



T.C.

AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ
ANABİLİM DALI

**YÜKSEK
LİSANS
TEZİ**

**ÖĞRENCİ BAŞARISINI KESTİRMEDE
MAKİNE ÖĞRENME ALGORİTMALARINA
DAYALI YENİ BİR SINIFLANDIRMA
MODELİ**

ŞERAFETTİN KUZUCUK

**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI**

Antalya, 2022

T.C
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI

ÖĞRENCİ BAŞARISINI KESTİRMEDE MAKİNE ÖĞRENME
ALGORİTMALARINA DAYALI YENİ BİR SINIFLANDIRMA MODELİ

YÜKSEK LİSANS TEZİ
Şerafettin KUZUCUK

Danışman:
Doç. Dr. Bilal Barış ALKAN

Antalya, 2022

DOĞRULUK BEYANI

Yüksek lisans tezi olarak sunduğum bu çalışmayı, bilimsel ahlak ve geleneklere aykırı düşecek bir yol ve yardıma başvurmaksızın yazdığımı, yararlandığım eserlerin kaynakçalardan gösterilenlerden oluştuğunu ve bu eserleri her kullanışında alıntı yaparak yararlandığımı belirtir; bunu onurumla doğrularım. Enstitü tarafından belli bir zamana bağlı olmaksızın, tezimle ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacağımı bildiririm.

20/06/2022

Şerafettin KUZUCUK

T.C.
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Şerafettin KUZUCUK'un bu çalışması 20/06/2022 tarihinde jürimiz tarafından **Eğitim Bilimleri** Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans Programında **Yüksek Lisans Tezi** olarak **oy birliği** ile kabul edilmiştir.

İMZA

Başkan : Prof. Dr. Bayram BIÇAK
Akdeniz Üniversitesi
Eğitim Bilimleri Bölümü

Üye : Doç. Dr. Pelin KASAP
Ondokuz Mayıs Üniversitesi
İstatistik Bölümü

Üye (Danışman) : Doç. Dr. Bilal Barış ALKAN
Akdeniz Üniversitesi
Eğitim Bilimleri Bölümü

YÜKSEK LİSANS TEZİNİN ADI: Öğrenci Başarısını Kestirmede Makine Öğrenme Algoritmalarına Dayalı Yeni Bir Sınıflandırma Modeli

ONAY: Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulunun tarihli ve sayılı kararıyla kabul edilmiştir.

Doç. Dr. Güçlü ŞEKERCİOĞLU
Enstitü Müdürü

TEŞEKKÜR

Bu süreçte desteğini hiçbir zaman esirgemeyen, sorularıma sabırla cevap veren, her şekilde iletişim kanallarını bana açık tutan çok kıymetli danışmanım Doç. Dr. Bilal Barış ALKAN'a, bana daima cesaret veren, beni destekleyen hayat arkadaşım Feride Gülşah KUZUCUK'a, bu süreçte zamanından çaldığım, biricik kızım Zeynep Derin KUZUCUK'a, uzakta olsalar da her zaman yanlarımda olduğunu hissettiğim annem ve babam Keziban-Abdullah KUZUCUK'a, kardeşlerim Mine ÇELİK ve Hasan Samet KUZUCUK'a, veri toplama sürecinde destek veren okul ve kurum müdürlerine teşekkürü borç bilirim.



ÖZET

ÖĞRENCİ BAŞARISINI KESTİRMEDE MAKİNE ÖĞRENME ALGORİTMALARINA DAYALI YENİ BİR SINIFLANDIRMA MODELİ

Kuzucuk, Şerafettin

Yüksek Lisans, Eğitim Bilimleri Ana Bilim Dalı,

Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Tez Danışmanı: Doç. Dr. Bilal Barış ALKAN

Haziran 2022, 94 Sayfa

Günümüzde büyük veri yığınlarından önemli bilgilerin ortaya çıkarılması için kullanılan veriden öğrenme temelli makine öğrenme algoritmalarının kullanımı artmaktadır. Makine öğrenme algoritmalarının kullanıldığı en önemli alanlardan birisi de eğitimidir. Modern eğitim sistemi anlayışına sahip olan ülkeler, eğitim alanında elde edilen verilerden yola çıkarak geleceğe yönelik planlamalar yaparken süreç içerisinde yer alan eksiklikleri de belirlemektedir. Bu çalışmada, öncelikle öğrencilerin sınav başarısını etkileyen değişkenler üzerine bir araştırma yapılmıştır. Öğrenci başarısına etki eden değişkenler belirlendikten sonra, bu değişkenler üzerinden bir madde havuzu oluşturulmuştur. Oluşturulan bu madde havuzu uzman görüşü ve ön uygulamaların ardından son halini alarak 84 maddeye düşürülmüştür. Antalya merkez ilçelerinde yer alan okullardan Liselere Giriş Sınav (LGS) puanı ile öğrenci alan okullara kayıtlı olan öğrenciler başarılı, sadece not ortalamasıyla kayıt olunan okullara devam eden öğrenciler başarısız olarak adlandırılmıştır. Toplamda 363 başarısız, 250 başarılı öğrenci olmak üzere 613 öğrenciden COVID-19 nedeniyle çevrimiçi anket yoluyla veriler toplanmıştır. Toplanan verilerle analiz kısmına geçilmiştir. Öncelikle maddelerden önem derecesi yüksek olan maddeler belirlenmiş, en yüksek önem derecesine sahip 30 maddeye verilen cevaplar üzerinden araştırmaya devam edilmiştir. “Öğrenciye ait kitap sayısı”, “öğrencinin sosyo-ekonomik durumu”, “ailenin okul etkinliklerine katılımı” gibi maddelerin öğrenci başarısında yüksek önem derecesine sahip olduğu görülmüştür. Çalışma kapsamında elde edilen veriler kategorik veriler olduğu için literatürde yapılan çalışmalar incelendiğinde, genelde kategorik verilerle daha iyi çalıştığı görülen C5.0, CART, DVM ve Rastgele Ormanlar algoritmaları verilerin analizinde kullanılmıştır. Veri seti eğitim (%80) ve test (%20) seti olmak üzere ikiye ayrılmıştır. Analizin hatalardan arınma oranını artırmak için tüm algoritmalar için çapraz doğrulama uygulanmıştır. Elde edilen model değerlendirme ölçütlerine göre Rastgele Ormanlar

Algoritmasının en başarılı tahmini yapan algoritma olduğu görülmüştür. Rastgele ormanlar algoritması üzerinden sınıflandırıcı modeli için model katsayılarının tahminleri belirlenmiştir. Bu yeni sınıflandırıcı modeli üzerinden, sürece sonradan dahil olacak olan öğrenci için başarı durumunu sınıflandırması yapılacaktır. Elde edilen bu yeni sınıflandırıcı modelinin kullanımı sadece LGS sınavına hazırlanan öğrencilerin nitelikli bir okula yerleşip yerleşmemesi değil, aynı zamanda eğitim dönemi içerisinde öğrencilerin eksikliklerinin erken dönemde tespit edilerek bu eksikliklerin giderilebileceği ve başarı anlamında olumlu yönde katkı sağlanacağı düşünülmektedir.

Anahtar Kelimeler: Makine Öğrenme Algoritmaları, Veri Madenciliği, Sınıflandırma Modeli



ABSTRACT

A NEW CLASSIFICATION MODEL BASED ON MACHINE LEARNING ALGORITHMS TO PREDICT STUDENT SUCCESS

Kuzucuk, Şerafettin

Master of Science, Department of Educational Sciences,
Department of Measurement and Evaluation in Education

Supervisor: Assoc. Prof. Dr. Bilal Barış ALKAN

June 2022, 94 Pages

Today, the use of data-based machine learning algorithms, which are used to reveal important information from large data piles, is increasing. One of the most important areas in which machine learning algorithms are used is education. Countries with the understanding of modern education system determine the deficiencies in the process while making plans for the future based on the data obtained in the field of education. In this study, first of all, a research was conducted on the variables that affect students' exam success. After determining the variables affecting student success, an item pool was created based on these variables. This item pool was finalized after expert opinion and preliminary practices, and was reduced to 84 items. Students enrolled in schools in the central districts of Antalya that accept students with the High School Entrance Examination (LGS) score are considered successful, while students who attend schools with only a grade point average are considered unsuccessful. Data were collected from 613 students, 363 unsuccessful and 250 successful students, through an online survey due to COVID-19. The analysis part was started with the collected data. First of all, the items with a high degree of importance were determined, and the research continued on the answers given to the 30 items with the highest degree of importance. Items such as "the number of books belonging to the student", "the socio-economic status of the student", "the participation of the family in school activities" have a high degree of importance in student success. Since the data obtained within the scope of the study are categorical data, when the studies in the literature are examined, C5.0, CART, SVM and Random Forests algorithms, which are seen to work better with categorical data, were used in the analysis of the data. The data set is divided into two as training (80%) and test (20%) sets. Cross validation was applied for all algorithms to increase the error-free rate of the analysis. According to the model evaluation criteria obtained, the Random Forests Algorithm was found to be the most successful estimation algorithm. The estimations of the model coefficients for the classifier model were determined through the

random forests algorithm. Through this new classifier model, the classification of success will be made for the student who will be involved in the process later. It is thought that the use of this new classifier model obtained is not only whether the students preparing for the LGS exam will be placed in a qualified school, but also that the deficiencies of the students can be detected early in the education period, these deficiencies can be eliminated and a positive contribution will be made in terms of success.

Key Words: Machine Learning Algorithms, Data Mining, Classification Model



İÇİNDEKİLER

TEŞEKKÜR.....	i
ÖZET.....	ii
ABSTRACT.....	iv
İÇİNDEKİLER.....	vi
TABLolar LİSTESİ.....	ix
ŞEKİLLER LİSTESİ.....	x
KISALTMALAR LİSTESİ.....	xi

BÖLÜM I

GİRİŞ

1.1 Problem Durumu.....	1
1.2 Araştırmanın Amacı ve Alt Problemleri.....	7
1.3 Araştırmanın Önemi.....	7
1.4 Araştırmanın Varsayımları.....	8
1.5 Araştırmanın Sınırlılıkları.....	8
1.6 Tanımlar.....	8

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

2.1 Veri Madenciliğinin İlişkili Olduğu Alanlar.....	9
2.1.1 Makine Öğrenme.....	9
2.1.2 Veri Tabanı Sistemleri.....	10
2.1.3 Bilgi Teknolojileri.....	10
2.1.4 Veri Görselleştirme.....	10
2.1.5 İstatistik.....	11
2.1.6 Diğer Disiplinler.....	11
2.2 Veri Madenciliği Süreçleri.....	11
2.2.1 Problemin Tanımlanması.....	12
2.2.2 Verilerin Tanımlanması.....	12
2.2.2.1 Veri Tiplerinin Dönüştürülmesi	13
2.2.2.2 Kesintisiz Sütun Değişkenleri.....	13
2.2.2.3 Gruplandırma	13

2.2.2.4 Veri Birleřtirme.....	13
2.2.2.5 Veri Azaltma.....	13
2.2.3 Modelin Kurulması ve Deęerlendirilmesi.....	14
2.2.4 Modelin Kullanımı.....	14
2.2.5 Modelin İzlenmesi.....	14
2.3 Veri Madencilięi Yöntemleri	14
2.3.1 Özellik Seçimi	13
2.3.2 Sınıflandırma	15
2.3.3 Birliktelik	15
2.3.4 Kümeleme.....	15
2.4 Makine Öğrenme Algoritmaları.....	15
2.4.1 Karar Ağaçları.....	16
2.4.2 ID3 Algoritması.....	18
2.4.3 C4.5 Algoritması.....	19
2.4.4 C5.0 Algoritması.....	20
2.4.5 CART Algoritması.....	20
2.4.6 Rastgele Ormanlar Algoritması.....	21
2.4.7 Destek Vektör Makinesi Algoritması.....	23
2.4.8 K-En Yakın Komşu Algoritması	24
2.4.9 Naive Bayes Algoritması.....	25
2.5 Sınıflandırma Kalitesinin Ölçümü.....	26
2.5.1 Verilerin Ayırıştırılması.....	26
2.5.2 Sınıflandırma Performansı Deęerlendirme Ölçütleri.....	27
2.6 Makine Öğrenme İle İlgili Yapılan Arařtırmalar.....	28

BÖLÜM III

YÖNTEM

3.1 Arařtırmanın Modeli	31
3.2 Çalışma Grubu.....	31
3.3 Veri Toplama Aracının Geliřtirilmesi.....	34
3.4 Verilerin Toplanması.....	34
3.5 Verilerin Analizi.....	35
3.5.1 Veri Tiplerinin Dönüřtürülmesi.....	35
3.5.2 Verilerin Seçilmesi.....	35

BÖLÜM IV

BULGULAR

4.1 Katılımcıların Başarı Durumlarına Göre Sınıflandırılması	39
4.2 C5.0 Algoritması Uygulaması ve Uygulama Sonuçları.....	39
4.3 CART Algoritması Uygulaması ve Uygulama Sonuçları	41
4.4 Destek Vektör Makinesi Algoritması Uygulaması ve Uygulama Sonuçları.....	43
4.5 Rastgele Ormanlar Algoritması Uygulaması ve Uygulama Sonuçları	44
4.6 Öğrenci Başarısının Model Olarak Gösterimi.....	46

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1 Sonuç ve Tartışma.....	51
5.2 Öneriler.....	52
KAYNAKÇA.....	53
EKLER	
EK 1-ETİK KURUL KARARI.....	65
EK 2-ANTALYA İL MİLLİ EĞİTİM MÜDÜRLÜĞÜ ARAŞTIRMA İZİN ONAYI.....	66
EK 3-GELİŞTİRİLEN ANKET FORMU.....	67
ÖZGEÇMİŞ.....	77
İNTİHAL RAPORU.....	78
BİLDİRİM.....	79

TABLÖLAR LİSTESİ

Tablo 2.1 İki Sınıflı Bir Karışıklık Matrisi.....	27
Tablo 3.1 Çalışma Grubunu Oluşturan Öğrencilerin İlçelere ve Okul Türlerine Göre Dağılımı.....	32
Tablo 3.2 Çalışma Grubundaki Öğrencilerin Sahip Oldukları Olanaklara İlişkin Dağılımları.....	32
Tablo 3.3 Çalışma Grubundaki Öğrencilerin Aile Özelliklerine İlişkin Dağılımlar	33
Tablo 3.4 Analiz İçin Kullanılacak Olan Anket Soruları	37
Tablo 4.1 Öğrencilerin Başarı Sınıflaması Sonucu Oluşan Gruplar.....	39
Tablo 4.2 C5.0 Algoritmasının Eğitim Veri Setine Ait Karışıklık Matrisi Tablosu	40
Tablo 4.3 C5.0 Algoritması Test Veri Setine Ait Tahmin Karışıklık Matrisi Tablosu	40
Tablo 4.4 C5.0 Algoritmasına Ait Model Doğruluğu Ölçütleri	41
Tablo 4.5 CART Algoritmasının Eğitim Veri Setine Ait Karışıklık Matrisi Tablosu	42
Tablo 4.6 CART Algoritmasının Test Veri Setine Ait Karışıklık Matrisi Tablosu.....	42
Tablo 4.7 CART Algoritmasına Ait Model Doğruluğu Ölçütleri.....	43
Tablo 4.8 DVM Algoritmasının Eğitim Veri Setine Ait Karışıklık Matrisi Tablosu.....	43
Tablo 4.9 DVM Algoritmasının Test Veri Setine Ait Karışıklık Matrisi Tablosu.....	43
Tablo 4.10 DVM Algoritmasına Ait Model Doğruluğu Ölçütleri.....	44
Tablo 4.11 Rastgele Ormanlar Algoritmasının Eğitim Veri Setine Ait Karışıklık Matrisi Tablosu.....	45
Tablo 4.12 Rastgele Ormanlar Algoritmasının Eğitim Veri Setine Ait Karışıklık Matrisi Tablosu.....	45
Tablo 4.13 Rastgele Ormanlar Algoritmasına Ait Model Doğruluğu Ölçütleri.....	46
Tablo 4.14 Dört Algoritmaya Ait Doğruluk Oranları Karşılaştırma Tablosu.....	46
Tablo 4.15 Öğrencilerin Verdiği Cevaplar Üzerinden Elde Edilen Katsayı Tahminleri.....	47

ŞEKİLLER LİSTESİ

Şekil 2.1 Veri Madenciliği ile İlgili Alanlar.....	9
Şekil 2.2 Veri Tabanlarında Bilginin Keşif Süreci.....	12
Şekil 2.3 Verilerin Durumuna Göre Makine Öğrenme Teknikleri	16
Şekil 2.4 Bir Karar Ağacının Yapısı	17
Şekil 2.5. Rastgele Orman Algoritmasının Görselleştirilmesi	22
Şekil 2.6 KNN Algoritmasının Görselleştirilmesi.....	25
Şekil 3.1 Araştırmaya Ait Görsel Akış Şeması.....	31
Şekil 3.2 Bağımsız Değişkenlerin Seçimi.....	36
Şekil 4.2.1 C5.0 Algoritmasına Ait Çapraz Doğrulama Sonuçları.....	41
Şekil 4.3.1 CART Algoritmasına Ait Çapraz Doğrulama Sonuçları.....	42
Şekil 4.4.1 DVM Algoritmasına Ait Çapraz Doğrulama Sonuçları.....	44
Şekil 4.5.1 Rastgele Ormanlar Algoritmasına Ait Çapraz Doğrulama Sonuçları.....	45

KISALTMALAR LİSTESİ

CART : Classification and Regression Tree

DVM: Destek Vektör Makinesi



BÖLÜM I

GİRİŞ

Bu bölümde yapılan araştırmanın problemine, amacına, önemine ve sınırlılıklarına yer verilmiştir.

1.1 Problem Durumu

Ülkelerin gelişmişlik düzeyi göstergelerinden birisi aynı zamanda ülkenin kültürel, sosyal ve ekonomi anlamında gelişmesinde de büyük bir paya sahip olan nitelikli insan gücüdür. Ülkeler toplumsal ihtiyaçlarına yönelik nitelikli insan gücünü eğitim sayesinde karşılar. Eğitimi genel anlamda “bireyde istendik davranış oluşturma ya da istendik davranış değiştirme süreci” olarak tanımlayabiliriz (Ertürk, 1973). Eğitim, insanın içinde yaşadığı toplum içinde bir değeri olan ve beceri, tutum ve davranışlarını geliştirebildiği süreçlerdir (Fidan, 1996). Eğitim önceden belirlenmiş ilkelere göre insan davranışlarında birtakım gelişmeler sağlayan planlı etkinlikler dizisidir (Oğuzkan, 1993). Bu planlar eğitim programları sayesinde gerçekleşebilir. Eğitim programı, öğrenene okulda veya okul dışında belli bir plan dahilinde etkinlikler sayesinde sağlanan ve öğrenilen yaşantıların mekanizmasıdır (Demirel, 1999). Türkiye’de yeni yaklaşımlar da göz önüne alınarak eğitim ve öğretimin merkezine öğrenci konmuştur. Bu süreçte hedefler, eğitim programı, ölçme ve değerlendirme gibi unsurlar sürekli etkileşim halindedirler. Ölçme ve değerlendirme eğitim öğretimin planlanarak yönlendirilmesinde (öğretim ortamlarının tasarlanması, öğrenim hedeflerinin belirlenmesi vb.) ve öğrencilerin başarısının ölçülmesinde aktif olarak kullanılmaktadır (Tekin, 2004).

Ölçme, bir ya da birden çok nesnenin belli bir niteliğe sahip olup olmadığının sahip olması durumunda sahip olma derecesinin gözlenerek, gözlem sonuçlarının sayı sembolleriyle ifade edilmesidir (Tekin, 2004). Değerlendirme işlemi ise elde edilen gözlem sonuçlarından bir yargıya varma, karar verme işlemidir. Bu yargı işleminde ölçme sonucu belli bir ölçütle kıyaslanır. Eğitimin hedefleri doğrultusunda beklenen davranış değişikliklerinin öğrencide hangi ölçüde gerçekleşebildiğinin ölçülerek değerlendirilmesi bir eğitim sistemi çerçevesinde uygulanan değerlendirme işlemlerinin en önemlileri arasında sayılabilir (Koç, 1981). Değerlendirme işleminin görevlerinden biri öğrenci başarı durumunu belirlemek diğer bir görevi de öğrencileri bulundukları okulu bitirdikten sonra ilgi ve yeteneklerine göre uygun okullara yöneltmesidir (Baykul & Turgut, 2010).

Türkiye’ de eğitim kademeleri şu şekildedir:

- Okul öncesi: Okul öncesi eğitim kademesi Eylül ayı sonu itibarıyla 36-68 ay arası eğitimin verildiği anaokulları ve Eylül ayı sonu itibarıyla 57-68 aylık çocukların eğitimi amacıyla örgün eğitim ve hayat boyu öğrenme kurumları bünyesinde açılan anasınıfları,
- İlköğretim kurumları: Resmî ve özel ilkokul, ortaokul ile eğitim, öğretim, yönetim ve bütçe ile ilgili iş ve işlemleri Din Öğretimi Genel Müdürlüğünce yürütülen imam-hatip ortaokulları,
- Ortaöğretim Kurumları: Mesleki ve Teknik Anadolu Lisesi, Mesleki ve Teknik Eğitim Merkezi, Çok Programlı Anadolu Lisesi, Özel Eğitim Meslek Lisesi, Mesleki Eğitim Merkezi, Fen Liseleri, Sosyal Bilimler Liseleri, Açık Lise gibi türleri olan 4 yıllık zorunlu eğitimin olduğu kurumlardır (MEB, 2013).
- Yükseköğretim Kurumları: Milli eğitim sistemi içinde, ortaöğretime dayalı, en az dört yarı yılı kapsayan her kademedeki eğitim -öğretimin tümüdür (YÖK, 1983).

Öğrenciler, sırasıyla zorunlu olmayan okul öncesi eğitim, 4 yıllık zorunlu ilkokul, 4 yıllık zorunlu ortaokul ardından da 4 yıllık zorunlu lise eğitimi almaktadır. Tüm kademedeki öğretim kurumlarının genel amacı öğrenciyi bir sonraki öğretim kurumlarına ve mesleğe, iş hayatına hazırlamaktır (Gedikoğlu, 2005). Bu yüzden özellikle öğrencilerin lise ve üniversite tercihlerinde bu kurumlara olan talepleri karşılayabilmek adına öğrencilerin bilgi durumlarını tarafsız bir şekilde ölçebilecek bir değerlendirme sistemine ihtiyaç bulunmaktadır (Şad & Şahiner, 2016).

Türkiye’de 8 yıllık kesintisiz ve zorunlu eğitimin ardından bu taleplerin karşılayabilmek için Liselere Giriş Sınavı (LGS) düzenlenmektedir. Bu sınavı Milli Eğitim Bakanlığı (MEB) düzenlemektedir. LGS sınavı sonucunda yüksek puan alan adaylar “nitelikli lise” olarak adlandırılan Fen Liseleri, Anadolu Liseleri veya Sosyal Bilimler Lisesi’ni tercih etmektedirler. Giderek bu tarz okulların sayısı artarken bu okullara oluşan talep de aynı şekilde artmaktadır. 2019 yılı ile karşılaştırıldığında 140 bin civarında olan nitelikli lise kapasitesi 2020 yılında %52,23’lük bir artışla 213 bin 128’ e çıkarılmıştır. Bu okullardan Anadolu Liseleri bir önceki yıla göre %69’luk bir kontenjan artışıyla 71 bin 494’e yükselirken, nitelikli lise kapsamında bulunan Mesleki ve Teknik Anadolu Liseleri’nde ise kontenjan yaklaşık %62 artmıştır. Fen ve Sosyal Bilimler Liseleri’nde %35, Anadolu İmam Hatip Liseleri’nde de %45’lik bir kontenjan artışı yapılmıştır. Nitelikli okul kontenjanlarının yaklaşık %61’ini Fen ve Sosyal Bilimler Liseleri oluşturmaktadır (MEB, 2020). Her ne kadar nitelikli okul sayısı ve kontenjanı artsa da, genç nüfusun fazla olması bu okullara olan talebi de arttırmaktadır. Özellikle ülkemizde

sınavlar genelde meslek edinme ve başarının net bir göstergesi olarak görüldüğünden doğal olarak öğrencilerden beklenti de fazla olmaktadır.

Eğitimsel anlamda bakıldığında başarıyı, programın hedefleriyle örtüşen davranışlar olarak tanımlayabiliriz (Demirtaş & Güneş, 2002). Bu durumdan yola çıkarak başarıyı etkileyen birçok değişken olduğunu söyleyebiliriz. Kimi zaman toplumsal, kimi zaman psikolojik, kimi zaman da fizyolojik nedenler bu değişkenlere etki eder. Bu yüzden değişkenler, öğrencinin başarı durumunda veya öğrenmesinde bazen olumlu bazen de olumsuz etki göstermektedir. (Uluğ, 2000). Her öğrencinin öğrenme yöntemi farklılık gösterdiğinden öğrencilerin okul başarısı durumlarında da farklılık gözlenmektedir. Bu farklılıklara sebep olan değişkenlere göz atmak faydalı olacaktır.

Bu değişkenleri okul içi ve okul dışı olarak iki farklı grupta inceleyebiliriz. Okul içi değişkenlere baktığımızda okulda öğrencinin karşılaştığı öğretmenlerin özellikleri, yöneticilerin özellikleri, okulun bulunduğu bölge ve bölgenin maddi durumu, okulun büyüklüğü, okulda eğitim için kullanılan araç ve gereçlerin niteliği ve sayısı, okul kütüphanesi, laboratuvar, atölye gibi bölümlerin varlık-yokluk durumu, okulun genel iklimi, okulun ailelerle birlikte gösterdiği işbirliği, öğrenci için birebir harcanabilecek eğitim süresi, öğretim ilke ve yöntemleri, okulun genel amaç ve hedeflerinin gerçekleşme düzeyi gibi maddeler karşımıza çıkar (Richards, 1986). Okul dışı değişkenleri öğrencinin duyuşsal, psikomotor ve bilişsel özellikleri, öğrencinin ailesinin sosyo-ekonomik durumu, çevre ve kültürel özellikler, öğrencinin akran çevresi olarak sıralanabilir.

Öğrencinin zeka düzeyi, yetenek, beceri ve ilgi gösterdiği alan bir önceki eğitim kurumunda göstermiş olduğu başarı düzeyi, çalışma alışkanlıkları, öğrencinin kendine ait bir çalışma ortamının olup olmadığı, ailenin ve yakın çevresinin öğrenciye gösterdiği ilgi ve destek, öğrenciye dair beklenti ve tutumlar öğrencinin başarı düzeyinde önemli değişkenlerdir (Richards, 1986). Bunların yanında öğrencinin kendine dair benlik algısı, öz-yeterlik algısı, ders motivasyonu, öğrencinin derslere dair tutumu da öğrenci ile ilgili değişkenler arasında yer almaktadır.

Bu değişkenlerden okul değişkeni, öğrencinin okulda karşılaştığı öğrenci, öğretmen ve diğer personeller ile okulun fiziksel olanaklarının, öğretim programlarının, uygulanan yöntemlerin, eğitim ve öğretim süreçlerinin ve örgütsel kültürün tamamını kapsayan bir kavramdır (Tural, 2002). Aynı sınıftaki öğrencilerin anlama, kavrama yöntemlerinin birbirinden farklılık gösterdiği doğal olarak öğrenci başarılarının da farklılık göstermesinin doğal olduğunu gözükmemektedir. Bazı çalışmalarda “etkili okul hareketi” kavramının okulun olanakları ne olursa olsun öğrenci başarısını artırdığı savunulmuştur. Dolayısıyla eğitim için

gerekli temel araç-gereç ve edinimlere sahip bir okulda, olumlu bir okul iklimi sağlaması ve öğrencilere ilgi gösterilmesi durumunda bireysel akademik farklılığın da minimize olacağı belirtilmiştir (Edmonds, 1979). Nitelikli eğitim için gereken araç ve gereçleri sağlayan, öğrenme merkezi olan bir okul oluşturmada öncü olan, öğretmenlerin gelişmesine katkıda bulunan ve destek veren, okul iklimine davranışlarıyla öncülük eden, yaptığı gözlemler ve verdiği kararlarla liderlik özelliğine sahip olan okul yöneticileri de öğrencilerin akademik başarısını hatırı sayılır bir düzeyde etkiler (Şişman, 2014).

Öğretmenlerin eğitimin temel yapıtaşlarından biri olduğu herkes tarafından kabul edilmektedir. Sınıf içinde öğrenci ile kurulan iletişim, öğretmenin bilgi ve beceri düzeyi ile yine öğretmenin ders ve öğrencilere karşı olan tutumuna bağlıdır (Şişman, 2002). Öğretmenin ders ile ilgili gerekli bilgi ve beceriye sahip olmasının yanında; ders sırasında kullandığı ders araç gereçleri, ders öğretiminde kullandığı yöntem ve teknikler de öğrenci başarısında önemli bir yer tutmaktadır. Öğretmenin mesleki gelişim açısından yaptığı çalışmalar, etkili iletişim becerilerine sahip olması öğrenci başarısına olumlu etki yapmaktadır. Öğretmenlerin öğrenciye karşı olan destekleyici tutum ve davranışlarının öğrencinin başarısını olumlu yönde etkilediği ve okulun genel durumunu güvenli bulan öğrencilerin akademik başarılarının daha yüksek olmaktadır (Kızıldağ, 2009).

Öğrenci başarısına etki eden önemli değişkenlerden birisi de ailenin sosyo-ekonomik düzeyidir. Ailenin sosyo-ekonomik düzeyinden kasıt, yaşanılan sosyal çevre, ebeveynlerin toplumdaki statüleri, sahip oldukları bilgisayar gibi değişkenlerdir (Özer & Anıl, 2011). Ebeveynlerin öğrencileri destekleyen tutum ve davranışlarının eğitim kalitesine olumlu etki yaptığı gözlenmiştir. Carlson'a (1999) göre erken çocukluk ve erişkinlik dönemindeki ailenin verdiği destek ve öğrenciye gösterdiği ilgi ile, öğrencinin ergenlik dönemindeki akademik başarısı ve okula uyumu arasında pozitif yönde bir ilişki bulunmaktadır. Ebeveynlerin çocuklarıyla okul hakkında yaptığı sohbet, ödev konusunda öğrencilere verdikleri destek, öğrencinin bulunduğu okulun etkinliklerine ve toplantılarına gösterdikleri katılım ve öğretmenlerle iletişim gibi birçok aile katılım uygulaması öğrenci başarısına olumlu yönde katkıda bulunmaktadır (Şad, 2012). Yüksek gelir düzeyine sahip olan ailelerin, çocuklarına kaynak sağlamada ve akademik anlamda çocuklarını teşvik etmede daha başarılı olmaktadır (Checchi, 2000). Sosyo-ekonomik düzey ile eğitim arasındaki ilişki alanındaki araştırmalar, sosyo-ekonomik düzeyi yüksek olan öğrencilerin olmayanlara göre daha başarılı olduğunu ortaya koymaktadır (Şirin, 2005). Türkiye, akademik başarı düzeyi ile sosyo-ekonomik durum arasındaki ilişkinin en güçlü olarak görüldüğü ilk üç ülkeden biridir (OECD, 2012).

Öğrenci başarısına etki eden önemli değişkenlerden bir diğeri de anne ve babanın eğitim durumlarıdır. Bu konudaki araştırmalarda anne ve babanın iyi bir eğitim geçmişine sahip olması durumunda öğrencilerin de gittikleri okula karşı bakış açısının olumlu olduğu ve okul içi etkinliklerine katılımda daha istekli olduğu görülmektedir. Böylece iyi bir eğitim geçmişine sahip olan anne ve babanın, iyi bir eğitim geçmişine sahip olmayan anne ve babaya göre daha ilgili olduğu ve öğrencilerin eğitimine daha fazla katkı sağladığını söyleyebiliriz (Carlisle, Stanley, & Kemple, 2005). Ebeveyn eğitimi ve aile geliri akademik başarının uzun dönemdeki belirleyicileridir ve genel anlamda öğrencilerin akademik başarısını artırmak için bu faktörleri etkileyecek eğitim politikaları oluşturulmalıdır (Cameron & Heckman, 2001).

Öğrencilerin akademik başarısında çevresel etmenlerin yanında içsel süreçler de önemli bir yer tutmaktadır. Benlik saygısı yani bir insanın kendini onaylaması, kendinden genel anlamda memnun olması, kişisel özelliklerini beğenmesi; öğrencinin akademik ve bilişsel yaşamlarını etkilemektedir (Choi, 2005). Öz yeterlilik, bireyin zorluklar karşısında yılmamasını, hedefine ulaşmak için gösterdiği çabayı, inancını, analitik düşünme yeteneklerini ifade etmektedir. Bu konudaki araştırmalar, bireyin öz-yeterlik algısı ile akademik başarısı arasında olumlu yönde bir ilişki olduğunu göstermektedir (Xing & Koirala, 2009). Motivasyon, canlıyı davranışa yönelten, farklı iç ve dış faktörler sayesinde canlının işleyişini sağlayan mekanizmalar olarak tanımlanabilir (Aydın, 2014). Bu konudaki araştırmalarda, motivasyon seviyesi üst seviyede olan öğrencilerin, benlik algısının da üst seviyede olduğu ve davranışlarının daha yaratıcı olduğu, dolayısıyla daha yüksek bir akademik başarı seviyesine ulaştıkları görülmektedir (Boyd, 2011).

Türkiye’de ve Dünya üzerindeki bilgi sistemlerinde geleneksel yöntemlerle anlaşılması güç olan birçok belirsiz bilgi karşımıza çıkmaktadır. Sahip olduğumuz verilerin analizinin yapılması artık çok daha elzem bir durum haline gelmiştir. Teknolojinin gelişmesiyle birlikte dünya üzerinde gerçekleşen birçok işlem de artık elektronik olarak kayıt altına alınmaktadır. Bu kayıtların depolanmasının kolay olması, istenilen zamanlarda erişimi de sağlamaktadır. Gün geçtikçe kayıt işlemleri de daha ucuza mal olmaktadır. Veri tabanlarında yer alan bu bilgilerin analizi, günümüzde kamu ve özel kurumların bu verilerin değerini anlamasıyla birlikte önem kazanmıştır. Gelecekte, çağın gerektirdiği durumlar için strateji geliştirmek için bu alanda ciddi yatırımlar yapılmaktadır. Doğru stratejinin geliştirilmesi verilerin doğru depolanması, sınıflanması, temizlenmesi, dönüştürülmesi, modellenmesi ve yorumlanmasına bağlıdır. Böylece çok sayıdaki veri arasından anlamlı bilgiler çıkarılması ve veriler arası ilişkilerin ortaya çıkması ihtiyaç haline gelmiştir. Bu ihtiyaçların karşılanması için bilginin elde edilmesi amacıyla kullanılan tekniklerin bütününe veri madenciliği adı verilmektedir. Veri madenciliği;

makine öğrenmenin, yapay zekanın, istatistiğin ve grafiklerin kullanıldığı disiplinler arası bir olgudur. Bu alanlardan istatistik, veri madenciliğinin yöntem ve araçlarının temelinde yer alır. Sezgisel yaklaşımları temel alan yapay zekanın prensibi ise insan gibi düşünebilme becerisidir. Problem çözümünde bu beceriden yararlanır. Makine öğrenmesi yapay zekanın insan gibi düşünme prensibi ve ileri düzey istatistik metotlarının karışımıdır. Makine öğrenme, istatistik ve yapay zeka algoritmalarının kullanılarak, bilgisayar sistemlerinde verilerden sonuçlar çıkarılmasına ve sonuçlara dayanan kararlar alınmasını mümkün kılmaktadır (Coşkun, 2010).

Veri madenciliğinin birçok tanımı bulunmaktadır. Anlamsız görünen veri yığınınından çeşitli işlemler sonucunda önemli ve anlamlı bilgilerin elde edilmesi, büyük verileri analiz etme yolu, büyük ve birbiriyle ilintili veri tabanlarından ilişki ve düzen bulan bir karar destek sistemi olarak tanımlanabilir (Ersöz, 2019). Veri madenciliği, verinin ham halinin sunamadığı bilgiyi ortaya çıkaran analiz süreci olarak tanımlanmaktadır (Jacobs, 1999). Doğan ve Türkoğlu (2007) ise veri madenciliğini büyük veri yığınlarının içinden gelecek konusunda tahmin yapmamızı sağlayan içeriklerin, bilgisayar uygulamaları aracılığıyla aranması işlemi olarak tanımlamaktadırlar. Büyük miktardaki veri yığınınından kıymetli ve kullanılabilir bilgilerin ortaya çıkarılması ve bu bilgilere göre ilgili yönetsel kararlar alınması, geleceğe dair kestirimler yapılmasını sağlayabilecek bağıntı ve kuralların keşif sürecidir (Gökçen, 2010). Veri madenciliği çok büyük bir veri ağında faydalı olduğu daha önce bilinmeyen, faydalı olma olasılığı olan bilginin diğer bilgilerinden ayırt edilerek meydana çıkarılmasıdır (Pujari, 2008). Veri madenciliği birçok bilginin keşfedilme süreci olarak tanımlandığı için, sınıflandırma, veriler arası bilgileri keşfetme, verilerdeki değişimin analizi, özetleme gibi tekniklere başvurulmaktadır. Bu teknikler sonucu yapılan keşiflere model adı verilir. Oluşturulan modeller iki amaç doğrultusunda kullanılmaktadır;

(i) Çevremizde gerçekleşenleri daha iyi kavramak

(ii) İleriye dönük kestirimlerde bulunmak (G.Williams, 2011)

Veri madenciliğinde verilerin önceden belirlenmiş olan sınıflara göre ayrılması, gerçekte gerçekleşmesi muhtemel sonuçların tahmin edilmesidir. İstatistiksel tahmin modelleri sayesinde sınıflandırma işlemleri yapılabilir (Silahtaroglu, 2008). Bu metotta, veri tabanındaki veriler önceden belirlenen bir sınıfta veya kategoride toplanır (Refaat, 2007). Sınıflandırmada, araştırmaya konu olan nesneler farklı niteliklerine göre ayrılarak farklı kategorilerde etiketlenir (Bramer, 2007). Sınıflandırma, veri madenciliğinde kestirimsel bir analiz yöntemi olarak kabul edilmektedir. Burada amaç, nesnelere ait verilerden hareketle, nesnelerin belli özelliklerine göre sınıflara ayrılmasını sağlamaktır. Sınıflama ve regresyon modelleri kestirici, kümeleme ve

birliktelik kuralları modelleri tanımlayıcı modellerdir (Chaudhuri, 1998). Diğer bir ifadeye göre sınıflandırma, farklı bölümlerdeki maddeleri birbirinden ayırmak olarak tanımlanır. Sınıflandırma yaparken öğelerin belli bir nitelik veya değeri varsa “denetimli sınıflandırma” yoksa “denetimsiz sınıflandırma” olarak adlandırılır.

1.2 Çalışmanın Amacı

Bu araştırmanın genel amacı, öğrenci başarısı üzerinde etkili olan değişkenlerin belirlenmesi ve nitelikli liselerde okuyan öğrencilerden ve nitelikli okula yerleşmek için yeterli puan alamayarak diğer liselerde öğrenim gören öğrencilerden toplanacak verilerin kullanılması ile oluşturulacak sınıflandırma modeli üzerinden nitelikli liselere giriş sınavına hazırlanan öğrencilerin bu sınava girmeden önce başarı sınıflandırma tahminlerinin elde edilebileceği bir model geliştirmektir. Bu amaçlar doğrultusunda cevabı aranan sorular şunlardır:

1. Akademik başarıyı etkileyen en önemli değişkenler nelerdir?
2. Farklı makine öğrenmesi sınıflandırma algoritmalarının öğrencileri başarı durumlarına göre genel doğru sınıflandırma oranları karşılaştırıldığında en iyi yöntem hangisidir? Nasıl bir model elde edilir?

1.3 Çalışmanın Önemi

Türkiye’de ilköğretim sonrasında nitelikli bir liseye yerleşebilmek lise sonrası süreç olan üniversiteye giriş sınavındaki başarıyı doğrudan etkilemektedir. Öğrencilerin sınavlar öncesindeki başarısızlıklarının önüne geçilebilmesi için başarısızlık üzerinde etkili faktörlerin belirlenmesi ve analizi oldukça önemlidir. Eğitim sisteminde öğrencilerin farklı niteliklerine göre hangi sınıflara yerleştirileceğini belirleyecek bir sınıflayıcı model oluşturarak erken uyarı sistemi gibi öğrencinin başarı durumuna ilişkin problem oluşturabilecek değişkenler belirlenerek birtakım önlemler alınabilir. Bu alanda yapılan çalışmalar genelde sınırlı sayıda bağımsız değişkenle sınırlandırılmıştır. Bu tür çalışmalar eğitim alanında oldukça yetersizdir.

Bu çalışma ile öğrenci başarısını değerlendirmede makine öğrenmesi algoritmalarına dayalı yeni bir sınıflandırıcı modelinin elde edilecek olması ve böyle bir modelin genellenebilirliği de düşünüldüğünde eğitimde ölçme ve değerlendirme alanına katkı sağlayıcı nitelikte olacağı öngörülmektedir.

1.4 Sayıtlar

Bu arařtırmada, öğrencilerin anketteki sorulara samimi ve doğru cevaplar verdikleri varsayılmıştır. Öğrencilere ve velilerine elde edilen bilgilerin hiç kimseyle paylaşılmayacağı bilgisi verilmiştir.

1.5 Sınırlılıklar

Bu araştırma öğrenci başarısı ile ilişkili olduğu düşünülen öğrencinin not ortalaması, sınava hazırlanma süresi, düzenli ders çalışma alışkanlığına sahip olma durumu, çalışma odası, çalışma masası, internet bağlantısı, bilgisayar, kitaplık, kardeş sayısı, ailenin ortalama toplam aylık geliri, annenin öğrenim durumu, babanın öğrenim durumu, annenin çalışma durumu, babanın çalışma durumu gibi değişkenlerle oluşturulmuş ancak; Covid-19'dan dolayı planlanan öğrenci sayısının altında öğrenciden bilgi toplanabilmiştir.

1.6 Tanımlar

Akademik Başarı: Bir öğrencinin, öğretmenin veya kurumun kısa veya uzun vadeli eğitim hedeflerine ulaşma derecesi.

Veri Madenciliği: Büyük ölçekli veriler arasından faydalı bilgiye ulaşma.

Karar Ağaçları: Sınıflandırma ve Regresyon problemlerinde kullanılan, ağaç tabanlı algoritmalar

Veri Analizi: İşlenmemiş verinin toplanması ve istatistik yöntemleri kullanarak anlamlı ve yararlı bilgiler haline getirilmesi işi.

BÖLÜM II

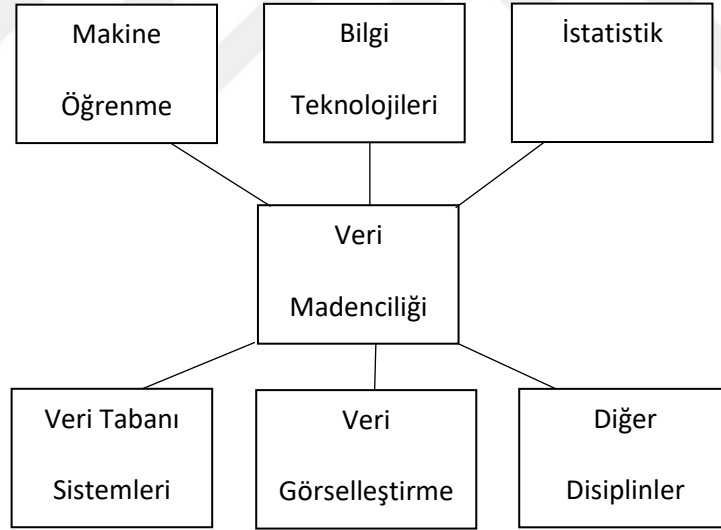
KURAMSAL ÇERÇEVE ve İLGİLİ ARAŞTIRMALAR

2.1 Veri Madenciliğinin İlişkili Olduğu Alanlar

Han vd. (2012) veri madenciliğini, bilgi teknolojilerinde meydana gelen doğal evrim sonucu ortaya çıktığını belirtmiş, bu dönüşümün üç aşamada meydana geldiğini ileri sürmüşlerdir;

- (1) 1960'lı yıllarda veri toplama ve veri tabanı oluşturma aşaması,
- (2) 1970 ve 1980'lerin başlarındaki veri yönetim aşaması,
- (3) 1980'lerden sonra ortaya çıkan veri madenciliği de dahil olmak üzere üst düzey veri analiz teknikleri aşaması.

Veri madenciliği disiplinler arası çalışmaların olduğu geniş bir alandır. İstatistik, makine öğrenme ve veri tabanlarından yararlanılarak ortaya atılan bir veri toplama metodudur (Holsheimer Vd., 2015).



Şekil 2.1 Veri Madenciliği ile İlişkili Alanlar (Han & Kamber, 2000)

2.1.1 Makine Öğrenme

Makine öğrenme, verilerin sistematik bir biçimde sınıflandırılabilmesi ve depolanmış olan verilerin değişik araçlarla hızlı şekilde işlenebilmesi olarak ifade edilebilir. Makine öğrenme; denetimsiz öğrenme, aktif öğrenme ve yarı denetimli olmak üzere 3 temel öğrenme modeli üzerine kurulmuştur. Denetimsiz öğrenmede sınıf etiketi olmadığından öğrenme süreci denetime tabi tutulamaz. Veri setinde sınıfları belirlemek için kümeleme metodundan

yararlanılır. Veri kümelerine ait etiketleme maliyetinin fazla olduğu durumlarda ya da sınıf etiketlerinin bilinmezliği durumlarda denetimsiz öğrenmeden faydalanılmaktadır (Vural, 2017). Yarı denetimli öğrenmede ise etiketlenmemiş veri ile birlikte etiketlenmiş veriler birlikte bulunmaktadır (Chapelle, Schölkopf, & Zien, 2006). Aktif öğrenme yaklaşımında ise bir alan uzmanından bir veriye etiket vermesi istenir. Buradaki amaç model kalitesinin en üst düzeye çıkarılmasıdır (Vural, 2017). Veri madenciliğinde çok önemli bir fonksiyonu vardır. Büyük veri gruplarından işe yarar gerekli bilgilerin elde edilmesine yardımcı olur. Makine öğrenmenin tarihi bilgisayarın icadına kadar ulaşır. Bilgisayarların nasıl öğrenebileceği üzerine yoğunlaşan makine öğrenme; bireylerin ilgi alanları ve eğilimlerinin belirlenerek bireylere yönelik olay ve durumların yansıtılması , tıp, enerji verimliliği gibi pek çok farklı alanda yararlanılan bir bölüm haline gelmiştir (Mitchell, 1997, s. 1). İlerleyen bölümlerde makine öğrenme algoritmaları üzerinde ayrıntılı olarak durulacaktır.

2.1.2 Veri Tabanı Sistemleri

Veri toplama ve işleme süreçlerinde önemli araçlar olarak ön plana çıkmaktadır. Veri tabanları ve veri ambarları olarak ikiye ayrılır. Veri tabanları; bilgisayarlarda erişilebilir, düzenlenebilir, taşınabilir ve birbirleriyle tanımlanmış ilişkiler bulunan bilgi öbekleri olarak tanımlanabilir (Gunjal & Koganurmath, 2003). Veri ambarları ise; farklı veri tabanlarından toplanan verilerin bir araya getirilerek bir bütün durumuna getirildiği veri kaynaklarıdır.

2.1.3 Bilgi Teknolojileri

Veri madenciliğinde ana kaynak olarak veri depolama sistemlerinin kabul görmesi, bilgiyi işleme ve depolama konusunda bilgi teknolojilerini vazgeçilmez bir alan olarak karşımıza çıkarmaktadır.

2.1.4 Veri Görselleştirme

Veri madenciliğinde, büyük verilerin görsel bir şekilde ele alınması bütün verinin tümel olarak görünür hale gelmesinde ayrıca verilerin kolay bir şekilde analiz edilmesinde ciddi bir şekilde kolaylık sağlamaktadır. Bu bakımdan veri görselleştirmenin de veri madenciliğinin önemli bir parçası olduğu söylenebilir (Uyan & Çay, 2008).

2.1.5 İstatistik

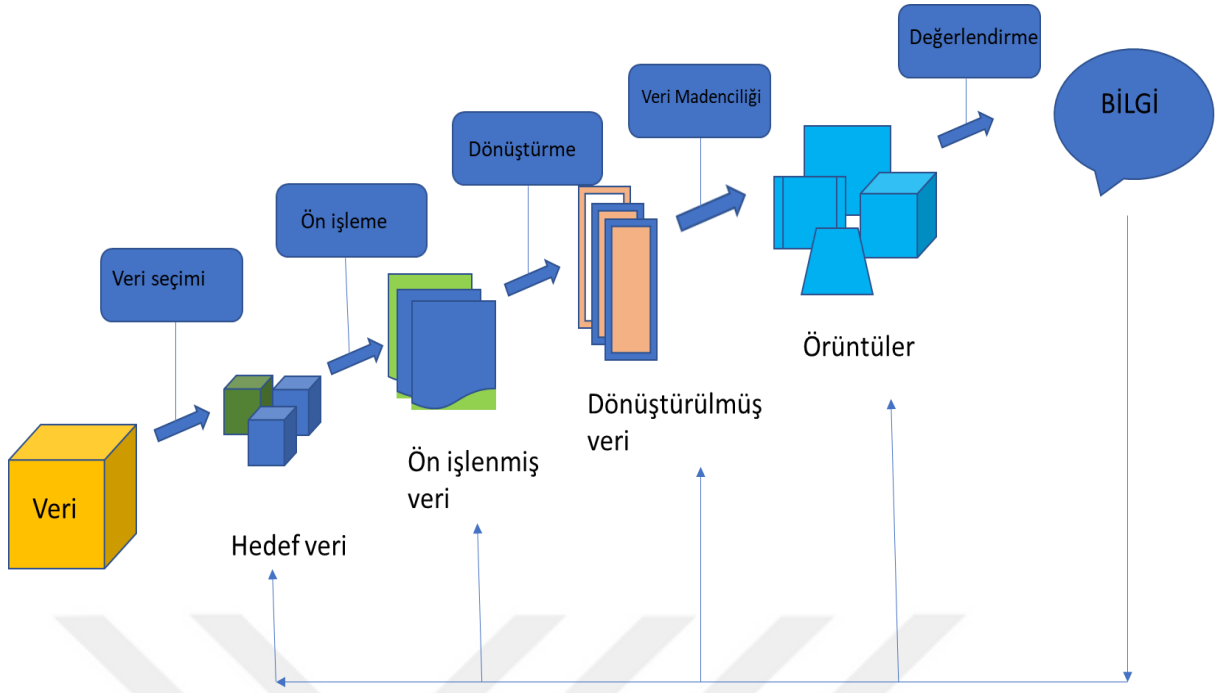
İstatistik, ampirik verileri toplamak, analiz etmek, yorumlamak ve sunmak için yöntemler geliştirmek ve incelemekle ilgilenen bilimdir. İstatistik, tüm veri madenciliği algoritmalarının en önemli parçası olan sayısal verilerin analizini ve sunumunu ifade eder. Büyük miktarda veriyle başa çıkmak için araçlar ve analitik teknikleri sağlar. İstatistik, planlama, tasarlama, bilgi toplama, analiz etme ve araştırma bulgularını raporlamayı içerir. Veriden öğrenme, anlamlı ve değerli bilgiye ulaşma, değişkenler arasındaki ya da verideki farklı örüntüleri keşfetme, analizlere dayalı gerçekleştirilen işlemler veri madenciliği ve istatistik sayesinde yapılabilmektedir (Emre & Erol, 2017).

2.1.6 Diğer Disiplinler

Makine öğrenme ile olan ilişkisinden dolayı veri madenciliği, yapay zeka çalışmaları ile de ilgili bir alandır (Uyan & Çay, 2008). Örüntü tanıma da veri madenciliği ile ilgili alanlardan biridir (Kumdereli, 2012).

2.2 Veri Madenciliği Süreçleri

Veri madenciliğinde süreç gereksiz olan bilgi kaynaklarının temizlenmesi ile başlar. Ardından süreç verinin toplanacağı kaynakların bir araya getirilmesi ile devam eder. Bir araya getirilen kaynaklardan amaçlar doğrultusunda faydalı olacağı düşünülen bilgiler toplanır. Toplanan bu bilgiler belirlenmiş olan ölçütlere göre sınıflara ayrılır. Sınıflandırmaya göre bilgi modelleri oluşturulur. Oluşturulan modeller amaçlara göre değerlendirilir. Değerlendirme sonucu elde edilen bilgi kişi ya da kurumlara iletilir (Han & Kamber, 1999). Veri madenciliği süreci 5 aşamadan oluşur (Shearer, 2000).



Şekil 2.2 *Veri Tabanlarında Bilginin Keşif Süreci*

2.2.1 Problemin Tanımlanması

Veri madenciliği sürecinin tam anlamıyla başarıya ulaşabilmesi için problemin tam anlamıyla tanımlanması, hedeflerin belirlenmesi, süreç sonunda ortaya çıkacak olan sonuçların nasıl değerlendirileceğinin net bir şekilde ortaya konması gerekmektedir. Bu süreç diğer basamaklar için de referans olma özelliği teşkil etmektedir. Bu süreçte belirlenen amaçların nasıl sonuçlar doğuracağına dair değerlendirmeler de yapılır. Yapılan bu değerlendirmeler sonucunda ihtiyaç duyulan veriler ve türleri tespit edilebilir. Bu süreç, genel süreç açısından gereksiz bilgilerin ayıklanmasına yönelik işlevsel bir süreçtir (Toksöz, 2016).

2.2.2 Verilerin Hazırlanması

Ham veriden, son aşamada veri setlerinin elde edilmesine kadar olan süreçlerde gerçekleştirilen bütün işlemleri kapsamaktadır. Veri hazırlamaya yönelik bu işlemler belli bir sıralama takip etmeksizin birçok defa yinelenen işlemleri ifade etmektedir (Jackson, 2002). Verilerin hazırlanması aşaması veri madenciliğinin en çok çaba gerektiren aşamasıdır. Bir sonraki aşama olan modelin kurulması aşamasında ortaya çıkması olası problemler, bu aşamaya yeniden dönülerek verilerin tekrar düzenlenmesine sebep olacaktır. Böyle bir durum araştırmacının genel süreç içerisinde enerji ve zamanının %50-%85'ini harcamasına neden olmaktadır (Piramuthu, 1998). Bu aşamada gerçekleştirilen en önemli işlemlerden biri verilerin temizlenip dönüştürme işlemidir. Böylece ham veri farklı değer ve formatlara dönüştürülür.

Veri temizliđi sayesinde ise veri madenciliđine uygun olmayan ya da hatalı olan veriler ıkarılmıř olur (Tekerek, 2011). Tang ve MacLennan (2005, ss.13-14) veri dnřtrme srecinde yer alan iřlemleri řu řekilde sıralamıřtır:

2.2.2.1 Veri Tiplerinin Dnřtrlmesi

Verilerin farklı biimlerde yeniden yapılandırılması anlamına gelir. Sayısal verilerin kategorik bir veriye dnřmesi rnek olarak verilebilir. Bu dnřtrmenin sebepleri bazı veri tiplerinin bazı algoritmalar ile olan uyumu ya da veri madenciliđine uygun olmayan verilerin dnřtrlerek uygun hale getirilmesidir.

2.2.2.2 Deđiřken Dnřmleri

Srekli deđiřken olarak ifade edilen verilerin gruplandırılarak kategorik veri haline getirilmesini ifade eder (Tang & MacLennan, 2005). Bu iřlemdeki ama, byk boyutlardaki verilerin kk boyutlara indirgenerek zerinde analiz ve yorum yapılabilmesini sađlamaktır.

2.2.2.3 Gruplandırma

Veri setinde birbirinden farklı yapıda veriler yer alabilir. Bu veriler, birkaç veri setinde toplanarak hem model karmařıklıđı giderilebilir hem de kolay bir analiz ve deđerlendirme yapılabilir. Yapılan bu veriyi yeniden yapılandırma iřlemine gruplandırma denir (Tang & MacLennan, 2005).

2.2.2.4 Veri Btnleřtirme

Veri btnleřtirme, aynı veri tabanlarında farklı zamanlarda veya farklı veri tabanlarında bulunan verilerin btnleřtirilerek analiz iin anlamlı duruma getirilmesidir. Bylece veriler arası uyumsuzluk giderilmiř olur (Argden & Erřahin, 2008).

2.2.2.5 Veri Azaltma

Veri azaltma srecinde nemli verilerin de kaybolma ihtimali bulunmaktadır. Bu yzden hesaplama ya da iřlem yapmak iin gerekli zaman, betimsel veya kestirisel dođruluk, modelin temsil durumu gz nnde bulundurulmalıdır (Kantardzic, 2011). Han vd.,(2012, ss.99-100) veri madenciliđinde veri azaltma yntemlerini  ana strateji altında toplandıđını belirtmiřlerdir. Bu stratejilerden ilki boyut azaltmadır. Boyut azaltmada; temel bileřenler analizi, dalgacık dnřmleri ve zellik alt kme seđimi yntemleri bulunmaktadır. İkinci

strateji verileri sayısal olarak azaltmadır. Bu stratejide veriyi temsil edebilecek şekilde daha küçük veri birimi kullanılır. Üçüncü strateji ise veri sıkıştırımadır. Bu stratejide ise veri dönüşümleri orijinal verinin azaltılmış ya da sıkıştırılmış bir temsilini ortaya koyacak biçimde belirlenir.

2.2.3 Modelin Kurulması ve Değerlendirilmesi

Tanımlanmış olan problemin çözümü için, çok sayıda modelin denenmesi gerekebilir. Bu yüzden veri hazırlama ve model kurma aşamaları, en iyi olduğu belirlenen modele kadar tekrarlanan bir süreçtir.

2.2.4 Modelin Kullanımı

En iyi olduğu belirlenen model farklı bir uygulamanın alt parçası olarak ya da doğrudan uygulanır.

2.2.5 Modelin İzlenmesi

Tüm sistem özelliklerinde ve sistemin ürettiği verilerde zamanla ortaya çıkan değişiklikler, modelin devamlı izlenmesine ve gerektiği takdirde yeniden düzenlenmesine neden olabilir.

2.3 Veri Madenciliği Yöntemleri

Bu bölümde veri madenciliği yöntemleri olan; özellik seçme, sınıflandırma, birliktelik ve kümeleme üzerinde durulacaktır.

2.3.1 Özellik Seçimi

Büyük veri yığınlarında gerçekleştirilen veri madenciliği yöntemlerinde hem gözlem hem de özellik sayıları çok fazla olabilmektedir. Böyle durumlarda öncelikle işe faydalı özelliklerin belirlenmesi gerekmektedir. Böylece kullanılan algoritmalar hem hızlı hem de etkili sonuçlar verecek ve sonuca daha hızlı ulaşılacaktır (Gorunescu, 2011). Bu yöntem için literatürde sarmal, karma ve filtreleme yöntemler adı altında uygulanan algoritmalar bulunmaktadır.

2.3.2 Sınıflandırma

Bu yöntem veri madenciliğinde geniş bir yer tutmaktadır. Sınıflandırma işleminin basit düzeydeki tanımı, özelliğin belli olduğu durumlarda sınıf düzeyini ya da değerini tahmin etmektir (Zaki & Jr, 2014). Klasik istatistikte yaygın olarak kullanılan regresyon analizi de bir sınıflandırma yöntemidir. Ancak regresyon analizi için birçok varsayım vardır (Gamgam & Altunkaynak, 2017).

2.3.3 Birliktelik

Birden fazla tercih durumunda bu tercihler arasındaki yapının incelenerek birliktelik yapısı oluşturma yöntemidir. “Dondurma alan kişilerin su alma olasılıkları nedir?” sorusu bu yöntem için güzel bir örnektir.

2.3.4 Kümeleme

Gözlem sonucu elde edilen birimlerdeki benzerlik durumuna göre gruplara ayırma işlemidir. Benzerlikler bulunurken gruplar arası uzaklığa bakılır. Bu uzaklık, oluşturulan uzaklık ölçütlerine göre değerlendirilir.

2.4 Makine Öğrenme Algoritmaları

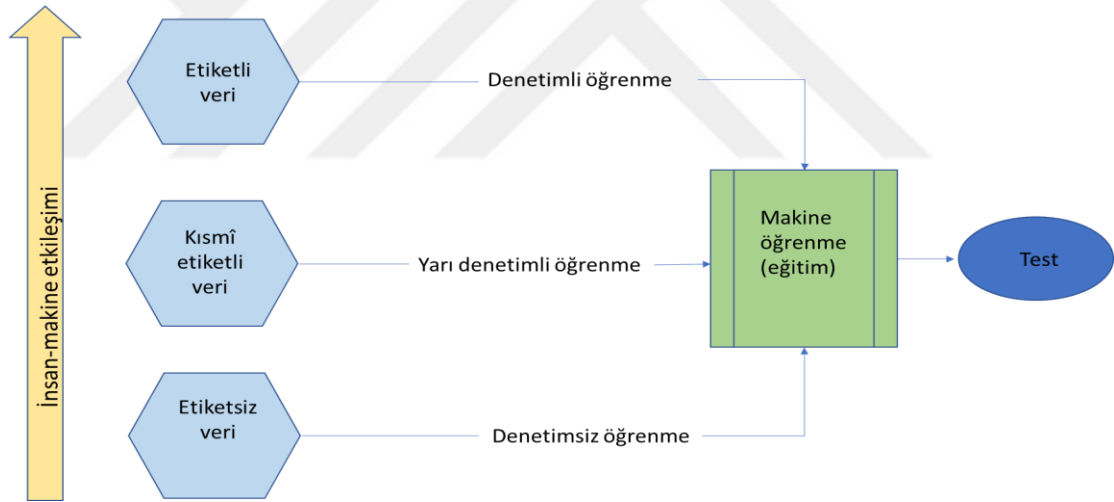
Teknolojinin zamanla gelişmesi veri artışını da beraberinde getirmiştir. Veri artması da geleneksel analiz yöntemlerini yetersiz ve zaman alıcı hale getirmiştir. Böylece yeni analiz yöntemleri arayışına girilmiş ve bu arayışın sonucunda da yapay zekanın alt branşı olan makine öğrenme ortaya çıkmıştır. Makine öğrenmede, veriler arasındaki bağlantı otomatik olarak algılanır ve geleceğe yönelik tahminde bulunulur. Bir makine öğrenimi algoritması, istenen bir görevi gerçekleştirmek adına girdi verilerini kullanarak hesaplama yapılan bir süreçtir. Bu algoritmalar, yapılarını tekrarlama yoluyla değiştirerek yeni durumlara uyarlarlar. Böylece gittikçe görevleri başarmada daha iyi bir duruma gelirler. Yani makine öğrenmede algoritmalar, yeni verileri işlediği ve hatalarından da öğrendiği için her zaman daha iyi sonuca doğru gider. Makine öğrenimine dayalı teknikler, örüntü tanıma, bilgisayarla görme, uzay aracı mühendisliği, finans, eğlence ve hesaplamalı biyolojiden biyomedikal ve tıbbi uygulamalara kadar çeşitli alanlarda başarıyla uygulanmıştır (El Naqa & Murphy, 2015).

Makine öğrenme, denetimsiz, denetimli ve yarı-denetimli olmak üzere 3 alt başlıkta incelenmektedir.

Denetimsiz öğrenme; herhangi bir kullanıcı müdahalesi olmadan etiketsiz veriler üzerinden gerçekleştirilen öğrenmedir. Burada amaç veriler hakkında bilgi sahibi olmak veri yapısını keşfetmektir. Veri seti sistemin belirlediği işleme yöntemi ile gruplara ayrılır. Herhangi bir sınıflandırmaya yönlendirme olmadığından sınıf sayısında herhangi bir sınır yoktur. En yaygın kullanılan teknik kümeleme tekniğidir (Ayaz, 2021, s. 18).

Yarı denetimli öğrenme; verilerin bir kısmının etiketlendiği bir kısmının etiketlenmediği öğrenme türüdür. Bu öğrenme türünde etiketlenmiş bir veri etiketlenmeyen verinin öğrenilmesinde yardımcı olarak kullanılabilir (El Naqa & Murphy, 2015, s. 5).

Denetimli öğrenme; en fazla kullanılan öğrenme yöntemidir. Bu öğrenme yönteminde amaç etiketlenmiş bir veri kümesi verildiğinde girdilerden çıktıya doğru bir eşleştirme yaratmaktır. Bu tür öğrenmelerde veri setinde yer alan girdiler, kategorik veya sayısal değerler almaktadır. Dolayısıyla çıktılar da sayısal veya kategorik olmakta, değişkenlerin türüne göre de sınıflama ve regresyon modelleri olarak iki bölümde incelenmektedir. Kategorik özelliklere sahip olan çıktılar sınıflama, sayısal olan çıktılar ise regresyon modeli olarak isimlendirilmektedir (Ayaz, 2021, s. 19).

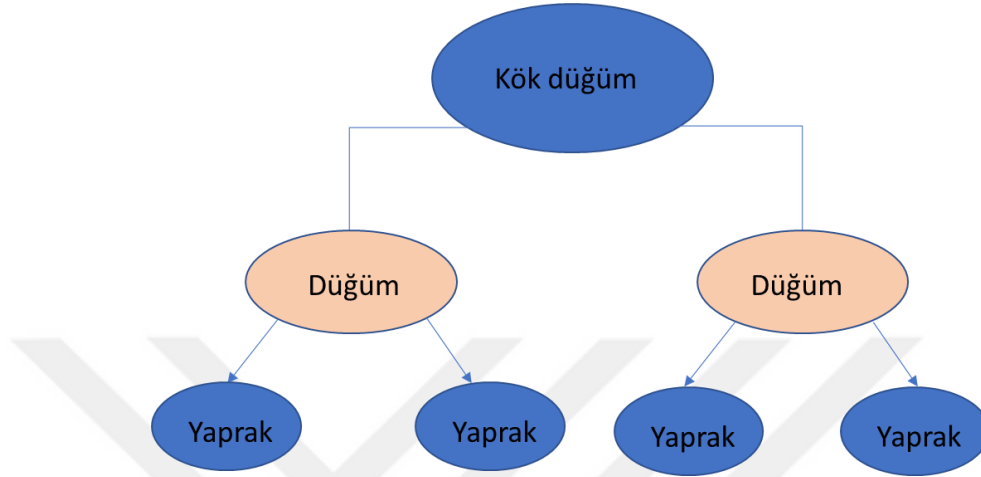


Şekil 2.3 Verilerin Durumuna Göre Makine Öğrenme Teknikleri

2.4.1 Karar Ağaçları

Karar ağaçları, tıp, ilaç sanayi, yönetim bilimleri ve eğitim gibi birçok alanda kullanılmaktadır. “Sınıflandırma ağaçları” ya da “regresyon ağaçları” olarak da isimlendirilmektedir. Veri madenciliğinin yapı taşları olarak kabul edilirler. Karar ağaçları klasik makine öğrenmesine dayalı veri madenciliği modellemelerinin en önemli elemanlarından biridir ve ilk kez 1980’lerde geliştirilmişlerdir. Bu denli yaygın olarak kullanılmalarının en önemli nedeni sonuç modelinin basit olmasından kaynaklanmaktadır. Karar ağaçları regresyonla birlikte sınıflandırma modellerini de temsil edebilir (Williams, 2011).

Bir sınıflandırma metodu ortaya koymak adına veriler basit anlamda iki ana kategoriye ayrılır. İlk kategori “eğitim seti” ikinci kategori ise “test seti” olarak adlandırılır. “Eğitim seti” sınıflandırma ölçütü olarak kabul edilirken, “test seti” ise ölçütün geçerliliğini test eder. Sınıflandırma ölçütüne ait geçerlik, doğru sınıflandırılan test verilerinin, bütün test verisine oranı ile ilgilidir (Pujari, 2008).



Şekil 2.4 Bir Karar Ağacının Yapısı

Bir karar ağacı “kök” adı verilen bir merkez noktası ile başlar. Bu noktaya “kök düğüm” adı verilmektedir (Han, Kamber, & Pei, 2012). Bu kök düğümünden başlayan dallarda karar düğümleri yer almaktadır. Karar düğümleri kökten başlayarak, aşağıya doğru bir yapı oluşturacak şekilde sonuç değişkenine doğru bir yol takip etmektedir. Karar ağaçlarında hedeflenen değişkene dair nihai bir karara ulaşılması beklenmektedir. Bu yüzden kökten dallarla ayrılmış olan karar düğümleri, son değişkene varıncaya dek devam ederler (Larose, 2005). Karar ağacının son aşamasında yer alan ve sınıf etiketlerini belirleyen düğümlere “uç düğüm”, uç düğümden önceki düğümlere ise “iç düğüm” denilmektedir (Han, Kamber, & Pei, 2012).

Karar ağaçları oluşturulmadan önce, verinin sınıflandırmaya uygunluğu analiz edilir. Bu sebeple karar ağacı modeli oluşturabilmek için verilerin bazı varsayımları karşılaması gerekmektedir:

- Verinin sınıflanabilir düzeyde olması gerekmektedir.
- Verinin konuyla ilgili olması ve gereksiz verilerin temizlenmiş olması gerekmektedir.
- Problemin iyi tanımlanmış olması gerekmektedir.
- Sonuç, üzerinde işlem yapılabilecek bir bulgu olmalıdır (Patel, Prajapati, & Lakhtaria, 2012).

- Karar ağacı oluştururken başvurulmuş olan yöntem büyük ölçekteki verileri de ele alabilmelidir.
- Kategorik nitelikteki verileri analiz etme durumunda yeterli olmalıdır (Pujari, 2008).

Karar ağacının oluşum süreci üç ana evrede gerçekleşmektedir;

İlk aşamada eğitim veri seti ile karar ağacı oluşturulur. Eğitim veri seti tekrarlanan şekilde bir bölme kriteri kullanılarak iki veya daha fazla kategoriye ayrılır. Bu ayırma işlemi sınıflandırma sonuç kriteri elde edilinceye kadar sürer. İkinci aşamada oluşturulmuş olan karar ağacının daha iyi performans gösterebilmesi için alt dal veya düğümlerden bazıları kaldırılır. Üçüncü aşamada ise karar ağacının daha anlaşılır ve net duruma gelmesi için budanmış olan ağaç işlenir. Karar ağacı algoritmalarından bazıları aşağıda tanımlanmıştır.

2.4.2 ID3 Algoritması

Quinlan (1986) tarafından karar ağaçları yoluyla sınıflandırma işlemi yapmak üzere geliştirilen bir algoritmadır. Bölünme değeri olarak entropiden yararlanır. ID3 ile yapılan karar ağacı oluşturma sürecinde gerçekleşecek bölünmenin hangi niteliğe göre yapılacağı konusunda başvuru bölünme değeri entropidir (Özkan, 2008). Bu algoritmada, karar ağacındaki her düğüm bölünmüş olan niteliğe, düğümlere işaret eden her bir ok ise o niteliğin bir değerine işaret etmektedir. Bir düğümün ne kadar bilgi içerdiği entropi değeri ile belirlenir. Bölünme iyiliği indeksi olarak ise bilgi kazanımı değeri temel teşkil etmektedir. En fazla bilgi kazanım değerine sahip nitelik bölünme özelliği olarak kabul edilerek, veri seti bu niteliğin tüm farklı değerlerine göre alt kategorilere ayrılır (Pujari, 2008).

ID3 algoritmasının dallandırma yapacağı özelliği seçerken kazanım değeri (2.1) eşitliği ile bulunur.

$$Kazanım(X_i) = H(Y) - \sum_{j=1}^{k_i} P(X_{ij})H\left(\frac{Y}{X_{ij}}\right); i = 1, 2, \dots, m \quad (2.1)$$

Formülde belirtilen değerlere ait tablo ise şu şekildedir:

$Kazanım(X_i)$	i .bağımsız değişken için kazanım değeri
$H(Y)$	Sınıfa ait <i>entropi</i> değeri
k_i	i .özelliğin kategori sayısı
$P(X_{ij})$	i .özelliğin j .düzeyinin ortaya çıkma olasılığı
$H(Y/X_{ij})$	i .özelliğin j .düzeyi için sınıfa ait <i>entropi</i> değeri
M	özellik sayısı

Kazanım değeri büyük olan özellik, sınıf üzerinde en fazla belirleyiciliğe sahip özellik olarak alınarak dallandırma için seçilir. Herhangi bir düğümde sınıf değeri tek bir kategoride

toplanıyorsa o düğüme saf düğüm denir. ID3 algoritmasında sınıflandırma bu şekilde son bulur. Yukarıda belirtilen entropi değeri, bir değişkendeki belirsizliği ölçmek için kullanılır. Değişkenin bütün değerleri birbirine eşit olduğunda belirsizlik yoktur ve entropi değeri sıfır olur. Ancak tamamen birbirinden farklı değerler alıyorsa entropi maksimum değerini alır. Bir değişkene ait k tane düzey için entropi hesabı (2.2) eşitliği ile hesaplanır:

$$H(Y) = H(p_1, p_2 \dots p_k) = \sum_{j=1}^k (p_j \log_b \left(\frac{1}{p_j} \right)) \quad (2.2)$$

Bu formülde p_j notasyonu j . düzeyin ortaya çıkma olasılığını gösterir. b ise logaritma tabanıdır. $b=2$ olarak alındığında elde edilen entropi değerine *Shannon entropisi* denir. Birimi bit olarak ifade edilir. Ancak bu entropi, değişkenin iki düzeye sahip olduğu durumlarda tercih edilir. Değişken ikiden fazla değere sahipse $b=10$ alınarak elde edilen *Hartley entropisi* kullanılmaktadır.

ID3 algoritmasının adımları şu şekildedir:

- Genel entropi değeri hesaplanır,
- Her bir özellik için kazanım değeri hesaplanır,
- En büyük kazanım değerine sahip özelliğin düzeylerine göre dallandırma gerçekleşir,
- Yeni oluşan ve saf olmayan düğümlerin her biri için veri indirgenir ve 1.adıma tekrar dönülür. Tüm düğümler saf olduğunda işlem biter.

2.4.3 C4.5 Algoritması

Temelinde ID3 algoritması olan ve Quinlan tarafından geliştirilen bir algoritmadır. Sayısal değeri olan özellikler de bu algoritma ile karar ağacı oluşturma imkânı sağlamıştır. Buna ek olarak bilinmeyen nitelik değerleri için örnek kümeleri ile karar ağacının nasıl oluşturabileceği hakkında bir yol sunar (Özkan, 2008). Bilinmeyen nitelik değerlerinin hesaplanması, sayısal ve sürekli olan değerler ile karar ağacı oluşturma, karar ağaçlarında budama işleminin gerçekleştirilmesi ve kural oluşturma olmak üzere dört temel işlevi yerine getirir. Karar ağacı oluştururken mevcut durumdaki veri eğitim setlerinde bilinmeyen özellik değerlerine dair sınıflandırmaların, “bilgi kazanımı” (information gain) ve “kazanç oranı” (gain ratio) gibi değerlere sahip olması gerekmektedir. C4.5 algoritması her boğuma ait farklı bölünmeler gerçekleştirebilir. Belirlenmiş olan bir özellik, bölünme özelliği olarak tercih edildiğinde, karar ağacında özelliğin tüm değerlerine ait bir dal bulunacaktır (Pujari, 2008). Kazanım oranı (2.3) eşitliği ile verilmiştir:

$$KazanımOranı(X_i) = \frac{Kazanım(X_i)}{H(X_i)}; i = 1, 2, \dots, m \quad (2.3)$$

Bu eşitlikte $H(X_i)$, X_i özelliğine ait entropiyi göstermektedir. Eğer bu özelliğin kategori sayısı fazla ise entropi değeri de büyük olacaktır. Eşitlik (2.3)'den de görüldüğü gibi X_i özelliğinin kazanım değeri entropi değerine oranlanmakta ve kategori sayısına göre bir düzeltme işlemi yapılmaktadır. Kazanım oranı büyük olan özellik sınıf üzerinde en fazla belirleyici özellik olarak alınıp, dallandırma için seçilir. Her ne kadar C4.5 algoritması kategorik özellikler için kullanılsa da bazı yazılımlar sayısal özellikler için belli eşik değerlerine göre kategorilere ayırma işlemini otomatik olarak yapmaktadır. Bu sayede sayısal özelliklerin yer aldığı durumlarda da bu algoritma uygulanabilmektedir.

2.4.4 C5.0 Algoritması

Büyük veri kümesinde geçerli bir sınıflandırma algoritması olan C5.0 algoritması aynı zamanda ID3 ve C4.5 algoritmasının bir uzantısıdır. Verimlilik açısından bakıldığında C4.5 algoritmasından daha iyidir. Bu model, örnekleri en büyük bilgi kazancına göre bölerek çalışır. Önceki bölümlerden alınan alt kümeler, örneklerden elde edilen alt kümeler bölünemeyecek hale gelene kadar bölünmeye devam eder. En son olarak modele herhangi bir katkısı olmayan örnek alt kümeler çıkarılır (Patil, Lathi, & Chitre, 2012).

2.4.5 CART (Classification and Regression Trees) Algoritması

Karar ağacı oluşturma sürecinde en çok kullanılan algoritmalarından bir tanesidir. 1984 yılında Breiman, Friedman, Olshen ve Stone tarafından geliştirilmiştir. CART algoritmasının yorumlanması daha zor olsa da, bu algoritma ile oluşturulan ağaçlar tek değişkenden oluşan karar ağaçlarına göre daha küçüktür ve daha doğru bilgiler içerir (Witten & Frank, 2005).

Hem kategorik, hem de sürekli değişkenlerden yararlanarak, sınıflandırma ve regresyon problemlerinin çözümünde karar ağaçlarının kullanıldığı, parametrik olmayan istatistiksel bir metottur. Bağımlı değişken eğer kategorik bir veri ise yöntem sınıflama ağaçları, sürekli bir veri ise regresyon ağaçları olarak isimlendirilir. Bu açıdan bakıldığında CART algoritması hem lojistik regresyon analizi hem de çoklu doğrusal regresyon analizini kapsamaktadır. (Güner, 2014).

CART algoritması ile oluşan karar ağaçları ikili karar ağacı özelliği taşır ve her karar boğumu iki dal ile birbirinden ayrılır. Hedeflenen bağımlı değişkendeki özellik durumuna göre CART karar ağacı sürekli olarak elde edilen kayıtları alt kategorilere bölmektedir (Larose D. T., 2005).

Diğer algoritmaların aksine CART algoritmasında ilk başta çok büyük bir karar ağacı oluşturulur. Daha sonra bu ağaç budanarak yanlış sınıflandırma hatası en az olacak seviyeye

indirilir. Bu algoritmada hata katsayılarının hesabında 10 katlı çapraz doğrulamadan faydalanılmaktadır (Loh, 2011). CART algoritması hesaplama yaparken Gini indeksinden yararlanır. Bu indeks ikili ayırmalara dayanan bir sınıflandırma metodudur (Özkan, 2008).

Sınıfın kategorik olduğu durumlarda genellikle twoing kriteri veya gini ölçüsü, sınıfın sayısal olduğu durumlarda dallandırma için hata kareler toplamı kullanılmaktadır. Twoing kriterine göre yapılan dallandırmalarda her bir bağımsız değişkenin bütün mümkün ikili ayıştırmaları dikkate alınır. Twoing kriteri (2.4) eşitliği ile verilmiştir:

$$\phi(D_i) = 2P_L P_R \sum_{j=1}^k [|P(Y_j/L) - P(Y_j/R)|] \quad (2.4)$$

Burada kullanılan ifadelerin anlamları şu şekildedir:

$\phi(D_i)$ = i .bölünme için twoing kriteri

P_L = D_i bölünmesi için bir gözlemin ağacın solunda kalma olasılığı

P_R = D_i bölünmesi için bir gözlemin ağacın sağında kalma olasılığı

$P(Y_j/L)$ = Ağacın solunda sınıfın j . düzeyinin olasılığı

$P(Y_j/R)$ = Ağacın sağında sınıfın j . düzeyinin olasılığı

Sınıfın kategorik olarak değerlendirildiği durumlarda kullanılan diğer bir ölçü olan Gini değeri ise (2.5) eşitliği ile elde edilir.

$$Gini(X_{ij}) = 1 - \sum_{m=1}^k [P\left(\frac{y_m}{x_{ij}}\right)]^2 \quad (2.5)$$

Burada k sınıfın düzey sayısını $P(Y_m/X_{ij})$ ifadesi i . özelliğin j . düzeyinde sınıfın m . düzeyinin görülme olasılığını gösterir. Gini ölçüsüne göre dallandırma özellikleri için Gini Ayırma değerine dayalıdır.

Bu ayırma değeri (2.6) eşitliğinde gösterilmiştir:

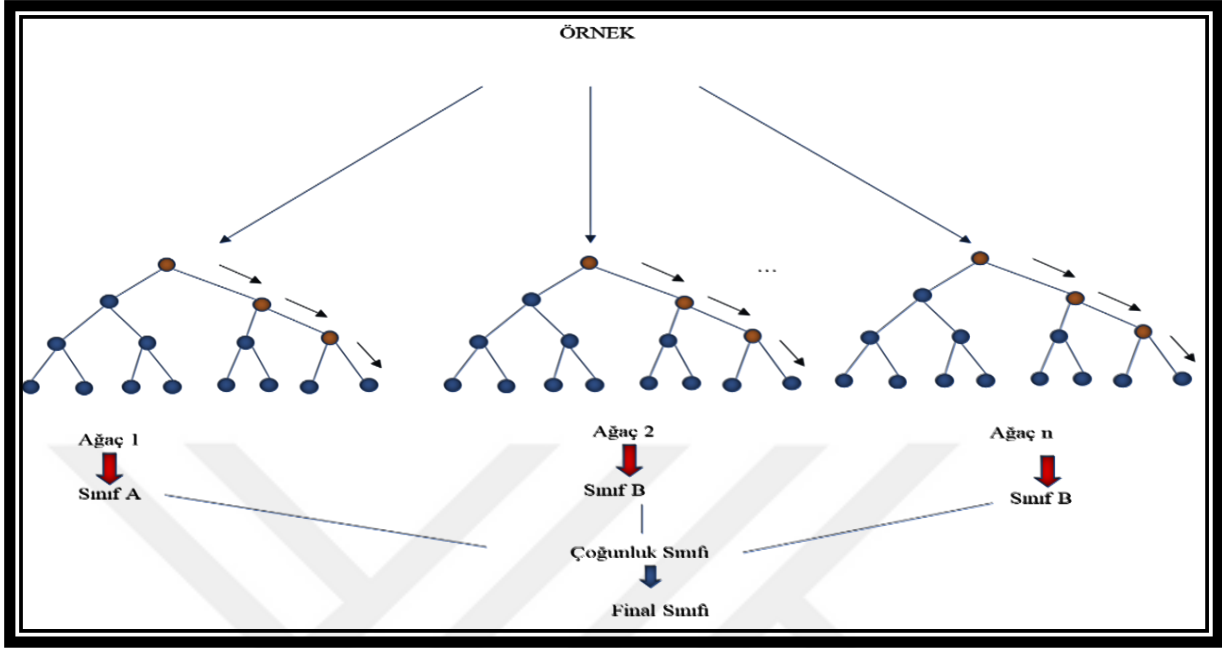
$$Gini_{split}(X_i) = \sum_{j=1}^k P(X_{ij}) Gini(X_{ij}) \quad (2.6)$$

Bu değerde X_i ifadesi i . özelliği, X_{ij} ifadesi i . özelliğin j . düzeyini, k özelliğin düzey sayısını, $P(X_{ij})$ i . özelliğin j . düzeyinin çıkma olasılığını gösterir. Gini ayırma değeri küçük olan özellik sınıf üzerinde en fazla belirleyiciliğe sahip olan özellik olarak alınarak dallandırma için seçilir.

2.4.6 Rastgele Ormanlar

2001 yılında Leo Breiman tarafından geliştirilmiştir. Breiman'ın geliştirdiği “Bagging” yöntemi ile Kim Ho tarafından geliştirilen “Random Subspace” yönteminin birleşimidir. Topluluk öğrenme yöntemidir. Sınıflandırma esnasında birden fazla karar ağacı üreterek, sınıflandırma kalitesini yükseltmeyi hedefleyen bir algoritmadır. Tek tek oluşturulmuş olan karar ağaçları bir araya gelerek bir karar ormanı oluşturmaktadır. Burada tek olarak bahsedilen

karar ağaçları veri setinden rastgele seçilmiş birer alt kümedir. Algoritmadaki ağaç sayısı ile sonuç arasında doğrudan bir ilişki bulunmamaktadır. Karar ağacı ile arasındaki fark kök düğümü bulma ve düğümleri bölme işleminin rastgele çalışıyor olmasıdır.



Şekil 2.5 Rastgele Orman Algoritmasının Görselleştirilmesi (Koehrsen, 2021)

Rastgele ormanların avantajları şu şekilde sıralanabilir:

- Hem sınıflandırma hem regresyon görevlerinde kullanılabilir.
- Rastgele ormanda yeteri kadar ağaç varsa aşırı uyum (over-fitting) probleminin ortaya çıkma ihtimali düşüktür.
- Kısa zamanda sonuç alınır.
- Gürültülü verilerden temizler.
- Diğer yöntemlere göre daha net sonuçlara sahiptir.

Dezavantajları ise şu şekilde sıralanabilir:

- Karmaşıklığı yüksek ve yorumlaması zor bir algoritmadır.
- Donanımsal olarak yüksek hesaplama gücü gerektirir.

Bankacılık, tıp, eğitim, borsa gibi birçok alanda kullanılan rastgele ormanların uygulaması kullanıcı tarafından bu algoritmaya 2 parametre verilmesiyle başlar. İlk olarak veri setinin %80'i eğitim veri seti olarak ayrılır. Bu set ile önyükleme örnekleri oluşmuş olur. Daha sonra, kalan %20 'lik kısım hataları test etmek amacıyla test verisi olarak kullanılır. Her bir düğümde, ağaç sayısı tüm değişkenler arasından rastgele seçilerek en iyi dal belirlenir. Sınıfların homojenliğini ölçmek için ise Gini indeksi kullanılır. Gini indeksi ne kadar düşük olursa sınıf

da o derece homojen olarak kabul edilir. Bu indeksin hesaplaması (2.7) eşitliğinde gösterilmiştir.

$$Gini = 1 - \sum_{i=1}^n (p_i)^2 \quad (2.7)$$

Bu hesaplamada n seçilen veri sayısını $(p_i)^2$ ifadesi ise; veri setinde yer alan her bir verinin küçük ve büyük olan tüm verilere bölümünün karesini belirtmektedir.

2.4.7 Destek Vektör Makinesi

Destek vektör makineleri (DVM), “istatistiksel öğrenme teorisini” temeline dayanan kontrollü bir sınıflandırma algoritması olarak tanımlanabilir. Önceleri, destek vektör makinelerine ait olan algoritmalar iki sınıf içeren doğrusal verilerin sınıflandırılması için ortaya çıkmış, sonraları bu algoritmaların kullanım alanı, çok sınıflı ve doğrusal olmayan verilerin de sınıflandırılmasını sağlayacak şekilde geliştirilmiştir. DVM çalışma prensibi iki sınıf arasındaki farklılığı ortaya çıkarabilecek en uygun karar fonksiyonun tanımlanması, farklı bir anlatımla iki sınıfı birbirinden uygun biçimde ayırabilen “hiper-düzlemin” (hyperplane) tanımlanması temeline dayanmaktadır (Vapnik, 1998). DVM’ler doğrusal olmayan regresyon analizi ve yapı sınıflandırma problemlerinde sıkça kullanılan sınıflandırma algoritmalarıdır. Sınıflandırma birçok alanda araştırma konusu olabileceğinden birçok farklı durum için kullanılan bir yöntem olarak kabul edilmektedir (Chu, Jin, & L.Wang, 2005). DVM’ler genel olarak destek vektör sınıflandırıcı (support vector classifier, SVC, DVS) ve destek vektör regresör (support vector regressor, SVR,DVR) olarak ikiye ayrılmaktadır. DVM’ler veri madenciliği algoritmalarının içinde en doğru ve etkili sınıflandırma araçlarından biri olarak kabul edilmektedir. Çok katmanlı girdi dönüşümlere ait genellemeler üzerinde oldukça başarılı sonuçlar elde edilmektedir (Cortes & Vapnik, 1995). DVM’ler köklü ve sağlam bir teorik temele sahip, makine öğrenme için çok az örnek gerekliliği olan ve çoğunlukla boyut sayılarına duyarsız kalabilen algoritmalarıdır. Son yıllarda hem kuramsal hem de pratik alanda hızlı bir gelişme göstermektedir (Kuzey, 2012). Destek vektör makineleri veriyi iki veya daha fazla sınıfa ayırma potansiyeline sahiptir. Bunu, iki boyutlu uzayda doğrusal, üç boyutlu uzayda düzlemsel ve çok boyutlu uzayda “hiperdüzlem”(hyperplane) şeklindeki ayırma mekanizmaları sayesinde gerçekleştirir (Güran, Uysal, & Doğrusöz, 2014). Genel anlamda destek vektör sınıflandırıcısı temsili (2.8) eşitliğindeki şekliyle gösterilebilir.

$$f(X) = b + \sum_{i \in S} \alpha_i K(x_i, x_i') \quad (2.8)$$

Burada; $K(x_i, x_i')$ iki gözlemin benzerliğini ölçen çekirdek fonksiyon,

S ; bu destek noktalarının endekslerinin toplamı,

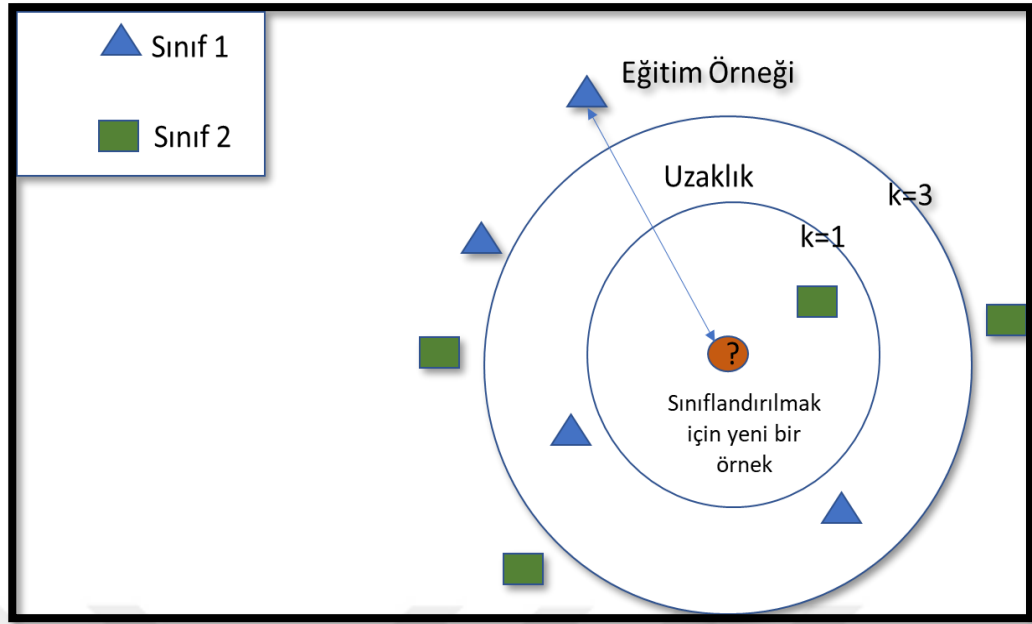
α_i ve b değeri ise; tahmin edilmesi gereken parametrelerdir.

2.4.8 K-En Yakın Komşu Algoritması (K-Nearest Neighbour-KNN)

1950’li yıllarda ortaya çıkan bu algoritma, 1960’lı yıllara kadar beklenen ilgiyi görememiş, hesaplama sistemlerinin gelişmesi ile birlikte kabul görerek günümüzde özellikle örüntü tanıma konusunda yaygın olarak kullanılmaktadır (Han, Kamber, & Pei, 2012). Makine öğreniminde, KNN, sınıflandırma ve regresyon problemlerini çözmek için yaygın olarak kullanılan parametrik olmayan bir tekniktir (Akhil, Deekshatulu, & Chandra, 2013). Temeli karşılaştırma yoluyla öğrenmeye dayanmaktadır. Etiket bilinmeyen teste ait bir verildiği zaman KNN algoritması eğitim veri kümesinde bulunan en “k” sayıdaki komşuya bakarak çoğunluk olan etikete göre atamayı gerçekleştirmektedir (Han, Kamber, & Pei, 2012). Yani bağımsız değişkenlerin sayısal olduğu durumlarda kullanılmakta ve gözlemler arasındaki uzaklıklara göre sınıflandırma yapmaktadır. Bu sınıflandırmayı yaparken şu adımlar takip edilir;

- “k” değeri (komşu sayısı) belirlenir,
- Gözlemler arası uzaklıklar hesaplanır,
- Uzaklığa göre gözlemler küçükten büyüğe doğru sıralanır,
- En küçük uzaklığa sahip olan “k” tane gözlemde en fazla tekrarlanan sınıfa atama yapılır.

Bu algorithmada en önemli sorunlardan birisi “k” değerinin seçimidir. “k” değeri küçük olarak belirlenirse sonuçlar aykırı değerlere aşırı duyarlı duruma gelecektir. “k” değerinin büyük olduğu durumlarda ise seçilen komşular birçok sınıf elemanlarından meydana gelecek ve sınıflandırma performansına olumsuz yönde etki edecektir. Bu yüzden alanyazında “k” değerinin seçimi konusunda ön çalışma yapılması ve çapraz geçerliliklerin (cross-validation) incelenmesi tavsiye edilmektedir. Ayrıca bu algoritmanın uygulama aşamasındaki performansını komşu sayısı, uzaklık ölçüsü ve verileri ağırlıklandırma gibi faktörler etkilemektedir.



Şekil 2.6 KNN Algoritmasının Görselleştirilmesi

2.4.9 Naive Bayes Sınıflandırma Algoritması

Tüm ölçütlerin sonuç üzerindeki etkilerinin olasılıksal olarak hesaplanmasına dayanan bir veri madenciliği modelidir. Bu algorithmada sonucun çıkma olasılığı o sonuca etki eden tüm faktörlerin o sonucu sağlama olasılıklarının çarpımı olarak hesaplanır (Bilekdemir, 2010, s. 34). Thomas Bayes'in teoremine dayanan bu sınıflandırıcılar istatistiksel sınıflandırıcılar olup, belli bir değişkenler grubunun sınıf üyeliği olasılığını tahmin ederler (Han, Kamber, & Pei, 2012, s. 350).

Sınıflandırma problemlerinde amaç belirli tahmin yapan değişkenler kümesinin verildiği durumlarda sınıfların üyelik olasılıklarına dair tahminde bulunmaktır. Bu tür olasılıklar da koşullu olasılık olarak tanımlanmaktadır. Bayes teoreminde, olasılık dağılımı içinde rastgele bir değişken için koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişki gösterilir. Bu iki olay arasındaki koşullu olasılığın gösterimi (2.9) eşitliği ile verilmiştir;

$$P(A|B) = \frac{P(A).P(B|A)}{P(B)} \quad (2.9)$$

Bu formülde $P(A)$ ifadesi A olayının marjinal olasılığını $P(B)$ ifadesi B olayının marjinal olasılığını, $P(A|B)$ ifadesi B olayı bilindiğinde A olayının koşullu olasılığını, $P(B|A)$ ifadesi de A olayı bilindiğinde B olayının koşullu olasılığını ifade etmektedir.

Naive Bayes sınıflandırma algoritması, tahmin için kullanılacak olan tüm özelliklerin kendi arasında bağımsız olduğunu ancak tahmin edilecek durumla bağımlı olduğunu

varsaymaktadır. Bu varsayıma göre; Naive Bayes sınıflandırıcı denklemi Bayes formülünden türetilmiştir. Bu eşitlik (2.10)' da verilmiştir.

$$P(C_i|X) = \frac{P(C_i).P(X|C_i)}{P(X)} \quad (2.10)$$

Bu denklemde X değeri birbirinden bağımsız olan özellikleri C_i ise sınıf sayısını göstermektedir. Denklemdaki $P(X)$ olasılığı her sınıf için eşittir. Naive Bayes sınıflandırma algoritması en yüksek olasılık değerine sahip sınıfı seçer.

2.5 Sınıflandırma Kalitesinin Ölçümü

Sınıflandırma amacıyla kullanılan algoritmalara ait bilgiler yukarıda belirtilmişti. Bu algoritmaların arasından seçim yapılırken bu algoritmaların ne derece iyi olduğuna ilişkin bir değerlendirmenin de yapılması gerekmektedir. Bu değerlendirme için iki aşamalı bir yaklaşım kullanılır. Bu aşamalardan ilki veri setinin eğitim (training data) ve test verisi (test data) olarak ikiye ayrılmasıdır. İkinci aşamada ise sınıflandırmaya dair performans ölçüleri hesaplanır. Bu ölçütler birbiriyle karşılaştırılarak kullanılacak olan sınıflandırma algoritması belirlenir.

2.5.1 Verinin Ayrıştırılması

Sınıflandırma kalitesinin ölçümünde ilk basamak olan veri setinin eğitim ve test verisi olarak ikiye bölünmesinde de farklı yaklaşımlar söz konusudur.

Bu yaklaşımlardan en yaygın olarak kullanılanlarından biri yüzdesel bölme (percentile split) olarak adlandırılır. Bu bölme işleminde verinin belli bir yüzdelik kısmı test verisi olarak ayrılmaktadır. Bu oran farklılıklar gösterebilir.

Bir diğer bölme yöntemi “*n-Bölme*”(n-fold) dur. Bu yöntemde veri rassal şekilde “*n*” parçaya ayrılır. Verinin ilk parçasından başlayarak *n*. parçasına kadar her bir parça sırayla test verisi olarak kullanılır.

Tekrarlı *n*-bölme yönteminde ise aynı *n*-bölme yönteminde olduğu gibi veri rassal olarak *n* parçaya ayrılır, verinin ilk parçasından başlayarak *n*.parçasına kadar her bir parça sırayla test verisi olarak kullanılır. Bu yöntemin farkı bu işlemlerin tekrar etmesidir.

1 eksiltme yöntemi ise *n*-bölme yönteminin özel bir halidir. Gözlem sayısı ve *n*-bölme yöntemindeki *n* sayısı eşit olarak alındığında yöntem 1 eksiltme yöntemine dönüşür. Bu yöntemde her bir gözlem test için ayrılır ve geriye kalan *n*-1 gözlem algoritmayı eğitim amaçlı kullanır.

2.5.2 Sınıflandırma Performansı Değerlendirme Ölçütleri

Sınıflandırma algoritmaları ile oluşturulan modeller değerlendirilirken, hangi sınıflandırma algoritmasının daha doğru sonuçlar ürettiğine belirlenmesinde bazı ölçütler kullanılmaktadır. Bu metrikler, genelde karışıklık matrisi olarak adlandırılan bir tabloya dayanmaktadır. Karışıklık matrisi tablosu sınıflandırıcının performansını görsel hale getirmek için kullanılmaktadır. Matriste bulunan satırlar gerçek değerleri, sütunlar ise tahmin değerlerini göstermektedir. İki sınıflı bir karışıklık matrisi aşağıdaki gibidir:

Tablo 2.1. *İki Sınıflı Bir Karışıklık Matrisi*

	Kestirilen Sınıf		
	Sınıf	Pozitif	Negatif
Gerçek Sınıf	Pozitif	DP	YN
	Negatif	YP	DN

Bu matriste belirtilen DP (doğru pozitif); sınıflandırıcı tarafından yapılan sınıflandırmada pozitif değere sahip olan verilerin kaç tanesinin doğru şekilde sınıflandırıldığını belirtir.

Matriste yer alan DN (doğru negatif); sınıflandırıcı tarafından yapılan sınıflandırmada negatif değere sahip olan verilerin kaç tanesinin doğru şekilde sınıflandırıldığını belirtir.

Matriste yer alan YP (yanlış pozitif); gerçekte negatif değere sahip bir verinin sınıflandırma sonucunda pozitif olarak etiketlenerek sınıflandırıldığı bölümdür.

Matriste yer alan YN (yanlış negatif); gerçekte pozitif değere sahip bir verinin sınıflandırma sonucunda negatif olarak etiketlenerek sınıflandırıldığı bölümdür.

Karışıklık matrisiyle birlikte kullanılan performans değerlendirme metriklerinden biri de doğruluktur (accuracy). Eğitim kümesi kullanılarak oluşturulan modelin test kümesindeki verileri doğru sınıflandırma oranı doğruluk metriğini verir. Doğruluk, (2.11) eşitliği yardımıyla hesaplanır;

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (2.11)$$

Duyarlılık (recall), diğer bir deyişle doğru pozitif oran (DPO) sınıflandırıcının pozitif sınıfa ait olan verilerin doğru olarak sınıflandırılmasıdır. (2.12) eşitliği yardımıyla hesaplanır;

$$\text{Duyarlılık}(DPO) = \frac{DP}{DP+YN} \quad (2.12)$$

Kesinlik (precision), pozitif tahmin değeri (PTD), sınıflandırma sonucu pozitif olarak sınıflandırılan verilerin ne oranda doğru tahmin edildiğini gösteren değerdir. (2.13) eşitliği yardımıyla hesaplanır;

$$Kesinlik(PTD) = \frac{DP}{DP+YP} \quad (2.13)$$

Kesinlik ve duyarlılık ölçülerinin tek başına anlamlı bir karşılaştırma sonucu çıkaramadığı durumlarda iki ölçütü bir arada değerlendiren f -ölçütü daha anlamlı sonuçlar çıkarabilmektedir. f -ölçütü 0 ile 1 arasında değer almaktadır. Bu değer 1'e yaklaştıkça sınıflandırıcının doğruluk tahmini de artmaktadır. f -ölçütü (2.14) eşitliği yardımıyla hesaplanır (Japkowicz, 2011).

$$f - \text{ölçütü} = 2 * \frac{Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık} \quad (2.14)$$

İşlem karakteristiği eğrisi olarak adlandırılan ROC eğrileri de istatistiksel model performanslarının değerlendirilmesinde kullanılan bir ölçüttür. ROC eğrilerinde doğru pozitif oranı Y ekseninde yanlış pozitif oranı da X ekseninde çizilerek grafiksel olarak gösterilir. ROC eğrisinin altında kalan alan algoritmanın sınıflandırma başarısını gösterir. Bu alana AUC (area under curve) denir. Bu alan 0 ile 1 arasında değerler almaktadır. Bu değer 1'e yaklaştıkça sınıflandırmanın başarısı artar. Geçerli bir sınıflandırma 0.5'ten yüksek değerler almalıdır (Kılıç, 2013).

2.6 İlgili Araştırmalar

Al-Radaideh, Ananbeh, ve Al-Shawakfa (2001), yaptıkları çalışmada öğrenciler için en uygun çalışma yöntemini tahmin etmek için bir karar ağacı modeli oluşturmuştur. 6 farklı okuldan 248 öğrencinin katıldığı bu çalışmada karar ağacının doğruluk oranı %87 civarında olmuştur.

Minaei-Bigdoli ve diğ. (2003) uzaktan eğitim gören öğrencilere ait verileri kullanarak akademik başarılarını tahmin etmeye yönelik bir çalışma yapmışlardır. Bu çalışmada 10 adet bağımsız değişken kullanılmış, çalışmaya 227 kişi katılım sağlamıştır. Sene sonuna ait not ortalamaları üç farklı şekilde derecelendirilmiştir. Veri indirgeme işleminde temel bileşenler analizinden faydalanılmıştır. Bu çalışmada, karar ağaçları, K-en yakın komşu, yapay sinir ağları ve genetik algoritmalar MATLAB programında kullanılmıştır. Çalışmada 10 kat çapraz doğrulama metodu kullanılmış, modelin doğruluğunun değerlendirme ölçütü olarak da ROC eğrisinden faydalanılmıştır. Genetik algoritmalar bu ölçüte göre en yüksek doğruluk derecesine sahip algoritma olarak belirlenmiştir.

İbrahim ve Rusli (2007) tarafından yapılan çalışmada, öğrenci başarısının tahmin etmek için karar ağaçları, yapay sinir ağları ve doğrusal regresyon yöntemleri karşılaştırılmıştır. Öğrencilerin akademik başarı durumları bağımlı değişken; teknolojiyi kullanabilme yeterliliği, öğrencinin bir önceki okulunun türü, şu an kayıtlı bulunduğu okul ve aile geliri ise bağımsız değişkenler olarak belirlenmiştir. Bu karşılaştırmada en iyi sonucu yapay sinir ağlarının verdiği gözlenmiştir.

Cortez ve Silva (2008), ortaokul öğrencilerinin akademik başarısını tahmin etmeye yönelik yaptıkları çalışmada öğrencilerin demografik, sosyo-ekonomik, okul ve geçmiş dönemlere ait not bilgilerini kullanmışlardır. 32 bağımsız değişkenin kullanıldığı bu çalışmada, hedef nitelik alanları olarak Matematik ve Portekizce derslerinin yıl sonu notları belirlenmiştir. Portekizce dersi için 649, Matematik dersi için 395 örnek veri setlerini oluşturmuştur. Model oluşturmak için karar ağaçları, destek vektör makineleri yapay sinir ağları ve rasgele orman algoritmaları R yazılımı üzerinde kullanılmıştır. Çapraz doğrulamadan yararlanılan çalışmada, modeli değerlendirmek için doğruluk ölçütü ve karmaşıklık matrisi kullanılmıştır. Bu ölçütlere göre Matematik dersinde Naive Bayes algoritması, Portekizce için ise Karar Ağaçları algoritması en yüksek doğruluk derecesine sahip algoritmalar olarak belirlenmiştir. Anne ve babanın eğitim durumu, meslekleri ve alkol tüketimleri ile öğrencilerin geçmiş döneme ait notlarının öğrencinin başarısına en fazla etki eden faktörler olduğu görülmüştür.

Ramaswami ve Bhaskaran (2009), 1969 katılımcının olduğu çalışmada 32 bağımsız değişken kullanmıştır. Demografik, sosyo-ekonomik ve akademik verilerden yola çıkılarak lise seviyesindeki başarı durumunun ikili sınıflandırma ile tahmin edilmesi hedeflenen bu çalışmada 5 farklı özellik seçim yöntemi kullanılarak bu yöntemlerin her birinden elde edilen modellerin sonuçları karşılaştırılmıştır. Doğruluk, F-Skoru ve ROC eğrisi çalışmada doğruluk ölçütleri olarak kullanılmıştır. Modelleme yöntemi olarak Naive Bayes algoritmasının kullanıldığı çalışmada OneR yöntemi en iyi sonucu veren özellik seçim yöntemi olmuştur. Çalışmada öğrenciye ait ortaokul not ortalaması, babanın mesleği, okulun konumu, öğrencinin okula ulaşım şekli, öğrenim dili ve cinsiyet özellikleri belirleyici değişkenler olmuştur.

Baradwaj ve Pal (2011), öğrencilerin haftalık notları, dönem notları, laboratuvar ödevleri, devam-devamsızlık, final ödevleri gibi çeşitli ölçütler ile 50 öğrencinin performanslarını 4 yıllık süreç içerisinde incelemişlerdir. En kolay öğrenme algoritmalarından birisi olan ID3 algoritmasını kullanarak, öğrencinin başarısız olduğu alanlara odaklanarak bu durumu azaltmayı hedeflemişlerdir.

Kabakchieva, Stefanova ve Kisimov (2011), OneR Rule Learner, Decision Tree, Neural Network ve K-Nearest Neighbour veri madenciliği algoritmalarını kullanarak sınıflandırma

modeli oluşturduğu bu araştırmada en doğru sonucu, sinir ağları ve ardından Karar Ağaçları ve K-En Yakın Komşu algoritmaları ile oluşturulan model vermiştir.

Kabra ve Bichkar (2012), 346 mühendislik öğrencisinin performansını karar ağaçları yaklaşımı ile analiz etti. Bu çalışmada karar ağaçları ile başarısız olma ihtimalleri olan öğrencileri önceden belirlediler. Veri setleri iki dersin sınav sonucu ile cinsiyet, olanaklar gibi kişisel bilgilerden oluşmaktaydı. WEKA programının yardımıyla öğrencilerin başarı durumunu tahmin eden bir model oluşturdular. Bu modelin doğruluk oranı %69,94 olmuştur.

Abeer ve Elaraby (2014), bir çalışma grubu ile 6 yıllık bir periyotta bireylerin eğitim kurumlarından elde ettikleri akademik performansı sınıflandırmak ve tahminde bulunmak için bir çalışma yaptılar. Bu çalışma sonucunda, öğrencilerin ilerideki bazı notlarını tahmin ederek, bazı zayıf alanlarda iyileştirme sağladılar.

Sin ve Muthu'ya (2015) ait çalışmada eğitim alanında büyük verinin uygulanması açıklanmıştır. Çalışmada, bir sistem performansının tahmini, risk tespiti, öğrencilerin başarı tahmini, öğrencilerin sınıflandırılması gibi konular büyük veri ve öğrenme analitiği teknikleri ile saptanmıştır. Öğrencilerin performans davranışı ve beceri tahminine odaklanılan bu çalışmada, tahmine dayalı analiz ile işlevsellik vurgulanmıştır.

Abu Saa (2016), öğrencinin performansının sadece öğreticiye bağlı olmadığı aynı zamanda kişisel, sosyal veya müfredat dışı sebeplere de bağlı olduğu sonucuna ulaşılmıştır. Çalışma için yapılan ankette toplanan verilere çeşitli veri madenciliği yöntemleri uygulanarak tahmin için sınıflandırma modelleri oluşturmuştur.

Ağaoğlu (2016), öğretim görevlilerinin performansını tahmin etmeye yönelik yaptığı çalışma için yapay sinir ağlarını, sınıflandırma algoritmalarını, karar ağaçlarını, destek vektör makinelerini ve ayırt edici analiz algoritmalarını kullanmıştır. Bu çalışmada en iyi tahmin yapan algoritmanın C5.0 olduğu belirlenmiştir.

Daud ve Abbas (2017) tarafından yapılan çalışmada dört farklı parametre üzerinden inceleme yapılmıştır. Bu parametreler şu şekildedir; öğrencinin kişisel bilgileri, aile harcama düzeyleri, aile geliri ve ailenin toplam varlığı. Yapılan çalışmada aile harcamaları ve kişisel bilgi özelliklerinin içgüdüsel nedenlerle öğrenci performansı üzerinde önemli bir etkiye sahip olduğu sonucuna ulaşılmıştır.

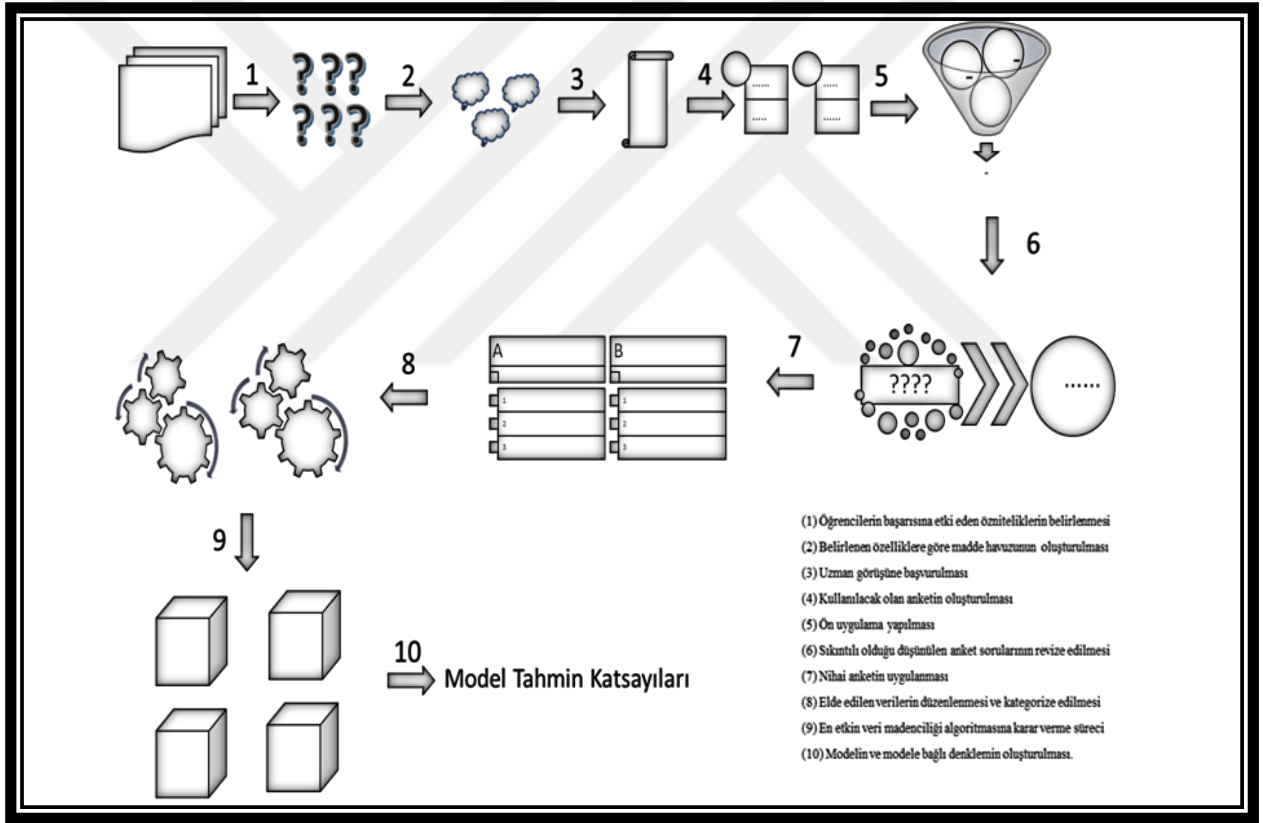
Kumar ve Singh (2017) tarafından yapılan, 412 öğrencinin katıldığı çalışmada öğrenci performansını tahmin etmek için Naive Bayes, Rastgele Orman, CART ve Bayes ağını kullandılar. Bu çalışmada en iyi yöntem Rastgele Orman olarak belirlenmiştir.

BÖLÜM III YÖNTEM

Bu bölümde; araştırma modeli, çalışma grubu, veri toplama aracı ve veri analizine ilişkin bilgilere yer verilmiştir.

3.1 Araştırma Modeli

Bu araştırma farklı makine öğrenmesi sınıflandırma yöntemlerinin sınıflandırma performanslarını karşılaştırmayı ve en iyi sınıflandırıcı yöntemi bularak bir model etmeyi amaçladığından temel bir araştırmadır. Temel araştırmaların amacı var olan bilgiye yenilerini eklemektir. Şekil 3.1’de araştırmaya ait görsel akış şeması sunulmuştur.



Şekil 3.1 Araştırmaya Ait Görsel Akış Şeması

3.2 Çalışma Grubu

Araştırmanın çalışma grubunu 2020-2021 eğitim öğretim yılında Antalya Merkez İlçelerindeki nitelikli ve niteliksiz liselere giden 613 öğrenci oluşturmaktadır. Çalışma grubunu oluşturan öğrencilerin okullara ve ilçelere göre dağılımı Tablo 3.1’de sunulmuştur.

Tablo 3.1 Çalışma Grubunu Oluşturan Öğrencilerin İlçelere ve Okul Türlerine Göre Dağılımı

İlçesi	Okul Türü	f	%
Döşemealtı	Anadolu Lisesi	74	69.8
	Anadolu Meslek Lisesi	15	14.1
	Anadolu İmam Hatip Lisesi	2	1.9
	Sosyal Bilimler Lisesi	15	14.1
	Toplam	106	100
Kepez	Güzel Sanatlar Lisesi	23	18.3
	Anadolu Lisesi	14	11.1
	Fen Lisesi	88	69.9
	Anadolu İmam Hatip Lisesi	1	0.7
	Toplam	126	100
Konyaaltı	Anadolu Lisesi	286	97.9
	Anadolu Meslek Lisesi	6	2.1
	Toplam	292	100
Muratpaşa	Anadolu Lisesi	79	87.8
	Anadolu İmam Hatip Lisesi	11	12.2
	Toplam	90	100
Toplam	Toplam	613	100

Tablo 3.1'e göre; Döşemealtı ilçesinden çalışmaya katılan 106 öğrenciden 74'ü (%69.8) Anadolu Lisesi'nde, 15'i (%14.1) Mesleki Teknik ve Anadolu Lisesi'nde, 2'si (%1.9) Anadolu İmam-Hatip Lisesi'nde, 15'i (%14.1) Sosyal Bilimler Lisesi'nde öğrenim görmektedir. Kepez ilçesinden çalışmaya katılan 126 öğrenciden 23'ü (%18,3) Güzel Sanatlar Lisesi'nde, 14'ü (%11.1) Anadolu Lisesi'nde, 88'i (%69.9) Fen Lisesi'nde, 1 (%0.7) kişi de Anadolu-İmam Hatip Lisesi'nde öğrenim görmektedir. Konyaaltı ilçesinden çalışmaya katılan 292 öğrenciden 286'sı (%97.9) Anadolu Lisesi'nde, 6'sı(%2.1) ise Mesleki Teknik ve Anadolu Lisesi'nde öğrenim görmektedir. Muratpaşa ilçesinden çalışmaya katılan 90 öğrenciden 79'u (%87.8) Anadolu Lisesi'nde, 11'i ise (%12.2) Anadolu İmam-Hatip Lisesi'nde öğrenim görmektedir.

Tablo 3.2 Çalışma Grubundaki Öğrencilerin Sahip Oldukları Olanaklara İlişkin Dağılımlar

Değişken	Düzy	F	%
Çalışma Odası	Var	520	84.8
	Yok	93	15.2
	Toplam	613	100
Çalışma Masası	Var	552	90
	Yok	61	10
	Toplam	613	100
Bilgisayar	Var	489	79.7
	Yok	124	20.3
	Toplam	613	100
İnternet	Var	595	97
	Yok	18	3
	Toplam	613	100
Kitaplık	Var	562	91.6
	Yok	51	8.4
	Toplam	613	100

Tablo 3.2'ye göre çalışma grubunda yer alan 613 öğrenciden 520'si (%84.8) çalışma odasına sahipken 93'ü (%15.2) sahip değildir. Çalışma grubunda yer alan 613 öğrencinin 552'si

(%90) çalışma masasına sahipken, 61'i (%10) sahip değildir. Çalışma grubunda yer alan 613 öğrencinin 489'ü (%79.7) bilgisayara sahipken, 124'ü (%20.3) sahip değildir. Çalışma grubunda yer alan 613 öğrencinin 595'i (%97) internet bağlantısına sahipken, 18'i (%3) sahip değildir. Çalışma grubunda yer alan 613 öğrencinin 562'si (%91.6) kitaplık sahibi iken, 51'i (%8.4) kitaplık sahibi değildir.

Tablo 3.3 *Çalışma Grubundaki Öğrencilerin Aile Özelliklerine İlişkin Dağılımlar*

Değişken	Düzy	F	%
Anne Eğitim Durumu	Okuma-yazma bilmiyor	5	0.81
	Okuma-yazma biliyor	17	2.87
	İlkokul	138	22.81
	Ortaokul	70	11.72
	Lise	203	33.72
	Lisans	140	22.83
	Yüksek Lisans	27	4.76
	Doktora	3	0.48
	Toplam	613	100.00
Baba Eğitim Durumu	Okuma-yazma bilmiyor	1	0.2
	Okuma-yazma biliyor	6	1.2
	İlkokul	84	13.8
	Ortaokul	96	15.7
	Lise	213	33.7
	Lisans	173	28.8
	Yüksek Lisans	28	4.6
	Doktora	10	2
	Toplam	613	100.00
Kardeş Sayısı	0	71	11.8
	1	254	41.8
	2	178	27.9
	3	79	12.9
	4	23	3.9
	5 ve üstü	8	1.7
	Toplam	613	100.00
Ailenin Ortalama Geliri	0-3000₺	164	26.8
	3001-6000₺	251	41.0
	6001-9000₺	120	19.6
	9001-12000	48	7.8
	12001 ve üstü	30	4.8
	Toplam	613	100.00

Tablo 3.3'e göre çalışma grubunda yer alan 613 öğrenciden; 5 (%0.81) öğrencinin annesinin okuma yazma bilmediği, 17 (%2.87) öğrencinin annesinin sadece okuma-yazma bildiği, 138 (%22.81) öğrencinin annesinin ilkokul mezunu olduğu, 70(%11.72) öğrencinin annesinin ortaokul mezunu olduğu, 203 (%33.72) öğrencinin annesinin lise mezunu olduğu, 140 (% 22.83) öğrencinin annesinin lisans mezunu olduğu, 27 (% 4.76) öğrencinin annesinin yüksek lisans mezunu olduğu, 3 öğrencinin annesinin (%0.48) ise doktora mezunu olduğu görülmüştür. Araştırmaya katılan çalışma grubunda yer alan 613 öğrenciden 1 (% 0.2)

öğrencinin babasının okuma yazma bilmediği, 6'sının(%1.2) sadece okuma yazma bildiği, 84 (%13.8) öğrencinin babasının ilkokul mezunu olduğu, 96(%15.7) öğrencinin babasının ortaokul mezunu olduğu, 213(%33.7) öğrencinin babasının lise mezunu olduğu, 173(%28.8) öğrencinin babasının lisans mezunu olduğu, 28(%4.6) öğrencinin babasının yüksek lisans, 10(%2) öğrencinin babasının da doktora mezunu olduğu görülmüştür. Çalışma grubunda yer alan 613 öğrenciden 71(%11.8) öğrencinin kardeşinin olmadığı, 254(%41.8) öğrencinin tek kardeşi olduğu, 178(%27.9) öğrencinin iki kardeşi olduğu, 79 öğrencinin (%12.9) üç kardeşi olduğu, 23 öğrencinin (%3.9) dört kardeşi olduğu, 8 öğrencinin (%1.7) ise 5 ve üstü kardeşe sahip olduğu görülmektedir. Çalışma grubunda yer alan 613 öğrenciden 164(%26.8) öğrencinin ailesinin ortalama geliri 0-3000₺ arasında, 251(%41) öğrencinin ailesinin ortalama geliri 3001-6000₺ arasında, 120(%19.6) öğrencinin ailesinin ortalama geliri 6001-9000₺ arasında, 48 (%7.8) öğrencinin ailesinin ortalama geliri 9001-12000₺ arasında, 30 (%4.8) öğrencinin ailesinin ortalama geliri ise 12001₺ ve üzerinde yer almaktadır.

3.3 Veri Toplama Aracının Geliştirilmesi

Araştırmada veri toplama aracı olarak; çalışma kapsamında geliştirilen “Ortaokul ve Lise Öğrencilerinin Akademik Başarısını Etkileyen Değişkenler” anketi kullanılmıştır. Anketin geliştirilmesi sürecinde ilk olarak alan yazın incelenmiş, ortaokul ve lise öğrencilerinin akademik başarılarına etki eden değişkenlerin neler olduğu belirlenmiştir. Bu değişkenlerden, öğrencinin kendisi, olanakları, ailesinin olanakları, genel ders başarıları, ders çalışma alışkanlıkları ile ilgili olanlar üzerinde durulmuştur. Öğrenci başarısına etki eden değişkenler belirlendikten sonra, bu değişkenler üzerinden 300 maddelik bir madde havuzu oluşturulmuştur. Oluşturulan bu madde havuzu uzman görüşü ve ön uygulamaların ardından son halini alarak 84 maddeye düşürülmüştür. Çalışma kapsamında geliştirilen anket formu Ek-1’de yer almaktadır.

3.4 Verilerin Toplanması

Çalışma grubunda yer alan öğrencilere geliştirilen anketin uygulanması için Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü ve Antalya İl Milli Eğitim Müdürlüğü arasında izin yazışmaları yapılmış, gereken izin alındıktan sonra yüz yüze eğitim yapılamadığından veri Google Forms aracılığıyla çevrimiçi olarak toplanmaya başlanmıştır. Alınan izin belgesi Ek2’de sunulmuştur. Çevrimiçi olarak toplanan bu form 3 bölümden oluşmuştur. İlk bölümde anket hakkında bilgi verilerek, katılımcılardan elde edilen verilerin gizliliğine dair açıklamalar yapılmıştır. Ankete katılımın zorunlu olmadığı belirtilmiştir. İkinci bölümde katılımcıların demografik özelliklerine dair sorular yer almış, üçüncü bölümde ise katılımcıların hayata, okula ve sosyal çevrelerine dair bakış açılarını ölçebilecek nitelikte sayısal değerlerin bulunduğu

sorular yer almıştır. Bu soruların cevapları “0-10 arası” değerler almıştır. Daha sonra cevaplar üzerinden bu sayısal değerler; “Kötü”, “Orta”, “İyi” ve “Çok iyi” şeklinde kategorize edilmiştir. Verilerin toplanma süreci 3 ay sürmüştür. Ankete toplam 613 katılım olmuştur.

3.5 Verilerin Analizi

Çalışma grubundaki katılımcılardan elde edilen verilerle katılımcılar bulundukları okullara göre “başarılı” ve “başarısız” olarak kategorize edilmiştir. Sınıflandırma yapılırken katılımcının MEB tarafından belirlenen herhangi bir proje okulu ya da nitelikli okul tanımına uyan bir okulda öğrenim görüp görmediği ölçüt olarak belirlenmiştir. Ankette cinsiyet, ailenin gelir durumu, derslere karşı ilgi, tutum ve becerileri, günlük hayatında olan planlamalar, katılımcının okul kültürüne olan aidiyet durumu ile ilgili sorular bağımsız değişken; katılımcının öğrenim gördüğü okula göre başarılı ya da başarısız olması durumu ise bağımlı değişken olarak kabul edilmiştir. Verilerin analizi “R” programı ile yapılmıştır.

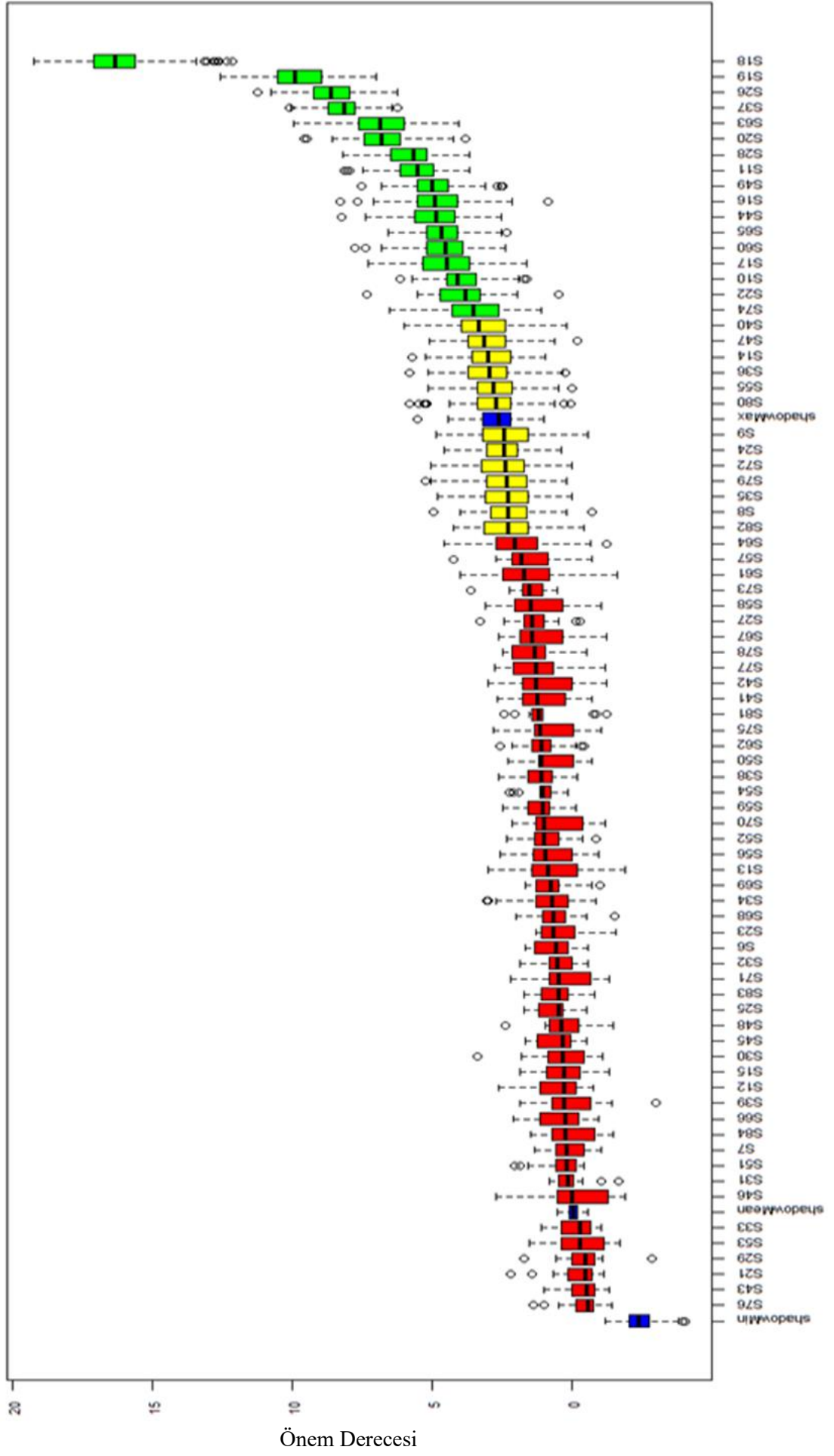
3.5.1 Veri Tiplerinin Dönüştürülmesi

“Ortaokul ve Lise Öğrencilerinin Akademik Başarısını Etkileyen Değişkenler” anketinde yer alan sorulara verilen yanıtlar veri madenciliğinde kullanılmak üzere kategorik hale getirilmiştir. Veri tipleri kategorik hale getirilirken literatürde yer alan ölçütlerden yararlanılmıştır.

3.5.2 Verilerin Seçimi

Ankete katılım sağlayanların cinsiyeti, okullarının bulunduğu ilçe ve okul türü maddeleri ankette çıkarılarak, diğer bağımsız değişkenlerin hangilerinin önemli olduğu R yazılımında yer alan “Boruta” paketi ile gerçekleştirilmiştir. Bu paket, bağımsız değişkenin önemli olup olmadığını belirlemekte, belli bir önem derecesinin üzerindeki değişkenler analizde kullanılmak üzere seçilmektedir. “Ortaokul ve Lise Öğrencilerinin Akademik Başarısını Etkileyen Değişkenler” anketinde yer alan soruların önem derecesi aşağıda belirtilen tablodaki şekliyle oluşmuştur.

Değişkenlerin Önem Dereceleri



Şekil 3.2 Bağımsız Değişkenlerin Seçimi

Şekil 3.2’den elde edilen bilgiler doğrultusunda analiz için kullanılacak olan maddeler Tablo 3.4’te verilmiştir.

Tablo 3.4 *Analiz İçin Kullanılacak Olan Anket Soruları*

Özellik adı	Özellik	Veri Tipi
Madde-1 En sevdiğiniz ders	Türkçe, Edebiyat, Sosyal Bilgiler, Tarih, Coğrafya, Felsefe, İnkılap Tarihi ve Atatürkçülük, Din Kültürü ve Ahlak Bilgisi, Matematik, Geometri, Fizik, Kimya, Biyoloji, Fen Bilimleri, Beden Eğitimi ve Spor, Müzik, Resim, Diğer	Kategorik
Madde-2 Çalışma odası	Var, Yok	Kategorik
Madde-3 Çalışma masası	Var, Yok	Kategorik
Madde-4 Bilgisayar	Var, Yok	Kategorik
Madde-5 Kardeş sayısı	Yok, 1-2 tane, 3-4 tane, 5 ve üstü	Kategorik
Madde-6 Dünya klasikleri okuma	Evet, Hayır, Bazen	Kategorik
Madde-7 Ailenizin aylık ortalama geliri	0-3000 TL, 3001-6000TL, 6001-9000TL, 9001-12000TL, 12001 TL ve üzeri	Kategorik
Madde-8 Kitap sayısı	0-15, 16-30, 31-100, 101-200, 201 ve üzeri	Kategorik
Madde-9 Anne Öğrenim Durumu	Okuma-yazma bilmiyor, Okuma-yazma biliyor, İlkokul , Ortaokul, Lise ,Lisans,Yüksek lisans Doktora	Kategorik
Madde-10 Baba Öğrenim Durumu	Okuma-yazma bilmiyor, Okuma-yazma biliyor, İlkokul , Ortaokul, Lise ,Lisans,Yüksek lisans Doktora	Kategorik
Madde-11 Anne meslek grubu	Sağlık, Eğitim, Gıda, Turizm, Tarım, Ticaret, Ulaşım, Savunma ve güvenlik, Meslek sahibi değil, Diğer	Kategorik
Madde-12 Baba meslek grubu	Sağlık, Eğitim, Gıda, Turizm, Tarım, Ticaret, Ulaşım, Savunma ve güvenlik, Meslek sahibi değil, Diğer	Kategorik
Madde-13 Güzel sanatlara yönelik kaynak olma durumu	Hayır, Evet	Kategorik
Madde-14 Spora ayrılan zaman	1 saatten az, 1-2 saat arası, 2 saatten fazla	Kategorik

Tablo 3.4 (devamı)

Madde-15 Arkadaşlarla geçirilen zaman	0-1 saat, 2-3 saat, 4-5 saat, 6-7 saat, 8 saat ve üzeri	Kategorik
Madde- 16 Televizyona ayrılan zaman	0-2saat, 2-3 saat, 4-5 saat, 6-7 saat, 8 saat ve üzeri	Kategorik
Madde-17 Kaynak kitap durumu	Var, Yok	Kategorik
Madde-18 Doğayla zaman geçirme, yolculuk yapma yeni yerler keşfetme durumu	Hayır, Evet	Kategorik
Madde-19 Konuşulan diller	Fransızca, Almanca, İngilizce, İtalyanca, İspanyolca, Rusça, Diğer	Kategorik
Madde- 20 Evdeki bilgisayar sayısı	Yok, 1-3 arası, 4-5 arası	Kategorik
Madde-21 Evdeki müzik aleti sayısı	Yok, 1-3 arası, 4-5 arası	Kategorik
Madde-22 Kendine güven	Evet, Hayır, Bazen	Kategorik
Madde-23 “Başarı duygusunu tatmak için okula gidiyorum” cümlesine katılma durumu	Kötü, Orta, İyi, Çok İyi	Kategorik
Madde-24 Matematik genel başarı durumu	Kötü, Orta, İyi, Çok İyi	Kategorik
Madde-25 Sosyal Bilimler genel başarı durumu	Kötü, Orta, İyi, Çok İyi	Kategorik
Madde-26 Pandeminin akademik başarıınızı etkileme durumu	Kötü, Orta, İyi, Çok İyi	Kategorik
Madde-27 Ailenizin okul etkinliklerine ilgi durumu	Kötü, Orta, İyi, Çok İyi	Kategorik
Madde-28 Okulda kendini yalnız hissetme durumu	Kötü, Orta, İyi, Çok İyi	Kategorik
Madde-29 Okulda arkadaş edinme becerisi	Kötü, Orta, İyi, Çok İyi	Kategorik
Madde-30 Okuldaki öğretmenlerin ders anlatımı ve hakimiyeti	Kötü, Orta, İyi, Çok İyi	Kategorik

BÖLÜM IV

BULGULAR

Bağımsız değişkenlerin belirlenmesi bağımlı değişkenin de iki kategoriye ayrılması ile R programı üzerinden makine öğrenme algoritmaları ile analiz yapılarak uygulanan algoritmaların başarı durumları rapor halinde bölüm sunulmuştur. Tezin problem ve alt problemlerinde belirlenen sorulara çözüm aranmıştır.

4.1 Katılımcıların Başarı Durumlarına Göre Sınıflandırılması

Bağımlı değişkende katılımcıların durumu öğrenim gördükleri okulların nitelik durumuna göre belirlenmiştir. Buna göre; çalışmaya katılan öğrencilerden proje okulu durumunda bulunan Anadolu Liseleri, belirli bir puan ile girilen liseler olan Fen Liseleri ve Sosyal Bilimler Liseleri, özel yetenek gerektiren Güzel Sanatlar Lisesi nitelikli okullar olarak kabul edilmiştir. Burada öğrenim gören öğrenciler başarılı, diğer okullarda öğrenim gören öğrenciler ise başarısız olarak belirlenmiştir. Bu sınıflandırma sonucu elde edilen tablo aşağıda gösterilmiştir.

Tablo 4.1 Öğrencilerin Başarı Sınıflaması Sonucu Oluşan Gruplar

Grup Adı	Öğrenci Sayısı
Başarılı	363
Başarısız	250
Toplam	613

Tablo 4'e göre yapılan sınıflandırma sonucu başarısız öğrenci sayısı 363, başarılı öğrenci sayısı ise 250 olarak belirlenmiştir. Verilerin analizi, veri madenciliğinin denetimli öğrenme uygulamalarından karar ağaçları (C5.0 algoritması, CART algoritması, Rastgele Ormanlar) ve denetimsiz öğrenme uygulamalarından Destek Vektör Makinesi algoritmaları kullanılarak gerçekleştirilmiştir.

4.2 C5.0 Algoritması Uygulaması ve Elde Edilen Bulgular

C5.0 algoritması ile yapılacak olan uygulama için analize dahil edilecek olan değişkenler Tablo 3.5'te sunulmuştur.

Öncelikle kodlar yardımı ile rastgele olarak tüm çalışma grubunun %80'ine denk gelen 491 rastgele gözlem eğitim kümesi, geriye kalan 122 gözlem ise test kümesi olarak seçilmiştir. Ardından yine komutlar yardımıyla R paketine C5.0 algoritması paketi yüklenmiş, programın kütüphanesinden işlemleri yapmak üzere çağrılarak, sınıflandırma gerçekleştirilmiştir. Oluşturulan eğitim kümesine ait hata oranı %18.9 olarak gerçekleşmiş olup, doğru tahmin değeri ise %81.1 olarak gerçekleşmiştir. Doğru ve yanlış sınıflamaya ait karışıklık matrisi tablosu Tablo 4.2 'de verilmiştir.

Tablo 4.2 C5.0 Algoritmasının Eğitim Veri Kümesine Ait Karışıklık Matrisi Tablosu

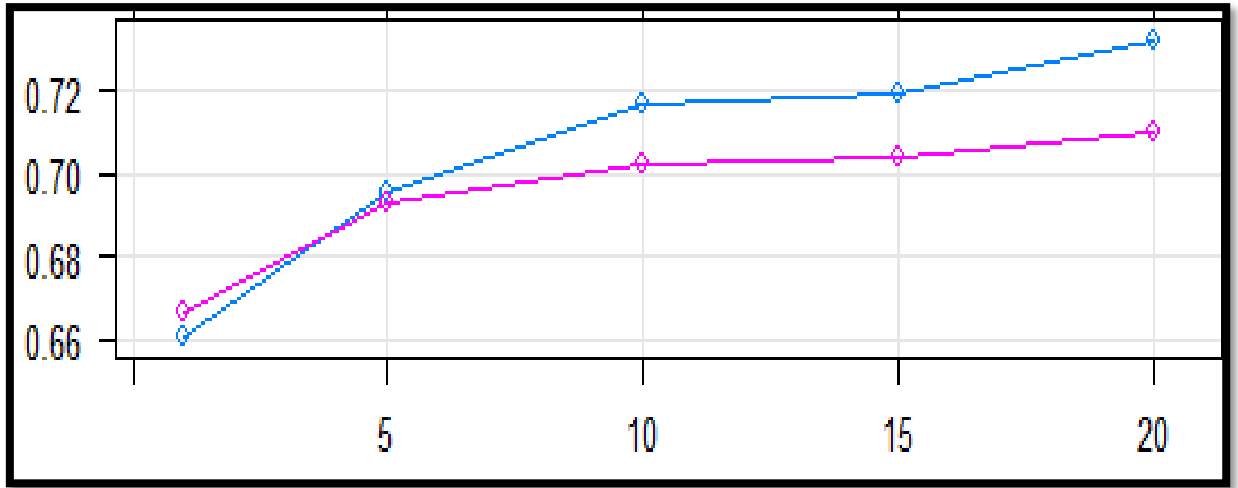
	Başarısız	Başarılı
Başarısız	265	28
Başarılı	65	133

Bu işlemin ardından, test veri kümesinin ne kadar iyi kestirim yapacağını bulmak için yazılım üzerinden daha önce indirilen kestirim kodları kütüphaneden tekrar çağrılarak test kestirimi gerçekleştirilmiştir. Bu kestirim sonucunda ortaya çıkan karışıklık matrisi tablosu Tablo 4.3'te verilmiştir.

Tablo 4.3 C5.0 Algoritması Test Veri Setine Ait Kestirim Karışıklık Matrisi Tablosu

	Başarısız	Başarılı
Başarısız	60	10
Başarılı	32	20

Tablo 4.4'te verilen bilgilere göre kurulan modelin test seti için tahmin başarısı % 68.47 olarak gerçekleşmiştir. Ardından eğitim ve test setindeki hata oranlarını daha iyi tahmin etmek için çapraz doğrulama yapılmıştır. Çapraz doğrulama sonucunda doğruluk oranı %73 olarak belirlenmiştir. Sonuçlar Şekil 4.2.1'de verilmiştir.



Şekil 4.2.1 C5.0 Algoritmasına Ait Çapraz Doğrulama Sonuçları

Modelin doğruluğunu değerlendirmek için kullanılan tüm ölçütlerin yer aldığı Tablo 4.4'te değerlendirme ölçütleri karşılaştırılmalı olarak verilmiştir.

Tablo 4.4 C5.0 Algoritmasına Ait Model Doğruluğu Ölçütleri

Değerlendirme Ölçütü	Eğitim Veri Kümesi	Test Veri Kümesi	Çapraz Doğrulama
Doğruluk (%)	81.1%	68.47%	73%
Doğru sınıflandırılan örnek sayısı	398	80	
Yanlış sınıflandırılan örnek sayısı	83	42	
TP oranı	0.810		
FP oranı	0.286		
Precision	0.739		
Recall	0.6717		
F-Skoru	0.703		
Kappa Değeri	0.569		

4.3 CART Algoritması Uygulaması ve Elde Edilen Bulgular

Veri kümesi kodlar yardımı ile %80 eğitim kümesi %20 test kümesi olarak ayrılmıştır. 491 gözlemden oluşan eğitim veri kümesi öncelikle hesaplanmıştır. Elde edilen sonuçlar Tablo 4.5' te gösterilmektedir.

Tablo 4.5 *CART Algoritması Eğitim Veri Kümesine Ait Karışıklık Matrisi Tablosu*

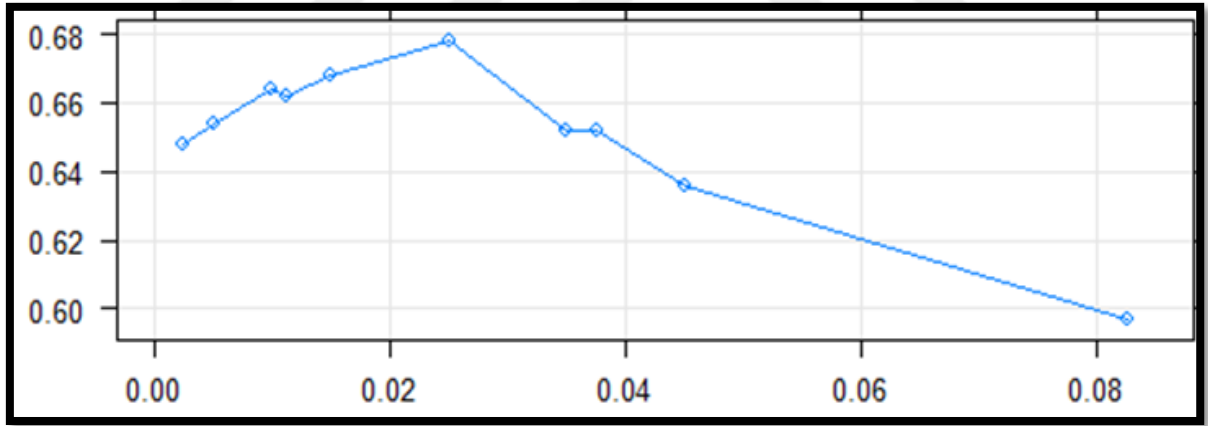
	Başarısız	Başarılı
Başarısız	241	40
Başarılı	50	160

Tablo 4.6' ya bakıldığında doğru kestirilen 401 gözlem varken, 90 gözlemin durumu yanlış kestirilmiştir. Oransal olarak eğitim veri kümesine ait doğruluk oranı %81.67 olmuştur. Ardından aynı işlemler bu kez test kümesi için gerçekleştirilmiştir. 118 gözlem için yapılan bu kestirime ait bilgiler Tablo 4.6'da gösterilmiştir.

Tablo 4.6 *CART Algoritması Test Veri Setine Ait Karışıklık Matrisi Tablosu*

	Başarısız	Başarılı
Başarısız	40	19
Başarılı	22	37

Tablo 4.6'ya göre test veri kümesinin doğruluk oranı %65.25 olarak gözlemlenmiştir. Bu sonuçların ardından eğitim ve test kümesinde yer alan hataları daha iyi tespit edebilmek için çapraz doğrulama yapılmıştır. Çapraz doğrulama sonucunu gösteren Şekil 4.3.1'e göre çapraz doğrulama ile doğruluk oranı %59.5 olmuştur.

**Şekil 4.3.1** *CART Algoritmasına Ait Çapraz Doğrulama Sonuçları*

Modelin doğruluğunu değerlendirmek için kullanılan tüm ölçütlerin yer aldığı Tablo 4.7'de değerlendirme ölçütleri karşılaştırılmalı olarak verilmiştir.

Tablo 4.7 *CART Algoritmasına Ait Model Doğruluğu Ölçütleri*

Değerlendirme Ölçütü	Eğitim Veri Kümesi	Test Veri Kümesi	Çapraz Doğrulama
Doğruluk (%)	81.67%	65,25%	59,5%
Doğru sınıflandırılan örnek sayısı	401	77	
Yanlış sınıflandırılan örnek sayısı	90	41	
TP oranı	0.816		
FP oranı	0.184		
Precision	0.76		
Recall	0.80		
F-Skoru	0.78		
Kappa Değeri	0.623		

4.4 Destek Vektör Makinesi Algoritması Uygulaması ve Uygulama Sonuçları

Veri kümesi kodlar yardımıyla %80'lik kısmı eğitim kümesi, %20'lik kısmı test kümesi olmak üzere ikiye ayrılmıştır. 491 gözlem eğitim kümesi, 122 gözlem ise test kümesi olmuştur. Öncelikle eğitim kümesine ait karışıklık tablosu hesaplanmıştır. Elde edilen veriler, Tablo 4.8'de gösterilmiştir.

Tablo 4.8 *DVM Algoritması Eğitim Veri Kümesine Ait Karışıklık Matrisi Tablosu*

	Başarısız	Başarılı
Başarısız	223	64
Başarılı	68	136

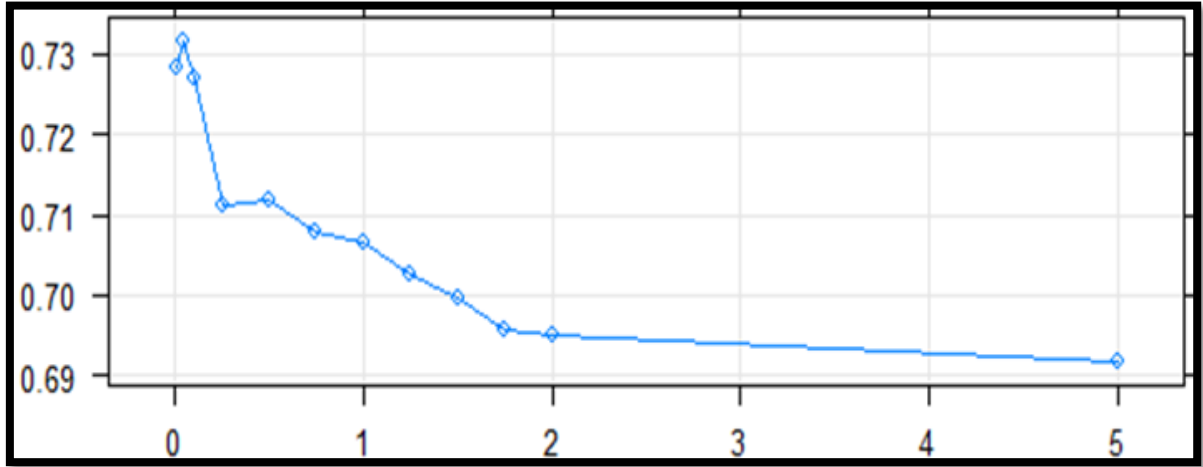
Tablo 4.8 incelendiğinde 491 öğrenciden kestirimi doğru yapılan gözlem sayısı 359 iken, yanlış kestirimde bulunan gözlem sayısı 132 olmuştur. Oransal olarak bakıldığında eğitim veri kümesine ait doğruluk oranının %73,11 olduğu görülmektedir. Bu işlemlerin ardından aynı işlemler test veri kümesi için yapılmıştır. Bu işlemlere ait karışıklık matrisi Tablo 4.9'da gösterilmiştir.

Tablo 4.9 *DVM Algoritması Test Veri Kümesine Ait Karışıklık Matrisi Tablosu*

	Başarısız	Başarılı
Başarısız	54	20
Başarılı	18	30

Tablo 4.9 incelendiğinde test veri kümesine ait 122 gözlemden doğru kestirilen gözlem sayısı 84 yanlış kestirilen gözlem sayısı ise 32 olmuştur. Oransal olarak bakıldığında test veri kümesine ait doğruluk oranının %68,85 olduğu görülmektedir. Bu sonuçların ardından eğitim

ve test kümesinde yer alan hataları daha iyi tespit edebilmek için çapraz doğrulama yapılmıştır. Çapraz doğrulama sonucu doğruluk oranı %69 olmuştur. Sonuçlar Şekil 4.4.1’de gösterilmiştir.



Şekil 4.4.1 DVM Algoritmasına Ait Çapraz Doğrulama Sonuçları

Modelin doğruluğunu değerlendirmek için kullanılan tüm ölçütlerin yer aldığı Tablo 4.10’da değerlendirme ölçütleri karşılaştırılmalı olarak verilmiştir.

Tablo 4.10 DVM Algoritmasına Ait Model Doğruluğu Ölçütleri

Değerlendirme Ölçütü	Eğitim Veri Kümesi	Test Veri Kümesi	Çapraz Doğrulama
Doğruluk (%)	73.11%	68.8%	69.2%
Doğru sınıflandırılan örnek sayısı	359	74	
Yanlış sınıflandırılan örnek sayısı	122	38	
TP oranı	0.731		
FP oranı	0.289		
Precision	0.717		
Recall	0.833		
F-Skoru	0.754		
Kappa Değeri	0.401		

4.5 Rastgele Ormanlar Algoritması Uygulaması ve Elde Edilen Bulgular

Veri kümesi kodlar yardımıyla %80’lik kısmı eğitim kümesi, %20’lik kısmı test kümesi olmak üzere ikiye ayrılmıştır. 490 gözlem eğitim kümesi, 123 gözlem ise test veri kümesi olmuştur. Eğitim veri seti üzerinden hesaplamalar yapılmıştır. Eğitim veri kümesine ait karışıklık tablosu Tablo 4.11’de verilmiştir.

Tablo 4.11 *Rastgele Ormanlar Algoritması Eğitim Veri Kümesine Ait Karışıklık Matrisi Tablosu*

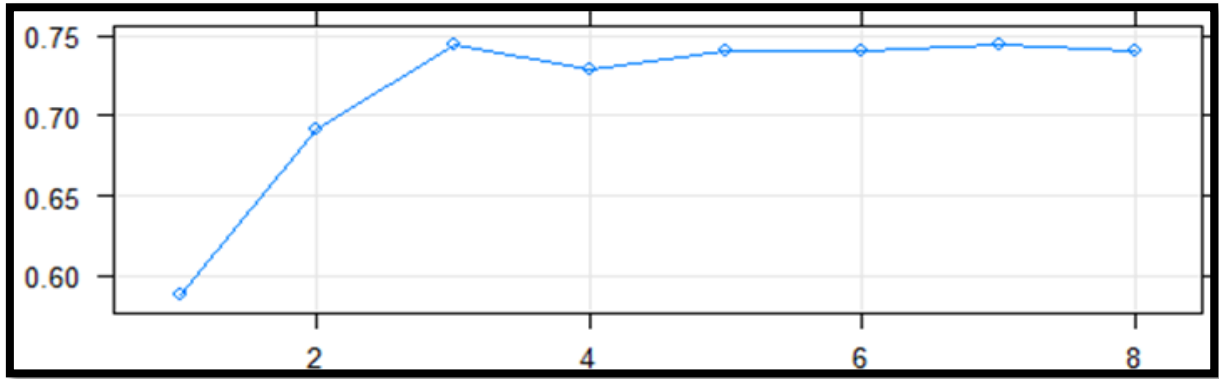
	Başarısız	Başarılı
Başarısız	232	59
Başarılı	64	135

Tablo 4.11 incelendiğinde 490 öğrenciden kestirimi doğru yapılan gözlem sayısı 367 iken, yanlış kestirimde bulunan gözlem sayısı 123 olmuştur. Oransal olarak bakıldığında eğitim veri kümesine ait doğruluk oranının %74,89 olduğu görülmektedir. Bu işlemlerin ardından aynı işlemler test veri kümesi için yapılmıştır. Test veri kümesine ait karışıklık tablosu Tablo 4.12’de verilmiştir.

Tablo 4.12 *Rastgele Ormanlar Algoritması Test Veri Kümesine Ait Karışıklık Matrisi Tablosu*

	Başarısız	Başarılı
Başarısız	58	16
Başarılı	14	35

Tablo 4.12 incelendiğinde 93 gözlemin başarılı kestirildiği gözlenmiştir. Dolayısıyla Rastgele Ormanlar algoritmasına ait doğruluk oranının %75,61 olduğu görülmektedir. Test veri kümesine ait doğruluk oranı belirlendikten sonra hataları daha iyi tespit edebilmek için çapraz doğrulama işlemi gerçekleştirilmiştir. Sonuçlar şekil 4.5.1’de gösterilmiştir.

**Şekil 4.5.1** *Rastgele Ormanlar Algoritmasına Ait Çapraz Doğrulama Sonuçları*

Şekil 4.5.1 incelendiğinde çapraz doğrulama sonrasında Rastgele Ormanlar algoritmasına ait çapraz doğrulama doğruluk oranının %74,3 olduğu gözlemlenmiştir. Modelin doğruluğunu değerlendirmek için kullanılan tüm ölçütlerin yer aldığı Tablo 4.13’te değerlendirme ölçütleri karşılaştırılmalı olarak verilmiştir.

Tablo 4.13 *Rastgele Ormanlar Algoritması Test Veri Kümesine Ait Model Doğruluğu Ölçütleri*

Değerlendirme Ölçütü	Eğitim Veri Kümesi	Test Veri Kümesi	Çapraz Doğrulama
Doğruluk (%)	74.89%	75.61%	74.3%
Doğru sınıflandırılan örnek sayısı	367	93	
Yanlış sınıflandırılan örnek sayısı	123	30	
TP oranı	0.766		
FP oranı	0.233		
Precision	0.720		
Recall	0.584		
F-Skoru	0.645		
Kappa Değeri	0.423		

Tablo 4.14'te tüm algoritmaların eğitim, test veri kümesi ve çapraz doğrulama sonucunda oluşan doğruluk oranları karşılaştırmalı olarak verilmiştir.

Tablo 4.14 *Dört Algoritmaya Ait Doğruluk Oranları Karşılaştırma Tablosu*

Algoritma Adı	Eğitim Kümesi	Test Kümesi	Çapraz Doğrulama
C5.0	81.1%	68.47%	73%
CART	81.67%	65.25%	59.5%
Destek Vektör Makinesi	73.11%	68.8%	69.2%
Rastgele Ormanlar	74.89%	75.61%	74.3%

Tablo 4.14 incelendiğinde sınıflandırmada C5.0 algoritması eğitim kümesinde %81.1, test kümesinde %68.47, çapraz doğrulama sonucunda ise %73 doğruluk oranı göstermiştir. CART algoritması eğitim kümesinde %81.67, test kümesinde %65.25, çapraz doğrulama sonucunda ise %59.5 doğruluk oranı göstermiştir. DVM algoritması eğitim kümesinde %73.11, test kümesinde %68.8, çapraz doğrulama sonucunda ise %69.2 doğruluk oranı göstermiştir. Rastgele ormanlar algoritması eğitim kümesinde %74.89, test kümesinde %75.61, çapraz doğrulama sonucunda ise %74.3 doğruluk oranı göstermiştir. C5.0, CART, DVM ve Rastgele Ormanlar algoritmaları karşılaştırıldığında hatalardan en fazla arınık olan ve en başarılı kestirim yapan algoritmanın Rastgele Ormanlar olduğu gözlemlenmiştir.

4.6 Öğrenci Başarısı İçin Önerilen Sınıflandırma Modeli

En başarılı kestirim yapan algoritmanın belirlenmesinin ardından bu algoritma üzerinden yeni bir sınıflandırma modeli oluşturulmuştur. Bu model, ankette yer alan her bir soru için bir katsayı tahmininin yer aldığı modeldir. Belirlenen bu model üzerinden, geliştirilen anket formunu cevaplayarak sisteme yeni girecek öğrencilerin cevaplarına göre öğrencinin başarılı ya da başarısız sınıflandırılacağına yönelik kestirim yapılmaktadır.

Sisteme yeni girecek öğrencinin ankete vereceği cevaplar üzerinden yapılacak analizlerle öğrencinin eksiklikleri belirlenerek, bu eksikliklere erken bir uyarı sistemi gibi eğitim süreci içerisinde müdahale edilebilecek, öğrencinin başarı durumu negatif bir durumdan pozitive dönüştürülmesinde atılacak adımların bir planlamasının yapılabilmesi mümkün olabilecektir. Yapılan hesaplamalar sonucunda rastgele orman algoritması kullanılarak elde edilecek bu katsayı tahminlerinin toplamı 0.50'nin üzerinde olan bir öğrencinin başarılı olarak sınıflandırılmaktadır. Anket sorularına verilen cevaplar üzerinden elde edilen katsayı tahminlerine ilişkin bilgiler Tablo4.15'te yer almaktadır.

Tablo 4.15 Öğrencilerin Verdiği Cevaplar Üzerinden Elde Edilen Katsayı Tahminleri

Madde adı ve numarası	Verilen Cevap ve Katsayı Tahminleri
1-En sevdiğiniz ders	1-Türkçe (0), 2-Edebiyat(-0,042), 3-SosyalBilgiler (-0,1), 4-Tarih(0,252), 5- Coğrafya(-0,263), 6-Felsefe (-0,469), 7-İnkılap Tarihi ve Atatürkçülük(-0,01), 8-Din Kültürü ve A.B. (0,144), 9-Matematik(-0,045), 10-Geometri(-0,237), 11-Fizik(-0,102), 12-Kimya(0,112), 13-Biyoloji(0,006), 14-Fen Bilimleri(-0,312), 15-Beden Eğitimi ve Spor(-0,093), 16-Müzik(0,094), 17-Görsel Sanatlar(0,09), 18-Diğer(-0,183)
2-Çalışma odası	0-Yok (0), 1-Var(0,0373)
3-Çalışma masası	0-Yok (0), 1-Var(0,0143)
4-Bilgisayar	0-Yok (0), 1-Var(0,106)
5-Kardeş sayısı	0-Yok(0), 1) 1 tane(-0,0721) ,2) 2 tane(-0,0192) ,3) 3tane(-0,0683), 4) 4 tane(-0,074), 5) 5 ve üstü (-0,23)
6-Dünya klasikleri okumaktan hoşlanma	0-Hayır(0), 1-Bazen(0,002), 2- Evet(0,059)
7-Ailenizin aylık ortalama geliri	1)0-3000 TL (0), 2) 3001-6000TL(0,008) ,3)6001-9000TL(0,0977), 4)9001-12000TL(0,1576),5) 12001TL ve üzeri(-0,22)
8-Kitap sayısı	1)0-15(0),2) 16-30(0,008), 3) 31-100(0,223), 4)101-200(0,0814), 5) 201 ve üzeri(0,1043)
9-Anne Öğrenim Durumu	1)Okuma-yazma bilmiyor(0), 2)Okuma-yazma biliyor(0,3) ,3)İlkokul(0,03) , 4)Ortaokul(0,058), 5)Lise(0,078) ,6)Lisans(0,189),7)Yüksek lisans(0,156) 8)Doktora(-0,568)
10-Baba Öğrenim Durumu	1)Okuma-yazma bilmiyor(0), 2)Okuma-yazma biliyor(-0,304), 3) İlkokul(0,234) , 4)Ortaokul(0,244), 5)Lise(0,293) ,6)Lisans(0,251),7)Yüksek lisans(0,123) 8)Doktora(0,301)
11-Anne meslek grubu	1)Sağlık(0),2) Eğitim(0,0827),3) Gıda(0,03), 4)Turizm(-0,0356), 5)Tarım(-0,065), 6)Ticaret(0,009), 7)Ulaşım(0),8) Savunma ve güvenlik(-0,124), 9)Meslek sahibi değil(0,1202),10) Diğer(0,014)
12-Baba meslek grubu	1)Sağlık(0), 2) Eğitim(-0,073),3) Gıda(-0,18), 4)Turizm(0,045), 5)Tarım(-0,401), 6)Ticaret(-0,107), 7)Ulaşım(0,025), 8) Savunma ve güvenlik(-0,018), 9)Meslek sahibi değil(-0,058),10) Diğer(-0,024)
13-Güzel sanatlara yönelik kaynak olma durumu	0)Hayır(0), 1)Evet(0,1362)
14-Spora ayrılan zaman	1)1 saatten az(0),1) 1-2 saat arası(-0,1087), 2) 2 saatten fazla(-0,0718)

Tablo 4.15 devamı

15-Arkadaşlarla geçirilen zaman	1)0-2saat(0),2) 2-3 saat(-0,0894),3) 4-5 saat(-0,0834), 4)6-7 saat(-0,1886), 8 saat ve üzeri(0,112)
16-Televizyona ayrılan zaman	1)0-2saat(0) 2),2-3 saat(-0,0717), 3) 4-5 saat(-0,012), 4)6-7 saat(-0,221), 5)8 saat ve üzeri(-0,643)
17-Kaynak kitap durumu	0)Yok(0), 1)Var(0,0885)
18-Doğayla zaman geçirme, yolculuk	0)Hayır, 1)Evet(-0,2054)
19-Konuşabilinen diller	1)Fransızca(0), 2)Almanca(-0,039), 3)İngilizce(-0,01), 4)İtalyanca(-0,472), 5)İspanyolca(-0,301),6)Rusça(-0,325), 7)Diğer(-0,164)
20-Evdeki bilgisayar sayısı	0)Yok(0), 1)1 tane(-0,035), 2)2 tane(-0,053), 3)3tane(-0,045), 4)4 tane(0,102), 5)5tane(-0,293)
21-Evdeki müzik aleti sayısı	0)Yok(0), 1)1 tane(-0,0977), 2)2 tane(-0,0378), 3)3tane(-0,0575), 4)4 tane(0,1434), 5)5tane(0,037)
22-Kendine güven	0-Hayır(-0,1187), 1)Bazen(-0,0381) 2)Evet(0)
23-“Başarı duygusunu tatmak için okula gidiyorum” cümlesine katılma durumu puan kategorisi	0-Kötü(0), 1-Kötü(-0,028), 2-Kötü(-0,005), 3-Kötü(-0,022), 4-Orta(-0,043), 5-Orta(0,008), 6-Orta(0,013), 7-İyi(-0,055), 8-İyi(0,129), 9-Çok iyi(0,158), 10-Çok İyi(0,207)
24-Matematik genel başarı durumu puan kategorisi	0-Kötü(0), 1-Kötü(-0,007), 2-Kötü(-0,013), 3-Kötü(0,0811), 4-Orta(0,0258), 5-Orta(0,0519), 6-Orta(-0,0957), 7-İyi(0,0983), 8-İyi(0,0715), 9-Çok iyi(-0,822), 10-Çok İyi(-0,0983)
25-Sosyal Bilimler genel başarı durumu puan kategorisi	0-Kötü(0), 1-Kötü(-0,196), 2-Kötü(0,022), 3-Kötü(-0,0126), 4-Orta(-0,203), 5-Orta(-0,213), 6-Orta(-0,076), 7-İyi(-0,054), 8-İyi(0,043), 9-Çok iyi(-0,14), 10-Çok İyi(-0,014)
26-Pandeminin akademik başarınızı etkileme durumu puan kategorisi	0-Kötü(0), 1-Kötü(0,232), 2-Kötü(0,164), 3-Kötü(0,131), 4-Orta(0,082), 5-Orta(0,021), 6-Orta(0,209), 7-İyi(0,274), 8-İyi(0,362), 9-Çok iyi(0,322), 10-Çok İyi(0,285)
27-Ailenizin okul etkinliklerine ilgi durumu puan kategorisi	0-Kötü(0), 1-Kötü(-0,056), 2-Kötü(0,189), 3-Kötü(-0,032), 4-Orta(-0,058), 5-Orta(0,22), 6-Orta(0,076), 7-İyi(-0,067), 8-İyi(-0,011), 9-Çok iyi(0,103), 10-Çok İyi(-0,034)
28-Okulda kendini yalnız hissetme durumu puan kategorisi	0-Kötü(0), 1-Kötü(-0,042), 2-Kötü(0,001), 3-Kötü(0,042), 4-Orta(0,039), 5-Orta(0,068), 6-Orta(0,063), 7-İyi(0,042), 8-İyi(-0,017), 9-Çok iyi(-0,024), 10-Çok İyi(-0,047)
29-Okulda arkadaş edinme becerisi puan kategorisi	0-Kötü(0), 1-Kötü(-0,0388), 2-Kötü(-0,0634), 3-Kötü(0,0306), 4-Orta(-0,0372), 5-Orta(-0,1264), 6-Orta(-0,1261), 7-İyi(-0,1158), 8-İyi(-0,0414), 9-Çok iyi(-0,1839), 10-Çok İyi(-0,0975)
30-Okuldaki öğretmenlerin ders anlatımı ve hakimiyeti puan kategorisi	0-Kötü(0), 1-Kötü(0,004), 2-Kötü(-0,107), 3-Kötü(0,195), 4-Orta(-0,03), 5-Orta(-0,078), 6-Orta(-0,032), 7-İyi(-0,01), 8-İyi(0,002), 9-Çok iyi(-0,014), 10-Çok İyi(-0,1291)

Tablo 4.15’te yer alan tahmin katsayılarına göre öğrencilerin seçtiği cevaplar üzerinden elde edilen tahmin katsayıları sonucunda öğrenci başarısını kestirmede önerilen model (4.1) eşitliğinde sunulmuştur.

$$Y = 0,076 + \sum_{i=1}^{30}[i. maddeye ait katsayı tahmini] \quad (4.1)$$

(4.1) eşitliğinde belirtilen “Y” öğrenciye ait toplam puanı göstermektedir. Öğrencinin ankete verdiği cevaplar üzerinden elde edilen toplam puan ile 0,076 sabit sayısının toplamı ≥ 0.50 olduğunda öğrenci başarılı olarak, < 0.50 olduğunda ise öğrenci başarısız olarak sınıflandırılmaktadır.

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1 Sonuç ve Tartışma

Bu araştırmada, öğrencilerin, başarısına etki eden değişkenlerin tespit edilerek bu değişkenler üzerinden makine öğrenme algoritmalarıyla öğrenci başarısının tahmin edilmesi amaçlanmıştır. Bu bölümde araştırmanın amacı bağlamında araştırmanın alt problemlerine dair sonuçlar tartışmalı olarak incelenecek ve önerilere yer verilecektir.

Araştırmaya ait ilk alt problem öğrencilerin akademik başarısını en çok etkileyen değişkenlerin belirlenmesine yöneliktir. Bu sorunun cevabı aranırken literatür detaylı bir şekilde taranmıştır. Yapılan araştırmaların genelde öğrencilerin okul notları üzerinden tahmini üzerine olduğu görülmüştür. Chang ve Kim'e (2021) ait araştırmada Kore'de öğrenim gören 26.345 ortaokul öğrencisinden çevrimiçi olarak veri toplanmış öğrencilerin geçmiş yıl notları, derse katılımları ve ders öğretmenleriyle ilgili sorular yöneltilmiştir. Uzaktan eğitim sisteminin ve öğrencilerin ders başarısının tahminine yönelik olan yapılan çalışmada lojistik regresyon analizi kullanılmıştır. Ancak öğrencilerin okul başarısının, sınav başarısını tek başına açıklayacak bir değişken olmadığı da bir gerçektir. Bu yüzden araştırmanın ilk alt sorusunu cevaplayabilmek için öğrenci başarısına etki eden değişkenlerin neler olduğunu inceleyen birçok kaynak taranarak 300 sorudan oluşan bir madde havuzu elde edilmiştir. Ardından ön uygulamalar yapılarak ve uzman görüşü alınarak madde sayısı 84'e indirilmiştir. Öğrencinin ve ailesinin sahip olduğu demografik yapı, sosyo-ekonomik durum, sahip olduğu olanaklar, derslere karşı geliştirdiği tutumlar, sorun çözme becerisi, anne ve babalarının eğitim durumu, anne ve babasının meslekleri, pandemiden etkilenme durumları, arkadaşlarıyla geçirdikleri zaman, TV izlemeye ayırdığı zaman, oynadığı bilgisayar oyunu çeşitleri, güzel sanatlara olan ilgi durumları, yeni yerler keşfetme isteği, okul iklimi, en sevdiği dersler, planlı ve sistemli çalışma becerisi, ailesinin gelir durumu, spora ayırdığı zaman, uyku ve beslenme düzeni, ailesiyle geçirdiği zaman dilimleri, anne ve babasının okul etkinliklerine katılma durumu, sosyalleşme durumu, sosyal medyaya ayırdığı zaman, sınavlar karşısında kaynağa ve desteğe ihtiyaç duyup duymadığı, el becerileri, benlik algıları, hayata bakış açılarına yönelik sorular öğrenciye yöneltilmiştir. Bu sorulara verilen cevaplar üzerinden nitelikli okullarda okuyan (başarılı) öğrencilerin değişkenlerden hangilerine ne derecede sahip olduklarına yönelik analizler "R"

programı üzerinden “Boruta” paketi ile yapılmıştır. Boruta paketinde değişken belirlenirken sırasıyla; tüm değişkenlerin birer kopyası (shadowmax) oluşturulur. Ardından rastgele bir orman sınıflandırıcısı eğitilir ve ortalama azalma doğrusu ile her bir özelliğin önem derecesi belirlenir. Her yinelemede gerçek bir değişkenin gölge özelliklerinin en iyilerinden daha yüksek bir öneme sahip olup olmadığı kontrol edilerek, önemsiz değerlere sahip değişkenler kaldırılır. Son olarak, algoritma ya tüm değişkenler onaylandığında veya reddedildiğinde ya da çalışma sınırına ulaştığında durur (Kursa & Rudnicki, 2010). Bu araştırmada kullanılan veri aracı için 84 değişkenden 54’ü bu özellik seçim yöntemi uygulanıp elenmiş ve 30 değişken üzerinden değerlendirme yapılmıştır. Değişkenlerden en yüksek önem derecesine sahip olanı “Öğrencinin sahip olduğu kitap sayısı” olmuştur. Anne-baba eğitim durumları, sahip olunan olanaklar, matematik ve sosyal bilgilere karşı geliştirilen tutumlar, güzel sanatlara yönelik ilgi durumu, en çok sevilen dersler, öğretmenlerin ders anlatımı ve hakimiyeti gibi değişkenler de önemli değişkenler olarak belirlenmiştir.

Öğrencilerin belirlenen 30 değişkene verdikleri cevaplar üzerinden analiz çalışmalarına devam edilmiştir. Ankette yer alan soruların cevapları kategorik olduğu için genelde kategorik verilerle daha iyi çalıştığı bilinen CART, C5.0, DVM ve Rastgele Ormanlar algoritmaları ile analizler yapılmıştır (López, Lara, & Romero, 2021). 613 veri %80’i eğitim test verisi, %20’si ise test veri seti olarak ayrılmış, veri eğitildikten sonra test verisi üzerinden tahminler gerçekleştirilmiştir. Bu tahminlerde Rastgele ormanlar algoritması %75,61, DVM algoritması %68,80, C5.0 algoritması %68,47, CART algoritması ise %65,25 başarı oranına ulaşmıştır. Ardından hatalardan arınıklığı artırmak için çapraz doğrulama yapılmıştır. Doğrulama sonucunda Rastgele Ormanlar algoritması %74,3, C5.0 algoritması %73, DVM algoritması %69,2, CART algoritması ise %59,5 oranında tahmin başarıları göstermişlerdir. Rastgele ormanlar algoritmasının en başarılı tahmin yapan algoritma olduğu görülmüş, algoritma üzerinden oluşturulan modele ait katsayılar belirlenmiştir. Öğrencilerin yaptıkları tercihlere yönelik katsayılar ile 0,076 sabit değeri toplanmış, elde edilen sonucun 0.50’nin üzerinde olması durumunda öğrencinin başarılı, altında olması durumunda ise başarısız olduğu görülmüştür. Bu durumdan yola çıkılarak excel üzerinden basit bir program geliştirilmiştir. Programda değişkenlere verilen cevaplar üzerinden sabit değerin eklenmesiyle öğrenciye ait toplam puan elde edilmekte öğrencinin başarılı ya da başarısız olacağına dair tahminde bulunmaktadır. Ayrıca öğrencinin hangi koşullarda başarılı olabileceği tespit edildiğinden sisteme yeni giren bir öğrencinin süreç içerisinde eksiklikleri giderilebilecek, sistem bu anlamda bir erken uyarı sistemi olarak da kullanılma durumu olacaktır.

5.2 Öneriler

- Araştırma Antalya'nın merkez ilçelerinde toplam 613 öğrenciden toplanan veriler ile yapılmıştır. Araştırma için çok daha büyük sayılarda çalışma grupları seçilerek öğrencilerin başarı durumlarının kestirimi için hata oranı daha düşük modeller elde edilebilir.
- Öğrencinin geçmiş sınıf başarıları, sözlü ve yazılı sınavlardan aldıkları puanlara bakılarak nümerik değerler üzerinden lojistik regresyon analizi yapılarak, bu araştırmanın sonucuyla birlikte değerlendirilebilir. Böylece öğrencinin gelecekteki başarı durum tahmini iki yönlü gözlemlenmiş olabilir.
- Eğitim ve test veri kümelerine ait %80 ve %20'lik oranlar değiştirilerek farklı oranlarda elde edilen doğruluk dereceleri karşılaştırılabilir.
- Kategorik verilerle iyi çalıştığı bilinen farklı makine öğrenme algoritmaları ile model oluşturma uygulaması yapılabilir.
- Değişkenlere madde havuzunda belirlenmiş olan değişkenler de eklenerek önem derecesi belirleme işlemi tekrarlanabilir.
- Okul türlerine ait genel bilgi haritası çıkarılarak okullara giren öğrencilerin genel özellikleri çıkarılarak kurumlara özgü değişken durumu belirlenebilir.

KAYNAKÇA

- Ağaoğlu, M. (2016). Predicting Instructor Performance Using Data Mining Techniques in Higher Education. *IEEE ACCESS* , vol.4, 2379-2387.
- Ahmed, A., & Elaraby, I. (2014). Data Mining: A Prediction for Student's Performance Using Classification Method. *World Journal of Computer Application and Technology*, 2(2), 43-47.
- Akhil, M., Deekshatulu, B. L., & Chandra, P. (2013). Classification of Heart Disease Using K-nearest neighbor. *Procedia Technology*,10, 85-94.
- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, Cilt 29, Sayı 1/Nisan, 1-22.
- Akpınar, H. (2014). *Data: Veri Madenciliği Veri Analizi*. İstanbul: Papatya Yayıncılık Eğitim.
- Al-Radaideh, Q. A., Ananbeh, A. A., & Al-Shawakfa, E. M. (2001). A classification model for predicting the suitable study track for school students. *International Journal of Research and Reviews in Applied Sciences*, vol. 8.
- Altunkaynak, B. (2019). Kural Temelli Algoritmalar. B. Altunkaynak içinde, *Veri Madenciliği Yöntemleri ve R Uygulamaları* (s. 110). Ankara: Seçkin Yayıncılık.
- Argüden, Y., & Erşahin, B. (2008). *Veri madenciliği: Veriden bilgiye, masraftan değere*. İstanbul: Arge Danışmanlık Yayınları.
- Ayaz, M. (2021, Temmuz). Makine öğrenmesi algoritmaları ile covid-19 hastalarının belirlenmesi. *Master's thesis*. Denizli, Türkiye: Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü.

- Aydın, A. (2014). *Eğitim Psikolojisi*. Ankara: Pegem Yayıncılık.
- Bahadır, İ. (2008). Bayes Teoremi ve Yapay Sinir Ağları Modelleriyle Borsa Gelecek Değer Tahmini Uygulaması. *Yüksek Lisans Tezi*. Ankara: TOBB Ekonomi ve Teknoloji Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı.
- Baradwaj, B., & S.Pal. (2011). Mining Educational Data to Analyze Students Performance. *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 6.
- Baykul, Y., & Turgut, M. (2010). *Eğitimde Ölçme ve Değerlendirme*. içinde Ankara: Pegem Akademi Yayıncılık.
- Bilekdemir, G. (2010). Veri Madenciliği Teknikleri Kullanarak Üretim Süresi Tahmini ve Bir Uygulama. İzmir: Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.
- Bilekdemir, G. (2010). Veri Madenciliği Teknikleri Kullanarak Üretim Süresi Tahmini ve Bir Uygulama. İzmir: Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.
- Boyd, F. B. (2011). Reading and Writing Quarterly: Overcoming Learning Difficulties. *Motivation to continue: enhancing literacy learning for struggling readers and writers* (s. 257-277). içinde
- Bramer, M. (2007). Principles of Data Mining. London: Springer-Verlag London Limited.
- Cameron, S., & Heckman, J. (2001). The Dynamics of Educational Attainment for Black, Hispanic, and White Males. S. V. Cameron, & J. J. Heckman içinde, *Journal of Political Economy* (s. 455-499). Chicago: The University of Chicago Press .
- Carlisle, E., Stanley, L., & Kemple, K. (2005). Opening Doors: Understanding School and Family Influences on Family Involvement. *Early Childhood Education Journal*, 155-162.

- Carlson, E. A. (tarih yok). Early Environmental Support and Elementary School Adjustment as Predictors of School Adjustment in Middle Adolescence. *Journal of Adolescent Research* (s. 72-94). içinde
- Cemaloğlu, N., & Duykuloğlu, A. (2020). Yapay Sinir Ağı Modelleri. *Sosyal Bilimlerde Veri Madenciliği* (s. 510). içinde Ankara: Pegem Akademi Yayıncılık.
- Chapelle, O., Schölkopf, B., & Zien, A. (2006). Semi-supervised learning: Adaptive computation and machine learning. *Massachusetts: Massachusetts Institute of Technology*, 2.
- Checchi, D. (2000). Does educational achievement help to explain income inequality? *Departmental working papers 2000-11*. Milano, İtalya: Department of Economics, Management and Quantitative Methods at Università degli Studi di Milano.
- Choi, N. (2005). Self-Efficacy and Self-Concept as Predictors of College Students' Academic Performance. *American Psychological Association*, 197-205.
- Chu, F., Jin, G., & L.Wang. (2005). *Cancer diagnosis and protein secondary structure prediction using support vector machines*. L.WANG (Ed.) *Support vector machines: Theory and applications*(s. 243-363). Berlin: Springer- Verlag Berlin Heidelberg.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning* (s. 273-297). içinde
- Cortez, P. v. (2008). Using Data Mining to Predict Secondary School Student Performance. *Proceedings of 5th Annual Future Business Technology Conference*, (s. 5-12). Porto.
- Coşkun, C. (2010). Veri Madenciliği Algoritmaları Karşılaştırılması. Diyarbakır: Dicle Üniversitesi Fen Bilimleri Enstitüsü.

- Çetin, M. (2009). *Bir Üretim İşletmesinde Veri Madenciliği Yöntemleri*. Sakarya Üniversitesi Fen Bilimleri Enstitüsü.
- Daud, A., & Abbas, F. (2017). Predicting Student Performance using Advanced Learning Analytics. *International World Wide Web Conference Committee (IW3C2)*, published under Creative Commons, 415-421.
- Demirel, Ö. (1999). *Kuramdan Uygulamaya Eğitimde Program Geliştirme*. içinde Ankara: Pegem A Yayıncılık.
- Demirtaş, H., & Güneş, H. (2002). *Eğitim Yönetimi ve Denetimi Sözlüğü*. Ankara: Anı Yayıncılık.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 15-24.
- Efe, Ö., & Kaynak, O. (2000). *Yapay sinir ağı ve uygulamaları*. İstanbul: Boğaziçi Üniversitesi Printhouse.
- El Naqa, I., & Murphy, M. (2015). *What is machine learning?. In machine learning in radiation oncology (pp. 3-11)*. Cham: Springer.
- Elmas, Ç. (2003). *Yapay Sinir Ağları*. Seçkin Yayıncılık.
- Emre, İ. E., & Erol, Ç. (2017). Veri analizinde istatistik mi veri madenciliği mi? *Bilişim Teknolojileri Dergisi (10)2*, 161-167.
- Ersöz, F. (2019). Veri Madenciliği ve Karar Destek Sistemi. F. Ersöz içinde, *Veri Madenciliği Teknikleri ve Uygulamaları* (s. 15). Ankara: Seçkin Yayıncılık.
- Ertürk, S. (1973). *Eğitimde Program Geliştirme*. içinde Ankara: Yelkentepe Yayınları.
- Fidan, N. (1996). *Eğitim Psikolojisi, Okulda Öğrenme ve Öğretme*. içinde Ankara: Alkım Yayınevi.

- G.Williams. (2011). Data Mining With Rattle and R: The art of excavating data for knowledge discovery. *Data Mining With Rattle and R* (s. 3). içinde New York: Springer Science and Business Media.
- Gamgam, H., & Altunkaynak, B. (2017). *Spss Uygulamalı Regresyon Analizi*.
- Gedikoğlu, T. (2005). Avrupa Birliği Sürecinde Türk Eğitim Sistemi: Sorunlar ve Çözüm Önerileri . *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 66-80.
- Gedleç, Ş.(2020,Mart8).<https://www.datascienceearth.com>:
<https://www.datascienceearth.com/karar-agaclarinda-algoritma-secimi/> adresinden alındı
- Gorunescu, F. (2011). *Data mining concepts,models and techniques*. Berlin: Springer.
- Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: a longitudinal study. *Journal of Educational Psychology*,93(1), 3-13.
- Gökçen, H. (2010). *Kamuda Karar Destek Sistemlerinin Kullanımı ve Bir Model*. Kamu Bilgi İşlem Merkezleri Yöneticileri Birliği Kamu Bilişim.
- Gunjal, B., & Koganurmath, M. (2003). Database management: Concepts and design. *Prooceedings of 24th IASLIC-SIG-2003*, 2-3.
- Güner, Z. (2014). Veri madenciliğinde cart ve lojistik regresyon analizinin yeri: İlaç provizyon sistemi verileri üzerine örnek bir uygulama. *Sosyal Güvençe Dergisi* (6), 53-99.
- Güran, A., Uysal, M., & Doğrusöz, Ö. (2014). Destek vektör makineleri parametre optimizasyonunun duygu analizi üzerindeki etkisi. *Doku Eylül Üniversitesi Mühendislik Fakültesi Mühendislik Bilimleri Dergisi* (16), 86-93.
- Han, J., & Kamber, M. (1999). *San Francisco : Morgan Kaufmann Publishers*, 7.

- Han, J., & Kamber, M. (2000). Data Mining: Concepts and techniques. *Massachusetts : Morgan Kaufmann Publishers*, 18.
- Han, J., & M., K. (1999). Data mining: Concepts and techniques. *Massachusetts: Morgan Kaufmann Publishers*, 7.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques (Third Edition)*. Massachusetts: Morgan Kaufmann Publishers.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques (3rd edition)*. Waltham: Morgan Kaufmann Publishers.
- Hasan Demirtaş, H. G. (2002). *Eğitim Yönetimi ve Denetimi Sözlüğü*. Ankara : Anı Yayıncılık.
- Holsheimer, M., Kersten, M., Mannila, H., & Toivonen, H. (2015). A perspective on databases and data mining. *Proceedings of KDD-95, The Association For The Advancement Of Artificial Intelligence*, 150-155.
- İbrahim, Z., & Rusli, D. (2007). Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression. Shangri-La Hotel, Kuala Lumpur: 21st Annual SAS Malaysia Forum.
- İnan, O. (2003). Veri Madenciliği. *Yüksek Lisans Tezi*. Konya: Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.
- Jackson, J. (2002). Data mining: A conceptual overview. *Communications of the Association for Information Systems*, 8, 267-296.
- Jacobs, P. (1999). Data Mining: What General Managers Need to Know. *Harvard Management Update* 4, 8.
- Japkowicz, N. (2011). *Performance evaluation for learning algorithms*. Cambridge: Cambridge University Press.

- Kabakchieva, D., Stefanova, K., & Kisimov, V. (2011). Analyzing University Data for Determining Student Profiles and Predicting Performance. *Conference Proceedings of the 4th International Conference on Educational Data Mining (EDM 2011)*, (s. 347-348). Eindhoven.
- Kabra, R., & Bichkar, R. (2012). Performance prediction of engineering students using decision trees. *International Journal of Computer Applications*, 36.11, 8-12.
- Kantardzic, M. (2011). *Data mining: Concepts, models, methods, and algorithms(2nd edition)*. New Jersey: JohnWiley & Sons, Inc.
- Kılıç, S. (2013). Klinik karar vermede ROC analizi. *Journal of Mood Disorders*, 3, 40.
- Kızıldağ, S. (2009). Akademik başarının yordayıcısı olarak yalnızlık, boyun eğici davranışlar ve sosyal destek. *Yayımlanmamış yüksek lisans tezi*. Ankara: Hacettepe Üniversitesi.
- Klomegah, R. (2007). Predictors of academic performance of university students: An application of the goal efficacy model. *College Student Journal*, 407-415.
- Koç, N. (1981). Liselerde Öğrencilerin Akademik Başarılarının. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Yayınları*, 104.
- Koehrsen, W. (2021, 12 2). *www.medium.com*. <https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d> adresinden alındı
- Kumar, M., & Singh, A. (2017). Evaluation of Data Mining Techniques for Predicting Student's Performance. *I.J. Modern Education and Computer Science*, 25-31.
- Kumdereli, Ü. (2012). Tıp Bilişimi ve veri madenciliği uygulamaları: Eeg sinyallerindeki epileptiform aktiviteye veri madenciliği uygulanması(Yayımlanmamış yüksek lisans tezi). Edirne: Trakya Üniversitesi.

- Kuzey, C. (2012). Veri Madenciliğinde destek vektör makineleri ve karar ağaçları yöntemlerini kullanarak bilgi çalışanlarının kurum performansı üzerine etkisinin ölçülmesi ve bir uygulama(Yayınlanmamış doktora tezi). İstanbul: İstanbul Üniversitesi.
- Larose, D. T. (2005). *Discovering knowledge in data: An introduction to data mining*. New Jersey: John Wiley and Sons.
- Loh, W. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery* 1 (January), 14-23.
- MEB.(2013,09,07). www.mevzuat.meb.gov.tr.
<https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=18812&MevzuatTur=7&Mevzuatertip=5> adresinden alındı
- MEB. (2020, Nisan 04). <https://www.hurriyet.com.tr/yazarlar/nuran-cakmakci/lgs-kontenjani-yuzde-50-artti-41486122>. adresinden alındı
- Minaei-Bidgoli, B. K. (2003). Predicting Student Performance: An Application of Data Mining Methods with an Educational Web-Based System. *In 33rd Annual Frontiers in Education, 1, T2A-13, IEEE*.
- Mitchell, T. (1997). *Machine Learning*. New York: Mc Graw-Hill Education.
- Namok, C. (2005). Self-Efficacy and Self-Concept as Predictors of College Students' Academic Performance. *American Psychological Association*, 197-205.
- OECD. (2012). <http://www.oecd.org/pisa/keyfindings/PISA-2012-results-turkey.pdf>.
<http://www.oecd.org>. adresinden alındı
- Oğuzkan, A. F. (1993). *Eğitim Terimleri*. içinde Ankara: Gül Yayınevi.
- Özer, Y., & Anıl, D. (2011). Öğrencilerin fen ve matematik başarılarını etkileyen faktörlerin yapısal eşitlik modeli ile incelenmesi. *Hacettepe Eğitim Fakültesi Dergisi*, 313-324.

- Özer, Y., & Anıl, D. (tarih yok). Öğrencilerin fen ve matematik başarılarını etkileyen faktörlerin yapısal eşitlik modeli ile incelenmesi. *Hacettepe Üniversitesi Eğitim b.*
- Özkan, Y. (2008). *Veri madenciliği yöntemleri*. İstanbul: Papatya Yayıncılık.
- Patel, B. N., Prajapati, S. G., & Lakhtaria, K. I. (2012). Efficient classification of data using decision tree. *Bonfring International Journal of Data Mining*, 2(1), 6-12.
- Patil, N., Lathi, R., & Chitre, V. (2012). Customer Card Classification Based on C5.0 & CART Algorithms. *International Journal of Engineering Research Vol. 2, Issue 4*, 164-167.
- Piramuthu, S. (1998). Evaluating Feature Selection Methods For Learning in Data Mining Applications. *Thirty-First Annual Hawaii International Conference on System Sciences* (s. 294). Kohala Coast Hawaii USA: IEEE Computer Society.
- Pujari, A. (2008). Data Mining techniques. *Hyderabad: Universities Press(India)Private Limited*, 45.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- Ramaswami, M. v. (2009). A Study on Feature Selection Techniques in Educational Data Mining. *Journal of Computing*, 1(1).
- Refaat, M. (2007). Data preparationvfor data mining using SAS. San Fransisco: Morgan Kaufmann Publishers.
- Richards, M. (1986). *Productive and Effective Schools*. Chicago,İllinois: The Annual Conference of The American Finance Association.
- Saa, A. A. (2016). Educational Data Mining & Student's Performance Prediction. *International Journal of Advanced Computer Science and Applications*, Vol.7, No.5.

- Shearer, C. (2000). The Crisp-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing, Cilt 5 No 4*, 13-23.
- Silahtaroglu, G. (2008). Kavram ve Algoritmalarıyla Temel Veri Madenciliği. Papatya Yayıncılık.
- Şad, S. N. (2012). Investigation of Parental Involvement Tasks as Preditors of Primary Students' Turkish, Math, and Science & Technology Achievement. *Eurasian Journal of Educational Research (EJER)*.
- Şad, S. N., & Şahiner, Y. K. (2016, 01 01). Temel Eğitimden Ortaöğretime Geçiş (TEOG) Sistemine İlişkin Öğrenci, Öğretmen ve Veli Görüşleri. Gaziantep, Türkiye: 3. Ulusal Eğitim Programları ve Öğretim Kongresi.
- Şirin, S. R. (2005). Socioeconomic Status and Academic Achievement:A Meta-Analytic Review of Research. *Review of Educational Research*, 417-453.
- Şişman, M. (2002). *Eğitimde Mükemmellik Arayışları*. Ankara: Pegem Yayıncılık.
- Şişman, M. (2002). *Eğitimde Mükemmellik Arayışları*. Ankara: Pegem Yayıncılık.
- Şişman, M. (2014). *Öğretim Liderliği*. Ankara: Pegem Yayıncılık.
- Tang, Z. H., & MacLennan, J. (2005). *Data mining with SQL server 2005*. Indianapolis: Wiley Publishing, Inc.
- Tatar, M. (2006). Okul ve Öğretmenin Öğrenci Başarısı Üzerindeki Etkisi. *Milli Eğitim*, 156-166.
- Tekerek, A. (2011). Veri madenciliği süreçleri ve açık kaynak kodlu veri madenciliği araçları. *Akademik Bilişim '11- XIII. Akademik Bilişim Konferansı Bildirileri* (s. 162). Malatya: İnönü Üniversitesi.

- Tekin, H. (2004). *Eğitimde Ölçme ve Değerlendirme*. içinde Ankara: Yargı Yayınevi.
- Toksöz, C. (2016). Yayımlanmamış yüksek lisans tezi. *Veri madenciliği süreçleri ile tavuk yumurtalarında cinsiyet ayrımı*. Isparta: Süleyman Demirel Üniversitesi.
- Tural, K. N. (2002). Öğrenci Başarısında Etkili Okul Değişkenleri ve Eğitimde Verimlilik. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 40-54.
- Uluğ, F. (2000). *Okulda Başarı*. İstanbul: Remzi Kitabevi.
- Uyan, M., & Çay, T. (2008). *Mekansal Uygulamalar İçin Veri Madenciliği Yaklaşımı, 2. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu UZAL-CBS 2008, Kayseri*. <http://uzalcbs.org>: <http://uzalcbs.org/gallery/ii-uzal-cbs-2008-erciyes-uni/> adresinden alındı
- Vapnik, V. (1998). *Statistical Learning Theory*. New York: John Wiley & Sons, Inc.
- Vural, T. (2017). Veri madenciliği kullanılarak mikrodalga verileri ile beyindeki kitlesel lezyonların teşhisi. *Yayımlanmamış yüksek lisans tezi*. İstanbul: Yıldız Teknik Üniversitesi.
- Wikipedia. (2021, Ocak 13). *R (programlama dili)*. [https://tr.wikipedia.org://tr.wikipedia.org/w/index.php?title=R_\(programlama_dili\)&oldid=24298848](https://tr.wikipedia.org://tr.wikipedia.org/w/index.php?title=R_(programlama_dili)&oldid=24298848) adresinden alındı
- Williams, G. (2011). *Data Mining with Rattle and R: The art of excavating data for knowledge discovery*. New York: Springer Science+Business Media.
- Witten, I., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques (2nd edition)*. San Francisco: Morgan Kaufmann Publishers.
- Xing, L., & Koirala, H. (2009). The Effect of Mathematics Self-Efficacy on Mathematics Achievement of High School Students. *NERA Conference Proceedings 2009*

https://opencommons.uconn.edu/nera_2009/30 (s. 30). Connecticut: Eastern Connecticut State University.

Yavuz, M. A. (2009). Kaçak Su Kullanımının Tespitinde Veri Madenciliği Yaklaşımı . Sakarya Üniversitesi Fen Bilimleri Enstitüsü.

YÖK.(1983,817).www.mevzuat.gov.tr.

<https://www.mevzuat.gov.tr/MevzuatMetin/1.5.2547.pdf> adresinden alındı

Zaki, M., & Jr, W. M. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.



TOPLANTI TARİHİ : 29.03.2021

TOPLANTI SAYISI : 05

KARAR SAYISI : 141

Universitemiz Eğitim Fakültesi Eğitim Bilimleri Bölümü öğretim üyesi **Doç. Dr. Bilal Barış ALKAN**'ın danışmanlığını, Şerafettin **KUZUCUK**'un araştırmacılığını üstlendiği, *"Öğrenci Başarısını Değerlendirmede Makine Öğrenmesi Algoritmalarına Dayalı Yeni Bir Sınıflandırıcı Modeli"* konulu çalışmanın, fikri hukuki ve telif hakları bakımından metod ve ölçeğine ilişkin sorumluluğun başvurucaya ait olmak üzere, proje süresince uygulanmasının etik olarak uygun olduğuna oy birliği ile karar verilmiştir. (2021G133)

Prof. Dr. Osman ERAVŞAR
Kurul Başkanı

Başkan
Prof. Dr.
Osman ERAVŞAR

Başkan Yrd.
Prof. Dr.
Bahattin ÖZDEMİR

Üye
Prof. Dr.
Hilmi DEMİRKAYA

Üye
Prof. Dr.
Mustafa ŞEKER

Üye
Prof. Dr.
Adnan DÖNMEZ

Üye
Prof. Dr.
Abdullah KARAÇAĞ

Üye
Prof. Dr.
Eyyup YARAŞ



T.C.
ANTALYA VALİLİĞİ
İl Millî Eğitim Müdürlüğü

Sayı : E-98057890-20-25656390
Konu : Anket Uygulaması

27.05.2021

İL MİLLÎ EĞİTİM MÜDÜRLÜĞÜNE
ANTALYA

İlgi : 21/01/2020 tarih ve 1563890 sayılı Millî Eğitim Bakanlığına Bağlı Okul ve Kurumlarda Yapılacak Araştırma, Yarışma ve Sosyal Etkinlik İzinlerine Yönelik İzin Uygulama Genelgesi. ve

Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğü, Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans Öğrencisi Şerafettin KUZUCUK'un "**Öğrenci Başarısını değerlendirme Makine Öğrenmesi Algoritmalarına Dayalı Yeni Bir Sınıflandırıcı Modeli**" adlı araştırmasını, İlimize Bağlı 5 (Beş) Merkez İlçedeki Ortaokullar ve Liselede uygulama isteği ile ilgili 07/05/2021 tarih ve 85370 sayılı başvurusu, Müdürlüğümüz ARGE Birimi Değerlendirme ve İnceleme Komisyonunca incelenmiş olup;

Adı geçenin ilgi Genelge kapsamında 2020-2021 Eğitim Öğretim Yılı içerisinde olmak üzere, İlimize Bağlı 5 (Beş) Merkez İlçedeki Ortaokullar ve Liselerde öğrenim gören öğrencilere yönelik araştırmasını, Okul Müdürlüklerinin sorumluluğunda Eğitim-Öğretim faaliyetlerini aksatmaksızın yürütmesi,

Söz konusu araştırmanın bitimine müteakip; sonuç raporunun bir örneğinin CD ortamında Müdürlüğümüz Ar-Ge bürosuna gönderilmesi kaydıyla uygulanması, Komisyonca uygun görülmüştür.

Makamlarınızca da uygun görüldüğü takdirde, Valilik Makamının 25/08/2020 tarih ve 24911 sayılı yetki devrine göre olurlarınıza arz ederim.

Mehmet KARAKAŞ
Müdür a.
Müdür Yardımcısı

OLUR
27.05.2021

Hüseyin ER
Vali a.
İl Millî Eğitim Müdürü

Adres : Antalya İl Millî Eğitim Müdürlüğü

Telefon No : 0 (242) 238 60 00
E-Posta: arge07@meb.gov.tr
Kep Adresi : meb@hs01.kep.tr

Bu belge güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Adresi : <https://www.turkiye.gov.tr/meb-ebys>

Bilgi için: A.ATICI

Unvan : Veri Hazırlama ve Kontrol İşletmeni
İnternet Adresi: Faks:2422386111



Bu evrak güvenli elektronik imza ile imzalanmıştır. <https://evraksorgu.meb.gov.tr> adresinden 85d4-04ac-3bd3-9ea8-994e kodu ile teyit edilebilir.

ANKET FORMU

Bu çalışmaya kendi rızam ve velimin onayı ile katılıyorum. İstedğim zaman çalışmadan ayrılabilceğimi bilerek verdiğim bilgilerin bilimsel amaçlarla kullanılmasını kabul ediyorum.

☐ Evet ☐ Hayır

Cinsiyetiniz*

☐ Kız ☐ Erkek

Veliniz

☐ Anne ☐ Baba ☐ Diğer

Okulunuzun Türü

☐ Ortaokul ☐ Anadolu Lisesi ☐ Fen Lisesi ☐ Sosyal Bilimler Lisesi
☐ Anadolu İmam-Hatip Lisesi ☐ Güzel Sanatlar Lisesi

Okulunuzun bulunduğu ilçeyi işaretleyiniz.

☐ Aksu ☐ Döşemealtı ☐ Kepez ☐ Konyaaltı ☐ Muratpaşa

Doğum Yerinizi;

Derslerinize belirli bir plan dahilinde çalışır mısınız?

☐ Evet ☐ Bazen ☐ Hayır

Dersin özelliğine göre bir çalışma metodu belirler misiniz?

☐ Evet ☐ Bazen ☐ Hayır

En sevdiğiniz dersin adını yazınız.

Kendinize ait bir çalışma odanız var mı?

☐Var

☐Yok

Kendinize ait bir çalışma masanız var mı?

☐Var

☐Yok

Evde kullanabileceğiniz bir bilgisayarınız var mı?

☐Var

☐Yok

Evde internet bağlantınız var mı?

☐Var

☐Yok

Evde kullanabileceğiniz bir kitaplık var mı?

☐Var

☐Yok

Kaç kardeşiniz var?

☐0

☐1

☐2

☐3

☐4

☐5 ve üstü

Türk klasiklerini okumaktan hoşlanır mısınız?

☐Evet

☐Bazen

☐Hayır

Dünya klasiklerini okumaktan hoşlanır mısınız?

☐Evet

☐Bazen

☐Hayır

Ailenizin ortalama aylık gelirini seçiniz.

☐0-3000 ₺

☐3001-6000 ₺

☐6001-9000 ₺

☐9001-12000 ₺

☐12001 ve üstü

Evinizde size ait olan kitap sayısını seçiniz.

☐ 0-15 ☐ 16-30 ☐ 31-100 ☐ 101-200 ☐ 201 ve üstü

Annenizin öğrenim durumunu belirtiniz.

☐ Okuma-yazma bilmiyor ☐ Okuma-yazma biliyor ☐ İlkokul ☐ Ortaokul ☐ Lise ☐
☐ Ön lisans-Lisans ☐ Yüksek lisans ☐ Doktora

Babanızın öğrenim durumunu belirtiniz.

☐ Okuma-yazma bilmiyor ☐ Okuma-yazma biliyor ☐ İlkokul ☐ Ortaokul ☐ Lise ☐
☐ Ön lisans-Lisans ☐ Yüksek lisans ☐ Doktora

Anneniz bir işte çalışıyor mu?

☐ Evet ☐ Hayır ☐ Pandemi döneminde işten ayrıldı.

Annenizin mesleği hangi meslek grubunda yer almaktadır? (Yukarıdaki soruyu hayır olarak işaretlediyseniz, soruyu boş bırakın.)

☐ Sağlık ☐ Eğitim ☐ Gıda ☐ Turizm ☐ Ticaret
☐ Tarım ☐ Ulaşım ☐ Savunma ve Güvenlik ☐ Diğer ☐ Meslek sahibi değil

Babanız bir işte çalışıyor mu?

☐ Evet ☐ Hayır

Babanızın mesleği hangi meslek grubunda yer almaktadır? (Yukarıdaki soruyu hayır olarak işaretlediyseniz, soruyu boş bırakın.)

☐ Sağlık ☐ Eğitim ☐ Gıda ☐ Turizm ☐ Ticaret
☐ Tarım ☐ Ulaşım ☐ Savunma ve Güvenlik ☐ Diğer ☐ Meslek sahibi değil

Güzel sanatlara ilgi gösteriyor musunuz?

☐ Evet ☐ Hayır

Evinizde güzel sanatlara yönelik bir kaynak (müzik aleti, resim gereçleri, dergiler, kitaplar vb.) bulunuyor mu?

☐ Evet ☐ Hayır

En sevdiğiniz spor türünü yazınız.

.....

Spor yapıyorsanız, spor yapmak için günlük olarak ayırdığınız zaman dilimini işaretleyiniz.

☐ 1 saatten az

☐ 1-2 saat arası

☐ 2 saatten fazla

Günlük ortalama kaç saat uyuyorsunuz?

☐ 5 saatten az

☐ 5-7 saat arası

☐ 7-9 saat arası

☐ 9-11 saat arası

☐ 11 saatten fazla

Yeterli ve dengeli beslendiğinizi düşünüyor musunuz?

☐ Evet

☐ Kısmen

☐ Hayır

Kahvaltınızı ve akşam yemeklerinizi tüm aile fertleri ile birlikte mi yersiniz?

☐ Evet

☐ Kısmen

☐ Hayır

Robotik ve kodlama alanı ile ilgileniyor musunuz?

☐ Evet

☐ Kısmen

☐ Hayır

Tiyatro, konser, sinema gibi etkinliklerden hoşlanır mısınız?

☐ Evet

☐ Kısmen

☐ Hayır

Twitter, Facebook, Instagram, Youtube gibi sosyal medya platformlarında günlük olarak ortalama kaç saat geçiriyorsunuz?

☐ 0-1 saat arası

☐ 1-2 saat arası

☐ 2-3 saat arası

☐ 4-5 saat arası

☐ 5 saatten fazla

Arkadaşlarınızla günlük olarak ortalama kaç saat birlikte vakit geçiriyorsunuz?

☐0-1 ☐2-3 ☐4-5 ☐6-7 ☐8 ve üzeri

Televizyon izleyerek günde ortalama kaç saat vakit geçiriyorsunuz?

☐0-1 ☐2-3 ☐4-5 ☐6-7 ☐8 ve üzeri

Evinizde derslerinize yardımcı kaynak kitaplar var mıdır?

☐Var ☐Yok

İnternet üzerinden yararlanabileceğiniz bir eğitim kursuna daha önce katıldınız mı?

☐Evet ☐Hayır

Eğitim hayatınız boyunca genel anlamda özel ders ve takviyelere ihtiyaç duydunuz mu?

☐Evet ☐Hayır

Yolculuk yapmayı, yeni yerler keşfetmeyi, doğa ile iç içe olmayı sever misiniz?

☐Evet ☐Hayır

Bir makineyi söküp yeniden takmak, bir makine ya da alet yapmak gibi el becerisi ve görsel düşünme becerisi gerektiren işlerle uğraşmayı sever misiniz?

☐Evet ☐Hayır

Bilgisayar oyunlarına günde ortalama kaç saat ayırıyorsunuz?

☐0-1 ☐2-3 ☐4-5 ☐6-7 ☐8 ve üzeri

Aşağıdaki bilgisayar oyunlarından oynadıklarınızı işaretleyiniz.

- ☐ Savaş ve Macera Oyunları (PubG, Counter-Strike: Global Offensive, League of Legends, Fortnitevb.)
- ☐ Strateji Oyunları (OffworldTradingCompany, Stellaris, XCOM 2, Homeworld:Deserts of Kharakvb.)
- ☐ Spor Oyunları (PES 21,NBA 2K20,Football Manager 2020,WWE 2K20,Fifa 2021vb.)
- ☐ Simülasyon oyunları (Euro Truck Simulator 2, American Truck Simulator, Bus Simulator vb.)
- ☐ Diğer

Konuşabildiğiniz dilleri seçiniz.

- ☐ Fransızca ☐ Almanca ☐ İngilizce ☐ İtalyanca ☐ İspanyolca ☐ Rusça ☐ Diğer

Ailenizin sahip olduğu araba sayısını işaretleyiniz.

- ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Evinizde çalışır durumdaki televizyon sayısını işaretleyiniz.

- ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Evinizde çalışır durumdaki bilgisayar sayısını işaretleyiniz.

- ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Evinizde çalışır durumdaki telefon sayısını işaretleyiniz.

- ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Evinizde bulunan müzik aleti sayısını işaretleyiniz.

- ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Hayata bakış açınızı nasıl değerlendirirsiniz?

- ☐ Olumlu ☐ Kısmen Olumlu ☐ Olumsuz

İnsanların sizinle birlikteyken genelde güzel zaman geçirdiğini düşünüyor musunuz?

☐ Evet ☐ Bazen ☐ Hayır

Zorluklar karşısında hemen pes eder misiniz?

☐ Evet ☐ Bazen ☐ Hayır

Beklenmedik durumlarla başa çıkabilmek için davranışlarınızı değiştirir misiniz?

☐ Evet ☐ Bazen ☐ Hayır

Bir ödev ya da görev verildiğinde bitirmek için gayret gösterir misiniz?

☐ Evet ☐ Bazen ☐ Hayır

Kendinize güveniyor musunuz?

☐ Evet ☐ Bazen ☐ Hayır

Diğer öğrencilerle rekabet içinde olmaktan keyif alır mısınız?

☐ Evet ☐ Bazen ☐ Hayır

Bazı zamanlarda kendinizi faydasız hissettiğiniz oluyor mu?

☐ Evet ☐ Bazen ☐ Hayır

Hayatınızdan memnuniyet derecenizi "0=hiç memnun değilim, 10= çok memnunuz" olmak üzere 0 ile 10 arasında derecelendiriniz.

☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

"İleride yüksek bir maaş alabilmek için okula gidiyorum." cümlesine ne derece katılıyorsunuz? (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

"Başarı duygusunu tatmak için okula gidiyorum." cümlesine ne derece katılıyorsunuz? (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☒ 10

Okuma alışkanlığınızı derecelendiriniz. (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Okuduğunuz metni anlama derecenizi değerlendiriniz. (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Matematik dersindeki genel başarı durumunuzu derecelendiriniz(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında).

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Fen Bilimleri(Biyoloji, Fizik, Kimya) alanındaki genel başarı durumunuzu derecelendiriniz(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında).

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Sosyal Bilimler(Tarih, Coğrafya, Felsefe) alanındaki genel başarı durumunuzu derecelendiriniz(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında).

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Başarılı olduğunuz alanlarda çalışmak motivasyonunuzu ne derecede etkiler? (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Aynı anda birden fazla zorlukla başa çıkabilme becerinizi değerlendiriniz. (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Yeni bir duruma hızlı bir şekilde uyum sağlama becerinizi derecelendiriniz. (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Derse devam durumunuzu derecelendiriniz. (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Baskı ya da stres altında olmanız sizi ne derece etkiler? (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Covid-19 salgını sürecinden fiziksel ve duygusal olarak ne derece etkilendiğinizi düşünüyorsunuz? (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Covid-19 salgının akademik başarınızı ne derecede etkilediğini veya etkileyeceğini düşünüyorsunuz?(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Covid-19 salgın sürecinin kaygı düzeyinizi ne derecede etkilediğini düşünüyorsunuz?(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Ailenizin okul ile ilgili etkinliklere karşı ilgisini derecelendiriniz. (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Aileniz eğitim-öğretim konusundaki çabalarınızı ne derecede destekler? (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Aileniz karşılaştığınız zorluklar karşısında sizi ne derece destekler? (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Bir sınav sonucunda sınav sonucunuz konusunda duyduğunuz kaygıyı derecelendiriniz. (0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Gireceğiniz sınava bilgi yönünden hazır olsanız da kaygılı hisseder misiniz? Kaygı durumunuzu derecelendiriniz.

(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Okulda kendinizi ne derece yalnız hissedersiniz?

(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Okulda çabuk arkadaş edinir misiniz? Derecelendiriniz.

(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Bilim ile ilgili video, belgesel, blog, dergi vb araçları ne derece takip edersiniz?

(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Okulunuz öğretmenlerinin genel olarak ders anlatımı ve derse hakimiyetini derecelendiriniz.

(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Okulunuzda ne derece huzurlu bir çalışma ortamı olduğunu düşünüyorsunuz?

(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

Genel sağlık durumunuzu derecelendiriniz.

(0= en az, 10 = en çok olmak üzere 0 ile 10 arasında)

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9 ☐10

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı-Soyadı: Şerafettin KUZUCUK

Doğum Yeri ve Tarihi:

Eğitim Bilgileri

Lisans: Selçuk Üniversitesi Sınıf Öğretmenliği

Yüksek Lisans: Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü Eğitimde Ölçme Değerlendirme Ana Bilim Dalı

İş Deneyimi

(2008-2019) (Milli Eğitim Bakanlığı) Antalya /Kaş/ Ova İlkokulu /Sınıf Öğretmeni

(2019- 2020) (Milli Eğitim Bakanlığı) (Antalya/Muratpaşa Merkez İlkokulu/Okul Müdür Yardımcısı)

(2020- 2021) (Milli Eğitim Bakanlığı) (Antalya/Muratpaşa Ahmet Ferda Kahraman /Okul Müdür Yardımcısı/ Okul Müdürü)

(2021- Devam ediyor)) (Milli Eğitim Bakanlığı) (Antalya/Muratpaşa Azize Kahraman Halk Eğitimi Merkezi /Kurum Müdür Yardımcısı)

Turnitin Orijinallık Raporu

İşleme kondu: 26-May-2022 15:11 +03

NUMARA: 1844573784

Kelime Sayısı: 16283

Gönderildi: 1

ÖĞRENCİ BAŞARISINI DEĞERLENDİRMEDE MAKİNE ÖĞR... Şerafettin Kuzucuk
tarafından

Kaynağa göre Benzerlik	
Benzerlik Endeksi	
%13	
İnternet Sources:	%12
Yayımlar:	%3
Öğrenci Ödevleri:	%6

[alıntılar çıkar](#) [bibliyografayı dahil et](#) [küçük eşleşmeleri çıkar](#)

mod: raporu hızlı görüntüle (klasik)

Change mode

[vazdır](#)

[yenile](#)

[indir](#)

1% match (16-May-2015 tarihli internet)

<http://acikarsiv.ankara.edu.tr>

1% match (13-Şub-2022 tarihli internet)

<https://dergipark.org.tr/tr/download/issue-file/33516>

1% match (09-Haz-2021 tarihli internet)

<https://www.imcofe.org/dosyalar/antalyaproceedingdandabstract.pdf>

<1% match (16-Ara-2021 tarihli internet)

https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/40057/yokAcikBilim_10327277.pdf?isAllowed=y&sequence=-1

<1% match (24-May-2022 tarihli internet)

https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/222834/yokAcikBilim_10297507.pdf?isAllowed=y&sequence=-1

<1% match (15-Nis-2022 tarihli internet)

https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/204255/yokAcikBilim_10082258.pdf?isAllowed=y&sequence=-1

<1% match (07-Mar-2022 tarihli internet)

https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/144365/yokAcikBilim_10142711.pdf?sequence=-1

<1% match (20-Ara-2021 tarihli internet)

https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/420128/yokAcikBilim_10324608.pdf?se=

<1% match (10-Şub-2022 tarihli internet)

https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/349833/yokAcikBilim_10215650.pdf?isAllowed=y&sequence=-1

<1% match (12-Oca-2022 tarihli internet)

https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/438409/yokAcikBilim_10024319.pdf?isAllowed=y&sequence=-1

<1% match (12-Oca-2022 tarihli internet)

https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/40174/yokAcikBilim_10246283.pdf?isAllowed=y&sequence=-1

<1% match (11-May-2022 tarihli internet)

https://acikbilim.yok.gov.tr/bitstream/handle/20.500.12812/400966/yokAcikBilim_10338705.pdf?sequence=-1

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kâğıt ve elektronik kopyalarının Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

☐ Tezimin/Raporumun tamamı her yerden erişime açılabilir.

☐ Tezim/Raporum sadece Akdeniz Üniversitesi yerleşkelerinden erişime açılabilir. Tezimin/Raporumun yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

20/06/2022

Şerafettin KUZUCUK