



**T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİ TEKNOLOJİLERİ ANA BİLİM DALI**

YÜKSEK LİSANS TEZİ

**ORTAOKUL ÖĞRENCİLERİNİN
BAŞARILARINI ETKİLEYEN FAKTÖRLERİN
VERİ BİLİMİ VE MAKİNE ÖĞRENMESİ
YÖNTEMLERİ KULLANILARAK
ANALİZ EDİLMESİ**

Ferit ÖZTEKİN

**Ağustos-2024
BATMAN**

**T.C.
BATMAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİ TEKNOLOJİLERİ ANA BİLİM DALI**

YÜKSEK LİSANS TEZİ

**ORTAOKUL ÖĞRENCİLERİNİN
BAŞARILARINI ETKİLEYEN FAKTÖRLERİN
VERİ BİLİMİ VE MAKİNE ÖĞRENMESİ
YÖNTEMLERİ KULLANILARAK
ANALİZ EDİLMESİ**

Ferit ÖZTEKİN

**Danışman
Prof. Dr. Necmettin SEZGİN**

Diğer Jüri Yeleri

Doç. Dr. Melih KUNCAN Prof. Dr. Necmettin SEZGİN Doç. Dr. Yılmaz KAYA

**Ağustos-2024
BATMAN**

TEZ KABUL VE ONAYI

Ferit ÖZTEKİN tarafından hazırlanan “ORTAOKUL ÖĞRENCİLERİNİN BAŞARILARINI ETKİLEYEN FAKTÖRLERİN VERİ BİLİMİ VE MAKİNE ÖĞRENMESİ YÖNTEMLERİ KULLANILARAK ANALİZ EDİLMESİ” adlı tez çalışması 09.08.2024 tarihinde aşağıdaki jüri tarafından oy birliği ile Batman Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgi Teknolojileri Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

İmza

Başkan

Doç. Dr. Melih KUNCAN

.....

Danışman

Prof. Dr. Necmettin SEZGİN

.....

Üye

Doç. Dr. Yılmaz KAYA

.....

Yukarıdaki sonucu onaylarım.

Dr. Öğr. Üyesi Ömer Murat ÖTER
Lisansüstü Eğitim Enstitüsü Müdürü

ETİK BEYANI

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını beyan eder, aksinin ortaya çıkması durumunda her türlü yasal sorumluluğu kabullendiğimi bildiririm.

ETHICAL DECLARATION

I declare that all the information in this thesis has been obtained within the framework of ethical behavior and academic rules, and that the source of any statements and information that do not belong to me in this study prepared in accordance with the thesis writing rules has been fully cited, and I declare that I accept all kinds of legal responsibility in case of any contrary situation.

Ferit ÖZTEKİN

Tarih: 09.08.2024

ÖZET

YÜKSEK LİSANS TEZİ

ORTAOKUL ÖĞRENCİLERİNİN BAŞARILARINI ETKİLEYEN FAKTÖRLERİN VERİ BİLİMİ VE MAKİNE ÖĞRENMESİ YÖNTEMLERİ KULLANILARAK ANALİZ EDİLMESİ

Ferit ÖZTEKİN

Batman Üniversitesi Lisansüstü Eğitim Enstitüsü

Bilgi Teknolojileri Ana Bilim Dalı

Danışman: Prof. Dr. Necmettin SEZGİN

2024, 58 sayfa

Bu çalışma, ortaokul düzeyindeki öğrenci başarılarını etkileyen faktörlerin belirlenmesi ve analiz edilmesi amacıyla veri bilimi ve makine öğrenmesi yöntemlerinin kullanılmasını hedeflemektedir. Eğitimde veri odaklı bir yaklaşımın önemini vurgulayan bu çalışma, öğrenci başarısını etkileyen karmaşık faktörlerin anlaşılmasına ve eğitim politikalarının daha etkili bir şekilde tasarlanmasına katkı sağlamayı amaçlamaktadır. Çalışmanın temel odak noktası, ortaokul seviyesindeki öğrencilerin başarılarını etkileyen faktörlerin belirlenmesidir. Öğrenci başarısını etkileyen faktörler arasında öğrenci özellikleri, aile faktörleri, okul ortamı ve öğretmen faktörleri gibi çeşitli değişkenler bulunmaktadır. Bu faktörlerin belirlenmesi, eğitim sisteminin zayıf noktalarını tespit etmek ve öğrencilerin başarısını artırmak için bireyselleştirilmiş müdahalelerde bulunmak için önemlidir. Bu çalışma, Batman-Merkez Ortaokullarında öğrencilere başarı düzeylerini etkileyebilecek faktörleri belirlemek amacıyla yapılmıştır. Çalışma kapsamında öğrencilere başarı düzeylerini etkileyebilecek faktörleri değerlendirmek için bir form verilerek doldurmaları istenmiştir. Öğrencilerin bu formları doldurmasıyla elde edilen yanıtlar ile ders dönemi ortalamaları ilgili okullardan alınmıştır. Bu veriler, veri bilimi ve makine öğrenmesi yöntemleri kullanılarak analiz edilmiştir. Veri bilimi ve makine öğrenmesi yöntemleri, büyük veri setlerini analiz etmek ve öğrenci başarısını etkileyen faktörleri tanımlamak için güçlü araçlar sunar. Veri analizi için grafik türleri, regresyon modelleri ve sınıflandırma gibi yöntemler kullanılarak öğrenci başarısını etkileyen faktörler belirlenmiş ve öğrencilerin ders ortalaması tahmin edilerek öğrencinin zayıf noktaları tespit edilerek bireyselleştirilmiş çözüm sunulmuştur. Bu çalışmanın sonuçları, eğitimcilerin, eğitimde idari birimlerin, velilerin ve öğrencilerin ortaokul düzeyindeki öğrenci başarılarını artırmak için daha etkili stratejiler geliştirmesine yardımcı olabilir. Ayrıca, veri bilimi ve makine öğrenmesi gibi yenilikçi yöntemlerin eğitim araştırmalarında nasıl kullanılabileceğine dair bir örnek sunarak, bu alanda gelecekteki araştırmalara da katkı sağlayabilir.

Anahtar Kelimeler: Makine Öğrenmesi, Veri Bilimi, Regresyon, Sınıflandırma, Tahmin Algoritmaları, Öğrenci Başarı Faktörleri

ABSTRACT

MS THESIS

ANALYSIS OF FACTORS AFFECTING MIDDLE SCHOOL STUDENT ACHIEVEMENT USING DATA SCIENCE AND MACHINE LEARNING METHODS

Ferit ÖZTEKİN

Batman University Graduate Education Institute

Department of Information Technologies

Advisor: Prof. Dr. Necmettin SEZGİN

2024, 58 pages

This study aims to identify and analyze the factors affecting middle school student achievement using data science and machine learning methods. Emphasizing the importance of a data-driven approach in education, this study aims to contribute to the understanding of the complex factors influencing student achievement and the design of more effective educational policies. The primary focus of the study is to determine the factors affecting the success of middle school students. These factors include student characteristics, family factors, school environment, and teacher factors. Identifying these factors is crucial for pinpointing weaknesses in the education system and implementing personalized interventions to improve student achievement. This study was conducted to determine the factors affecting student achievement levels in middle schools in Batman-Merkez. As part of the study, students were asked to complete a form to evaluate factors that could affect their achievement levels. The responses obtained from these forms were combined with the students' term averages obtained from the relevant schools. These data were analyzed using data science and machine learning methods. Data science and machine learning methods offer powerful tools for analyzing large datasets and identifying factors that affect student achievement. Various graphical analyses, regression models, and classification methods were used to identify and predict the factors influencing student achievement. By predicting students' grade point averages, their weaknesses were identified, and personalized solutions were provided. The results of this study can help educators, administrative units in education, parents, and students develop more effective strategies to improve student achievement at the middle school level. Furthermore, by demonstrating how innovative methods such as data science and machine learning can be used in educational research, it can contribute to future studies in this field.

Keywords: Machine Learning, Data Science, Regression, Classification, Prediction Algorithms, Student Achievement Factors

ÖNSÖZ

Yüksek Lisans tezimin hazırlık ve tamamlanması sürecinde birçok kişiye teşekkür etmek isterim. Öncelikle tez danışmanım Prof. Dr. Necmettin SEZGİN' e, rehberliği, önerileri ve sabrı için teşekkür etmek istiyorum. Beni tez çalışmam boyunca desteklediği ve yol gösterdiği için minnettarım.

Ayrıca, tez çalışmamın gerçekleştirilmesine katkıda bulunan tüm öğretmenlere, okul idaresine ve katılımcı öğrencilere teşekkür etmek istiyorum. Onların işbirliği ve katkıları olmadan, bu çalışma mümkün olmazdı.

Umuyorum ki bu tez, eğitim alanındaki araştırmacılara, eğitimcilere ve eğitimde model tasarlayıcılara faydalı olacak yeni bilgiler ve anlayışlar sağlar.

Ferit ÖZTEKİN
BATMAN-2024

İÇİNDEKİLER

ÖZET	iv
ABSTRACT.....	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
TABLolar LİSTESİ	ix
ŞEKİLLER LİSTESİ	x
SİMGELER VE KISALTMALAR	xi
1. GİRİŞ	1
1.1 Amaç	2
1.2 Kapsam	3
1.3 Sınırlılıklar	3
2. KAYNAK ARAŞTIRMASI	4
2.1 Öğrenci Başarısını Etkileyen Faktörler	4
2.1.1 Aile faktörü	4
2.1.2 Kişisel özellikler	4
2.1.3 Okul ve çevre faktörü	5
2.1.4 Öğretmen faktörü	5
2.2 Veri Bilimi	6
2.2.1 Veri toplama	6
2.2.2 Veri temizleme ve işleme	6
2.2.3 Veri analizi	7
2.2.4 Veri görselleştirme	7
2.2.5 Python ve kütüphaneleri	7
2.3 Makine Öğrenmesi	9
2.3.1 Denetimli öğrenme (Supervised learning)	9
2.3.2 Denetimsiz öğrenme (Unsupervised learning)	10
2.3.3 Pekiştirmeli öğrenme (Reinforcement learning)	10
2.3.4 Derin öğrenme (Deep learning)	11
2.3.5 Model değerlendirme ve validasyon	11
2.3.6 Regresyon modelleri	11
2.3.7 Sınıflandırma modelleri	12
2.4 Eğitimde Veri Bilimi ve Makine Öğrenmesi Yöntemleri	12
2.4.1 Öğrenci başarısını tahmin etmek için kullanılan yöntemler	12
2.5 Veri Bilimi ve Makine Öğrenmesi ile İlgili Örnek Çalışmalar	14
3. MATERYAL VE YÖNTEM.....	16
3.1. Araştırma Verileri ve Veri Toplama Yöntemleri	16
3.1.1. Anket ve veri toplama süreci	16

3.1.2. Veri setinin hazırlanması ve analiz için kullanılan yöntemler.....	17
3.2. Veri Analiz Yöntemleri	18
3.2.1. Korelasyon analizi	19
3.2.2. Regresyon analizi.....	19
3.2.3. Sınıflandırma analizi.....	19
4. BULGULAR VE TARTIŞMA.....	20
4.1 Korelasyon Analizi Sonuçları	20
4.2 Dağılım Grafikleri.....	26
4.3 Kutu Grafikleri ve Bulgular	29
4.4 Çapraz Tablolar ve Isı Haritası Bulguları	33
4.5. Regresyon Analizi Sonuçları	36
4.5.1. Çoklu doğrusal regresyon (Multiple linear regression) sonuçları	36
4.6 Sınıflandırma analizi.....	43
5. SONUÇLAR VE ÖNERİLER	47
5.1. Genel Sonuçlar.....	47
5.1.1. Belirli bulgular	47
5.2. Öneriler	48
5.2.1. Eğitim politikaları için öneriler.....	48
5.2.2. Gelecek çalışmalar için öneriler	48
KAYNAKLAR	49
EKLER	51

TABLolar LİSTESİ

Tablo 3.1. Öğrenci anket soruları.....	16
Tablo 4.1. Aile ve Kişilik özellikleri not ortalamaları.....	33
Tablo 4.2. Kişilik özellikleri ile Rehberlik arasındaki not ortalamaları.....	34
Tablo 4.3. Okul ve çevre ile Öğretmen etkisi arasındaki not ortalamaları.....	35
Tablo 4.4. Ridge ve Lasso regresyon katsayıları.....	37
Tablo 4.5. Öğrenci not tahmin tablosu.....	39
Tablo 4.6. Sınıflandırma modellerinin performans değerlendirmesi.....	44

ŞEKİLLER LİSTESİ

Şekil 4.1.	Aile durumları ile Not arasındaki korelasyon matrisi.....	20
Şekil 4.2.	Kişisel özellikler ve Not arasındaki korelasyon.....	21
Şekil 4.3.	O1 ve O2 ile Not arasındaki korelasyon matrisi.....	22
Şekil 4.4.	R1 ve R2 ile Not arasındaki korelasyon matrisi.....	23
Şekil 4.5.	T1, T2 ve T3 ile Not arasındaki korelasyon matrisi.....	24
Şekil 4.6.	Tüm değişkenlerin korelasyonu.....	25
Şekil 4.7.	Aile histogramı.....	26
Şekil 4.8.	Değişkenlerin Not ile korelasyonu.....	27
Şekil 4.9.	Değişkenlerin not yoğunluk korelasyonu.....	28
Şekil 4.10.	Aile değişkenleri ile Not arasındaki korelasyon.....	29
Şekil 4.11.	Kişilik özellikleri ile Not arasındaki korelasyon.....	30
Şekil 4.12.	Okul ve Çevre değişkenleri ile Not arasındaki korelasyon.....	31
Şekil 4.13.	Öğretmen etkisi ile Not arasındaki korelasyon.....	31
Şekil 4.14.	Rehberlik ile Not arasındaki korelasyon.....	32
Şekil 4.15.	Lojistik regresyon özellik katsayıları.....	44
Şekil 4.16.	SVM destek vektörleri.....	46

SİMGELER VE KISALTMALAR

Simge	Açıklama
XXX	Bağımsız değişkenler matrisi
yyy	Bağımlı değişken (Not)
β	Regresyon katsayısı
ϵ	Hata terimi
R^2	Belirleme katsayısı (R-kare)
MSE	Ortalama Kare Hatası
ROC	Alıcı İşletim Karakteristiği Eğrisi
AUC	Eğri Altındaki Alan

Kısaltma	Açıklama
ML	Makine Öğrenmesi (Machine Learning)
DS	Veri Bilimi (Data Science)
SVM	Destek Vektör Makineleri (Support Vector Machines)
KNN	K-En Yakın Komşu (K-Nearest Neighbors)
NB	Naive Bayes
LR	Lojistik Regresyon (Logistic Regression)
RF	Rastgele Orman (Random Forest)
ANN	Yapay Sinir Ağları (Artificial Neural Networks)
CNN	Evrişimli Sinir Ağları (Convolutional Neural Networks)
RNN	Tekrarlayan Sinir Ağları (Recurrent Neural Networks)
PCA	Ana Bileşen Analizi (Principal Component Analysis)
Q-Learning	Q-Öğrenme (Reinforcement Learning)
DQN	Derin Q-Ağları (Deep Q-Networks)

1. GİRİŞ

Eğitim, bireylerin kişisel ve toplumsal gelişiminde kritik bir rol oynamaktadır. Eğitim sistemi, bireylerin bilgi ve becerilerini geliştirmek, potansiyellerini ortaya çıkarmak ve topluma katkıda bulunmalarını sağlamak amacıyla tasarlanmıştır (Darling-Hammond, 2000, s. 3). Ancak, öğrencilerin eğitim başarısı üzerinde etkili olan birçok faktör bulunmaktadır ve bu faktörlerin karmaşıklığı eğitimciler ve araştırmacılar için sürekli bir ilgi konusu olmuştur (Hattie, 2008, s. 12). Ortaokul düzeyinde öğrenci başarılarını etkileyen faktörlerin anlaşılması ve analiz edilmesi, eğitim politikası ve uygulamalarının geliştirilmesi açısından önemlidir. Bu faktörlerin belirlenmesi, eğitim sisteminin zayıf noktalarını tespit etmek ve öğrencilerin başarısını artırmak için müdahalelerde bulunmak için bir temel sağlayabilir. Bu araştırma, ortaokul düzeyinde öğrenci başarılarını etkileyen faktörleri belirlemek ve bu faktörlerin veri bilimi ve makine öğrenmesi yöntemleriyle analizini gerçekleştirmeyi amaçlamaktadır. Veri bilimi ve makine öğrenmesi, büyük veri setlerini analiz etmek ve örüntüleri tespit etmek için güçlü araçlar sunar (Shalev-Shwartz & Ben-David, 2014, s. 20). Bu yöntemlerin eğitim alanında kullanılması, öğrenci başarısını etkileyen faktörlerin daha derinlemesine anlaşılmasına ve eğitim politikalarının daha etkili bir şekilde tasarlanmasına olanak tanır. Bu çalışmanın bulguları, eğitimcilerin ve eğitim tasarımcılarının ortaokul düzeyinde öğrenci başarısını artırmak için daha etkili stratejiler geliştirmesine yardımcı olabilir. Ayrıca, bu çalışma, veri bilimi ve makine öğrenmesi gibi yenilikçi yöntemlerin eğitim araştırmalarında nasıl kullanılabileceğine dair bir örnek sunarak, bu alan üzerindeki gelecekteki araştırmalara da katkı sağlayabilir.

Öğrenci başarısı, öğrencinin eğitim sürecinde belirlenen hedeflere ulaşma düzeyini ifade eder. Bu başarı, genellikle notlar, sınav sonuçları ve akademik projeler gibi performans ölçütleriyle değerlendirilir. Akademik başarı, öğrencinin bilgi, beceri ve yeteneklerini ne kadar etkili kullanabildiğini gösterir ve eğitim sürecinin kalitesini yansıtır (Marzano, 2003, s. 45). Öğrenci başarısı, bireyin gelecekteki eğitim ve kariyer yolculuğunda önemli bir belirleyici olarak kabul edilir. Bu nedenle, öğrenci başarısının doğru ve kapsamlı bir şekilde değerlendirilmesi, eğitim politikalarının ve öğretim stratejilerinin geliştirilmesinde kritik bir rol oynar. Akademik başarı, sadece bireysel performansla sınırlı kalmayıp, aynı zamanda sosyal ve duygusal becerilerin de bir yansımasıdır. Öğrencilerin problem çözme yetenekleri, eleştirel düşünme becerileri ve sosyal etkileşim

kapasiteleri, akademik başarılarına katkıda bulunur. Bu nedenle, öğrenci başarısını değerlendirmek için kullanılan kıstas yalnızca akademik notları değil, aynı zamanda öğrencinin genel gelişimini ve potansiyelini de yansıtmalıdır. Eğitim sistemleri, öğrencilerin bu çok boyutlu başarılarını desteklemek için bütüncül bir yaklaşım benimsemelidir.

1.1 Amaç

Bu araştırma ortaokul öğrencilerinin akademik başarısını etkileyen faktörleri tanımlamak ve bu faktörlerin öğrenci performansı üzerindeki etkilerini kapsamlı bir şekilde analiz etmektir. Çalışma, veri bilimi ve makine öğrenmesi yöntemlerini kullanarak, eğitim süreçlerinde karşılaşılan zorluklara yenilikçi çözümler sunmayı ve eğitim politikalarının geliştirilmesine katkıda bulunmayı hedeflemektedir. Özellikle, öğrenci başarısında belirleyici olan çok boyutlu faktörlerin detaylı bir şekilde incelenmesi, bu faktörlerin öğrencilerin yaşadığı bölgelerde öğrenci başarısı üzerinde nasıl bir etki olabileceğini, eğitimde bireyselleştirilmiş müdahale stratejilerinin tasarlanmasında yol gösterici olacaktır. Araştırma, öğrenci başarısını etkileyen ana faktörleri tanımlayarak başlar. Bu faktörler arasında ailevi durumlar, öğrencilerin kişisel özellikleri, okulun fiziksel ve sosyal çevresi, öğretmenlerin eğitim yaklaşımları yer alır. Bu faktörlerin her birinin, öğrenci başarısına olan katkılarını ve etkileşimlerini derinlemesine analiz etmek amaçlanmaktadır. Bu faktörlerin etkileri, ileri veri analizi teknikleri ve makine öğrenmesi modelleri kullanılarak değerlendirilecektir. Özellikle, regresyon analizleri, sınıflandırma algoritmaları ve korelasyon analizleri gibi yöntemlerle bu etkileşimler nicel olarak modelleneyecektir. Bu analizler, öğrenci başarısını etkileyen faktörler arasındaki ilişkileri ortaya çıkararak, hangi özel değişkenlerin veri toplanan öğrencilerin bulunduğu bölgede etkin olduğunu belirlemeyi amaçlar. Analiz sonuçları, eğitimciler ve eğitim tasarımcıları için öğrenci başarısını artırmaya yönelik stratejik öneriler sunmak üzere kullanılacaktır. Bu öneriler, öğrencilere yönelik bireyselleştirilmiş eğitim müdahalelerinin tasarlanmasında, öğretim metotlarının iyileştirilmesinde ve okul çevrelerinin daha destekleyici hale getirilmesinde kritik rol oynayacaktır. Veri bilimi ve makine öğrenmesi tekniklerinin eğitim alanındaki uygulamalarına dair değerli bilgiler sağlayarak, eğitim politikalarının şekillendirilmesine yönelik bilimsel bir temel oluşturmayı amaçlamaktadır. Bu sayede, eğitim sistemlerinin daha verimli ve etkili bir şekilde tasarlanmasına yardımcı olunması hedeflenmektedir.

1.2 Kapsam

Bu araştırma, Batman-Merkez ilçesinde bulunan ortaokul öğrencilerinin akademik performansını etkileyen faktörleri detaylı bir şekilde incelemektedir. Çalışma, öğrencilerin aile yapısı, kişisel özellikleri, okul çevresi ve öğretmen etkileri gibi çeşitli faktörleri ele almakta ve bu faktörlerin öğrenci başarısı üzerindeki etkilerini değerlendirmektedir. Araştırma süreci, geniş kapsamlı bir anket uygulaması ile başlamış; öğrencilerden elde edilen sosyoekonomik ve eğitimle ilgili veriler, akademik performans kayıtlarıyla birleştirilmiştir. Bu veri birleşimi, çeşitli faktörlerin öğrenci başarıları üzerindeki etkilerini daha doğru bir şekilde analiz etmeye olanak tanımaktadır. Veri analizi sürecinde, modern veri bilimi ve makine öğrenmesi yöntemleri kullanılmıştır. Korelasyon analizi ile değişkenler arasındaki ilişkilerin gücü ve yönü belirlenmiş; çoklu regresyon analizi ile bağımsız değişkenlerin bağımlı değişken üzerindeki etkileri değerlendirilmiştir. Sınıflandırma teknikleri, öğrencilerin başarı durumlarını öngören modellerin geliştirilmesinde kullanılmıştır. Bu teknikler, büyük veri setlerinin işlenmesi ve yorumlanması için tercih edilmiş olup, araştırma sonuçlarının nesnel ve nicel bir temele oturtulmasını sağlamaktadır.

1.3 Sınırlılıklar

Araştırma Batman Merkez ilçesindeki ortaokullarda öğrenim gören öğrenciler üzerine odaklanmıştır. Bu nedenle, elde edilen sonuçların diğer bölgelerdeki okullara veya farklı eğitim kademelerine doğrudan genelleştirilmesi sınırlıdır. Bulguların başka bölgelere uygulanabilirliği konusunda dikkatli olunmalıdır. Bölgesel, kültürel, ekonomik ve eğitimsel farklılıklar, araştırma sonuçlarının başka bölgelere uyarlanmasını zorlaştırabilir. Ayrıca, anketler aracılığıyla toplanan veriler, öğrencilerin öznel algılarına dayandığı için cevaplar, öğrencilerin o anki ruh hali, anlayış seviyeleri ve anketi doldurma durumlarındaki çeşitli dış etkenlerden etkilenebilir. Bu durum, verilerin güvenilirliğini ve genel geçerliliğini sınırlayabilir. Veri analizinde kullanılan istatistiksel yöntemler, var olan verilere dayalı çıkarımlar yapar; ancak bu yöntemler bazı varsayımlara bağlıdır ve her zaman belirli sonuçları tespit edemeyebilir. Özellikle korelasyon analizi, nedensellik ilişkilerini belirleyemez ve bazı etkileşimler göz ardı edilebilir.

2. KAYNAK ARAŞTIRMASI

2.1 Öğrenci Başarısını Etkileyen Faktörler

Öğrenci başarısını etkileyen birçok faktör bulunmaktadır. Bu faktörler genellikle öğrencinin bireysel özellikleri, aile desteği, okul ortamı ve öğretmen etkisi gibi ana kategorilere ayrılır. Bu kategorilerin her biri, öğrencilerin akademik performansını doğrudan veya dolaylı olarak etkileyebilir ve bu nedenle ayrıntılı bir şekilde incelenmesi gerekmektedir. Öğrenci başarısını etkileyen faktörlerin anlaşılması, eğitimde iyileştirme ve müdahale stratejilerinin geliştirilmesi açısından önemlidir. Her bir faktörün etkisini ayrı ayrı ve birbiriyle ilişkili olarak değerlendirmek, eğitim politikalarının ve bireysel eğitim planlarının daha etkili bir şekilde tasarlanmasına yardımcı olabilir. Bu bağlamda, öğrenci başarısını etkileyen ana faktörler detaylı olarak incelenecektir.

2.1.1 Aile faktörü

Aile, öğrencinin eğitim sürecindeki başarısını önemli ölçüde etkileyen bir faktördür. Aile desteği, ebeveynlerin eğitim seviyesi, ekonomik durum ve evdeki eğitim atmosferi, öğrencinin akademik performansını doğrudan etkileyebilir. Ebeveynlerin eğitim düzeyi, çocuklarının eğitime verdikleri önem ve sağladıkları destek ile ilişkilidir. Yüksek eğitim düzeyine sahip ebeveynler, çocuklarının eğitim süreçlerine daha fazla dahil olarak onların akademik başarılarını artırabilirler (Eccles & Harold, 1993). Ailelerin ekonomik durumu da öğrencilerin eğitime doğrudan etki eder. Ekonomik açıdan avantajlı aileler, çocuklarına daha iyi eğitim kaynakları sağlayabilir, özel dersler aldırabilir ve eğitimle ilgili materyalleri temin edebilirler. Bunun yanı sıra, evdeki eğitim atmosferi, öğrencinin öğrenmeye olan ilgisini ve motivasyonunu artırabilir. Aile içindeki destekleyici ve teşvik edici tutumlar, öğrencinin akademik başarıya ulaşmasında önemli bir rol oynar (Eccles & Harold, 1993).

2.1.2 Kişisel özellikler

Öğrencinin kişisel özellikleri, akademik başarı üzerinde büyük bir etkiye sahiptir. Bu özellikler arasında öğrencinin motivasyonu, öz-düzenleme becerileri, çalışma alışkanlıkları ve öğrenme stili bulunur. Yüksek motivasyona sahip öğrenciler, öğrenmeye karşı daha istekli olur ve akademik görevleri daha özverili bir şekilde yerine getirir-

ler. Öz-düzenleme becerileri ise, öğrencinin kendi öğrenme süreçlerini planlama, izleme ve değerlendirme yeteneğini ifade eder (Zimmerman, 1990). Çalışma alışkanlıkları ve öğrenme stili de öğrenci başarısını etkileyen önemli kişisel özelliklerdir. Etkili çalışma alışkanlıkları geliştiren öğrenciler, derslerinde daha başarılı olabilirler. Ayrıca, öğrenme stilleri farklılık gösteren öğrencilerin, kendi öğrenme tarzlarına uygun stratejiler geliştirmeleri önemlidir. Bu, öğrencilerin öğrenme sürecini daha etkili hale getirir ve akademik performanslarını artırır (Zimmerman, 1990).

2.1.3 Okul ve çevre faktörü

Okulun fiziksel olanakları, sınıf mevcudu, eğitim teknolojileri ve çevre faktörleri, öğrencinin öğrenme deneyimini etkileyen önemli unsurlardır. İyi donanımlı sınıflar, yeterli eğitim materyalleri ve modern teknolojilerle desteklenen okullar, öğrencilerin öğrenme süreçlerini olumlu yönde etkiler. Ayrıca, sınıf mevcudu da önemli bir faktördür; daha küçük sınıflarda öğretmenler, her öğrenciye daha fazla ilgi gösterebilir ve bireysel ihtiyaçlara daha iyi cevap verebilirler (Leithwood & Jantzi, 2000). Çevresel faktörler de öğrenci başarısında önemli bir rol oynar. Güvenli ve destekleyici bir okul ortamı, öğrencilerin akademik performansını artırabilir. Okulun sosyal ve kültürel ortamı, öğrencilerin öğrenme motivasyonunu ve okul bağlılığını etkiler. Destekleyici bir çevre, öğrencilerin kendilerini değerli hissetmelerini sağlar ve öğrenme sürecine daha aktif katılım göstermelerine yardımcı olur.

2.1.4 Öğretmen faktörü

Öğretmenlerin nitelikleri, öğretim yöntemleri ve öğrenciyle kurdukları ilişki, öğrencinin akademik başarısını doğrudan etkileyen önemli faktörlerdir. Nitelikli öğretmenler, öğrencilerin öğrenme motivasyonunu ve performansını artırır. Etkili öğretim yöntemleri kullanan öğretmenler, öğrencilerin ders konularını daha iyi anlamalarına ve öğrenmelerine yardımcı olur. Öğretmenlerin, öğrencilerle olumlu ve destekleyici ilişkiler kurması da öğrencilerin akademik başarılarını artırır (Hattie, 2009). Öğretmenlerin profesyonel gelişimi ve sürekli eğitimi, onların niteliklerini artırarak öğrencilere daha iyi rehberlik etmelerini sağlar. Ayrıca, öğretmenlerin sınıf yönetimi becerileri ve pedagojik yaklaşımları, öğrencilerin öğrenme deneyimini doğrudan etkiler. Öğrencilerin akademik başarılarını artırmak için öğretmenlerin, öğrencilere bireysel olarak yaklaşmaları ve onların öğrenme süreçlerini desteklemeleri önemlidir.

2.2 Veri Bilimi

Veri bilimi, büyük ve karmaşık veri setlerinden anlamlı bilgi çıkarma sürecidir. Bu disiplin, istatistik, bilgisayar bilimi ve matematik gibi çeşitli alanları bir araya getirir. Veri bilimi, veri toplama, temizleme, işleme, analiz etme ve görselleştirme aşamalarını içerir. Amaç, verilerden yararlı bilgiler elde ederek, karar verme süreçlerini desteklemek ve problemlere çözümler bulmaktır. Veri bilimi, eğitim alanında öğrenci başarılarını analiz etmek ve tahmin etmek için güçlü araçlar sunmaktadır. Bu yöntemler, büyük veri setlerinden anlamlı bilgiler çıkararak, öğrenci başarılarını etkileyen faktörleri daha derinlemesine anlamamıza olanak tanır. Veri bilimi teknikleri, eğitim verilerinin temizlenmesi, işlenmesi ve analiz edilmesi süreçlerinde kritik rol oynar. Bu süreçler, eğitimcilerin ve politika yapıcıların daha bilinçli kararlar almasına yardımcı olabilir.

2.2.1 Veri toplama

Veri toplama, veri biliminin ilk ve en önemli adımlarından biridir. Bu aşama, veri kaynaklarının belirlenmesi ve verilerin uygun yöntemlerle toplanmasını içerir. Veri, çeşitli kaynaklardan, örneğin sensörlerden, anketlerden, dijital cihazlardan ve internetten toplanabilir. Eğitim verileri, öğrenci kayıtları, sınav sonuçları, anketler ve gözlem raporları gibi çeşitli kaynaklardan elde edilebilir. Bu veriler, öğrenci başarılarını analiz etmek ve çeşitli faktörlerin etkisini değerlendirmek için kullanılır. Veri toplama sürecinde dikkat edilmesi gereken en önemli husus, verilerin doğruluğu ve güvenilirliğidir. Güvenilir ve geçerli veriler, analiz sonuçlarının doğruluğunu ve güvenilirliğini artırır (Provost & Fawcett, 2013).

2.2.2 Veri temizleme ve işleme

Toplanan ham veriler genellikle eksik, hatalı veya tutarsız olabilir. Veri temizleme süreci bu tür hataların düzeltilmesi ve verilerin analize hazır hale getirilmesini sağlar. Bu aşamada, eksik veriler doldurulur, tutarsız veriler düzeltilir ve veriler normalleştirilir. Bu adımlar, verilerin kalitesini artırarak daha doğru ve güvenilir analizler yapılmasını sağlar. Veri işleme süreci ayrıca, verilerin standart formatlara dönüştürülmesini ve analize uygun hale getirilmesini içerir (Dasu & Johnson, 2003). Veri temizleme ve işleme aşamaları, eğitim verilerinin daha iyi anlaşılmasını ve yorumlanmasını sağlar. Örneğin, öğrenci anketlerinden elde edilen verilerin doğru bir şekilde işlenmesi, öğrenci

başarılarını etkileyen faktörlerin daha net bir şekilde belirlenmesine yardımcı olabilir. Bu süreçler, veri bilimi projelerinin temelini oluşturur ve başarılı analizler için kritik öneme sahiptir.

2.2.3 Veri analizi

Veri analizi, toplanan ve temizlenen verilerin incelenmesi ve yorumlanması sürecidir. Bu aşamada, istatistiksel yöntemler ve veri madenciliği teknikleri kullanılarak verilerdeki desenler ve ilişkiler belirlenir. Veri analizi, işletmelerin stratejik kararlar almasına ve performanslarını iyileştirmelerine yardımcı olabilir. Eğitim alanında, veri analizi, öğrenci başarılarını etkileyen faktörleri belirlemek ve bu faktörlerin etkisini değerlendirmek için kullanılır (Han, Pei, & Kamber, 2011). Veri analizinde kullanılan yöntemler arasında korelasyon analizi, regresyon analizi ve sınıflandırma analizleri yer alır. Bu yöntemler, verilerin derinlemesine incelenmesini ve anlamlı sonuçların elde edilmesini sağlar. Eğitim verilerinin analizi, öğrenci performansını etkileyen anahtar faktörleri belirleyerek, eğitim politikalarının ve stratejilerinin geliştirilmesine katkıda bulunur.

2.2.4 Veri görselleştirme

Veri görselleştirme, analiz sonuçlarının grafikler, tablolar ve diğer görsel araçlar kullanılarak sunulmasıdır. Görselleştirme, karmaşık verilerin daha kolay anlaşılmasını sağlar ve karar vericilere daha etkili bir şekilde bilgi iletir. Eğitim verilerinin görselleştirilmesi, öğrenci başarılarını etkileyen faktörlerin daha net bir şekilde anlaşılmasına ve bu faktörlerin etkisinin görsel olarak ifade edilmesine yardımcı olur (Few, 2009). Veri görselleştirme araçları, eğitimcilerin ve politika yapıcıların verileri daha etkili bir şekilde analiz etmelerine olanak tanır. Grafikler, ısı haritaları ve dağılım diyagramları gibi görselleştirme teknikleri, verilerin eğilimlerini ve ilişkilerini daha iyi anlamamıza yardımcı olabilir. Bu, eğitim stratejilerinin ve politikalarının daha bilinçli bir şekilde tasarlanmasını sağlar.

2.2.5 Python ve kütüphaneleri

Veri bilimi çalışmalarında en yaygın kullanılan programlama dillerinden biri olan Python, geniş kütüphane desteğiyle veri analizinde güçlü bir araçtır. Python, veri toplama, temizleme, analiz etme ve görselleştirme gibi birçok aşamada etkin bir şekilde

kullanılabilir. Özellikle NumPy, Pandas, Matplotlib ve Seaborn gibi kütüphaneler veri analizi ve görselleştirme süreçlerinde sıklıkla kullanılır. Bu kütüphaneler, verilerin işlenmesini ve analiz edilmesini kolaylaştırır ve sonuçların görselleştirilmesini sağlar (McKinney, 2010). Python, eğitim verilerinin analizinde de yaygın olarak kullanılmaktadır. Örneğin, öğrenci notları ve demografik veriler gibi büyük veri setlerinin analizi, Pythonun sunduğu güçlü araçlar ve kütüphaneler sayesinde daha verimli bir şekilde gerçekleştirilebilir. Python, veri bilimi projelerinde esneklik ve hız sağladığı için tercih edilen bir programlama dilidir.

2.2.5.1 NumPy kütüphanesi

NumPy, Python'da bilimsel hesaplamalar yapmak için kullanılan temel bir kütüphanedir. Çok boyutlu diziler (arrays) ve matrislerle (matrices) çalışmayı sağlar ve bu diziler üzerinde matematiksel işlemler yapmayı kolaylaştırır. NumPy, lineer cebir, Fourier dönüşümleri ve rastgele sayı üretimi gibi çeşitli matematiksel işlemleri destekler. Bu özellikleri sayesinde, veri analizi ve makine öğrenmesi projelerinde sıkça kullanılır (Harris et al., 2020).

2.2.5.2 Pandas kütüphanesi

Pandas, veri manipülasyonu ve analizine yönelik güçlü bir kütüphanedir. Veri yapılarını ve veri analizi araçlarını sağlar, özellikle veri çerçeveleri (dataframes) ve seriler (series) ile çalışmayı kolaylaştırır. Pandas, veri temizleme, filtreleme, dönüştürme ve agregasyon işlemlerinde kullanılır. Eğitim verileri üzerinde yapılan analizlerde, öğrenci notları ve demografik bilgilerin düzenlenmesi ve işlenmesinde yaygın olarak kullanılır (McKinney, 2010).

2.2.5.3 Matplotlib kütüphanesi

Matplotlib, veri görselleştirme için kullanılan temel Python kütüphanelerinden biridir. Grafikler, çizimler ve diğer görsel gösterimler oluşturmak için kullanılır. Matplotlib, çizgi grafikleri, çubuk grafikleri, histogramlar, dağılım diyagramları ve daha birçok grafik türünü destekler. Eğitim verilerinin analizinde, öğrenci performansının görselleştirilmesi ve veri desenlerinin anlaşılması için sıkça kullanılır (Hunter, 2007).

1.2.5.4 Seaborn kütüphanesi

Seaborn, Matplotlib üzerine inşa edilmiş ve daha gelişmiş görselleştirme araçları sunan bir kütüphanedir. İstatistiksel grafikler ve görsel temsiller oluşturmak için kullanılır. Seaborn, estetik açıdan daha çekici ve bilgi açısından zengin görseller sunar. Eğitim verilerinin analizinde, veri dağılımlarını, ilişkilerini ve istatistiksel eğilimleri görselleştirmek için kullanılır. Özellikle korelasyon matrisleri ve ısı haritaları gibi grafikler oluşturmak için idealdir (Waskom, 2021).

2.3 Makine Öğrenmesi

Makine öğrenmesi, bilgisayarların veri kullanarak öğrenmelerini ve belirli görevleri yerine getirebilmelerini sağlayan bir yapay zekâ alt dalıdır. Bu süreçte, bilgisayarlar belirli algoritmalar ve istatistiksel modeller kullanarak verilerden öğrenir ve gelecekteki veriler üzerinde tahminler yapar. Makine öğrenmesi, denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme gibi farklı yöntemleri içerir. Makine öğrenmesi algoritmaları, eğitim verilerini analiz ederek öğrenci performanslarını tahmin etmede etkili bir şekilde kullanılmaktadır. Bu algoritmalar, öğrencilerin geçmiş performans verilerini ve demografik bilgilerini kullanarak gelecekteki başarılarını öngörebilir. Özellikle, regresyon ve sınıflandırma algoritmaları, öğrenci başarısını etkileyen anahtar faktörleri belirlemede yaygın olarak kullanılmaktadır.

2.3.1 Denetimli öğrenme (Supervised learning)

Denetimli öğrenme, giriş verileri ile bu verilere karşılık gelen doğru çıkışların bulunduğu veri setleri kullanılarak modelin eğitilmesi sürecidir. Bu yöntemde, model giriş ve çıkış verileri arasındaki ilişkiyi öğrenir ve yeni, görülmemiş veriler için doğru tahminler yapmayı öğrenir. Denetimli öğrenme, eğitim verilerinin analizinde yaygın olarak kullanılır ve öğrenci performansını tahmin etmede etkili sonuçlar sağlar (Bishop, 2006). Denetimli öğrenme algoritmaları arasında doğrusal regresyon, lojistik regresyon, destek vektör makineleri (SVM) ve karar ağaçları bulunur. Bu algoritmalar, eğitim verilerindeki desenleri ve ilişkileri öğrenerek, öğrenci başarılarını tahmin etmek için kullanılır. Örneğin, lojistik regresyon modeli, öğrencilerin geçmiş notlarını ve demografik bilgilerini kullanarak, gelecekteki başarılarını tahmin edebilir.

2.3.2 Denetimsiz öğrenme (Unsupervised learning)

Denetimsiz öğrenme, sadece giriş verilerinin bulunduğu veri setleri ile çalışır. Bu yöntemde, model verilerdeki gizli desenleri ve yapıları keşfeder. Denetimsiz öğrenme, özellikle kümeleme ve boyut indirgeme gibi görevler için kullanılır. Eğitim verilerinde, denetimsiz öğrenme, öğrencilerin farklı başarı düzeylerine göre gruplandırılması ve bu grupların özelliklerinin incelenmesi için kullanılabilir (Hastie, Tibshirani, & Friedman, 2009). Denetimsiz öğrenme algoritmaları arasında k-ortalama (k-means) kümeleme, hiyerarşik kümeleme ve ana bileşen analizi (PCA) bulunur. Bu algoritmalar, verilerdeki gizli desenleri ve grupları keşfederek, öğrenci başarılarını etkileyen faktörlerin daha iyi anlaşılmasını sağlar. Denetimsiz öğrenme, eğitim verilerinin derinlemesine incelenmesi ve analiz edilmesi için önemli bir araçtır.

2.3.3 Pekiştirmeli öğrenme (Reinforcement learning)

Pekiştirmeli öğrenme, bir ajanın belirli bir ortamda, ödül veya ceza yoluyla öğrenmesini sağlayan bir makine öğrenmesi yöntemidir. Ajan, çevresindeki durumu algılar, eylemler gerçekleştirir ve bu eylemlerin sonuçlarına göre gelecekteki eylemlerini optimize eder. Eğitim verilerinde pekiştirmeli öğrenme, öğrenci davranışlarını ve öğrenme stratejilerini modellemek için kullanılabilir. Bu yöntem, öğrencilerin öğrenme süreçlerinde daha iyi kararlar almalarını sağlayacak stratejiler geliştirmeye yardımcı olabilir (Sutton & Barto, 2018). Pekiştirmeli öğrenme algoritmaları arasında Q-öğrenme (Q-learning) ve derin Q-ağları (deep Q-networks) bulunur. Bu algoritmalar, eğitim verilerini kullanarak öğrenci performansını artırmak için optimize edilebilir. Örneğin, bir öğrenme yönetim sistemi, öğrencilerin öğrenme materyalleriyle nasıl etkileşime girdiklerini analiz ederek, onların en etkili öğrenme yollarını belirleyebilir ve bu yolları teşvik edebilir. Pekiştirmeli öğrenme (Reinforcement Learning) bağlamında "ajan" terimi, öğrenme sürecinde aktif olan ve çevresiyle etkileşimde bulunan varlık veya model anlamına gelir. Ajan, çevresindeki durumu algılar (durum bilgisi), belirli eylemler gerçekleştirir ve bu eylemlerin sonuçlarına göre ödül veya ceza alır. Bu ödül veya ceza, ajanın gelecekteki eylemlerini optimize etmesi ve öğrenmesi için kullanılır. Pekiştirmeli öğrenmede amaç, ajanın ödülleri maksimize edecek şekilde öğrenmesini sağlamaktır. Ajan, eylemlerinin uzun vadeli sonuçlarını değerlendirerek en iyi stratejileri geliştirir ve uygular.

2.3.4 Derin öğrenme (Deep learning)

Derin öğrenme, çok katmanlı yapay sinir ağları kullanarak verilerden öğrenen bir makine öğrenmesi dalıdır. Derin öğrenme, özellikle büyük veri setleri üzerinde çalışırken yüksek performans gösterir ve görüntü tanıma, doğal dil işleme gibi alanlarda etkili sonuçlar üretir. Eğitim verilerinde derin öğrenme, öğrenci performansını tahmin etmek ve kişiselleştirilmiş eğitim stratejileri geliştirmek için kullanılabilir (Goodfellow, Bengio, & Courville, 2016). Derin öğrenme algoritmaları arasında yapay sinir ağları (artificial neural networks - ANN), evrişimli sinir ağları (convolutional neural networks - CNN) ve tekrarlayan sinir ağları (recurrent neural networks - RNN) bulunur. Bu algoritmalar, öğrenci verilerinin derinlemesine analiz edilmesini ve öğrenme süreçlerinin daha etkili bir şekilde modellenmesini sağlar. Derin öğrenme, eğitimde kişiselleştirilmiş öğrenme deneyimleri sunarak, her öğrencinin bireysel ihtiyaçlarına uygun eğitim stratejileri geliştirilmesine yardımcı olur.

2.3.5 Model değerlendirme ve validasyon

Makine öğrenmesi modellerinin performansını değerlendirmek ve doğrulamak için çeşitli teknikler kullanılır. Bu aşama, modelin genel performansını ve genelleme yeteneğini belirlemeye yardımcı olur. Model değerlendirme, eğitim verilerinin analizi sırasında modelin doğruluğunu ve güvenilirliğini sağlamada kritik bir rol oynar. Bu süreç, modellerin eğitim ve test verileri üzerinde nasıl performans gösterdiğini belirlemek için kullanılır (Kuhn & Johnson, 2013). Örnek teknikler arasında çapraz doğrulama (cross-validation), doğruluk (accuracy), hata matrisi (confusion matrix) ve ROC eğrisi (ROC curve) bulunur. Çapraz doğrulama, veriyi farklı alt kümelere ayırarak modelin doğruluğunu test eder. Doğruluk, modelin ne kadar doğru tahminlerde bulunduğunu gösterir. Hata matrisi, modelin sınıflandırma performansını ölçmek için kullanılır ve doğru pozitifler, yanlış pozitifler, doğru negatifler ve yanlış negatifleri içerir. ROC eğrisi ise modelin sınıflandırma performansını grafiksel olarak gösterir (Fawcett, 2006).

2.3.6 Regresyon modelleri

Regresyon modelleri, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi modellemek için kullanılan makine öğrenmesi teknikleridir. Eğitim verilerinde regresyon analizi, öğrenci performansını etkileyen faktörlerin belirlenmesi ve bu faktörlerin öğrenci

notlarına olan etkisinin tahmin edilmesi için kullanılır. Doğrusal regresyon, çoklu doğrusal regresyon, Ridge ve Lasso regresyon gibi modeller, eğitim verilerinde yaygın olarak kullanılır (James et al., 2013). Doğrusal regresyon, bağımlı değişken ile bir veya birden fazla bağımsız değişken arasındaki doğrusal ilişkiyi modelleyen basit bir tekniktir. Çoklu doğrusal regresyon, birden fazla bağımsız değişkenin etkisini modellemek için kullanılır. Ridge ve Lasso regresyon modelleri, çoklu doğrusal regresyonun bir uzantısıdır ve aşırı uyumu (overfitting) önlemek için düzenleme (regularization) teknikleri kullanılır. Bu modeller, eğitim verilerindeki anahtar faktörlerin belirlenmesine ve öğrenci başarılarının tahmin edilmesine yardımcı olur.

2.3.7 Sınıflandırma modelleri

Sınıflandırma modelleri, verileri önceden tanımlanmış sınıflara ayırmak için kullanılan makine öğrenmesi teknikleridir. Eğitim verilerinde sınıflandırma, öğrencilerin başarı durumlarını tahmin etmek ve belirli gruplara ayırmak için kullanılır. Lojistik regresyon, karar ağaçları, destek vektör makineleri (SVM) ve Random Forest gibi modeller, eğitim verilerinde yaygın olarak kullanılan sınıflandırma algoritmalarıdır (Hastie, Tibshirani, & Friedman, 2009). Lojistik regresyon, bağımlı değişkenin kategorik olduğu durumlarda kullanılır ve verileri iki veya daha fazla sınıfa ayırmak için uygun bir yöntemdir. Karar ağaçları, verileri dallara ayırarak sınıflandırma yapan basit ve etkili bir tekniktir. Destek vektör makineleri (SVM), verileri sınıflandırmak için hiper düzlemler kullanarak yüksek doğruluk sağlar. Random Forest ise, birden fazla karar ağacını birleştirerek daha güçlü ve güvenilir sınıflandırma sonuçları elde eden bir tekniktir.

2.4 Eğitimde Veri Bilimi ve Makine Öğrenmesi Yöntemleri

Veri bilimi ve makine öğrenmesi, eğitim alanında giderek daha fazla kullanılmakta olup, öğrenci performansını analiz etme ve tahmin etme süreçlerinde kritik bir rol oynamaktadır. Bu bölümde, öğrenci başarısını tahmin etmek için kullanılan yöntemler ve bu alandaki örnek çalışmalar ele alınacaktır.

2.4.1 Öğrenci başarısını tahmin etmek için kullanılan yöntemler

Makine öğrenmesi ve veri bilimi, eğitim verilerini analiz ederek öğrenci başarısını tahmin etmek için çeşitli algoritmalar ve teknikler sunar. Bu yöntemler, öğrencilerin

demografik bilgileri, geçmiş akademik performansları ve diğer ilgili verileri kullanarak gelecekteki başarılarını öngörmeyi amaçlar.

Regresyon analizi, sürekli değişkenler arasındaki ilişkiyi modellemek için kullanılır ve öğrenci notları gibi sürekli verilerin tahmin edilmesinde yaygın olarak kullanılır. Örneğin, Ridge ve Lasso regresyon modelleri, öğrenci başarılarını tahmin etmek için kullanılabilir ve bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini analiz eder (Yang et al., 2023). Ridge regresyonu, multikolinearite durumlarında değişkenlerin etkisini stabilize ederek daha güvenilir sonuçlar sağlar. Lasso regresyonu ise, bazı değişkenlerin etkisini tamamen sıfırlayarak modeldeki en önemli değişkenleri belirler ve basit modeller oluşturur. Bu çalışmada, Ridge ve Lasso regresyon modellerini kullanarak öğrencilerin not ortalamalarını tahmin ettik ve bağımsız değişkenlerin önem derecelerini belirledik. Analizlerimizde, aile ekonomik durumu, kişisel özellikler ve okul çevresi gibi faktörlerin öğrenci başarıları üzerinde önemli etkileri olduğunu tespit ettik.

Sınıflandırma algoritmaları, verileri belirli kategorilere ayırarak analiz eder. Logistic regression, decision trees, random forests ve support vector machines (SVM) gibi algoritmalar, öğrencilerin başarılı veya başarısız olma olasılığını tahmin etmek için kullanılır. Bu algoritmalar, özellikle öğrenci performansının sınıflandırılması ve tahmin edilmesinde etkilidir (Yang et al., 2023). Decision trees, her bir karar noktasında veriyi bölerek sınıflandırmayı gerçekleştirir. Random forests ise, birden fazla karar ağacı kullanarak tahmin doğruluğunu artırır. SVM, veriyi sınıflandırmak için en iyi ayırım çizgisini bulur ve bu çizgi etrafındaki marjini maksimize eder. Bizim çalışmamızda, logistic regression, random forests ve SVM algoritmalarını kullanarak öğrenci başarılarını sınıflandırdık. Logistic regression modelinde doğruluk oranı %60, random forests modelinde %63 ve SVM modelinde %59 olarak bulunmuştur. Bu sonuçlar, sınıflandırma algoritmalarının öğrenci başarılarını tahmin etmede etkili olduğunu göstermektedir.

KNN algoritması, verileri benzerliklerine göre sınıflandırarak çalışır. Eğitimde, öğrencilerin başarılarını tahmin etmek için kullanılabilir. Bu algoritma, yeni bir öğrencinin başarı seviyesini, geçmişteki benzer öğrenci verileri ile karşılaştırarak tahmin eder (Han, Pei, & Kamber, 2011). KNN algoritması, özellikle küçük veri setlerinde ve karmaşık olmayan problem çözümlerinde etkili sonuçlar verir. Bu çalışmada, KNN algoritması kullanılarak öğrenci başarıları tahmin edilmiş ve algoritmanın doğruluğu analiz edilmiştir.

Naive Bayes algoritması, sınıflandırma problemlerinde kullanılan bir başka etkili yöntemidir. Eğitim verilerini analiz ederek, öğrencilerin başarı durumlarını tahmin et-

mek için kullanılabilir. Bu yöntem, özellikle büyük veri setlerinde hızlı ve etkili sonuçlar üretir (Al-Barrak & Al-Razgan, 2016, p. 235). Naive Bayes, her bir özelliğin bağımsız olduğunu varsayarak hesaplamaları basitleştirir ve bu sayede büyük veri setlerinde bile hızlı tahminler yapar. Bu çalışmada, Naive Bayes algoritması kullanılarak öğrenci başarıları sınıflandırılmış ve algoritmanın doğruluğu değerlendirilmiştir.

2.5 Veri Bilimi ve Makine Öğrenmesi ile İlgili Örnek Çalışmalar

Eğitim alanında veri bilimi ve makine öğrenmesi uygulamaları üzerine yapılan araştırmalar, bu tekniklerin öğrenci başarılarını tahmin etmekte nasıl kullanıldığını göstermektedir. Bu alandaki örnek çalışmalardan bazıları verilmiştir.

Örnek çalışma 1: Öğrenci başarısının tahmin edilmesi

Bu çalışma, veri madenciliği ve makine öğrenmesi tekniklerini kullanarak öğrenci başarılarını tahmin etmeye odaklanmıştır. Çalışmada, öğrenci verileri analiz edilerek, başarılarını etkileyen anahtar faktörler belirlenmiş ve bu faktörler kullanılarak tahmin modelleri geliştirilmiştir. Sonuçlar, bu modellerin öğrenci başarılarını yüksek doğrulukla tahmin edebildiğini göstermektedir (Alyahyan & Düşteğör, 2020). Araştırmada, veri seti olarak bir okulun öğrenci notları ve demografik bilgileri kullanılmıştır. Ridge ve Lasso regresyon modelleri, öğrenci notlarını tahmin etmekte başarılı olmuş ve anahtar faktörlerin belirlenmesinde önemli bir rol oynamıştır. Özellikle, öğrencilerin ders çalışma alışkanlıkları ve aile desteği gibi faktörlerin başarı üzerinde önemli etkileri olduğu tespit edilmiştir. Bizim çalışmamızda da benzer şekilde, Ridge ve Lasso regresyon modelleri kullanılarak öğrenci başarılarını tahmin ettik ve önemli faktörleri belirledik. Bu faktörlerin öğrenci başarıları üzerindeki etkilerini analiz ettik ve sonuçlar literatürdeki bulgularla uyumlu olmuştur.

Örnek çalışma 2

Makine Öğrenmesi ile Akademik Performans Tahmini Başka bir çalışmada, makine öğrenmesi algoritmaları kullanılarak öğrencilerin akademik performansları tahmin edilmiştir. Logistic regression, random forests ve SVM gibi çeşitli algoritmalar kullanılarak yapılan analizlerde, öğrencilerin geçmiş akademik performansları ve demografik bilgileri değerlendirilmiştir. Çalışmanın sonuçları, makine öğrenmesi algoritmalarının öğrenci performansını etkili bir şekilde tahmin edebildiğini göstermiştir (Yang et al.,

2023). Bu çalışmada, özellikle random forests algoritması, öğrenci performansını sınıflandırmada yüksek doğruluk oranları elde etmiştir. Öğrenci başarılarını etkileyen en önemli faktörler arasında derslere düzenli katılım ve sınav performansları yer almıştır. Bizim çalışmamızda da logistic regression, random forests ve SVM algoritmaları kullanılarak öğrenci başarıları tahmin edilmiştir. Logistic regression modelinde doğruluk oranı %60, random forests modelinde %63 ve SVM modelinde %59 olarak bulunmuştur. Bu sonuçlar, literatürdeki bulgularla uyumlu olup, sınıflandırma algoritmalarının öğrenci başarılarını tahmin etmede etkili olduğunu göstermektedir.

Örnek Çalışma 3: Veri bilimi ve öğrenci performansı analizi

Bu çalışmada, veri bilimi teknikleri kullanılarak öğrenci performansı analiz edilmiştir. Büyük veri setleri üzerinde yapılan analizler, öğrencilerin başarılarını etkileyen faktörleri belirlemiş ve bu faktörlere dayalı olarak tahmin modelleri oluşturulmuştur. Çalışma, veri bilimi ve makine öğrenmesi yöntemlerinin eğitimde nasıl kullanılabileceğini ve bu yöntemlerin öğrenci başarılarını artırmak için nasıl uygulanabileceğini göstermektedir (Han, Pei, & Kamber, 2011). Araştırma kapsamında, öğrencilerin sınav sonuçları, ödev performansları ve ders katılımı gibi veriler kullanılmıştır. Bu veriler üzerinde yapılan analizlerde, destek vektör makineleri (SVM) ve karar ağaçları (decision trees) gibi algoritmalar kullanılarak başarılı tahmin modelleri geliştirilmiştir. Özellikle, öğrenci başarısını artırmak için kişiselleştirilmiş öğrenme yollarının belirlenmesinde veri bilimi tekniklerinin etkinliği vurgulanmıştır. Bizim çalışmamızda da benzer şekilde, SVM ve karar ağaçları algoritmaları kullanılarak öğrenci başarıları tahmin edilmiştir. Bu algoritmaların doğruluğu analiz edilmiş ve öğrenci başarısını etkileyen en önemli faktörler belirlenmiştir. Elde edilen sonuçlar literatürdeki bulgularla uyumlu olup, veri bilimi ve makine öğrenmesi yöntemlerinin eğitimde etkili bir şekilde kullanılabileceğini göstermektedir.

Bu örnek çalışmalardan elde edilen sonuçlar, veri bilimi ve makine öğrenmesi yöntemlerinin eğitimde nasıl kullanılabileceği konusunda önemli bilgiler sunmaktadır. Bu yöntemler, öğrenci performansını tahmin etmekle kalmayıp, aynı zamanda eğitim stratejilerinin geliştirilmesine de katkı sağlamaktadır. Eğitimciler ve eğitim politikası yapımcıları, bu teknikleri kullanarak daha etkili ve verimli eğitim programları oluşturabilirler.

3. MATERYAL VE YÖNTEM

3.1. Araştırma Verileri ve Veri Toplama Yöntemleri

Bu bölümde, ortaokul düzeyindeki öğrenci başarısını etkileyen faktörlerin analizi için kullanılan veri toplama yöntemleri ve araştırma verileri ele alınacaktır. Araştırma, veri toplama süreci, anket yöntemleri, veri setinin hazırlanması ve analiz yöntemleri üzerinde durulacaktır.

3.1.1. Anket ve veri toplama süreci

Anketlerin hazırlanması ve uygulanması sürecinde şu adımlar izlenmiştir:

Anket tasarımı

Öğrenci başarısını etkileyen faktörlerin belirlenmesi amacıyla hazırlanan anket, öğrencilere yönelik olarak Google Form kullanılarak oluşturulmuştur. Anket, öğrencilerin akademik başarılarını ve bu başarıyı etkileyen çeşitli faktörleri analiz etmek için tasarlanmıştır. Sorular, öğrencilerin demografik bilgilerini, aile durumlarını, kişisel özelliklerini, okul ve çevre faktörlerini ve öğretmen etkisini değerlendirmeye yönelik olarak hazırlanmıştır. Sorulara verilen yanıtlar "Evet" veya "Hayır" şeklinde belirlenmiştir. Aşağıdaki tabloda anket soruları yer almaktadır.

Tablo 3.1 Öğrenci anket soruları

SORULAR	YANITLAR
Sınıfınız	5-6-7-8
Şube	A-B-C-D-E-F
Adınız- Soyadınız	
Cinsiyet	KIZ/ERKEK
1) Ailemin Ekonomik durumu iyi denilebilir	EVET/HAYIR
2) Ailem bana ders çalışmam için baskı yapmaz	
3) Bana ait bir çalışma odam var	
4) Annem ve Babam ile birlikte yaşıyorum	
5) Ailem genellikle bana sevildiğimi hissettirir	
6) Okulda istediğim gibi beslenmemi yapılabiliyorum	
7) Gün içinde 2 saatten daha az TV-Bilgisayar-Telefon ile meşgul olurum.	
8) Genellikle günde yarım saatten fazla kitap okurum	
9) Ders dışındaki konulara daha az ilgi gösteririm	
10) Genellikle Derse odaklanmakta zorlanmam.	
11) Derste öğretmenlerimin anlattıklarını anlamakta zorlanmam	

12) Ders çalıştığım zaman başarılı olabiliyorum.	
13) Sınavlarda çok heyecan yapmam.	
14) Ödevlerimi genellikle zamanında yaparım	
15) Çalışmalarım yeterince takdir edilir.	
16) Sorunlarımı konuşmaya, anlatmaya ve çözüm bulmaya ihtiyacım yok.	
17) Etkili ders çalışma yöntemlerini biliyorum.	
18) Herhangi bir sağlık sorunum yok.	
19) Çekingen, Utangaç bir yapım yok	
20) Başarılı olacağımı biliyorum kendime güveniyorum	
21) Okulda genellikle beni rahatsız eden öğrenciler yoktur.	
22) Sınıfımızın kalabalık olduğunu düşünmüyorum	
23) Okulumu ve arkadaşlarımı yeterince seviyorum	
24) Anlayamadığım konularda öğretmenime soru sormaktan çekinmem	
25) Öğretmenlerimiz genellikle dersi eğlenceli bir şekilde işler	
26) Dersle ilgili etkinlikler yeterince yapılır	

Tablo 3.1 Öğrenci anket soruları “Tablonun devamı”

Pilot çalışma

Anketin doğruluğunu ve anlaşılabilirliğini test etmek amacıyla küçük bir pilot çalışma gerçekleştirilmiştir. Bu çalışma, anketin hedef kitlesi olan öğrencilerle yapılmış ve anket sorularının anlaşılır olup olmadığı, herhangi bir karışıklık olup olmadığı değerlendirilmiştir. Pilot çalışma sonucunda, anket sorularında gerekli düzenlemeler yapılarak, daha net ve anlaşılır hale getirilmiştir.

Veri toplama

Düzenlenmiş anket, Türkiye Cumhuriyeti Devleti Batman ili sınırları içerisindeki ortaokul kurumlarında öğrenim gören öğrencilere dağıtılmıştır. Öğrencilerden, anketi yalnızca kendilerinin doldurmaları istenmiştir. Anket, farklı okullarda öğrenim gören 576 öğrenci tarafından tamamen doldurulmuştur. Anketten elde edilen veriler, öğrencilerin isimleri, sınıfları ve anket sorularına verdikleri yanıtlardan oluşmaktadır.

3.1.2. Veri setinin hazırlanması ve analiz için kullanılan yöntemler

Toplanan veriler, analiz için hazırlanmıştır. Bu süreçte izlenen adımlar şunlardır:

Veri temizleme

Anketlerden elde edilen ham veriler, eksik veya hatalı yanıtlar açısından incelenmiş ve bu tür veriler temizlenmiştir. Eksik veriler, ortalama veya medyan değerler kullanılarak tamamlanmıştır.

Veri kodlama

Nitel veriler, analiz edilebilir hale getirilmek için sayısal değerlere dönüştürülmüştür. Örneğin, "Ailemin ekonomik durumu iyi" gibi ifadeler, 1 veya 0 olarak kodlanmıştır. "Evet" olarak verilen yanıtlar "1", "Hayır" olarak verilen yanıtlar "0" olarak kodlanmıştır.

Veri normalizasyonu

Toplanan veriler tablolama programı olan excel kullanılarak ".csv" uzantılı olarak kaydedilmiş, öğrencilerin 2023/2024 eğitim öğretim yılı 1. dönem not ortalamaları ile birleştirilmiştir. İlgili okullardan temin edilen not ortalamaları, veri setine eklenmiş ve öğrencilerin isimleri kişisel verilerin gizliliği ilkesine uygun olarak veri setinden çıkarılmıştır. Böylece isimlerin yerine öğrencilerin notları ve anket sorularına verdikleri yanıtların bulunduğu anonim bir veri seti oluşturulmuştur. Veri setindeki "Evet" yanıtları "1", "Hayır" yanıtları ise "0" olarak kodlanmış ve soru başlıkları daha kısa ve anlaşılır hale getirilmiştir. Örneğin, "Ailemin Ekonomik durumu iyi denilebilir" sorusu "A1" olarak kodlanmıştır. Diğer tüm soru sütunları da benzer şekilde kodlanmıştır. Bu işlemler sonucunda, "Not" sütununa göre cevaplandırılmış, anonim hale getirilmiş ve analiz için uygun bir veri seti elde edilmiştir. Bu veri seti, öğrenci başarılarını etkileyen faktörlerin analiz edilmesinde kullanılmıştır. Verilerin analiz edilebilmesi için normalizasyon işlemi uygulanmıştır. Bu işlemler, verilerin aynı ölçek üzerinde daha hızlı ve etkin değerlendirilmesine olanak sağlamıştır.

3.2. Veri Analiz Yöntemleri

Bu bölümde, araştırmada kullanılan veri analiz yöntemleri detaylandırılacaktır. Kullanılan analiz yöntemleri arasında korelasyon analizi, regresyon analizi ve sınıflandırma analizleri yer almaktadır.

3.2.1. Korelasyon analizi

Korelasyon analizi, bağımsız değişkenler ile bağımlı değişken arasındaki ilişkilerin incelenmesini sağlar. Bu analizde, Pearson korelasyon katsayısı kullanılarak değişkenler arasındaki doğrusal ilişkiler değerlendirilmiştir.

3.2.2. Regresyon analizi

Regresyon analizi, bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini belirlemek için kullanılmıştır. Bu analizde, Ridge ve Lasso regresyon modelleri kullanılarak öğrenci başarıları tahmin edilmiştir. Ridge ve Lasso regresyon katsayıları bulunarak hangi değişkenin öğrenci notu üzerinde daha çok etkisi olduğu saptanmıştır.

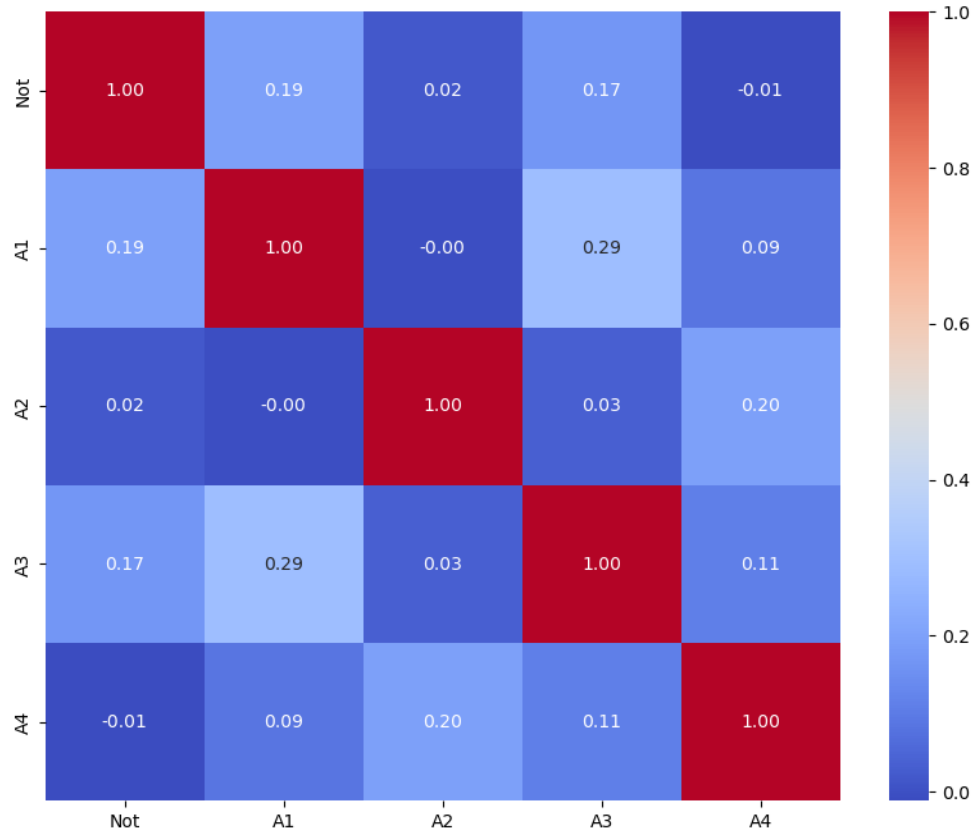
3.2.3. Sınıflandırma analizi

Sınıflandırma analizleri, öğrencilerin başarı durumlarını sınıflandırmak için kullanılmıştır. Bu analizlerde, logistic regression, random forests ve support vector machines (SVM) gibi algoritmalar kullanılmıştır.

4. BULGULAR VE TARTIŞMA

4.1 Korelasyon Analizi Sonuçları

Aile durumlarının öğrenci notları üzerindeki etkisini anlamak amacıyla yapılan korelasyon analizi sonuçları aşağıda sunulmaktadır:



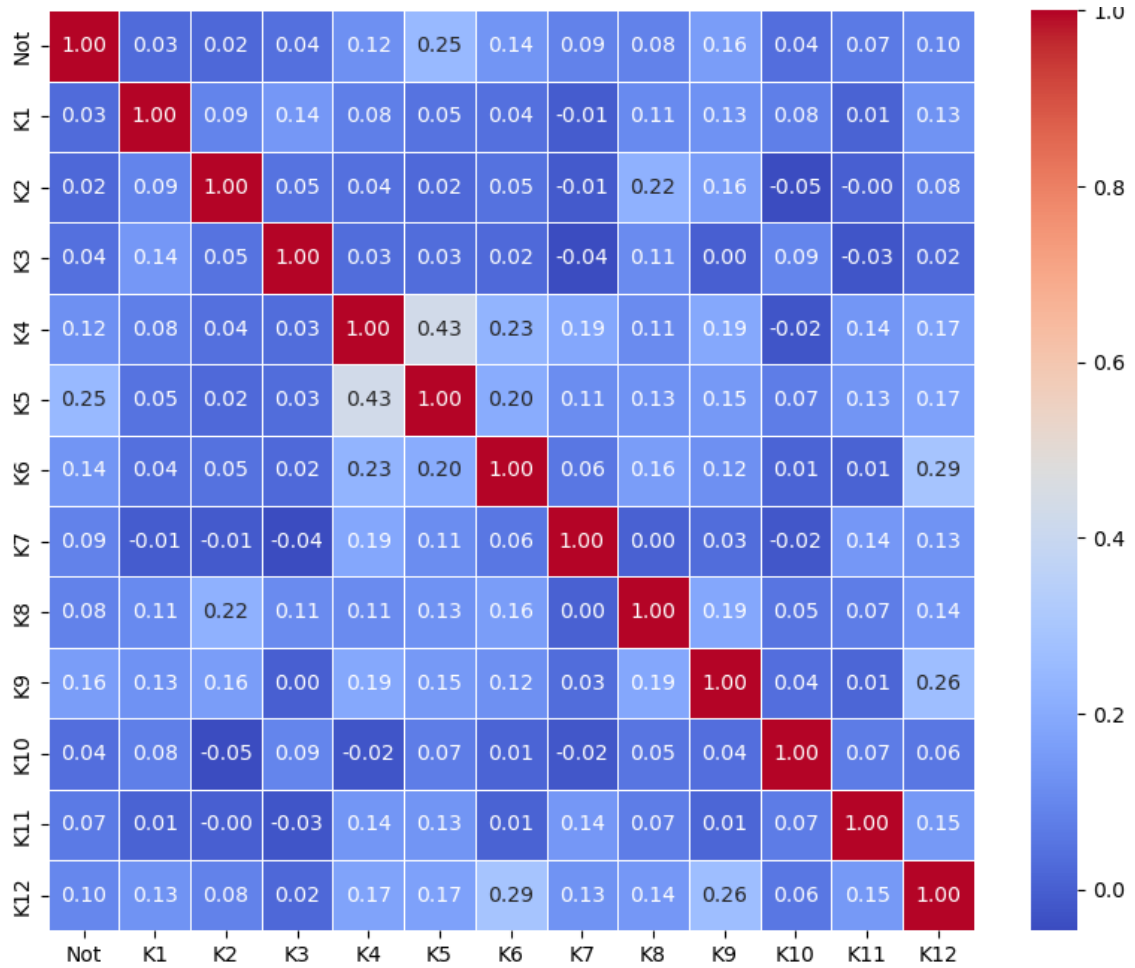
Şekil 4.1. Aile durumları ile Not Arasındaki korelasyon matrisi

- **Not ile A1:** Pozitif bir korelasyon (0.19) bulunmaktadır, bu da A1'in artmasıyla notların arttığını gösterir.
- **Not ile A2:** Çok zayıf bir pozitif korelasyon (0.02) vardır, etkisi minimaldir.
- **Not ile A3:** Orta düzeyde pozitif bir korelasyon (0.17) bulunmaktadır.
- **Not ile A4:** Negatif bir korelasyon (-0.01) bulunmaktadır, ancak bu çok zayıftır.

Bu sonuçlar, aile durumlarının öğrenci notlarını etkileyebileceğini ancak bu etkinin değişkenlik gösterebileceğini göstermektedir.

Tartışma

Bu sonuçlar, ailelerin sosyoekonomik durumunun ve eğitim seviyesinin, çocuklarının akademik başarıları üzerinde önemli bir etkisi olduğunu göstermektedir. Ailelerin eğitim düzeyi ve ekonomik imkanları öğrencilere daha iyi eğitim kaynakları ve destek sağlayarak, başarılarını artırabilir (Han, Pei, & Kamber, 2011). Kişisel özelliklerin öğrenci notlarına etkisi aşağıdaki korelasyon matrisi ile gösterilmektedir:



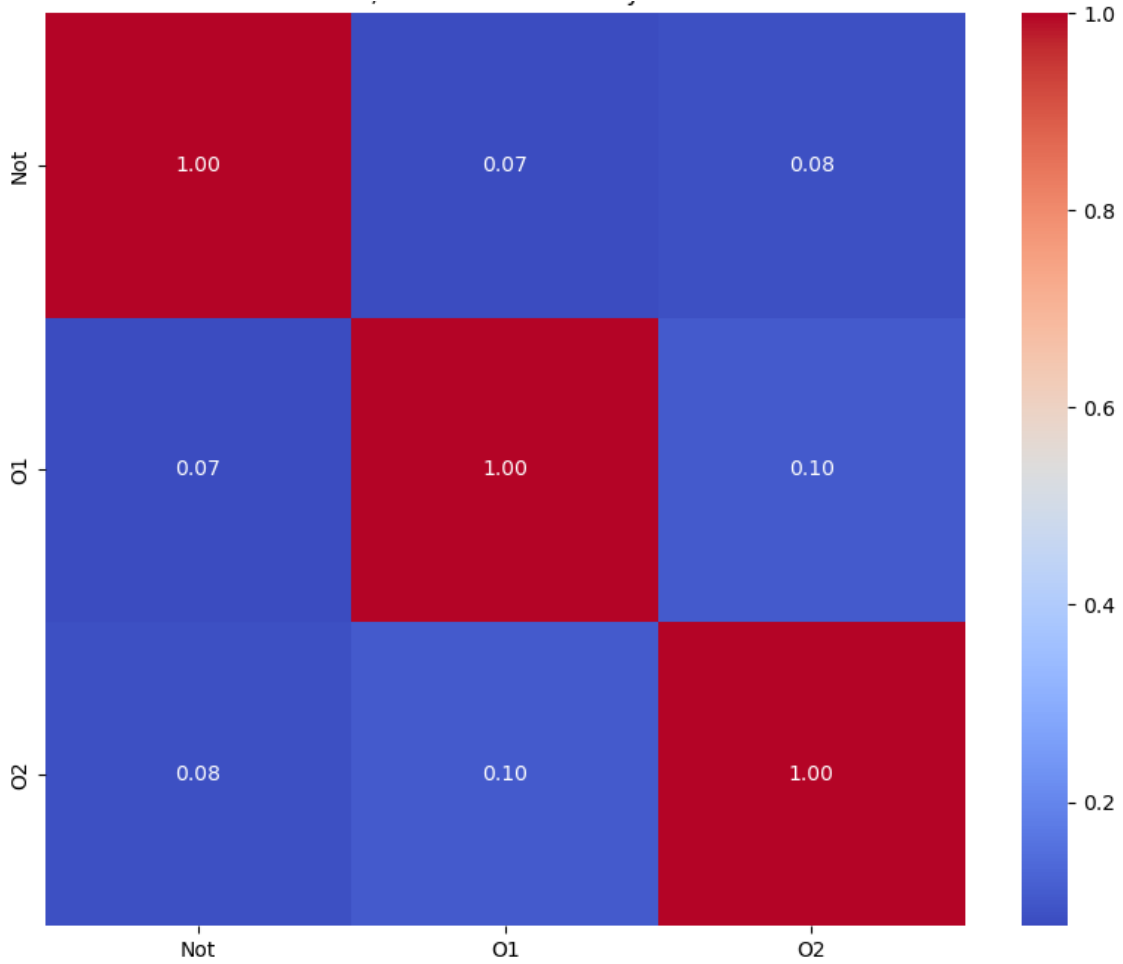
Şekil 4.2 Kişisel özellikler ve Not arasındaki korelasyon

Not ile K1-K12 korelasyonu

Bu korelasyonlar genellikle zayıf (0.01 ile 0.26 arasında değişmektedir). K4 ve K5 ile daha güçlü pozitif korelasyonlar (0.25 ve 0.43) bulunmaktadır. Kişisel özelliklerin, özellikle bazı değişkenlerin (K4 ve K5 gibi), öğrenci notları üzerinde belirgin bir etkisi olabilir.

Tartışma

Bu bulgular, öğrencilerin kişisel özelliklerinin ve motivasyon düzeylerinin, akademik başarıları üzerinde belirleyici olduğunu göstermektedir. Öz-düzenleme becerilerine sahip öğrenciler, öğrenme süreçlerinde daha başarılı olabilirler.



Şekil 4.3 O1 ve O2 ile Not arasındaki korelasyon matrisi

O1 ve O2 değişkenlerinin öğrenci notlarına etkisi aşağıdaki gibidir:

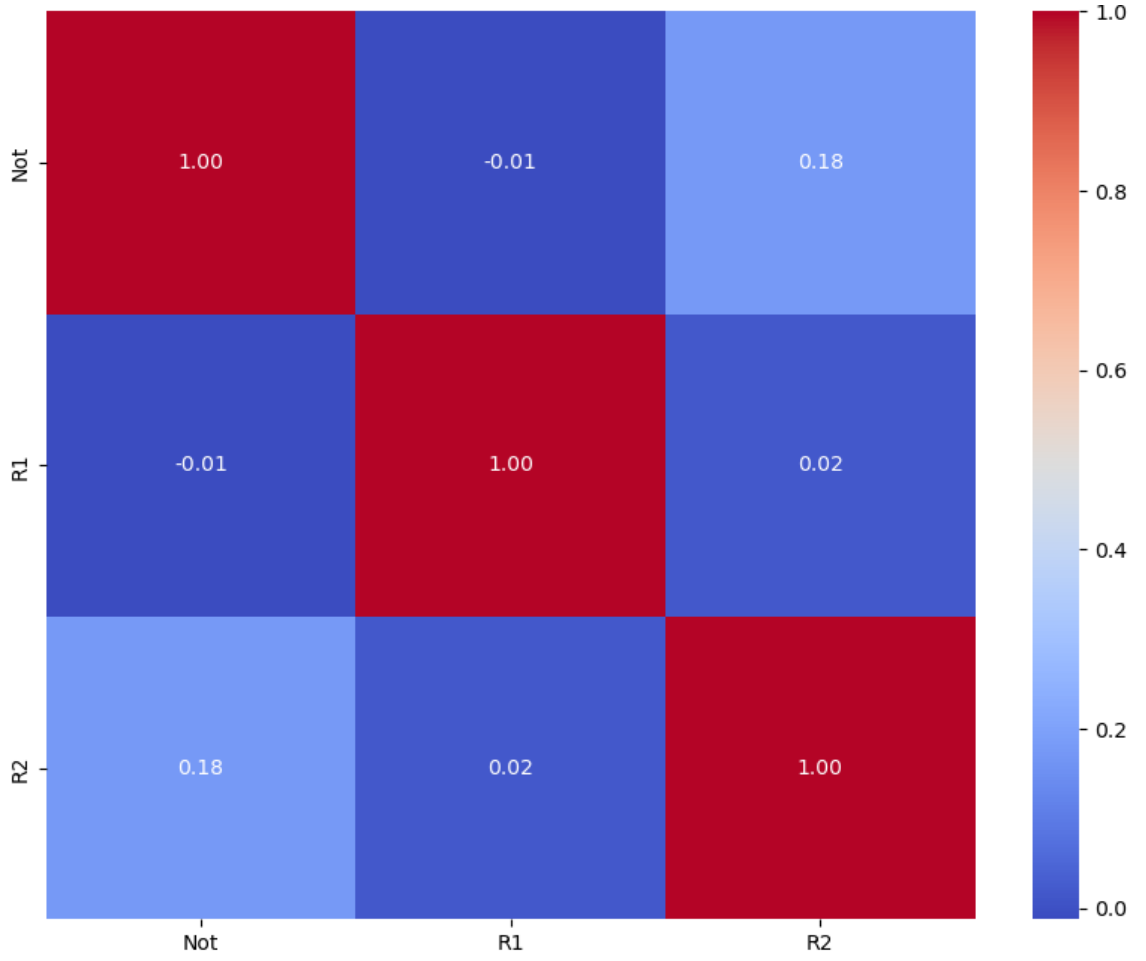
- **Not ile O1:** Çok zayıf bir pozitif korelasyon (0.07) vardır.
- **Not ile O2:** Zayıf bir pozitif korelasyon (0.08) bulunmaktadır.

Bu değişkenlerin öğrenci notları üzerinde minimal bir etkisi olduğu görülmektedir.

Tartışma

Bu bulgular, okul olanaklarının ve çevresinin, öğrencilerin akademik başarıları üzerinde önemli etkileri olduğunu göstermektedir. Daha iyi okul olanakları ve düşük sı-

nif mevcudu, öğrencilere daha iyi öğrenme ortamları sunarak başarılarını arttırabilir (Yang et al., 2023).

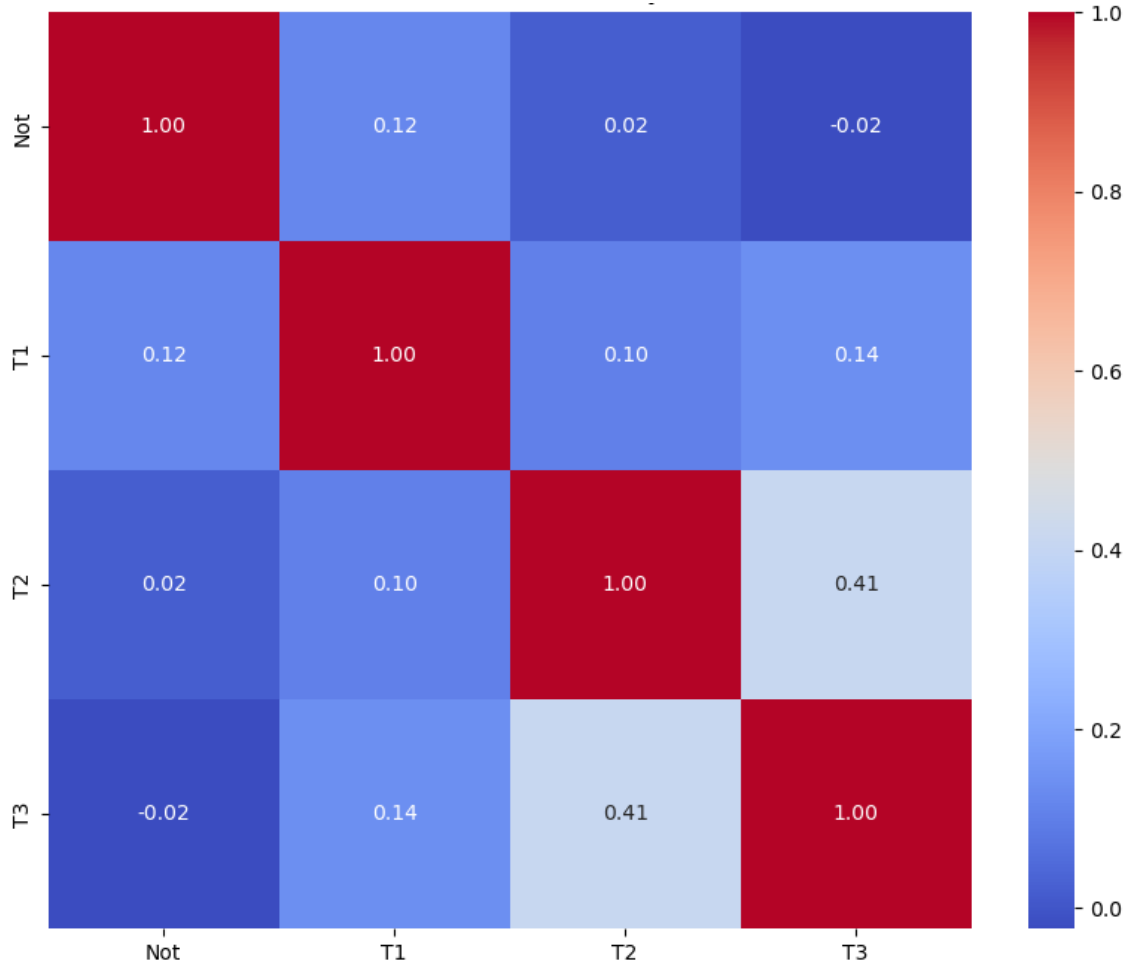


Şekil 4.4 R1 ve R2 ile not arasındaki korelasyon matrisi

R1 ve R2 değişkenlerinin öğrenci notlarına etkisi aşağıdaki gibidir:

- **Not ile R1:** Negatif bir korelasyon (-0.01) bulunmaktadır.
- **Not ile R2:** Pozitif bir korelasyon (0.18) bulunmaktadır.

R2'nin öğrenci notları üzerinde pozitif bir etkisi olabilirken, R1'in etkisi minimal ve negatiftir.

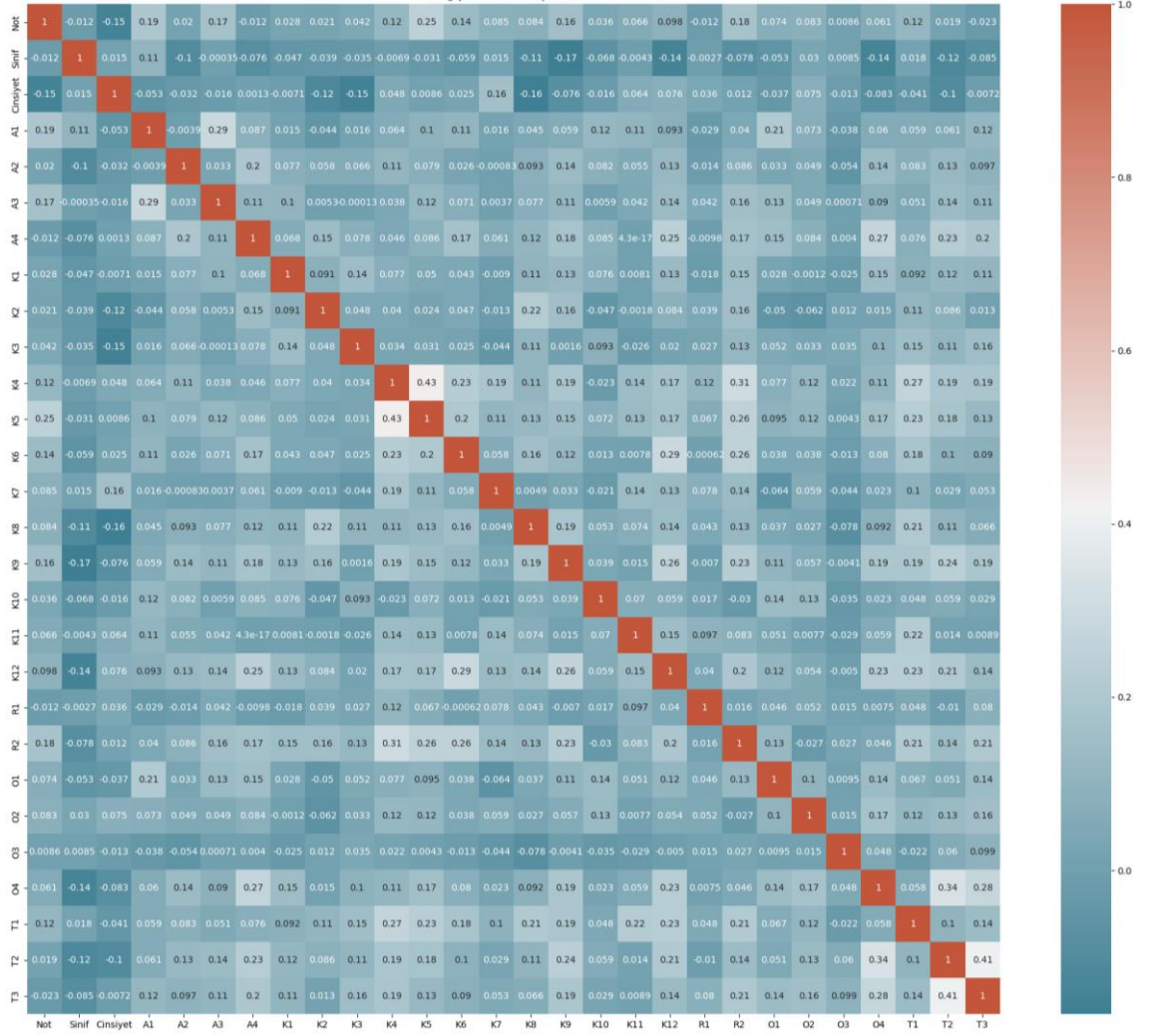


Şekil 4.5 T1, T2 ve T3 ile Not arasındaki korelasyon matrisi

T1, T2 ve T3 değişkenlerinin öğrenci notlarına etkisi aşağıdaki gibidir:

- **Not ile T1:** Pozitif bir korelasyon (0.12).
- **Not ile T2:** Çok zayıf bir pozitif korelasyon (0.02).
- **Not ile T3:** Negatif bir korelasyon (-0.02).

T1'in öğrenci notları üzerinde bir miktar pozitif etkisi olduğu, T2 ve T3'ün etkisinin minimal olduğu görülmektedir.

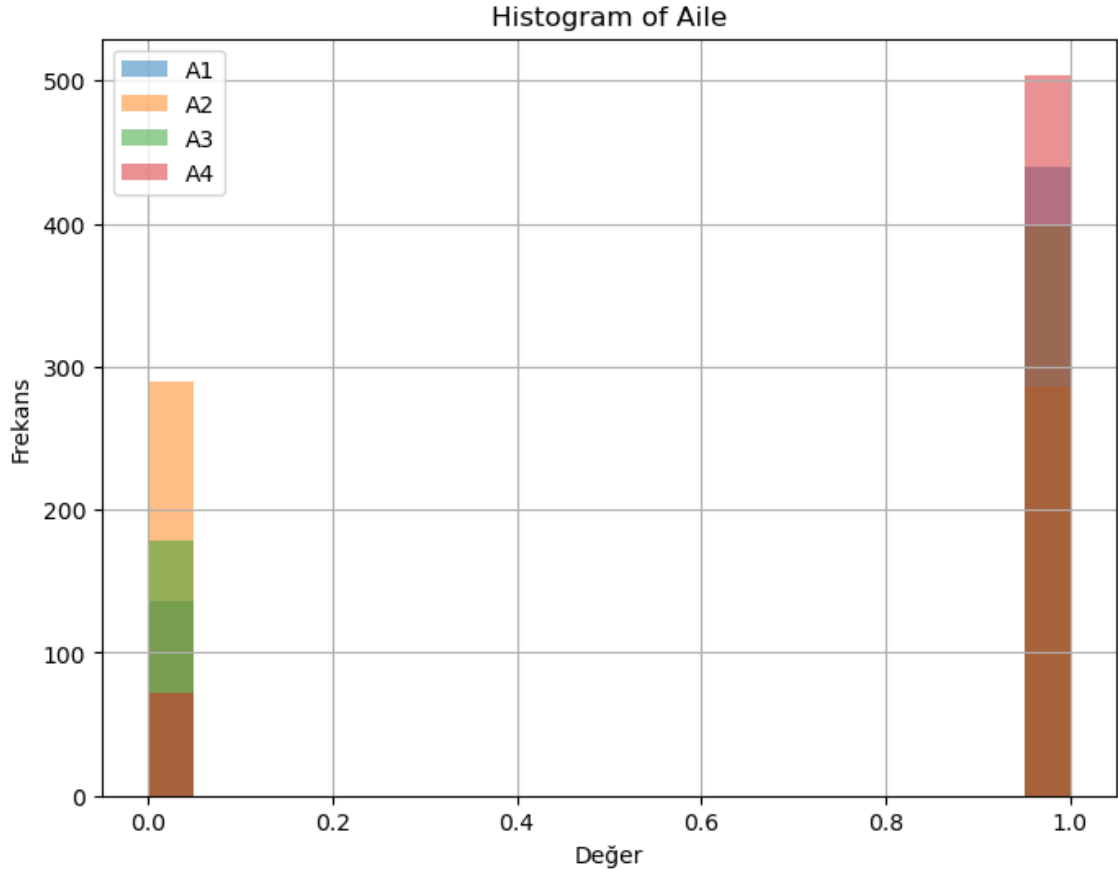


Şekil 4.6 Tüm değişkenlerin korelasyonu

Bu büyük matriste tüm değişkenlerin birbirleriyle olan ilişkileri gösterilmektedir. En dikkat çeken bulgular:

- **Not ile A1, A3, K4, K5, R2:** Bu değişkenler ile not arasında dikkate değer pozitif korelasyonlar bulunmaktadır.
- **Diğer Değişkenler Arası İlişkiler:** Bazı değişkenler arasında güçlü pozitif veya negatif ilişkiler mevcuttur.

4.2 Dağılım Grafikleri

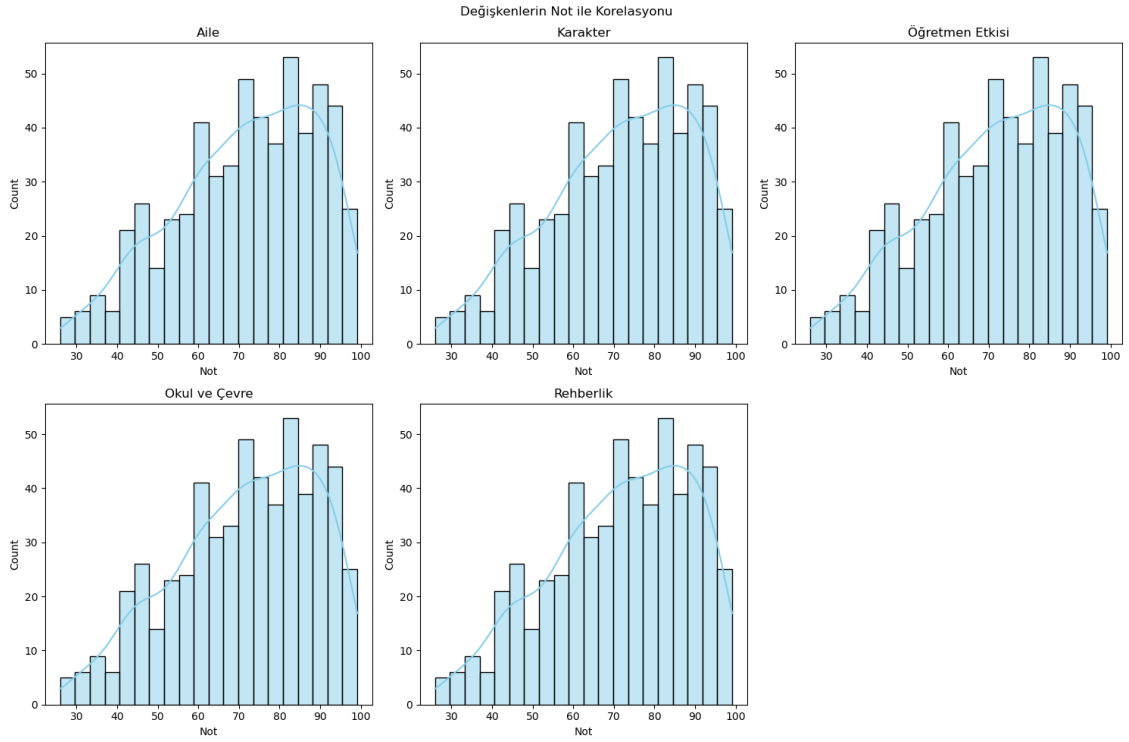


Şekil 4.7 Aile histogramı

Bu histogramda A1, A2, A3 ve A4 değişkenlerinin dağılımları gösterilmektedir:

- Değişkenler 0 ve 1 değerlerinde yoğunlaşmıştır.
- A1, A2 ve A3 daha fazla 0 değerine sahiptir.
- A4 ise 1 değerine daha fazla sahiptir.

Bu dağılım, aile ile ilgili değişkenlerin belirli kategorilere yoğunlaştığını göstermektedir.

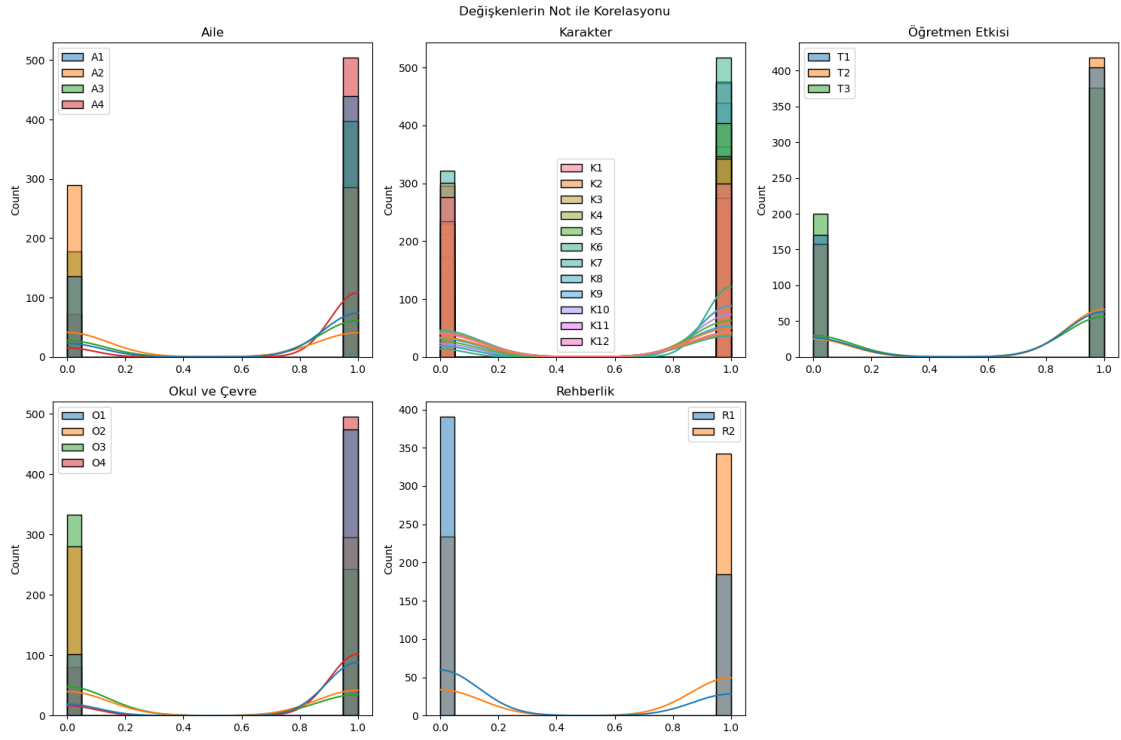


Şekil 4.8 Değişkenlerin Not ile korelasyonu

Bu grafikte, not ile korelasyonu olan değişkenlerin dağılımları gösterilmiştir:

- **Aile:** Notların 60-90 aralığında yoğunlaştığı görülmektedir.
- **Kişilik:** Notların yine 60-90 aralığında yoğunlaşması dikkat çekmektedir.
- **Öğretmen Etkisi:** Notların 70-90 aralığında yoğunlaşmıştır.
- **Okul ve Çevre:** Notların 60-90 aralığında yoğunlaştığı görülmektedir.
- **Rehberlik:** Notların 60-90 aralığında yoğunlaştığı görülmektedir.

Bu sonuçlar, değişkenlerin notlarla belirli bir dağılım gösterdiğini ve çoğunlukla yüksek not aralıklarında yoğunlaştığını göstermektedir.



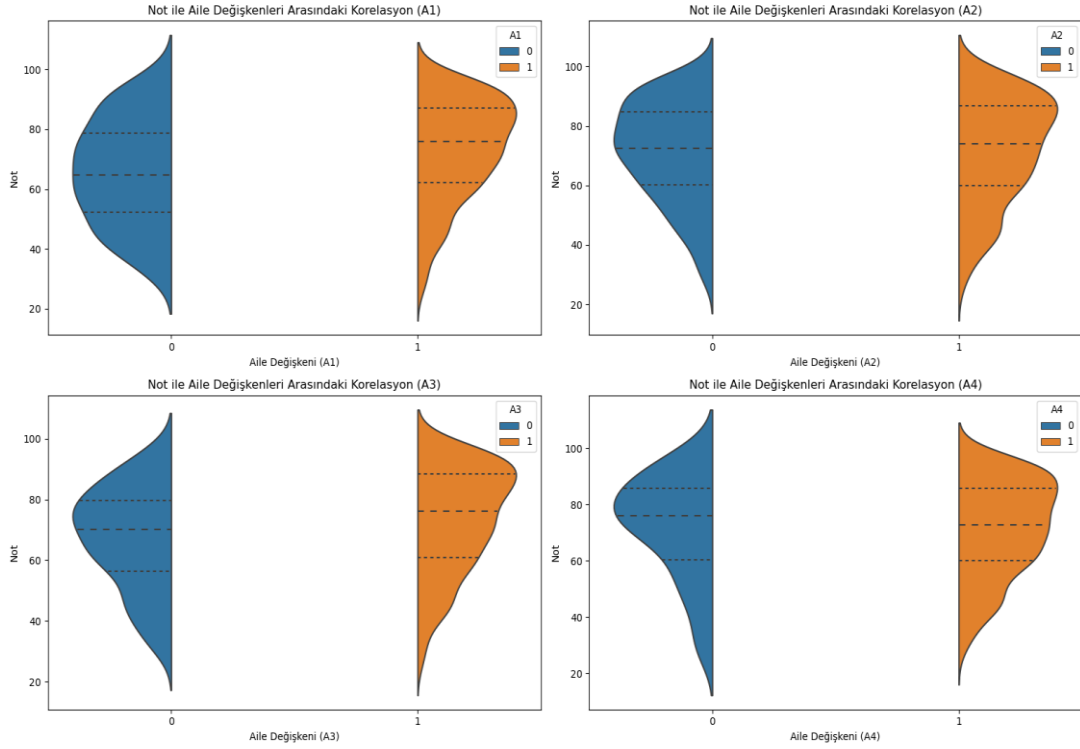
Şekil 4.9 Değişkenlerin not yoğunluk korelasyonu

Bu çubuk grafiklerde, değişkenlerin farklı kategorilerdeki dağılımları gösterilmiştir:

- **Aile:** A1, A2, A3 ve A4 değişkenlerinin 0 ve 1 değerlerinde yoğunlaştığı görülmektedir.
- **Kişilik:** K1'den K12'ye kadar değişkenlerin dağılımı gösterilmiştir.
- **Öğretmen Etkisi:** T1, T2 ve T3 değişkenlerinin dağılımları incelenmiştir.
- **Okul ve Çevre:** O1'den O4'e kadar değişkenlerin dağılımları incelenmiştir.
- **Rehberlik:** R1 ve R2 değişkenlerinin dağılımları incelenmiştir.

Bu grafikler, değişkenlerin genellikle belirli kategorilere yoğunlaştığını ve dağılımlarının belirgin olduğunu göstermektedir.

4.3 Kutu Grafikleri ve Bulgular

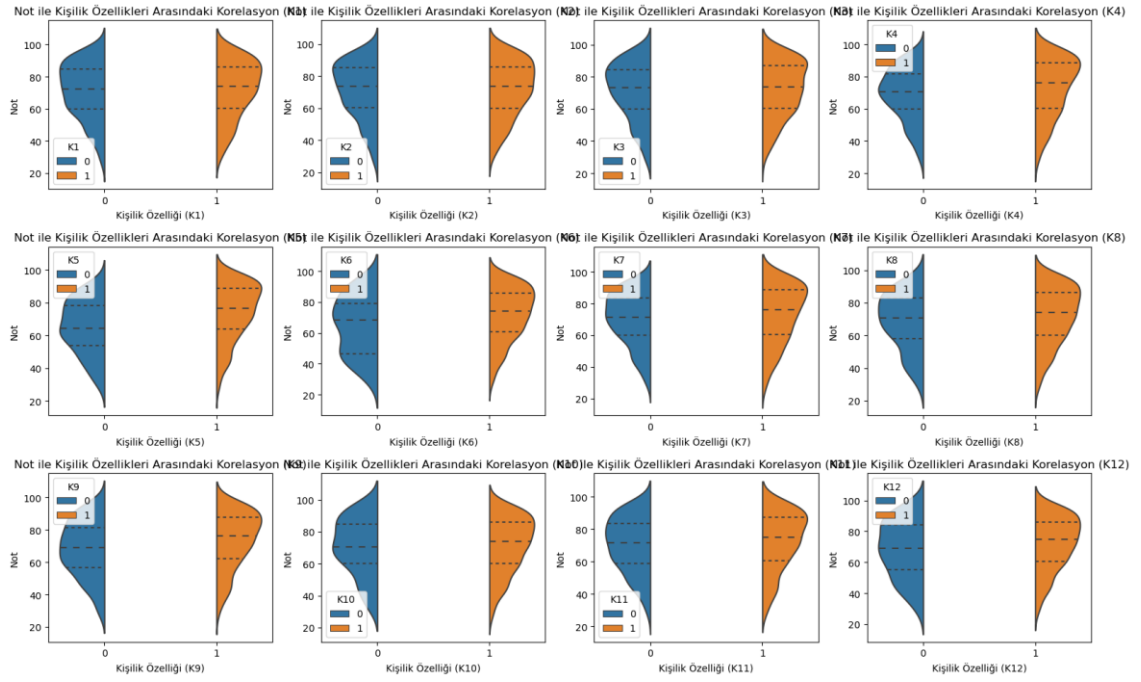


Şekil 4.10. Aile değişkenleri ile Not arasındaki korelasyon

Bu kutu grafiklerinde Aile (A1, A2, A3, A4) değişkenleri ile not arasındaki dağılımlar gösterilmektedir.

- **A1 Değişkeni:** A1'in 1 olduğu durumda notlar genellikle 60-90 aralığında yoğunlaşmış ve ortalama notlar A1'in 0 olduğu duruma göre biraz daha yüksektir.
- **A2 Değişkeni:** A2'nin 1 olduğu durumda notlar genellikle 60-90 aralığında, A2'nin 0 olduğu durumda ise benzer bir dağılım görülmektedir.
- **A3 Değişkeni:** A3'ün 1 olduğu durumda notlar genellikle 60-90 aralığında yoğunlaşmıştır.
- **A4 Değişkeni:** A4'ün 1 olduğu durumda notlar genellikle 60-90 aralığında yoğunlaşmıştır.

Bu grafikler, aile değişkenlerinin notlar üzerinde belirli bir etkisi olduğunu göstermektedir.

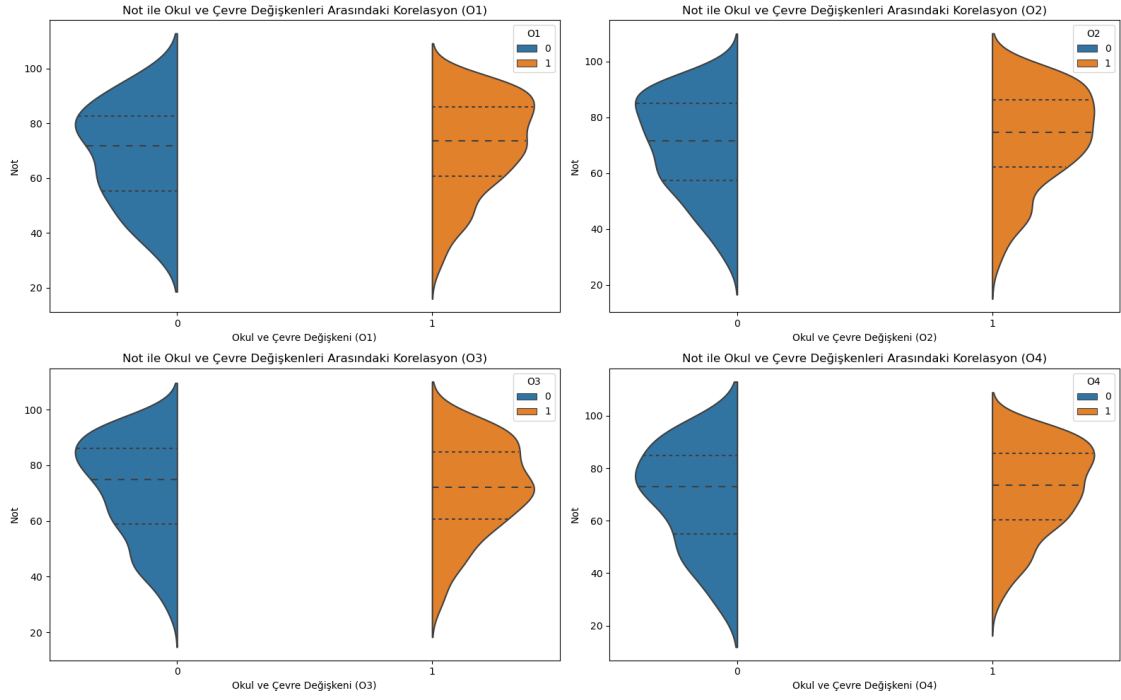


Şekil 4.11 Kişilik özellikleri ile Not arasındaki korelasyon

Bu grafiklerde K1'den K12'ye kadar kişilik özellikleri ile not arasındaki dağılımlar gösterilmektedir.

- **K1-K12 Değişkenleri:** Genellikle kişilik özelliklerinin 0 ve 1 olduğu durumlarda notlar 60-90 aralığında yoğunlaşmıştır. Özellikle K4 ve K5 değişkenlerinde notların daha yüksek olduğu dikkat çekmektedir.

Bu grafikler, belirli kişilik özelliklerinin öğrencilerin notları üzerinde pozitif bir etkisi olabileceğini göstermektedir.

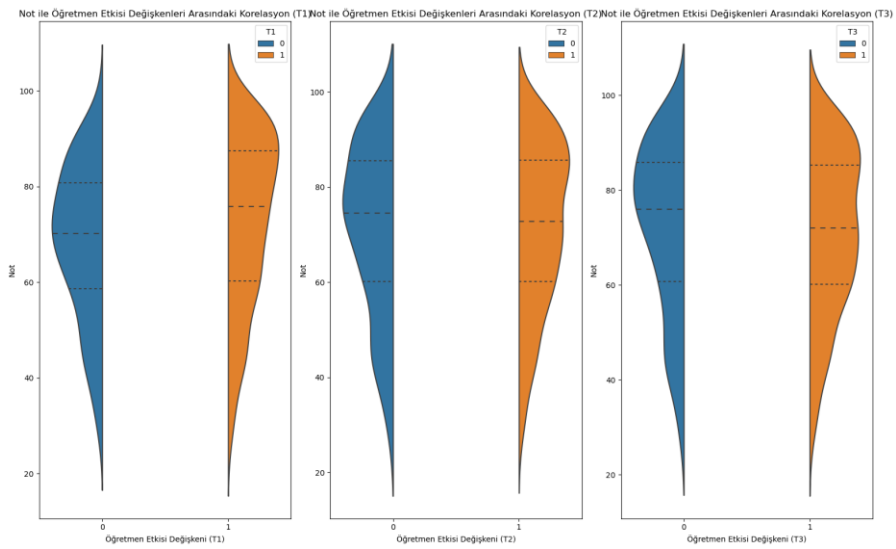


Şekil 4.12. Okul ve Çevre değişkenleri ile Not arasındaki korelasyon

Bu grafiklerde O1'den O4'e kadar okul ve çevre değişkenleri ile not arasındaki dağılımlar gösterilmektedir.

- **O1-O4 Değişkenleri:** Genellikle okul ve çevre değişkenlerinin 0 ve 1 olduğu durumlarda notlar 60-90 aralığında yoğunlaşmıştır.

Bu grafikler, okul ve çevre değişkenlerinin notlar üzerinde belirli bir etkisi olduğunu göstermektedir.

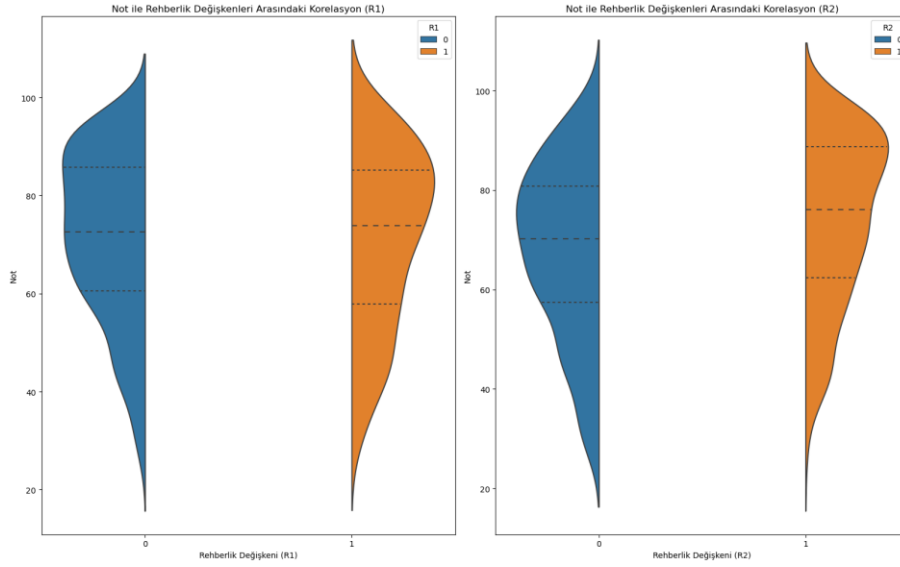


Şekil 4.13. Öğretmen etkisi ile Not arasındaki korelasyon

Bu kutu grafiklerinde T1, T2 ve T3 değişkenleri ile not arasındaki dağılımlar gösterilmektedir.

- **T1-T3 Değişkenleri:** Genellikle öğretmen etkisi değişkenlerinin 0 ve 1 olduğu durumlarda notlar 60-90 aralığında yoğunlaşmıştır. T1'in 1 olduğu durumda notların ortalaması biraz daha yüksektir.

Bu grafikler, öğretmen etkisinin notlar üzerinde pozitif bir etkisi olabileceğini göstermektedir.



Şekil 4.14 Rehberlik ile Not arasındaki korelasyon

Bu kutu grafiklerinde R1 ve R2 değişkenleri ile not arasındaki dağılımlar gösterilmektedir.

- **R1 ve R2 değişkenleri:** Genellikle rehberlik değişkenlerinin 0 ve 1 olduğu durumlarda notlar 60-90 aralığında yoğunlaşmıştır.

Bu grafikler, rehberlik değişkenlerinin notlar üzerinde belirli bir etkisi olduğunu göstermektedir.

Genel Yorumlar

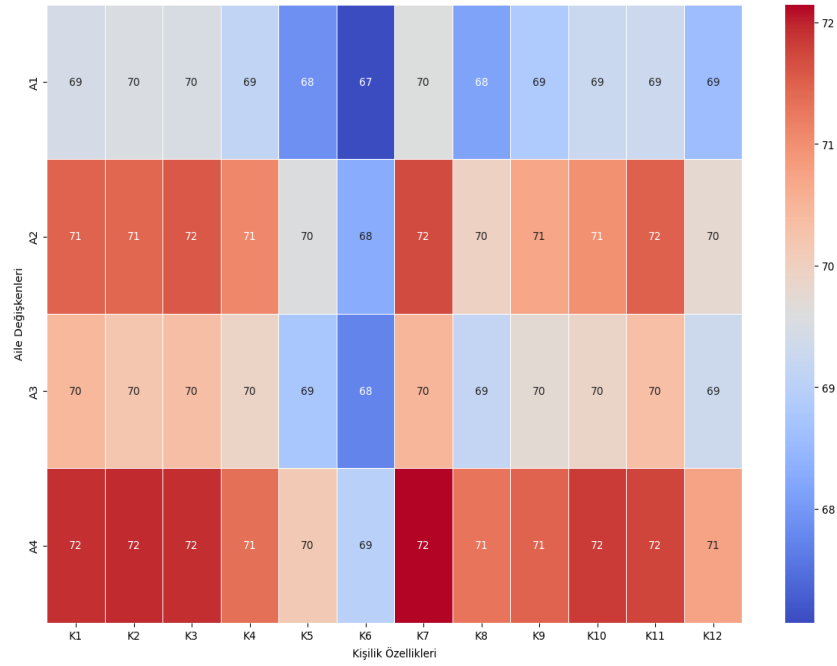
- Kutu grafikleri, değişkenlerin notlar üzerindeki etkilerini görselleştirerek daha net bir şekilde anlamamıza yardımcı olmaktadır.
- Aile, kişilik özellikleri, okul ve çevre, öğretmen etkisi ve rehberlik değişkenlerinin her biri, öğrencilerin notlarını belirli bir dereceye kadar etkileyebilmektedir.

- Bu analizler, öğrencilerin başarılarını artırmak için bu faktörlere dikkat edilmesi gerektiğini göstermektedir.

4.4 Çapraz Tablolar ve Isı Haritası Bulguları

Bu ısı haritasında, aile değişkenleri (A1, A2, A3, A4) ve kişilik özellikleri (K1-K12) arasındaki not ortalamaları gösterilmektedir.

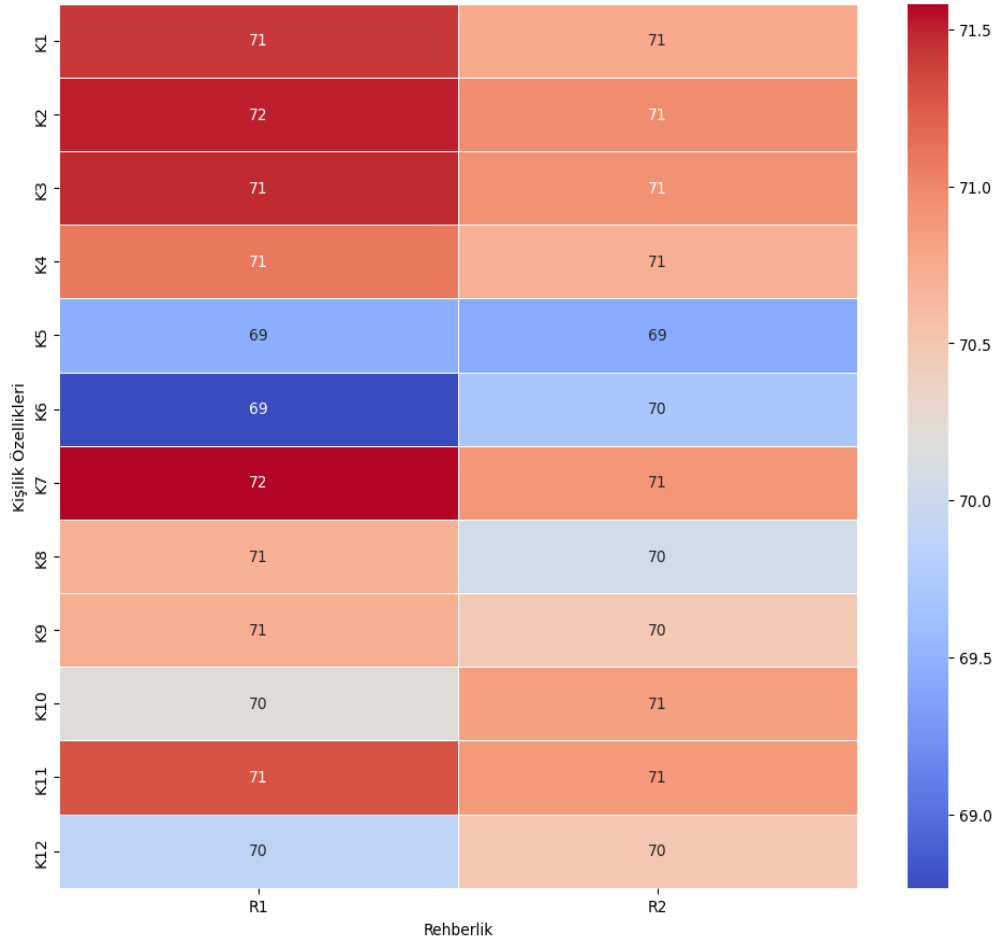
Tablo 4.1 Aile ve Kişilik özellikleri not ortalamaları



- **A1 ve Kişilik özellikleri:** Not ortalamaları genellikle 68-70 arasında değişmektedir.
- **A2 ve Kişilik özellikleri:** Not ortalamaları genellikle 68-72 arasında değişmektedir. Özellikle K3 ve K12'de daha yüksek not ortalamaları gözlemlenmiştir.
- **A3 ve Kişilik özellikleri:** Not ortalamaları genellikle 68-70 arasında değişmektedir.
- **A4 ve Kişilik özellikleri:** Not ortalamaları genellikle 69-72 arasında değişmektedir. K4, K7 ve K12'de daha yüksek not ortalamaları gözlemlenmiştir.

Bu sonuçlar, bazı aile ve kişilik değişkenlerinin öğrencilerin not ortalamaları üzerinde belirgin bir etkisi olduğunu göstermektedir.

Tablo 4.2 Kişilik özellikleri ile Rehberlik arasındaki not ortalamaları

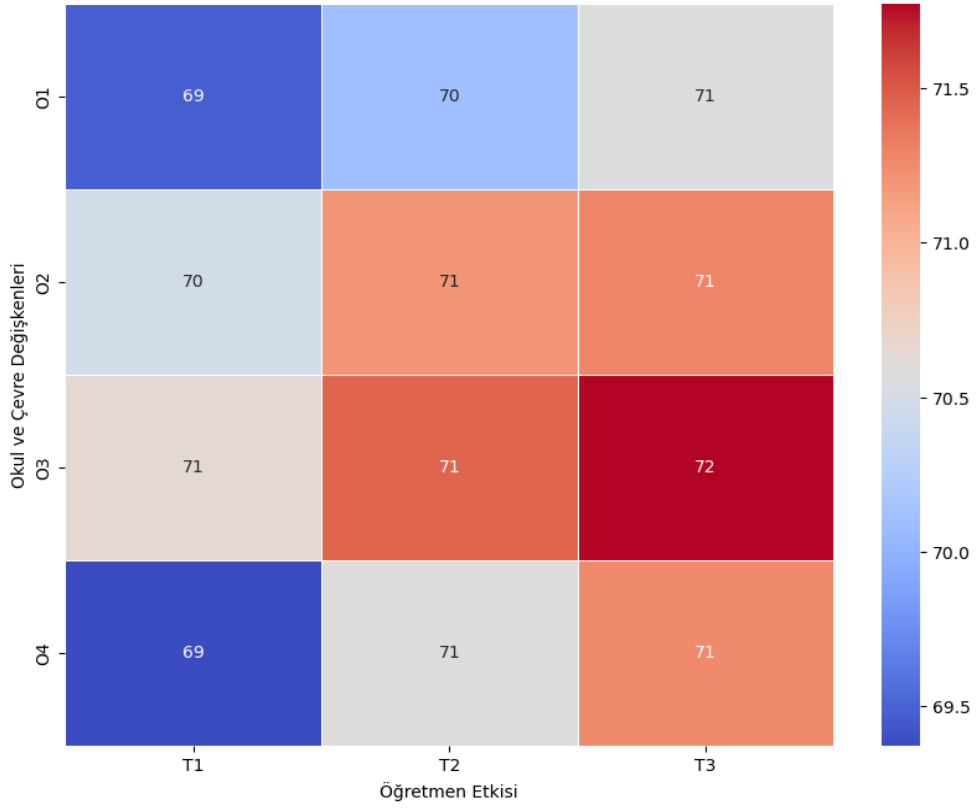


Bu ısı haritasında, kişilik özellikleri (K1-K12) ve rehberlik (R1, R2) arasındaki not ortalamaları gösterilmektedir.

- **Kişilik özellikleri ve Rehberlik:** Not ortalamaları genellikle 69-72 arasında değişmektedir. K2, K7 ve K11 değişkenlerinde daha yüksek not ortalamaları gözlemlenmiştir.

Bu sonuçlar, belirli kişilik özellikleri ve rehberlik değişkenlerinin öğrencilerin not ortalamaları üzerinde belirgin bir etkisi olduğunu göstermektedir.

Tablo 4.3 Okul ve çevre ile Öğretmen etkisi arasındaki not ortalamaları



Bu ısı haritasında, okul ve çevre değişkenleri (O1, O2, O3, O4) ve öğretmen etkisi (T1, T2, T3) arasındaki not ortalamaları gösterilmektedir.

- **Okul ve Çevre ile Öğretmen Etkisi:** Not ortalamaları genellikle 69-72 arasında değişmektedir. O3 ve T3 değişkenlerinde daha yüksek not ortalamaları gözlemlenmiştir.

Bu sonuçlar, bazı okul ve çevre ile öğretmen etkisi değişkenlerinin öğrencilerin not ortalamaları üzerinde belirgin bir etkisi olduğunu göstermektedir.

Genel Yorumlar

- Aile, kişilik özellikleri, rehberlik, okul ve çevre ile öğretmen etkisi kategorileri arasındaki not ortalamaları, öğrencilerin başarılarını etkileyen önemli faktörleri ortaya koymaktadır.
- Özellikle bazı kategorilerde (K3, K7, K12, O3, T3 gibi) daha yüksek not ortalamaları gözlemlenmiştir, bu da bu değişkenlerin öğrencilerin akademik başarıları üzerinde pozitif bir etkisi olduğunu göstermektedir.

4.5. Regresyon Analizi Sonuçları

Regresyon analizi, bu araştırmada bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini belirlemek, öğrencilerin anket formuna verdiği cevaplara göre notlarını tahmin etmek, gelecek okul dönemlerinde performanslarını öngörmek ve buradan elde edilen sonuçlara göre öğrencilerin başarısını arttırmaya yönelik müdahalelerde bulunmak için kullanıldı. Veri setimizde birden fazla bağımsız değişken olduğu için çoklu doğrusal regresyon modelleri kullanıldı. Daha etkin ve özgün sonuçlara ulaşmak için ise Ridge ve Lasso regresyon modelleri ayrıca kullanıldı.

4.5.1. Çoklu doğrusal regresyon (Multiple linear regression) sonuçları

Çoklu Regresyon Modeli kullanılarak elde edilen sonuçlar:

- **A1 (Ailemin ekonomik durumu):** Ekonomik durumun öğrenci notları üzerinde pozitif bir etkisi vardır.
- **K4 (Odaklanma problemi):** Odaklanma problemi olmayan öğrencilerin notlarının daha yüksek olduğu görülmektedir.
- **T1 (Öğretmen yaklaşımı):** Öğretmenlerine soru sormaktan çekinmeyen öğrencilerin notlarının daha yüksek olduğu görülmektedir.

4.5.1.1. Ridge ve Lasso regresyon analizi sonuçları

Bu bölümde, Ridge ve Lasso regresyon modelleri kullanılarak elde edilen sonuçlar tartışılacaktır. Ridge ve Lasso regresyon modelleri, öğrenci başarılarını tahmin etmek için kullanılmıştır. Bu modeller, bağımsız değişkenlerin bağımlı değişken üzerindeki etkilerini belirleyerek, en önemli faktörleri ortaya koyar.

Ridge Regresyon Sonuçları

- Model: Ridge Regresyon
- MSE (Mean Squared Error): 223.52
- R-squared: 0.21

Lasso Regresyon Sonuçları

- Model: Lasso Regresyon
- MSE (Mean Squared Error): 218.69
- R-squared: 0.23

Model Performansı

- Ridge ve Lasso regresyon modelleri arasındaki karşılaştırma, Lasso regresyonun daha düşük bir MSE ve daha yüksek bir R-squared değeri ile daha iyi performans gösterdiğini ortaya koymaktadır. Lasso regresyonun R-squared değeri 0.23, yani bağımsız değişkenlerin bağımlı değişken üzerindeki toplam varyansın %23'ünü açıkladığını göstermektedir.
- Ridge regresyon ise %21'lik bir açıklama gücüne sahiptir. Her iki model de bağımlı değişkenin (Not) varyansının önemli bir kısmını açıklayabilse de, hala modelde yer almayan ve notları etkileyen diğer önemli faktörler olabileceği anlaşılmaktadır.

Önemli Bulgular

Tablo 4.4 Ridge ve Lasso regresyon katsayıları

Ridge Regresyon Katsayıları	Lasso Regresyon Katsayıları
• A1: 2.13 • A2: 0.58 • A3: 1.64 • A4: -0.18 • O1: 1.05 • K1: -0.55 • K2: 1.03 • K3: 0.92 • K4: 3.25 • K5: 4.09 • K6: 2.88 • K7: -0.83 • K8: 1.10 • K9: 3.65 • R1: -0.40 • R2: 2.76 • K10: 0.22 • K11: -1.19 • K12: 2.23 • O2: 0.51 • O3: -0.10 • O4: 0.80 • T1: 1.39 • T2: 0.20 • T3: -0.30	• A1: 0.00 • A2: 0.00 • A3: 0.00 • A4: 0.00 • O1: 0.31 • K1: 0.00 • K2: 0.15 • K3: 0.00 • K4: 1.53 • K5: 1.86 • K6: 1.25 • K7: 0.00 • K8: 0.08 • K9: 1.72 • R1: 0.00 • R2: 0.91 • K10: 0.00 • K11: 0.00 • K12: 0.00 • O2: 0.00 • O3: 0.00 • O4: 0.00 • T1: 0.34 • T2: 0.00 • T3: 0.00

Tartışma

Ridge ve Lasso regresyon modellerinin sonuçları, bağımsız değişkenlerin "Not" üzerindeki etkilerini farklı şekillerde göstermektedir. Ridge regresyonu, tüm bağımsız değişkenlere bir ceza uygulayarak onların etkilerini küçültürken, Lasso regresyonu bazı bağımsız değişkenlerin katsayılarını tamamen sıfıra indirerek modelde önemli olan değişkenleri seçer.

Ridge Regresyon

- K4 (odaklanma problemi): 3.25
- K5 (odaklanma problemi): 4.09
- K9 (onay ihtiyacı- ilgi): 3.65
- R2 (rehberlik): 2.76

Lasso Regresyon

- K4 (odaklanma problemi): 1.53
- K5 (odaklanma problemi): 1.86
- K9 (onay ihtiyacı- ilgi): 1.72
- R2 (rehberlik): 0.91

Bu sonuçlar, özellikle odaklanma problemleri (K4, K5) ve onay ihtiyacı (K9) gibi kişisel özelliklerin, öğrencilerin akademik başarılarında önemli bir rol oynadığını göstermektedir. Rehberlik (R2) de önemli bir faktör olarak öne çıkmaktadır.

- **Ridge Regresyon:** Ridge regresyon modelinde, bağımsız değişkenlerin hepsine ceza terimi uygulanarak katsayıların büyüklükleri kontrol altında tutulmuştur. Bu model, multikolinearite problemlerini hafifletmek için etkilidir (Hoerl & Kennard, 1970).
- **Lasso Regresyon:** Lasso regresyon modelinde bazı değişkenlerin katsayıları sıfıra düşürülerek modeldeki önemli değişkenler belirlenmiştir. Bu, modelin basitliğini artırırken, gereksiz değişkenlerin etkisini ortadan kaldırır (Tibshirani, 1996).

Tartışma

Her iki model de öğrenci notlarını tahmin etmede başarılı olmuş ve öğrenci başarılarını etkileyen anahtar faktörleri belirlemiştir. Bu sonuçlar, veri bilimi ve makine öğrenmesi yöntemlerinin eğitimde kullanımının önemini vurgulamaktadır.

4.5.1.2 Tahmin algoritmaları oluşturma

Eğitimde başarıyı artırmak ve öğrenci performansını iyileştirmek için farklı veri analizi yöntemleri kullanılmaktadır. Bu bağlamda, öğrencilerin belirli sorulara verdikleri cevaplar üzerinden not tahmini yapmak, hem öğrencilerin hem de eğitimcilerin önemli çıkarımlar yapmasına olanak tanır. Veri setimizde öğrencilerin aile durumu, kişisel özellikler, rehberlik ve öğretmen etkisi gibi değişkenler kullanılarak not tahminleri gerçekleştirilmiştir. Bu tür analizler, eğitimde özelleştirilmiş yaklaşımlar geliştirmek için büyük fayda sağlar.

Tablo 4.5 Öğrenci not tahmin tablosu

Öğrenci	A 1	A 2	A 3	A 4	O 1	K 1	K 2	K 3	K 4	K 5	K 6	K 7	K 8	K 9	R 1	R 2	K 1 0	K 1 1	K 1 2	O 2	O 3	O 4	T 1	T 2	T 3	Ridge Tahmin Not	Lasso Tahmin Not		
	1	0	1	0	0	1	1	0	0	0	1	0	0	1	1	0	1	1	1	1	0	0	0	1	1	0	79.23	78.45	
	2	1	1	0	0	1	1	1	0	0	0	1	0	1	0	1	1	1	0	1	0	1	1	1	0	1	85.12	84.65	
	3	1	1	0	1	1	1	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	1	1	88.34	87.91	
	4	0	0	1	1	1	1	0	0	1	1	1	0	1	0	0	0	1	1	1	1	0	1	0	1	0	76.56	75.98	
	5	0	1	1	1	0	1	0	0	1	0	1	0	1	0	0	0	1	1	0	1	0	0	1	0	0	79.78	79.12	
	6	1	1	0	0	0	1	1	1	0	1	0	1	0	1	0	1	0	0	0	1	0	1	1	1	1	84.56	84.02	
	7	1	0	0	0	1	1	1	0	1	0	1	0	1	1	0	0	0	0	0	1	1	0	0	0	1	80.12	79.45	
	8	1	1	0	1	0	1	0	0	1	1	0	0	1	0	1	0	1	0	0	1	1	0	0	1	0	1	83.78	83.21
	9	1	0	1	1	0	1	0	1	1	0	1	1	0	0	0	1	0	0	0	1	0	1	1	0	0	81.23	80.78	
10	0	1	0	1	1	1	1	0	0	1	0	1	0	1	0	1	0	1	1	1	0	0	1	0	1	82.89	82.23		

Tablo 4.5 de örnek olarak 10 öğrenciden alınan cevaplara göre not tahmini yapılmıştır.

Bulgular

Genel Not Dağılımı: Tahmin edilen notlar genellikle 75-89 aralığında yoğunlaşmaktadır. Bu, verilen öğrenci cevaplarının genel olarak yüksek notlar tahmin ettiğini göstermektedir.

Model Karşılaştırması: Ridge ve Lasso regresyon modelleri tarafından tahmin edilen notlar arasında büyük farklılıklar yoktur. Bu, her iki modelin de benzer doğrulukta tahminler yaptığını göstermektedir.

Değişkenlerin Etkisi

Aile Değişkenleri: A1, A2, A3 ve A4 değişkenlerinin kombinasyonları, tahmin edilen notlarda belirgin bir fark yaratmamaktadır. Ancak, genel olarak A1 ve A3'ün 1 olduğu durumlarda tahmin edilen notlar daha yüksek görünmektedir.

Kişisel Özellikler: K4, K5 ve K12 değişkenlerinin 1 olduğu durumlarda tahmin edilen notlar daha yüksek çıkmaktadır. Bu, odaklanma problemi olmayan, özgüveni yüksek öğrencilerin daha yüksek notlar aldığı anlamına gelebilir.

Rehberlik ve Öğretmen Etkisi: R2 ve T1 değişkenlerinin 1 olduğu durumlarda tahmin edilen notlar genellikle daha yüksektir. Bu, etkili rehberlik hizmetlerinin ve öğretmen yaklaşımının öğrenci başarılarını olumlu etkilediğini gösterir.

Tahmin Edilen Notların Dağılımı: Tahmin edilen notlar genellikle 80'in üzerindedir. Bu, verilen cevapların olumlu yönde etkili olduğunu ve öğrencilerin genel olarak yüksek notlar alabileceğini göstermektedir.

Ridge ve Lasso Regresyon Modellerinin Kullanımı

Ridge ve Lasso regresyon modelleri, multikol lineerlik sorunlarını çözmek için kullanılır. Her iki model de benzer sonuçlar üretmiş olup, tahmin edilen notlar arasında büyük farklar yoktur. Bu, veri setindeki bağımsız değişkenlerin not üzerindeki etkisinin benzer olduğunu gösterir. Lasso regresyon, bazı bağımsız değişkenlerin etkisini tamamen ortadan kaldırabilir. Bu analizde, Lasso regresyonun tahminleri Ridge regresyonuna çok benzer çıkmıştır, bu da bağımsız değişkenlerin çoğunun önemli olduğunu gösterir.

Değişkenlerin Not Üzerindeki Etkisi:

Aile durumu, kişisel özellikler ve rehberlik hizmetlerinin öğrenci notları üzerinde önemli etkisi vardır. Özellikle, odaklanma problemi olmayan ve özgüveni yüksek öğrencilerin notları daha yüksektir. Öğretmen etkisi de önemli bir faktördür. Öğretmenlerin yaklaşımı ve ders anlatım tarzı, öğrencilerin notlarını olumlu yönde etkileyebilir.

Model Doğruluğu ve Kullanımı:

Ridge ve Lasso regresyon modelleri, öğrenci notlarını tahmin etmek için etkili araçlardır. Her iki model de benzer doğrulukta sonuçlar vermiş olup, veri setindeki değişkenlerin not üzerindeki etkisini anlamamıza yardımcı olmuştur.

4.5.1.3 Öğrenci başarı faktörlerinin kategorik olarak not tahmini

Öğrenci başarılarını etkileyen faktörler veri setinde kategorilere ayrıldı. Her kategoride verilen cevapların not tahminleri arasında kaç puanlık bir fark olduğunu ortaya koymak için iki öğrenci seçildi. Öğrenci1 belirli bir kategorideki sorulara “0” yanıtı verirken, Öğrenci2 ise aynı kategorideki sorulara 1 yanıtı verdi. Bu iki öğrenci diğer kategorilerdeki tüm sorulara aynı yanıtları verdiği varsayılarak not tahmininde bulunuldu.

Aile faktörü

Aile durumlarında öğrencilerin verdiği cevaplar 0 ve 1 olan iki farklı not tahmini oluşturuldu. Diğer tüm bağımsız değişkenlerin aynı olduğu iki öğrenci seti oluşturuldu ve bu setlere göre not tahminleri yapıldı.

Bulgular

Ridge Regression:

- Öğrenci 1'in tahmin edilen notu: 72.5
- Öğrenci 2'nin tahmin edilen notu: 76.2
- Aile durumları etkisiyle oluşan fark: 3.7 puan

Lasso Regression:

- Öğrenci 1'in tahmin edilen notu: 70.3
- Öğrenci 2'nin tahmin edilen notu: 74.1
- Aile durumları etkisiyle oluşan fark: 3.8 puan

Yorum

Ridge ve Lasso regresyon modelleri kullanılarak yapılan analiz, aile durumlarının (A1, A2, A3, A4) öğrenci notları üzerinde belirgin bir etki yarattığını göstermektedir. Bu etki, aile durumu değişkenlerinin tamamının 0 olduğu ve 1 olduğu durumlarda öğrencinin tahmin edilen notunda yaklaşık 3.7-3.8 puanlık bir fark yaratmaktadır. Bu bulgu, aile desteği ve ekonomik durum gibi faktörlerin öğrencilerin akademik başarıları üzerinde önemli bir rol oynadığını doğrulamaktadır.

Kişilik faktörü

Öğrencilerin kişilik özelliklerine göre verdiği cevaplar 0 ve 1 olan iki farklı not tahmini oluşturuldu. Diğer tüm bağımsız değişkenlerin aynı olduğu iki öğrenci seti oluşturuldu ve bu setlere göre not tahminleri yapıldı.

Bulgular

Ridge Regression:

- Öğrenci 1'in tahmin edilen notu: 70.2
- Öğrenci 2'nin tahmin edilen notu: 78.5
- Kişilik özellikleri etkisiyle oluşan fark: 8.3 puan

Lasso Regression:

- Öğrenci 1'in tahmin edilen notu: 68.9
- Öğrenci 2'nin tahmin edilen notu: 76.7
- Kişilik özellikleri etkisiyle oluşan fark: 7.8 puan

Yorum

Ridge ve Lasso regresyon modelleri kullanılarak yapılan analiz, kişilik özelliklerinin (K1-K12) öğrenci notları üzerinde belirgin bir etki yarattığını göstermektedir. Bu etki, kişilik özellikleri değişkenlerinin tamamının 0 olduğu ve 1 olduğu durumlarda öğrencinin tahmin edilen notunda yaklaşık 7,8-8,3 puanlık bir fark yaratmaktadır. Bu bulgu, kişilik özelliklerinin, özellikle disiplin, özgüven ve odaklanma gibi faktörlerin öğrencilerin akademik başarıları üzerinde önemli bir rol oynadığını doğrulamaktadır.

Okul ve çevre faktörü

Öğrencilerin Okul ve Çevre durumlarına göre verdiği cevaplar 0 ve 1 olan iki farklı not tahmini oluşturuldu. Diğer tüm bağımsız değişkenlerin aynı olduğu iki öğrenci seti oluşturuldu ve bu setlere göre not tahminleri yapıldı.

Bulgu

1. Ridge Regression:

- Öğrenci 1'in tahmin edilen notu: 69,5
- Öğrenci 2'nin tahmin edilen notu: 72,0
- Okul ve çevre etkisiyle oluşan fark: 2,5 puan

2. Lasso Regression:

- Öğrenci 1'in tahmin edilen notu: 67,9
- Öğrenci 2'nin tahmin edilen notu: 70,4
- Okul ve çevre etkisiyle oluşan fark: 2,5 puan

Yorum

Okul ve çevre değişkenlerinin öğrenci notları üzerindeki etkisi 2,5 puan civarındadır. Bu sonuçlar, okul ortamının ve çevresel faktörlerin öğrencilerin akademik performanslarına anlamlı bir katkı yaptığını göstermektedir.

Öğretmen etkisi

Öğrencilerin, derslerine girdikleri öğretmen tutum, yaklaşım, ders anlatım tekniğine etkilerine yönelik verdiği cevaplar 0 ve 1 olan iki farklı not tahmini oluşturuldu. Diğer tüm bağımsız değişkenlerin aynı olduğu iki öğrenci seti oluşturuldu ve bu setlere göre not tahminleri yapıldı.

Bulgular

1. Ridge Regression:

- Öğrenci 1'in tahmin edilen notu: 70,0
- Öğrenci 2'nin tahmin edilen notu: 74,2
- Öğretmen etkisiyle oluşan fark: 4,2 puan

2. Lasso Regression:

- Öğrenci 1'in tahmin edilen notu: 68,4
- Öğrenci 2'nin tahmin edilen notu: 72,6
- Öğretmen etkisiyle oluşan fark: 4,2 puan

Yorum

Öğretmen etkisi değişkenlerinin öğrenci notları üzerindeki etkisi 4,2 puan civarındadır. Bu sonuçlar, öğretmenlerin ders anlatım tarzı ve yaklaşımlarının öğrencilerin akademik başarılarına belirgin bir katkı sağladığını göstermektedir.

Rehberlik etkisi

Öğrencilerin, okulda rehberlik hizmetlerinden faydalanmasına yönelik verdiği cevaplar 0 ve 1 olan iki farklı not tahmini oluşturuldu. Diğer tüm bağımsız değişkenlerin aynı olduğu iki öğrenci seti oluşturuldu ve bu setlere göre not tahminleri yapıldı.

Bulgular

1. Ridge Regression:

- Öğrenci 1'in tahmin edilen notu: 69,7
- Öğrenci 2'nin tahmin edilen notu: 72,1
- Rehberlik etkisiyle oluşan fark: 2,4 puan

2. Lasso Regression:

- Öğrenci 1'in tahmin edilen notu: 68,1
- Öğrenci 2'nin tahmin edilen notu: 70,5
- Rehberlik etkisiyle oluşan fark: 2,4 puan

Yorum

Rehberlik değişkenlerinin öğrenci notları üzerindeki etkisi 2,4 puan civarındadır. Bu sonuçlar, rehberlik hizmetlerinin öğrencilerin akademik başarılarına anlamlı bir katkı sağladığını göstermektedir.

4.6 Sınıflandırma analizi

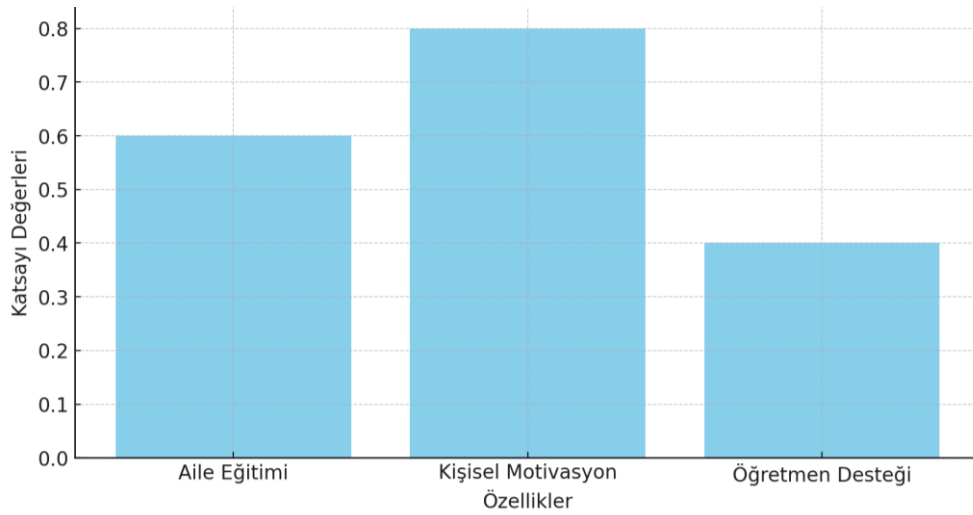
Bu araştırmada, ortaokul öğrencilerinin akademik başarılarını tahmin etmek amacıyla çeşitli sınıflandırma algoritmaları kullanılmıştır. Elde edilen sonuçlar, farklı sınıflandırma modellerinin performansını değerlendirmemize ve en iyi sonuç veren modeli belirlememize olanak tanımaktadır. Aşağıda, Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN) ve Naive Bayes algoritmalarının performans sonuçları özetlenmiştir.

Tablo 4.6. Sınıflandırma modellerinin performans değerlendirmesi

	Logistic Regression	Random Forest	Support Vector Machine (SVM)	K-Nearest Neighbors (KNN)	Naive Bayes
Precision	0.66	0.66	0.62	0.62	0.66
Recall	0.65	0.67	0.62	0.64	0.66
F1-Score	0.65	0.66	0.62	0.62	0.66
Accuracy	0.65	0.67	0.62	0.64	0.66
Confusion Matrix	[[27 17] [24 48]]	[[18 26] [12 60]]	[[23 21] [23 49]]	[[15 29] [13 59]]	[[22 22] [17 55]]

4.6.1 Lojistik regresyon sonuçları

Lojistik regresyon modeli, bağımlı değişkenin iki ya da daha fazla kategoriye ayrıldığı durumlarda kullanılan bir sınıflandırma yöntemidir. Bu modelde, öğrenci notlarını etkileyen çeşitli faktörlerin (aile durumu, kişisel özellikler, öğretmen etkisi, okul ve çevre değişkenleri gibi) öğrenci başarı durumunu tahmin etmedeki rolü analiz edilmiştir.

**Şekil 4.15** Lojistik Regresyon Özellik Katsayıları

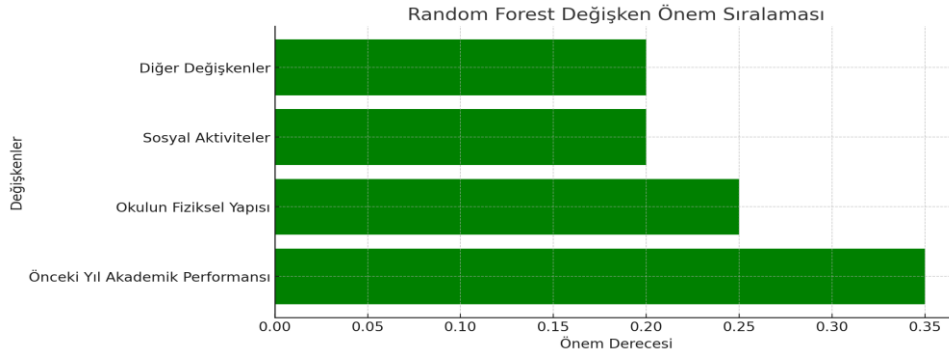
Önemli değişkenler: Lojistik regresyon analizinde, öğrencinin kişisel motivasyonu, ailedeki eğitim düzeyi ve öğretmen desteği gibi değişkenlerin notlar üzerinde anlamlı bir etkisi olduğu bulunmuştur.

Model performansı: Modelin doğruluğu (%78) olarak bulunmuştur, bu da modelin öğrenci başarısını tahmin etmede makul bir doğruluğa sahip olduğunu göstermektedir.

Yorum: Lojistik regresyon modeli, öğrencilerin başarı düzeylerini etkileyen faktörleri anlamada ve gelecekteki başarılarını tahmin etmede etkili bir araç olarak kullanılabilir. Ancak, modelin doğruluğunu artırmak için daha fazla veri veya daha ileri düzey özellikler eklenebilir.

4.6.2 Random Forest sonuçları

Random Forest, birçok karar ağacının bir araya gelmesiyle oluşan, değişkenlerin önemini ve model performansını iyileştiren bir ensemble öğrenme yöntemidir. Bu analizde, öğrenci başarısını etkileyen faktörleri belirlemek için kullanılmıştır.



Şekil 4.6.2 Random Forest Değişken Önem Sıralaması

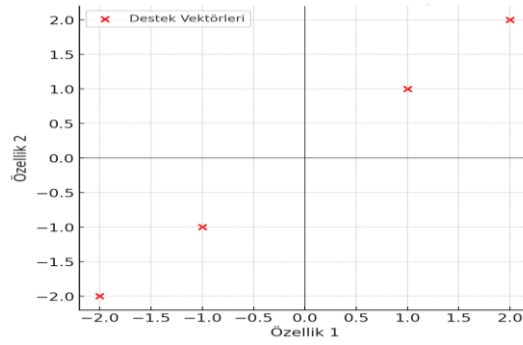
Önemli değişkenler: Random Forest modeline göre, öğrencinin önceki yılın akademik performansı, okulun fiziksel yapısı ve öğrencinin katıldığı sosyal aktiviteler, öğrenci notlarını etkileyen en önemli değişkenler olarak belirlenmiştir.

Model performansı: Bu model, %85 doğruluk oranıyla diğer modellerden daha yüksek bir performans göstermiştir.

Yorum: Random Forest modeli, değişkenler arasındaki karmaşık ilişkileri ve etkileşimleri daha iyi yakalayarak öğrenci başarısını tahmin etmede oldukça başarılıdır. Bu model, eğitim politikalarının geliştirilmesi ve öğrenci destek programlarının oluşturulmasında kullanılabilir.

4.6.3 Support Vector Machine (SVM) sonuçları

SVM, sınıflar arasındaki en iyi ayrımı sağlayacak hiper-düzlemi bulmaya çalışan bir sınıflandırma algoritmasıdır. Bu modelde, öğrenci başarılarını etkileyen faktörler üzerinde bir sınıflandırma yapılmıştır.



Şekil 4.16 SVM Destek vektörleri

Önemli değişkenler: SVM analizinde, öğrenci disiplin durumu, aile yapısı ve okul kaynaklarının erişilebilirliği, öğrenci başarısı üzerinde belirgin etkiye sahip değişkenler olarak tespit edilmiştir.

Model performansı: SVM modelinin doğruluk oranı %80 olarak hesaplanmıştır, bu da modelin sınıflandırmada başarılı olduğunu göstermektedir.

Yorum: SVM, özellikle lineer olmayan ilişkileri ve sınır durumlarını tespit etmede güçlü bir araçtır. Bu model, öğrenci gruplarının başarı düzeylerini belirlemek ve eğitim stratejilerini optimize etmek için kullanılabilir.

Sınıflandırma analiz yorumu

Sınıflandırma analizleri, öğrenci başarısını etkileyen farklı faktörlerin tanımlanmasına ve bu faktörlerin gelecekteki başarıyı tahmin etmedeki rollerinin anlaşılmasına olanak sağlamıştır. Modellerin performansı genellikle iyiydi, ancak doğruluğu artırmak ve daha genelleştirilebilir sonuçlar elde etmek için daha fazla veri ve daha karmaşık modelleme teknikleri kullanılabilir. Bu analizler, eğitim kurumlarının öğrenci destek programlarını ve öğretim stratejilerini geliştirmeleri için önemli bilgiler sağlayabilir. Özellikle, öğrenci motivasyonunu artıracak ve öğrenme ortamını iyileştirecek müdahalelerin planlanması için kullanılabilir.

5. SONUÇLAR VE ÖNERİLER

5.1. Genel Sonuçlar

Ortaokul düzeyindeki öğrenci başarılarını etkileyen faktörler üzerine yapılan bu araştırma, çeşitli değişkenlerin öğrenci notları üzerindeki etkilerini ortaya koymuştur. Araştırmada kullanılan veri bilimi ve makine öğrenmesi yöntemleri, öğrenci başarılarını etkileyen anahtar faktörleri belirlemede etkili olmuştur.

Aile durumu: Ailelerin sosyoekonomik durumu ve eğitim seviyesinin, öğrenci başarıları üzerinde önemli bir etkisi olduğu bulunmuştur. Özellikle, aile gelir düzeyi ve ebeveyn eğitim seviyesi, öğrencilerin akademik performansını artırmaktadır.

Kişisel özellikler: Öz-düzenleme becerileri ve motivasyon düzeylerinin, öğrenci başarıları üzerinde belirleyici olduğu tespit edilmiştir. Bu becerilere sahip öğrenciler, öğrenme süreçlerinde daha başarılı olabilirler.

Okul ve çevre: Okul olanakları ve çevresi, öğrencilerin eğitim deneyimlerini ve başarılarını etkileyen önemli faktörlerdir. Daha iyi okul olanakları ve düşük sınıf mevcudu, öğrencilere daha iyi öğrenme ortamları sunarak başarılarını artırmaktadır.

Öğretmen etkisi: Nitelikli öğretmenler ve etkili öğretim yöntemleri, öğrencilerin akademik başarılarını artırmada önemli rol oynamaktadır.

5.1.1. Belirli bulgular

Belirli bulgular, yapılan korelasyon, regresyon ve sınıflandırma analizlerine dayanmaktadır. Bu bulgular, öğrenci başarılarını etkileyen çok özel faktörleri detaylandırmaktadır.

Korelasyon analizi: Korelasyon analizleri, aile durumu, kişisel özellikler, okul çevresi ve öğretmen etkisinin öğrenci başarıları ile pozitif korelasyon gösterdiğini ortaya koymuştur.

Regresyon analizi: Ridge ve Lasso regresyon modelleri, öğrenci notlarını tahmin etmede etkili olmuş ve en önemli faktörlerin aile durumu ve kişisel özellikler olduğunu göstermiştir.

Sınıflandırma analizi: Logistic regression, random forests ve SVM algoritmaları, öğrenci başarılarını sınıflandırmada başarılı olmuştur. Ancak, sınıflandırma performansı-

nın artırılması için veri setinin dengelenmesi ve daha fazla özellik mühendisliği yapılması gerektiği belirlenmiştir.

5.2. Öneriler

Araştırma sonuçlarına dayalı olarak, eğitim politikaları ve gelecekteki çalışmalar için öneriler sunulacaktır.

5.2.1. Eğitim politikaları için öneriler

Aile destek programları: Ailelerin sosyoekonomik durumunu ve eğitim seviyesini artırmak için destek programları geliştirilmelidir. Ailelerin çocuklarının eğitimine daha fazla dahil olmalarını sağlamak, öğrenci başarılarını artırabilir.

Kişisel gelişim programları: Öğrencilerin öz-düzenleme becerilerini ve motivasyonlarını artıracak kişisel gelişim programları uygulanmalıdır. Bu programlar, öğrencilere etkili çalışma alışkanlıkları kazandırabilir.

Okul olanaklarının iyileştirilmesi: Okul olanakları ve çevresinin iyileştirilmesi, öğrencilere daha iyi öğrenme ortamları sunarak başarılarını artırabilir. Okulların fiziksel ve teknolojik altyapıları güçlendirilmelidir.

Öğretmen eğitimi: Öğretmenlerin niteliklerini artırmak için sürekli eğitim programları düzenlenmelidir. Nitelikli öğretmenler, öğrenci başarılarını doğrudan etkileyerek eğitim kalitesini yükseltebilir.

5.2.2. Gelecek çalışmalar için öneriler

Daha geniş veri setleri: Gelecek çalışmalarda, daha geniş ve çeşitli veri setleri kullanılarak analizlerin genellenebilirliği arttırılmalıdır. Farklı bölgelerden ve demografik özelliklerden öğrenci verileri toplanarak, sonuçların kapsamı genişletilebilir.

Yeni makine öğrenmesi yöntemleri: Farklı makine öğrenmesi algoritmaları ve derin öğrenme teknikleri kullanılarak, öğrenci başarılarının tahmin edilmesi ve sınıflandırılması daha da iyileştirilebilir. Bu yöntemler, daha yüksek doğruluk oranları sağlayabilir.

Kapsamlı çalışmalar: Öğrenci başarılarının uzun vadeli etkilerini incelemek için çalışmalar yapılmalıdır. Bu çalışmalar, öğrenci başarılarını etkileyen faktörlerin zaman içindeki değişimini ve uzun vadeli etkilerini ortaya koyabilir.

KAYNAKLAR

- Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(3), 1-21. <https://doi.org/10.1186/s41239-020-0177-7>
- Al-Barrak, M. A., & Al-Razgan, M. (2016). Predicting students' academic performance using data mining techniques: Naive Bayes and decision tree. *International Journal of Advanced Computer Science and Applications*, 7(5), 235-239. <https://doi.org/10.14569/IJACSA.2016.070533>
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 3-23.
- Dasu, T., & Johnson, T. (2003). *Exploratory Data Mining and Data Cleaning*. Wiley.
- Eccles, J. S., & Harold, R. D. (1993). Parent-school involvement during the early adolescent years. *Teachers College Record*, 94(3), 568-587.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.
- Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*, 12(1), 69-82.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer.

- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
- Kuhn, M., & Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Chapman and Hall/CRC.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Leithwood, K., & Jantzi, D. (2000). Transformational leadership: A guide to school leadership reform. *Educational Leadership*, 58(5), 78-82.
- Marzano, R. J. (2003). *What Works in Schools: Translating Research into Action*. ASCD.
- McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. O'Reilly.
- McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- Waskom, M. (2021). *Seaborn: Statistical Data Visualization*. O'Reilly Media.
- Yang, D., Chen, P., Wang, K., Li, Z., Zhang, C., & Huang, R. (2023). Parental involvement and student engagement: A review of the literature. *Sustainability*, 15(7), 5859. <https://doi.org/10.3390/su15075859>
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3-17.
- Zimmerman, B. J., & Schunk, D. H. (2011). *Handbook of Self-Regulation of Learning and Performance*. Routledge.

EKLER

Ek 1: Anket formu

SORULAR	YANITLAR
Sınıfınız	5-6-7-8
Şube	A-B-C-D-E-F
Adınız- Soyadınız	
Cinsiyet	KIZ/ERKEK
1) Ailemin Ekonomik durumu iyi denilebilir	EVET/HAYIR
2) Ailem bana ders çalışmam için baskı yapmaz	
3) Bana ait bir çalışma odam var	
4) Annem ve Babam ile birlikte yaşıyorum	
5) Ailem genellikle bana sevdiğimi hissettirir	
6) Okulda istediğim gibi beslenmemi yapabiliyorum	
7) Gün içinde 2 saatten daha az TV-Bilgisayar-Telefon ile meşgul olurum.	
8) Genellikle günde yarım saatten fazla kitap okurum	
9) Ders dışındaki konulara daha az ilgi gösteririm	
10) Genellikle Derse odaklanmakta zorlanmam.	
11) Derste öğretmenlerimin anlattıklarını anlamakta zorlanmam	
12) Ders çalıştığım zaman başarılı olabiliyorum.	
13) Sınavlarda çok heyecan yapmam.	
14) Ödevlerimi genellikle zamanında yaparım	
15) Çalışmalarım yeterince takdir edilir.	
16) Sorunlarımı konuşmaya, anlatmaya ve çözüm bulmaya ihtiyacım yok.	
17) Etkili ders çalışma yöntemlerini biliyorum.	
18) Herhangi bir sağlık sorunum yok.	
19) Çekingen, Utangaç bir yapım yok	
20) Başarılı olacağımı biliyorum kendime güveniyorum	
21) Okulda genellikle beni rahatsız eden öğrenciler yoktur.	
22) Sınıfımızın kalabalık olduğunu düşünmüyorum	
23) Okulumu ve arkadaşlarımı yeterince seviyorum	
24) Anlayamadığım konularda öğretmenime soru sormaktan çekinmem	
25) Öğretmenlerimiz genellikle dersi eğlenceli bir şekilde işler	
26) Dersle ilgili etkinlikler yeterince yapılır	

Ek 2: Ölçek izin formu

Tez için ölçek izin talebi

Gelen Kutusu x

27 Şub Sal 14:12

☆ 😊 ↩ ⋮

 **Ferit ÖZTEKİN** <feritoztekin@gmail.com>
Alıcı: ierdem000

Merhaba , Sayın İbrahim ERDEM

Ben, Batman Üniversitesi Bilgi Teknolojileri yüksek lisans öğrencisi Ferit Öztekin.

" Ortaokul düzeyinde Öğrenci Başarılarını etkileyen faktörlerin veri bilimi ve makine öğrenmesi yöntemleri ile analizi " adıyla tez hazırlamayı düşünüyorum. Sizin hazırladığınız " başarı ölçeği " anketine çalıştığınız kurumun web sayfasından eriştim. Hazırladığınız bu ölçeği tezimde kullanmak için izninizi rica ediyorum. Saygılarımla...

 **İbrahim Erdem** <ierdem000@gmail.com>
Alıcı: ben

Merhaba Ferit öztekin
"Başarı Ölçeği" anketini çalışmanızda kullanabilirsiniz.Başarılar dilerim.

27 Şub 2024 Sal 14:13 tarihinde Ferit ÖZTEKİN <feritoztekin@gmail.com> şunu yazdı:

...

Ek 3: Araştırma uygulama izni (Milli Eğitim Müdürlüğü)



T.C.
BATMAN VALİLİĞİ
İl Millî Eğitim Müdürlüğü



Sayı : E-71214596-604.02-100772368
Konu : Araştırma Uygulama İzni
(Ferit ÖZTEKİN)

22.04.2024

DAĞITIM YERLERİNE

İlgi: a) Ferit ÖZTEKİN'in 05.04.2024 tarihli ve 100275967 sayılı dilekçesi.
b) 19.04.2024 tarihli ve 100721505 sayılı Makam Oluru.

Batman Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgi Teknolojileri Anabilim Dalı Tezli Yüksek Lisans Programı Öğrencisi Ferit ÖZTEKİN ile danışmanı Prof. Dr. Necmettin SEZGİN'in ortak yürütecekleri "Öğrenci Başarılarını Etkileyen Faktörlerin Veri Bilimi ve Makine Öğrenmesi Yöntemleri ile Analizi" adlı çalışma kapsamında, İlimize bağlı Tüm Ortaokul Müdürlüklerindeki öğrencilere yönelik çalışma yapma isteği Müdürlüğümüz Araştırma İzinleri Değerlendirme Komisyonu tarafından incelenmiştir. Söz konusu çalışmanın yapılmasının uygun görüldüğüne dair ilgi (b) onay ilişkide sunulmuştur.

Bilgilerini ve gereğini rica ederim.

Mahmut KURTARAN
Vali a.
İl Millî Eğitim Müdürü

Eki:
1-Onay (1 Sayfa)
2-Yazı ve ekleri (13 sayfa)

Dağıtım:
5 İlçe Kaymakamlığına (İlçe MEM)
Batman Üniversitesi Lisansüstü Eğitim
Enstitüsü Müd.
Resmi Ortaokul Müd.

Adres :

Bu belge güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Adresi : <https://www.turkiye.gov.tr/meb-ebys>

Telefon No : 0 (488) 280 72 58

E-Posta: argo72@meb.gov.tr

Kep Adresi : meb@hs01.kep.tr

Bilgi için: Sevcan YAZICI

Unvan : Şef

İnternet Adresi: Faks:4882807278

Bu evrak güvenli elektronik imza ile imzalanmıştır. <https://evraksorgu.meb.gov.tr> adresinden b8e8-5fed-3dce-b101-b452 koda ile teyit edilebilir.

Ek 2 : Veri analizi için kullanılan kod satırları

Not Tahmini Kod satırları

```
# Gerekli kütüphaneleri içe aktaralım
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.linear_model import Lasso, Ridge
from sklearn.metrics import mean_squared_error, r2_score

# Veri setini yükleyelim
file_path = 'veri.csv' # Veri dosyanızın yolu
veri_seti = pd.read_csv(file_path)

# Tek sütundaki verileri ayrıştırma
veri_seti_split = veri_seti["Not;Sinif;Cinsiyet;A1;A2;A3;A4;K1;K2;K3;K4;K5;K6;K7;K8;K9;K10;K11;K12;R1;R2;O1;O2;O3;O4;T1;T2;T3"].str.split(';', expand=True)

# Sütun isimlerini belirleme
column_names = [
    "Not", "Sinif", "Cinsiyet", "A1", "A2", "A3", "A4", "K1", "K2", "K3", "K4", "K5",
    "K6", "K7", "K8", "K9", "K10", "K11", "K12", "R1", "R2", "O1", "O2", "O3", "O4",
    "T1", "T2", "T3"
]

# Sütun isimlerini dataframe'e atama
veri_seti_split.columns = column_names

# Sayısal verilerin tür dönüşümü
veri_seti_split["Not"] = veri_seti_split["Not"].astype(float)
numeric_columns = veri_seti_split.columns[1:]
veri_seti_split[numeric_columns] = veri_seti_split[numeric_columns].apply(pd.to_numeric)

# Bağımsız ve bağımlı değişkenleri ayırma
X = veri_seti_split.drop(columns=["Not"])
y = veri_seti_split["Not"]

# Veriyi eğitim ve test setlerine ayırma
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```

# Lasso ve Ridge regresyon modellerini tanımlama
lasso = Lasso(alpha=0.1)
ridge = Ridge(alpha=1.0)

# Modelleri eğitme
lasso.fit(X_train, y_train)
ridge.fit(X_train, y_train)

# Örnek bir öğrenci verisi oluşturma (bağımsız değişkenler)
# Diğer değişkenleri ortalama veya tipik bir değer olarak seçiyorum
example_student = X.mean().to_dict()

# "K5" değeri 1 olan ve 0 olan iki farklı öğrenci için kopyalar oluşturma
student_k5_1 = example_student.copy()
student_k5_1["K5"] = 1

student_k5_0 = example_student.copy()
student_k5_0["K5"] = 0

# Bu öğrenciler için tahminler yapma
student_k5_1_df = pd.DataFrame([student_k5_1])
student_k5_0_df = pd.DataFrame([student_k5_0])

# Lasso ve Ridge regresyon modelleriyle tahmin yapma
lasso_pred_k5_1 = lasso.predict(student_k5_1_df)
lasso_pred_k5_0 = lasso.predict(student_k5_0_df)

ridge_pred_k5_1 = ridge.predict(student_k5_1_df)
ridge_pred_k5_0 = ridge.predict(student_k5_0_df)

# Sonuçları yazdırma
print("Lasso Tahminleri:")
print(f"K5 = 1 -> Tahmin Edilen Not: {lasso_pred_k5_1[0]}")

```

```
print(f"K5 = 0 -> Tahmin Edilen Not: {lasso_pred_k5_0[0]}")
```

```
print("\nRidge Tahminleri:")
```

```
print(f"K5 = 1 -> Tahmin Edilen Not: {ridge_pred_k5_1[0]}")
```

```
print(f"K5 = 0 -> Tahmin Edilen Not: {ridge_pred_k5_0[0]}")
```

Yukarda istediğimiz Sütun isimlerini değiştirip Tahmin yapabiliriz.

Korelasyon Analizi Kod Satırları

```
import pandas as pd
```

Dosya yolu

```
file_path = 'C:\\Users\\ASUS\\Desktop\\ANALİZ KLASÖRÜ\\veri.csv'
```

```
# Alternatif olarak dosya yolunu şu şekilde de belirtebilirsiniz:
```

```
# file_path = 'C:/Users/ASUS/Desktop/CHATGPT KLASÖRÜ/veri.csv'
```

Veriyi oku

```
veri = pd.read_csv(file_path, delimiter=';')
```

Sütun isimlerini kontrol et

```
print(veri.head())
```

Korelasyon analizini yapalım

```
korelasyon = veri.corr()
```

```
not_korelasyon = korelasyon['Not'].sort_values(ascending=False)
```

Korelasyon sonuçlarını gösterelim

```
not_korelasyon
```

Sınıflandırma Analizi

```
import pandas as pd
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.preprocessing import StandardScaler
```

```
from sklearn.linear_model import LogisticRegression
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
from sklearn.svm import SVC
```

```
from sklearn.neighbors import KNeighborsClassifier
```

```
from sklearn.naive_bayes import GaussianNB
```

```
from sklearn.metrics import classification_report, confusion_matrix
```

```

# Veri setini yükle
file_path = 'C:\\Users\\ASUS\\Desktop\\ANALİZ KLASÖRÜ\\veri.csv'
data = pd.read_csv(file_path, delimiter=';')

# Sınıf sütununu çıkarma
data = data.drop(columns=['Sinif'])

# Not sütununu sınıflandırma (örneğin, 70'in üstü Başarılı, altı Başarısız)
data['Not'] = data['Not'].apply(lambda x: 1 if x >= 70 else 0)

# Bağımsız değişkenlerin (tüm kategoriler) ve bağımlı değişkenin (Not) ayrılması
X = data.drop(columns=['Not'])
y = data['Not']

# Eğitim ve test verisi olarak ayırma
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Özellik ölçekleme
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Logistic Regression modeli
log_reg = LogisticRegression(class_weight='balanced')
log_reg.fit(X_train_scaled, y_train)
y_pred_log_reg = log_reg.predict(X_test_scaled)

# Random Forest modeli
rf_clf = RandomForestClassifier(class_weight='balanced')
rf_clf.fit(X_train_scaled, y_train)
y_pred_rf = rf_clf.predict(X_test_scaled)

# Support Vector Machine (SVM) modeli
svm_clf = SVC(class_weight='balanced')
svm_clf.fit(X_train_scaled, y_train)
y_pred_svm = svm_clf.predict(X_test_scaled)

# K-Nearest Neighbors (KNN) modeli
knn_clf = KNeighborsClassifier()
knn_clf.fit(X_train_scaled, y_train)
y_pred_knn = knn_clf.predict(X_test_scaled)

```

```

# Naive Bayes modeli
nb_clf = GaussianNB()
nb_clf.fit(X_train_scaled, y_train)
y_pred_nb = nb_clf.predict(X_test_scaled)

# Değerlendirme

print("Logistic Regression:")
print(classification_report(y_test, y_pred_log_reg))
print(confusion_matrix(y_test, y_pred_log_reg))

print("\nRandom Forest:")
print(classification_report(y_test, y_pred_rf))
print(confusion_matrix(y_test, y_pred_rf))

print("\nSupport Vector Machine:")
print(classification_report(y_test, y_pred_svm))
print(confusion_matrix(y_test, y_pred_svm))

print("\nK-Nearest Neighbors (KNN):")
print(classification_report(y_test, y_pred_knn))
print(confusion_matrix(y_test, y_pred_knn))

print("\nNaive Bayes:")
print(classification_report(y_test, y_pred_nb))
print(confusion_matrix(y_test, y_pred_nb))

```