



Yapay Zekâ Teknikleriyle Yükseköğretim Kurumları Sınavı (YKS) Puanlarının Tahmini

Betül ERSÖZ*  Halil İbrahim BÜLBÜL 

^{1,2} Gazi Üniversitesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Ankara, Türkiye

Makale Bilgisi

Araştırma makalesi
Başvuru: 02/07/2024
Düzeltilme: 30/08/2024
Kabul: 15/12/2024

Anahtar Kelimeler

Yapay zekâ
Makine öğrenme
Üretken yapay zeka
YKS
TYT
AYT

Article Info

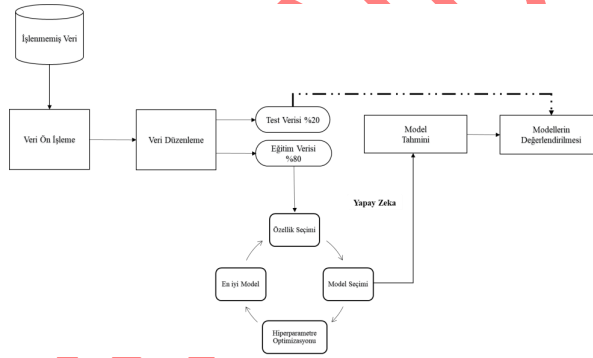
Research article
Received: 02/07/2024
Revision: 30/08/2024
Accepted: 15/12/2024

Keywords

Artificial intelligence
Machine learning
Generative artificial
intelligence
YKS
TYT
AYT

Grafik Özet (Graphical/Tabular Abstract)

Bu çalışmada, Yükseköğretim Kurumları Sınavı (YKS) sınavına giren öğrencilerin deneme sınavı netlerine (doğru/yanlış) göre YYS puanlarının tahmini yapılmıştır. İki aşamalı olan YYS sınavında TYT (Temel Yeterlilik Testi) ve AYT (Alan Yeterlilik Testi) puanlarının tahmin edilmesinde Yapay zekâ (YZ) tekniklerinden, Makine Öğrenme (ML) algoritmalarından Lineer Regresyon (LinReg), Çok Katmanlı Algılayıcı (MLP), K-En Yakın Komşu (KNN), Rastgele Orman (RF) gibi dört farklı model kullanılmıştır. / In this study, YYS scores of students who participated the Higher Education Institutions Examination (YKS) were predicted according to their net exam results (true/false). Four different models such as Linear Regression (LinReg), Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), Random Forest (RF) from Artificial Intelligence (AI) techniques, Machine Learning (ML) algorithms were used to predict TYT (Basic Proficiency Test) and AYT (Field Proficiency Test) scores in the two-stage YYS exam.



Şekil A: Çalışma Akış Şeması /Figure A: Work Flow Chart

Highlights (Önemli noktalar)

- Makine öğrenme algoritmalarıyla YYS (TYT/AYT) puanlarının tahmin edilmesi/ Predicting YYS (TYT/AYT) scores with machine learning algorithms
- Geliştirilen modelin ticari bir dijital öğrenme platformun üretken yapay zeka aracına entegrasyonu/Integration of the developed model into the generative AI tool of a commercial digital learning platform.
- Öğrencilerin YYS puan tahminlerine yönelik öngörü sunan bir ürün geliştirilmesi/ Developing of a product that provides insight into the students' YYS score predictions

Amaç (Aim): Geliştirilen model, öğrencilerin deneme sınav verilerinden YYS puan (TYT/AYT) sonuçlarını ML teknikleri ile tespit ederek ticari bir dijital öğrenme platformun üretken yapay zeka aracına entegre edilmiştir. / The developed model is integrated into the generative artificial intelligence tool of a commercial digital learning platform by predicting YYS score (TYT/AYT) results from the students' test data with ML techniques.

Özgünlük (Originality): Yapay zekâ yöntemleriyle YYS puanlarını tahmin eden çalışmaların veya uygulamaların bulunmaması ve modelin ürün olarak öğrenme platformuna entegre edilmesidir. / There are no studies or applications that predict YYS scores using artificial intelligence methods, and the model is integrated into the learning platform as a product.

Bulgular (Results): ML algoritmaları içerisinde TYT' de en iyi performans gösteren MLP, R-Kare (0.999), MAE (0.056) ve RMSE (0.447) ve AYT' de en iyi performans gösteren Lineer regresyon R-Kare (0.999), MAE (0.214) ve RMSE (0.0413) değerleri bulunmuştur. / ML algorithms, MLP performed the best for TYT with R-Square (0.999), MAE (0.056) and RMSE (0.447) and Linear regression performed the best for AYT with R-Square (0.999), MAE (0.214) and RMSE (0.0413).

Sonuç (Conclusion): Bu çalışma sonucunda, ticari bir dijital öğrenme platforma entegre edilen yapay zekâ tabanlı YYS puan tahmin modeli geliştirilmiştir./ As a result of this study, an artificial intelligence-based YYS score prediction model integrated into a commercial digital learning platform has been developed.



Yapay Zekâ Teknikleriyle Yükseköğretim Kurumları Sınavı (YKS) Puanlarının Tahmini

Betül ERSÖZ* Halil İbrahim BÜLBÜL

¹Gazi University, Department of Computer and Instructional Technologies Education, Ankara, Türkiye

Makale Bilgisi

Araştırma makalesi
Başvuru: 02/07/2024
Düzeltilme: 30/08/2024
Kabul: 15/12/2024

Anahtar Kelimeler

Yapay zekâ
Makine öğrenme
Üretken yapay zeka
YKS
TYT
AYT

Öz

Yükseköğretim programlarına yerleştirmeler öncelikle öğrencilerin akademik başarılarına ve tercihlerine göre belirlenir. Yükseköğretim Kurumları Sınavı (YKS) giren öğrenciler başta YKS puanı, Ortaöğretim Başarı Puanı (OBP) ve tercih sıralamalarına göre kariyer hedeflerine uygun yükseköğretim bölümlerine yerleştirilir. YKS ile yerleşmede en önemli faktör sınav puanıdır. Bu sebeple öğrenciler, sınav öncesi netlerinden puanlarının belirlenmesi için öneri sistemlerine ihtiyaç duymaktadır. Deneme sınavları netlerinden öğrencilerin YKS puanlarını formüller yardımıyla hesaplayarak tahmin eden hâlihazırda çeşitli sistemler mevcuttur. Ancak yapay zekâ yöntemleriyle puanları tahmin eden uygulamalar bulunmamaktadır. Bu çalışmada YKS'ye giren öğrencilerin deneme sınavı netlerine göre YKS puanlarının tahmini yapılmıştır. Araştırmada, iki aşamalı olan YKS sınavında TYT (Temel Yeterlilik Testi) ve AYT (Alan Yeterlilik Testi) puanlarının tahmin edilmesinde Yapay zekâ (YZ) tekniklerinden olan Makine Öğrenme (ML) algoritmalarından Lineer Regresyon (LinReg), Çok Katmanlı Algılayıcı (MLP), K-En Yakın Komşu (KNN), Rastgele Orman (RF) gibi dört farklı model kullanılmıştır. ML algoritmaları içerisinde TYT' de en iyi performans gösteren MLP, R-Kare (0.999), MAE (0.056) ve RMSE (0.447) değerleri bulunmuştur. AYT' de en iyi performans gösteren Lineer regresyon R-Kare (0.999), MAE (0.214) ve RMSE (0.0413) değerleri bulunmuştur. Geliştirilen model, öğrencilerin YKS puan sonuçlarını ML teknikleriyle tespit ederek ticari bir dijital öğrenme platformun üretken yapay zeka aracına entegre edilmiş olup, puan tahminlerinde destek sağlamaktadır.

Prediction of Higher Education Institutions Examination (YKS) Scores with Artificial Intelligence Techniques

Article Info

Research article
Received: 02/07/2024
Revision: 30/08/2024
Accepted: 15/12/2024

Keywords

Artificial intelligence
Machine learning
Generative artificial
intelligence
YKS
TYT
AYT

Abstract

Students' academic success and preferences primarily determine placements in higher education programs. Students who take the Higher Education Institutions Exam (YKS) are placed in higher education departments suitable for their career goals, primarily according to their YKS score, Secondary Education Success Score (OBP), and preference rankings. The most crucial factor in placement with YKS is the exam score. Therefore, students need recommendation systems to determine their net scores before the exam. Similar to this study, various systems calculate students' YKS scores using formulas from net scores of practice exams. However, it has been observed that no application estimates score with artificial intelligence methods. In this study, the YKS scores of students who participated to the YKS are estimated according to their net scores of practice exams. In the study, four different models such as Linear Regression (LinReg), Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), Random Forest (RF) from Machine Learning (ML) algorithms, which are among the Artificial Intelligence (AI) techniques, were used in estimating the scores of TYT (Basic Proficiency Test) and AYT (Field Proficiency Test) in the two-stage YKS exam. Among the ML algorithms, the best-performing MLP for TYT was found to be R-Square (0.999), MAE (0.056), and RMSE (0.447). The best-performing Linear Regression for AYT was found to be R-Square (0.999), MAE (0.214), and RMSE (0.0413). The developed model is integrated into the generative artificial intelligence tool of a commercial digital learning platform by using ML techniques to predict students' YKS score results and provides assistance in score predictions.

1. GİRİŞ (INTRODUCTION)

Türkiye’de yükseköğretime geçiş süreci olan Yükseköğretim Kurumları Sınavı (YKS), milyonlarca öğrencinin üniversite eğitimine adım atmasında kritik bir rol oynamaktadır. Öğrenciler, bu sınavdan aldıkları puanlarla hayal ettikleri üniversitelere yerleşmeye çalışırken, tercih sürecinde doğru yönlendirilmek, hem hedeflerine ulaşma açısından hem de başarılı bir eğitim hayatına adım atma bakımından büyük önem taşır. Ancak, tercih süreci karmaşık ve yorucudur. Dolayısıyla öğrencilerin yüzlerce seçenek arasından doğru üniversite ve bölümü seçebilmeleri için rehberlik almaları gereklidir. Bu doğrultuda, Yükseköğretimde eğitime devam etmek isteyen öğrenciler Yükseköğretim Kurumları Sınavına (YKS) girerek Temel Yeterlilik Testi (TYT), Alan Yeterlilik Testi (AYT) ve Yabancı dil olmak üzere üç farklı sınav oturumunda orta öğretimde almış oldukları temel derslerden sınava tabi tutulmaktadır. TYT, tüm adaylar için zorunluken, AYT ve Yabancı dil isteğe bağlı olarak adayın tercih ettiği bölüme göre seçilmektedir. Türkiye’de birçok aile ve öğrenci, iyi bir üniversiteye ve bölüme yerleşmenin gelecekteki başarı ve yaşam standardı için çok önemli olduğuna inanmaktadır. Dolayısıyla YKS, öğrencilerin mesleki hayatlarındaki başarı hedeflerine ulaşmaları için kullandıkları en önemli sınavlardan biridir. Her yıl düzenli olarak yapılan ve ortama üç milyon öğrencinin katılım sağladığı bu sınavlarda öğrencilerin YKS puan sıralamaları, YKS sınavlarında derslere ait testlerdeki net sayılarına ve Ortaöğretim Başarı Puanı (OBP) hesaplanarak yapılmaktadır [1]. Öğrencilerin YKS’de başarısını etkileyen birçok faktör bulunmaktadır. Bu faktörler arasında öğrencinin akademik yeteneği, sınavlara daha iyi hazırlayan eğitim sistemi, sınav teknikleri, soru çözme becerileri, çalışma disiplini, sınavdaki performansını olumsuz etkileyen stres ve kaygı olarak sıralanabilir. Öğrencilerin bu faktörlerin bilincinde olması ve olumlu yönde etkileyecek şekilde çalışması, YKS’de başarı şansını artırabilmektedir. Bu doğrultuda öğrencilerin sınav kaygılarının azalmasında, geldikleri seviyeleri belirlemelerinde ve eksiklerini değerlendirmelerinde bir ölçme aracı olarak deneme

sınavları etkin bir şekilde kullanılmaktadır. Deneme sınavları ile öğrenciler, sınav öncesinde başarılarını ölçmek için sınavlarda yapmış oldukları doğru ve yanlış sorulardan net puanlarını analiz ederek girmek istedikleri yükseköğretim programının yerleştirme puanını hesaplayabilirler [2-4]. Bilim ve teknolojideki hızlı ilerlemeler, ilköğretimden yükseköğretime kadar tüm eğitim kademelerini olumlu yönde etkilemektedir. Bununla birlikte, eğitim teknolojilerinin yaygınlaşmasıyla akıllı telefon, tablet, bilgisayar gibi teknolojik araçların kullanımı öğrenciler tarafından son yıllarda hızla yaygınlaşmıştır [5]. Bu sayede uygulamaları artan eğitim platformları, öğrenci-öğretmen arasında dijital ortamlarda etkileşim, erişilebilirlik ve etkileşimi artırmaktadır. Özellikle son yıllarda, öğretim yönetimi otomasyonu, dil modelleri (ChatGPT, Gemini, Falcon vb.) ile geri bildirim ve değerlendirme, sanal ve artırılmış gerçeklik uygulamaları ve YZ uygulamaları öğrencilerin dijital öğrenme platformlarında oldukça etkin olarak kullanılmaktadır [6, 7]. Son dönemlerde YZ’ nin bir alt dalı olan Makine öğrenme (ML) algoritmaları, büyük miktarda veriler işleyerek anlamlı bilgiler çıkarma ve bu bilgileri tahminlerde bulunmak ve kararlar almak için kullanılmaktadır. Bu sayede öğrenci verilerinden anlamlı çıkarımlar ve modeller öğrenme ortamlarının iyileştirilmesinde destek sağlamaktadır [8]. İlgili literatür incelendiğinde: Yapılan araştırmalarda [9] okullardaki erken aşamalarda öğrencilerin zayıf ve güçlü yönlerini tahmin etmek için (karar ağacı, rastgele orman, destek vektör makinesi) gibi makine öğrenme algoritmalarını kullanarak RF (Rastgele Orman) algoritmasıyla %94 doğruluk oranıyla öğrenci notunu tahmin eden kapsamlı bir web uygulaması oluşturulmuştur. Bir diğer çalışmada [10] her disiplindeki öğrencilerin üniversiteye giriş sınavı puanlarını tahmin etmek için bir makine öğrenmesi algoritması kullanılmış ve öğrencilerin ilgileri, baskıya dayanıklılıkları ve kişilik özellikleri gibi çeşitli etki faktörleri dikkate alınmıştır. RF algoritması ile tahmin doğruluğu % 80’in üzerindeyken Adaboost algoritmasının doğruluğu % 90’ın üzerinde bulunmuştur. Öğrencilerin performansını etkili bir şekilde geliştirmek ve eğitimcilerin öğrencilerin öğrenme durumlarını

izlemelerine yardımcı olmak için bir çalışmada [11], öğrencilerin akademik performansını etkileyen faktörleri araştıran ve akademik performansını tahmin eden sistemde destek vektör makine (DVM) modelinin en kararlı etkiye sahip olduğunu göstermiştir. Yapılan diğer çalışmada [12], üniversitenin akademik itibarı, üniversitenin bulunduğu şehrin olanakları, üniversitenin olanakları ve kültürel olanaklara ilişkin verilere dayanarak öğrenci yerleştirme yüzdesini tahmin etmeye yönelik makine öğrenmesi algoritmaları geliştirilmiştir. Yapılan çalışmada model doğruluğu performans metriklerine göre değerlendirildiğinde Extreme Gradient Boosting (XGBoost) algoritması diğer makine öğrenimi algoritmalarına göre daha yüksek tahmin doğruluğu göstermiştir. Yapılan bir çalışmada [13] lise öğrencilerinin üniversiteye kabul edilebilmek için Matura sınavı ML teknikleriyle değerlendirilmiştir. Bu sınavla ilgili yapılan çalışmada toplam on altı algoritma uygulanmıştır. Sonuçlara göre RepTree algoritması, 0,90 Gerçek Pozitif (TP) oranı ve 0,14 Yanlış Pozitif (FP) oranıyla en iyi performansı göstermiştir. Şanlıurfa ilinde [14] 2018 yılında YKS' ye giren öğrencilerden 12.000 öğrencinin TYT puanları analiz edilmiştir. Doğrusal Regresyon, Destek Vektör Regresyonu ve epsilon parametrelili, Rastgele Orman Regresyonu kullanılmıştır. Tüm regresyon yöntemleriyle Korelasyon Katsayısı değerlerinin TYT puanına etkisi tespit edilmiştir.

Literatürde, YKS yerleştirmeleri ile ilgili oldukça sınırlı sayıda akademik çalışma olduğu görülmektedir. Ancak dershaneler, özel okullar ve bunlara kaynak üreten ticari işletmeler tarafından bir önceki yıllardaki YKS net sayısı, ortalaması, standart sapması gibi istatistiksel veriler hesaplanarak kural tabanlı uygulamalar bulunmaktadır. Yapılan bu çalışmada YZ tabanlı dört ML yöntemi olan Lineer Regresyon (LinReg), Multi-Layer Perceptron (MLP), K-Nearest Neighbors, (KNN), Random Forest (RF) gibi dört farklı Makine Öğrenme (ML) algoritması kullanılarak YKS puan tahmin modeli geliştirilmiştir. Geliştirilen model, öğrencilerin deneme sınav verilerinden YKS puan (TYT/AYT) sonuçlarını ML teknikleri ile tespit edilerek YZ tabanlı ticari bir dijital öğrenme

platformun üretken yapay zeka aracına entegre edilmiştir.

2. MATERYAL VE YÖNTEM (MATERIAL AND METHOD)

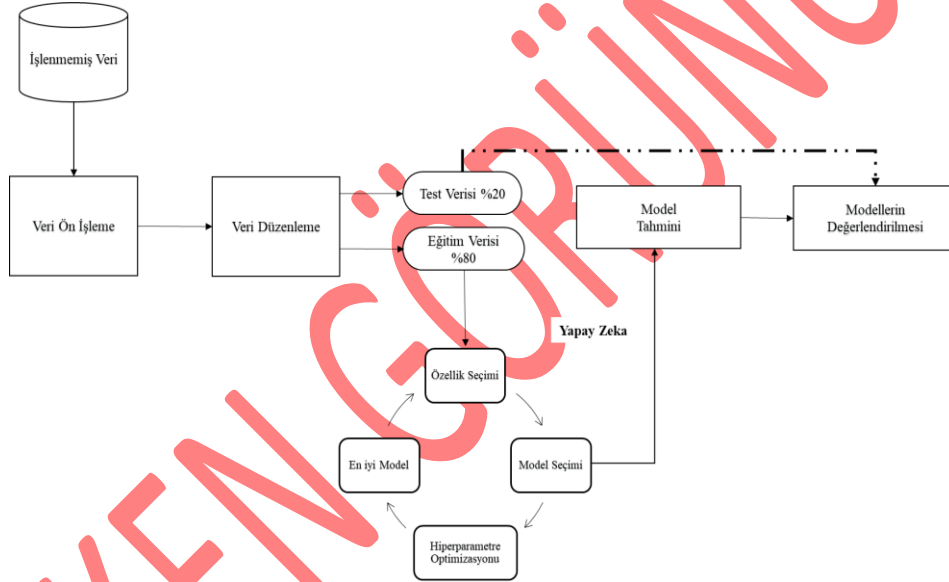
Bu bölümde veri seti, kullanılan yöntem, geliştirilen modele yönelik bir akış diyagramı, kullanılan araçlar ve modelin eğitim ve test olarak geliştirilmesine yönelik bilgiler verilmiştir.

Çalışma kapsamında, veri seti olarak bir firma tarafından üretilen veriler kullanılmıştır. Veri seti 2023-2024 eğitim-öğretim yılında YKS deneme sınavına giren öğrencilere ait 56.000 TYT ve AYT verisinden oluşmaktadır. Veri setindeki hatalı ve eksik veriler yeniden düzenlenerek 40.863 TYT ve AYT veri seti çalışmada kullanılmak üzere yeniden düzenlenmiştir. TYT sınavı Türkçe (40), Sosyal Bilimler (20), Temel Matematik (40) ve Fen Bilimleri (20) sınav sorularından oluşmaktadır. AYT sınavında puan hesaplamalarında bölüm tercihlerine göre, Sayısal (SAY), Eşit Ağırlık (EA) ve Sözel (SÖZ) puan alanları yer almaktadır. AYT sınavı Sayısal puanı için TYT netlerine ek olarak, AYT sınavı soruları Matematik (40), Fizik (14), Kimya (13), Biyoloji (13) netleriyle puan hesaplanmaktadır. Aynı şekilde AYT Eşit ağırlık puanı, TYT netlerine ek olarak, AYT sınavı soruları Türk Dili ve Edebiyatı (24), Tarih-1 (10), Coğrafya-1 (6) ve Matematik (40) netleriyle hesaplanmaktadır. AYT Sözel puanı için, TYT netlerine ek olarak, AYT sınavı soruları, Türk Dili ve Edebiyatı (24), Tarih-1 (10), Coğrafya-1 (6), Tarih-2 (11), Coğrafya-2 (11), Felsefe Grubu (12) ve Din Kültürü ve Ahlak Bilgisi / Ek Felsefe Grubu (6) netleriyle hesaplanmaktadır.

Çalışma kapsamında, işlenmemiş veri seti öncelikle ön işlemden geçirilerek eksik ve hatalı olanlar çıkarılmış olup bazı veriler için boşluklar sıfır (0) ile doldurulmuştur. Veri seti %80 eğitim ve %20 test olarak ayrılmıştır. Yapılan çalışmada, YKS puan tahminlerinde LinReg, MLP, KNN, RF olmak üzere dört farklı ML algoritması kullanılarak modeller geliştirilmiştir. **LinReg**, ML teknikleri içerisinde en temel ve yaygın kullanılan modellerinden biridir. Bu model, giriş ve çıkış değişkenleri arasındaki ilişkiyi modellemek ve giriş değişkeninin değerini değerine göre tahmin etmek için kullanılır. Basitliği, yorumlanabilirliği ve farklı problemlere uygulanabilirliği nedeniyle yaygın olarak kullanılmaktadır. **MLP**, sinir ağının eğitimi için

ileri beslemeli olarak bilinen ANN' nin denetimli bir öğrenme tekniği olarak bilinir. MLP karmaşık problemleri çözmek ve yüksek doğruluk elde etmek için kullanılırlar. **KNN**, benzer özelliklere sahip veri noktalarının gruplandığı sınıflandırma ve regresyon için kullanılan bir algoritmadır. Aynı zamanda denetimli bir öğrenme algoritmasıdır ve verileri en yakın komşusunun hiperparametresine göre sınıflandırmak için de kullanılır. **RF**, birçok karar ağacının bir araya gelerek tahmin yapmasına dayalı bir tahmin yapmaktadır [15-21]. Çalışma, Python programlama diliyle Scikit-Learn, NumPy, Matplotlib, Pandas gibi kütüphaneleri ve Jupyter, Spyder editörleri kullanılarak modeller geliştirilmiştir.

Modeller, deneme sınavı netleri ve puanları arasındaki ilişkiyi öğrenmek için eğitilmiştir. Bu aşamada, modelin hiperparametreleri en iyi sonuca göre ayarlanmıştır. Veri seti ölçeklendirilmiştir. Çalışmada kullanılan bir yöntem olan K-Fold çapraz doğrulama, ML modellerinin performansını değerlendirmek ve aşırı öğrenme (overfitting) sorununu azaltmak için yaygın olarak kullanılmaktadır. K-Fold, modelin farklı veri bölümleri üzerinde nasıl performans gösterdiğini değerlendirir ve bu sayede modelin genelleme yeteneğini daha iyi ölçmektedir. Modelin doğruluğunu kontrol etmek için, eğitim verisinden ayrı bir test veri kümesi kullanılmaktadır [22, 23]. Modelin performansını ölçmek için R-Kare skoru, MAE ve RMSE metrikleri kullanılmıştır. Yapılan çalışmaya ilişkin akış diyagramı Şekil 1 'de gösterilmiştir.



Şekil 1. Çalışma Akış Şeması (Study Flowchart)

Sayısal bir değer tahmin edilmesini gerektiren tahmine dayalı regresyon modelinin performansını değerlendirmek ve raporlamak için sıklıkla kullanılan hataların mutlak değerlerinin ortalaması (MAE) hata ölçüsüdür. MAE, bir modelin tahmin edilen ve gerçek değerleri arasındaki hataların mutlak değerlerinin ortalamasını hesaplayan basit ve güçlü bir ölçüm aracıdır. Bir modelin ne kadar iyi genelleştirildiğini ve tahminlerinin gerçek değerlerle ne kadar iyi eşleştiğini değerlendirmektedir. Bu çalışmada deneysel sonuçlara en yakın tahmine dayalı modellemeyi bulmayı, deneysel değerlerden sapmaları görmeyi ve bağımlı değişken ile bağımsız değişken

arasındaki ilişkileri yakalamak amacıyla MAE kullanılmıştır. MAE aşağıdaki formülle hesaplanmaktadır [23]:

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - Y_{tahmin(i)}| \quad (1)$$

R-Kare (R^2): Modelin giriş değişkenleri tarafından belirlenen çıkış değişkenindeki değişim oranını ölçmektedir. Ayrıca bir modelin bağımlı değişkenindeki toplam varyansın ne kadarını açıkladığını gösteren bir metriktir. Bir regresyon modeli eğitim verilerine iyi uyum sağlayabilir ancak çok fazla bağımsız giriş değişkene sahipse test verilerine uymayabilir. Modele eklenen ek

bağımsız değişkenler dikkate alınmaktadır. Gerçek ve tahmin edilen değerler arasındaki ilişkiyi ifade eden R-Kare değeri 0 ile 1 arasında değişmektedir. R^2 değeri 1'e ne kadar yakınsa model o kadar başarılı demektir [24]. Aşağıda R^2 denklem yer almaktadır.

$$R^2 = 1 - \frac{\sum_i^N (Y_{tahmin(i)} - Y_i)^2}{\sum_i^N (Y_i - Y_{mean})^2} \quad (2)$$

Hataların Karesinin Ortalamasının Karekökü (RMSE): MSE' nin kareköküdür. tahmin edilen değerlerle gerçek değerler arasındaki farkların büyüklüğünü ölçmek için kullanılan bir hata metriğidir [25].Aşağıda RMSE denklemi yer almaktadır.

$$RMSE = \sqrt{\frac{1}{N} \sum_i^N (Y_i - Y_{tahmin(i)})^2} \quad (3)$$

Burada N toplam veri sayısını, y_i hesaplanan değerler, $Y_{tahmin(i)}$ ise tahmin edilen değerlerdir.

3. BULGULAR (RESULTS)

Bu bölümde istatistiksel analiz ve makine öğrenme yöntemiyle gerçekleştirilen analizlerin sonuçları yer almaktadır.

Tablo 1'de gösterilen istatistiksel analiz sonucunda, her konu için 40.863 gözlemden oluşan örneklem büyüklüğüne dayanarak TYT (Temel Yeterlilik Testi) ve AYT (Alan Yeterlilik Testi) sınavlarında çeşitli konulardaki öğrenci performansına ilişkin istatistiksel bir genel bakış sunmaktadır. TYT ve AYT 'de farklı konulardaki değişen performans seviyeleri ve dağılım özellikleri incelendiğinde: TYT Türkçe, daha yüksek ortalama puanlar ve daha tipik bir dağılım gösterirken, AYT'deki Felsefe ve Coğrafya 2 aşırı çarpıklık ve basıklık sergiliyor, bu da birçok düşük puana ve az sayıda yüksek aykırı değere işaret etmektedir.

Tablo 1. Veri Seti İstatistik Özellikleri (Data Set Statistical Properties)

Sınavlar	Maksimum	Minimum	Ortalama	Standart Sapma	Çarpıklık	Basıklık	Veri seti
Fen Bilimleri TYT	20.0	-4.0	7.69	5.84	0.22	-1.12	40863
Matematik TYT	40.0	-4.0	15.72	10.26	0.35	-0.86	40863
Türkçe TYT	40.0	-6.75	25.35	7.92	-0.87	0.71	40863
Sosyal Bilimler TYT	20.0	-3.75	10.57	4.5	-0.45	-0.32	40863
Türk Dili AYT	24.0	-4.75	3.47	6.05	1.58	1.16	40863
Felsefe AYT	12.0	-3.0	0.32	1.3	4.72	23.7	40863
Coğrafya 1 AYT	6.0	-1.5	0.94	1.75	1.66	1.41	40863
Din Kültürü ve Ahlak Bilgisi AYT	6.0	-1.5	0.36	1.05	3.49	12.35	40863
Matematik AYT	40.0	-5.75	15.66	11.15	0.42	-0.94	40863
Fizik AYT	14.0	-3.5	4.38	4.28	0.54	-0.93	40863
Kimya AYT	13.0	-3.25	4.98	4.53	0.28	-1.37	40863
Cografya2_AYT	11.0	-2.75	0.45	1.61	4.1	17.27	40863
Biyoloji AYT	13.0	-3.25	3.3	3.61	0.81	-0.42	40863
Tarih1 AYT	10.0	-2.5	1.69	2.86	1.39	0.46	40863
Tarih2 AYT	11.0	-2.75	0.49	1.66	3.75	14.28	40863

Tablo 2 'de çalışmada kullanılan TYT veri seti için dört makine öğrenme (ML) modelinin R-Kare, MAE ve RMSE değerleri verilmiştir. En yüksek R-Kare, 0.99996 MAE 0.05635ve RMSE 0.447350 değerleri ile MLP modelin iken, R-Kare, 0.99885 MAE 1.78563 ve RMSE 2.679079 değerleri ile RF modeli daha düşük bir başarı göstermiştir. Yüksek R-Kare değeri 1 yakınlığı ile birlikte düşük MAE

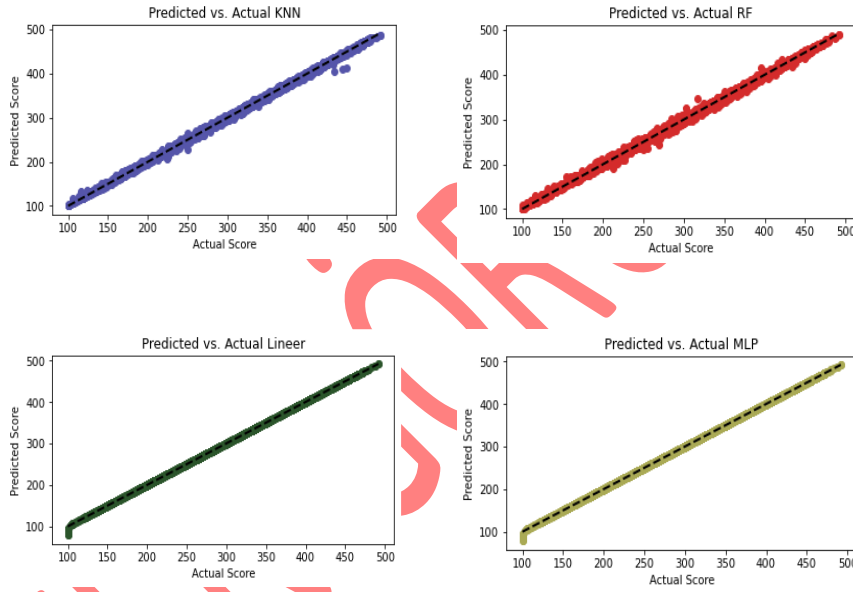
ve RMSE değerleri, modellerin tahminlerinin genel olarak daha doğru ve düşük hata olduğunu gösterir. Bununla birlikte R-Kare, MAE ve RMSE açısından en iyi performansa sahip TYT için MLP modelidir. Bu model, veri setindeki derslere ait net sayıları olan bağımlı değişkenleri en iyi şekilde açıklayabilen ve tahmin edebilen model olarak öne çıkmaktadır.

Table 2. Öğrencilerin YKS (TYT) Deneme Sınavı Netlerinin LinReg,MLP,RF ve KNN Algoritmalarıyla Tahmini (Prediction of Students' YKS (TYT) Test scores with LinReg, MLP, RF and KNN Algorithms)

Model	R-Kare	MAE	RMSE
LinReg	0.99996	0.06090	0.444093
MLP	0.99996	0.05635	0.447350
KNN	0.99923	1.51495	2.193468
RF	0.99885	1.78563	2.679079

Şekil 2’de TYT için R-kare değerlerinin grafiksel gösterimi yer almaktadır. Çalışmada R-kare değerlerinin grafiği incelendiğinde: LinReg ve MLP modelleri, en iyi performansı sergileyen modeller olarak öne çıkmaktadır. Tahmin edilen puanlar, gerçek puanlarla neredeyse mükemmel bir uyum içindedir. KNN ve RF modelleri de iyi performans

sergilemektedir ancak bazı sapmalar mevcuttur. Bu modellerin performansı, doğruluk gerektiren uygulamalarda biraz daha düşük olabilir. Genel olarak, tüm modeller belirli bir doğruluk seviyesinde iyi performans göstermiş, ancak regresyon ve MLP modelleri en yüksek doğruluğu sunmaktadır.



Şekil 2. Öğrencilerin YKS (TYT) Deneme Sınavı Netlerinin LinReg,MLP,RF ve KNN Algoritmalarıyla R² Tahmin Grafiği (R² Prediction Graph of Students' YKS (TYT) Test scores with LinReg, MLP, RF and KNN Algorithms)

AYT netlerinden puan tahmini incelendiğinde Tablo 3’de gösterildiği üzere veri setindeki bağımlı değişkeni tahmin etmek için en iyi modelin Regresyon olduğu söylenebilir. MLP modeli de oldukça başarılı sonuçlar vermektedir. Ancak, KNN

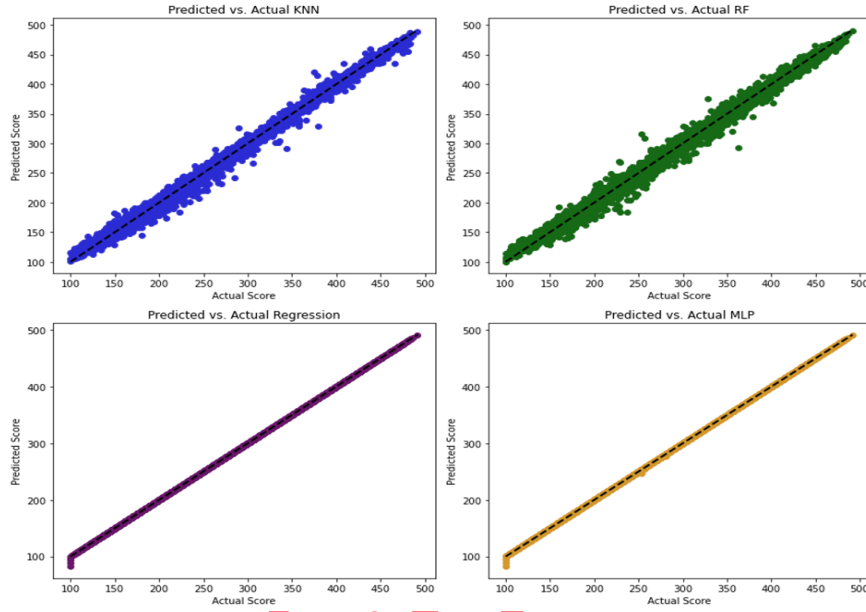
ve RF modelleri, diğer iki modele kıyasla çok daha yüksek hata oranlarına sahip olup, veri seti üzerinde daha az başarılı tahminler yapmaktadırlar. LinReg modeli 0.99998, 0.041368, 0.214139 sırasıyla R-Kare, MAE ve RMSE değerlerine sahiptir.

Table 3. Öğrencilerin YKS (AYT) Deneme Sınavı Netlerinin LinReg,MLP,RF ve KNN Algoritmalarıyla Tahmini (Prediction of Students' YKS (AYT) Test scores with LinReg, MLP, RF and KNN Algorithms)

Model	R-Kare	MAE	RMSE
LinReg	0.99998	0.041368	0.214139
MLP	0.99998	0.055249	0.235580
KNN	0.99052	4.203871	6.031329
RF	0.98619	4.469268	6.780396

Regresyon Modeli, R-Kare, MAE ve RMSE açısından en iyi performansa sahip modeldir. Bu model, veri setindeki bağımlı değişkeni en iyi şekilde açıklayabilen ve tahmin edebilen model olarak öne çıkmaktadır. MLP Modeli, performansı regresyon modeline çok yakındır. KNN ve RF Modelleri performansları Regresyon ve MLP'ye göre daha düşüktür. Özellikle MAE ve RMSE değerleri, bu modellerin tahminlerinde daha fazla

hata yaptığını göstermektedir. Bu sonuçlara göre, veri setindeki bağımlı değişkeni tahmin etmek için en iyi modelin Regresyon olduğu söylenebilir. Diğer modeller (özellikle KNN ve RF) daha yüksek hata oranlarına sahip olup, veri seti üzerinde daha az başarılı tahminler yapmaktadırlar. Şekil 3'de AYT R-kare tahmin edilen ve gerçek değerler grafiği yer almaktadır.



Şekil 3. Öğrencilerin YKS (AYT) Deneme Sınavı Netlerinin LinReg,MLP,RF ve KNN Algoritmalarıyla R^2 Tahmin Grafiği (R^2 Prediction Graph of Students' YKS (AYT) Test scores with LinReg, MLP, RF and KNN Algorithms)

4. SONUÇ (CONCLUSION)

Türkiye'de yükseköğretime geçiş sınavı olan Yükseköğretim Kurumları Sınavı (YKS), öğrencilerin üniversite yerleştirmelerinde belirleyici bir rol oynamaktadır. Öğrencilerin tercih yaparken dikkat ettikleri puan türleri ve sıralamaları, eğitim ve kariyer hedeflerine ulaşmalarında büyük bir önem taşımaktadır. Bu çalışmada, YKS puanlarını dört farklı algoritma ile analiz ederek öğrenciler için öneri geliştiren bir yapay zekâ modelinin tasarımı ve uygulaması gerçekleştirilmiştir. Elde edilen sonuçlar, her bir algoritmanın YKS tercih sürecine olan katkısını göstermiş ve öğrencilerin üniversite seçimlerinde daha bilinçli kararlar alabilmeleri için bu tür bir modelin faydalı olabileceğini ortaya koymuştur. Bu çalışma kapsamında, ticari bir dijital öğrenme platformuna entegre edilen yapay zekâ tabanlı sınav puan tahmin modeli geliştirilmiştir. Çalışmada,

YKS' ye giren öğrencilerin sınav puanlarından (doğru/yanlış) net puanları elde edilerek TYT ve AYT puan tahminleri yapılmıştır. Standart sapmadan dolayı her yıl öğrencilerin puan tahmini değişmektedir. Geleneksel formüller belirli varsayımlar ve sabit katsayılar kullanmaktadır. Ancak, bu yöntemler her öğrencinin performansını ve sınavın dinamiklerini tam olarak yansıtmayabilir. YZ algoritmaları, daha yüksek doğruluk ve hassasiyet sunarak öğrenci puanlarını daha isabetli bir şekilde tahmin edebilir. YZ, geniş veri setlerini işleyebilme yeteneğine sahiptir. Formüllerle yapılan tahminler eksik ya da fazla puan tahminleri yapmaktadır. Çalışmada LinReg ve MLP algoritmaları gerçek değere en yakın R-Kare değeri ve MAE, RMSE gibi hata metrikleri ile en düşük hata oranlarına sahiptir. Geliştirilen modelin üretken yapay zeka aracıyla entegre olan öğrenme platformuna entegre edilmesi, öğrencilerin puan

tahminlerine destek sağlayarak YZ tabanlı yenilikçi bir yaklaşım sunmaktadır.

5. TARTIŞMA (DISCUSSION)

YZ teknikleriyle öğrencilerin geçmiş performans verilerini, deneme sınav sonuçlarını ve diğer ilgili verileri kullanarak daha kapsamlı analizler yapılabilmektedir. YZ, verilerdeki gizli kalıpları ve eğilimleri ortaya çıkarmaktadır. Bu, öğrenci performansının daha derinlemesine anlaşılmasını sağlamaktadır. Bu çalışmadan farklı olarak, daha farklı yöntemler ve algoritmalarla performans değerlendirmeleri yapılabilir. Yapılan bu çalışmanın devamında öğrencilerin ilerlemesine göre dinamik olarak uyarlanan öğrenme planları ve öneriler sunulabilir. Bu, öğrencilerin ve öğretmenlerin daha hızlı ve verimli kararlar almasını sağlayabilir. Deneme sınavı sonuçları kısa sürede analiz edilerek öğrencilere hızlı geri bildirim sağlanabilmektedir.

Bu model, sadece YKS gibi benzer sınav ve tercih süreçlerinde de rehberlik edici bir araç olarak kullanılabilir. Modelin geliştirilmesi ve optimize edilmesi, öğrenci yönlendirme sistemlerinde yenilikçi bir yaklaşım sunmakta ve eğitim teknolojilerinde ileri bir adım temsil etmektedir. Geliştirilen modelin daha geniş kitlelere ulaşması ve farklı sınav türlerine uyarlanarak uygulanması, eğitim ve kariyer danışmanlığı alanında daha iyi bir gelecek için umut vaat etmektedir.

TEŞEKKÜRLER (ACKNOWLEDGMENTS)

Bu çalışma 102275 STB koduyla BK Mobil İletişim Hizmetleri ve Bilişim A.Ş. tarafından Ar-Ge projesi olarak desteklenmiştir. (This study has been supported as an R&D project by BK Mobil İletişim Hizmetleri ve Bilişim A.Ş. with the code 102275 STB.)

ETİK STANDARTLARIN BEYANI (DECLARATION OF ETHICAL STANDARDS)

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler. (The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.)

YAZARLARIN KATKILARI (AUTHORS' CONTRIBUTIONS)

Yazarlar tarafından eşit şekilde sonuçlar analiz etmiş ve makalenin yazım işlemi gerçekleştirilmiştir. (The results were analyzed equally by the author and the article was written.)

ÇIKAR ÇATIŞMASI (CONFLICT OF INTEREST)

Bu çalışmada herhangi bir çıkar çatışması yoktur. (There is no conflict of interest in this study.)

6. KAYNAKLAR (REFERENCES)

- [1] H. Atılgan, "Türkiye'de kademeler arası geçiş: Dünü-bugünü ve bir model önerisi," *Ege Eğitim Dergisi*, vol. 19, no. 1, pp. 1-18, 2018.
- [2] Ş. S. Gacanoğlu and C. Nakiboğlu, "Yükseköğretim Kurumları Sınavında Yer Alan Kimya Sorularının 2018 Yılı Kimya Dersi Öğretim Programı Kazanımlarına Göre Analizi," *Türkiye Kimya Derneği Dergisi Kısım C: Kimya Eğitimi*, vol. 7, no. 2, pp. 217-242, 2022.
- [3] Y. Uzunpolat and A. Çakmak, "Yükseköğretim Kurumları Sınavında Çıkan DKAB Sorularının Ortaöğretim DKAB Öğretim Programı Çerçevesinde Analizi," *Sirnak University Journal of Divinity Faculty/Sirnak Üniversitesi İlahiyat Fakültesi Dergisi*, no. 32, 2023.
- [4] N. Çokişler, "Turist Rehberliği Programlarına Yerleşen Adayların Üniversite Sınavı Başarı İstatistikleri Üzerine Betimsel Bir Analiz (2021)," *Yükseköğretim ve Bilim Dergisi*, vol. 12, no. 3, pp. 621-632, 2022.
- [5] B. Ersöz, H. İ. Bülbül, and Ş. Sağıroğlu, "Using LXP for Green Deal: A New Approach," in *2023 12th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2023: IEEE, pp. 524-529.
- [6] H. Bülbül and B. Ersöz, "Eğitimde yapay zekâ sanal gerçeklik ve sanal evren (Metaverse)," *Yapay zekâ ve büyük veri kitap serisi (4. Baskı, s. 149-183) içinde. Nobel Akademik Yayıncılık*, 2022.
- [7] F. Tahiru, "AI in education: A systematic literature review," *Journal of Cases on Information Technology (JCIT)*, vol. 23, no. 1, pp. 1-20, 2021.
- [8] A. Öter, B. Ersöz, H. İ. Bülbül, and Ş. Sağıroğlu, "Using Generative Artificial Intelligence in Exams: A Research on KPSS with ChatGPT," *International Journal of Educational Research Review*, vol. 9, no. 4, pp. 269-274.
- [9] H. Kirthika, S. Mayasre, and M. Balamurugan, "Score Predicting Web Application Using

- Machine Learning Techniques," in *2022 1st International Conference on Computational Science and Technology (ICCST)*, 2022: IEEE, pp. 157-161.
- [10] Z. Wang and Y. Shi, "Prediction of the admission lines of college entrance examination based on machine learning," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, 2016: IEEE, pp. 332-335.
- [11] D. Wang, D. Lian, Y. Xing, S. Dong, X. Sun, and J. Yu, "Analysis and prediction of influencing factors of college student achievement based on machine learning," *Frontiers in Psychology*, vol. 13, p. 881859, 2022.
- [12] E. Çakıt and M. Dağdeviren, "Predicting the percentage of student placement: A comparative study of machine learning algorithms," *Education and Information Technologies*, vol. 27, no. 1, pp. 997-1022, 2022.
- [13] B. Ujkani, D. Minkovska, and L. Stoyanova, "A machine learning approach for predicting student enrollment in the university," in *2021 XXX International Scientific Conference Electronics (ET)*, 2021: IEEE, pp. 1-4.
- [14] S. Tankuş, "Üniversite sınavına giren öğrencilerin sınav sonuçlarının yapay zeka ile analizi ve değerlendirmesi: 2018 yılı Şanlıurfa ili örneği/Analysis and evaluation of the exam results of the students attending the university examination with artificial intelligence: 2018 Şanlıurfa province," 2020.
- [15] E. Erdem and F. Bozkurt, "A comparison of various supervised machine learning techniques for prostate cancer prediction," *Avrupa Bilim ve Teknoloji Dergisi*, no. 21, pp. 610-620, 2021.
- [16] İ. Koyuncu and S. Gelbal, "Comparison of data mining classification algorithms on educational data under different conditions," *Journal of Measurement and Evaluation in Education and Psychology*, vol. 11, no. 4, pp. 325-345, 2020.
- [17] H. Jin, Y.-G. Kim, Z. Jin, A. A. Rushchitc, and A. S. Al-Shati, "Optimization and analysis of bioenergy production using machine learning modeling: Multi-layer perceptron, Gaussian processes regression, K-nearest neighbors, and Artificial neural network models," *Energy Reports*, vol. 8, pp. 13979-13996, 2022.
- [18] M. A. Jassim, "Analysis of the performance of the main algorithms for educational data mining: a review," in *IOP conference series: materials science and engineering*, 2021, vol. 1090, no. 1: IOP Publishing, p. 012084.
- [19] M. A. Baig, S. A. Shaikh, K. K. Khatri, M. A. Shaikh, M. Z. Khan, and M. A. Rauf, "Prediction of Students Performance Level Using Integrated Approach of ML Algorithms," *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 1, pp. 216-234, 2023.
- [20] S. A. Alwarthan, N. Aslam, and I. U. Khan, "Predicting student academic performance at higher education using data mining: a systematic review," *Applied Computational Intelligence and Soft Computing*, vol. 2022, 2022.
- [21] Ö. Ali, "Automatic Detection of Epileptic Seizures from EEG Signals Using Artificial Intelligence Methods," *Gazi University Journal of Science Part C: Design and Technology*, pp. 1-1, 2024.
- [22] M. A. A. Walid, S. M. Ahmed, M. Zeyad, S. S. Galib, and M. Nesa, "Analysis of machine learning strategies for prediction of passing undergraduate admission test," *International Journal of Information Management Data Insights*, vol. 2, no. 2, p. 100111, 2022.
- [23] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *Peerj computer science*, vol. 7, p. e623, 2021.
- [24] U. Verma, C. Garg, M. Bhushan, P. Samant, A. Kumar, and A. Negi, "Prediction of students' academic performance using Machine Learning Techniques," in *2022 International Mobile and Embedded Technology Conference (MECON)*, 2022: IEEE, pp. 151-156.
- [25] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geoscientific Model Development Discussions*, vol. 2022, pp. 1-10, 2022.