



## Investigation of Demographic Variables Affecting the Academic Success of Students via Educational Data Mining

Neytullah KARAKURT <sup>1</sup> Habib ÖZKAN <sup>2</sup>

### To cite this article:

Karakurt, N. & Özkan, H. (2024). Investigation of Demographic Variables Affecting the Academic Success of Students via Educational Data Mining [Öğrencilerin Akademik Başarısını Etkileyen Demografik Değişkenlerin Eğitsel Veri Madenciliği İle İncelenmesi]. *Electronic Journal of Education Sciences*, [Elektronik Eğitim Bilimleri Dergisi], 13(25), 16-33. DOI: 10.55605/ejedus.1351750

Research article

Received: 2023-08-29

Accepted: 2023-12-19

### Abstract

The aim of this research is to predict and model the exam success of 8th grade students by using data mining methods, taking into account the relevant factors in the literature. Educational data mining techniques were used in the research designed in relational scanning model. The population of the research consists of 8th grade students studying in 42 secondary schools located in the central district of Adıyaman. The study group of the research consists of 135 8th grade students in eight secondary schools with a low socio-economic environment in the central district of Adıyaman province. The study group of the research was determined by the Stratified Sampling method. In the analysis of the data obtained in the research, the CRISP-DM (Cross Industry Standard Process for Data Mining) process model, which is widely accepted in the international data mining literature, was used. In the analysis of the data collected from the participants, KNN, J48, Random Forest, and Naive Bayes methods, which are classification methods, were used via WEKA software. As a result of the research, it has been determined that some demographic variables such as students going to a private course, coming to school by having breakfast at home, not living in a rented house have an effect on their academic achievement. Among the algorithms used in predicting and modeling academic success, the Random Forest algorithm gave the most accurate result. With the results of the research, the demographic variables that affect the exam success in disadvantaged schools have been revealed and it is thought that the results will allow the necessary measures to be taken by estimating the exam success of the students.

**Keywords:** Education management, educational data mining, academic success, middle school students.

<sup>1</sup> Adıyaman Provincial Directorate of National Education, Teacher, neytullahkarakurt@gmail.com ORCID NO: 0000-0001-8798-3350

<sup>2</sup> Gaziantep University, Prof. Dr., habibozyan@gmail.com ORCID NO: 0000-0002-0167-7342



## Öğrencilerin Akademik Başarısını Etkileyen Demografik Değişkenlerin Eğitsel Veri Madenciliği İle İncelenmesi

Neytullah KARAKURT<sup>3</sup> Habib ÖZKAN<sup>4</sup>

### Atıf:

Karakurt, N. & Özkan, H. (2024). Investigation of Demographic Variables Affecting the Academic Success of Students via Educational Data Mining [Öğrencilerin Akademik Başarısını Etkileyen Demografik Değişkenlerin Eğitsel Veri Madenciliği İle İncelenmesi]. *Electronic Journal of Education Sciences*, [Elektronik Eğitim Bilimleri Dergisi], 13(25), 16-33. DOI: 10.55605/ejedus.1351750

Araştırma Makalesi

Geliş Tarihi: 2023-08-29

Kabul Tarihi: 2023-12-19

### Öz

Bu araştırmanın amacı, 8. Sınıfta öğrenim gören öğrencilerin, literatürde yer alan ilgili faktörleri göz önüne alarak, sınav başarılarını veri madenciliği yöntemleriyle tahmin etmek ve modellemektir. İlişkisel tarama modelinde tasarlanan araştırmada eğitsel veri madenciliği teknikleri kullanılmıştır. Araştırmanın evrenini Adıyaman ili Merkez ilçesinde yer alan 42 ortaokulda öğrenim gören 8. sınıf öğrencileri oluşturmaktadır. Araştırmanın çalışma grubunu Adıyaman ili Merkez ilçesinde yer alan düşük sosyo-ekonomik çevreye sahip sekiz ortaokuldaki 8. sınıfta öğrenim gören 135 öğrenci oluşturmaktadır. Araştırmanın çalışma grubu Tabakalı Örneklem Alma yöntemi ile belirlenmiştir. Araştırmada elde edilen verilerin analizinde uluslararası veri madenciliği literatüründe yaygın kabul gören CRISP-DM (Cross Industry Standart Process for Data Mining) süreç modeli kullanılmıştır. Katılımcılardan toplanan verilerin analizinde sınıflandırma yöntemlerinden KNN, J48, Random Forest, ve Naive Bayes yöntemleri WEKA yazılımı aracılığıyla kullanılmıştır. Araştırma sonucunda, öğrencilerin özel bir kursa gitmesi, evde kahvaltı yaparak okula gelmesi, kiralık evde oturmaması gibi bazı demografik değişkenlerin akademik başarıları üzerinde etkisi olduğu tespit edilmiştir. Akademik başarıyı tahmin ve modellemede kullanılan algoritmalarından en doğru sonucu veren, Random Forest algoritması olmuştur. Araştırma sonuçları ile dezavantajlı okullardaki sınav başarısını etkileyen demografik değişkenlerin neler olabileceği ortaya konulmuş olup sonuçların öğrencilerin sınav başarısının tahmin edilerek gerekli önlemlerin alınmasına olanak sağlanacağı düşünülmektedir.

**Anahtar Sözcükler:** Eğitim yönetimi, eğitsel veri madenciliği, akademik başarı, ortaokul öğrencileri.

<sup>3</sup> Adıyaman İl Milli Eğitim Müdürlüğü, Öğretmen, neytullahkarakurt@gmail.com ORCID NO: 0000-0001-8798-3350

<sup>4</sup> Gaziantep Üniversitesi, Prof. Dr., habibozgan@gmail.com ORCID NO: 0000-0002-0167-7342

## Giriş

Her öğrenci akademik başarılarının ölçülmesi amacıyla eğitim-öğretim sürecinde çeşitli sınavlara tabi tutulur. Bu sınavlar öğrencilerin başarılarını ölçmenin yanında seçmek istedikleri ortaöğretim ve yükseköğretim kurumları kararlarını da etkilemektedir (Karaçor, 2020). Bir eğitim kurumu olarak okulların amaçlarını gerçekleştirme düzeyleri çoğunlukla öğrencilerin okulların amaçları ekseninde elde ettikleri kazanımlarla belirlenir. Bu anlamda okulların başarısının öğrencilerin akademik başarılarına dayandığı söylenebilir (Altun, 2019). Öğrencilerin akademik başarılarının yordanması, eğitim-öğretim alanında yapılan çok sayıda araştırmanın odağında yer almaktadır (Koyuncu, 2018). Akademik başarının üzerinde; zeka, öğrenme hızı gibi öğrencinin zihinsel özelliklerinin; kişilik, öz-yeterlik algısı, benlik saygısı, ders çalışma alışkanlıkları ve motivasyon gibi duyuşsal özelliklerinin; ailenin sosyo-ekonomik seviyesi ve ana-baba tutumunun, öğretmen ve okul yöneticilerinin yeterlilikleri ve öğrenciye yönelik tutumlarının etkili olduğu birçok farklı çalışmada ortaya konulmaktadır (Sarier, 2016).

Bilgi teknolojilerinin hızlı gelişimine paralel olarak internetin de hayatımıza yoğun etkisi, büyük verilerin depolanmasını kolaylaştırmasının yanında çok farklı türlerde oldukça büyük veri yığınlarının da oluşmasına sebep olmuştur (Akgün ve Bulut Özek, 2020). Bu veri yığınları ilk bakışta herhangi bir anlam ifade etmeyen, işlenmemiş ve yapılandırılmamış verilerdir. Veri Madenciliği yöntemleri kullanılarak, çıplak gözle anlamsız görünen bu yapılandırılmamış verilerin içinden, birbiriyle ilişkili, anlamlı gizli bilgiler çıkarılabilir (Ersöz, 2017). Veri madenciliği kısaca, büyük miktarda anlamsız veriden ilginç, önemsiz, gizli ve bilinmeyen anlamlı verilerin çıkarılması olarak tanımlanabilir (Mining, 2006).

Bilgi teknolojilerindeki son yıllarda yaşanan hızlı gelişmeler, okulların kurdukları bilgi sistemleriyle öğrencilere dair demografik ve akademik verileri sanal ortamda depolama ve saklama imkanı sağlamıştır. Bu durum, eğitimcilerin ve okul yöneticilerinin öğrencilere ait istedikleri verilere kolay bir şekilde ulaşabilmesini, verileri değerlendirebilmesini ve bu verileri bilgilendirme faaliyetlerinde kullanabilmesine olanak vermiştir (Polat, 2021). Toplanan eğitsel veri yığınlarını öğrenci başarılarını yordamak amacıyla analiz etmek son zamanlarda oldukça önem arz etmektedir (Koyuncu, 2018). Son yıllarda ortaya çıkan ve 2008 yılında senelik Uluslararası Eğitimde Veri Madenciliği Konferansı ve Eğitimde Veri Madenciliği Dergisi'nin kurulmasıyla da zirveye ulaşan Eğitimde Veri Madenciliği (EVM), bağımsız bir araştırma yöntemi olarak eğitsel başarıları yordamada kullanılmaya başlanmıştır (Baker, 2010). Eğitimde Veri Madenciliği; Pedagojik unsurları da göz önünde bulundurarak, eğitim-öğretim sürecindeki ham verilere (öğrencinin başarı durumu, devamsızlık bilgileri, sosyo-demografik bilgileri, hazırbulunuşluk seviyeleri, güdülenmesi vb.) veri madenciliği yöntemlerinin uygulanarak elde edilen bilgilerin de yine eğitim-öğretim sürecinde girdi haline getirilmesi ve eğer gerekirse veri madenciliği yöntemleri kullanılarak tekrar analiz edilmesini ifade etmektedir (Özdemir, 2016). EVM, öğrenci başarısını etkileyen etmenlerin tespit edilmesinde oldukça etkili ve başarılı sonuçlar ortaya koymaktadır (Kurt ve Erdem, 2012). EVM'nin kullanım alanları; öğrencileri başarı ve başarısızlık nedenlerinin tespiti, eğitim-öğretim ortamlarında ortaya çıkabilecek aksaklıkların tespiti, öğrenci verilerinin analizi, öğrencilerin akademik başarılarının artırılması ve öğrenci çalışmalarının özgünlüğünün tespiti olarak listelenebilir (Cemaloğlu, 2021). Her yıl 1.5 milyona yakın öğrenci ortaöğretim kurumlarına kayıt yaptırıyor. Ancak bu ortaöğretim kurumlarından Anadolu ve Fen liselerine ise sadece sınav puanlarına göre yerleşilebiliyor. Bu bağlamda öğrencilerin sınavda başarılı olarak Anadolu ve Fen liselerine yerleşmeleri önemli bir durum olmakla birlikte, öğrencilerin çeşitli demografik bilgileri üzerinden elde edilecek

gizli örüntüler sayesinde sınav başarılarını tahmin edebilmek eğitimciler ve yöneticiler için oldukça önem arz etmektedir. Bu çalışmada en başarılı sonucu veren algoritma kullanılarak geliştirilecek bir yazılım ile öğrencilerin hangi lise türüne yerleşecekleri tahmin edilebilecek ve böylece başarıyı artırmak için gerekli önlemler alınabilecektir. Bu bağlamda çalışmadan elde edilen sonuçların önem arz edeceği söylenebilir.

Araştırmanın temel problemi, 8. Sınıf öğrencilerinin eğitim-öğretim sürecinde etkili olan demografik bilgileri ile ilgili faktörleri göz önünde bulundurarak, sınav başarılarının veri madenciliği yöntemleriyle tahmin edilip edilemeyeceğinin araştırılmasıdır. Bu temel problem çerçevesinde aşağıda yer alan alt problemler belirlenmiştir:

1. 8. Sınıf öğrencilerinin sınav başarılarını tahmin etmede kullanılan sınıflandırma modellerinin başarısı nasıldır?
2. Veri setindeki niteliklerin öğrenci sınav başarısının tahminindeki bilgi kazanç değerleri nasıldır?
3. Veri seti üzerinde gerçekleştirilen analiz sonucunda oluşan karar ağacı modeli nasıldır?
4. Sınav başarılarını tahmin etmede kullanılan sınıflandırma modellerinden hangisi daha başarılıdır?

## Yöntem

### Araştırmanın Modeli

8. sınıf öğrencilerinin demografik özelliklerinin akademik başarıları ile ilişkisini incelemeyi amaçlayan bu çalışma, nicel araştırma yöntemlerinden ilişkisel tarama modelinde bir çalışmadır. İlişkisel tarama modeli, iki veya ikiden fazla değişken arasındaki mevcut ilişkiyi incelemek amacıyla yapılan çalışmalardır (Büyüköztürk vd., 2013).

### Evren/Örneklem

Araştırmanın evrenini 2021-2022 eğitim-öğretim yılında Adıyaman il merkezinde yer alan 42 ortaokulda öğrenim gören 8. Sınıf öğrencileri oluşturmaktadır. Araştırmaya tabakalı örnekleme yöntemi kullanılarak belirlenen 8 ortaokulda öğrenim gören 135 öğrenci dahil edilmiştir. Araştırma örneklemini belirlemek amacıyla çalışma evreni düşük, orta ve yüksek sosyo-ekonomik düzeyde bulunan okullar olmak üzere üç gruba ayrılmıştır. Gruplamalar yapılırken, okulun bulunduğu bölge, öğrenci devamsızlık durumları ve okulun akademik başarıları kriterleri göz önünde bulundurulmuştur. Böylece araştırma örneklemine düşük sosyo-ekonomik çevrede yer alan sekiz ortaokul seçilmiştir. Araştırmaya dahil edilen okulların, Adıyaman İl Milli Eğitim Müdürlüğü'nün 2021-2022 yılında düşük sosyo-ekonomik çevrede bulunan okullar üzerinde yürüttüğü Başarı İçin El Ele Projesi'nde de yer almış olması, alınan örneklemin uygunluğu açısından olumlu bir gösterge olmuştur.

### Veri Toplama Süreci

8. Sınıf öğrencilerine çalışmacı tarafından hazırlanan Demografik Bilgiler Formu uygulanarak veriler toplanmış ve bilgisayar ortamına aktarılmıştır. Verilerin toplanmasında etik ilkelere riayet edilmiştir. Bu kapsamda çalışmaya katılan öğrencilerin velilerinden izin alınmıştır. Araştırma için etik kurul izni ve Adıyaman İl Milli Eğitim Müdürlüğü'nden de gerekli araştırma izinleri alınmıştır. İzinler alındıktan sonra katılımcı öğrencilerden veriler toplanmıştır. Veri seti içerisindeki katılımcılara ait sınav bilgileri ise Adıyaman Ölçme Değerlendirme Birimi'nden talep edilmiştir. Veri setindeki sınav analiz raporlarında öğrencilerin elde ettikleri puanların LGS karşılığı da yer almaktadır. Veriler analiz edilirken öğrencilerin LGS puanları baz alınmıştır. Araştırma verileri aynı sosyo-ekonomik çevreye sahip

olan 8 okulda toplanmıştır. Veri toplama aracı Basit Seçkisiz Yöntemle seçilen örneklem gruba elden teslim edilmiştir. Okul yöneticisinin bilgisi ve ders öğretmenin koordinesiyle gerçekleştirilen veri toplama süreci, katılımcıların veri toplama aracını doldurup teslim etmesiyle sona ermiştir.

Araştırmada eğitsel veri madenciliği teknikleri kullanılmıştır. Veri madenciliği, elde edilen verilerin içerisinde bulunan gizli örüntüleri ortaya koyan ve ayrıca henüz keşfedilmemiş enformasyonları açığa çıkaran, özellikle son yıllarda oldukça popüler olan araştırma tekniğidir (Olgun, 2021). Eğitsel veri madenciliğinde, Veri yığınları içerisinde anlamlı veriler çıkarabilmek amacıyla kullanılacak çok sayıda analiz yöntemi ile bu yöntemlerin uygulanabileceği yazılımlar mevcuttur. RapidMiner, Knime, R, Orange ve Weka veri madenciliğinde en yaygın kullanılan yazılımlar olarak ön plana çıkmaktadır. Verilerin analizinde Weka 3.9.6 yazılımı kullanılmıştır. WEKA, içerisinde makine öğrenme algoritmalarını barındıran ve veri madenciliği çalışmaları için kullanılan açık kaynak kodlu bir yazılımdır. Yazılım, içerisinde regresyon, veri ön işleme, özellik seçimi, sınıflandırma, kümeleme, birliktelik kuralları ve görselleştirme için araçlar barındırmaktadır (Weka, 2022).

### Verilerin Analizi

Araştırmanın veri analizinde izlenen aşamalar CRISP-DM veri madenciliği süreç modeline göre oluşturulmuştur. CRISP-DM, NCR Corporation ve Daimler Chrysler (SPSS Inc.) tarafından 1996 yılında geliştirilen ve tanımlan adaptif ve iteratif bir süreçtir (Özdemir, 2016). Veri madenciliği için endüstriden bağımsız bir süreç modeli olan CRISP-DM, İş Anlama, Veriyi Anlama, Veri Hazırlama, Modelleme, Değerlendirme ve Dağıtım olmak üzere altı yinelenmeli aşamadan oluşur (Schröer, Kruse ve Gomez, 2021).

#### 1. İş Anlama:

İlk aşama, organizasyonun perspektifinden araştırmanın hedeflerini ve ihtiyaçlarını anlamaya odaklanır. Daha sonra bu bilgiyi bir veri madenciliği problem tanımına ve hedeflere ulaşabilmek amacıyla hazırlanmış bir ön plana dönüştürmeye odaklanır (Chapman vd., 2000). Bu aşamada Eğitsel Veri Madenciliği için araştırmanın amacı “8.sınıf öğrencilerinin eğitim-öğretim sürecinde etkili olan demografik bilgiler ile ilgili faktörleri göz önüne alarak, sınav başarılarını veri madenciliği yöntemleriyle tahmin etmek ve modellemek” olarak belirlenmiştir.

#### 2. Veriyi Anlama:

İlk olarak veri toplama ile başlayan bu aşama, verilere aşina olmak, verinin kalitesine engel olabilecek sorunları belirlemek, verileri keşfetmek ve açıklamak veya gizli örüntüleri ortaya çıkarmak için hipotezler oluşturmak amacıyla ilgili alt kümeleri tespit etmek gibi faaliyetlerle devam eder (Chapman vd., 2000). Bu aşamada, Excel’de .csv uzantılı bir dosyada kaydedilen ham veriler, daha sonra .arff uzantılı bir dosya haline getirilerek WEKA ortamına aktarılmıştır. Böylece, veri setinde yer alan değişkenler ve özelliklerin keşfedilmesine olanak sağlanmıştır. Veri setinde, demografik bilgilerden oluşan toplamda 40 değişken ve bir hedef özellik yer almıştır. Ayrıca, veri setinde yer alan kayıp veriler ve aykırı veriler tespit edilmiştir.

#### 3. Veri Hazırlama:

Veri hazırlama aşaması, toplanan ilk işlenmemiş ham verilerden, modelleme için kullanılacak nihai veri setinin oluşturulmasına dönük her türlü faaliyeti kapsar. Birden çok defa gerçekleştirilebilen bu aşamada verilerin dönüştürülmesi ve temizlenmesi gibi faaliyetler de yer alabilmektedir (Chapman vd., 2000). Araştırmanın bu aşamasında veri setindeki eksik ve kayıp verilere SPSS 25.0 paket program ile veri ataması gerçekleştirilmiştir. Böylece düzensiz veri

seti problemi giderilmiştir. Son olarak WEKA programında dengesiz veri seti problemine çözüm aranmıştır. Makine öğrenmesinde önemli bir problem olan veri dengesizliğinin etkisini en aza indirmek veya bu etkiyi ortadan kaldırmak gerekmektedir (Ataseven ve Yıldırım, 2022). Araştırmadaki dengesiz veri setleri problemini ortadan kaldırmak için, literatürde sıklıkla kullanılmakta olan “yeniden örnekleme” yöntemlerinden biri olan SMOTE kullanılmıştır.

#### **4. Modelleme:**

Bu aşamada, çeşitli modelleme teknikleri seçilir ve uygulanır ve parametreleri optimal değerlere kalibre edilir. Tipik olarak, aynı veri madenciliği problem tipi için birkaç teknik vardır. Bazı tekniklerin veri formuyla ilgili özel gereksinimleri vardır. Bu nedenle, veri hazırlama aşamasına geri dönmek çoğu zaman gereklidir (Chapman vd., 2000). Veri Madenciliği sürecinde mevcut yazılımların içerisinde regresyon, yapay sinir ağları, karar ağaçları, k-en yakın komşu ve Naive Bayes gibi çok sayıda algoritma bulunmaktadır (Topuz, 2021). Katılımcılardan toplanan nicel verilerin analizinde sınıflandırma yöntemlerinden KNN, J48, Random Forest, ve Naive Bayes yöntemleri kullanılmıştır.

#### **5. Değerlendirme:**

Veri madenciliğinde modellerin performansını ölçen ölçme aracı olarak tanımlanan değerlendirme ölçütleri en uygun sınıflandırıcıyı elde etmek için kritik bir rol oynar (Hossin ve Sulaiman, 2015). Bu çalışmada sınıflandırma amacıyla geliştirilen modellerin performansları değerlendirirken aşağıdaki ölçütler kullanılmıştır:

- Doğru Sınıflama Oranı (DSO)
- Kappa değeri
- Pozitif Tahmin Oranı
- Negatif Tahmin Oranı
- Kesinlik (Precision)
- F-Measure ölçütü
- MCC değeri

Ayrıca araştırmanın bu aşamasında hata matrisleriyle analiz sonucu oluşan sınıflandırmanın veri setinde yer alan farklı sınıfları ne kadar başarılı bir şekilde sınıflandırdığı gösterilmiştir. Hata matrisinde yer alan doğru ve yanlış sınıflandırılan örnek sayıları analiz sonucunda geliştirilen tahmin modelinin ne kadar başarılı olduğunu ortaya çıkarmaktadır (Polat, 2021).

#### **6. Dağıtım:**

Araştırma sürecinde yapılan çalışmaların raporlandığı dağıtım aşamasında, süreçte elde edilen bilgilerin sunumu da yapılır. (Polat, 2021). Dağıtım aşaması, bir rapor oluşturmak kadar basit veya kuruluş genelinde tekrarlanabilir bir veri madenciliği sürecini uygulamak kadar karmaşık olabilir (Chapman vd., 2000). Araştırmanın bu aşamasında, analiz sonucunda elde edilen bilgilerin sunulduğu Bulgular ve Yorum ile literatürdeki araştırma sonuçlarıyla karşılaştırmalar yapıldığı Tartışma bölümü yer almaktadır. Ayrıca araştırmacılar için çeşitli önerilerde bulunulmuştur.

### **Bulgular**

8.sınıf öğrencilerinin eğitim-öğretim sürecinde etkili olan demografik bilgiler ile ilgili faktörleri göz önüne alarak, sınav başarılarını veri madenciliği yöntemleriyle tahmin etmeyi ve



modellemeyi amaçlayan araştırmanın bu bölümünde, veri seti üzerinde gerçekleştirilen veri madenciliği analizlerinin sonucu yer almaktadır.

## Birinci Alt Probleme İlişkin Bulgular

### 1. Sınıflandırma Analizi Bulguları

Sınıflandırma teknikleri, bir veri setindeki her öğeyi, daha önce tanımlanmış gruplar veya sınıflardan birine sınıflandırmak amacıyla kullanılan ve oldukça geniş uygulama alanı olan önemli bir veri madenciliği tekniğidir (Vijayarani ve Muthulakshmi, 2013). Sınıflandırma analizleri için J48, Random Forest, IBK ve Naive Bayes algoritmaları kullanılmıştır. Veri setlerinin analizinde WEKA 3.9.6 yazılımı kullanılmıştır. Veri madenciliğinde modelleme yapılırken yapılan modellemenin başarısını değerlendirmede; hata oranları, duyarlılık, F-ölçütü, kesinlik gibi değerler dikkate alınır (Coşkun ve Baykal). Analiz sonucunda yapılan modellemenin başarısını değerlendirmede, Doğru Değeri (DSO), Kappa değeri, Pozitif ve negatif tahmin oranları, Kesinlik değeri, F-Measure değeri ve MCC değeri dikkate alınmıştır. Ayrıca analizlere ait hata matrisleri ile doğru yerleştirme oranlarını görme imkanı elde edilmiştir. Sınıflandırma analizi, 10-kat çapraz geçirme (cross validation) ile oluşturulan eğitim ve test setleri üzerinden gerçekleştirilmiştir.

### 2. J48 Analizi Bulguları

J48, oldukça sık kullanılan ve popüler olan C4.5 algoritmasını temel alan, J.ross Quinlan tarafından geliştirilen bir algoritmadır (Alan ve Yeşilyurt, 2018). İkili bir ağaç oluşturan J48, modelleme yapmak için oluşturduğu karar ağacındaki bütün dallarda en tepeden en aşağıya doğru muhtemel nitelik etkileşimlerini bulmaya çalışır. Bu yönüyle J48 algoritması WEKA programındaki en yüksek bilgi kazancına sahip olan karar ağacı algoritmalarından biridir (Omende Kahudi, 2021). Tablo 1’de veri setine uygulanan J48 algoritmasının analiz sonuçları gösterilmektedir.

Tablo 1. J48 Analiz Sonuçları

Sınıflandırma Algoritması	DSO (%)	Kappa	Pozitif Tahmin Oranı	Negatif Tahmin Oranı	Kesinlik	F-Measure	MCC
J48	92.0904	0.8813	0,921	0,039	0,921	0,921	0,882

Tablo 1’de yer alan bilgilere göre model %92.09 oranında doğru tahminde bulunmuştur. Analiz sonucunda elde edilen Kesinlik (precision) değeri 0.921, F-Measure değeri 0.921, MCC (Matthews Korelasyon Katsayısı) değeri ise 0.882 olarak gerçekleşmiştir. Modelin öğrencilerin yerleştiği okul türlerini doğru tahmin edebilme oranı 0.921, yanlış tahmin edebilme oranı ise 0.039 olarak hesaplanmıştır. Ayrıca J48 analizine ait Kappa değerinin de 0,881 olduğu görülmektedir. Landis ve Koch (1977)’ye göre elde edilen bu değer, geliştirilen modelin oldukça başarılı bir model olduğunu ortaya koymaktadır. J48 analizine ait hata matrisi Tablo 2’de ayrıntılı bir şekilde verilmiştir.

Tablo 2. J48 Analizine Ait Hata Matrisi

Tahmin Edilen Sınıflar	Gerçekte olan Sınıflar		
	Düz	Fen	Anadolu
Düz	99	8	6
Fen	6	114	0
Anadolu	8	0	113

Tablo 2’de yer alan verilere göre; geliştirilen model, düz liseye yerleşen öğrencilerin 99’ünü, fen lisesine yerleşen öğrencilerin 114’ünü ve Anadolu lisesine yerleşen öğrencilerin ise 113’ünü doğru tahmin etmiştir.

### 3. *Random Forest Analizi Bulguları*

Random Forest, ilk olarak Breiman (2001) tarafından geliştirilen bir topluluk sınıflandırma yöntemidir (Ozgis vd. 2020). Random Forest, gürültüye dayanıklı bir yapısının olması, basit bir şekilde ağaçlara ayrılabilmesi, hata tahminlerini doğru ve hızlı bir şekilde gerçekleştirmesi ve kullanıcıya istediği kadar ağaç oluşturabilme olanağı vermesi gibi avantajlardan dolayı veri madenciliğinde sıklıkla tercih edilmektedir (Üre, 2021). Tablo 3’de veri setine uygulanan Random Forest algoritmasının analiz sonuçları gösterilmektedir.

Tablo 3. Random Forest Analiz Sonuçları

Sınıflandırma Algoritması	DSO (%)	Kappa	Pozitif Tahmin Oranı	Negatif Tahmin Oranı	Kesinlik	F-Measure	MCC
Random Forest	95.1977	0.928	0,952	0,023	0,956	0,953	0,930

Tablo 3’de yer alan bilgilere göre model %95.19 oranında doğru tahminde bulunmuştur. Analiz sonucunda elde edilen Kesinlik (precision) değeri 0.956, F-Measure değeri 0.953, MCC (Matthews Korelasyon Katsayısı) değeri ise 0.930 olarak gerçekleşmiştir. Modelin öğrencilerin yerleştiği okul türlerini doğru tahmin edebilme oranı 0.952, yanlış tahmin edebilme oranı ise 0.023 olarak hesaplanmıştır. Ayrıca Random Forest analizine ait Kappa değerinin de 0,928 olduğu görülmektedir. Landis ve Koch (1977)’ye göre elde edilen bu değer, geliştirilen modelin oldukça başarılı bir model olduğunu ortaya koymaktadır. Random Forest analizine ait hata matrisi Tablo 4’de ayrıntılı bir şekilde verilmiştir.

Tablo 4. Random Forest Analizine Ait Hata Matrisi

Tahmin Edilen Sınıflar	Gerçekte olan Sınıflar		
	Düz	Fen	Anadolu
Düz	111	1	1
Fen	7	113	0
Anadolu	8	0	113

Tablo 4’de yer alan verilere göre; geliştirilen model, düz liseye yerleşen öğrencilerin 111’ini, fen lisesine yerleşen öğrencilerin 113’ünü ve Anadolu lisesine yerleşen öğrencilerin ise 113’ünü doğru tahmin etmiştir.

### 4. *IBK (KNN) Analizi Bulguları*

IBK, aynı mesafe metriğini kullanan bir k-en yakın komşu sınıflandırıcıdır (Vijayarani ve Muthulakshmi, 2013). IBK, oldukça sık kullanılan ve popüler olan C4.5 algoritmasını temel alan, J.ross Quinlan tarafından geliştirilen bir algoritmadır (Alan ve Yeşilyurt, 2018). İkili bir ağaç oluşturan J48, modelleme yapmak için oluşturduğu karar ağacındaki bütün dallarda en tepeden en aşağıya doğru muhtemel nitelik etkileşimlerini bulmaya çalışır. Bu yönüyle J48 algoritması WEKA programındaki en yüksek bilgi kazancına sahip olan karar ağacı algoritmalarından biridir (Omende Kahudi, 2021). Tablo 5’de veri setine uygulanan IBK algoritmasının analiz sonuçları gösterilmektedir.



Tablo 5. IBK (KNN) Analiz Sonuçları

Sınıflandırma Algoritması	DSO (%)	Kappa	Pozitif Tahmin Oranı	Negatif Tahmin Oranı	Kesinlik	F-Measure	MCC
IBK	88.7006	0.8302	0,887	0,057	0,887	0,885	0,832

Tablo 5’de yer alan bilgilere göre model %88.70 oranında doğru tahminde bulunmuştur. Analiz sonucunda elde edilen Kesinlik (precision) değeri 0.887, F-Measure değeri 0.885, MCC (Matthews Korelasyon Katsayısı) değeri ise 0.832 olarak gerçekleşmiştir. Modelin öğrencilerin yerleştiği okul türlerini doğru tahmin edebilme oranı 0.887, yanlış tahmin edebilme oranı ise 0.057 olarak hesaplanmıştır. Ayrıca IBK analizine ait Kappa değerinin de 0.830 olduğu görülmektedir. Landis ve Koch (1977)’ye göre elde edilen bu değer, geliştirilen modelin oldukça başarılı bir model olduğunu ortaya koymaktadır. IBK analizine ait hata matrisi Tablo 6’de ayrıntılı bir şekilde verilmiştir.

Tablo 6. IBK (KNN) Analizine Ait Hata Matrisi

Tahmin Edilen Sınıflar	Gerçekte olan Sınıflar		
	Düz	Fen	Anadolu
Düz	86	8	19
Fen	6	114	0
Anadolu	6	1	114

Tablo 6’da yer alan verilere göre; geliştirilen model, düz liseye yerleşen öğrencilerin 86’sını, fen lisesine yerleşen öğrencilerin 114’ünü ve Anadolu lisesine yerleşen öğrencilerin ise 114’ünü doğru tahmin etmiştir.

### 5. Naive Bayes Analizi Bulguları

Bayes sınıflandırıcılar, bir grubun belirli bir sınıfa dahil edilme olasılığı gibi sınıfa katılım olasılıklarını tahmin edebilen ve özellikle büyük veri setleri üzerinde çalışıldığında hızlı ve doğru sonuçlar veren istatistiksel sınıflandırıcılardır (Polat, 2021). Tablo 7’de veri setine uygulanan Naive Bayes algoritmasının analiz sonuçları gösterilmektedir.

Tablo 7. Naive Bayes Analiz Sonuçları

Sınıflandırma Algoritması	DSO (%)	Kappa	Pozitif Tahmin Oranı	Negatif Tahmin Oranı	Kesinlik	F-Measure	MCC
Naive Bayes	94.3503	0.9153	0,944	0,027	0,945	0,944	0,916

Tablo 7’de yer alan bilgilere göre model %94.35 oranında doğru tahminde bulunmuştur. Analiz sonucunda elde edilen Kesinlik (precision) değeri 0.945, F-Measure değeri 0.944, MCC (Matthews Korelasyon Katsayısı) değeri ise 0.916 olarak gerçekleşmiştir. Modelin öğrencilerin yerleştiği okul türlerini doğru tahmin edebilme oranı 0.944, yanlış tahmin edebilme oranı ise 0.027 olarak hesaplanmıştır. Ayrıca Naive Bayes analizine ait Kappa değerinin de 0.915 olduğu görülmektedir. Landis ve Koch (1977)’ye göre elde edilen bu değer, geliştirilen modelin oldukça başarılı bir model olduğunu ortaya koymaktadır. Naive Bayes analizine ait hata matrisi Tablo 8’de ayrıntılı bir şekilde verilmiştir.

Tablo 8. Naive Bayes Analizine Ait Hata Matrisi

Tahmin Edilen Sınıflar	Gerçekte olan Sınıflar		
	Düz	Fen	Anadolu
Düz	106	3	4
Fen	4	116	0
Anadolu	9	0	112

Tablo 8’de yer alan verilere göre; geliştirilen model, düz liseye yerleşen öğrencilerin 106’sını, fen lisesine yerleşen öğrencilerin 116’sını ve Anadolu lisesine yerleşen öğrencilerin ise 112’sini doğru tahmin etmiştir.

### İkinci Alt Probleme İlişkin Bulgular

#### *Veri Setindeki Niteliklerin Bilgi Kazanç Değerlerine İlişkin Bulgular*

Veri madenciliğinde yapılan analizler sonucu geliştirilen her bir modelde, veri setindeki niteliklerin bilgi kazanç değerleri önem arz etmektedir. Zira niteliklerden elde edilen bilgi kazançları, karar ağaçlarındaki düğümlerin oluşmasında önemli bir faktör olmaktadır (Özdemir, 2016). Araştırmada kullanılan niteliklerin bilgi kazanç değerlerini elde etmek için WEKA yazılımında information gain yöntemi kullanılmıştır. Bu yöntem için WEKA yazılımındaki Select Attributes özelliğinin altında yer alan “InfoGainAttributeEval” operatörü ile “Ranker” metodu kullanılmıştır. Analiz sonucu niteliklerin bilgi kazanç değerlerine ilişkin elde edilen bulgular Tablo 9’da verilmiştir.

Tablo 9. Bilgi Kazanç Değerleri

Önem Sırası	Nitelik	Önem Değeri
1	A_EGITIM	0.67106
2	KURS	0.66586
3	EV_DURUM	0.58206
4	B_EGITIM	0.58127
5	H_MESLEK	0.57435
6	K_SAYI	0.52914
7	G_DSURE	0.52042
8	G_TV	0.46964
9	YAS	0.41206
10	BOY	0.3248
11	B_YAS	0.31277
12	OKUL_SURE	0.29815
13	KAHVALTI	0.28593
14	C_MASA	0.26079
15	KILO	0.24277
16	B_MESLEK	0.23782
17	C_ODA	0.23092
18	G_TEL	0.2134
19	CINSIYET	0.19702
20	A_YAS	0.19361
21	AILE_DESTEK	0.161
22	A_GELIR	0.15453
23	A_DYER	0.11338
24	A_HASTALIK	0.11046
25	B_DYER	0.10911

26	O_SEVGI	0.09542
27	G_KITAP	0.09293
28	UNI_ISTEK	0.07274
29	E_BILG	0.0549
30	A_MESLEK	0.05227
31	B_HASTALIK	0.03724
32	S_HASTALIK	0.03289
33	B_SAG	0.02525
34	ANASINIF	0.02148
35	EV_ISI	0.01424
36	A_SAG	0.00467

### Üçüncü Alt Probleme İlişkin Bulgular

#### *Karar Ağaçları Analizi Sonucu Oluşan Kurallara İlişkin Bulgular*

Sınıflandırma analizlerinin en önemli çıktılarından biri karar ağaçlarıdır. J48 algoritmasıyla yapılan analiz sonucunda geliştirilen karar ağacının boyutu 20 dal ve 11 yapraktan oluşmaktadır. Oluşturulan ağaç yapısıyla bazı kurallar dizisi de oluşturulabilmektedir. Şekil 1’de karar ağacı sonucu oluşan kurallar dizisi gösterilmektedir.

```

KURS = H
| KAHVALTI = E
| | EV_DURUM = E
| | | A_GELIR = KOTU: DUZ (9.0/1.0)
| | | A_GELIR = IYI: DUZ (2.0)
| | | A_GELIR = ORTA
| | | | G_KITAP <= 15: DUZ (4.0)
| | | | G_KITAP > 15: ANADOLU (114.0/1.0)
| | EV_DURUM = H: DUZ (32.0/5.0)
| KAHVALTI = H: DUZ (52.0/2.0)
KURS = E
| EV_DURUM = E
| | KAHVALTI = E
| | | G_TEL <= 1.49671: FEN (3.0/1.0)
| | | G_TEL > 1.49671: DUZ (3.0)
| | KAHVALTI = H: DUZ (12.0/1.0)
| EV_DURUM = H
| | B_MESLEK = CALISMIYOR: FEN (116.0/1.0)
| | B_MESLEK = CALISIYOR: DUZ (7.0/1.0)

Number of Leaves :    11
Size of the tree :    20
    
```

Şekil 1. Veri Setine Ait Karar Ağacı Yapısı

Şekil 1 incelendiğinde, ağaç yapısındaki ilk dallanmanın, öğrencinin okul dışı zamanlarında özel ders veya kurs alma durumunu belirten KURS niteliği ile başladığı görülmektedir. Bu durum ağaç yapısının oluşumundaki en önemli niteliğin KURS olduğunu ortaya koymaktadır. Ağaç yapısında ortaya çıkan bir diğer önemli nitelik KAHVALTI olarak görülmektedir.

### Dördüncü Alt Probleme İlişkin Bulgular

Araştırmada belirlenen amaca ulaşabilmek için veri madenciliği yöntemlerinden J48, Random Forest, IBK(KNN) ve Naive Bayes kullanılmıştır. Analiz sürecinde Veri madenciliği yazılımlarından olan WEKA 3.9.5 paket programı kullanılmıştır. Veri setindeki okul türü hedef nitelik olarak belirlenmekle birlikte sınavla öğrenci alan Anadolu lisesi, Fen lisesi ve sınavsız öğrenci alan Düz lise olmak üzere 3 kategoride incelenmiştir. Verilerin analizinde kullanılan veri madenciliği algoritmalarından elde edilen bulgular karşılaştırmalı bir şekilde Tablo 10’da yer almaktadır.

Tablo 10. Naive Bayes Analiz Sonuçları

Sınıflandırma Algoritması	DSO (%)	Kappa	Pozitif Tahmin Oranı	Negatif Tahmin Oranı	Kesinlik	F-Measure	MCC
J48	92.0904	0.8813	0,921	0,039	0,921	0,921	0,882
Random Forest	95.1977	0.928	0,952	0,023	0,956	0,953	0,930
IBK	88.7006	0.8302	0,887	0,057	0,887	0,885	0,832
Naive Bayes	94.3503	0.9153	0,944	0,027	0,945	0,944	0,916

Tablo 10’da yer alan bilgilere göre en yüksek doğruluk oranına sahip algoritmanın Random Forest olduğu görülmektedir. Benzer şekilde Kappa istatistiği, pozitif tahmin oranı gibi istatistiklerde de en yüksek değeri veren Random Forest algoritmasının, öğrencilerin hangi lise türüne yerleşeceğini tahmin etmede kullanılabilir en iyi algoritma olduğu söylenebilir. Sara vd., (2015), Danimarkalı lise öğrencilerinin liseyi terk etme oranlarını tahmin ettiği araştırmasında Random Forest algoritması ile %93.47 gibi yüksek bir doğruluk değeri elde etmiştir. Ayrıca Aksu (2018)’in PISA başarısını tahmin etmeyi amaçladığı araştırmada, Sıtkı (2020)’nin hastalıkları tahmin ettiği araştırmada ve Ceyhan (2020)’nin sınıflandırma algoritmalarını karşılaştırdığı araştırmada da Random Forest algoritmasının en yüksek doğruluk değerlerini verdiği tespit edilmiştir. Kullanılan sınıflandırma yöntemlerinin negatif tahmin oranları ise 0.2-0.6 arasında değişmektedir (Tablo 10). Veri madenciliği sürecinde bu oran oldukça düşük olmakla birlikte algoritma performanslarının oldukça yüksek olduğunu ortaya koymaktadır.

### Tartışma, Sonuç ve Öneriler

Bu araştırmada 8. Sınıf öğrencilerinin akademik başarılarını etkilediği düşünülen demografik değişkenler ile sınav başarılarını tahmin etmek ve modellemek amaçlanmıştır. Araştırmanın bu bölümünde, yapılan veri madenciliği analizlerinden elde edilen bulgular ile literatürde yer alan araştırmalar arasında karşılaştırmalar ve öneriler yer almaktadır.

Veri setinden elde edilen sonuçlara göre, öğrencilerin akademik başarılarının bazı değişkenlerle pozitif bir ilişkiye sahip olduğu tespit edilmiştir. Bu değişkenlerden ilki olan öğrencilerin özel kurslara gitme durumları, akademik başarı üzerinde en önemli belirleyici faktör olmuştur. Buna göre özel kurslara giden öğrencilerin akademik başarıları gitmeyen öğrencilere göre daha yüksektir. Yavuz (2020), sekizinci sınıf öğrenciler üzerinde yaptığı araştırmada, kursların öğrencilerin LGS puanlarının ve akademik başarılarının yüksek olmasına katkı sağladığını tespit etmiştir.

Öğrencilerin akademik başarılarını pozitif yönde etkileyen ikinci en önemli değişken ise kahvaltıdır. Öğrencilerin kahvaltı yapma durumlarının akademik başarılarını tahmin etmede önemli bir nitelik olduğu ve kahvaltı yapan öğrencilerin akademik başarılarının, kahvaltı

yapmayan öğrencilere göre daha yüksek olduğu tespit edilmiştir. Budak vd., (2005)'in üniversite öğrencileri üzerinde yapmış olduğu araştırmada kahvaltı yapan öğrencilerin akademik başarısının kahvaltı yapmayan öğrencilere göre daha yüksek olduğunu bulmuştur. Benzer şekilde Baysal (1999), Rampersaud vd., (2005) ve Diremler (2009) da kahvaltı yapmanın akademik başarı üzerinde pozitif yönde bir etkisi olduğunu saptamıştır.

Öğrencilerin kendi evlerinde yaşamaları, akademik başarı üzerinde pozitif bir etkiye sahiptir. Yavuz (2020) de yaptığı araştırmada kendi evlerinde yaşayan öğrencilerin kiralık evde yaşayan öğrencilere göre daha yüksek LGS başarısı elde ettiğini bulmuştur. Günlük cep telefonu kullanım oranı, günlük okunan kitap sayfa sayısı babanın mesleği gibi değişkenlerin de akademik başarı üzerinde etkisi bulunmaktadır. Veri setinden elde edilen bulgulara göre günlük cep telefonu kullanım oranı 1.5 saatten az olan, günde 15 sayfadan fazla kitap okuyan öğrencilerin akademik başarıları daha yüksek olmaktadır. Ayrıca günlük kitap okuma oranı 15 sayfa üzeri olan öğrenciler diğer öğrencilere göre daha yüksek bir akademik başarı göstermektedir. Bardak ve Bardak (2016), üniversite öğrencileri üzerinde yaptığı araştırmada günlük 5 saat ve üzeri cep telefonu kullanımının öğrencinin akademik başarısı üzerinde olumsuz etkiye sahip olduğunu tespit etmiştir. Yılmaz (2012), sekizinci sınıf öğrencileri üzerinde yaptığı araştırmada kitap okuma sıklığı arttıkça 6. Ve 7. Sınıf not ortalamalarının arttığını tespit etmiştir. Bu bulgu araştırma sonuçlarıyla uyum göstermektedir.

Araştırma verileri üzerinde yapılan analizler ile, demografik bilgilerin akademik başarıyla ilişkisine ve başarının tahmin edilebilmesine yönelik bazı kurallar oluşturulmuştur. Bu kurallar şöyle sıralanabilir:

- ✓ EĞER öğrenci okul dışında özel ders/kurs almıyorsa ve okula gelmeden önce kahvaltı yapmıyorsa puansız öğrenci alan Anadolu Liselerine yerleşecek bir başarı elde eder.
- ✓ EĞER öğrenci okul dışında özel ders/kurs almıyorsa ve okula gelmeden önce kahvaltı yapıyorsa ve kaldıkları ev kiralık değil ise ve ailesinin aylık geliri orta düzey ise ve günde 15 sayfadan fazla kitap okuyorsa Anadolu Lisesi'ni kazanabilecek başarı elde eder.
- ✓ EĞER öğrenci okul dışında özel ders/kurs alıyorsa ve oturdukları ev kiralık değil ise ve okula gelmeden kahvaltı yapıyorsa ve günlük telefon kullanımını 1.5 saatin altındaysa Fen Lisesi'ni kazanabilecek başarı elde eder.
- ✓ EĞER öğrenci okul dışında özel ders/kurs alıyorsa ve oturdukları ev kiralık değil ise ve okula gelmeden kahvaltı yapmıyorsa puansız öğrenci alan Anadolu Liselerine yerleşecek bir başarı elde eder.

Veri seti üzerinde j48, Random Forest, IBK (KNN) ve Naive Bayes algoritmaları kullanılarak analizler yapılmıştır. Yapılan analizler sonucunda algoritmaların en düşük 88.70 ve en yüksek 95.19 gibi doğruluk değerleri vererek oldukça başarılı performans gösterdikleri tespit edilmiştir. Yapılan analizler sonucunda en yüksek doğruluk oranına sahip algoritmanın Random Forest olduğu görülmektedir. Benzer şekilde Kappa istatistiği, pozitif tahmin oranı gibi istatistiklerde de en yüksek değeri veren Random Forest algoritmasının, öğrencilerin hangi lise türüne yerleşeceğini tahmin etmede kullanılabilir en iyi algoritma olduğu söylenebilir. Sara vd., (2015), lise öğrencileri üzerinde yaptığı araştırmada Random Forest algoritması ile %93.47 gibi yüksek bir doğruluk değeri elde etmiştir. Ayrıca Aksu (2018)'in PISA başarısını tahmin etmeyi amaçladığı araştırmada, Sıtkı (2020)'nin hastalıkları tahmin ettiği araştırmada ve Ceyhan (2020)'nin sınıflandırma algoritmalarını karşılaştırdığı araştırmada da Random Forest algoritmasının en yüksek doğruluk değerlerini verdiği tespit edilmiştir.

Araştırma sonuçları, öğrencilerin bazı demografik değişkenlerinin akademik başarıları üzerinde pozitif yönde etkisi olduğunu göstermiştir. Öğrencilerin okul dışında özel ders veya kurs almaları, kendilerine ait bir evde oturmaları ve baba mesleği gibi sosyo-ekonomik değişkenlerin akademik başarıyı etkilediği ortaya koyulmuştur. Bu bağlamda sosyo-ekonomik anlamda avantajlı durumda olan öğrencilerin, diğer öğrencilere nazaran daha başarılı olduğu söylenebilir. Literatürde araştırma sonuçlarıyla tutarlı çeşitli çalışmalar mevcuttur. Hakbilen (1984), okul başarısını etkileyen faktörleri incelediği çalışmada da sosyo-ekonomik düzeyi orta ve yüksek olan öğrencilerin sosyo-ekonomik düzeyi düşük olan öğrencilerden daha başarılı notlar aldığını tespit etmiştir. Savaş, Taş ve Duru (2010), Matematikte öğrenci başarısını etkileyen faktörleri incelediği çalışmada özel kurslara giden öğrencilerin gitmeyen öğrencilere nazaran Matematik başarısının daha yüksek olduğu bulgusuna ulaşmıştır. Çömlekçioglulları (2020), yapmış olduğu çalışmada anne-babanın eğitim durumu ve öğrencilerin kendi evlerinde oturma değişkenlerinin öğrenci başarısını pozitif anlamda etkilediğini tespit etmiştir.

Akademik başarı bir eğitim kurumu olan okulların en temel amacıdır. Her okulun yöneticisi öğrenci başarısını artırmak için çeşitli planlamalar yapmaktadır. Öğrenci başarısını etkileyen çok sayıda faktör var iken okul yöneticisinin öğrenci başarısızlığının kaynaklarını tespit etmesi oldukça zor olmaktadır. Bu bağlamda yaptığımız araştırma okul yöneticisine, öğrenci başarısını etkileyen demografik değişkenlerin neler olduğuna dair bazı göstergeler sunmuştur. Sosyo-ekonomik düzey, kahvaltı yapma alışkanlığı, günlük cep telefonu kullanım sıklığı, ikamet edilen evin kira olması gibi değişkenlerin öğrenci başarısı üzerinde etkisinin olması, ortaya konulan göstergelerden bazılarıdır. Okul yöneticileri bu göstergeleri göz önünde bulundurarak yapacağı iyileştirme ve geliştirme çalışmalarıyla öğrencinin akademik başarısını artıracaktır. Özellikle her öğrencinin okula kahvaltı yaparak gelmesini sağlamak ve okul dışı zamanlarda öğrencilere yönelik takviye dersler ve kurslar düzenlemek, okul yöneticilerinin öğrenci başarısını artırma noktasında önemli kazanımlar edinmesini sağlayacaktır.

Araştırmadan elde edilen en önemli çıktılardan biri de öğrencilerin demografik değişkenlerine göre okul yöneticisinin öğrenci başarısını veya başarısızlığını tahmin edebilmesi mümkün olmaktadır. Araştırma sürecinde hazırlanan tahmin modelleri kullanılarak öğrencilerin hangi liseye yerleşebileceğine dair okul yöneticilerinin tahminlerde bulunması mümkün olmaktadır. Bu sonuçlar, okul yöneticisi için oldukça kıymetli olmakla birlikte okullarındaki başarıyı nasıl artırabileceği ya da başarısızlığı nasıl azaltabileceği hususunda önemli bir yol haritası olacaktır.

Araştırmadan elde edilen sonuçlardan hareketle şu önerilerde bulunulabilir:

Yapılan araştırmanın ortaokul öğrencileriyle sınırlandırılmış olduğu göz önünde bulundurulduğunda, ilkokul ve lise öğrencilerine yönelik benzer çalışmaların daha büyük örneklem gruplarıyla yapılması literatüre önemli katkılar sunacaktır.

Araştırma, akademik başarı üzerinde etkili olan demografik bilgilerin tespitine yöneliktir. Bu bağlamda ortaokul öğrencilerinin akademik başarısını etkileyen okulla ve çevreyle ilgili değişkenlerin de tespitine yönelik benzeri araştırmalar yapılabilir.

Sosyo-ekonomik düzey yükseldikçe öğrencilerin akademik başarısı da artmaktadır. Politika yapıcıların sosyo-ekonomik düzeyi düşük bölgelerdeki okullara yönelik çalışmalarını yoğunlaştırması gerekmektedir. Okullarda ise sosyo-ekonomik düzeyi düşük öğrencilere yönelik farkındalığın artırılarak ilave çalışmaların yapılması, öğrenciler arasındaki başarı farkını da azaltacaktır.



Günlük cep telefonu kullanımının akademik başarı üzerindeki olumsuz etkisini ortadan kaldırmak amacıyla, okullarda bağımlılıkla mücadele çalışmalarının her kademedede etkin ve yoğun bir şekilde yapılması faydalı olacaktır.

Kitap okuma oranının akademik başarı üzerinde doğrudan pozitif yönde bir etkisi bulunmaktadır. Öğrencilerin günlük okudukları kitap sayısının artırılmasına yönelik, kitap okuma projelerinin artırılması, ebeveynlerin bu konuda bilinçlendirilmesi, kitap okumanın teşvik edilmesi gibi faaliyetler yapılabilir.

Eğitsel veri madenciliği yeni bir uygulama alanı olması sebebiyle literatürde bu alanla ilgili yapılmış araştırma sayısı sınırlıdır. Bu bağlamda eğitsel veri madenciliği yöntemleri kullanılarak yapılacak çalışmaların sayısı artırılabilir.

Literatür tarandığında veri madenciliği ile ilgili ortaöğretim düzeyindeki öğrencilerin örnekleme dahil edildiği araştırma sayısı diğer kademelere oranla daha az kalmaktadır. Bu örneklem grubuna yönelik daha fazla araştırma yapılmalıdır.

### Lisans Bilgileri

Elektronik Eğitim Bilimleri Dergisi'nde yayımlanan eserler Creative Commons Atıf-Gayri Ticari 4.0 Uluslararası Lisansı ile lisanslanmıştır.

### Copyrights

The works published in Electronic Journal of Education Sciences are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

#### Etik Beyanname

Bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında belirtilen kurallara uyulduğunu ve “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbirini gerçekleştirmediğimizi beyan ederiz. Aynı zamanda yazarlar arasında çıkar çatışmasının olmadığını, tüm yazarların çalışmaya katkı sağladığını ve her türlü etik ihlalinde sorumluluğun makale yazarlarına ait olduğunu bildiririz.

#### Etik Kurul İzin Bilgileri

Etik kurul adı: Gaziantep Üniversitesi Sosyal ve Beşeri Bilimler Etik Kurulu

Etik kurul karar tarihi: 06.06.2022

Etik kurul belgesi sayı numarası: E-87841438-604.01.01-195371

### Kaynakça

- Akgün, K. & Bulut Özek, M. (2020). “Eğitsel Veri Madenciliği Yöntemi İle İlgili Yapılmış Çalışmaların İncelenmesi: İçerik Analizi.” *Uluslararası Eğitim Bilim ve Teknoloji Dergisi*, 6(3): 197-213.
- Aksu, G. (2018). *PISA Başarısını Tahmin Etmede Kullanılan Veri Madenciliği Yöntemlerinin İncelenmesi*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Alan, M. & Yeşilyurt, C. (2018). “Farklı Veri Setleri Üzerinde Smo ve j48 Algoritmalarının Sınıflandırma Sonuçlarının Karşılaştırılması.” *İşletme Bilimi Dergisi*, 6 (3) , 199-213 . DOI: 10.22139/jobs.487388

- Altun, M. (2019). *Öğrenci Akademik Performansının Kestirilmesine İlişkin Bir Model Önerisi: Veri Madenciliğine Dayalı Bir Çalışma*. Yayınlanmamış doktora tezi, Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü, Antalya.
- Ataseven, U., & Yıldırım, P. (2022). Lise Öğrencilerinin Alan Seçimlerinin Eğitsel Veri Madenciliği İle Tahmini. 2nd International Congress of Engineering and Natural Sciences Studies, 07-09 May 2022, Ankara.
- Babaoğlu, A. (2015). *Veri Madenciliği Yöntemleri Ve Bir Uygulama*. Yayınlanmamış yüksek lisans tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Baker, R. S. J. D. (2010). Data Mining For Education. *International Encyclopedia of Education*, 7(3), 112-118.
- Bardak, S. & Bardak, T. (2016). Cep Telefonu Kullanım Sıklığının Öğrencilerin Akademik Başarı Üzerindeki Etkisi: Bartın Meslek Yüksekokulu Örneği. 3. Ulusal Meslek Yüksekokulları Sosyal ve Teknik Bilimler Kongresi, 28-30 Nisan 2016, Aydın.
- Baysal, A. (1999). Kahvaltı ve Okul Başarısı. *Beslenme Ve Diyet Dergisi*, 28(1), 1-3.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32.
- Budak, N., Özer, E., Kovalı, S., & İnceiş, N. (2005). Kahvaltının Öğrencilerin Beslenmesine Katkısı ve Akademik Başarıya Etkisi. *Beslenme Ve Diyet Dergisi*, 33(1), 47-54.
- Büyüköztürk, Ş., Çakmak, E.K., Akgün, Ö.E., Karadeniz, Ş. & Demirel, F. (2013). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-Step Data Mining Guide. SPSS inc. 15 Nisan 2023 tarihinde <http://www.statoo.com/CRISP-DM.pdf> adresinden alınmıştır.
- Cemaloğlu, N. (Ed.). (2021). *Veriye Dayalı Yönetim*. Ankara: Pegem Akademi.
- Ceyhan, G. (2020). *Sınıflandırmada Kullanılan Veri Madenciliği Yöntemlerinin Performanslarının Veri Seti Özelliklerine Göre Karşılaştırılması*. Yayınlanmamış doktora tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Coşkun, C., & Baykal, A. (2011). Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması. *Akademik Bilişim*, 11, 51-58.
- Çömlekçioğulları, A. (2020). *Öğrenci Başarısı İle Ailelerin Sosyo-Ekonomik Düzeyleri Arasındaki İlişki (Denizli İli Örneği)*. Yayınlanmamış tezsiz yüksek lisans projesi, Pamukkale Üniversitesi Eğitim Bilimleri Enstitüsü, Denizli.
- Diremler, C. (2009). *Antalya'da Bulunan Turizm Ve Otelcilik Meslek Liselerinde Okuyan Öğrencilerin Beslenme Durumlarının Akademik Başarılarına Etkisi Üzerine Bir Araştırma*. Yayınlanmamış yüksek lisans tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Ersöz, A. R. (2017). *Eğitsel Veri Madenciliği İle Öğrenci Profillerinin Belirlenmesi*. Yayınlanmamış yüksek lisans tezi. Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü, Bursa.
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning From Class-Imbalanced Data: Review of Methods and Applications. *Expert systems with applications*, 73, 220-239.

- Hakbilen, N. (1984). *İlkokul Çağı Çocuklarının Okul Başarılarını Etkileyen Faktörlerin İncelenmesi*, Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
- Hossin, M., & Sulaiman, M.N., (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process (IJDMP)*, 5(2), 1-11.
- Karaçor, M. (2020). *Üniversite Sınavına Hazırlanan Öğrenciler Arasında Duygu Düzenleme, Anne-Baba Tutumu Ve Sınav Kaygısı Arasındaki İlişkinin İncelenmesi*. Yayınlanmamış yüksek lisans tezi. Hasan Kalyoncu Üniversitesi Sosyal Bilimler Enstitüsü, Gaziantep
- Koyuncu, İ. (2018). *Öğrencilerin PISA Matematik Başarılarının Yordanmasında Veri Madenciliği Yöntemlerinin Karşılaştırılması*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Kurt, Ç. & Erdem, O. A. (2012). Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi. *Politeknik Dergisi*, 15(2), 111-116.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.
- Mining, W. I. D. (2006). Data Mining: Concepts and Techniques. *Morgan Kaufmann*, 10, 559-569.
- Olgun, K. B. (2021). *Ters Yüz Sınıflardaki Video İzleme Davranışları İncelenerek Veri Madenciliği İle Başarının Tahmin Edilmesi*. Yayınlanmamış yüksek lisans tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Omende, Kahudi, J., D. (2021). *Karar Ağaçları Kullanılarak Trafik Kazalarının Nedenlerinin Araştırılması: Sakarya İli Vaka Çalışması*. Yayınlanmamış yüksek lisans tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya.
- Ozgis, M. S., Kaduk, J. D., Jarvis, C. H., da Conceição Bispo, P., & Balzter, H. (2020). Detection Of Oil Pollution Impacts On Vegetation Using Multifrequency Sar, Multispectral Images With Fuzzy Forest And Random Forest Methods. *Environmental pollution (Barking, Essex : 1987)*, 256, 113360. <https://doi.org/10.1016/j.envpol.2019.113360>
- Özdemir, Ş. (2016). *Eğitimde Veri Madenciliği Ve Öğrenci Akademik Başarı Öngörüsüne İlişkin Bir Uygulama*. Yayınlanmamış doktora tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Polat, A. (2021). *Açık Öğretim Liseleri Öğrencilerinin Okul Terki ve Mezuniyet Durumlarının Eğitsel Veri Madenciliği ile İncelenmesi*. Yayınlanmamış doktora tezi, Sakarya Üniversitesi Eğitimleri Enstitüsü, Sakarya.
- Rampersaud, G. C., Pereira, M. A., Girard, B. L., Adams, J., & Metz, J. D. (2005). Breakfast Habits, Nutritional Status, Body Weight, And Academic Performance In Children And Adolescents. *Journal of the american dietetic association*, 105(5), 743-760.
- Sara, N. B., Halland, R., Igel, C., ve Alstrup, S. (2015). High-School Dropout Prediction Using Machine Learning: A Danish Large-Scale Study. In M. Verleysen (Ed.), Proceedings. ESANN 2015: 23rd European Symposium on Artificial Neural Networks,

- Computational Intelligence and Machine Learning, 319-324. 20 Nisan 2023 tarihinde <https://core.ac.uk/download/pdf/269280541.pdf> adresinden erişim sağlanmıştır.
- Sarıer, Y. (2016). Türkiye'de Öğrencilerin Akademik Başarısını Etkileyen Faktörler: Bir Meta-analiz Çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 31(3), 609 - 627.
- Savaş, E., Taş, S. & Duru, A. (2010). Matematikte Öğrenci Başarısını Etkileyen Faktörler. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 11 (1) , 113-132 .
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review On Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526-534.
- Sıtkı, Y. H. (2020). *Tıp Bilişiminde Veri Madenciliği Yöntemleri Kullanılarak Hastalıkların Tahmin Edilmesi*. Yayınlanmamış yüksek lisans tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Topuz, S. (2021). *Eğitsel Verilerde Weka Ve Orange Veri Madenciliği Yazılımlarından Elde Edilen Analiz Sonuçlarının Karşılaştırılması*. Yayınlanmamış yüksek lisans tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Üre, N. (2021). *Veri Madenciliği Sınıflandırma Algoritmalarının Performans Karşılaştırması: Tiroid Hastalığının Tahmini Üzerinde Bir Uygulama*. Yayınlanmamış yüksek lisans tezi, Kafkas Üniversitesi Sosyal Bilimler Enstitüsü, Kars.
- Vijayarani, S., & Muthulakshmi, M. (2013). Comparative Analysis Of Bayes And Lazy Classification Algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(8), 3118-3124.
- Weka, (2022). Weka 3: Machine Learning Software in Java. 15 Ocak 2023 tarihinde <https://www.cs.waikato.ac.nz/ml/weka/> adresinden alınmıştır.
- Yavuz, A. (2020). *Ortaöğretime Geçiş Sınavında Öğrenci Başarısını Etkileyen Etmenler*. Yayınlanmamış tezsiz yüksek lisans projesi, Pamukkale Üniversitesi Eğitim Bilimleri Enstitüsü, Denizli.
- Yılmaz, B. (2012). Okuma Alışkanlığının Okul Başarısına Etkisi: Ankara Keçiören Atapark İlköğretim Okulu Öğrencileri Üzerine Bir Araştırma. Külcü, Ö., Çakmak. T., Özel, N. (Editörler), Prof. Dr. K. Gülbün Baydur'a armağan (210- 218). Ankara: Özyurt Matbaacılık. 1 Nisan 2023 tarihinde <https://www.openaccess.hacettepe.edu.tr/xmlui/bitstream/handle/11655/11695/Y%c4%b1lmaz.pdf?sequence=1&isAllowed=y> adresinden erişim sağlanmıştır.