HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

PISA 2018 TÜRKİYE ÖRNEKLEMİNDE OKUMA OKURYAZARLIK DÜZEYLERİNİN FARKLI VERİ MADENCİLİĞİ SINIFLANDIRMA YÖNTEMLERİ İLE İNCELENMESİ

Emrah BÜYÜKATAK

Doktora Tezi

Ankara, 2022

Liderlik, araştırma, inovasyon, kaliteli eğitim ve değişim ile

Daha ileriye... En İyiye...

HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

PISA 2018 TÜRKİYE ÖRNEKLEMİNDE OKUMA OKURYAZARLIK DÜZEYLERİNİN FARKLI VERİ MADENCİLİĞİ SINIFLANDIRMA YÖNTEMLERİ İLE İNCELENMESİ

EXAMINATION OF READING LITERACY LEVELS IN PISA 2018 TURKEY SAMPLE WITH DIFFERENT DATA MINING CLASSIFICATION METHODS

Emrah BÜYÜKATAK

Doktora Tezi

Ankara, 2022

Kabul ve Onay

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Emrah BÜYÜKATAK'IN hazırladığı "PISA 2018 Türkiye Örnekleminde Okuma Okuryazarlık Düzeylerinin Farklı Veri Madenciliği Sınıflandırma Yöntemleri ile İncelenmesi" başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Yüksek Lisans/Doktora Tezi** olarak kabul edilmiştir.

Jüri Başkanı	Prof.Dr. Nizamettin KOÇ	İmza
Jüri Üyesi (Danışman)	Prof. Dr. Duygu ANIL	İmza
Jüri Üyesi	Prof.Dr. Devrim Alıcı	İmza
Jüri Üyesi	Doç.Ergül DEMİR	İmza
Jüri Üyesi	Doç.Dr. Burcu ATAR	İmza

Enstitü Yönetim Kurulunun
…./…/…. Tarihli ve ……
sayılı kararı.

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından ..... / ..... / ..... tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca ..... / ..... / ..... tarihi itibarıyla kabul edilmiştir.

Prof. Dr. Selahattin GELBAL
Eğitim Bilimleri Enstitüsü Müdürü

Öz

Bu araştırmanın amacı PISA 2018 Türkiye örneklemine dayalı olarak öğrencilerin okuma becerileri başarısını etkileyen faktörlere ve başarı puanlarına göre başarı durumlarının ve okuma becerileri yeterlilik düzeylerinin Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri ile sınıflama doğruluklarının belirlenmesi ve başarı gruplarının genel karakteristiğinin incelenmesidir. Araştırmada PISA 2018 Türkiye uygulamasına katılan 6890 öğrenci anketi kullanılmıştır. Birinci aşamada kayıp veri incelenmiş ve eksik veriler tamamlanmıştır. İkinci aşamada alanyazın, PISA 2018 Teknik Rapor ve veriler incelenerek okuma becerileri başarısını etkilediği düşünülen 24 indis değişken belirlenmiştir. Üçüncü aşamada alt problemler dikkate alınarak öğrenciler PISA 2018 okuma becerileri başarı testi puanlarına göre "Başarılı-Başarısız" olarak 2 ve yeterlik düzeylerine göre "Düzey-1", "Düzey-2" ve "Düzey-3" olarak 3 kategoride ölçeklenmiştir. Verilerin istatistiksel çözümlemeleri SPSS MODELER programı ile yapılmıştır. Araştırma sonunda başarı puanlarına göre; Karar Ağaçları C5.0 algoritmasının %89.6 ile en yüksek, QUEST algoritmasının %75 ile en düşük sınıflama oranına sahip olduğu, genel karakteristiğin incelenmesinde iki aşamalı kümeleme analizi yöntemiyle gruplandırılması sonucunda dağılımları oransal olarak birbirine yakın dört küme elde edildiği, başarı düzeylerine göre de Karar Ağaçları C5.0 algoritmasının %88.6 ile en yüksek, QUEST algoritması da %61.7 ile en düşük sınıflama oranına sahip olduğu, genel karakteristiğin incelenmesi sonucunda dağılımları oransal olarak birbirine yakın olmayan üç küme elde edildiği belirlenmiştir. Her iki kümeleme analizinde de 0,1 olarak hesaplanan Sihoutte Katsayısının 0 değerinden büyük olmasından dolayı veri setlerinin kümeleme yapılmaya elverişli olduğu ifade edilebilir. Hem başarı puanlarına hem de düzeylerine göre bütün veri madenciliği yöntemlerinin rastgele sınıflamanın ötesinde doğru sınıflandırma yapabilmesi sebebiyle öğrencileri sınıflandırmada kullanılabileceği sonucuna varılabilir.

Anahtar sözcükler: PISA, okuma becerileri, veri madenciliği, kümele analizi, sınıflandırma.

Abstract

The purpose of research is to determine the classification accuracy of students' success status and reading skills proficiency levels according to the factors affecting the success of students' reading skills and their success scores based on the PISA 2018 Turkey sample by using Artificial Neural Networks, Decision Trees, K-Nearest Neighborhood and Naive Bayes methods and to examine the general characteristics of success groups. In the research, 6890 student questionnaires were used. Firstly, the missing data were examined and completed. Secondly, 24 index variables were determined by examining the literature, PISA 2018 Technical Report and data. Thirdly, the students were scaled in 2 categories as “Successful-Unsuccessful” according to the scores of PISA 2018 reading test and in 3 categories as “Level-1”, “Level-2” and “Level-3” according to their proficiency levels. Statistical analysis was conducted with SPSS MODELER. At the end of the research, Decision Trees C5.0 had the highest classification rate with 89.6%, QUEST had the lowest classification rate with 75%, and four clusters were obtained with the Two-Step Clustering analysis method to according to the success scores. C5.0 had the highest classification rate with 88.6% and the QUEST had the lowest classification rate with 61.7%, and three clusters whose distributions are not proportionally close to each other were obtained. It can be said that the data sets are suitable for clustering and according to both their achievement scores and their levels, all data mining methods can be used to classify students because of their ability to correctly classify beyond random classification.

Keywords: PISA, reading skills, data mining, cluster analysis, classification.

Teşekkür

Doktora eğitimim boyunca yakın ilgi ve desteğini gördüğüm, bana sabırla yardımcı olan, yol gösterici olan ve güler yüzünü hiç eksik etmeyen, her zaman yanımda olan ve her türlü fedakârlığı sağlayan çok kıymetli hocam, danışmanım Prof. Dr. Duygu ANIL'a,

Doktora eğitim boyunca ve akademik çalışmalarımda bana her zaman katkıda bulunan, motivasyon sağlayan ve beni destekleyen çok değerli hocam Doç. Dr. Ergül DEMİR'e,

Akademik alanda bana gelişmemde yol gösterici olan değerli hocam Doç. Dr. Burcu ATAR'a,

Tezimin gelişmesinde çok önemli katkıları olan kıymetli hocalarım Prof.Dr. Nizamettin KOÇ'a ve Prof.Dr. Devrim ALICI'ya,

Bu süreçte desteğini her zaman hissettiren, yardımını esirgemeyen ve tezimi tamamlayabilmem için her zaman bana destek olan canım eşim Filiz BÜYÜKATAK'a ve ona baktığımda manevi desteğini hissettiğim kuzucuğum Metehan'ıma,

Benden desteğini hiç esirgemeyen, doktora sürecine ve tezime katkıda bulunan arkadaşlarım Sevil ÇINAR ve Cansu AKDAĞ'a,

Tez sürecinde ve özellikle de teknik ve programlama açısından tezime desteğini esirgemeyen can arkadaşım Gökhan KARADENİZ'e,

Doktora eğitimi ve tez sürecinde her konuda ve özellikle de öğrenci işleri konusunda her zaman yardımcı olan Müjgan KAHVECİ ve Pelin KÖROĞLU'na,

Yetişmemde emeği geçen, bilgisini benimle paylaşan, beni destekleyen herkese çok teşekkür ederim.

İçindekiler

Tablolar Dizini

Şekiller Dizini

Simgeler ve Kısaltmalar Dizini

EVM : Eğitimsel Veri Madenciliği

PISA (Programme for International Student Assessment): Uluslararası Öğrenci Değerlendirme Programı

PIRLS (Progress in International Reading Literacy Study): Uluslararası Okuma Becerilerinde Gelişim Projesi

TIMSS (Trends in International. Mathematics and Science Study): Uluslararası Matematik ve Fen Eğilimleri Araştırması

YSA: Yapay Sinir Ağları

KNN: K-En Yakın Komşuluk

ÇKA: Çok Katmanlı Algılayıcı

RTF: Radyal Tabanlı Fonksiyon

AUROC (Area Under The ROC Curve): ROC Eğrisi Altında Kalan Alan

Bölüm 1
Giriş

Teknoloji ve üretimin artmasına paralel olarak eğitimin de küreselleştiği modern bilgi ve veri çağına uygun özellikte ve başarılı insanlar yetiştirmek, ülkelerin eğitim politikalarının ana amaçlarından birisi olmaktadır. Öğretim programları öğrencilerin okulda sadece ne öğrendiğinden ziyade öğrendiklerini kullanarak neler yapabildikleri üzerine odaklanmaktadır. Artık ülkelerin eğitim sistemleri, bu hedeflere paralel olarak ulusal ve uluslararası düzeydeki performanslarla da değerlendirilmeye başlanmıştır. Bu kapsamda son yıllarda, eğitim sistemlerinin değerlendirilmesi için geniş ölçekli sınavlar uygulanmakta, farklı akademik alanlarda ve türdeki okullardaki öğrenci sonuçları değerlendirilerek karşılaştırma yapılmaktadır. Farklı sınıf düzeyi ve alanlarda belirlenmiş bilgi ve becerileri kapsayan, birden çok alt test ve boyuttan oluşan başarı testleri geniş ölçekli testler olarak ifade edilmektedir.

Ülkeler eğitim sistemlerinin dünya çapındaki seviyesini değerlendirmeye yönelik uluslararası çalışmalara katılmakla birlikte, uyguladıkları eğitim sisteminin ülke vizyonu doğrultusundaki işlerliğini görebilmek için de uluslararası sınavlara ve uygulamalara katılarak değerlendirme yapmaktadır. Ülkelerin kendi eğitim sistemlerini izleyip değerlendirebilmek adına uluslararası izleme ve değerlendirme çalışmalarında yararlandıkları uygulamalardan birisi olan PISA, OECD'nin düzenlediği 15 yaş grubu öğrencilerin matematik, fen ve okuma alanlarındaki performanslarını belirlemek ve onların sahip oldukları motivasyonları, kendileri ile ilgili düşünceleri, öğrenme yöntemleri, eğitim gördükleri şartlar, bulundukları ortamlar ve kendi aileleri hakkında veriler toplayarak sadece OECD ülkelerinin değil, bununla birlikte dünya genelinde ülkelerin eğitim sistemlerinin değerlendirildiği uygulama ve programdır. PISA'da matematik, fen ve okuma alanlarındaki öğrencilerin sahip oldukları becerilerin tespit edilmesinden ziyade matematik, fen ve okuma okuryazarlığını gerçek durumlarda karşılaşacakları sorunları ve problemleri çözmede ne derece kullanabildiklerinin belirlenmesi hedeflenmektedir (OECD, 2013a, 2013b).

Her gün her alanda iş, okul, toplum, sosyal hayat, bilim ve günlük yaşamın hemen hemen her alanında çok büyük miktarda ve çeşitlilikte veriler depolama

cihazlarına aktarılmaktadır. Günümüzde veri tabanları artık terabaytlarla ifade edilebilecek büyüklüğü de geçmiştir. Bu kadar çok miktardaki veri içinden verimli olanları sistemli olarak ortaya çıkarmak ve bunları organize edilmiş verilere ve bunun sonucunda da bilgiye dönüştürmek için kuvvetli sistemlere ve vasıtalara ihtiyaç duyulmuştur. 1700'li ve 1800'li yıllara kadar dayanan uzun bir geçmişi olan veri madenciliği, bu gelişmelerle birlikte 1980'li yıllarda ön plana çıkmaya, daha popüler olmaya başlamıştır. Gittikçe artan veri hacmi ve bilgisayar kullanımının yaygınlaşması veri madenciliğinin daha fazla ön plana çıkmasına neden olmaktadır.

Veri madenciliği büyük veri tabanlarında gizli ilişkilerin ve genel örüntülerin araştırılması olarak da tanımlanmaktadır (Holsheimer ve Siebes, 1994). Bu yöntem büyük hacimdeki veriler içinde saklı bulunan önemli ticari ve bilimsel bilgilerin ortaya çıkarılmasını mümkün kılmaktadır (Jain ve Dubes, 1998; Wu, 2012). Bu doğrultuda veri madenciliği teknikleri PISA, TIMMS gibi çok miktarda verilere ait parametreler arasındaki ilişkinin ortaya çıkarılmasını mümkün kılmaktadır.

Veri madenciliğini eğitim yönüyle inceleyen Eğitimsel Veri Madenciliği (EVM) ise eğitimdeki alanındaki zamanla fazlalaşan büyük hacimli verileri ortaya çıkarmak için yöntem oluşturan, sonucunda da bunları öğrencileri ve eğitim ortamlarını kapsamlı şekilde anlamak amacıyla uygulayan yeni disiplin olarak görülmektedir. Eğitimsel veri madenciliği son dönemde önemli hale gelen bir araştırma sahasıdır ve araştırma problemlerinin ve sorunlarının çözümü için bu alanlarda oluşan özgün veri setlerinin analizini amaçlamaktadır (Baker ve Yacef, 2009).

Problem Durumu

Toplumun, bireylerin, ekonomik durumların gelişiminde ve değişiminde okuma becerileri en önemli öğelerden birisidir. Bilgi edinmede en temel yollardan biri okuma becerileridir ve bu alandaki üst düzey yeterlik yaşamın bütün ortamlarında başarı sağlayabilir (Coşkuner, 2013). Buna paralel olarak üst düzey düşünme becerilerini ölçen PISA'da sadece okuma becerileri, matematik ve fen bilimleri bilgi ve becerilerinin değil, aynı zamanda Türkçe, matematik ve fene karşı tutumların da ele alınıp, okullarda kazandıkları bilimsel yeterliklerin onlara ne gibi imkânlar yaratacağının farkında olup olmadığı da değerlendirilmektedir (Anıl, 2009).

Okuma becerileri veya okuma okuryazarlığı, bireysel amaçlara sahip olma, belirli alanlardaki sahip olunan bilgiyi artırma, aktif birey olma ve metinleri anlama

ve metne ilgi duyma olarak belirtilmektedir (OECD, 2010a). PISA'daki okuryazarlık kavramı ise gerekli bilgi, beceri ve kavramlara odaklanmakta ve topluma dahil olmak için gerekli bilgi ve becerilerle birlikte üst düzey okur yazarlığı da gerekli kılmaktadır.

Okuryazarlık kavramı yalnızca okuldaki değil bununla birlikte etkileşim içinde olunan aile bireyleri, akran, iş arkadaşları gibi büyük gruplar içerisinde de kişileri etkileyen ve değişim getiren ve hayat boyu süregelen bir süreç olarak ifade edilmektedir. Okuryazarlık, sosyal, kültürel ve siyasi hayata katılımı sağlamak, bireysel tatmini yaşamak, kişisel gelişi sağlamak, istihdam imkanlarını artırmak ve doğal olarak da başarı için önem bir süreçtir (İş, 2003). Böylelikle okuma becerileri ve okur yazarlığı yaşamın bir sürü alanını etkilemektedir. Bununla birlikte diğer değişkenlerden de etkilenebildiği, okuma becerileri başarısına tesir eden falı türlerde yapıların bulunduğu görülebilmektedir. Bundan dolayı okuma becerileri başarısını etkileyen yapıların etkilerinin araştırılması okuma becerileri başarısının anlaşılmasında ve artırılmasında önem arz etmektedir. Bu amaçla PISA'da öğrencilerin sınav puanları ile birlikte öğrenci başarısını etkilediği değerlendirilen diğer değişkenler arasındaki ilişkiler de belirlenmeye çalışılarak akademik başarının daha iyi yordanması sağlanmaktadır.

PISA'daki hem test puanlarındaki başarı puanları verileri hem de okuma becerileri başarısını etkilediği değerlendirilen değişkenlerin tespit edilmesi amacıyla oluşturulan anketlerden ortaya çıkan ve yaklaşık yarım milyondan fazla öğrenciden elde edilen bu bilgiler ise büyük veri yığınını oluşturmaktadır. Ancak bu kadar çok verinin olduğu durumlarda önemli olan ise karar sürecinde hangisinin anlamlı, hangisinin anlamsız olduğunun tespitidir. Veri tabanındaki verilerden anlamlı bilgiler çıkarılması amacıyla, öncelikle veri kütleleri içinden ulaşılması istenen ve gerekli veriler alınarak sınıflandırma yapılır ve ardından da bu veriler işlenir. Elde edilen verilerin çok miktarda olması sebebiyle bilgisayar programı kullanılarak gelecek için tahmin etmemizi sağlayabilecek bağıntılar elde edilmesi veri madenciliği yöntemleri ile yapılmaktadır.

Veri madenciliği, veri tabanındaki verilerden bilginin otomatik bir şekilde meydana getirilmesi ve analizinde birden çok bilgisayar öğrenme tekniklerinin uygulanma sürecidir (Roiger, 2017). Ayrıca bu süreç, geçerli tahminlerde bulunmak amacıyla verilerdeki örüntüleri ve veriler arasındaki ilişkiyi ortaya koymak için birden çok veri analiz araçlarının kullanılmasıdır. PISA verilerinin büyük veri olmasından

dolayı bu verilerin veri madenciliğinde kullanılabileceği ortaya çıkmıştır (Nisbet, Elder ve Miner, 2009).

Bütün alanlarda olduğu gibi veri madenciliğinin eğitim alanına yönelik çalışmalar da yapılmaktadır. Çünkü veri madenciliğinin, eğitim alanı faktörlerinden olan öğrencilere, öğretmenlere ve sorumlulara etkili analiz sonuçları sağlayarak önemli çıktılar ortaya koyacağı bilinmektedir. Diğer sektörlere oranla veri madenciliğinin eğitim alanında çok sayıda çalışma ve uygulaması bulunmamaktadır. Ancak özellikle son dönemde eğitim alanındaki ölçme sonuçları dikkate alınarak yapılan seçmede ve sınıflamada veri madenciliğinin kullanışlı olacağı düşünülmeye başlanmıştır. Böylece hangi küme veya sınıfta hangi değişkenin etkili olabileceği tespit edilerek öğrencilerin öğrenme seviye ve davranışı daha iyi kavranabilecektir. Bunun sonucu olarak başarı ve bu başarıya tesir eden etkenlerin tespiti için yapılan yordama çalışmaları çoğalmıştır (Anıl, 2008; Gelbal, 2008; Erdil, 2010; Anıl, 2011; Özer ve Anıl, 2011).

Eğitimde bu derecede büyük verilerin kullanılmasının amaçlayan EVM, büyük hacimli verilere sahip olmasından dolayı analizi zor olan çok miktardaki eğitim verileri örüntülerini belirlenmesi amacıyla bilgisayarda analiz metotları geliştirme ve uygulamayla ilgilenmektedir (Romero ve Ventura, 2007). Toplanan büyük eğitim verilerindeki gizli örüntülerin ve fonksiyonların bulunarak, bunların bilgiye dönüştürülmesi aşamasında faydalanılan matematiksel ve istatistiksel yöntemlerdir ve eğitim sürecinin hemen her ortamında karşılık bulmaktadır. Bu kapsamda eğitimde kaliteyi ve verimliliği artırmak için veri madenciliği tekniklerini kullanılması yararlı olmaktadır. Böylelikle başarıya tesir eden etkenlerin belirlenerek, öğrencilerin sahip oldukları performansların artırılması için uygulamalar geliştirilmektedir.

Araştırmanın Amacı ve Önemi

Öğrencilerin hedeflerine ulaşması, bilgilerini ve mevcut sahip oldukları düzeylerini geliştirmesi ve sosyal çevreye katılması için farklı tarzlarda verilen metinleri anlaması, bu metinleri kullanması, metinlerin kendileri değerlendirmesi, birbiriyle veya başka metinlerle ilişkilendirmesi ve bu metinler hakkında derinlemesine düşünmesi okuma becerileri olarak ifade edilmektedir. Okuma becerisinin önemi, son yıllarda günümüz hızla değişen teknolojinin etkisiyle önemini artırmaya başlamış, başarı, bireysel gelişim ve ekonomi vb. alanlarda eskiden

gerekli olan okuma becerisi ile şu an gerekli olan okuma becerisi arasında nitelik ve içerik bakımından farklılıklar olmaya başlamıştır. Okuma, daha önce olmadığı kadar pratik gereklilik olmuş, farklı kaynaklardan elde edilen bilgilerin çözümlenmesi, analiz edilmesi, bunların birleştirilmesi ve yorumlanması işlemleri okumanın gündelik süreci olmuştur (MEB, 2019). Böylelikle birden çok kaynak ve yöntem kullanarak bilgiye ulaşma, elde edilen ve ulaşılan bilgiyi anlama ve nihai olarak da bilgilerin değerlendirilmesi okuma becerilerinde ölçülen temel özellikler olmuştur. PISA 2018 uygulamasında metindeki bilgilerin taranması ve bulunması amacıyla ilgili metinleri arama ve seçme, elde edilen bilgileri anlama ve çıkarımlar oluşturma, bilgilerin değerlendirme ve derinlemesine düşünme süreçlerinin ölçülmesi vasıtasıyla okuma becerileri temel özellikleri ölçülmektedir.

PISA 2018 kapsamında okuma becerisi, metinleri sesli şekilde ifade etmenin ilerisinde belli bir amaca yönelik sağlanan bir ya da daha çok metindeki bilgi ile benzer ilişki oluşturmasını sağlayacak yeterliklerin tümünü ifade etmektedir. Okumada bazı yeterliklere sahip olmanın ötesinde öğrencilerin çeşitli hedefler doğrultusunda okuma yapmaları ve yüksek motivasyona sahip olmaları beklenmektedir (OECD, 2019b). Ortaya konan bu özellikler PISA'da ölçülen özellikler ile okuma becerisi arasında yakın ilişki olduğunu göstermektedir.

Bu büyük veri setlerine ilişkin parametreler arasındaki ilişkiler veri madenciliği yöntemleri kullanılarak yapılmaktadır. PISA uygulamalarının geniş ölçekli sınavlar olması sebebiyle bütün bu süreçler sonucunda PISA'da farklı formatlardan ve kaynaklardan büyük miktarda veriler toplanmaktadır. Okuma becerilerinin ölçülmesi sürecinde PISA'da bu büyük verilerin elde edilmesi, PISA sonuçlarının analizinde ve bununla birlikte okuma becerilerinin ölçülmesinde veri madenciliği yöntemlerinin kullanılabileceğini göstermektedir.

Bu çalışmada PISA 2018 Türkiye örneklemine dayalı olarak öğrencilerin okuma becerileri başarısını etkileyen faktörlere ve başarı puanlarına göre başarı durumlarının ve okuma becerileri yeterlilik düzeylerinin Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri ile sınıflama doğruluklarının belirlenmesi ve başarı gruplarının genel karakteristiğinin incelenmesi amaçlanmıştır.

PISA uygulamalarında, öğrencilerin akademik başarı düzeylerini ölçmek amacıyla hazırlanan bilişsel testin sonuçları, öğrenci, veli ve okul yöneticisi anketlerinden oluşan bütünleşik veri toplanmaktadır. Uygulanan bu anketler ile öğrenci başarısına etki eden değişkenlerin tespit edilebilmesi amaçlanmaktadır. Anket sonuçları ile test sonuçları arasında yapılan istatistiksel analizler, öğrenci başarısının yordanması açısından önemli bilgiler sağlamaktadır. EVM, öğrenme verilerinin etkisini ve nasıl kullanılabileceği göstererek, makro boyutta eğitim sürecine yardımcı karar destek ve tavsiye sistemlerinin geliştirilmesini de sağlamaktadır (Huebner, 2013). Bu şekilde devletlerin eğitim politikalarını gözden geçirmesine ve iyileştirmesine hizmet etmektedir. Bu açıdan bakıldığında EVM, PISA'da elde edilmesi amaçlanan hedeflere ulaşmada bir vasıta veya yöntem olarak hizmet etmektedir.

Eğitimde veri madenciliği uygulamalarında farklı istatistiksel modele dayanan farklı yöntemler kullanılmaktadır. Bu alanda çok yöntem olmasından dolayı hangi yöntemlerin uygulamada daha etkili kestirimde bulanacağı ve düşük hataya dayalı hesaplama yaptığının tespiti sonuçların güvenirliği ve geçerliği konusunda önem taşımaktadır.

Son dönemde çok çalışma yapılan yordama ve sınıflamada alan yazında farklı uygulamalar kullanılmaktadır. Bu uygulamalarda kullanılan modellerin kendine özgü bir algoritması bulunmaktadır. Kullanılan algoritmaların karşılaştırılması ve böylelikle değerlendirmeye tabi tutulması, algoritmaların hangi koşullarda başarılı olduğunun ortaya çıkarılması ve performanslarının artırılması bakımdan önemlidir. Bu kapsamda araştırmanın, 2018 PISA uygulamasındaki öğrenci başarısı üzerinde etkili olan değişkenlerin bulunması, öğrencilerin gelecekteki başarılarının tahmin edilmesinde kullanılan YSA, Karar Ağaçları, KNN ve Naive Bayes yöntemlerinin sınıflama doğruluğunun incelenmesi ve bununla birlikte bu yöntemlerin performanslarının karşılaştırılması açısından alan yazına katkı getireceği düşünülmektedir. Bununla birlikte alan yazında yapılan çalışmalar genel olarak Yapay Sinir Ağları ve Karar ağaçları ile sınırlı kalmaktadır. Ancak bu çalışma öğrencilerin gelecekteki başarısını tahmin etmeyi sağlayacak sınıflandırma modellerinin elde edilmesinde YSA, Karar Ağaçları, KNN ve Naive Bayes yöntemlerinin de kullanılması ve birlikte karşılaştırılması açısından da önemlidir. Bu sebeple çalışmada en çok kullanılan sınıflandırma yöntemleri dışında alan yazına

önemli katkılar sağlayan diğer veri madenciliği sınıflandırma yöntemlerinin de incelenmesi çalışmanın önemini artırmaktadır.

Alan yazında veri madenciliği çalışmaları çoğunlukla 2 kategorili (başarılı:1, başarısız:0) olacak şekilde yapılmış, sınıflama doğruluğu öğrencilerin başarı durumlarına göre incelenmiştir. Bu çalışmada PISA 2018 verilerine dayalı olarak elde edilen sonuçlara göre öğrencilerin okuma başarı puanları yeterlik düzeylerine göre belirlenmiş ve 6 yetenek düzeyi 3 kategoriye indirilerek (1 ve 2. Düzey "Düzey-1", 3 ve 4. Düzey "Düzey-2", ve 5 ve 6. Düzey "Düzey-3) sınıflama doğruluğu öğrencilerin okuma başarısı yetenek düzeylerine göre de incelenmiştir. Bu kapsamda çalışma alan yazında en çok kullanılan başarı durumlarına göre 2 kategorili sınıflama doğruluğu analiz sonuçlarını göstermesi ile birlikte yetenek düzeylerine göre 3 kategorili sınıflama doğruluğu sonuçlarını da göstererek alan yazına katkı sağlayacak ve çalışmanın önemi artacaktır.

Veri madenciliği yöntemlerinin özellikle sanayii, bankacılık gibi sektörel bazda yoğun olarak kullanımına karşın, eğitim alanında çok az çalışmalar yapılmıştır. Oysa eğitim alanında derlenen verilerin kullanılması, bu alanda başarı elde edilebilmesi ve öğrenci başarısının arttırılmasında merkezi bir öneme sahiptir. Teknolojik gelişmelerle birlikte eğitim alanında da daha fazla verinin toplanması sonucunda, veri madenciliği yöntemlerinin alışagelmiş sektörel baz dışında eğitim alanlarında da kullanılmasını incelemesi açısından bu araştırma önem arz etmektedir.

Bu araştırmada İki Aşamalı Kümeleme kullanılarak, büyük hacimli PISA 2018 veri setindeki öğrencilere ait verilerin benzer niteliklere göre aynı kümede toplanarak farklı grupların belirlenmesi ve bu gruplar üzerindeki değişkenlerin önemini görülmesi de çalışmanın diğer bir önemli yönünü göstermektedir.

Araştırma Problemi

PISA 2018 verilerine dayalı olarak öğrencilerin okuma becerileri başarısını etkileyen faktörlere ve okuma becerileri başarı puanlarına göre Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri öğrencilerin başarı durumlarını ve okuma yeterlik düzeylerini hangi oranda doğru sınıflama yapmaktadır.

Alt problemler. Araştırma problemine ait alt problemler aşağıda sunulmaktadır:

1. Öğrencilerin 2018 PISA okuma becerileri başarısını etkileyen faktörlere ve okuma becerilerine ilişkin başarı puanlarına göre Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri öğrencileri başarı durumlarına göre hangi doğruluk oranında sınıflandırmaktadır?

2. Öğrencilerin 2018 PISA okuma becerileri başarısını etkileyen faktörlere ve okuma becerilerine ilişkin başarı puanlarına göre başarı gruplarının genel karakteristiği nedir?

3. Öğrencilerin 2018 PISA okuma becerileri başarısını etkileyen faktörlere ve okuma becerilerine ilişkin başarı puanlarına göre Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri öğrencileri okuma becerileri yeterlik düzeylerine göre hangi doğruluk oranında sınıflandırmaktadır?

4. Öğrencilerin 2018 PISA okuma becerileri başarısını etkileyen faktörlere ve okuma becerilerine ilişkin başarı puanlarına göre okuma becerileri yeterlik düzeylerinin genel karakteristiği nedir?

5. Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemlerinin öğrencileri başarı durumlarına ve okuma becerileri yeterlik düzeylerine göre genel doğru sınıflandırma oranlarının karşılaştırılmasına ilişkin sonuçlar nasıldır?

Sınırlılıklar

1. PISA 2018 örnekleminde yer alan öğrencilerin öğrenci anketinden seçilen okuma becerileri başarısına ait maddeler ve okuma becerileri başarı puan ortalamasına ait maddelerle sınırlıdır.

2. Karar Ağaçları veri madenciliği yöntemi analizinde sadece C5.0, CHAID, C&RT ve QUEST algoritmaları kullanılmıştır.

Bölüm 2
Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Araştırmanın bu bölümünde kuramsal temel ve ilgili araştırmalar hakkında bilgi verilmektedir.

Araştırmanın Kuramsal Temeli

Kuramsal temelde; Okuma Becerileri, Veri Madenciliği, Eğitimsel Veri Madenciliği, Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk, Naive Bayes yöntemi ile İki Aşamalı Kümeleme Analizine ait kavramsal çerçeveye yer verilmiştir.

Okuma Becerileri

İnsanlar hayat boyu sürekli her konuda bir şeyler okurlar. Bu sürekli okumalar sayesinde kişiler kendini geliştirmektedir. Bu noktada okuma becerisinin önemi ortaya çıkmaktadır. Okuma becerisi ayrıca, öğrencilerin sosyal hayata dahil olmak amacıyla, verilen yazılı metinleri anlayarak değerlendirmesi ve bunların hakkında detaylı düşünüp ilişki kurabilmesini, sahip olduğu bilgilerle halihazırdaki potansiyelini geliştirerek gerçek şartlarda ve ortamlarda kullanabilme becerisi olarak da ifade edilmektedir (OECD, 2019). Okuma becerisi, yalnızca sözcüklerin bilinmesi değil; bununla birlikte dilbilimsel ve metinsel yapı bilgisini bulundurmayı, çözümlemede uygun metotlar uygulayabilmeyi, okumanın etkin bir şekilde belirli bir amaca ve göreve göre yapılmasını, metni anlamayı, kullanmayı, metne ilgi duymayı (Ülper, 2010; Çiftçi, 2007), sabit şekilde metni okurken metni anlayabilmeyi gerektirmektedir.

Okuma becerisi anlama ile doğrudan ilgilidir. Anlama görülen, işitilen, okunanların algılanması ve kavranmasıdır. Okuma becerisi açısından anlama, metinlerin iletisini algılama eylemidir. Bilginin kazanılması ve yorumlanması, doğruluğunun ve yararlılığının sorgulanması süreçlerinde ve bu becerilerin kazanılmasında okuma becerisi önem kazanmaktadır (Sadioğlu ve Bilgin, 2008). Bu kapsamda bireyin kişisel hedeflerini gerçekleştirebilmesi ve bilgi çağını yakalayabilmesi için okuma becerisinin geliştirilmesi son derece önemlidir (Sadioğlu ve Bilgin, 2008).

Okuma becerileri üzerine yapılan araştırma çalışmaları incelendiğinde, bu becerileri etkileyen birçok faktörün olduğu gözlemlenmiştir. Bu faktörler öğrenciden kaynaklı, aile kaynaklı, öğretmen ve okul kaynaklı ve öğretmenin kullandığı strateji kaynaklı olduğu yapılan araştırmaların bulgularında görülmüştür.

PİSA 2000 verilerine dayanılarak yapılan araştırmalarda düşük okuma becerilerinin öğrencilerin aile yapısına, kardeşlerinin sayısına, göçmenlik durumlarına ve sosyo–ekonomik faktörlerine bağlı olduğu (Linnakylä, Malin ve Taube, 2004), aile eğitim seviyesi ve evde bulunan kitap sayısından pozitif olarak etkilendiği, ancak evdeki birden çok televizyon sayısından da negatif olarak etkilendiği belirtmiştir (Van Ours, 2008). Eccles ve Wigfield (2002) tarafından yapılan bir araştırmada ise metnin okunmasının ardından öğrenmenin olabilmesi için mevcut metne ilgi duyulması gerektiği ifade edilmiştir. Baker ve Wigfiield (1999) ise okumaya zevk amaçlı tahsis edilen zamanın, öğrencilerdeki okuma ilgisinin göstergesi olduğunu ve hangi sıklıkla okuma yapılmasının ise okuduğunu anlama ile yakın ilişkili olduğunu belirtmiştir.

Bu araştırmalar bir bütün olarak incelendiğinde bu becerinin niteliğini; cinsiyetin, evdeki aile desteğinin ve aile-okul işbirliğinin, öğrencilerin özel bir dershaneye gidip gitmemelerinin, Türkçe dersinden özel ders alıp almamalarının, öğrencinin kendine ait bir çalışma odasının olup olmamasının, ailenin sosyo-ekonomik yapısının altını çizme, okunan metnin kenarına not alma, kelimelerin değil düşüncelerin takibini yapma gibi okuma stratejilerinin öğretiminin), okudukları metin türünün, öğrencinin okuma sıklığının ve eve gazete-dergi alınıp alınmamasının öğretmen-öğrenci ilişkisinin ve sınıflardaki disiplin anlayışının, hikâye anlatma yönteminin sınıf içinde kullanılmasının gibi bir çok etkenin etkilediği görülmüştür (Aydın, Erdağ ve Taş, 2011; Coşkun, 2003; OECD, 2010a; Yılmaz, 2008).

Bu değerlendirmeler dikkate alındığında okuma becerileri başarısına tesir eden birden fazla yapı olduğu görülmektedir. Bu sebeple bu yapıların etki ve sonuçlarının incelenmesinin okuma becerileri başarısının etkilerinin kavranmasında önemli olacağı düşünülmektedir. Bu sonuçla okuma becerileri, hızla değişen teknolojinin etkisiyle beraber, gün geçtikçe değişmekte ve farklılaşmaktadır. Bunun yanı sıra hala yaşamımızdaki en önemli beceridir.

PISA okuma becerileri değerlendirme çerçevesinin, okuma becerileri alanındaki soruların değerlendirme çalışmalarına dayanak oluşturduğu ve değerlendirme sonuçlarının raporlaştırılması için bir ölçüt sağladığı açıklanmaktadır. Bu açıdan PISA uygulamasındaki okuma becerileri; öğrencilerin hedeflerini gerçekleştirmek, sahip oldukları bilgi düzeyini ve potansiyellerini geliştirmek ve günlük hayatla birlikte topluma katılmada sunulan yazılı formattaki metinlerin anlaşılması, kullanılması, bunların düşünülüp ilgilenilmesi olarak belirtilmektedir (Gürtekin, 2021).

Teknolojideki sürekli meydana gelen değişimler evde, okulda ve işyerinde bilgiyi okuma ve aktarma yöntemlerinin de değişmesine sebep olmuştur. PISA 2018 uygulama sonuçları dikkate alındığında bireylerin boş zamanlarında kitap, dergi ve gazete okuma sıklıklarının kayda değer oranda azaldığı; öğrencilerin okumanın yerine farklı internet sayfalarını kullanarak sohbet, haber ya da kısa formattaki bilgilendirme amaçlı metinleri tercih ettikleri görülmektedir (Doğaç, 2021).

Veri Madenciliği

Teknolojide meydana gelen hızlı gelişmeler neticesinde çok büyük hacim ve türdeki verilerin depolanması mümkün hale gelmektedir. Depolanan verilerin çığ gibi artması ve karmaşık hale gelmesiyle veri tabanlarındaki gizli ilişkilerin ve genel örüntülerin keşfedilmesinde büyük zorluklarla karşılaşılmıştır. Bu zorlukların aşılmasında veri madenciliği yöntemleri önemli hale gelmiştir. Verilerden örüntülerin tespit edilmesi amacıyla algoritmaların kullanılması (Fayyad ve diğ., 1996) olarak tanımlanan veri madenciliği ile ilgili birçok tanımlama yapılmaktadır. Bu tanımlamalardan bazıları aşağıda alıntılanmıştır.

En basit tanımıyla veri madenciliği, büyük veri tabanlarında gizli ilişkilerin ve genel örüntülerin araştırılmasıdır (Holsheimer ve Siebes, 1994). Diğer bir ifadeyle büyük veri setlerinden bilginin elde edilmesi sorununun çözülebilmesi için veri tabanları, makine öğrenmesi, örüntü tanıma, görselleştirme ve istatistik gibi çeşitli teknikleri bir araya getiren disiplinler arası bir çalışma sahasıdır (Cabena ve diğ., 1998).

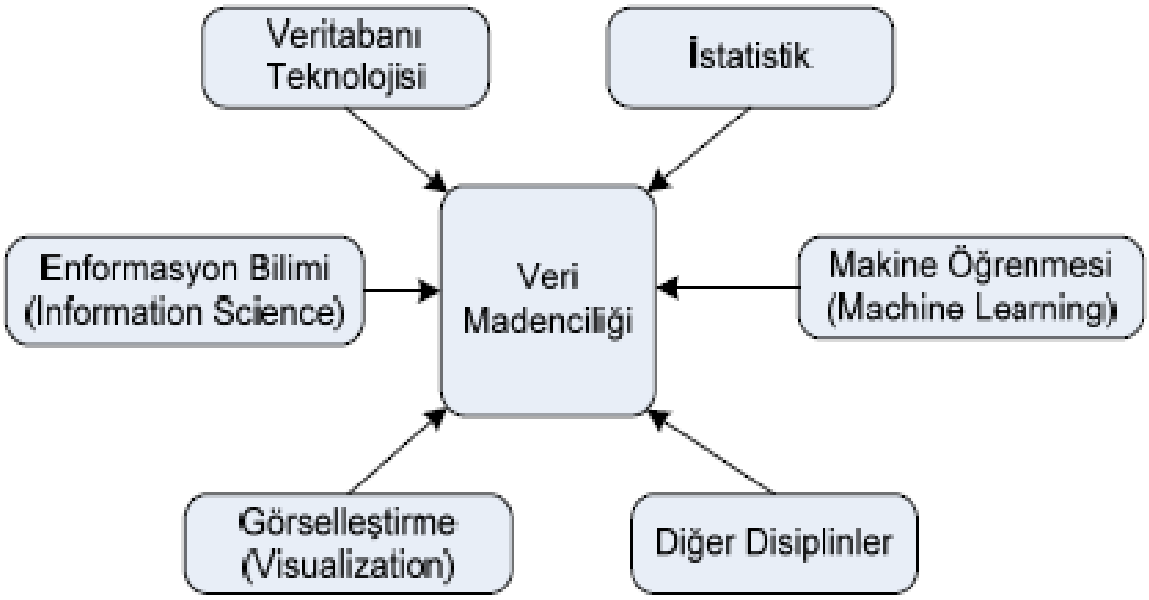
Veri madenciliği verideki örüntülerin geçerli, özgün, kullanışlı ve yeteri kadar anlaşılır biçimde tanımlanması süreci olarak da tanımlanmaktadır (Fayyad ve diğ., 1996). Özellikle istatiksel ve matematik yöntemlerin kullanılarak, anlamlı yeni

korelasyonların, örüntülerin ve trendlerin depolanmış olan büyük miktarlardaki veriden elenerek keşfedilmesi sürecidir (Gartner Group, 2013).

Veri madenciliği maden ortaya çıkarmadaki madenciliğe benzemektedir. Örnek olarak altın elde etmek için tonlarca madde ayrıştırılıp, altının saf hali elde edilmeye çalışılır. Veri madenciliği hammadde olarak veriyi, madeni ya da başka bir deyişle ürünü ise bilgiyi kullanmaktadır. Ortaya çıkan ürün bilgidir ancak sürecin bilgi madenciliği yerine veri madenciliği olarak adlandırılması veri büyüklüğünden kaynaklanmaktadır (Aydın, 2007).

Veri madenciliği yakın zamanda ortaya çıkan yeni bir disiplindir ve geniş bir alanda uygulanmaktadır. Veri madencilik yöntemleri günümüzde sigortacılık, eğitim, pazarlama, bankacılık, borsa, tıp, endüstri sektörleri gibi birçok alanda yaygındır ve karar vermeye ihtiyaç duyulduğu alanlarda sık kullanılmaktadır.

Şekil 1'de görüldüğü üzere veri madenciliği veri tabanı, makine öğrenmesi, görselleştirme, istatistik ve enformasyon disiplinlerini kapsayan disiplinler arası bir alandır (Han ve Kamber, 2001).



Şekil 1. Veri madenciliği disiplinleri.

Veri madenciliği sınıflandırılması. Veri madenciliği sistemleri çeşitli ölçütlere göre sınıflandırılabilir. Bu sınıflandırmalar aşağıdaki gibi yapılmaktadır (Han ve Kamber, 2001).

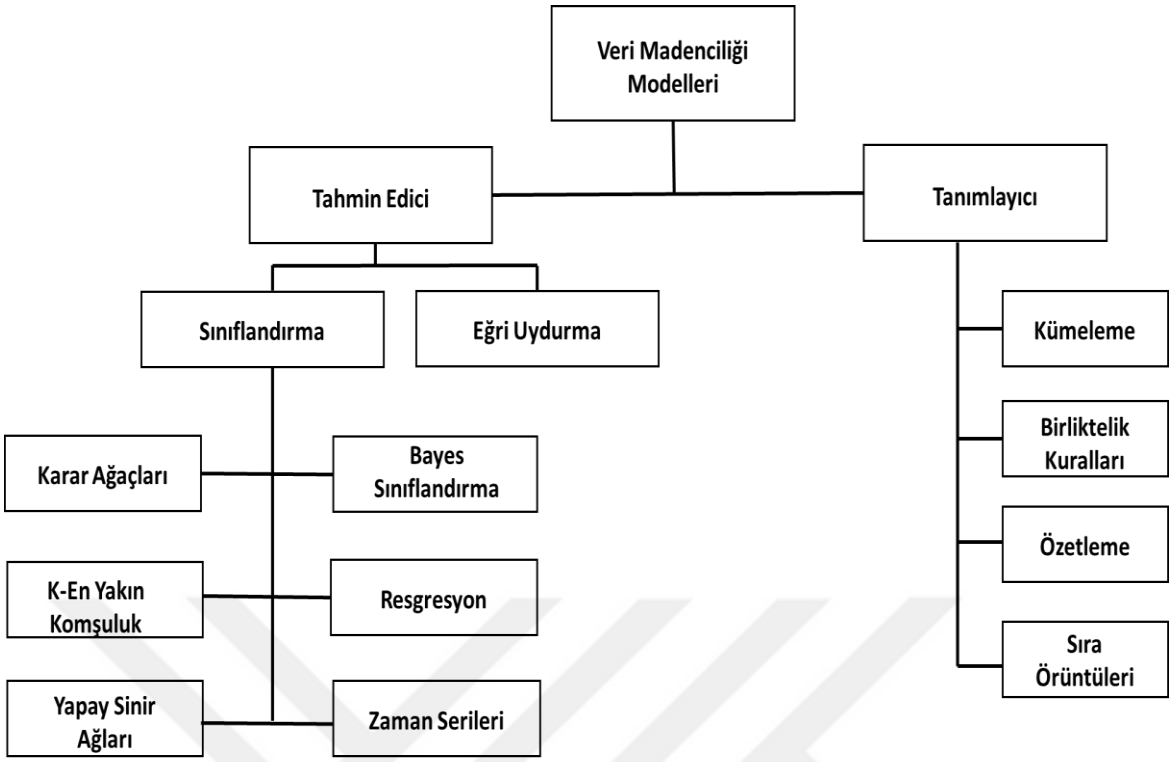
• Veri tabanı kapsamında: Veri tabanı yönetim sistemleri; veri modeli, tipi ve uygulama alanı gibi farklı niteliklere göre aralarında sınıflandırılmaktadır ve kendi özelliklerine has veri madencilik teknikleri uygulamayı gerektirmektedir. Örnek olarak veri tabanı modellerine göre sınıflandırılma yapıldığında harekete dayalı, nesne-ilişkisel veya veri ambarı, ilişkisel ve nesneye dayalı olarak ayrılırlar. Eğer veriler özel nitelikte ise metin, uzaysal, çoklu ortam, zaman serileri veya web madenciliği olarak sınıflandırılma yapılmalıdır.

• Bilgi türü kapsamında: Bilgi türüne göre sınıflama, aykırı değer analizi ve kümeleme gibi sınıflandırma yapılmaktadır. Geniş kapsamdaki veri madencilik sistemleri birden çok fonksiyon yaptığından dolayı birden fazla fonksiyonun birleştirildiği teknikleri de bulunmaktadır.

• Teknik kapsamda: Veri madencilik sistemleri uyguladıkları belirli tekniklere göre sınıflandırılmaktadır. İstatistik, yapay sinir ağları, makine öğrenmesi ve örüntü tanımlama gibi veri analiz metoduna ve uygulayıcının müdahale seviyesine göre tanımlanabilir.

• Uygulama alanı kapsamında: Bu sistemlerin uygulandıkları alana göre sınıflandırılması da mümkündür. e-posta, iletişim, borsa, finans, DNA gibi alanlara yönelik özellikle hazırlanan sistemler bulunmaktadır. Bu sebeple genel amaca yönelik tasarlanan sistemin özel bir alanda uygulanacak veri madenciliğinde kullanılması uygun değildir.

Veri madenciliği modelleri. Veri madenciliğinde en temel düzeyde betimleme (descriptive) ve tahmin yapma (predictive) olmak üzere 2 model kullanılmaktadır. Tahmin modeli, değişik veri setlerinden sağlanan sonuçları kullanarak yeni veri değerleri ile ilgili tahmin yapma, tanımlayıcı modelde ise veride bulunan desen ve ilişkiler tespit edilir. Tanımlayıcı modelde tahmin edicinin tersi olarak yeni özellikleri tahmin edilmez, ancak üzerinde çalışma yapılan veri özellikleri keşfi için yollar sunulmaktadır (Dunham, 2003). Tanımlayıcı modellerde, karara yardımcı olan verilerde bulunan örüntüler tanımlanmaktadır. Tahmin edici modelde ise sonuçları belli olan veride modelin geliştirilme ve bu model sayesinde sonucu tam bilinmeyen veri kümesine sonuç değerinin tahmini sağlanmaktadır.



Şekil 2. Veri madenciliği modelleri.

Bu modellerin altında yer alan işlevler için aşağıdaki tanımlamalar verilmiştir (Dunham, 2003; Tan ve diğ., 2006; Arabacı, 2007; Altıntaş, 2006; Özekeş, 2003).

Sınıflandırma. Nesneleri önceden tanımlı kategorilere atamaya dayalı, çok çeşitli uygulamaları barındırabilecek nitelikteki geniş ölçekli bir problemdir. Sınıflandırma, veriyi önceden tanımlı sınıflara veya gruplara eşlemek, veriyi sınıflara ya da kavramlara tanımlayacak ve ayırabilecek bir modelin ya da fonksiyonun bulunması sürecidir. Veri incelenmesi yapmadan mümkün olduğu kadar sınıflar karar verilirse, sınıflandırma denetimli öğrenim şeklinde adlandırılmaktadır. Sınıflandırma için veri üzerinde bilgiye ihtiyaç duyulur. Sınıflandırmada ihtiyaç duyulan parametreleri geliştirmede çoğunlukla eğitim seti kullanılır.

Sınıflandırma algoritması ve öğrenme verileri birlikte mevcut sınıflar tespit edilir ve bunlara girmede hangi niteliklerin gereli olduğu otomatik şekilde keşfedilerek daha önce belirlenen gruplardan birisine dâhil edilmesi sağlanabilmektedir. Test veri setiyle de öğrenme testi yapılarak çıkarımı sağlanan kurallar mümkün olan en çok sayıya ulaştırılır. Sınıflandırma algoritması, denetimli öğrenmede bir öğrenme yöntemi olup öğrenme ve test verilerini girdi ve çıktı olacak şekilde kullanır. Sınıflandırma yöntemlerinden istatistik temelli olanlar; diskriminant,

lineer regresyon, lojistik regresyon ve bayes yöntemleri, makine öğrenmesi temelli yöntemler ise karar ağaçları, yapay sinir ağları, destek vektör makineleri ve k-en yakın komşuluk yöntemidir.

Regresyon. Regresyon öngörü yöntemi olup tahmin edilen değişken süreklidir. Regresyonda veri öğesi, gerçek değerli tahmin edilen değişkene eşlemede kullanılır ve fonksiyonun öğrenilmesini içerir. Regresyonda amaç hedef verinin bilinen fonksiyonlardan birisine (lineer, lojistik) uymasını sağlamaktır. Regresyon ardından veriyi en iyi modelle yapan fonksiyona karar verir. Hangi fonksiyonun en iyi olduğu hata analizi ile karar verilir.

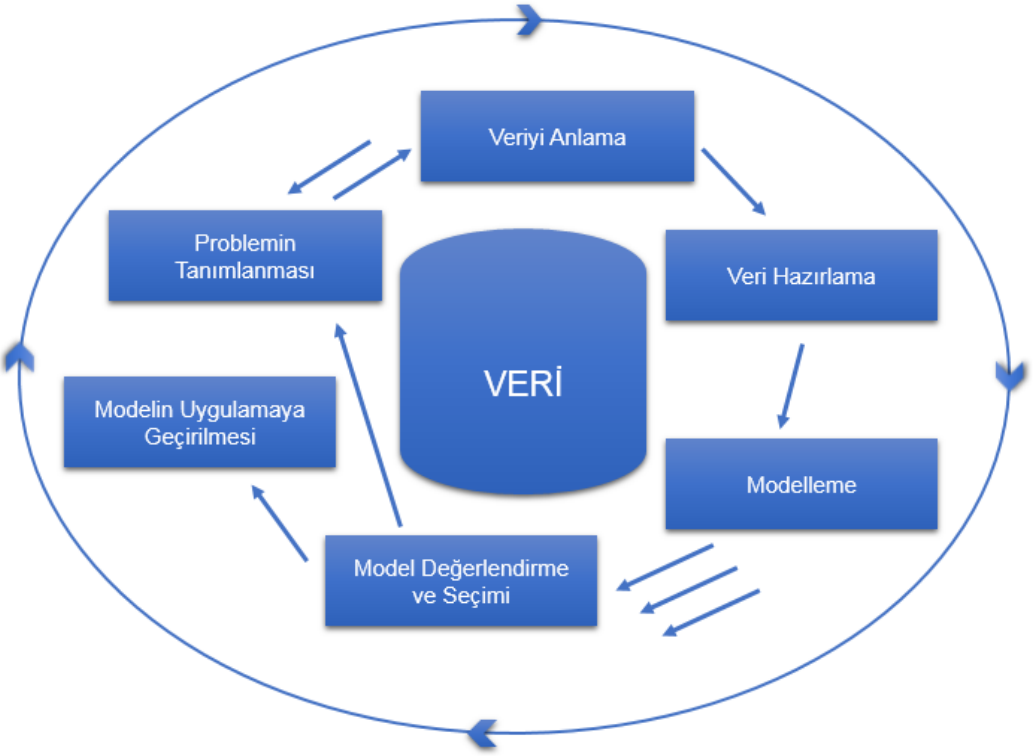
Kümeleme. Grupların (mümkün sınıf değerlerinin) önceden belirlenmesi haricinde sınıflandırmaya benzer ve bu grupların ne olacağı veride belli olur. Kümeleme denetimsiz öğrenim olup, veri setinin farklı gruplara bölünmesidir. Önceden belirlenen özelliklere göre veriler arasındaki benzerlikler ile karar verilerek kümeleme sağlanmaktadır. Verilerden en çok benzeyenler aynı kümede gruplanır. Kümelerin daha önce tanımlı olmaması sebebiyle ortaya çıkan kümelerin anlamı uzman görüşü ile yorumlanır. Kümeleme analizi ise nesnelerin bir altdizine gruplandırmasını sağlayan işlemdir. Bu şekilde nesneler, örnekleme yapılan grup özelliklerini iyi yansıtan temsil gücüne sahip olmaktadır. Sınıflamanın tersine, kümeleme yeniden tanımlanan sınıflara dayalı olmayıp denetimsiz öğrenme yöntemi olarak tanımlanır.

Birliktelik kuralları. Veri içindeki ilişkilerin açığa çıkarılması birliktelik kuralları olarak adlandırılır ve veri içindeki birliktelik ilişkisi ortaya çıkarılan modeldir. Bu yöntem genellikle satış sektörü alanında birlikte alınan ürünlerin belirlenmesinde kullanılmaktadır. Büyük hacimli verilerde birliktelik kuralını belirlemek, karar sürecinde işlemlerin daha etkili olmasını sağlamaktadır. Bu kurallar veri içindeki ilişkilerin nedensel açıklamasını sağlayamamakta ve gerçek dünyada veri yapısındaki ilişkiler temsil edilmemektedir. Bununla birlikte mevcut birlikteliklerin gelecekte de geçerli olacağı garanti edilmemektedir.

Veri madenciliği süreci. Veri Madenciliği çalışmalarında izlenmesi gereken birtakım temel adımlar bulunmaktadır. Veri madenciliğinde yazılım geliştiren birçok firma kullananlara yol göstermede bazı uygulama süreç modelleri önermektedir.

Modeller çoğunlukla kullanıcıları ardışık basamakların uygulanmasıyla hedefe götürmeyi amaçlamaktadır.

Bu kapsamda alanyazında en çok kullanılan uygulama süreçlerinden birisi CRISP-DM uygulama sürecidir. CRISP-DM bu alan uygulamaları için geliştirilmiş yaygın kullanılan bir modeldir ve kullanıcıların temel basamakları anlamasına yardım eden bir başlangıçtır. Uygulama süreci, gerekli görevleri ve bunlar arasındaki ilişkileri kapsamaktadır. CRISP-DM uygulama süreç adımları Şekil 3'te gösterilmiştir.



Şekil 3. CRISP modelinin adımları.

Diğer bir veri madenciliği uygulama süreci ise Two Crows şirketi tarafından oluşturulan süreçtir. Özellikle banka ve sigorta alanı, devlet uygulamaları, telekomünikasyon, bilgi sistemleri ve perakendecilik için uygulama adımları belirlenmiştir. Bu sürecin adımları aşağıdaki sıralanmaktadır.

1. Problemin tanımlanması
2. Veri madenciliği veri tabanın oluşturulması
3. Verinin incelemesi
4. Model için veri hazırlama

5. Modelin oluşturulması

6. Modelin değerlendirilmesi

7. Modelin uygulanması ve sonuçların izlenmesi

Eğitimsel Veri Madenciliği (EVM)

Her eğitim kademesinde öğrencilere ait kişisel bilgiler, ders durumları, notları gibi birçok bilgi büyük veri tabanlarında saklanmaktadır. Bu veri yığınlarından anlamlı ilişkiler araştırılabilir ve önemli bilgiler elde edilerek eğitimdeki aksaklıklara sebep veren problemler tespit edilebilir. Böylelikle eğitimin kalitesi de bu şekilde arttırılabilir. Veri madenciliği yoluyla eğitim alanında bu veriler analiz edilerek verilerin arasındaki örüntülerin keşfedilir. Özellikle eğitimdeki öğrencilerden, öğretim ortamlarından, öğretmenlerden, ölçme sonuçlarından sağlanan büyük hacimli veri yığınları düşünüldüğünde; bu verilerde saklı bilgi ve örüntünün ortaya çıkarılması ve bunların eğitimin kalitesi ile verimliliğinin artırılmasında kullanılmasının önemi her geçen gün artmaktadır. Bu kapsamda eğitimsel veri madenciliği yöntem ve algoritmaları, eğitimin etkililiğinin ve verimliliğinin artırılmasında gerekli bilgileri sağlamada kullanılabilir. Bu süreçte amaç “büyük hacimli” verilerdeki saklı-gizli ve kolay tespit edilemeyen bilgilerin ortaya çıkarılmasıdır.

Öğrenme ortamlarında sağlanan veri yığınındaki ilişkileri ve örüntüleri ortaya çıkarmak için yöntemler geliştirmek, bunları öğrenci ve onların öğrenme ortamlarını daha iyi anlamak için kullanan disiplin eğitim veri madenciliğidir (Siemens & Baker, 2012). Veri madenciliği yöntemleri, çeşitli eğitim ortamlarından gelen verilerden sonuçlar çıkarılması ve bilginin elde edilmesi ile ilgilenmektedir.

Veri madenciliği yöntemlerinin birçok sektörde yoğun olarak kullanılmasına rağmen, eğitim alanında kullanımı çok sınırlı kalmıştır. Aslında eğitim alanında derlenen verilerin kullanılması öğrenci başarısının arttırılmasında önemlidir. EVM sadece eğitim ortamlarından alınan verinin analizi ile bir takım anlamlı sonuçlar çıkarılmasını içermemekte, eğitmenlere; öğrenme verilerinin etkisi ve nasıl kullanılabileceği gösterilebilmekte, makro boyutta eğitim sürecine yardımcı karar destek ve tavsiye sistemlerinin geliştirilmesini sağlamaktadır. (Huebner, 2013). Ayrıca eğitim politikalarında gözden geçirilmesi ve iyileştirilmesi için de EVM

kullanılmaktadır. Eğitimde veri madencilik yöntemleri kullanılarak yapılan çalışmalar gittikçe artmaktadır.

Siemens ve Long (2011) eğitim alanındaki verilerin kullanılarak değerlendirilmesi ile ilgili aşağıda belirtilen konularda önemli iyileştirme çalışmalarının yapılabileceğini belirtmiştir.

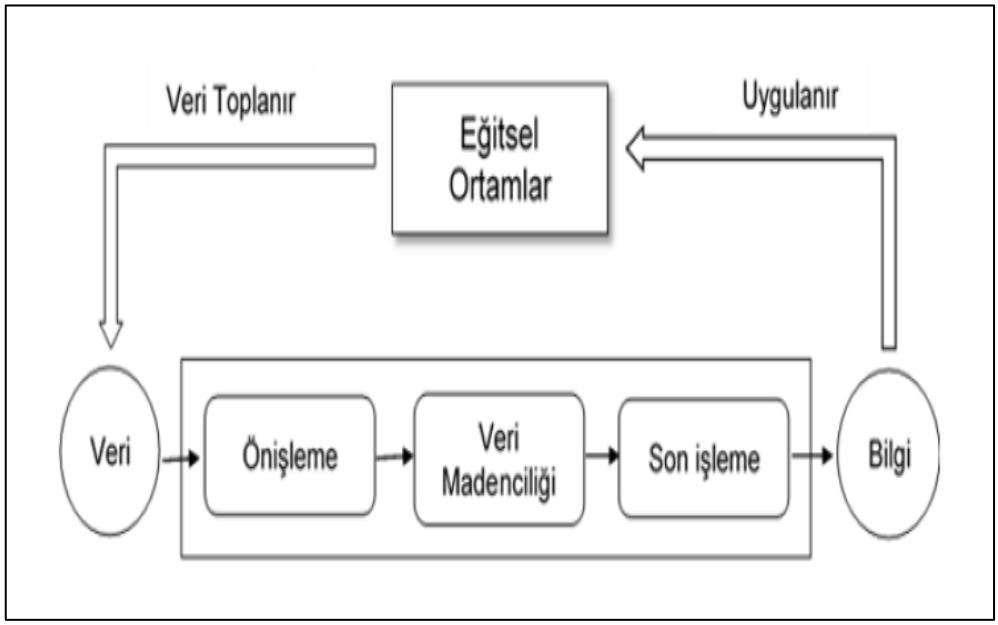
-Yönetsel karar-vermenin iyileştirilmesi,

-Eğitim kurumlarının başarılarının ve karşılaştıkları zorlukların anlaşılabilmesi,

-Çeşitli öğrenme zorlukları yaşayan bireylerin tespit edilmesi, nasıl destekleneceklerinin belirlenmesi ve bu sayede başarıda artışın sağlanması,

-Bireylerin öğrenme alışkanlıklarının tespit edilmesi ve bu alışkanlıkların iyileştirilmesi.

Eğitimsel veri madenciliği süreci. García ve diğ. (2011) tarafından Eğitimsel Veri Madenciliği, eğitsel sistemlerden sağlanan ham verinin eğitim yazılımlarının, öğretmen ve araştırmacıların kullanacağı bilgiye dönüştürme süreci olarak ifade edilmektedir ve Şekil 4'teki gibi belirtilmektedir.



Şekil 4. Eğitimsel veri madenciliği süreci (García ve diğ., 2011).

Eğitsel ortamlar. EVM için elde edilen verilerin büyük çoğunluğu bilgisayar ve ağ tabanlı sistemlerden elde edilen verilerdir. Ayrıca uyarlanabilir ortamlar,

çevrimiçi değerlendirme sistemleri, geleneksel sınıf ortamları, forumlar vb. ağ araçları da çalışmalarda veri sağlanan ortamlardır.

Veri kaynakları. EVM'de veri kaynakları gözlem verileri, ölçek verileri, loglar, davranışa ait veriler, etkileşim ve kontrollü deneylerden sağlanan verilerdir (Cristóbal, Sebastián, Mykola ve Ryan, 2010). Ayrıca son dönemde metrik veriler (göz hareketleri gibi) ya da öğrencilerin yazdığı mesaj ve sınav cevapları da EVM için veri kaynağı olarak kullanılmaktadır.

Ön işleme süreci. Eğitim veri madenciliğinde analize geçmeden önce en gerekli ve önemli aşama ön işlem sürecidir. Bu süreç, veri kalitesinin artırılması ve en iyi değişkenleri elde etmek için gereklidir. Ayrıca anlaşılabilir sonuçların üretilmesi açısından da önemlidir. Bu açıdan kalitesiz veriler olması sonuçların da kalitesiz çıkmasına sebep olabilir. Verilerin ön işlemden geçirilmesi veri kümesinin büyük olmasından dolayı en çok zaman alan aşamadır.

Veri temizleme. Bu adım ilgisiz veri silinmesi, kayıp veya aykırı verilerin tespitidir. Veri temizlemede eksik değerler tamamlanır, gürültülü veriler düzeltilir, aykırı veriler tanımlanır ya da çıkarılır, böylece verinin kalitesinin arttırılması sağlanır.

Veri birleştirme. Çalışmalarda farklı veri tabanlarından sağlanan verilerin bir araya getirilmesi gerekebilir. Bu adım verilerin tek bir ortamda birleştirilmesi, analiz edilmesi ve böylece hazır duruma getirilmesidir.

Veri dönüştürme. Veri madenciliği çalışmalarında ön işlem sürecinde sık yapılan işlemlerden birisidir. Orijinal verilerin içerdiği bilgiler madencilik algoritmalarına uygun veri olmayabilir. Bunların veri madencilik analizine uygun gerekli formlara dönüştürülmesi gerekir. Veri dönüştürme bir araya getirme, normalleştirme, düzeltme, özellik oluşturma ve genelleme işlemleri yapılarak sağlanır.

Özellik oluşturma. İlgisiz ve fazla değişkenlerin çıkarılarak analiz için kullanılması planlanan değişkenlerin seçilmesi süreci olarak tanımlanır (Piramuthu, 2004).

Model oluşturma-analizler. Ön işlemeden sağlanan verilere veri madenciliği tekniklerinin uygulandığı aşama olup en uygun veri madenciliği yöntemi belirlenir ve analizler yapılır. Veri madenciliğinde farklı görevler için birçok farklı algoritmayı

kullanılmaktadır. Algoritmaların veriyi incelemesi ve verinin özelliğine en uygun modelin belirlenmesi sağlanır. EVM araştırmalarında sınıflama, kümeleme, regresyon ve birliktelik kuralları en çok tercih edilen veri madenciliği yöntemlerindendir.

Sonuçların ve modelin değerlendirilmesi. Modelin oluşturulmasının ardından sonuçların değerlendirildiği ve yorumlandığı aşamadır. Bu süreç analizde sağlanan bilgi, model ve örüntülerin yorumlanıp karar sürecinde veya eğitim ortamlarının iyileştirilmesi için yorumlanma aşaması olarak tanımlanır (García ve diğ., 2011; Romero & Ventura, 2007). Ayrıca elde edilen modellerin yalnızca kendi içlerinde karşılaştırması değil, aynı zamanda tercih yapılan modelin kullanılması ile sağlanacak faydaların da karşılaştırılması gereklidir.

Modelin uygulanması. Modelin oluşturulup, geçerliliğinin kabul edilmesinin ardından uygulamaya geçilir. Seçilen modelin uygulanmasının ardından sistemin ne derece iyi çalıştığının belirlenmesi ve ölçülmesi gerekmektedir. Model çok iyi çalışsa da performansının daima izlenmesi önemlidir.

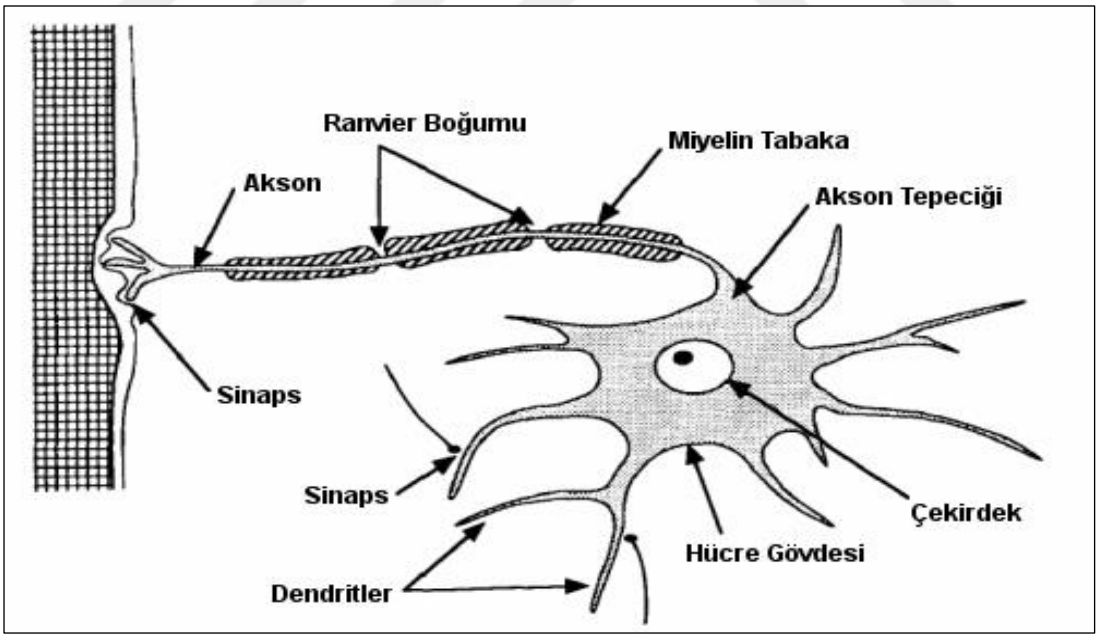
Yapay Sinir Ağları (YSA)

İnsanın öğrenme, düşünme ve yorumlama yeteneklerine makinelerinde sahip olması için insanın sahip olduğu sinir hücresi biyolojik formatı taklit edilerek bu yeteneklerin kazandırılması sağlanmaktadır. Yapay Sinir Ağları, biyolojik sinir sisteminden esinlenerek problemlere çözüm olabilecek modeller üretmeye yarayan bir yöntem, biyolojik sinir ağları gibi çalışarak belirli performans niteliklerine sahip bilgi işleme sistemidir. Böylece Yapay Sinir Ağlarındaki sistem, insan sinir ağı yapısıyla aynı şekilde çalışmaktadır. Yapay sinir ağları, insanın sahip olduğu öğrenme yolu ile yeni bilgiler yaratabilme, ortaya çıkarabilme ve keşfedebilme gibi yetenekleri dışarıdan yardım almaksızın otomatik şekilde gerçekleştirmek için geliştirilen bilgisayar sistemleri olarak bilinmektedir. YSA insanın öğrenmesinden temel alınarak oluşturulmuştur. Bu sebeple öğrenilecek konu ilgili bilgi ve verilere ihtiyaç duyulmaktadır. Bu süreç insanın deneyim kazanarak öğrenmesi ve çıkarımlarda bulunması, ardında da kararlar alma sürecine benzemektedir. Matematiksel bir model olan YSA, kümeleme, sınıflandırma, tahmin, regresyon ve örüntü problemlerinin çözümünde başarı ile uygulanmaktadır (Priddy ve Keller, 2005).

YSA alanyazında mühendislik ve fen bilimlerinde, tıp ve sosyal bilimlerde çok fazla kullanım sahası bulmaktadır. Bu alanlar endüstriyel alandaki uygulamalar, finansal uygulamalar, askeri ve savunma alanlarındaki uygulamalar, sağlık alanındaki uygulamalardır. Genel olarak durum sınıflandırma, sınıflama/kümeleme, fonksiyon yaklaşımı, tahmin, optimizasyon, veri ilişkilendirme ve kontrol amacıyla kullanılmaktadır.

Yapay sinir ağlarının paralellik, genelleme, öğrenme, doğrusal olmama, uyarlanabilirlik, bilginin saklanması, hata toleransı ve eksik verilerle çalışma olarak sıralanan özellikleri bulunmaktadır. Yapay sinir ağları genel olarak, sınıflandırma, tahmin, veri filtreleme, veri ilişkilendirme, tanıma, teşhis, yorumlama ve eşleştirme fonksiyonları gerçekleştirmek amacıyla uygulanmaktadır.

Yapay sinir hücresi. Biyolojik sinir ağlarında ana birim, biyolojik sinir hücreleridir. Sinir hücreleri beyin korteks kısmında yer alır ve yaklaşık sayısı 1011 olup, başka hücreler ile de ilişki içerisinde bulunmaktadır. Bir sinir ağı milyarca sinir hücresinin bir araya gelmesiyle oluşmaktadır ve birbirleriyle bağlanarak işlevlerini yerine getirmektedirler. Bir sinir hücresi kesiti Şekil 5'te görüldüğü gibidir.



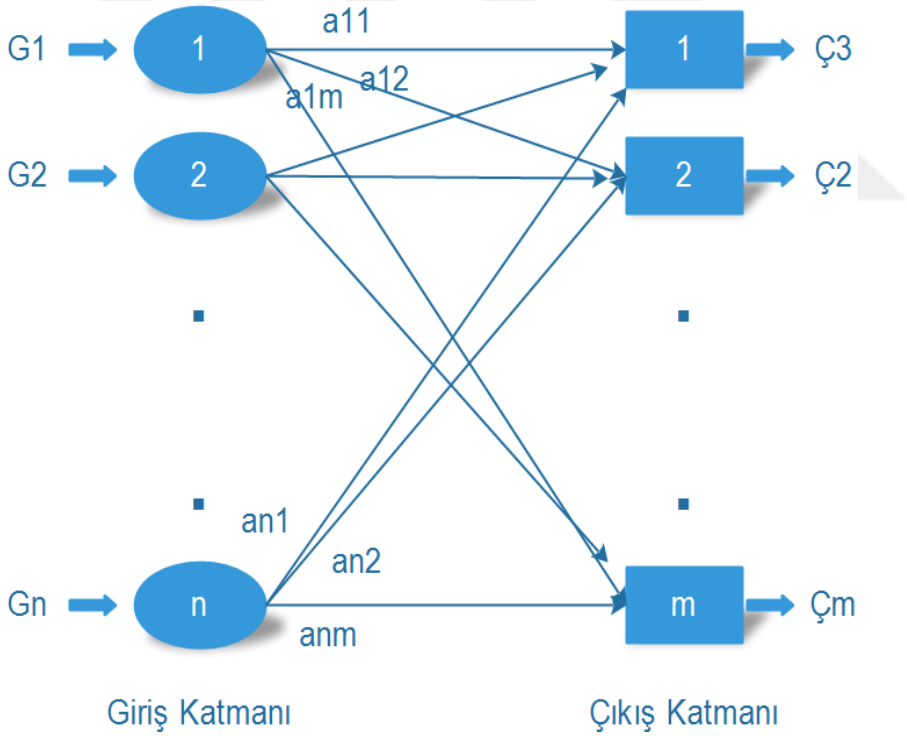
Şekil 5. Bir sinir hücresi kesiti (Fenokulu, 2021).

Sinir hücreleri genel olarak dendritler, soma, akson ve sinapslardan oluşmaktadır. Dışarıdan gelen uyarılar (sinyaller) sinir hücresinde görülen dendiritden hücrenin aksonuna doğru iletilmektedir. Bu iletim sırasında doğrusal

olmayan; karmaşık işlemler meydana gelmektedir. Bu işlemler sonrasında sinapsa ulaşan bilgiler (sinyaller) başka sinirlere sinaps sayesinde gönderilmektedir.

Diğer sinir hücresinde gelen elektrokimyasal sinyali somaya ulaştıran parçalar dentrit olarak adlandırılır. Soma ise sinir hücresinin gövdesidir. Alınan bilgilerin, birleştirilmek suretiyle anlamlandırıldığı, işlendiği ve sağlanan çıktıların aksona iletildiği mikro işlem parçasıdır. Akson da sinir hücresinin sağladığı elektrokimyasal sinyalleri diğer sinir hücrelerine ve çıktı elemanlarına gönderen birimdir. Sinapslar; parçası oldukları aksonun, diğer sinir hücrelerine ait dentritlere bağlanmasını sağlayan birleşme noktalarıdır (Öztemel, 2006; Burmaoğlu, 2009).

Yapay sinir ağlarını da aynı biyolojik sinir hücresi gibi yapay sinir hücreleri oluşturmaktadır. Bu ağların temel birimi olan yapay sinir hücresi, biyolojik sinir hücresinin 4 temel elemanı ifade etmektedir. Bir yapay sinir hücresinin genel gösterimi Şekil 6 'da görüldüğü gibidir.



Şekil 6. Yapay sinir hücresi (Şen, 2004).

Sinir ağlarındaki sinir hücreleri bir ya da birden çok girdi alır ancak bir çıktı verir. Çıktı girdi olarak ağın dışına veya başka yapay sinir hücresine de verilmektedir. Sinir hücresinin girdilerinin her biri bir bağlantı ağırlığıyla çarpıldıktan sonra toplama fonksiyonuna ulaştırılarak hesaplanmaktadır.

Girdi, ağa sunmak üzere dış ortamdan bilgiler sağlamaktadır. Bu girdiler sinir hücresinin öğrenmesi istenen örneklerden oluşmaktadır.

Ağırlıklar, hücrelere gelen bilginin önem derecesini ve hücre üzerindeki etkisini belirleyen değerlere denir. Yapay sinir ağlarında öğrenme, bu ağırlıkların değiştirilmesi ile oluşmaktadır.

Toplama Fonksiyonu, yapay sinir hücresine gelen net girdiyi hesaplamak için kullanılan fonksiyondur. Toplama fonksiyonu, hücreye gelen tüm girdilerin ağırlıklı toplamını hesaplamaktadır. Alanyazında en yaygın olarak kullanılan fonksiyon, girdilerin kendi ağırlıkları ile çarpımının toplamını ifade eden ağırlıklı toplamdır. Oluşturulacak sinir ağı için optimum toplama fonksiyonunu tespit etmede geliştirilmiş bir metot yoktur. Ancak çoğunlukla deneme yanılma yöntemi kullanılarak en optimum toplama fonksiyonu belirlenmektedir (Öztemel, 2006; Baş, 2006).

Aktivasyon fonksiyonu, toplama fonksiyonundan gelen ağırlıklı toplamı algoritmik bir süreç yardımıyla çıktıya dönüştürmektedir. Aktivasyon fonksiyonunda ağırlıklı toplam hücrenin çıktısını belirlemek için belirli bir eşik değer ile karşılaştırılmaktadır. Ağırlıklı toplam eşik değerden büyük ise bir sinyal üretmektedir. Ağırlıklı toplam eşik değerden büyük değilse işlem elemanı sinyal üretmeyecektir (Anderson ve McNeill, 1992).

Çıktı, bir işlemci elemanın birden fazla girdisi olmasına rağmen sadece tektir. Normal olarak bu çıktı doğrudan aktivasyon fonksiyonunun sonucuna eşittir. Bu değer ya dış ortama ya da başka bir yapay sinir hücresine girdi olarak gönderilmektedir (Anderson ve McNeill, 1992). Sonuç olarak yapay sinir hücresine gelen “girdiler”, bir “toplama fonksiyonu” ile “net girdiye” dönüştürülür ve “aktivasyon fonksiyonuna” gönderilir. Aktivasyon fonksiyonuna gelen net girdi işlenerek “çıktı” oluşur.

Yapay sinir ağının genel yapısı. Yapay Sinir Ağı birbirine hiyerarşik bağlı ve paralel çalışan yapay (suni) sinir hücrelerden oluşmaktadır. Bir sinir ağının en başlıca görevi ona sunulan girdi setine karşılık olacak çıktı seti belirlemektir. Bunun sağlanması için ağın, söz konusu olayın örnekleri ile eğitilmesi (öğrenme) suretiyle genelleme yapacak yeteneğe kavuşturulur. Bu şekilde sinir hücreleri bir araya gelip yapay sinir ağını oluşturmaktadır. Sinir hücrelerinin bir araya gelmesi rasgele

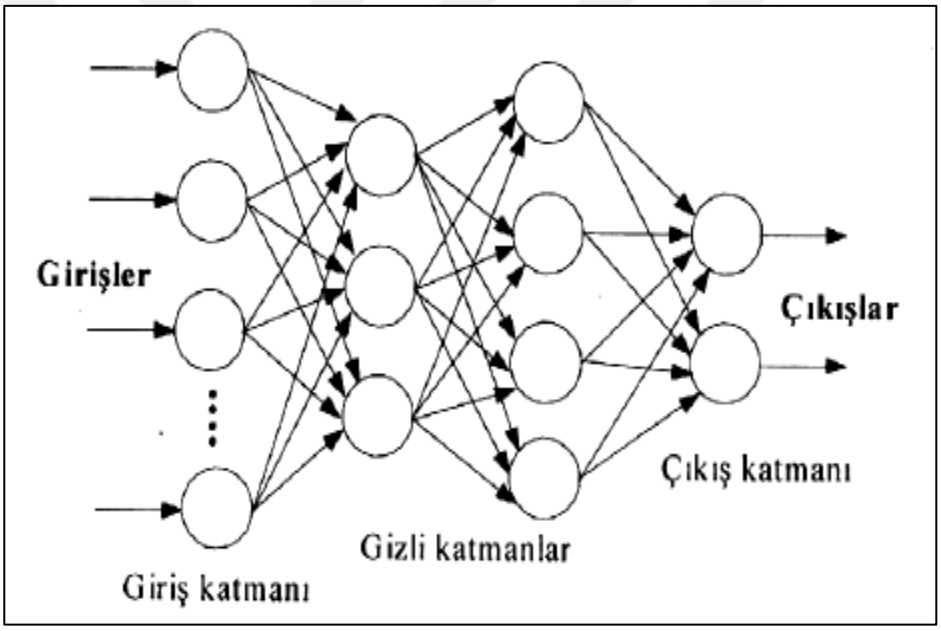
olmamaktadır. Hücreler 3 katman halindedir ve her bir katmanda paralel olarak bir araya gelerek ağı oluşturmaktadırlar. Bu katmanlar:

• Girdi katmanı: Bu katmandaki süreç elemanları dış dünyadan bilgileri alarak ara katmanlara transfer etmekle sorumludurlar.

• Ara katmanlar: Girdi katmanından gelen bilgiler işlenerek çıktı katmanına gönderirler. Bu bilgilerin işlenmesi ara katmanlarda gerçekleştirilir.

• Çıktı katmanı: Bu katmandaki süreç elemanları ara katmandan gelen bilgileri işleyerek ağın girdi katmanından sunulan girdi seti (örnek) için üretmesi gereken çıktıyı üretirler. Üretilen çıktı dış dünyaya gönderilir.

YSA'da katmanlardaki düğümler yalnızca kendisinden önceki katmandaki düğümlerden giriş almaktadır. YSA katmanları Şekil 7'de gösterilmektedir.



Şekil 7. YSA katmanları (Şen, 2004).

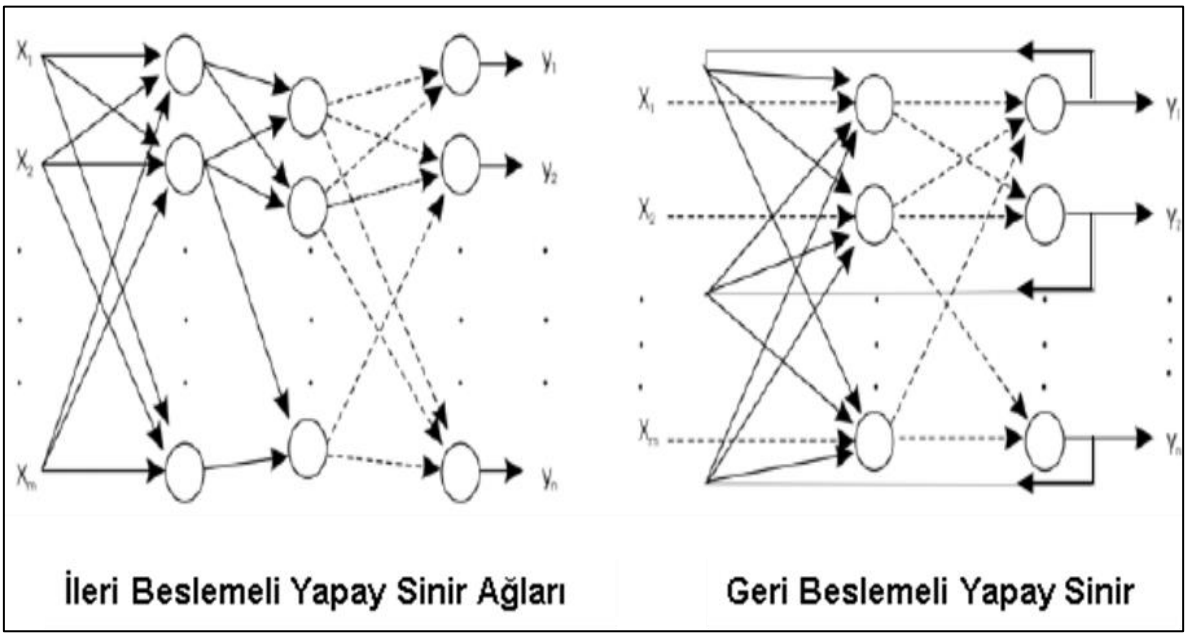
Yapay sinir ağlarının sınıflandırılması. Alanyazında pek çok yapay sinir ağı sınıflandırması mevcuttur. Bu modelleri içinde, bu çalışmadaki sınıflandırma problemleri için sıklıkla kullanılan model "Çok Katmanlı Algılayıcı (Multilayer Perceptron) Model"dir. Çok Katmanlı Algılayıcı Modeli yapısına göre iki grupta sınıflandırılmaktadır.

İleri beslemeli yapay sinir ağları. Bilgilerin girdi katmanından çıktı katmanına iletildiği, geri beslemenin olmadığı ağlardır ve çok katmanlı bir yapıya sahiptirler. Yapay sinir ağlarının çok katmanlı bu yapısı biyolojik sinir sisteminin

fonksiyonlarını yerine getirmesi amacıyla tasarlanmıştır. Girişler bir katmandan diğer katmana tek yönlü bağlantı ile gönderilmektedir. Bağlantının tek yönlü olmasından dolayı diğer katmandan sağlanan çıkışların geriye tekrar giriş olarak dönmesi mümkün değildir.

İleri beslemeli yapay sinir ağlarında bilgiler, bir ara katman aracılığıyla girdi katmanından çıktı katmanına iletilmektedir. Ara katmanın fonksiyonu girdi katmanından gelen bilgiyi işleme tabi tutmaktır. Ara katman gizli katman olarak da ifade edilebilmektedir. Çünkü ara katmandaki hücre sayısı kesin olarak bilinememektedir. Çok Katmanlı Algılayıcı Modelin çok yaygın olarak kullanılan bu çeşidi, bir girdi katmanı, bir veya birden çok gizli katman ve bir çıktı katmanından meydana gelir. Bu katmanların her birinde, bir veya birden fazla yapay sinir hücresi bulunabilir.

Geri beslemeli yapay sinir ağları. Geri beslemeli ağlarda nöronlar ileri beslemeli ağlardaki gibi katmanlarda bulunmaktadır. Fakat katmanlar arasında bulunan bağlantılar tek değil de çift yönlü bağlantıdır. Böylelikle ara veya çıkış katmanı tarafından sağlanan çıkışların önceki katmana giriş olacak şekilde gönderilebilmesi mümkündür. Geri besleme, bir katmandaki işlemci elemanlar arasında olduğu gibi katmanlar arsındaki işlemci elemanlar arasında da olabilmektedir.



Şekil 8. İleri ve geri beslemeli yapay sinir ağları (Yurtoğlu, 2005).

Dolayısıyla örnek bilgiler hem ileri yönde hem de geri yönde iletilebilmektedir. Geri beslemeli yapay sinir ağları sahip oldukları bu yapı sayesinde doğrusal olmayan dinamik davranış sergilerler. Böylelikle geri besleme türüne göre farklı yapı ve davranışta geri beslemeli yapay sinir ağları elde edilebilmektedir (Abraham, 2005).

Yapay sinir ağlarında öğrenme. Öğrenme yeteneğine sahip olması yapay sinir ağlarının en ayırt edici ve önemli özelliğidir. Yapay sinir ağı, belirli bir görevi yerine getirmek amacıyla uygun bir ağın tasarlanması ve eğitilmesiyle oluşturulmaktadır. Yapay sinir ağının tasarlanması kullanılacak katman sayısının, her katmanda kullanılacak hücre sayısının, katmanlar ve hücreler arasındaki bağlantı örüntüsünün, her bir hücrede kullanılacak olan düğüm fonksiyonunun ve ağın işlem modelinin belirlenmesini kapsamaktadır (Roy, 2000). Yapay sinir ağının eğitimi ise eşik değerlerin ve bağlantı ağırlıklarının belirlenmesini kapsamaktadır.

Farklı problemler için farklı ağlar geliştirilmiştir. Bu ağların eğitilmesinde de çok sayıda öğrenme kuralı uygulanmaktadır. Bu kurallar, ağların yapısına ve problemin türüne göre farklılık göstermektedirler. En eski ve en iyi bilinen öğrenme kuralı Hebb kuralıdır. Diğer bütün öğrenme kuralları Hebb kuralı temel alınarak geliştirilmiştir. Hebb'den sonra birçok araştırmacı, Hebb kurallarını esas alarak bir takım öğrenme kuralları geliştirmişlerdir. Bu öğrenme kuralları, denetimli ve denetimsiz olmak üzere iki grupta toplanmıştır. Ayrıca denetimli öğrenme yöntemine benzeyen pekiştirmeli öğrenme yöntemi de nadir de olsa kullanılmaktadır.

Denetimli öğrenme. Bu öğrenme yönteminde yapay sinir ağının kullanılmadan önce eğitilmesi gerekmektedir. Eğitim için giriş bilgileri ile birlikte çıkış bilgileri de verilir. Birçok uygulama için ağa gerçek örnek kümesi verilme zorunluluğu bulunmaktadır. Verilen örnek kümesi ile ağ eğitilir ve istenen istatistiksel doğruluk elde edildiğinde eğitme tamamlanmış olur. Böylelikle ağ kullanılmaya başladığında eğitim işlemi sonucunda elde edilen ağırlık değerleri genellikle sabit kalır, bir daha değiştirilmez. Burada temel olarak yapay sinir ağının ürettiği çıktı değerleri ile beklenen değerler birbiriyle karşılaştırılmaktadır. Bu yöntemde yapay sinir ağına ürettiği çıktının ne ölçüde yanlış veya doğru olduğunu söyleyen bir denetim olduğundan "denetimli öğrenme" olarak adlandırılmaktadır.

Denetimsiz öğrenme. Bu yöntemde, ağın öğrenmesine yardımcı olacak bir denetim yoktur. Ağa yalnızca girdi değerleri verilmektedir. Bu girdi değerlerine karşılık gelen çıktı değerleri için beklenen değerler bulunmamaktadır. Bir denetime ihtiyaç duyulmadığından kendiliğinden öğrenme gerçekleşir. Bu yöntem çoğunlukla sınıflandırmada kullanılır ve yapay sinir ağının öğrenme aşamasında "eğitim veri seti"nin özelliklerini keşfetmesi beklenir. Yapay sinir ağı bu özellikleri keşfederek, girdileri ayırt edici özelliklerine göre sınıflandırır. Bu noktada yapay sinir ağı, beklenen çıktı değerleri ile değil, ağa girilen verilerle öğrenmiş olur (Öztemel, 2006; Burmaoğlu, 2009).

Karar Ağaçları (KA)

Karar ağaçları, verileri farklı niteliklere göre, bulunduğu değerlerle sınıflandıran, sabit karar adımları kullanarak, veriyi küçük alt gruplara bölmede kullanılan ve bölme işlemlerinden sonra gruplarda oluşan öğeleri birbirine daha da benzer duruma getiren yöntem olarak ifade edilmektedir (Berry & Linoff, 2004; Sun & Li, 2008).

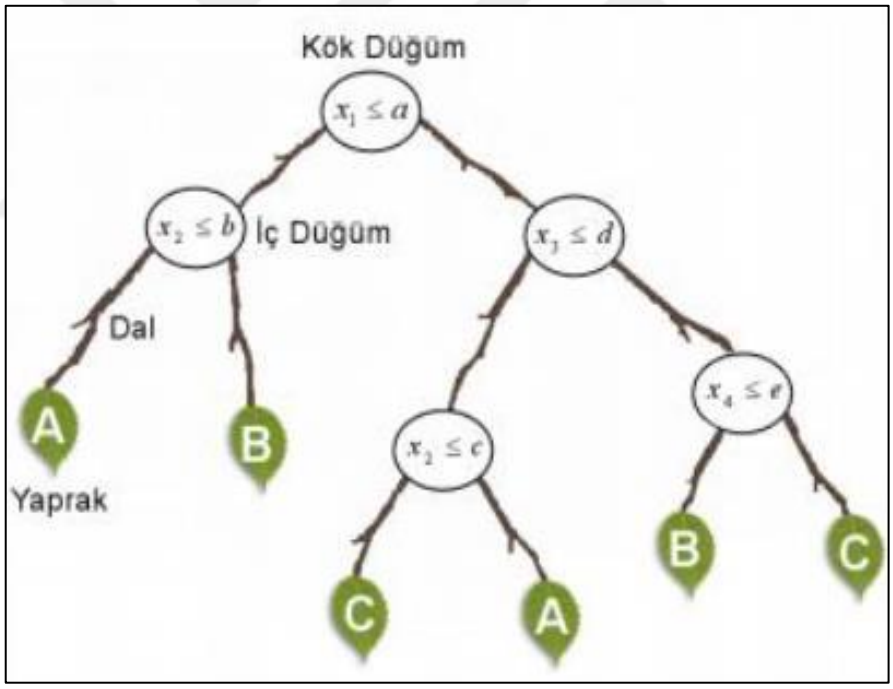
Karar ağacı istatistik, işletme, mühendislik, tıp, ekonometri, yönetim, muhasebe, bilişim, bankacılık gibi farklı alanlarda farklı uygulamalara sahiptir. Karar ağaçları kullanılarak farklı alanlarda gerçekleştirilen uygulamalar bir sınıf veya grubun üyeleri belirlemek, alt grupların ilişkilerini tespit etmek, kategorilere atamak, kategorileri birleştirmek ve sürekli değişkenlerin dönüştürülmesi, bir gruptan en önemli değişkenleri seçmek, vaka tahmininde kurallar oluşturmak, tıbbi verileri kullanarak en iyi kararları almak olarak sıralanabilir.

Karar Ağaçları, verilen bir problemin yapısına bağlı olarak bir ağaç yapısı şeklinde sınıflandırma ve regresyon modeli oluşturmaktadır. Kuralların anlaşılır oluşu metodun kullanımını basit ve uygulanabilir olmasını sağlamaktadır. Karar Ağacı, problem çözümü sırasında, karar verme işlemini, çok aşamalı ve sıralı işlemleri müteakip yapmaktadır. Karar ağaçları veri madenciliğindeki sınıflandırma modellerine nazaran oluşturulması ve anlaşılması kolay bir yöntemdir ve bu sebepten dolayı sınıflandırmada en çok tercih edilen modellerden birisidir.

Karar ağaçları, belli bir amaç doğrultusunda seçenekler ile olasılıklara bağlı durumları aşamalı olarak irdeler ve hiyerarşik yapılarla grafiksel bir gösterime sahiptir. Ardışık kararların ve durumların analizinde yaygın olarak kullanılan KA,

karar probleminde meydana gelebilecek tüm muhtemel senaryoların kararlarının ve rassal olayların sonuçlarını gösterir (Marshall, 1995). Karar ağacının en etkileyici tarafı, YSA aksine, KA kurallarını temsil etmesidir. Diğer bir deyişle, kuralları yorumlamak daha kolaydır.

Karar ağacı yapısı. Bir karar ağacı kök, dal ve yapraklardan meydana gelmektedir. Bir ağaç yapısı; bir kök düğümden, iç düğümlerden ve uç düğümlerden oluşmaktadır. Karar ağacının gövdesini düğümler oluşturmaktadır ve her karar düğümü dallara ayrılmaktadır, bu dallarda da yapraklar bulunmaktadır. Karar ağacı yapısı Şekil 9'da gösterilmiştir. Eğitilen bir ağacın tahmini gerçekleştirmesi kök, düğüm ve yapraklar aracılığıyla gerçekleşmektedir. Yaprak tahmin değerini içerir. Yukarıdan aşağıya her adımda düğüm değer belirleyicileri kontrol edilmelidir. Karar ağacı, kök düğümden başlayarak devam eden ve deney ünitelerine uygulanan evet-hayır gibi cevaplara göre oluşan yollardan oluşur.



Şekil 9. Karar ağacı yapısı (Silahtaroğlu, 2013).

Karar ağacı, "öğrenme" ve "sınıflandırma" olarak iki yapılandırılmış bir sistemde çalışmaktadır. Öğrenme aşamasında, model eğitme verilerinin, model oluşturmayı sağlamak için sınıflandırma algoritması ile analizi sağlanmaktadır. Öğrenilen model, sınıflama kuralları veya karar ağacı olarak belirtilir. Sınıflandırmada ise, sınıflandırma kuralının ya da karar ağacı doğruluğunun

belirlenmesi maksadıyla test verileri kullanılmaktadır. Analiz yapıldıktan sonraki doğruluk oranı uygun ise, kurallar sonraki verileri sınıflandırmak için kullanılır.

Karar ağacı algoritmaları. Karar ağacı yöntemlerinde cevap değişkenini açıklayıcı değişkenlere göre sınıflandırma ilkesine dayanan genel olarak farklı algoritmalar bulunsa da, bu yöntemlerin her biri farklı amaçlara göre çalışmaktadır. Algoritmalar dallanmanın nasıl olacağını tespit ederek, farklı yapılarda sınıflandırma olacaktır (Silahtaroğlu, 2013). Alanyazında bir sürü algoritma bulunmakla birlikte bu çalışmada C5.0, CART, CHAID ve QUEST algoritmaları ile analiz yapıldığı için bu dört algoritmadan bahsedilecektir.

C5.0. C5.0 algoritması en sık kullanılan algoritmalardan birisidir. ID3 kullanılarak elde edilmiştir ve C4.5 algoritması gelişmiş şeklidir (Hou ve diğerleri, 2014). C5.0 karar ağacı algoritması, C4.5 algoritmasına göre daha hızlı ve doğru sonuçlar vermektedir ve hafızayı daha verimli kullanmaktadır. Özellikle büyük veri setlerinde C5.0 algoritması başarılı ağaçlar oluşturmaktadır. C5.0 ve C4.5 algoritmaları sonuçları benzer olarak görülse de C5.0 algoritması biçimsel olarak daha iyi karar ağaçlarına sahip olmamızı sağlamaktadır.

C5.0'te türetme algoritması, bir düğümle başlar, ardından en uygun sınıflayıcı belirlenmesi için "bilgi kazanımı" olarak adlandırılan entropi tabanlı ölçü kullanmaktadır. Karar ağaçlarını en asgariye indirgemek ve karar kurallarını oluşturma bakımından daha iyi sonuçlar sağlamaktadır.

***C&RT* (classification & regression trees).** Breiman, Friedman, Olshen, Stone tarafından 1984 yılında geliştirilen C&RT algoritmasında, her seviyede grubun kendinden daha benzer 2 alt gruba bölünmesi sağlanır (Akpınar, 2000). C&RT, kümedeki verileri daha homojen olacak biçimde alt kümelere bölmektedir ve bu işlemi homojenliğe varılıncaya kadar yinelemeli şekilde tekrarlayan sınıflandırma algoritmasıdır.

C&RT algoritması, sürekli ve kategorik verileri kullanabilmektedir. CART, ağaç modelinde ayırma kriterini hesaplama sırasında eksik gözlemleri önemsemeyen bir algoritmadır. Yordanan değişkenin kategorik ise olduğunda yöntem "sınıflandırma ağaçları", sürekli ise "regresyon ağaçları" olarak adlandırılmaktadır.

C&RT verileri her defasında iki alt kümeye ayırır. Maksimum büyüklükteki ağacı türetmeye çalışır ve böylece saf, en iyi bölünmeyi sağlamaya çalışır. Bu işlemi yaparken ağaç sürekli şekilde bölünerek büyümektedir ve yeni bölünme oluşuncaya kadar bu bölünmeye devam edilmektedir (Sezer ve diğerleri, 2010).

CHAID. İlk defa Kass (1980) tarafından önerilmiş CHAID (CH-i-squared Automatic Interaction Detection), entropi ya da Gini yerine Ki-kare kullanmaktadır. Veriyi daha benzer alt gruplara bölmek en önemli amaçtır. Hedef değişken ölçüm seviyesine göre kullanılan istatistiksel test değişmektedir. Hedef değişken sürekli ise F testi, kategorik olması durumunda Ki-Kare testi uygulanmaktadır. Bu algoritmayı diğer algoritmalardan ayıran özellik ise "C4.5 ve C&RT algoritmaları" ikili dallanma üretirken, bu algoritma çoklu dallanma üretebilmesidir. Bu algoritmada yordanan değişken kategoriktir (niteliksel, sıralı, sınıflayıcı). Yordanan değişken sürekli olduğunda ise bu değişken sıralı değişkene dönüştürülür (Kass, 1980). CHAID algoritması ilişki düzeyine göre farklılık bulunan grupları ayrı sınıflandırmaktadır. Karar ağacı yaprakları, ikili dallanma yapmayıp veride bulunan farklı unsur sayısına göre dallanmaktadır (Altan ve ark., 2015).

QUEST. QUEST (Quick, Unbiased, Efficient, Statistical Tree) algoritması tek değişkenli ve doğrusal bileşimli bölmeleri/sınıflandırmaları destekler (Loh ve Shih, 1997). QUEST hızlı ve yansız istatistiksel ağaç şeklinde tanımlanmaktadır. C&RT algoritması gibi ikili ağaç oluşturmak üzerine kurulan bir yapıda bulunmaktadır. Bu algoritma C&RT'a göre daha yüksek bir verimlilik sağladığı ve C&RT'ın dezavantajlarını ortaya çıkarttığı gözlenmiştir. Bu algoritma, ikili karar ağaçlarında olduğu gibi; bölme, durdurma ve budama gibi işlemlere izin verir (Loh ve Shih, 1997; Lim ve ark., 2000). QUEST algoritmasında değişkenler açısından yordanan tek ve sınıflı değiken, yordayıcı ise bir veya daha çok sayıda olmakta ve böyle sürekli, sıralı ve sınıflı yapıda olmaktadır. QUEST, eğer yordayıcı değişken kategorik olduğunda, büyük hacimli ve karmaşık yapıda veri seti olduğunda ve ağacın ikili bölünmeyle sınırlandırılması söz konusu olduğunda tercih edilmektedir.

K-En Yakın Komşu Yöntemi (K-Nearest Neighbors -KNN)

1951 yılında örüntü tanımada kullanılmak için Fix ve Hodges (1951) tarafından nonparametrik yöntem olarak meydana çıkan K-En Yakın Komşuluk yöntemi daha sonra Cover ve Hart (1967) tarafından geliştirilmiştir. Bu yöntem

mesafeye dayalı olarak sınıflandırma yapan, en kolay ve yorumlanması basit denetimli makine öğrenmesi olarak tanımlanmaktadır. K-En Yakın Komşuluk Algoritması (K-NN), yapılmasının basit ve kolay olmasından, öğrenme sürecinin ise güçlü ve kullanışlı olması sebebiyle sınıflandırmada sık olarak kullanılmaktadır.

Bu yöntem, en yakın komşuyu ve gözlemlerin bulunacağı sınıfı, k-değerini baz alarak belirleyen sınıflama yöntemi olarak tanımlanmaktadır. Gözlemler ya da nesneler arasında bulunan uzaklık esas alınarak sınıflandırma yapan algoritmalardan biridir. Makine öğrenmesi, veri madenciliği gibi çok çeşitli alanlarda uygulanmaktadır. K-En Yakın Komşuluk algoritması, eğitime ihtiyaç duyulmaması, oluşturulmasının kolay olması, analitik izlenebilmesi, yerel bilgilere adapte edilebilir, gürültülü eğitim veril setlerine dirençli olması vb. avantajları sayesinde sınıflandırmada özellikle tercih edilen algoritmalardan birisidir (Bhatia, 2010).

K-En Yakın Komşuluk algoritması amacı; bireyleri veya nesneleri, sahip oldukları özelliklerden faydalanarak, önceden tespit edilen sınıflara ya da gruplara en doğru olarak atama yapmaktır. K-En Yakın Komşuluk algoritması bununla birlikte yeni gözlemin de sınıflandırılmasını sağlamaktadır. Sınıflandırılacak gözlem, öğrenme verileri kullanılarak, en yakındaki k tane gözlemden en çok benzeyenlerle aynı veri setinde sınıflandırılacak şekilde ayarlanır. Bir modelin oluşumunda kullanılacak olan verilerin oluşturduğu veri setine öğrenme veri seti denilmektedir (Fix ve Hodges, 1951; Cover ve Hart, 1967; Harrington, 2012).

K-En Yakın Komşuluk algoritması, örnek tabanlı öğrenme algoritmaları arasında yer almaktadır. Örnek tabanlı algoritmalarda, öğrenim süreci eğitim setlerinde bulunan verilere göre gerçekleştirilir. İlk defa karşılaşılan örnek, eğitim veri setindeki örnekler ile arasında bulunan benzerliğe göre sınıflandırılma yapılmaktadır. K-En Yakın Komşuluk algoritmasında, eğitim verilerindeki örnekler n boyutlu sayısal niteliklerle gösterilmektedir. Her bir örnek n boyutlu uzayda bir noktayı belirtecek şekilde tüm eğitim verilerinin örnekleri n boyutlu örnek uzayda tutulmaktadır. Bilinmeyen örnekle karşılaşılması durumunda, eğitim setinden söz konusu örneğe en yakın k adet örnek tespit edilerek yeni örneğin sınıfı, k-en yakın komşusu sınıf etiketi çoğunluk oylamasına göre atanmaktadır (Kamber ve Pei, 2006).

K-En Yakın Komşuluk algoritmasında bütün örneklemler örüntü uzayında bulunmaktadır. Bu algoritma, aşina olunmayan örneklemin hangi sınıfta olduğunu tespit etmek amacıyla örüntü uzayını araştırır, böylelikle bilinmeyen örnekleme en yakın k örneklemi bulunur. En yakın komşular belirlenirken, seçilen örnek ile eğitim setindeki örnekler arasındaki mesafe ölçülür. Uzaklık hesabı tespitinde korelasyon, öklit, kosinüs ve hamming kullanılır. Ardından, bilinmeyen örneklem, yakın komşulardan en çok benzeyen sınıfa atanmaktadır K-En Yakın Komşuluk algoritması, ayrıca bilinmeyen bir örneklem için gerçek değerin tahmin edilmesinde de uygulanabilir.

Bu yöntemdeki performansı iki önemli öğe belirlemektedir. Birinci husus test örneğine en yakındaki eğitim örnekleri hesabındaki uzaklık hesaplama metodudur. İkinci husus ise tespit edilecek k parametresi değeridir (Aydemir, 2013).

K-en yakın komşu yöntemi avantaj ve dezavantajları. Bu yöntemin avantajları ve dezavantajları şu şekilde sıralanabilir. Avantajları; K-en yakın komşuluk yönteminde komşunun ortalaması alınması sebebiyle gürültüye sahio verilerden daha az etkilenmektedir, uygulaması ve anlaşılması kolaydır. Dezavantajları ise en yakın komşu miktarı k parametresinden, belirlenen uzaklık kriterinden fazla etkilenmesidir ve model eğitme verisi çok olduğunda verimli olmasıdır.

K-en yakın komşu yöntemi adımları. Bu yöntem gözlemleri diğer olgulara benzerliklerine göre sınıflandırmada kullanılır. Benzer gözlemler birbirine komşu, benzer olmayan gözlemler birbirinden uzaktır. İki gözlem arasındaki uzaklık, birbirine aykırılığı belirleyen bir kriterdir. Yeni bir gözlemin modeldeki gözlemlerden uzaklıkları hesaplanır. Bu gözlem, en fazla tekrar eden/benzer kategoriye atanır (Fix ve Hodges, 1951; Cover ve Hart, 1967).

K-en yakın komşu algoritması, isteğe bağlı olarak veriyi; “öğrenme” ve “test” veri seti olmak üzere ikiye ayrılmaktadır. Modelin oluşumunda öğrenme veri seti, modelin bağımsız olarak değerlendirilmesinde ise test veri seti kullanılır. Kategorik değişkenlerde eksik gözlemlerin değerlendirilmesi seçeneği bulunurken, sürekli değişkenlerdeki eksik gözlemler ise göz ardı edilebilir. Benzer kategoriler birleştirilerek veya model uygulanmadan önce az gözlenen kategoriler çıkarılarak kategori sayısı azaltılabilir. Ayrıca, aykırı gözlemler de modelden çıkarılabilir

(Fix ve Hodges, 1951; Cover ve Hart, 1967). K-en yakın komşu algoritması, sürekli yapıdaki gözlemleri sınıflandırmak için kullanıldığında, en yakın komşuların yaklaşık bir değeri yeni bir gözlemin sınıflandırma tahmininde kullanılır. K-en yakın komşu algoritması isteğe bağlı olarak değişkenleri yeniden ölçeklendirerek (normalize ederek) modeli eğitmeden önce ölçek süreklilik tahmini yapar.

K-en yakın komşu algoritmasında algoritmaya değişken seçiminde ileriye doğru (forward) seçim yöntemi kullanılmaktadır. Değişkenler sırayla seçilir ve her adımda seçilen değişken, hata oranının veya hata kareler toplamının minimum olmasını sağlayan değişkendir (Cover ve Hart, 1967; Cunningham ve Delany, 2007). Modele yeni bir değişken eklenmesiyle modelin daha fazla geliştirilemeyeceği anlaşıldığı durumda algoritma durur (Cover ve Hart, 1967).

Naive Bayes Yöntemi

İstatistiksel sınıflandırıcı olarak görülen Bayes yöntemi, verilerin üyelik olasılıklarını yani belirli bir sınıfa ait olma olasılıklarını tahmin etmektedir. Temeli istatistiksel Bayes teorisine dayanır ve sınıflandırma yapacak durumları birbirinden bağımsız şekilde ele alır. Bu yöntem, bağımsız değişkenler ile hedef değişken arasındaki ilişkiyi analiz ederek tahmin ve tanımlama yapan bir sınıflama algoritmasıdır.

Naive Bayes algoritmasında her kriterin sonuca olan etkilerinin olasılık olarak hesaplanması temeline dayanmaktadır. Naive Bayes yönteminde sınıflandırma yapılırken özellik değerleri birbirinden bağımsız olarak hesaba katılmaktadır. Veri setindeki örnekler üzerinden gerçekleştirilen olasılık hesapları yardımıyla veri dosyasında daha önce yer almayan yeni örneğin daha önce elde edilmiş olasılık değerlerine göre hangi sınıfta yer alacağı tespit edilmeye çalışılır.

Olasılık temelli sınıflandırma olan Navie Bayes, bir verinin sınıfının olasılığını tahmin etmek için verilen bu verideki herhangi bir bilginin sınıfının koşullu olasılıklarını kullanmaktır. Örnek olarak belge sınıflandırmada bu yöntem sıklıkla tercih edilir. Bu yaklaşımın “Naive” kısmı içindeki kelime bağımsızlığı varsayımından kaynaklanmaktadır. Çünkü kelime kombinasyonlarının olasılıklarını tahminci olarak kullanmaz. Bu sebeple Karar Ağacı gibi algoritmalardan daha verimli bir yaklaşımdır. Genellikle sonuçları Yapay Sinir Ağları ve Karar Ağacı ile

karşılaştırılmaktadır. Bayes Teoremi, olasılıkların doğruluk oranını bulunması için uygulanmaktadır (Abdullahi, 2018).

Naive Bayes, modelin öğrenilmesi esnasında, her çıktının öğrenme kümesinde kaç kere meydana geldiğini hesaplar. Bulunan bu değer, öncelikli olasılık olarak adlandırılmıştır. Naive Bayes aynı zamanda her bağımsız değişken/bağımlı değişken kombinasyonunun meydana gelme sıklığını bulur. Bu sıklıklar öncelikli olasılıklarla birleştirilmek suretiyle tahminde kullanılır (Hudairy, 2004).

Olasılık teorisinde incelenen bir olay olarak, koşullu bir olayın olasılık değeri (B olayı bilindiğinde A olayı), B olayı için olasılık değerinden koşullu olarak farklıdır (A olayı bilindiğinde B olayıdır). Ancak, bu iki farklı koşul arasında belirli bir ilişki vardır.

Naive Bayes; tahmin edici ve tanımlayıcı sınıflama yöntemidir. Makine öğrenmesi ve veri madenciliği konusunda en etkili tümevarımsal öğrenme yöntemlerinden biri olarak görülmektedir. Bağımsızlık varsayımı realitede çok seyrek ortaya çıksa da, bu varsayım genellikle gerçek olmasa da, bu yöntemin sınıflandırma yapmadaki başarısı ve diğer birçok sınıflama yöntemlerinden üstünlüğü farklı çalışmalarda ortaya çıkmaktadır.

Sınıflandırma algoritmalarının karşılaştırılması çalışmalarında, Bayes algoritması performansı karar ağaçları ve yapay sinir ağları algoritmalarının performanslarıyla karşılaştırılmaktadır. Bayes'in sınıflandırıcı büyük veri tabanları üzerinde uygulandığında yüksek doğruluk ve hız sergilediği görülmüştür.

İki Aşamalı Kümeleme (Two-Step Cluster) Analizi

Alanyazın incelendiğinde kümeleme yöntemlerinin genel anlamda hiyerarşik (hierarchical), bölümlemeli (partitional), yoğunluk (density-based) ve ızgara (grid-based) tabanlı olarak dörde ayrıldığı görülmektedir. Bununla birlikte bunların dışında pek çok farklı yöntem ve algoritma da geliştirilmiştir. Bunlara örnek olarak öngörülen kümeleme, iki modlu kümeleme yöntemleri örnek olarak verilebilir.

İki Aşamalı Kümeleme yöntemi hem sayısal hem de kategorik verileri kümeleyebilen, çok büyük boyutlu veri setlerinde etkili olarak çalışabilen bir algoritma olup BIRCH algoritmasına benzemektedir. Hem kategorik hem de sürekli değişkenlerin uzaklıkları İki Aşamalı Kümelemede model tabanlı uzaklık ölçüsü

kullanılarak ölçülebilmektedir. Küme sayısı, otomatik olarak algoritma aracılığıyla hesaplanabileceği gibi kullanıcı tarafından da girilebilmektedir. Bu algoritmada, veri yapısına uygun olarak bölümlendirilmiş veriler kullanılabilmekte ve sayısal alanlar standartlaştırılabilmektedir. Uzaklık ölçüsü olarak olasılık temelli Log-likelihood ya da Öklid uzaklığı seçilebilmektedir. Kümeleme kriteri olarak BIC (Schwarz's Bayesian Criterion) ya da AIC (Akaike's Information Criterion) seçilebilmektedir. Bu seçimler yapıldıktan sonra algoritma çalıştırıldığında kısa bir süre içerisinde küme sayısı, kümelerin boyutları, dağılımları ve özellikleri elde edilmektedir.

İki Aşamalı Kümeleme yöntemi ön kümeleme aşaması ve kümeleme aşaması olmak üzere iki adımdan meydana gelmektedir. İlk aşama olan ön kümeleme aşamasında küçük alt kümeler halinde ilk kümeleme işlemi gerçekleştirilir. Bu alt kümeler ayrı gözlemler olarak ele alınmaktadır. Bu aşama önceden oluşturulan kümelenme ile gözlemin birleştirilip birleştirilmediği ya da yeniden bir kümelenme meydana getirilip getirilmeyeceğinin kararının verildiği aşamadır. Uzaklık kriteri göz önünde bulundurularak bu yeni gözlemlerin gruplandırılması işlemi hiyerarşik kümeleme yöntemiyle gerçekleştirilir. İkinci adım, ön kümeleme sonucu elde edilen alt kümelerin analizin temelini oluşturduğu ve alt kümelerin gerekli sayıda kümeye ayrıldığı yönlendirme işlemidir. Alt kümelerin sayısı gözlem sayısından önemli ölçüde daha küçük olduğu için, geleneksel gruplama yöntemlerinin kullanımı daha kolaydır. Alt küme sayısı ne kadar fazla ise, yöntem o derece hassastır (Cameron ve Miller, 2015; Zhang ve ark. 1996).

İki aşamalı kümeleme algoritmasında Minkowski, Karl-Pearson (Standartlaştırılmış öklid uzaklığı), Chebyshev, Log-likelihood ve Öklid gibi birçok uzaklık ölçüsü kullanılabilmektedir. Bununla birlikte en çok tercih edilen uzaklık ölçüleri ise Log-likelihood ve Öklid'dir. İki küme arasında bulunan Öklid uzaklığı, kümelerin merkezleri esas alınarak hesaplanır.

İlgili Araştırmalar

Subbanarasimha, Arinzeb, & Anandarajanb (2000) tarafından yapılan çalışmada öğrenci akademik performansını tahmin etmek için iki farklı küme kullanılmıştır. Veri madenciliği tekniklerinden yapay sinir ağları ile regresyon karşılaştırılmıştır. Çalışmada yapay sinir ağları yönteminin regresyona göre daha yüksek doğru tahmin etme yüzdesine sahip olduğu ortaya konmuştur.

Aydın (2007) yaptığı çalışmada Uzaktan Eğitim Sistemi planlamasına fayda sağlayan öğrencilerin performans tahminine yönelik model geliştirmiş ve mezunların profilini belirlemek için kümeleme çalışması yapmıştır. Çalışmasında C5.0 karar ağacı algoritmasını kullanan tahmin modeli uygulamıştır. Mezun olan öğrencilere ilişkin "K-means" algoritması ile 5 küme elde etmiştir. Sonuç olarak kümeleme analizi ile sağlanan bilgilerin öğrenci başarısı ve bilgisayar kullanımı arasındaki ilişkiyi doğruladığı görülmüştür.

Tosun (2007) çalışmasında öğrenci başarısı ile ilgili veri madenciliği yöntemlerinden karar ağacı ve yapay sinir ağlarını karşılaştırmıştır. Çalışma sonucunda yapay sinir ağları doğru sınıflandırma başarı oranı yaklaşık %92, karar ağaçları başarı oranı ise %86 olarak bulunmuştur.

İbrahim ve Ruslı (2007) çalışmalarında öğrenci başarılarını tahmin ederken karar ağaçlarını, yapay sinir ağlarını ve doğrusal regresyonu kullanmış ve bu yöntemleri karşılaştırmıştır. Akademik başarının genel tahmin analizinde yapay sinir ağlarının daha yüksek sonuçlar gösterdiği görülmüştür.

Oladokun, Adebanjo ve Charles-Owaba (2008) tarafından yapılan çalışmada öğrenci başarısındaki etkili değişkenlerin belirlenmesi ve öğrenci performansının tahmininde yapay sinir ağlarının test edilmesi amaçlanmıştır. Modele dahil edilecek yeni öğrencinin ilerideki başarısının tahmin edilmesine yönelik, yapay sinir ağları doğru tahmin etme oranı %74 olarak bulunmuştur.

Kelley-Winstead (2010) çalışmasında sınıf tekrarı yapacak öğrencileri tahmin etmiştir. Çalışmada sınıf tekrarı yapan 1570 öğrenci dahil toplam 10140 öğrenci kullanılmıştır. Çalışmada sınıf tekrar etmede etkili olabilecek ailesel geçmiş, sosyo demografik ve okula ait faktörler ortaya konmuştur. Çalışmada lojistik regresyon ve karar ağaçları kullanılmış, analiz sonucunda karar ağaçlarının daha yüksek doğru sınıflandırma oranına sahip olduğu gözlenmiştir.

Tsai, Tsai, Hung ve Hwang (2011) tarafından yapılan çalışmada yeterlilik sınavından başarısız olacak öğrenciler tahmin edilmiştir. Araştırma Tayvan'daki bir üniversitede yürütülmüştür. Öğrencileri kümelere ayırmak için iki aşamalı kümeleme, k-ortalamalar, öz düzenleme haritaları teknikleri uygulanmıştır. En iyi kümeleme bulunmasının ardından karar ağacı en yüksek sonucu veren algoritma olarak bulunmuştur.

Bhardwaj ve Pal (2011) tarafından yapılan çalışmada, öğrenci bilgileri kullanılarak bilgisayar uygulamaları dersi başarıları veri madenciliği sınıflandırma yöntemleri ile tahmin edilmiştir. Çalışmada Bilgisayar Uygulamaları Bölümü'nde öğrenim gören 300 öğrenci verileri, sınıflandırmada yöntem ise Bayesian kullanılmıştır. Araştırmanın sonunda akademik performansların yalnızca kendi çalışmalarına bağlı kalmadığı ortaya konulmuştur. Bu araştırmada öğrenci ders başarısında eğitim ortamı, öğrenci alışkanlıkları, yaşam yeri, aile yıllık geliri, anne nitelikleri, aile statüsü etkili olan değişkenler olmuştur.

Şen, Uçar ve Delen (2012) çalışmalarında orta öğretime yerleştirme testi puanlarını tahmini için model geliştirmiş ve başarıda etkili olabilecek faktörleri tespit etmişlerdir. Bu araştırmada Türkiye'den orta öğretim geçiş sınavından büyük veri seti uygulanmıştır. Veri madenciliği karar ağacı algoritmalarından C5.0 en iyi tahmin eden algoritma olmuştur. C5.0 algoritmasının doğru tahmin etme oranı ardından Destek Vektör Makineleri ve Yapay Sinir Ağları gelmiştir. Araştırmada tahmin etmede kullanılan genel nitelikler: özür durumu, cinsiyet, burs durumu, özel ders durumu, kardeş sayısı, ebeveynlerin yaşama/boşanma durumu, özel/devlet okulu durumu, çalışma durumu olarak belirlenmiştir. Çalışma sonucunda kardeş sayısının, öğrencinin burs durumunun, daha önceki test deneyiminin bir sene önceki yılın ortalama puanını tahmin etmeyi etkileyen en önemli özelliklerden olduğu ortaya konmuştur. Ebeveynlerin evlilik durumunun, cinsiyetin ve çalışma durumunun öteki özellikler kadar önemli olmadığı belirtilmiştir. Bu özelliklerin ortaya çıkarılmasının başarıyı artırmada öğrencilere, öğretmenlere ve ailelere yardımcı olacağı önerilmiştir. Bu tarz analizlerin, standart yapılan okul giriş test yapısının anlaşılması ve daha eşit değerlendirme araçları ortaya koyması açısından faydalı olacağı ifade edilmiştir.

Olgun ve Özdemir (2012) istatistiksel süreç kontrolünde kullanılan Shewhart kontrol grafikleri ile ilgili yaptıkları çalışmalarında zaman içinde meydana gelebilecek değişimlerin tespiti, süreç kontrolü ve önlem almak için süreç içerisinde anormal değişim örüntülerinin tanımlanmasına ilişkin Yapay Sinir Ağları ve Bayes sistemleri oluşturmuşlardır. Meydana getirilen örüntü tanımlayıcıların sınıflandırma performanslarını da ölçmüşlerdir. Doğru sınıflandırma performansının artırılması amacıyla örüntü gözlem değerlerinden, 6 istatistiksel özellik oluşturulmuş ve bunların sınıflandırma performansları karşılaştırılmıştır. Bayes ve yapay sinir

ağlarının, ilgili nitelikler belirlendikten sonra daha iyi performans gösterdiği tespit edilmiştir. Çalışmada sonucunda Bayesin yapay sinir ağlarına göre daha yüksek sınıflandırma performansı gösterdiği sonucuna varılmıştır.

Lopez ve diğerleri (2012) çalışmalarında Moodle öğrenme yönetim sistemi forum verilerinin ders başarısında önemli olup olmadığını belirlemişledir. Diğer araştırma probleminde de sınıf değişkeni olan ders başarısının bilinmediğinde kümeleme algoritmaları kullanılarak aynı sonucun çıkıp çıkmayacağı test edilmiştir. Araştırmada öğrencilerin forumlara katılımlarına yönelik 8 değişken belirlenmiş ve bunlar kullanılarak öğrencinin ders başarısı (kaldı veya geçti) tahmin edilmeye çalışılmıştır. Buna yönelik birden fazla sınıflandırma algoritması karşılaştırılmış ve doğru sınıflama oranı kapsamında performanslar belirlenmiştir. Bulgulara göre 8 değişken ile oluşturulan sınıflamada en yüksek sınıflama oranına BayesNet algoritması (%87,7), nitelik seçme işlemi sonucunda tespit edilen 6 değişken ile oluşturulan sınıflandırmada ise Naive Bayes Simple algoritması (%89,4) en yüksek sınıflama oranına sahip olmuştur.

Şengür ve Tekin (2013) tarafından yapılan çalışmada karar ağaçları algoritması ve yapay sinir ağları yöntemleri kullanılarak Fırat Üniversitesinde 127 Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü öğrencisinin mezuniyet notları tahmin edilmeye çalışılmıştır. Çalışma sonucunda yapay sinir ağlarının, karar ağaçları algoritmasına nazaran daha yüksek tahmin başarı gösterdiği ortaya konmuştur.

Aksu ve Güzeller (2016) yaptıkları çalışmada Türkiye'de 2012 yılında PISA'ya katılan öğrencilerden elde edilen verilerden yapılan analizde veri madenciliği yöntemleri ve karar ağacı algoritmalarından CHAID kullanılmıştır. Çalışma sonucunda öğrencileri başarılı ve başarısız olacak şekilde sınıflandırmada öz yeterlik algısının, çalışma disiplininin ve derse ilişkin tutumun en önemli duyuşsal özellikler olduğu ortaya konmuştur. Bununla birlikte veri J.48 karar ağacı algoritması doğru sınıflama yüzdesinin CHAID algoritması sınıflandırma oranına çok yakın olduğu görülmüştür. Bu sonuçlara göre CHAID analizinin veri madenciliği karar ağacı algoritmalarında alternatif olarak görülebileceği ortaya konmuştur. Araştırmadaki bulgulara göre öz yeterlik algısının, çalışma disiplininin ve derse ilişkin tutum ile kaygı durumlarının Türkiye örnekleminde matematik okuryazarlığında özellikle üzerinde durulmasının gerektiği belirtilmiştir.

Taşcı ve Onan (2016) tarafından yapılan çalışmada K-En Yakın Komşuluk yöntemi parametrelerinin sınıflandırmadaki performansına etkisi incelenmiştir. Araştırmacılar, UCI Machine Learning Repository'deki altı farklı gerçek dünya verisinde K-en yakın komşuluk parametresinin sınıflandırma üzerine etkisini incelemişler ve tartışmışlardır. Çalışmada, ağırlıklandırma ölçütleri, k komşu sayısı ve uzaklık parametreleri kullanılarak farklı parametrelerde sınıflandırıcının performansı test edilerek doğruluk yüzdesi belirlenmiştir.

Yılmaz (2012) tarafından üniversite öğrencileri internet kullanımı profillerinin belirlenmesi ve internet ilgilerine göre profillerinin farklılaşıp farklılaşmadığı amacıyla yapılan çalışmada internet kullanımının eğlence ve iletişim amacı olmak üzere iki kümeye ayrıştığı belirlenmiştir. Ayrıca bölünmede internet ilgisinin büyük öneme sahip olduğu belirlenmiştir. Bu bulgulara göre birinci kümenin çoğunlukla erkek öğrenci olduğu, interneti yoğun kullandıkları ve büyük önem verdikleri; ikinci kümenin de ilk kümedekilere nazaran interneti daha az kullanan öğrencilerden oluştuğu ve internete orta düzeyde önem verdikleri belirlenmiştir.

Güldal ve Çakıcı (2017) yaptıkları çalışmada 70 öğrenci ile veri madenciliğindeki farklı sınıflandırma yöntem algoritmalarından Karar Ağacı (C4.5), k-en yakın komşuluk yöntemi k=1, 3 ve 5 değeri ve Naive Bayes doğru sınıflandırma oranlarını karşılaştırmıştır. Araştırmada en yüksek doğru sınıflandırma oranına sırasıyla k-en yakın komşuluk (k=3 değeri için), J48, k-en yakın komşuluk (k=1 değeri için), k-en yakın komşuluk (k=3 değeri için) ve Naive Bayes yöntemleri tarafından sağlanmıştır. Araştırmada farklı algoritma kullanılarak başarılı ve başarısız olacak şekilde öğrencilerin doğru sınıflandırma oranlarının %55.7 ile %64.3 arasında farklılık gösterdiği ortaya konmuştur.

Önen (2018) tarafından yapılan çalışmada TIMSS-2015 matematik başarısında etkili olduğu varsayılan öğretmen ve öğrenciye ait özellikler ile öğretim özellikleri açısından 4. ve 8. sınıf öğrencileri kümelere ayrılmıştır. Böylelikle her kümeye ait öğrenci profilleri belirlemiştir. Kümeleme analizi bulgularına göre 4. sınıf düzeyinde 3 küme, 8. sınıf düzeyinde 2 küme oluşmuştur. Bulgularda 4. sınıf düzeyinde öğrenci ve öğretmen özellikleri ile matematik başarısı kümelerin oluşmasında en etkili özelliklerin olduğu, öğretimsel niteliklerin de en az etkisi olan özellikler olduğu görülmüştür.

İlgili çalışmalar bir bütün olarak incelendiğinde yapay sinir ağları ve karar ağaçlarının en çok kullanılan veri madenciliği yöntemleri olduğu görülmektedir. Yapay sinir ağları diğer yöntemlere göre daha yüksek oranda doğru sınıflandırma yapmaktadır. Ayrıca araştırmaların çoğunda sadece iki farklı yöntemin incelenmiş ve karşılaştırılmıştır. Buna karşılık ikiden fazla sayıda yöntemin incelendiği ve karşılaştırılma yapıldı araştırma sayısı çok azdır. Kümeleme çalışmaları değerlendirildiğinde “K-means” algoritması daha sık kullanılan algoritma olmuştur. Bazı çalışmalarda hem kümeleme hem de sınıflandırma yapıldığı görülmüştür.

Bölüm 3
Yöntem

Bu bölümde araştırmanın türü, araştırmanın evreni ve örneklemi, veri toplama süreci, veri toplama araçları ve verilerin analizine ilişkin bilgilere yer verilmektedir.

Bu araştırma PISA 2018 Türkiye örneklemine dayalı olarak öğrencilerin okuma becerileri başarısını etkileyen faktörlere ve okuma becerileri başarı puanlarına göre öğrencilerin okuma okuryazarlık düzeylerinin Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri ile sınıflama doğruluklarının belirlenmesine yöneliktir. Çalışma okuma becerileri başarısını etkileyen faktörlerin ve okuma becerileri başarı puanlarının öğrencilerin başarı durumlarını ve okuma becerileri yeterlik düzeylerini Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri ile ne seviyede doğru sınıflama yaptığının tespit edilmesine yönelik var olan durumu tanımlaması açısından betimsel araştırma olarak görülmektedir (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2008).

Araştırmanın Evreni ve Örneklemi

Araştırma çalışma grubu PISA 2018 sınavına katılan 15 yaş grubu 79 ülkeden toplam 600.000 civarında öğrenciden oluşmaktadır. PISA 2018 Türkiye uygulaması ulaşılabilir öğrenci evreni 884971'dir. PISA 2018 uygulamasına Türkiye'yi 186 okul ve 6890 öğrenci temsil etmiştir. Türkiye'de eğitim gören 15 yaş grubu öğrencilerin temsil edecek öğrenciler, Uluslararası Merkez tarafından seçkisiz-olasılığa dayalı olacak şekilde seçilmiştir (MEB, 2019). Madde sayısı ve örneklem büyüklüğü arasındaki ilişki incelendiğinde 79 maddeye karşılık örneklem sayısı 6890 ile kabul edilebilir bir oranda ilişki bulunmuştur.

PISA 2018 uygulamasına katılan öğrencilerden Anadolu Lisesinde eğitim alanların oranı %43,7, Mesleki ve Teknik Anadolu Lisesinde eğitim alanların oranı %31,1 ve Anadolu İmam Hatip Lisesinde oran ise %13,7'dir. Türkiye örnekleminin %11,2'si Çok Programlı Anadolu, Fen, Sosyal Bilimler ve Anadolu Güzel Sanatlar Liselerinden oluşmaktadır. Uygulamaya katılanların %0,3'ü ortaokul seviyesinde öğrenim görmektedir. Örneklemin %49,6'sını kız, %50,4'ünü ise erkek öğrencilerden oluşmaktadır. Çalışmada çoğunlukla cevaplandırılmayan veya hiç

cevap girilmeyen maddeler veri setinden çıkarıldığında araştırmanın örneklemi 6431 öğrenciden oluşmaktadır.

Veri Toplama Süreci

Bu çalışmada kullanılan veriler 2020 yılında OECD'nin internet sayfasından (https://www.oecd.org/pisa/data/2018database/) elde edilmiştir. Veriler 2018 PISA uygulamasındaki öğrenci anketine Türkiye'den katılan öğrencilerin SPSS formatındaki cevaplarından oluşmaktadır.

Veri Toplama Araçları

Araştırmada, PISA 2018 uygulamasındaki ölçme araçları verileri kullanılmıştır. Her PISA uygulamasında matematik, okuma ve fen alanlarından biri ağırlıklı alan olarak belirlenmektedir. 2006 ve 2015 yıllarında Fen, 2003 ve 2012 yıllarında matematik ve 2000 ve 2009 yıllarında ise ağırlıklı alanı olarak okuma belirlenmiştir. PISA 2018 uygulamasındaki ağırlıklı alan da okuma becerileridir. Ağırlıklı alanın okuma becerileri olması PISA 2018 sonuçlarının matematik okuryazarlığı ve fen okuryazarlığından çok, okuma becerilerine odaklandığı ifade etmektedir. PISA araştırması, 2000, 2003, 2006 ve 2009 yıllarında kâğıt-kalem testi şeklinde uygulanmıştır. İlk kez 2012 yılında matematik okuryazarlığı alanında bilgisayar temelli uygulamaya imkân tanınmıştır. PISA 2018 uygulamalarına Türkiye bilgisayar tabanlı uygulama türünde katılmıştır (MEB, 2015; MEB, 2017).

PISA 2018 uygulamasında, önceki yıllardaki uygulamaların aksine sabit sorular yerine dinamik yapıda sorulardan oluşan bireyselleştirilmiş testler kullanılmıştır. Öğrenciler temelde bölümdeki sorulara verdiği cevapların doğruluğuna göre yeni sorularla karşılaşmaktadır. Bireyselleştirilmiş testler sayesinde, ortalamanın çok altında veya çok üstünde performans gösteren öğrencilerin kendi düzeylerine uygun sorularla test edilebilmesi sağlanmaktadır (OECD, 2019).

PISA 2018 öğrencilerin akademik performanslarını ölçmeyi amaçlayan bilişsel testler ile öğrenciyi bir bütün olarak değerlendirmek amacıyla hazırlanmış öğrenci ve okul anketlerini içermektedir.

Bilişsel testler iki saat süren ve her biri 30 dakikalık 4 alt testten meydana gelmektedir. Matematik ve fen okuryazarlığını ölçmek için kullanılan testlerde 6 alt test bulunmaktadır. Okuma becerileri okuryazarlığı ise alt test yerine 15 ünite olarak tasarlanmıştır. Ayrıca bilgisayar tabanlı olarak katılmayan ülkeler için de okuma, matematik ve fen okuryazarlığı alanında 30 maddelik kağıt-kalem testi uygulanmıştır.

Öğrencilerden içinde kendisi, ailesi ve evi, okuldaki dil öğrenimi, okulda öğrendiği Türkçe/Türk Dili ve Edebiyatı Dersi, hayatı hakkındaki düşüncesi, okulu, okul programı ve öğrenme süreleri konuları hakkındaki sorulardan oluşan anketi cevaplaması beklenmiştir. Bilgisayar tabanlı olarak icra edilen ana öğrenci anketi 79 sorudan oluşmaktadır ve 35 dakika sürmüştür. Öğrenci ve okul müdürü anketi dışında 5 anket bölümü daha bulunmaktadır. Bunlar; Bilgisayar aşinalık anketi, sosyal refah anketi, mesleki kariyer anketi, aile anketi ve öğretmen anketidir. Ayrıca tercihli olarak finansal okuryazarlık anketi de yapılmıştır. Çalışmada kullanılan veriler öğrenci anketi sorularından oluşmaktadır ve OECD'nin internet sitesinden indirilmiştir.

Verilerin Analizi

Çalışmada analiz yapılmadan önce önemli bir sorun olarak görülen kayıp veriler incelenmiştir (Demir ve Parlak, 2012). Yapılan inceleme sonucunda 459 öğrencinin verisi çoğunlukla cevaplandırılmadığı veya hiç cevap verilmediği için veri setinden çıkarılmıştır. Analize 6431 öğrenci verisi ile devam edilmiştir.

Ayrıca maddeleri yanıtlayanlar ile yanıtlamayanlar arasındaki sistematik farklılıklar yanlılık kaynağı olabileceğinden veri grubunda kayıp veri olup olmadığını belirlemek için kayıp veri analizi yapılmıştır. Kayıp veri analizindeki Little'ın MCAR testi sonucunda kayıp veri türünün (MAR-Missing At Random) olduğu görülmüştür. Analiz sonucunda 1678 rastgele eksik veri tespit edilmiştir. Kayıp veri oranı her bir değişkende yüzde 5'in altındadır. Kayıp verilerin incelenmesi sonucunda kayıp verilerin özellikle MAR varsayımının sağlandığı durumlarda daha iyi kestirimler ürettiği belirtilen ML (maximum likelihood) kestirimlerinden Beklenti Maksimizasyon (EM) yöntemi ile kayıp veri tamamlaması yapılmıştır. Kayıp veri tamamlanmadan önce yapılan her bir araştırma probleminin analizi ile kayıp tamamlandıktan sonra yapılan analiz sonuçları arasında bazı algoritmalarda yüzde 2-3, bazı algoritmalarda

ise yüzde 15-20 arasında değişim tespit edilmiş, kayıp tamamlandıktan sonra sınıflama doğruluğu oranının arttığı görüşmüştür.

Analiz, SPSS paket programı ve veri madenciliğinde eski adı CLEMENTINE paket programı olan SPPS Modeler ile gerçekleştirilmiştir. SPSS Modeler veri madenciliği için geliştirilen modelleme analiz programıdır. Veriye kolay erişim, verinin modelleme için hazırlanması, model oluşturma, birden çok modelin ardışık şekilde uygulanması ve farklı modellerin sonuçlarının kolay karşılaştırması sağlanmaktadır. SPSS Modeler, bütün veri madenciliği uygulamasında kullanılabilmekte, mümkün bütün model metotlarını içermekte, daha nitelikli sonuçların elde edilmesi için birbirine ardışık şekilde kullanımı mümkün kılmaktadır. SPSS Modeler, çok miktarda istatistiksel model algoritma ve grafik içermektedir.

Araştırmanın birinci ve üçüncü alt problemi ile ikinci ve dördüncü alt problemi analizleri aynı olup birinci alt problemde “başarı durumu bağımlı değişken, 24 indis değişken ise bağımsız değişken, üçüncü alt problemde ise “okuma becerileri yeterlik düzeyi” bağımlı değişken, 24 indis değişken ise bağımsız değişkendir. Buna paralel olarak ikinci alt problemde bağımlı değişken “başarı durumu” iken, dördüncü alt probleminde bağımlı değişken “okuma becerileri yeterlik düzeyi”dir. Bu sebeple verilerin analizi yapılırken sıralama yerine ilk önce birinci ve üçüncü alt problemin analizi, ardından da ikinci ve dördüncü alt problemin analizi açıklanmıştır.

Araştırmada Karar Ağaçları veri madenciliği yöntemi algoritmaları seçilirken alanyazın ve Karar Ağaçları algoritmaları özellikleri ile yapısı incelenmiştir. Yapılan inceleme sonucunda PISA 2018 Türkiye örneklemi için en uygun olan ve eğitim alanındaki geniş ölçekli testlerde ve okuma becerileri ile ilgili yapılan veri madenciliği çalışmalarında en sık kullanılan yöntem olduğu düşünülen C5.0, CHAID, C&RT ve QUEST algoritmaları kullanılmıştır.

Araştırmanın birinci alt probleminde 2018 PISA okuma becerileri başarısını etkileyen faktörlere ve okuma becerilerine ilişkin başarı puanlarına göre Yapay Sinir Ağlarının, Karar Ağacı algoritmalarının, K-En Yakın Komşuluk ve Naive Bayes analizlerinin öğrencileri başarı durumlarına göre hangi doğruluk oranında sınıflandırdığı bulmak için öncelikle öğrenciler, PISA 2018 okuma becerileri alanındaki başarı testinden aldıkları puanlara göre “Başarılı-Başarısız” olmak üzere iki kategoride ölçeklenmiştir. Bunun için veri setindeki 10 okuma başarı puanının

(Plausible Value; PV1READ, PV2READ….PV10READ) ortalaması alınarak "ortalama okumabaşarı puanı" değişkeni oluşturulmuştur. Ardından 6431 öğrencinin "ortalama okumabaşarı puanı" değişkeninin ortalaması hesaplanmıştır. Bu ortalamaya göre öğrencilerden "ortalama okumabaşarı puanı" 469,9'un altında olanlar "başarısız-0", üstünde olanlar "başarılı-1", olacak şekilde "başarıdurumu" değişkeni oluşturulmuştur. Bu düzenlemeler ışığında toplam 6431 öğrencinin katıldığı PİSA 2018 Türkiye uygulamasında "başarılı-1" öğrenci sayısı %49.9 ile 3212, "başarısız-0" öğrenci sayısı da %50.1 ile 3219'dur.

Araştırmanın üçüncü alt problemi kapsamında MEB PISA 2018 Türkiye Ön Raporunda belirtilen "Okuma Becerileri Yeterlik Düzeyleri" dikkate alınarak öğrenciler, PISA bilişsel başarı testlerinden aldıkları puanlara göre 6 kategoride ölçeklenmiştir. Öğrencilerin hangi yeterlik düzeyine girdiğini tespit etmek için öğrencilerin testten aldıkları 10 okuma başarı puanının (Plausible Value; PV1READ, PV2READ..PV10READ) ortalaması alınarak "ortalama okumabaşarı puanı" değişkeni oluşturulmuştur. Ardından öğrencilerin hepsinin "ortalama okumabaşarı puanı" ortalaması alınmıştır. 6 kategoride belirtilen Türkiye'den PISA'ya katılan öğrencilerin yeterlik düzeyleri 3 kategoriye indirgenmiştir. Bu kategorileşme yapılırken Okuma Becerileri Yeterlik Düzeyleri incelenerek 1 ve 2. Düzeyi "Düzey-1", 3 ve 4. Düzeyi "Düzey-2", ve 5 ve 6. Düzeyi "Düzey-3" olacak şekilde oluşturulmuştur. Okumabaşarı puanına göre düzey değişkeni kapsamında öğrencilerin okuma becerileri yeterlik düzeyleri 3 düzey baz alınarak belirlenmiş, böylece "Düzey" değişkeni oluşturulmuştur. Bu veri düzenlenmelerinin ardından değişkenlerden "Düzey" bağımlı değişken, 24 indis değişken ise bağımsız değişken olarak analize dahil edilmiştir. Bu düzenlemeler ışığında toplam 6431 öğrencinin katıldığı PİSA 2018 Türkiye uygulamasında "Düzey-1" öğrenci sayısı %54.6 ile 3512, "Düzey-2" öğrenci sayısı %43 ile 2763 ve "Düzey-3" öğrenci sayısı da %2.4 ile 156'dır. Bu çalışmada veri madenciliği analizi sonucunda belirlenen sınıflandırma doğruluğu yüzdeleri, nisbî şans kriterleri ile karşılaştırılarak değerlendirilmiştir.

Analiz boyunca ve analizler yapıldıktan sonra model içi değerlendirme sürecinde kullanılan veri madenciliği yönteminde ya da algoritmalarda en uygun modelin bulunabilmesi için birçok modelin oluşturularak test edilmesi gerekir. Bu sebeple en iyi modele ulaşıncaya kadar model oluşturma tekrarlanmaktadır. Ancak model oluşturma denetimli ve denetimsiz öğrenme modellerinde farklıdır. Bu model

oluşturduktan sonra diğer bir önemli aşama ise geçerliliğin test edilmesi ve incelenmesidir. Denetimli ve denetimsiz öğrenme modellerde geçerlilik yöntemleri modele göre farklılık göstermektedir. Sınıflama, regresyon gibi tahmin edici modeller ilk olarak bir miktar veri ile uygulanır, bu modelin eğitilmesi olarak adlandırılmaktadır. Ardından verinin kalanı kullanılarak model test edilip, doğrulanmaktadır. Eğitim ve test döngüsü bitirilince model sağlanır. Test kümesi kullanılarak model doğruluk derecesi tespit edilir. Model doğruluğu testinde çok farklı geçerlilik yöntemleri bulunmaktadır.

Bir sınıflama modelinin doğruluğunun belirlenmesinde kullanılan en temel yöntem basit geçerlilik yöntemdir. Verinin %5 ile %30 arasında bir kısmı test verisi olarak ayrılır ve modelin eğitilmesinde herhangi bir şekilde kullanılmaz (Akpınar, 2000). Tüm hesaplamalar için verinin bu şekilde ayrılmasında seçimi tesadüfi olarak yapan yöntemler kullanılır. Bu şekilde eğitim ve test veri setleri modeli oluşturan tüm veriyi temsil etmektedir. Bu yöntemde doğru sınıflanan veya tahmin edilen verinin toplam test örnek sayısına oranı doğruluk oranını vermektedir.

Çapraz geçerlilik yöntemi ise modelin oluşturulmasında tüm verinin kullanılmasına olanak sağlayan yöntemdir. Çapraz geçerlilikte veri kümesi rastgele iki eşit kümeye ayrılmaktadır. Öncelikle alt kümelerden birisi eğitim diğeri ise test için seçilir. Model oluşturularak test için veri üzerinde basit geçerlilik yönteminde olduğu gibi hata ve doğrulama oranı hesaplanır. Daha sonra test ve eğitim kümelerinin rolleri değiştirilerek aynı işlemler tekrarlanır. Elde edilen iki bağımsız doğrulama değerinin ortalaması alınarak modelin doğruluk oranı hesaplanır (Berthold ve Hand, 2003). Her verinin bir kez eğitim ve bir kez de test olarak kullanıldığı çapraz geçerlilik yönteminin genelleştirilmiş hali “n-katlı çapraz geçerlilik yöntemi” adını almaktadır. n-katlı çapraz geçerlilik yönteminde veri tesadüfi olarak eşit n adet gruba ayrılmaktadır. Veriler n gruba bölündükten sonra bir grup test olarak belirlenir ve kalan verinin n-1 grubu da eğitim için kullanılır. Bu şekilde n adet basit geçerlilik süreci tekrarlanarak her grup bir kez test için kullanılmış olmaktadır. Elde edilen n adet bağımsız hata oranının ortalaması oluşturulan modelin hata oranı olarak kullanılmaktadır.

Bootstrap yönteminde ise hata oranının tahmin edilmesinde kullanılan ve çapraz geçerlilik de olduğu gibi tüm veri model oluşturmada kullanılır. Bootstrap olarak adlandırılan örnek veri kümeleri asıl veri kümesinden örneklenerek

oluşturulur. Bu veri kümeleri “eğitim veri kümesi”, bu veri kümesi dışındaki veriler ise “test veri kümesi” olarak kullanılmaktadır. Bu şekilde her veri en az bir kez eğitim verisinde yer almaktadır. Bazen 1000 tekrar yapılmaktadır. Oluşturulan bu modelin hata oranı her Bootstrap örneğinin hata oranının ortalaması ile hesaplanmaktadır.

Birinci ve üçüncü araştırma probleminde de SPPS Modeler programında bulunan hem Çapraz Geçerlilik hem de Bootstrap yöntemi kullanılmıştır. Yapılan analizlerden önce veri seti %70 eğitime, %30 test verisine ayrılmıştır. Yöntemlerin ve algoritmaların doğruluk oranlarının arttırılması için analiz Boosting ile çalıştırılmıştır. Modellerin geliştirilmesinde ayrıca 10 katlı çapraz geçerlilik tekniği kullanılmıştır.

Birinci ve üçüncü araştırma problemi analizleri sonucunda her bir modelin ve algoritmanın eğitim setinin, test veri setinin ve tüm verinin genel doğru sınıflandırma oranları hesaplanmıştır. Tüm veri setinin genel doğru sınıflandırma oranı aşağıdaki eşitlik kullanılarak hesaplanmıştır.

$$\text{Doğru Sınıflandırma Oranı} = \frac{\text{Doğru Sınıflandırılmış Örnek Sayısı}}{\text{Toplam Örnek Sayısı}}$$

Yapay Sinir ağları analizlerinde “Çok Katmanlı Algılayıcı” modeli kullanılmıştır. Bu modele 1 bağımlı ve 24 bağımsız indis değişken dahil edilmiştir. Veri setinin tamamı kullanılarak verilerin %70.8’i (n=4556) eğitim setine, %29.2’si (n=1875) ise test setine ayrılmıştır.

Yapay sinir ağları modellerinin analizinde ROC analizleri yapılmaktadır. ROC eğrisi, olası kesim noktalarının bir araya getirilmesi ile oluşmaktadır. ROC eğrisi, referans eğrisi olarak adlandırılan y=x doğrusuna ne kadar uzaksa tahminin doğru sınıflandırma performansının o denli güçlü olduğu, y=x doğrusuna ne kadar yakınsa doğru sınıflandırma performansının o kadar zayıf olduğunu, sınıflandırmanın başarılı olmadığını göstermektedir. ROC eğrisinin altında kalan alan ‘AUROC’ olarak adlandırılmaktadır ve modelin güvenirliğinin tespit edilmesine yardımcı olan bir ölçümdür. İki olasılık dağılımının ayrılabilirlik ölçümünü veren gösteren ‘AUROC’ değeri, 0.5 ila 1.0 arasında değerler alan bir olasılık gibi yorumlanabilir. AUROC’

indeksine ait olasılıklar 1'e ne kadar yakınsa sonuç o kadar başarılı olacaktır (Bayru, 2007). Alan yazında göre tahmin modelinin ayırma yeteneği aşağıdaki şekilde sınıflandırılabildiği belirtilmektedir.

'AUROC' =0.5 Tahmin olasılığı mevcut değil, dolayısıyla ayrım yok

0.7≤ 'AUROC'≤ 0.8 istatistiksel olarak kabul edilebilir ayrım

0.8 ≤ 'AUROC' ≤0.9 istatistiksel olarak mükemmel ayrım

'AUROC'>0.9 istatistiksel olarak olağanüstü.

Karar Ağacı C5.0, CHAID, C&RT ve QUEST algoritmalarında, K-En Yakın Komşuluk ve Naive Bayes yöntemleri analizlerinde veri setinin tamamı kullanılarak verilerin %69.6'sı (n=4476) eğitim, %30.4'ü (n=1955) ise test için belirlenmiştir.

Bir modelde sınıflandırmanın doğruluğu testinde, maksimum şans ve nisbi şans kriterlerinin hesaplanarak karşılaştırılma yapılması gerekmektedir. Birinci araştırma problemi örneklemde "başarılı-1" öğrenci sayısı %49.9 ile 3212, "başarısız-0" öğrenci sayısı da %50.1 ile 3219'dur. Bu bilgiler ışığında örneklemin maksimum şans kriteri 0,501'dir. Nisbi şans kriteri ise $(0.501)^2$ +$(0.499)^2$ işlemi sonucunda 0,49.9 olarak hesaplanmıştır. Üçüncü araştırma problemi örnekleminde "Düzey-1" öğrenci sayısı %54.6 ile 3512, "Düzey-2" öğrenci sayısı %43 ile 2763 ve "Düzey-3" öğrenci sayısı da %2.4 ile 156'dır. Bu bilgiler ışığında örneklemin maksimum şans kriteri 0.55'dir. Nisbi şans kriteri ise $(0.546)^2$ +$(0.43)^2$ $(0.024)^2$ işlemi sonucunda 0.485 olarak hesaplanmıştır.

Araştırmanın ikinci ve dördüncü alt probleminde gruplanmamış verileri benzerliklerine göre gruplandırmak amacıyla kümeleme analizi yapılmıştır. Kümeleme yöntemleri hiyerarşik, bölümlemeli, yoğunluk tabanlı ve ızgara tabanlı olmak üzere genellikle dört grup altında toplanmaktadır. Bu yöntemlerin dışında da bazı farklı yöntemler de bulunmaktadır.

Büyük hacimli verilerde hem kategorik ve hem de sürekli verilerden oluşan büyük ve yüksek boyutlu veri İki Aşamalı Kümeleme algoritması tercih edilmektedir. İki Aşamalı Kümeleme algortiması hem kategorik hem de sayısal verileri kümeleyebilen, büyük boyutlu veri setlerinde etkili olarak çalışabilen bir algoritmadır. Bu yöntem hiyerarşik "Ward'ın En Küçük Varyans" teknikleri ile hiyerarşik olmayan "K-Ortalamalar" yöntemlerinin birleştirilmesi ile ortaya çıkan hibrid kümeleme tekniği

olarak ifade edilmektedir. Klasik algoritmalarla kıyaslandığında bu yöntem daha öznitelikli kategoriler vermesinden dolayı farklı alanlardaki birçok araştırmacı tarafından uygulama alanı bulmuştur (Taşkın ve Emel, 2010).

Bu teknikte küme sayısı, hem otomatik olarak algoritma tarafından hesaplanabilmektedir hem de gibi kullanıcı tarafından da girilebilmektedir. Veri yapısına uygun olarak bölümlendirilmiş veriler kullanılabilmektedir ve sayısal alanlar standartlaştırılabilmektedir.

İki Aşamalı Kümeleme algoritması iki adımdan oluşmaktadır. İlk adım verilerin pek çok alt kümeye ayrıldığı ön-kümeleme aşaması, ikinci adım ise bu alt kümelerin belirlenen sayıdaki küme içine yerleştirildiği kümeleme aşamasıdır. Kümeleme analizlerinde küme miktarı ile ilgili ön bilgi olmaması durumunda İki Aşamalı Kümeleme Analizi çoğunlukla tercih edilmektedir. Bu analizde en uygun sayıda küme yöntem vasıtasıyla belirlenmektedir. Küme sayısını otomatik olarak belirleyen Bayesçi Bilgi Ölçütü (BIC) veya Akaike Bilgi Ölçütü (AIC) kullanılmaktadır.

Bu çalışmada, İki Aşamalı Kümeleme analizinin kullanılmasının temel amacı büyük hacimli PISA 2018 veri setindeki öğrencilere ait verileri benzer özelliklere (değişkenlere) göre aynı kümede toplayarak farklı grupları belirlemek ve bu gruplar üzerindeki değişkenlerin önemini görmektedir. Bu veriler ışığında çalışmanın ikinci alt problemde değişkenlerden “başarıdurumu” bağımlı değişken, 24 indis değişken ise bağımsız değişken olarak, dördüncü alt probleminde de “düzey” bağımlı değişken, 24 indis değişken ise bağımsız değişken olarak SPSS programı vasıtasıyla İki Aşamalı Kümeleme analizi yapılmıştır.

Araştırmanın beşinci alt probleminde öğrencilerin okuma becerileri puanları dikkate alındığında hem başarı durumu hem de okuma becerileri yeterlik düzeylerine göre yapıla sınıflandırma sonuçları değerlendirilmiştir. Yapılan analizleri performanslarının karşılaştırılması gösterilmiştir.

Değişkenlerin seçimi. Bu çalışmada PISA 2018 verilerine dayalı olarak değişken seçiminde öğrencilerin okuma becerileri başarısını etkileyen değişkenlerin seçilmesine dikkat edilmiştir. Değişken seçimi kapsamında alan yazın, PISA 2018 Teknik Rapor ve PISA 2018 verileri incelenmiştir. Yapılan inceleme sonucunda bu çalışmada öğrencilerin okuma becerileri başarısını etkilediği düşünülen 24 indis değişken belirlenmiştir. Çalışma kapsamında kullanılan değişkenler Tablo 1’de

gösterilmiştir. Bu değişkenlerin yapısı ve kapsamı aşağıda belirtilmiştir. Araştırmanın her iki alt probleminde de aynı değişkenler kullanılmıştır.

Tablo 1

Çalışma Kapsamında Kullanılan Değişkenlere İlişkin Tanımlayıcı Bilgiler

Değişkenin Adı	Kodu
Sosyo Ekonomik Durum İndeksi	ESCS
Ailenin Mal Varlığı	WEALTH
Anlama ve Hatırlama	UNDREM
Özetleme	METASUM
Okuma ve Stratejileri Kullanma	METASPAM
Okumayı Sevme	JOYREAD
Disiplin Ortamı	DISCLIMA
Evdeki Eğitim Kaynakları	HEDRES
Evdeki Eğitimsel Eşyalar	HOMEPOS
Bilgi İletişim Teknolojileri (BİT) Kaynakları	ICTRES
Kültürel Sahiplik (Edebiyat, Şiir, Müzik, Resim, Sanat Kitabı)	CULTPOSS
Öğretmen Desteği	TEACHSUP
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki	STIMREAD
Okuma Yeterlik Algısı	SCREADCOMP
Okuma Zorluk Algısı	SCREADDIFF
PISA Zorluk Derecesini Algılama	PISADIFF
Öğrenci Tarafından Algılanan Aile Desteği	EMOSUPS
Algılanan Geri Bildirim	PERFEED
Kişisel Refah- Pozitif Etki	SWBP
Okuldaki İş Birliği Algısı	PERCOOP
Okula Ait Hissetme	BELONG
Okul Dışı BİT Kullanma	ENTUSE
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	HOMESCH
BİT Okulda Genel Olarak Kullanma	USESCH

PISA 2018 Öğrenci Anketi, MTK ölçeklendirmesine dayalı 39 türetilmiş değişken için veri sağlamıştır. PISA 2018 anketinde öğrencilerin okuma üst biliş kuramı kapsamındaki bilgisini değerlendiren üç senaryo içermektedir. "Anlamak ve hatırlamak" (UNDREM), "Özetlemek" (METASUM) ve "Güvenilirliği değerlendirmek" (METASPAM). UNDREM ve METASUM daha önce PISA 2009'da uygulanmıştır. METASPAM, PISA 2018 için yeni geliştirilmiştir. PISA 2018'de öğrenciler, ülke bağlamında aile refahının yerel ölçüsü olarak görülen ülkeye özgü üç ev eşyası dâhil olmak üzere evlerinde 16 ev eşyasının mevcudiyetini belirtmişlerdir. Ayrıca öğrenciler evdeki eşya ve kitapların miktarını da bildirmiştir. Bu öğelerden beş endeks türetilmiştir: Ailenin mal varlığı (WEALTH), kültürel sahiplik (CULTPOSS), evde eğitim kaynakları (HEDRES), Bilgi İletişim Teknolojileri (BİT) kaynakları (ICTRES) ve evdeki eğitimsel eşyalar (HOMEPOS).

Sınıftaki disiplin ortamını (DISCLIMA) değerlendirmek için öğrenciler, "Her ders", "Çoğu ders", "Bazı dersler" ve "Asla veya hemen hemen hiç” olacak şekilde dörtlü likert tipi madde cevaplandırmışlardır. Öğrencilere öğretmen desteği (TEACHSUP) sorulmuş, öğrenciler, "Her ders", "Çoğu ders", "Bazı dersler" ve "Asla veya hemen hemen hiç" kategorileriyle dörtlü likert ölçeğinde yanıt vermişlerdir. PISA 2009'da kullanılan öğretmenlerin okuma ve öğretme stratejilerini teşvik etme ölçeği (STIMREAD), öğretmenlerin öğrencilerin okuma katılımını ve okuma becerilerini nasıl teşvik ettikleri hakkında bilgi sağlamaktadır. Dört yanıt kategorisi "Asla veya neredeyse hiç", "Bazı derslerde", "Çoğu derste" ve "Tüm derslerde" arasında değişmektedir.

PISA 2009'da da sorulan JOYREAD değişkeni gene aynı şekilde okuma zevkini ölçmek için beş maddeye dayalı olarak “Kesinlikle katılmıyorum”, “Katılmıyorum”, “Katılıyorum” ve “Kesinlikle katılıyorum” arasında değişen dört yanıt kategorisi ile sorulmuştur. 6 maddelik bir ölçek kullanılarak, okuma görevlerini gerçekleştirirken öğrencilerden okuma ile ilgili benlik kavramlarını iki açıdan derecelendirmeleri istenmiştir: Yeterlik algıları (SCREADCOMP) ve zorluk algıları (SCREADDIFF). Ayrıca öğrencilere PISA testinin zorluk algıları (PISADIFF) sorulmuştur. Yanıtlar “Kesinlikle katılmıyorum”, “Katılmıyorum”, “Katılıyorum” ve “Kesinlikle katılıyorum” kategorilerinde dörtlü likert ölçeğinde verilmiştir. Öğrencilere, “Kesinlikle katılmıyorum”, “Katılmıyorum”, “Katılıyorum” ve “Kesinlikle

katılıyorum” yanıt kategorileriyle dörtlü likert ölçeği kullanılarak ebeveynlerinden algıladıkları duygusal destek (EMOSUPS) sorulmuştur. Bu soru aynı zamanda PISA 2015'te de kullanılmıştır. "Asla veya neredeyse hiç", "Bazı dersler", "Birçok ders" ve "Her ders veya hemen hemen her ders" kategorileriyle dörtlü likert ölçeği kullanılarak öğrencilerden algılanan öğretmen geri bildirimini (PERFEED) değerlendirmeleri istenmiştir.

Kişisel refah-pozitif etki (SWBP), öğrencilerin sahip olabileceği farklı duyguları sorarak öğrencilerin olumlu etkilerini ölçmektedir. Üç maddeden oluşur ve “Asla”, “Nadiren”, “Bazen” ve “Her Zaman” arasında değişen dörtlü likert ölçeği uygulamaktadır. Öğrencilerin okullarındaki öğrenciler arasındaki iş birliğini (PERCOOP) nasıl algıladıkları hakkında "Hiç doğru değil", "Biraz doğru", "Çok doğru" ve "Son derece doğru" arasında değişen dörtlü likert ölçeğine sahip üç madde bulunmaktadır. PISA 2018, öğrencilere daha önce PISA 2012 ve PISA 2015'te kullanılan altı maddeyi kullanarak “Kesinlikle katılıyorum”, “Katılıyorum”, “Katılmıyorum” ve “Kesinlikle katılmıyorum” yanıt kategorili likert ölçek ile okula aidiyet duygusu (BELONG) sorulmuştur.

Bilgi iletişim teknolojileri (BİT) aşinalık anketindeki üç soru ile dijital cihazların okul dışında boş zaman etkinlikleri, okul dışında okul çalışmaları ve okuldaki etkinlikler için ne sıklıkla kullanıldığı sorulmuştur. Üç soru için yanıt kategorileri "Asla veya neredeyse hiç", "Ayda bir veya iki kez", "Haftada bir veya iki kez", "Hemen hemen her gün" ve "Her gün" arasında değişmektedir. Bu konudaki ilgili değişkenler boş zaman etkinlikleri (ENTUSE), okul dışında okul çalışmaları (HOMESCH) ve okulda BİT kullanımı (USESCH) olarak ayarlanmıştır.

Bölüm 4
Bulgular ve Yorumlar

Bu bölümde, çalışmada yer alan alt problemlere göre elde edilmiş bulgu ve tablolara yer verilmiştir.

Araştırmanın Birinci Alt Problemine İlişkin Bulgular ve Yorumlar

Araştırmanın birinci alt problemi öğrencilerin 2018 PISA okuma becerileri başarısını etkileyen faktörlere ve okuma becerilerine ilişkin başarı puanlarına göre Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri öğrencileri başarı durumlarına göre hangi doğruluk oranında sınıflandırmaktadır?

Bu kapsamda birinci alt problem dört bölümden oluşmaktadır. Bu nedenle her bölüme ilişkin elde edilen bulgular ve yorumlar dört ayrı başlık altında incelenmiştir.

Yapay sinir ağlarına ilişkin bulgular ve yorumlar. Bu modele 1 bağımlı ve 24 bağımsız indis değişken dahil edilmiştir. Okuma becerileri başarısı yordanırken en iyi performansı veren ağın mimarisini bulabilmek için 50 adet deneme yapılmıştır. Yapılan uygulamalarda değişik miktardaki oluşturulan gizli katmanı ve her katmanda farklı sayıda sinir hücreleri mevcut ağ mimarileri birbirleriyle karşılaştırılmıştır. Bununla birlikte katmanlardaki en iyi sonucu ortaya koyan aktivasyon fonksiyonlarının belirlenmesi amacıyla, her mimari için farklı şekilde oluşturulan fonksiyonlar sınanmıştır. Yapılan bu denemeler sonucunda oluşturulan ağ üç katmandan meydana gelmektedir. Birinci katmanda (girdi katman) olan 24, ikinci katman olan gizli katmanda 7 adet yapay sinir hücresi, son katmanda ise iki sinir hücresi bulunmaktadır. Gizli katmanda “Hiperbolik Tanjant Fonksiyon” ve çıktı katmanında ise “Softmax Fonksiyonu” uygulanmıştır.

Yapılan analiz sonucunda yapay sinir ağı ile okuma becerileri başarısına ilişkin analiz sonuçları Tablo 2’de gösterilmiştir.

Tablo 2’e göre yapay sinir ağı, eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %76.6’lık, test örneklemindeki öğrencilerinkini ise %72.5’lik bir performansla doğru tahmin etmiştir. Ayrıca eğitim veri setindeki başarılı öğrencilerin %75.4’ünü, test veri setindeki başarılı öğrencilerin ise %71.5’ini doğru sınıflandırmıştır. Eğitim veri setindeki başarısız öğrencilerin %77.8’i doğru

sınıflandırılırken, test veri setindeki başarısız öğrencilerin %73.5'i doğru sınıflandırılmıştır.

Tablo 2

Yapay Sinir Ağları Okuma Becerileri Başarısı Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen		
		Başarısız	Başarılı	Sınıflandırma Doğruluğu (%)
Eğitim	Başarısız	1776	507	77.8
	Başarılı	553	1691	75.4
	Toplam %	% 51.4	% 48.6	76.6
Test	Başarısız	688	248	73.5
	Başarılı	276	692	71.5
	Toplam %	% 50.6	% 49.4	72.5

Eğitim ve test veri setlerinin beraber genel doğru sınıflandırma oranı ise doğru sınıflandırılmış örneklem sayısının toplam örneklem sayısına bölünmesi ile hesaplanmaktadır. Tablo 2'de görüldüğü üzere doğru sınıflandırılmış örneklem sayısı 4847, toplam örneklem sayısı 6431'dir. Bu sonuca göre yapay sinir ağının genel doğru sınıflandırma oranı %75.4 olarak hesaplanmıştır.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem dereceleri Tablo 3'te gösterilmiştir.

Tablo 3 incelendiğinde, okuma becerileri başarısına ilişkin en önemli girdi değişkenlerinin "Evdeki Eğitimsel Eşyalar (0.118)" ve "Ailenin Mal Varlığı (0.097)" olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, "Sosyo-Ekonomik Durum İndeksi", "PISA Zorluk Derecesini Algılama", "Okul Dışı BİT Kullanma ", "BİT Okulda Genel Olarak Kullanma", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Okuma ve Stratejileri Kullanma", "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Özetleme", "Disiplin Ortamı", "Algılanan Geri Bildirim", "Okumayı Sevme" , "Kültürel Sahiplik", "Anlama ve Hatırlama", "Okuma Yeterlik Algısı", "Okuma Zorluk Algısı", "Okula Ait Hissetme", "Öğretmen Desteği", "Öğrenci Tarafından Algılanan Aile Desteği", "Okuldaki İş Birliği Algısı", "Evdeki Eğitim Kaynakları", "Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki" ve "Kişisel Refah- Pozitif Etki" değişkenleri takip etmektedir.

Tablo 3

Yapay Sinir Ağları Bağımsız Değişken Önem Dereceleri

Bağımsız değişken	Önem
Evdeki Eğitimsel Eşyalar	.118
Ailenin Mal Varlığı	.097
Sosyo Ekonomik Durum İndeksi	.073
PISA Zorluk Derecesini Algılama	.062
Okul Dışı BİT Kullanma	.062
BİT Okulda Genel Olarak Kullanma	.061
Bilgi İletişim Teknolojileri (BİT) Kaynakları	.060
Okuma ve Stratejileri Kullanma	.054
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	.045
Özetleme	.040
Disiplin Ortamı	.038
Algılanan Geri Bildirim	.035
Okumayı Sevme	.035
Kültürel Sahiplik	.030
Anlama ve Hatırlama	.028
Okuma Yeterlik Algısı	.022
Okuma Zorluk Algısı	.021
Okula Ait Hissetme	.020
Öğretmen Desteği	.019
Öğrenci Tarafından Algılanan Aile Desteği	.018
Okuldaki İş Birliği Algısı	.018
Evdeki Eğitim Kaynakları	.017
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı Teşviki	.017
Kişisel Refah- Pozitif Etki	.012

Oluşturulan yapay sinir ağının çıktısı okuma becerileri başarısı olmasından dolayı, bağımsız değişkenlerin önem dereceleri incelendiğinde “Evdeki Eğitimsel Eşyalar” ve “Ailenin Mal Varlığı” değişkenlerinin okuma becerileri başarısında en önemli belirleyiciler olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenlerin ise “Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki (0.017)” ve “Kişisel refah- Pozitif Etki (0.012)” olduğu ifade edilebilir.

Nisbi ve maksimum şans kriterleri dikkate alındığında örneklemdeki başarı durumlarına göre “başarılı” ve “başarısız” öğrencilerin oranı dikkate alındığında maksimum şans kriteri 0.51 ve nisbi şans kriteri ise 0.499 olarak hesaplanmıştır. Bu bulgulara göre yapay sinir ağları genel sınıflandırma oranı değeri (%75.4) bu değerlerin üstündedir. Bu bulgu yapay sinir ağlarının bu modelde sınıflandırmada başarılı olarak kullanılabileceğini göstermektedir.

Yapay sinir ağları bağımsız değişken önem dereceleri bulguları değerlendirildiğinde “Evdeki Eğitimsel Eşyalar”daki değişkenliğin okuma becerileri başarısını büyük ölçüde etkilediği sonucuna varılabilir. Ayrıca, “Kişisel refah- Pozitif Etki” değişkeni, yani öğrencilerin kendilerini değerlendirdiklerinde sahip olabileceği farklı duygular (sevinçli, neşeli ve mutlu) okuma becerileri başarısı üzerinde en etkisiz bağımsız değişkendir. Bu durum; öğrencilerin sahip olduğu olumlu etkilerin ve duyguların okuma başarıları için önemli olmadığını, okuma başarısı için öğrencilerin mutluluk durumunun başarılarına katkı sağlamadığını göstermektedir.

YSA analizlerinde ROC analizleri yapılmaktadır. Yapılan analiz sonucunda modele ilişkin ROC eğrisi altında kalan alanın değerleri Tablo 4’te gösterilmiştir.

Tablo 4

ROC Analizi Sonucu Eğri Altında Kalan Alan Değerleri

		Eğri Altında Kalan Alan
Başarı Durumu	Başarısız	0.837
	Başarılı	0.837

Tablo 4’te da görüldüğü üzere model tarafından 0,837 ile istatistiksel olarak mükemmel bir ayırma yeteneği sergilenmiştir. Bu analizle modelin performansı da test edilmiştir.

Karar ağaçlarına ilişkin bulgular ve yorumlar. Öğrencilerin okuma becerileri başarısını yordamak amacıyla, ikinci bölümde yapısı ve çalışma prensipleri ayrıntılı olarak açıklanan dört karar ağacı algoritması analiz sonuçları incelenmiştir. Karar ağaçları algoritması analizine 1 bağımlı ve 24 bağımsız indis değişken dahil edilmiştir.

C5.0 algoritmasına ilişkin bulgular ve yorumlar. Çalışmada C5.0 algoritması, doğruluk oranının arttırılması için "Boosting" ile çalıştırılmıştır. Boosting tekniğinde sırasıyla birden çok model üretilmektedir. Birinci aşamada standart C5.0 algoritması ile model üretilmekte, ikinci aşamada ise ilk modelin sınıflayamadığı veriler üzerine odaklanılmaktadır. Üçüncü ve son aşamada ise ikinci modeldeki hatalarının giderildiği bir model oluşturulmaktadır. Bu modelde geçerlilik sınaması olarak 10 katlı çapraz geçerlilik testi kullanılmıştır. Modelin oluşturulması aşamasında veri 10 parçaya bölünerek her bir parça üzerinde model türetilip diğer kümelerde test edilmiştir. Modelin çapraz geçerlilik yöntemi ile belirlenen doğruluk oranı ortalama %74.9 ve standart sapması %0.7'dir. Model üretildikten sonra tüm veri üzerinde yapılan testteki doğruluk oranı %89.6 ve modelin karar ağacının derinliği 21 olarak gerçekleşmiştir.

C5.0 algoritması ile okuma becerileri başarısı analiz sonuçları Tablo 5'te gösterilmiş, eğitim ve test verilerinin frekans ve doğruluk oranları detaylı olarak belirtilmiştir.

Tablo 5

C5.0 Algoritması Okuma Becerileri Başarısı Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen		
		Başarısız	Başarılı	Sınıflandırma Doğruluğu (%)
Eğitim	Başarısız	2,014	219	91
	Başarılı	259	1984	88.5
	Toplam%	50.78%	49.2%	89. 32
Test	Başarısız	895	91	90.8
	Başarılı	96	873	90
	Toplam%	50.7%	49.3%	90.43

Tablo 5 incelendiğinde, C5.0 algoritması eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %89.32'lik bir performansla, test örneklemindeki öğrencilerin okuma becerileri başarısını ise %90.43'lük bir performansla doğru tahmin etmiştir. Ayrıca eğitim veri setindeki başarılı öğrencilerin %88.5'ini, test veri setindeki başarılı öğrencilerin ise %90'nını doğru sınıflandırmıştır. Eğitim veri setindeki başarısız öğrencilerin %91'i doğru sınıflandırılırken, test veri setindeki başarısız öğrencilerin %90.8'i doğru sınıflandırılmıştır. Bu sonuca göre yapay sinir ağlarında olduğu gibi özellikle başarısız öğrencilerin yordanmasında iyi sonuçlar verdiği sonucuna varılabilir.

Tablo 5'te görüldüğü üzere doğru sınıflandırılmış örneklem sayısı 5766, toplam örneklem sayısı 6431'dir. Bu sonuca göre C5.0 algoritmasının genel doğru sınıflandırma oranı %89.6 olarak hesaplanmıştır. Örneklemin maksimum ve nisbi şans kriterileri sırasıyla 0.501 ve 0.499'dur. C5.0 algoritmasının sınıflandırma oranı olan %89.6 bu değerlerin üstündedir. Bu sonuca göre C5.0 algoritmasının bu modelde sınıflandırmada başarılı olarak kullanılabileceği sonucuna varılabilir.

C5.0 algoritması analizinde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem dereceleri belirlenerek Tablo 6'da gösterilmiştir.

Bağımsız değişkenlerin önem dereceleri incelendiğinde "Okuma Zorluk Algısı (0.047)" değişkeninin okuma becerileri başarısının en önemli belirleyicisi olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişken ise "Sosyo-Ekonomik Durum İndeksi (0.037)" değişkeni olarak görülmektedir. Her zaman etkisi yüksek olan Sosyo-Ekonomik Durum İndeksinin etkisinin düşük olması diğer değişkenlerle birlikte değerlendirilmesinden kaynaklanmaktadır ve bu durum manidar olmamaktadır. "Okuma Zorluk Algısı"ndaki değişkenliğin okuma becerileri başarısını büyük ölçüde etkilediği görülmektedir. Ayrıca, "Sosyo-Ekonomik Durum İndeksi" değişkeni, yani öğrencilerin sosyo-ekonomik durumunun okuma becerileri başarısı üzerindeki en etkisiz bağımsız değişkendir.

Tablo 6

C5.0 Algoritması Bağımsız Değişkenlerinin Önem Dereceleri

Bağımsız değişken	Önem
Okuma Zorluk Algısı	0.0475
Bilgi İletişim Teknolojileri (BİT) kaynakları	0.0427
Okuldaki İş Birliği Algısı	0.0424
Öğrenci Tarafından Algılanan Aile Desteği	0.0424
BİT Okulda Genel Olarak Kullanma	0.0423
Kültürel Sahiplik (Edebiyat, Şiir, Müzik, Resim, Sanat Kitabı)	0.0423
Kişisel Refah- Pozitif Etki	0.0423
Okul Dışı BİT Kullanma	0.0423
Ailenin Mal Varlığı	0.0423
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı Teşviki	0.0423
Disiplin Ortamı	0.0423
Evdeki Eğitim Kaynakları	0.0423
Algılanan Geri Bildirim	0.0423
Okula Ait Hissetme	0.0421
PISA Zorluk Derecesini Algılama	0.0420
Evdeki Eğitimsel Eşyalar	0.0420
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	0.0420
Anlama ve Hatırlama	0.0418
Okuma ve Stratejileri Kullanma	0.0398
Özetleme	0.0396
Öğretmen Desteği	0.0394
Okumayı Sevme	0.0394
Okuma Yeterlik Algısı	0.0391
Sosyo-Ekonomik Durum İndeksi	0.0371

CHAID algoritmasına ilişkin bulgular ve yorumlar. CHAID algoritması, karar ağacının ilk dalını meydana getirmek için birinci tahmin edici özelliği seçmektedir. Alt daldaki her bir düğüm seçilen değişkenin homojen grubundan oluşturulur. Bu tekrarlama süreci karar ağacın hepsi geliştirilinceye kadar sürer. Analiz parametreleri en büyük ağaç derinliği 10, kategorik hedef için ki-kare hesaplama metodu Pearson, durdurma kriterleri kök düğüm için %2, alt düğüm için %1 ve en büyük yineleme 100'dür.

CHAID algoritması ile okuma becerileri başarısına ilişkin elde edilen analiz sonuçları Tablo 7'de gösterilmiştir.

Tablo 7

CHAID Algoritması Okuma Becerileri Başarısı Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen		
		Başarısız	Başarılı	Sınıflandırma Doğruluğu (%)
Eğitim	Başarısız	1830	403	82
	Başarılı	387	1856	82.7
	Toplam%	%49.53	%50.47	82.35
Test	Başarısız	669	317	68
	Başarılı	291	678	70
	Toplam%	%49.10	%50.90	69

Tablo 7 incelendiğinde, CHAID algoritması eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %82.35'lik, test örneklemindeki öğrencilerin okuma becerileri başarısını da %69'luk bir performansla doğru tahmin etmiştir. Ayrıca eğitim veri setindeki başarılı öğrencilerin %82.7'sini, test veri setindeki başarılı öğrencilerin ise %70'ni doğru sınıflandırmıştır. Eğitim veri setindeki başarısız öğrencilerin %82'si doğru sınıflandırılırken, test veri setindeki başarısız öğrencilerin %68'i doğru sınıflandırılmıştır. Bu sonuca göre CHAID algoritması C5.0 algoritmasının tersine özellikle başarılı öğrencilerin yordanmasında iyi sonuçlar verdiğini söylemek mümkündür.

Tablo 7'de görüldüğü üzere doğru sınıflandırılmış örneklem sayısı 5033, toplam örneklem sayısı 6431'dir. Bu sonuca göre CHAID algoritmasının genel doğru

sınıflandırma oranı %78.2 olarak hesaplanmıştır. Örneklemdeki maksimum ve nisbi şans kriteri 0.501 ve 0.499'dur. %78.2 olarak hesaplanan CHAID algoritması sınıflandırma oranı bu maksimum ve nisbi şans kriteri üstündedir. Bu değerlere göre CHAID algoritmasının bu modelde sınıflandırmada başarılı olarak kullanılabileceği sonucuna varılabilir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 8'de gösterilmiştir.

Tablo 8 incelendiğinde, okuma becerileri başarısına yönelik oluşturulan CHAID Algoritması için en önemli girdi değişkenlerinin "Okuma ve Stratejileri Kullanma (0.18)" ve "Özetleme (0.11)" olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, "PISA Zorluk Derecesini Algılama", "Evdeki Eğitimsel Eşyalar", "Sosyo Ekonomik Durum İndeksi", "Anlama ve Hatırlama", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Ailenin Mal Varlığı" "Okuma Zorluk Algısı", "Kültürel Sahiplik (Edebiyat, Şiir, Müzik, Resim, Sanat Kitabı)", "Evdeki Eğitim Kaynakları", "Okumayı Sevme", "BİT Okulda Genel Olarak Kullanma", "Öğrenci Tarafından Algılanan Aile Desteği", "Okuma Yeterlik Algısı", "Disiplin Ortamı", "Okul Dışı BİT Kullanma", "Okula Ait Hissetme", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki", "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Okuldaki İş Birliği Algısı", "Algılanan Geri Bildirim", "Kişisel Refah- Pozitif Etki" ve "Öğretmen Desteği" değişkenleri takip etmektedir.

Bağımsız değişkenlerin önem dereceleri incelendiğinde, "Okuma ve Stratejileri Kullanma (0.18)" değişkeninin okuma becerileri başarısında en önemli belirleyici olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenin ise "Öğretmen Desteği (0.004)" olduğu ifade edilebilir. Bu sonuçla "Okuma ve Stratejileri Kullanma"daki değişkenliğin okuma becerileri başarısını büyük ölçüde etkilediği, "Öğretmen Desteği"nin, yani öğrencilerin öğretmenlerinden aldığı desteğin ise okuma becerileri başarısı üzerinde en etkisiz olduğu görülmektedir. Bu durum; öğretmenlerin öğrencilere öğrenme konusundaki yardımının, bir konuyu anlamadaki desteğinin okuma başarıları için önemli olmadığını, okuma başarısı için öğretmen desteğinin öğrenme ve anlamada başarılarına katkı sağlamadığını göstermektedir.

Tablo 8

CHAID Algoritması Bağımsız Değişkenlerinin Önem Dereceleri

Bağımsız değişken	Önem
Okuma ve Stratejileri Kullanma	.1838
Özetleme	.1128
PISA Zorluk Derecesini Algılama	.0966
Evdeki Eğitimsel Eşyalar	.0759
Sosyo Ekonomik Durum İndeksi	.0736
Anlama ve Hatırlama	.0712
Bilgi İletişim Teknolojileri (BİT) Kaynakları	.0596
Ailenin Mal Varlığı	.0484
Okuma Zorluk Algısı	.0447
Kültürel Sahiplik	.0442
Evdeki Eğitim Kaynakları	.0413
Okumayı Sevme	.0398
BİT Okulda Genel Olarak Kullanma	.0220
Öğrenci Tarafından Algılanan Aile Desteği	.0177
Okuma Yeterlik Algısı	.0160
Disiplin Ortamı	.0158
Okul Dışı BİT Kullanma	.0094
Okula Ait Hissetme	.0083
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki	.0066
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	.0040
Okuldaki İş Birliği Algısı	.0034
Algılanan Geri Bildirim	.0024
Kişisel Refah- Pozitif Etki	.0020
Öğretmen Desteği	.0004

C&RT algoritmasına ilişkin bulgular ve yorumlar. C&RT algoritması en iyi bölme yapmak için bölme sonrası oluşacak katışıklık ölçümündeki azalmayı dikkate alır. Bu algoritmada her düğüm iki alt gruba bölünür ve bu bölme işlemi durma kriterlerinden biri sağlanıncaya kadar devam etmektedir. Analiz parametreleri en büyük ağaç derinliği 5, en büyük vekil (proxy) sayısı 0 (veri setinde eksik değer olmadığını göstermektedir), kategorik hedef alanı için katışıklık ölçümü Gini, durdurma kriterleri kök düğüm için %2, alt düğüm için %1'dir. C&RT algoritması ile okuma becerileri başarısına ilişkin elde edilen analiz sonuçları Tablo 9'da gösterilmiş, eğitim ve test verilerinin frekans ve doğruluk oranları detaylı olarak belirtilmiştir.

Tablo 9

C&RT Algoritması Okuma Becerileri Başarısı Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen		
		Başarısız	Başarılı	Sınıflandırma Doğruluğu (%)
Eğitim	Başarısız	1734	799	68.4
	Başarılı	488	1755	78.2
	Toplam%	%46.52	%53.48	77.95
Test	Başarısız	736	250	74.6
	Başarılı	264	705	72.7
	Toplam%	%51.15	%48.85	73.71

Tablo 9 incelendiğinde, C&RT algoritması eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %77.95'lik bir performansla, test örneklemindeki öğrencilerin okuma becerileri başarısını ise %73.71'lik bir performansla doğru tahmin etmiştir. Ayrıca eğitim veri setindeki başarılı öğrencilerin %78.2'ini, test veri setindeki başarılı öğrencilerin ise %72.7'sini doğru sınıflandırmıştır. Eğitim veri setindeki başarısız öğrencilerin %68.4'ü doğru sınıflandırılırken, test veri setindeki başarısız öğrencilerin %74.6'sı doğru sınıflandırılmıştır. Bu sonuca göre C&RT algoritması, CHAID algoritması ile aynı şekilde özellikle başarılı öğrencilerin yordanmasında iyi sonuçlar verdiğini söylemek mümkündür.

Tablo 9'da belirtildiği üzere doğru sınıflandırılmış örneklem sayısı 4930, toplam örneklem sayısı 6431'dir. Bu sonuca göre C&RT algoritmasının genel doğru sınıflandırma oranı %76.6 olarak hesaplanmıştır. Örneklemin maksimum ve nisbi

şans kriterleri sırasıyla 0.501 ve 0.499'dur. %76.6 olarak hesaplanan C&RT algoritması sınıflandırma oranı örneklemin maksimum ve nisbi şans kriterleri üstündedir. Bu sonuca göre C&RT algoritmasının bu modelde sınıflandırmada başarılı olarak kullanılabileceği görülmektedir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 10'da gösterilmiştir.

Tablo 10 incelendiğinde, okuma becerileri başarısına yönelik oluşturulan C&RT Algoritması için en önemli girdi değişkenlerinin "Okuma ve Stratejileri Kullanma (0.18)" ve "Özetleme (0.10)" olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, "Sosyo Ekonomik Durum İndeksi", "PISA Zorluk Derecesini Algılama", "Evdeki Eğitimsel Eşyalar", "Anlama ve Hatırlama", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Ailenin Mal Varlığı", "Okuma Zorluk Algısı", "Kültürel Sahiplik", "Evdeki Eğitim Kaynakları", "Okumayı Sevme", "BİT Okulda Genel Olarak Kullanma", "Öğrenci Tarafından Algılanan Aile Desteği", "Okuma Yeterlik Algısı", "Disiplin Ortamı", "Okul Dışı BİT Kullanma", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki", "Okula Ait Hissetme", "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Algılanan Geri Bildirim", "Okuldaki İş Birliği Algısı", "Kişisel Refah-Pozitif Etki" ve "Öğretmen Desteği" değişkenleri takip etmektedir.

Bağımsız değişkenlerin önem derecelerine göre "Okuma ve Stratejileri Kullanma (0.18)" değişkeninin okuma becerileri başarısında en önemli belirleyici olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenin ise "Öğretmen Desteği (0.005)" olduğu görülmektedir. "Okuma ve Stratejileri Kullanma"daki değişkenliğin okuma becerileri başarısını büyük ölçüde etkilemektedir. Bunun yanı sıra "Öğretmen Desteği"nin, yani öğrencilerin öğretmenlerinden aldığı desteğin ise okuma becerileri başarısı üzerinde en etkisiz olduğu görülmektedir. Bu durum; öğretmenlerin öğrencilere öğrenme konusundaki yardımının, bir konuyu anlamadaki desteğinin okuma başarıları için önemli olmadığını, okuma başarısı için öğretmen desteğinin öğrenme ve anlamada başarılarına katkı sağlamadığını göstermektedir.

Tablo 10

C&RT Algoritması Bağımsız Değişkenlerinin Önem Dereceleri

Bağımsız değişken	Önem
Okuma ve Stratejileri Kullanma	.1813
Özetleme	.1025
Sosyo Ekonomik Durum İndeksi	.0888
PISA Zorluk Derecesini Algılama	.0886
Evdeki Eğitimsel Eşyalar	.0792
Anlama ve Hatırlama	.0691
Bilgi İletişim Teknolojileri (BİT) Kaynakları	.0652
Ailenin Mal Varlığı	.0515
Okuma Zorluk Algısı	.0445
Kültürel Sahiplik	.0433
Evdeki Eğitim Kaynakları	.0418
Okumayı Sevme	.0329
BİT Okulda Genel Olarak Kullanma	.0263
Öğrenci Tarafından Algılanan Aile Desteği	.0177
Okuma Yeterlik Algısı	.0162
Disiplin Ortamı	.0152
Okul Dışı BİT Kullanma	.0128
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki	.0076
Okula Ait Hissetme	.0054
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	.0049
Algılanan Geri Bildirim	.0022
Okuldaki İş Birliği Algısı	.0016
Kişisel Refah- Pozitif Etki	.0007
Öğretmen Desteği	.0005

QUEST algoritmasına ilişkin bulgular ve yorumlar. Bölmeleri belirlemek için QUEST algoritmasında kuadratik ayırma analizi (quadratic discriminant analysis) kullanılmış, her düğüm 2 alt gruba bölünmüştür. Analiz parametreleri en büyük ağaç derinliği 10, en büyük vekil (proxy) sayısı 0 (veri setinde eksik değer olmadığını göstermektedir), Alfa (bölme için) 0.05, durdurma kriterleri kök düğüm için %2, alt düğüm için %1'dir.

Yapılan analiz sonucunda QUEST algoritması ile okuma becerileri başarısına ilişkin analiz sonuçları Tablo 11'de sunulmuştur.

Tablo 11

QUEST Algoritması Okuma Becerileri Başarısı Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin edilen		
		Başarısız	Başarılı	Sınıflandırma Doğruluğu (%)
Eğitim	Başarısız	1715	518	76.8
	Başarılı	555	1688	75.2
	Toplam%	%50.71	%49.29	76.3
Test	Başarısız	732	254	74.2
	Başarılı	268	701	72.3
	Toplam%	%51.15	%48.85	73.3

Tablo 11 incelendiğinde, QUEST algoritması eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %76.3'lük, test örneklemindekilerin okuma becerileri başarısını da %73.3'lük performansla doğru tahmin etmiştir. Ayrıca eğitim veri setindeki başarılı öğrencilerin %75.2'ini, test veri setindeki başarılı öğrencilerin ise %72.3'ünü doğru sınıflandırmıştır. Eğitim veri setindeki başarısız öğrencilerin %76.8'i doğru sınıflandırılırken, test veri setindeki başarısız öğrencilerin %74.2'si doğru sınıflandırılmıştır. Bu sonuca göre QUEST algoritması, C5.0 ile aynı şekilde özellikle başarısız öğrencilerin yordanmasında iyi sonuçlar verdiğini söylemek mümkündür.

Tablo 11'e göre doğru sınıflandırılmış örneklem sayısı 4836, toplam örneklem sayısı 6431, QUEST algoritmasının genel doğru sınıflandırma oranı ise %75 olarak hesaplanmıştır. Örneklemin maksimum ve nisbi şans kriterleri sırasıyla

0.501 ve 0.499'dur. %75 olarak hesaplanan QUEST algoritması sınıflandırma oranı örneklemin maksimum ve nisbi şans kriterleri üstündedir. Bu sonuç QUEST algoritmasının bu modelde sınıflandırmada başarılı olarak kullanılabileceğini göstermektedir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 12'de gösterilmiştir.

Tablo 12 incelendiğinde, okuma becerileri başarısına yönelik oluşturulan QUEST Algoritması için en önemli girdi değişkenlerinin "Okuma ve Stratejileri Kullanma (0.196)" ve "Özetleme (0.115)" olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, "Sosyo Ekonomik Durum İndeksi" "PISA Zorluk Derecesini Algılama", "Anlama ve Hatırlama", "Evdeki Eğitimsel Eşyalar", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Kültürel Sahiplik", "Okuma Zorluk Algısı", "Ailenin Mal Varlığı", "Okumayı Sevme", "Evdeki Eğitim Kaynakları", "BİT Okulda Genel Olarak Kullanma", "Öğrenci Tarafından Algılanan Aile Desteği", "Okuma Yeterlik Algısı", "Disiplin Ortamı", "Okul Dışı BİT Kullanma", "Algılanan Geri Bildirim", "Okula Ait Hissetme", "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki", "Okuldaki İş Birliği Algısı", "Öğretmen Desteği" ve "Kişisel Refah- Pozitif Etki" değişkenleri takip etmektedir.

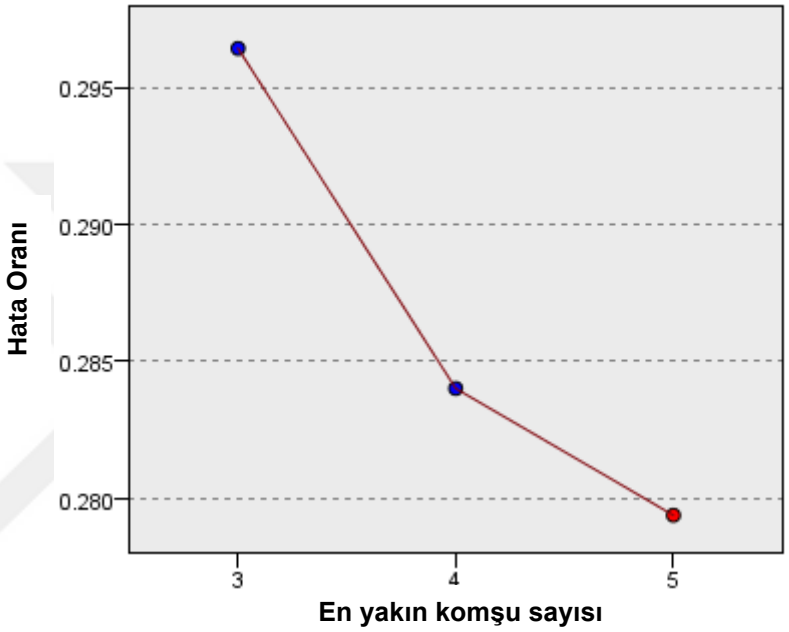
QUEST algoritması analizi bağımsız değişkenlerin önem dereceleri incelendiğinde, "Okuma ve Stratejileri Kullanma (0.195)" değişkeninin okuma becerileri başarısında en önemli belirleyici olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişken ise "Kişisel Refah- Pozitif Etki (0.0001)" olduğu görülmektedir. Böylelikle "Okuma ve Stratejileri Kullanma"daki değişkenliğin okuma becerileri başarısını büyük ölçüde etkilemektedir. Bunun yanı sıra "Kişisel Refah- Pozitif Etki" değişkeni, yani öğrencilerin kendilerini değerlendirdiklerinde sahip olabileceği farklı duygular (sevinçli, neşeli ve mutlu) okuma becerileri başarısını çok az etkilemektedir. Bu durum; öğrencilerin sahip olduğu olumlu etkilerin ve duyguların okuma başarıları için önemli olmadığını, okuma başarısı için öğrencilerin mutluluk durumunun başarılarına katkı sağlamadığını göstermektedir.

Tablo 12

QUEST Algoritması Bağımsız Değişkenlerinin Önem Dereceleri

Bağımsız değişken	Önem
Okuma ve Stratejileri Kullanma	.1966
Özetleme	.1155
Sosyo Ekonomik Durum İndeksi	.0904
PISA Zorluk Derecesini Algılama	.0875
Anlama ve Hatırlama	.0774
Evdeki Eğitimsel Eşyalar	.0721
Bilgi İletişim Teknolojileri (BİT) Kaynakları	.0563
Kültürel Sahiplik (Edebiyat, Şiir, Müzik, Resim, Sanat Kitabı)	.0477
Okuma Zorluk Algısı	.0470
Ailenin Mal Varlığı	.0468
Okumayı Sevme	.0371
Evdeki Eğitim Kaynakları	.0351
BİT Okulda Genel Olarak Kullanma	.0214
Öğrenci Tarafından Algılanan Aile Desteği	.0163
Okuma Yeterlik Algısı	.0160
Disiplin Ortamı	.0129
Okul Dışı BİT Kullanma	.0089
Algılanan Geri Bildirim	.0043
Okula Ait Hissetme	.0037
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	.0032
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki	.0025
Okuldaki İş Birliği Algısı	.0012
Öğretmen Desteği	.0002
Kişisel Refah- Pozitif Etki	.0001

K-en yakın komşuluk analizine ilişkin bulgular ve yorumlar. Analizde doğruluk oranını artırması nedeniyle Manhattan (City Blok) uzaklık ölçüsü kullanılmıştır. 10 katlı çapraz geçerlilik sınaması sonunda en küçük hata oranını sağlayan k değeri 5'tir. Şekil 10'da k değerlerine ait hata grafiği sunulmuştur. Şekil 10'a göre Küme sayısının artırılması sonucunda sınıflama hatasının azalacağı ifade edilebilir.



Şekil 10. Seçilen k değerlerine ait hata oranları grafiği.

K-En Yakın Komşuluk analizi ile okuma becerileri başarısına ilişkin elde edilen analiz sonuçları Tablo 13'te gösterilmiştir.

Tablo 13 incelendiğinde, K-En Yakın Komşuluk yöntemi eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %81.28'lik bir performansla, test örneklemindeki öğrencilerin okuma becerileri başarısını ise %82.15'lik bir performansla doğru tahmin etmiştir. Ayrıca eğitim veri setindeki başarılı öğrencilerin %81.9'unu, test veri setindeki başarılı öğrencilerin ise %83'ünü doğru sınıflandırmıştır. Eğitim veri setindeki başarısız öğrencilerin %80.6'sı doğru sınıflandırılırken, test veri setindeki başarısız öğrencilerin %81.2'si doğru sınıflandırılmıştır. Bu sonuca göre K-En Yakın Komşuluk yönteminin özellikle başarılı öğrencilerin yordanmasında iyi sonuçlar verdiğini söylemek mümkündür.

Tablo 13

K-En Yakın Komşuluk Yöntemi Okuma Becerileri Başarısı Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin edilen		
		Başarısız	Başarılı	Sınıflandırma Doğruluğu (%)
Eğitim	Başarısız	1801	432	80.6
	Başarılı	406	1837	81.9
	Toplam%	%49.31	%50.69	81.28
Test	Başarısız	801	185	81.2
	Başarılı	164	805	83
	Toplam%	%49.36	%50.64	82.15

Tablo 13'e göre doğru sınıflandırılmış örneklem sayısı 5244, toplam örneklem sayısı 6431'dir. K-En Yakın Komşuluk yönteminin genel doğru sınıflandırma oranı %81.5 olarak hesaplanmıştır. Örneklemin maksimum ve nisbi şans kriterleri sırasıyla 0.501 ve 0.499'dur. %81.5 olarak hesaplanan K-En Yakın Komşuluk yöntemi sınıflandırma oranı örneklemin maksimum ve nisbi şans kriterleri üstündedir. Bu değerler K-En Yakın Komşuluk yönteminin bu modelde sınıflandırmada başarılı olarak kullanılabileceğini göstermektedir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 14'te gösterilmiştir.

Tablo 14 incelendiğinde, okuma becerileri başarısına yönelik oluşturulan K-En Yakın Komşuluk yöntemi için en önemli girdi değişkenlerinin "Okuma ve Stratejileri Kullanma (0.0438)" ve "Özetleme (0.0433)" olduğu görülmektedir. Bu iki değişkenin etkisi birbirine çok yakındır. Bu girdi değişkenlerini sırasıyla, "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Kişisel Refah- Pozitif Etki", "Algılanan Geri Bildirim", "PISA Zorluk Derecesini Algılama", "Kültürel Sahiplik", "Okul Dışı BİT Kullanma", "Evdeki Eğitim Kaynakları", "Öğrenci Tarafından Algılanan Aile Desteği", "Disiplin Ortamı", "Okuma Zorluk Algısı", "Sosyo Ekonomik Durum İndeksi", "Okula Ait Hissetme", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Okumayı Sevme", "Anlama ve Hatırlama", "Okuma Yeterlik Algısı", "BİT Okulda Genel Olarak Kullanma", "Evdeki Eğitimsel Eşyalar", "Ailenin Mal Varlığı", "Öğretmenin Öğrenci

Tarafından Algılanan Okuma Katılımı Teşviki", "Öğretmen Desteği" ve "Okuldaki İş Birliği Algısı" değişkenleri takip etmektedir.

Tablo 14

K-En Yakın Komşuluk Yöntemi Analizi Bağımsız Değişkenlerinin Önem Dereceleri

Bağımsız değişken	Önem
Okuma ve Stratejileri Kullanma	.0438
Özetleme	.0433
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	.0423
Kişisel Refah- Pozitif Etki	.0423
Algılanan Geri Bildirim	.0420
PISA Zorluk Derecesini Algılama	.0418
Kültürel Sahiplik	.0416
Okul Dışı BİT Kullanma	.0416
Evdeki Eğitim Kaynakları	.0416
Öğrenci Tarafından Algılanan Aile Desteği	.0416
Disiplin Ortamı	.0415
Okuma Zorluk Algısı	.0415
Sosyo Ekonomik Durum İndeksi	.0415
Okula Ait Hissetme	.0415
Bilgi İletişim Teknolojileri (BİT) Kaynakları	.0414
Okumayı Sevme	.0414
Anlama ve Hatırlama	.0413
Okuma Yeterlik Algısı	.0413
BİT Okulda Genel Olarak Kullanma	.0413
Evdeki Eğitimsel Eşyalar	.0412
Ailenin Mal Varlığı	.0411
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki	.0410
Öğretmen Desteği	.0409
Okuldaki İş Birliği Algısı	.0409

K-En Yakın Komşuluk yöntemi analizi bağımsız değişkenlerin önem dereceleri incelendiğinde, “Okuma ve Stratejileri Kullanma (0.0438)” değişkeninin okuma becerileri başarısında en önemli belirleyici olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenin ise “Okuldaki İş Birliği Algısı (0.0409)” olduğu ortaya çıkmaktadır. Değişkenlerin etkileri birbirine çok yakındır ancak “Okuma ve Stratejileri Kullanma”daki değişkenliğin okuma becerileri başarısını büyük ölçüde etkilemektedir. Bunun yanı sıra “Okuldaki İş Birliği Algısı” değişkeni, yani öğrencilerin öğrenmede kendi aralarındaki iş birliği okuma becerileri başarısı üzerindeki en etkisiz bağımsız değişkendir. Bu durum; öğrencilerin arasındaki iş birliğinin, iş birliğine önem vermenin okuma başarıları için önemli olmadığını, okuma başarılarına katkı sağlamadığını göstermektedir.

Naive bayes analizine ilişkin bulgular ve yorumlar. Yapılan analiz sonucunda Naive Bayes analizi ile okuma becerileri başarısı analiz sonuçları Tablo 15’te gösterilmiştir.

Tablo 15

Naive Bayes Yöntemi Okuma Becerileri Başarısı Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen		
		Başarısız	Başarılı	Sınıflandırma Doğruluğu (%)
Eğitim	Başarısız	1716	517	76.8
	Başarılı	517	1726	76.9
	Toplam%	%49.89	%50.11	76.9
Test	Başarısız	757	229	76.7
	Başarılı	225	744	76.7
	Toplam%	%50.23	%49.77	76.78

Tablo 15 incelendiğinde, Naive Bayes yöntemi eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %76.9’luk bir performansla, test örneklemindeki öğrencilerin okuma becerileri başarısını ise %76.78’lik bir performansla doğru tahmin etmiştir. Ayrıca eğitim veri setindeki başarılı öğrencilerin %76.9’unu, test veri setindeki başarılı öğrencilerin ise %76.7’sini doğru sınıflandırmıştır. Eğitim veri setindeki başarısız öğrencilerin %76.8’i doğru sınıflandırılırken, test veri setindeki başarısız öğrencilerin %76.7’si doğru sınıflandırılmıştır. Bu sonuca göre K-En Yakın Komşuluk yönteminin başarılı ve

başarısız öğrencilerin yordanmasında yakın sonuçlar verdiğini söylemek mümkündür.

Tablo 15'e göre doğru sınıflandırılmış örneklem sayısı 4943, toplam örneklem sayısı 6431'dir. Naive Bayes yönteminin genel doğru sınıflandırma oranı %76.8 olarak hesaplanmıştır. Analizde kullanılan örneklemin maksimum ve nisbi şans kriterleri sırasıyla 0.501 ve 0.499'dur. %76.8 olarak hesaplanan Naive Bayes yöntemi sınıflandırma oranı örneklemin maksimum ve nisbi şans kriterleri üstündedir. Bu sonuca göre Naive Bayes yönteminin bu modelde sınıflandırmada başarılı olarak kullanılabileceği sonucuna varılabilir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 16'da gösterilmiştir.

Tablo 16 incelendiğinde, okuma becerileri başarısına yönelik oluşturulan Naive Bayes yöntemi için en önemli girdi değişkenlerinin "Disiplin Ortamı (0.667)" ve "PISA Zorluk Derecesini Algılama (0.623)" olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, "Okuma Zorluk Algısı", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Kişisel Refah-Pozitif Etki", "Ailenin Mal Varlığı", "Okula Ait Hissetme", "Öğretmen Desteği", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı Teşviki", "Öğrenci Tarafından Algılanan Aile Desteği", "Okuma Yeterlik Algısı", "Kültürel Sahiplik", "Anlama ve Hatırlama", "Sosyo-Ekonomik Durum İndeksi", "Evdeki Eğitimsel Eşyalar", "Evdeki Eğitim Kaynakları", "Okuldaki İş Birliği Algısı", "Algılanan Geri Bildirim", "Okumayı Sevme", "Okul Dışı BİT Kullanma", "Özetleme", "Okuma ve Stratejileri Kullanma", "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma" ve "BİT Okulda Genel Olarak Kullanma" değişkenleri takip etmektedir.

Naive Bayes Komşuluk yöntemi analizi bağımsız değişkenlerin önem dereceleri incelendiğinde, "Disiplin Ortamı (0.667)" değişkeninin okuma becerileri başarısına en önemli belirleyicisi olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenin ise "BİT Okulda Genel Olarak Kullanma (0.367)" olduğu görülmektedir. Bu durum; öğrencilerin okulda bilgi iletişim teknolojilerini kullanmasının, ondan faydalanmasının okuma başarıları için önemli olmadığını, okuma başarılarına katkı sağlamadığını göstermektedir.

Tablo 16

Naive Bayes Yöntemi Analizi Bağımsız Değişkenlerinin Önem Dereceleri

Bağımsız değişken	Önem
Disiplin Ortamı	.6677
PISA Zorluk Derecesini Algılama	.6239
Okuma Zorluk Algısı	.6029
Bilgi İletişim Teknolojileri (BİT) Kaynakları	.5809
Kişisel Refah-Pozitif Etki	.5785
Ailenin Mal Varlığı	.5594
Okula Ait Hissetme	.5415
Öğretmen Desteği	.5352
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki	.5244
Öğrenci Tarafından Algılanan Aile Desteği	.5165
Okuma Yeterlik Algısı	.5065
Kültürel Sahiplik	.5040
Anlama ve Hatırlama	.4977
Sosyo Ekonomik Durum İndeksi	.4949
Evdeki Eğitimsel Eşyalar	.4915
Evdeki Eğitim Kaynakları	.4790
Okuldaki İş Birliği Algısı	.4596
Algılanan Geri Bildirim	.4465
Okumayı Sevme	.4388
Okul Dışı BİT Kullanma	.4170
Özetleme	.4153
Okuma ve Stratejileri Kullanma	.4038
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	.3898
BİT Okulda Genel Olarak Kullanma	.3607

Bu bulgular ışığında örneklemin maksimum ve nisbi şans kriterleri sırasıyla 0.501 ve 0.499 olarak belirlenmiştir. Öğrencilerin 2018 PISA okuma becerileri başarısını etkileyen faktörlere ve okuma becerilerine ilişkin başarı durumlarına göre yapılan analiz sonuçlarına göre, Yapay sinir ağları, Karar Ağaçları algoritmaları, K-En Yakın Komşuluk ve Naive Bayes analizi sonucundaki sınıflandırma oranları örneklemin maksimum ve nisbi şans kriterleri değerlerinin üstündedir. Buna sonuçlara göre Yapay sinir ağlarının, Karar Ağaçları algoritmalarının, K-En Yakın Komşuluk ve Naive Bayes yöntemlerinin öğrencileri başarı durumlarına göre sınıflandırmada başarılı olarak kullanılabileceği görülmektedir. Birinci araştırma probleminde Karar Ağaçları C5.0 algoritması %89.6 ile en yüksek, QUEST algoritması da %75 ile en düşük sınıflama oranına sahiptir. Diğer yöntem ve algoritmaların sınıflama oranlarının birbirine yakın olduğu görülmektedir. K-En Yakın Komşuluk yöntemi, Naive Bayes yöntemi başta olmak üzere diğer yöntemlerden daha yüksek sınıflama oranına sahip olarak ikinci en yüksek orana sahiptir.

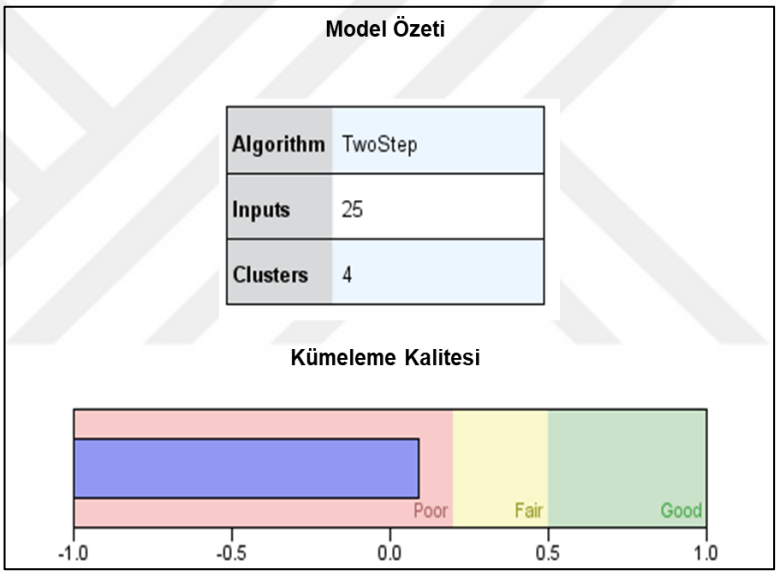
Bu sonuçlar Çalış, Kayapınar ve Çetinyokuş (2014) çalışmalarında veri madenciliğinde sınıflandırma için karar ağaçlarını kullanarak kişilerin demografik yapılarına göre sınıflama doğruluğunu dört karar ağacı algoritmasında test edip, C5.0'in doğru sınıflandırma oranını diğer algoritmalardan daha yüksek değere sahip olduğunu belirttiği çalışma ile paralellik göstermektedir. Liu ve Schumann (2005) tarafından YSA, M5, LR ve KNN yöntemleri karşılaştırılarak kredi puanları hesaplanmış ve en yüksek sınıflama doğruluğu K-En Yakın Komşuluk (KNN) yöntemi ile elde edilmiştir. İstatistiksel süreç kontrolünde kullanılan Shewhart kontrol grafikleri ile ilgili oluşturulan örüntü tanıyıcılarının sınıflandırma performanslarının ölçülmesi için yapılan çalışmada Naive Bayes yöntemi yapay sinir ağlarından daha iyi sınıflandırma yapmıştır (Olgun ve Özdemir, 2012).

Araştırmanın İkinci Alt Problemine İlişkin Bulgular ve Yorumlar

Araştırmanın ikinci alt probleminde öğrencilerin 2018 PISA okuma becerileri başarı durumlarına göre öğrencilere ait verilerin benzer özellikler açısından aynı kümede toplayarak farklı grupları belirlenmesine, gruplardaki değişkenlerin öneminin tespitine çalışılmıştır. Öğrencilerin, okuma becerileri başarılarına göre

kümeleme analizi yöntemiyle gruplandırılması İki Aşamalı Kümeleme algoritması kullanılarak gruplar elde edilmiştir.

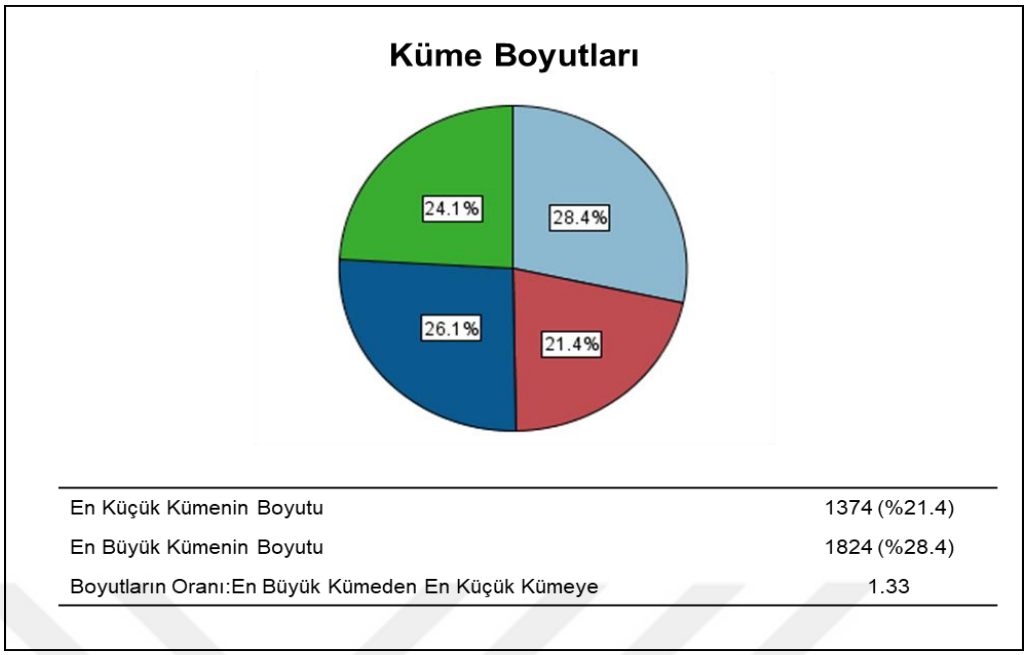
1 bağımlı ve 24 bağımsız indis girdi değişkeninin kullanıldığı iki aşamalı kümeleme analizinde sihoutte katsayısı 0.1 olarak hesaplanmış ve Silhouette katsayısına endeksli kümeleme kalitesi Şekil 11'de gösterilmiştir. Alanyazında Sihoutte katsayısı ile ilgili olarak değerlendirmelerde tam bir kesinlik, tam bir eşik değeri tanımlanmamaktadır. Ancak katsayı değerinin 0'dan büyük olmasının kümele için yeterli olduğu, katsayının ne kadar büyükse kümelemenin kalitesinin o kadar iyi olduğu belirtilmektedir. Bu kapsamda çalışmada iki aşamalı kümeleme analizinde Sihoutte katsayısının 0.1 değeri her ne kadar çok iyi kümeleme yapıldığına işaret etmese de kümeleme yapılması için yeterli olduğu sonucuna varılabilir.



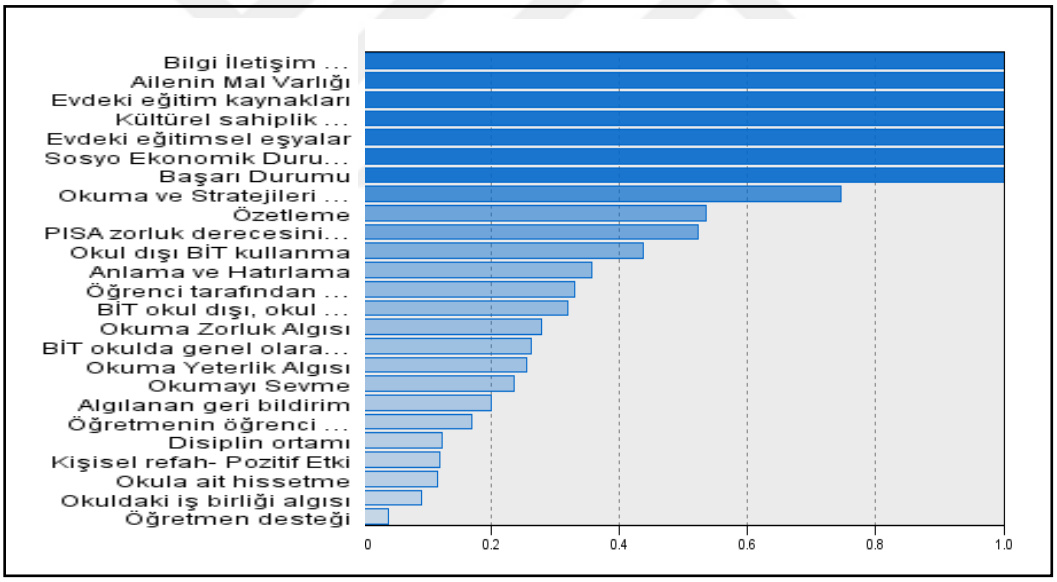
Şekil 11. Silhouette katsayısına endeksli kümeleme kalitesi.

Yapılan kümeleme analizi sonucunda Şekil 12'de görülen dört küme elde edilmiştir ve bu kümelerin dağılımlarının oransal olarak birbirine yakın olduğu belirlenmiştir.

En büyük kümeden en küçük kümeye oran 1.33 olarak bulunmuştur. Bu oranın 2'den küçük olması gerekmektedir. Bu kapsamda kümelerin boyutu ve En büyük kümeden en küçük kümeye oranın uygun olduğu görülmektedir. Kümeleme analizinde önem derecesine göre değişkenler Şekil 13'te gösterilmiştir.



Şekil 12. Kümeleme analizi sonucunda elde edilen kümelerin boyutları.



Şekil 13. Kümeleme analizi bağımsız değişkenlerinin önem dereceleri.

Küme içindeki dağılımlar ve değişkenler incelendiğinde:

Küme 1'de, sadece başarılı öğrencilerin olduğu, "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Algılanan Geri Bildirim", "Okuldaki İş Birliği Algısı" değişkenlerinin en düşük, "Aile Mal Varlığı" "Evdeki Eğitimsel Eşyalar", "Bilgi İletişim Teknolojileri Kaynakları", "Kültürel Sahiplik", "Sosyo Ekonomik Durum İndeksi", "Öğrenci Tarafından Algılanan Aile Desteği", "BİT Okulda Genel Olarak Kullanma",

“Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki” ve “Öğretmen Desteği” değişkenlerinin ortalamanın biraz altında, “Evdeki Eğitim Kaynakları”, “PISA Zorluk Derecesini Algılama”, “Okul Dışı BİT Kullanma”, “Anlama ve Hatırlama”, “Okuma Zorluk Algısı”, “Okuma Yeterlik Algısı”, “Kişisel Refah- Pozitif Etki”, “Okumayı Sevme”, “Disiplin Ortamı” ve “Okula Ait Hissetme” değişkenlerinin ortalama düzeyde, “Okuma ve Stratejileri Kullanma” ve “Özetleme” değişkenlerinin ortalamanın üstünde performans sergilediği görülmektedir.

Küme 2’de, sadece başarılı öğrencilerin olduğu, “BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma” değişkeninin ortalamaya çok yakın olarak üzerinde, “Okuldaki İş Birliği Algısı”, “Kişisel Refah- Pozitif Etki” değişkenlerinin ortalamanın üstünde, “Ailenin Mal Varlığı”, “Bilgi İletişim Teknolojileri (BİT) Kaynakları”, “Evdeki Eğitim Kaynakları”, “Disiplin Ortamı”, “Evdeki Eğitimsel Eşyalar”, “Kültürel Sahiplik”, “Sosyo Ekonomik Durum İndeksi”, “Okuma ve Stratejileri Kullanma”, “Özetleme”, “Okul Dışı BİT Kullanma”, “Öğrenci Tarafından Algılanan Aile Desteği”, “Anlama Ve Hatırlama”, “Okumayı Sevme”, “Algılanan Geri Bildirim”, “Öğretmen Desteği”, “Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki” ve “Okula Ait Hissetme” değişkenlerinin de ortalamanın üstünde, 4 küme içinde bu değişkenlerin en iyi performansı sergilediği, “BİT Okulda Genel Olarak Kullanma” ve “Okuma Yeterlik Algısı” değişkenlerinin ortalama düzeyde, “PISA Zorluk Derecesini Algılama” ve “Okuma Zorluk Algısı” değişkenlerinin ortalamanın çok altında ve 4 küme içinde en düşük performans sergilediği görülmektedir.

Küme 3’te, sadece başarısız öğrencilerin olduğu, “Okuma ve Stratejileri Kullanma”, “Özetleme”, “Anlama ve Hatırlama”, “Okumayı Sevme” ve “Disiplin ortamı” değişkenlerinin ortalamanın altında ancak 4 küme içerisinde en düşük performansı sergilediği, “Okula Ait Hissetme”, “Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki”, “Okuma Yeterlik Algısı”, “Okuma Zorluk Algısı”, “Öğrenci Tarafından Algılanan Aile Desteği”, PISA Zorluk Derecesini Algılama” değişkenlerinin ortalama, “Ailenin Mal Varlığı”, “Bilgi İletişim Teknolojileri (BİT) Kaynakları”, “Evdeki Eğitim Kaynakları”, “Evdeki Eğitimsel Eşyalar”, “Kültürel Sahiplik”, “Sosyo Ekonomik Durum İndeksi”, “Okul Dışı BİT Kullanma”, “Algılanan Geri Bildirim”, “Kişisel Refah- Pozitif Etki”, “Okuldaki İş Birliği Algısı” ve “Öğretmen Desteği” değişkenlerinin ortalamanın üstünde, “BİT Okul Dışı, Okul Aktiviteleri İçin

Kullanma” ve “BİT Okulda Genel Olarak Kullanma” değişkenlerinin ortalamanın üstünde ve 4 küme içinde en iyi performansı sergilediği görülmektedir.

Küme 4’te, sadece başarısız öğrencilerin olduğu, “Okuma Ve Stratejileri Kullanma”, “Anlama Ve Hatırlama”, “BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma” , “Okumayı Sevme”, “Öğretmenin Öğrenci Tarafından Algılanan Okumaya Katılımını Teşviki”, “Özetleme”, “Disiplin Ortamı” ve “Öğretmen Desteği” değişkenleri ortalamanın altında, ancak “Ailenin Mal Varlığı”, “Bilgi İletişim Teknolojileri (BİT) Kaynakları”, “Evdeki Eğitimsel Eşyalar”, “Evdeki Eğitim Kaynakları”, “Kültürel Sahiplik”, “Sosyo Ekonomik Durum İndeksi”, “Okul Dışı BİT Kullanma Öğrenci Tarafından Algılanan Aile Desteği” “Okuma Yeterlik Algısı”, “Kişisel Refah- Pozitif Etki”, “Okula Ait Hissetme” ve “Okuldaki İş Birliği Algısı” değişkenlerinin ortalama altında ve 4 küme içerisinde en düşük performansı sergilediği, “Algılanan Geri Bildirim” değişkenin ortalama düzeyde, “BİT Okulda Genel Olarak Kullanma” değişkenin ortalamanın üstünde ancak “PISA Zorluk Derecesini Algılama” ve “Okuma Zorluk Algısı” değişkenlerinin ortalama üstünde ancak 4 küme içinde en iyi performansı sergilediği görülmektedir.

Bu bulgular ışığında iki aşamalı kümeleme analizindeki Sihoutte Katsayısının 0.1 olarak hesaplanması sonucunda veri setinin kümeleme yapılmaya elverişli olduğu ortaya çıkmaktadır. En büyük kümeden en küçük kümeye oranın 1.33 olarak hesaplanması kümelemenin uygun olduğunu göstermektedir. Bulgulara göre başarılı öğrencilerin 1. ve 2. Kümede toplandığı, değişkenler açısından “BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma”, “Algılanan Geri Bildirim”, “Okuldaki İş Birliği Algısı”, “PISA Zorluk Derecesini Algılama” ve “Okuma Zorluk Algısı” değişkenlerinin başarılı öğrencilerde önemli etki yaratmadığı, “Başarı Durumu”, “Okuma ve Stratejileri Kullanma”, “Özetleme” “Ailenin Mal Varlığı”, “Bilgi İletişim Teknolojileri (BİT) Kaynakları”, “Evdeki Eğitim Kaynakları” değişkenlerinin iyi performans göstererek anlamlı etki yarattığı görülmektedir. Başarısız öğrencilerin ise 3. ve 4. Kümede toplandığı, değişkenler açısından da “Okuma ve Stratejileri Kullanma”, “Özetleme”, “Anlama ve Hatırlama” ve “Okumayı Sevme” değişkenlerinin başarısız öğrencilerde önemli etki yaratmadığı, “BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma” ve “BİT Okulda Genel Olarak Kullanma” değişkenlerinin de iyi performans göstererek anlamlı etki yarattığı ortaya çıkmaktadır.

Araştırmanın Üçüncü Alt Problemine İlişkin Bulgular ve Yorumlar

Araştırmanın birinci alt problemi öğrencilerin 2018 PISA okuma becerileri başarısını etkileyen faktörlere ve okuma becerilerine ilişkin başarı puanlarına göre Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri öğrencileri okuma becerileri yeterlik düzeylerine göre hangi doğruluk oranında sınıflandırmaktadır?

Bu kapsamda üçüncü alt problem dört bölümden oluşmaktadır. Bu nedenle her bölüme ilişkin elde edilen bulgular ve yorumlar dört ayrı başlık altında incelenmiştir.

Yapay sinir ağlarına ilişkin bulgular ve yorumlar. Bu modele 1 bağımlı ve 24 bağımsız indis değişken dahil edilmiştir. Okuma becerileri başarısı yordanırken en iyi performansı veren ağın mimarisini bulabilmek için 50 adet deneme yapılmıştır. Yapılan uygulamalarda değişik miktardaki oluşturulan gizli katmanı ve her katmanda farklı sayıda sinir hücreleri mevcut ağ mimarileri birbirleriyle karşılaştırılmıştır. Bununla birlikte katmanlardaki en iyi sonucu ortaya koyan aktivasyon fonksiyonlarının belirlenmesi amacıyla, her mimari için farklı şekilde oluşturulan fonksiyonlar sınanmıştır. Yapılan bu denemeler sonucunda oluşturulan çok katmanlı algılayıcı ağ üç katmandan meydana gelmektedir. Girdi katmanı olan birinci katmanda 24 adet, ikinci katmanda, yani gizli katmanda 6 adet yapay sinir hücresi bulunmaktadır. Son katman olan çıktı katmanın da ise, bağımlı değişkenin her bir düzeyini (Düzey-1, Düzey-2 ve Düzey-3) temsil eden iki yapay sinir hücresi bulunmaktadır. Gizli katmanda aktivasyon fonksiyonu olarak “Hiperbolik Tanjant Fonksiyon” ve çıktı katmanında ise “Softmax Fonksiyonu” uygulanmıştır.

Yapılan analiz sonucunda yapay sinir ağı ile okuma becerileri yeterlik düzeylerine göre analiz sonuçları Tablo 17’de gösterilmiştir.

Tablo 17

Yapay Sinir Ağları Okuma Becerileri Yeterlik Düzeyleri Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen			
		Düzey 1	Düzey 2	Düzey 3	Sınıflandırma Doğruluğu %
Eğitim	Düzey 1	2104	338	0	86.1
	Düzey 2	458	1473	1	76.2
	Düzey 3	0	92	10	9.8
	Toplam%	%57.2	%42.5	% 0	80
Test	Düzey 1	857	213	0	80.1
	Düzey 2	271	559	1	67.2
	Düzey 3	1	53	0	0
	Toplam%	% 57.7	% 42.1	% 00	72.4

Tablo 17 incelendiğinde, yapay sinir ağı düzeylerine göre eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %80.1'lik, test örneklemindekilerin okuma becerileri başarısını da %72.4'lük bir performansla doğru tahmin etmiştir. Eğitim veri setindeki Düzey 1'deki öğrencilerin %86.1'ini, Düzey 2'deki öğrencilerin %76.2'sini, Düzey 3'teki öğrencilerin %9.8'ini doğru sınıflandırmıştır. Test veri setindeki Düzey 1'deki öğrencilerin %80.1'ini, Düzey 2'deki öğrencilerin %67.2'sini doğru sınıflandırırken, Düzey 3'teki öğrencilerin hiç birini doğru sınıflandıramamıştır. Bu sonuca göre yapay sinir ağının özellikle Düzey 3'teki öğrencilerin yordanmasında iyi sonuçlar veremediğini söylemek mümkündür. Bunun sebebi olarak da Düzey 3'te daha az öğrenci olması olarak ifade edilebilir.

Tablo 17'ye göre doğru sınıflandırılmış örneklem sayısı 5003, toplam örneklem sayısı 6431'dir. Bu sonuca göre genel doğru sınıflandırma oranı %78 olarak hesaplanmıştır. Örneklemin düzey durumlarına göre "Düzey-1", "Düzey-2" ve "Düzey-3" öğrencilerin oranı olan maksimum ve nisbi şans kriterleri sırasıyla 0.55 ve 0.485'dir. %78 olarak hesaplanan Yapay sinir ağları sınıflandırma oranı maksimum ve nisbi şans kriterleri üstündedir ve bu sonuçla oluşturulan modelin sınıflandırma yapmada başarılı bir şekilde kullanılabileceği ifade edilebilir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 18'de gösterilmiştir.

Tablo 18 incelendiğinde, okuma becerileri düzeylerine için meydana getirilen ağ için en önemli girdi değişkenlerinin "Evdeki Eğitimsel Eşyalar (%100)" ve "Ailenin Mal Varlığı (%91.2)" olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, "Sosyo Ekonomik Durum İndeksi", "Okuma ve Stratejileri Kullanma", "PISA Zorluk Derecesini Algılama", "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Okul Dışı BİT Kullanma", "Özetleme", "Disiplin Ortamı", "Okula Ait Hissetme", "BİT Okulda Genel Olarak Kullanma", "Algılanan Geri Bildirim", "Anlama ve Hatırlama", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı", "Okuma Yeterlik Algısı", "Kültürel Sahiplik", "Öğretmen Desteği", "Kişisel Refah- Pozitif Etki", "Okumayı Sevme", "Okuma Zorluk Algısı", "Öğrenci Tarafından Algılanan Aile Desteği", "Evdeki Eğitim Kaynakları" ve "Okuldaki İş Birliği Algısı" değişkenleri takip etmektedir.

Bağımsız değişkenlerin önem dereceleri incelendiğinde, "Evdeki Eğitimsel Eşyalar (%100)" ve "Ailenin Mal Varlığı (%91.2)" değişkenlerinin okuma becerileri başarısında en önemli belirleyiciler olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenler ise "Evdeki Eğitim Kaynakları" ve "Okuldaki İş Birliği Algısı"dır. Bu bulgulara göre "Evdeki Eğitimsel Eşyalar"daki değişkenlik okuma becerileri başarısını büyük ölçüde etkilemektedir. "Okuldaki İş Birliği Algısı" değişkeni, yani öğrencilerin öğrenmede kendi aralarındaki iş birliği ise okuma becerileri başarısı üzerindeki en etkisiz bağımsız değişkendir. Bu durum; öğrencilerin arasındaki iş birliğinin, iş birliğine önem vermenin okuma başarıları için önemli olmadığını, okuma başarılarına katkı sağlamadığını göstermektedir.

Tablo 18

Yapay Sinir Ağları Bağımsız Değişkenlerin Önem Dereceleri

Bağımsız değişken	Önem	Normalleştirilmiş Önem
Evdeki Eğitimsel Eşyalar	0.122	100.0%
Ailenin Mal Varlığı	0.111	91.2%
Sosyo Ekonomik Durum İndeksi	0.089	73.0%
Okuma ve Stratejileri Kullanma	0.064	52.6%
PISA Zorluk Derecesini Algılama	0.062	50.8%
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	0.052	43.0%
Bilgi İletişim Teknolojileri (BİT) Kaynakları	0.051	42.2%
Okul Dışı BİT Kullanma	0.042	34.2%
Özetleme	0.038	31.0%
Disiplin Ortamı	0.038	31.3%
Okula Ait Hissetme	0.033	26.9%
BİT Okulda Genel Olarak Kullanma	0.033	27.0%
Algılanan Geri Bildirim	0.032	25.9%
Anlama ve Hatırlama	0.025	20.7%
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı	0.025	20.6%
Okuma Yeterlik Algısı	0.025	20.8%
Kültürel Sahiplik	0.024	19.4%
Öğretmen Desteği	0.024	19.3%
Kişisel Refah- Pozitif Etki	0.021	17.0%
Okumayı Sevme	0.020	16.4%
Okuma Zorluk Algısı	0.020	16.2%
Öğrenci Tarafından Algılanan Aile Desteği	0.018	14.8%
Evdeki Eğitim Kaynakları	0.016	13.4%
Okuldaki İş Birliği Algısı	0.016	13.1%

Yapılan analiz sonucunda modele ilişkin ROC eğrisi altında kalan alanın değerleri Tablo 19'da gösterilmektedir.

Tablo 19

ROC Analizi Sonucunda Eğri Altında Kalan Alan Değerleri

		Eğri Altında Kalan Alan
Düzey	Düzey 1	0.828
	Düzey 2	0.795
	Düzey 3	0.912

Model tarafından Tablo 19'da görüldüğü gibi en düşük 0.795 ile istatistiksel anlamda kabul edilen ayırma yeteneği ortaya çıkmıştır. Bu analiz sonucu ile model performansının da test edildiği görülmektedir.

Karar ağaçlarına ilişkin bulgular ve yorumlar. Öğrencilerin düzeylerine göre okuma becerileri başarısını yordamak amacıyla, ikinci bölümde yapısı ve çalışma prensipleri ayrıntılı olarak açıklanan dört karar ağacı algoritması analiz sonuçları incelenmiştir. Karar ağaçları algoritması analizine 1 bağımlı ve 24 bağımsız indis değişken dahil edilmiştir.

C5.0 algoritmasına ilişkin bulgular ve yorumlar. Modelde geçerlilik sınaması olarak 10 katlı çapraz geçerlilik testi kullanılmıştır. Modelin oluşturulması aşamasında veri 10 parçaya bölünerek her bir parça üzerinde model türetilip diğer kümelerde test edilmiştir. Modelin çapraz geçerlilik yöntemi ile belirlenen doğruluk oranı ortalama %72.2 ve standart sapması %0.5'dir. Model üretildikten sonra tüm veri üzerinde yapılan testteki doğruluk oranı % 88.5 ve modelin karar ağacının derinliği 21 olarak gerçekleşmiştir.

Yapılan analiz sonucunda C5.0 algoritması ile okuma becerileri yeterlik düzeylerine göre analiz sonuçları Tablo 20'de gösterilmiştir.

Tablo 20

C5.0 Algoritması Okuma Becerileri Yeterlik Düzeyleri Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen			
		Düzey 1	Düzey 2	Düzey 3	Sınıflandırma Doğruluğu %
Eğitim	Düzey 1	2280	162	0	93.3
	Düzey 2	266	1664	2	86.1
	Düzey 3	3	62	37	36.2
	Toplam%	%56.9	%42.1	% 0.01	88.9
Test	Düzey 1	992	77	1	92.7
	Düzey 2	129	701	1	84.3
	Düzey 3	0	36	18	33.3
	Toplam%	% 57.3	% 41.6	% 0.01	87.5

Tablo 20 incelendiğinde, yapay sinir ağı düzeylerine göre eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %88.9'luk bir performansla, test örneklemindeki öğrencilerin okuma becerileri başarısını ise %87.5'lik bir performansla doğru tahmin etmiştir. Eğitim veri setindeki Düzey 1'deki öğrencilerin %93.3'ünü, Düzey 2'deki öğrencilerin %86.1'ini, Düzey 3'teki öğrencilerin %36.2'sini doğru sınıflandırmıştır. Test veri setindeki Düzey 1'deki öğrencilerin %92.7'sini, Düzey 2'deki öğrencilerin %84.3'ünü doğru sınıflandırırken, Düzey 3'teki öğrencilerin %33.3'ünü doğru sınıflandırmıştır. Bu sonuçlara göre Düzey 1'deki öğrencilerin diğer düzeylere göre daha yüksek oranda sınıflandırıldığı görülmektedir.

Tablo 20'ye göre doğru sınıflandırılmış örneklem sayısı 5692, toplam örneklem sayısı 6431'dir. Bu sonuca göre yapay sinir ağının genel doğru sınıflandırma oranı %88.5 olarak hesaplanmıştır. Örneklemin maksimum ve nisbi şans kriterleri sırasıyla 0.55 ve 0.485'dir. %88.5 olarak hesaplanan C5.0 Algoritması sınıflandırma oranı örneklemin maksimum ve nisbi şans kriterleri üstündedir ve böylece modelin sınıflandırma yapmada başarılı bir şekilde kullanılabileceği ifade edilebilir.

Analizde kullanılan bağımsız değişkenlerin düzeylerine göre okuma becerileri başarısı üzerindeki önem dereceleri belirlenerek Tablo 21'de gösterilmiştir.

Tablo 21

C5.0 Algoritması Bağımsız Değişkenlerin Önem Dereceleri

Bağımsız değişken	Önem
Sosyo Ekonomik Durum İndeksi	.0490
Okul Dışı BİT Kullanma	.0460
BİT Okulda Genel Olarak Kullanma	.0448
Anlama ve Hatırlama	.0445
Okuma ve Stratejileri Kullanma	.0438
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	.0417
Disiplin Ortamı	.0417
PISA Zorluk Derecesini Algılama	.0416
Bilgi İletişim Teknolojileri (BİT) Kaynakları	.0414
Okuldaki İş Birliği Algısı	.0410
Ailenin Mal Varlığı	.0410
Öğretmen Desteği	.0410
Evdeki Eğitim Kaynakları	.0410
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı	.0410
Kültürel Sahiplik	.0410
Okuma Yeterlik Algısı	.0410
Okumayı Sevme	.0410
Öğrenci Tarafından Algılanan Aile Desteği	.0409
Okula Ait Hissetme	.0409
Evdeki Eğitimsel Eşyalar	.0409
Kişisel Refah- Pozitif Etki	.0392
Algılanan Geri Bildirim	.0392
Okuma Zorluk Algısı	.0382
Özetleme	.0381

Tablo 21 incelendiğinde, düzeylerine göre okuma becerileri başarısına yönelik oluşturulan C5.0 Algoritması için en önemli girdi değişkenlerinin “Sosyo Ekonomik Durum İndeksi (0.049)” ve “Okul Dışı BİT Kullanma (0.046)” olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, “BİT Okulda Genel Olarak Kullanma”, “Anlama ve Hatırlama”, “Okuma ve Stratejileri Kullanma”, “BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma”, “Disiplin Ortamı”, “PISA Zorluk Derecesini Algılama”, “Bilgi İletişim Teknolojileri (BİT) Kaynakları”, “Okuldaki İş Birliği Algısı”, “Ailenin Mal Varlığı”, “Öğretmen Desteği”, “Evdeki Eğitim Kaynakları”, “Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı”, “Kültürel Sahiplik”, “Okuma Yeterlik Algısı”, “Okumayı Sevme”, “Öğrenci Tarafından Algılanan Aile Desteği”, “Okula Ait Hissetme”, “Evdeki Eğitimsel Eşyalar”, “Kişisel Refah- Pozitif Etki”, “Algılanan Geri Bildirim”, “Okuma Zorluk Algısı” ve “Özetleme” değişkenleri takip etmektedir.

Bu sonuçlara göre oluşturulan C5.0 algoritması analizi bağımsız değişkenlerin önem dereceleri incelendiğinde “Sosyo Ekonomik Durum İndeksi (0.049)” değişkenin okuma becerileri başarısının en önemli belirleyicisi olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenin ise “Özetleme” (0.038)” olduğu ifade edilebilir. “Sosyo Ekonomik Durum İndeksi”ndeki değişkenlik okuma becerileri başarısını büyük ölçüde etkilemektedir. Bunun yanı sıra ““Özetleme” değişkenin, yani öğrencilerin okuduğunu anlamada bir metnin özetini çıkarmasının okuma başarıları için önemli olmadığı, başarılarına katkı sağlamadığı görülmektedir.

CHAID algoritmasına ilişkin bulgular ve yorumlar. Analiz parametreleri en büyük ağaç derinliği 10, kategorik hedef için ki-kare hesaplama metodu Pearson, durdurma kriterleri kök düğüm için %2, alt düğüm için %1 ve en büyük yineleme 100’dür.

Yapılan analiz sonucunda CHAID algoritması ile okuma becerileri yeterlik düzeylerine göre analiz sonuçları Tablo 22’de sunulmuştur.

Tablo 22

CHAID Algoritması Okuma Becerileri Yeterlik Düzeyleri Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen			
		Düzey 1	Düzey 2	Düzey 3	Sınıflandırma Doğruluğu %
Eğitim	Düzey 1	2013	428	1	82.4
	Düzey 2	553	1363	2	71
	Düzey 3	12	66	24	23.5
	Toplam%	%57.4	%41.4	% 0.01	76
Test	Düzey 1	772	296	2	72.1
	Düzey 2	322	497	12	59.8
	Düzey 3	9	39	6	11.1
	Toplam%	% 56.4	% 42.5	% 0.01	65.2

Tablo 22 incelendiğinde, CHAID algoritması eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %76'lık, test örneklemindekilerin okuma becerileri başarısını da %65.2'lik performansla doğru tahmin etmiştir. Eğitim veri setindeki Düzey 1'deki öğrencilerin %82.4'ünü, Düzey 2'deki öğrencilerin %71'ini, Düzey 3'teki öğrencilerin %23.5'ini doğru sınıflandırmıştır. Test veri setindeki Düzey 1'deki öğrencilerin %72.1'ini, Düzey 2'deki öğrencilerin %59.8'ini doğru sınıflandırırken, Düzey 3'teki öğrencilerin %11.1'ini doğru sınıflandırmıştır. Bu sonuçlara göre Düzey 1'deki öğrencilerin diğer düzeylere göre daha yüksek oranda sınıflandırıldığı görülmektedir. Bu sonuca göre CHAID algoritmasının da C5.0 algoritması gibi Düzey 1'deki öğrencilerin yordanmasında iyi sonuçlar verdiğini söylemek mümkündür.

Tablo 22'e göre doğru sınıflandırılmış örneklem sayısı 4675, toplam örneklem sayısı 6431'dir. Bu sonuca göre CHAID algoritmasının genel doğru sınıflandırma oranı %72.7 olarak hesaplanmıştır. 0.55 ve 0.485'dir. %72.7 olarak hesaplanan CHAID sınıflandırma oranı örneklemin maksimum ve nisbi şans

kriterleri üstündedir ve böylece CHAID algoritması modelinin sınıflandırma yapmada başarılı şekilde kullanılabileceği görülmektedir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 23'te gösterilmiştir.

Tablo 23 incelendiğinde, okuma becerileri başarısına yönelik oluşturulan CHAID algoritması için en önemli girdi değişkenlerinin "Okuma ve Stratejileri Kullanma (0.23)" ve "Özetleme (0.10)" olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, "Sosyo Ekonomik Durum İndeksi", "PISA Zorluk Derecesini Algılama", "Evdeki Eğitimsel Eşyalar", "Anlama ve Hatırlama", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Kültürel Sahiplik", "Okuma Zorluk Algısı", "Ailenin Mal Varlığı", "Okumayı Sevme", "Evdeki Eğitim Kaynakları", "BİT Okulda Genel Olarak Kullanma", "Okuma Yeterlik Algısı", "Öğrenci Tarafından Algılanan Aile Desteği", "Disiplin Ortamı", "Okul Dışı BİT Kullanma", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımın", "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Okula Ait Hissetme", "Algılanan Geri Bildirim", "Okuldaki İş Birliği Algısı", "Kişisel Refah-Pozitif Etki" ve "Öğretmen Desteği" değişkenleri takip etmektedir.

CHAID algoritması analizi bağımsız değişkenlerin önem dereceleri incelendiğinde "Okuma ve Stratejileri Kullanma (0.23)" değişkeni okuma becerileri başarısında en önemli belirleyicidir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenin ise "Öğretmen Desteği (0.0002)" olduğu ifade edilebilir. "Okuma ve Stratejileri Kullanma"daki değişkenliğin okuma becerileri başarısını büyük ölçüde etkilediği görülmektedir. Ayrıca "Öğretmen Desteği" değişkeni, yani öğrencilerin öğretmenlerinden aldığı destek okuma becerileri başarısı üzerindeki en etkisiz bağımsız değişkendir. Bu durum; öğretmenlerin öğrencilerin öğrenme konusundaki yardımının, öğretmenin öğrencilerin bir konuyu anlamadaki desteğinin okuma başarıları için önemli olmadığını, okuma başarısı için öğrenme ve anlamda öğretmen desteğinin başarılarına katkı sağlamadığını göstermektedir.

Tablo 23

CHAID Algoritması Bağımsız Değişkenlerin Önem Dereceleri

Bağımsız değişken	Önem
Okuma ve Stratejileri Kullanma	0.2314
Özetleme	0.1031
Sosyo Ekonomik Durum İndeksi	0.0908
PISA Zorluk Derecesini Algılama	0.0828
Evdeki Eğitimsel Eşyalar	0.073
Anlama ve Hatırlama	0.057
Bilgi İletişim Teknolojileri (BİT) Kaynakları	0.0558
Kültürel Sahiplik	0.0509
Okuma Zorluk Algısı	0.0446
Ailenin Mal Varlığı	0.0429
Okumayı Sevme	0.04
Evdeki Eğitim Kaynakları	0.0383
BİT Okulda Genel Olarak Kullanma	0.0243
Okuma Yeterlik Algısı	0.0176
Öğrenci Tarafından Algılanan Aile Desteği	0.0154
Disiplin Ortamı	0.0103
Okul Dışı BİT Kullanma	0.0056
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımın	0.0048
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	0.0048
Okula Ait Hissetme	0.0028
Algılanan Geri Bildirim	0.0016
Okuldaki İş Birliği Algısı	0.0014
Kişisel Refah- Pozitif Etki	0.0004
Öğretmen Desteği	0.0002

C&RT algoritmasına ilişkin bulgular ve yorumlar. Analiz parametreleri en büyük ağaç derinliği 5, en büyük vekil (proxy) sayısı 0 (veri setinde eksik değer olmadığını göstermektedir), kategorik hedef alanı için katışıklık ölçümü Gini, durdurma kriterleri kök düğüm için %2, alt düğüm için %1'dir.

Yapılan analiz sonucunda C&RT algoritması ile okuma becerileri yeterlik düzeylerine göre analiz sonuçları Tablo 24'te sunulmuştur.

Tablo 24

C&RT Algoritması Okuma Becerileri Yeterlik Düzeyleri Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen			
		Düzey 1	Düzey 2	Düzey 3	Sınıflandırma Doğruluğu %
Eğitim	Düzey 1	1923	528	1	78.4
	Düzey 2	515	1387	30	71.7
	Düzey 3	6	72	24	23.5
	Toplam%	%54.6	%44.3	% 1.2	74.2
Test	Düzey 1	782	286	2	73
	Düzey 2	271	552	8	66.4
	Düzey 3	8	41	5	9.2
	Toplam%	% 54.2	% 45	% 0.01	68.4

Tablo 24 incelendiğinde, C&RT algoritması eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %74.2'lik, test örneklemindeki öğrencilerin okuma becerileri başarısını da %68.4'lük performansla doğru tahmin etmiştir. Eğitim veri setindeki Düzey 1'deki öğrencilerin %78.4'ünü, Düzey 2'deki öğrencilerin %71.7'sini, Düzey 3'teki öğrencilerin %23.5'ini doğru sınıflandırmıştır. Test veri setindeki Düzey 1'deki öğrencilerin %73'ünü, Düzey 2'deki öğrencilerin %66.4'ünü doğru sınıflandırırken, Düzey 3'teki öğrencilerin %9.2'sini doğru sınıflandırmıştır. Bu sonuçlara göre Düzey 1'deki öğrencilerin diğer düzeylere göre daha yüksek oranda sınıflandırıldığı görülmektedir. Bu sonuca göre C&RT algoritmasının da CHAID ve C5.0 algoritması gibi Düzey 1'deki öğrencilerin yordanmasında iyi sonuçlar verdiğini söylemek mümkündür.

Tablo 24'e göre doğru sınıflandırılmış örneklem sayısı 4673, toplam örneklem sayısı 6431'dir. Bu sonuca göre C&RT algoritmasının genel doğru sınıflandırma oranı %72.6 olarak hesaplanmıştır. sırasıyla 0.55 ve 0.485'dir. %72.6 olarak hesaplanan C&RT algoritması sınıflandırma oranı örneklemin maksimum ve nisbi şans kriterleri üstündedir ve bu sonuç C&RT algoritması modelin sınıflandırma yapmada başarılı şekilde kullanılabileceğini göstermektedir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 25'te gösterilmiştir.

Tablo 25 incelendiğinde, okuma becerileri başarısına yönelik oluşturulan C&RT algoritması için en önemli girdi değişkenlerinin "Okuma ve Stratejileri Kullanma (0.22)" ve "Özetleme (0.10)" olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, "PISA Zorluk Derecesini Algılama", "Anlama ve Hatırlama", "Sosyo Ekonomik Durum İndeksi", "Evdeki Eğitimsel Eşyalar", "Okumayı Sevme", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Kültürel Sahiplik", "Ailenin Mal Varlığı", "Okuma Zorluk Algısı", "Evdeki Eğitim Kaynakları", "Okuma Yeterlik Algısı", "BİT Okulda Genel Olarak Kullanma", "Disiplin Ortamı", "Öğrenci Tarafından Algılanan Aile Desteği", "Okul Dışı BİT Kullanma", BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı", "Öğretmen Desteği" "Okula Ait Hissetme", "Okuldaki İş Birliği Algısı", "Kişisel Refah- Pozitif Etki" ve "Algılanan Geri Bildirim" değişkenleri takip etmektedir.

Bu sonuçlara göre oluşturulan C&RT algoritması analizi bağımsız değişkenlerin önem derecelerine bakarak "Okuma ve Stratejileri Kullanma (0.22)" değişkeni okuma becerileri başarısının en önemli belirleyicisidir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişken ise "Algılanan Geri Bildirim (0.001)" olduğu ifade edilebilir. "Okuma ve Stratejileri Kullanma"daki değişkenliğin okuma becerileri başarısını büyük ölçüde etkilediği görülmektedir. Bunun yanı sıra "Algılanan Geri Bildirim" değişkeni, yani öğretmen tarafından öğrencilere sağlanan geri bildirimi okuma becerileri başarısı üzerindeki en etkisiz bağımsız değişkendir. Bu durum; öğretmenler tarafından öğrencilere hangi alanda kendilerini ve performanslarını geliştirebilecekleri konusundaki yönlendirme yapılmasının okuma başarıları için önemli olmadığını, okuma başarısı için geri bildirimin başarılarına katkı sağlamadığını göstermektedir.

Tablo 25

C&RT Algoritması Bağımsız Değişkenlerin Önem Dereceleri

Bağımsız değişken	Önem
Okuma ve Stratejileri Kullanma	.221
Özetleme	.103
PISA Zorluk Derecesini Algılama	.091
Anlama ve Hatırlama	.076
Sosyo Ekonomik Durum İndeksi	.075
Evdeki Eğitimsel Eşyalar	.068
Okumayı Sevme	.051
Bilgi İletişim Teknolojileri (BİT) Kaynakları	.049
Kültürel Sahiplik	.046
Ailenin Mal Varlığı	.044
Okuma Zorluk Algısı	.042
Evdeki Eğitim Kaynakları	.031
Okuma Yeterlik Algısı	.022
BİT Okulda Genel Olarak Kullanma	.020
Disiplin Ortamı	.016
Öğrenci Tarafından Algılanan Aile Desteği	.014
Okul Dışı BİT Kullanma	.008
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	.007
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı	.006
Öğretmen Desteği	.004
Okula Ait Hissetme	.003
Okuldaki İş Birliği Algısı	.002
Kişisel Refah- Pozitif Etki	.002
Algılanan Geri Bildirim	.001

QUEST algoritmasına ilişkin bulgular ve yorumlar. Bu algoritmada her düğüm 2 alt gruba bölünmektedir. Analiz parametreleri en büyük ağaç derinliği 10, en büyük vekil (proxy) sayısı 0 (veri setinde eksik değer olmadığını göstermektedir), Alfa (bölme için) 0.05, durdurma kriterleri kök düğüm için %2, alt düğüm için %1'dir.

Yapılan analiz sonucunda QUEST algoritması ile okuma becerileri yeterlik düzeylerine göre analiz sonuçları Tablo 26'da sunulmuştur.

Tablo 26

QUEST Algoritması Okuma Becerileri Yeterlik Düzeyleri Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen			
		Düzey 1	Düzey 2	Düzey 3	Sınıflandırma Doğruluğu %
Eğitim	Düzey 1	1832	513	97	75
	Düzey 2	665	889	378	46
	Düzey 3	3	27	72	70.5
	Toplam%	%55.8	%31.9	% 12.2	62.4
Test	Düzey 1	806	208	56	75.3
	Düzey 2	293	342	196	41.1
	Düzey 3	4	22	28	51.8
	Toplam%	% 56.4	% 29.2	% 14.3	60.15

Tablo 26 incelendiğinde, QUEST algoritması eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %62.4'lük, test örneklemindeki okuma becerileri başarısını da %60.15'lik performansla doğru tahmin etmiştir. Eğitim veri setindeki Düzey 1'deki öğrencilerin %75'ini, Düzey 2'deki öğrencilerin %46'sını, Düzey 3'teki öğrencilerin %70.5'ini doğru sınıflandırmıştır. Test veri setindeki Düzey 1'deki öğrencilerin %75.3'ünü, Düzey 2'deki öğrencilerin %41.1'ini doğru sınıflandırırken, Düzey 3'teki öğrencilerin %51.8'ini doğru sınıflandırmıştır. Bu sonuçlara göre Düzey 1'deki öğrencilerin diğer düzeylere göre daha yüksek oranda sınıflandırıldığı ve diğer C&RT, CHAID ve C5.0 algoritmaları gibi Düzey 1'deki öğrencilerin yordanmasında iyi sonuçlar verdiğini söylemek mümkündür. Ancak QUEST algoritması diğer karar ağacı algoritmalarından farklı olarak Düzey 2 en

düşük doğru sınıflandırmayı yaparken, Düzey 3'te de karar ağaçlarından en yüksek doğru sınıflandırmayı yapmıştır.

Tablo 26'ya göre doğru sınıflandırılmış örneklem sayısı 3969, toplam örneklem sayısı 6431'dir. Bu sonuca göre QUEST algoritmasının genel doğru sınıflandırma oranı %61.7 olarak hesaplanmıştır. QUEST algoritması örnekleminin maksimum ve nisbi şans kriterleri sırasıyla 0.55 ve 0.485'dir. Çok yüksek olmasa da %61.7 olarak hesaplanan QUEST algoritması sınıflandırma oranı maksimum ve nisbi şans kriterleri üstündedir ve bu durum modelin sınıflandırma yapmada başarılı şekilde kullanılabileceğini göstermektedir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 27'de gösterilmiştir.

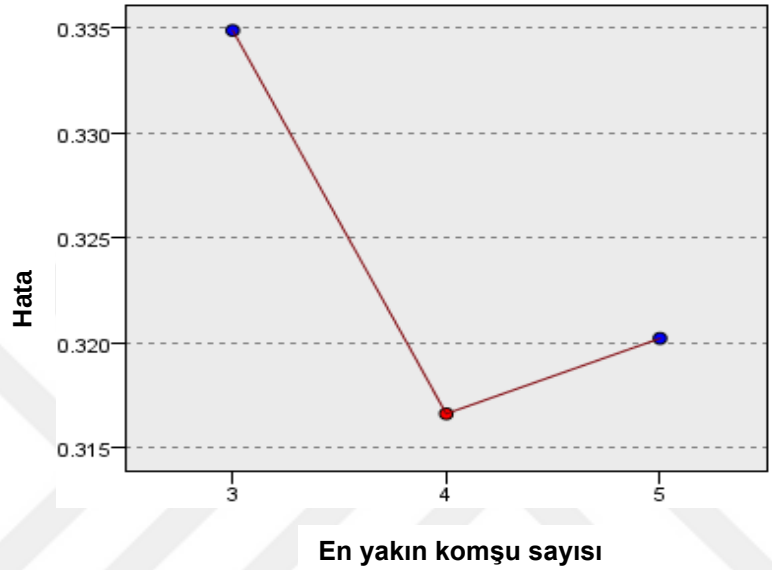
QUEST algoritması analizi bağımsız değişkenlerin önem dereceleri incelendiğinde "Okuma ve Stratejileri Kullanma (0.151)" değişkeninin okuma becerileri başarısının en önemli belirleyicisi olduğunu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenin ise "Kişisel Refah- Pozitif Etki (0.002)" olduğu ifade edilebilir. "Okuma ve Stratejileri Kullanma"daki değişkenlik okuma becerileri başarısını büyük ölçüde etkilemektedir. Bunun yanı sıra "Kişisel refah- Pozitif Etki" değişkeni, yani öğrencilerin kendilerini değerlendirdiklerinde sahip olabileceği farklı duygular (sevinçli, neşeli ve mutlu) okuma becerileri başarısı üzerindeki en etkisiz bağımsız değişkendir. Bu durum; öğrencilerin sahip olduğu olumlu etkilerin ve duyguların okuma başarıları için önemli olmadığını, okuma başarısı için öğrencilerin mutluluk durumunun başarılarına katkı sağlamadığını göstermektedir.

Tablo 27

QUEST Algoritması Bağımsız Değişkenlerin Önem Dereceleri

Bağımsız değişken	Önem
Okuma ve Stratejileri Kullanma	0.151
Özetleme	0.151
Sosyo Ekonomik Durum İndeksi	0.138
Evdeki Eğitimsel Eşyalar	0.105
Ailenin Mal Varlığı	0.079
Bilgi İletişim Teknolojileri (BİT) Kaynakları	0.075
PISA Zorluk Derecesini Algılama	0.062
Kültürel Sahiplik	0.056
Evdeki Eğitim Kaynakları	0.049
Anlama ve Hatırlama	0.043
Okuma Zorluk Algısı	0.021
Okumayı Sevme	0.021
Okuma Yeterlik Algısı	0.013
Öğrenci Tarafından Algılanan Aile Desteği	0.010
Okul Dışı BİT Kullanma	0.004
BİT Okulda Genel Olarak Kullanma	0.004
Disiplin Ortamı	0.003
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı	0.002
Okula Ait Hissetme	0.001
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	0.001
Okuldaki İş Birliği Algısı	0.001
Algılanan Geri Bildirim	0.001
Öğretmen Desteği	0.004
Kişisel Refah- Pozitif Etki	0.002

K-en yakın komşuluk analizine ilişkin bulgular ve yorumlar. Analizde Manhattan (City Blok) uzaklık ölçüsü kullanılmıştır. 10 katlı çapraz geçerlilik test sonucunda en düşük hata oranını sağlayan k değeri olarak 4 bulunmuştur. Şekil 14'te seçilen çeşitli k değerlerine ait hata grafiği verilmiştir. Şekil 14'e göre Küme sayısının artırılması sonucunda sınıflama hatasının azalacağı ifade edilebilir.



Şekil 14. Seçilen k değerlerine ait hata oranları grafiği.

K-En Yakın Komşuluk yöntemi ile okuma becerileri yeterlik düzeylerine göre analiz sonuçları Tablo 28'de sunulmuştur.

Tablo 28 incelendiğinde, K-En Yakın Komşuluk yöntemi eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %79.11'lik bir performansla, test örneklemindeki öğrencilerin okuma becerileri başarısını ise %79.28'lik bir performansla doğru tahmin etmiştir. Eğitim veri setindeki Düzey 1'deki öğrencilerin %92'sini, Düzey 2'deki öğrencilerin %65.8'ini, Düzey 3'teki öğrencilerin %20.5'ini doğru sınıflandırmıştır. Test veri setindeki Düzey 1'deki öğrencilerin %91.4'ünü, Düzey 2'deki öğrencilerin %67.6'sını doğru sınıflandırırken, Düzey 3'teki öğrencilerin %16.6'sını doğru sınıflandırmıştır. Bu sonuçlara göre Düzey 1'deki öğrencilerin diğer düzeylere göre daha yüksek oranda sınıflandırıldığı görülmektedir.

Tablo 28

K-En Yakın Komşuluk Yöntemi Okuma Becerileri Yeterlik Düzeyleri Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen			
		Düzey 1	Düzey 2	Düzey 3	Sınıflandırma Doğruluğu %
Eğitim	Düzey 1	2248	190	4	92
	Düzey 2	658	1272	2	65.8
	Düzey 3	9	72	21	20.5
	Toplam%	%65.1	%34.2	% 0.06	79.11
Test	Düzey 1	979	91	0	91.4
	Düzey 2	265	562	4	67.6
	Düzey 3	8	37	9	16.6
	Toplam%	% 64	% 35.2	% 0.06	79.28

Tablo 28'e göre doğru sınıflandırılmış örneklem sayısı 5091, toplam örneklem sayısı 6431'dir. Bu sonuca göre K-En Yakın Komşuluk yönteminin genel doğru sınıflandırma oranı %79.1 olarak hesaplanmıştır. Örneklemin sırasıyla 0.55 ve 0.485'dir. %79.1 olarak hesaplanan K-En Yakın Komşuluk yöntemi sınıflandırma oranı maksimum ve nisbi şans kriteri üstündedir ve bu sonuçla sınıflandırmada modelin başarılı şekilde kullanılabileceği ifade edilebilir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 29'da gösterilmiştir.

Tablo 29

K-En Yakın Komşuluk Yöntemi Bağımsız Değişkenlerin Önem Dereceleri

Bağımsız değişken	Önem
Okuma ve Stratejileri Kullanma	0.0443
Özetleme	0.0430
Sosyo Ekonomik Durum İndeksi	0.0423
Bilgi İletişim Teknolojileri (BİT) Kaynakları	0.0420
Ailenin Mal Varlığı	0.0419
Evdeki Eğitimsel Eşyalar	0.0418
Okuma Zorluk Algısı	0.0418
Okul Dışı BİT Kullanma	0.0417
BİT Okulda Genel Olarak Kullanma	0.0417
PISA Zorluk Derecesini Algılama	0.0417
Öğretmen Desteği	0.0416
Anlama ve Hatırlama	0.0416
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı	0.0416
Okula Ait Hissetme	0.0415
Okumayı Sevme	0.0415
Kültürel Sahiplik	0.0415
Algılanan Geri Bildirim	0.0414
Evdeki Eğitim Kaynakları	0.0413
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	0.0413
Okuma Yeterlik Algısı	0.0412
Kişisel Refah- Pozitif Etki	0.0411
Disiplin Ortamı	0.0409
Okuldaki İş Birliği Algısı	0.0408
Öğrenci Tarafından Algılanan Aile Desteği	0.0404

Bu sonuçlara göre oluşturulan K-En Yakın Komşuluk yöntemi analizi bağımsız değişkenlerin önem dereceleri incelendiğinde “Okuma ve Stratejileri Kullanma (0.0443)” değişkenin okuma becerileri başarısının en önemli belirleyicisi olduğu ifade edilebilir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişken ise “Öğrenci Tarafından Algılanan Aile Desteği (0.0404)” değişkenidir. “Okuma ve Stratejileri Kullanma”daki değişkenliğin okuma becerileri başarısını büyük ölçüde etkilediği görülmektedir. Ancak değişkenlerin değerleri birbirine çok yakındır. Bunun yanı sıra “Öğrenci Tarafından Algılanan Aile Desteği” değişkeni, yani öğrencilerin ailelerinin onların başarılarını ve çabalarını desteklemesi, zorluklar karşısında cesaretlendirmesi okuma becerileri başarısı üzerindeki en etkisiz bağımsız değişkendir. Bu durum; aileler tarafından öğrencilere sağlanan desteğin okuma başarıları için önemli olmadığını, okuma başarısı için öğrencilerin mutluluk durumunun başarılarına katkı sağlamadığını göstermektedir.

Naive bayes analizine ilişkin bulgular ve yorumlar. Yapılan analiz sonucunda Naive Bayes yöntemi ile okuma becerileri yeterlik düzeylerine göre analiz sonuçları Tablo 30’da gösterilmiştir.

Tablo 30

Naive Bayes Yöntemi Okuma Becerileri Yeterlik Düzeyleri Sınıflama Doğruluğu Yüzdeleri

Örneklem	Gözlenen	Tahmin Edilen			
		Düzey 1	Düzey 2	Düzey 3	Sınıflandırma Doğruluğu %
Eğitim	Düzey 1	1925	514	3	78.8
	Düzey 2	518	1377	37	71.2
	Düzey 3	0	56	46	45
	Toplam%	%54.5	%43.4	% 1	74.8
Test	Düzey 1	851	218	1	79.5
	Düzey 2	229	586	16	70.5
	Düzey 3	1	32	21	38.8
	Toplam%	% 55.2	% 42.7	% 0.1	74.6

Tablo 30 incelendiğinde, Naive Bayes Yöntemi yöntemi eğitim örneklemindeki öğrencilerin okuma becerileri başarısını %74.8'lik, test örneklemindeki okuma becerileri başarısını da %74.6'lIk bir performansla doğru tahmin etmiştir. Düzeylerine göre Naive Bayes Yöntemi analizi sonuçlarına göre eğitim ve test verilerinin sınıflandırma doğruluk oranları birbirine çok yakındır. Eğitim veri setindeki Düzey 1'deki öğrencilerin %79.5'ini, Düzey 2'deki öğrencilerin %70.5'ini, Düzey 3'teki öğrencilerin %45'ini doğru sınıflandırmıştır. Test veri setindeki Düzey 1'deki öğrencilerin %79.5'ini, Düzey 2'deki öğrencilerin %70.5'ini doğru sınıflandırırken, Düzey 3'teki öğrencilerin %38.8'ini doğru sınıflandırmıştır.

Tablo 30'a göre Naive Bayes yönteminin genel doğru sınıflandırma oranı %74.7 olarak hesaplanmıştır. Örneklemin maksimum ve nisbi şans puanları sırasıyla 0.55 ve 0.485'dir. %74.7 olarak hesaplanan Naive Bayes algoritması sınıflandırma oranı maksimum ve nisbi şans puanları üstündedir ve bu sonuç modelin sınıflandırmada başarılı olarak kullanılabileceğini göstermektedir.

Analizde kullanılan bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem derecelerinin belirlenerek Tablo 31'de gösterilmiştir.

Tablo 31 incelendiğinde, okuma becerileri başarısına yönelik oluşturulan Naive Bayes yöntemi için en önemli girdi değişkenlerinin "Disiplin Ortamı (0.589)" ve "Ailenin Mal Varlığı" (0.562)" olduğu görülmektedir. Bu girdi değişkenlerini sırasıyla, "Okuma Zorluk Algısı", "PISA Zorluk Derecesini Algılama", "Kişisel Refah-Pozitif Etki", "Okuma Yeterlik Algısı", "Okula Ait Hissetme", "Öğretmen Desteği", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı", "Öğrenci Tarafından Algılanan Aile Desteği", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Okuma ve Stratejileri Kullanma", "Anlama ve Hatırlama", "Sosyo-Ekonomik Durum İndeksi", "Kültürel Sahiplik", "Evdeki Eğitim Kaynakları", "Okuldaki İş Birliği Algısı", "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma", "Evdeki Eğitimsel Eşyalar", "Okumayı Sevme", "Özetleme", "BİT Okulda Genel Olarak Kullanma", "Algılanan Geri Bildirim" ve "Okul Dışı BİT Kullanma" değişkenleri takip etmektedir.

Tablo 31

Naive Bayes Yöntemi Bağımsız Değişkenlerin Önem Dereceleri

Bağımsız değişken	Önem
Disiplin Ortamı	.5893
Ailenin Mal Varlığı	.5620
Okuma Zorluk Algısı	.5587
PISA Zorluk Derecesini Algılama	.5501
Kişisel Refah-Pozitif Etki	.5412
Okuma Yeterlik Algısı	.5144
Okula Ait Hissetme	.5111
Öğretmen Desteği	.5032
Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımı	.4744
Öğrenci Tarafından Algılanan Aile Desteği	.4698
Bilgi İletişim Teknolojileri (BİT) Kaynakları	.4645
Okuma ve Stratejileri Kullanma	.4519
Anlama ve Hatırlama	.4423
Sosyo Ekonomik Durum İndeksi	.4332
Kültürel Sahiplik	.4305
Evdeki Eğitim Kaynakları	.4285
Okuldaki İş Birliği Algısı	.4217
BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma	.4197
Evdeki Eğitimsel Eşyalar	.4184
Okumayı Sevme	.4170
Özetleme	.4138
BİT Okulda Genel Olarak Kullanma	.4102
Algılanan Geri Bildirim	.3658
Okul Dışı BİT Kullanma	.3307

Naive Bayes Komşuluk yöntemi analizi bağımsız değişkenlerin önem dereceleri incelendiğinde, “Disiplin Ortamı (0.589)” değişkeninin okuma becerileri başarısında en önemli belirleyici olduğu görülmektedir. Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişkenin ise “Okul Dışı BİT Kullanma (0.330)” olduğu ifade edilebilir. Bu durum; öğrencilerin okulda dışında bilgi iletişim teknolojilerini kullanmasının, ortak çevrimiçi oyunlar oynama, e-posta kullanma, sosyal ağlara katılma, internette haber okuma, internetten pratik bilgiler edinme gibi aktivitelerinin ondan okuma başarıları için önemli olmadığını, okuma başarılarına katkı sağlamadığını göstermektedir.

Bu bulgular ışığında örneklemin maksimum ve nisbi şans kriterleri sırasıyla 0,55 ve 0,485 olarak belirlenmiştir. Öğrencilerin 2018 PISA okuma becerileri başarısını etkileyen faktörlere ve okuma becerilerine ilişkin başarı düzeylerine göre yapılan analiz sonuçlarına göre, Yapay sinir ağları, Karar Ağaçları algoritmaları, K-En Yakın Komşuluk ve Naive Bayes analizi sonucundaki sınıflandırma oranları örneklemin maksimum ve nisbi şans kriterleri değerlerinin üstündedir. Buna sonuçlara göre Yapay sinir ağlarının, Karar Ağaçları algoritmalarının, K-En Yakın Komşuluk ve Naive Bayes yöntemlerinin öğrencileri başarı durumlarına göre sınıflandırmada başarılı olarak kullanılabileceği görülmektedir. Üçüncü araştırma probleminde Karar Ağaçları C5.0 algoritması %88.5 ile en yüksek, QUEST algoritması da %72.7 ile en düşük sınıflama oranına sahiptir. Diğer yöntem ve algoritmaların sınıflama oranlarının birbirine yakın olduğu görülmektedir. K-En Yakın Komşuluk yöntemi, Naive Bayes yöntemi başta olmak üzere diğer yöntemlerden daha yüksek sınıflama oranına sahip olarak ikinci en yüksek orana sahiptir.

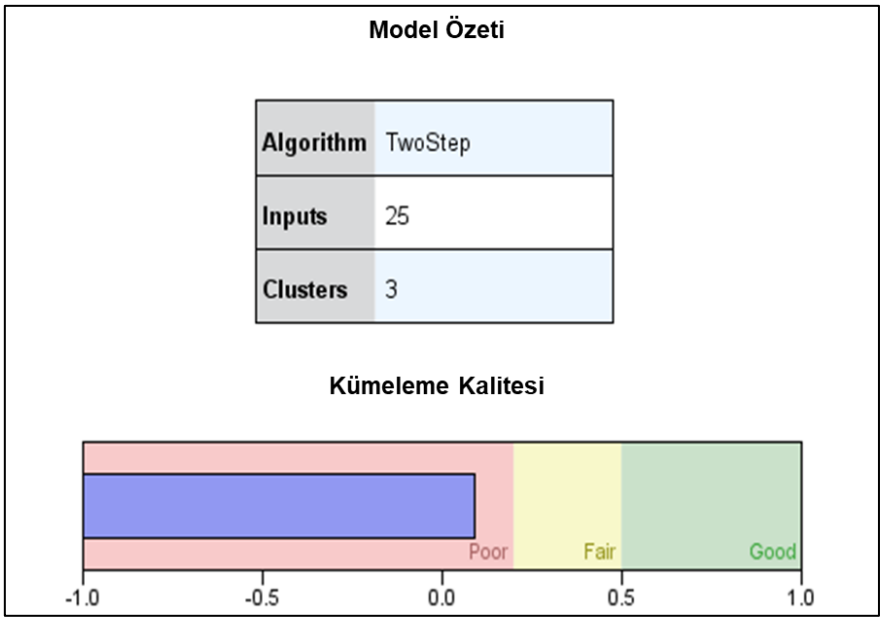
Üçüncü araştırma problemi ile ilgili elde edilen bulgular Destek Vektör Makineleri yöntemi, Yapay Sinir Ağları ve Karar Ağaçları algoritmaları ile yapılan sınıflandırma çalışmasında OKS testi puanı tahmin eden bir modelde C5 karar ağacı algoritması en iyi tahmin edici çıkan çalışmanın bulgularıyla benzerlik göstermektedir (Şen, Uçar ve Delen, 2012). Bu araştırma probleminde ortaya konan bulgular 29 tahmine dayalı değişkenden faydalanarak iki ortaokulun öğrencilerinin başarılı olup olmadığını tahmin için yapılan ve en çok kullanılan veri madenciliği algoritmalarından dördünün sınıflama doğruluğu incelendiğinde karar ağacı algoritmalarının sınıflama doğruluğu en yüksek olarak belirlenen çalışma ile de

örtüşmektedir (Cortez ve Silva, 2008). Kredi kart sahiplerine ilişken gerçek kredi kartı işlemleri üzerine Naive Bayesian yöntemi, Destek Vektör yöntemi ve K-En Yakın Komşuluk (KNN) veri madenciliği ile ilgili çalışmada K-En Yakın Komşuluk yönteminin en yüksek sınıflama doğruluğuna sahip olduğu belirtilmektedir (Hazım, 2018).

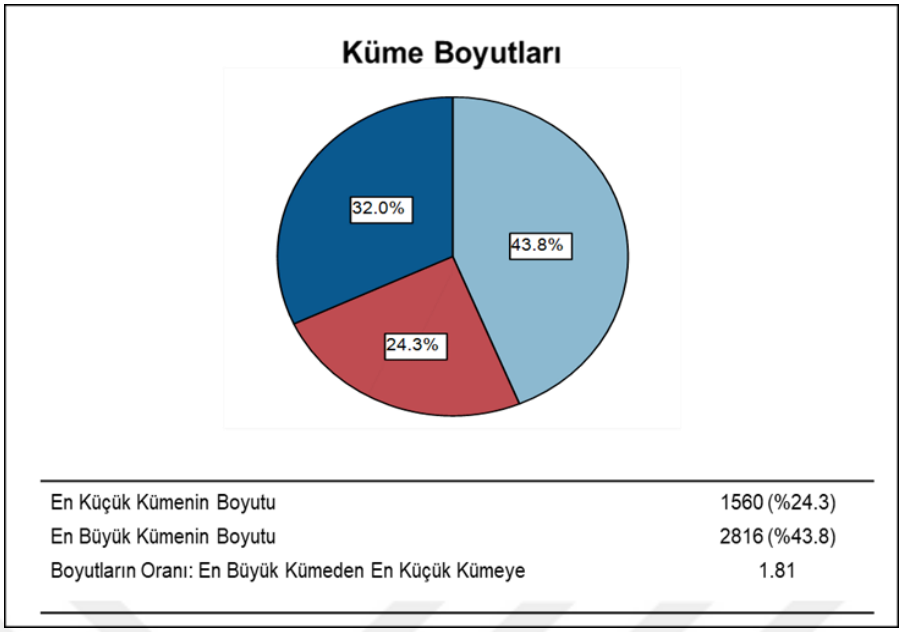
Araştırmanın Dördüncü Alt Problemine İlişkin Bulgular ve Yorumlar

Araştırmanın dördüncü alt probleminde öğrencilerin 2018 PISA okuma becerileri yeterlik düzeylerine göre öğrencilere ait verilerin benzer özelliklere açısından aynı kümede toplayarak farklı grupları belirlenmesine ve bu gruplarda değişkenlerin öneminin tespitine çalışılmıştır. Öğrencilerin, okuma becerileri başarılarına göre kümeleme analizi yöntemiyle gruplandırılması İki Aşamalı Kümeleme algoritması kullanılarak gruplar elde edilmiştir.

Düzeylerin gösterildiği 1 bağımlı ve 24 bağımsız indis girdi değişkeninin kullanıldığı iki aşamalı kümeleme analizinde sihoutte katsayısı 0.1 Şekil 15'te gösterilmiştir. Katsayı değerinin 0'dan büyük olmasının kümele için yeterli olduğu, katsayının ne kadar büyükse kümelemenin kalitesinin o kadar iyi olduğu belirtilmektedir. Bu kapsamda analizindeki Sihoutte katsayısının 0.1 değeri her ne kadar çok iyi kümeleme yapıldığına işaret etmese de kümeleme yapılması için yeterlidir.



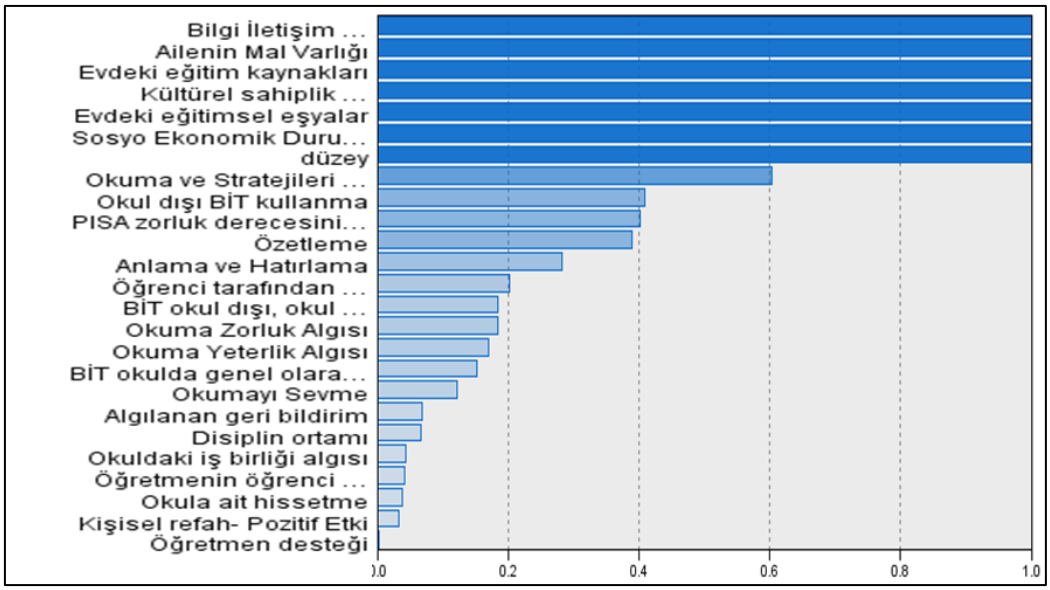
Şekil 15. Silhouette katsayısına endeksli kümeleme kalitesi.



Şekil 16. Kümeleme analizi sonucunda elde edilen kümelerin boyutları.

Kümeleme analizi sonuçlarına göre Şekil 16'da görüldüğü gibi 3 adet küme elde edilmiş, ancak elde edilen kümelerin dağılımları oransal açıdan birbirlerine yakın değildir.

En büyük kümeden en küçük kümeye oran 1.81 olarak bulunmuştur. Bu oranın 2'den küçük olması gerekmektedir. Bu kapsamda kümelerin boyutu ve En büyük kümeden en küçük kümeye oranın uygun olduğu görülmektedir. Kümeleme analizinde önem derecesine göre değişkenler Şekil 17'de gösterilmiştir.



Şekil 17. Kümeleme analizi bağımsız değişkenlerinin önem dereceleri.

Kümeleme analizinde önem derecesine göre değişkenler incelendiğinde ikinci araştırma probleminde olduğu gibi “Bilgi İletişim Teknolojileri (BİT) Kaynakları”, “Ailenin Mal Varlığı”, “Evdeki eğitim Kaynakları”, “Kültürel Sahiplik”, “Evdeki Eğitimsel Eşyalar”, “Sosyo Ekonomik Durum İndeksi” ve “Düzey” değişkenleri önem dereceleri 1 olarak en ayırt edici değişkenler olarak ortaya çıkmıştır. Ancak diğer değişkenlerin önem dereceleri ikinci araştırma problemindekinden farklı bulunmuştur. “Okuldaki İş Birliği Algısı”, “Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki”, “Okula Ait Hissetme”, “Kişisel Refah- Pozitif Etki” ve “Öğretmen Desteği” en düşük ayırt ediciliğe sahip değişkenler olarak belirlenmiştir. Ancak “Öğretmen Desteği”nin önem derece değeri yok denecek kadar az bulunmuştur.

Küme içindeki dağılımlar ve değişkenler incelendiğinde:

Küme 1’de, sadece Düzey-1 ve en fazla öğrencinin olduğu, “Okul Dışı BİT Kullanma”, “Okumayı Sevme”, “Disiplin Ortamı”, “Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki”, “Okula Ait Hissetme”, “Öğretmen Desteği” değişkenlerinin ortalamanın biraz altında, “Aile Mal Varlığı”, “Bilgi İletişim Teknolojileri Kaynakları”, “Evdeki Eğitimsel Eşyalar”, “Evdeki Eğitim Kaynakları”, “Kültürel Sahiplik”, “Sosyo Ekonomik Durum İndeksi”, “Özetleme”, “Anlama ve Hatırlama”, “Öğrenci Tarafından Algılanan Aile Desteği”, “Okuma Yeterlik Algısı”, “Okuldaki İş Birliği Algısı” değişkenlerinin ortalamanın altında ve 3 küme içerisinde en düşük performansı sergilediği, “Okuma ve Stratejileri Kullanma”, BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma”, “Algılanan Geri Bildirim”, “Kişisel Refah- Pozitif Etki” değişkenlerinin ortalama, “PISA Zorluk Derecesini Algılama”, “Okuma Zorluk Algısı”, “BİT Okulda Genel Olarak Kullanma” değişkenlerinin ortalamanın üstünde performans sergilediği görülmektedir.

Küme 2’de, her üç düzeyden de ve en az öğrencinin olduğu, “PISA Zorluk Derecesini Algılama” değişkeni ortalama altında, “BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma”, “BİT Okulda Genel Olarak Kullanma”, “Okumayı Sevme”, “Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki”, “Okula Ait Hissetme”, “Öğretmen Desteği”, “Okuma ve Stratejileri Kullanma”, “Özetleme” değişkenlerinin ortalamanın üstünde, “Ailenin Mal Varlığı”, “Bilgi İletişim Teknolojileri (BİT) Kaynakları”, “Evdeki Eğitim Kaynakları”, “Evdeki Eğitimsel Eşyalar”, “Kültürel Sahiplik”, “Sosyo Ekonomik Durum İndeksi”, “Okul Dışı BİT Kullanma”, “Öğrenci

Tarafından Algılanan Aile Desteği", "Algılanan Geri Bildirim", "Okuldaki İş Birliği Algısı", "Kişisel Refah- Pozitif Etki" değişkenlerinin ortalamanın üstünde ve 3 küme içinde bu değişkenlerin en iyi performansı sergilediği, "Okuma Zorluk Algısı", "Anlama Ve Hatırlama", "Okuma Yeterlik Algısı" , "Disiplin Ortamı" değişkenlerinin de ortalama performans sergilediği görülmektedir.

Küme 3'te, sadece Düzey-2 öğrencilerin olduğu, "Ailenin Mal Varlığı", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Kültürel Sahiplik", "BİT Okulda Genel Olarak Kullanma", "Algılanan Geri Bildirim", "Sosyo Ekonomik Durum İndeksi", "BİT Okul Dışı, Okul Aktiviteleri İçin Kullanma" değişkenlerinin ortalamanın altında, "Evdeki Eğitim Kaynakları", "Okuma Zorluk Algısı", "PISA Zorluk Derecesini Algılama", "Okuma Yeterlik Algısı", "Okuldaki İş Birliği Algısı", "Disiplin ortamı", "Okula Ait Hissetme", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki", "Okul Dışı BİT Kullanma", "Evdeki Eğitimsel Eşyalar", "Kişisel Refah- Pozitif Etki" ve "Öğretmen Desteği" değişkenlerinin ortalama düzeyde, "Öğrenci Tarafından Algılanan Aile Desteği", "Okumayı Sevme" değişkenlerinin ortalamanın üstünde, "Özetleme", "Anlama ve Hatırlama", "Okuma ve Stratejileri Kullanma" değişkenlerinin ortalamanın üstünde ve 3 küme içinde en iyi performansı sergilediği görülmektedir.

Bu bulgular ışığında iki aşamalı kümeleme analizindeki Sihoutte Katsayısının 0.1 olarak hesaplanması sonucunda veri setinin kümeleme yapılmaya elverişli olduğu ortaya çıkmaktadır. En büyük kümeden en küçük kümeye oranın 1.81 olarak hesaplanması kümelemenin uygun olduğunu göstermektedir. Bulgulara göre Düzey-1 öğrencilerin 1. ve 2. Kümede toplandığı, değişkenler açısından, "PISA Zorluk Derecesini Algılama", "Okuma Zorluk Algısı", "BİT Okulda Genel Olarak Kullanma" değişkenlerinin iyi performans göstererek anlamlı etki yarattığı görülmektedir. Düzey-2 öğrencilerin ise 2. ve 3. Kümede toplandığı, değişkenler açısından da "Özetleme", "Anlama ve Hatırlama", "Okuma ve Stratejileri Kullanma" değişkenleri önemli etki göstermektedir. Düzey-3 öğrencilerin ise 2. Kümede toplandığı, "Ailenin Mal Varlığı", "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Evdeki Eğitim Kaynakları", "Evdeki Eğitimsel Eşyalar", "Kültürel Sahiplik", "Sosyo Ekonomik Durum İndeksi", "Okul Dışı BİT Kullanma", "Öğrenci Tarafından Algılanan Aile Desteği", "Algılanan Geri Bildirim", "Okuldaki İş Birliği Algısı", "Kişisel Refah- Pozitif

Etki” değişkenlerinin de iyi performans göstererek anlamlı etki yarattığı görülmektedir.

Araştırmanın Beşinci Alt Problemine İlişkin Bulgular ve Yorumlar

Araştırmanın beşinci alt probleminde Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri analizi sonuçlarına göre öğrencileri başarı durumuna ve okuma becerileri yeterlik düzeylerine göre doğru sınıflandırma oranlarının karşılaştırılmasına ait sonuçlar ortaya konmuştur. Analiz yöntemleri ve algoritmaları tarafından kullanılan matematiksel modellerin ve sınıflandırma problemine yaklaşım metotları değerlendirildiğinde, karşılaştırma esnasında yalnızca sınıflandırmada ortaya çıkan başarıların dikkate alınması sonuçların karşılaştırılması bakımından önemli olarak görülmüştür.

Öğrencilerin başarı puanları dikkate alınarak, “başarı durumların”a göre doğru sınıflandırma oranları Tablo 32’de gösterilmiştir.

Tablo 32

Başarı Durumlarına Göre Sınıflandırma Sonuçları

Yöntem		Sınıflandırma Oranı
Yapay Sinir Ağları		75.4
Karar Ağacı	C5.0	89.6
	CHAID	78.2
	C&RT	76.6
	QUEST	75
K-En Yakın Komşuluk		81.5
Naive Bayes		76.8

Tablo 32 incelendiğinde araştırma içinde kullanılan veri madenciliği yöntemleri ve algoritmaları dikkate alındığında yapılan bütün yöntemlerin ve algoritmaların analizin başarılı olduğu görülmektedir. Karar Ağaçları yönteminde C5.0 algoritmasının %89.6 ile en yüksek sınıflandırma oranı ortaya çıkmaktadır. İkinci en yüksek oran %81.5 ile K-En Yakın Komşuluk yöntemidir. QUEST algoritması %75 ile en düşük sınıflama oranına sahiptir. Bununla birlikte diğer yöntem ve algoritmaların sınıflama oranlarının birbirine yakın olduğu görülmektedir.

PISA 2018 uygulamasına katılan Türk öğrencilerin başarı puanları dikkate alınarak, okuma becerileri yeterlik düzeylerine göre doğru sınıflandırma oranları Tablo 33'te gösterilmiştir.

Tablo 33

Okuma Becerileri Yeterlik Düzeylerine Göre Sınıflandırma Sonuçları

Yöntem		Sınıflandırma Oranı
Yapay Sinir Ağları		78
Karar Ağacı	C5.0	88.5
	CHAID	72.7
	C&RT	72.6
	QUEST	61.7
K-En Yakın Komşuluk		79.1
Naive Bayes		74.7

Tablo 33 incelendiğinde araştırma içinde kullanılan veri madenciliği yöntemleri ve algoritmaları dikkate alındığında yapılan bütün yöntemlerin ve algoritmaların analizin başarılı olduğu görülmektedir. Karar Ağaçları yönteminde C5.0 algoritması %88.5 ile en yüksek sınıflandırma oranına sahiptir. İkinci en yüksek oran %79.1 ile K-En Yakın Komşuluk yöntemidir. QUEST algoritması %61.7 ile en düşük sınıflama oranına sahiptir.

Bu bulgulara göre son dönemde yöntem ve algoritma karşılaştırılması için yapılan çalışmalara yönelik çeşitli eleştiriler olmakla Tablo 32 ve Tablo 33 karşılaştırıldığında Karar Ağaçları yönteminde C5.0 algoritmasının en yüksek, QUEST algoritmasının da en düşük oranda sınıflandırma yaptığı ortaya çıkmaktadır. Ayrıca PISA 2018 uygulamasına katılan Türk öğrencilerin başarı durumlarına ve başarı düzeylerine göre yapılan analiz sonuçları örneklemlerin maksimum şans kriterinin ve nisbi şans kriterinin üstündedir ve bu sonuçlar kullanılan modellerin sınıflandırmada başarılı olarak kullanılabileceğini göstermektedir. Sigorta riskini tahmin performansları karşılaştırılması amacıyla yapılan Yapay Sinir Ağları ve Karar Ağaçları yöntemleriyle elde edilen modellerin karşılaştırıldığı çalışmada her iki yöntemin de kabul edilebilir seviyede olmakla birlikte Karar Ağaçları yönteminin tahmin başarısı, bu çalışmada ortaya çıkan sonuçla paralel olarak daha yüksek bulunmuştur (Şahin, 2018). Biyoteknoloji çalışmalarında, proteinlerin hücre tahminini kapsamında geliştirilen modelin başarı

oranının incelenmesi sonucunda K-En Yakın Komsu (KNN) analizi başarı oranı diğer veri madenciliği algoritmalarına göre çok daha yüksek bulunmuştur (Cai ve Chou, 2003)

Bölüm 5
Sonuç, Tartışma ve Öneriler

Bu bölümde araştırma bulgularına ve yorumlara dayalı ortaya konan sonuçların özeti ve bu sonuçlara göre oluşturulan önerilere yer verilmiştir.

Sonuç ve Tartışma

Bu çalışmada PISA 2018 Türkiye örneklemine dayalı olarak öğrencilerin okuma becerileri başarısını etkileyen faktörlere ve okuma becerileri başarı puanlarına göre öğrencilerin başarı durumlarının ve okuma yeterlik düzeylerinin Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk ve Naive Bayes yöntemleri ile sınıflama doğrulukları incelenmiştir. Bununla birlikte büyük hacimli PISA 2018 veri setindeki öğrencilere ait verilerin benzer niteliklere göre benzer kümede toplanarak farklı grupların belirlenmesi ve bu gruplardaki değişkenlerin önemi belirlenmiştir.

Yapılan analizler sonucundaki bulgular değerlendirildiğinde öğrencilerin başarı durumlarına göre Karar Ağaçları C5.0 algoritmasının %89.6 ile en yüksek, QUEST algoritmasının da %75 ile en düşük sınıflama oranına sahip olarak diğer yöntem ve algoritmaların sınıflama oranlarının birbirine yakın olduğu görülmektedir. K-En Yakın Komşuluk yöntemi, Naive Bayes yöntemi başta olmak üzere diğer yöntemlerden daha yüksek sınıflama oranına sahip olarak ikinci en yüksek orana sahiptir. Bu bulgular öğrenci kimlik bilgileri, önceki başarı durumları ve elektronik öğrenme verileri girdi olarak uygulanan tahmin modellerinin C5.0, Logistic Regresyon, Yapay Sinir Ağları, Karar Ağacı algoritmaları çalıştırılarak elde edilen analiz sonuçlarındaki C5.0 ile ortaya konan karar ağacı modelinin en iyi tahmin yapan olarak seçilen çalışma ile örtüşmektedir (Aydın, 2007). Hayashi, Hsieh ve Setiono (2009) tarafından yapılan, karar ağacı ve yapay sinir ağlarının yakın oranda sınıflama yaptığı ancak değişkenler ile ilgili olarak daha çok açıklayıcı olması sebebiyle karar ağaçlarının yapısının ve performanslarının incelenmesinin önemli görüldüğü çalışma sonuçları ile paralel bulunmaktadır.

Örneklemin başarı durumlarına göre maksimum ve nisbi şans puanı sırasıyla 0.501, ise 0.499'dur. Analiz sonucundaki sınıflandırma oranlarının örneklemin maksimum şans kriteri ve nisbi şans kriteri değerlerinde yüksek olmasından dolayı bu Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk (KNN) ve Naive Bayes

yöntemlerinin öğrencileri başarı durumlarına göre sınıflandırma yapmada kullanılabileceği ve oluşturulan modellerin şansın ötesinde doğru sınıflandırma yaptığı görülmektedir. Okuma başarı puanlarının öğrencileri başarı durumlarına göre ayırmada etkili olduğu, anlamlı bir fark yarattığı sonucu ortaya konulmuştur.

Başarı durumlarına göre bağımsız değişkenlerin okuma becerileri başarısı üzerindeki önem dereceleri incelendiğinde "Sosyo-Ekonomik Durum İndeksi" değişkeninin C5.0 algoritması analizinde Okuma becerileri başarısı üzerinde en az etkiye sahip bağımsız değişken olduğu görülmektedir. Etkisi her zaman yüksek olan "Sosyo-Ekonomik Durum İndeksi"nin etkisinin düşük olmasının diğer değişkenlerle birlikte değerlendirilmesinden kaynaklandığı ve bu durumun manidar olmadığı sonucuna varılabilir. Bu doğrultuda başarı durumlarına göre "Öğretmen Desteği" değişkenin CHAID, C&RT, QUEST ve K-En Yakın Komşuluk analizlerinde okuma becerileri başarısı üzerindeki önem derecesinin düşük kaldığı ve en etkisiz bağımsız değişken olduğu sonucu ortaya çıkmaktadır. Ancak her zaman önemli ve yüksek etkiye sahip "Öğretmen Desteği" değişkenin diğer değişkenlerle beraber değerlendirilmesi sonucunda düşük etkiye sahip olduğu ifade edilebilir.

Öğrencilerin 2018 PISA okuma başarı durumlarına göre başarı gruplarının genel karakteristiği incelenmesi için gruplandırılması sonucunda dört küme ortaya çıkmış ve dört kümenin oransal olarak dağılımları birbirine yakın çıkmıştır. Sihoutte Katsayısı (0.1) 0 değerinden büyük olduğu için veri setinin kümeleme yapılmaya elverişli ve uygun olduğu görülmektedir. Önem derecesine göre değişkenler incelendiğinde "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Ailenin Mal Varlığı", "Evdeki eğitim Kaynakları", "Kültürel Sahiplik", "Evdeki Eğitimsel Eşyalar", "Sosyo Ekonomik Durum İndeksi" ve "Başarı Durumu" değişkenlerinin önem dereceleri 1 olarak kümelemede etkili olduğu ve anlamlı fark yaratarak en ayırt edici değişkenler olduğu ortaya konmuştur. "Disiplin Ortamı", "Kişisel Refah- Pozitif Etki", "Okula Ait Hissetme", "Okuldaki İş Birliği Algısı" ve "Öğretmen Desteği" değişkenlerinin ise en düşük ayırt ediciliğe sahip olarak ayırmada etkili olmadığı ve anlamlı fark yaratmadığı ortaya çıkmıştır.

Tsai, Tsai, Hung ve Hwang (2011) tarafından yeterlilik sınavından başarısız olacak öğrenciler tahmin edilmesi için yapılan ve öğrencileri kümelere ayırmada iki aşamalı kümeleme, k-ortalamalar, öz düzenleme haritaları teknikleri uygulanan çalışmada en iyi kümeleme yönteminin iki aşamalı kümeleme analizi olduğu tespit

edilmiştir. Bu çalışmada bulduğumuz sonuçları ile Tsai, Tsai, Hung ve Hwang (2011) tarafından yapılan çalışmanın sonuçları benzerlik göstermektedir. Eğitim alanında ve başarı durumları ile ilgili konulardaki çalışmalarda iki aşamalı kümeleme analizinin daha yüksek sonuçlar verdiği ifade edilebilir. Ayrıca Önen (2018) tarafından yapılan ve matematik başarısının kümelerin oluşmasında en etkili özelliklerin olduğu, öğretimsel niteliklerin de en az etkisi olan özellikler olduğu görülen çalışma sonuçları ile bu çalışmada bulunan sonuçlar örtüşmektedir. Başarı durumunun öğrencileri kümelere ayırmada en etkili değişkenlerden biri olduğu görülmektedir.

Öğrencilerin okuma becerilerine yeterlik düzeylerine göre Karar Ağaçları C5.0 algoritmasının %88.6 ile en yüksek, QUEST algoritması da %61.7 ile en düşük sınıflama oranına sahip olarak, diğer yöntem ve algoritmaların sınıflama oranlarının birbirine yakın olduğu görülmektedir. Bu bulgular Fransa ve İtalya'daki iki üniversitede gerçekleştirilen öğrenci memnuniyeti için ana faktörlerin tanımlandığı veri madenciliği yöntemlerinin uygulanabilirliğinin araştırıldığı ve bu amaçla karar ağacı algoritmalarının, lojistik regresyon ve diğer bir çok veri madenciliği tekniğinin uygulandığı, analiz sonucunda da karar ağacı algoritmalarından C5.0 algoritmasının en yüksek sınıflama doğruluğuna sahip olduğunun belirtildiği çalışma ile benzerlik göstermektedir (Dejaeger, Goethals, Giangreco, Mola ve Baesens, 2012).

Okuma becerilerine yeterlik düzeylerine göre örneklemin maksimum ve nisbi şans kriteri sırasıyla 0.55 ve 0.485 olarak hesaplanmıştır. Analizde hesaplanan sınıflandırma oranlarının bu değerlerin üzerinde olmasından dolayı Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk (KNN) ve Naive Bayes yöntemlerinin öğrencileri başarı düzeylerine göre sınıflandırma yapmada kullanılabileceği ve oluşturulan modellerin şansın ötesinde doğru sınıflandırma yaptığı görülmektedir. Okuma başarı puanlarının öğrencileri başarı düzeylerine göre ayırmada etkili olduğu, anlamlı bir fark yarattığı sonucu ortaya konulmuştur.

Okuma becerilerine yeterlik düzeylerine göre genel karakteristiğin incelenmesi sonucunda üç küme elde edilmiştir. Ancak 3 kümenin oransal olarak dağılımları birbirine yakın değildir. 0,1 olarak hesaplanan Sihoutte Katsayısının 0 değerinden büyük olduğu için veri setinin kümelemeye elverişli olduğu ve uygun olduğu ifade edilebilir. Önem derecesine göre değişkenler incelendiğinde "Bilgi İletişim Teknolojileri (BİT) Kaynakları", "Ailenin Mal Varlığı", "Evdeki eğitim

Kaynakları", "Kültürel Sahiplik", "Evdeki Eğitimsel Eşyalar", "Sosyo Ekonomik Durum İndeksi" ve "Düzey" değişkenleri önem dereceleri 1 olarak kümelemede etkili olduğu ve anlamlı fark yaratarak en ayırt edici değişkenler olduğu ortaya konmuştur. "Okuldaki İş Birliği Algısı", "Öğretmenin Öğrenci Tarafından Algılanan Okuma Katılımını Teşviki", "Okula Ait Hissetme", "Kişisel Refah- Pozitif Etki" ve "Öğretmen Desteği" değişkenlerinin ise en düşük ayırt ediciliğe sahip olarak ayırmada etkili olmadığı ve anlamlı fark yaratmadığı ortaya çıkmıştır.

Son dönemde veri madenciliği çalışmalarının birbirlerinden farklı sonuçlar ortaya koyduğu görülmektedir. Bu nedenle yöntem ve algoritma karşılaştırılması amacıyla yapılan çalışmalar çeşitli eleştirilere maruz kalmaktadır. Bu alandaki algoritma karşılaştırması ile ortaya çıkan sonuçların doğru olmayacağı, çalışmaların gerçekte illüzyon yarattığı, deneysel çalışma sonuçlarının gerçekle bağdaşmayabileceği belirtilmektedir (Hand, 2006). Modelde ortaya konan başarı oranlarının yanlı olabileceği, verilerin işlenmesi aşamasındaki işlemlerin ve algoritma parametrelerinin sonuçlarda farklı tesir yaratabilecektir. Michie ve Spiegelhalter (1994) kitaplarında çalışmaların sonuçlarını yayınlamış, benzer ve aynı verilerde bazı algoritmaların daha etkili olduğunu göstermişlerdir. Ancak hangi algoritma ve yöntemin daha iyi sonuçlar veren model ürettiğini araştırmak için başka veri setlerinde fazla sayıda algoritma yöntemle çalışılarak sınıflandırılması ve karşılaştırılması gerekmektedir. Veri madenciliği çalışmalarında en iyi analiz sonuçlarını sağlayacak algoritma seçimi zordur ve önemlidir (Romero ve Ventura, 2007). Bundan dolayı en çok tercih edilen metot, farklı sınıflandırmama algoritmalarını kullanıp, en yüksek doğru sınıflandırma oranını sağlayan algoritmanın seçimidir (Lopez, Luna, Romerove ve Ventura, 2012; Aydın, 2017).

Veri madenciliği sınıflandırma yöntemlerinin karşılaştırması amacıyla Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk (KNN) ve Naive Bayes yöntemlerinin öğrencileri başarı durumlarına ve okuma becerileri yeterlik düzeylerine göre sınıflandırma sonuçları değerlendirildiğinde son dönemde yöntem ve algoritma karşılaştırılması için yapılan çalışmalara yönelik çeşitli eleştiriler olmakla birlikte bu çalışmada Karar Ağaçları yönteminde C5.0 algoritmasının en yüksek, QUEST algoritmasının da en düşük oranda sınıflandırma yaptığı ortaya çıkmaktadır. Bu bulgu Cristobal ve Sebastian (2013) tarafından sınıflama modelleri kullanılarak, öğrencilerin eğitimlerinin ilk yıllarında başarısız olma risklerinin

tanımlanması üzerine yapılan çalışmada Logistik Regresyon, Naive Bayes Yöntemi, Karar Ağaçları Algoritmaları, Sinir Ağları ve K-En Yakın Komşuluk (KNN) analizleri sonucunda Karar Ağaçları ve K-En yakın Komşuluk yöntemlerinin en yüksek sınıflandırma oranına sahip olduğu yönünde belirtilen sonuçla örtüşmektedir. Bu çalışmada bulunan sonuçlar Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk (KNN) ve Naive Bayes yöntemlerinin eğitimde de başarılı olarak kullanılabileceğini ifade etmektedir.

Öneriler

Uygulamaya dönük öneriler. Bu çalışmada 2018 PISA okuma becerileri başarı puanlarına göre veri madenciliği sınıflama yöntemlerinden Yapay Sinir Ağları, Karar Ağaçları, K-En Yakın Komşuluk (KNN) ve Naive Bayes incelenmiştir. Benzer şekilde aynı veri setine diğer sınıflama yöntemleri ile inceleme yapılabilir.

Çalışmada ilk aşamada eksik ve kayıp veriler tamamlanmış, ardından analizler yapılmıştır. Veri madenciliğinde sınıflandırma yöntemleri analizleri eksik verilerle de gerçekleştirilebilmektedir. Ancak kayıp verilerin tamamlanmasından sonra analiz sonuçlarının daha yüksek olduğu görülmüştür. Bu doğrultuda veri madenciliği yöntemleri ile çalışırken daha yüksek geçerlik ve doğrulukta sınıflama yapmak için kayıp verilerin analizi yapılıp tamamlanması önerilebilir.

Çalışmada yapılan inceleme sonucunda öğrencilerin okuma becerileri başarısını etkilediği düşünülen 24 indis değişken belirlenerek, analizler bu değişkenlerle gerçekleştirilmiştir. Araştırmalarda farklı sayıda değişkenlerin kullanıldığı analizler yapılabilir.

Çalışmadaki YSA modellerinde Çok Katmanlı Algılayıcı (ÇKA) ağ kullanılmıştır. SPSS Modeler yazılımının sunduğu bu ağ dışında sık kullanılan Radyal Tabanlı Fonksiyon (RTF) ile diğer modellerin kullanılması sonucu sağlanan sonuçlar kıyaslanabilir. Ancak alanyazındaki çalışmaların sonuçları göz önüne alındığında, ÇKA'nın RTF'ler ile karşılaştırıldığında yüksek kestirim oranına sahip olmasından dolayı bu tür çalışmalarda ÇKA kullanılması önerilebilir.

Çalışmadaki yöntemlerin analizinde her uygulamada bir bölümün eğitim ve bir bölümün test seti şeklinde belirleme yapılması sonucunda sınıflama performansları farklılık göstermektedir. Bu durum göz önüne alındığında çok fazla deneme yapılarak en yüksek performansa ulaşılabileceği ifade edilebilir.

Çalışmada 4 farklı veri madenciliği sınıflama yöntemi kullanılmıştır. Girdi değişkenlerin farklı yöntemlerle elde edilen sonuçlarında farklılık olduğu görülmüştür. Bu durum göz önüne alındığında eğitim ile ilgili yapılan veri madenciliği çalışmalarında 3-4 farklı yöntem kullanarak elde edilen sonuçların geçerliğine ilişkin kanıt sunulmalıdır.

Alanyazında yapılan çalışmalar çoğunlukla 2 kategorili (başarılı:1, başarısız:0) olacak şekilde yapılmış, sınıflama doğruluğu öğrencilerin başarı durumlarına göre incelenmiştir. Bu çalışmada da olduğu gibi kategori sayısının artırılarak üç, dört veya beş kategorili sınıflama doğruluğu çalışmaları yapılabilir.

Örneklemin fazla olduğu durumlarda bu çalışmada olduğu gibi test edilen veri setinin %30 ve üzeri olması önerilebilir. Bu hususun sağlanmaması durumunda tutarlı sonuçlar elde edilemeyebileceği dikkate alınmalıdır.

Araştırmacılara dönük öneriler. Analizler için çalışmada SPSS Modeler programı kullanılmıştır. Veri madenciliği analiz programları çok fazladır. Programların karşılaştırılması amacıyla aynı veri setleri kullanılarak veri madenciliği yöntemleri ve analiz programlarının karşılaştırılması yapılabilir. Bununla birlikte TIMMS, PIRLS vb. gibi farklı verilerin olduğu sınavlar ile ilgili de benzer çalışmalar yapılabilir.

Bu çalışmada 2018 PISA okuma becerileri başarı puanlarına göre veri madenciliği sınıflama yöntemleri incelenmiştir. Benzer şekilde 2018 PISA matematik ve fen başarı puanlarına göre veri madenciliğindeki diğer yöntem ve algoritmaların performanslarını ölçen araştırmalar yapılabilir.

Bu araştırma ile aynı b çalışmalar farklı eğitim kademelerinde uygulanarak araştırmacılar kendi çalışmalarında da bu hususlara dikkat edebilirler.

Veri madenciliğine yönelik özellikle sanayii, bankacılık gibi sektörel bazda yoğun olarak çalışmalar yapılmaktadır. Eğitim alanında derlenen verilerin kullanılması, bu alanda başarı elde edilebilmesi ve öğrenci başarısının arttırılmasında merkezi bir öneme sahip olmasına rağmen eğitim alanında veri

madenciliği az araştırma yapılan alanlardandır ve eğitim alanındaki araştırmaların yıllardır ihmal edildiği görülmektedir. Veri madenciliği konusunda eğitim alanında diğer sektörlere oranla çok sayıda çalışma ve uygulama bulunmamaktadır. Veri madenciliği yöntemlerinin alışagelmiş sektörel baz dışında eğitim alanında daha çok çalışma yapılması gerekmektedir.

Veri madenciliğinde oldukça çok çalışma yapılan yordama ve sınıflamada farklı yöntemler ve algoritmalar bulunmaktadır. Ancak yapılan çalışmalar incelendiğinde Yapay Sinir Ağlarının ve Karar Ağaçlarının en çok çalışılan yöntemler olduğu ortaya çıkmaktadır. Aynı ya da benzer örneklem üzerinde diğer araştırmalarda diğer sınıflama yöntemleri (Regresyon Destek Vektör Makineleri, K-Ortalamalar, Zaman Serisi Analizleri,) vasıtasıyla başarılar tahmin edilebilir.

Veri madenciliğinde sınıflandırma yöntemleri analizleri eksik verilerle de gerçekleştirilebilmektedir. Bu kapsamda aynı veya benzer veri setleri diğer sınıflama yöntemleri analizlerinin kayıp verilerde nasıl performans gösterdiği incelenebilir.

Eğitim alanında yapılan araştırmalar farklı örneklem büyüklüklerinde, farklı değişkenler dahil edilerek gerçekleştirilmektedir. Bu sebeple kullanılan örneklemde hangi yöntem ve algoritmanın daha iyi performans gösterdiğinin belirlenmesi için çok sayıda yöntem ve algoritmanın kullanılarak sınıflandırılma ve karşılaştırılma yapılması gerekmektedir.

Kaynaklar

Abdullahi, A. H. (2018). *An intrusion detection approach based on binary particle swarm optimization and Naive Bayes*. (Doktora tezi). Selçuk Üniversitesi, Konya.

Abraham, A. (2005). Artificial neural networks. *Handbook of measuring system design*.

Akpınar, H. (2000). Veri tabanlarında bilgi keşfi ve veri madenciliği. *İÜ İşletme Fakültesi Dergisi*, *29*(1), 1-22.

Akpınar, H. (2014). *Data: Veri madenciliği veri analizi*. İstanbul: Papatya Yayıncılık Eğitim.

Aksu, G. ve Güzeller, C. O. (2016). PISA 2012 matematik okuryazarlığı puanlarının karar ağacı yöntemiyle sınıflandırılması: Türkiye örneklemi. *Eğitim ve Bilim*, *41*(185), 101-122.

Altan, Ş., Atan, M. ve Kızılkaya, S. (2015). Genel sağlık durumunu etkileyen faktörlerin CHAID analizi yöntemi ile incelenmesi, ODTÜ örneği. *Social Sciences, 10*(3), 92-106.

Altıntaş, T. (2006). *Veri madenciliği metotlarından kümeleme algoritmalarının uygulamalı etkinlik analizi* (Yüksek lisans tezi). Sakarya Üniversitesi, Sakarya.

Anderson, D., & McNeill, G. (1992). Artificial neural networks technology. *Kaman Sciences Corporation, 258*(6), 1-83.

Anıl, D. (2008). The analysis of factors affecting the mathematical success of Turkish students in the PISA 2006 evaluation program with structural equation modeling. *American-Eurasian Journal of Scientific Research*, *3*(2), 222-227.

Anıl, D. (2009). Uluslararası öğrenci başarılarını değerlendirme programında (PISA) Türkiye'deki öğrencilerin fen bilimleri başarılarını etkileyen faktörler. *Eğitim ve Bilim*, *34*(152), 87-100.

Anıl, D. (2011). Türkiye'nin PISA 2006 fen bilimleri başarısını etkileyen faktörlerin yapısal eşitlik modeli ile incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, *11*(3), 1253-1266.

Arabacı, G. (2007). *Veri madenciliğinde appriori, tahminci appriori ve tertius algoritmalarının WEKA ve YALE programları ile karşılaştırılması ve bir uygulama* (Yüksek lisans tezi). İstanbul Ticaret Üniversitesi, İstanbul.

Aydemir, Ö. (2013). *İmlecin iki boyutlu hareketinin hayali sırasında kaydedilmiş EEG işaretlerinin karar ağaç yapısı esaslı sınıflandırılması* (Doktora tezi). Karadeniz Teknik Üniversitesi, Trabzon.

Aydın, A., Erdağ, C. ve Taş, N. (2011). 2003-2006 PISA okuma becerileri sonuçlarının karşılaştırmalı olarak değerlendirilmesi: en başarılı beş ülke ve Türkiye. *Kuramdan Uygulamaya Eğitim Bilimleri, 11*(2), 651-673.

Aydın, S. (2007). *Veri madenciliği ve Anadolu Üniversitesi uzaktan eğitim sisteminde bir uygulama* (Yayınlanmamış doktora tezi). Anadolu Üniversitesi, Eskişehir.

Baker, L. and Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly, 34* (1), 452–497

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of educational data mining*, *1*(1), 3-17.

Baş, N. (2006). *Yapay sinir ağları yaklaşımı ve bir uygulama*. (Yüksek Lisans Tezi). Mimar Sinan Güzel Sanatlar Üniversitesi, İstanbul.

Bayru, P. (2007). *Elektronik basında tüketici tercihleri analizi: Yapay sinir ağları ile lojit modelin performans değerlendirilmesi (*Doktora Tezi). İstanbul Üniversitesi, İstanbul.

Berry, M. J. & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. New Jersey: John Wiley & Sons.

Berthold, M., & Hand, D. J. (2003). *Intelligent data analysis* (Vol. 2). Berlin: Springer.

Bhatia, N. (2010). Vandana: Survey of nearest neighbor techniques. *International Journal of Computer Science and Information Security*, *8*(2), 302-305.

Bhardwaj, B. K., & Pal, S. (2011). Data Mining: A prediction for performance improvement using classification. *International Journal of Computer Science and Information Security 9*(4), 136-140.

Burmaoğlu, S. (2009). *Birleşmiş Milletler kalkınma programı beşerî kalkınma endeksi verilerini kullanarak diskriminant analizi, lojistik regresyon analizi ve yapay sinir ağlarının sınıflandırma başarılarının değerlendirilmesi.* (Doktora tezi). Atatürk Üniversitesi, Erzurum.

Büyüköztürk, Ş., Kılıç, E. K., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2009). *Bilimsel araştırma yöntemleri* (4. Basım) Ankara: Pegem A Yayıncılık.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering data mining: from concept to implementation*. Prentice-Hall, Inc.

Cai, Y. D. & Chou, K. C. (2003). Nearest neighbour algorithm for predicting protein subcellular location by combining functional domain composition and pseudo-amino acid composition. *Biochemical and Biophysical Research Communications*, *305*(2), 407-411.

Cameron, A. C. & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317-372.

Cortez, P. and Silva, A. (2008). Using data mining to predict secondary school student performance. http://www3.dsi.uminho.pt/pcortez/student.pdf (25.04.2020 tarihinde erişilmiştir).

Coşkun, E. (2003). Çeşitli değişkenlere göre lise öğrencilerinin etkili okuma becerileri ve bazı değişkenler. *TÜBAR 13*(1),101-130.

Coşkuner, T. (2013). *Uluslararası Öğrenci Başarı Değerlendirme Programı (PISA) 2009 uygulaması okuma becerileri okuryazarlığını etkileyen faktörler.* (Yüksek lisans Tezi). Akdeniz Üniversitesi, Antalya.

Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, *13*(1), 21-27.

Cristobal, R., Sebastián, V., Mykola, P. & Ryan, S. (2010). *Introduction handbook of educational data mining.* Florida: CRC Press.

Cristobal, R. & Sebastian, V. (2013). Data mining in education. *Wiley ınterdisciplinary reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27.

Cunningham, P. & SJ, D. (2007). k-Nearest neighbour classifiers. *Cornell University* *3*(1), 15-22.

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.

Çiftçi, Ö. (2007). *İlköğretim 5. sınıf öğrencilerinin Türkçe öğretim programında belirtilen okuduğunu anlamayla ilgili kazanımlara ulaşma düzeyinin belirlenmesi.* (Yayınlanmamış Doktora Tezi). Gazi Üniversitesi, Ankara.

Cortez, P. and Silva, A. (2008). Using data mining to predict secondary school student performance. http://www3.dsi.uminho.pt/pcortez/student.pdf (25.04.2020 tarihinde erişilmiştir).

Demir, E. ve Parlak, B. (2012). Türkiye'de eğitim araştırmalarında kayıp veri sorunu. *Journal of Measurement and Evaluation in Education and Psychology*, *3*(1), 230-241.

Dejaeger, K., Goethals, F., Giangreco, A., Mola, L. & Baesens, B. (2012). Gaining insight into student satisfaction using comprehensible data mining techniques. *European Journal of Operational Research*, *218*(2), 548-562.

Doğaç, A. (2021). *PISA 2018 okuma becerilerini açıklayan değişkenlerin çok düzeyli yapısal eşitlik modeli ile incelenmesi.* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.

Dunham, M. H. (2003). *Data mining introductory and advanced topics.* New Jersey: Prentice Hall,

Eccles, J.S. and Wigfield, A. (2002). Motivational beliefs, values and goals., *Annu. Rev. Psychol*., *1*(53), 109–32.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37-37.

Fenokulu, (2015). Fenokulu, https://fenokulu.net adresinden 2021 yılında edinilmiştir.

Fix, E. & Hodges, J. (1951). An important contribution to nonparametric discriminant analysis and density estimation. *International Statistical Review*, *3*(57), 233-238.

García, E., Romero, C., Ventura, S. & De Castro, C. (2011). A collaborative educational association rule mining tool. *The Internet and Higher Education*, *14*(2), 77-88.

Gartner IT Glossary (2021) https://www.gartner.com/en/information-technology/glossary/big-data adresinden 2021 yılında edinilmiştir.

Gelbal, S. (2010). Sekizinci sınıf öğrencilerinin sosyoekonomik özelliklerinin Türkçe başarısı üzerinde etkisi. *Eğitim ve Bilim*, *33*(150), 1-13.

Güldal, H., ve Çakıcı, Y. (2017) Ders Yönetim Sistemi Yazılımı Kullanıcı Etkileşimlerinin Sınıflandırma Algoritmaları ile Analizi, *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *21*(4),1355-1367.

Gürtekin, E. (2021). *Türkiye'deki öğrencilerin PISA 2015 ve PISA 2018 okuma becerilerinin bazı değişkenler açısından incelenmesi: İkincil veri analiz çalışması*. (Yüksek lisans tezi). Akdeniz Üniversitesi, Antalya.

Han, J. & Kamber, M. (2001). *Data mining: Concepts and techniques.* USA: Academic Press.

Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science*, *21*(1), 1-14.

Harrington, P. (2012). *Machine learning in action*. New York: Manning Publications.

Hayashi, Y., Hsieh, M-H., & Setiono, R. (2009). Predicting consumer preference for fast-food franchises: A data mining approach. *The Journal of the Operational Research Society, 60*(9), 1221-1229.

Hazım, L. R. (2018). *Four classification methods Naïve Bayesian, support vector machine, K-nearest neighbors and random forest are tested for credit card fraud detection*. (Yüksek lisans tezi). Altınbaş Üniversitesi, İstanbul.

Holsheimer, M. & Siebes, A. P. (1994). *Data mining: The search for knowledge in databases*. Amsterdam: Centrum voor Wiskunde en Informatica.

Huebner, R. A. (2013). A survey of educational data-mining research. *Research in Higher Education Journal*, *19*.

Hudairy H. (2004) *Data mining and decisionmaking support in the governmental sector.* (Yüksek lisans tezi). Faculty of Graduate School of The University of Louisville, Kentucky.

Ibrahim, Z., & Rusli, D. (2007, September). Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. In *21st Annual SAS Malaysia Forum, 5th September*.

İş, Ç. (2003). *Uluslararası öğrenci başarı belirleme programına göre (PISA) matematik okuryazarlığını belirleyen faktörlerin kültürler arası karşılaştırılması.* (Yüksek lisans tezi). Orta Doğu Teknik Üniversitesi, Ankara.

Jain A. K. & Dubes R. C. (1988) *Algorithms for clustering data.* Englewood Cliffs, NJ: Prentice Hall.

Kamber, M. & Pei, J. (2006). Data mining techniques and concepts. USA: Academic Press.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series 29*(2), 119-127.

Kelley-Winstead, D. (2010). *New directions in education research: using data mining techniques to explore predictors of grade retention. (*Doktora Tezi). George Mason University Education, Fairfax, VA.

Lim, T. S., Loh, W. Y. & Shih, Y. S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine learning*, *40*(3), 203-228.

Linnakylä, P., Malin, A. & Taube, K. Factors behind low reading literacy achievement. *Scandinavian Journal of Educational Research*, *48*(3), 231–249.

Liu, Y., & Schumann, M. (2005). Data mining feature selection for credit scoring models. *The Journal of the Operational Research Society*, *56* (9), 1099-1108.

Loh, W. Y. & Shih, Y. S. (1997). Split selection methods for classification trees. *Statistica Sinica*, 815-840.

Lopez, M. I., Luna, J. M., Romero, C. & Ventura, S. (2012). Classification via clustering for predicting final marks based on student participation in forums. *International Educational Data Mining Society*.

Marshall, K. T. (1995). Decision Making and Forecasting: with emphasis on model building and policy analysis. New York: McGraw-Hill.

MEB (2015). *PISA 2012 araştırması ulusal nihai raporu*. Ankara: İşkur Matbaacılık.

MEB (2017). *PISA 2015 ulusal raporu*. Ankara.

MEB (2019). PISA 2018 Türkiye Ön Raporu. http://www.meb.gov.tr/meb_iys_dosyalar/201912/03105347PISA_2018_Turkiye_On_Raporu.pdf adresinden edinilmiştir.

Michie D, Spiegelhalter DJ & Taylor CC (1994). *Machine learning, neural and statistical classification*. New York: Prentice Hall.

Nisbet, R., Elder, J. & Miner, G. (2009). *Handbook of statistical analysis and data mining applications.* Burlington: Academic press.

OECD (2010a). PISA *2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*. PISA. OECD Publishing.

OECD (2013a). *PISA 2012 results: What students know and can do – student performance in mathematics, reading, and science (Volume I).* Paris. OECD Publications.

OECD (2013b). *PISA 2012 results: What students know and can do – student performance in mathematics, reading, and science (Volume II)*. PISA. OECD Publishing

OECD (2019). *PISA 2018 Assessment and Analytical Framework*. PISA. OECD Publishing

OECD (2019b). *PISA 2018 results (Volume I): What students know and can do.* Paris. OECD Publishing

Oladokun, V. O., Adebanjo, A. T., & Charles-Owaba, O. E. (2008). Predicting students academic performance using artificial neural network: A case study of an engineering course. *The Pacific Journal of Science and Technology 9*(1), 72-79

Olgun, M. ve Özdemir, G. (2012). İstatistiksel özellik temelli bayes siniflandirici kullanarak kontrol grafiklerinde örüntü tanıma. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, *27*(2), 303-311.

Önen, E. (2018). Öğrenci, öğretmen ve öğretimsel nitelikler açısından TIMSS-2015'e dayalı olarak öğrencilerin sınıflandırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 9(1), 64-84.

Özekeş, S. (2003). Veri madenciliği modelleri ve uygulama alanları. *İstanbul Ticaret Üniversitesi Dergisi*.

Özer, Y. ve Anıl, D. (2011). Öğrencilerin fen ve matematik başarılarını etkileyen faktörlerin yapısal eşitlik modeli ile incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,* 41, 313-324.

Öztemel, E. (2006). *Yapay sinir ağları*. İstanbul: Papatya Yayıncılık.

Piramuthu, S. (2004). Evaluating feature selection methods for learning in data mining applications. *European Journal of Operational Research, 156*(2), 483-494.

Priddy, K. L., & Keller, P. E. (2005). *Artificial neural networks: An introduction*. Bellingham: SPIE Press.

Roiger, R. J. (2017). *Data mining: a tutorial-based primer.* Florida: CRC press.

Romero, C. & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications, 33*(1), 135-146.

Roy, A. (2000). Artificial neural networks: a science in trouble. *Explorations Newsletter*, *1*(2), 33-38.

Sadioğlu, Ö., ve Bilgin, A. (2008). İlköğretim öğrencilerinin eleştirel okuma becerileri ile cinsiyet ve anne-baba eğitim durumu arasındaki ilişki. *İlköğretim Online, 7*(3), 814-822.

Sezer, E. A., Bozkır, A. S., Yağız, S. ve Gökçeoğlu, C. (2010). Karar ağacı derinliğinin CART algoritmasında kestirim kapasitesine etkisi: Bir tünel açma makinesinin ilerleme hızı üzerinde uygulama. *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, Kayseri*.

Siemens, G. & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 30.

Siemens, G., & Baker, R. S. D. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (252-254).

Silahtaroğlu, G. (2013). *Veri madenciliği: Kavram ve algoritmaları*. İstanbul: Papatya Yayıncılık.

Sun, J. & Li, H. (2008). Data mining method for listed companies' financial distress prediction. *Knowledge-Based Systems*, *21*(1), 1-5.

Subbanarasimha, P. N., Arinzeb, B., & Anandarajanb, M. (2000). The predictive accuracy of artificial neural networks and multiple regression in the case of skewed data. *Exploration of Some Issues. Expert Systems with Applications, 19*(1), 117-123.

Şahin, M. (2018). *Karar ağaçları ve yapay sinir ağları kullanılarak kasko sigortalarında risk değerlendirme* (Yüksek lisans tezi). Yıldız Teknik Üniversitesi İstatistik Bölümü Anabilim Dalı, İstanbul.

Şen, Z. (2004). *Yapay Sinir Ağları İlkeleri*, Su Vakfı Yayınları, İstanbul.

Şen, B., Uçar, E. ve Delen, D. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, *39*(10), 9468-9476.

Şengür, D. ve Tekin, A., 2013, Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini. *International Journal Of Informatics Technologies*, 7-16.

Tan, P. N., Steinbach, M. & Kumar, V. (2006). Introduction to Data Mining, Boston: Person Education.

Taşcı, E. ve Onan, A. (2016). K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. *Akademik Bilişim, 1*(1), 4-18.

Taşkın, Ç. ve Emel, G. G. (2010). Veri madenciliğinde kümeleme yaklaşımları ve kohonen ağlari ile perakendecilik sektöründe bir uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *15*(3), 395-409.

Tosun, S. (2007). *Sınıflandırmada yapay sinir ağları ve karar ağaçları karşılaştırması: Öğrenci başarıları üzerine bir uygulama*. (Yüksek lisans tezi). İstanbul Teknik Üniversitesi, İstanbul.

Tsai, C. F., Tsai, C. T., Hung, C. S. & Hwang, P. S. (2011). Data mining techniques for identifying students at risk of failing a computer proficiency test required for graduation. *Australasian Journal of Educational Technology*, *27*(3).

Two Crows Corporation. (1999). *Introduction to data mining and knowledge discovery*. Potomac: Two Crows.

Ülper, H. (2010). *Okuma ve anlamlandırma becerilerinin kazandırılması*. Ankara: Nobel Yayınevi.

Van Ours, J.C. (2008). When do children read books? *Education Economics*, *16*(4), 313–328.

Wu, J. (2012). Advances in k-mean clustering: A data mining thinking. Hedilberg: Springer Science & Business Media.

Yılmaz, M. (2008) Türkçe'de okuduğunu anlama becerilerini geliştirme yolları. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 5* (9), 131-139.

Yılmaz, M. B. (2012). Profiles of university students according to internet usage with the aim of entertainment and communication and their affinity to internet. *International Online Journal Education Science, 4*(1), 225-242.

Yurtoğlu, H. (2005). *Yapay Sinir Ağları Metodolojisi İle Öngörü Modellemesi: Bazı Makroekonomik Değişkenler İçin Türkiye Örneği* (Uzmanlık Tezi). Devlet Planlama Teşkilatı, Ankara.

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: An efficient data clustering method for very large databases. *ACM sigmod record*, *25*(2), 103-114.

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu

F46

25 / 10 / 2021

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına

Tez/Araştırma Başlığı	PISA 2018 OKUMA BAŞARI PUANLARININ VERİ MADENCİLİĞİ SINIFLANDIRMA YÖNTEMLERİ İLE İNCELENMESİ

Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.
4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir.
5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.

Çalışmada kullanacağım veriler:
(X) Kamusal erişime açık
OECD'nin internet sayfasından (https://www.oecd.org/pisa/data/2018database/) elde edilmiştir
( ) Özel izin ve onaya tabi (buraya yazınız): .....
( ) Üretilmiş veri (buraya yazınız): .....
( ) Diğer (buraya yazınız): .....

Yükseköğretim Kurumları Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Emrah BÜYÜKATAK

Araştırmacı Bilgileri

Adı Soyadı	Emrah BÜYÜKATAK
Öğrenci ise No	N16142434
Ana Bilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme Bilim Dalı
Statüsü	☐ Yüksek Lisans X Doktora ☐ Bütünleşik Dr. ☐ Diğer

Danışman Görüşü ve Onayı*

Prof.Dr. Duygu ANIL

*Tez ve tezden üretilen yayınlarda gerekli

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü. Beytepe Yerleşkesi, 06800, Çankaya / ANKARA
Telefon: 0(312) 297 85 72 Belgegeçer: 0(312) 297 85 66 e-Ağ: http://ebe.hacettepe.edu.tr/ e-Posta: ebe@hacettepe.edu.tr

B: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin bütününü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

25 / 01 / 2022

Emrah BÜYÜKATAK

EK-C: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu

06/03/2022

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı : PISA 2018 Türkiye Örnekleminde Okuma Okuryazarlık Düzeylerinin Farklı Veri Madenciliği Sınıflandırma Yöntemleri İle İncelenmesi

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
06/03/2022	124	189619	28/01/2022	10	1777543318

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Emrah BÜYÜKATAK

Öğrenci No.: N16142434

Ana Bilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Statüsü: ☐ Y.Lisans X Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.
Prof.Dr. Duygu ANIL

EK-Ç: Thesis/Dissertation Originality Report

06/03/2022

HACETTEPE UNIVERSITY
Graduate School of Educational Sciences
To The Department of Educational Sciences

Thesis Title: Examination Of PISA 2018 Reading Scores By Data Mining Classification Methods

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
06/03/2022	124	189619	28/01/2022	10	1777543318

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

NameLastname: Emrah BÜYÜKATAK

Student No.: N16142434 Signature

Department: Educational Sciences

Program: Measurement and Evaluation in Education

Status: ☐ Masters X Ph.D. ☐ Integrated Ph.D.

ADVISOR APPROVA

APPROVED
Prof.Dr. Duygu ANIL

EK-D: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan **"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge"** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. (1)
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren … ay ertelenmiştir. (2)
- Tezimle ilgili gizlilik kararı verilmiştir. (3)

25 /01 /2022

Emrah BÜYÜKATAK

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

(1) Madde 6. 1. Lisansüstü tezle ilgili patent başvurusu yapılması veya patent alma *sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü Üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.*

(2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanın önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.

(3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü Üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.*
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

** Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir*