wilcoxon W değeri 15571,000, z skoru ise -5,004 olarak hesaplanmıştır. Video sonuna gelme sayısına göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık bulunmuştur ($p=0,000$).

4.3.9. Toplam Atlama Yapma Sayısı ile Akademik Başarı

Toplam atlama yapma sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.43’de bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.43: Toplam atlama yapma sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	166,00	145,159	2	803

Tablo 4.43'e göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklemin toplam atlama yapma sayısı ortalaması 166,00 ve standart sapması 145,159 olarak hesaplanmıştır. Örneklem içerisinde toplam atlama yapma sayısı 2 olan öğrenci ya da öğrenciler minimum atlama yapma, en çok atlama yapan öğrenci ya da öğrenciler ise 803 atlama ile maksimum toplam atlama yapma sayısını temsil etmektedir.

Tablo 4.44'de uygulanan testin toplam atlama yapma sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.44: Toplam atlama yapma sayısına ait Mann-Whitney testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
	Başarısız	102	185,03	18873,00
Değerler	Başarılı	302	208,40	62937,00
	Toplam	404		

Tablo 4.44'e bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin toplam atlama yapma sayısına göre sıralama ortalamaları (mean-rank) 185,03 ve başarılı öğrencilerin toplam atlama yapma sayısına göre sıralama ortalamaları (mean-rank) ise 208,40 olarak bulunmuştur. Başarılı öğrencilerin toplam atlama yapma sayısına göre sıralama toplamı (sum of rank) 62937,00 değeri ile başarısız öğrencilerin toplam atlama yapma sayısına göre sıralama toplamı (sum of rank) değerinden (18873,00) daha yüksektir.

Tablo 4.45'de Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.45: Toplam atlama yapma sayısına ait p değeri.

	Değer
Mann-Whitney U	13620,000

Wilcoxon W	18873,000
Z Skoru	-1,748
P değeri	,081

Tablo 4.45'e göre yapılan testin Mann-Whitney U değeri 13620,000, Wilcoxon W değeri 18873,000, z skoru ise -1,748 olarak hesaplanmıştır. Toplam atlama yapma sayısına göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık tespit edilememiştir ($p=0,081$).

4.3.10. İleri Atlama Yapma Sayısı ile Akademik Başarı

İleri atlama yapma sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.46'da bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.46: İleri atlama yapma sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	101,08	104,949	0	706

Tablo 4.46'ya göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklem ileri atlama yapma sayısı ortalaması 101,08 ve standart sapması 104,949 olarak hesaplanmıştır. Örneklem içerisinde ileri atlama yapma sayısı 0 olan öğrenci ya da öğrenciler minimum atlama sayısını, en çok ileri atlama yapan öğrenci ya da öğrenciler ise 706 atlama sayısı ile maksimum ileri atlama yapma sayısını temsil etmektedir.

Tablo 4.47'de uygulanan testin ileri atlama yapma sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.47: İleri atlama yapma sayısına ait Mann-Whitney testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
	Başarısız	102	181,90	18553,50
Değerler	Başarılı	302	209,46	63256,50
	Toplam	404		

Tablo 4.47'ye bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin ileri atlama yapma sayısına göre sıralama ortalamaları (mean-rank) 181,90 ve başarılı öğrencilerin ileri atlama yapma sayısına göre sıralama ortalamaları (mean-rank) ise 209,46 olarak bulunmuştur. Başarılı öğrencilerin ileri atlama yapma sayısına göre sıralama toplamı (sum of rank) 63256,50 değeri ile başarısız öğrencilerin ileri atlama yapma sayısına göre sıralama toplamı (sum of rank) değerinden (18553,50) daha yüksektir.

Tablo 4.48'de Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.48: İleri atlama yapma sayısına ait p değeri.

	Değer
Mann-Whitney U	13300,500
Wilcoxon W	18553,500
Z Skoru	-2,061
P değeri	,039

Tablo 4.48'e göre yapılan testin Mann-Whitney U değeri 13300,500, Wilcoxon W değeri 18553,500, z skoru ise -2,061 olarak hesaplanmıştır. İleri atlama yapma sayısına göre başarılı

ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık bulunmuştur ($p=0,039$).

4.3.11. Geri Atlama Yapma Sayısı ile Akademik Başarı

Geri atlama yapma sayısına göre başarı durumunun anlamlı bir farklılık gösterip göstermediği incelendiğinde ilk olarak Tablo 4.49'da bu niteliğe ait açıklayıcı bulgulara yer verilmiştir.

Tablo 4.49: Geri atlama yapma sayısına ait açıklayıcı analiz sonuçları.

	N	Ortalama	Standart Sapma	Minimum	Maksimum
Değer	404	63,00	70,261	0	552

Tablo 4.49'a göre toplam örneklem sayısı (N) 404 olarak bulunmuştur. Bu örneklem geri atlama yapma sayısı ortalaması 63,00 ve standart sapması 70,261 olarak hesaplanmıştır. Örneklem içerisinde hiçbir geri atlama yapmayan (0) öğrenci ya da öğrenciler minimum geri atlama yapma sayısını, videolarda en çok atlama yapan öğrenci ya da öğrenciler ise 552 geri atlama sayısı ile maksimum video geri atlama yapma sayısını temsil etmektedir.

Tablo 4.50'de uygulanan testin geri atlama yapma sayısına ve başarı durumuna göre sonuçları listelenmektedir.

Tablo 4.50: Geri atlama yapma sayısına ait Mann-Whitney testi sonuçları.

	Başarı Durumu	N	Sıralama Ortalaması	Sıralama Toplamı
	Başarısız	102	192,75	19660,00
Değerler	Başarılı	302	205,79	62150,00
	Toplam	404		

Tablo 4.50'ye bakıldığında başarısız öğrenci sayısının 102, başarılı öğrenci sayısının ise 302 olduğu görülmektedir. Başarısız öğrencilerin geri atlama yapma sayısına göre sıralama ortalamaları (mean-rank) 192,75 ve başarılı öğrencilerin geri atlama yapma sayısına göre sıralama ortalamaları (mean-rank) ise 205,79 olarak bulunmuştur. Başarılı öğrencilerin geri atlama yapma sayısına göre sıralama toplamı (sum of rank) 62150,00 değeri ile başarısız öğrencilerin geri atlama yapma sayısına göre sıralama toplamı (sum of rank) değerinden (19660,00) daha yüksektir.

Tablo 4.51'de Mann-Whitney u testinin sonuçlarına göre p değeri verilmektedir.

Tablo 4.51: Geri atlama yapma sayısına ait p değeri.

	Değer
Mann-Whitney U	14407,000
Wilcoxon W	19660,000
Z Skoru	-,976
P değeri	,329

Tablo 4.51'e göre yapılan testin Mann-Whitney U değeri 14407,000, Wilcoxon W değeri 19660,000, z skoru ise -,976 olarak hesaplanmıştır. Geri atlama yapma sayısına göre başarılı ve başarısız öğrenciler arasında yapılan hesaplamalarda istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p=0,329$). |

5. TARTIŞMA VE SONUÇ

Öğretim sürecindeki değişimler çerçevesinde öğrenmenin daha verimli gerçekleşmesi amacıyla çevrimiçi öğrenme ortamlarının kullanımı ve önemi artmıştır. Bu ortamlarda öğrenme ve öğretmeyi kolaylaştıracak birçok araç geliştirilmiştir. Geliştirilen araçlar öğrencilerin çevrimiçi ortamlardaki davranışları hakkında büyük miktarda veri üretmektedir. Çevrimiçi öğrenme ortamlarındaki bu verilerin öğrenme ve öğretme süreçlerinin iyileştirilmesi için nasıl kullanılabileceği sorusu eğitimcileri ve araştırmacıları veriler ile çalışmaya yöneltmiştir. Elde edilen verilerin doğru analiz edilmesi ve içerisindeki örüntülerin ortaya çıkarılarak eğitsel problemlere çözümler üretmek için veri madenciliği tekniklerinin ve öğrenme analitiği yaklaşımlarının kullanılması ihtiyaç haline gelmiştir. Bu ihtiyaç öğretim ortamlarından elde edilen verilerin kullanılmasıyla eğitsel veri madenciliği alanını doğurmuştur. Eğitsel veri madenciliği çalışmaları kullanılan öğrenme ortamlarının iyileştirilmesi amacıyla gerçekleştirildiği gibi öğrencilerin risk durumlarının ve başarılarının tahmin edilmesi amacıyla da yürütülmektedir. Öğrencilerin büyük çaptaki veri setlerinin eğitsel veri madenciliği alanında veri madenciliği algoritmaları ile analiz edilmesinin yüksek oranda tahmin başarısına sahip modeller geliştirmede büyük önemi vardır (Chung & Sunbok, 2019). Öğrenciler hakkında büyük verilerin elde edilmesi için genellikle veri tabanları, öğrenci bilgi sistemleri ya da çevrimiçi öğrenme ortamları kullanılmaktadır. Buralardan elde edilen veriler kullanılarak tahmin başarısı yüksek modeller geliştirmek pek mümkün olmamaktadır. Tek başına öğrenme yönetim sistemlerinden elde edilen veriler kullanılarak yapılan tahminlere oranla elde edilecek demografik ve kurumsal verilerin de dahil edilmesiyle yapılan tahminler daha yüksek doğruluk sağlayabilir (Yu, et al., 2020). Kurumsal veri tabanlarındaki verilerin kurum tarafından araştırmacılarla paylaşılmasındaki zorluklar ve etik kaygılar bu verilerin çalışmalara dahil edilmesinde ve alanın gelişmesinde engel teşkil etmektedir. Bunun yerine araştırmalardaki analizler çevrimiçi öğrenme ortamlarından gelen öğrencilerin dijital ayak izi verilerinin yanında demografik veriler ile desteklenmektedir.

Çevrimiçi ortamlardan genellikle öğrencilerin bu ortamlarda gerçekleştirdikleri etkileşim verileri elde edilmekte olup geliştirilen bazı araçlar ile sınav esnasındaki fare hareketleri ya da video izleme davranışları gibi daha ayrıntılı veriler de kaydedilebilmektedir. Öğrenci başarısının tahmin edilmesi için genellikle öğrencilerin önceki dönem başarı durumlarına ait verilerin kullanıldığı görülse de dersler bazında erken dönem tahminleri geliştirebilmek için

çevrimiçi ortamlardan elde edilen farklı etkileşim verilerinin kullanılmaya başlandığı görülmektedir. Bu çalışmada da öğrenme süreçlerindeki problemlerden biri olan öğrenci başarısının tahmin edilmesi ele alınmıştır.

Kullanılan veriler bir öğrenme yönetim sistemi üzerinden yayınlanan videolardan Bayazıt ve Akçapınar (2018) tarafından geliştirilen araç sayesinde elde edilen öğrencilerin video izleme davranışlarına ait verilerdir. Araştırmacı, ters yüz öğrenme modeli ile yürüttüğü Temel Bilgisayar Uygulamaları dersine kayıtlı toplam 404 öğrencinin 4 dönem boyunca video izleme verilerini kaydetmiştir. 14 hafta süren her bir dönem için araştırmacı tarafından hazırlanmış toplam 11 ders videosu kullanılmıştır ve bu süreçte öğrencilerin öğrenme süreçleri ile ilişkisi olduğu düşünülen, öğrenme yönetim sistemi üzerinden 11 adet, toplanan demografik veriler ve uygulanan ölçekten 6 adet olmak üzere toplam 17 adet nitelik üretilmiştir. Bu nitelikler veri madenciliği teknikleri ile analiz edilmiş ve öğrencilerin ders sonu başarı performansı modellenmeye çalışılmıştır.

Araştırmada ilk olarak alt problemlere yanıt olacak, çevrimiçi öğrenme ortamından elde edilen her bir niteliğin öğrenci başarısı ile olan ilişkisi araştırılmıştır. Yapılan analizlerde daha fazla toplam etkileşim, görüntülenen farklı video, video oynatma, videonun sonuna gelme, ileri atlama yapma sayılarına sahip ve video izlerken daha fazla zaman harcayan öğrencilerin başarı performanslarının daha iyi olduğu ortaya konmuştur. Temel olarak çevrimiçi öğrenme ortamlarında etkileşim sayıları yüksek olan öğrencilerin derslerden başarılı olması beklenen bir durumdur. Başarı performansı ile ilişki tespit edilen nitelikler incelendiğinde ileri atlama yapma sayısı hariç diğerlerinin öğrencinin öğrenme süreci içerisinde daha fazla zaman geçirdiğini göstermektedir. Daha fazla ileri atlama gerçekleştiren öğrencilerin başarılı olması ise öğrencilerin videoları tekrar izlerken özüksedikleri konuların anlatıldığı süreleri geçmek istemesi ile açıklanabilir. Buradan yola çıkarak öğrencilerin video izleme davranışlarından elde edilen 11 adet nitelikten 6 tanesinin öğrenci başarısı ile ilişkili olduğu söylenebilir. Buna istinaden sınıflandırma algoritmalarında demografik veriler ile beraber bu 6 nitelik ile analizler gerçekleştirilmiştir. Fakat sınıflandırma algoritmalarının performanslarının daha yüksek olması amacıyla mümkün olduğunca çok nitelik ve veriye ihtiyaç duyulduğu yapılan çalışmalardan gözlemlenerek daha sonra tüm niteliklerin dahil olduğu veri seti üzerinde analiz işlemleri tekrar gerçekleştirilmiştir.

Araştırmada ana problem, öğrencilerin akademik başarılarının video izleme davranışlarından elde edilen veriler ile tahmin edilmesi üzerinedir. Yapılan analizlerde benzer araştırmalarda en çok kullanılan veri madenciliği sınıflandırma algoritmaları tercih edilerek eldeki veriler üzerinde en iyi sonucu veren algoritmanın keşfedilmesi amaçlanmıştır. Analizlerde öğrenciler başarı durumu tahmin etme amacıyla başarılı ve başarısız olmak üzere ikiye ayrılmıştır. Verilerin dengesiz dağılımı görülmüş ve SMOTE tekniği ile sentetik veriler üretilerek verilerdeki dengesiz dağılım giderilmiştir. Alanda yapılan çalışmalarda en çok kullanılan üç sınıflandırma algoritması tercih edilmiş ve her biri için sonuçlar incelenmiştir.

Genel tahmin performansı en iyi olan algoritma %83,11 ile Rastgele Orman algoritması olarak bulunmuştur. Ayrıca başarısız öğrencileri en iyi tahmin etmede %87,67 ve başarılı öğrencileri en iyi tahmin etmede %78,67 oranlarıyla yine Rastgele Orman algoritması en iyi performansı gösteren algoritma olarak öne çıkmıştır. Diğer iki algoritmadan Naive Bayes algoritmasının tahmin oranları Destek Vektör Makinelerine göre daha başarılı olsa da Rastgele Orman algoritmasının performansına erişememiştir.

Analiz sonuçlarına göre öğrencilerin derslerden dönem sonunda başarılı olup olamayacaklarını final sınavına girmeden önce video izleme davranışlarını analiz ederek %83,11 oranında doğru tahmin edebilecek bir model geliştirilmiştir. Modelin özellikle başarısız öğrencileri %87,67 gibi bir oranda doğru tahmin ediyor olması bu öğrencilerin sınav öncesi gerekli önlemler alınarak ders başarısını arttırması açısından oldukça önemli görülmektedir.

Araştırmada kullanılan verilerin öğrenci başarısını etkileyebilecek daha fazla nitelik ile desteklenerek daha yüksek tahmin performansları elde edilmesi mümkündür. Bu çalışmanın ileride gerçekleştirilecek öğrenci başarısı tahmini ile alakalı ve özellikle günümüz şartlarında zorunlu olarak yürütülen uzaktan eğitim sürecinde elde edilebilecek veriler kullanılarak gerçekleştirilebilecek diğer araştırmalara ilham vereceği düşünülmektedir.

Araştırmada öğrenci başarısı ile ilişkisi bulunamayan niteliklerin ileriki çalışmalarda farklı araştırma sorularına yanıt bulmak amacıyla incelenmesi mümkündür. Ayrıca öğrencilerin video izleme davranışları ile ilgili zamansal kayıtlar da veri olarak tutulmuş fakat araştırmanın amacı doğrultusunda bu veriler sayısal değerlere dönüştürülerek kullanılmıştır. Daha sonraki çalışmalarda öğrencilerin video izleme örüntüleri zamansal olarak çıkarılarak başarıları ile ilişkisi araştırılabilir.

Ters yüz sınıf ortamında gerçekleştirilen bu araştırmada çevrimiçi ortamdan elde edilen veriler kullanılırken yüz yüze dersler de yapılmıştır. Dolayısıyla uzaktan eğitim sürecinde sadece çevrimiçi ortamda yürütülen derslerden elde edilen veriler analiz edilerek yeni bir çalışma yapılabilir.

Araştırmada tek bir ders için 4 farklı dönem boyunca veri toplanmıştır. Veri toplama süresi oldukça uzunken ders içeriğinin de öğrenci başarısı üzerinde etkisi olabileceği hesaba katılarak farklı alanlardaki birden fazla dersten toplanan veriler analiz edilerek çalışmalar gerçekleştirilebilir.

Öğrencilerden toplanan video izleme verileri öğrenciler hakkında çıkarımlar yapılabilmesinin yanında eğitimcilerin video ders hazırlama becerilerinde iyileştirmeler yapılabilmesi için de kullanılabilir. |

6. KAYNAKLAR

- Açıřlı, S., 2016. Sınıf Öğretmeni Adaylarının Öğrenme Stilleri ile Eleřtirel Düşünme Eğilimlerinin İncelenmesi. *İlköğretim Online*, 15(1), pp. 273-285.
- Akar, Ö. & Güngör, O., 2012. Classification of Multispectral Images Using Random Forest Algorithm. *Journal of Geodesy and Geoinformation*, 1(2), pp. 105-112.
- Akçapınar, G., Altun, A. & Ařkar, P., 2015. Modeling Students' Academic Performance Based on Their Interactions in an Online Learning Environment. *İlköğretim Online*, 14(3), pp. 815-824.
- Akçapınar, G. & Bayazıt, A., 2019. MoodleMiner: Moodle Öğrenme Yönetim Sistemi için Veri Madencilięi Analiz Aracı. *İlköğretim Online*, 18(1), pp. 406-415.
- Akçapınar, G. & Cořgun, E., 2019. Öğrencilerin STEM Kariyer Tercihlerinin Veri Madencilięi Yaklařımı ile Tahmin Edilmesi. *Eğitim Teknolojisi Kuram ve Uygulama*, 9(1), pp. 73-88.
- Akçapınar, G. ve dięerleri, 2019. Developing an Early-Warning System for Spotting at-Risk Students by Using eBook Interaction Logs. *Smart Learning Environments*, 6(1), pp. 1-15.
- Akçayır, G. & Akçayır, M., 2018. The Flipped Classroom: A Review of Its Advantages and Challenges. *Computers & Education*, 126(2018), pp. 334-345.
- Aksoy, T., Çelik, S. & Gülseçen, S., 2020. Data Pre-processing in Text Mining. %1 içindeS. Gülseçen, S. Sharma & E. Akadal, düz. *Who Runs The World: Data*. İstanbul: İstanbul University Press, p. 124.
- 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(2019), pp. 13-49.
- Algayres, M. & Triantafyllou, E., 2020. Learning Analytics in Flipped Classrooms: A Scoping Review. *Electronic Journal of e-Learning*, 18(5), pp. 397-409.

- Asif, R., Merceron, A., Ali, S. A. & Haider, N. G., 2017. Analyzing Undergraduate Students' Performance Using Educational Data Mining. *Computers & Education*, 113(2017), pp. 177-194.
- Baker, R., 2010. Data Mining for Education. *International Encyclopedia of Education*, 7(3), pp. 112-118.
- Batory, D. S. ve diğeri, 1988. Genesis: An Extensible Database Management System. *IEEE Transactions on Software Engineering*, 14(11), pp. 1711-1730.
- Bayazıt, A. & Akçapınar, G., 2018. Çevrimiçi Dersler İçin Video Analitik Aracının Tasarlanması ve Geliştirilmesi. *İlköğretim Online*, 17(1), pp. 14-25.
- BBC, 2018. [www.bbc.com](https://www.bbc.com/turkce/haberler-dunya-43463045). [Çevrimiçi]
Available at: <https://www.bbc.com/turkce/haberler-dunya-43463045>
[Erişildi: 2 Mayıs 2020].
- Berka, P. & Bruha, I., 1998. *Discretization and Grouping: Preprocessing Steps for Data Mining*. Berlin, Springer, pp. 239-245.
- Berry, M. J. A. & Linoff, G. S., 2004. *Data Mining Techniques For Marketing, Sales and Customer Relationship Management*. 2. dü. Indianapolis: Wiley Computer Publishing.
- Bosancic, B., 2020. Information, Data, and Knowledge in the Cognitive System of the Observer. *Journal of Documentation*, 76(4), pp. 893-894.
- Bousbia, N. & Belamri, I., 2014. Which Contribution Does EDM Provide to Computer-Based Learning Environments?. %1 içindeA. Pena-Ayala, dü. *Educational Data Mining*. basım yeri bilinmiyor:Springer, pp. 3-28.
- Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), pp. 5-32.
- Bulathwela, S. ve diğeri, 2020. *Predicting Engagement in Video Lectures*. Virtual Conference, International Educational Data Mining Society, pp. 50-60.
- Calders, T. & Pechenizkiy, M., 2012. Introduction to The Special Section on Educational Data Mining. *SIGKDD Explorations*, 13(2), pp. 3-6.

- Canobi, K. H., 2009. Concept-Procedure Interactions in Children's Addition and Subtraction. *Journal of Experimental Child Psychology*, 102(2), pp. 131-149.
- Chung, J. Y. & Sunbok, L., 2019. Dropout Early Warning Systems For High School Students Using Machine Learning. *Children and Youth Services Review*, 96(2019), pp. 346-353.
- Demirci, Ç., 2020. *Dijital Çağda Seçin Özgürlüğü: Facebook ve Büyük Veri*, İstanbul: yazarı bilinmiyor
- Denkel, A., 1998. *Bilginin Temelleri*. İstanbul: Metris Yayınları.
- Devasia, T., T P, V. & Hedge, V., 2016. *Prediction of Students Performance Using Educational Data Mining*. Ernakulam, IEEE, pp. 91-95.
- Doğan, K. & Arslantekin, S., 2016. Büyük Veri: Önemi, Yapısı ve Günümüzdeki Durum. *Ankara Üniversitesi Dil ve Tarih - Coğrafya Fakültesi Dergisi*, 56(1), pp. 15-36.
- Duan, Y. ve diğerleri, 2017. Data, Information and Knowledge-Driven Manipulation Between Strategical Planning and Technical Implementation for Wireless Sensor Network Construction. *International Journal of Distributed Sensor Networks*, 13(11), pp. 1-16.
- Duan, Y. ve diğerleri, 2019. Modeling Data, Information and Knowledge for Security Protection of Hybrid IoT and Edge Resources. *IEEE Access*, Cilt 7, pp. 99161 - 99176.
- Duman, B., 2008. Öğrencilerin Benimsedikleri Eğitim Felsefeleriyle Kullandıkları Öğrenme Strateji ve Öğrenme Stillerinin Karşılaştırılması. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 17(1), pp. 203-224.
- Ekici, G., 2013. Gregorc ve Kolb Öğrenme Stili Modellerinne göre Öğretmen Adaylarının Öğrenme Stillerinin Cinsiyet ve Genel Akademik Başarı Açısından İncelenmesi. *Eğitim ve Bilim*, 38(167), pp. 211-225.
- Evin Gencel, İ., 2007. KOLB'un Deneyimsel Öğrenme Kuramına Dayalı Öğrenme Stilleri Envanteri-III'ü Türkçeye Uyarlama Çalışması. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(2), pp. 120-139.

- Fernandes, E. ve diğerleri, 2019. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, 94(2019), pp. 335-343.
- Ferrier, A. & Stangret, C., 1982. Heterogeneity in the Distributed Database Management System Sirius-Delta. *VLDB*, pp. 45-53.
- Flanagan, B. ve diğerleri, 2019. *Knowledge Map Creation for Modeling Learning Behaviors in Digital Learning Environments*. Arizona, SoLAR, pp. 428-436.
- Flipped Learning Network, 2014. *Definition of Flipped Learning*. [Çevrimiçi] Available at: <https://flippedlearning.org/definition-of-flipped-learning/> [Erişildi: 14 Kasım 2020].
- Gren, L., 2020. A Flipped Classroom Approach to Teaching Empirical Software Engineering. *IEEE Transactions on Education*, 63(3), pp. 155-163.
- Güldal, H. & Çakıcı, Y., 2017. *Eğitsel Veri Madenciliği*. Edirne, Trakya Üniversitesi Yayınları, pp. 135-143.
- Gülseçen, S., 2014. Bir Değer Olarak Bilgi ve Bilginin Yönetimi. *Türk Kütüphaneciliği*, 28(1), pp. 62-68.
- Güven, M. & Kürüm, D., 2004. *Öğrenme Stilleri ve Eleştirel Düşünme Arasındaki İlişkiye Genel Bir Bakış*. Malatya, Pegem.
- Hand, D. J., Mannila, H. & Smyth, P., 2001. *Principles of Data Mining*. 1. dü. Cambridge: MIT Press.
- Han, J., Kamber, M. & Pei, J., 2006. *Data Mining Concepts and Techniques*. 2. dü. San Francisco: Morgan Kaufmann.
- Hanna, M., 2004. Data Mining in the E-learning Domain. *Campus-Wide Information Systems*, 21(1), pp. 29-34.
- Hasnine, M. N. ve diğerleri, 2018. *Towards Final Scores Prediction over Clickstream Using Machine Learning Methods*. Philippines, Asia-Pacific Society for Computers in Education.

Haye, A. & Torres-Sahli, M., 2017. To Feel is to Know Relations: James' Concept of Stream of Thought and Contemporary Studies on Procedural Knowledge. *New Ideas in Psychology*, Cilt 46, pp. 46-55.

Hewitt, S., 2019. Data, Information, and Knowledge. *Journal of Histochemistry & Cytochemistry*, 67(4), pp. 227-228.

Hicham, A., Jeghal, A., Sabri, A. & Tairi, H., 2020. A Survey on Educational Data Mining [2014-2019]. Fez, Morocco, IEEE, pp. 1-6.

Hung, H.-G., Liu, I.-F., Liang, C.-T. & Su, Y.-S., 2020. Applying Educational Data Mining to Explore Students' Learning Patterns in the Flipped Learning Approach for Coding Education. *Symmetry*, 12(2), p. 213.

Kalıpsız, O., 2001. *Bilgisayar Veri Tabanı Sistemleri*. 1. Baskı dü. İstanbul: Der Yayınları.

Kanehisa, M. ve diğerleri, 2014. Data, Information, Knowledge and Principle: Back to Metabolism in KEGG. *Nucleic Acids Research*, 42(1), p. 1.

Kelleher, J. & Tierney, B., 2018. *Data Science*. 1 dü. Massachusetts: MIT Press.

Kettinger, W. & Li, Y., 2010. The Infological Equation Extended: Towards Conceptual Clarity in the Relationship Between Data, Information and Knowledge. *European Journal of Information Systems*, 19(4), pp. 409-421.

Knobbout, J. & Stappen, E. v. d., 2020. Where is the Learning in Learning Analytics? A Systematic Literature Review on the Operationalization of Learning-Related Constructs in the Evaluation of Learning Analytics Interventions. *Transactions on Learning Technologies*, 13(3), pp. 631-645.

Kokoç, M., Ilgaz, H. & Altun, A., 2018. *Understanding How Video Interaction Data Predicts Academic Performance: A Preliminary Study*. Kansas City, ERIC, pp. 99-103.

Ktona, A., Xhaja, D. & Ninka, I., 2014. *Extracting Relationships between Students' Academic Performance and Their Area of Interest Using Data Mining Techniques*. Bhopal-India, Communication Systems and Networks, pp. 6-11.

Kulaklı, G., 2020. *Twitter Veri Paylaşım Politikasını Güncelledi: Kullanıcı Verileri Satılacak*, İstanbul: yazarı bilinmiyor

Lage, M. J., Platt, G. & Treglia, M., 2000. Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment. *The Journal of Economic Education*, 31(1), pp. 30-43.

Lal, P., 2014. Designing Online Learning Strategies through Analytics. %1 içindeJ. Garcia-Penalvo & A. M. S. Pardo, düz. *Online Tutor 2.0: Methodologies and Case Studies for Successful Learning*. basım yeri bilinmiyor:IGI Global, pp. 1-15.

Landis, R. & Koch, G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), pp. 159-174.

Lenz, K., Dreher, A., Holzapfel, L. & Wittmann, G., 2019. Are conceptual knowledge and procedural knowledge empirically separable? The Case of Fractions. *British Journal of Educational Psychology*, pp. 1-21.

Lim, K., Ahmed, P. & Zairi, M., 1999. Managing for Quality Through Knowledge Management. *Total Quality Management & Business Excellence*, 10(4-5), pp. 615-621.

Linan, L. C. & Perez, A. A. J., 2015. Educational Data Mining and Learning Analytics: differences, similarities, and yime evolution. *International Journal of Educational Technology in Higher Education*, 12(3), pp. 98-112.

Liu, S., 2020. *statista.com*. [Çevrimiçi]
Available at: <https://www.statista.com/statistics/809750/worldwide-popularity-ranking-database-management-systems/>
[Erişildi: 02 Temmuz 2020].

Merceron, A. & Yacef, K., 2005. *Educational Data Mining: a Case Study*. Amsterdam, International Conference on Artificial Intelligence in Education, pp. 1-8.

Monino, J. L. & Sedkaoui, S., 2016. *Big Data, Open Data and Data Development*. 3. dü. London: ISTE.

Munguia, P., Brennan, A., Taylor, S. & Lee, D., 2020. A Learning Analytics Journey: Bridging the Gap Between Technology Services and the Academic Need. *The Internet and Higher Education*, 46(2020), pp. 1-9.

Ok, K., 2013. Bilgi ve Bilgi Yönetimine Giriş. %1 içindeS. Gülseçen, dü. *Bilgi ve Bilginin Yönetimi*. İstanbul: Papatya Yayıncılık, pp. 20-21.

Olney, A., 2020. *Journal of Educational Data Mining*. [Çevrimiçi] Available at: <https://jedm.educationaldatamining.org/index.php/JEDM> [Erişildi: 2 Ekim 2020].

Özbay, Ö., 2015. Veri Madenciliği Kavramı ve Eğitimde Veri Madenciliği Uygulamaları. *Uluslararası Eğitim Bilimleri Dergisi*, 2(5), pp. 262-272.

Özbay, Ö. & Ersoy, H., 2017. Öğrenme Yönetim Sistemi Üzerindeki Öğrenci Hareketliliğinin Veri Madenciliği Yöntemleriyle Analizi. *Gazi Eğitim Fakültesi Dergisi*, 37(2), pp. 523-558.

Özçakır, F., 2013. Yeni Bilgi Keşfi: Veri Madenciliği. %1 içinde *Bilgi ve Bilginin Yönetimi*. İstanbul: Papatya Yayıncılık Eğitim, pp. 72-73.

Özekes, S., 2003. Veri Madenciliği Modelleri ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Dergisi*, 2003(3), pp. 65-82.

Padhy, N., Mishra, P. & Panigrahi, R., 2012. The Survey of Data Mining Applications And Feature Scope. *International Journal of Computer Science*, 2(3), pp. 43-58.

Paquette, L. ve diğerleri, 2020. Who's Learning? Using Demographics in EDM Research. *Journal of Educational Data Mining*, 12(3), pp. 1-30.

Pehlivan, F. & Arabacıoğlu, T., 2020. 2010-2018 Yılları Arasında Türkiye'de Yapılmış Dönüştürülmüş Sınıf Araştırmalarının İçerik Analizi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 21(2), pp. 894-911.

Pena-Ayala, A., 2018. Learning Analytics: A Glance of Evolution, Status, and Trends According to a Proposed Taxonomy. *Data Mining & Knowledge Discovery*, 8(3), pp. 1-29.

- Qu, X. S. & Jiang, Z., 2019. A Time-Based Dynamic Synchronization Policy for Consolidated Database Systems. *MIS Quarterly*, 43(4), pp. 1041-1057.
- Razak, R. A., Omar, M. & Ahmad, M., 2018. A Student Performance Prediction Model Using Data Mining Technique. *International Journal of Engineering & Technology*, 7(2.15), pp. 61-63.
- 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*, 40(6), pp. 601-618.
- Romero, C. & Ventura, S., 2020. Educational Data Mining and Learning Analytics: An Updated Survey. *Data Mining and Knowledge Discovery*, 10(3), pp. 1-21.
- Romero, C., Ventura, S. & Garcia, E., 2008. Data Mining in Course Management Systems: Moodle Case Study and Tutorial. *Computers & Education*, Issue 51, pp. 368-384.
- Russell, J.-E., Smith, A. & Larsen, R., 2020. Elements of Success: Supporting at-risk Student Resilience Through Learning Analytics. *Computers & Education*, 152(2020), p. 103890.
- Sachin, B. & Vijay, S., 2012. *A Survey and Future Vision of Data Mining in Educational Field*. Rohtak, Institute of Electrical and Electronics Engineers.
- Sanjeev, A. & Zytow, J., 1995. *Discovering Enrollment Knowledge in University Databases*. Montreal, Canada, AAAI Press, pp. 246-251.
- Santamaria-Bonfil, G. ve diğerleri, 2020. Learning Analytics for Student Modeling in Virtual Reality Training Systems: Lineworkers Case. *Computers & Education*, 151(2020), p. 103871.
- Savaş, S., Topaloğlu, N. & Yılmaz, M., 2012. Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), pp. 1-23.
- Siemens, G. & Gasevic, D., 2012. Guest Editorial - Learning and Knowledge Analytics. *Journal of Educational Technology & Society*, 15(3), pp. 1-2.

Society, I. E. D. M., 2020. *educationaldatamining.org*. [Çevrimiçi]
Available at: <https://educationaldatamining.org/conferences/>
[Erişildi: 02 Ekim 2020].

Strelan, P., Osborn, A. & Palmer, E., 2020. The Flipped Classroom: A Meta-analysis of Effects on Student Performance Across Disciplines and Education Levels. *Educational Research Review*, 30(2020), p. 100314.

Sun, D. & Li, Y., 2019. Effectiveness of Digital Note-Taking on Students' Performance in Declarative, Procedural and Conditional Knowledge Learning. *International Journal of Emerging Technologies in Learning*, 14(18), pp. 108-119.

Şahin, M. & Yurdugül, H., 2020. Educational Data Mining and Learning Analytics: Past, Present and Future. *Bartın University Journal of Faculty of Education*, 9(1), pp. 121-131.

Şengür, D. & Tekin, A., 2013. Öğrencilerin Mezuniyet Notlarının Veri Madenciliği Metotları ile Tahmini. *Bilişim Teknolojileri Dergisi*, 6(3), pp. 7-16.

Tekin, A., 2014. Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach. *Eurasian Journal of Educational Research*, 14(54), pp. 207-226.

Tekin, A. & Öztekin, Z., 2018. Eğitsel Veri Madenciliği ile İlgili 2006-2016 Yılları Arasında Yapılan Çalışmaların İncelenmesi. *Eğitim Teknolojisi Kuram ve Uygulama*, 8(2), pp. 67-89.

TRT, 2019. www.trthaber.com. [Çevrimiçi]
Available at: <https://www.trthaber.com/haber/dunya/twitterda-veri-skandali-434720.html>

Uçak, N. Ö., 2010. Bilgi: Çok Yüzlü Bir Kavram. *Türk Kütüphaneciliği*, 24(4), pp. 705-722.

Wang, D. & Han, H., 2020. Applying Learning Analytics Dashboards Based on Process-Oriented Feedback to Improve Students' Learning Effectiveness. *Journal of Computer Assisted Learning*, pp. 1-13.

Wang, J., Yang, Y. & Xia, B., 2019. A Simplified Cohen's Kappa for Use in Binary Classification Data Annotation Tasks. *IEEE Access*, Cilt 7, pp. 164386-164397.

Wuryaningrum, R., Bektiarso, S. & Imam, S., 2020. The Effect of Knowledge-Transforming Text on Elementary Students' Declarative, Procedural Knowledge, and Motivation in Environmental Learning. *International Journal of Instruction*, 13(1), pp. 567-586.

Yalçın, Ş., Çeçen, S. & Erçetin, G., 2015. Dil Öğrenme Yatkınlığının Tanımlanmasında İşler Bellek ve Açık/Örtük Bilgi Türlerinin Rolü. *Boğaziçi Üniversitesi Eğitim Dergisi*, 32(2), pp. 1-17.

Yıldırım, B. F. & Özdemirci, F., 2019. Kurumlarda Örtük Bilginin Yapay Zeka Destekli Tavsiye Sistemleri Aracılığıyla Ortaya Çıkarılması. *Bilgi Yönetimi Dergisi*, 2(1), pp. 34-43.

Yu, R. ve diğerleri, 2020. *Towards Accurate and Fair Prediction of College Success: Evaluating Different Sources of Student Data*. Sanal Konferans, International Educational Data Mining Society, pp. 292-301.

Zander, U. & Kogut, B., 1995. Knowledge and the Speed of the Transfer and Imitation of Organizational Capabilities: An Empirical Test. *Organization Science*, 6(1), pp. 76-92.

Zhao, Y., Xu, Q., Chen, M. & Weiss, G. M., 2020. *Predicting Student Performance in a Master of Data Science Program using Admissions Data*. Sanal Konferans, International Educational Data Mining Society, pp. 325-333.

Zins, C., 2007. Conceptual Approaches for Defining Data, Information, and Knowledge. *Journal of the American Society for Information Science and Technology*, 58(4), pp. 479-493.

Zou, D., Luo, S., Xie, H. & Hwang, G.-J., 2020. A Systematic Review of Research on Flipped Language Classrooms: Theoretical Foundations, Learning Activities, Tools, Research Topics and Findings. *Computer Assisted Language Learning*.

7. EKLER

GÖNÜLLÜ KATILIM / ONAM FORMU

1. Araştırmayla İlgili Bilgiler:

- Araştırmanın Adı: Öğrencilerin Video İzleme Davranışlarının Makine Öğrenmesi Yöntemleri ile İncelenerek Başarı Düzeylerinin Modellenmesi
- Araştırmanın Öngörülen Süresi: 14 Hafta
- Araştırmaya Katılması Beklenen Katılımcı/Gönüllü Sayısı: 450
- Araştırmada İzlenecek Deneyisel İşlemler: Ön-Test, Son-Test, Gözlem

2. Gönüllülük, Çalışmayı Reddetme ve Çalışmadan Çekilme Hakkı, Çalışmadan Çıkarılma:

- Araştırmaya hiçbir baskı ve zorlama altında olmaksızın gönüllü olarak katılıyorum.
- Araştırmaya katılmayı reddetme hakkına sahip olduğum bana bildirildi.
- Sorumlu araştırmacıya haber vermek kaydıyla, hiçbir gerekçe göstermeksizin istediğim anda bu çalışmadan çekilebileceğimin bilincindeyim.
- Çalışmanın yürütücüsü olan araştırmacı ya da destekleyen kuruluş, çalışma programının gereklerini yerine getirmedeki ihmali nedeniyle ya da araştırma prosedürüne bağlı olarak onayımı almadan beni çalışma kapsamından çıkarabilir.

3. Gizlilik:

Çalışma süresince tutulan bütün kayıtlar ve dosya bilgileri araştırma kapsamında kullanılacak ve gerektiğinde kamuoyu ile paylaşılabilir. Ancak, bu tür durumlarda kimliğim kesin olarak gizli tutulacaktır.

4. Çalışmaya Katılma Onayı:

Yukarıda yer alan ve araştırmadan önce gönüllüye / katılımcıya verilmesi gereken bilgileri gösteren Aydınlatılmış Onam Formu adlı metni kendi anadilimde okudum ya da bana okunmasını sağladım. Bu bilgilerin içeriği ve anlamı, yazılı ve sözlü olarak açıklandı. Aklıma gelen bütün soruları sorma olanağı tanındı ve sorularıma doyurucu cevaplar aldım. Çalışmaya katılmadığım ya da katıldıktan sonra çekildiğim durumda, hiçbir yasal hakkımdan vazgeçmiş olmayacağım. Bu koşullarla, söz konusu araştırmaya hiçbir baskı ve zorlama olmaksızın gönüllü olarak katılmayı kabul ediyorum.

Gönüllünün / katılımcının	Açıklamaları yapan araştırmacının
Adı - Soyadı:	Adı - Soyadı: Kadir Burak Olgun
İmzası:	İmzası:
Tarih:	Tarih:

Ad-Soyad :

KOLB ÖĞRENME STİL ENVANTERİ

Aşağıda her biri için 4'er cümle bulunan 12 durum söz konusudur. Her durum için size en uygun olan cümleye 4, ikinci uygun olana 3, üçüncü uygun olana 2 ve en az uygun olana 1 yazınız. Her bir durum için 1-4 arası tüm sayıların kullanılması zorunludur. Teşekkürler.

4	En uygun olan	3	İkinci uygun olan
2	Üçüncü uygun olan	1	En az uygun olan

1. Öğrenirken ...,

- Duygularımı da öğrenmeye katarım.
 ----- Öğrendiğim fikirler üzerinde düşünmeyi severim.
 ----- Bir şeyler yapıyor olmaktan hoşlanırım.
 ----- İzlemekten ve dinlemekten hoşlanırım.

2. En iyi öğrenme yolum...,

- Dikkatle dinlemek ve izlemektir.
 ----- Kendi mantığımla yorumlamaktır.
 ----- Duygularıma ve sezgilerime güvenmektir.
 ----- Çok çalışıp bir şeyleri başarmaktır.

3. Öğrenirken...,

- Mantığıma uygun olan sonucu bulmaya çalışırım.
 ----- Öğrenmede sorumlu olduğumu hissedirim.
 ----- Derse katılmadan sessizce izlerim.
 ----- Derse yoğun bir şekilde katılırım.

4. En iyi...,

- Duygularımı öğrenirim.
 ----- Yaparak öğrenirim.
 ----- İzleyerek öğrenirim.
 ----- Fikirler üzerinde düşünerek öğrenirim.

5. Öğrenirken...,

- Konuyla ilgili yeni bilgilere/fikirlere açığım.
 ----- Konuyu her yönüyle/ayrıntılarıyla ele alırım.
 ----- Konuyu kendi içinde küçük bölümlere ayırırım.
 ----- Konuyla ilgili öğrendiğim şeyleri yapmaktan/uygulamaktan hoşlanırım.

6. Öğrenirken...,

- Gözlem yapan biriyim.
 ----- Öğrenmeye katılan biriyim.
 ----- Duygularıyla hareket eden biriyim.
 ----- Mantıklı davranan biriyim.

Ad-Soyad :

7. En iyi öğrenme yolum...,
----- Konuyla ilgili gözlem yapmaktır.
----- İnsanlarla konuyla ilgili konuşmak, iletişim kurmaktır.
----- Konunun dayandığı temel fikirleri düşünmektir.
----- Konuyla ilgili deneme ve uygulama yapmaktır.
8. Öğrenirken...,
----- Çalışmamın sonuçlarını görmekten hoşlanırım.
----- Konuyla ilgili temel fikirleri düşünmeyi severim.
----- Acele etmekten hoşlanmam.
----- Kendimi tamamen öğrenme işinin içinde hissederim.
9. En iyi öğrenme yolum...,
----- İzlemektir.
----- Hissettiklerimi dikkate almaktır.
----- Öğrendiklerimi uygulamaktır.
----- Kendi düşüncelerimi dikkate almaktır.
10. Öğrenirken...,
----- Çekingen biri olurum.
----- Öğrendiklerimi sorgulamadan kabul ederim.
----- Sorumluluklarını bilen biriyim.
----- Öğrendiğim şeyler üzerinde düşünen biriyim.
11. Öğrenirken...,
----- Derse katılırım.
----- Derse katılmadan izlerim.
----- Öğrendiklerimi değerlendiririm.
----- Aktif olmaktan hoşlanırım.
12. En iyi öğrenme yolum...,
----- Anlatılan fikirleri (konuları) tek tek ele almaktır.
----- Yeni fikirleri öğrenmeye açık olmaktır.
----- Dikkatli olmaktır.
----- Anlatılanları uygulamaktır.

8. ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Kadir Burak Olgun
Doğum Yeri	
Doğum Tarihi	
Uyruğu	☒ T.C. ☐ Diğer:
Telefon	
E-Posta Adresi	
Web Adresi	

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

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

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

Makale ve Bildiriler	
Kızılkaya Cumaoglu, G., Kusbeyzi Aybar, I., Olgun, K. B. (2013). University Students in a Non-Linear Environment Need Non-Linear Higher Education System. In Proceedings of 6th Chaotic Modeling and Simulation International Conference. Istanbul: Yeditepe University. Olgun, K. B., Kızılkaya Cumaoglu, G., Gülseçen, S. (2014). Effects of programming course on middle school students' reflective thinking skills towards problem solving. In Y. Gülbahar, E. Karataş & M. Adnan (Eds.), Local Proceedings of 7th International Conference on Informatics in Schools: Situation, Evolution and Perspectives (pp. 115-116). Istanbul: Istanbul University.	

- Olgun, K. B. (2014). Programlama eğitiminin ortaokul öğrencilerinin düşünme stilleri üzerine etkisi. Yayımlanmamış Yüksek Lisans Tezi, İstanbul Üniversitesi, İstanbul.
- Aksaya, H., Arşık, A., Elibol, M., Erdoğan, A., Olgun, K. B., Özcan, M., Ayvaz Reis, Z. (2015). Hobim İçin Öğreniyorum: Akvaristler Ne Kadar Memnun. In Proceedings of International Educational Technology Conference. Istanbul: Istanbul University
- Olgun, K. B. (2020). Kelime İşlemci Programları. S. Bayram, G. K. Cumaoğlu, İ. K. Aybar(Ed.) Bilişim Teknolojileri: Kuramsal Yaklaşım ve Uygulamalar içinde. Nobel Yayıncılık, Ankara. Ss. 143-171.

