



Student Success Prediction with Machine Learning: Modeling and Performance Analysis of Factors Affecting Academic Success

Volkan Göreke^{1,a,*}

¹ Sivas Teknik Bilimler Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, Sivas, 58140, Türkiye

*Corresponding author

Research Article

History

Received: 27/10/2024

Accepted: 08/11/2024

Copyright



This work is licensed under
Creative Commons Attribution 4.0
International License

ABSTRACT

This study aims to analyze the factors affecting students' academic success and develop prediction models based on these factors. Many features such as "Studying Hours," "Absence," and "Family Income" that affect students' performance during the education process were evaluated. In the study, student success predictions were made using machine learning models such as Random Forest, Ridge, and Deep Neural Networks and the performance of these models were compared. The importance levels of the features were calculated with the Random Forest algorithm and the most important features were determined. As a result of the analysis, the effect of removing low-importance features from the model to increase prediction performance was also examined. The findings of the study contribute to educators making data-driven decisions and developing strategies that can increase students' success.

Keywords: Artificial Intelligence, Academic performance, Regression, Important features, Root mean square error

Makine Öğrenimi ile Öğrenci Başarı Tahmini: Akademik Başarıyı Etkileyen Faktörlerin Modellenmesi ve Performans Analizi

Süreç

Geliş: 27/10/2024

Kabul: 08/11/2024

Öz

Bu çalışma, öğrencilerin akademik başarısını etkileyen faktörleri analiz etmeyi ve bu faktörlere dayalı tahmin modelleri geliştirmeyi amaçlamaktadır. Öğrencilerin eğitim sürecindeki performansını etkileyen "Çalışma Saatleri", "Devamsızlık" ve "Aile Geliri" gibi birçok özellik değerlendirilmiştir. Çalışmada, Rastgele Orman, Ridge ve Derin Sinir Ağları gibi makine öğrenimi modelleri kullanılarak öğrenci başarı tahminleri yapılmış ve bu modellerin performansı karşılaştırılmıştır. Özniteliklerin önem düzeyleri Rastgele Orman algoritması ile hesaplanmış ve en önemli özellikler belirlenmiştir. Analiz sonucunda, düşük öneme sahip özelliklerin modelden çıkarılmasının tahmin performansını artırma etkisi de incelenmiştir. Çalışmanın bulguları, eğitimcilerin veri odaklı kararlar almasına ve öğrencilerin başarısını artıracak stratejiler geliştirmesine katkıda bulunmaktadır.

Anahtar Kelimeler: Yapay Zeka, Akademik performans, Regresyon, Önemli özellikler, Kök ortalama karesel hata

^a v goreke@cumhuriyet.edu.tr

0000-0002-2418-8373 |

How to Cite: Göreke V (2024) Student Success Prediction with Machine Learning: Modeling and Performance Analysis of Factors Affecting Academic Success, Journal of Engineering Faculty, 2(2): 137-143

Giriş

Bir ülkenin en değerli kaynağı olan insan gücünü, topluma katkı sunacak bireylere dönüştürmek için eğitim büyük önem taşır. Eleştirel düşünebilen, yaratıcı ve sorun çözüme becerisi olan bireyler yetiştirmek adına eğitim sistemlerinde köklü değişiklikler yapılmalıdır. Başarı, belirli bir hedefe ulaşma ve istenilen sonuca varma olarak tanımlanır [1]. Eğitimde başarı ise, programın hedeflediği davranışları sergileyen öğrencilerle ilişkilidir ve genellikle derslerde gösterilen performans, notlar ya da test puanları ile değerlendirilir. "Akademik başarı" terimi, öğrencilerin bilgi ve beceri kazanımlarını ifade eder. Bu başarı, öğrencilerin mesleki ve toplumsal yaşama donanımlı olarak hazırlanmasına katkı sağladığı için aileler ve çevre açısından da büyük önem taşır [2]. Araştırmalar, öğrencilerin akademik başarılarını etkileyen çeşitli faktörler olduğunu göstermektedir. Bu faktörler genellikle öğrencinin içsel özellikleri (zekâ, sağlık durumu, ilgi düzeyi, kaygı) ve çevresel koşullarla ilişkilidir [3]. Bu koşullar ders çalışma saatleri, kaynaklara erişim, ders dışı etkinlikler, uyku saatleri, önceki sınav notları, motivasyon düzeyi, internet erişim, özel ders, aile geliri, öğretmen kalitesi, okul türü, akran etkisi, fiziksel etkinlik, öğrenme engelleri, ebeveyn eğitim seviyesi, evden uzaklık ve cinsiyet gibi parametreler ile belirlenmektedir [4]. Boyd, bu parametrelerden motivasyon, bir bireyin belirli bir hedefe ulaşmak veya bir görevi başarmak için harekete geçmesini sağlayan içsel veya dışsal bir itici güçtür. İnsanları belirli davranışlara yönelten, istek ve çaba gösterme sürecidir. Motivasyon, kişisel ilgi ve ihtiyaçlar, hedefler, beklentiler veya ödüller gibi çeşitli faktörlerle şekillenebilir ve yoğunluğu kişinin amaca ulaşmadaki kararlılığını etkileyebilir. Yüksek motivasyon bireyin akademik başarısında etkili olduğu yapılan akademik çalışmalar ile ortaya konmuştur [4]. Ayrıca anne baba eğitim düzeyi ve ekonomik durum gibi ailesel faktörlerin de başarı üzerinde etkisi olan ve yapılan akademik araştırmalar ile kanıtlanmış bir parametredir [3]. Linda ve arkadaşlarının yaptıkları çalışma internet kullanımının akademik başarıyı etkileyen bir parametre olduğu ortaya konmuştur. [5]. Askerc ve Gasparic, yaptıkları çalışma ile özel derse yönelik birçok eleştiriye rağmen, kesinlikle bazı olumlu etkileri olduğunu, örneğin bireye daha fazla odaklanma ve onların özel öğrenme özellikleri ve ihtiyaçlarının dikkate alınması, daha ilgi çekici öğretim, daha hızlı geri bildirim, daha iyi öğrenci ilerlemesi gibi avantajlar sağladığını tespit etmişlerdir [6]. Akademik performansla ilişkili iç mekanizmaların analizinde, fiziksel aktivite eksikliğinin veya hareketsizliğin şu anda en büyük halk sağlığı sorunlarından biri haline geldiği bilinmektedir [7]. Bu anlamda, fiziksel aktivite bilişsel süreç iyileştirmesiyle ilişkilendirilmiştir, çünkü dopamin veya serotonin gibi nörotransmitterler spor pratiği sırasında salgılanır ve bunlar konsantrasyonu ve motivasyonu iyileştirir [8]. Pedagojik model veya öğretmenin talepleri gibi eğitim sürecindeki bir takım parametreler de öğrencilerin akademik başarısında etkin olan faktörler arasındadır [9].

Makine öğrenmesi, verilerden öğrenme ve tahmin yapma yeteneği sunan bir yapay zeka alanıdır. Regresyon algoritmaları, bu alanda sıkça kullanılan tekniklerdir ve

özellikle sürekli değerlerin tahmin edilmesi amacıyla geliştirilmiştir. Regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi modelleyerek, bir bağımlı değişkenin (tahmin edilmek istenen değer) bir veya daha fazla bağımsız değişken (öznitelik) tarafından nasıl etkilendiğini anlamaya çalışır [10]. Basit regresyon, en temel regresyon türü olup yalnızca bir bağımsız değişken ile bir bağımlı değişken arasındaki ilişkiyi inceler. Bu modelde veriler bir doğru ile temsil edilir ve en küçük kareler yöntemi kullanılarak doğru en iyi şekilde uydurulur. Çoklu regresyon ise birden fazla bağımsız değişkenin etkisini incelemek amacıyla kullanılır; bu tür regresyon, karmaşık ilişkileri modellemek için daha fazla esneklik sağlar ve daha doğru tahminler yapılmasına olanak tanır. Lojistik regresyon, gerçek değerler yerine kategorik sonuçların tahmini için kullanılır ve bir olayın olasılığını tahmin etmek amacıyla sigmoid fonksiyonunu kullanarak sonuçları 0 ve 1 arasında bir değerle temsil eder. Ridge ve Lasso regresyon ise çoklu regresyon modelinin aşırı uyum (overfitting) sorununu çözmek için kullanılan düzenleme yöntemleridir. Ridge regresyon, modelin karmaşıklığını azaltmak için L2 normunu kullanırken, Lasso regresyonu L1 normunu kullanarak bazı katsayıları sıfıra indirir, bu da değişken seçimini kolaylaştırır. Karar ağaçları, regresyon işlemlerinde sürekli bir bağımlı değişkenin tahmini için kullanılan popüler bir yöntemdir. Bu yöntem, verileri dallara ayırarak ve her dalda belirli bir karar kılarak çalışır. Her yaprak düğümü, tahmin edilen değeri temsil ederken, iç düğümler ise verinin özelliklerine dayalı olarak bölünme kriterlerini belirler. Karar ağaçları, basit yapıları ve yorumlanabilirlikleri nedeniyle regresyon problemlerinde etkili bir seçenek sunar, ancak aşırı uyum riskine karşı dikkatli olunmalıdır.

Bu çalışmanın amacı, öğrenci performansını etkileyen faktörleri kullanarak öğrenci notlarını tahmin etmek için yapay zeka tekniklerine dayalı bir yöntem geliştirmektir. Bu yöntem sayesinde eğitimciler öğrencilerin başarısını tahmin edebilecek ve başarı tahmini düşük çıkan öğrenci profili üzerinde başarıyı artıracak yönde yaklaşımlar gösterebilecektir. Ayrıca bu çalışma ile eğitimciler, öğrencilerin okullarda daha iyi performans sağlamaları için öğrenci üzerinde hangi faktörlere odaklanmaları gerektiğine karar verebilecektir.

Literatür

Hussain ve Khan [10], öğrencilerin notlarını tahmin etmek amacıyla makine öğrenimini kullanan bir sistem geliştirmiştir. Bu çalışmada, Pakistan eğitim kurulundan elde edilen veriler kullanılmıştır. Yöntem; veri ön işleme, öznitelik seçimi, not tahmini için bir regresyon modelinin eğitilmesi ve not sınıflandırması için bir karar ağacı sınıflandırıcısı adımlarını içermektedir. Öznitelik seçimi için Genetik Algoritma kullanılsa da bu yöntem, büyük veri kümelerinde yüksek hesaplama karmaşıklığına neden olmaktadır. Kullandıkları farklı Karar Ağacı (Decision-Tree -DT) 8.23, K-En Yakın Komşular (K-Nearest Neighbors- K-NN 27.66, Genetik Algoritma (GA) + Karar Ağacı 5.34 GA + K-NN 24.31 Kök Ortalama Kare Hatası (Root Mean Square Error -RMSE) değerleri elde etmişlerdir. Jayaprakash ve arkadaşları [11],

öğrenci performans tahmini için geliştirilmiş gelişmiş bir Rastgele Orman sınıflandırıcısı önermişlerdir. Yazarların önerdikleri model güçlü bir doğruluğa sahip olmuştur, ancak en iyi performansı elde etmek için öznelik seçimi ve hiperparametre ayarlaması çok önemlidir. Kumar ve arkadaşları [12], Naive Bayes, Rastgele Orman ve Lojistik Regresyon algoritmalarını birleştirerek çok seviyeli bir topluluk öğrenme algoritması (NB-RF-LR-SEMod) geliştirmiş ve bir veri kümesinde %88,3 doğruluğa ulaşmıştır. Topluluk teknikleri, tek başına kullanılan modellerden daha yüksek hassasiyet sağlayabilmektedir; ancak bu yöntemler oldukça karmaşıktır ve öğrencilerin özenle seçilmesini gerektirir. Ofori ve arkadaşları [13], öğrencilerin performans tahmininde makine öğreniminin kullanımını inceleyen bir literatür çalışması yapmıştır. Çalışmada, erken tahminin öğrenme sürecini geliştirmede kritik bir rol oynadığı ve en iyi modeli seçmenin zorlukları vurgulanmıştır. Bulguları, sosyoekonomik faktörlerin tahmin doğruluğunun düşük olmasının nedeni olabileceğini göstermektedir. Asselman ve arkadaşları [14], öğrencilerin performansını tahmin etmek için bir topluluk öğrenme yöntemi olan XGBoost algoritmasının etkinliğini incelemiştir. Beckham ve arkadaşları yaptıkları çalışma ile öğrenci performansına etki eden faktörleri çeşitli makine öğrenme teknikleri kullanarak incelemişlerdir. 12 nöronlu çok katmanlı algılayıcı ağ ile 4.32 RMSE değeri elde etmişlerdir [4].

Materyal ve Metod

Bu çalışmada Kaggle sitesinden elde edilen veri seti kullanılmıştır. Bu veri seti, sınavlarda öğrenci performansını etkileyen çeşitli faktörlere ilişkin kapsamlı bir genel bakış sunar. Çalışma alışkanlıkları, katılım, ebeveyn katılımı ve

akademik başarıyı etkileyen diğer yönler hakkında bilgi içerir. Bu çalışmada başarıyı etkileyen faktörler öznelik olarak nitelendirilmiştir. Kullanılan veri setine ait öznelik isimleri ve açıklamaları aşağıdaki Çizelge 1 ile verilmiştir. Veri seti 6607 örnek ve 17 öznelik ve sınav notu değerlerini içermektedir. Bu çalışmada önerilen yöntem Visual Studio 2022 Community IDE üzerinde Python programlama dili kullanılarak geliştirilmiştir. Ayrıca bu çalışmada bağımsız değişkenler olan özneliklerin bağımlı değişken sınav notu üzerindeki etki dereceleri de araştırılmıştır. Bu araştırma sonucunda öznelik seçme işlemi gerçekleştirilmiştir. Rastgele orman (Random Forest, RF), gereksiz ve alakasız öznelikleri ortadan kaldırma ve makine öğrenimi sınıflandırıcılarının performansını etkili bir şekilde iyileştirmek için en uygun öznelikleri belirleme avantajına sahip en yaygın kullanılan algoritmalarından biridir [16]. Öznelik seçme aşamasında Rastgele Orman yöntemi kullanılmıştır. Çalışmada ilk önce yapay zeka sistemleri üzerinde öznelik seçmeden tahmin uygulaması yapılmıştır. Daha sonra öznelik seçme işlemi uygulanarak sistemler üzerinde tahmin işlemi tekrarlanmış ve sonuçlar karşılaştırılmıştır.

Çalışmada performans kriterini ölçmek için Kök Ortalama Kare Hatası (Root Mean Square Error - RMSE) metriği kullanılmıştır. Bu metrik, bir modelin tahmin ettiği değerler ile gerçek değerler arasındaki hata miktarını ölçmek için kullanılan yaygın bir performans değerlendirme metriğidir. Özellikle regresyon problemlerinde, tahminlerin doğruluğunu değerlendirmek amacıyla kullanılır. RMSE değeri, hatanın büyüklüğünü (ölçekli hata) gösterir ve genellikle "Ortalama Karesel Hata"nın (MSE) karekökü alınarak elde edilir. RMSE'nin düşük olması, modelin daha doğru tahminler yaptığını gösterir.

Çizelge 1. Öğrenci başarısını etkileyen öznelikler

Table 1. Features affecting student success

| Öznelik | Açıklama |
|-------------------|---|
| Çalışma_Saati | Haftada harcanan çalışma saati sayısı. |
| Devamsızlık | Katılılan derslerin yüzdesi. |
| Veli_Katılım | Velilerin öğrencinin eğitimine katılım düzeyi (Düşük, Orta, Yüksek). |
| Kaynak_Erişimi | Eğitim kaynaklarının erişilebilirliği (Düşük, Orta, Yüksek). |
| Dış_Activity | Kısa süreli aktivitelerdeki katılım (Evet, Hayır). |
| Uyku_Saati | Gece başına ortalama uyku saati sayısı. |
| Önceki_Puanlar | Önceki sınavlardan alınan puanlar. |
| Motivasyon_Düzeyi | Öğrencinin motivasyon düzeyi (Düşük, Orta, Yüksek). |
| İnternet_Erişimi | İnternet erişiminin varlığı (Evet, Hayır). |
| Özel_Ders_Sayısı | Ayda katılılan özel ders sayısı. |
| Aile_Geliri | Aile gelir seviyesi (Düşük, Orta, Yüksek). |
| Öğretmen_Kalitesi | Öğretmenlerin kalitesi (Düşük, Orta, Yüksek). |
| Okul_Türü | Öğrenim görülen okulun türü (Devlet, Özel). |
| Akran_Etkisi | Akranların akademik başarı üzerindeki etkisi (Olumlu, Nötr, Olumsuz). |
| Fiziksel_Aktivite | Haftada ortalama fiziksel aktivite saati sayısı. |
| Öğrenme_Güçlüğü | Öğrenme güçlüğü varlığı (Evet, Hayır). |
| Cinsiyet | Öğrencinin cinsiyeti (Erkek, Kadın). |
| Sınav_Puanı | Final sınav puanı. |

Bir veri setindeki n sayıda örneğin gerçek değerleri y_i ve model tarafından tahmin edilen değerleri \hat{y}_i olmak üzere RMSE aşağıdaki Eşitlik 1 ile hesaplanır.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

Ayrıca bu çalışmada, derin sinir ağları (DNN) kullanarak bir regresyon modeli oluşturulmuş ve eğitilmiştir. Bu şamada öncelikle veri seti yüklenir, varsa eksik veriler temizlenir ve öznelikler (X) ile etiketler (y) ayrılır. Model, veri setindeki bağımlı değişkeni (bu çalışmada, öğrenci notları) sürekli bir değer olarak tahmin etmeyi amaçlamaktadır. Veriler, eğitim ve test setlerine ayrıldıktan sonra ölçeklendirilir. DNN modeli, iki gizli katmanda "ReLU" aktivasyonu kullanarak öğrenir; son

katmanda ise sürekli tahminler için doğrusal (linear) aktivasyon kullanılır. Model "Adam" optimizasyon algoritmasıyla derlenir ve 50 epok boyunca eğitimden geçer. Modelin performansı, Ortalama Kare Hatası (MSE), Kök Ortalama Kare Hatası (RMSE), Ortalama Mutlak Hata (MAE) ve R-kare (R^2) ile değerlendirilir.

Sonuç

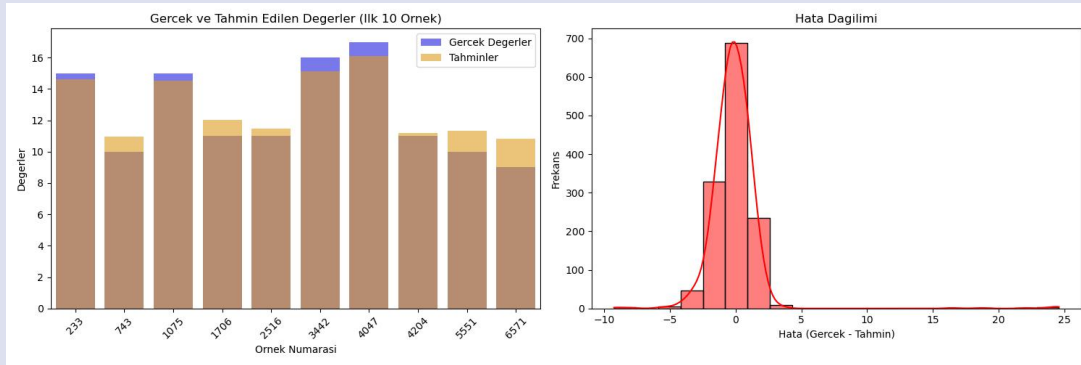
İlk aşamada tahmin sisteminde kullanılan yöntemlerde veri seti üzerinde öznelik seçme işlemi uygulanmamıştır. Elde edilen sonuçlar Çizelge 2 ile verilmiştir.

Çizelge 2 göz önünde bulundurulduğunda Rastgele Orman regresyon tekniği en düşük RMSE değerine sahiptir. Bu sonuca göre en yüksek tahmin başarımı Rastgele Orman tekniği ile elde edilmiştir.

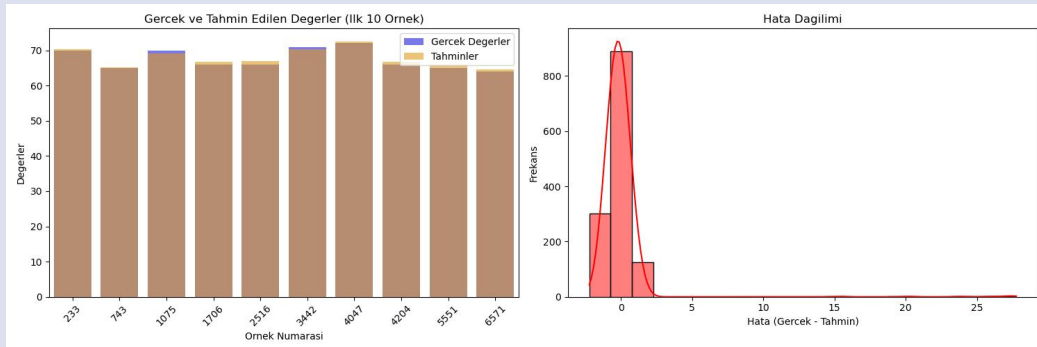
Yöntemlere ait grafiksel sonuçlar aşağıda verilmiştir.

Çizelge 2. Tahmin yöntemlerinin başarımları
Table 2. Performance results of prediction methods

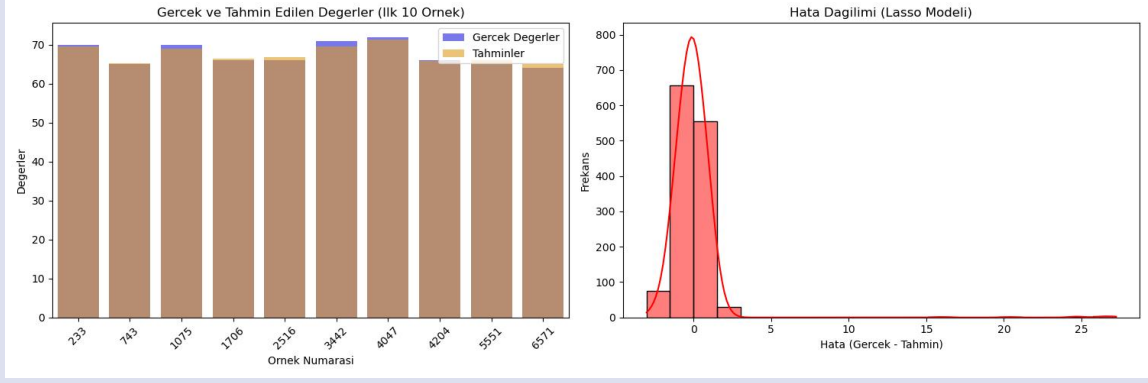
| Yöntem | RMSE |
|-----------------|--------------------|
| Rastgele Orman | 2.0880059652073077 |
| Ridge | 1.9078711855834931 |
| Lasso | 1.9551066974885811 |
| Karar Ağacı | 3.3723759900783428 |
| Derin Sinir Ağı | 2.4868690299085996 |



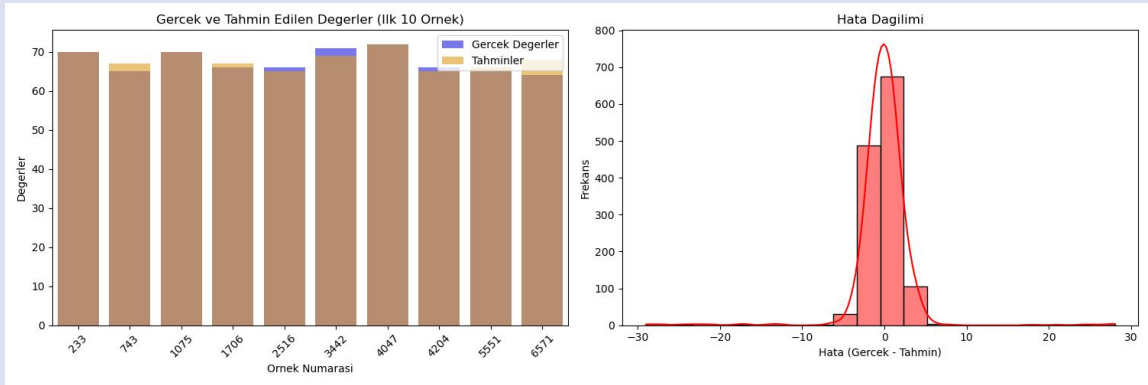
Resim 1. Rastgele Orman yöntemine ait grafiksel sonuçlar
Figure 1. Graphical results of the Random Forest method



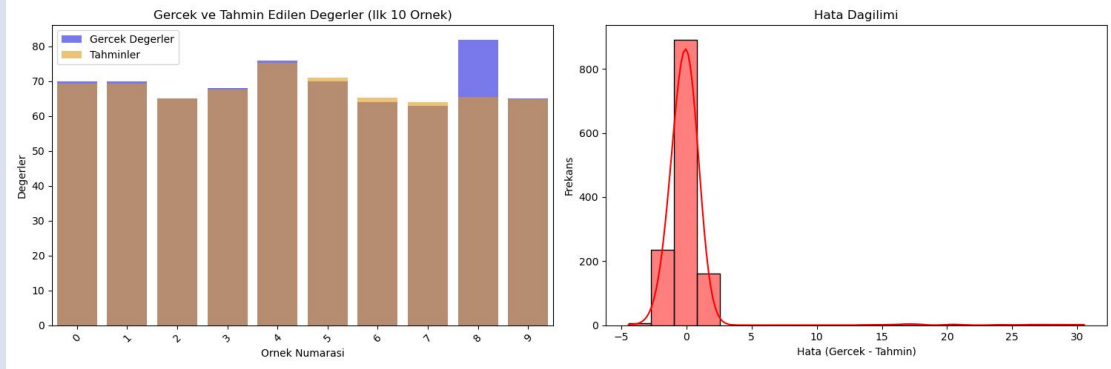
Resim 2. Ridge yöntemine ait grafiksel sonuçlar
Figure 2. Graphical results of the Ridge method



Resim 3. Lasso yöntemine ait grafiksel sonuçlar
Figure 3. Graphical results of the Lasso method



Resim 4. Karar Ağacı yöntemine ait grafiksel sonuçlar
Figure 4. Graphical results of the Decision Tree method



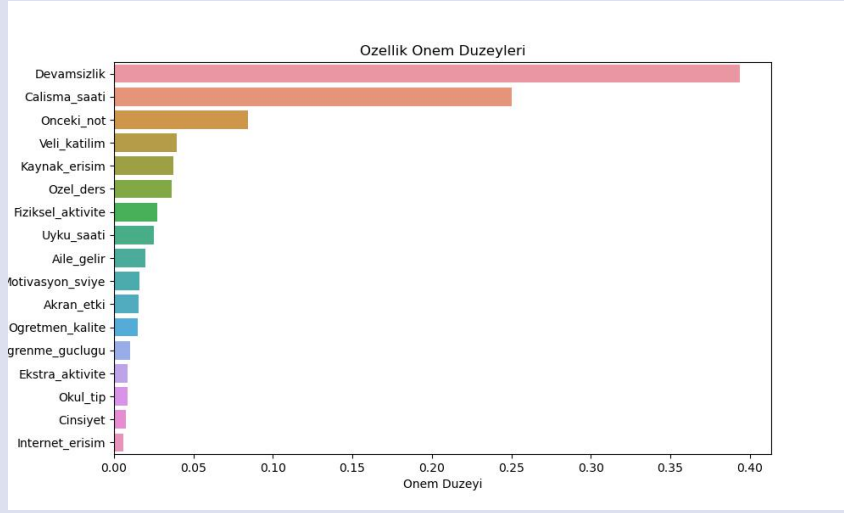
Resim 5. Derin Sinir Ağları yöntemine ait grafiksel sonuçlar
Figure 5. Graphical results of the Deep Neural Networks method

Hata dağılımı grafiği, modelin tahminlerindeki doğruluğunu incelemek için önemlidir. Histogram şeklinde dağılan hata değerleri, modelin gerçek ve tahmin edilen değerler arasındaki farkların (hataların) dağılımını gösterir. Normal dağılım gösteren, sıfıra yakın bir hata dağılımı modelin iyi bir performans sergilediğini, hataların çoğunun küçük olduğunu ve tahminlerin genellikle doğru olduğunu gösterir[16].

Bu çalışmada önemliliklerin önem derecesi, Rastgele Orman algoritması tarafından hesaplanan önemlilik

önemleri kullanılarak belirlenmiştir. Rastgele Orman modelindeki önemlilik önem derecesi (feature importance), modelin her bir önemliliğin hedef değişken üzerindeki göreceli etkisini anlamak için kullanılır. Öznitelik önem derecesi, genellikle Gini impurity veya mean decrease in impurity (MDI) gibi kriterlere dayanır [15].

Özniteliklerin tahmin başarısına etkisi grafiksel olarak Şekil 6 ile verilmiştir.



Resim 6. Özniteliklerin tahmin başarımına etkisi
Figure 6. Effect of features on prediction performance

Çizelge 3. Özniteliklerin önem dereceleri

Table 3. Importance levels of features

| Öznitelik | Önem Derecesi | Öznitelik | Önem Derecesi |
|-------------------|---------------|-------------------|---------------|
| Devamsizlik | 0.3939 | Motivasyon_seviye | 0.0158 |
| Calisma_saati | 0.2505 | Akran_etki | 0.0155 |
| Onceki_not: | 0.0843 | Ogretmen_kalite | 0.0151 |
| Veli_katirim | 0.0394 | Ogrenme_guclugu | 0.0099 |
| Kaynak_erisim | 0.0374 | Ekstra_aktivite | 0.0086 |
| Ozel_ders | 0.0362 | Okul_tip | 0.0084 |
| Fiziksel_aktivite | 0.0274 | Cinsiyet | 0.0072 |
| Uyku_saati | 0.0250 | Internet_erisim | 0.0058 |
| Aile_gelir | 0.0196 | | |

Çizelge 4. Tahmin yöntemlerinin başarımları

Table 4. Performance results of prediction methods

| Yöntem | RMSE |
|-------------------|-------------------|
| Rastgele Orman | 2.083993406275049 |
| Ridge | 1,903771185583493 |
| Lasso | 1,951006697488581 |
| Karar Ağacı | 3,368275990078342 |
| Derin Sinir Ağacı | 2,482769029908599 |

Çizelge 5. Literatür karşılaştırması

Table 5. Literature comparison

| Referans | Yöntem | RMSE |
|-------------------|-----------------------------|-------------|
| [10] | Karar Ağacı | 8,23 |
| [4] | Yapay Sinir Ağı | 4,32 |
| Bu Çalışma | Rigde-Rastgele Orman | 1,90 |

Özniteliklerin Rastgele Orman yöntemi ile elde edilen önem dereceleri Çizelge 3 ile verilmiştir.

Çizelge 3'e göre en düşük etki gösteren öznitelikler cinsiyet ve internet erişimi olarak tespit edilmiştir. Bu iki öznitelik veri boyutunu azaltmak için veri setinden çıkarılarak aynı yapay zeka teknikleri ile öğrenci başarımı tahmin edilmiştir. Elde edilen sonuçlar Çizelge 4 ile verilmiştir.

Çizelge 4 ile verilen sonuçlar göz önüne alındığında öznitelik seçme işlemi ile tahmin işleminde kullanılan yapay zeka tekniklerinin başarımı az da olsa artış göstermiştir.

Bu çalışmada Ridge ve Rastgele Orman teknikleri kullanılarak oluşturulan model en başarılı model olarak ortaya konmuştur. Önerilen yöntemin başarımı literatürdeki benzer çalışmalar ile karşılaştırılmış ve sonuçlar Çizelge 5 ile verilmiştir.

Tartışma

Çalışma sonuçlarına göre, Ridge ve Rastgele Orman modellerinin tahmin doğruluğunun diğer yöntemlere kıyasla daha yüksek olduğu gözlemlenmiştir. Özellikle Rastgele Orman algoritması, özneliklerin önem derecelerini belirlemede etkili olmuş ve modelin performansını artırmak amacıyla öznelik seçimini başarılı bir şekilde sağlamıştır. Strobl ve arkadaşlarının çalışmasında (2007), Rastgele Orman algoritmasının öznelik önem sıralamasında yanlışlıkları önleyici özelliklerinin altı çizilmiş ve bu algoritmanın öznelik seçimi için uygun bir yöntem olduğu belirtilmiştir. Çalışmada, en önemli öznelik olarak "Devamsızlık" belirlenmiş ve bu özneliğin öğrenci başarısı üzerindeki yüksek etkisi literatürdeki birçok çalışmayla tutarlılık göstermiştir. Bu durum, okul devam oranının akademik başarıyla doğrudan ilişkili olduğunu savunan Quigley ve MacKay-Lyons'un (2020) araştırmalarıyla da desteklenmektedir. Ayrıca, düşük etkili öznelikler olarak belirlenen "Cinsiyet" ve "İnternet Erişimi" gibi değişkenlerin çıkarılmasıyla yapılan tekrar tahminlerde, model performansında hafif bir artış gözlenmiştir. Bu sonuç, düşük etkili özneliklerin modelden çıkarılmasının gereksiz veri karmaşıklığını azaltarak tahmin doğruluğunu artırabileceğini ortaya koymaktadır.

Sonuç olarak, bu çalışma, eğitim ortamında karar verme süreçlerine rehberlik edebilecek öngörücü modeller geliştirmek için makine öğrenimi yöntemlerinin önemli bir araç olduğunu göstermektedir. Elde edilen bulgular, eğitimcilerin öğrencilerin başarısını etkileyen faktörlere odaklanarak bireysel destek stratejileri geliştirmeleri için veri odaklı yaklaşımlar kullanmalarını önermektedir. Bu çalışma, öğrenci başarı tahmin modellerinin daha geniş eğitim politikaları ve bireysel rehberlik planlarına entegre edilmesi için güçlü bir temel sağlamaktadır.

Referanslar

- [1] Demirtaş H, Güneş H. Eğitim yönetimi ve denetimi sözlüğü. Ankara: Anı Yayıncılık; 2002.
- [2] Sarier Y. Türkiye'de öğrencilerin akademik başarısını etkileyen faktörler bir meta-analiz çalışması. Hacettepe University Journal of Education. 2016; 31(3):1-1. doi:10.16986/HUJE.2016015868.
- [3] Arıkan Nİ. Akademik başarı ile seçilmiş sosyoekonomik değişkenler arasındaki ilişki üzerine bir araştırma. USBAD Uluslararası Sosyal Bilimler Akademi Dergisi. 2019;1(2):14-29.
- [4] Boyd FB. Motivation to continue: enhancing literacy learning for struggling readers and writers. Reading and Writing Quarterly. 2002; 18:257-277.
- [5] Jackson LA, Eye A, Witt EA, Zhao Y, Fitzgerald HE. A longitudinal study of the effects of Internet use and videogame playing on academic performance and the roles of gender, race, and income in these relationships. Computers in Human Behavior. 2011; 27:228-239. doi:10.1016/j.chb.2010.08.001.
- [6] Aškerc K, Gasparič RP, Bray M. Confronting the shadow education system: what government policies for what private tutoring? C.E.P.S. Journal. 2012; 2(4).
- [7] Quigley A, MacKay-Lyons M, Eskes G. Effects of exercise on cognitive performance in older adults: a narrative review of the evidence, possible biological mechanisms, and recommendations for exercise prescription. J Aging Res. 2020; 2020.
- [8] Erickson KI, Voss MW, Prakash RS, Basak C, Szabo A, Chaddock L, Kramer AF. Exercise training increases size of hippocampus and improves memory. Proc Natl Acad Sci U S A. 2011; 108(7):3017-3022. doi:10.1073/pnas.1015950108.
- [9] Kim HY, LaRusso MD, Hsin LB, Harbaugh AG, Selman RL, Snow CE. Social perspective-taking performance: construct, measurement, and relations with academic performance and engagement. J Appl Dev Psychol. 2018; 57:24-41.
- [10] Hussain S, Khan MQ. Student-performulator: predicting students' academic performance at secondary and intermediate level using machine learning. Annals of Data Science. 2023; 10:637-655.
- [11] Jayaprakash S, Krishnan S, Jaiganesh V. Predicting students academic performance using an improved random forest classifier. In: 2020 International Conference on Emerging Smart Computing and Informatics (ESCI); 2020. p. 238-243.
- [12] Kumar M, Bajaj K, Sharma B, Narang S. A comparative performance assessment of optimized multilevel ensemble learning model with existing classifier models. Big Data. 2022; 10:371-387.
- [13] Ofori F, Maina E, Gitonga R. Using machine learning algorithms to predict students' performance and improve learning outcomes: a literature-based review. J Inf Technol. 2020; 4:33-55.
- [14] Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interact Learn Environ. 2023; 31:3360-3379.
- [15] Long K, Guo D, Deng L, Shen H, Zhou F, Yang Y. Cross-combination analyses of random forest feature selection and decision tree model for predicting intraoperative hypothermia in total joint arthroplasty. J Arthroplasty. 2024; 1-9.
- [16] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res. 2005; 30(1):79-82.
- [17] Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources, and a solution. BMC Bioinformatics. 2007; 8(1):1-21.