



T.C.
KONYA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



**BÜTÜNLEME SINAVINA GİRECEK
ÖĞRENCİ SAYISININ KARAR AĞACI VE
YAPAY SİNİR AĞLARI İLE TAHMİN
EDİLMESİ**

Miyase Nur ŞENKAYA

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Ocak-2024
KONYA
Her Hakkı Saklıdır
TEZ KABUL VE ONAYI

Miyase Nur ŞENKAYA tarafından hazırlanan “**Bütünleme Sınavına Girecek Öğrenci Sayısının Karar Ağacı ve Yapay Sinir Ağları ile Tahmin Edilmesi**”adlı tez çalışması 25 /01/2024 tarihinde aşağıdaki jüri tarafından oy birliği ile Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

İmza

Başkan

Unvanı Adı SOYADI
(Üniversitesi)

.....

Danışman

Unvanı Adı SOYADI
(Üniversitesi)

.....

Üye

Unvanı Adı SOYADI
(Üniversitesi)

.....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Mevlüt UYAN
Enstitü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

İmza

Miyase Nur Şenkaya

Tarih: 25/01/2024

ÖZET

YÜKSEK LİSANS TEZİ

BÜTÜNLEME SINAVINA GİRECEK ÖĞRENCİ SAYISININ KARAR AĞACI VE YAPAY SİNİR AĞLARI İLE TAHMİN EDİLMESİ

Miyase Nur ŞENKAYA

Konya Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Semiye DEMİRCAN
2024, 44 Sayfa

Jüri

Dr. Öğr. Üyesi Semiye Demircan
Diğer Üyenin Unvanı Adı SOYADI
Diğer Üyenin Unvanı Adı SOYADI

Bütünleme sınavı, genel sınav sonucunda herhangi bir dersten başarısız olan öğrenciye dersten geçebilmesi için ikinci bir şans tanınması amacıyla yapılan ve genel sınav yerine geçen bir sınav türüdür. Ancak çeşitli nedenlerden öğrencilerin dikkate değer bir kısmı bütünleme sınavı hakkını kullanmamakta ve bu sınava girmemektedir. Bu tez çalışmasında sınava girmeyecek öğrenci sayısının öngörülmesi ile gereksiz sayıda soru kâğıdı basılması, sınav salonlarının gereksiz yere açılması ve gözetmenlerin gereksiz yere görevlendirilmesi gibi kaynak, enerji ve iş gücünün gereksiz harcanmasının önüne geçilmesi hedeflenmiştir. Bu tez çalışmasında kaynakların ve iş gücünün gereksiz kullanımını önlemek amacıyla sınava girmeyecek öğrenci sayısı öngörülme çalışılmıştır. Çalışma için her bir öğrenciye ait kişisel olmayan verilerin özellikleri (cinsiyet, genel ağırlıklı not ortalaması, dönem not ortalaması, dersin ara sınav notu, genel sınav notu vb.) belirlenmiştir. Belirlenen bu özelliklere bakılarak derslere ait veri setleri oluşturulmuş bu veri setleri kullanılarak Yapay Sinir Ağları ve Karar Ağacı algoritması ile uygulamalar geliştirilmiştir. Bu iki sınıflandırma yönteminden elde edilen sonuçlar kıyaslandığında Yapay Sinir Ağı ile 88,70 doğruluk oranı elde edilmişken Karar Ağacı algoritmasında 87,96 doğruluk oranında elde edilmiştir.

Anahtar Kelimeler: Yapay Sinir Ağları, Karar Ağacı, Makine Öğrenmesi Algoritmaları, Eğitsel Veri Madenciliği

ABSTRACT

MS THESIS

PREDICTION OF THE NUMBER OF STUDENTS WHO WILL TAKE THE MAKE-UP EXAM USING DECISION TREES AND ARTIFICIAL NEURAL NETWORKS

Miyase Nur ŞENKAYA

**Konya Technical University
Institute of Graduate Studies
Department of Computer Engineering**

Dr. Öğr. Üyesi Semiye DEMİRCAN

2024, 44 Pages

Jury

**Dr. Öğr. Üyesi Semiye Demircan
Diğer Üyenin Unvanı Adı SOYADI
Diğer Üyenin Unvanı Adı SOYADI**

Make-up exam is a type of exam held to give a second chance to the student who has failed a course as a result of the final exam to pass the course. However, for various reasons, a significant number of students do not exercise their right and do not take this exam. In this thesis, it is aimed to prevent unnecessary waste of resources, energy and labor, such as printing excessive exam papers, opening redundant exam halls and assigning invigilators unnecessarily, by predicting the number of students who will not take the exam. In this thesis, an attempt was made to predict the number of students who will not take the exam in order to prevent unnecessary use of resources and workforce. For the study, the characteristics of each student's non-personal data (gender, overall weighted grade point average, semester grade average, course midterm exam grade, final exam grade, etc.) were determined. Considering these determined features, data sets for some courses were created and applications were developed with Artificial Neural Networks and Decision Tree algorithms using these data sets. When the results obtained from these two classification methods are compared; The accuracy rate was obtained as 88.70 with the Artificial Neural Network and 87.96 with the Decision Tree algorithm.

Keywords: Artificial neural networks ,Decision trees, Machine Learning Algorithms, Educational Data Mining;

ÖNSÖZ

Doğal ve sınırlı kaynakların kötü kullanılması ve israf edilmesi çağımızın en büyük sorunlarından biridir. Kağıt israfı, birçok ülkede olduğu gibi Türkiye'de de önemli bir çevresel sorun ve sürdürülebilirlik meselesi haline gelmiştir.

Bu tez çalışmasında bütünleme sınavına girmeye hakkı olmasına girmeyen öğrencilerinin sayılarının tahmin edilerek kağıt, toner ve akademik çalışanların zaman israfı gibi problemlere çözüm bulunması amaçlanmıştır.

Tez çalışmasında verilerin elde edilmesine izin veren kurula teşekkür ederim. Tez döneminde benden yardımlarını esirgemeyen Prof. Dr. Mustafa Servet KIRAN'a teşekkürlerimi sunarım.

Tezin gelişmesine yönlendirici görüş ve önerileriyle yardımcı olan danışmanım Dr. Öğr. Üyesi Semiye Demircan'a

Tezimi çalışmalarım boyunca bana destek olan eşim Alican'a ve oğlum Ömer Giray'a ve aileme ithaf ediyorum.

Miyase Nur ŞENKAYA
KONYA-2024

İÇİNDEKİLER

ÖZET	iv
ABSTRACT.....	iv
ÖNSÖZ	v
İÇİNDEKİLER	iv
SİMGELER VE KISALTMALAR	v
1. GİRİŞ	6
1.1. Tezin Amacı.....	6
1.2. Tezin Önemi	7
1.3. Tezin Organizasyonu	8
2. KAYNAK ARAŞTIRMASI	9
3. MATERYAL VE YÖNTEM.....	12
3.1. Veri Setinin Özellikleri	12
3.2. Veri Setinin Nitelikleri.....	12
3.3. Kullanılan Yöntemler	15
3.3.1 Yapay Sinir Ağı Algoritmalarının Tarihçesi	15
3.3.2 Yapay Sinir Ağı Algoritmalarının Çalışma Prensipleri	16
3.3.3 Geri Yayılım Algoritması	18
3.3.4 Karar Ağacı Algoritmalarının Tarihçesi	19
3.3.5 Karar Ağacı Algoritmalarının Çalışma Prensipleri	20
3.3.6 C4.5 Algoritması.....	23
3.4 Beklenti Maksimizasyonu (EM)	23
3.5 Karmaşıklık Matrisi	24
3.6 Kappa İstatistiği.....	Hata! Yer işareti tanımlanmamış.
4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA.....	27
5. SONUÇLAR.....	35
6.KAYNAKLAR	Hata! Yer işareti tanımlanmamış.

SİMGELER VE KISALTMALAR

Kısaltmalar

EVM: Eğitsel Veri Madenciliği

YSA: Yapay Sinir Ağları

KA: Karar Ağacı

EM: Beklenti Maksimizasyonu

GWO: Gri Kurt Optimizasyon

ELM: Aşırı Öğrenme Makinesi

EVM: Eğitimsel Veri Madenciliği

NB: Naive Bayes

RO: Rastgele Orman

DVM: Destek Vektör Makinesi

GYA: Geri Yayılım Algoritması

1. GİRİŞ

Doğal kaynakların kötü kullanılması ve israf edilmesi, çağımızın en büyük sorunlarından biridir. Ülkemizde de israfına en az dikkat ettiğimiz ürünlerin başında maalesef kağıt gelmektedir. Kağıt israfı da doğanın denge unsuru olan, erozyonu engelleyen, yaşam döngüsünde oksijeni üreten, nemi dengeleyen, hava kirliliğini önleyen ve daha birçok faydası bulunan ağaçların yok edilmesine neden olmaktadır. Bu tezde amaç gereksiz kağıt israfını önlemek için bilgisayar destekli algoritmalarla bütünleme sınavına girme hakkı olmasına rağmen sınava katılmayacak öğrencilerin tahmini sayı değerlerini almaktır.

Araştırmacıların buldukları bilimsel sonuçlara göre elli yaşındaki bir ağacın katma değeri aşağıda verilmiştir; (Özdemir, 2014)

- Hava Kirliliğini Önleme 64,750 Dolar
- Nemi Dengeleme 32,530 Dolar
- Oksijen Üretimi 32,240 Dolar
- Toprağı Geliştirme 32,240 Dolar
- Protein Üretimi 2.250 Dolar
- Doğayı Koruma Erozyonu Önleme 32,240 Dolar
- Toplam 196,250 Dolar

1. 200.000 sayfa A4 kağıdı için 20 tane ağaç kesilmesi gerekmektedir. Türkiye genelinde yıllık kişi başı kağıt tüketimi 70 kilogramdır bu da yıllık 7 adet ağaca denk gelmektedir. Yetişkin bir ağaç günde ortalama 1000 kişinin karbondioksitini oksijene çevirebilir. Bununla birlikte 7 kilogram tozu ve 300 kilogram zehri dönüştürür. 1 hektar alan içinde 10.000 ağaç bulunur bu ağaçlar 30.000 insanın günlük tükettiği oksijeni üretir. Dünyada herkes kağıt ve dolayısıyla ağaç israfının önlemek için kağıt tüketimini yarıya indirse yılda 8 milyon hektar orman korunabilir(Larton 2023).

1.1. Tezin Amacı

Bu tezin amacı üniversitemizde ve Türkiye’de birçok üniversitede yapılan bütünleme sınavına girme hakkı olmasına rağmen sınava katılmayan öğrencilerin sayılarının kullanılan algoritmalarla tahmin edilmesidir. Bu tahmin sayesinde sınava

girecek öğrenci sayısı kadar sınav kâğıdı basılarak kağıt, toner, elektrik gibi diğer israflarında önüne geçilmesi istenmektedir.

Ayrıca bu sınavlarda görevli olan akademik çalışanların sayısı bütünleme sınavına girmeye hakkı olan öğrenci sayılarına bakılarak belirlenmektedir. Bu sayının da normale göre daha fazla olması fazladan görevlendirilen akademisyenler için zaman kaybı olup akademik çalışmalarını kesintiye uğratmaktadır.

1.2. Tezin Önemi

Bütünleme sınavı, genel sınav sonucunda herhangi bir dersten başarısız olan öğrenciye dersten geçebilmesi için ikinci bir şans tanınması amacıyla yapılan ve genel sınav yerine geçen sınav türüdür. Ancak bütünleme sınavı zorunlu değildir ve derslerden başarısız olan bazı öğrenciler çeşitli gerekçelerle kendisine tanınan bu hakkı kullanmamakta ve bütünleme sınavına girmemektedir. Bunun nedeni öğrencinin bütünleme sınavında da başarısız olacağı düşüncesi, dönem tatilini bütünleme sınavlarıyla geçirmek istememesi ya da üniversitelerin mevzuatlarının öğrencilere tanıdığı başka haklar olabilmektedir. Örneğin final sınavı sonucunda harf notu DD veya DC olan bir öğrenci bütünleme sınavına girmese dahi dönem ağırlıklı not ortalaması 2 ve üzeri olması durumunda dersten başarıyla sayılmaktadır. Buna ek olarak üniversitemize dört yıllık örgün eğitimi boyunca toplamda 6 adet dersten DC notu bulunan öğrenciler diğer tüm başarı kıstaslarını sağlamış olmaları durumunda bu derslerden başarılı sayılmaktadırlar ve öğrenciler bazı derslerde bu hakkını kullanarak bütünleme sınavına girmemektedir. Ancak öğrenci bilgi sistemlerinde FF, FD, DD ve DC notunu alan tüm öğrenciler bütünleme sınavına girecek öğrenci olarak görüntülenmekte ve sınav öncesi hazırlıklar bu öğrenci sayılarına göre yapılmakta dolayısıyla soru kâğıtları bu sayıya göre çoğaltılmakta, sınav salonları ve gözetmenler bu sayıya göre ayarlanmaktadır. Bu durumda gereksiz kaynak, iş gücü ve zaman sarfıyatı ortaya çıkmaktadır. Bu gereksiz sarfıyatlarına önlenmesi amacıyla öğrencilerin ders notları, öğrencinin geçmiş dönem başarıları, vize, final, notları, alttan aldığı ders sayısı, dönemi giriş yılı, cinsiyeti ve kaldığı bütünleme sayısı dikkate alarak yapay sınır ağları ve karar ağacı yöntemleriyle bu sınavlara girecek öğrenci sayılarının tespiti üzerine çalışmalar yürütülmüştür. Doğru bir öngörü neticesinde gereksiz kaynak, iş gücü ve zaman sarfıyatının önüne geçilecektir.

1.3. Tezin Organizasyonu

Gerçekleştirilen tez çalışması 5 bölümden oluşmaktadır. İlk bölümde tezin konusu tanıtılmıştır.

İkinci bölümde ise literatürde yapılan bu teze kaynak oluşturabilecek çalışmalar ile ilgili bilgiler verilmiştir.

Üçüncü bölümde tez çalışmasında kullanılan veri setlerinin özellikleri, nitelikleri verilmiş tez çalışmasında kullanılan yöntemler anlatılmıştır.

Dördüncü bölümde ilk olarak tez çalışmasında kullanılan veri setleri üzerinde yapılan ön işlemler anlatılmıştır. İkinci bölümünde ise tez çalışmasında gerçekleştirilen iki uygulama açıklanmıştır ve elde edilen sonuçlar analiz edilmiştir.

Beşinci ve son bölümde tez çalışmasının sonuçlarının yorumları verilmiştir.

2. KAYNAK ARAŞTIRMASI

2016 yılında Eğitimsel Veri Madenciliği (EVM) öğrencilerin performanslarının tahmin edilmesi, öğrenci verilerinin analizi ve görselleştirilmesi gibi çeşitli yaklaşımlarla ilgilenerek öğrencilerin gelecekteki öğrenme sonuçlarını tahmin etmek üzere bir veri modeli geliştirmiştir. Bu modelin sınıflandırma algoritmalarından Karar Ağaçlarını (KA) kıyaslamada kullandıkları J48 algoritmasının sonuca en uygun algoritma olduğunu görmüşlerdir (Sumitha ve Vinothkumar, 2016).

2016 yılında yaptıkları çalışmada EVM ‘nin hedeflerinden biri olan kişisel, sosyo-ekonomik, psikolojik ve diğer çevresel özellikler dikkate alındığında öğrencinin akademik performansının nasıl tahmin edileceğini daha iyi anlamak için KA, Naive Bayes (NB) ve Rastgele Orman (RO)) sınıflandırma algoritmalarını kullanmışlardır. Öğrenci verilerinin tahmin doğruluğunu iyileştirmişler. Ayrıca kurala dayalı tekniklerin (Apriori, Filtered Associator ve Tertius) bir kombinasyonunu kullanarak öğrenci sonuçlarını etkileyen birliktelik kurallarını da belirlemişlerdir (Satyanarayana ve Nuckowski, 2016).

2014 yılında öğrenci veri tabanı kullanarak KA yöntemiyle öğrencinin performansını tahmin etmek için bir çalışma yapmışlardır. Öğrencilerin 6 yıl süresince verilerini kaydederek analiz etmişler. Bu analizlerle seçilen bir ders programındaki öğrenci verilerinden final notunun bir dereceye kadar tahmin etmişler ve öğrencilerin başarısızlık oranını azaltmak ve özel özen gerektiren öğrencileri ayırt etmeye yardımcı olmuşlardır (Ahmed,. ve Elaraby 2014).

2013 yılında yaptıkları çalışmada 2011 yılında mezun olan 127 öğrencinin 4 yıl boyunca almış olduğu 49 derse ait yılsonu notlarını kullanarak mezuniyet notlarını tahmin etmeyi amaçlamışlar. Bulunan sonuçları kıyasladıklarında KA’nın YSA’ya göre daha kötü tahmin sonuçları çıkardığını görmüşlerdir (Şengür ve Tekin, 2013).

KA, RO, Destek Vektör Makinesi (DVM), lojistik regresyon algoritmalarını kullanarak Malezya Politeknik Üniversitesindeki 489 öğrencin daha önceki not verilerinden final notlarını tahmin etmeye çalışmışlardır. Bu çalışmada elde edilen sonuçlara göre en yüksek doğruluk oranını veren algoritma KA algoritması (J48) algoritması olmuştur (Bujang ve ark., 2021).

2019 yılında Open University Learning Analytics Dataset (OULAD) ile 32593 öğrencinin akademik başarısını tahmini için öğrencinin önceki notları, demografik verileri ve öğrencilerin katılımı gibi bilgileriyle farklı regresyon ve sınıflandırma algoritmaları kullanmışlar En yüksek doğruluk oranı, algoritmalar arasından YSA algoritması kullanılarak elde etmişlerdir. (Tomasevic ve ark., 2019).

2019 yılında Akdeniz üniversitesinde okuyan 578 öğrencinin verileriyle mezuniyet notlarını tahmin etmek için bir çalışma yürütmüşler. Çoklu doğrusal regresyon analizi ve YSA kullandıkları çalışmada %0.13 gibi küçük bir oranda olsa YSA'nın daha yüksek sonuç verdiğini gözlemlemişlerdir (Altun ve ark., 2019).

2016 yılında Akademisyenlerin performansını değerlendirmek için bir çalışma yapmış. Marmara Üniversitesinde görev yapan akademisyenlerin performanslarını 26 nitelik üzerinde dört farklı sınıflandırma tekniği (KA algoritmaları, DVM, YSA ve Diskriminant analizi) ile çalıştırmıştır. Bu sınıflandırma tekniklerinden en başarılı olanın C5.0 algoritması olduğu ifade etmişlerdir (M. Ağaoglu, 2016).

Eğitmen performansını tahmin etmek için 10 nitelik üzerinde NB sınıflandırma, KA ve YSA kullanmışlar. Bu çalışmanın sonucu olarak en iyi doğruluk tahmininde bulunan algoritmanın KA olduğu sonucuna ulaşmıştır (Sanjay ve Keshav, 2017).

2017 yılında 1387 öğrencinin verisiyle akademik başarıyı görmek için bir çalışma yapmıştır. Bu çalışmada beş farklı algoritma kullanmış mezuniyet yılına göre en iyi başarıyı J4.8 (KA) ve NB algoritmalarından elde etmiştir(Aydemir, 2017).

2015 yılında yavaş öğrenen öğrencilerin tespiti için 152 öğrencilik anket verisiyle bir çalışma yapmışlar. Sınıflandırma tekniklerinden YSA, NB, SMO, J48(KA) ve REPTree algoritmaları uygulamışlardır.Elde edilen sonuçlara göre en iyi sonucu %75 doğruluk oranıyla YSA ile ulaşmışlardır (Kaura ve ark., 2015).

2017 yılında üniversite öğrencisinin performansları sınıflandırma algoritmaları olan J48 (KA),RO ,NB, NB Multinomial, K-star kullanarak test edilmiş. Performans testi için 16 parametrelilik (cinsiyet, uyruk, doğum yeri, ders notları, derse katılımı) 480 öğrenci verisi kullanıldı. Performans alt, orta ve üst düzey olarak sınıflandırıldı. Algoritmaların başarı oranları %67 ile %76 arasında bir dağılım gösterdi. En iyi sonucu veren algoritmaların RO ve J48 (KA) olduğunu gözlemlediler (Kapur ve ark., 2017).

2015 yılında 2002 ile 2015 yılları arasında öğrenci performans analizlerinde kullanılan değişkenleri ve tahmin yöntemlerini incelemek için IEEE Xplore, Springerlink, ScienceDirect, ACM dijital kütüphane veri tabanlarındaki çalışmalar

incelenmiş. Çalışmaya göre öğrenci performans tahmininde en iyi metotların YSA ile KA olduğu sonucuna varmışlardır (Shahiria ve ark., 2015).

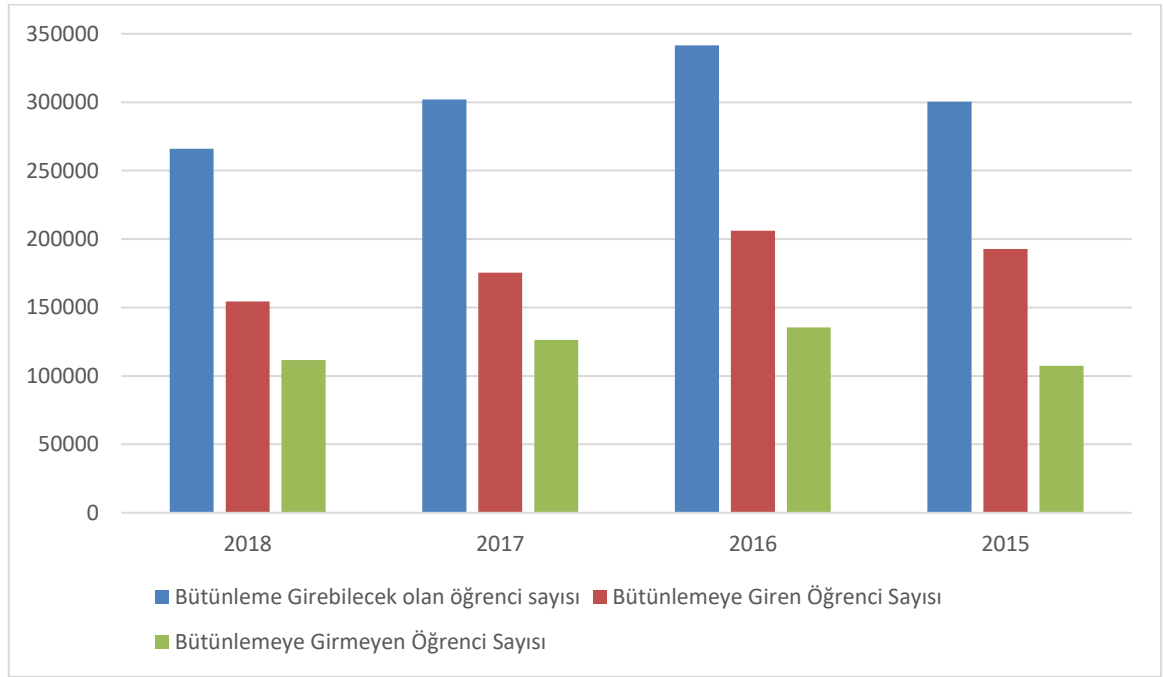
2023 yılında yaptığı parkinson hastalığının tespit edilmesi için makine öğrenmesi tabanlı RO, Lojistik Regresyon, Gaussion NB, KA, DVM, aşırı gradyan arttırma ve YSA'dan oluşan yedi algoritmayı karşılaştırmış. Çalışmasında en yüksek doğruluk değerine YSA ile ulaşmıştır (Yusuf Karagöz, 2023) .

2022 yılında 10 adet ders veri kümesini kullanarak bütünleme sınavına girecek öğrenci sayılarının tahmin etmek için aşırı öğrenme makinesi tabanlı yaklaşımları kullanmışlardır. Kullandığı GYA, ELM ve metasezgisel algoritmaları birbiriyle eğitim doğruluğu, test doğruluğu ve eğitim süresi olarak kıyasladığında GYA eğitim doğruluğu olarak bakıldığında karşılaştırmalı algoritmalarından daha yüksek doğruluk değerine olduğunu, GWO ile eğitilen YSA'nın test doğruluğu açısından karşılaştırmalı algoritmalarından daha yüksek olduğu sonucuna varmıştır. Kullandığı algoritmalarda aldığı sonuçların eğitim sürelerini kıyasladığında ise ELM'nin diğerlerinden daha hızlı olduğunu gözlemlemiştir (Kıran ve ark., 2022).

3. MATERYAL VE YÖNTEM

3.1. Veri Setinin Özellikleri

Tez çalışması için kullanılan veriler 09/06/2022-E301179 sayılı belge ile temin edilmiştir. Bu verilere göre, Selçuk Üniversitesinde 2015-2018 yılları arasında, toplam öğrenci sayılarından bütünlleme sınavı hakkı olan, sınav hakkı olup sınava giren ve sınava girmeyen öğrenci sayıları şekil-3.1 ‘de gösterilmiştir.



Şekil 3.1. Bütünleme Sınavına Hakkı Olan/Giren/Girmeyen Sayıları

Şekil 3.1’de gösterilen sayıların oran karşılıkları hesaplandığında 2018 yılında bütünlleme sınavına girebilecek olduğu halde girmeyen öğrenci sayısı oranı yaklaşık %41.92 olmuştur. Aynı oran 2017 yılında %41.86, 2016 yılında %39.67 ve 2015 yılında %35.78 olmuştur. Bu 4 yılın ortalaması %39.81 olmuştur, yani her 100 öğrenciden 39’u bütünlleme sınavına girebilecek olduğu halde sınava girmemektedir.

3.2. Veri Setinin Nitelikleri

Tezde kullanılan veri setleri Bilgisayar Mühendisliği Bölümünde okutulmuş olan derslerin bilgilerinden seçilmiştir. Bu dersler ; Bilgisayar Mühendisliğine Giriş, Atatürk İlke ve İnkılap Tarihi, Fizik 1, İşletim Sistemleri, Matematik 1, Matematik 2, Türk Dili 1, Türk Dili 2 ve Veri Yapılarıdır.

Bilgisayar Mühendisliğine Giriş dersinin verileri Veri Seti 1 olarak tanımlanmıştır ve 124 veriden oluşmaktadır. Atatürk İlke ve İnkılap Tarihi dersinin verileri Veri Seti 2 olarak tanımlanmıştır ve 71 veriden oluşmaktadır. Fizik 1 dersinin verileri Veri Seti 3 olarak tanımlanmıştır ve 391 veriden oluşmaktadır. İşletim Sistemleri dersinin verileri Veri Seti 4 olarak tanımlanmıştır ve 196 veriden oluşmaktadır. Matematik 1 dersinin verileri Veri Seti 5 olarak tanımlanmıştır ve 391 veriden oluşmaktadır. Matematik 2 dersinin verileri Veri Seti 6 olarak tanımlanmıştır ve 497 veriden oluşmaktadır. Türk Dili 1 dersinin verileri Veri Seti 7 olarak tanımlanmıştır ve 84 veriden oluşmaktadır. Türk Dili 2 dersinin verileri Veri Seti 8 olarak tanımlanmıştır ve 62 veriden oluşmaktadır. Veri Yapıları dersinin verileri Veri Seti 9 olarak tanımlanmıştır ve 292 veriden oluşmaktadır.

Veri setleri bu dersleri almış ve bütünleme sınavına girmesi gereken öğrencilerin Kişisel Verilen Korunması Kanunu (KVKK)'na uygun olan verilerinden oluşmaktadır. Tez çalışmasında yukarıda açıklanan dokuz veri setinin yanı sıra tüm derslerin bir araya getirilmesiyle oluşturulan, tüm dersler veri seti (Veri Seti 10) kullanılmıştır. Çizelge 3.1'de kullanılan veri setlerinin adları, tanımları ve veri sayıları gösterilmiştir. Tüm dersler veri setindeki eksik verileri tamamlamak için Beklenti Maksimizasyonu (Expectation Maximization-EM) algoritması kullanılmıştır.

Çizelge 3.1. Tez çalışmasında kullanılan veri setleri

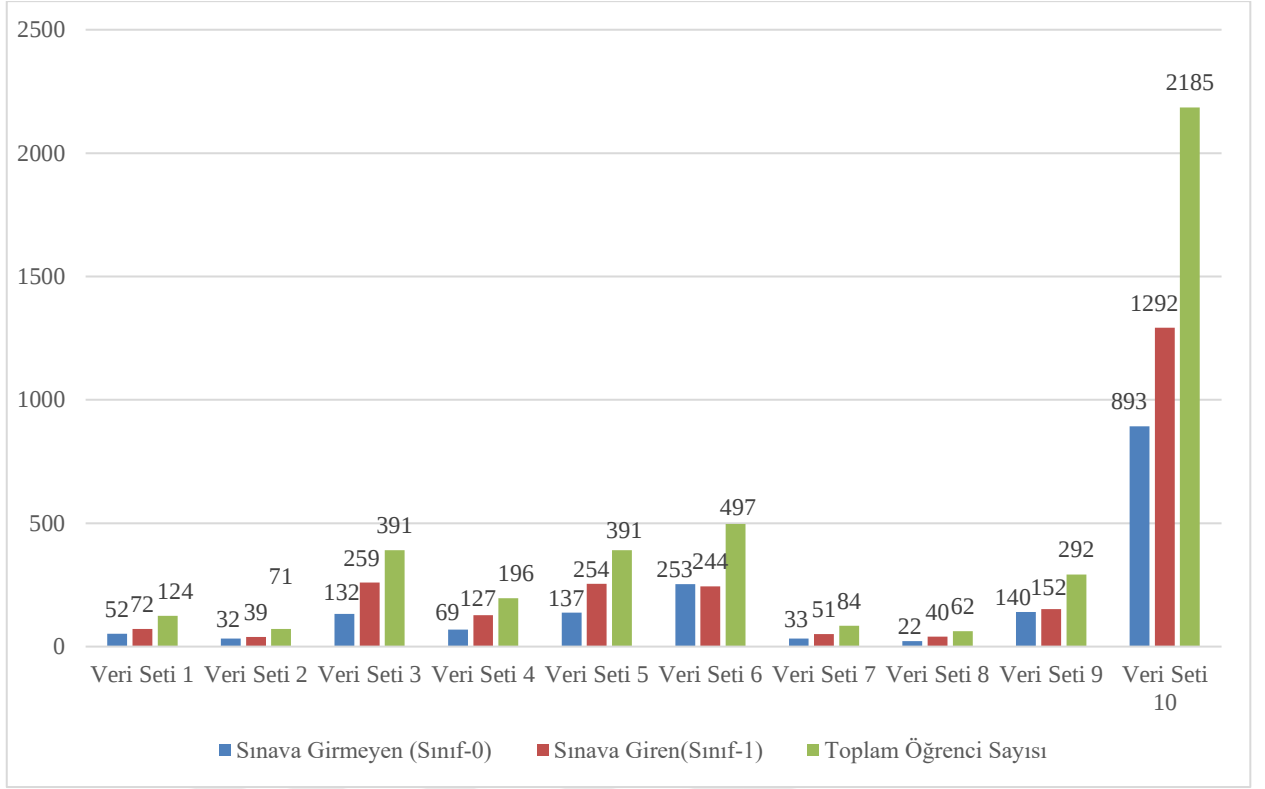
Dersin Adı	Dersin Tanımı	Veri Setindeki Veri Sayısı
Bilgisayar Mühendisliğine Giriş	Veri Seti 1	124
Atatürk İlke ve İnkılap Tarihi	Veri Seti 2	71
Fizik 1	Veri Seti 3	391
İşletim Sistemleri	Veri Seti 4	196
Matematik 1	Veri Seti 5	391
Matematik 2	Veri Seti 6	497
Türk Dili 1	Veri Seti 7	84
Türk Dili 2	Veri Seti 8	62
Veri Yapıları	Veri Seti 9	292
Tüm Dersler	Veri Seti 10	2185

Her bir veri seti 12 nitelikten oluşmaktadır. Bu 12 nitelik sırasıyla;

1. Öğrencinin genel not ortalaması,
2. Öğrencinin aldığı ders sayısı,
3. Öğrencinin ara sınav notu,
4. Öğrencinin genel sınav notu,
5. Öğrencinin ders notunun harf karşılığı,
6. Öğrencinin alttan aldığı derslerin sayısı,
7. Öğrencinin ders dönemi,
8. Öğrencinin giriş yılı,
9. Öğrencinin cinsiyeti,
10. Dersin CC harf notu alt sınırı,
11. Dersin dönemi,
12. Öğrencinin kaldığı bütünleme sayısıdır.

Veri kümelerinde 2 adet sınıflandırma değeri kullanılmıştır. Sınıf değeri 0 olan öğrenciler bütünleme sınavına girebilecek olmasına rağmen sınava girmeyen öğrencilerdir ve bu öğrenciler sınıf-0 olarak gösterilmiştir. Sınıf değeri 1 olan öğrenciler bütünleme sınavına giren öğrencilerdir ve bu öğrenciler sınıf-1 olarak gösterilmiştir.

Şekil 3.2’de veri kümelerindeki sayılar verilmiştir. Buna göre Veri Seti 1’de toplam 124 öğrencinin 72’si sınıf-1’deyken, 52’si sınıf-0’da, Veri Seti 2’de toplam 71 öğrencinin 39’u sınıf-1’deyken, 32’si sınıf-0’da, Veri Seti 3’de toplam 391 öğrencinin 259’u sınıf-1’deyken 132’si 0 sınıfında, Veri Seti 4’de toplam 196 öğrencinin 127’si sınıf-1’deyken 69’u sınıf-0’da, Veri Seti 5’de toplam 391 öğrencinin 254’ü sınıf-1’deyken 137’si sınıf-0’da, Veri Seti 6’da toplam 497 öğrencinin 244’ü sınıf-1’deyken 253’ü sınıf-0’da, Veri Seti 7’de toplam 84 öğrencinin 51’i 1 sınıf-1’deyken 33’ü sınıf-0’da, Veri Seti 8’de toplam 62 öğrencinin 40’ı 1 sınıf-1’deyken 2’si sınıf-0’da, Veri Seti 9’da toplam 292 öğrencinin 152’si sınıf-1’deyken 140’ı sınıf-0’da, Veri Seti 10’da toplam 2185 öğrencinin 1292’si sınıf-1’deyken 893’ü sınıf-0’dadır.



Şekil 3.2 Veri kümelerinin içerdiği veri sayıları

3.3. Kullanılan Yöntemler

3.3.1 Yapay Sinir Ağı Algoritmalarının Tarihçesi

İlk olarak 1943 yılında Warren McCulloch ve Walter Pitts tarafından YSA modeli geliştirilerek YSA'nın temeli atılmışlardır. (McCulloch ve Pitts, 1943) Donald Hebb 1949 yılında yayımlanan "The Organization of Behaviour" adlı kitabında hücresel seviyede biyolojik öğrenmeden bahsetmiştir. Hebb, bu öğrenme sinir ağının bağlantı sayısı ile uyum sağlayabilme ve öğrenmenin ilişkili olduğunu tespit etmiştir.

1950'lerde tek katmanlı ilk YSA modeli Frank Rosenblatt tarafından kullanılmıştır. (Widrow ve Hoff, 1960) 1960'lı yıllarda ise Widrow ve Hoff tek katmanlı nöron modelini kurmuşlar, istenen çıkış ile gerçek çıkış arasındaki farka eşit bir hata terimi kullanarak bağ ağırlıklarının değiştirildiği bir öğrenme kuralını ortaya koymuşlardır.

1980'lerde bilgisayar bilimindeki hızlı ve büyük gelişmeler YSA'nın da dönüm noktası olmuştur. (Kohonen, 1982) 1982'de "Kendi kendine öğrenme nitelik haritaları" adlı çalışmasını yayınlanmıştır. Yine 1982'de "Neural Networks and Physical Systems" adlı çalışma J.J. Hopfield tarafından yayınlanmış ve böylece çağdaş YSA devri

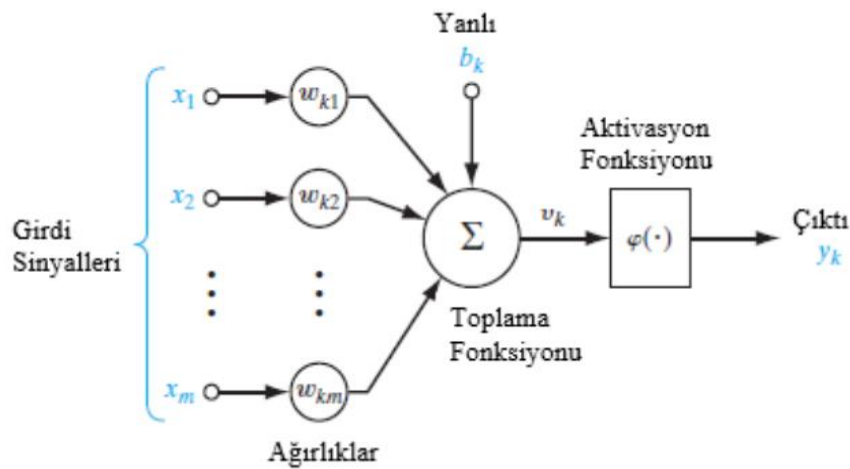
başlamıştır. Hopfield doğrusal olmayan bir enerji fonksiyonu kullanarak nöronların birbirleriyle etkileşimlerine dayanarak bir hesaplama modeli geliştirmiştir. Çalışma sonrasında pratik optimizasyon problemlerinde kullanılabilecek geri beslemeli bir YSA modelini ortaya çıkarmıştır (Hopfield, 1982-a, 1982-b).

1986 yılında “Parallel Distributed Processing” adlı kitaplarında Rumelhart ve arkadaşları ileri beslemeli ağlar için yeni bir öğrenme modeli geliştirmişlerdir. Bu modele “Hatanın Geriye Yayılması Algoritması” adını vermişlerdir(Rumelhart ve McClelland, 1986).

3.3.2 Yapay Sinir Ağı Algoritmalarının Çalışma Prensipleri

Yapay Sinir Ağları, insan beynindeki sinirlerin birbirine bağlı ve uyumlu çalışmasına benzer bir mantıkla çalışan yapay sinir hücrelerinden oluşan yapılardır. Beynin en küçük ögesi olan nöronlar (neuron), birbirine dendrit (dendrite) ve aksonlar (axon) ile bağlanır. Bu bağlantı noktaları sinaps (synapse) olarak adlandırılır. Burada bir nöronun aksonundan aldığı elektriksel uyarı, diğer nöronun dendritine aktarır. Hücre gövdesi denilen kısımda, dendritlerden gelen tüm sinyaller toplanır. Sinyaller hücrenin eşik değerini geçtiği takdirde elektrik akımı oluşturulur.

YSA’da aynı olay örgüsü gerçekleştirilir. Burada dendritin görevini toplama fonksiyonu, hücre gövdesinin görevini aktivasyon fonksiyonu (activation function), sinapsların görevini ağırlıklar (weight) yapmaktadır. Biyolojik hücrede oluşan elektrik akımının yapay hücredeki karşılığı ise çıktıdır (output). Şekil 3.3’de YSA hücresinin nöron modeli yapısı verilmiştir.



Şekil 3.3: Yapay sinir ağları nöron modeli yapısı (Haykin, 2009)

Bir yapay sinir ağı hücresi 5 bölümden oluşur;

1.Girdiler: İnsan sinir hücrelerindeki gibi nöronlara gelen verileri nöron çekirdeğine gönderir.

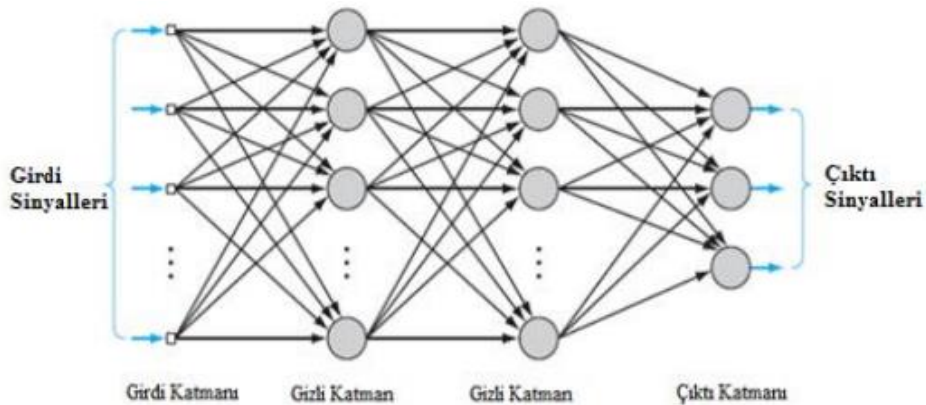
2.Ağırlıklar: Nöronlara gelen veriler nöron çekirdeğine iletilmeden önce sinapsislerin ağırlıklarıyla çarpılıp öyle iletilir.

3.Toplama Fonksiyonu (Birleştirme Fonksiyonu): Hücreye gelen ve ağırlıklarla çarpılan girdilerin toplan ve net girdilerin hesaplanmasını sağlar.

4.Aktivasyon fonksiyonu: Genellikle doğrusal olmayan aktivasyon fonksiyonları girdideki tüm ağırlıklarının alıp bir çıkış değeri üretip bunu bir sonraki katmana ileten bir yapıdır.

5.Çıktılar: Aktivasyon fonksiyonundan elde edilen sonuçlar hücrelerin çıktı değerleridir. Hücrelerde birden çok girdi olmasına rağmen bir tek çıktı olabilir.

YSA'da üç farklı katman türünün üzerine kurulmuş nöronlarla çalışan bir yapıdır. Bu üç katman girdi katmanı, gizli katman ve çıktı katmanıdır. Bilgiler ağı girdi katmanından giriş yapar ara katmanda ağırlık değerleri kullanarak işlenip sonrasında çıktı katmanına gönderilir. Yapay sinir ağı tek katmandan oluşuyorsa tek katmanlı sinir ağı olarak ifade edilirken. Girdi, gizli ve çıktı katmanlarının hepsini bir arada barındıran yapay sinir ağlarına çok katmanlı yapay sinir ağları denir. Çok katmanlı olan sinir ağlarında ağların yerel sonuçlara takılmadan devam edebilmesi için momentum katsayısı kullanılır.



Şekil 3.4. Çok Katmanlı YSA (Haykin, 2009)

Aşağıda verilen denklem şekil 3.4’ de ağ yapısının formüle edilişi gösterilmiştir. Bu denklemde;

X_i : girdiler,

W_i : girdilere ait ağırlıklar,

b_i : sapma terimi,

ϕ : aktivasyon fonksiyonu,

y : çıktı ifade etmektedir.

$$y = \sum_{i=1}^N (X_i W_i + b_i) \quad (3.1)$$

YSA ağ modellerine göre, ileri beslemeli ağlar ve geri beslemeli ağlar olarak iki grupta incelenebilir (Nasuhoğlu, 2019). İleri beslemeli ağlarda hücre katmanları arasında sadece ileri doğru hareket edip bir katmana ait çıkış değeri kendinden sonra gelen katmanın ağırlıkları üzerinden o katmana giriş olarak gönderilir. Geri beslemeli ağlarda ise bir hücreye ait çıkış verisi kendi bulunduğu kendinden önceki veya kendinden sonraki katmanlara iletilebilir.

3.3.3 Geri Yayılım Algoritması

Yapay sinir ağı tasarımına başlanırken ağırlıklara rastgele bazı değerler verilir. Bu verilen rastgele değerlerle alınan sonuçların hata değerini minimize etmek için geri yayılım algoritması(GYA) kullanılır. Bu yayılım iki aşamadan meydana gelir; ilki ağa giriş verilerinin geldiği ileri besleme, ikincisi hatanın geriye doğru yayılımıdır. Geri beslemeli Yapay Sinir Ağları'nda, bilgi çıktı veya ara katmanlardaki nöronlardan alınarak, bu bilgi çıktılarına, girişe veya önceki ara/gizli katmanlardaki nöronlara geri iletilmektedir. Bu şekilde bilgi, ileri ve geri yönde aktarılabilir.(Güvenç, 2023) İleri besleme sonucunda elde edilen çıkış değerleri hata fonksiyonuna girilir, ağırlıkların güncellenmesi için hatalar geriye yayılır.

GYA'nın hatanın geriye nasıl yayıldığını hesaplama mantığında başlangıçta ağa giriş verileri uygulandıktan sonra her katmandaki düğümler üzerinde aktivasyon fonksiyonu kullanılarak çıkışlar hesaplanır. Bu çıkışlar gerçek değerle kıyaslanarak hata hesaplanır. Hata, ağ içinde geriye doğru yayıldıkça, ağın parametreleri güncellenir. İlk olarak, çıkış katmanındaki hata hesaplanır. Daha sonra, her katmandaki nöronların katkısıyla bir önceki katmandaki hata hesaplanır. Hatanın parametrelere göre türevleri alınarak gradyan hesaplanır bu hesaplamayla her parametrenin hataya olan katkısı belirlenir. Hesaplanan gradyanlara göre ağın ağırlıklar ve biasları güncellenir. Girişlerin

ileri yayılımı ve hatanın geri yayılımı adımları birçok iterasyon boyunca tekrarlanır. Her iterasyonda ağırlıklar ve biaslar güncellenirken ağ daha yüksek doğruluk değerinde bir tahmin değeri vermesi sağlanır.

3.3.4 Karar Ağacı Algoritmalarının Tarihçesi

İlk olarak 1963 yılında Morgan ve Sonquist adlı araştırmacılar “Automatic Interaction Detector (AID)” algoritmasını temel alarak KA tabanlı ilk algoritmayı tasarlamışlardır. Bu algoritmanın analiz yöntemi başlangıç seçeneğinin seçilmesi sonrasında tüm şans ve karar değişkenlerinin de seçilmesini varsayar. Ağacın en sonunda seçilen şans ve karar düğümleri ilk adımda analiz edilip bu işlemler başlangıç düğümüne varıncaya kadar devam ettirilir (Gordon ve Pressman, 1983).

1984 yılında Leo Breiman, Jerry Friedman, R. Olshen ve Charles J. Stone araştırmacılar “Classification And Regression Trees” adlı kitapta CART algoritması olarak kitapla aynı isimle bahsedilen yeni bir KA yordamından söz etmişlerdir (Leo Breiman, ve Jerome Friedman, 1984).

1986 yılında KA’larına J.R. Quinlan adlı araştırmacı tarafından ID3 algoritması eklenmiştir. (Binokay, 2018) 1993 yılında ise ID3 geliştirilmiş hali olan C4.5 KA algoritmasını ortaya koymuştur. Bazı KA algoritmaları ve özellikleri Çizelge 3.2’de verilmiştir.

Çizelge 3.2 : Bazı Karar Ağacı Algoritmaları Özellikleri (Bounsaythip,ve ark., (2001))

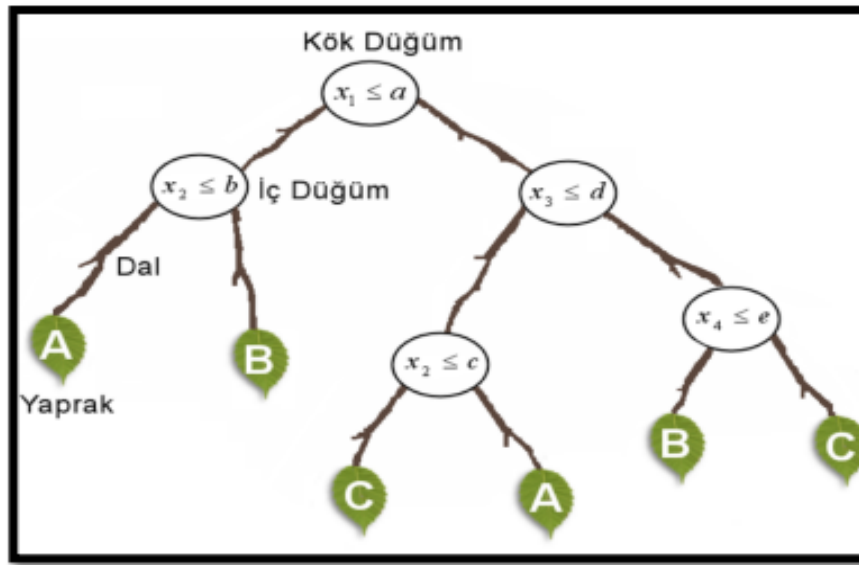
KARAR AĞACI ALGORİTMASI	ÖZELLİKLERİ
CART	Gini tabanlı ikili bölme işlemi mevcuttur. Son veya uç olmayan her bir düğümde iki adet dal bulunmaktadır. Budama işlemi, ağacın karmaşıklık seviyesine bağlıdır ve sınıflandırma ile regresyonu destekleyen bir yapıya sahiptir. Bu yöntem, sürekli hedef değişkenleri ile çalışabilir, ancak verinin hazırlanmasına ihtiyaç duyar.
C4.5 ve C5.0 (ID3 karar ağacı algoritmasının ileri versiyonları)	Çoklu dallardan oluşan bir ağaç oluşturur, her düğümden birden fazla dal çıkar. Dalların sayısı, tahmin edicinin kategori sayısına eşittir. Tek bir sınıflayıcıda birden çok karar ağacını birleştirir. Ayırma işlemi için bilgi kazancını kullanır ve budama işlemi, her yapraktaki hata oranına dayanır.
CHAID (Chi-Squared Automatic Interaction Detector)	Bölme işlemi gerçekleştirilmek için ki-kare testleri kullanır. Dalların sayısı, iki ile tahmin edicinin kategori sayısı arasında değişebilir.
SLIQ (Supervised Learning In Quest)	Hızlı bir şekilde ölçeklenebilen bir sınıflayıcıdır ve hızlı ağaç budama algoritmasını içerir.
SPRINT (Scalable Parallelizable Induction of Decision Trees)	Büyük veri kümeleri için uygun olan bir sınıflayıcıdır. Bölme işlemi, yalnızca tek bir niteliğin değerine dayanır. Tüm bellek sınırlamaları dahilinde nitelik listesi veri yapısını kullanarak işlem gerçekleştirir.

3.3.5 Karar Ağacı Algoritmalarının Çalışma Prensipleri

Karar ağacı, belirlenmiş bir soruna yönelik bu sorunun çözümüne dair olası tüm seçeneklerin eldeki verilere dayanarak kare, daire, oval gibi çeşitli akış şemalarıyla gösteren bir diyagramdır.

Karar ağaçları ağaç formunda gösterilen, karar vermek için kullanılan karar düğümleri sonrasında bu düğümlerin sonucunu ifade eden yaprak düğümleri ve bunları birbirine bağlayan dallardan meydana gelen bir sınıflandırma yöntemidir. Bu düğümlerden iç düğüm olanlar bir özellik veya niteliği temsil ederken dallar yaprak düğümlerin sonucunu veya karar düğümlerinin kurallarını temsil ederler.

Ağaçların gösteriminde en üste bulunan düğüm kök düğümdür. Bu kök düğüm dışındaki düğümler iç düğüm denir. İç düğümler alt düğümlere ayrılabilir. Alt düğümü olmayanlar yaprak düğüm olarak ifade edilir. Şekil 3.5’de üç sınıftan oluşan bir karar ağacı yapısı gösterilmiştir.



Şekil 3.5: Üç sınıftan oluşan karar ağacı yapısı (Kavzaoglu, 2010)

Bir karar ağacının 3 bölümü şöyle ifade edilebilir;

Kök Düğüm: Ağacın başlangıç noktası olarak kabul edilir ve tüm popülasyon kümesini temsil eder. Kök düğüm dengeli ve eşit olarak maksimum bilgi kazancı veya minimum Gini Impurityye göre iki veya daha fazla bölünür.

İç Düğüm: Kökün altında yer alan iç düğümler popülasyondaki veriyi ayıran özelliklerdir. Özelliklerin değerlerine göre ağacı bölme işlemi gerçekleştirilir ve ağacın izleyeceği yol belirlenir. İç düğümler özelliklerine göre çok fazla alt düğüme bölünebilir.

Yaprak Düğüm: Ağaçta popülasyondaki veriye göre daha fazla bölünme gerçekleştiremeyen düğümlere yaprak düğüm denir. Yaprak düğümler ağacın sınıflandırması bittiğinde oluşur ve son tahminleri içerirler.

KA algoritması çalışırken ilk önce popülasyondaki veri kümesinin kategorisini bulmak için ağacın temelini yani kök düğümü bulunur. Kök düğümünün özellikleri ile

veri kümesinin özellikleri karşılaştırılır bunun sonucuna göre bir dal oluşturulup atama yapılır. Yaprğa ulaşınca kadar bu karşılaştırma ve dal atama işlemi devam edilir. Bu kök düğüm, iç, alt düğümler ve yapraklar bulunurken bazı değerler kullanılır.

Gini (Alt kümenin Sıfık Değeri): Gini hesaplamasına göre 0 ile 1 değerleri arasında değışen sonuçlar çıkaran ve sonuç değeri 0'a olan yakınlığına bakarak ayırım yapılan bir hesaplama yöntemidir. Ağacı oluştururken gini değeri sağ ve sol değerler olarak ayrı ayrı hesaplanır. Denklem 3.2' de gini hesaplaması gösterilmiştir. Denklemdaki formülde P_x , değeri x sınıfının olabilme olasılığıdır. 0' a yakın olan taraf seçilir.

$$Gini = 1 - \sum_x (P_x^2) \quad (3.2)$$

Entropi: Rastgeleliği, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir. Entropy de Gini hesaplaması gibi bilgi kazancını ölçmek için kullanılan bir formüldür. Bu formül Denklem 3.3'de aşağıda gösterilmiştir. İki sınıflı veri setleri için gruplarının tamamı dengeli bir dağılım göstermişse entropi 0 olur. Değerlerin birbirine eşit olması durumunda entropi 1 olur. Değerler 1'e yaklaştıkça sonuç çıkarmak zorlaşır.

$$Entropi(T) = - \sum P_i \log_2(P_i) \quad (3.3)$$

Entropiyle gini arasında kayda değeri bir farklılık yoktur. Gini frenkası sık tekrarlanana göre ayırtırmaya gitme eğilimindeyken, entropi dengeli bir ağaç oluşturmaya çalışır. Bu tez çalışmasında entropy hesaplaması kullanıldı.

Information Gain (Bilgi Kazanımı): Bir veri setinde çıktı değışkenlerine göre yapılması gereken düğüm bölümlenmesinin nasıl olması gerektiğini bulmak için kullanılır. Veri bölümlendiğinde hedef değışkendeki belirsizliğin ne kadar değıştiğinin bir ölçüsüdür. Bu hesaplama aşağıda denklem 3.4'te gösterilmiştir.

$$BilgiKazancı(S,A) = Entropi(S) - \sum P_v Entropi(P_v) \quad (3.4)$$

Budama: Eğitim verilerinin fazla takılması önlemek için az bilgi veren karar düğümünü ve buna bağlı olan düğümleri kaldırma işlemidir.

3.3.6 C4.5 Algoritması

Veri setleri üzerinde KA sınıflandırma algoritmalarından biri olan C4.5 sınıflandırma algoritması kullanılmıştır. C4.5 algoritması, sadece kategorik veriler üzerinde çalışan ID3 algoritmasının nümerik özellik içeren verilerinde çalıştırılabilecek halde geliştirilmiştir. Her iterasyonun başlangıcında, veri örneklerine ait sınıf bilgilerini içeren vektörün entropisi belirlenir. Daha sonra, özellik vektörlerinin sınıfa bağımlı entropileri hesaplanarak ilk adımda elde edilen entropiden çıkartılır. Bu yöntemle elde edilen değer, ilgili özellik vektörü için kazanç değerini temsil eder. En yüksek kazançla sahip olan özellik vektörü, ağacın o iterasyonda belirlenen dallanmasını gerçekleştirir.

C4.5 algoritmasının diğer KA algoritmalarından farklı olan nitelikleri; Kayıp veri özellikleriyle baş etmesi, bölünme-dallanma bilgisi ve sayısal özellik içeren veri setlerini de kullanabilmesi gösterilebilir.

Kategorik bir özelliğin çok sayıda farklı değer çeşitliliği içermesi durumunda, bu özellikten elde edilen bilgi kazancı gereksiz yüksek olabilir ve bu durum ağacın doğruluğunu olumsuz yönde etkileyebilir. Bu olumsuzluğu gidermek için değer çeşitliliği fazla olan özelliklerin bilgi kazancını azaltarak algoritmanın gereksiz bazı çıkarımlar yapmasını bölünme bilgisi kavramıyla önüne geçilir.

3.4 Beklenti Maksimizasyonu (EM)

EM algoritması, maksimum olabilirlik tahmininde oldukça sık kullanılan bir yöntemdir (Dempster, 1977). Bu algoritma matematiksel işlemler, tıp alanındaki çalışmalar, sinyal işleme ve yapay zeka ile öğrenme birçok alanda, sayısal optimizasyon yöntemlerinin yerine tercih edilir (Özkan, 2019).

EM algoritması, tam olmayan veri problemlerini çözmek için maksimum olasılık tahminlerini yapan tekrarlı bir algoritmadır. EM Algoritmasının her tekrarı iki adımda gerçekleşir. Bu adımlar, bekleneni bulma (E-Adımı) ve maksimizasyon (M-Adımı) olarak adlandırılır (Baygöl, 2007). E-adımında gözlenen verilerin parametrelerine ait kestirimler kullanılarak bilinmeyen (kayıp) veri ile ilgili en iyi olasılıklar tahmin edilirken,

M-Adımında ise tahmin edilen kayıp veri yerine konulup bütün veri üzerinden maksimum olabilirlik hesaplanarak parametrelerin yeni kestirimleri elde edilir (Bruzzone, 2002).

EM Algoritması Adımları:

Parametre Başlangıç Tahmini: Başlangıçta parametrelere rastgele tahmini değerler atanır.

Beklenti Adımı: Rastgele atanan başlangıç parametreleri üzerinde olasılık dağılımı hesabı yapılır.

Maksimizasyon Adımı: Beklenti adımında alınan olasılık değerlerine bakılarak modelin parametrelerini en üst düzeye çıkarmak amacıyla bir güncelleme uygulanır. Sonrasında eksik veri olması durumunda tekrar değişkenlere bağlı olasılıkların dağılımı tahmin edilip, bu değişkenlerin olası değerlerine dair tahminler iyileştirilip modelin uygunluğu artırılır.

Parametre Güncelleme: Maksimizasyon adımında alınan parametreler kullanılarak modele ait parametreler güncelleştirilir.

3.5 Karmaşıklık Matrisi

Sınıflandırma algoritmalarında alınan sonuçların karşılaştırılması için nesnel değerlendirme yöntemlerine ihtiyaç duyulmaktadır. Sınıflandırmanın ne kadar başarılı olduğunu anlamak için karmaşıklık matrisi kullanıldı (Cihan ve ark., 2020).

Karmaşıklık matrisinde bulunan sütunlar verilerin gerçek değerlerine satırları ise sınıflandırma da alınan sonuçlara karşılık gelmektedir. Çizelge 3.3 karmaşıklık matrisinin ifade etmektedir.

Çizelge 3.3. Tez çalışmasında kullanılan veri setleri

Tahmin Edilen Değerler		Gerçek değerler	
		Pozitif	Negatif
P		Doğru Pozitif (GP)	Yanlış Pozitif (YP)
	N	Yanlış Negatif (YN)	Doğru Negatif (GN)

Çizelge 3.3'deki Doğru Pozitif (GP), algoritma verideki bilgiyi doğru tahmin ederse buna gerçek pozitif denir. Doğru Negatif (GN), algoritma kontrol grubuna ait bir veriyi algoritma doğru tahmin ederse buna gerçek negatif denir. Yanlış Pozitif (YP) veri

setindeki veri setindeki o sınıfa ait olmayan bir veriyi algoritma o sınıfta diye tahmin ederse buna yanlış pozitif denir. Yanlış Negatif (YN), verilerdeki o sınıfa ait bir verinin algoritma sınıflandırırken bu sınıfa ait değildir diye tahmin ederse bu yanlış negatif olarak değerlendirilir. Doğruluk değeri, modelde doğru tahmin ettiğimiz alanların toplam veri kümesine oranı ile hesaplanmaktadır. Eşitlik denklem 3.5’de doğruluk denklemi gösterilmiştir (Powers W. ve Ailab A., 2008).

$$Doğruluk = \frac{GP+GN}{GP+YP+GN+YN} \quad (3.5)$$

Kesinlik, pozitif olarak tahmin edilen değerlerin gerçekten kaç adedinin pozitif olduğunu ifade eder. Eşitlik denklem 3.6’da kesinlik denklemi gösterilmiştir (Gülcan, 2020).

$$Kesinlik = \frac{GP}{GP+YP} \quad (3.6)$$

Duyarlılık ise pozitif olarak tahmin ettiklerimizin, gerçek değeri pozitif olanlara oranını gösteren bir metriktir (Yargı V., 2021) Eşitlik denklem 3.7’de duyarlılık denklemi verilmiştir.

$$Duyarlılık = \frac{GP}{GP+YN} \quad (3.7)$$

F1_skor, Kesinlik ve Duyarlılık değerlerinin harmonik ortalamasını göstermektedir. Harmonik ortalama kullanılmasının sebebi çok uç durumları göz önünde bulundurma zorunluluğudur. Basit bir aritmetik ortalama bizi yanıltabilir (Cohen, 1960) Eşitlik denklem 3.8’de verilmiştir.

$$F1 - Skor = 2 * \frac{Keskinlik * Duyarlılık}{Keskinlik + Duyarlılık} \quad (3.8)$$

3.6 Kappa İstatistiği

Kappa istatistiği 1960’ da Cohen tarafından sınıflama seviyesinde puanlama yapan iki değerlendirici arasındaki uyumu ölçmek amacıyla geliştirilmiştir (Cohen, 1960). Cohen tarafından geliştirilen bu istatistik sadece iki puanlayıcıyla sınırlıyken 1971

yılında Fleiss ikiden daha fazla puanlayıcıyla çalışabilecek şekilde geliştirmiştir (Fleiss, 1971).

Kappa katsayısı istatistiksel olarak 0 ile 1 arasın da değışen değeri alır. Bu değlendirmede katsayı sıfıra yaklaştıkça zayıf uyum bire yaklaştıkça tam uyum olarak değlendirilir. Kappa değeri 0.4'ün üzerinde olanlar kabul edilebilir değlerdir (Landis ve Koch, 1977).

Kappa katsayısı hesaplanırken $Pr(a)$ ve $Pr(e)$ olmak üzere iki farklı olasılık hesaplanır. Eşitlik denklem 3.9'da gösterilen $Pr(a)$, iki değlendirici arasındaki gözlemlenen uyumların toplam oranını temsil ederken, $Pr(e)$ ise bu uyumun şansa bağı ortaya çıkma olasılığını ifade eder.

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (3.9)$$

Uygulamada kullanılan her bir veri kümesinin iki farklı sınıflandırıcı YSA ve KA algoritmalarıyla alınan kappa istatistiğı oranları çizelge 3.4'de verilmiştir.

Çizelge 3.4 Veri Setleri Üzerindeki Algoritmaların Kappa İstatistiğı Oranları

Dersin Tanımı	Karar Ağacı Kappa İstatistiğı %	YSA Kappa İstatistiğı %
Veri Seti 1	0,51	0,54
Veri Seti 2	0,60	0,53
Veri Seti 3	0,68	0,51
Veri Seti 4	0,74	0,78
Veri Seti 5	0,69	0,76
Veri Seti 6	0,68	0,48
Veri Seti 7	0,66	0,72
Veri Seti 8	0,83	0,81
Veri Seti 9	0,60	0,64
Veri Seti 10	0,52	0,58

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

Bu tez çalışmasında YSA ve KA algoritmaları kullanılarak bütünlemeye girecek öğrenci sayılarının tahmini yapılmıştır. Bu sayılar tahmin edilirken Selçuk Üniversitesi Bilgisayar Mühendisliği Bölümü'nde 2015-2018 yılları arasında verilen dokuz derse ait verilerden oluşan veri setleri kullanılmıştır. Tez çalışmasında dokuz derse ait veri setinin yanı sıra tüm derslerin bir araya getirilmesiyle oluşturulan, tüm dersler veri seti kullanılmıştır. Bu şekilde on farklı veri seti oluşturulmuştur.

Bu veri setlerindeki eksik olan verilerin doldurulması için EM algoritması kullanılmıştır. Eksik verilerin doldurulması işleminden sonra on veri setinin tamamında normalizasyon işlemi uygulanmıştır. Normalizasyon işlemi değişkenlerin 0 ve 1 arasındaki değerlere atanmasıdır.

Çizelge 4.1'de Veri Seti 10'da bulunan cinsiyet özelliğinin sayı dağılımı, cinsiyet niteliğine göre sınava girme sayısı ve yüzdesi gösterilmiştir.

Çizelge 4.1 Cinsiyetlere göre popülasyon ve sınava girme yüzdeleri

Cinsiyet	Kadın	Erkek
Kişi Sayısı	663	1523
Popülasyon Dağılım Yüzdesi	30,32	69,68
Sınava Giren Sayısı	444	848
Sınava Giren Yüzdesi	66,96	55,67

Çizelge 4.2'de Veri Seti 10'da bulunan bütünlemeye kaldığı ders sayısı 5'ten az olanlar ve 5 ve 5'ten fazla sayıda dersten bütünlemeye kalmış olanlar olarak gruplanmıştır. Bu yapılan gruplamaya göre sınava giren yüzdesi bulunup tabloda gösterilmiştir.

Çizelge 4.2 Kalınan ders sayısına göre sınava girme yüzdeleri

Kaldığı Bütünleme Sayısı	5 'ten Az Olan	5 ve 5'ten Fazla Olan
Kaldığı Ders Sayısı	1438	748
Sınava Giren	902	390
Sınava Giren Yüzdesi	62,72	52,13

Uygulamaya başlanırken ilk olarak eksik verisi doldurulmuş ve nominalleştirilmiş olan veri seti on kullanılarak kalan dokuz ders veri setinde nasıl bir YSA modeli kullanılacağına belirlenmesi için öğrenme hızı (Learning Rate), iterasyon sayısı, çapraz doğrulama (cross validation) ve momentum katsayısı gibi YSA parametreleri değiştirilerek en iyi sonucu alacak bir YSA modeli kurgulanmıştır. Çizelge 4.3’de alınan sonuçlar gösterilmiştir.

Çizelge 4.3 YSA’nın farklı parametrelerle çalıştırılması

Çalıştırma Numarası	Çapraz Doğrulama	Öğrenme Hızı	Momentum	İterasyon Sayısı	Sonuç (%)
1	5	0.6	0.3	3000	48.1922
2	5	0.6	0.3	500	51.8993
3	5	0.3	0.2	3000	55.3776
4	5	0.3	0.2	500	48.2838
5	10	0.3	0.2	500	58.9931
6	10	0.6	0.3	500	50.0686
7	10	0.3	0.2	3000	63.3867
8	10	0.6	0.3	3000	49.9771

Çizelge 4.3’de görüldüğü gibi öğrenme hızı, çapraz doğrulama, momentum ve iterasyon sayısı değiştirilerek alınan sonuçlarda en iyi sonuç Çizelge 4.3’de gösterilen 7.çalıştırma satırında elde edilmiştir. Geriye kalan dokuz veri setinde de bu parametreler kullanılmıştır.

On ders veri setine ait nominalleştirilmiş veriler KA ve YSA algoritmalarıyla çalıştırılmış çıkan sonuçlara ait karmaşıklık matrisleri Bölüm 4.1’de gösterilmektedir.

4.1.Karar Ağacı ve Yapay Sinir Ağları Karmaşıklık Matrisleri

On derse ait veri setleri KA ve YSA ile çalıştırılmasıyla elde edilen sınıflandırmaların doğruluğunun hesaplamak için en basit ve metriklerden olan karmaşıklık matrisi kullanılmıştır. Bu karmaşıklık matrisleri Tablo 4.4 ile Tablo 4.13'te arasındaki tablolarda gösterilmiştir.

	(Sınıf-0 Sınava girmeyen)	(Sınıf-1 Sınava giren)
(Sınıf-0 Sınava girmeyen)	Doğru Pozitif (GP)	Yanlış Pozitif (YP)
(Sınıf-1 Sınava giren)	Yanlış Negatif (YN)	Doğru Negatif (GN)

KararAğacı

Yapay Sinir Ağları

Tablo 4.4 Veri Seti 1 Karmaşıklık Matrisleri

Tahmin Sınıfı(a)				Tahmin Sınıfı(b)			
Gerçek Sınıf	Sınıf-0		Sınıf-1	Gerçek Sınıf	Sınıf-0		Sınıf-1
	Sınıf-0	5	47		Sınıf-0	33	19
	Sınıf-1	4	68		Sınıf-1	22	50

Tablo 4.4'de gösterilen karmaşıklık matrisleri, Veri Seti 1'in KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 58,87 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 66,93 elde edilmiştir.

Tablo 4.5 Veri Seti 2 Karmaşıklık Matrisleri

Tahmin Sınıfı(a)				Tahmin Sınıfı(b)			
Gerçek Sınıf	Sınıf-0		Sınıf-1	Gerçek Sınıf	Sınıf-0		Sınıf-1
	Sınıf-0	16	16		Sınıf-0	21	11
	Sınıf-1	8	31		Sınıf-1	11	28

Tablo 4.5’de gösterilen karmaşıklık matrisleri, Veri Seti 2’nin KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 66,19 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 69,14 elde edilmiştir.

Tablo 4.6 Veri Seti 3 Karmaşıklık Matrisleri

<i>Tahmin Sınıfı(a)</i>				<i>Tahmin Sınıfı(b)</i>			
<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1	<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1
	Sınıf-0	60	73		Sınıf-0	57	96
	Sınıf-1	34	224		Sınıf-1	37	201

Tablo 4.6’de gösterilen karmaşıklık matrisleri, Veri Seti 3’ün KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 72,63 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 65,98 elde edilmiştir.

Tablo 4.7 Veri Seti 4 Karmaşıklık Matrisleri

<i>Tahmin Sınıfı(a)</i>				<i>Tahmin Sınıfı(b)</i>			
<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1	<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1
	Sınıf-0	56	13		Sınıf-0	57	12
	Sınıf-1	23	104		Sınıf-1	18	109

Tablo 4.7’de gösterilen karmaşıklık matrisleri, Veri Seti 4’ün KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 81,63 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 84,69 elde edilmiştir.

Tablo 4.8 Veri Seti 5 Karmaşıklık Matrisleri

<i>Tahmin Sınıfı(a)</i>				<i>Tahmin Sınıfı(b)</i>			
<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1	<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1
	Sınıf-0	102	56		Sınıf-0	109	49
	Sınıf-1	28	282		Sınıf-1	47	263

Tablo 4.8’de gösterilen karmaşıklık matrisleri, Veri Seti 5’ün KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 79,48 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 82,51 elde edilmiştir.

Tablo 4.9 Veri Seti 6 Karmaşıklık Matrisleri

<i>Tahmin Sınıfı(a)</i>				<i>Tahmin Sınıfı(b)</i>			
<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1	<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1
	Sınıf-0	211	43		Sınıf-0	217	37
	Sınıf-1	65	178		Sınıf-1	172	71

Tablo 4.9’de gösterilen karmaşıklık matrisleri, Veri Seti 6’ün KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 78,48 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 57,94 elde edilmiştir.

Tablo 4.10 Veri Seti 7 Karmaşıklık Matrisleri

<i>Tahmin Sınıfı(a)</i>				<i>Tahmin Sınıfı(b)</i>			
<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1	<i>Gerçek Sınıf</i>	Sınıf-0		Sınıf-1
	Sınıf-0	19	14		Sınıf-0	26	7
	Sınıf-1	10	41		Sınıf-1	8	43

Tablo 4.10’de gösterilen karmaşıklık matrisleri, Veri Seti 7’ün KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 71,42 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 82,14 elde edilmiştir.

Tablo 4.11 Veri Seti 8 Karmaşıklık Matrisleri

		<i>Tahmin Sınıfı(a)</i>				<i>Tahmin Sınıfı(b)</i>	
<i>Gerçek Sınıf</i>		Sınıf-0	Sınıf-1	<i>Gerçek Sınıf</i>		Sınıf-0	Sınıf-1
	Sınıf-0	19	3		Sınıf-0	17	5
	Sınıf-1	4	36		Sınıf-1	3	37

Tablo 4.11’de gösterilen karmaşıklık matrisleri, Veri Seti 8’in KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 88,70 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 87,96 elde edilmiştir.

Tablo 4.12 Veri Seti 9 Karmaşıklık Matrisleri

		<i>Tahmin Sınıfı(a)</i>				<i>Tahmin Sınıfı(b)</i>	
<i>Gerçek Sınıf</i>		Sınıf-0	Sınıf-1	<i>Gerçek Sınıf</i>		Sınıf-0	Sınıf-1
	Sınıf-0	96	44		Sınıf-0	105	35
	Sınıf-1	57	95		Sınıf-1	34	118

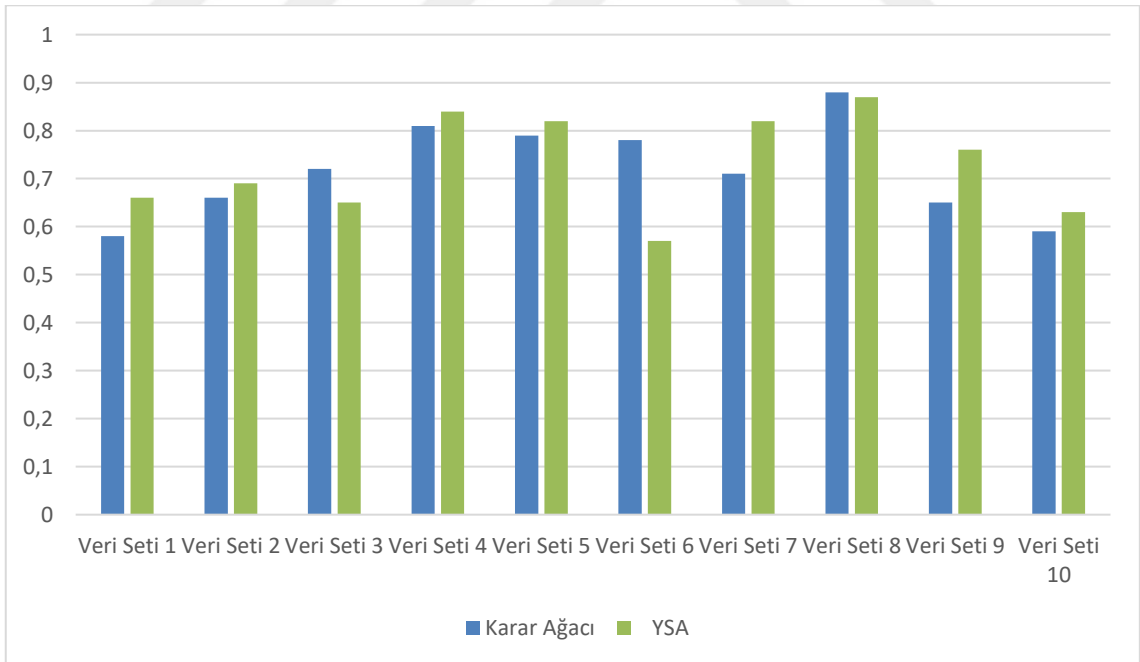
Tablo 4.12’de gösterilen karmaşıklık matrisleri, Veri Seti 9’un KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 65,41 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 76,36 elde edilmiştir.

Tablo 4.13 Veri Seti 10 Karmaşıklık Matrisleri

		<i>Tahmin Sınıfı(a)</i>				<i>Tahmin Sınıfı(b)</i>	
<i>Gerçek Sınıf</i>		Sınıf-0	Sınıf-1	<i>Gerçek Sınıf</i>		Sınıf-0	Sınıf-1
	Sınıf-0	431	462		Sınıf-0	401	492
	Sınıf-1	416	876		Sınıf-1	308	984

Tablo 4.13’de gösterilen karmaşıklık matrisleri, Veri Seti 10’un KA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı(a) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri olarak 59,81 elde edilmiştir. Aynı veri seti YSA ile çalıştırılmasından elde edilen sonuçlar tahmin sınıfı (b) karmaşıklık matrisi ile gösterilmiş olup doğruluk değeri 63,38 elde edilmiştir.

Şekil 4.1’de Gerçekleştirilen çalışmalardan elde edilen duyarlılık değerleri toplu bir şekilde gösterilmiştir.



Şekil 4.1. Sınıf-0 ve Sınıf-1 için Karmaşıklık Matrisi Doğruluk Değerleri

4.2.Karar Ağacı ve Yapay Sinir Ağları ile Edilen Sonuçlar

Bu çalışmada, on veri seti için EM algoritması kullanılarak eksik verileri doldurulmuştur Eksik verileri giderilmiş veri setleri üzerinde sınıflandırma sonuçlarını iyileştirmek için normalizasyon işlemi yapılmıştır. On veri seti içinde en fazla veriye sahip olan Veri Seti 10 üzerinde YSA'ya ait çeşitli parametreler değiştirilerek sınıflandırma başarısı en yüksek olan hiper parametre değerleri kullanılmıştır. Tüm veri setleri YSA ile çalıştırılırken on çapraz doğrulama kullanılmıştır.Tüm veri setleri YSA ve KA algoritmasında çalıştırılarak Çizelge 4.4'te sonuçlar elde edilmiştir.

Çizelge 4.4 Veri Setleri Üzerindeki Algoritmaların Sınıflandırma Doğruluk Oranları

Dersin Tanımı	Karar Ağacı %	YSA %
Veri Seti 1	58,87	66,93
Veri Seti 2	66,19	69,14
Veri Seti 3	72,63	65,98
Veri Seti 4	81,63	84,69
Veri Seti 5	79,48	82,51
Veri Seti 6	78,26	57,94
Veri Seti 7	71,42	82,14
Veri Seti 8	88,70	87,96
Veri Seti 9	65,41	76,36
Veri Seti 10	59,81	63,38

5. SONUÇLAR

Bu çalışma kapsamında Selçuk Üniversitesi'nde 2015-2018 yılları arasında okutulan dokuz derse ait öğrenci verilerini barındıran veri setleri kullanılmıştır. Bu dokuz veri seti birleştirilerek onuncu veri seti oluşturulmuştur. Bu veri setleri iki farklı algoritma ile çalıştırılmış elde edilen sonuçlarla bütünleme sınavına giren öğrenci sayısının tahmini amaçlanmıştır.

Tez çalışmasında ön çalışma olarak veri setlerindeki eksik olan veriler EM algoritmasıyla doldurulmuş. Tüm veri setlerinde sınıflandırma sonuçlarını iyileştirmek için normalizasyon işlemi yapılmıştır.

Tez çalışmasının birinci uygulaması için veri setleri üzerinde çalıştırılacak YSA algoritmasında daha iyi sonuçlar alabilmek için veri seti on üzerinde YSA algoritmasının öğrenme hızı iterasyon sayısı, çapraz doğrulama ve momentum katsayısı gibi çeşitli parametreleri değiştirilerek testler yapılmış. Testlerin sonucunda en yüksek doğruluğu veren parametreler kullanarak geriye kalan dokuz veri seti üzerinde çalışmalar yapılmıştır. Tüm veri setlerinde on çapraz doğrulama yapılmıştır.

Veri seti on, diğer dokuz veri seti birleştirilerek oluşturulmuştur. Yapılan çalışmalarda diğer dokuz veri setinde alınan sonuçların onuncu veri setine göre daha iyi olmasının nedeni birden fazla dersten kalan öğrencilerin bazı dersleri seçerek bütünleme sınavına girmiş olması olarak değerlendirilmiştir. Bu veri setinde dersleri değiştirerek daha yüksek bir sonuç alınabilir. Ayrıca kullanılan veri setleri Bilgisayar Mühendisliği ders havuzunda genellikle ilk iki yılda okutulan derslerdir. Bu ders veri setine gelecekteki uygulamalarda diğer yılların verileri veya başka bölümlere okutulan derslerde eklenerek daha yüksek bir sonuç elde edilebilir.

Çalışmada ikinci uygulama için de veri setleri üzerinde çalıştırılacak KA algoritmasının parametrelerine karar verebilmek farklı çapraz doğrulama parametreleri ile veri seti on üzerinde testler yapılmış. Testlerin sonucunda en yüksek doğruluğu veren parametreyle geriye kalan dokuz veri seti çalıştırılmıştır.

Veri setlerinin KA ve YSA algoritmalarıyla çalıştırılmasıyla alınan sonuçların sınıflandırmaların doğruluğunu hesaplamak için karmaşıklık matrisi kullanılmıştır. Karmaşıklık matrisi sonrası elde edilen doğruluk değerleri tablolarda gösterilmiştir.

KA ve YSA algoritmalarının veri setlerinin üzerinde çalıştırılması ile elde edilen sonuçlar Çizelge 4.4'te gösterilmiş. Alınan sonuçlar kıyaslandığında YSA algoritması ile

büyük oranda KA algoritmasından daha üstün bir sınıflandırma başarısı elde edilebileceği sonucuna varılmıştır.



6.KAYNAKLAR

1. Özdemir, E. (2014). Elektronik belge yönetim sisteminin üniversiteler açısından tasarruf boyutu. XIX. Türkiye'de İnternet Konferansı. Sözlü Bildiri. 27-29 Kasım, 2014. Yaşar Üniversitesi, İzmir.
2. <https://larton.com.tr/rakamlarla-karton-kagit-ve-agac-tuketimi/> Larton, erişim tarihi: 6 Kasım 2023
3. Sumitha, R. and E.S. Vinothkumar, Prediction of Students Outcome Using Data Mining Techniques. International Journal of Scientific Engineering and Applied Science (IJSEAS), 2016. 2(6): p. 8.
4. Satyanarayana A, Nuckowski M. "Data mining using ensemble classifiers for improved prediction of student academic performance". Spring 2016 Mid-Atlantic ASEE Conference, GWU, 8-9 April 2016.
5. Ahmed, A.B.E.D. and Elaraby, I.S., 2014. Data Mining: A prediction for Student's Performance Using Classification Method. World Journal of Computer Application and Technology, 2(2), pp.43-47.
6. Şengür, D. ve Tekin, A. (2014). Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini. Bilişim Teknolojileri Dergisi, 6(3), 7-16.
7. Bujang, S. D. A., Selamat, A., & Krejcar, O. (2021, February). A Predictive Analytics Model For Students Grade Prediction By Supervised Machine Learning. In Iop Conference Series: Materials Science And Engineering (Vol. 1051, No. 1, P. 012005). Iop Publishing.
8. Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An Overview And Comparison Of Supervised Data Mining Techniques For Student Exam Performance Prediction. Computers & Education, 143, 103676
9. Altun, M. , Kayıkçı, K. & Irmak, S. (2019). Sınıf Öğretmenliği Öğrencilerinin Mezuniyet Notlarının Regresyon Analizi ve Yapay Sinir Ağları Yöntemleriyle Tahmini / Estimation of Graduation Grades of Primary Education Students by Using Regression Analysis and Artificial Neural Networks . e-Uluslararası Eğitim Araştırmaları Dergisi , 10 (3) , 29-43 . DOI: 10.19160/ijer.624839.
10. M. Ağaoglu, "Predicting instructor performance using data mining techniques in higher education," IEEE Access, c. 4, ss. 2379–2387, 2016.
11. Sanjay, S. S., & Keshav, B. B. (2017). Teacher's performance analyzer. The International Journal on Emerging Trends in Technology, 4(1), 178- 180
12. Aydemir, B. 2017. Veri Madenciliği Yöntemleri Kullanarak Meslek Yüksek Okulu Öğrencilerinin Akademik Başarı Tahmini." Master's Thesis., Pamukkale Üniversitesi Fen Bilimleri Enstitüsü.

13. Kaura P, Singhb M, Josanc G.S., "Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector", In 3rd International Conference on Recent Trends in Computing (ICRTC-2015), Procedia Computer Science 57:500-508v, (2015).
14. Kapur B., Ahluwalia N., Sathyaraj R., "Comparative Study on Marks Prediction using Data Mining and Classification Algorithms", International Journal of Advanced Research in Computer Science, Vol. 8 Issue 3, p632- 636. 5p., (2017).
15. Shahiria A.M., Husaina W., Rashida N.A., "A Review on Predicting Student's Performance using Data Mining Techniques", The Third Information Systems International Conference, Procedia Computer Science 72 ,414 – 422, (2015).
16. Karagöz, Yusuf. "Makine Öğrenmesi Algoritmaları ile Ses Sinyallerinden Parkinson Hastalığının Tespit Edilmesi." (2023).
17. Kiran, M. S., Siramkaya, E., Esme, E., & Senkaya, M. N. (2022). Prediction of the number of students taking make-up examinations using artificial neural networks. *International Journal of Machine Learning and Cybernetics*, 13(1), 71-81.
18. McCulloch, W. S., & Pitts, W. A. (1943). A logical calculus of the ideas immanent in nervous activity. *Buttetin of Mathematics and Biophysics*, 5, 115-133.
19. Transactions on Neural Networks, 2(6), 568-576. Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. WESTCON Convention, Record Part IV, 96-104.
20. Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.
21. Hopfield, J. J. (1982a). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National, Academy of Sciences*, 79, 2554-2558.
22. Hopfield, J. J. (1982b). Neurons with graded response have collective computational properties like those of two state neurons. *Proceedings of the National, Academy of Sciences*, 81, 3088-3092
23. Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing, explorations in the microstructure of cognition*. Foundations, MIT Press Cambridge, MA, 1
24. Haykin, S. S. (2009), 'Neural networks and learning machines', third edn, Pearson Education

25. Nasuhoğlu H., Eczacılık Sektöründe Yapay Sinir Ağları Ve Zaman Serileri Analizi İle Talep Tahmini, Yüksek Lisans Tezi, Maltepe Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2019, 611168.
26. Güvenç, Uğur, and Burak KOÇAK. "Kademeli İleri Geri Yayılım ve Gauss Fonksiyon Modelleri ile Pomza ve Diatomit İçeren Çimento Harçlarının Basınç Dayanımlarının Tahmini." Düzce Üniversitesi Bilim ve Teknoloji Dergisi 11.2 (2023): 682-698.
27. Gordon, G. ve Pressman, I. (1983). Quantitative Decision-Making For Business. İkinci Baskı, USA: Prentice Hall International, Inc
28. Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees, Wadsworth International Group, 1984.
29. Binokay H. (2018) Lojistik Regresyon ve Farklı Sınıflama Modellerinin Performanslarının Karşılaştırılması Yüksek Lisans Tezi, Çukurova Üniversitesi Biyoistatistik Anabilim Dalı, Adana, 149s
30. Bounsaythip, Catherine ve Esa Rinta-Runsala, "Overview of Data Mining For Customer Behavior Modeling", VTT Information Technology Research Report, Version:1, 2001, s. 21.
31. Kavzaoğlu, T., Çölkesen, İ., (2010). Karar Ağaçları ile Uydu Görüntülerinin Sınıflandırılması Kocaeli Örneği. Harita Teknolojileri Elektrik Dergisi. 36-45 e-ISSN: 1309-3983
32. Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data Via the EM Algorithm, Journal of the Royal Statistical Society. Series B (Methodological), 39(1), 1-38.
33. Özkan, Emre. "Beklenti Maksimizasyonu ile Genişletilmiş Hedef İzleme." Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi 34.1 (2019): 145-154.
34. Baygül, Arzu , Kayıp Veri Analizinde Sıklıkla Kullanılan Etkin Yöntemlerin Değerlendirilmesi, İstanbul Üniversitesi Sağlık Bilimleri Enstitüsü Yüksek Lisans Tezi 2007.
35. Bruzzone, Lorenzo, and D. Fernandez Prieto. "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images." IEEE Transactions on image processing 11.4 (2002): 452-466.
36. Cihan Pınar, Oya Kalipsiz, and Erhan Gökçe. "Yenidoğan kuzularda bilgisayar destekli tanı." Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi 26.2 (2020): 385-391.

37. Powers W., Ailab A., Evaluation: from precision, recall and F-measure to ROC informedness, markedness and correlation, J. Mach. Learn. Technolgy, 2008, 2: 2229-3981.
38. Gülcan O., Doğruluk (Accuracy) Kesinlik (Precision) Duyarlılık (Recall) F1 Score, <https://medium.com/@gulcanogundur/do%C4%9Fruluk-accuracy-kesinlik-precision-duyarl%C4%B1l%C4%B1k-recall-ya-da-f1-score-300c925feb38>, erişim tarihi: 6 Kasım 2023.
39. Yargı, Veysel, and Seda Postacıoğlu. "EEG işareti kullanılarak bağımlılığa yatkınlığın makine öğrenmesi teknikleri ile analizi." El-Cezeri 8.1 (2021): 142-154.
40. Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46
41. Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378- 382.
42. Landis, J. R. ve Koch, G. G., 1977, The Measurement of Observer Agreement for Categorical Data, Biometrics, 33 (1), 159-174

YAYINLAR

1. Kiran, M. S., Siramkaya, E., Esme, E., & Senkaya, M. N. (2022). Prediction of the number of students taking make-up examinations using artificial neural networks. *International Journal of Machine Learning and Cybernetics*, 13(1), 71-81.

