

T.C.
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI



LİSANSÜSTÜ PROGRAMLARA ÖĞRENCİ KABULÜNDE
KULLANILAN KRİTER PUANLARIN İŞLEVSELLİĞİ
ÜZERİNE BİR VERİ MADENCİLİĞİ ÇALIŞMASI

YÜKSEK LİSANS TEZİ

SELCAN KESER

TEZ DANIŞMANI

Prof. Dr. ZEKERİYA NARTGÜN

BOLU, OCAK - 2023

KABUL VE ONAY SAYFASI

Selcan KESER tarafından hazırlanan “**LİSANSÜSTÜ PROGRAMLARA ÖĞRENCİ KABULÜNDE KULLANILAN KRİTER PUANLARIN İŞLEVSELLİĞİ ÜZERİNE BİR VERİ MADENCİLİĞİ ÇALIŞMASI**” adlı tez çalışması jürimiz tarafından Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme Ve Değerlendirme Bilim Dalı’nda Yüksek Lisans Tezi olarak oy birliği ile kabul edilmiştir. 13/01/2023

Jüri Üyeleri

İmza

Danışman
Prof. Dr. Zekeriya NARTGÜN
Bolu Abant İzzet Baysal Üniversitesi

.....

Üye
Dr. Öğr. Üyesi Süleyman DEMİR
Sakarya Üniversitesi

.....

Üye
Dr. Öğr. Üyesi Alperen YANDI
Bolu Abant İzzet Baysal Üniversitesi

.....

Lisansüstü Eğitim Enstitüsü Onayı

Prof. Dr. İbrahim KÜRTÜL
Lisansüstü Eğitim Enstitüsü Müdürü

ETİK BEYAN

Bolu Abant İzzet Baysal Üniversitesi, Lisansüstü Eğitim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu bildirir,

aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Teze ilişkin 15/02/2023 tarihinde Turnitin adlı intihal tespit programından enstitü müdürlüğünce belirlenen filtrelemeler uygulanarak alınmış olan benzerlik raporuna göre, tezin benzerlik oranı % 21 olarak tespit edilmiştir.

Bu çalışma için Sosyal Bilimlerde İnsan Araştırmaları Etik Kurulundan 2021/341 sayısı ile etik izin alınmıştır.

.....
SELCAN KESER

ÖZET

**LİSANSÜSTÜ PROGRAMLARA ÖĞRENCİ KABULÜNDE
KULLANILAN KRİTER PUANLARIN İŞLEVSELLİĞİ ÜZERİNE BİR
VERİ MADENCİLİĞİ ÇALIŞMASI
YÜKSEK LİSANS TEZİ
SELCAN KESER
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI
(TEZ DANIŞMANI: PROF. DR. ZEKERİYA NARTGÜN)
BOLU, OCAK - 2023
XII + 49**

Bu çalışma lisansüstü eğitim programlarına öğrenci kabulünde yaygın olarak kullanılan ALES, yabancı dil, bilim sınavı ve mezuniyet puanı kriterlerinin işlevselliğini bir başka deyişle mezun olma durumunu yordayıp yordamadığını belirlemek amacı ile yapılmıştır. Çalışma kapsamında lisansüstü programlara kabul edilmiş öğrencilere ait kriter puanların, öğrencilerinin programdan mezun olma durumları üzerinde etkili bir yordayıcı olup olmadığı incelenmiştir. Bahsi geçen puanlar Lojistik regresyon, Karar ağaçları (C4.5), K en yakın komşu (KNN) gibi veri madenciliğine ait sınıflandırma (Tahmin) algoritmalarında kullanılarak, bu algoritmalarla öğrencilerin kabul edildikleri programları bitirip bitirmeme durumlarını tahmin etmeleri beklenmiştir. Analizler Weka 3.9.4 programı aracılığı ile gerçekleştirilmiştir. Araştırmada veri seti olarak, Bolu Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsündeki tezsiz yüksek lisans, tezli yüksek lisans ve doktora programlarına 2011-2016 yılları arasında kayıt yaptırmış 740 öğrenciye ait ALES, yabancı dil, Bilim sınavı ve Mezuniyet puanları ile programlardan mezun olma durumları kullanılmıştır. Veri setinde bulunan 740 bireyin %66'sına ait veriler algoritmalar tarafından öğrenme verisi (Learning data) olarak kullanılarak mezuniyet durumunu açıklayan modeller oluşturulmuş, geri kalan %33 'üne ait veriler ise oluşturulan modelleri test etmek üzere kullanılmıştır. Araştırma sonucu elde edilen bulgulara göre ALES, yabancı dil, mezuniyet ve bilim sınavı puanlarından biri ya da birkaçı kullanılarak öğrencilerin programlardan mezun olma durumlarını açıklayabilen bir model her üç sınıflandırma algoritması tarafından da oluşturulamamıştır. Bu bulgular ışığında lisansüstü programlara öğrenci kabulünde kullanılan kriter puanların işlevsel olmadığı bir başka deyişle öğrencilerin program sürecindeki mezun olma durumlarını yordayamadığı sonucuna ulaşılmıştır.

ANAHTAR KELİMELER: Eğitsel Veri Madenciliği, Veri Madenciliği, Weka, Sınıflandırma Algoritmaları, Öğrenci Kabul Kriterleri

ABSTRACT

**TITLE OF THE THESIS A DATA MINING STUDY ON THE
FUNCTIONALITY OF THE CRITERIA SCORES USED IN STUDENT
ADMISSION TO GRADUATE PROGRAMS**

MA THESIS

SELCAN KESER

BOLU ABANT İZZET BAYSAL UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF EDUCATIONAL SCIENCES

EDUCATIONAL MEASUREMENT AND EVALUATION

(SUPERVISOR: PROF. DR. ZEKERİYA NARTGÜN)

BOLU, JANUARY 2023

XII + 49

This study aims to examine to what extent the criteria of ALES (Academic Staff and Graduate Education Entrance Examination), foreign language proficiency, graduate entrance exam and undergraduate GPA, widely used in the admission of students to graduate programs, allow an effective evaluation process. The study investigates whether the criterion scores of the students admitted to the graduate programs are an effective predictor of their graduation from the program. Classifications (Prediction) algorithms for data mining such as Logistic regression, Decision trees (C4.5), K-nearest neighbors (KNN) were run on the aforementioned scores, with the expectation that these algorithms would predict whether students would complete the programs they were admitted to. The analyzes were carried out through the Weka 3.9.4 software. As the data set, the study uses ALES, foreign language proficiency, Graduate entrance exam, and GPA scores of 740 students admitted in non-thesis master's, master's and doctoral programs at Bolu Abant İzzet Baysal University, Institute of Educational Sciences between 2011-2016, as well as their graduation status. While the data from 66% of the 740 individuals in the data set were used by algorithms as learning data to create models that explain the graduation status, and the data from the remaining 33% were used to test the models created. The findings obtained from the study shows that none of the three classification algorithms managed to create a model that could explain the graduation status of students from the programs by using one or several of ALES, foreign language proficiency, GPA and graduate entrance exam scores. This finding leads us to conclude that the criterion scores used in the admission of students to graduate programs fail to predict the student graduate during the program process.

KEYWORDS: Educational Data Mining, Data Mining, Weka, Classification Algorithms, Student Admission Criteria

İÇİNDEKİLER

	<u>Sayfa</u>
KABUL VE ONAY SAYFASI	iii
ETİK BEYAN.....	iv
ÖZET.....	v
ABSTRACT	vi
İÇİNDEKİLER	vii
ŞEKİL LİSTESİ.....	ix
TABLO LİSTESİ	x
KISALTMA VE SEMBOLLER LİSTESİ	xi
TEŞEKKÜR.....	xii
1. GİRİŞ.....	1
1.1 Problem Cümlesi	4
1.2 Alt Problemler	4
1.3 Araştırmanın Önemi	5
1.4 Sayıtlar.....	5
1.5 Sınırlılıklar.....	5
1.6 Tanımlar.....	6
2. ARAŞTIRMANIN KURAMSAL TEMELİ VE İLGİLİ	
ARAŞTIRMALAR.....	7
2.1 Araştırmanın Kuramsal Temelleri	7
2.1.1 Lisansüstü Eğitim	7
2.1.2 Veri Madenciliği.....	7
2.1.3 Eğitsel Veri Madenciliği.....	9
2.1.4 Veri Madenciliği Modelleri	9
2.1.5 Sınıflama Algoritmaları Ve Model Başarım Ölçütleri	12
2.2 İlgili Çalışmalar	21
3. YÖNTEM.....	26
3.1 Araştırmanın Modeli.....	26
3.2 Çalışma Grubu	26
3.3 Veri Toplama	27
3.4 Verilerin Analizi	27
4. BULGULAR	29
4.1 Birinci Alt Probleme İlişkin Bulgular	30
4.2 İkinci Alt Probleme İlişkin Bulgular	31
4.3 Üçüncü Alt Probleme İlişkin Bulgular	33
5. TARTIŞMA	38
6. SONUÇ VE ÖNERİLER	40
6.1 Sonuç	40
6.2 Öneriler.....	41

6.2.1 Arařtırma Bulgularına Dayalı Öneriler	41
6.2.2 Diğer Arařtırmacılar İçin Öneriler	41
7. KAYNAKLAR.....	42
8. EKLER.....	46



ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Veri Madenciliği Disiplinleri (Han ve Kamber, 2006).....	8
Şekil 2.2. Veri Madenciliği Modelleri	10
Şekil 2.3. Örnek Küme Yapısı	11
Şekil 2.4. Örnek karar ağacı yapısı	13
Şekil 2.5. Knn algoritması çalışma prensibi örneği	16
Şekil 2.6. Karışıklık matrisi örneği	17
Şekil 4.1. ALES puanları ve yabancı dil puanlarına göre mezuniyet durumu..	35
Şekil 4.2. ALES sınavı puanları ve bir önceki programdan mezuniyet puanlarına ait plot grafiği	36
Şekil 4.3. Mezuniyet durumlarına göre; bilim sınavı puanları, ales, yabancı dil sınavı puanlarına ait üç boyutlu plot grafiği.....	37

TABLO LİSTESİ

Sayfa

Tablo 2.1. Kappa istatistiği uyum gücü değerleri	19
Tablo 4.1. Veri setine ait istatistiksel bilgiler.....	29
Tablo 4.2. Veri setine ilişkin cinsiyet, program türü ve mezunluk durumu bilgileri	29
Tablo 4.3. KNN algoritması karışıklık matrisi.....	30
Tablo 4.4. KNN algoritmasına ait model başarımlar ölçütü sonuçları.....	31
Tablo 4.5. Karar ağaçları (C4.5) algoritmasına ait karışıklık matrisi.....	32
Tablo 4.6. Karar ağaçları (C4.5) algoritmasına ait model başarımlar ölçütü sonuçları	32
Tablo 4.7. Lojistik regresyon algoritması karışıklık matrisi	33
Tablo 4.8. Lojistik Regresyon Algoritmasına Ait Model Başarımlar Ölçütü.....	34
Tablo 4.9. Üç algoritmaya algoritmaya ait analiz sonuçları.....	35

KISALTMA VE SEMBOLLER LİSTESİ

ALES	: Akademik Personel Ve Lisansüstü Eğitime Giriş Sınavı
C4.5	: Yapay Sinir Ağları (J48)
K	: Kappa İstatistiği
K-NN	: K- Nearest Neighbour (K En Yakın Komşu)
LR	: Lojistik Regresyon
MAE	: Mean Absolute Error (Mutlak hata ortalaması)
RAE	: Relative Absolute Error (Göreceli mutlak hata)
RMSE	: Root Mean Squared Error (Hata kareler ortalaması kökü)
RRSE	: Root relative squared error (Bağıl karekök hatası)
WEKA	: Waikato Environment For Knowledge Analysis

TEŞEKKÜR

Bu çalışmayı yürütürken desteğini esirgemeyen, değerli bilgilerini benimle paylaşan, zamanını ayırıp sabırla ve büyük bir ilgiyle bana faydalı olabilmek için elinden geleni yapan, güler yüzünü ve samimiyetini benden esirgemeyen kıymetli danışman hocam sayın Prof. Dr. Zekeriya NARTGÜN'e teşekkürlerimi sunarım.

Araştırmamda kullandığım: bilgi ve belgeler için Abant İzzet Baysal Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğü'ne teşekkür ederim.

Her zaman ilgi ve desteklerini benden esirgemeyen, beni bu günlere özenle yetiştirerek getiren annem Nezahat KESER ve babam Kemal KESER'e teşekkür ederim.

Her an yanımda hissettiğim, desteğini benden esirgemeyen tüm zorlukları benimle birlikte göğüsleyen kardeşim Oğuzhan KESER'e çok teşekkür ederim.

Değerli arkadaşım Duygu ÖZDEMİR'e bana bu süreçte verdiği destek için beni yüreklendirdiği ve her ihtiyacım olduğu anda yanımda bulabildiğim için teşekkür ederim.

1. GİRİŞ

Ülkemizde lisans düzeyinde eğitimini başarıyla tamamlayan birey sayısının çoğalması: bireyleri daha üst düzey eğitim kademelerine devam etmeye ve farklı birçok niteliklerle donanmaya mecbur bırakmıştır (Bozan, 2012). Bu anlamda bireylerin kendilerini geliştirmek, seçtikleri alanda uzmanlaşmak ve iş dünyasında kendilerine yer edinmek ya da gelirini ve sosyal statüsünü artırmak amacıyla lisansüstü öğretime yatırımı bir zorunluluk olarak gördükleri söylenebilir (Çıkrıkçı-Demirtaşlı, 2002). Sadece ülkemizde değil lisansüstü eğitim dünyada da bir alanda uzmanlaşmak için zorunluluk haline gelmiştir. Lisansüstü eğitim: eğitimde lisans kademesini izleyen bir üst dereye götüren, bilim ve teknolojiyi araştırmalar yoluyla üreten ve bunun yanında ülkelerin kalkınması gibi bir çok önemli duruma yön veren, alanında uzman bireyler yetiştirmeyi amaç edinen eğitim programıdır (Karaman ve Bakırcı, 2010). Lisansüstü eğitimin kalitesinin ve verimini artırması, tüm bu hedeflerine ulaşılabilmesi için lisansüstü eğitime kabul edilecek bireylerin seçimi de oldukça önemlidir.

Ülkemizde bu uygulamalarda genellikle öğrencinin akademik hazır bulunuşlukları konusunda bilgi verdiği düşünülen bazı puanların belirli kriterlere göre ağırlıklandırılması ile elde edilmiş yeni bir ağırlıklandırılmış puan kullanılır. Kullanılan puanlar ve ağırlıklandırma şekilleri yükseköğretim kurumlarının senato kararları ile esnetilebilmekle birlikte, yaygın olarak ALES puanı, bilim sınavı puanı, yabancı dil puanı ve bir önceki eğitim programından elde ettiği mezuniyet puanlarının kullanıldığı söylenebilir.

Ülkemizde genel olarak, ALESten yüksek lisans için en az 60, doktora için ise en az 70 puan beklemektedir. Yabancı dil puanı ise yüksek lisans için en az 50, doktora programı için en az 60 olarak belirlenmiştir. Yüksek lisans için lisans ortalamasının en az 2.70-3.00 arası olması istenirken doktora programı için yüksek lisans mezuniyet ortalamasının beklenen değerleri 3.00 – 3.50 arasında değişmektedir. Bazı üniversitelerde mülakat yöntemi de kullanılırken bunlara ek olarak referans mektubu istenmektedir. Yurt dışındaki üniversitelere bakıldığında ise yüksek lisans için aranan şartlar genel olarak 2.5 üstü not ortalaması, yabancı dil için TOEFL 90 ve üzeri veya IELTS Akademik modülden alınan her puan türünün 6'nın üzerinde olması, en az 2 akademik referans, profesyonel ve akademik kariyeri gösteren cv ve motivasyon mektubu (personal statement)

beklenmektedir. Doktora programları için ise İngilizce yeterlilik belgesi; IELTS'den en az 6.5, TOEFL'dan en az 90 puan şartı aranmaktadır. Akademik referans ve yayınlanmış eser ya da makalenizin olması gerekmektedir. Yapılacak araştırmanın konusu (Research Proposal) önceden belirlenmiş olmalı ve bir supervisorle görüşülmüş olması gerekmektedir. Bunlara ek olarak bazı üniversitelerin kendi açtıkları alanda yeterlilik sınavlarında bulunmaktadır. Tüm bu puanların öğrencilerin sayısal ve sözel becerilerini, alan ile ilgili bilgi beceri düzeylerini belirli ölçüde temsil ettiği söylenebilir. Ancak programa kabul durumunda öğrencilerin başarı düzeyleri üzerinde başka özellik ve değişkenlerin etkili olup olmadığının sınanması gerekmektedir (Özdemir vd., 2018). Bu durum bahsi geçen kriter puanların öğrencilerin kabul edildiği programdan mezun olma durumunu yordayıp yordamadığının tespit edilmesi ile mümkündür. Lisansüstü öğrenci kabulünde kullanılan ölçme sonuçlarının yordama geçerliğinin tespit edilmesi için birçok farklı yöntem kullanılabilir. Bu çalışmada bu sınama veri madenciliği algoritmaları kullanılarak yapılmıştır.

Veri madenciliği, verilerin içindeki gizli kalıpları ve ilişkileri keşfetmek için kullanılan bir dizi araç ve yöntem olarak tanımlanabilir (Artsın, 2019). Veri madenciliği büyük veri setlerindeki bilinmeyen ve ilgi çeken oluşumların keşfedilmesini sağlarken bununla birlikte büyük veri setleri içerisinde gelecek durumları ilgilendiren tahminlerin yapılması, denklem kurma ve kurulan bağıntıların bilgisayar programları kullanarak ortaya çıkarılması imkanını da sağlar (Hand, 2007). Veri madenciliği: yeni bir disiplin olmasına rağmen uygulandığı alanların geniş çaplı olduğu görülmektedir. Veri madenciliğinde; ekonomi, finans, sağlık çalışmaları, güvenlik çalışmaları, sosyal hizmetler, e-devlet, telekomünikasyon ve nakliye gibi bir çok alanda gerçekleştirilmiş uygulamaları bulunmaktadır (Aydın, 2007).

Birçok alanda eldeki verilerden yola çıkarak yeni tahminlerde bulunmak için kullanılan veri madenciliği uygulamaları günümüzde gelişen teknolojinin sayesinde çevrimiçi sınavların da artmasıyla büyük eğitimsel veri depolarının oluşmasını sağlarken eğitim alanında da çokca kullanılmaya başlanmıştır (Öztürk, 2018).

Eğitim alanında yapılan veri madenciliği alanyazında Eğitsel Veri Madenciliği olarak adlandırılmaktadır (Berland vd., 2014). Eğitsel veri madenciliği, eğitsel bilgileri, sınav sonuçları ve uzaktan öğretim sistemlerinde

kullanılan bilgileri veri olarak kullanır. Eğitsel veri madenciliğinin amacı bu bilgileri kullanarak belirlenecek çıktıların formülleştirilmesidir (Aydemir, 2017). Eğitimde veri madenciliği öğrenenleri ve öğrenmeyi anlamlandırmak için veri havuzlarını kullanmayı ve eğitimde fayda sağlayacak şekilde pratiğe dönüştürmek için verileri ve teorileri birleştiren hesaplamalar geliştirmeye çalışır. Bununla birlikte eğitimde veri madenciliği bize elde edilen veriler ile yeni bilgileri keşfetme, eğitim kalitesini bir çok yönden iyileştirme, eğitim sistemlerini değerlendirme ve daha etkili bir öğrenme alanları için zemin hazırlama imkanı sağlamaktadır (Romero vd., 2004).

Uluslararası eğitsel veri madenciliği derneği eğitimde veri madenciliğini eğitim alanlarından elde edilen günden güne artan veri setlerini araştırmak için yöntemler geliştirmekle ilgi olan, bu yöntemleri öğrencileri ve öğrenme yöntemlerini daha verimli anlamlandırmak için kullanılan bir disiplin olarak tanımlamaktadır (Romero vd., 2010a). Eğitimde veri madenciliği; eğitim sistemini verimli hale getirmek, eğitim kalitesini artırmak ve öğrenme sürecini daha etkili hale getirmek ve eğitim çıktılarını yordamak için kullanılabilir (Romero vd., 2010b).

Veri havuzlarından elde edilen büyük verileri anlamlı hale getirmek için bilgisayar yazılımlarına ihtiyaç duyulur. Veri madenciliği için sıklıkla kullanılan Açık kaynak kodlu uygulamalar; WEKA, RapidMiner, Orange, KNIME ve R, ticari uygulamalar arasında ise SPSS Clementine, SPSS, SAS, Angoss, KXEN, SQL Server ve MATLAB sıralanabilir (Danacı vd., 2010). Veri madenciliği algoritmalarının en yaygın işlevleri: sınıflandırma ve regresyon, kümeleme ile birliktelik kurallarıdır. Sınıflama ve regresyon modelleri tahmin edici model, kümeleme ile birliktelik kuralları ise tanımlayıcı modellerdir (Özbay, 2015). Sınıflandırma algoritmaları: verileri analiz ederek her örneği önceden tanımlanmış kategorilere atar. Eldeki verilerin belirli bir kısmı: (genellikle % 66) öğrenme verisi olarak kullanılır. Algoritmalar: öğrenme verilerini işleyerek sınıflandırma için modeller oluşturur. Verinin geri kalan kısmına test verisi denir. Oluşturulan modeller test verileri üzerinde sınanır ve modellerin başarısına ilişkin bilgiler elde edilir (Turgut, 2019).

Sınıflandırma algoritmalarının işlevi: veri setinin tahmin sınıfı ile olan ilişkisini açıklayan modeller oluşturmaktır. Bu algoritmalarının başarısı kullanılan değişkenlerin tahmin edilecek sınıf ile ne ölçüde ilişkili olduğuna bağlıdır

(Silahtaroglu, 2008). Tahmin sınıfı nominal olduğunda sıklıkla kullanılan algoritmalarından bazıları karar ağaçları, K en yakın komşu (KNN) ve lojistik regresyondur. Bu çalışmada elde edilen sonuçlara daha sağlıklı dayanak olması adına bu üç algoritmanın tümü kullanılmıştır. Algoritmaların model geliştirme becerilerinin karşılaştırılması bu çalışmanın amaçları arasında yer almamaktadır.

Çalışmada lisansüstü eğitime öğrenci alımında kullanılan ALES, bilim sınavı, yabancı dil ve önceki mezuniyet puanları tahmin için kullanılan değişkenleri oluştururken, öğrencilerin programdan mezun olup olmama durumu ise tahmin edilecek sınıfı oluşturmaktadır. Kullanılacak 3 algorithmadan bu değişkenler ile mezuniyet durumunu açıklayan anlamlı bir model oluşturması beklenmektedir. Bu analizler öncesi algoritmalara bahsi geçen puanların hangi ağırlıklarda kullanılması gerektiğine dair bir komut verilmemektedir. Bu sebeple puan türlerinin hangi ağırlıklarda kullanılacağına ya da kullanılıp kullanılmayacağına algoritmalar karar vermektedir.

Analiz çıktılarında algoritmaların anlamlı modeller kurabilmesi 4 kriter puan türünün öğrencinin program başarısını yordadığı şeklinde yorumlanacaktır. Sınıflandırma algoritmalarının anlamlı modeller oluşturması durumunda, bu modeller sayesinde programlara başvuran öğrencilerin kayıt öncesi kabul durumundaki olası başarıları üzerinde öngörüler oluşturulabilecektir. Böylelikle lisansüstü eğitime kabulünde kullanılan kriter puanların kullanımı daha yordayıcı ve anlamlı hale getirilmesine katkı sağlayabilir.

1.1 Problem Cümlesi

Lisansüstü eğitime öğrenci kabulünde kullanılan kriter puanlar, programa kabul görmüş öğrencilerin mezun olup olmama durumunu yordama bakımından bir veri madenciliği uygulaması olan K-NN, C4.5 ve lojistik regresyon Algoritmaları ile yapılan analizlere göre işlevsel midir?

1.2 Alt Problemler

- 1) Lisansüstü eğitime öğrenci kabulünde kullanılan kriter puanlar, programa kabul görmüş öğrencilerin mezun olup olmama durumunu yordama bakımından K-NN algoritmasına göre işlevsel midir?
- 2) Lisansüstü eğitime öğrenci kabulünde kullanılan kriter puanlar, programa kabul görmüş öğrencilerin mezun olup olmama durumunu yordama bakımından C4.5 algoritmasına göre işlevsel midir?

- 3) Lisansüstü eğitime öğrenci kabulünde kullanılan kriter puanlar, programa kabul görmüş öğrencilerin mezun olup olmama durumunu yordama bakımından lojistik regresyon algoritmasına göre işlevsel midir?

1.3 Araştırmanın Önemi

Bu araştırmanın temel amacı, lisansüstü eğitime öğrenci kabulünde kullanılan kriter puanların işlevselliğinin bir başka deyişle öğrencinin kabul edildiği programdan mezun olma durumunu yordayıp yordamadığının eğitsel veri madenciliği uygulamaları yardımı ile tespit edilmesidir.

Lisansüstü eğitim: dünyadaki toplumların ihtiyacı olan araştırmacıları, bilim insanlarını nitelikli insan gücünü yetiştirmeyi amaçlamasının yanında geleceğin bilim insanlarını ve araştırmacılarını yetiştiriyor olması lisanüstü eğitimin önemini ortaya koymaktadır. Lisanüstü üstü eğitimden beklenen verimi almanın yolu: eğitim ortamına dahil edilecek bireylerin alanlarında yeterli ve alanlarına katkı sağlayacak bireyler olması gerektiği söylenebilir. Bu durum bize lisansüstü eğitime öğrenci seçiminin tüm bu durumları sağlayabilmek için ne kadar önemli olduğunu göstermektedir. Bu araştırma lisansüstü eğitime öğrenci seçiminde kullanılan kabul kriterlerinin öğrencinin başarısını yordayıp yordamadığını bize veri madenciliği algortimaları ile destekleyerek gösterdiği için önemlidir. Eğitsel veri madenciliği uygulamalarının yaygınlaşması ile bu çalışmalardan elde edilen sonuçlar: ülkemiz de uygulanan birçok programa giriş sınavlarına alternatif yeni yöntemler geliştirilmesinin gerektiğini bize göstermektedir. Eğitimde veri madenciliği uygulamalarının çok fazla olmadığı, daha önce lisansüstü eğitime kabul kriterlerinin incelendiği çalışmalar olsa da veri madenciliği ve weka uygulamaları içeren çalışmaların sınırlı sayıda olduğu görülmektedir. Bu nedenle eğitsel veri madenciliği ve eğitim bilimleri alan yazınına katkıda bulunacağı düşünülmektedir.

1.4 Sayıtlar

Araştırmada yer alan çalışma grubunun kabul edildikleri programları verilen süre zarfında tamamlamayanların mezun olamadığı varsayılmıştır.

1.5 Sınırlılıklar

Araştırmada kullanılan veri seti Abant İzzet Baysal Üniversitesi Lisanüstü Eğitim Fakültesinde 2011-2016 yılları arasında eğitim gören 740 öğrencilerden elde edilen cinsiyet, ALES, bilim sınavı, mezuniyet not ortalaması, yabancı dil sınavı puanları ve mezuniyet durumları bilgileri ile sınırlıdır. Bununla birlikte

arařtırmada kullanılan tahmin ve sınıflandırma modelleri WEKA programında ve lojistik regresyon, k-en yakın komřu ve C4.5 algoritmaları ile sınırlıdır.

1.6 Tanımlar

Lisansüstü eğitim: 2011-2016 yılları arasında Bolu Abant İzzet Baysal Üniversitesi Eğitim Bilimleri Enstitüsü'nde lisans eğitimine baęlı olan yüksek lisans ve doktora eğitimiyle birlikte bilimsel araştırma ve uygulamalardan oluşan eğitim kademesidir.

Lisansüstü eğitime kabul kriterleri: Bolu Abant İzzet Baysal Üniversitesi Eğitim Bilimleri Enstitüsü bünyesinde bulunan lisansüstü eğitim programlarına başvuran adayların seçiminde ölçüt olarak kullanılan ALES puanı, Lisans/ Yüksek lisans Mezuniyet puanı, Dil sınavı puanı ve Bilim sınavı puanlarıdır.

Lisansüstü eğitime kabulde kullanılan kriterlerin işlevsellięi: Lisansüstü eğitime öğrenci kabulünde kullanılan kriterlerin, öğrencilerin kabul edildikleri programdan mezun olup olmadıklarını yordama durumudur.

Mezun: Lisansüstü eğitimi başarıyla tamamlama durumu

Mezun deęil: Lisansüstü eğitimi başarıyla tamamlayamama durumu

2. ARAŞTIRMANIN KURAMSAL TEMELİ VE İLGİLİ ARAŞTIRMALAR

Bu bölümde: araştırmanın kuramsal temelleri ve ilgili araştırmalardan bahsedilmiştir.

2.1 Araştırmanın Kuramsal Temelleri

2.1.1 Lisansüstü Eğitim

Türkiye’de lisansüstü eğitim, lisans eğitime bağlı olan yüksek lisans ve doktora eğitimi, sanat alanında yapılan sanatta yeterlik çalışması ve tıp alanında yapılan tıpta uzmanlık eğitimiyle birlikte bilimsel araştırma ve uygulamalardan oluşan eğitim kademesi olarak tanımlanmıştır (Lisansüstü Öğretim Yönetmeliği, Mad. 2; Yükseköğretim Kanunu, Mad.3). Başka bir tanımla lisansüstü eğitim: bireylere başarıyla tamamlamış oldukları lisans programlarına dayalı, aynı zamanda ilgi duydukları bilimsel alanlarda yüksek lisans ya da doktora eğitimi olarak uzmanlaşma olanağı sunan eğitim kademesi olarak tanımlanabilirken diğer yandan araştırma yapabilen, üretken, yaratıcı bireyleri yetiştiren eğitim programı şeklinde de tanımlanabilir (Karaman ve Bakırcı, 2010).

Lisansüstü eğitim: dünyadaki toplumların ihtiyacı olan araştırmacıları, bilim insanlarını ve yüksek nitelikli insan gücünü yetiştirmeyi amaçlamaktadır. Buradan yola çıkarak lisansüstü eğitimin geleceğin bilim insanlarını ve araştırmacılarını yetiştiriyor olması ülkelerin: hedefledikleri amaçlara, teknolojik, ekonomik ve kültürel gelişimlerine katkı sağladığı söylenebilir (Bülbül, 2003).

Türkiye’de: 2021-2022 öğretim yılı itibarıyla toplam yüksek lisans öğrenci sayısı 358271, doktora eğitimi gören öğrenci sayısı ise 109540 olarak belirlenmiştir (<https://istatistik.yok.gov.tr>, 2023). 2012 yılında 25 bin 704 olan yüksek lisans mezun sayısı yıldan yıla arttığı, 2021 yılına gelindiğinde 70 bin 396’ya ulaştığı görülmektedir. Doktora mezun sayısı ise 2021 yılında 8 bin 857’dir. Türkiye’de 2012 yılından bu yana artarak gelen yükseköğretim mezunu oranı 2020 yılında %35 iken 2021 yılında artarak % 40’a yükseltmiştir (Türkoğlu vd. , 2022).

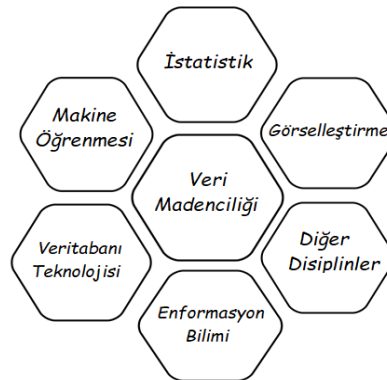
2.1.2 Veri Madenciliği

Her alanda günden güne giderek artan veri yığınları ve veri çeşitliliği: gelişen teknoloji sayesinde daha kolay depolanabildiği gibi bu verilere ulaşım imkânı da oldukça kolay hale gelmiştir. Teknolojinin gelişmesi bunlarla birlikte

bu verilerin anlamlandırılmasını ve işlenmesini de ucuz ve basit hale getirmiştir (Turgut ve Baykul, 2012). Günümüzde üretilen bu büyük verilerin anlamlı ve yararlı bilgi olarak kullanılmasında zorluklar yaşanmaktadır. Geleneksel istatistik yöntemleri büyük boyuttaki veri yığınlarını anlamlandırmak ve işlemekte yetersiz kalmaktadır. Bu sebeple verileri anlamlandırmak ve işlemek için daha ayrıntılı analiz yapan yöntemlere ihtiyaç hissedilmiştir. İstatistik, makine öğrenmesi, veritabanı teknolojileri ve diğer teknikleri bir araya getiren veri madenciliği ve yöntemleri de bu gereksinim sayesinde ortaya çıkmıştır (Tüzüntürk, 2010).

Veri madenciliğini; Tüzüntürk (2010), verilerin içindeki gizli kalıpları ve ilişkileri keşfetmek için kullanılan bir dizi araç ve yöntem olarak tanımlarken, Aydın (2007), veri madenciliğini hedeflenen sonuçlara ulaşmak için, elde edilmiş verilere algoritmaların uygulandığı bir yeni bilgileri keşfetme süreci olarak tanımlamaktadır. Veri madenciliğinin başka bir tanımı ise büyük veri tabanlarındaki veri yığınlarından faydalanarak, gizli bağlantıların ve genel örüntülerin ortaya çıkarılmasıdır (Holsheimer ve Siemes, 1994). Veri madenciliğini Witten vd. (2005): “Büyük veri setlerinde bulunan verilerin otomatik veya yarı otomatik şekilde aralarındaki bağıntının keşfedilme serüveni” olarak tanımlamıştır.

Veri madenciliğinde amaç, anlamlandırılması ve işlenmesi zor ve karmaşık olan büyük veri kümelerini analiz ederek işe yarar ve anlamlı bilgiler haline dönüştürmek; bu bilgileri içinde içeren bir denklem ortaya çıkararak yeni oluşmuş veri kümeleri hakkında yorum yapmayı ve tahminde bulunmayı sağlamaktır (Coşkun ve Baykal, 2011).



Şekil 2.1. Veri Madenciliği Disiplinleri (Han ve Kamber, 2006)

2.1.3 Eğitsel Veri Madenciliği

Veri madenciliği birçok alanda sağladığı faydalı bilgilerin yanında eğitime de verimli katkılarda bulunmaktadır. Veri madenciliği yoluyla kurulan bağlantılar, daha sonra veri madenciliği uygulamalarıyla bireysel davranışlar ve bilişsel özellikleri kestirmede kullanılabilmektedir(Akpınar, 2000).

Birçok araştırmacı tarafından eğitim sürecinde kullanılması önemi vurgulanan ve günden güne uygulamaları artan eğitsel veri madenciliği; uluslararası Eğitsel veri madenciliği topluluğu tarafından, eğitim alanlarında elde edilen birbirinden farklı veri türlerini araştırmak, bu veri madenciliği yöntemlerini eğitimin aktif elemanı olan öğrencileri ve öğrenme alanlarını daha iyi anlamak ve geliştirmek olarak tanımlamaktadır (Romero ve Ventura, 2010a). Eğitsel veri madenciliği: “Eğitim alanlarının veri tabanlarından elde edilen verilere eğitsel hedefler doğrultusunda veri madenciliği yöntemlerinin uygulanmasıyla ortaya çıkan bir araştırma alanıdır.” şeklinde de tanımlamıştır (Narlı vd., 2014).

Eğitsel veri madenciliği, eğitim alanlarından elde edilen farklı veri türlerini keşfetmek için yeni yöntemler geliştirmek ve bu yöntemleri; eğitimin elemanlarını geliştirmek ve eğitim elemanlarına sağlanacak öğrenme alanlarını daha iyi anlamlandırmak amacıyla kullanan, günden güne gelişmekte olan disiplindir (Baker ve Yacef, 2009). Başka bir tanımla; Eğitsel veri madenciliği öğrencinin mevcut bilgisini, motivasyonunu, tutumlarını ve bilişsel gelişimini gösteren modellerin geliştirilmesi şeklinde tanımlanmaktadır (Mohamad ve Tasir, 2013).

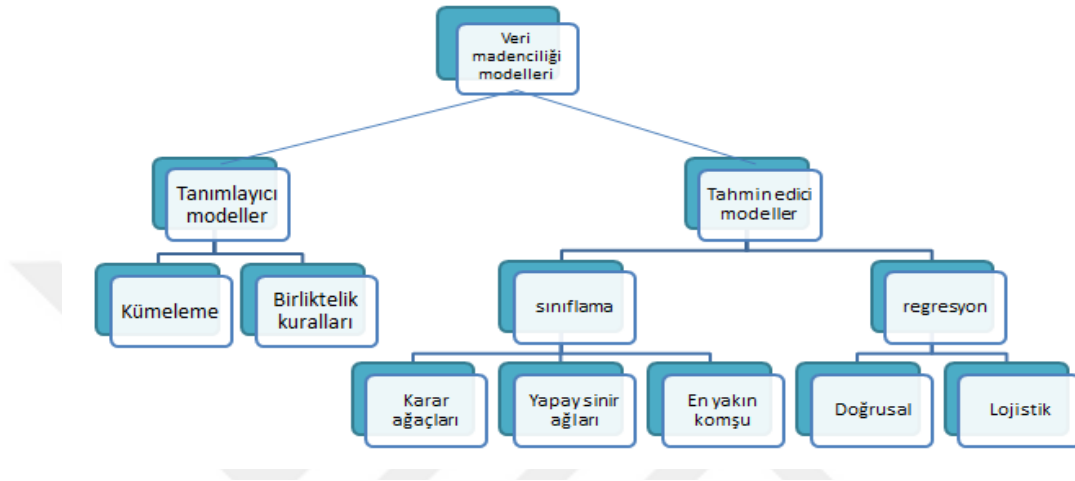
2.1.4 Veri Madenciliği Modelleri

Veri madenciliği modelleri tahmin edici (predictive) ve tanımlayıcı (descriptive) olmak üzere ikiye ayrılır:

- Tahmin edici modellerde, yapılan öncelikli çalışma sonuçları daha önceden bilinen verilerden yola çıkarak bir denklem kurulmasıdır. Daha sonra kurulan bu denklemden faydalanarak sonuçları bilinmeyen veri kümeleri için sonuçların tahmin edilmesi amacı vardır (Özekes, 2003). Tahmin edici modeller özellikle karar verme aşamasında etkili bir yöntemdir.
- Tanımlayıcı modellerse karar verme aşamasına öncülük etmede kullanılabilecek eldeki verilerin örüntülerinin tanımlanması sağlanmaktadır (Silahtaroglu, 2008).

Veri madenciliği yöntemleri işlevlerine göre üç ana başlığa ayrılır:

1. Kümeleme (Clustering),
2. Birliktelik Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns)
3. Sınıflama (Classification) ve Regresyon (Regression),



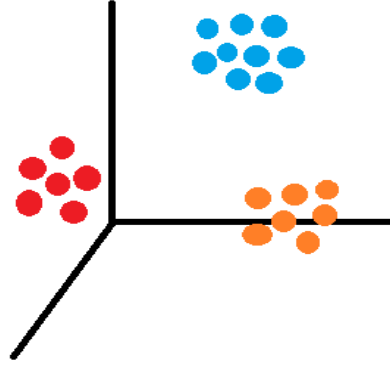
Şekil 2.2. Veri Madenciliği Modelleri

Şekil 2. 2. de görüldüğü gibi sınıflama ve regresyon tahmin edici model iken kümeleme ve birliktelik kuralları ise tanımlayıcı modellerdir.

2.1.4.1 Kümeleme

Kümeleme: verileri sınıflandırma ve kümelere ayırma işlemidir. Kümeleme modelinde amaç değişkenin değerini belirlemeye yönelik tahmin etme veya kestirim yapmak değil, verinin tamamını gruplara ayırmak için birbirine benzer özellikte alt gruplar araştırmaktır. Bu alt gruplara ayırma uygulaması yapılırken kümeler içindeki verilerin ortak noktaları dikkate alınır (Şengür ve Tekin, 2013).

Birbirine benzer ya da ilişkili olan veri grupları aynı kümede birleşirken farklı veri ya da birbiriyle ilişkili olmayan veriler farklı bir küme de birleşir (Yılmaz ve Patır, 2011).



Şekil 2.3. Örnek Küme Yapısı

Kümeleme uygulamalarına örnek olarak: marketlerde müşterilerin tercihlerinin keşfedilmesi, alışveriş ilişkilerinin ortaya çıkarılması, biyolojik alanlarda bitki ve hayvan türlerinin sınıflandırmaları ve benzer genlerin işlevlerine göre sınıflandırılması gibi uygulamalar verilebilir (Kaya ve Köymen, 2008).

2.1.4.2 Birliktelik Kuralları Ve Ardışık Zamanlı Örüntüler

Birliktelik kuralları: büyük veri kümeleri arasındaki birliktelik ilişkilerini ortaya çıkarırlar (Hark, 2013). Veri yığınlarının her geçen gün katlanarak artmasıyla: çoğu şirket veri tabanlarındaki verilerin birliktelik kurallarını ortaya çıkarmak istemektedir. Büyük veri yığınlarındaki birliktelik ilişkilerinin keşfedilmesi, şirketlere daha verimli karar alma süreçleri sağlamaktadır (Taş, 2018).

En yaygın ve genel kullanım alanı market sepet analizi olan Birliktelik kuralları analizine “Düşük yağlı peynir ve yağsız süt alan müşteriler %85 olasılıkla diyet süt alırlar.” Ve “Bir süper markette yumurta ve peynir satın alınan satış hareketlerinin %75’inde zeytin de satın alınmıştır.” cümleleri örnek olarak verilebilir (Akpınar, 2000).

Ardışık zamanlı örüntüler: ardışık zamanlı gerçekleşen ve birbiriyle bağlantısı olan ilişkileri tanımlamak için kullanılır (Hark, 2013).

2.1.4.3 Sınıflama Ve Regresyon

Sınıflama: veri yığınlarını tanımlayan ve ayırtıran bir model bulma sürecidir. Model hangi sınıfa ait olduğu bilinen verileri kullanarak sınıf etiketi bilinmeyen verilerin ait olduğu sınıfı tahmin etmek için kullanılan bir dizi analize dayalı olarak türetilmiştir. Regresyon ise bağımlı değişken ile bağımsız

değişkenler arasındaki en iyi ilişkiyi gösteren fonksiyonu elde etmek için uygulanan istatistiksel yöntemlerdir (Coşkun ve Baykal, 2011). Sınıflama kategorik değerlerin tahmin etmek için kullanılırken regresyon sürekli değerlerin tahmin edilmesi için kullanılır (Hand, 2007). Bu yöntem veri tabanındaki, gizli bağlantıları veya örüntüleri ortaya çıkarma amacıyla kullanılmaktadır (Özkan, 2014). Örneğin, bankalarda kredi uygulamalarının güvenli veya güvenli değil olarak sınıflamak amacıyla kurulan model sınıflama, mesleği ve aylık ortalama geliri bilinen müşterilerin e-ticaret ürünleri için yapacakları ortalama harcamaları tahmin etmek için ise regresyon modeli kurulabilir (Özekes, 2003).

Sınıflama ve regresyon modellerinde kullanılan algoritmalar:

1. Karar Ağaçları (Decision Trees)
2. Yapay Sinir Ağları (Artificial Neural Networks)
3. K-En Yakın Komşu (K-Nearest Neighbor)
4. Genetik Algoritmalar (Genetic Algorithms)
5. Bellek Temelli Nedenleme (Memory Based Reasoning)
6. Naive-Bayes
7. Lojistik regresyon (Logistic regression)

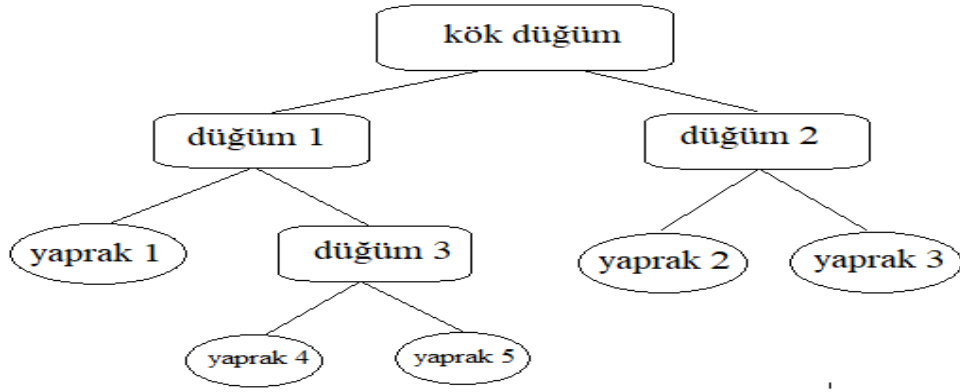
2.1.5 Sınıflama Algoritmaları Ve Model Başarım Ölçütleri

2.1.5.1 Sınıflama Algoritmaları

2.1.5.1.1 Karar Ağaçları (Decision Tree Learning)

Karar ağaçları, sınıflandırma algoritmaları arasında kolay yorumlanabilen ve anlaşılabilen algoritmalarından biri olması, veri tabanları ile entegrasyonun kolaylığı, güvenilirliklerinin yüksek olması sebebiyle sınıflama modelleri içerisinde en yaygın kullanıma sahip teknik olmuştur (Brodley ve Friedl, 1997). Karar ağaçları doğrusal olmayan ilişkileri ele alabilir, hem kategorik hem sayısal verilerle çalışmaya izin verdiği gibi kayıp verilere de izin verir (Altaş ve Gürpınar, 2012).

Silahtaroglu'na (2008) göre, karar ağacı algoritmaları ilk adımda karar ağacı oluşturulması, ikinci adımda ise sınıflandırma olmak üzere iki aşamadan oluşur. Öğrenme basamağında veri setinin önceden belirlenen yüzdesi: eğitim verisi olarak adlandırılır ve sınıflama algoritması tarafından model oluşturmak amacıyla analiz edilir. Bu analiz sonucunda ortaya çıkan değerler karar ağaçlarının yapısını oluştur (Çalış vd., 2014). Karar ağaçlarının yapısı düğüm, dal ve yapraklardan oluşmaktadır (Albayrak ve Yılmaz-Koltan, 2009).



Şekil 2.4. Örnek karar ağacı yapısı

Sınıflama basamağında ise eğitim verisi, karar ağacının doğruluğunu ve sınıflama kurallarını ortaya çıkarmak amacıyla kullanılır. Eğer doğruluk kabul edilebilir seviyede ise kurallar, elde edilen yeni verilerin sınıflanması için kullanılır. Eğitim verisindeki kullanılacak alanlar ve bu alanların hangi sırada kullanılarak ağacın oluşturulacağı belirlenmelidir (Alan, 2012).

ID3, C4.5, C5.0, CART ve CHAID karar ağaçlarında en çok tercih edilen algoritmalarındandır.

2.1.5.1.1.1 C4.5 Algoritması

J.R. Quinlan, tarafından 1986 yılında Karar ağaçlarını oluşturmak için geliştirilen ID3 algortimasında kökten alt dallara doğru (top-down) ve sonuca en yakın durum (greedy search) tekniklerini kullanılır (Sayad, 2022). ID3 algoritmasının kategorik değişkenlerle sınırlı olma durumunu gidermek hedefiyle C4.5 algoritması 1993 yılında Quinlan tarafından önerilmiştir. C4.5 algoritmasının en önemli gelişimleri kategorik ve sürekli değişkenlerden ikisiyle de çalışılabilmesi ve veri seti içinde ki kayıp verileri ayıklayarak hesaplamaya katmadan analizi yapması olduğu görülmektedir (Gümüşçü vd. , 2016).

Algoritmada kullanılan ölçümlerden en çok tercih edileni Entropi ölçümüdür. Entropi; bir değişkendeki belirsizliğin ölçülmesi için kullanılmaktadır. Daha basit bir tanımla entropi bir veri grubundaki düzensizliktir denebilir. Entropi değeri 0 ile 1 arasında değerler alır, değişkenin tüm değerleri eşit ise entropi (belirsizlik) yoktur (Aydemir, 2017). Değerin 1'e yakın olması belirsizliğin yüksek olduğunu gösterirken, değer 0 olması veri grubunun düzenli olduğunu gösterir. Bu nedenle, karar ağacı algoritması uygulanırken kökünde Entropi

ölçüsü en az olan kullanılır (Aydemir, 2017). Örneğin; eğitim düzeyinin tahmin edildiği bir veri grubunda herkesin üniversite mezunu olması entropinin 0 olduğu anlamına gelir. Veri setindeki her bir bireyin farklı eğitim düzeyine ait olması entropi değerinin 1 olduğu anlamına gelir diyebiliriz.

Entropi değerinin formülü ;

$$H(S) = \sum_{j=1}^k \left(p_j \log_b (1/p_j) \right) \quad (2.1)$$

n : Herhangi Y değişkeninin düzey sayısı

P_j : j. Düzeyin ortaya çıkma olasılığı

b : Logaritma tabanını ifade eder.

Entropi değeri hesaplandıktan sonra, ID3 Algoritmasında bağımsız değişkenlerin her biri için “gain” değeri hesaplanır. Bu değer bağımsız değişkenin entropi değerinden bağımlı değişkenin entropi değerinin çıkarılması ile elde edilir.

$$Gain(X_i) = H(S) - H(S, X_i) \quad (2.2)$$

Bağımsız değişkenlerin her biri için gain değeri belirlenir, belirlenen gain değerlerinden en yüksek olan değişken üzerinden dallandırma yapılır (Altunkaynak, 2017).

C4.5 algoritması, ID3 algoritmasında kullanılan gain değerinin yerine gain oranını kullanır. Gain oranı, gain değerinin entropi değerine bölünmesi ile elde edilir. Gain oranı, değerlerin kategori sayısına göre düzenlenmesini için kullanılmaktadır.

$$GainRatio(X_i) = \frac{Gain(X_i)}{H(X_i)} \quad (2.3)$$

C4.5 Algoritmasının lisansı telif ücretine tabi olduğu için bu algoritma yerine J48 algoritması alternatif olarak kullanılır.

2.1.5.1.2 K- En Yakın Komşu (KNN, K- Nearest Neighborhood)

K-NN Algoritması: öğrenme verisi olarak kullanılan veri kümesin ve en yakın komşu elemanının, komşuluk değeri olarak ifade edilen k değerine göre belirlendiği bir makine öğrenme yöntemi olarak ifade edilmektedir. K-NN Algoritmasının bilinen algoritmalar arasında kullanılan en eski ve en iyi sınıflandırma algoritmalarından biri olarak kabul edilir. Örnek tabanlı öğrenme algoritmaları arasındaki en temel algoritma olduğu söylenebilir (Şengür ve Tekin, 2013). Örnek tabanlı öğrenme algoritmalarında: yeni bir model kurmak için bağıntıyı keşfetme işlemi örnek verilerden yararlanarak gerçekleştirilir, yeni karşılaşılan bir değer ise örnek alınan veri setinde yer alanlar ile arasındaki benzerliğe göre sınıflandırılmaktadır.

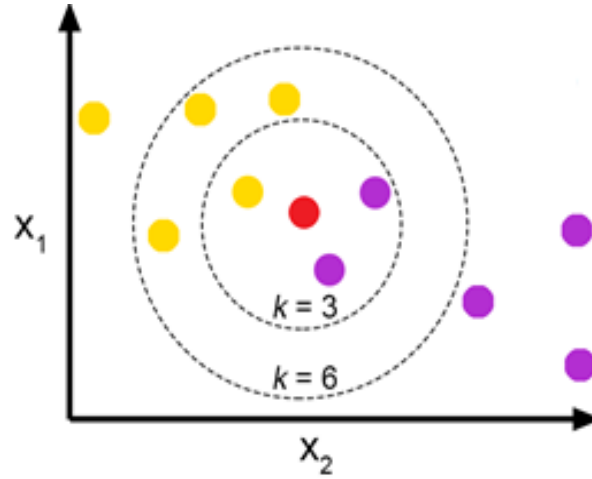
K-NN algoritması için ilk yapılması gereken k değerini belirlemektir, k değeri algoritmaya öğrenme verisi olarak verilen bir değere en yakın komşuların sayısıdır (Kutlugün vd., 2018). K değeri belirlendikten sonra yapılacak şey: sınıfı belirlenmek istenen verinin komşu noktalara uzaklıklarının hesaplanmasıdır. Bulunan uzaklık değerleri arasından değeri en az olan k tanesi alınır. Alınan bu k tane değerin hangi komşu sınıfına ait oldukları belirlenir ve en çok tekrarlanan sınıf değeri seçilir.

Verilerin komşu noktalara uzaklık değerlerini hesaplamak için kullanılan ölçülerden biri: Öklid ölçüsüdür. *Öklid ölçüsü*: iki nokta arası verilen mesafenin hesaplaması olarak ifade edilmektedir.

N boyutlu uzayda Euclidean (Öklidyen) Uzaklık hesaplaması aşağıdaki gibidir:

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2.4)$$

Aşağıda ki şekilde KNN algoritmasının çalışma prensibi ile ilgili bir örnek verilmiştir:



Şekil 2.5. Knn algoritması çalışma prensibi örneği

Şekilde her nokta bir bireyi temsil etmek üzere her bireye ait x_1 ve x_2 değişkenlerindeki değerlerine göre iki boyutlu düzlem kullanılmıştır. Sarı noktalar A sınıfındaki bireyleri, mor noktalar B sınıfındaki bireyleri temsil etmektedir. Sarı ve mor bireyler sınıfları daha önceden belirlenmiş bireylerken, kırmızı renkli bireyin hangi sınıfa ait olduğu K-NN algoritması ile belirlenmeye çalışılmaktadır. Kullanıcı tarafından $k=3$ değeri seçildiğinde kırmızı nokta ile temsil edilen bireye en yakın konumlanmış 3 birey incelemeye alınır. Bu üç bireyden 2'si mor renkli bir başka deyişle B sınıfına ait olduğu için, kırmızı nokta ile gösterilen bireyin B sınıfına ait olduğu tahmini yapılır. Bu örnek için kullanıcı $k=6$ değerini belirlerse algoritma sınıflandırılacak bireye en yakın 6 bireyin sınıflarını inceleyecek ve dört birey sarı renkli bir başka deyişle A sınıfına ait olduğu için, bireyin sınıfını A olarak atayacaktır.

2.1.5.1.3 Lojistik Regresyon (Lojistik Regression)

Lojistik regresyon, kategorik olarak ikili, üçlü ve çoklu kategorilerde bağımlı değişkenler bulunduğunda, bağımlı değişkenler ile bağımsız değişkenler arasındaki sebep-sonuç ilişkisini ortaya çıkarmak için kullanılan bir yöntemdir (Özdamar, 2004). Lojistik regresyon ile bağımlı değişken üzerinde etkisi bulunan bağımsız değişkenler ve bu değişkenlerin bağımlı değişkenler üzerindeki etkileri olasılık olarak hesaplanır (Şekerçioğlu, 2010). Lojistik regresyonun amacı; bağımlı ve bağımsız değişkenler arasındaki ilişkiyi en açık şekilde ifade eden geçerli bir model oluşturmaktır (Bircan, 2004).

Lojistik regresyon modeli aşağıdaki şekilde açıklanır;

$$P = \frac{e^{\beta_0 + (\beta_1 x_1) + \dots + (\beta_n x_n)}}{1 + e^{\beta_0 + (\beta_1 x_1) + \dots + (\beta_n x_n)}} \quad (2.5)$$

P ; incelenen olayın gözlenme olasılığı;

β_0 ; sabit

$\beta_0, \beta_1, \dots, \beta_n$; bağımsız değişkenlerin regresyon katsayıları

$x_1, x_2, \dots, x_n$; bağımsız değişkenleri

N ; bağımsız değişkenlerinin sayısını ifade etmektedir (Özdamar, 2004).

2.1.5.2 Model Başarım Ölçütleri

Kullanılan algoritmalarından elde edilen sınıflama performanslarının kalitesini ölçmek için: başarı değerlendirme ölçütleri kullanılmaktadır. Tahmin sonuçlarının değerlendirme göstergeleri olarak: doğru sınıflandırılmış örnek oranları, bağıl mutlak hata, Kappa İstatistiği kullanılabilir.

2.1.5.2.1 Doğru Sınıflandırılmış Örnek Oranları

Sınıflandırma modellerinin başarı düzeyi: doğru sınıfa dahil edilen örnek sayısı ve yanlış sınıfa dahil edilen örnek sayısı ile belirlenir. Algoritmanın uygulanması sonucunda elde edilen bu çıktıların başarı bilgileri karışıklık matrisi ile gösterilir (Bilgin, 2017).

		GERÇEK	
		Doğru (+)	Yanlış (-)
TAHMİN	Doğru (+)	Doğru pozitif (DP)	Yanlış pozitif (YP)
	Yanlış (-)	Yanlış negatif (YN)	Doğru negatif (DN)

Şekil 2.6. Karışıklık matrisi örneği

Bu matrizen elde edilecek: doğru pozitif (DP), yanlış pozitif (YP), yanlış negative (YN) ve doğru negative (DN) değerleri çeşitli model başarı ölçütlerine dönüştürülebilir. Bu ölçütler doğruluk, hata oranı, duyarlık, özgüllük, kesinlik ve F-ölçütüdür (Coşkun ve Baykal, 2011).

- doğruluk ölçütü = $\frac{DP+DN}{DP+YP+YN+DN}$

Doğru sınıflandırılmış örnek değerlerin (DP +DN): örnek değerlerin toplamına (DP+DN+YP+YN) oranı doğruluk ölçütünü verir.

Doğruluk: algoritma tarafından doğru yapılan sınıflama sayısı ile algoritmanın kalitesi ile ilgili bilgi verir. Yanlış sınıflandırma yapılmış örnek değerlerin örnek değerlerin toplamına oranı: hata oranı ölçütüdür. Ayrıca hata oranı ölçütü: doğruluk ölçütünü 1'e tamamlayan değerdir (Coşkun ve Baykal, 2011).

- Duyarlık ölçütü = $\frac{DP}{DP+YN}$

- Özgüllük ölçütü = $\frac{DN}{YP+DN}$

- Kesinlik ölçütü = $\frac{DP}{DP+YP}$

- F ölçütü = $\frac{2 \times \text{duyarlık} \times \text{özümlülük}}{\text{duyarlık} + \text{kesinlik}}$

Doğru sınıflandırılmış örnek oranları: doğru sınıflandırılmış örnek sayısının tüm örnek sayısına bölünmesiyle hesaplanan doğru sınıflandırılmış örnek oranları model başarısı hakkında önemli göstergelerdir. Fakat bir zayıf noktası sınıf değişkeni iki veya üç kategoriye sahip olduğunda (bu çalışmada olduğu gibi), rastgele bir yordayıcı bile yaklaşık %30 - %50 doğru sınıflandırılmış örnek oranı gerçekleştirebilir olmasıdır. Bu zayıf noktadan dolayı: doğru sınıflandırılmış örnek oranları tek başına yetersizdir denebilir. Bu nedenle: farklı göstergelere ihtiyaç duyulabilir.

Sınıflama performans kalitesi ölçümünde kullanılan alternatif ölçütlerden biri kappa istatistiğidir.

2.1.5.2.2 Kappa İstatistiği (κ)

Cohen (1960) tarafından sınıflama algoritmasında puanlama yapan iki puanlayıcı arasındaki tutarlılığın seviyesini belirlemek için geliştirilen kappa istatistiği: Fleiss (1971) tarafından ikiden fazla puanlayıcı arasındaki tutarlılığın belirlenmesi için geliştirilmiştir. Kappa istatistiğinin hesaplanmasının basit olması

ve pratik yorumlanmasının yanında getirdiği avantajlarından en önemlisi: tesadüfe dayalı oluşan benzerliği düzeltmeyi temel almasıdır.

Kappa istatistiği formülü aşağıda verilmiştir:

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \quad (2.6)$$

$\bar{P}$: gözlenen tutarlılığın oranı,

$\bar{P}_e$: tesadüfi/şans tutarlılık oranını ifade eder.

Kappa istatistiğinin aldığı değerler: -1 ile +1 arasında yer almaktadır (Fleiss, 1971, akt. Bıkmaz-Bilgen ve Doğan, 2017).

Kappa istatistiğinin de;

- Puanlayıcılar arasındaki tesadüfi tutarlılığın beklenen tutarlılıktan fazla olması pozitif κ değerlerine sahip olduğunu
- Puanlayıcılar arasındaki tutarlılığın tesadüfi tutarlılıktan az olması negatif κ değerlerine sahip olduğunu gösterir.

Landis ve Koch (1977), tarafından önerilen kappa istatistiği tutarlılık düzeyleri aşağıda verilmiştir;

Tablo 2.1. Kappa istatistiği uyum gücü değerleri

K	<i>Tutarlılık Gücü</i>
<0,00	Zayıf
0,00 – 0,20	Önemsiz
0,21 – 0,40	Düşük
0,41 - 0,60	Orta
0,61 – 0,80	Önemli
0,81 - 1,00	çok yüksek

Modellerin sınıflama kalitelerinin belirlenmesinde kullanılan diğer ölçüm yöntemleri ise şunlardır:

- Ortalama Mutlak Hata/ Mean Absolute Error (MAE)
- Ortalama Karekök Hatası/ Root Mean Squared Error (RMSE),

- Göreceli Mutlak Hata / Relative Absolute Error (RAE)

2.1.5.2.3 Ortalama Mutlak Hata/ Mean Absolute Error (MAE)

Mutlak hata ortalaması (MAE), tahmin edilen değerlerin gerçek değerlerden uzaklık seviyesini bulmak için kullanılır P_i = Tahmini değer, R_i =Gerçek değer ve n =örneklem sayısı olmak üzere MAE formülü aşağıdaki gibidir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - A_i| \quad (2.7)$$

2.1.5.2.4 Ortalama Karekök Hatası/ Root Mean Squared Error (RMSE)

Ortalama Karekök Hatası, tahmin edilen değer ile gerçek değer arasındaki farkın standart sapmasını hesaplar. P_i = Tahmini değer, R_i =Gerçek değer ve n =örneklem sayısı olmak üzere RMSE formülü aşağıdaki gibidir(Gök, 2017).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - A_i)^2} \quad (2.8)$$

2.1.5.2.5 Bağlı Mutlak Hata / Relative Absolute Error (RAE)

Bağlı mutlak hata (RAE), doğru sınıflandırılma oranında şans başarısını ortadan kaldırmak için geliştirilmiştir (Ebren Kara ve Şamlı, 2021). RAE değerini hesaplamada kullanılan yöntem açıklamalarla paylaşılmıştır.

$$RAE = \frac{\text{Model hata oranı}}{\text{Basit tahminleyicinin hata oranı}} \quad (2.9)$$

Yukarıdaki eşitlikte model hata oranı yanlış sınıflandırılmaların oranını temsil eder. Basit tahminleyicinin hata oranı ise en ilkel sınıflandırma siteminde elde edilecek hata oranını gösterir. Örneğin öğrenme verisinde 60 erkek 40 kadın bulunan bir grupta cinsiyet sınıflandırması yapılacak olsun. Bu durumda

sınıflandırılacak tüm bireyler erkek olarak sınıflandırılırsa yaklaşık %60'lık bir başarı oranı ortaya çıkaracaktır. %60 doğru sınıflandırma %40 hata oranı anlamına gelir. Bu durumda örnek gruptaki basit tahminleyicinin hata oranı %40 olarak kabul edilir. Eğer aynı grupta cinsiyet sınıflandırılması için kurulan bir model %40 hata oranına sahipse;

$$RAE = \frac{\%40}{\%40} = 1 = 100 \text{ olarak hesaplanır. Bu işlemin amacı oluşturulan}$$

modelin hata oranını en basit sınıflandırma sisteminin (rastgele belirleyici) hata oranı ile karşılaştırma yaparak kullanılan modelin başarısını olabildiğince net ortaya çıkarmaktır. RAE değeri artması modelin başarısının azaldığını göstermektedir şeklinde yorum yapılabilir. RAE değeri %100 olarak hesaplanırsa oluşturulan modelin rassal tahminden farkı olmadığı bir başka deyişle tamamen anlamsız olduğu yorumu yapılabilir.

2.2 İlgili Çalışmalar

Bülbül (2003), tarafından yapılan çalışma lisansüstü öğretim programların öğrenci kabul etme sürecine ilişkin öğretim üyelerinin görüşlerini belirlemeyi amaçlamıştır. 2002-2003 öğretim yılında Ankara Üniversitesi, Eğitim Bilimleri Fakültesi'nde yapılan araştırmaya daha önce öğrenci kabul sürecine jüri üyesi olarak bulunmuş 13 öğretim üyesi katılmıştır. Öğretim üyelerinin lisansüstü öğretime öğrenci kabul süreciyle ilgili olarak sundukları öneriler hali hazırda uygulanmaya devam edilen süreçten vazgeçilerek, üniversitelerin kendi bünyelerinde hazırlayacakları yeni bir sınav yönteminin düzenlenmesi gerektiği, kullanılacak merkezi bir sınav sonucunun başarı puanındaki ağırlığının düşük tutulması gerektiğini , öğrenci seçme sürecinde yazılı sınav ve mülakatın birlikte kullanılmasının daha sağlıklı olabileceği şeklindedir. Öğretim üyeleri uygulanan sürecin üzerinde çalışılması gerektiğini savunmakta oldukları görülmüştür.

Tokat (2004) bu araştırmada, Lisansüstü Eğitimi Giriş Sınavı, mülakat puanı, lisans diploma notu ve bu puanları içeren birleştirilmiş bir değerlendirme sonucunun, yüksek lisans akademik not ortalamasını ne derecede yordadığını belirlemeye çalışmıştır. Araştırma: enstitüde lisansüstü eğitim gören 251 yüksek lisans öğrencinin bilgileri kullanılarak yapılmıştır. Yapılan analizler sonucu: lisansüstü eğitim gören bu öğrencilerin ALES puanı ile yüksek lisans akademik başarı not ortalamaları arasındaki ilişki anlamsız çıkmıştır. Bundan yola çıkarak: lisansüstü programlara öğrenci kabulünde kriter puan olarak yüksek ağırlıkla

kullanılan ALES puanlarının yüksek lisans başarı not ortalaması ile anlamlı bir ilişki göstermediği bulunmuştur. Lisans mezuniyet notu ile yüksek lisans akademik başarı not ortalaması arasında tezsiz yüksek lisans grubunda anlamlı bir ilişki ortaya çıktığı görülmezken: tezli yüksek lisans grubu için düşük düzeyde anlamlı bir ilişkinin ortaya çıktığı görülmüştür. Kullanılan puanları içeren birleştirilmiş değerlendirme sonucu ile yüksek lisans akademik başarı not ortalaması arasında düşük düzeyde anlamlı ilişki bulunurken: yüksek lisans başarı not ortalamalarını yordama da bütün gruplar için, ALES ve mülakat puanlarının; tezli grup için ise, sadece mülakattan alınan puanların anlamlı düzeyde katkı sağladığı sonucuna varılmıştır.

Kilmen (2007) tarafından yapılan bu çalışmada: lisansüstü eğitimi giriş sınavı'ndan alınan puanlar, lisans mezuniyet puanı ile yüksek lisans akademik başarı puan ortalaması arasındaki ilişkiler ve Lisansüstü Eğitimi Giriş Sınavı puanlarıyla lisans diploma notunun yüksek lisans başarı puan ortalamasını yordama gücü belirlenmeye çalışılmıştır. Lisansüstü Eğitimi Giriş Sınavı puanları, lisans mezuniyet puanı ile yüksek lisans akademik başarı puan ortalamaları arasındaki ilişkiler pearson momentler çarpım korelasyon katsayısı ile, lisans başarı ortalamasının ve lisansüstü eğitimi giriş sınavı puanlarının, yüksek lisans akademik başarı ortalamasını yordama gücü ise çoklu regresyon analizi ile hesaplanmıştır. Öğrencilerin lisansüstü eğitimi giriş sınavı puanları ile yüksek lisans başarı ortalamaları arasında anlamlı bir ilişki bulunmazken: lisans mezuniyet notları ile yüksek lisans akademik başarı ortalaması arasında, bulunan anlamlı ilişkinin orta düzeyde olduğu görülmüştür.

Şengür ve Tekin (2013) tarafından Fırat Üniversitesi'nde yapılan bir çalışmada; BÖTE bölümünden 2011 yılında mezun olan 127 öğrencinin eğitim süreci boyunca katılmış bulunduğu 49 dersin sene sonu not ortalamaları kullanılarak mezuniyet not ortalamaları yapay sinir ağları ve karar ağaçları algoritmaları yardımı tahmin edilmesi amaçlanmıştır. Yapay sinir ağları algoritması karar ağaçlarına göre daha başarılı olduğu görülürken: ikinci durumda her iki algoritmada da birinci duruma göre daha başarılı olduğu görülmüştür.

Koğar ve Sayın (2014) tarafından yapılan bir araştırmada: Gazi Üniversitesinde yüksek lisans programına başvuran 2566 öğrenci üzerinde çalışılmıştır. Çalışmanın amacı öğrencilerin yüksek lisans programına yerleştirilmesine ilişkin sürecin ne derece geçerli ve güvenilir olduğunu ortaya

koymak, yüksek lisans programına öğrenci kabulünde kullanılan kriterlerin ne derece geçerli olduğunu belirlemektir. Öğrencilerin yüksek lisans programlarına yerleştirilmesinde kullanılan ALES, lisans mezuniyet puanı, yabancı dil puanı ve mülakat puanı arasında düşük düzeyde bulunan ilişkinin anlamlı olduğu, yüksek lisansa yerleştirilme sürecinde kullanılan puanlardan en etkilisinin ise mülakat puanı olduğu sonucunda ulaşılmıştır. Ayrıca, sınıflama değerleri incelendiğinde yüksek lisans programına başvuran öğrencilerin doğru sınıflanma oranının %82 olduğu sonucuna varılmıştır. Son olarak öğrencilerin yüksek lisans programına yerleşmelerinde kullanılan: ALES, Lisans Mezuniyet Puanı, Yabancı Dil Puanı ve Mülakat Puanı arasında düşük düzeyde olduğu belirlenen ilişkinin anlamlı olduğu görülmüştür.

Ankara Üniversitesi'nde Karakul ve Karakütük (2014) tarafından öğretim elemanlarının, lisansüstü öğretime öğrenci seçiminde kullanılan ölçütlere ilişkin görüşlerini ortaya koymak amacıyla yapılan bu çalışma, 2007-2008 eğitim öğretim yılında Ankara Üniversitesi'nde görev yapmakta olan 274 öğretim elemanı katılımıyla yapılmıştır. Araştırmadan elde edilen sonuçlara göre öğretim elemanlarının: lisansüstü öğretime öğrenci kabulüne ilişkin ölçütlerin yetersizlikler ve olumsuzluklar barındırdığını düşündükleri görülmüştür. Çoğu araştırma görevlisi: yabancı dil bilme durumunun lisansüstü düzey için için önemli bir koşul olarak görmesine rağmen merkezi olarak yapılan yabancı dil sınavlarının yabancı dil bilme yeterliğini yeterli seviyede ölçtüğüne inanmadıkları görülmüştür. Bunun yanında ALES'in akademisyen olmak için gereken yetenek ve bulunması gereken davranış ve tutumları ölçtüğüne inanmadıkları görülmüştür. Çoğu araştırma görevlisinin: lisansüstü öğretime öğrenci kabulünde akademik not ortalamasının kullanılması gerektiğini düşündüğü görülürken ve mülakat yönteminin de dahil edilmesi gerektiğini düşündükleri görülmektedir.

Doğan vd. (2016) tarafından 2011 ile 2016 yılları arasında Hacettepe Üniversitesi lisansüstü programlarına başvuru yapan toplam 4990 adayın programlara kabul durumlarını etkileyen değişkenleri lojistik regresyon analizi ile yordamak amacıyla yapılan çalışmada, tezli yüksek lisans ve doktora başvuran aday öğrencilerin başvuru yaptıkları anabilim dalı, program, yıl ve yarıyıl, cinsiyet, medeni durumu, yabancı dil puanı, diploma notu, ALES puanı, ALES alanı, mezun olduğu okul, yaşı ve programa kabul durumları kullanılarak

oluşturulan modeller karşılaştırmalı olarak incelenmiştir. Çalışma sonucunda elde edilen hassaslık değerlerinin yüksek lisans için %28 ve doktora için %35 olması programa kabul etme açısından verilen kararın yeterli düzeyde hassas olmadığını göstermiştir. Yine bu çalışma da ALES puanının değerlendirilmedeki ağırlığının yüksek olmasına rağmen diğer değişkenlere göre katkısının daha az olması ALES'in değerlendirmelerdeki önem düzeyinin tartışılması gereken bir nokta olduğunu göstermektedir.

Aydemir (2017) tarafından Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Bölümünde yapılan bir çalışmada; 2009 - 2010 yılları arasında Meslek Yüksek Okullarına kayıt olan 1387 öğrencinin, 13 sorudan oluşan bir anket ile akademik bilgilerinin yanında ailevi bilgilerinden de yararlanılarak akademik başarılarını tahmin edilmeye çalışılmış ve en başarılı sınıflama algoritması belirlenmiştir. Bu tahmin için: öğrencilerin not ortalamaları ile mezun oldukları yıl kullanılmıştır. Not ortalamasına göre başarı tahmininde doğru sınıflandırma oranının %55 - %59 aralığında bulunduğu görülmüş ve en iyi sonucu Sıralı Minimum Optimizasyon(SMO) algoritmasının, Mezun olunan yıla göre başarı tahmininde doğru sınıflandırma oranı %78 - %81 aralığında bulunduğu görülmüş ve en iyi sonucu J4.8 ve NaiveBayes algoritmalarının verdiği belirlenmiştir.

Gök (2017) tarafından yapılan bir çalışmada 6., 7. ve 8. sınıf öğrencilerinin demografik bilgileri kullanılarak; Türkçe, matematik dersleri ve dönem sonu genel başarı ortalaması makine öğrenmesi algoritmaları yardımıyla tahmin edilmiştir. Kullanılan veriler sosyo demografik özellikleri farklı üç ortaokulda, 6, 7 ve 8. Sınıf öğrencilerine 24 soruluk bir anket uygulanarak elde edilmiştir. Elde edilen veriler kullanılarak: Türkçe, Matematik dersleri ve dönem sonu genel başarı ortalamalarını tahmin etmek için regresyon ve sınıflandırma yöntemi kullanılmıştır. Genel başarı ortalamasına ait not ve puan tahmininde regresyon ve sınıflandırma yöntemlerinin başarılarının yüksek olduğu görülürken: puan tahmini rastgele orman regresyon yöntemi ile not tahmini ise (korelasyon tabanlı öz nitelik alt kümesi) KÖAK yöntemi ve lojistik sınıflandırma algoritması beraber uygulandığında en iyi tahmini sonuç elde edildiği görülmüştür.

Akgün ve Özek (2020) tarafından yapılan çalışmada veri madenciliği uygulamaları ile ilgili yayınlanmış çalışmaların irdelenmesi yapılmıştır. Araştırmada 102 çalışma incelenmiştir. 2003-2019 yılları arasında konu ile ilgili

yapılan alıřmaların artıř gsterdięi belirlenmiř olup, alıřmaların ana odaęının farklı veri tabanları ile ęrencilerin gelecek dnemlerini oluřturan akademik bařarılarını tahmin etmeye ynelik alıřmalar olduęu gzlenmiřtir. İncelenen alıřmaların odaklandığı dięer nokta ise ęrenci tutum ve davranıřlarını belirlenmesi olduęu gzlemlenmiřtir. İlgili alıřmalarda oęunlukla kullanılan veri madencilięi uygulaması WEKA olurken en fazla kullanılan veri madenncilięi algoritmalarının: karar aęaları ve yapay sinir aęları olduęu grlmřtir.



3. YÖNTEM

Bu bölümde araştırmada kullanılan veri seti ve analiz yöntemlerine ilişkin bilgiler sunulmuştur.

3.1 Araştırmanın Modeli

Bu araştırmada, Abant İzzet Baysal Üniversitesi Lisansüstü Programlara öğrenci alımında kullanılan puanların ne kadar işlevsel olduğu bir başka deyişle öğrencilerin mezun olup olamama durumunu yordayıp yordamadığı saptanmaya çalışılmıştır.

Bu araştırmada genel tarama modellerinden ilişkisel tarama modeli kullanılmıştır. Geçmişte ya da hala var olan bir durumu, süregelen halini değiştirmeden betimlemeyi amaç edinen bir araştırma yaklaşımına tarama modeli denir. Araştırmaya konu olan olay, birey ya da nesne, sahip olduğu koşullara her hangi bir müdahalede bulunulmadan tanımlanmaya çalışılır. Koşullar herhangi bir şekilde değiştirilmez ve etkilenmez (Karasar, 2005). İlişkisel tarama modeli ise, iki ve daha fazla sayıdaki değişken arasındaki ortak değişimin varlığını belirlemeyi amaçlayan tarama yaklaşımına denir. İlişkisel tarama modelinde, değişkenlerin birlikte değişip değişmediği, değişme varsa bunun nasıl olduğu belirlenmeye çalışılır (Karasar, 2005).

Bu araştırma mevcut bilgilere yeni bilgiler eklemeyi amaçladığı için temel araştırma türündedir. Temel araştırmalar: halihazır şekilde elde edilmiş bilgilere yeni bilgileri eklemeyi amaçlayan araştırmalardır (Karasar, 2005).

Enstitüden elde edilen veri setindeki sayısal verilerle istatistiksel analizler kullanıldığı ve değişkenler arasındaki ilişkiyi keşfetmeye çalıştığı için ise nicel bir çalışma olduğu gibi değişkenler arası ilişkileri de belirleyip açıklamaya çalıştığı için ilişkisel araştırmadır. Nicel araştırmalar, değişkenler arası ilişkileri keşfedip bu ilişkileri açıklamaya amaçlayan araştırmalardır (Fraenkel ve Wallen, 2006). İlişkisel araştırmalar ise değişkenler arasındaki ilişkilerin açığa çıkarılması, açığa çıkarılan bu ilişkinin düzeyinin belirlenmesi için yapılacak yeni araştırmalara öngörü sağlar (Büyüköztürk vd., 2013).

3.2 Çalışma Grubu

Araştırmanın çalışma grubu: 2011-2016 yılları arasında Abant İzzet Baysal Üniversitesi Eğitim Bilimleri Enstitüsü yüksek lisans ve doktora programlarına kayıt olmuş 740 öğrenciden oluşmaktadır.

Çalışma grubunu oluşturan öğrencilerin %90,3'ü yüksek lisans, %9,7'sinin ise doktora düzeyinde öğrenim gördükleri belirlenmiştir. Öğrencilerin öğrenim gördükleri lisans üstü eğitim türlerinin dağılımına bakıldığında ise %93,2'sinin tezli, %6,2'sinin ise tezsiz olduğu gözlemlenmiştir. Çalışmada yer alan öğrencilerin %20,8'i mezun durumunda olduğu görülürken %79,2'sinin mezun değil durumunda olduğu görülmüştür. Örneklemin cinsiyet dağılımına bakıldığında ise %59,6'sını kadınların oluşturduğu görülürken %44,4'ünü ise erkeklerin oluşturduğu görülmüştür.

3.3 Veri Toplama

Araştırmada veri seti olarak, Bolu Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsündeki tezsiz yüksek lisans, tezli yüksek lisans ve doktora programlarına 2011 -2016 yılları arasında kayıt yaptırmış 740 öğrenciye ait veriler kullanılmıştır. Veri setinde her bireye ait ALES, yabancı dil, bilim sınavı ve mezuniyet puanları yer almaktadır. Ayrıca her bireyin kabul edildiği programdan mezun olup olmadığını gösteren kategorik bir değişken de veri setinde bulunmaktadır.

3.4 Verilerin Analizi

Araştırma kapsamında veri madenciliği algoritmalarından faydalanılmıştır. İlgili veri madenciliği algoritmaları: Weka 3.9.4 programı kullanılarak uygulanmıştır. Weka 3.9.4 programı: veri madenciliği ve makine öğrenmesi algoritmalarının kullanımını sağlayan, Waikato üniversitesinde geliştirilmiş bir yazılımdır. Weka programı: analiz öncesi verilerin düzenlenmesi, sınıflandırma (classification) algoritmaları, kümeleme (clustering) analizleri, birliktelik analizleri ve veri görselleştirme gibi işlevleri gerçekleştirebilmektedir. Weka Programı: csv, arff, xarff gibi formatları desteklemektedir.

Veriler weka programında kullanılabilmek için analize uygun hale getirilmiştir. Cinsiyet değişkeni “kadın/erkek” değerlerini, tezlilik değişkeni için “tezli, tezsiz” değerleri, öğrenim değişkeni için “yükseklisana/ doktora” değerleri, mezuniyet durumu değişkeni için ise “mezun/mezundeğil” değerleri kullanılırken, ALES, yabancı dil, bilim sınavı ve lisans mezuniyet puanı değişkenleri ise 1’den 100’e kadar sayısal değerler almıştır.

Bu çalışmada Weka 3.9.4’de sunulan sınıflandırma algoritmalarından K en yakın komşu (KNN), karar ağaçları (C4.5) ve lojistik regresyon yöntemleri kullanılmıştır. Kullanılan üç algoritmanın tercih edilmesinin sebepleri ; sınıflama algoritmalarının arasında güvenilirlikleri en yüksek algrtimalardan olmaları, çıktıların kolay yorumlanabilir ve anlaşılabilir olması, veri tabanları ile entegrasyonlarının kolaylığı, hem kategorik hem sayısal verilerle eş zamanlı çalışmaya izin vermesi ve kayıp verileri ayıklayarak analiz yapabilme imkanını sağlamasıdır.

Kullanılan algortimaların sınıflama performans kalitesini ölçmek için kullanılan başarı değerlendirme ölçütleri;

- Doğru Sınıflandırılmış Örnek Oranları:

Doğru sınıflandırılmış örnek oranı; modelin başarısını yüzde olarak ifade eder. %100’e yaklaştıkça modelin dopru sınıflama başarısı artar.

- Kappa İstatistiği:

Kappa değeri -1 ve +1 arasında değerler alır. Kappa değerlerinin negatif olması tutarlılık gücünün zayıf olduğunu gösterirken 0,00 – 0,20 aralığında önemsiz, 0,21 – 0,40 aralığında düşük, 0,41 - 0,60 aralığında orta, 0,61 – 0,80 aralığında önemli, 0,81 - 1,00 aralığında ise tutarlılık gücünün çok yüksek olduğunu gösterir.

- Bağlı Mutlak Hata (RAE) :

Bağlı mutlak hata değeri 0 ile sonsuz arasında değer alır. 0 ideal değere karşılık gelmektedir bir başka deyişle bu değerin artması modelin başarısının azaldığını gösterir.

4. BULGULAR

Bu bölümde: araştırmada WEKA programı yardımıyla ele alınan K-NN, Lojistik regresyon ve C4.5 algoritmalarından elde edilen analiz sonuçları incelenmiştir. Araştırmada kullanılan veri setine ait istatistiksel bilgilere ve her bir alt probleme ilişkin bulgulara yer verilmiştir.

Araştırmada kullanılan veri setine ait ALES, yabancı dil, bilim sınavı puanları ve mezuniyet not ortalamalarına ait istatistiksel bilgiler aşağıdaki tabloda gösterilmiştir.

Tablo 4.1. Veri setine ait istatistiksel bilgiler

	<i>Ortalama</i>	<i>Standart Sapma</i>	<i>Max</i>	<i>Min</i>
ALES Puanı	78,35	8,71	99	36
Bilim sınavı Puanı	66,09	17,71	100	5
Yabancı Dil Puanı	61,04	18,47	99	0
Mezuniyet Puanı	78,32	8,19	100	53

Veri seti istatistiksel bilgileri incelendiğinde ALES ortalamasının 78,35 , bilim sınavı ortalamasının 66,9, yabancı dil puanı ortalamasının 61,04 ve mezuniyet puanı ortalamasının ise 78,32 olduğu görülmektedir.

Veri setine ilişkin cinsiyet, program türü ve mezunluk durumu bilgileri aşağıdaki tabloda verilmiştir.

Tablo 4.2. Veri setine ilişkin cinsiyet, program türü ve mezunluk durumu bilgileri

	<i>f</i>	<i>%</i>
Yüksek Lisans	668	90,3
Doktora	72	9,7
Tezli	964	93,8
Tezsiz	46	6,2
Mezun	154	20,8
Mezun Değil	586	79,2
Kadın	441	59,6
Erkek	329	44,4

Tablo 4. 2. incelendiğinde veri setine dahil edilen öğrencilerden 441 (%59,6)' i kadın, 329 (%44,4)' u ise erkektir. Yüksek lisans programlarına kayıtlı öğrenci sayısı 668 iken doktora programlarına kayıtlı öğrenci sayısı 72' dir. Bunlardan 694 öğrenci tezli programlarda yer alırken 46 öğrenci tezsiz programlarda yer almaktadır. Kayıtlı bulunduğu programlardan mezun olan öğrenci sayısı 154 iken, mezun değil durumunda bulunan öğrenci sayısının 586 olduğu görülmektedir.

4.1 Birinci Alt Probleme İlişkin Bulgular

Araştırmanın birinci alt probleminde lisanüstü eğitime kabul kriterlerinin işlevselliğine bir başka deyişle öğrencilerin mezun olup olmama durumlarını yordayıp yordamadığına KNN Algoritması analizleri ile bakılmıştır. Veri kümesine KNN algoritmasının uygulanması sonucunda elde edilen çıktılar aşağıdaki tabloda gösterilmiştir. Yapılan denemeler sonucunda k değeri 19 olarak alınmıştır.

K-NN algoritmasının analiz sonuçlarından elde edilen karışıklık matrisi aşağıda verilmiştir.

Tablo 4.3. KNN algoritması karışıklık matrisi

=== Karışıklık Matrisi ===		
a	b	
2	44	a=mezun
1	205	b=mezundegil

Analiz sonuçlarından elde edilen karışıklık matrisi incelendiğinde gerçekte mezun durumunda olan öğrencilerden 2' si mezun durumunda olarak tahmin edilirken 44' ü mezun durumunda olmasına rağmen algoritma tarafından mezun değil olarak tahin edilmiştir. Bunun yanında 1 öğrenci mezun değil durumunda olmasına rağmen mezun sınıfında tahmin edilirken, mezun değil durumunda olan 205 öğrenci algoritma tarafından mezun değil durumunda tahmin edilmiştir.

KNN algoritmasına ait model başarımlı ölçütü sonuçları aşağıda verilmiştir.

Tablo 4.4. KNN algoritmasına ait model başarımlı ölçütü sonuçları

<i>K- En Yakın Komşu</i>	
Doğru sınıflandırma oranı	%82.14
Kappa istatistiğı	0.0606
Mutlak hata ortalaması(MAE)	0.2929
Ortalama Karekök Hatası(RMSE)	0.3753
Bağıl mutlak hata(RAE)	%90.45
Bağıl karekök hatası (RRSE)	%96.64

Tablo 4. 4. incelendiğinde K-NN algoritmasının %82.14 oranında doğru sınıflandırma yaptığı görülmektedir. Kappa istatistiğı değeri .0606 iken, Mutlak hata ortalaması (MAE) değeri .2929 ve Hata kareler ortalaması kökü (RMSE) değerinin ise .3753 olduğu sonucuna varılmaktadır. Göreceli mutlak hata (RAE) değeri %90.45 iken Bağıl karekök hatası (RRSE) değerinin %96.64 olduğu görülmektedir.

Birinci alt probleme ilişkin yapılan analizlerden elde edilen karışıklık matrisi incelendiğinde K-NN Algoritması sınav puanları ve mezuniyet durumları arasında anlamlı bir ilişki modeli kuramadıkları için, test verisindeki bireylerin çoğunluğunu frekansı en yüksek kategori olan mezun değil sınıfına dahil ettiği görülmektedir. Bu sınıflandırma modeli için elde edilen Kappa değerinin .0606 olduğu görülmektedir. Elde edilen kappa değerinin pozitif olması puanlayıcılar arasındaki tesadüfi uyumun beklenen uyumdan fazla göstermesine rağmen kappa değerinin .00 ile .20 arasında bulunmuş olması bu uyumun önemsiz olduğu göstermektedir. KNN algoritmasından elde edilen relative absolute error değeri ise %90.45 olarak hesaplanmıştır. Bu değerin %100'e oldukça yakın olarak bulunması kurulan modelin önemsiz olduğunu ve rassal tahminden neredeyse farksız olduğunu göstermektedir. Bu durum kullanılan sınav puanlarının mezuniyet durumu üzerinde yordayıcı bir gücü olmadığı şeklinde yorumlanabilir.

4.2 İkinci Alt Probleme İlişkin Bulgular

Araştırmanın ikinci alt probleminde lisanüstü eğitime kabul kriterlerinin işlevselliğine bir başka deyişle öğrencilerin mezun olup olmama durumlarını yordayıp yordamadığına C4.5 Algoritması analizleri ile bakılmıştır. Veri kümesine C4.5 algoritmasının uygulanması sonucunda elde edilen çıktılar aşağıdaki çizelgede gösterilmiştir.

C4.5 algoritmasının analiz sonuçlarından elde edilen karışıklık matrisi aşağıda verilmiştir.

Tablo 4.5. Karar ağaçları (C4.5) algoritmasına ait karışıklık matrisi

=== Karışıklık matrisi ===		
a	b	
1	45	a=mezun
0	206	b=mezundegil

C4.5 algoritmasının uygulanması ile elde edilen karışıklık matrisi incelendiğinde gerçekte mezun durumda olan öğrencilerden sadece 1'i mezun durumunda tahmin edilirken, gerçekte mezun durumda olan 45 öğrenci mezun değil durumunda şeklinde sınıflandırılmıştır. Gerçekte mezun değil durumda olan 206 öğrenci mezun değil sınıfında yer alırken, gerçekte mezun değil durumda yer alıp mezun olarak sınıflandırılan hiçbir öğrenci bulunmamaktadır.

Tablo 4.6. Karar ağaçları (C4.5) algoritmasına ait model başarımlar ölçütü sonuçları

<i>Karar ağaçları (C4.5)</i>	
Doğru sınıflandırma oranı	%82.14
Kappa istatistiği	0.0351
Mutlak hata ortalaması(MAE)	0.3128
Ortalama Karekök Hatası(RMSE)	0.3847
Bağıl mutlak hata(RAE)	%96.61
Bağıl karekök hatası (RRSE)	%99.06

Tablo 4. 6 incelendiğinde C4.5 algoritmasının %82.14 oranında doğru sınıflandırma yaptığı görülmektedir. Kappa istatistiği değeri .0351 iken, Ortalama Mutlak Hata (MAE) değeri .3128 ve Ortalama Karekök Hatası (RMSE) değerinin ise .3847 olduğu sonucuna varılmaktadır. Göreceli mutlak hata (RAE) değeri %96.61 iken Bağıl karekök hatası (RRSE) değerinin %99.06 olduğu görülmektedir.

İkinci alt probleme ilişkin yapılan analizlerden elde edilen karışıklık matrisi incelendiğinde C4.5 Algoritması sınav puanları ve mezuniyet durumları

arasında anlamlı bir ilişki modeli kuramadıkları için, test verisindeki bireylerin tümünü frekansı en yüksek kategori olan mezun değil sınıfına dâhil ettiği görülmektedir. Bu sınıflandırma modeli için elde edilen Kappa değerinin .0351 olduğu görülmektedir. Elde edilen kappa değerinin pozitif olması puanlayıcılar arasındaki tesadüfi uyumun beklenen uyumdan fazla göstermesine rağmen kappa değerinin .00 ile .20 arasında ve 0'a oldukça yakın bulunmuş olması bu uyumun önemsiz olduğu göstermektedir. C4.5 algoritmasından elde edilen relative absolute error değeri ise 96.61% olarak hesaplanmıştır. Bu değer %100'e oldukça yakın olarak bulunması kurulan modelin önemsiz olduğunu gösterirken kabul kriterleri olarak kullanılan sınav puanlarının mezuniyet durumu üzerinde yordayıcı bir gücü olmadığı göstermektedir.

4.3 Üçüncü Alt Probleme İlişkin Bulgular

Araştırmanın üçüncü alt probleminde lisanüstü eğitime kabul kriterlerinin işlevselliğine bir başka deyişle öğrencilerin mezun olup olmama durumlarını yordayıp yodamadığına lojistik regresyon analizleri ile bakılmıştır. Veri kümesine lojistik regresyon algoritmasının uygulanması sonucunda elde edilen çıktılar aşağıdaki çizelgede gösterilmiştir.

Lojistik regresyon algoritmasının analiz sonuçlarından elde edilen karışıklık matrisi aşağıda verilmiştir.

Tablo 4.7. Lojistik regresyon algoritması karışıklık matrisi

=== Karışıklık Matrisi ===		
a	b	
4	42	a=mezun
3	203	b=mezundegil

Lojistik regresyon analizinden elde edilen karışıklık matrisi incelendiğinde gerçekte mezun durumunda olan 4 öğrencinin mezun olarak sınıflandırıldığı görülürken 42 öğrencinin gerçekte mezun durumda olmasına rağmen mezun değil durumunda sınıflandırıldığı görülmektedir. Bunun yanında gerçekte mezun değil durumunda olan 3 öğrencinin tahmini de mezun olarak sınıflandırılırken 203 öğrencinin mezun değil olarak sınıflandırıldığı görülmektedir.

Lojistik regresyon algoritmasına ait model başarımlı ölçütü sonuçları aşağıda verilmiştir.

Tablo 4.8. Lojistik Regresyon Algoritmasına Ait Model Başarımlı Ölçütü

<i>Lojistik regresyon</i>	
Doğru sınıflandırma oranı	%82.14
Kappa istatistiğı	0.1079
Mutlak hata ortalaması(MAE)	0.3051
Ortalama Karekök Hatası(RMSE)	0.3817
Bağıl mutlak hata(RAE)	%94.22
Bağıl karekök hatası (RRSE)	%98.29

Tablo 4. 8. incelendiğinde lojistik regresyon algoritmasının %82.14 oranında doğru sınıflandırma yaptığı görülmektedir. Kappa istatistiğı değeri .1079 iken, Ortalama Mutlak Hata (MAE) değeri .3051 ve Ortalama Karekök Hatası (RMSE) değerinin ise .3817 olduğu sonucuna varılmaktadır. Göreceli mutlak hata (RAE) değeri %94.22 iken Bağıl karekök hatası (RRSE) değerinin %98.29 olduğu görülmektedir.

Üçüncü alt probleme ilişkin yapılan analizlerden elde edilen karışıklık matrisi incelendiğinde lojistik regreyon algoritmasında da test verisindeki bireylerin neredeyse tümünü frekansı en yüksek kategori olan mezun değil sınıfına dahil etmesi, sınav puanları ve mezuniyet durumları arasında anlamlı bir ilişki modeli kuramadığını göstermektedir. Bu sınıflandırma modeli için elde edilen Kappa değerinin .0351 olduğu görülmektedir. Elde edilen kappa değeri puanlayıcılar arasındaki tesadüfi uyumun beklenen uyumdan fazla olduğunu gösterirken bunun yanında bu uyumun önemsiz olduğunu da göstermektedir. Lojistik Resresyon algoritmasından elde edilen relative absolute error değeri ise 94.22% olarak hesaplanmıştır. Bu değeri oldukça yüksek olması kurulan modelin anlamsız olduğunu göstermektedir.

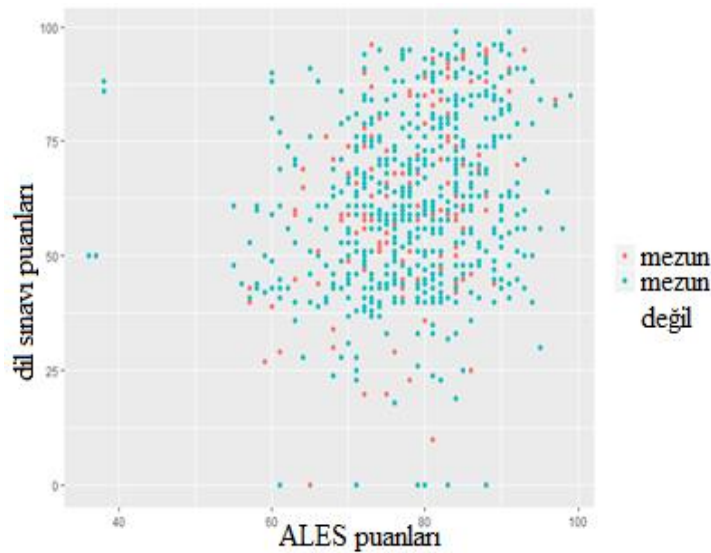
Araştırmada kullanılan üç sınıflandırma algoritmasına ait analiz sonuçları ve sınıflandırma sonuçlarının değerlendirilmesi için kullanılan göstergeler aşağıdaki tabloda gösterilmiştir.

Tablo 4.9. Üç algoritmaya algoritmaya ait analiz sonuçları

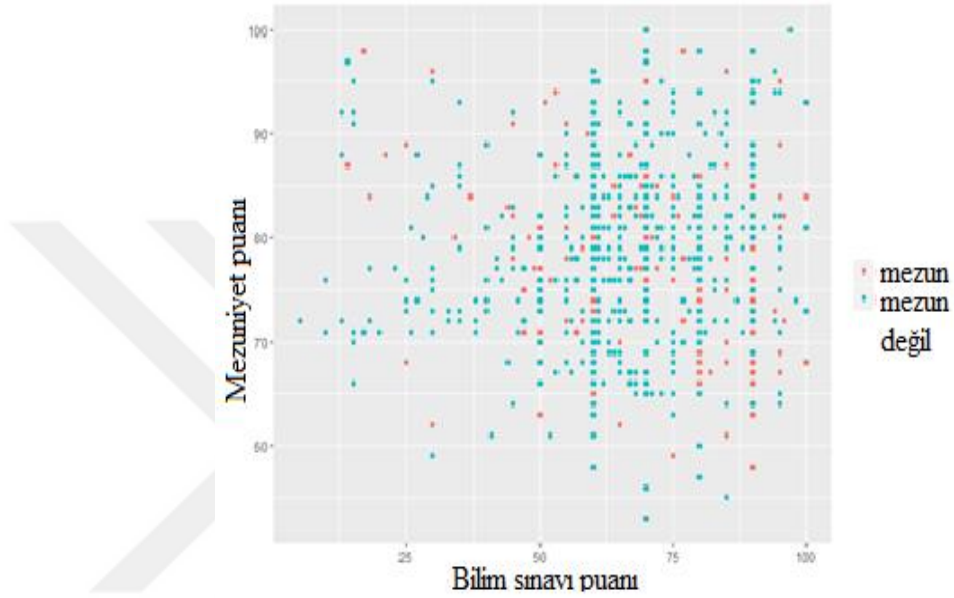
	<i>K-NN</i>	<i>C4.5</i>	<i>LR</i>
Doğru sınıflama oranı	%82.14	%82.14	%82.14
Kappa istatistiği	0.0606	0.0351	0.1079
Mutlak hata ortalaması(MAE)	0.2929	0.3128	0.3051
Ortalama Karekök Hatası(RMSE)	0.3753	0.3847	0.3817
Bağıl mutlak hata(RAE)	%90.45	%96.61	%94.22
Bağıl karekök hatası (RRSE)	%96.64	%99.06	%98.29

Bu üç algoritmaya ait doğru sınıflama oranları K-NN, C4.5 ve Lojistik regresyon algoritmalarında % 82.14 olarak bulunmuştur. Kappa istatistiği ise K-NN algoritmasında .0606, C4.5 algoritmasında .0351 ve lojistik regresyon algoritmasında ise .1079 olarak bulunmuştur. Mutlak Hata Ortalaması sırasıyla .2929, .3128 ve .3051 olarak bulunurken Ortalama Karekök Hatası sırasıyla .5516, .3847 ve .3817 olarak bulunmuştur. Bağıl karekök hatası ise sırasıyla %96.64, %99.06, %98.29 olarak bulunmuştur.

Üç algoritmadan elde edilen analiz sonuçları incelendiğinde de mezuniyet durumlarına göre sınav puanlarının rassal dağıldığı yorumu yapılabilir. Bu yorumu desteklemek amacı ile şekil 6. 1 ve şekil 6. 2’de sınav puanlarına ait plot grafiklerinin mezuniyet durumuna göre renklendirilmiş biçimleri sunulmuştur.

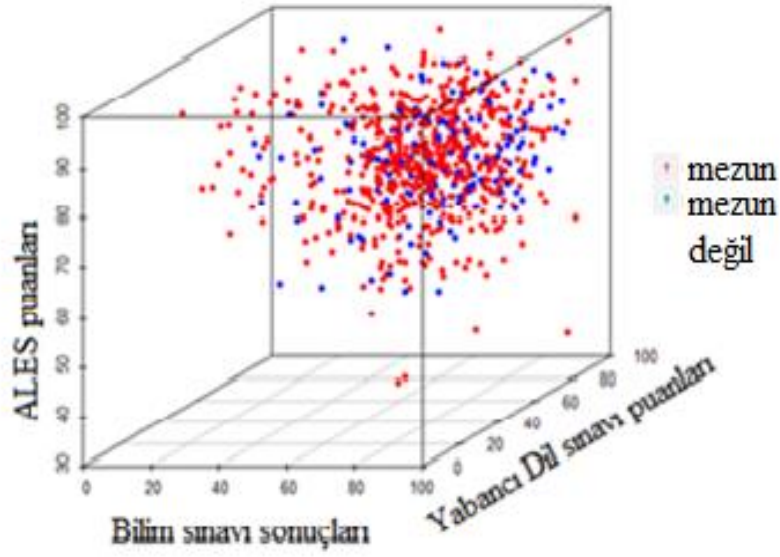
**Şekil 4.1.** ALES puanları ve yabancı dil puanlarına göre mezuniyet durumu

Şekil 4. 1. incelendiğinde ALES ve yabancı dil sınavı puanlarını gösteren grafiğe göre mezun öğrencilere ait noktaların kitlenin içerisinde rassal olarak dağıldığı görülmektedir. Her hangi bir sınav puanının belirli değer aralıklarına göre mezun öğrenci yoğunluğunun artış ya da azalış göstermemesi, sınav puanları ile mezuniyet durumu arasında bir yordayıcı ilişkisinin olmadığını göstergesi olarak kabul edilebilir.



Şekil 4.2. ALES sınavı puanları ve bir önceki programdan mezuniyet puanlarına ait plot grafiği

Şekil 4. 1.'deki duruma benzer olarak bilim sınavı puanları ve bir önceki programdan mezuniyet puanlarına ait plot grafiğinde de mezun öğrencilere ait noktaların kitlenin içerisinde homojen bir dağılım gösterdiği görülmektedir.



Şekil 4.3. Mezuniyet durumlarına göre; bilim sınavı puanları, ales, yabancı dil sınavı puanlarına ait üç boyutlu plot grafiği

Ek olarak bilim sınavı puanları, ALES, yabancı dil sınavı puanlarına ait 3 boyutlu plot grafiği de öğrencilerin mezuniyet durumuna göre renklendirilmiştir. Bu grafikte mezun öğrencilerin kitlenin içerisinde homojen olarak dağıldığı, 3 boyutlu dağılım grafiğinin her hangi bir bölgesinde yoğunlaşmadığı görülmüştür.

5. TARTIŞMA

Lisansüstü eğitime öğrenci kabulünde kullanılan kriter puanların işlevselliğinin eğitsel veri madenciliği uygulamaları yardımı ile tespit edilmesi amacı ile yapılan bu çalışmada edilen bulgular incelendiğinde kullanılan 3 farklı sınıflandırma algoritmasının oluşturduğu modellerin basit tahminleyici model ile aynı başarı düzeyinde sınıflandırmalar yaptığı görülmektedir. Yapılan araştırmadan elde edilen sonuçların basit tahminleyici model ile benzer olması oluşturulan modellerin tamamen anlamsız olduğu ya da kullanılan kriter puanlarının mezuniyet durumu üzerinde herhangi bir işlevselliğe sahip olmadığı, diğer bir deyişle ilgili kriter puanların öğrencilerin mezuniyet durumunun iyi birer yordayıcısı olmadığı şeklinde yorumlanabilir.

Araştırmada ulaşılan bulgular literatürdeki bazı çalışmalarla benzerlikler göstermektedir. Örneğin; Doğan vd. (2016) tarafından 2011 ile 2016 yılları arasında Hacettepe Üniversitesi lisansüstü programlarına başvuru yapan toplam 4990 adayın programlara kabul durumlarını etkileyen değişkenleri lojistik regresyon analizi ile yordamak amacıyla yapılan çalışmada, tezli yüksek lisans ve doktora başvuran aday öğrencilerin başvuru yaptıkları anabilim dalı, program, yıl ve yarıyıl, cinsiyet, medeni durumu, yabancı dil puanı, diploma notu, ALES puanı, ALES alanı, mezun olduğu okul, yaşı ve programa kabul durumları kullanılarak oluşturulan modeller karşılaştırmalı olarak incelenmiştir. Çalışma sonucunda elde edilen hassaslık değerlerinin yüksek lisans için %28 ve doktora için %35 olması programa kabul durumunda verilen kararın yeterli düzeyde hassas olmadığını göstermiştir. Yine bu çalışma da ALES puanının değerlendirmedeki ağırlığının yüksek olmasına rağmen diğer değişkenlere göre daha az katkı sağlaması ALES'in değerlendirmedeki öneminin tartışılması gereken bir nokta olduğunu göstermektedir. Bunun yanında Bülbül (2003) tarafından öğrenci seçme sürecine ilişkin öğretim üyelerinin görüşlerini belirlemek amacıyla yapılan bir çalışmada süreç ile ilgili verdikleri önerilerin başında uygulanmakta olan süreçten vazgeçilmesi, üniversitelerin kendi bünyelerince hazırlanan yeni bir sınav sisteminin geliştirilmesine ihtiyaç duyulduğu, eğer merkezi bir sınav sonucu kullanılmak isteniyorsa bunun başarı puanındaki ağırlıklarının çok düşük olması, seçim sürecinde yazılı ve mülakatın birlikte kullanılması gerektiği görüşü paylaşılmaktadır.

Lisansüstü Eğitimi Giriş Sınavı, mülakat puanı, lisans diploma notu ve bu puanları içeren birleştirilmiş bir değerlendirme sonucunun, yüksek lisans akademik not ortalamasını ne derecede yordadığını belirlemeye çalışan Tokat (2004), analizler sonucunda lisansüstü eğitim gören öğrencilerin, Lisansüstü Eğitimi Giriş Sınavı puanı ile yüksek lisans akademik başarı not ortalamaları arasındaki ilişkiyi anlamsız bulmuştur. Lisans mezuniyet notu ile yüksek lisans akademik başarı not ortalaması arasında; tezsiz yüksek lisans grubunda anlamlı bir ilişki ortaya çıktığı görülmezken, tezli yüksek lisans grubu için düşük düzeyde anlamlı bir ilişkinin olduğu görülmüştür. Kullanılan puanları içeren birleştirilmiş değerlendirme sonucu ile yüksek lisans akademik başarı not ortalaması arasında düşük düzeyde anlamlı ilişki bulunurken, yüksek lisans başarı not ortalamalarını yordama da bütün gruplar için, ALES ve mülakat puanlarının; tezli grup için ise, sadece mülakattan alınan puanların anlamlı düzeyde katkı sağladığı sonucuna varılmıştır. Ankara Üniversitesi'nde çalışan öğretim elemanlarının, lisansüstü öğretime öğrenci seçiminde kullanılan ölçütlere ilişkin görüşlerini ortaya koymak amacıyla Karakul ve Karakütük (2014) tarafından yapılan çalışmada öğretim elemanlarının, ALES'in akademisyen olmak için gereken yetenek ve bulunması gereken davranış ve tutumları ölçtüğüne inanmadıkları görülmüştür. Çoğu araştırma görevlisinin lisansüstü öğretime öğrenci kabulünde akademik not ortalamasının kullanılması gerektiğini düşündüğü görülürken, mülakat yönteminin de dahil edilmesi gerektiğini düşündükleri görülmektedir. Bunların yanında lisansüstü öğretime öğrenci kabulüne ilişkin ölçütlerin yetersizlikler ve olumsuzluklar barındırdığını düşündükleri görülmüştür.

Özetle; araştırma kullanılan 3 algorithmadan elde edilen analiz sonuçları incelendiğinde lisansüstü eğitime giriş için kullanılan kabul kriterlerinin işlevsiz olduğu bir başka deyişle diğer bir deyişle öğrencilerin mezun olup olmama durumlarını yordayamadığı görülmüştür. Alanyazın da benzer konular üzerine yapılan çalışmalar incelendiğinde bir çoğunda kabul kriterlerinin işlevselliği ve mezun olup olmama durumunu yordayıcılığı ile ilgili problemler olduğu üzerinde hem fikir olunduğu ve bu konuda düzenleme yapılması gerektiği sonucuna varılmıştır.

6. SONUÇ VE ÖNERİLER

6.1 Sonuç

Araştırmanın birinci alt problemine ilişkin elde edilen sonuçlar incelendiğinde KNN Algoritması ile kriter puanlar ve mezuniyet durumları arasında anlamlı bir ilişki modeli kurulamadığı görülmüştür. Elde edilen değerlere bakıldığında kullanılan kriter puanların işlevsel olmadığı bir başka deyişle mezuniyet durumu üzerinde yordayıcı bir gücü olmadığı görülmüştür.

Araştırmanın ikinci alt problemine ilişkin elde edilen sonuçlar incelendiğinde C4.5 Algoritması ile kriter puanlar ve mezuniyet durumları arasındaki uyumun önemsiz olduğu görülmüştür. Elde edilen değerler kullanılan kriter puanların işlevsel olmadığı bir başka deyişle mezuniyet durumu üzerinde yordayıcı bir gücü olmadığı göstermektedir.

Araştırmanın üçüncü alt problemine ilişkin elde edilen sonuçlara bakıldığında lojistik regresyon Algoritmasının kriter puanlar ve mezuniyet durumları arasında anlamlı bir ilişki modeli kurulamadığı görülmüştür. Elde edilen değerler ise kullanılan kriter puanların işlevsel olmadığı bir başka deyişle mezuniyet durumu üzerinde yordayıcı bir gücü olmadığı görülmüştür.

Çalışma sonucunda ALES, yabancı dil, bilim sınavı ve mezuniyet puanlarının öğrencilerin lisansüstü programdan mezun olma durumunu açıklamadığı görülmüştür. Lisansüstü eğitimin bireyler için zorunlu olmaması, iş bulmak için zorunlu bir koşul teşkil etmemesinin program başarısında bazı duyuşsal ve psikolojik değişkenlerin etkili olabileceği şeklinde yorumlanabilir. Bu durum lisansüstü program başarısının yalnızca bireylerin bilişsel ve akademik düzeyleri ile açıklanamayacağını göstermektedir. Lisansüstü program başarısının bireylerin akademik tutum, motivasyon, öz saygı, yaratıcılık, öz yeterlik algısı gibi psikolojik değişkenlerden etkilenebileceği düşünülmektedir.

Bu sonuçlar lisansüstü öğrenci alımında bahsedilen 4 puan türünün öğrenci düzeyi temsil eden bu kriterlerin öğrencilerin duyuşsal ve psikolojik durumlarını temsil eden farklı ölçme sonuçları ile desteklenmesi gerektiğine işaret etmektedir.

6.2 Öneriler

6.2.1 Araştırma Bulgularına Dayalı Öneriler

Lisanüstü eğitime öğrenci kabulünde kullanılan kriterlerin öğrencilerin mezun olma durumunu yordayamadığı, başka bir deyişle öğrenci başarısını belirlemede yetersiz olduğu ve bu kriterlerin karar alıcılar tarafından gözden geçirilerek; akademik tutum, motivasyon, öz yeterlilik gibi psikolojik bazı ölçeklerin de öğrenci kabulünde kullanılabileceği gerektiği düşünülmektedir. Bunun yanında objektif bir yüz yüze mülakat uygulamasının öğrencilerin bahsi geçen psikolojik özelliklerinin ortaya konmasında etkili olabileceği düşünülmektedir.

6.2.2 Diğer Araştırmacılar İçin Öneriler

- Araştırmadan elde edilen sonuçlarının daha anlamlı ve güvenilir olabilmesi için, farklı ve daha büyük veri setleriyle tekrar edilmesi faydalı olacaktır.
- Araştırma da kullanılan K en yakın komşu, karar ağaçları (C4.5) ve lojistik regresyon algoritmalarının yanında bu Algoritmaların sayısı artırılarak veya algoritmalar çeşitlendirilerek çalışma geliştirilebilir.
- Bu araştırmanın sınırlılığı olan WEKA programının kullanılmasının yanı sıra diğer veri madenciliği programlarının da kullanılması araştırmayı daha verimli hale getirecektir.
- Bu çalışmadan elde edilen sonuçların yanında diğer paket programların içerdiği algoritmalar ve gelecekte yapılacak araştırmaların bir bütün halinde ele alınarak daha verimli kriterler oluşturulması gerektiği düşünülmektedir.
- Bunların yanında veri setine dahil edilen katılımcıların bir önceki mezuniyet tarihinin üzerinden geçen süre ve cinsiyet durumuna bağlı değişiklikler özellikle incelenebilir.

Tüm bunların yanında, Bolu Abant İzzet Baysal Üniversitesi ile sınırlı kalan bu araştırma farklı üniversitelerden elde edilecek daha büyük veri setleri ile çok daha verimli ve faydalı hale getirilebilir.

7. KAYNAKLAR

(Bu tez çalışmasında APA 7th Edition atıf sistemi kullanılmıştır.)

- Akgün, K., ve Bulut Özek, M. (2020). Eğitsel veri madenciliği yöntemi ile ilgili yapılmış çalışmaların incelenmesi: içerik analizi. *Uluslararası Eğitim Bilim ve Teknoloji Dergisi*, 6(3), 197–213. <https://doi.org/10.47714/uebt.753526>
- Akpınar, H. (2000). Veri tabanlarında bilgi keşfi ve veri madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29(1), 1-22.
- Alan, M. A. (2012). Veri madenciliği ve lisansüstü öğrenci verileri üzerine bir uygulama. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (33), 165-174.
- Albayrak, A. S., ve Yılmaz-Koltan, S. (2009). Veri madenciliği: karar ağacı algoritmaları ve imkb verileri üzerine bir uygulama. *Süleyman Demirel University Journal of Faculty of Economics ve Administrative Sciences*, 14(1), 31 – 52.
- Altaş, D., ve Gürpınar, V. (2012). Karar ağaçları ve yapay sinir ağlarının sınıflandırma performanslarının karşılaştırılması: Avrupa Birliği örneği. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 14(1), 1–22.
- Altunkaynak, B. (2017). Veri madenciliği yöntemleri ve R uygulamaları. Seçkin Yayıncılık
- Artsın, M. Veri madenciliği ve bilgi keşfi (Kitap özeti). *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 5(3), 174-180.
- Aydemir, B. (2017). Veri madenciliği yöntemleri kullanarak meslek yüksek okulu öğrencilerinin akademik başarı tahmini [Yükseklisans tezi, Pamukkale Üniversitesi]. YÖK Tez Merkezi.
- Aydın, S. (2007). Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama [Doktora tezi, Anadolu Üniversitesi]. YÖK Tez Merkezi.
- Baker, R. S. J. D., ve Yacef, K. (2009). The state of educational data mining in 2009 : A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–16.
- Berland, M., Baker, R. S., ve Blikstein, P. (2014). Educational data mining and learning analytics: applications to constructionist research. *Technology, knowledge and learning*, 19(1–2), 205–220. <https://doi.org/10.1007/s10758-014-9223-7>
- Bıkmaz-Bilgen, Ö. ve Doğan, N. (2017). Puanlayıcılar arası güvenirlik belirleme tekniklerinin karşılaştırılması. *Journal of Measurement and Evaluation in Education and Psychology*, 8 (1) , 63-78. DOI: 10.21031 /epod.294847
- Bilgin, M. (2017). Gerçek veri setlerinde klasik makine öğrenmesi yöntemlerinin performans analizi. *Breast*, 2(9), 683.
- Bircan, H. (2004). Lojistik regresyon analizi: Tıp verileri üzerine bir uygulama. *Kocaeli Üniversitesi Sosyal Bilimler Dergisi*, (8), 185-208.
- Bozan, M. (2012). Lisansüstü eğitimde nitelik arayışları. *Sosyal ve Beşeri Bilimler Dergisi*, 4(2), 177-187.
- Brodley, C. E., ve Friedl, M. A. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3), 399–409. [https://doi.org/10.1016/S0034-4257\(97\)00049-7](https://doi.org/10.1016/S0034-4257(97)00049-7)
- Bülbül, T. (2003). Ankara üniversitesi eğitim bilimleri fakültesinde görev yapan öğretim üyelerinin lisansüstü öğretime öğrenci seçme sürecine ilişkin görüşleri. *Ankara Üniversitesi*

- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2013). *Bilimsel araştırma yöntemleri* (15. Baskı). Ankara: Pegem Akademi.
- Çalış, A., Kayapınar, S., ve Çetinyokuş, T. (2014). Veri madenciliğinde karar ağacı algoritmaları ile bilgisayar ve internet güvenliği üzerine bir uygulama. *Endüstri Mühendisliği*, 25(3), 2-19.
- Çıkrıkçı-Demirtaşlı, N. (2002). Lisansüstü eğitim programlarına girişte lisansüstü eğitime giriş sınavı (İS) sonucunun ve diğer ölçütlerin kullanımına ilişkin bir tarama. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 001–010. https://doi.org/10.1501/egifak_00000000064
- Coşkun, C., ve Baykal, A. (2011). Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması. *Akademik Bilişim*, 11, 51-58.
- Danacı, M., Çelik, M., ve Akkaya, A. E. (2010). Veri madenciliği yöntemleri kullanılarak meme kanseri hücrelerinin tahmin ve teşhisi. *Akıllı Sistemlerde Yenilikler ve Uygulama Sempozyumu*, 21(24), 9-12.
- Doğan, N., Koyuncu, İ., Gökdemir, P., ve Kahveci, M. (2016). Öğrencilerin eğitim bilimleri enstitüsü lisansüstü programlarına kabul durumlarının yordanması. *The Journal of Educational Research*, 2(2), 115-131.
- Fraenkel, J. R., ve Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill, 471-473.
- Gök, M. (2017). Makine öğrenmesi yöntemleri ile akademik başarının tahmin edilmesi. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, 5(3), 139-148. <https://dergipark.org.tr/gujsc/issue/31140/311082>
- Gümüşçü, A., Taşaltın, R., ve Aydılek, İ. B. (2016). C4. 5 decision tree pruning using genetic algorithm. *Dicle Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 5(2), 77-80.
- Han, J., ve Kamber, M. (2006). *Data mining concepts and techniques* (2nd ed.): Morgan Kaufmann Publishers.
- Hand, D. J. (2007). Principles of data mining. *Drug Safety*, 30(7), 621–622. <https://doi.org/10.2165/00002018-200730070-00010>
- Hark, C. (2013). Öğrencilerin akıllı tahtaya ilişkin tutumlarının incelenmesine yönelik bir veri madenciliği uygulaması. [Yüksek lisans tezi, Fırat Üniversitesi]. YÖK Tez Merkezi.
- Holsheimer, M., ve Siebes, A. (1994). *Data mining: The search for knowledge in databases*. Amsterdam: Centrum voor Wiskunde en Informatica.
- Ebren Kara, Ş., ve Şamlı, R. (2021). Yazılım projelerinin maliyet tahmini için WEKA’da makine öğrenmesi algoritmalarının karşılaştırmalı analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (23), 415-426.
- Karakul, A. K., ve Karakütük, K. (2014). Views of academics on student selection criteria for graduate education. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 47(2), 179-200.
- Karaman, S., ve Bakırcı, F. (2010). Türkiye’de lisansüstü eğitim: Sorunlar ve çözüm önerileri. *Sosyal Bilimler Araştırmaları Dergisi*, 5(2), 94-114.
- Karasar, N. (2005). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayın Dağıtım
- Kaya, H., ve Köymen, K. (2008). Veri madenciliği kavramı ve uygulama alanları, *Doğu Anadolu Bölgesi Araştırmaları*, 2008, 159-164.
- Kilmen, S. (2007). Lisansüstü eğitimi giriş sınavının ve lisans diploma notunun yüksek lisans

- başarısını yordama gücü. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 7(2), 177-189.
- Koğar H., & Sayin, A. (2014). Analysis in the accuracy of classification in students graduate school entrance. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 47(1), 171-186. https://doi.org/10.1501/Egifak_0000001322
- Kutlugün, M. A., Çakir, M. Y., ve Kiani, F.(2017). Yapay Sinir Ağları ve K-En Yakın Komşu Algoritmalarının Birlikte Çalışma Tekniği (Ensemble) ile Metin Türü Tanıma. *İstanbul Sabahattin Zaim Üniversitesi*.
- Landis, J. R., ve Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363-374.
- Mohamad, S. K., ve Tasir, Z. (2013). Educational Data Mining: A Review. *Procedia - Social and Behavioral Sciences*, 97, 320–324. <https://doi.org/10.1016/j.sbspro.2013.10.240>
- Narli, S., Aksoy, E., ve Ercire, Y. E. (2014). Investigation of prospective elementary mathematics teachers' learning styles and relationships between them using data mining. *International Journal of Educational Studies in Mathematics*, 1(1), 37-57.
- Özbay, Ö. (2015). Veri madenciliği kavramı ve eğitimde veri madenciliği uygulamaları. *The Journal of International Educational Sciences*, 2(5), 262–262. <https://doi.org/10.16991/inesjournal.162>
- Özdamar, K. (2004) Paket Programlar ile İstatiksel Veri Analizi 1, Kaan Kitabevi.
- Özdemir, A., Saylam, R., ve Betül BİLEN, B. (2018). Eğitim sisteminde veri madenciliği uygulamaları ve farkındalık üzerine bir durum çalışması. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi Ekim*, 22, 2159–2172.
- Özdemir, Ş., ve Balaban, E. (2016). Eğitimde veri madenciliği ve öğrenci akademik başarı öngörüsüne ilişkin bir uygulama. [Doktora Tezi, *İstanbul Üniversitesi*]. YÖK Tez Merkezi.
- Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 2(3), 65–82.
- Özkan, M. (2014). Veri madenciliğinin finansal kararlarda kullanımı. *Çankırı Karatekin Üniversitesi İİBF Dergisi*, 4(1), 59–82.
- Öztürk, A. (2018). Açık ve uzaktan öğrenme ortamlarında eğitsel veri madenciliği. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 4(2), 10–13. <http://educationaldatamining.org>
- Romero, C., Ventura, S., ve Bra, P. D. (2004). Knowledge discovery with genetic programming for providing feedback to courseware authors. *User Modeling and User-Adapted Interaction*, 14, 425-464.
- Romero, C., ve Ventura, S. (2010a). 04th International Conference on Educational Data Mining EDM 2011. In *Proceeding 390p* (Issue December).
- Romero, C., ve Ventura, S. (2010b). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)*, 40(6), 601-618.
- Sayad, S.(2022). Erişim tarihi: 19 Ocak 2023, http://www.saedsayad.com/decision_tree.htm
- Silahtaroglu, G. (2008). *Kavram ve algoritmalarıyla temel veri madenciliği*. Papatya Yayıncılık Eğitim.172.
- Şekeroğlu, S. (2010). Hizmet sektöründe bir veri madenciliği uygulaması. [Doktora Tezi, *İstanbul Teknik Üniversitesi*]. YÖK Tez Merkezi.

- Şengür, D., ve Tekin, A. (2013). Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini. *Bilişim Teknolojileri Dergisi*, 6(3), 7–16.
- Taş, Y. (2018). Birliktelik kuralları madenciliği bir uygulama. [Yüksek lisans Tezi, Cumhuriyet Üniversitesi]. YÖK Tez Merkezi.
- Tokat, E. (2004). Lisansüstü eğitimi giriş sınavı ve diğer kabul ölçülerin yordama geçerliğine ilişkin bir araştırma. [Yayımlanmamış yüksek lisans tezi, Ankara Üniversitesi]. YÖK Tez Merkezi.
- Turgut, H. (2019). Veri madenciliği süreci kullanılarak alzheimer hastalığı teşhisine yönelik bir uygulama. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Turgut, M. F., ve Baykul, Y. (2012). Eğitimde ölçme ve değerlendirme. *Pegem Akademi*.
- Türkoğlu A., Çiftçi M., Çiçek M. ve Günay A. “ Üniversite izleme ve değerlendirme genel raporu 2022”. Yüksek Öğretim Kurumu, 2022.
- Tüzüntürk, S. (2010). Veri madenciliği ve istatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 1(2001), 65–90.
- Witten, I. H., Hall, M. A., ve Frank, E. (2005). Data mining: practical machine learning third edition. (*Morgan Kaufmann Series in Data Management Systems*). *Morgan Kaufmann*, 104(June), 113.
- Yılmaz, Ş. K., ve Patır, S. (2011). Kümeleme analizi ve pazarlamada kullanımı. *Akademik Yaklaşımlar Dergisi*, 2(1), 91–113.
- Yüksek öğretim bilgi yönetim sistemi. Erişim Tarihi: 01.02.2023, <https://istatistik.yok.gov.tr/> .

8. EKLER

EK 1 KNN Algoritması Analiz Çıktıları

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
\\weka.core.EuclideanDistance -R first-last\""

Relation: selcan data new kategorik

Instances: 740

Attributes: 9

id

cinsiyet

ALES

dil

Bilim

lisanspuan

Tezliklik

Dokt

Mezuniyet

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

IB1 instance-based classifier

Time taken to build model: 0 seconds

using 19 nearest neighbour(s) for classification

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,043	0,005	0,667	0,043	0,082	0,138	0,690	0,303	mezun
0,995	0,957	0,823	0,995	0,901	0,138	0,690	0,899	mezundegil

EK 2 C4.5 Algortiması Analiz Çıktıları

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: selcan data new kategorik

Instances: 740

Attributes: 9

id
cinsiyet
ALES
dil
Bilim
lisanspuan
Tezlilik
Dokt
Mezuniyet

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree

id <= 617

| Bilim <= 75

| | Tezlilik = tezli: mezundegil (411.0/81.0)

| | Tezlilik = tezsiz

| | | id <= 27

| | | | id <= 22

| | | | | dil <= 60: mezundegil (3.0)

| | | | | dil > 60: mezun (2.0)

| | | | | id > 22: mezun (5.0)

| | | | id > 27: mezundegil (16.0)

| Bilim > 75

| | Tezlilik = tezli

| | | cinsiyet = erkek

| | | | Bilim <= 92

| | | | | id <= 331

| | | | | | Bilim <= 86

| | | | | | Bilim <= 81

| | | | | | | id <= 220: mezundegil (4.0/1.0)

| | | | | | | id > 220: mezun (5.0/1.0)

| | | | | | Bilim > 81: mezun (7.0)

| | | | | Bilim > 86

| | | | | | id <= 137: mezun (5.0/1.0)

| | | | | | id > 137: mezundegil (6.0)

| | | | | id > 331: mezundegil (20.0/4.0)

| | | | Bilim > 92: mezun (7.0/1.0)

| | | cinsiyet = kadin

| | | | lisanspuan <= 68

| | | | | lisanspuan <= 67

| | | | | | lisanspuan <= 66

| | | | | | | lisanspuan <= 63: mezundegil (3.0/1.0)

| | | | | | | lisanspuan > 63: mezun (4.0)

| | | | | | | lisanspuan > 66: mezundegil (2.0)

| | | | | | | lisanspuan > 67: mezun (3.0)

| | | | | | | lisanspuan > 68: mezundegil (87.0/23.0)

| | | Tezlilik = tezsiz

| | | | cinsiyet = erkek: mezundegil (14.0/1.0)

| | | | cinsiyet = kadin

| | | | | dil <= 31: mezun (2.0)

| | | | | dil > 31: mezundegil (5.0)

| id > 617: mezundegil (129.0/6.0)

Number of Leaves : 21

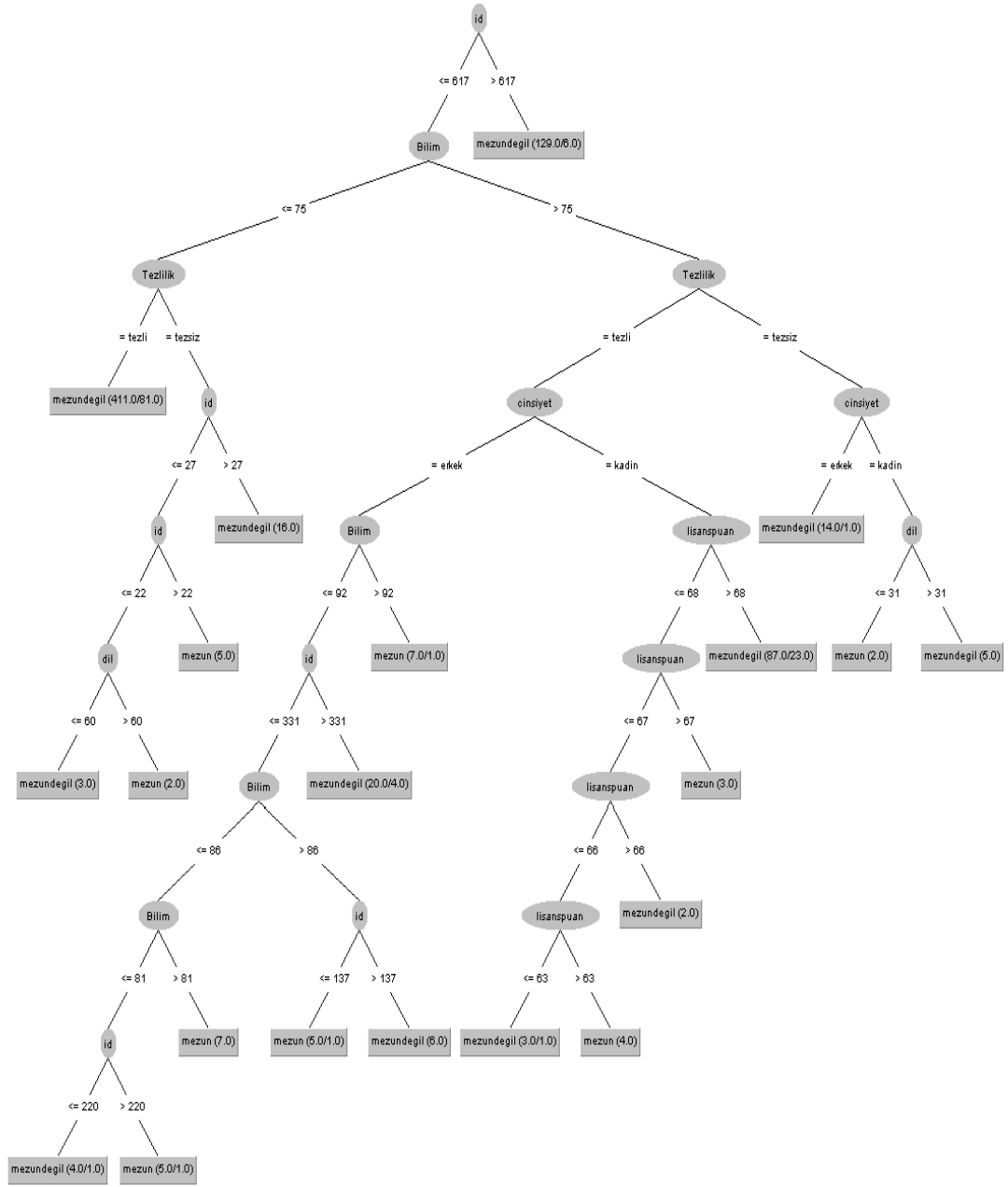
Size of the tree : 41

Time taken to build model: 0.08 seconds

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,022	0,000	1,000	0,022	0,043	0,134	0,528	0,214	mezun

EK 3 C4.5 Karar Ağacı Yapısı



EK 4 Lojistik Regresyon Algortiması Analiz Çıktıları

=== Run information ===

Scheme: weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4

Relation: selcan data new kategorik

Instances: 740

Attributes: 9

id
cinsiyet
ALES
dil
Bilim
lisanspuan
Tezlilik
Dokt
Mezuniyet

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===

Logistic Regression with ridge parameter of 1.0E-8

Coefficients...

Variable	Class mezun
id	-0.0028
cinsiyet=kadin	0.0749
ALES	-0.0156
dil	0.0091
Bilim	0.0167
lisanspuan	-0.0082
Tezlilik=tezsiz	-0.6343
dokt=doktora	0.1724
Intercept	-0.1931
Odds Ratios...	

Variable	Class mezun
id	0.9972
cinsiyet=kadin	1.0778
ALES	0.9845
dil	1.0092
Bilim	1.0168
lisanspuan	0.9918
Tezlilik=tezsiz	0.5303
dokt=doktora	1.1882

Time taken to build model: 0.02 seconds

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,087	0,015	0,571	0,087	0,151	0,170	0,643	0,318	mezun
0,985	0,913	0,829	0,985	0,900	0,170	0,643	0,896	mezundegil