

	ADİYAMAN ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ DERGİSİ
	ISSN: 1308-9196 / e-ISSN:1308-7363
	Yıl: 18 Sayı: 50 Ağustos 2025
	Yayın Geliş Tarihi: 26.01.2025 Yayına Kabul Tarihi: 11.08.2025
	DOI Numarası: https://doi.org/10.14520/adyusbd.1627396 Makale Türü: Araştırma Makalesi/Research Article Atıf/Citation: Metin, S. (2025). Öğrencilerin Akademik Not Ortalamalarının Makine Öğrenmesi Yöntemleri ile Tahmini. <i>Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi</i> , (50), 886-910.

ÖĞRENCİLERİN AKADEMİK NOT ORTALAMALARININ MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNİ

Serkan METİN*

Öz

Bu çalışma, öğrencilerin akademik not ortalamalarını tahmin etmek için farklı makine öğrenmesi yöntemlerini kullanmıştır. Çalışmada, Gradient Boosting Regressor (GBR), Random Forest Regressor, Feedforward Neural Networks (FFNNs), XGBoost ve diğer modeller uygulanmış ve performansları MAE, RMSE, MAPE ve R^2 gibi metriklerle değerlendirilmiştir. Random Forest modeli, %100 R^2 ile en yüksek doğruluğu sağlamış ve en düşük hata oranlarına ulaşmıştır. Diğer modeller arasında Gradient Boosting ve XGBoost da yüksek doğruluk oranlarıyla öne çıkmıştır. Araştırma, öğrencilerin günlük çalışma saatleri, sosyal ve fiziksel aktiviteler ile stres seviyeleri gibi değişkenlerin akademik başarı üzerindeki etkilerini analiz etmiştir. Günlük çalışma saatleri, %73'lük pozitif korelasyonla başarı üzerindeki en güçlü etkiye sahip faktör olarak belirlenmiştir. Stres seviyesinin başarıya ölçülü bir şekilde pozitif etkisi olduğu görülürken, fiziksel aktivitelerin başarıyı az da olsa olumsuz etkilediği tespit edilmiştir

Anahtar Kelimeler: Akademik başarı, Makine öğrenmesi, Regresyon analizi, Tahmin



* Dr. Öğr. Üyesi, Malatya Turgut Özal Üniversitesi, Sosyal ve Beşerî Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, serkan.metin@ozal.edu.tr, Malatya/Türkiye

PREDICTION OF STUDENTS' ACADEMIC GRADE POINT AVERAGES USING MACHINE LEARNING METHODS

Abstract

This study utilized various machine learning methods to predict students' academic grade point averages. The models applied in the study included Gradient Boosting Regressor (GBR), Random Forest Regressor, Feedforward Neural Networks (FFNNs), XGBoost, among others, and their performances were evaluated using metrics such as MAE, RMSE, MAPE, and R^2 . The Random Forest model achieved the highest accuracy with an R^2 of 100% and the lowest error rates. Among the other models, Gradient Boosting and XGBoost also demonstrated high accuracy levels. The research analyzed the impact of variables such as daily study hours, social and physical activities, and stress levels on academic success. Daily study hours were identified as the most influential factor, with a 73% positive correlation to academic performance. While stress levels showed a moderate positive effect on success, physical activities were found to have a slightly negative impact.

Keywords: Academic achievement, Machine learning, Regression analysis, Prediction

1. GİRİŞ

Eğitim, bilgi aktarımı, beceri ve karakter gelişimiyle ilgilenen bir süreçtir (Parkavi vd., 2024). Bu nedenle, bireylerin öğrenme süreçlerini daha etkili bir şekilde anlamak ve geliştirmek, yalnızca bireysel başarıyı artırmakla kalmaz, aynı zamanda toplumların genel refahına da katkıda bulunur. Günümüzde teknolojinin hızla ilerlemesi ve yaşamın her alanının ayrılmaz bir parçası olması (Acem vd., 2024), eğitim sistemlerinde de köklü değişimlere neden olmuş ve eğitim alanında veri madenciliği, yapay zeka, akıllı öğretim sistemleri gibi uygulamaların yaygın olarak kullanılmasına olanak sağlamıştır (Chen vd., 2020). Bu teknolojiler, eğitim alanında bireylerin öğrenme süreçlerini geliştirmek ve

daha etkili bir eğitim sistemi oluşturmak için çok önemli fırsatlar sunmaktadır. Özellikle veri madenciliği ve yapay zeka gibi ileri teknolojiler, eğitim süreçlerini analiz etme, öğrenci performansını öngörme ve öğretim yöntemlerini iyileştirme gibi alanlarda çığır açan çözümler sunarak eğitimde dönüşümün temel unsurları haline gelmiştir (Ouyang ve Jiao, 2021). Ayrıca, bu tür veri analitiği araçları, öğretmenlerin sınıf yönetiminde daha bilinçli kararlar almasını da desteklemektedir. Eğitim sistemlerinde biriken büyük miktarda veri, geleneksel yöntemlerle analiz edilmek için çok karmaşık ve büyük olduğundan dolayı bu verileri anlamlı hale getirmek için veri madenciliği teknikleri kullanılmaktadır. Bu yöntemler ile anlamlı hale getirilen bilgiler, eğitim sistemlerinin iyileştirilmesine de katkıda bulunmaktadır. Aynı şekilde, veri madenciliği araçları, öğrencilerin öğrenme güçlüklerini öngörmek ve bu sorunları çözmek içinde etkili çözümler sunabilmektedir.

Yapay zeka, veri madenciliği tekniklerini daha ileri bir düzeye taşıyarak veriler üzerinde daha ayrıntılı analizler yapılmasına olanak sağlamaktadır. Yapay zeka sistemleri, eğitim alanında öğrencilerin öğrenme alışkanlıklarını, akademik başarılarını ve hatta davranışlarını analiz ederek daha kişiselleştirilmiş öğrenme deneyimleri sunmak için gerekli veri analiz sonuçlarının elde edilmesini sağlar. Derin öğrenme algoritmaları, daha karmaşık veri analiz işlemlerinde kullanılmaktadır. Yapay sinir ağları temelinde geliştirilen bu algoritmalar, öğrencilerin sınav performanslarını, ders katılım oranlarını ve öğrenme alışkanlıklarını analiz ederek yüksek doğruluk oranlarına sahip tahminler yapabilmektedir. Derin öğrenme modelleri, bir öğrencinin geçmiş akademik verilerini analiz ederek öğrencinin gelecek dönemlerde hangi derslerde zorlanabileceği gibi zorlu tahmin işlemlerini yerine getirebilme kabiliyetlerine de sahiptirler. Bu tür karmaşık analizler, yöneticilerin öğrenci başarılarını artırmak için stratejik karar vermelerinde önemli bir yer tutmaktadır. Ayrıca, derin öğrenme algoritmaları, öğrencilerin öğrenme süreçlerindeki bireysel

farklılıklarını da daha iyi anlamak için de güçlü bir araçtır. Eğitime bu modellerin bir diğer önemli katkısı, öğrencilerin uzun vadeli kariyer planlamalarına rehberlik etmektir. Bu tür analizler, sadece bireysel başarıyı artırmakla kalmaz, aynı zamanda toplumların genel eğitim ve iş gücü kalitesini iyileştirilmesine çok önemli katkılar sağlarlar.

Eğitim alanının farklı dallarında makine öğrenme yöntemleri kullanılarak eğitim alanında birçok farklı çalışmalar yapılmıştır. Bu alanda yayınlanan makaleler araştırılarak benzer çalışmalar incelenmiştir.

Ashenafi vd., (2015), çoklu doğrusal regresyon modeli kullanarak ders notlarını tahmin etmek için öğrencilerin tartışma sorularına verdikleri cevapları derecelendirerek özellikler çıkarmışlardır. Gök (2017), 24 sorudan oluşan ve ortaöğretim 6,7 ve 8. sınıf öğrencilerini kapsayan çalışmada çok sınıflı makine öğrenmesi yöntemini kullanmıştır. Okubo vd., (2017) eğitim sistemlerinde depolanan günlük verileri ile tekrarlayan sinir ağı modeli ile öğrencilerin final notlarını tahmin etmek için çalışma yapmıştır. Ahmad ve Shahzadi (2018), makine öğrenimine dayalı bir model ile öğrencilerin öğrenme becerilerini, çalışma alışkanlıklarını ve akademik etkileşim özelliklerini kullanarak %85'lik bir sınıflandırma doğruluğuna sahip bir tahmin algoritması oluşturmuşlardır. Wang vd., (2018), Geleneksel Makine Öğrenme Yöntemlerine Dayalı Not Tahmini modeli ile öğrencilerin cinsiyetini, eğitim geçmişi verilerini, konu geçmişini ve çevrimiçi öğrenme davranışı verilerini birleştirerek bunları bir karar ağacına girdi olarak verip akademik performansı tahmin etmişlerdir.

Fernandes vd., (2019), öğrencilerin akademik başarısını Gradient Boosting Machine (GBM) temelli sınıflandırma algoritması kullanarak tahmin etmişlerdir. Xu vd., (2019), üniversite öğrencilerinin internet kullanım davranışları ile akademik performansları arasındaki ilişkiyi belirlemek için makine öğrenimi yöntemleri kullanmışlar ve yüksek bir doğruluk oranıyla tahmin etmeyi

başarmışlardır. Qu vd., (2019), kısmi zamanlı davranışların ve davranış kalıplarının ağırlıklarını ayarlamak için dikkat mekanizmasını kullanarak bir not tahmin modeli oluşturdular.

Rastrollo-Guerrero ve diğerleri (2020), öğrenci başarısını tahmin etme konusunda 64 yeni makale üzerinde nitel bir araştırma çalışması gerçekleştirerek kullanılan teknikleri özetlemişlerdir. Salloum vd., (2020), öğrenci başarısıyla ilgili yaptıkları çalışmada makine öğrenimi tekniklerinin daha başarılı olduğunu göstermişlerdir. Rebai vd., (2020), okulların akademik performansını etkileyen faktörler arasındaki ilişkiyi göstermek için makine öğrenmesi modelleri önermişlerdir. Kalyani vd., (2020), derin öğrenme yöntemlerinden CNN algoritmasını öğrenci başarı durumunu tahmin etmek için kullanmışlardır. Rodríguez-Hernandez vd., (2021), Kolombiya'daki özel ve kamu üniversitelerinden her iki cinsiyetten 162.030 öğrenciyi içeren veritabanı üzerinden öğrencilerin akademik performansını yapay sinir ağlarını kullanarak tahmin etmişlerdir. Sekeroglu vd., (2021), 2010 ile 2020 yıllarını kapsayan öğrenci performansı tahmin çalışmaları ile ilgili 297 makaleyi inceleyerek öğrenci başarısını belirlemede kullanılan yöntemleri sınıflandırmışlardır. Martins vd., (2021), öğrencilerin akademik başarılarını tahmin etmek için çeşitli makine öğrenimi modellerini karşılaştırmış ve rastgele orman sınıflandırıcısının tahmin doğruluğunun diğer modellerden daha iyi performans gösterdiğini belirlemişlerdir. Ahmed vd., (2021) üniversite öğrencilerinin final sınavlarındaki performanslarını tahmin etmek için GBDT algoritmasının kullanılmasını önermişlerdir. Yağcı (2022), 1.854 öğrenciden oluşan bir veri kümesine k-en yakın komşu, rastgele orman ve lojistik regresyon modellerini uygulayarak öğrencilerin final sınav notlarını tahmin etmiştir.

Batool vd., (2023), öğrenci performansı tahmini üzerine 260 çalışmanın kapsamlı bir incelemesini gerçekleştirmişlerdir. Yapay sinir ağları ve rastgele orman

sınıflandırıcılarının en yaygın kullanılan veri madenciliği araçları olduğunu tespit etmişlerdir. Zilyas ve Atınç (2023), 519 farklı öğrenci üzerine yaptıkları çalışmada Türkçe notu üzerinden eğitim başarısının tahmini için makine öğrenmesi yöntemlerini kullanmışlardır. Adak ve Duralioğlu (2023), öğrencilerin bir önceki sınav notlarını kullanarak yapılacak diğer sınavdaki başarılarını tahmin etmek için çok çıkışlı regresyonu destekleyen doğrusal regresyon, k-en yakın komşu ve karar ağacı algoritmaları kullanılmışlardır. Pallathadka vd., (2023) tarafından yapılan bu çalışmada, öğrencilerin benzer derslerdeki önceki başarılarını analiz ederek ders performansını tahmin etmek için veri madenciliği tekniklerini kullanmışlardır. Wenjin vd., (2024), mevcut öğrenci bilgi verilerine dayalı PyTorch tensörlerini kullanarak eğitim ve test veri kümeleri oluşturan bir Evrişimsel Sinir Ağı (CNN) tahmin modeli oluşturdular. Yauri vd., (2024), makine öğrenimi algoritmasının yardımıyla öğrencilerin başarısızlık ve başarı oranlarını tahmin eden bir model geliştirdiler. Çalışmada, Nijerya'nın Adamawa Eyaletindeki 720 öğrenciye ait bilgiler kullanılmıştır. Geliştirilen sinir ağı modeli %97,36'lık doğruluk oranına ulaşmıştır. Hussain vd. (2024) çalışmalarında derin öğrenme yöntemlerinden Levenberg Marquardt Algoritmasını kullanarak %88.6'lık doğruluk oranı ile öğrencilerin Final notlarını tahmin etmişlerdir.

2. YÖNTEM

2.1. Araştırma Modeli

Bu çalışmada, bağımsız değişkenlerin hedef değişken üzerindeki etkilerini anlamak ve tahmin gücünü artırmak için yapay zeka ve makine öğrenimi yöntemleri kapsamlı bir şekilde uygulanmıştır. Araştırma modeli, Gradient Boosting Regressor ve Yapay Sinir Ağları (FFNN) modelleme yöntemlerini kullanarak değişkenler arasındaki karmaşık ilişkileri analiz etmeye odaklanmaktadır. Bu yöntemler, yüksek doğruluk ve genelleştirme kapasitesi sunmalarıyla öne çıkmaktadır.

2.2. Evren ve Örneklem

Araştırmının çalışma grubu, Google Form anketi aracılığıyla 01.08.2023 – 01.05.2024 tarihleri arasında Hindistan’daki öğrencilerin yaşam tarzları hakkında bilgileri içeren 2.000 öğrenci verisinden oluşmaktadır. Veriseti açık olarak yayınlanan Kaggle-Dataset adresinden indirilmiştir. Daily Lifestyle and Academic Performance of Students başlıklı veri setine ait alan adları Tablo 1’de verilmiştir.

Tablo 1. Veriseti Alan Bilgileri

Orijinal Alan Adları	Türkçeleştirilmiş Alan Adları
Student_ID	Öğrenci İd
Study_Hours_Per_Day	Günlük Çalışma Saatleri
Extracurricular_Hours_Per_Day	Ders Dışı Günlük Aktivite Saati
Sleep_Hours_Per_Day	Günlük Uyku Saati
Social_Hours_Per_Day	Günlük Sosyal Faaliyet Saati
Physical_Activity_Hours_Per_Day	Günlük Fiziksel Aktivite Saati
GPA	Not Ortalaması
Stress_Level	Stres Seviyesi

2.3. Feedforward Neural Networks (FFNNs)

Feedforward Neural Networks (FFNNs), yapay sinir ağlarının temel bir yapısıdır ve genellikle regresyon ve sınıflandırma problemlerinde kullanılır. Bu ağ, veri akışını giriş katmanından gizli katmanlara ve ardından çıkış katmanına yönlendiren ileri yönlü bir bilgi akışı modeline dayanır. Her nöron, kendisine gelen girişleri ağırlıklandırır, bir aktivasyon fonksiyonuyla işler ve sonucu bir sonraki katmana iletir. Gizli katmanlardaki nöron sayısı ve aktivasyon fonksiyonları, modelin karmaşıklığını ve doğrusal olmayan ilişkileri öğrenme kapasitesini belirler. FFNN'ler, özellikle çok katmanlı yapılarıyla, karmaşık ve doğrusal olmayan veri ilişkilerini öğrenebilir.

2.4. Gradient Boosting Regressor (GBR)

Gradient Boosting Regressor, regresyon problemleri için kullanılan güçlü bir makine öğrenmesi modelidir. Model, ensemble öğrenme tekniğini kullanarak bir dizi zayıf tahminciyi bir araya getirir. Bu süreçte, her bir zayıf model, önceki modelin hatalarından öğrenir ve adım adım daha iyi tahminler yapar. Eğitim sırasında, her ardışık model, hataları optimize etmek için bir kayıp fonksiyonunu minimize eder. GBR, hem doğrusal hem de doğrusal olmayan veri ilişkilerini öğrenme kapasitesine sahiptir. Model, doğrusal olmayan ilişkileri öğrenme yeteneğiyle, karmaşık veri kümelerinde yüksek performans gösterir ve genellikle manuel özellik mühendisliği ihtiyacını azaltır.

2.5. Random Forest Regressor

Random Forest Regressor, bir ensemble (topluluk) öğrenme yöntemidir ve regresyon problemlerinde tahmin doğruluğunu artırmak için kullanılır. Bu model, birden fazla karar ağacı oluşturur ve her bir ağacın tahminini birleştirerek sonuç üretir. Tahmin, regresyon problemlerinde genellikle bu ağaçların tahmin değerlerinin ortalaması alınarak hesaplanır. Model, bootstrap örnekleme yöntemiyle farklı veri alt kümeleri üzerinde çalışır ve rastgele özellik seçimi yaparak her bir ağaç için farklı kararlar oluşturur. Bu yaklaşım, overfitting'i azaltır ve modelin genelleme yeteneğini artırır. Bu modelin avantajlarından biri, doğrusal olmayan ilişkileri kolaylıkla öğrenebilmesidir. Aynı zamanda, veri içindeki eksikliklere ve aykırı değerlere karşı dayanıklıdır.

2.6. XGBoost

XGBoost (Extreme Gradient Boosting), gradient boosting algoritmasının hızlı ve optimize edilmiş bir versiyonudur. Makine öğrenmesi yarışmalarında ve büyük veri projelerinde sıkça tercih edilir. XGBoost, gradyan artırımı ağaç oluşturma

yöntemiyle çalışır ve hem hız hem de doğruluk açısından etkileyici performans sunar. Model, veri setindeki hataları minimize etmek için ardışık olarak birden fazla karar ağacı oluşturur ve her adımda hataları düzelterek tahmin doğruluğunu artırır.

2.7. LightGBM

LightGBM (Light Gradient Boosting Machine), gradient boosting algoritmasına dayanan hızlı, verimli ve güçlü bir makine öğrenmesi kütüphanesidir. Büyük veri setleri ve yüksek boyutlu veriler üzerinde çalışmak için optimize edilmiştir. LightGBM'nin temel avantajları arasında hız, düşük bellek kullanımı ve paralel işlem yapabilme yeteneği yer alır. Bu model, histogram tabanlı bir algoritma kullanır; özellik değerlerini bir histogramda gruplandırarak işlemleri hızlandırır. LightGBM, büyük veri setlerinde etkileyici bir hız sağlar ve eksik verilerle çalışabilme yeteneğine sahiptir.

2.8. Support Vector Regressor (SVR)

Support Vector Regressor (SVR), doğrusal olmayan ilişkileri modelleyebilen güçlü bir regresyon yöntemidir. SVR, Support Vector Machines'in (SVM) regresyon problemlerine uyarlanmış halidir. Modelin temel prensibi, tahmin edilen değerler ile gerçek değerler arasındaki hatayı minimize ederken, tahmin doğruluğunu artıracak bir hiper düzlemi bulmaktır. SVR, doğrusal olmayan veri yapılarını modellemek için genellikle çekirdek (kernel) fonksiyonları kullanır. Çekirdek fonksiyonları arasında doğrusal (linear), polinomsal (polynomial) ve RBF (Radial Basis Function) gibi seçenekler yer alır.

2.9. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN), denetimli öğrenme algoritmalarından biridir ve regresyon veya sınıflandırma problemlerinde kullanılır. Bu algoritma, bir veri

noktasının sonucunu tahmin ederken, ona en yakın k komşusunun etiketlerine dayanır. KNN'nin temel prensibi, benzer veri noktalarının genellikle benzer sonuçlara sahip olacağı varsayımına dayanır. Tahmin yapılırken, her bir veri noktasının uzaklığı hesaplanır ve belirli bir k değeri seçilir.

3. BULGULAR

Verilere ilişkin en küçük (min), en yüksek (max) değerler, ortalama (mean) ve standart sapma (std), 25% (Birinci Çeyrek, Q1), 50% (İkinci Çeyrek, Q2) ve 75% (Üçüncü Çeyrek, Q3) değerleri Tablo 2’de verilmiştir.

Tablo 2. Veri Setine Ait İstatistiksel Özellikler

Değişkenler	mean	std	min	25%	50%	75%	max
Günlük Çalışma Saatleri	7.48	1.42	5.00	6.30	7.40	8.70	10.00
Ders Dışı Günlük Aktivite Saati	1.99	1.16	0.00	1.00	2.00	3.00	4.00
Günlük Uyku Saati	7.50	1.46	5.00	6.20	7.50	8.80	10.00
Günlük Sosyal Faaliyet Saati	2.70	1.69	0.00	1.20	2.60	4.10	6.00
Günlük Fiziksel Aktivite Saati	4.33	2.51	0.00	2.40	4.10	6.10	13.00
Not Ortalaması	3.12	0.30	2.24	2.90	3.11	3.33	4.00
Stres Seviyesi	2.37	0.73	1.00	2.00	3.00	3.00	3.00

İstatistiksel analizler incelendiğinde, öğrencilerin günlük çalışma saatlerinin ortalaması 7.48 saat olup, 5 ile 10 saat arasında değişmektedir. Çalışma saatlerindeki standart sapmanın 1.42 olması, bireyler arasında orta derecede bir farklılık olduğunu ortaya koymaktadır. Ders dışı günlük aktivite saatleri ortalama 1.99 saat olarak hesaplanmıştır. Bu süre 0 ile 4 saat arasında değişmektedir. Günlük uyku saatleri ortalama 7.5 saat ile sağlıklı bir düzeyde gözlemlenmiştir. Uyku süresi minimum 5, maksimum 10 saat arasında değişmekte ve bireylerin uyku düzenlerinde standart sapma 1.46 olarak hesaplanmıştır. Günlük sosyal faaliyetlere ayrılan süre ortalama 2.7 saat olup, 0 ila 6 saat arasında

değişmektedir. Günlük fiziksel aktivite süresi ise ortalama 4.32 saat olarak hesaplanmıştır ve 0 ile 13 saat arasında değişmektedir. Not ortalamaları 2.24 ile 4 arasında değişmekte olup, ortalama 3.11 olarak kaydedilmiştir. Standart sapma 0.29 olduğundan, öğrencilerin başarı düzeyleri arasında çok büyük farklar olmadığı görülmektedir. Stres seviyesi ise genellikle orta ve yüksek düzeylerde seyretmektedir. Ortalama stres seviyesi 2.36 olup, minimum 1 ve maksimum 3 olarak gözlemlenmiştir. Bu durum, öğrencilerin genel olarak yüksek stres altında olduklarını göstermektedir.

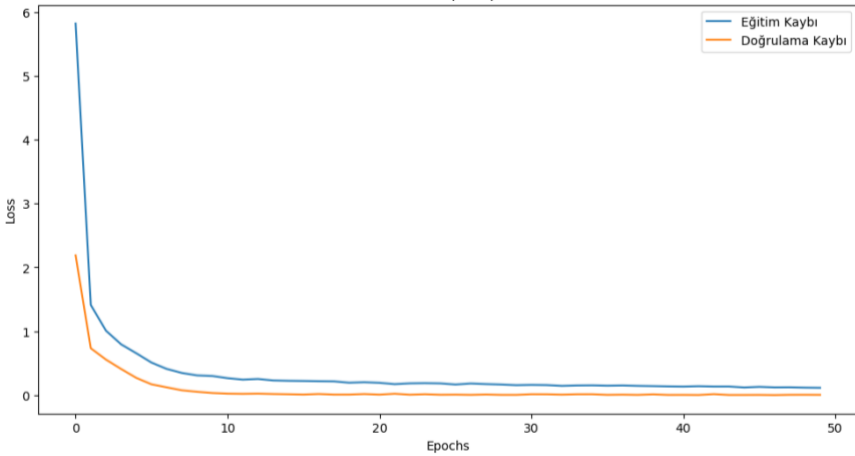
Araştırmanın temelini oluşturan veri seti, öğrenci yaşam tarzını ve akademik performansı yansıtan birden fazla değişkenden oluşmaktadır. Veri setinin analizine başlamadan önce eksik değerler kontrol edilmiş ve mean yöntemiyle doldurulmuştur. Aykırı değerlerin tespiti ve temizlenmesi amacıyla Çeyrekler Arası Aralık (Interquartile Range - IQR) yöntemi kullanılmıştır. Bu süreçte, veriler, alt ve üst limitlerin dışında kalan uç değerlerden arındırılmıştır. Modelleme öncesinde bağımsız değişkenler olarak belirlenen veriler, Standartlaştırma (StandardScaler) yöntemiyle normalize edilmiştir. Veri setindeki değişkenler farklı ölçeklerde ölçülmüştür. Örneğin, Günlük Çalışma Saatleri değerleri 5 ile 10 arasında değişirken, Not Ortalaması değerleri 2.24 ile 4.00 arasında değişmektedir. Tablo 3'te normalizasyon sonucunda elde edilen verilere ait özet bilgiler verilmiştir.

Tablo 3. Veri Setinin Normalizasyon Sonrası Özeti

Günlük Çalışma Saatleri	Ders Dışı Günlük Aktivite Saati	Günlük Uyku Saati	Günlük Sosyal Faaliyet Saati	Günlük Fiziksel Aktivite Saati	Not Ortalaması	Stres Seviyesi
-0.404	1.566	0.821	0.057	-1.006	-0.422	-0.503
-1.528	1.307	0.341	0.886	-0.528	-1.226	-1.878
-1.669	1.653	1.163	-0.891	0.108	-1.494	-1.878
-0.685	0.095	-0.206	-0.595	0.864	-0.790	-0.503

0.438	-1.203	-0.686	-0.299	0.904	1.320	0.872
-1.037	0.095	0.341	-1.424	1.302	-0.891	-0.503
0.368	-1.116	-1.507	1.774	-0.011	-0.120	0.872
0.649	-0.165	-1.302	0.175	0.347	0.281	0.872
-1.599	1.393	-0.822	0.767	0.227	-0.991	-1.878
0.157	-1.116	1.574	1.064	-1.205	-1.192	-0.503
1.562	1.393	0.341	-0.121	-1.642	1.052	0.872
-0.404	-0.770	1.095	-0.358	0.188	-0.489	-0.503

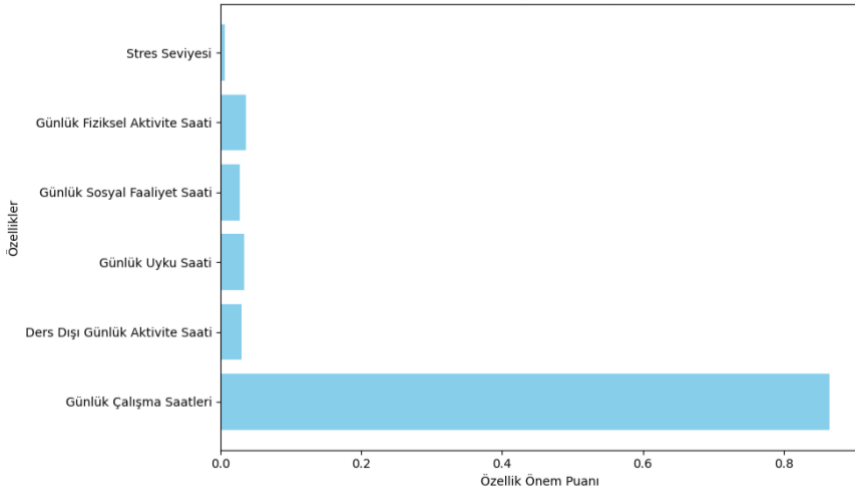
Normalizasyon işlemlerinden sonra modellerin uygulamasına geçilmiştir. Bu aşamada verilerin %80'i eğitim ve %20'si test için ayrılmıştır. FFNN modeli, iki gizli katmandan oluşur. İlk katman 64, ikinci katman 32 nörona sahiptir. Aktivasyon fonksiyonu olarak ReLU kullanılmıştır ve aşırı öğrenmeyi önlemek için dropout katmanları eklenmiştir. Şekil 1, Feedforward Neural Networks modelinin eğitim sürecindeki kayıp fonksiyonunun (loss) epoch'lara göre değişimini göstermektedir.



Şekil 1. Feedforward Neural Networks modelinin eğitim sürecindeki kayıp değişimi

Şekil 1 incelendiğinde, Eğitim ve doğrulama kaybı başlangıçta yüksek değerlerden başlayarak hızla azalmış ve epoch sayısı arttıkça sabit bir seviyeye ulaşmıştır. Eğitim ve doğrulama kayıplarının birbirine yakın seyretmesi, modelin genelleme yeteneğinin iyi olduğunu ve aşırı öğrenme (overfitting) yaşanmadığını göstermektedir.

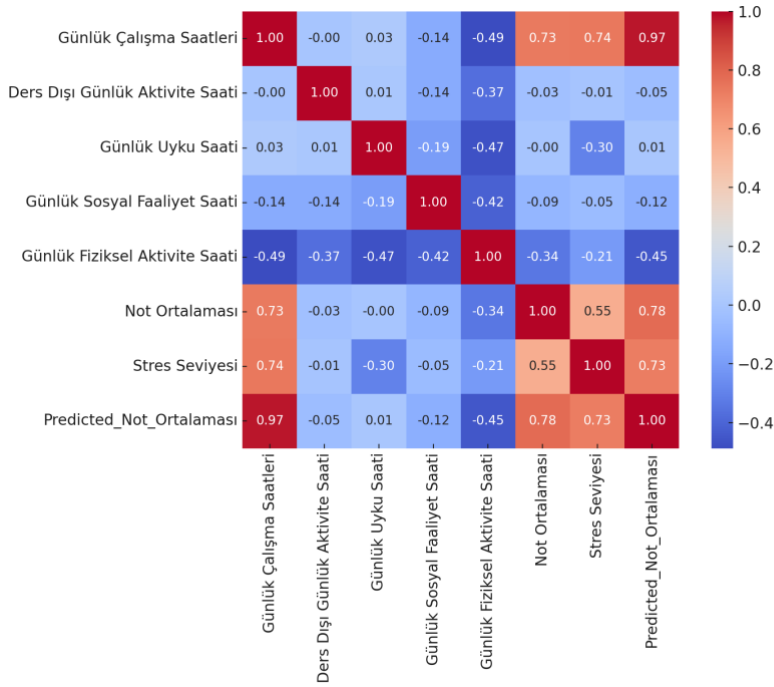
Gradient Boosting Regressor modelinde, gradyan artırma algoritmasını kullanılmıştır. Gradient Boosting Regressor modelinde bağımsız değişkenlerin bağımlı değişken olan Not Ortalaması üzerindeki etkilerinin önem dereceleri Şekil 2’de verilmiştir.



Şekil 2. Gradient Boosting Regressor Özellik Önem Dereceleri

Şekil 2 incelendiğinde öğrencinin başarısını etkileyen en kritik faktörün Günlük Çalışma Saatleri olduğu; fiziksel aktivite, uyku ve sosyal faaliyetlerin daha düşük etkilerinin olmasına rağmen, dengeli bir yaşamın önemini de dikkate alınması gerektiğini göstermektedir.

Veriseti içerisindeki değişkenler arasındaki ilişkileri analiz etmek için Şekil 3’deki korelasyon matrisi oluşturulmuştur.



Şekil 3. Korelasyon Matrisi

Korelasyon matrisi, değişkenler arasındaki ilişkileri net bir şekilde ortaya koymaktadır. En dikkat çekici bulgu, Günlük Çalışma Saatleri ile Not Ortalaması arasındaki 0.73 gibi yüksek bir pozitif korelasyondur. Bu, öğrencilerin çalışma sürelerinin akademik başarı üzerindeki güçlü etkisini doğrular. Ayrıca, Stres Seviyesi ile Not Ortalaması arasında da 0.55’lik pozitif bir ilişki bulunmaktadır. Bu durum, yüksek stres seviyesinin bazı durumlarda öğrencilerin daha fazla çalışmasına ve başarıya ulaşmasına katkıda bulunabileceğini düşündürmektedir. Günlük Fiziksel Aktivite Saati ile Not Ortalaması arasında -0.34’lük bir negatif korelasyon bulunmaktadır. Bu, daha fazla fiziksel aktiviteye zaman ayırmanın akademik başarıda bir miktar azalmaya neden olabileceğini gösterir. Günlük Uyku Saati ve diğer değişkenler arasındaki düşük korelasyonlar, uyku düzeninin başarı üzerindeki etkisinin sınırlı olabileceğini işaret eder. Son olarak, tahmin edilen değerler ile gerçek Not Ortalaması arasında 0.97’lik yüksek bir korelasyon

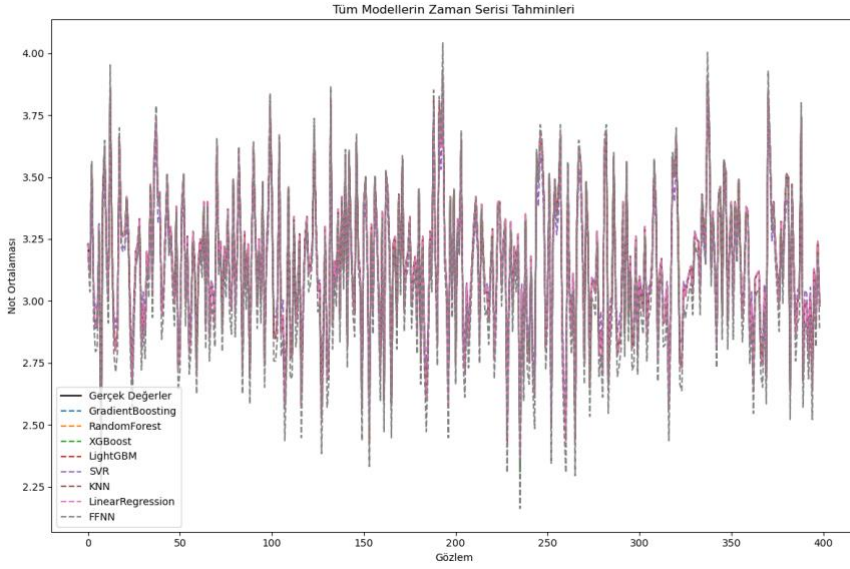
bulunması, modelin doğruluğunu kanıtlar. Bu, Gradient Boosting modelinin başarılı bir şekilde kullanılabileceğini göstermektedir.

Modellerin başarılarını ölçmek için MAE, RMSE, MAPE ve R^2 metrikleri hesaplanmıştır. Bu metrikler, modellerin tahmin performansını değerlendirmek için yaygın olarak kullanılmaktadır. Modellere ait eğitim metrik değerleri Tablo 4'te verilmiştir.

Tablo 4. Modellerin Tahmin Sonuçlarının Metrik Ölçümleri

Metric	KNN	FFNNs	SVR	GBR	LightGBM	Random Forest	XGBoost
MAE	0,0005	0,0419	0,0653	0,0012	0,0019	0,0002	0,0008
RMSE	0,0029	0,0437	0,0714	0,0024	0,0107	0,0014	0,0028
MAPE	0,0159	1,3777	2,1017	0,0376	0,0584	0,0068	0,0261
R^2	0,9999	0,9793	0,9447	0,9999	0,9988	1,0000	0,9999

Tabloda 4'deki metrikler, farklı modellerin tahmin performansını karşılaştırmaktadır. Random Forest modeli, tüm metriklerde en iyi sonuçları elde etmiştir. Bu model, özellikle Ortalama Mutlak Hata (MAE) ve Kök Ortalama Kare Hata (RMSE) değerlerinde en düşük hata oranlarını 0.0002 ve 0.0014 sağlayarak, tahminlerde oldukça yüksek bir doğruluğa ulaşmıştır. Aynı zamanda Ortalama Yüzde Hata (MAPE) değeri de 0.0068 ile diğer modellere göre en düşük seviyededir. KNN ve XGBoost modelleri, özellikle hata metriklerinde düşük değerler elde ederek güçlü alternatifler olarak öne çıkmaktadır. KNN'nin MAPE değeri 0.0159, XGBoost'un ise 0.0261 ile oldukça düşüktür ve her iki model de $R^2 = 0.9999$ ile veri uyumunda başarılı olmuştur. Buna karşılık, FFNNs ve SVR modelleri, diğer modellere göre daha yüksek hata oranlarına sahiptir. Özellikle MAPE değerleri sırasıyla 1.3777 ve 2.1017 olması bu modellerin tahminlerde daha fazla hata yaptığını göstermektedir. Modellere ait tahmin ve gerçek değer grafiği Şekil 4'de verilmiştir.



Şekil 4. Tüm Modellerin Tahmin

Modellerin tahmin ettiği değerlere ait veri setinden rastgele seçilmiş özet veriler Tablo 5'te verilmiştir.

Tablo 5. Modellerin Tahmin Sonuçlarının Sayısal Değerleri

Gerçek Değerler	FFNNs	KNN	SVR	LightGBM	XGBoost	Random Forest	GBR
3.11	3.02	3.11	3.11	3.11	3.11	3.11	3.11
3.55	3.44	3.55	3.45	3.55	3.55	3.55	3.55
3.01	2.93	3.01	3.05	3.01	3.01	3.01	3.01
2.89	2.82	2.89	2.97	2.89	2.89	2.89	2.89
2.90	2.83	2.90	2.97	2.90	2.90	2.90	2.90
3.31	3.21	3.31	3.23	3.31	3.31	3.31	3.31
2.42	2.28	2.43	2.46	2.43	2.41	2.42	2.41
3.44	3.34	3.44	3.34	3.44	3.44	3.44	3.44
3.62	3.50	3.62	3.53	3.62	3.62	3.62	3.62
3.20	3.10	3.20	3.16	3.20	3.20	3.20	3.20
3.00	2.92	3.00	3.04	3.00	3.00	3.00	3.00
3.86	3.72	3.85	3.81	3.81	3.85	3.86	3.85

3.22	3.12	3.22	3.17	3.22	3.22	3.22	3.22
2.88	2.81	2.88	2.96	2.88	2.88	2.88	2.88
2.80	2.74	2.80	2.89	2.80	2.80	2.80	2.80
2.89	2.82	2.89	2.97	2.89	2.89	2.89	2.89
3.66	3.54	3.67	3.58	3.66	3.66	3.66	3.66
3.29	3.19	3.29	3.22	3.29	3.29	3.29	3.29
3.26	3.16	3.26	3.20	3.26	3.26	3.26	3.26

4. TARTIŞMA ve SONUÇ

Eğitimde teknolojik yapıların kullanılması, bireysel öğrenme süreçlerini iyileştirmenin yanı sıra, genel eğitim sistemlerini optimize etmek içinde çok önemlidir. Bu teknolojiler, öğrenci başarısını tahmin etmek, öğrenme süreçlerini kişiselleştirmek ve eğitim politikalarını geliştirmek için etkili araçlar sunmaktadır.

Bu çalışma, öğrencilerin akademik not ortalamalarını çeşitli makine öğrenmesi yöntemleri kullanarak tahmin etmeyi amaçlamaktadır. Araştırma kapsamında kullanılan modeller arasında Feedforward Neural Networks (FFNNs), Gradient Boosting Regressor (GBR), Random Forest Regressor, XGBoost (Extreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), Support Vector Regressor (SVR) ve K-Nearest Neighbors (KNN) algoritmaları yer almıştır.

Araştırma, öğrencilerin günlük çalışma saatleri, sosyal aktiviteler, fiziksel aktiviteler ve stres seviyelerinin akademik başarı üzerindeki etkilerini incelemiş ve bu değişkenlerin akademik başarıyı anlamlı bir şekilde etkilediğini ortaya koymuştur. Özellikle günlük çalışma saatleri ile not ortalaması arasındaki güçlü pozitif korelasyon (%73), öğrencilerin çalışma sürelerinin başarı üzerinde kritik bir etkiye sahip olduğunu doğrulamaktadır. Öğrencilerin akademik not ortalamalarını tahmin etmek için çeşitli kullanılan makine öğrenmesi modelleri arasında en iyi performansı, tüm metriklerde üstünlük sağlayan ve %100 R² ile en yüksek doğruluk oranına ulaşan Random Forest modeli göstermiştir.

Gradient Boosting Regressor (GBR) ve XGBoost modelleri de yüksek doğruluk oranlarıyla sırasıyla %99.99 ve %99.99 R^2 dikkat çekmiş; özellikle doğrusal olmayan ilişkileri etkili bir şekilde öğrenme kabiliyetleriyle öne çıkmıştır. Feedforward Neural Networks (FFNNs) modeli, doğrusal olmayan ilişkileri öğrenme kapasitesine rağmen, diğer modellere kıyasla daha yüksek hata oranlarıyla sınırlı bir performans sergilemiştir. Sonuçlar, öğrencilerin günlük çalışma saatleri ile akademik başarı arasında güçlü bir pozitif ilişki olduğunu ve stres seviyesinin başarı üzerinde ölçülü bir pozitif etki yarattığını ortaya koymaktadır.

Bu çalışmada elde edilen bulgular, öğrencilerin akademik başarılarını tahmin etme konusunda önemli içgörüler sunmakla birlikte bazı kısıtlamalar da taşımaktadır. İlk olarak, çalışma verileri belirli bir coğrafi bölge ve kültürel bağlamdan (Hindistan'daki bir öğrenci veri seti) elde edilmiştir. Bu durum, sonuçların Türkiye gibi farklı eğitim sistemlerine sahip ülkelerde genellenebilirliğini sınırlandırabilir. İkinci olarak, kullanılan veri seti, öğrencilerin akademik başarılarını etkileyebilecek diğer psikolojik, sosyo-ekonomik ve çevresel faktörleri kapsamamaktadır. Örneğin, aile desteği, öğretmen-öğrenci etkileşimi veya okulların altyapısı gibi unsurlar analiz dışında kalmıştır. Ayrıca, makine öğrenmesi modelleri uygulanırken yalnızca belirli algoritmalar kullanılmış; daha geniş bir model yelpazesi veya farklı hiperparametre ayarları çalışmaya dahil edilmemiştir. Son olarak, çalışma gelecekteki dönemlere yönelik tahminlerde bulunmak yerine, mevcut veriler üzerinde tahmin doğruluğunu artırmaya odaklanmıştır. Son olarak, farklı seviyelerdeki öğrencilerin başarılarını tahmin etmeye yönelik bir sistem geliştirilerek, ilk öğretimden üniversite düzeyine kadar eğitim kademelerinde makine öğrenmesi uygulamalarının etkinliği araştırılabilir. Bu tür çalışmalar, daha etkili eğitim politikalarının geliştirilmesine ve bireysel öğrenme süreçlerinin iyileştirilmesine katkıda bulunacaktır.

Çıkar Çatışması Bildirimi:

Yazar, bu makalenin araştırılması, yazarlığı ve yayınlanmasına ilişkin herhangi bir potansiyel çıkar çatışması olmadığını beyan eder.

Destek/Finansman Bilgileri:

Yazar, bu makalenin araştırılması, yazarlığı ve yayınlanması için herhangi bir finansal destek almadığını beyan eder.

Etik Kurul Kararı:

Bu araştırma için etik kurul kararına ihtiyaç olmadığını beyan eder.

KAYNAKÇA

- Acem, Y., Arslantaş, K., Bişirici, M., & Erdoğan, K. (2024). Öğretmenlerin Eğitimde Yapay Zeka Kullanımına Yönelik Tutumlarının İncelenmesi. *International Journal of New Trends in Education and Social Sciences*, 1(2), 12-23 <https://doi.org/10.5281/zenodo.11113077>
- Adak, M.F., & Duralioğlu, Ö. (2023). Makine Öğrenmesi Yöntemleri Kullanılarak Öğrencilerin Kazanım Bilgileri ile Sınavlardaki Başarı Durumunun Tahmini. *Zeki Sistemler Teori ve Uygulamaları Dergisi* 6(1), 43-51 doi: 10.38016/jista.1183353
- Ahmad, Z., & Shahzadi, E. (2018). Prediction of students' academic performance using artificial neural network. *Bulletin of Education and Research*, 40(3), 157–164.
- Ahmed, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2021). Predicting university's studentsperformance based on machine learning techniques. *IEEE International Conference onAutomatic Control &*

- Intelligent Systems (I2CACIS)*. doi:10.1109/I2CACIS52118.2021.9495862
- Ashenafi, M. M., Riccardi, G., & Ronchetti, M. (2015). Predicting students' final exam scores from their courseactivities. *Frontiers in Education Conference (Fie)*, 372–380. DOI: <https://doi.org/10.1109/FIE.2015.7344081>
- Batool, S., Rashid, J., Nisar, M.W., Kim, J., Kwon, H., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, 28(1), 905–971. <https://doi.org/10.1007/s10639-022-11152-y>
- Chen, X., Xie, H., & Hwang, G. J. (2020). A multi-perspective study on artificial intelligence in education: Grants, conferences, journals, software tools, institutions, and researchers. *Computers & Education: Artificial Intelligence*, 1, <https://doi.org/10.1016/j.caeai.2020.100005>
- Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Erven, V.G. (2019). Educational data mining : Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, 335–343. <https://doi.org/10.1016/j.jbusr>
- Gök, M. (2017). Makine Öğrenmesi Yöntemleri İle Akademik Başarının Tahmin Edilmesi. *Gazi Üniversitesi Fen Bilimleri Dergisi*, 5(3): 139-148.
- Hussain, M.M., Akbar, S., Hassan, S.A., Aziz, M.W., & Urooj, F. (2024). Prediction of Student's Academic Performance through Data Mining Approach. *Journal of Informatics and Web Engineering*, 3(1) <https://doi.org/10.33093/jiwe.2024.3.1.16>
- Kalyani, B. S., Harisha, D., Ramyakrishna, V., & Manne, S. (2020). Evaluation of students performance usingneural networks. *Intelligent Computing, Information and Control Systems, Iccics 2019 [J]*. 1039:499–505.
- Martins, M.V., Tolledo, D., Machado, J., Baptista, L.M., & Realinho, V. (2021). Early prediction of student's performance in higher education: A case

- study. Paper presented at the World Conference on Information Systems and Technologies, pp. 166–175. https://doi.org/10.1007/978-3-030-72657-7_16
- Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. A. (2017). Neural network approach for students' performance prediction. *Seventh International Learning Analytics & Knowledge Conference*. 598–599. <https://doi.org/10.1145/3027385.3029479>
- Parkavi, R., Karthikeyan, P., Sujitha, S., & Abdullah, A.S. (2024). Enhancing Educational Assessment: Predicting and Visualizing Student Performance using EDA and Machine Learning Techniques. *Journal of Engineering Education Transformations*, 37, 240–245
- Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-Lopez, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials Today: Proceedings*, 80, 3782–3785. <https://doi.org/10.1016/j.matpr.2021.07.382>
- Qu, S. J., Li, K., Wu, B., Zhang, S. H., & Wang, Y. C. (2019). Predicting student achievement based on temporal learning behavior in MOOCs. *Appl. Sciences-Basel*, 9(24). <https://doi.org/10.3390/app9245539>
- Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020. <https://doi.org/10.1016/j.caeai.2021.100020>
- Rastrollo-Guerrero, J., Gómez-Pulido, J.A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied Sciences*, 10(3), 1–16. <https://doi.org/10.3390/app10031042>
- Rebai, S., Ben Yahia, F., & Essid, H. (2020). A graphically based machine learning approach to predict secondary schools performance in Tunisia. *Socio-*

- Economic Planning Sciences*,
<https://doi.org/10.1016/j.seps.2019.06.009>
- Rodríguez-Hernandez , C.F., Musso , M., Kyndt , E., & Cascallar E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence* 2- 100018
<https://doi.org/10.1016/j.caeai.2021.100018>
- Salloum, S.A., Alshurideh, M., Elnagar, A., Shaalan, K., & Tolba, F.M. (2020). Mining in educational data: Review and future directions. *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, 1153, 92–102. https://doi.org/10.1007/978-3-030-44289-7_9
- Sekeroglu, B., Abiyev, R., Ilhan, A., Arslan, M., & Idoko, J.B. (2021). Systematic Literature Review on Machine Learning and Student Performance Prediction: Critical Gaps and Possible Remedies. *Applied Sciences*, 11(22), 10907. <https://doi.org/10.3390/app112210907>
- Wang, G. H., Zhang, J., & Fu, G. S. (2018) Predicting student behaviors and performance in online learning using decision tree. *2018 Seventh International Conference of Educational Innovation through Technology*, 214–219. Doi: 10.1109/EITT.2018.00050
- Xu, W., Gao, C., & Yang, H. (2024). Research on Student Performance Prediction Based on Deep Learning. *Research Square*.
<https://doi.org/10.21203/rs.3.rs-4967448/v1>
- Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, 166–173.
<https://doi.org/10.1016/j.chb.2019.04.015>
- Yagci, M. (2022). Educational data mining: Prediction of students' academic

performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. z

Yauri, R.A., Suru, H.U., Afrifa, J., & Moses, H.G. (2024). A Machine Learning Approach in Predicting Student's Academic Performance Using Artificial Neural Network. *Journal of Computational and Cognitive Engineering*, 3(2), 203–212. <https://doi.org/10.47852/bonviewJCCE3202470>

Zilyas, D., & Yılmaz, A. (2023). Makine Öğrenmesi Yöntemleri İle Eğitim Başarısının Tahmini Modeli. *Dicle University Journal of Engineering*, 14(3), 437-447 <https://doi.org/10.24012/dumf.1322273>

EXTENDED ABSTRACT

Introduction

The rapid advancement of technology has transformed educational systems worldwide, introducing novel methods for evaluating and improving academic outcomes. In this context, machine learning (ML) has emerged as a powerful tool for predicting student performance, analyzing learning behaviors, and enhancing educational policies. This study explores the use of machine learning algorithms to predict students' academic grade point averages (GPAs) based on various lifestyle and educational factors. The research identifies key determinants of academic success, such as daily study hours, extracurricular activities, sleep patterns, social activities, physical activities, and stress levels. By leveraging advanced predictive models, this study aims to provide actionable insights that can enhance individual learning experiences and optimize broader educational strategies.

Method

The research employed a dataset comprising responses from 2,000 students collected through a structured Google Form survey between August 2023 and May 2024. The dataset, sourced from Kaggle, included variables such as daily study hours, extracurricular activity hours, sleep hours, social activity hours, physical activity hours, stress levels, and GPA. Before analysis, the data underwent preprocessing steps, including missing value imputation using mean

substitution, outlier removal with the Interquartile Range (IQR) method, and standardization via StandardScaler to normalize variable scales. A range of machine learning algorithms was implemented to predict student GPAs. These models included Random Forest Regressor, Gradient Boosting Regressor (GBR), XGBoost, Feedforward Neural Networks (FFNNs), LightGBM, Support Vector Regressor (SVR), and K-Nearest Neighbors (KNN). The models were evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2 . The dataset was divided into training and testing subsets, with 80% allocated for model training and 20% for evaluation. The FFNN model utilized two hidden layers with 64 and 32 neurons, respectively, and employed ReLU as the activation function. Dropout layers were added to prevent overfitting. Gradient Boosting models relied on ensemble learning techniques, combining weak learners to improve prediction accuracy. Random Forest, known for its robustness, aggregated predictions from multiple decision trees. Advanced techniques such as XGBoost and LightGBM optimized gradient boosting for faster computations and higher accuracy.

Findings (Results)

The analysis revealed that Random Forest achieved the highest accuracy among all models, with an R^2 of 1.0000 and the lowest error rates (MAE: 0.0002, RMSE: 0.0014). This performance underscores the model's ability to effectively capture complex relationships in the data while maintaining robustness. Gradient Boosting Regressor and XGBoost also performed exceptionally well, achieving R^2 values of 0.9999, with low MAE and RMSE scores, demonstrating their capability to model non-linear relationships effectively. The correlation analysis indicated that daily study hours were the most significant predictor of GPA, showing a strong positive correlation of 0.73. Stress levels exhibited a moderate positive correlation (0.55) with GPA, suggesting that manageable stress may serve as a motivational factor. In contrast, physical activity hours showed a negative correlation (-0.34), implying that higher physical activity levels could slightly detract from academic performance. Other variables, such as sleep hours and social activity hours, exhibited weaker correlations with GPA, indicating a limited direct influence. Model comparison highlighted the superior generalization capabilities of Random Forest and Gradient Boosting models, which maintained high accuracy across training and testing data. In contrast, Feedforward Neural Networks and SVR displayed higher error rates and lower accuracy, suggesting these models were less effective for the dataset's characteristics.

Conclusion and Discussion

This study highlights the effectiveness of machine learning in predicting academic performance, offering valuable insights for educators, policymakers, and students. The superior performance of Random Forest and Gradient Boosting models demonstrates their potential for educational applications, particularly in identifying key factors influencing student success. The findings underscore the importance of daily study hours as the most critical determinant of academic achievement, emphasizing the need for structured study schedules.

The observed positive correlation between stress levels and GPA suggests that moderate stress can enhance performance, likely by motivating students to engage more with their studies. However, excessive stress could have adverse effects, warranting strategies to help students manage stress effectively. The negative impact of physical activities on GPA underscores the importance of balancing academic and extracurricular commitments to optimize performance.

Despite its strengths, the study has limitations. The dataset is specific to a particular geographical and cultural context (India), which may limit the generalizability of the findings to other regions. Moreover, the analysis excluded psychological and socio-economic variables, such as family support and school infrastructure, which could significantly influence academic outcomes. Future research should incorporate more diverse datasets and additional predictors to enhance model robustness and applicability across different educational contexts.

Further exploration of time-series analyses and longitudinal studies could provide deeper insights into how academic performance evolves over time. Additionally, integrating these models into personalized learning platforms could enable real-time predictions and tailored interventions, ultimately improving individual learning outcomes. By addressing these avenues, future studies can contribute to the development of innovative and data-driven educational practices that support students' academic and personal growth.