



**ÖĞRENCİLERİN LGS TERCİHLERİNE GÖRE
ORTAÖĞRETİM KURUMLARINA YERLEŞME
DURUMLARININ VERİ MADENCİLİĞİ
YAKLAŞIMI İLE MODELLENMESİ**

YÜKSEK LİSANS TEZİ

Ümit ÇELİK

Danışman

Doç. Dr. Fatih ÖZDİNÇ

BİLGİSAYAR ANABİLİM DALI

Şubat 2024

AFYON KOCATEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ÖĞRENCİLERİN LGS TERCİHLERİNE GÖRE ORTAÖĞRETİM
KURUMLARINA YERLEŞME DURUMLARININ VERİ
MADENCİLİĞİ YAKLAŞIMI İLE MODELLENMESİ

Ümit ÇELİK

Danışman

Doç. Dr. Fatih ÖZDİNÇ

BİLGİSAYAR ANABİLİM DALI

Şubat 2024

TEZ ONAY SAYFASI

Ümit ÇELİK tarafından hazırlanan “Öğrencilerin LGS Tercihlerine Göre Ortaöğretim Kurumlarına Yerleşme Durumlarının Veri Madenciliği Yaklaşımı ile Modellenmesi” adlı tez çalışması lisansüstü eğitim ve öğretim yönetmeliğinin ilgili maddeleri uyarınca 21/02/2024 tarihinde aşağıdaki jüri tarafından **oy birliği** ile Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü **Bilgisayar Anabilim Dalı’nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Doç. Dr. Fatih ÖZDİNÇ

Başkan : Doç. Dr. Erhan ÜNAL
Afyon Kocatepe Üniversitesi, Eğitim Fakültesi

Üye : Doç. Dr. Fatih ÖZDİNÇ
Afyon Kocatepe Üniversitesi, İ.İ.B Fakültesi

Üye : Doç. Dr. Gökhan AKÇAPINAR
Hacettepe Üniversitesi, Eğitim Fakültesi

Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü Yönetim Kurulu’nun
...../...../..... tarih ve
..... sayılı kararıyla onaylanmıştır.

.....
Prof. Dr. Bekir YALÇIN
Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİM SAYFASI
Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlâk kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı,

beyan ederim.

21 / 02 /2024

Ümit ÇELİK

ÖZET

Yüksek Lisans Tezi

ÖĞRENCİLERİN LGS TERCİHLERİNE GÖRE ORTAÖĞRETİM KURUMLARINA YERLEŞME DURUMLARININ VERİ MADENCİLİĞİ YAKLAŞIMI İLE MODELLENMESİ

Ümit ÇELİK

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Anabilim Dalı

Danışman: Doç. Dr. Fatih ÖZDİNÇ

Bu araştırma e-Okul sistemindeki verileri kullanarak her yıl bir liseye yerleşmek için tercih yapan öğrencilerin LGS tercihlerine göre yerleşmesinin VM yöntemleri ile modellenmesini amaçlamaktadır. Bu kapsamda araştırma grubunu 2022 yılında Ankara ilinde LGS tercihi yapmış olan 6000 8. sınıf öğrencisi oluşturmaktadır. LGS sonucunda öğrenciler sınavlı ve sınavsız okullara yerleştirilmekte farklı nedenlerden dolayı ise bir kısım öğrenci hiçbir okula yerleşememektedir. Algoritmaların daha iyi bir test ve tahminleme yapabilmesi için yerleşme durumlarına göre hiçbir okula yerleşmeyen, sınavlı okula yerleşen ve sınavsız bir okula yerleşen olmak üzere 2000’ er tane kayıt rastgele olarak seçilmiştir. Verilerin analizinde orange veri madenciliği yazılımı kullanılmıştır.

Araştırma sonucunda veri setimizden information gain yöntemi ile veri başlıklarından en iyi 10 tanesi seçilmiş ve bu seçime göre öğrencilerin bir liseye yerleşme durumları (yerleşemedi, sınavlı bir okula yerleşti, sınavsız bir okula yerleşti) VM yöntemleri ile modellenmeye çalışılmıştır. Verideki öğrencinin akademik başarısını gösteren yılsonu başarı puanlarının daha etkin olduğu bunun dışındaki demografik verilerin ise anne-babanın eğitim durumu, öğrencinin yapmış olduğu tercihlere göre ikinci ve üçüncü sıradaki tercihleri, tercih sayısı, burs alıp almadığı ve kayıtlı olduğu okul türü(devlet ortaokulu, özel ortaokul, imam hatip ortaokulu) bu kriterlerden sonra geldiği sınıflama algoritmaları açısından incelediğimizde ise en yüksek doğruluk oranları yakalanan algoritmalarda %67 ile lojistik regresyon öne çıkmakta ve bunları %64 ile yapay sinir ağları ve naive bayes algoritmaları izlemektedir. Bu çalışma neticesinde öğrencilerin akademik ve demografik

verileri kullanarak tercih öncesi bir tercih robotu gibi bir uygulama sunulabilir ve farklı akademik ve demografik veri başlıkları kullanılarak EVM uygulamaları yapılabilir.

2024, xi+67 sayfa

Anahtar Kelimeler: E-Okul, LGS, Veri madenciliği, Eğitsel veri madenciliği, Özellik seçimi, Tahminleme



ABSTRACT

M. Sc. Thesis

MODELING OF PLACEMENT SITUATIONS IN SECONDARY EDUCATION INSTITUTIONS ACCORDING TO LGS PREFERENCES WITH DATA MINING APPROACH

Ümit ÇELİK

Afyon Kocatepe University

Graduate School of Natural and Applied Sciences

Department of Computer

Supervisor: Asst. Prof. Fatih ÖZDİNÇ

This research aims to model the placement of students in high schools each year based on their preferences in the LGS using DM methods. The study group consists of 6000 8th-grade students who made their LGS preferences in Ankara province in 2022. Following the LGS results, students are placed in schools with exams and without exams, while some students fail to be placed in any school due to various reasons. To enable algorithms to conduct better testing and prediction, 2000 records were randomly selected based on their placement status: those who were unable to be placed in any school, those placed in a school with exams, and those placed in a school without exams. Orange data mining software was used for data analysis.

The research selected the top 10 data headings from the dataset using the information gain method, and attempts were made to model students' placement statuses (unplaced, placed in a school with exams, placed in a school without exams) using DM methods based on this selection. It was found that end-of-year academic success scores were more effective in indicating placement, while other demographic data such as parents' education level, students' second and third preferences, number of preferences, whether they receive scholarships, and the type of school they are enrolled in (public middle school, private middle school, religious middle school) followed in importance. When examined in terms of classification algorithms, logistic regression stands out with the highest accuracy rates at 67%, followed by artificial neural networks and naive Bayes algorithms at 64%. As a result of this study, an application similar to a preference robot could be offered to

students based on their academic and demographic data, and different academic and demographic data headings could be used for further EDM applications.

2024, xi+67 pages

Keyword: E-school, LGS, Data mining, Educational data mining, Feature selection, Prediction



TEŞEKKÜR

Bu araştırma sürecinde çalışmaların yönlendirilmesi, sonuçların değerlendirilmesi ve yazımı aşamasında yapmış olduğu büyük katkılarından dolayı tez danışmanım Sayın Doç. Dr. Fatih ÖZDİNÇ' e en içten teşekkürlerimi sunarım.

Bu tezi yapmaya karar vermemde bana verilen görev ve sorumluluklar ve güvenleri nedeni ile Milli Eğitim Bakanlığı Bilgi İşlem Genel Müdürümüz Sn. Özgür TÜRK' e, Daire Başkanımız Sn. Mustafa Bülent TURHAN' a, mesai arkadaşlarım ve yüksek lisansa başlangıcımдан tez sürecime kadar bana desteğini esirgemeyen değerli meslektaşım Adem KOÇ' a desteklerinden dolayı teşekkür ederim.

Bin bir zorlukla beni bu günlere getiren anne ve babama, varlıklarıyla beni her daim güçlü kılan sevgili kızlarım ve eşime teşekkür ederim.

Ümit ÇELİK
Afyonkarahisar 2024

İÇİNDEKİLER DİZİNİ

	Sayfa
ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER DİZİNİ.....	vi
KISALTMALAR	viii
ŞEKİLLER DİZİNİ	ix
ÇİZELGELER DİZİNİ.....	x
RESİMLER DİZİNİ	xi
EKLER DİZİNİ	xi
1. GİRİŞ	1
1.1 Problem Durumu	1
1.2 Çalışmanın Amacı	3
1.3 Problem Cümlesi	3
1.3.1 Alt Problemler	4
1.4 Çalışmanın Önemi	4
1.5 Çalışmanın Sınırlılıkları.....	4
2. LİTERATÜR BİLGİLERİ.....	5
2.1 e-Okul	5
2.2 LGS Süreci.....	9
2.3 Öğrenme Analitiği	12
2.4 Veri Madenciliği	13
2.5 Eğitsel Veri Madenciliği.....	14
2.6 Veri Madenciliği Modelleri	15
2.7 Sınıflandırma Kalitesi	23
2.8 İlgili Araştırmalar	24
3. MATERYAL VE YÖNTEM.....	28
3.1 Veri Seti	28
3.2 Öznitelikler	29
3.3 Veri Analizi	32
3.4 Özellik Seçimi	36
4. BULGULAR.....	40
4.1 Tercih Sayılarının Yerleşmeye Etkisine Dair Bulgular	40

4.2 Arařtırma Sorularına Yönelik Bulgular	40
5. TARTIřMA VE SONUÇ	50
6. KAYNAKLAR	53
EKLER	58
ÖZGEÇMİř.....	67



KISALTMALAR

Kısaltmalar

ABİDE	Akademik Becerilerin İzlenmesi ve Değerlendirilmesi
AİHL	Anadolu İmam Hatip Lisesi
AL	Anadolu Lisesi
AMP	Anadolu Meslek Programı (Meslek Lisesi)
AUC	Area Under the Curve (ROC Eğrisi Altında Kalan Alan)
EVM(EDM)	Eğitsel Veri Madenciliği
BÖTE	Bilgisayar ve Öğretim Teknolojileri Eğitimi
CA	Classification Accuary(Doğruluk)
DVM (SVM)	Destek Vektör Makinesi (Support Vector Machine)
İOKBS	İlköğretim ve Ortaöğretim Kurumları Bursluluk Sınavı
LGS	Liselere Giriş Sınavı
MCC	Matthews korelasyon katsayısı
MEB	Millî Eğitim Bakanlığı
MEBBİS	Millî Eğitim Bakanlığı Bilişim Sistemleri
OBP	Ortaöğretim Başarı Puanı
OKS	Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı
SBS	Seviye Belirleme Sınavı
VM(DM)	Veri Madenciliği
YBP8	8. Sınıf Yıl Sonu Başarı Puanı
YBP7	7. Sınıf Yıl Sonu Başarı Puanı
YBP6	6. Sınıf Yıl Sonu Başarı Puanı
YKS	Yüksek Öğretim Kurumları Sınavı
YSA(ANN)	Yapay Sinir Ağları

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 1.1 Geçmişten günümüze ülkemizde uygulanan orta öğretime geçiş sınavları...	2
Şekil 2.1 LGS Süreci.....	9
Şekil 2.2 Makine Öğrenmesi ve Eğitimsel veri madenciliği	13
Şekil 2.3 EVM süreci.....	14
Şekil 2.4 Veri madenciliği modelleri	16
Şekil 2.7 a- Basit YSA örneği, (b) Çok katmanlı ileri yayımlı YSA örneği	17
Şekil 2.8 k-NN algoritması örneği	19
Şekil 2.5 Örnek karar ağacı yapısı	21
Şekil 2.6 Destek vektör makineleri yapısı	22
Şekil 3.1 Cinsiyete göre mezun olunan ortaokullar	28
Şekil 3.2 Orange data mining ekran görüntüsü.....	32
Şekil 3.3 Veri analizi modeli	33
Şekil 3.4 Özellik seçim yöntemlerine göre değişkenler.....	38
Şekil 4.1 Lojistik regresyon algoritması karmaşıklık matrisi	42
Şekil 4.2 YSA karmaşıklık matrisi	43
Şekil 4.3 Naive Bayes karmaşıklık matrisi	44
Şekil 4.4 Rastgele Orman algoritması karmaşıklık matrisi	45
Şekil 4.5 kNN algoritması karmaşıklık matrisi.....	46
Şekil 4.6 Karar ağacı algoritması karmaşıklık matrisi.....	47
Şekil 4.7 DVM algoritması karmaşıklık matrisi	48
Şekil 4.8 Özellik seçim yöntemlerine göre önemli değişkenler	49

ÇİZELGELER DİZİNİ

	Sayfa
Çizelge 2.1 Karmaşıklık matrisi örneği	23
Çizelge 3.1 Cinsiyete göre mezun olunan okul dağılımı.....	29
Çizelge 3.2 Yerel tercihlerin örnek kodlaması	31
Çizelge 3.3 Çok sınıflı bir sınıflandırma problemi için bir sınıfa ait karmaşıklık matrisi	33
Çizelge 4.1 Tercih sayılarına göre yerleşen öğrencilerin dağılımı.....	40
Çizelge 4.2 Information gain yöntemine göre seçilen en önemli 10 değişken.....	41
Çizelge 4.3 Information gain yöntemi ile seçilen değişkenlere göre tahmin sonuçları	41

RESİMLER DİZİNİ

Resim 2.1	e-Okul Giriş Ekranı.....	5
Resim 2.2	e-Okul İlkokul Ortaokul Öğrenci işlemleri.....	6
Resim 2.3	e-Okul öğrenci genel bilgileri ekran görüntüsü	7
Resim 2.4	e-Okul baba bilgileri ekranı	7
Resim 2.5	e-Okul ortaöğretim kurum işlemleri ekranı	8
Resim 2.6	LGS Tercih ekranı	10

EKLER DİZİNİ

Ek 1	Veri seti başlıkları.....	58
Ek 2	Öğrencilerin tercih dağılımı.....	61
Ek 3	Araştırma İzni	66

1. GİRİŞ

Türkiye'deki e-Dönüşüm çabaları ve Kamu Yönetimi Reformu kapsamında, Şubat 2003 tarihinde başlatılan e-Dönüşüm Türkiye projesi, bilgi toplumu faaliyetlerini destekleme ve koordine etme amacını taşımıştır. Bu inisiyatifin bir parçası olarak, Millî Eğitim Bakanlığı merkezi ve taşra örgütleri ile tüm eğitim kurumlarını entegre eden bir çatı projesi olan Millî Eğitim Bakanlığı Bilişim Sistemleri (MEBBİS)' i geliştirmiştir. MEBBİS, bir web tabanlı yönetim bilişim sistemi olarak tasarlanmış ve e-Okul yazılımını da içermektedir. e-Okul yazılımı, öğrenci bilgileri, notlar, sosyal etkinlikler ve devamsızlık gibi önemli verileri dijital bir platformda entegre bir şekilde yöneten web tabanlı bir yazılımdır. Bu, eğitim süreçlerini daha verimli hale getirerek eğitim yönetiminde daha modern ve etkili bir yaklaşım sunar. MEB, bu yazılımı kullanıma açarak, 2007'den itibaren ilköğretim kurumlarına ve 2008'den itibaren ortaöğretim kurumlarına yaygın bir şekilde uygulanmasını sağlamaktadır. MEBBİS projesi ise eğitim kurumları arasında koordinasyonu artırmayı, veri yönetimini standartlaştırmayı ve geleneksel yöntemlerle ortaya çıkan uyumsuzlukları gidermeyi amaçlamaktadır. Bu tür bilişim sistemleri, merkezi veritabanları aracılığıyla tutarlılık sağlayarak bilgi paylaşımını kolaylaştırır ve eğitim yönetiminde daha etkili bir altyapı oluşturur (Bağlıbel vd. 2014).

Geride kalan 16 yılda e-Okul sisteminde önemli ölçüde bir veri kaynağı oluşmuştur. Bu tezin amacı da e-Okul sisteminde yer alan bu verilerden faydalanarak LGS tercihi yapan öğrenciler için rehberlik yapacak bir çalışma ortaya koymaktır.

1.1 Problem Durumu

Millî Eğitim Temel Kanununa göre eğitim ülkemizde örgün eğitim, okul öncesi eğitimi, ilköğretim, ortaöğretim ve yükseköğretim kurumlarını kapsar. Okul öncesi eğitimi, mecburi ilköğrenim çağına gelmemiş çocukların eğitimini kapsar. Bu eğitim isteğe bağlıdır. Mecburi ilköğretim çağı, 6-14 yaş grubundaki çocukları kapsar. İlkokulların birinci sınıflarına o yılın 31 Aralık tarihinde 72 ayını dolduran çocukların kaydı yapılır. Ancak çocuğun gelişim durumuna bağlı olarak okula erken başlaması veya kaydının ertelenmesi ile ilgili hususlar yönetmelikle düzenlenir. Ortaöğretim; ilköğretime dayalı

dört yıllık zorunlu örgün veya yaygın öğrenim veren genel, mesleki ve teknik eğitim kurumları ile mesleki eğitim merkezlerinin tümünü kapsar.

Şekil 1.1 de gösterildiği üzere ortaöğretime geçişte ülkemizde yıllara bağlı olarak farklı sistemler uygulanmıştır.



Şekil 1.1 Geçmişten günümüze ülkemizde uygulanan orta öğretime geçiş sınavları (Güler vd. 2019)

Şekil 1.1’ de gösterilen sistemler 1997 yılında yapılan düzenlemelerle LGS ile başlamış ve 2004-2005 eğitim öğretim yılına kadar uygulanmıştır. Bu eğitim öğretim yılında ise yerini Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı olarak isimlendirilen sınava bırakmıştır. OKS 2007-2008 eğitim-öğretim yılına kadar uygulanmış olup sonrasında SBS olarak uygulanmaya başlamıştır. 2007- 2008 eğitim-öğretim yılında yapılan Seviye SBS sadece 8. sınıfı tamamlayan öğrencilere uygulanmakla kalmayıp 6. ve 7. sınıfı tamamlayan öğrencilere de uygulanmıştır. 2010-2011 eğitim-öğretim yılında ise değişikliğe gidilmiş ve Seviye Belirleme Sınavı (SBS) bu kez sadece 8. sınıfı tamamlayan öğrencilere uygulanmıştır. 2013-2014 eğitim öğretim yılına kadar da bu sistem uygulanmış olup devamında Temel Öğretimden Ortaöğretime Geçiş Sınavı uygulamaya konulmuştur. En son 2017-2018 eğitim öğretim yılında uygulanan TEOG sistemi sona ermiş ve 2018-2019 eğitim öğretim yılından bu yana uygulanmakta olan sisteme

geçilmiştir. Bu sistemde nitelikli okullar ve adrese dayalı öğrenci alan okullar olmak üzere iki tip yerleştirme modeline geçilmiştir. Öğrencilerin yaklaşık %10 u yapılan LGS sınavı ile nitelikli okullara yerleştirilirken geri kalanı ise adrese dayalı sistemle okullara yerleştirilmektedir (Vural 2020).

Adrese dayalı yerleştirme işlemleri öğrencilerin ulusal adres veritabanındaki adres bilgileri esas alınarak yapılmaktadır. Bu kapsamda illerde öğrenci yoğunluğu dikkate alınarak kayıt alanları oluşturulmaktadır. Bu kayıt alanlarında öğrencilerin adreslerinin ilişkili olduğu ortaokullar ile ortaöğretim kurumlarından oluşan kayıt alanı havuzları oluşmakta ve öğrenciler bu kayıt alanına göre tercihlerini yapmaktadırlar.

Bu tez kapsamında öğrencilerin LGS tercihlerine göre bir ortaöğretim kurumuna yerleşip yerleşemeyeceği veri madenciliği yöntemleri ile ele alınacak ve öğrencilere önceden tercih rehberliği yaparak bir ortaöğretim kurumuna (lise) yerleşememe riskini en aza indirmeye çalışacak bir model ortaya koyabilme yönünde veri madenciliği yaklaşımları üzerinde durulacaktır.

1.2 Çalışmanın Amacı

Bu tezde problem durumunu ortaya çıkarabilecek en iyi veri madenciliği algoritmasının tespit edilmesi amaçlanmaktadır. Bunun için 8. Sınıf öğrencilerinin yıl sonu başarı puanları ve demografik verileri ve LGS tercihlerini kullanarak öğrencilerin yerleşecekleri okul türünün tahminlenmesi üzerine çalışılacaktır.

Ayrıca bu araştırma ülkemizde uygulanmakta olan 4+4+4 eğitim sisteminin son basamağı olan 8. Sınıfta eğitim görmekte ve lise eğitimini başlayacak öğrencilerin geleceğini şekillendireceği bir ortaöğretim kurumuna geçişte yapacakları tercihlerin sonucunda yerleşme durumlarının veri madenciliği yöntemleri modellenmesi ve başarılı bir model ortaya çıkması durumunda tercih rehberliği için kullanılabilmesidir.

1.3 Problem Cümlesi

Öğrencilerin akademik ve demografik verilerinin bir ortaöğretim kurumuna yerleşme durumlarının VM yöntemleri ile modellenerek tahminlenmesinin araştırılmasıdır.

1.3.1 Alt Problemler

- i. Farklı VM yöntemlerinin öğrencilerin bir ortaöğretim kurumuna yerleşmelerinin tahminlenmesindeki performansı nasıldır?
- ii. Özellik seçim yöntemlerine göre tahminleme için hangi değişkenler daha önemlidir?

1.4 Çalışmanın Önemi

Ülkemizde eğitim gören öğrencilerin geleceğini ve eğitim hayatını etkileyen iki büyük sınav yapılmaktadır. Bu sınavlar LGS ve üniversiteye giriş için yapılan Yükseköğretim Kurumları Sınavlarıdır. Bu sınavlardan LGS öğrencinin bir defa girebileceği bir sınav olup bu sınav sonucuna göre öğrenim göreceği liseyi seçen öğrenciler bu eğitimlerinin sonunda üniversite eğitimi görmek için YKS sınavına girmekte veya bir mesleğe adım atmaktadır. Gerek yükseköğretim yapacak gerekse mesleki yaşama geçecek öğrenciler için LGS son derece önemli bir sınavdır. Ancak bir ortaöğretim kurumuna yerleşmede bu sınav başlı başına belirleyici değildir. Bu sınav sonucunda öğrenciler LGS kapsamında yapılan merkezi sınav puanlarına göre nitelikli bir okula yerleşebilir veya ortaokuldaki ders başarısı ve yerleşim yerine göre sınavsız bir okula yerleşmektedirler (MEB 2022).

Bu tez kapsamında öğrencilerin akademik ve demografik verileri ile tercih verileri bir araya getirilerek durum tahminlemesi modellenmekte, bu sayede bir çeşit erken uyarı sistemi modeli ortaya konulmaya çalışılarak tercih yapan öğrencilere mevcut verileri ve yapacakları tercihler ile bir ortaöğretim kurumuna (liseye) yerleşip yerleşemeyeceği konusunda VM yöntemleri ile ne kadar cevap bulunabilir ve bu yöntemleri kullanırken hangi değişkenler daha etkilidir ortaya konmaya çalışılacaktır.

Bu tez ile öğrencilerin bir ortaöğretim kurumuna yerleşip yerleşemeyeceği ile ilgili bir model ortaya koyulmaya çalışılacaktır. Ayrıca eğitsel veri madenciliği kapsamında 18 yıllık bir veri arşivine sahip e-Okul sisteminde bulunan veriler aracılığı ile farklı çalışmalar konusunda da yeni fikirler ortaya koyabilecektir.

1.5 Çalışmanın Sınırlılıkları

2022 yılında bir ortaöğretim kurumuna yerleşmek isteyen yaklaşık bir milyon öğrenci bulunmaktadır. Bu öğrenciler Türkiye’ nin 81 ilinden tercih yapmaktadır. Her ilin kendi

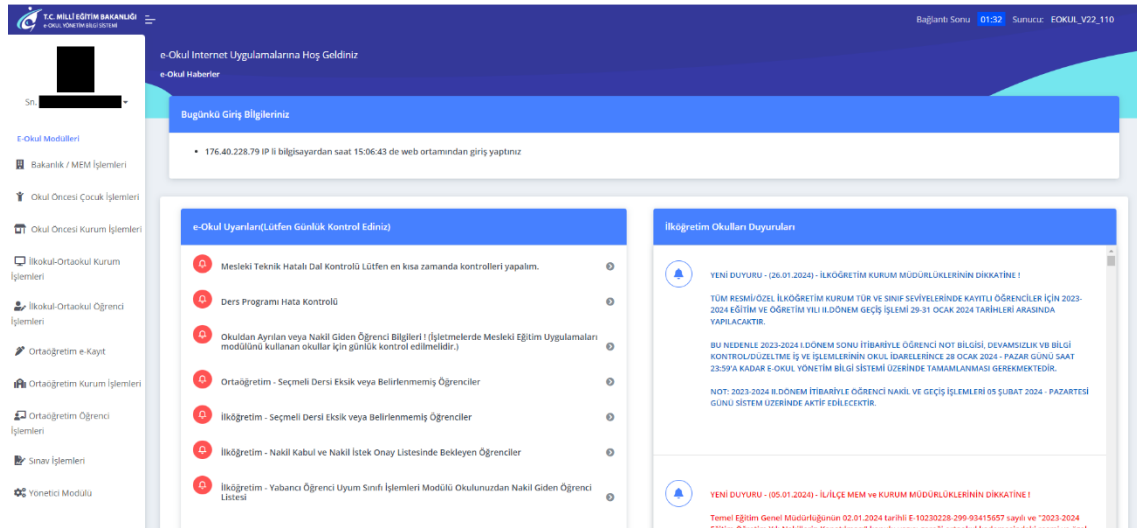
dinamiği, okulların ulaşım imkânı ve akademik başarı farkları söz konusudur. Bu anlamda tercih yapan öğrencilerin ve tercih yapılan okulların ulaşım kolaylığı açısından coğrafi verileri bulunmadığından tercihlerde bu etki gözlemlenmemektedir. Ayrıca tahminleme için kullanılacak verinin sadece Ankara ilinde bulunan 6000 öğrenciye ait verileri içermesi bu çalışmanın sınırlılıklarındandır.

2. LİTERATÜR BİLGİLERİ

Literatür bilgileri kapsamında e-Okul, LGS süreci, öğrenme analitiği, veri madenciliği, eğitsel veri madenciliği, veri madenciliği modelleri, sınıflandırma kalitesi ve ilgili araştırmalar alt başlıkları ele alınmıştır.

2.1 e-Okul

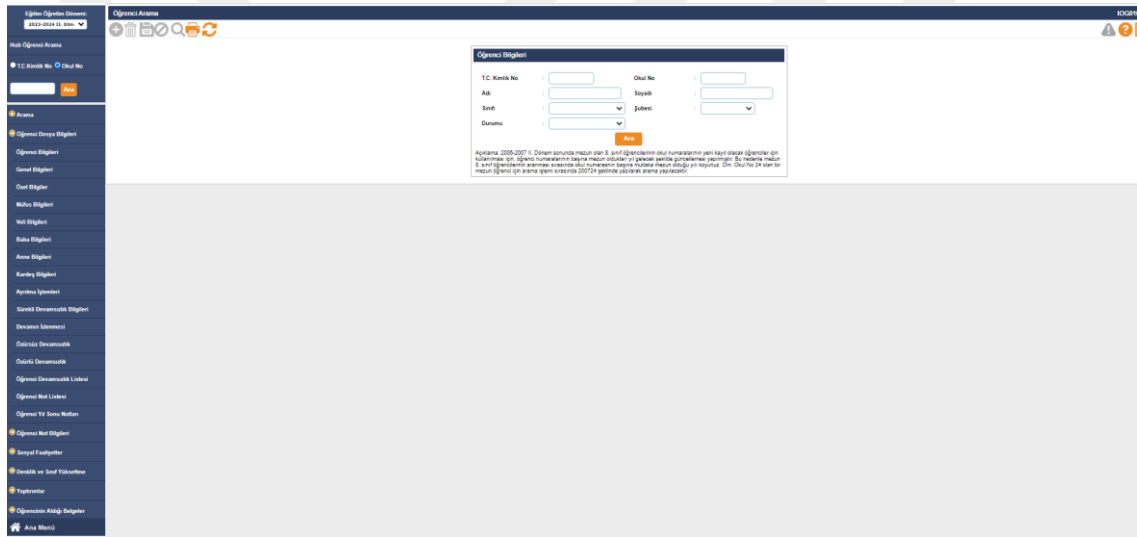
e-Okul MEB tarafından 2006 yılında geliştirilmeye başlanmış bir okul bilgi yönetim sistemidir. e-Okul sistemi içinde birçok modül barındıran web tabanlı bir otomasyon olup sisteme e-okul.meb.gov.tr internet adresinden giriş yapılmaktadır. Giriş işleminin ardından Resim 2.1’ de gösterilen ekran görüntüsündeki gibi kullanıcıları bir ekran karşılamaktadır. Bu ekranda kullanıcılar yetkisine göre menüleri ve duyuruları görmektedir.



Resim 2.1 e-Okul Giriş Ekranı.

Kullanıcılar Resim 2.1’deki menülerden eğer bir okulöncesi kurum ise “Okul Öncesi Çocuk İşlemleri”, “Okul Öncesi Kurum İşlemleri” ve “Sınav İşlemleri” menülerini

görmektedir. İlkokullar ve ortaokullar “İlkokul-Ortaokul Kurum İşlemleri”, “İlkokul Ortaokul Öğrenci İşlemleri” ve “Sınav İşlemleri” menülerini görmekte iken liseler ise “Ortaöğretim Kurum İşlemleri”, “Ortaöğretim Öğrenci İşlemleri” ve “Sınav İşlemleri” menülerini görmektedir. Okullar kademelerine göre yani okulöncesi, ilkokul, ortaokul ve lise olma durumlarına göre duyuru panelleri kullanıcıya gösterilmektedir. Ayrıca giriş ekranında bakanlık birimleri ve İl-İlçe Milli Eğitim Müdürlüklerinin yapacakları İşlemler için de “Bakanlık/MEM İşlemleri” menüsü yer almaktadır. e-Okul sistemindeki alt menülerde her eğitim kademesi için yer alan modüller eğitim kademelerine göre ayrılan bazı farklı modüller içerse de genel olarak aynı alt menüleri içermektedir. Örnek olarak “İlkokul Ortaokul Öğrenci İşlemleri” altındaki menüler Resim 2.2’ de paylaşılmıştır.



Resim 2.2 e-Okul İlkokul Ortaokul Öğrenci işlemleri.

Resim 2.2’ de paylaşılan “İlkokul Ortaokul Öğrenci İşlemleri” menüsünde birçok alt menü yer almakta olup burada öğrenciye ait birçok bilgi başlığı tutulmaktadır. Bunlar öğrencinin anne-baba bilgileri, veli bilgileri (veli anne baba dışında farklı birisi ise), devam-devamsızlık bilgileri, kardeş bilgileri, not bilgileri, almış olduğu belgeleri disiplin cezaları ve öğrenciye ait demografik bilgileri içeren Resim 2.3 de gösterilen genel bilgiler gibi veri başlıklarıdır.

Öğrenci Genel Bilgileri			
Kiminle oturuyor?	: Alisiyle	Geçirdiği kaza	: Kaza geçirmec
Oturduğu ev kira mı?	: Kendilerinin	Geçirdiği ameliyat	: Ameliyatı yok
Kendi odası var mı?	: Var	Kullandığı cihaz protez	: Cihaz protez y
Ev ne ile ısıtıyor?	: Kalorifer	Geçirdiği hastalık	: Hastalık geçirir
Okula nasıl geliyor?	: Alisiyle	Sürekli hastalığı	: Sürekli hastalı
Bir işte çalışıyor mu?	: Hayır	Sürekli kullandığı ilaç	: Kullandığı ilaç
Aile Dışında Kalan Var mı?	: Yok	Kardeş sayısı	: 1
Boy	: 130 cm	Diğer Sürekli hastalığı	: Sürekli hastalı
Kilo	: 25 kg		
☐ Yurtdışından Geldi Taşınma Nedeni :		☐ Burslu ☐ 2828 veya 5395 sayılı kanunlarına tabii	
Şartlı Eğitim Yardımı Uygulaması Başvuru No :			
Not: "Şartlı Eğitim Yardımı Uygulaması" kapsamında, yardım almayı hak kazanan öğrenci için başvuru numarası girilecektir. Yardım alma hakkını kaybeden öğrenci için bu başvuru numarasının silinerek kaydedilmesi yeterli olacaktır.			
Özel Eğitim Durumu	:	Açıklama : Özel eğitim ihtiyacı olan öğrencilerden Kaynaştırma/Butunleştirme/Özel Eğitim Sınıfı ve Özel Eğitim Okulunda eğitim alan öğrenciler için "Okulda eğitim alıyor" seçeneği seçilecektir.	
Hanede yaşayan kişi sayısı	: 4		
☐ Sporcu Lisansı var Öğrenci Aktif Dönemde Taşıma Uygulaması Kapsamında ☐		Taşınma Nedeni :	
Öğrenci Taşıma Planlama			
Taşıma Planlamaya Dahil Edilsin Mi? ☐		Taşınması Planlanan Eğitim Öğretim Yılı :	

Resim 2.3 e-Okul öğrenci genel bilgileri ekran görüntüsü.

Resim 2.3' de e-Okul sisteminde yer alan genel bilgiler ekranında öğrencinin kiminle oturduğu, oturduğu evin kiramı yoksa kendilerinin mi olduğu, kendine ait bir odasının olup olmadığı, okula nasıl gidip geldiği, herhangi bir hastalık geçirip geçirmediği, boy, kilo gibi bilgileri tutulmaktadır.

Öğrencinin anne, baba ve velisine ait bilgilere ait ekran görüntüsü ise Resim 2.4' de yer almaktadır.

Baba Bilgileri			
T.C. Kimlik No *****		Uyruğu T.C.	
Adı:			
Soyadı:			
Sağ / Diğer:		Birlikte / Ayrı:	
Sürekli Hastalığı:		Doğum Yeri:	
Engel Durumu:		Doğum Tarihi:	
Tel. (Ev):		Öğrenim Durumu:	
Tel. (Cep):		Mesleki:	
Tel. (İş):		E-posta Adresi:	

Resim 2.4 e-Okul baba bilgileri ekranı.

e-Okul sisteminde Resim 2.4 de verilen ekran görüntüsünde olduğu gibi anne, baba ve veliye ait bilgilerinde tutulduğu ekranlar yer almaktadır. Bu ekranlarda anne babanın sağ olup olmadığı, birlikte olup olmadığı, öğrenim durumu ve meslek gibi bilgileri tutulmaktadır.

e-Okul sisteminde öğrenci işlemleri menüleri altında genel olarak bu başlıklar yer alırken “Kurum İşlemleri” menüleri altında yer alan menüler için de “Ortaöğretim Kurum İşlemleri” menülerine ait ekran görüntüleri ise Resim 2.5’ de paylaşılmıştır.

Okul İşlemleri	
Kurum Kodu	Yaratılan Yıl
Kurum Adı	Durumu
Kurum Türü	Kapalı Kurum
İli	Adresi
İşlet	Telefon No
Bölge	Faks No
Köy	Web Sayfası
Şeyhlik Kodu	E-Posta
Kurum Tipi	Parayol Kapasite ETRİK
Kurum Müdürü	Parayol Kapasite KIZ

Resim 2.5 e-Okul ortaöğretim kurum işlemleri ekranı.

e-Okul siteminde Resim 2.5’ de gösterilen ekran görüntüsünde okulların işlemlerini yürüttüğü modül olan “Ortaöğretim Kurum İşlemleri” yer almaktadır. Bu ekranlarda okullar başlıca şube işlemlerini, nakil işlemlerini, not işlemlerini, devamsızlık girişlerini, karne işlemlerini, öğrenci fotoğraf işlemlerini ve mesleki okulların alan-dal işlemlerini gerçekleştirebilmektedirler.

2.2 LGS Süreci

Şekil 1.1 de gösterilmiş olan ortaöğretime geçiş sistemlerine göre halen uygulanmakta olan sınav ve ortaöğretime geçiş sistemi için Millî Eğitim Bakanlığı tarafından her yıl ortaokuldan bir ortaöğretime kuruma geçiş kılavuzu yayınlanmakta ve tüm süreç bu kılavuza göre yürütülmektedir. MEB (2022), “Ortaöğretime Geçiş Tercih ve Yerleştirme Kılavuzu” na göre süreç Şekil 2.1’ de gösterilmektedir.



Şekil 2.1 LGS Süreci (MEB 2022).

Şekil 2.1 de verilen LGS süreci öncelikle LGS’ na katılmak isteyen öğrencilerin başvuruları ile başlamaktadır. Bu süreç MEB tarafından her yıl oluşturulan “Sınavla Öğrenci Alacak Ortaöğretim Kurumlarına İlişkin Merkezî Sınav Başvuru ve Uygulama Kılavuzu” na göre yürütülmektedir. Sınava katılmak isteyen öğrenciler e-Okul sisteminden başvurularını yapmakta ve daha sonra kayıtlı oldukları okulları tarafından başvuruları onaylandıktan sonra sınav başvuru süreci tamamlanmaktadır. İkinci adımda yer alan LGS yapıldıktan sonra sınav sonuçları ilan edilmektedir. Öğrenciler bu sınavdan almış oldukları puan ile sınavlı okullara tercih süreçleri boyunca başvurabilmektedir. Dördüncü adımda yer alan sınavlı ve sınavsız okullara tercihlerin alınma sürecinde öğrenciler hem sınavlı hemde sınavsız okulları aynı anda tercih edebilmektedir. Tercihlerle ilgili örnek ekran görüntüsü Resim 2.6’ da paylaşılmıştır.

Yerel Yerleştirme Okul Tercihiniz

Tercihlerinizi değiştirilebilir. Tercihlerinizi kaydetmek için sayfayı (sac komutu) **Kaydet** butonuna basınız. Kaydet butonuna basmadan tercihleri yapılan hiçbir değişiklik **geçerli olmayacaktır**.

Tercih Sırası	Tercih Kodu	Okul	Okul Türü	Alan	Yabancı Dil	Öğretim Şekli	Kontenjan Tipi	Pansiyon	Öğretim Süresi	Kayıt Alanı	Boş Kontenjan	Yükarı Tık	Aşağı Tık
1	70831	ANKARALITINDAĞAĞahmet Hulusi Dursun Mesleki ve Teknik Anadolu Lisesi	Anadolu Mesleki Program	İngilizce	Normal Öğr.	KIZ/ERKEK	Pansiyon yok	4 yıl	Kayıt Alanında	33			
2	70628	ANKARALITINDAĞAĞaziz Mesleki ve Teknik Anadolu Lisesi	Anadolu Mesleki Program	İngilizce	Normal Öğr.	KIZ/ERKEK	Pansiyon yok	4 yıl	Kayıt Alanında	166			

Yerel Yerleştirme Okul Tercih Listesi

İ: İçe: Okul Türü: Kayıt Alanı:

70 Adet Tercih Listelendi

Tercih Kodu	Okul	Okul Türü	Alan	Yabancı Dil	Öğretim Şekli	Kontenjan Tipi	Pansiyon	Öğretim Süresi	Kayıt Alanı	Boş Kontenjan	Tercihleme Durumu
66684	ANKARALITINDAĞAĞahmet Yesevi Mesleki ve Teknik Anadolu Lisesi	Anadolu Mesleki Program	İngilizce	Normal Öğr.	KIZ/ERKEK	Pansiyon yok	4 yıl	Kayıt Alanında	9		Tercihleme Bile
70937	ANKARALITINDAĞAĞatındağ Mesleki ve Teknik Anadolu Lisesi	Anadolu Mesleki Program	İngilizce	Normal Öğr.	KIZ/ERKEK	Pansiyon yok	4 yıl	Kayıt Alanında	178		Tercihleme Bile
65795	ANKARALITINDAĞAĞanfatlar Mesleki ve Teknik Anadolu Lisesi	Anadolu Mesleki Program	İngilizce	Normal Öğr.	KIZ/ERKEK	Pansiyon yok	4 yıl	Kayıt Alanında	128		Tercihleme Bile
73581	ANKARALITINDAĞAĞankara Mesleki ve Teknik Anadolu Lisesi	Anadolu Mesleki Program	İngilizce	Normal Öğr.	KIZ/ERKEK	Pansiyon yok	4 yıl	Kayıt Alanında	140		Tercihleme Bile
69632	ANKARALITINDAĞAĞaziz Mesleki ve Teknik Anadolu Lisesi	Anadolu Mesleki Program	İngilizce	Normal Öğr.	KIZ/ERKEK	Pansiyon yok	4 yıl	Kayıt Alanında	0		Tercihleme Bile

Resim 2.6 LGS Tercih ekranı.

İlk tercihlerde 10 adet sınavlı okulla birlikte beş adet sınavsız öğrenci alan okul ve beş adet de sınavsız öğrenci alan pansiyonlu okul tercihi olmak üzere toplamda 20 adet tercih yapabilmektedir. Sınavlı okullara yerleştirme işlemleri öğrencinin LGS’ dan almış olduğu puana göre yapılmakta sınavsız okullara yerleştirme işlemleri ise İl Milli Eğitim Müdürlükleri tarafından oluşturulan kayıt bölgelerine göre adrese dayalı bir yerleştirme işlemidir. Yerleştirme işlemlerinde öğrenci LGS’ na katılmış ise ve bir yerleştirme puanı

var ise sınavlı okul tercih ekranları kendisine açılmaktadır. Eğer bir sınav puanı yok ise sınavlı okul tercihi yapamamaktadır. Sınavsız okul tercihi ise her öğrenci için ilk tercihlerde zorunludur. Sınavsız okul tercihi yapmayan bir öğrenci LGS puanı olsa bile sınavlı ve pansiyonlu bir okul tercihi yapamaz.

Sınavsız okullara tercih yaparken öğrenciler ilk üç tercihini kayıt alanlarından yapmak zorundadırlar ve aynı zamanda bir okul türünden en fazla üç adet tercih yapabilirler. Örneğin dört tane Anadolu Lisesi, dört tane Mesleki Lise veya dört tane Anadolu İmam Hatip Lisesi tercihi yapamaz. Üçten fazla tercih yaptığında mutlaka en az 2 farklı okul türü yapmalıdır. Kayıt alanları ise öğrencilerin adreslerine göre gidebilecekleri en yakın okullar gözetilerek oluşturulmaktadır.

Beşinci adımda ilk tercihlerin açıklanması süreci yer alır. Burada yerleştirme işlemlerinde öğrencilerin varsa önce sınavlı tercihlerine bakılır. Eğer öğrenci sınavlı bir okula yerleşiyorsa diğer sınavsız tercihlerine bakılmaz yani her ne kadar tercihler sırasında sınavsız tercihler zorunlu olsada öncelik sınavlı okul tercihleridir. Sınav puanına ve yaptığı tercihlere göre sınavlı bir okula yerleşemeyen öğrencilerin sırası ile önce sınavsız sonra da pansiyonlu tercihlerine bakılır ve hangisine yerleşiyorsa yerleştirme işlemi tamamlanmış olur.

Altıncı adımda öğrencilerden tekrar bir tercih alınır ve bu tercihlerde her tercih türünden (sınavlı, sınavsız, pansiyonlu) öğrenciler sadece üçer tercih yapabilirler. Bu yerleştirme işleminin adı “1. Nakil Yerleştirmesi” dir. Hiçbir okula yerleşemeyen öğrenciler ile birlikte ilk yerleştirmede yerleşen öğrencilerde bu tercih sürecinde tercih yapabilirler. Ancak nakil yerleştirmesinde kritik bir detay bulunmaktadır. Daha önce bir okula yerleşmiş olan öğrenciler herhangi bir kısıtlamaya tabi olmaksızın nakil tercihinde bulunabilmektedir. Yani ilk yerleştirmede olduğu gibi sınavlı bir okul tercihi yapmak için sınavsız okul tercihinde bulunmak, sınavsız okullardan belli sayıda kayıt alanındaki okullardan tercih yapmak gibi bir kısıtlama yoktur. Bu öğrenciler bu tercihlerine yerleşmeleri durumunda ilk yerleştirmedeki kazanmış oldukları okullardaki haklarından vazgeçmiş sayılarak yeni kazandığı okula yerleştirilirler. Ancak ilk yerleştirme sürecinde bir okula yerleşemeyen öğrenciler sınavlı bir okul tercih etmek için önce sınavsız tercih yapmak zorunda olup, sınavsız okul tercihlerinde ise ilk iki tercihini kayıt alanlarından

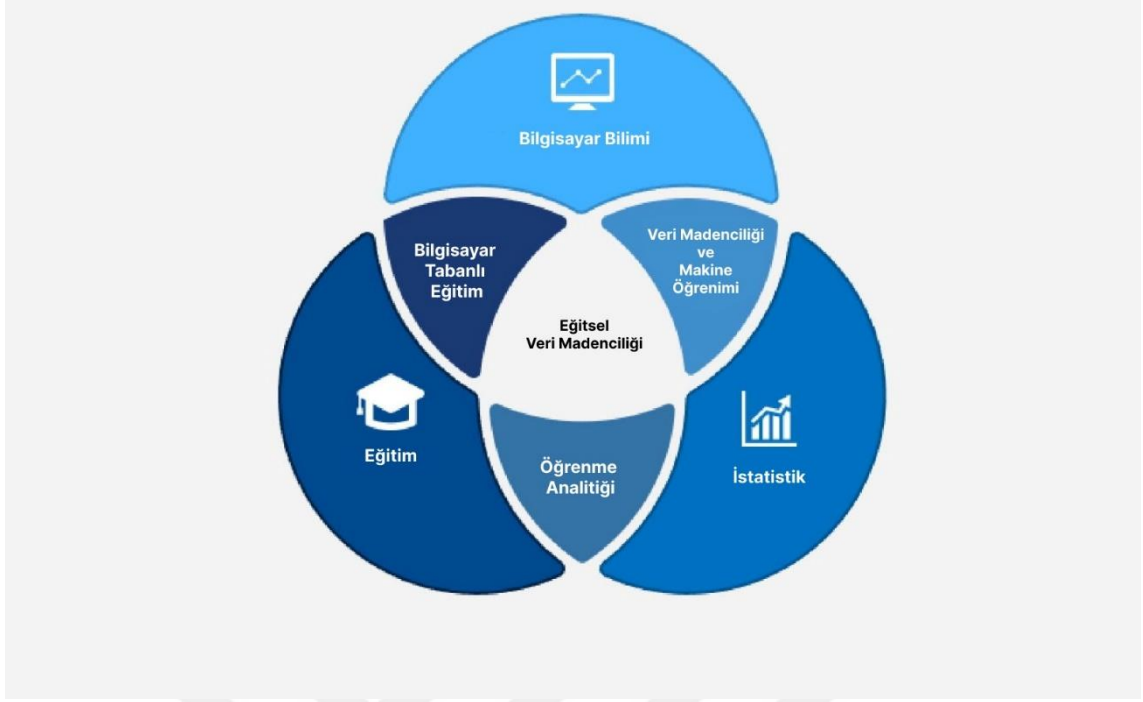
seçmek şartıyla bir okul türünden (Anadolu Lisesi, Mesleki ve Teknik Anadolu Lisesi, Anadolu İmam Hatip Lisesi) en fazla iki okul seçebilmektedirler.

Yedinci adımdaki “1. nakil tercih sonuçlarının açıklanması” sürecinden sonra sekizinci adımda “2. nakil yerleştirmesi” sürecine geçilir. Bu süreçte de yine hiçbir okula yerleşemeyen öğrencilerle birlikte ilk yerleştirme ve birinci nakil yerleştirmesi sürecinde bir liseye yerleştirmesi yapılmış olan öğrencilerin tercihleri birlikte alınır. Buradada yine daha önce bir okula yerleştirmesi yapılan öğrenciler herhangi bir tercih kısıtlaması olmadan tercih yapabilmekte iken bu sürece kadar herhangi bir okula yerleşemeyen öğrenciler “1. nakil yerleştirmesi” sürecinde olduğu gibi kısıtlamalara tabidir. Bu süreç tamamlandıktan sonra “2. nakil tercih sonuçlarının açıklanması” süreci ile öğrencilerden merkezi tercih alma süreci sona erer.

Yaklaşık bir aylık bir tercih ve yerleştirme süreci sonunda İl ve İlçe Milli Eğitim Müdürlüklerinde oluşturulan “Öğrenci Yerleştirme ve Nakil Komisyonları” yerleşemeyen öğrencilerin başvurularını alarak boş kalan okulları öğrenci ve velilere göstererek tercihlerini almakta ve bu okullara yerleştirmede kullanılan sıralama kriterleri gereği başarı durumları gözetilerek doğrudan okullara kayıtlarını yapmaktadırlar. Bu süreç ile birlikte yaklaşık bir buçuk aylık bir sürede öğrencilerin okullara yerleştirme işlemleri tamamlanmış olur.

2.3 Öğrenme Analitiği

Yeni bir alan olan öğrenme analitiği gelişmiş analiz araçlarının öğrenim ve öğretim amacıyla kullanıldığı bir alandır. Bu alan öğrenen ve öğrenme sürecinde elde edilen verilerin nasıl analiz ile ilgilenmektedir. İş zekâsı, web analitikleri, akademik analitikler, eğitsel veri madenciliği ve eylem analitiği gibi birçok alanla da ilişkilidir (Bozkurt 2016). Öğrenme analitiği, dijital eğitim ortamlarındaki verileri analiz eder ve öğrenme analitiğini "öğrenciler ve öğrenmenin gerçekleştiği ortamları anlamak ve optimize etmek için bağlamları hakkında verilerin ölçülmesi, toplanması, analizi ve raporlanması" olarak da tanımlanabilir (Ouhaichi vd. 2023). Öğrenme analitiği, eğitsel veri madenciliği ve diğer kavramların birbirleri ile ilişkisi Şekil 2.2 de ifade edilmiştir.



Şekil 2. 2 Makine Öğrenmesi ve Eğitimsel veri madenciliği (Pallathadka vd. 2023).

Şekil 2.2’ de eğitim, bilgisayar bilimi ve istatistik gibi üç disiplinin ortak noktası eğitsel veri madenciliğidir ve makine öğrenimi burada kritik bir öneme sahiptir. Bu üç disiplini bir araya getirerek eğitimde tahmin yapmaya olanak tanır (Pallathadka vd. 2023).

2.4 Veri Madenciliği

Veri madenciliğinin geçmişine indiğimizde ilk olarak 1960’lı yıllarda verilerin bilgisayar ortamında depolanmasıyla başladığı söylenebilir (Ledley 1960). Basit veri setlerinden başlanılarak ilişkisel ve neneye dayalı veri setlerine doğru bu süreç evrilmiştir. Özellikle 2000’ li yıllarda internetin yayınlaşması katlanarak artan bu verinin analiz gerekliliği ortaya çıkarmış ve bunun sonucu olarak da veri madenciliği kavramı ortaya çıkmıştır (Efron ve Tibshirani 1990). Veri madenciliği ile çeşitli algoritmalar ve bu algoritmaları kullanan çeşitli yazılımlar ortaya çıkmıştır. Bu yazılımlardan bazıları Weka, Orange, R ve Rapidminer olup bunun dışında Colab gibi online uygulamalar da mevcuttur (Ersöz 2017).

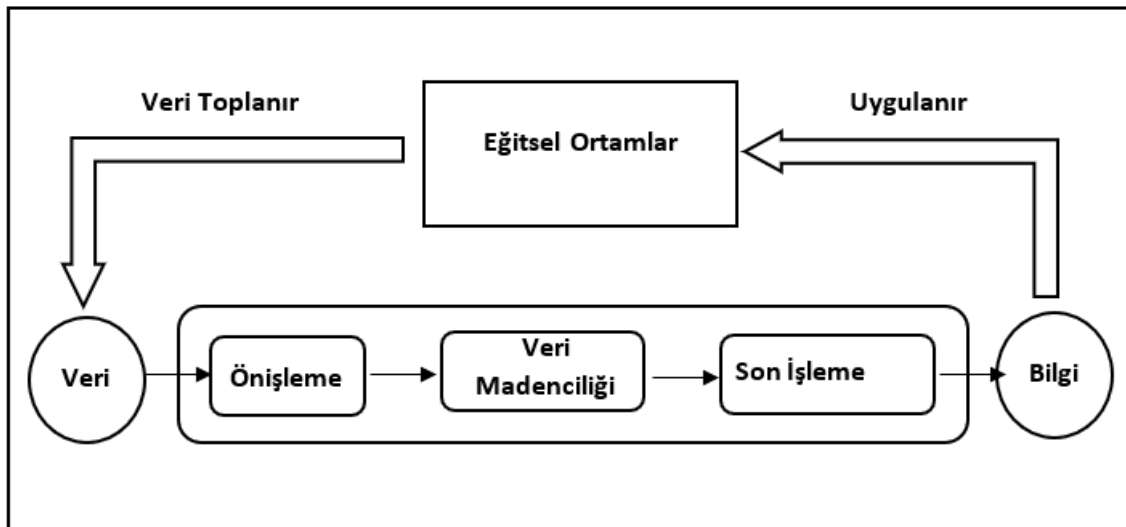
Savaş vd. (2012) veri madenciliğini işlenmemiş verinin tek başına ortaya koyamadığı bilgiyi çıkaran bir veri analizi süreci, büyük veri yığınları arasından geleceğe yönelik

tahminleme yapabilmemizi sağlayacak bağlantıların bilgisayar programı kullanarak aranması, veritabanı teknolojisi, istatistik, örüntü tanıma ve makine öğrenmesi ile önceden tahmin edilemeyen ilişkilerin ikincil analizi ve anahtar değişkenlerin binlerce potansiyel değişkenden izole edilmesi şeklinde toparlamıştır.

2.5 Eğitsel Veri Madenciliği

Veri madenciliği yöntemlerinin eğitim verilerinde kullanılmaya ihtiyaç duyulması ile eğitsel veri madenciliği (EVM) kavramını ortaya çıkarmıştır. Eğitim ile ilgili verilerle öğrencilerin daha iyi anlama ve daha iyi öğrenmeleri için çalışma yapılabilmesi amacıyla veri madenciliği teknikleri yapılmasına eğitsel veri madenciliği denilmektedir (Romero ve Ventura 2007). Eğitimle ilgili veriler okullar, e-öğrenme ortamları ve akademik çalışmalar sonucu ortaya çıkmıştır. Bu analiz yöntemi 2000-2007 yılları arasında popülaritesini arttırmıştır (Ersöz 2017).

EVM, öğrenci ve öğrenme ortamlarının daha iyi anlaşılması için öğrenme ortamlarından toplanan veriden faydalanılması sürecidir (Siemens ve Baker 2012). Bu süreçte toplanan ham veriler yazılımların, öğretmenlerin ve araştırmacıların kullanabileceği hale getirilmekte olup süreç Şekil 2.3 de gösterilmiştir.



Şekil 2.3 EVM süreci (Akçapınar 2014).

EVM süreci eğitsel ortamlardan bilginin elde edilmesi ile başlayan bir süreç olup tezimizin verileri e-Okul sisteminden alındığından eğitsel ortamımız olarak e-Okul olarak

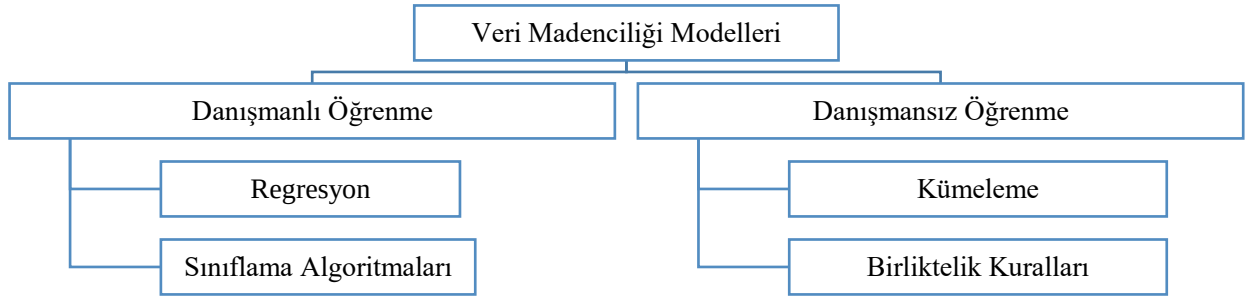
örneklendirilebilir. Ek-3 de yer alan araştırma izni ile tezimizde kullandığımız veriler Milli Eğitim Bakanlığı Bilgi İşlem Genel Müdürlüğü'nden 13.06.2023 tarih ve 78199263 sayılı resmi yazı ile elde edilmiştir. Bu yazı ile elde edilen veriler sürecimizdeki veri basamağını oluşturmakta ve öğrencilerin demografik verileri ile LGS tercih verilerini içermektedir. Önişleme sürecinde veriler analize hazır hale getirilerek kalitesi artırılmaktadır. Tez verilerimizde doğrudan işlenecek veri e-Okul sistemindeki veriler olduğundan ve yapılacak işleme uygun bir şekilde veri hazırlandığından önişleme sürecinde çok fazla işleme gerek kalmamıştır. Veri madenciliği süreci ile yöntemleri uygulanmakta ve sonuçların yorumlanması sürecine geçilir (Akçapınar 2014).

EVM, eğitim alanındaki veri madenciliği uygulamalarını ifade eder ve aynı zamanda öğretme-öğrenme sürecini geliştirmek için öğrencilerin akademik performansı, okulu bırakma oranları ve öğretmenlerin performansı gibi konuları sınıflandırmak, analiz etmek ve tahmin etmek için kullanılmaktadır. Bu teknikler, öğrencilerin davranışlarını ve öğrenme süreçlerini daha iyi anlamak için kullanılır. Ayrıca, öğretmenlerin öğrencilerin öğrenme ihtiyaçlarına göre daha iyi bir öğrenme ortamı yaratmalarına yardımcı olur. Bu nedenle, EVM, eğitim alanındaki verileri daha anlamlı ve kullanışlı hale getirmek için önemli bir araçtır (Dol ve Jawandhiya 2023).

Eğitimde teknoloji kullanımının sağladığı en önemli olanaklardan biri öğrencilere ait katılım ve etkileşim bilgilerinin bir veritabanında saklanmasıdır. Bu verilerin barındırdığı örüntülerin keşfedilmesi eğitsel veri madenciliği konu alanına girmekte olup keşfedilen bu örüntülerin eğitim planlamasında (öğrenme analitikleri) kullanılması günümüzde popüler bir konu haline almıştır. Bu süreç problemin ortaya konması, veri setinin tanımlanması, veriyi işleme, dönüştürme, analiz, değerlendirme ve sunum işlem basamaklarına sahiptir (Han ve Kanber 2006). Eğitim sistemlerinde kullanılan veri madenciliği teknikleri genellikle sınıflandırma, kümeleme, istatistik, tahminleme, metin madenciliği, görselleştirme gibi tekniklerdir (Romero ve Ventura 2007).

2.6 Veri Madenciliği Modelleri

Veride madenciliği teknikleri hedef değişken olup olmama durumuna göre danışmanlı ve danışmansız olmak üzere Şekil 2.4 de gösterildiği gibi iki ana başlıkta sınıflandırılmaktadır.



Şekil 2.4 Veri madenciliği modelleri (Şimşek 2022).

Şekil 2.4 de verilen veri madenciliği modellerinden danışmanlı öğrenme modelinde öğrenen bir danışman yardımı ile olayı öğrenir (Bilgin ve Şentürk 2019). Danışmansız öğrenmede ise veriler özetlenerek öğrenme yapılmaktadır (Akşehirli vd. 2014). Bu araştırmada kullanılan veri madenciliği yöntemleri aşağıdaki gibi özetlenmiştir.

Lojistik Regresyon:

Lojistik regresyon analizinde bağımlı yani yordayıcı değişkenin yapısı analizin sınıflandırılma sürecinde önemlidir. Lojistik Regresyon Analizi yordayıcı değişkenin yapısına göre üçe ayrılmaktadır. Kategorik yordayıcı değişkenin ikili olduğu durumlarda “İkili Lojistik Regresyon” kullanılır. Örneğin cinsiyet (kadın-erkek), medeni durum (bekar-evli) gibi değişkenlerin bulunduğu durumlardır. Kategorik yordayıcı değişkenin çok kategorili olduğu durumlarda “Multinomial Lojistik Regresyon Modeli” kullanılır. Örneğin eğitim durumu (ilkokul, ortaokul, lise, üniversite), gelir durumu (kötü, orta, iyi) gibi değişkenlerin bulunduğu durumlardır. Çok kategorili ve sıralı bir yapı olduğu durumlarda “Sıralı (Ordinal) Lojistik Regresyon Modeli” kullanılır. Likert tipi ölçekler buna örnek gösterilebilir. Lojistik regresyon analizinin bağımsız yani yordanan değişken sayısına göre sınıflandırıldığı durumlar da ise tek bağımsız değişken olması durumunda “Tek Değişkenli Lojistik Regresyon”, iki veya daha fazla bağımsız değişken bulunması durumunda ise “Çoklu Lojistik Regresyon Analizi” olarak adlandırılmaktadır (Şenel vd. 2014).

Lojistik regresyon için çeşitli örneklendirmeler yapılacak olursa spam e-posta, diyabet tahmini belirli ürün alım tahminlemeleri şeklinde yapılabilir. Yapısı nedeni ile uygulaması kolay denilebilecek bir yöntem olup en yüksek olasılığa göre tahminleme

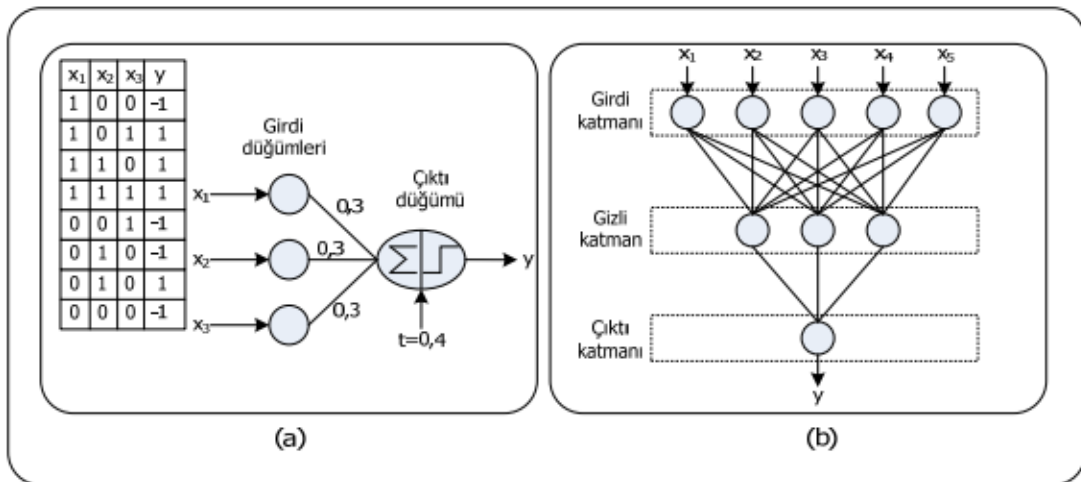
yapılır. Kategorik özellikler fazla ise işlemede sıkıntı yaşanabilir ancak genelde yüksek hesap gücü gerekmez (Can 2021).

Yapay Sinir Ağları:

Yapay sinir ağları (artificial neural networks- ANN), adındaki benzerlikten anlaşılacağı üzere insan beyninin fizyolojisi model alınarak geliştirilmiş bir veri madenciliği modelidir. Çalışma mantığı bu şekilde kurgulanmıştır. Bir yapay sinir ağı modeli katman ve hesaplama elemanlarından oluşur. Bu katmanlar n adet olarak ifade edilmekte olup her katmanda farklı sayılarda hesaplama elemanı bulunmaktadır. Bu hesaplama elemanları ise yapay sinir hücresi, düğümler ve işlem elemanları olarak adlandırılmaktadır (Topuz 2021).

Sınıflama ve regresyon analizleri için lojistik veya doğrusal fonksiyonları kullanır. Uzman olmayan kişilerce ifade edilmesi zordur. Daha çok ses tanıma, görüntü işleme ve örüntü tanıma gibi işlemlerde tercih edilmektedir (Akçapınar 2014).

Başka bir ifade ile yapay sinir ağları karmaşık hesaplamaları gerçekleştiren biyolojik sinir sistemlerinin bir simülasyonu olarak ifade edilebilir ve burada öğrenme sinir hücreleri arasındaki etkileşim ile sağlanmaktadır. Yapay sinir ağları denetimli ve denetimsiz öğrenme için kullanılmaktadır. Yapay sinir ağları ile ilgili örnek modelleme Şekil 2.5 de gösterilmiştir.



Şekil 2.5 a- Basit YSA örneği, (b) Çok katmanlı ileri yayımlı YSA örneği (Aydın 2007).

Şekil 2.5 de gösterilen yapı giriş ve çıktı düğümlerinden meydana gelirken bu düğümlerin arasındaki bağlantılar sinaptik bağlantıları ve her girdinin ağırlığını gösterir. Şekildeki T değeri öngörü değeridir. Şekildeki örnekte;

$$\hat{y} = \text{sign} (0,3x_1 + 0,3x_2 + 0,3x_3 - 0,4) \quad (2.1)$$

ifade edilir ve bu örnekteki katsayılar model öğrenirken en iyi değer tekrarlı bir algoritma ile hesap edilmeye çalışılır (Aydın 2007).

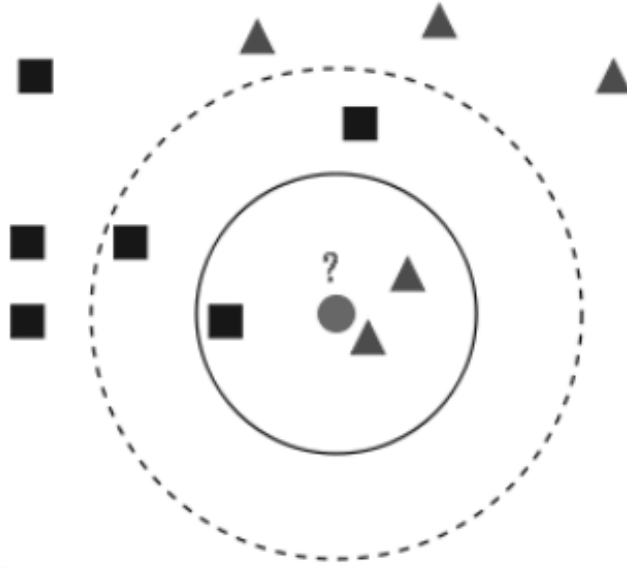
k-NN (k-En yakın Komşu):

k-NN algoritması, önceden belirlenmiş bir numune kümesinde bulunan belirli sınıflara ait numunelerin gözlem değerlerini kullanarak, yeni bir gözlemin hangi sınıfa ait olduğunu belirlemek için kullanılan bir sınıflandırma yöntemidir.

Bu yöntem, sonradan belirlenen bir gözlem değeri için, bu değeri içeren numune kümesindeki her bir gözlemin arasındaki mesafeyi hesaplar. Ardından en yakın K sayıda gözlemi seçer ve bu gözlemlerin hangi sınıfa ait olduğunu belirler.

Bu işlemde Öklit mesafesi formülü kullanılabilir. Öklit mesafesi, iki nokta arasındaki uzaklığın hesaplanması için yaygın olarak kullanılan bir metriktir. Aşağıdaki Öklit mesafesi formülü, hesaplanacak "x" ve "y" noktaları için kullanılabilir.

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (2.2)$$



Şekil 2.6 k-NN algoritması örneği (Ajaj 2021).

Şekil 2.6 da gösterilen k-NN algoritması örneğinde kareler ve üçgenlerden oluşan iki sınıf bulunmaktadır. Düz çizgili olan birinci çemberde sınıfını belirlemek istediğimiz verimiz bulunmaktadır. $K = 3$ olarak seçildiğinde, çemberimize yakın olan iki üçgen ve bir kare olduğundan üçgen sınıfını seçmemiz gerekir. Ama $K=5$ olarak seçersek, çemberimize üç kare ve iki üçgen yakın olduğu için kare sınıfını seçmemiz gerekir. Bu sebeple, K için belirlenen değer çok önemlidir (Ajaj 2021).

Naive Bayes:

Naive Bayes algoritması, 18. yüzyılda yaşamış olan Thomas Bayes 'in geliştirdiği Bayes teoremine dayalı bir sınıflandırıcıdır. Bayes teoremi temelde, herhangi bir B olayının koşullu bir A olayı ile A olayının koşullu bir B olayı olasılıklarının birbirinden farklılığının ilişkisini ifade eder. Bu teorem, Naive Bayes algoritmasının temel matematiksel çerçevesini oluşturan prensiplerden biridir.

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (2.3)$$

Burada, $P(A | B)$, B olayının gözlemlendiği durumda A olayının koşullu olasılığını temsil eder. $P(A)$ A olayının önceden bilinen (marjinal) olasılığıdır ve $P(B)$ B olayının önceden bilinen (marjinal) olasılığını temsil eder. $P(B | A)$ ise A olayının gözlemlendiği durumda B olayının koşullu olasılığını ifade eder. Bu denklemin aracılığıyla, gözlemlenmiş bir B

olayının varlığı durumunda, A için B' nin koşullu olasılığı hesaplanabilir. Koşullu bağımsızlık varsayımı gerçek dünyada nadiren sağlansa da bu varsayım aynı zamanda algoritmanın öğrenme sürecini hızlandırabilir.

Naive Bayes sınıflandırmasında, eğitim verilerinden elde edilen olasılıklara dayalı bir algoritma oluşturulur. Bu algoritma, test verisi veya yeni gelen verideki özelliği olasılıklara dayalı olarak sınıflandırır. Sınıflandırmanın amacı, bilinen bir bağımsız değişken vektörü ile bağımlı değişkenin değerini tahmin etmektir. Bağımlı değişkenin tahmini, bayes olasılığının en büyük değeri taşıyan sınıfın seçilmesiyle gerçekleştirilir. Yukarıdaki formülde yer alan A değeri bağımsız bir değişken olarak kabul edildiğinde, daha önce gözlemlenen B olayının gerçekleşme durumu maksimize edilir. Bu maksimizasyon, yine formüle dayanarak, B olayının A'da gerçekleşme olasılığının yani $P(B | A)$ 'nin en yüksek değeri elde edilerek gerçekleştirilir. Bu nedenle sınıflandırma için en büyük sonsal değer (Maximum A Posteriori, MAP) seçilmelidir. En büyük sonsal değer;

$\arg \max \{P(B | A)P(A)\}$ şeklinde ifade edilir (Topuz 2021).

Genellikle "Naive Bayes" olarak anılan sınıflandırma algoritması, temelde Bayes teoremine dayalı basit bir olasılık tabanlı sınıflandırma metodudur. Bu yöntem, tahmin temelli modellemeler için basit ama güçlü bir algoritma olarak kullanılmaktadır. Bu nedenle özellikle sinyal işleme ve görüntü işleme gibi alanlarda sıkça kullanılan sınıflandırma ve tahmin algoritmalarından biridir. Bu algoritmada, bir sınıftaki belirli özelliklerin diğer tüm özelliklerle ilişkili olmadığı varsayımı yapılmaktadır (Harman 2021).

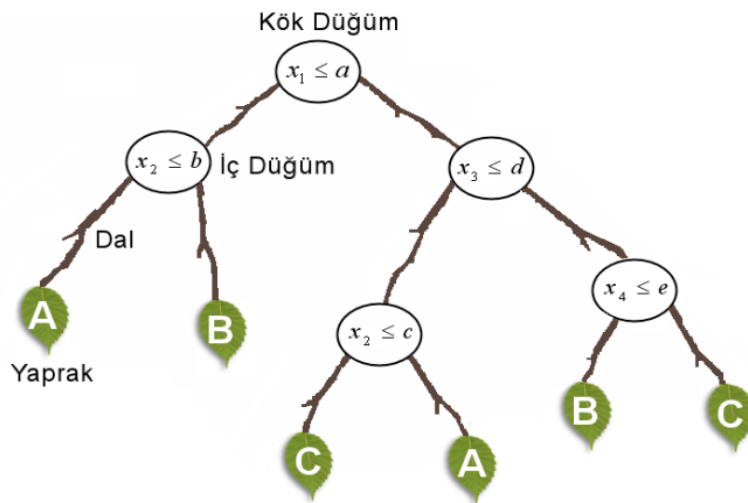
Karar Ağaçları:

Sınıflama algoritmaları içinde en çok tercih edilen algoritmaların başında gelir. Bunun sebebi ise kolayca oluşturulması ve yorumlanabilmesidir (Friedl ve Brodley 1997). Karar ağaçlarının tahmin edici özellikleri olduğu gibi tanımlayıcı özellikleride mevcuttur. Sınıflama algoritmaları içinde sıkça tercih edilmesinin nedenleri olarak veritabanlarına kolayca entegre edilebilmeleri, yüksek güvenilirliğin olması gibi nedenler gösterilebilir.

Düzgün (2017) e göre Karar ağaçları olası senaryoları karar vericiye grafiksel olarak sunabilen bir tekniktir.

Karar ağacı algoritmaları verileri analiz ederek kurallar oluşturur ve bu kurallara göre ağaç veri yapısı şeklinde bir şema oluşturur. Bu algoritmalar yüksek performanslı algoritmalardır. Bu algoritma yapı itibarı ile bir ağacı andırır ve kök, düğüm, dal ve yapraklardan oluşur. Karar ağaçları, karar verme adımları uygulanarak büyük veri setlerini küçük parçalara bölerek sonuca ulaşmayı temel alır. Büyük boyutlu verileri basit karar mekanizmaları ile küçük kayıt gruplarına bölerek kullanılan bir yapıdır (Sönmez 2015).

Örnek karar ağacı yapısı Şekil 2.7 de gösterilmiştir.



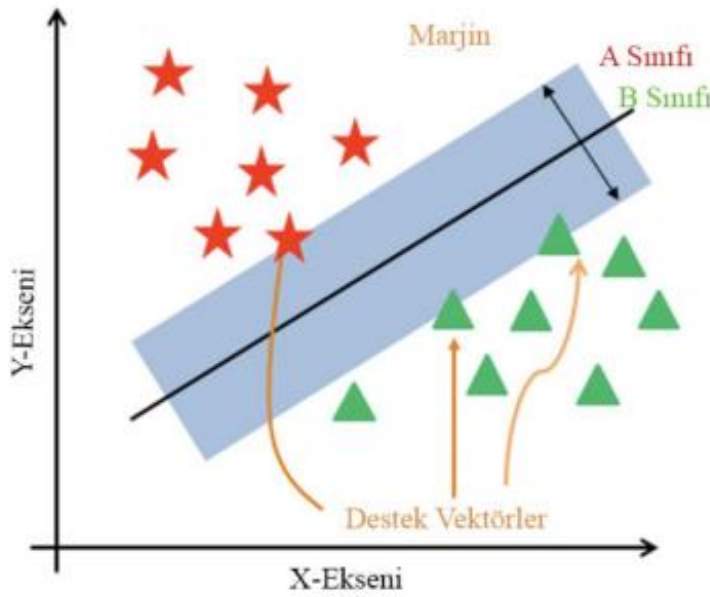
Şekil 2.7 Örnek karar ağacı yapısı (Kavzoğlu ve Çölkesen 2010).

Destek Vektör Makineleri (SVM):

İlk defa Vapnik tarafından 1995 yılından sınıflandırma ve regresyon tip problemlerin çözümü için önerilen destek vektör makineleri güçlü istatistiki teoriler üzerine kurulmuş bir makine öğrenmesi yöntemlerinden biridir. Destek vektör makineleri diğer makine öğrenme yöntemlerinin daha fazla miktarda eğitim verisi ihtiyacının olması, düşük yakınsama oranı ve eksik-fazla uyum problemleri yaşamalarının aksine risk minimizasyonu temelinde çalışarak daha iyi bir seçenek olarak öne çıkmıştır. Destek

vektör makineleri yüksek boyut ve az sayıda veri içeren uygulamalarda daha başarılıdır. (Güner ve Çomak 2011).

Destek vektör makineleri çok boyutlu bir uzayda farklı sınıflardan hiperdüzlemler oluşturarak çalışan bir sınıflandırma metodudur. Doğrusal olmayan karar sınıfları oluşturmak suretiyle regresyon ve sınıflandırma işlemlerini yaparak sınıflandırma ve regresyon işlemlerini yaparken büyük esneklikler sağlar (Kuzey 2012). Destek vektör makinelerinin yapısı Şekil 2.8 de gösterilmiştir.



Şekil 2.8 Destek vektör makineleri yapısı (Can 2021).

Rastgele Orman:

İstatistik ve makine öğrenimi disiplinlerinde sıklıkla kullanılan bir algoritmadır ve geniş bir veri yelpazesine uygulanabilir. Bu model, birçok karar ağacı modelinden oluşur. Her bir ağaç, veriye özgü sınıf tahminlerini üretir ve sonuç olarak, tüm ağaçların tahminlerinin oylaması sonucunda en fazla oy alan sınıf rastgele orman modelinin tahminini belirler. Bu yaklaşım, genellikle yüksek doğruluk ve güvenilirlik sağlama amacıyla tercih edilir (Breiman 2001).

2.7 Sınıflandırma Kalitesi

Veri madenciliği bağlamında, sınıflandırma algoritmalarının etkinliğinin değerlendirilmesi, hangi yöntemin diğerine göre daha üstün olduğunun belirlenmesi amacıyla büyük bir önem taşır. Sınıflandırma işleminin değerlendirilmesi için temel bir ölçüt, doğru ve yanlış sınıflandırılan gözlemlerin sayılarının incelendiği karışıklık matrisi (confusion matrisi) ile gerçekleştirilir. Bu matris, sınıflandırma sonuçlarının gerçek değerlerle karşılaştırılmasına dayanarak algoritmanın performansının detaylı bir şekilde analiz edilmesini sağlar (Altunkaynak 2017).

Çizelge 2.1 Karmaşıklık matrisi örneği (Topuz 2021) .

		Gerçek Durum	
		kazandı (+)	kazanamadı (-)
Yöntemin Tahmini	kazandı (+)	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	kazanamadı (-)	Yanlış Negatif (YN)	Doğru Negatif (DN)

Çizelge 2.1’ de yer alan karışıklık matrisinde yer alan karmaşıklık matrisini veri setimizdeki hedef değişkenimiz olan ve “kazanamadı, sınavlı ve sınavsız” olarak üç farklı değer alan YERLESTIRME_GRUBU üzerinden örneklendirdiğimizde hiçbir okula yerleşemeyen öğrencileri kazanamadı, sınavlı ve sınavsız bir okula yerleşenleri ise kazandı olarak iki boyutta kategorize edersek;

Doğru Pozitif (DP) : Bir liseye yerleşen kazandı olarak sınıflandırdığımız ve gerçekte de kazanan öğrenciler.

Yanlış Pozitif (YP) : Bir liseye yerleşen kazandı olarak sınıflandırdığımız ancak gerçekte kazanamayan yani bir liseye yerleşemeyen öğrenciler.

Yanlış Negatif (YN) : Bir liseye yerleşemeyen kazanamadı olarak sınıflandırdığımız ancak gerçekte kazanan yani bir liseye yerleşen öğrenciler.

Doğru Negatif (DN) : Bir liseye yerleşemeyen kazanamadı olarak sınıflandırdığımız ve gerçekte de kazanamayan yani bir liseye yerleştirmesi yapılmayan öğrenciler.

Bu değerler çeşitli formüller aracılığıyla farklı doğruluk ölçütlerine dönüştürülebilir. Bu dönüşümler arasında en temel ve basit olanı doğruluk (accuracy) oranıdır. Doğruluk oranı, sınıflandırma sonuçlarının toplam doğru tahmin edilen gözlem sayısına bölünmesi ile hesaplanır. Bu oran, sınıflandırma algoritmasının ne kadar doğru sonuçlar ürettiğini değerlendirmek amacıyla kullanılır.

$$\text{Doğruluk} = DP + DN / N \quad (2.4)$$

Doğru sınıflandırılmış gözlem sayısının tüm gözlem sayısına bölünmesi ile hesaplanan doğruluk oranı, bir algoritmanın doğru sınıflandırma yeteneği ve kalitesi hakkında değerli bir bilgi sunar. Bu oran, algoritmanın gerçek değerleri ne kadar doğru tahmin ettiğini ölçmekte kullanılır. Ayrıca, doğruluk oranını 1'e tamamlayan hata oranı da algoritmanın yanlış sınıflamaları hakkında fikir verir (Topuz 2021). Bu metriklerin yanı sıra, karışıklık matrisinden türetilen değerler kullanılarak hesaplanan duyarlılık (sensitivity), özgüllük (specificity), kesinlik (precision), F ölçüsü ve Matthews korelasyon katsayısı gibi metrikler de kullanılmaktadır. Bu metrikler, sınıflandırma algoritmasının performansını daha ayrıntılı bir şekilde değerlendirmemize yardımcı olur. Duyarlılık, algoritmanın gerçek pozitifleri ne kadar iyi tespit ettiğini, özgüllük ise gerçek negatifleri ne kadar doğru belirlediğini gösterir. Kesinlik, pozitif tahminlerin ne kadar doğru olduğunu yansıtırken, F ölçüsü hem hassasiyeti hem de kesinliği birleştirerek dengeli bir değerlendirme sağlar. Matthews korelasyon katsayısı ise sınıflandırma sonuçlarının gerçek değerlerle ne kadar iyi ilişkilendiğini ifade eder (Coşkun ve Baykal 2011).

$$\text{Duyarlılık} = DP / (DP + YN)$$

$$\text{Özgüllük} = DN / (YP + DN) \quad (2.5)$$

$$\text{Kesinlik} = DP / (DP + YP)$$

$$F \text{ değeri} = 2 * \text{duyarlılık} * \text{özgüllük} / (\text{duyarlılık} + \text{kesinlik})$$

2.8 İlgili Araştırmalar

Akçapınar (2014), çevrimiçi öğrenme ortamındaki etkileşim verileri kullanılarak öğrencilerin akademik performanslarının veri madenciliği yöntemleri ile modellenmesi

üzerine bir doktora tezi çalışma yapmış ve bu çalışmasını 2013-2014 güz döneminde Hacettepe Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi bölümü 2. sınıf öğrencilerinden Bilgisayar Donanımı dersine kayıtlı 76 ile öğrenci gerçekleştirmiştir. 14 haftalık bir veri toplama sürecinde öğrenciler yüz yüze derslerle birlikte ders dışı etkinlikleri gerçekleştirmek için araştırmacılar tarafından tasarlanan ve geliştirilen çevrimiçi öğrenme ortamını kullanmışlardır. Bu ortamdan toplanan veriler ile 28 adet değişken belirlenmiş ve sonuçta geçti veya kaldı olarak değerlendirmeye tabi tutulmuşlardır. Araştırma sorularının cevaplanması için yapılan analizlerde CN2 kuralları ve kNN %86 lık bir doğrulama oranı ile geçen-kalan öğrencilerin tahminlemiştir. Daha erken evrede öğrenci başarı tahminleme için yapılan çalışmalarda ise üçüncü haftada %74 lük bir oranla öğrencilerin geçeceği veya kalacağı tahminlemiştir. 14 haftalık süreç göz önüne alındığında üçüncü haftada bu tahminleme değerine ulaşılması oldukça erken sayılacak bir evrede iyi bir tahminleme değeri yakalandığını göstermektedir.

Kaur vd. (2025), veri madenciliği yöntemleri ile öğrenciler arasında yavaş öğrenenleri tahminlemeye ve sınıflandırmaya yönelik çalışmasında açık kaynaklı Weka yazılımı ile farklı sınıflandırma algoritmalarını kullanmış çok katmanlı algılayıcılar yöntemi %75 performansla en iyi sonucu vermiştir.

Özdemir (2016), İstanbul'da sosyo-demografik farklılıklara sahip farklı ilçelerdeki liselerden 2371 öğrenciden demografik değişkenler, (yaş, cinsiyet, İstanbul'da ikamet süresi, anne-baba birlikteliği, annenin eğitim durumu, babanın eğitim durumu, annenin çalışma durumu, babanın çalışma durumu, algılanan maddi gelir düzey, günlük ortalama ders çalışma süresi, günlük ortalama internet kullanım süresi, günlük ortalama televizyon izleme süresi, eğitim hayatında sınıf tekrarı yapmış olma durumu, yükseköğretime devam etme isteği, örnek aldığı bir rol modelin varlığı, anne ile ilişki düzeyi, baba ile ilişki düzeyi vb.), kaygı, tükenme, akademik güdülenme, iletişimde olduğu öğretmenlerin depresyon düzeyi gibi faktörler ile yılsonu başarı ortalaması ve devamsızlık bilgilerini içeren verileri toplamıştır. Verisi toplanan 2371 öğrencinin 887'sinin erkek, 819'unun kadın olmak üzere 1706 öğrencinin verisinin kullanılabilir olduğu anlaşılmıştır. Sınıflandırma tekniği olarak k-En Yakın Komşu, Naive Bayes, C4.5 Karar Ağacı Algoritması, Lojistik Regresyon Analizi ve Destek Vektör Makineleri kullanılarak farklı

modeller oluşturmuş ve C4.5 Karar Ağacı Algoritmasının akademik başarısının öngörülmesine ilişkin daha başarılı olduğu sonucuna varmıştır.

Ersöz (2017), Eğitsel Veri Madenciliği ile Öğrenci Profillerinin Belirlenmesi başlıklı yüksek lisans çalışmasında Uludağ Üniversitesi Eğitim Fakültesi öğrencilerinin tamamı ile çalışmış ve Veri Madenciliği yöntemlerinden Sınıflandırma Analizi yöntemlerini kullanılmıştır. Çalışma sonucunda öğrenci başarılarının önem sırasına göre bölüm, akademik dönem, geliş şekli ve cinsiyet değişkenlerine göre farklılıklar gösterdiği tespit edilirken sınıflandırma analizi sonucunda ise doğum yılı ve akademik dönem değişkenleri düğüm noktaları olarak belirlenmiştir.

Altun vd. (2019), Akdeniz Üniversitesi Eğitim Fakültesi Sınıf Öğretmenliği Bölümü'nden 2012 ile 2017 yılları arasında mezun olan 578 öğrenci ile yaptığı çalışmasında çoklu doğrusal regresyon analizi ve yapay sinir ağları ile oluşturulan modellerin çapraz değerlendirme sonuçlarında öğrenci başarısını tahminlemeye çalışmış ve %94 gibi birbirine yakın değerler elde etmiştir.

Topuz (2021), MEB tarafından yapılan ABİDE sınavına ilişkin sonuçları kullanmış ve veri madenciliğinde kullanılan Weka ve tezimizde kullandığımız Orange yazılımlarını kullanmıştır. Yapılan çalışmada öğrencilerin Türkçe dersi puanlarına göre sınıflandırılmış ve k-en yakın komşu, rastgele orman, destek vektör makinesi, naive bayes ve yapay sinir ağları yöntemleri kullanılmıştır. Çalışma sonucunda; k-en yakın komşu ve yapay sinir ağları algoritmalarında Orange, destek vektör makinesi ve naive bayes algoritmasında ise Weka'nın daha iyi sonuçlar ürettiği görülmüştür. Rastgele orman algoritmasının ise ikili ve beşli sınıflamada sırasıyla, Weka ve Orange paket programında daha yüksek sonuçlar elde edilmiştir. Son olarak ise ikili ve beşli olarak sınıflandırılmış puanlarda en yüksek doğru sınıflama oranını, yapay sinir ağları algoritması elde etmiştir.

Can (2021), karar ağaçları sınıflandırması, doğrusal regresyon, rastgele orman algoritması, destek vektör makineleri, k-en yakın komşu algoritması ve Gaussian NB algoritması kullanarak 1979-2020 yılları arasında üniversite sınavına katılan 677 kişi ile bir anket aracılığı topladığı veriler üzerinde bir tahminleme çalışması gerçekleştirmiş ve %73,77 lik bir doğruluk oranı ile en yüksek değere Gaussian NB algoritmasında ulaşmıştır.

Ekinci (2022), İnönü Üniversitesi otomasyon siteminden 223.279 öğrencinin verisini içeren bir veri seti ile çalışmış ve yaptığı çalışma neticesinde öğrenci mezuniyet süresinin tahminlenmesinde lojistik regresyon modelinin %76,80 ile yüksek düzeyde performans gösterdiği, elde edilen bu sonuç ile de öğrencilerin üniversitedeki kişisel ve akademik bilgileri ile mezuniyet süresinin tahminlenebileceği ve bu değişkenler kullanılarak öğrencilere yönelik gerekli önlemlerin alınabileceği sonucuna ulaşılmıştır. Aynı şekilde Bilgisayar II dersinde lojistik regresyon modelinin %79,34 oranında bir doğruluğa ulaştığını tespit etmiştir.

Adak ve Duralioğlu (2023), makine öğrenmesi ile öğrencilerin kazanım bilgileri ile sınav verilerine göre sonraki sınavlardaki başarısını tahminlemeye çalışmıştır. Çalışmasında 2021-2022 eğitim öğretim yılının ilk döneminde, İstanbul'da Ataşehir ilçesinde bulunan Dr. Nureddin Erk-Perihan Erk Mesleki ve Teknik Anadolu Lisesinin Bilişim Teknolojileri alanındaki öğrenim gören 10 ve 11'inci sınıfta okuyan 87 öğrencinin Nesne Tabanlı ve Programlama dersinde uygulanan üç sınavdaki puanlarını kullanmış ve en iyi sonucu kNN algoritması ile elde etmiştir.

Abbasoğlu (2020), e-Okul sistemindeki verilerden faydalanarak eğitsel veri madenciliği ile yapmış olduğu çalışmasında ortaokul öğrencilerinin akademik başarılarını tahminlemeye çalışmıştır. Bu çalışmasında Yalova ilinden 5,6,7 ve 8. Sınıf öğrencisi olan 1395 öğrenci ile çalışmış ve lojistik sınıflandırma algoritması en iyi sonucu elde etmiştir.

Suna vd. (2020), e-Okul sistemindeki verilerden faydalanarak fen lisesi öğrencilerinin sosyoekonomik arkaplanını ortaya koymuş ve fen lisesi mezunlarının yükseköğretim kurumlarına yönelik tercihlerini ve yerleştikleri program türlerinin yıllara göre değişimini incelemiştir.

İlgili araştırmalar genel olarak değerlendirildiğinde EVM kapsamında birçok farklı çalışma yapılmıştır. Bu çalışmaların ortak özelliği ise öğrencinin başarı durumlarının VM yöntemleri ile daha erken evrede tespit edilip edilemeyeceğinin ortaya konmasıdır. Bu araştırmalarda dikkat çeken bir diğer husus ise farklı algoritmaların başarılı olmasıdır. VM' de doğru algoritma ve doğru yöntem veri setine göre değişiklik gösterebilir. Bir algoritma bir veri setinde çok iyi sonuçlar verebilirken başka veri setinde çok kötü sonuç verebilir (Aksu 2018). Elde edilen sonuçlar ve literatürde yapılan araştırmalarda da farklı algoritmaların öne çıktığı görülmüştür.

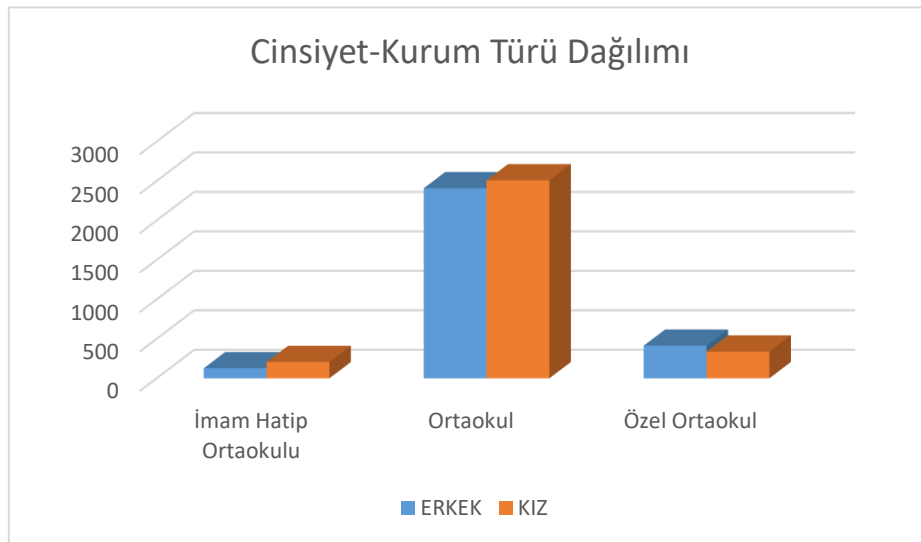
3. MATERYAL VE YÖNTEM

Bu bölümde veri seti, öznitelikler, veri analizi ve özellik seçimleri başlıkları anlatılmaktadır. Bu tez çalışmasında veri madenciliğın kullanım alanları içerisinde yer alan eğitsel veri madenciliğı ile danışmanlı veri madenciliğı teknikleri içerisindeki sınıflama ve regresyon yöntemleri ile tahminlemeler yapılarak verilerin işlenmesinde Orange yazılımı kullanılmıştır.

3.1 Veri Seti

Bu tezin araştırma grubunu 2022 yılında bir ortaöğretim kurumuna yerleşmek için Ankara ilinde öğrenim gören ve 2022 yılında LGS tercihinde bulunan 6000 8. Sınıf öğrencisi oluşturmaktadır. Bu tez kapsamında kullanılan veriler Millî Eğitim Bakanlığı Bilgi İşlem Genel Müdürlüğü'nden 13.06.2023 tarih ve 78199263 yazılı izni ile alınmıştır.

Bu öğrenciler tahminleme yapılacak hedef değişken olan kazanma bilgisine göre bir liseyi kazanamayan, sınavlı kazanan ve sınavsız kazanan olmak üzere her birinden 2000 adet olmak üzere rastgele örneklem alınarak seçilmişlerdir. Veri setine ait bazı istatistikî bilgilere aşağıda yer verilmiştir. Öğrencilerin öğrenim gördükleri kurum türüne göre dağılım grafiğı Şekil 3.1 de, grafiğı oluşturan veri tablosu ise Çizelge 3.1 de verilmiştir.



Şekil 3.1 Cinsiyete göre mezun olunan ortaokullar.

Şekil 3.1 de araştırma grubunu oluşturan öğrencilerin grafiksel dağılım gösterilmektedir. Bu dağılıma göre araştırma grubunun çoğunluğunu bir devlet ortaokulunda öğrenim gören öğrenciler oluşturmaktadır. Grafiğe ait veri tablosu Çizelge 3.1 de verilmiştir.

Çizelge 3.1 Cinsiyete göre mezun olunan okul dağılımı.

Mezun Olunan Okul Türü	Erkek Öğr. Sayısı	Kız Öğr. Sayısı	Toplam
İmam Hatip Ortaokulu	127	206	333
Ortaokul	2408	2508	4916
Özel Ortaokul	413	338	751

Çizelge 3.1 de verilen cinsiyete göre mezun olunan okulların dağılımında araştırma grubunu oluşturan 6000 öğrencinin 333'ü İmam Hatip Ortaokulu mezunu, 4916 'sı bir devlet ortaokulu mezunu iken 751'i ise Özel bir ortaokul mezunudur.

3.2 Öznitelikler

Araştırma grubundaki öğrencilere ait analiz verileri Ek 1 de verilmiştir. Bu verilerde veri başlığı bölümündeki;

OGRENCI_ID anonim olarak temin edilen verideki her bir öğrenci için verilmiş benzersiz bir numaradır ve her öğrenci için farklı verilmiştir.

YERLESTIRME_GRUBU veri başlığımız hedef değişkenimiz olup kazanamadı, sınavlı ve sınavsız olmak üzere üç farklı değer almaktadır. Herhangi bir liseye yerleşemeyen öğrenciler kazanamadı olarak, sınavlı bir okula yerleşen öğrenciler sınavlı, sınavsız bir okula yerleşen öğrenciler ise sınavsız olarak kategorize edilmiştir.

KAYITLI_TUR veri başlığında öğrencinin 8. Sınıfta kayıtlı olduğu okul türünü ifade etmektedir. Bu alan kategorik bir alan olup Ortaokul (Devlet Ortaokulu), İmam Hatip Ortaokulu ve Özel Ortaokul olmak üzere üç farklı değer alabilmektedir.

GELIR_DUZEYI veri başlığı öğrencinin ailesinin gelir düzeyini ifade etmekte olup çok kötü, düşük, orta, iyi, çok iyi ve belirsiz olmak üzere 6 farklı değerle kategorize edilmiştir.

CINSIYET veri başlığında öğrencilerin cinsiyetlerini ifade etmekte olup Erkek ve Kız olarak kategorize edilmiştir.

YBP veri başlıkları sırası ile YBP6, YBP7 ve YBP8 olarak öğrencinin 6. sınıf, 7. Sınıf ve 8. Sınıftaki yıl sonu başarı puanlarını ifade etmekte olup 0-100 aralığında değer almaktadır.

ANNE_EGITIM ve BABA_EGITIM veri başlıkları öğrencinin anne ve babasının eğitim durumunu gösteren 3 Yıllık Eğitim Enstitüsü, Doktora, İlkokul, İlköğretim, Açık İlköğretim Öğrencisi, Lisans, Lisansüstü, Lise, Okuma Yazma Biliyor, Okuma Yazma Bilmiyor, Ortaokul, Yüksek Okul (2 Yıllık) ve BELİRSİZ olarak değerler almaktadır.

SINAVSIZ_TERCİH veri başlıkları beş başlıktan oluşup öğrencilerin yerel yerleştirme kapsamında sorası ile 1. Tercihden 5. Tercihe kadar olan tercihlerini ifade etmektedir. Bu alan kategorik bir alan olup Anadolu Lisesi, Anadolu İmam Hatip Lisesi ve Anadolu Meslek Lisesi olmak üzere üç farklı değer almaktadır.

TERCİH SAYISI öğrencinin LGS için kaç tercih yaptığını gösteren sayısal alandır ve SINAVSIZ_TERCİH başlıklarındaki dolu alan sayısını ifade etmekte olup 1-5 aralığından değer alabilir.

BURS_DURUMU veri başlığında MEB tarafından her yıl yapılmakta olan İlköğretim ve Ortaöğretim Kurumları Bursluluk Sınavı (İOKBS)' nda başarılı olarak burs almaya hak kazanan öğrenci olup olmadığını ifade etmektedir. Bu alan 0 burs almıyor, 1 ise burs alıyor olarak veri setimizde kodlanmıştır.

YABANCI_UYRUKLU veri başlığında öğrencinin Türk vatandaşı olup olmadığı bilgisi yer almaktadır. Türk vatandaşı olmayanlar için bir 1 (evet), 0 (hayır) olarak veri setimizde kodlanmıştır.

ETKİNLİK veri başlığında öğrencinin 2022-2023 eğitim öğretim yılında herhangi bir sosyal etkinliğe katılıp katılmadığı verisi bulunmaktadır. Bu veri e-Okul sistemindeki sosyal etkinlik modülünden alınmış olup bu modüle veri girişleri MEB Sosyal etkinlikler yönetmeliğine göre yapılmaktadır (MEB 2017).

AİLESİYLE_OTURAN veri başlığında öğrencinin kiminle yaşadığı verisi bulunmaktadır. Bu alan kategorik bir alan olup -1 (Bilgi Yok), 0 (Ailesiyle), 1 (Annesiyle), 2 (Babasıyla), 3 (Velisiyle) olmak üzere beş farklı değer almaktadır.

KENDI_EVINDE_OTURAN veri başlığında öğrencinin kendi sahip oldukları evde mi yoksa kira veya lojmanda mı oturdukları verisi bulunmaktadır. Bu veri alanı kategorik bir alan olup 1(Kira), 2 (Lojman), 3 (Kendi evi) olmak üzere 3 farklı değer almaktadır.

KAYNASTIRMA veri başlığında ise öğrencinin kaynaştırma öğrencisi olup olmadığı bilgisi yer almaktadır. Bu veri alanı kategorik bir alan olup -1 (Kaynaştırma Öğrencisi Değil), 1 (Okulda eğitim alıyor-Kaynaştırma Öğrencisi), 2 (Hastanede eğitim alıyor-Kaynaştırma Öğrencisi), 3 (Evde eğitim alıyor-Kaynaştırma Öğrencisi) ve dört (Okulda Grup Eğitimine Hazırlık Alıyor -Kaynaştırma Öğrencisi) olmak üzere beş farklı değer almaktadır.

Veri setimizde farklı analizlerin yapılabilmesi için tercih bilgilerinde kodlama yapılmıştır. Bu kodlamada Anadolu Liseleri; AL, Anadolu Meslek Programı; AMP, Anadolu İmam Hatip Lisesi ise AİHL olarak kodlanmıştır ve daha sonra öğrenci tercihleri ile bir araya getirilerek yanyana + sembolü ile birleştirilmiştir. Örnek kodlama Çizelge 3.2 de paylaşılmıştır.

Çizelge 3.2 Yerel tercihlerin örnek kodlaması.

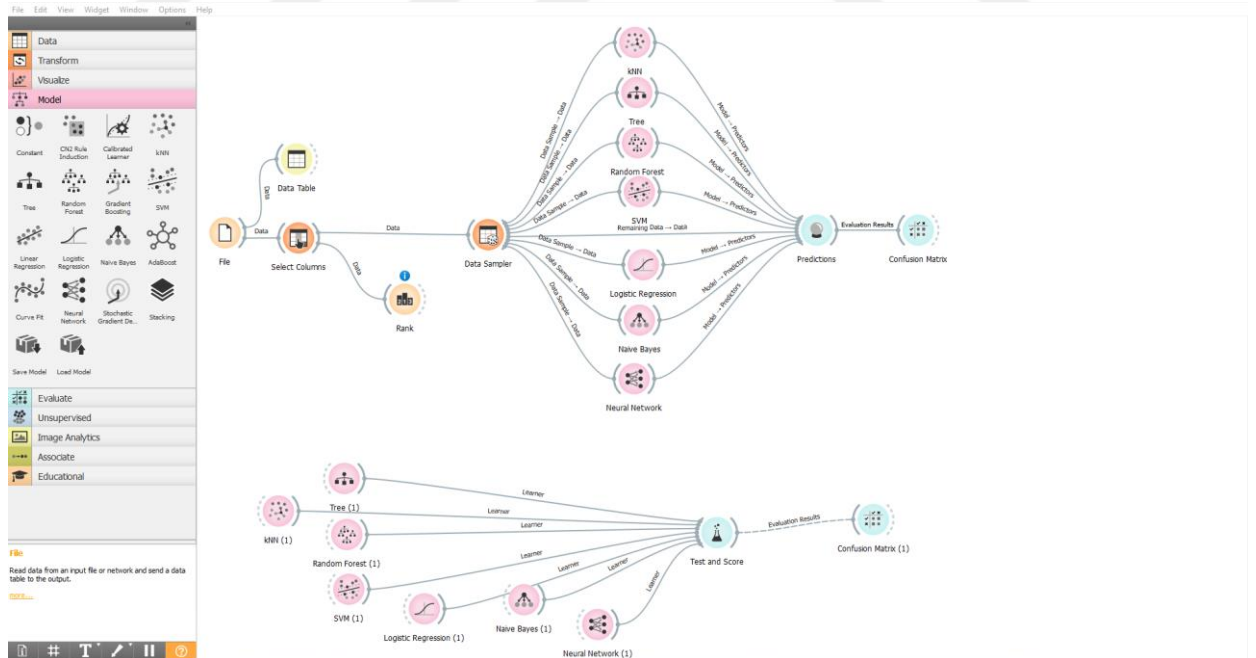
1. TERCİH	2. TERCİH	3. TERCİH	4. TERCİH	5. TERCİH	TERCİH KODLAMASI
Anadolu Lisesi					AL
Anadolu Lisesi	Anadolu Lisesi	Anadolu Meslek Programı			AL+AL+AMP
Anadolu Meslek Programı	Anadolu Lisesi	Anadolu Meslek Programı	Anadolu İmam Hatip Lisesi		AMP+AL+AMP+AİHL
Anadolu Lisesi	Anadolu Lisesi	Anadolu İmam Hatip Lisesi	Anadolu Lisesi	Anadolu Meslek Programı	AL+AL+AİHL+AL+AMP
Anadolu İmam Hatip Lisesi	Anadolu İmam Hatip Lisesi	Anadolu Meslek Programı	Anadolu Meslek Programı	Anadolu Meslek Programı	AİHL+AİHL+AMP+AMP+AMP

Çizelge 3.2 de tek tercihini Anadolu Lisesi yapan bir öğrenci AL olarak kodlanmış, ilk iki tercihi Anadolu lisesi olup 3. Tercihi Anadolu İmam Hatip Lisesi olan bir öğrenci ise AL+AL+AİHL şeklinde kodlanmıştır. Buradaki amaç beş farklı tercihi tek bir veri alanında birleştirerek ortaya çıkan dağılımı izlemektir. Bu dönüşüm sonucunda öğrencilerin tercih sayıları ve tercihlerinin sıralaması ile ilgili oluşan veriler Ek 2’de verilmiştir.

3.3 Veri Analizi

Bu araştırmada Şekil 3.2 de ekran görüntüsü verilen Orange Data Mining yazılımında K-en yakın komşu (k-nearest neighbors) algoritması, karar ağacı algoritması (tree), rastgele orman (random forest) algoritması, destek vektör makineleri (SVM), lojistik regresyon, naive bayes ve yapay sinir ağları (naurel network) algoritmaları kullanılmıştır. Toplamda 6000 öğrenci verisi üzerinde inceleme yapılmıştır. Veri analizi modeli Şekil 3.3’te gösterilmiştir.

Verinin %70’i yani 4200 satırı eğitim veri seti, %30 u yani 1800 satırı ise tahminlemede kullanılmıştır.



Şekil 3.2 Orange data mining ekran görüntüsü.

T10: Gerçekte sınavlı bir okulu kazanan ve kazanamadı olarak tahminlenen öğrenci sayısı.

T11: Gerçekte sınavlı bir okulu kazanan ve sınavlı kazandı olarak doğru tahminlenen öğrenci sayısı.

T12: Gerçekte sınavlı bir okulu kazanan ve sınavsız kazandı olarak tahminlenen öğrenci sayısı.

T20: Gerçekte sınavsız bir okulu kazanan ve kazanamadı olarak tahminlenen öğrenci sayısı.

T21: Gerçekte sınavsız bir okulu kazanan ve sınavlı kazandı olarak tahminlenen öğrenci sayısı.

T22: Gerçekte sınavsız bir okulu kazanan ve sınavsız kazandı olarak doğru tahminlenen öğrenci sayısı.

Çok sınıflı karmaşıklık matrisinde tanımlar iki boyutlu karmaşıklık matrisi ile aynıdır. Farklı olarak ise her sınıf için hesaplama yapılır

Doğru Pozitif (DP):

DP değerleri her sınıf için ortaktır.

Kazanamadı sınıfımız için DP değeri T00 ile ifade edilmiştir.

Sınavlı sınıfı için DP değeri T11 ile ifade edilmiştir.

Sınavsız sınıfı için DP değeri T22 ile ifade edilmiştir.

Yanlış Negatif (YN):

$YN = T11 + T12 + T21 + T22$ “sınavsız” sınıfı için YN değeri hesaplaması

$YN = T00 + T02 + T20 + T22$ “sınavlı” sınıfı için YN değeri hesaplaması

$YN = T00 + T01 + T10 + T11$ “sınavsız” sınıfı için YN değeri hesaplaması

Yanlış Pozitif (YP):

$YP = T10 + T20$ “sınavsız” sınıfı için YP değeri hesaplaması

YP =T01+T21 “sınavlı” sınıfı için YP değeri hesaplaması

YP =T02+T12 “sınavsız” sınıfı için YP değeri hesaplaması

Yanlış Negatif(YN):

YN =T01+T02 “sınavsız” sınıfı için YN değeri hesaplaması

YN =T10+T12 “sınavlı” sınıfı için YN değeri hesaplaması

YN =T20+T21 “sınavsız” sınıfı için YN değeri hesaplaması

Doğruluk:

Doğruluk, doğru sınıflandırma sayısının toplam sınıflandırma sayısına oranı olarak hesaplanır. Karışıklık matrisimize göre, doğru sınıflandırmalar her sınıf için DP ve toplam sınıflandırma sayısı, DP de dahil olmak üzere karışıklık matrisindeki her değer toplamıdır.

$$Doğruluk(CA) = \frac{T00 + T11 + T22}{T00 + T01 + T02 + T10 + T11 + T12 + T20 + T21 + T22} \quad (3.1)$$

Doğruluk gibi tahminleme işlemlerinde kullanılan kavramlar bulunmaktadır. Bulgular bölümünde ve tahminleme de kullanılan bazı kavramlar aşağıdaki gibi tanımlanmaktadır.

Duyarlık (Sensitivity/Recall):

Doğru sınıflandırılmış pozitif örnek sayısının toplam pozitif örnek sayısına oranıdır.

$$\begin{aligned} Duyarlık(kazanamadı) &= \frac{T00}{T00 + T01 + T02} \\ Duyarlık(sınavlı) &= \frac{T11}{T10 + T11 + T12} \\ Duyarlık(sınavsız) &= \frac{T22}{T20 + T21 + T22} \end{aligned} \quad (3.2)$$

Kesinlik (Precision):

Sınıfta doğru olarak tahminlenmiş DP örnek sayısının, kendi sınıfı olarak tahminlenmiş tüm örnek sayısına oranıdır.

$$\begin{aligned} Kesinlik(kazanamadı) &= \frac{DP}{DP + YP} = \frac{T00}{T00 + T10 + T20} \\ Kesinlik(sınavlı) &= \frac{DP}{DP + YP} = \frac{T11}{T11 + T01 + T21} \\ Kesinlik(sınavsız) &= \frac{DP}{DP + YP} = \frac{T22}{T22 + T02 + T12} \end{aligned} \quad (3.3)$$

F1(F-Score, F-ölçütü)

Kesinlik ve duyarlılık ölçütleri tek başına anlamlı bir karşılaştırma yapmak için yeterli değildir. Her iki ölçütün bir arada değerlendirilmesi, daha doğru sonuçlar elde etmemize yardımcı olur. Bu amaçla, f-ölçütü (f-score) tanımlanmıştır. F-ölçütü, kesinlik ve duyarlılık ölçütlerinin harmonik ortalamasını temsil eder. Bu ölçüt, sınıflandırma modelinin hem kesinlik hem de duyarlılık performansını dengeleyerek değerlendirilmesine yardımcı olur. F-1 ölçütü her bir sınıf için aşağıdaki formülle hesaplanmaktadır.

$$F - 1 = \frac{2x(Kesinlik)x(Duyarlık)}{(Kesinlik) + (Duyarlık)} \quad (3.4)$$

3.4 Özellik Seçimi

Özellik seçimi (feature selection), aslen veri kümesinin en iyi temsilini sağlayacak olan altkümenin seçilmesi olarak tanımlanır. Aynı zamanda nitelik seçimi veya değişken seçimi olarak da adlandırılan bu süreç, kullanılan algoritmaya bağlı olarak, veri setinde bulunan n adet özelliği değerlendirerek en iyi k adet özelliği seçme işlemini ifade eder. Bu yöntem, veri setinin boyutunu azaltarak gereksiz veya düşük bilgi taşıyan özelliklerin elemine edilmesini ve modelin performansını artırarak gereksiz gürültüyü azaltmayı amaçlar (Forman 2003).

Bu tez kapsamında Orange yazılımında bulunan farklı özellik seçicilerinden üç tanesine ait değerler karşılaştırılmış ve genellikle (Şekil 3.4) aynı değişkenlerin ön plana çıktığı görülmüş bu kapsamda information gain yönetimine göre tahminleme işlemleri yapılmıştır.

Bilgi Kazancı (Information Gain):

Bu özellik seçim yönteminde Y özelliğini anlamak için gereken bilgi ile X özelliği kullanıldığında Y özelliğini anlamak için gereken bilgi arasındaki farkı ölçen bir skordur ve hesaplamasında entropi modeli kullanılır. Entropi, bir sistemdeki belirsizliğin veya öngörülemezliğin derecesini temsil eder. X özelliği altında Y özelliğinin entropi değerindeki azalmayı gösteren Bilgi Kazancı, aşağıdaki formülle hesaplanır:

$$H(Y) = - \sum_{y \in Y} p(y) \log_2(p(y))$$

$$H(Y|X) = - \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2(p(y|x)) \quad (3.5)$$

$$\text{Bilgi Kazancı} = H(Y) - H(Y|X)$$

Bilgi kazancı, simetrik bir ölçüttür ve bu ölçüte göre, X gözlemlendikten sonra Y hakkında kazanılan bilgi ile Y gözlemlendikten sonra X hakkında kazanılan bilgi birbirine eşittir. Bununla birlikte, bu yöntemin bir zayıf yönü vardır; o da daha fazla bilgiye sahip olmayan ancak çeşitli değerlere sahip olan özelliklerin lehine önyargılı bir şekilde sonuçlar üretebilmesidir. Bu durum, eşit olmayan dağılıma sahip veri setlerinde ve özellikler arasında farklı frekansta gözlenen değerlerde problem yaratabilir.

Kazanç Oranı (Gain Ratio):

Bu özellik seçim yönteminde simetrik olmayan bir ölçüttür ve çeşitli değerlere sahip özelliklerin seçilme eğilimini dengelemek amacıyla kullanılır. Bu ölçüt, özellik seçimi sırasında sınıflandırıcı veya modelin performansını daha iyi değerlendirmek için kullanılır. Kazanç Oranı, 0 ile 1 arasında değer alır. Eğer oran 1'e eşitse, X bilgisi tamamen Y bilgisini tahmin edebildiği anlamına gelir. Eğer oran 0'a eşitse, X ile Y

arasında herhangi bir ilişki olmadığını ifade eder. Bu nedenle, Kazanç Oranı, özelliklerin birbiriyle ilişkisini ve seçilme değerini değerlendirmek için kullanışlı bir metriktir ve aşağıdaki formüller hesaplanmaktadır (Budak 2018).

$$Kazanç\ Oranı = \frac{Bilgi\ Kazancı}{H(X)} \quad (3.6)$$

Gini İndeksi:

Gini indeks değeri, veri setindeki öğelerin dağılımını ifade eden bir ölçüdür. İki farklı öğe türünün (sınıf, kategori vb.) gini indeks değeri eşitse, bu durumun sonuçların benzer bir dağılıma sahip olduğu anlamına geldiği kabul edilir. Eğer veri setinde bir nitelikte üç veya daha fazla seçenek varsa ve bu niteliği bölümlere ayırırken yalnızca iki bölüme izin veriliyorsa, yakın seçenekler benzer gruplara ayrılacaktır. Bu, özellikle aynı türden öğelerin benzerliklerini korumak ve daha iyi ayırtırmak amacıyla yapılan bir yöntemdir (Gökbay vd. 2021).

Bu üç özellik seçimine seçilen özelliklere değişkenlerin aldığı değerle Şekil 3.4 de verilmiştir.

Information Gain				Gain Ratio				Gini			
		#	Info. gain			#	Gain ratio			#	Gini
1	N	YBP8	0.328	1	C	BURS_DURUMU	0.206	1	N	YBP8	0.142
2	N	YBP6	0.301	2	N	YBP8	0.164	2	N	YBP6	0.129
3	N	YBP7	0.175	3	N	YBP6	0.150	3	N	YBP7	0.076
4	C	BABA_EGITIM	0.087	4	N	YBP7	0.087	4	C	BABA_EGITIM	0.041
5	C	SINAVSIZ_TERCIH3	0.083	5	C	KAYITLI_TUR	0.060	5	N	TERCIH_SAYISI	0.039
6	C	SINAVSIZ_TERCIH2	0.082	6	C	SINAVSIZ_TERCIH2	0.058	6	C	SINAVSIZ_TERCIH3	0.038
7	N	TERCIH_SAYISI	0.081	7	C	SINAVSIZ_TERCIH1	0.057	7	C	SINAVSIZ_TERCIH2	0.037
8	C	BURS_DURUMU	0.079	8	C	SINAVSIZ_TERCIH3	0.052	8	C	ANNE_EGITIM	0.037
9	C	ANNE_EGITIM	0.079	9	C	SINAVSIZ_TERCIH4	0.043	9	C	BURS_DURUMU	0.037
10	C	KAYITLI_TUR	0.051	10	N	TERCIH_SAYISI	0.043	10	C	KAYITLI_TUR	0.025
11	C	SINAVSIZ_TERCIH1	0.050	11	C	SINAVSIZ_TERCIH5	0.037	11	C	SINAVSIZ_TERCIH1	0.022
12	C	SINAVSIZ_TERCIH4	0.040	12	C	BABA_EGITIM	0.032	12	C	SINAVSIZ_TERCIH4	0.018
13	C	GELIR_DUZEYI	0.024	13	C	ANNE_EGITIM	0.029	13	C	GELIR_DUZEYI	0.011
14	C	SINAVSIZ_TERCIH5	0.021	14	C	YABANCI_UYRUKLU	0.021	14	C	SINAVSIZ_TERCIH5	0.010
15	C	YABANCI_UYRUKLU	0.003	15	N	KAYNASTIRMA	0.016	15	C	YABANCI_UYRUKLU	0.001
16	C	CINSIYET	0.002	16	C	GELIR_DUZEYI	0.012	16	C	CINSIYET	0.001
17	C	KENDI_EVINDE_OTURAN	0.002	17	C	CINSIYET	0.002	17	C	KENDI_EVINDE_OTURAN	0.001
18	N	AILESIYLE_OTURAN	0.001	18	N	AILESIYLE_OTURAN	0.002	18	N	AILESIYLE_OTURAN	0.001
19	N	KAYNASTIRMA	0.001	19	C	KENDI_EVINDE_OTURAN	0.002	19	N	KAYNASTIRMA	0.000
20	C	ETKINLIK	0.000	20	C	ETKINLIK	0.001	20	C	ETKINLIK	0.000

Şekil 3.4 Özellik seçim yöntemlerine göre değişkenler.

Yöntemlere göre deęişkenler Şekil 3.4 de görüldüğü gibi farklı önem derecelerine göre sıralanabilmektedir. Seçilen deęerler her üç yöntemdede benzer bir dağılım göstermektedir. Özellik seçiminde yine üç yöntemde de YBP puanlarının ön plana çıktığı ve kırılma noktalarının bu deęerlerden sonra başladığı görülmektedir. Ancak öğrencilerin demografik ve tercih verilerinde sonuca etkilerini görmek için information gain yöntemine göre en iyi 10 deęişken alınarak analizler yapılmış ve sonuçlar aşağıda paylaşılmıştır.

Sonuçlar deęerlendirilirken kesinlik ve duyarlık ölçütleri ile algoritma performansları ile ilgili bilgi sahibi olunabilse de kesin bir sonuç üretmez. Bu konuda daha iyi bir deęerlendirme yapmak için duyarlık ve kesinlik deęerlerinden faydalanılarak hesaplanan F1-skoru ölçütüne bakmak daha doğru olacaktır (Coşkun ve Baykal 2021).

4. BULGULAR

Bu bölümde, analiz sonucu elde edilen bulgular VM yöntemleri ile analiz edilerek mevcut veri setine göre en başarılı tahminleme algoritması ve yöntemi tespit etmek için elde edilen bulgular değerlendirilecektir. Orange yazılımı ile veri madenciliği teknikleri uygulanmadan önce veriler üzerinde yapılan ön analizde aşağıdaki bulgulara ulaşılmıştır.

4.1 Tercih Sayılarının Yerleşmeye Etkisine Dair Bulgular

Tercih sayılarına bağlı yerleşme Çizelge 4.1 de paylaşılmıştır.

Çizelge 4.1 Tercih sayılarına göre yerleşen öğrencilerin dağılımı.

Tercih Sayısı	Yerleşemeyen		Sınavlı Bir Okula Yerleşen		Sınavsız Bir Okula Yerleşen		Genel Toplam
	n	%	n	%	n	%	
1	291	14,55%	679	33,95%	185	9,25%	1155
2	357	17,85%	411	20,55%	304	15,20%	1072
3	1086	54,30%	676	33,80%	895	44,75%	2657
4	149	7,45%	113	5,65%	296	14,80%	558
5	117	5,85%	121	6,05%	320	16,00%	558
Genel Toplam	2000		2000		2000		6000

Çizelge 4.1 incelendiğinde Çizelge 3.2’ de paylaşılan örnek kodlamaya göre yapılan analizde yerleşemeyen öğrencilerin %86,70 i en fazla 3 tercih yapmış ve bu tercihlerin %67,6 sı Anadolu Lisesi şeklindedir. Tercih sayısı arttıkça ve özellikle üç tercihe kadar tercih yapan öğrencilerin yerleşme oranları oldukça yüksek görünmektedir.

4.2 Araştırma Sorularına Yönelik Bulgular

Öğrencilerin akademik ve demografik verilerinin bir ortaöğretim kurumuna yerleşme durumlarının VM yöntemleri ile modellenerek tahminlenmesi sonucunda en iyi performans lojistik regresyon algoritmasından elde edilmiştir. Bu algoritmayı yakın sonuçlar üreten YSA, navie bayes ve rastgele orman algoritmaları izlemiştir.

Tahminleme Algoritmalarının Sonuçlarına Dair Bulgular

Şekil 3.4 de paylaşılan özelli seçim yöntemlerinden information gain yöntemi ile seçilen ilk 10 değişkene göre seçilen özellikler Çizelge 4.2’de verilmiştir.

Çizelge 4. 2 Information gain yöntemine göre seçilen en önemli 10 değişken.

Model	Bilgi Kazancı
YBP8	0.328
YBP6	0.301
YBP7	0.175
BABA_EGITIM	0.087
SINAVSIZ_TERCIH3	0.083
SINAVSIZ_TERCIH2	0.082
TERCIH_SAYISI	0.081
BURS_DURUMU	0.079
ANNE_EGITIM	0.079
KAYITLI_TUR	0.051

Çizelge 4.2’ye göre kazanç oranı en iyi 10 değişkende öğrencilerin akademik veriler en üst sırada yer almaktadır. Ancak demografik verileri ve tercih verilerinin de etkilerini gözlemlemek için en iyi 10 değişken seçilmiştir. Onuncu değişkeni izleyen değişkenler yakın kazanç oranlarına sahip olduklarından listeye dahil edilmemişlerdir.

Information gain yöntemine göre seçilen değişkenlerin farklı tahminleme algoritmaları ile tahminlenmesi sonucunda elde edilen değerler Çizelge 4.3’de verilmiştir.

Çizelge 4.3 Information gain yöntemi ile seçilen değişkenlere göre tahmin sonuçları.

Model	Doğruluk	Kesinlik	Duyarlık	MCC	F-Ölçütü
Lojistik Regresyon	0.670	0.664	0.670	0.506	0.665
Yapay Sinir Ağları	0.643	0.646	0.643	0.468	0.641
Naive Bayes	0.642	0.640	0.642	0.465	0.638
Rastgele Orman	0.627	0.630	0.627	0.441	0.627
kNN	0.597	0.612	0.597	0.409	0.589
Karar Ağacı	0.575	0.573	0.575	0.363	0.574
Destek Vektör Makineleri	0.534	0.562	0.534	0.305	0.543

Çizelge 4.3’te sonuçları verilen information gain yöntemi ile Şekil 3.4 de seçilen ilk 10 değişkene göre tahminleme yapılmış ve en iyi skor 0,665 olarak lojistik regresyon yönteminde elde edilmiştir. Bunu 0,641 ile yapay sinir ağları, 0,638 ile naive bayes, 0,627

ile rastgele orman algoritması, 0,589 ile k-en yakın komşu, 0,574 ile karar ağaçları ve son olarak da 0,543 ile destek vektör makinesi algoritması izlemiştir. Eğitim veri setinden elde edilen tahminlemeye göre kalan 1800 satır veri üzerindeki tahminleme sonuçlarına göre bu algoritmaların karmaşıklık matrisleri ise aşağıda Şekil 4.1, Şekil 4.2, Şekil 4.3, Şekil 4.4, Şekil 4.5, Şekil 4.6 ve Şekil 4.7’de paylaşılmıştır.

		Predicted			
		kazanamadı	sınavlı	sınavsız	Σ
Actual	kazanamadı	421	45	123	589
	sınavlı	41	488	86	615
	sınavsız	172	127	297	596
	Σ	634	660	506	1800

		Predicted			
		kazanamadı	sınavlı	sınavsız	Σ
Actual	kazanamadı	71.5 %	7.6 %	20.9 %	589
	sınavlı	6.7 %	79.3 %	14.0 %	615
	sınavsız	28.9 %	21.3 %	49.8 %	596
	Σ	634	660	506	1800

Şekil 4.1 Lojistik regresyon algoritması karmaşıklık matrisi.

Şekil 4.1’de gösterilen Lojistik Regresyon algoritması ile tahminleme için kullanılan 1800 öğrencinin verilerine göre;

Herhangi bir okula yerleştirmesi yapılmamış (kazanamadı) 589 öğrencinin 421’ i (%71,5) doğru, 45’i (%7,6) sınavlı bir okulu kazandı olarak, 123’ ü (%20,9) ise sınavsız bir okulu kazandı olarak tahmin etmiştir. Sınavlı bir okul kazanan 615 öğrencinin 41’ini (%6,7) kazanamadı, 488’ ini (%79,3) doğru tahmin etmiş, 86’sını (%14) ise sınavsız bir okul kazandı olarak tahminlemiştir. Sınavsız bir okulu kazanan 596 öğrenciden 172’ sini (%28,9) kazanamadı, 127’ sini (%21,3) sınavlı bir okul kazandı olarak, 297’ sini (%49,8) ise doğru olarak tahminlemiştir.

		Predicted			Σ
		kazanamadı	sınavlı	sınavsız	
Actual	kazanamadı	438	23	128	589
	sınavlı	57	441	117	615
	sınavsız	224	93	279	596
Σ		719	557	524	1800

		Predicted			Σ
		kazanamadı	sınavlı	sınavsız	
Actual	kazanamadı	74.4 %	3.9 %	21.7 %	589
	sınavlı	9.3 %	71.7 %	19.0 %	615
	sınavsız	37.6 %	15.6 %	46.8 %	596
Σ		719	557	524	1800

Şekil 4.2 YSA karmaşıklık matrisi.

Şekil 4.2’de gösterilen YSA algoritması ile tahminleme için kullanılan 1800 öğrencinin verilerine göre;

Herhangi bir okula yerleştirmesi yapılmamış (kazanamadı) 589 öğrencinin 438’ i (%74,4) doğru, 23’ü (%3,9) sınavlı bir okulu kazandı olarak, 128’i (%21,7) ise sınavsız bir okulu kazandı olarak tahmin etmiştir. Sınavlı bir okul kazanan 615 öğrencinin 57’ sini (%9,3) kazanamadı, 441’ini (%71,7) doğru tahmin etmiş, 117’sini (%19) ise sınavsız bir okul kazandı olarak tahminlemiştir. Sınavsız bir okulu kazanan 596 öğrenciden 224’ ünü (%37,6) kazanamadı, 93’ünü (%15,6) sınavlı bir okul kazandı olarak, 279’unu (%46,8) ise doğru olarak tahminlemiştir.

		Predicted			Σ
		kazanamadı	sınavlı	sınavsız	
Actual	kazanamadı	417	58	114	589
	sınavlı	64	453	98	615
	sınavsız	204	107	285	596
Σ		685	618	497	1800

		Predicted			Σ
		kazanamadı	sınavlı	sınavsız	
Actual	kazanamadı	70.8 %	9.8 %	19.4 %	589
	sınavlı	10.4 %	73.7 %	15.9 %	615
	sınavsız	34.2 %	18.0 %	47.8 %	596
Σ		685	618	497	1800

Şekil 4.3 Naive Bayes karmaşıklık matrisi.

Şekil 4.3 de gösterilen Naive Bayes algoritması ile tahminleme için kullanılan 1800 öğrencinin verilerine göre;

Herhangi bir okula yerleştirmesi yapılmamış (kazanamadı) 589 öğrencinin 417' si (%70,8) doğru, 58'i (%9,8) sınavlı bir okulu kazandı olarak, 114'ü (%19,4) ise sınavsız bir okulu kazandı olarak tahmin etmiştir. Sınavlı bir okul kazanan 615 öğrencinin 64'ü (%10,4) kazanamadı, 453'ünü (%73,7) doğru tahmin etmiş, 98' ini (%15,9) ise sınavsız bir okul kazandı olarak tahminlemiştir. Sınavsız bir okulu kazanan 596 öğrenciden 204' ü (%34,2) kazanamadı, 107'sini (%18) sınavlı bir okul kazandı olarak, 285'ini (%47,8) ise doğru olarak tahminlemiştir.

		Predicted			
		kazanamadı	sınavlı	sınavsız	Σ
Actual	kazanamadı	400	31	158	589
	sınavlı	68	429	118	615
	sınavsız	190	107	299	596
	Σ	658	567	575	1800

		Predicted			
		kazanamadı	sınavlı	sınavsız	Σ
Actual	kazanamadı	67.9 %	5.3 %	26.8 %	589
	sınavlı	11.1 %	69.8 %	19.2 %	615
	sınavsız	31.9 %	18.0 %	50.2 %	596
	Σ	658	567	575	1800

Şekil 4.4 Rastgele Orman algoritması karmaşıklık matrisi.

Şekil 4.4'te gösterilen rastgele orman algoritması ile tahminleme için kullanılan 1800 öğrencinin verilerine göre;

Herhangi bir okula yerleştirmesi yapılmamış (kazanamadı) 589 öğrencinin 400' ü (%67,9) doğru, 31'i (%5,3) sınavlı bir okulu kazandı olarak, 158'i (%26,8) ise sınavsız bir okulu kazandı olarak tahmin etmiştir. Sınavlı bir okul kazanan 615 öğrencinin 68'i (%11,1) kazanamadı, 429'unu (%69,8) doğru tahmin etmiş, 118'ini (%19,2) ise sınavsız bir okul kazandı olarak tahminlemiştir. Sınavsız bir okulu kazanan 596 öğrenciden 190' ını (%31,9) kazanamadı, 107'sini (%18) sınavlı bir okul kazandı olarak, 299'unu (%50,2) ise doğru olarak tahminlemiştir.

		Predicted			Σ
		kazanamadı	sınavlı	sınavsız	
Actual	kazanamadı	447	20	122	589
	sınavlı	110	429	76	615
	sınavsız	321	76	199	596
Σ		878	525	397	1800

		Predicted			Σ
		kazanamadı	sınavlı	sınavsız	
Actual	kazanamadı	75.9 %	3.4 %	20.7 %	589
	sınavlı	17.9 %	69.8 %	12.4 %	615
	sınavsız	53.9 %	12.8 %	33.4 %	596
Σ		878	525	397	1800

Şekil 4.5 kNN algoritması karmaşıklık matrisi.

Şekil 4.5'te gösterilen kNN algoritması ile tahminleme için kullanılan 1800 öğrencinin verilerine göre;

Herhangi bir okula yerleştirmesi yapılmamış (kazanamadı) 589 öğrencinin 447'si (%75,9) doğru, 20'si (%3,4) sınavlı bir okulu kazandı olarak, 122'si (%20,7) ise sınavsız bir okulu kazandı olarak tahmin etmiştir. Sınavlı bir okul kazanan 615 öğrencinin 110'u (%17,9) kazanamadı, 429'unu (%69,8) doğru tahmin etmiş, 76'sını (%12,4) ise sınavsız bir okul kazandı olarak tahminlemiştir. Sınavsız bir okulu kazanan 596 öğrenciden 321'ini (%53,9) kazanamadı, 76'sını (%12,8) sınavlı bir okul kazandı olarak, 199'unu (%33,4) ise doğru olarak tahminlemiştir.

		Predicted			
		kazanamadı	sınavlı	sınavsız	Σ
Actual	kazanamadı	358	53	178	589
	sınavlı	61	425	129	615
	sınavsız	210	134	252	596
	Σ	629	612	559	1800

		Predicted			
		kazanamadı	sınavlı	sınavsız	Σ
Actual	kazanamadı	60.8 %	9.0 %	30.2 %	589
	sınavlı	9.9 %	69.1 %	21.0 %	615
	sınavsız	35.2 %	22.5 %	42.3 %	596
	Σ	629	612	559	1800

Şekil 4.6 Karar ağacı algoritması karmaşıklık matrisi.

Şekil 4.6’da gösterilen karar ağacı algoritması ile tahminleme için kullanılan 1800 öğrencinin verilerine göre;

Herhangi bir okula yerleştirmesi yapılmamış (kazanamadı) 589 öğrencinin 358’i (%60,8) doğru, 53’si (%9) sınavlı bir okulu kazandı olarak, 178’si (%30,2) ise sınavsız bir okulu kazandı olarak tahmin etmiştir. Sınavlı bir okul kazanan 615 öğrencinin 61’i (%9,9) kazanamadı, 425’ini (%69,1) doğru tahmin etmiş, 129’ unu (%21) ise sınavsız bir okul kazandı olarak tahminlemiştir. Sınavsız bir okulu kazanan 596 öğrenciden 210’unu (%35,2) kazanamadı, 134’ünü (%22,5) sınavlı bir okul kazandı olarak, 252 sini (%42,3) ise doğru olarak tahminlemiştir.

		Predicted			
		kazanamadı	sınavlı	sınavsız	Σ
Actual	kazanamadı	303	27	259	589
	sınavlı	96	362	157	615
	sınavsız	219	80	297	596
	Σ	618	469	713	1800

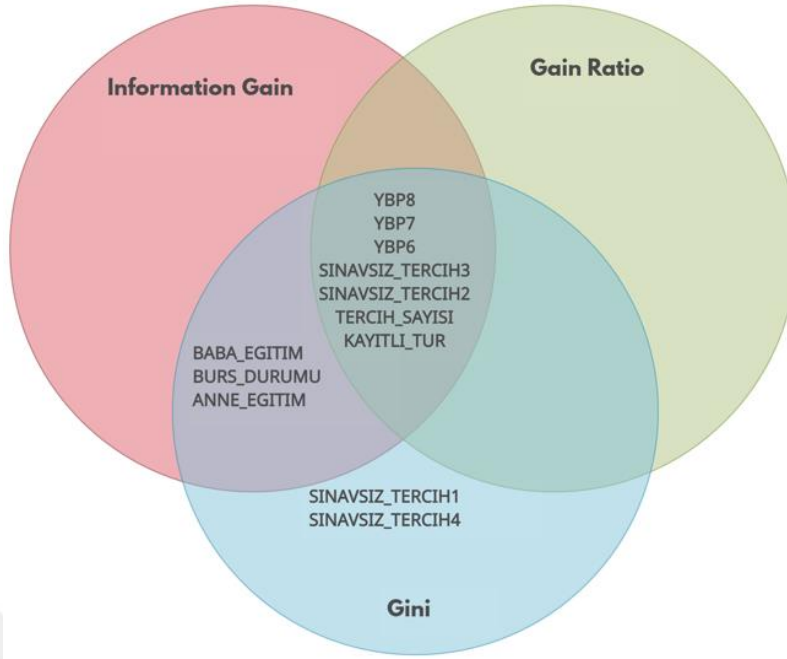
		Predicted			
		kazanamadı	sınavlı	sınavsız	Σ
Actual	kazanamadı	51.4 %	4.6 %	44.0 %	589
	sınavlı	15.6 %	58.9 %	25.5 %	615
	sınavsız	36.7 %	13.4 %	49.8 %	596
	Σ	618	469	713	1800

Şekil 4.7 DVM algoritması karmaşıklık matrisi.

Şekil 4.7’de gösterilen DVM algoritması ile tahminleme için kullanılan 1800 öğrencinin verilerine göre bu algoritma;

Herhangi bir okula yerleştirmesi yapılmamış (kazanamadı) 589 öğrencinin 303’ü (%51,4) doğru, 27’si (%4,6) sınavlı bir okulu kazandı olarak, 259’u (%44) ise sınavsız bir okulu kazandı olarak tahmin etmiştir. Sınavlı bir okul kazanan 615 öğrencinin 96’sı (%15,6) kazanamadı, 362’ sini (%58,9) doğru tahmin etmiş, 157’sini (%25,5) ise sınavsız bir okul kazandı olarak tahminlemiştir. Sınavsız bir okulu kazanan 596 öğrenciden 219’ unu (%36,7) kazanamadı, 60’ini (%13,4) sınavlı bir okul kazandı olarak, 297’sini (%49,8) ise doğru olarak tahminlemiştir.

Özellik seçim yöntemlerine göre en önemli değişkenler Şekil 4.8’de şema olarak gösterilmiştir.



Şekil 4.8 Özellik seçim yöntemlerine göre önemli değişkenler.

Şekil 4.8’ de verilen şemaya bakıldığında YBP8, YBP7, YBP6, SINAVSIZ_TERCIH3, SINAVSIZ_TERCIH2, TERCIH_SAYISI ve KAYITLI_TUR değişkenleri her 3 yöntemde de en iyi on değişken arasında yer almaktadır. BABA_EGITIM, BURS_DURUMU ve ANNE_EGITIM değişkenlerinin, information gain ve gini özellik seçim yöntemlerinde ortak olduğunu, SINAVSIZ_TERCIH1 ve SINAVSIZ_TERCIH4 değişkenlerinin ise sadece gini özellik seçim yönteminde en önemli değişkenler arasında olduğu görülmektedir. Çizelge 4.2’ de verilen information gain yöntemine göre seçilen en önemli on değişkenin ilk üç sırasında öğrencilerin akademik verilerinin olduğu görülmektedir.

5. TARTIŞMA VE SONUÇ

2006 yılından bu yana kullanılmakta olan e-Okul sisteminde birçok eğitsel verinin yanında öğrencilere ait demografik verilerde bulunmaktadır. Bu öğrenci verileri ile çok sayıda çalışma bulunmakta olup e-Okul sistemi dışında, üniversitelerin öğrenci bilgi yönetim sistemlerinde ve anket vb. yollarla veri toplanarak da yapılan birçok bilimsel çalışma mevcuttur. Bu çalışmalar ve bulunan sonuçlarla ilgili bazıları ise şöyledir:

Abbasoğlu (2020), 2019- 2020 eğitim-öğretim yılında Yalova ilinde 1395 öğrencinin e-Okul sisteminden alınan veriyle yapmış olduğu çalışmada ortaokul öğrencilerinin akademik başarılarını EVM yöntemleri ile tahminlemeye çalışmış bu çalışmada kullandığı değişkenler arasında önemli öznelik olarak bu tezimizle ortak olan anne-babanın birlikte veya ayrı olma durumu, anne-babanın eğitim durumu özelliklerini kullanmıştır. Şekil 4.8 Özellik seçim yöntemlerine göre önemli değişkenlerimizde information gain ve gini özellik seçim yöntemlerimizde ortak olarak seçilen değişkenlerle uyumluluk göstermektedir. En iyi performansı ise bu tezimizde olduğu gibi lojistik regresyon yönteminde elde etmiştir.

Göker (2012), Weka yazılımı ile üniversite giriş sınavında öğrenci başarılarını tahminleme için naïve bayes algoritmasını tercih etmiştir. Tahminleme için 39 veri başlığı içinden özellik seçimi ile bu sayı 20 başlığına indirmişdir. Seçilen özelliklerde ilk sıralarda öğrencilerin akademik verileri tezimizde olduğu gibi öne çıkarken baba mesleği ve aile gelir düzeyi önemli değişken arasındadır. Ancak seçilen özelliklerden cinsiyet verisi tezimizde sonucu etkileyen bir faktör olmamasına karşın bu çalışmada önemli özellikler arasında yer aldığı görülmektedir.

Özdemir (2016) akademik başarı öngörüsüne ilişkin Rstudio programlama ortamında yapmış olduğu doktora çalışmada en yüksek performansı karar ağacı yönteminden elde etmiştir. Analizinde seçmiş olduğu özellikler ise anne ve baba eğitim durumu tezimizde de kullandığımız özelliklerinden olup, anne-babanın ayrı olmadığı ve gelir durumu veri başlıklarında ayrılmaktadır.

Topuz (2021), eğitsel verilerle ilgili yapmış olduğu tez çalışmasında Orange yazılımı ile en yüksek doğruluğa yapay sinir ağıları yönteminde elde etmiştir. Veri analizinde seçimi olduğu veri başlıklarında ise anne eğitim durumu ortaklık göstermektedir.

Şenligil (2023), lise giriş sınavında öğrenci başarılarının makine öğrenmesi teknikleri ile analizinde chi-square ve anova yöntemlerini kullanarak özellik seçimi yapmıştır. Bu seçimlerde e-Okul sisteminden elde ettiği verilerde anne eğitim bilgisi ve cinsiyet bilgisinin tezimizle ortaklı gösterdiği görülmüş ve tek tabanlı modelde elde edilen sonuçlarda yapay sinir ağıları yöntemi ile en iyi sonuca ulaşmıştır.

Maimon ve Rokach (2010) a göre veri madenciliği sayesinde yapılan analizlerle ilk bakışta anlamsız gözükken veri yığınlarından anlamlı ve daha önce öngörülmemiş bilgiler ortaya çıkabilmektedir. Bu çalışma akademik veriler kadar demografik verilerinde tahminlemede etkisi olduğu öğrencinin kayıtlı olduğu okul türü, annenin eğitim durumu gibi veri başlıkları akademik veriler kadar olmasada tahminlemede etkili birer değişken olduğu görülmüştür. Benzer çalışmalarda da anne eğitim durumu ve anne-babanın birlikteliği ortak bir etkili değişken olarak göze çarpmaktadır.

e-Okul sisteminin 2006 yılından bu yana kullanılmakta olduğu ve bu doğrultuda öğrencinin eğitim süresinin 12 yıl olduğu göz önüne alınırsa ilkökul birinci sınıftan başlayıp liseden mezun olana kadar e-Okul sisteminde tüm sınıflarda verisi olan verisi olan çok sayıda öğrenci bulunmaktadır. Bu eğitsel veri madenciliği için çok kıymetli sayılabilecek bir veridir. Şekil 1.1'de verilen geçmişten günümüze ülkemizde uygulanan orta öğretime geçiş sınavlarında 2018 yılından bu yana uygulanmakta olan mevcut liselere yerleştirme sisteminde MEB verilerine göre 2022-2023 eğitim öğretim yılında yaklaşık bir milyon ikiyüzbin sekizinci sınıf öğrencisi bulunmaktadır. Bu öğrencilerin bir liseye devam edecekleri ve EK-3 araştırma izni ile talep edilen verilere göre yine bu tercih yapan öğrencilerin yaklaşık %96 sının Şekil 2.1'de yer alan ilk yerleştirme sürecinde bir liseye yerleştiği, kalan %4 lük yani 48 bine yakın öğrencinin ise yerleşememe stresi yaşadığı değerlendirilmektedir.

Bu tez VM yöntemleri ile tercih konusunda öğrencilere rehberlik edebileceğini ortaya koymakla birlikte e-Okul sisteminde bulunan farklı akademik ve demografik verilerle desteklenerek ve geçmiş yıllar da düşünüldüğünde daha fazla öğrenme verisi ile

öğrencilere tercih öncesi bir tercih robotu gibi bir uygulama konusunda fikir üretebilir. Bunun dışında özellik seçiminde ve benzer çalışmalarda öne çıkan anne eğitim durumu, anne-babanın birlikte olması durumları detaylandırılarak aile yaşamının veriye etkileri de incelenebilir.



6. KAYNAKLAR

- Abbasoğlu B, 2020, Ortaokul Öğrencilerinin Akademik Başarılarının Eğitsel Veri Madenciliği Yöntemleri ile Tahmini, Veri Bilimi, 3(1), 1–10.
- Adak M F, Duralioğlu Ö, 2023, Makine Öğrenmesi Yöntemleri Kullanılarak Öğrencilerin Kazanım Bilgileri ile Sınavlardaki Başarı Durumunun Tahmini, Journal of Intelligent Systems: Theory and Applications, 6(1), 43–51.
- Ajaj I A A, 2021, K-nn Sınıflandırma Yöntemine Dayalı Gri Seviye Eşdizimlilik Matrisi ve Momentum Özellikleri ile Yüzdeki Duyguyu tanıma, Kastamonu Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 49s, Kastamonu.
- Akçapınar G, 2014, Çevrimiçi Öğrenme Ortamındaki Etkileşim Verilerine Göre Öğrencilerin Akademik Performanslarının Veri Madenciliği Yaklaşımı ile Modellenmesi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Doktora Tezi, 96s, Ankara
- Akçapınar G, Coşgun, E, 2019, Öğrencilerin STEM Eğitimi Tercihlerinin Veri Madenciliği Yaklaşımı ile Tahmin Edilmesi, Eğitim Teknolojisi Kuram ve Uygulama, 9(1), 73-88.
- Aksu G, 2018, Pısa Başarısını Tahmin Etmede Kullanılan Veri Madenciliği Yöntemlerinin İncelenmesi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Doktora Tezi, 162s, Ankara
- Akşehirli Ö, Ankaralı H, Cangür Ş, Sungu M A, 2014, Breiman Algoritması Kullanılarak Homojen Alt Grupların Belirlenmesi: Bir Uygulama, Bilişim Teknolojileri Dergisi, 7(1), 19-24.
- Altun M, Kayıkçı K, Irmak S, 2019, Sınıf Öğretmenliği Öğrencilerinin Mezuniyet Notlarının Regresyon Analizi ve Yapay Sinir Ağları Yöntemleriyle Tahmini, e-Uluslararası Eğitim Araştırmaları Dergisi, 10(3), 29 – 43.
- Altunkaynak B, 2017, Veri Madenciliği Yöntemleri ve R Uygulamaları, Seçkin Yayıncılık, Ankara.

- Aydın S, 2007, Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama, Anadolu Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Ana Bilim Dalı, Doktora Tezi, 178s, Eskişehir.
- Bağlıbel M, Samancıoğlu M, Summak M, 2014, Okul Yöneticileri Tarafından E-Okul Uygulamasının Genişletilmiş Teknoloji Kabul Modeline Göre Değerlendirilmesi, Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 13, 331– 348
- Bilgin M, Şentürk İ F, 2019, Danışmanlı ve Yarı Danışmanlı Öğrenme Kullanarak Doküman Vektörleri Tabanlı Tweetlerin Duygu Analizi, Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 21(2), 822-839.
- Breiman L, 2001, Random Forests, Machine Learning, 45, 5-32.
- Bozkurt A, 2016, Öğrenme Analitiği: e-öğrenme, Büyük Veri ve Bireyselleştirilmiş Öğrenme, Açıköğretim Uygulamaları ve Araştırmaları Dergisi, 2(4), 55 – 81.
- Budak H, 2018, Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 22, 21– 31.
- Can S, 2021, Lise Öğrencilerinin Üniversiteye Giriş Başarılarının Eğitsel Veri Madenciliği ile Tahmin Edilmesi, Beykent Üniversitesi, Lisansüstü Eğitim Enstitüsü, Yüksek Lisans Tezi, 123s, İstanbul
- Coşkun C, Baykal A, 2011, Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması, Akademik Bilişim Konferansı, 1– 8.
- Dol S M, Jawandhiya P M, 2023, Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining A survey, Engineering Applications of Artificial Intelligence, 122, 106071.
- Düzgün O, 2017, Veri Madenciliği Yöntemleri ve İşkur İçin Uygulamaya Yönelik Model Önerisi, Çalışma ve Sosyal Güvenlik Bakanlığı, Türkiye İş Kurumu Genel Müdürlüğü, Uzmanlık Tezi, 105s, Ankara.
- Efron B, Tibshirani R, 1991, Statistical Data Analysis in The Computer Age, Science, 253(5018), 390–395.

- Ekinci H, 2022, Eğitsel Veri Madenciliğinin Öğrenci Başarısının Kestirimine Yönelik Kullanımı, İnönü Üniversitesi, Eğitim Bilimleri Enstitüsü, Yüksek Lisans Tezi, 98s, Malatya
- Ersöz A R, 2017, Eğitsel Veri Madenciliği ile Öğrenci Profillerinin Belirlenmesi, Bursa Uludağ Üniversitesi, Eğitim Bilimleri Enstitüsü, Doktora Tezi, 78s, Bursa.
- Forman G, 2003, An Extensive Empirical Study of Feature Selection Metrics for Text Classification, Journal of Machine Learning Research, 3, 1289 –1305.
- Friedl M, Brodley C, 1997, Decision Tree Classification Of Land Cover From Remotely Sensed Data, Remote Sensing of Environment, 61(3), 399 – 409.
- Gökbay İ Z, Oyacı Y, Pahlivan S, 2021, Gen Varyantlarının Madde Kullanım Bozukluklarındaki Davranış Eğilimlerine Etkisinin Gini İndeksi ve Bilgi Kazanım Sınıflandırması Algoritmalarıyla Analizi, IV International Conference on Data Science and Applications (ICONDATA'21), June 4-6-Türkiye, 669 – 679.
- Göker H, 2012, Üniversite Giriş Sınavında Öğrencilerin Başarılarının Veri Madenciliği Yöntemleri ile Tahmin Edilmesi, Yüksek Lisans Tezi, Gazi Üniversitesi, Bilişim Enstitüsü, Ankara
- Güler M, Arslan Z, Çelik D, 2018, Liselere Giriş Sınavına İlişkin Matematik Öğretmenlerinin Görüşleri, Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, 16(1), 337-363.
- Güner N, Çomak E, 2011, Mühendislik Öğrencilerinin Matematik I Derslerindeki Başarısının Destek Vektör Makineleri Kullanılarak Tahmin Edilmesi, Pamukkale University Journal of Engineering Sciences, 17(2), 87 – 96.
- Han J, Kanber M, 2006, Data Mining: Concepts and Techniques, Morgan Kaufmann.
- Harman G, 2021, Destek Vektör Makineleri ve Naive Bayes Sınıflandırma Algoritmalarını Kullanarak Diabetes Mellitus Tahmini, Avrupa Bilim ve Teknoloji Dergisi, 32, 7 – 13.

- Kaur P, Singh M, Josan G S, 2015, Classification And Prediction Based Data Mining Algorithms To Predict Slow Learners In Education Sector, *Procedia Computer Science*, 57, 500 – 508.
- Kavzoğlu T, Çölkesen İ, 2010, Karar Ağaçları ile Uydu Görüntülerinin Sınıflandırılması, *Harita Teknolojileri Elektronik Dergisi*, 2(1), 36 – 45.
- Kuzey C, 2012, Veri Madenciliğinde Destek Vektör Makinaları ve Karar Ağaçları Yöntemlerini Kullanarak Bilgi Çalışanlarının Kurum Performansı Üzerine Etkisinin Ölçülmesi ve Bir Uygulama, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, 341s, İstanbul.
- Ledley R S, 1960, Using Electronic Computers in Medical Diagnosis, *IRE Transactions on Medical Electronics*, (4), 274-280.
- Maimon O, Rokach L, 2005, *Data Mining And Knowledge Discovery Handbook* (Vol. 2, No. 2005), New York: Springer.
- MEB, 2022, Sınavla Öğrenci Alacak Ortaöğretim Kurumlarına İlişkin Merkezî Sınav Başvuru ve Uygulama Kılavuzu
- MEB, 2022, Ortaöğretime Geçiş Tercih ve Yerleştirme Kılavuzu
- MEB, 2023, Milli Eğitim İstatistikleri Örgün Eğitim 2022-23, Ankara.
- Ouhaichi H, Spikol D, Vogel B, 2023, Research Trends In Multimodal Learning Analytics: A Systematic Mapping Study, *Computers and Education: Artificial Intelligence*, 100136.
- Özdemir Ş, 2016, Eğitimde Veri Madenciliği ve Öğrenci Akademik Başarı Öngörüsüne İlişkin Bir Uygulama, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 191s, İstanbul
- Pallathadka H, Wenda A, Ramirez-Asís E, vd., 2023, Classification And Prediction Of Student Performance Data Using Various Machine Learning Algorithms, *Materials today: proceedings*, 80, 3782 – 3785.
- Romero C, Ventura S, 2007, Educational Data Mining: A Survey From 1995 to 2005, *Expert Systems with Applications*, 33(1), 135 – 146.

- Savaş S, Topaloğlu N, YILMAZ M, 2012, Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 11(21), 1 – 23.
- Siemens G, Baker R S D, 2012, Learning Analytics And Educational Data Mining: Towards Communication and Collaboration, In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 252 – 254.
- Şimşek A, 2022, Yeni Medyanın Yükselişi, Nobel akademi Yayıncılık, 476s, Ankara.
- Suna H E, Gelbal S, Özer M, 2020, Fen lisesi öğrencilerinin sosyoekonomik arkaplanı ve yükseköğretime geçişteki tercihleri, Yükseköğretim Dergisi, 10(3), 356 – 370.
- Sönmez F, 2015, Kredi Skorunun Belirlenmesinde Yapay Sinir Ağları ve Karar Ağaçlarının Kullanımı Bir Model Önerisi, Anadolu Bil Meslek Yüksekokulu Dergisi, 40, 1 – 22.
- Şenel S, Alatlı B, 2014, Lojistik regresyon analizinin kullanıldığı makaleler üzerine bir inceleme, Journal of Measurement and Evaluation in Education and Psychology, 5(1), 35 – 52.
- Şenligil M, 2023, Lise Giriş Sınavında Öğrenci Başarılarının Makine Öğrenmesi Teknikleri ile Analizi, Çankırı Karatekin Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 83s, Çankırı
- Topuz S, 2021, Eğitsel Verilerde Weka ve Orange Veri Madenciliği Yazılımlarından Elde Edilen Analiz Sonuçlarının Karşılaştırılması, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Yüksek Lisans Tezi, 79s, Ankara
- Vural C, 2020, Son 10 Yılda Yapılan Liselere Giriş Sınavlarında (Sbs, Teog ve Lgs) Yer Alan Türkçe Dersi Sorularının Yenilenmiş Bloom Taksonomisi Açısından Değerlendirilmesi, Gaziantep Üniversitesi, Eğitim Bilimleri Enstitüsü, Türkçe ve Sosyal Bilimler Eğitimi Ana Bilim Dalı, Yüksek Lisans Tezi, 92s, Gaziantep

EKLER

Ek 1 Veri seti başlıkları

Veri	Gösterim	Aldığı Değer	Tür
Öğrenci Id	OGRENCI_ID	Benzersiz Sayı	Meta veri
Öğrencinin Kayıtlı Olduğu Ortaokulun Türü	KAYITLI_TUR	Ortaokul İmam Hatip Ortaokulu Özel Ortaokul	Kategorik
Öğrenci Cinsiyeti	CINSIYET	Kız Erkek	Kategorik
6. Sınıf Yıl Sonu Başarı Puanı	YBP6	0-100 arasında	Sayısal
7. Sınıf Yıl Sonu Başarı Puanı	YBP7	0-100 arasında	Sayısal
8. Sınıf Yıl Sonu Başarı Puanı	YBP8	0-100 arasında	Sayısal
1. Sıradaki Yerel Okul Tercih Türü	SINAVSIZ_TERCIH1	Anadolu Lisesi Anadolu İmam Hatip Lisesi Anadolu Meslek Lisesi	Kategorik
2. Sıradaki Yerel Okul Tercih Türü	SINAVSIZ_TERCIH2	Anadolu Lisesi Anadolu İmam Hatip Lisesi Anadolu Meslek Lisesi	Kategorik
3. Sıradaki Yerel Okul Tercih Türü	SINAVSIZ_TERCIH3	Anadolu Lisesi Anadolu İmam Hatip Lisesi Anadolu Meslek Lisesi	Kategorik
4. Sıradaki Yerel Okul Tercih Türü	SINAVSIZ_TERCIH4	Anadolu Lisesi Anadolu İmam Hatip Lisesi Anadolu Meslek Lisesi	Kategorik
5. Sıradaki Yerel Okul Tercih Türü	SINAVSIZ_TERCIH5	Anadolu Lisesi Anadolu İmam Hatip Lisesi Anadolu Meslek Lisesi	Kategorik
TERCİH SAYISI	TERCİH SAYISI	1-5 arasında	Sayısal

Ek 1 (Devam)Veri seti başlıkları

Veri	Gösterim	Aldığı Değer	Tür
Devlet Bursu Alıp Almadığı	BURS_DURUMU	0 (Almıyor) 1 (Burs Alıyor)	Kategorik
Ailenin Gelir Düzeti	GELIR_DUZEYI	Belirsiz Çok İyi Çok Kötü Düşük İyi Orta	Kategorik
Annenin ve Babanın Eğitim Durumu	ANNE_EGITIM BABA_EGITIM	Belirsiz 3 Yıllık Eğitim Enstitüsü Doktora İlkokul İlköğretim İlköğretim Öğrencisi Lisans Lisansüstü Lise Okuma Yazma Biliyor Okuma Yazma Bilmiyor Ortaokul Yüksek Okul (2 Yıllık)	Kategorik
Yabancı Uyruklu Öğrenci Olup Olmadığı	YABANCI_UYRUKLU	0 (Hayır) 1 (Evet)	Kategorik
Sosyal etkinlik yapıp yapmadığı	ETKINLIK	0 (Hayır) 1 (Evet)	Kategorik
Ailesi ile oturup oturmadığı	AILESIYLE_OTURAN	-1 (Bilgi Yok) 0 (Ailesiyle) 1 (Annesiyle) 2 (Babasıyla) 3 (Velisiyle)	Kategorik
Kendi evinde oturup oturmadığı	KENDI_EVINDE_OTURAN	1 (Kira) 2 (Lojman) 3 (Kendi evi)	Kategorik

Ek 1 (Devam)Veri seti başlıkları

Veri	Gösterim	Aldığı Değer	Tür
Kaynaştırma Öğrencisi Olup Olmadığı	KAYNASTIRMA	-1 (Kaynaştırma Öğrencisi Değil) 1 (Okulda eğitim alıyor- Kaynaştırma Öğrencisi) 2 (Hastanede eğitim alıyor- Kaynaştırma Öğrencisi) 3 (Evde eğitim alıyor- Kaynaştırma Öğrencisi) 4 (Okulda Grup Eğitimine Hazırlık Alıyor - Kaynaştırma Öğrencisi)	Kategorik

Ek 2 Öğrencilerin tercih dağılımı

Yerleşme Durumu	Yerleşemeyen	Sınavlı Bir Okula Yerleşen	Sınavsız Bir Okula Yerleşen	Genel Toplam
AL+AL+AL	880	505	526	1911
AL+AL	260	334	174	768
AL	212	610	100	922
AMP+AMP+AMP	86	41	147	274
AMP	77	24	71	172
AMP+AMP	40	17	73	130
AL+AL+AL+AMP	40	16	71	127
AMP+AL+AL	37	7	19	63
AL+AL+AMP	33	42	72	147
AMP+AL	32	7	19	58
AL+AL+AMP+AL	29	17	40	86
AMP+AMP+AMP+AL+AL	27	19	59	105
AL+AMP	25	16	28	69
AMP+AL+AL+AL	23	4	19	46
AL+AL+AL+AMP+AMP	22	22	89	133
AL+AMP+AL	19	6	14	39
AMP+AMP+AL	15	3	26	44
AL+AMP+AL+AL	14	4	7	25
AL+AMP+AMP	13	12	35	60
AMP+AMP+AMP+AL	11	7	29	47
AL+AL+AMP+AMP+AMP	11	7	20	38
AMP+AMP+AL+AL+AL	9	4	5	18
AL+AL+AMP+AL+AMP	8	7	8	23
AL+AL+AMP+AMP	7	5	23	35
AMP+AL+AL+AL+AMP	6	2	9	17
AMP+AMP+AL+AL	5	9	7	21
AMP+AL+AL+AMP+AMP	5	2	3	10
AL+AMP+AMP+AMP	4	4	17	25
AMP+AL+AL+AMP	4	2	10	16
AL+AL+AMP+AMP+AL	4	1	7	12
AL+AMP+AMP+AL	4	2	5	11
AMP+AMP+AL+AMP+AL	4	2	2	8
AMP+AL+AMP+AL+AL	4		3	7
AL+AMP+AMP+AL+AL	3	3	9	15
AL+AL+AMP+AMP+AL	4	1	7	12
AMP+AMP+AL+AMP	3	2	10	15
AMP+AMP+AL+AMP	3	2	10	15

Ek 2 (Devam) Öğrencilerin tercih dağılımı

Yerleşme Durumu	Yerleşemeyen	Sınavlı Bir Okula Yerleşen	Sınavsız Bir Okula Yerleşen	Genel Toplam
AİHL	2	45	14	61
AMP+AL+AMP	2	7	23	32
AL+AMP+AMP+AMP+AL	2		10	12
AL+AMP+AL+AL+AMP	2	1	6	9
AMP+AL+AMP+AL	2	1	6	9
AMP+AL+AL+AMP+AL	2	2	1	5
AL+AMP+AL+AMP	2	1	1	4
AMP+AL+AMP+AMP	1	3	13	17
AİHL+AL+AL	1	11	3	15
AL+AMP+AMP+AL+AMP	1		7	8
AL+AL+AMP+AL+AİHL	1	1	3	5
AL+AİHL		13	3	16
AİHL+AİHL		13	2	15
AL+AL+AİHL+AL		10	3	13
AL+AİHL+AMP		5	7	12
AİHL+AL		9	2	11
AL+AL+AİHL		9	2	11
AL+AMP+AİHL		5	6	11
AL+AL+AL+AİHL+AİHL		7	2	9
AL+AL+AL+AİHL+AMP		1	7	8
AL+AL+AL+AMP+AİHL		2	5	7
AL+AİHL+AL		6		6
AL+AL+AİHL+AL+AİHL		4	2	6
AL+AL+AL+AİHL		2	4	6
AL+AMP+AL+AMP+AL		2	4	6
AMP+AMP+AİHL		2	4	6
AL+AMP+AMP+AİHL			6	6
AİHL+AL+AL+AL		5		5
AL+AL+AİHL+AMP+AMP		4	1	5
AİHL+AİHL+AİHL		3	2	5
AMP+AİHL		2	3	5
AİHL+AL+AL+AMP		4		4
AL+AL+AMP+AİHL+AL		2	2	4
AMP+AL+AİHL		2	2	4
AL+AL+AMP+AMP+AİHL		1	3	4
AMP+AMP+AMP+AL+AİHL			4	4

Ek 2 (Devam) Öğrencilerin tercih dağılımı

Yerleşme Durumu	Yerleşemeyen	Sınavlı Bir Okula Yerleşen	Sınavsız Bir Okula Yerleşen	Genel Toplam
AİHL+AL+AMP	3			3
AL+AİHL+AİHL	2	1		3
AL+AMP+AL+AİHL	2	1		3
AİHL+AİHL+AİHL+AL+AL	1	2		3
AL+AL+AİHL+AMP	1	2		3
AL+AMP+AİHL+AL+AL	1	2		3
AMP+AMP+AİHL+AMP	1	2		3
AMP+AMP+AL+AİHL	1	2		3
AL+AL+AMP+AİHL		3		3
AL+AMP+AL+AİHL+AMP		3		3
AMP+AİHL+AMP		3		3
AMP+AL+AL+AL+AİHL		3		3
AİHL+AİHL+AL	2			2
AİHL+AMP+AL	1	1		2
AL+AİHL+AL+AL+AMP	1	1		2
AL+AMP+AİHL+AMP	1	1		2
AL+AMP+AMP+AMP+AİHL	1	1		2
AMP+AİHL+AL+AL+AL	1	1		2
AMP+AL+AMP+AL+AMP	1	1		2
AMP+AMP+AMP+AİHL+AİHL	1	1		2
AL+AİHL+AİHL+AL+AL		2		2
AL+AİHL+AMP+AMP		2		2
AL+AL+AİHL+AİHL		2		2
AL+AMP+AL+AMP+AMP		2		2
AMP+AL+AMP+AİHL		2		2
AMP+AL+AMP+AİHL+AMP		2		2
AMP+AMP+AİHL+AMP+AL		2		2
AMP+AMP+AMP+AİHL		2		2
AİHL+AİHL+AİHL+AL	1			1
AİHL+AİHL+AL+AL+AL	1			1
AİHL+AİHL+AMP	1			1
AİHL+AİHL+AMP+AMP	1			1
AİHL+AL+AİHL	1			1
AİHL+AL+AİHL+AL+AMP	1			1
AİHL+AL+AL+AİHL	1			1
AİHL+AL+AL+AİHL+AİHL	1			1

Ek 2 (Devam) Öğrencilerin tercih dağılımı

Yerleşme Durumu	Yerleşemeyen	Sınavlı Bir Okula Yerleşen	Sınavsız Bir Okula Yerleşen	Genel Toplam
AİHL+AL+AL+AMP+AİHL	1			1
AİHL+AL+AL+AMP+AMP	1			1
AİHL+AL+AMP+AL+AL	1			1
AL+AİHL+AİHL+AİHL	1			1
AL+AİHL+AİHL+AL	1			1
AL+AİHL+AL+AİHL	1			1
AL+AİHL+AL+AL	1			1
AL+AİHL+AMP+AL+AL	1			1
AL+AMP+AİHL+AMP+AMP	1			1
AMP+AİHL+AL+AİHL+AMP	1			1
AMP+AİHL+AL+AMP	1			1
AMP+AL+AİHL+AL	1			1
AMP+AL+AİHL+AMP+AL	1			1
AMP+AL+AL+AİHL+AİHL	1			1
AMP+AMP+AİHL+AL+AL	1			1
AMP+AMP+AİHL+AL+AMP	1			1
AMP+AMP+AMP+AİHL+AL	1			1
AİHL+AİHL+AL+AL			1	1
AİHL+AİHL+AL+AMP+AL			1	1
AİHL+AİHL+AMP+AL+AMP			1	1
AİHL+AL+AMP+AİHL+AMP			1	1
AİHL+AMP+AMP			1	1
AL+AİHL+AL+AMP+AMP			1	1
AL+AL+AİHL+AL+AMP			1	1
AL+AMP+AMP+AİHL+AMP			1	1
AMP+AİHL+AL			1	1
AMP+AİHL+AMP+AİHL			1	1
AMP+AİHL+AMP+AL			1	1
AMP+AİHL+AMP+AL+AMP			1	1
AMP+AİHL+AMP+AMP+AİHL			1	1
AMP+AL+AİHL+AL+AL			1	1
AMP+AL+AİHL+AMP			1	1
AMP+AL+AL+AİHL			1	1
AMP+AL+AL+AİHL+AL			1	1
AMP+AL+AMP+AMP+AİHL			1	1
AMP+AMP+AİHL+AİHL+AMP			1	1

Ek 2 (Devam) Öğrencilerin tercih dağılımı

Yerleşme Durumu	Yerleşemeyen	Sınavlı Bir Okula Yerleşen	Sınavsız Bir Okula Yerleşen	Genel Toplam
AMP+AMP+AİHL+AL			1	1
AMP+AMP+AL+AL+AİHL			1	1

Ek 3 Arařtırma İzni



T.C.
MİLLÎ EđİTİM BAKANLIđI
Bilgi İřlem Genel M¼d¼rl¼đ¼

Sayı : E-65968543-622.03-78199263
Konu : Veri Talebi

13.06.2023

Sayın Ümit ÇELİK

İlgi : 09.06.2023 tarihli ve 78002754 sayılı dilekçeniz.

İlgi dilekçeniz incelenmiş olup yüksek tez lisansınızda kullanmak üzere talep ettiđiniz veriler, MEB Bulut sistemi üzerinden "~~Ümit Çelik~~" adresiyle paylaşılmıştır.

Bu kapsamda yapacağımız ulusal, uluslararası makale, bildiri ve tez çalışmalarında sınırlı kalmak koşulu ile verinin tarafınızca kullanılmasında herhangi bir sakınca bulunmamaktadır.
Bilgilerini rica ederim

Özg¼r TÜRK
Bakan a.
Bilgi İřlem Genel M¼d¼r¼