

Üniversite Yerleşme Başarısını Etkileyen Faktörlerin Yapay Zekâ Yöntemleriyle Araştırılması

Investigating Factors Affecting Success in University Placement Using Artificial Intelligence Methods

Ayhan Doğan, Cihan Ünal

Yazar Bilgileri

Ayhan Doğan

Dr., Hacettepe Üniversitesi,
Başkent OSB Teknik Bilimler
Meslek Yüksekokulu,
aydogan@hacettepe.edu.tr

Cihan Ünal

Dr., Hacettepe Üniversitesi,
Başkent OSB Teknik Bilimler
Meslek Yüksekokulu,
cihan.unal@hacettepe.edu.tr

ÖZ

Öğrencilerin akademik başarılarını etkilediği düşünülen birçok faktör bulunmaktadır. Bu çalışmada özellikle üniversite yerleşme başarısını etkileyen faktörler araştırılmıştır. Uzmanların görüşleri alınarak kırk soruluk bir anket hazırlanmış ve çeşitli alanlarda eğitim görmekte olan üniversite öğrencilerine uygulanmıştır. Anketten elde edilen veriler kullanılarak başarıyı en çok etkileyen faktörler önem sırasına göre yapay zekâ yöntemleriyle tahmin edilmiştir. Bu tahminlerin başarıları performans ölçüm metrikleriyle değerlendirilmiştir. Çalışmada dört farklı makine öğrenimi kullanılmıştır. Tahminlere ait performans metriklerine göre en başarılı yöntemler değişmektedir. Metrikler incelendiğinde en iyi sonuçlar, 4,95 MSE ve 2,22 RMSE değerleriyle Rastgele Orman yöntemi, 1,60 MAE değeriyle Aşırı Gradyan Yükseltme yöntemi, 0,36 MAPE değeriyle Lineer Regresyon yöntemidir. Destek Vektör Makineleri yönteminin başarısı ise tüm metriklere göre diğer yöntemlerden görece daha düşüktür. Çalışma sonucunda önem sırasına göre üniversite yerleşme başarısına etki eden faktörler dikkate alınarak öğrencilerin başarısını artırmanın mümkün olabileceği düşünülmektedir. Böylece eğitimciler, aileler, danışmanlar ve öğrenciler bu önem sırasını dikkate alacak şekilde mümkün olan alanlarda iyileştirmeler ve çalışmalar yapabilir, gerekli önlemleri alabilirler.

Makale Bilgileri

Anahtar Kelimeler

Makine öğrenimi
Yapay zekâ
Üniversite yerleşme başarısı
Başarıyı etkileyen faktörler

Keywords

Machine learning
Artificial intelligence
University placement success
Factors affecting success

Makale Geçmişi

Geliş: 17.02.2024
Kabul: 28.08.2024

ABSTRACT

There are several factors affecting students' success. This study investigates the factors that affect university placement success. A forty-question questionnaire, prepared with the opinions of experts, was applied to university students from various fields. The data obtained from the questionnaire was analyzed using artificial intelligence methods to predict the most important factors affecting success. The study evaluated the success of the prediction models with performance measurement metrics using four different machine learning methods. It is important to note that the most successful method varied depending on the performance metrics used for evaluation. The Random Forest method had the best results with 4.95 MSE and 2.22 RMSE values, followed by the Extreme Gradient Boosting method with a 1.60 MAE value and the Linear Regression method with a 0.36 MAPE value. Based on all metrics, the success rate of the Support Vector Machines method was relatively lower than that of other methods. The study suggests that by considering the factors affecting university placement success in order of importance, it can be possible to increase the students' success. Therefore, educators, families, counselors, and students can make improvements, conduct studies in the relevant areas, and take necessary measures to account for this order of importance.

Makale Türü

Araştırma

Önerilen Atıf

Doğan, A. & Ünal, C. (2024). Üniversite yerleşme başarısını etkileyen faktörlerin yapay zekâ yöntemleriyle araştırılması. *TEBD*, 22(3), 1678-1698. <https://doi.org/10.37217/tebd.1438947>

Giriş

Türkiye’de özellikle son dönemlerde eğitime önemli yatırımlar yapılmıştır. Ancak henüz bu yatırımların istenen sonuçları tam olarak alınamamıştır. Bu durum öğrencilerin yurt içinde ve yurt dışındaki sınav başarılarında görülmektedir. Başarıyı olumlu olumsuz etkileyen pek çok faktör bulunmaktadır. Öğrencilerin akademik başarılarının tahmin edilmesi eğitim kurumları için çok önemlidir. Çeşitli çalışmalarda öğrenci performansını etkilediği düşünülen parametreler araştırılmıştır. Bu çalışmalardan birisinde öğrencilerin demografik, sosyal, duygusal durumlarını kapalı uçlu sorulardan oluşan 32 soruluk anket kullanarak 4 farklı öğrenci grubuna uygulanmıştır. Anket sonuçlarından elde edilen veriler çeşitli makine öğrenimi yöntemleri kullanılarak tahmin edilmiştir. Kullanılan yöntemler içerisinde rastgele orman (RO) yöntemi en iyi sonuçları vermiştir (Selvi, 2020). Diğer bir çalışmada öğrencilerin mevcut sınav verilerine göre daha sonra girecekleri sınavlardaki performansları tahmin edilmiştir. En iyi performans değerleri K-En Yakın Komşu (KNN) yöntemi ile elde edilmiştir (Adak ve Duralioğlu, 2023). Güvenç vd. (2022) yaptıkları çalışmada öğrencilerin akademik başarılarını tahmin etmek ve eksiklerini gidermek amacıyla öğrencilerin dönem başı bilgisayar bilgi düzeylerini ve dönem sonu başarı notlarını makine öğrenimi yöntemleri kullanılarak araştırmışlardır. Aydoğan ve Karcı (2018) Bingöl Üniversitesi öğrencilerine uyguladıkları anket verilerini kullanarak makine öğrenimi ile öğrencilerin başarı durumlarına etki eden etkenleri analiz etmişler ve akademik performanslarını ortaya koymaya çalışmışlardır. Yapılan bir tez çalışmasında öğrencilerin akademik başarısını etkileyen faktörlerin tespit edilebilmesi için iki farklı veri seti kullanılarak RO, aşırı gradyan yükseltme (AGY) ve destek vektör makineleri (DVM) ile sınıflandırma modelleri kullanılmıştır (Kaya, 2022).

Günümüzde eğitim alanında dijitalleşmeye ilgi arttıkça öğrenme yönetim sistemleri (ÖYS) kullanımı da hızla yaygınlaşmaktadır. Öğrenciler bu ortamlarla girdikleri etkileşimlerde önemli ölçüde veri üretmekte ve bu veri üzerinde yapay zeka algoritmaları kullanılarak öğrenme sürecini anlamaya yönelik modeller geliştirilebilmektedir (Yavuzarslan ve Çiğdem, 2022). Bu alanda yapılan çok sayıda çalışma bulunmaktadır. Bu çalışmalardan bazıları ÖYS kullanılarak yapılan erken uyarı sistemleri hakkındadır (Kaban ve Bilen, 2021). Eğitim alanında yapılan çalışmalarda makine öğrenimi kullanımı gün geçtikçe artmaktadır. Bu çalışmalarda çeşitli makine öğrenimi yöntemleri tek başına veya birkaçı birlikte kullanılmıştır. Bu çalışmalardan yaygın olarak kullanılanlar makine öğrenimi yöntemlerine göre gruplandırılabilir.

- RO yönteminin kullanıldığı çalışmalar (Agustiniingsih vd., 2023; Chen vd., 2021; Elbadrawy vd., 2016; He vd., 2018; Hill vd., 2018; Samson vd., 2017; Sandoval vd., 2018),
- DVM yönteminin kullanıldığı çalışmalar (Chen vd., 2021; Dewan vd., 2015; Hussain vd., 2019; Marbouti vd., 2016),

- Lojistik Regresyon Modeli (LRM) yönteminin kullanıldığı çalışmalar (Agustiniingsih vd., 2023; Arnold, 2017; Chen vd., 2021; Hill vd., 2018; Hussain vd., 2019; Waddington vd., 2016),
- KNN yönteminin kullanıldığı çalışmalar (Dewan vd., 2015; Marbouti vd., 2015; Marbouti vd., 2016),
- Kümeleme Analizi (KA) yönteminin kullanıldığı çalışmalar (Brown vd., 2018; Hung vd., 2019; Stapel vd., 2016),
- Lineer Regresyon (LR) yönteminin kullanıldığı çalışmalar (Jokhan vd., 2019; Sletten vd., 2023; Yao ve Deng, 2020),
- AGY yönteminin kullanıldığı çalışmalar (Agustiniingsih vd., 2023; Nadar, 2023; Su vd., 2023).

Web of Science veri tabanı verilerine göre 2015-2020 yılları arasında bu alanda yapılan çalışmalarda %22,5 oranında karar ağaçları, %17,8 destek vektör makinesi, %14,2 navie bayes makine öğrenimi yöntemlerinin kullanıldığı görülmektedir. Makalelerde kullanılan örneklem profilleri incelendiğinde lisans öğrencileri %35,3 oranla en yüksek düzeydedir. Yapılan çalışmalardaki örneklem sayısı incelendiğinde en yüksek oran %29,7 ile 101-1000 kişi arasındadır (Tosunoğlu vd., 2021).

Bu çalışmada üniversite yerleşme başarısını etkileyen faktörleri belirlemek için bir anket çalışması yapılmış, elde edilen verilerle başarıyı en çok etkileyen faktörler önem sırasına göre makine öğrenimi yöntemleriyle tahmin edilmiştir.

Yöntem

Yapay zekâ, insan gibi düşünebilmesi için programlanmış, verileri kullanarak öğrenen, verilerden anlamlar çıkaran, kararlar üretebilen ve problem çözebilen yazılımlardır. Makine öğrenimi, konuya uygun veriler kullanılarak bilgisayarlara öğrenme ve tahmin yeteneği kazandırabilen bir yapay zekâ çeşididir. Bu yöntem sayesinde bilgisayarlar çeşitli görevleri veya problemleri çözmek için optimize edilirler. Problemleri geleneksel programlama yöntemleriyle çözmek zor olduğunda tercih edilen bir yapay zekâ yöntemidir. Bunun için veriler kullanılarak eğitimler gerçekleştirilir ve eğitim modelleri üretilir. Model eğitimleri sırasında veriler, eğitim, test ve doğrulama verileri olarak kullanılır. Doğrulama ve test işlemleri sayesinde modelin başarısı çeşitli ölçüm metrikleriyle ölçülebilir. Modellerin başarısı kullanılan verinin temsil kabiliyeti, sayısı ve kullanılan yöntemlere bağlıdır. Ancak yine de gerçek dünyada modelleri sürekli olarak optimize etmek ve başarılarını artırmak için yeni verilerle beslenmesi önemlidir (Jariwala vd., 2023).

Makine öğrenimi gerçekleştirilirken problemin şekli, mevcut veriler ve amaçlar doğrultusunda çeşitli öğrenme yöntemleri kullanılır. Bunlar, denetimli öğrenme, denetimsiz öğrenme, pekiştirmeli öğrenme ile yarı denetimli ve güçlendirilmiş öğrenme şeklinde gruplanabilir (Acı ve

Doğansoy, 2022). Tüm bu öğrenme yöntemlerini gerçekleştirmek için birçok algoritma geliştirilmiştir. Bunlara örnek olarak karar ağaçları (KA), LR, DVM, yapay sinir ağları (YSA), RO ve gauss süreç regresyonu (GSR) verilebilir (Doğan vd., 2023).

Bu çalışmada, öğrencilerin üniversite yerleşme başarısını etkileyen faktörler, bir anket aracılığıyla belirlenmiş ve makine öğrenimi yöntemleri kullanılarak değerlendirilmiştir. Araştırmanın amacı, üniversiteye yerleşmeye çalışan belirli bir öğrenci grubunun yerleşme başarısını etkileyen faktörleri önem sırasına göre ortaya koymaktır. Anket soruları, öğrencilerin akademik performansı, sosyo-ekonomik durumu, aile desteği, çalışma alışkanlıkları ve psikolojik durumları gibi çeşitli konuları kapsamaktadır. Elde edilen veriler, makine öğrenimi algoritmaları kullanılarak analiz edilmiş ve başarı üzerinde en büyük etkiye sahip faktörler önem sırasına göre belirlenmiştir. Bu tek durumlu, betimleyici nitel durum çalışması, öğrencilerin üniversite yerleşme sürecini etkileyen kritik faktörleri derinlemesine anlamamıza olanak sağlamaktadır. Bununla birlikte, araştırmada bazı nicel yöntemler de kullanılmıştır. Çalışma, belirli bir öğrenci grubunun geri bildirimlerine dayalı olarak gerçekleştirildiği için sonuçlar genel geçerlik taşımamakla birlikte, benzer bağlamlarda yapılan diğer araştırmalar için değerli ipuçları sunmaktadır. Çalışmada gerçekleştirilen işlem aşamaları genel olarak aşağıdaki şekildedir.

Veri Toplama Süreci

Üniversite yerleşme başarısını etkileyen faktörleri belirlemek amacıyla veri toplayabilmek için anket kullanılmıştır. Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi'nden elde edilen veriler kullanılarak üniversite yerleşme başarısını en çok etkileyen faktörler önem sırasına göre yapay zekâ yöntemleriyle tahmin edilmiştir. Üniversite yerleşme başarısını en çok etkileyen faktörlerin önem sırasını yapay zeka yöntemleriyle tahmin edebilen bu çalışmada veri toplama aracı olarak Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi geliştirilmiştir. Üniversiteye yerleşen bir öğrencinin başarısını etkileyen faktörlerin önem sırasını yapay zeka yöntemleri kullanarak tahmin edebilmek için geliştirilen ankette toplam 40 soru bulunmaktadır.

Anket tekniği, belirli bir konu hakkında bir evreni ya da örnekleme oluşturan bireylerden bilgi sağlamak için kullanılan sistemli veri toplama tekniğidir. Veri toplamak için yapılan anketler kişilerin, başarı ve yaşam düzeylerini, durumun etkilerinin oranını saptama, genel bilgiler elde edebilmek için bireylerin bilgilerini alma, mevcut durum içerisindeki riskleri tanımlama vb. gibi durumları saptamak için kullanılabilir (Yılmaz ve Aktürk, 2021). Veri toplama süreci aşağıdaki adımlardan oluşmaktadır.

Anket Tasarımı

Üniversite yerleşme başarısını etkileyen potansiyel faktörleri belirlemek ve ölçmek için kapsamlı bir anket tasarlanmıştır. Anket soruları, öğrenci profili, eğitim geçmişi, sosyoekonomik

durum ve öğrenme alışkanlıkları gibi çeşitli konuları içermektedir. Sorular rehberlik ve psikolojik danışmanlık uzmanları, eğitimciler ve sınavlara hazırlık konusunda çalışan uzmanların görüşleri alınarak hazırlanmıştır. Uygulanacak anket için Hacettepe Üniversitesi Rektörlüğü Sosyal ve Beşeri Bilimler Araştırma Etik Kurulundan izin alınmıştır.

Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi

Araştırmada veri toplama aracı olarak “Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi” geliştirilirken araştırmacılar tarafından anket soruları oluşturulmuştur. Oluşturulan bu sorularla beraber, yönerge ve demografik değişkenler bilgileriyle “anket taslak formu” haline getirilmiştir. Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi taslak formu, üniversite yerleşme başarısı üzerine araştırmalar gerçekleştirmiş uzman kişilerin (alanında beş uzman) görüşlerine sunulmuştur. Uzman kişiler, anket sorularına ilişkin kapsam geçerliği, içerik, yapı, uygulanabilirlik ve anlam yönünden değerlendirmeler gerçekleştirmiştir. Uzmanlar değerlendirmelerini gerçekleştirirken her bir soruya 1=uygun değil, 2=çok düzeltilmesi gerekli, 3=az düzeltilmesi gerekli, 4=çok uygun kullanarak 1-4 arası puan aralığında değerlendirmeler yapmışlardır. Uzman görüşlerinde uyumun varlığı, Kendall Uyuşum katsayısı değerine bakılarak kestirilmiştir. Uzmanlardan yaptıkları değerlendirmelerle Kendall’s analizi sonucunda uzmanların görüşleri arasında istatistiksel olarak anlamlı bir fark bulunmamıştır (Kendall’s $W=0.465$; $p=0.512$). Taslak anket, dil, anlatım ve ifade bakımından iki dil bilim uzmanı tarafından incelenmiştir. Dil bilim uzmanlarının yaptığı öneriler doğrultusunda soruların bir kısmına, yazım, biçim, dil ve anlatım bakımından revizyonlar yapılmıştır. Yapılan bütün bu işlemler sonucunda Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi yapılandırılmıştır. Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi’nin, 30 kişilik bir pilot uygulaması gerçekleştirilmiştir.

Pilot uygulamada katılımcılardan anket sorularının anlaşılabilirliğine ilişkin değerlendirmeler alınmıştır. Anket formu, bu pilot uygulamalardan gelen geri bildirimlere göre yeniden revize edilmiştir. En son aşamada Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi son haline getirilmiştir. Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi’nin çevrim içi formu da oluşturulmuştur.

Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi regresyon analizi nicel çalışmadır. İlişkisel korelasyon desen yöntemi kullanılmıştır. Bağımlı değişken öğrencinin sınav başarısıdır. Bağımsız değişkenler ise veri toplama aracında belirlenen 40 sorudur.

Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi’nin iç tutarlılığı (Cronbach Alfa Katsayısı - α) anket geliştirme sürecinde maddelerin bir yapıya ilişkin ölçümler gerçekleştirirken birbirleriyle uyumunu kestirmek amacıyla yani anketin iç tutarlılığını test etmek amacıyla kullanılır. Bu yaklaşımın anketteki benzer yapıyı ölçen ve benzer cevaplandırma gerektiren soruları içinde

kullanılabileceğine ilişkin görüşler mevcuttur. Bu nedenle Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi'nin soruları yapı bakımından sekiz alt boyuta ayrılmıştır. Bu alt boyutlar Tablo 1'de verilmiştir.

Tablo 1. Alt Boyutlar

Sıra No	Alt Boyut
1	Ailenin Eğitim ve Gelir Durumu
2	Okul Ortamı
3	Ders Çalışma ve Motivasyon
4	Lise ve Öncesi Başarı Durumu
5	Ailenin Yaklaşımı, Rehberlik, Dersane, Etüd ve Özel Ders
6	Okul Dışı Zamanı Değerlendirme
7	Ev Ortamı
8	Çevre İlişkileri

Bu boyutlarda sorulara verilen cevaplar benzer olduklarından dolayı her bir alt boyuttaki soruların tutarlığı, Cronbach Alfa katsayısı yöntemi ile yapı geçerliliği test edilmiştir.

Güvenilirlik Çalışması

Güvenilirlik; bir testin veya ölçeğin ölçmek istediği şeyi tutarlı ve istikrarlı bir biçimde ölçme derecesidir. Başka bir ifadeyle, bireylerin test maddelerine verdikleri yanıtlar arasındaki tutarlılıktır. Üniversite yerleşme başarısını etkileyen faktörler anket uygulamasından elde edilen veriler SPSS 20.0 paket programı kullanılarak analiz edilmiştir. Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi'ne ait iç tutarlılık anlamında Cronbach Alfa (α) güvenilirliğine bakılmıştır. İç tutarlılık anlamında güvenilirlik, tek bir ölçme aracının tek bir uygulama yapılarak anket içinde yer alan maddelerin kavramsal yapıyı tutarlı bir şekilde ölçüp ölçmediğinin belirlenmesi için yapılmaktadır. Ölçme aracının Cronbach Alfa (α) katsayısının çok yüksek düzeyde olması sadece ölçme aracının güvenilirliğini değil aynı zamanda yapı geçerliliğine de işaret etmektedir (Baykul, 1979). Güvenilir ölçme araçları, kendisini oluşturan maddeleri arasındaki yüksek düzeyde tutarlılık gerektirmektedir. Geliştirilmeye çalışılan Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi'nin güvenilirlik düzeyi, Cronbach Alfa güvenilirlik yöntemi kullanılarak hesaplanmıştır. Hesaplanan Cronbach Alfa güvenilirlik katsayısı her alt boyut için Tablo 2'de verilmiştir. Hesaplanan alt boyutların Alfa güvenilirlik katsayıları iç tutarlılık olduğunu ifade etmektedir. Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi'ne ilişkin elde edilen Cronbach Alfa katsayısı değerleri, ölçme aracının homojenliğinin bir göstergesi olarak kabul edilmekte olup hesaplanan Cronbach Alfa katsayısı 1'e yaklaştığı derecede ölçme aracının tek boyutlu bir yapıya sahip olduğu düşünülebilir. Özdamar (1999) güvenilirlik katsayısına ilişkin ölçüt değerleri aşağıda olduğu gibi ifade etmektedir. $0,00 \leq \alpha < 0,40$ olduğu zaman anket güvenilir değildir. $0,40 \leq \alpha < 0,60$ olduğu zaman anket düşük güvenilirliktedir. $0,60 \leq \alpha < 0,80$ olduğu zaman anket oldukça güvenilirdir. $0,80 \leq \alpha \leq 1,00$ olduğu zaman ise anket yüksek düzeyde güvenilir bir ankettir. Yukarıda verilen ölçüt değerlere bakıldığında anketin iyi düzeyde bir güvenilirlik katsayısına

sahip olduğu ve ölçeğin maddelerinin belirlenmesinde güvenilirlik düzeyi yüksek ölçme sonuçlarının elde edilebileceğini göstermektedir. Aynı zamanda yapı geçerliliğinin de yüksek olduğunu söylemek doğru olacaktır. Bağımsız değişkenler ankette kullanılan sorulardır. Ankette farklı alt boyutlar söz konusudur. Her alt boyut için ayrı ayrı güvenilirlik çalışmaları yapılmış, her alt boyut için Cronbach Alfa katsayıları Tablo 2’de verilmiştir.

Tablo 2. Alt Boyutlar ve Cronbach Alfa Katsayıları

<i>Sıra No</i>	<i>Alt Boyut</i>	<i>Cronbach Alfa Katsayısı(α)</i>
1	Ailenin Eğitim ve Gelir Durumu	0,725
2	Okul Ortamı	0,568
3	Ders Çalışma ve Motivasyon	0,652
4	Lise ve Öncesi Başarı Durumu	0,691
5	Ailenin Yaklaşımı, Rehberlik, Dersane, Etüd ve Özel Ders	0,529
6	Okul Dışı Zamanı Değerlendirme	0,643
7	Ev Ortamı	0,716
8	Çevre İlişkileri	0,649

Katılımcı Seçimi

Çalışmanın hedef kitleleri üniversite sınavına girmiş olan mevcut üniversite öğrencilerini kapsamaktadır. Bu öğrenciler arasından rastgele seçilmiş bir örneklem grubu üzerinde anket uygulanmıştır. Bu örneklem, genel öğrenci popülasyonunu temsil etme amacını taşıdığından ön lisans ve lisans öğrencilerinin farklı sınıflarında eğitim gören öğrencilerden oluşmaktadır.

Anket Uygulaması

Seçilen örneklem grubuna uygulanan anket, çevrim içi anket platformları kullanılarak gerçekleştirilmiştir. Katılımcılar ön lisans ve lisans eğitimi almakta olan üniversite öğrencileridir. Bu öğrenciler ağırlıklı olarak Hacettepe Üniversitesi öğrencilerinden oluşmakta olup az sayıda başka üniversitelerde eğitim gören öğrenciler de ankete katılmıştır. Ankete toplam 309 kişi katılmıştır. Bu sayı Web of Science veri tabanı verilerine göre 2015-2020 yılları arasında yapılan benzer araştırmalardaki en yüksek orana sahip olan örneklem sayı aralığındadır (Tosunoğlu vd., 2021).

Veri Analizi ve İşleme

Anket sonuçları elde edildikten sonra, veri analizi ve veri işleme adımları aşağıdaki şekilde gerçekleştirilmiştir.

Veri Temizleme

Anketten elde edilen veriler, eksik veya tutarsız değerler açısından kontrol edilmiş, çok az sayıda olan birkaç eksik veri tamamlanmış, tamamlanamayan veya tutarsız olan birkaç veri ise çalışmaya dahil edilmemiştir.

Veri Normalizasyonu

Veri setindeki değişkenleri benzer ölçeklere getirmek veya belirli bir aralığa sıkıştırmak için veri standardizasyonu ve normalizasyonu kullanılır. Normalleştirmenin amacı, veri setindeki farklı özelliklerin büyüklükleri veya birimleri arasındaki farklılıkları gidermektir. Böylece model performansları artar ve farklı ölçeklerin modeller üzerindeki olumsuz etkileri giderilmiş olur. Bu çalışmada makine öğrenimi modellerinin daha iyi performans göstermesi için veriler normalleştirilmiştir.

Makine Öğrenimi İşlemleri

Veri analizi sürecinin ardından, makine öğrenimi yöntemleri kullanılarak eğitimler, testler ve tahminler gerçekleştirilmiştir. Bu işleme ait aşamalar şu şekildedir.

Yöntem Seçimi

Çalışmada kullanılacak makine öğrenimi yöntemleri benzer problemlerde yaygın olarak başarıyla kullanılanlar arasından seçilmiştir. Bu çalışmada iki ayrı tahmin görevi gerçekleştirilmiştir. Bunlardan birincisi girdi verilerinden istenilen çıktının tahmin performansının makine öğrenimi yöntemleriyle gerçekleştirilmesi ve performans metrikleriyle bu başarının ölçülmesidir. İkincisi ise tahmin başarısına en çok katkısı bulunan girdi özelliklerinin belirlenmesidir. Yani başka bir deyişle üniversite başarısının tahmin edilebilmesi için en çok katkısı bulunan soruların tespit edilmesi işlemidir. Bu çalışmada DVM, LR, RO ve AGY makine öğrenimi yöntemleri tercih edilmiştir.

Model Eğitimi

Seçilen makine öğrenimi yöntemleri kullanılarak hazırlanan veri seti üzerinde eğitimler gerçekleştirilmiştir. Eğitim için verilerin %80'i kullanılmıştır.

Model Testi

Eğitilen modeller, test veri seti üzerinde değerlendirilmiştir. Test için verilerin %20'si kullanılmıştır.

Sonuçların Değerlendirilmesi

Elde edilen tahmin sonuçları ve modelin başarısı performans metrikleri kullanılarak analiz edilmiştir. Bu çalışmada ortalama mutlak hata (MAE), ortalama kare hata (MSE), ortalama mutlak yüzde hata (MAPE), kök ortalama kare hata (RMSE) metrikleri kullanılmıştır. Çalışmanın genel amaçları ve hipotezleri göz önünde bulundurularak elde edilen sonuçlar yorumlanmıştır.

Destek Vektör Makineleri

DVM, sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılan güçlü bir makine öğrenimi yöntemidir. İlk olarak Vapnik tarafından kullanılan bu yöntem, özellikle öğrenme modelinin genelleme yeteneğini iyileştirmek amacıyla tasarlanmıştır. Sınıflar arasında en iyi ayrımı sağlayacak

bir karar sınırı belirlemeye çalışır. Bunu da destek vektörleri kullanarak gerçekleştirir. Destek vektörleri, belirlenen sınıf sınırlarına en yakın veri noktalarını temsil eder ve genellikle modele olan katkıları nedeniyle kritik öneme sahiptir. Bu yöntemde çeşitli çekirdek fonksiyonları, veri noktalarını yüksek boyutlu uzaylara taşıyarak doğrusal olmayan ilişkilere olan duyarlılığı artırır (Doğan, 2023; Korkmaz vd., 2022; Soman vd., 2009).

DVM'nin geniş bir uygulama alanı vardır. Finansal analiz, metin madenciliği, görüntü tanıma ve daha birçok alanda başarıyla kullanılmaktadır.

Lineer Regresyon

LR, istatistik ve makine öğrenimi alanlarında yaygın olarak kullanılan temel bir modelleme yöntemidir. Bağımlı bir değişkenin bağımsız değişkenlerle doğrusal ilişkisinin modellenmesidir. Bu model, veri setindeki eğilimleri ve ilişkileri analiz etmek, öngörülerde bulunmak ve değişkenler arasındaki etkileşimleri anlamak için kullanılır. Bir doğru veya düzlem üzerinde en iyi uyan hatayı en aza indiren katsayıları bulmaya çalışır. Bunun için, genellikle en küçük kareler yöntemi ile kullanılır. Modelin performansı, regresyon katsayıları ve belirlilik katsayısı gibi istatistiksel ölçütlerle değerlendirilir. LR, basitliği ve etkililiği nedeniyle geniş bir uygulama yelpazesine sahiptir (Doğan vd., 2023; Korkmaz vd., 2023; Luu vd., 2021; Montgomery vd., 2021).

Rastgele Orman

RO, makine öğrenimi alanında özellikle sınıflandırma ve regresyon problemlerinde etkili olan topluluk öğrenme yöntemlerinden birisidir. Birden çok karar ağacını bir araya getirerek bir model oluşturur ve bu ağaçlar arasındaki çeşitliliği arttırmak için rastgelelik ilkesini kullanır. Her bir ağacın eğitim verilerinin bir alt kümesi üzerinde eğitilmesi ve ağaçların özelliklerinin farklı alt kümelerini kullanması ilkesine dayanır. Bu rastgelelik, her bir ağacın birbirinden bağımsız olmasını sağlar ve böylece genel modelin güçlü ve sağlam olmasına katkıda bulunur. Ağaçların sonuçlarını birleştirirken sınıflandırma problemlerinde oylama ve regresyon problemlerinde ortalama kullanarak bir tahmin yapar (Breiman, 2001; Rodriguez-Galiano vd., 2015; Wei vd., 2019).

Bu yöntemin başarısı, yüksek boyutlu ve gürültülü veri setlerinde etkili olabilme özelliğinden gelir. Aşırı öğrenmeye karşı dirençli olması, gürültülü verilerin etkisini azaltması önemli avantajlarından. Özellikle büyük ve karmaşık veri setleriyle çalışılacaksa en çok tercih edilen yöntemler arasındadır.

Aşırı Gradyan Yükseltme

AGY, özellikle sınıflandırma ve regresyon problemlerinde yüksek performans sağlayan bir makine öğrenimi algoritmasıdır. Gradyan yükseltme yöntemlerinin gelişmiş bir türü olan AGY, güçlü tahmin yetenekleri ve geniş uygulama alanları ile öne çıkmaktadır. Zayıf öğrenen modelleri bir araya

getirerek güçlü bir tahmin modeli oluşturur. Öğrenme sürecindeki ağaçların sayısını ve karmaşıklığını optimize etmek için aşırı gradyan yükseltme tekniğini kullanır. Bu, her bir ağacın önceki ağaçtan gelen hataları düzeltmeye odaklanarak, modelin tahmin doğruluğunu artırır. Özellik seçimi ve ağaç büyüklükleri gibi parametrelerin optimize edilebilirliği, bu yöntemin esnekliğini arttırarak farklı veri setleri ve problemlerine uyarlanabilirliğini sağlar (Doğan, 2023; Möller vd., 2016; Tamayo vd., 2016).

AGY'nin performans avantajları, büyük veri setleri ve karmaşık yapıdaki veri özellikleri ile başa çıkabilmesinden kaynaklanır. Ayrıca, aşırı öğrenmeye karşı direnç göstermesi ve genelleme yeteneği, modelin güvenilirliğini artırır.

Bulgular

Üniversite öğrencilerine uygulanan anket sorularından elde edilen veriler işlenmiş ve makine öğreniminde kullanılmıştır. Önce her bir makine öğrenimi yöntemi kullanılarak ayrı ayrı eğitimler gerçekleştirilmiş ve eğitimleri tamamlanan modeller test edilmiştir. Bu testlerden performans metrikleriyle elde edilen sonuçlar Tablo 3 ve Tablo 4'te verilmiştir. Tablo 3'te üniversite sınavında alınan sonucu tahmin etme başarısının performans metrikleriyle elde edilmiş sonuçları verilmektedir. Tablo 4'te ise tahmin başarısına en çok katkısı bulunan anket sorularının katkısına göre en büyükten en küçüğe doğru sıralanmış bir listesi verilmektedir. Tablo 3 üzerinde her bir performans metriği için en iyi sonuçlar koyu renkle gösterilmektedir. Buna göre MSE ve RMSE metriklerine göre en iyi sonuçlar RO yöntemiyle, MAE metriğine göre en iyi sonuç AGY yönteminde, MAPE metriğine göre en iyi sonuç LR yöntemiyle elde edilmiştir. Bununla birlikte bazı değerler birbirine yakın olduğundan genel anlamda en başarılı kabul edilebilecek yöntemler AGY ve RO yöntemleridir. Bunları sırasıyla LR ve DVM yöntemleri takip etmektedir.

Tablo 3. Makine Öğrenimi Yöntemlerine Ait Tahminlerin Test Performans Sonuçları

<i>Yöntem</i>	<i>MAE</i>	<i>MSE</i>	<i>MAPE</i>	<i>RMSE</i>
RO	1,74	4,95	0,40	2,22
DVM	1,82	5,47	0,46	2,34
LR	1,78	5,42	0,36	2,33
AGY	1,60	4,99	0,41	2,23

Tablo 4 incelendiğinde öğrencinin daha önceki başarı sıralaması, toplam çalışma süresi, çözdüğü soru sayısı, motivasyon durumu gibi faktörlerin başarısını yüksek oranda etkilediği görülmektedir. Bunun yanında kronik hastalık, özel okul veya özel ders alma, işte çalışma durumu, anne ve/veya babanın hayatta olup/olmama durumu gibi faktörlerin başarı üzerindeki etkisinin az olduğu görülmektedir.

Tablo 4. Makine Öğrenimi Yöntemleri Kullanılarak Üniversite Yerleşme Başarısına Etki Eden Faktörlerin Önem Sırasına Göre Sıralanmış Tahmin Sonuçları

Önem Sırası	Anket Sorusu
1	Liseye Giriş Sınavı başarı sıralamanız nedir?
2	Üniversite hazırlık döneminde toplam çözdüğünüz deneme kitapçık sayısı kaçtı?
3	Lise döneminde cep telefonunuzla günlük toplam geçirdiğiniz süre ne kadardı?
4	Üniversite sınavına hazırlık döneminde günlük ortalama çözdüğünüz soru sayısı kaçtı?
5	Babanızın eğitim durumu nedir?
6	Lise döneminizde genellikle hangi saatte yatarsınız?
7	Ailenizin aylık geliri nedir?
8	Üniversite sınavına hazırlık döneminde motivasyonunuzu bozan veya düşüren bir faktör var mıydı?
9	Lise döneminde sosyal medyada günlük toplam geçirdiğiniz süre ne kadardı?
10	Lise döneminde arkadaşlarınızla ders dışında ortalama günlük beraber geçirdiğiniz süre nedir?
11	Lise başarı ortalamanız nedir?
12	İlkokul eğitiminizde kaç tane sınıf öğretmeniniz oldu?
13	Üniversite sınavına hazırlandığınız dönemde günlük ortalama çalışma saatiniz kaçtı?
14	Lise son sınıftayken sınıf mevcudunuz kaçtı?
15	Annenizin eğitim durumu nedir?
16	Lisedeki devamsızlık durumunuz nasıldı?
17	Lise döneminizde evinizde sizin dışındaki kişilerin en geç yatış saati kaçtı?
18	Annenizin çalışma durumu nedir?
19	Üniversite sınavına hazırlık döneminde sizi daha çok kim motive etti?
20	Lise döneminde günlük televizyon izleme süreniz ne kadardı?
21	Üniversite sınavına hazırlandığınız dönemde nerede yaşıyordunuz?
22	Lise öğrenimi boyunca kaç okul değiştirdiniz?
23	Ailenizin haftalık misafir ağırlama veya misafirlığe gitme sıklığı nedir?
24	Lise döneminde dersane veya etüt merkezi eğitimlerine katıldınız mı?
25	Lise döneminizde kardeşlerinizden üniversite mezunu olan veya üniversitede okuyan var mıydı?
26	Cinsiyetiniz nedir?
27	Üniversite sınavına hazırlanırken anneniz veya babanız ders başarılarınızı ve devamsızlığınızı takip etti mi?
28	Lise döneminde aile bireylerinden sürekli tedavi gerektiren bir hastalığı, vefatı veya ailecek etkilendiğiniz doğal afet gibi üzücü bir olay yaşadınız mı?
29	Çocukluk döneminizden itibaren belirli bir meslek hedefiniz var mıydı?
30	Üniversite sınavına hazırlandığınız dönemde kendinize ait çalışma odanız var mıydı?
31	Lisede tek cinsiyetli bir sınıfta/okulda mı okudunuz?
32	Lise döneminde özel ders aldınız mı?
33	Lise döneminde yazları, hafta sonu, sürekli veya yarı zamanlı bir işte çalıştınız mı?
34	Sizin kronik bir rahatsızlığınız var mı?
35	Devlet lisesinden mi, özel liseden mi mezun oldunuz?
36	Üniversite sınavına hazırlandığınız dönemde rehberlik veya psikolojik destek aldınız mı?
37	Lise eğitiminde veya liseden önceki eğitimlerinizde hiç sınıfta kaldınız mı?
38	Babanız yaşıyor mu?
39	Anne ve babanızın birliktelik durumu nedir?
40	Anneniz yaşıyor mu?

Tartışma ve Sonuç

Tablo 3 incelendiğinde çeşitli makine öğrenimi yöntemlerinin tahmin başarılarının farklı olduğu görülmektedir. Aynı zamanda bu farklılık kullanılan farklı performans ölçüm metrikleriyle de ilgili olabilir. Bu nedenle makine öğrenimi çalışmalarında tek bir makine öğrenimi yöntemi ve tek bir performans ölçüm metriğinin kullanımının tahmin başarısını etkileyebileceği ve tahmin başarısı

üzerinde yanıtıcı olabileceğini göstermektedir. Öğrenci başarısını araştıran benzer çalışmalarda farklı makine öğrenimi yöntemlerinin sonuçlarının da değiştiği görülmektedir (Murat, 2017).

Tablo 4 incelendiğinde ilginç sonuçlar ortaya çıkmaktadır. Bu tabloda genel kaniya göre başarıya etkisinin yüksek olduğu düşünülen bazı soruların yanında hiç beklenmeyen bazı sonuçlar da ortaya çıkmıştır. Örneğin; lise giriş sınavındaki başarı sıralaması, çözülen toplam soru sayısı gibi başarıya doğrudan etkisinin olduğu düşünülen konular çalışma sonuçlarında da kendini göstermiştir. Fakat baba eğitim durumu, ailenin aylık geliri, ilkokulda sınıf öğretmeni değişikliği gibi etkisinin az olması beklenen bazı durumların da başarı üzerinde etkili olduğu görülmektedir. Bunun yanında lisede sınıfta kalıp/kalmadığı, devlet ya da özel lisede eğitim alma durumu, özel ders alma durumu gibi başarıda etkisinin önemli olduğu sanılan bazı konularda ise tam tersi bir durumun olduğu sonucu ortaya çıkmıştır. Güneş vd. (2012) yaptıkları çalışmada üniversite sınav başarısını etkileyen faktörler içerisinde öğretmen sayısının artması, okul donanım ölçütlerinin kalitesinin artması ve öğrencilerin dersaneler ile destek almasının başarıyı olumlu etkilediği sonucuna varmışlardır. Gürbüz vd. (2015) Adıyaman il örneğinde “ÖSYS başarısını etkileyen faktörler” konusunda yaptıkları çalışma sonucunda kesintisiz eğitim, sınavda uygulanan katsayılar, dersanelerin ve öğretmenlerin etkisi, sınav sistemi, sosyo-ekonomik durum, öğrenci profili ve teknolojinin öğrenci başarısını etkilediği sonuçlarına ulaşmışlardır.

Yapılan çalışmalardan bazıları bu çalışmada incelenen konuyla tam olarak aynı olmasa da başarıyı etkileyen faktörleri araştırmış ve çeşitli sonuçlara ulaşmışlardır. Çağdaş ve Erdem (2012) çalışmalarında veri madenciliği kullanarak ekonomik, sosyal, kişisel ve çevresel değişkenleri araştırmışlar, araştırmacı kişiliğin ve yüksek lise mezuniyet ortalamasının başarıyı olumlu etkilediğini, mezuniyet sonrası iş bulamama kaygısının ve istemediği bölümde okumasının başarıyı olumsuz etkilediğini, cinsiyetin ise belirleyici olmadığı sonucuna varmışlardır. Aydoğan ve Karcı (2018) “Meslek Yüksekokulu Öğrencilerinin Başarı Performanslarının Makine Öğrenmesi Yöntemleri ile Analizi” isimli çalışmalarında öğrencilerin orta eğitimi aldıkları ilde ön lisans programına devam etmelerinin ve gelir durumlarının iyi olmasının başarıyı artırdığını, sınavsız geçişin ise başarıyı düşürdüğü sonucuna ulaşmışlardır. Rençber (2012) yaptığı çalışmada “İstatistik ve Ulaşım Teknolojisi” derslerindeki başarının nedenlerini araştırmış, derse devam sayısı ile öğrenci başarısı arasında güçlü bir ilişki olduğunu fakat lise not ortalaması ile düşük ilişkili olduğunu bulmuştur. Özer ve Sarı (2016) öğrencinin dinlenme süresinin, sınıfının, bölümünün, okulu sevip sevmediğinin, derslere devamsızlığının, sınava hazırlandığı notların öğrenci başarısı üzerinde %5 önem düzeyinde istatistik bakımdan anlamlı olduğu sonucuna ulaşmışlardır.

Geçmişte yapılan birçok çalışmada öğrenci başarısını araştıran çalışmalar yapılmış olmasına rağmen bu çalışmada ortaya konulduğu şekliyle başarıyı etkileyen faktörler ve bunların önem sırası

birçok çalışmada ortaya konmamıştır. Bu konuda Selvi (2020) yaptığı çalışmada ilköğretimden liseye geçiş yapan öğrenciler üzerinde uyguladığı anket çalışmasında demografik, sosyal/duygusal ve kapalı uçlu sorular sormuş ve verilen cevapları kullanarak çeşitli makine öğrenimi işlemleri gerçekleştirmiştir. Bu çalışmasında öğrencilerin lise türlerinin tahmini ve farklı makine öğrenimi yöntemlerinin başarıları karşılaştırılmış, öğrencinin başarısına etki eden faktörler ortaya konmamıştır. Gülçin ve Çokluk-Bökeoğlu (2012) üniversite öğrencilerinin akademik başarılarını etkileyen değişkenler anketi yaparak uyguladıkları makine öğrenimi çalışmalarıyla öğrencilerin başarılı olup olamayacaklarını tahmin etmeye çalışmışlar, ancak başarıyı etkileyen değişkenlerle ilgili bir sonuca ulaşmamışlardır. Kaya (2022) sosyo-ekonomik, demografik ve sosyolojik bazı faktörlere dayalı veriler kullanarak yaptığı çalışmadan elde ettiği sonuçlar içerisinde başarıyı etkileyen nedenler bulunmamakta bunun yerine makine öğrenimi yöntemlerinin başarıları karşılaştırılmaktadır.

Çalışmada uzman görüşlerinde uyumun varlığı, Kendall Uyuşum katsayısı değerine bakılarak kestirilmiştir. Uzmanlardan yaptıkları değerlendirmelerle Kendall's analizi sonucunda uzmanların görüşleri arasında istatistiksel olarak anlamlı bir fark bulunmamıştır. Üniversite Yerleşme Başarısını Etkileyen Faktörler Anketi'nin soruları yapı bakımından sekiz alt boyuta ayrılmıştır. Her bir alt boyuttaki soruların tutarlığı, Cronbach Alfa katsayısı yöntemi ile yapı geçerliliği ve güvenilirlik testiyle değerlendirilmiştir. Alt boyutlara göre Cronbach Alfa katsayısı sonuçları; "Ailenin Eğitim ve Gelir Durumu" için 0,725, "Okul Ortamı" için 0,568, "Ders Çalışma ve Motivasyon" için 0,652, "Lise ve Öncesi Başarı Durumu" için 0,691, "Ailenin Yaklaşımı, Rehberlik, Dersane, Etüd ve Özel Ders" için 0,529, "Okul Dışı Zamanı Değerlendirme" için 0,643, "Ev Ortamı" için 0,716, "Çevre İlişkileri" için 0,649 çıkmıştır. Bu sonuçlara göre güvenilirlik iki alt boyutta düşük çıkarken altı alt boyut için oldukça güvenilir çıkmıştır.

Tüm makine öğrenimi çalışmalarından elde edilen tahminlerde olduğu gibi elbette ki bu çalışmadan elde edilen sonuçlar da kesinlik ifade etmemektedir ve ankete katılan örneklem grubunun temsil kabiliyeti ve sayısı ile sınırlıdır. Ancak yine de bu sonuçlar birçok insan tarafından dile getirilen bazı varsayımların sorgulanması ve daha derinlemesine araştırılması gerektiğini ortaya koymaktadır.

Sonuçlar kullanılan makine öğrenimi yöntemlerinin bu alanda başarıyla kullanılabileceğini göstermektedir. Bununla birlikte insan davranışları ve bu davranışlar sonucunda ortaya çıkan sonuçların tam olarak sadece bu çalışmada kullanılan anket soruları ve yöntemlerle tespit edilebilmesi elbette mümkün değildir. Yine de bu ve benzer çalışmalar öğrenci davranışlarını analiz etmek, takip etmek ve başarıyı artırabilmek için önemli bilgiler sunmaktadır. Öğrenci davranışlarını ve başarısını analiz edebilecek anketler hazırlamak, bu anketler için etik onay almak ve yeterli örneklem sayısına ulaşmak oldukça zor ve zaman alan işlemlerdir. Bununla birlikte ileride daha büyük örneklem sayısına ulaşarak ve yeni makine öğrenimi yöntemlerini de deneyerek konunun daha derinlemesine

araştırılması düşünülmektedir. Böylece belki de gelecekte anket çalışmalarına daha az gereksinim duyulacak ve makine öğrenimi yöntemleri kullanılarak daha başarılı tahminler yapılabilecektir. Bu sayede zaman tasarrufu elde edilebileceği, öğrenci başarısının artırılacağı ve alanyazına yeni katkılar sağlanabileceği düşünülmektedir.

Kaynaklar

- Acı, M. & Doğanşoy, A. G. (2022). Makine öğrenmesi ve derin öğrenme yöntemleri kullanılarak e-perakende sektörüne yönelik talep tahmini. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 37(3), 1325-1340. <https://doi.org/10.17341/gazimmfd.944081>
- Adak, M. F. & Duralioğlu, Ö. (2023). Makine öğrenmesi yöntemleri kullanılarak öğrencilerin kazanım bilgileri ile sınavlardaki başarı durumunun tahmini. *Journal of Intelligent Systems: Theory and Applications*, 6(1), 43-51. <https://doi.org/10.38016/jista.1183353>
- Agustiningih, A., Findawati, Y., & Kautsar, I. A. (2023). Classification of vocational high school graduates' ability in industry using extreme gradient boosting (xgboost), random forest, and logistic regression. *Jurnal Teknik Informatika (Jutif)*, 4(4), 977-985. <https://doi.org/10.52436/1.jutif.2023.4.4.945>
- Arnold, K. E. (2017). *The effects of educational technology usage profiles and legally protected bio-demographic data on behaviorally-based predictive student success models in learning analytics: An exploratory study* (Doktora Tezi). ProQuest Dissertations and Theses database (UMI No. 10269454).
- Aydoğan, M. & Karcı, A. (2018). *Meslek yüksekokulu öğrencilerinin başarı performanslarının makine öğrenmesi yöntemleri ile analizi*. 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies'de sunulmuş bildiri, Ekim 2018, Kızılcahamam, Ankara. <https://www.ismsitconf.org/?go=ismsit2018> sayfasından erişilmiştir.
- Baykul, Y. (1979). *Örtük özellikler ve klasik test kuramları üzerine bir karşılaştırma* (Doktora Tezi). <https://tez.yok.gov.tr> sayfasından erişilmiştir.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Brown, M., DeMonbrun, R. M., & Teasley, S. (2018). Taken together: conceptualizing students' concurrent course enrollment across the post-secondary curriculum using temporal analytics. *Journal of Learning Analytics*, 5(3), 60-72. <http://dx.doi.org/10.18608/jla.2018.53.5>
- Chen, K., Huang, M., Zhu, X., & Wang, G. (2021). *Learning disability early warning system based on classification algorithm*. 2nd International Conference on Information Science and Education'da sunulmuş bildiri, Kasım 2021, Chongqing, Çin. <https://ieeexplore.ieee.org/document/9742515> sayfasından erişilmiştir.

- Çağdaş, K. & Erdem, O. A. (2012). Öğrenci başarısını etkileyen faktörlerin veri madenciliği yöntemleriyle incelenmesi. *Politeknik Dergisi*, 15(2), 111-116.
- Dewan, M. A. A., Lin, F., Wen, D., & Kinshuk. (2015). *Predicting dropout-prone students in e-learning education system*. 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops'da sunulmuş bildiri, Ağustos, 2015, Pekin, Çin. <https://ieeexplore.ieee.org/abstract/document/7518496> sayfasından erişilmiştir.
- Doğan, A. (2023). Makine öğrenimi yöntemleri kullanılarak Türkiye'nin kuzeybatısı için deprem tahmini. *Yerbilimleri*, 44(2), 166-178. <https://doi.org/10.17824/yerbilimleri.1325321>
- Doğan, A., Korkmaz, M., & Kirmaci, V. (2023). Estimation of ranque-hilsch vortex tube performance by machine learning techniques. *International Journal of Refrigeration*, 150, 77-88. <https://doi.org/10.1016/j.ijrefrig.2023.01.021>
- Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. (2016). Predicting student performance using personalized analytics. *Computer*, 49(4), 61-69. <https://doi.org/10.1109/MC.2016.119>
- Gülçin, U. & Çokluk-Bökeoğlu, Ö. (2012). Yükseköğretimde öğrenci başarılarının sınıflandırılmasında yapay sinir ağları ve lojistik regresyon yöntemlerinin kullanılması. *Akdeniz İnsani Bilimler Dergisi*, 3(2), 71-79.
- Güneş, S., Görmüş, Ş., Yeşilyurt, F., & Tuzcu, G. (2012). ÖSYS başarısını etkileyen faktörlerin analizi. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*(11), 71-81.
- Gürbüz, R., Erdem, E., Temurtaş, A., & Aygen, K. (2015). ÖSYS başarısını etkileyen faktörler: Adıyaman ili örneği. *Adıyaman University Journal of Educational Sciences*, 5(1), 49-71.
- Güvenç, E., Sakal, M., Çetin, G., & Özkaraca, O. (2022). Öğrencilerin dersteki niteliklerinin makine öğrenmesi teknikleri kullanılarak sınıflandırılması. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 10(3), 1359-1371. <https://doi.org/10.29130/dubited.1017202>
- He, L., Levine, R. A., Bohonak, A. J., Fan, J., & Stronach, J. (2018). Predictive analytics machinery for STEM student success studies. *Applied Artificial Intelligence*, 32(4), 361-387. <https://doi.org/10.1080/08839514.2018.1483121>
- Hill, F., Fulcher, D., Sie, R., & De Laat, M. (2018). *Balancing accuracy and transparency in early alert identification of students at risk*. International Conference on Teaching, Assessment, and Learning for Engineering'da sunulmuş bildiri, Aralık 2018, Wollongong, NSW, Avustralya. <https://ieeexplore.ieee.org/abstract/document/8615370> sayfasından erişilmiştir.

- Hung, J.-L., Shelton, B. E., Yang, J., & Du, X. (2019). Improving predictive modeling for at-risk student identification: a multistage approach. *IEEE Transactions on Learning Technologies*, 12(2), 148-157.
- Hussain, M., Zhu, W., Zhang, W., Abidi, S. M. R., & Ali, S. (2019). Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review*, 52(1), 381-407. <https://doi.org/10.1007/s10462-018-9620-8>
- Jariwala, N., Putta, C. L., Gatade, K., Umarji, M., Ruhina-Rahman, S. N., Pawde, D. M., ..., & Shunmugaperumal, T. (2023). Intriguing of pharmaceutical product development processes with the help of artificial intelligence and deep/machine learning or artificial neural network. *Journal of Drug Delivery Science and Technology*, 87, 104751. <https://doi.org/10.1016/j.jddst.2023.104751>
- Jokhan, A., Sharma, B., & Singh, S. (2019). Early warning system as a predictor for student performance in higher education blended courses. *Studies in Higher Education*, 44(11), 1900-1911. <https://doi.org/10.1080/03075079.2018.1466872>
- Kaban, A. & Bilen, Ö. (2021). Eğitimde kullanılan erken uyarı sistemleri konusunda yapılmış çalışmaların incelenmesi. *MANAS Sosyal Araştırmalar Dergisi*, 10(2), 788-797.
- Kaya, F. H. (2022). *Makine öğrenmesi algoritmaları kullanarak öğrencilerin akademik başarısını etkileyen faktörlerin tespit edilmesi* (Yüksek Lisans Tezi). <http://tez.yok.gov.tr> sayfasından erişilmiştir.
- Korkmaz, M., Dogan, A., & Kirmaci, V. (2022). Performance analysis of counterflow ranque-hilsch vortex tube with linear regression, support vector machines and gaussian process regression method. *Gazi J. Eng. Sci*, 8(2), 361-370. <https://doi.org/10.30855/gmbd.0705015>
- Korkmaz, M., Doğan, A., & Kırmaç, V. (2023). Thermal temperature estimation by machine learning methods of counterflow ranque-hilsch vortex tube using different fluids. *Heat Transfer Research*, 54(12), 61-79. <https://doi.org/10.1615/HeatTransRes.2023046884>
- Luu, Q.-H., Lau, M. F., Ng, S. P., & Chen, T. Y. (2021). Testing multiple linear regression systems with metamorphic testing. *Journal of Systems and Software*, 182, 111062. <https://doi.org/10.1016/j.jss.2021.111062>
- Marbouti, F., Diefes-Dux, H., & Madhavan, K. (2015). *Predictive modeling for identifying at-risk students using course performance data*. The 6th Research in Engineering Education Symposium'da sunulmuş bildiri, Temmuz 2015, Dublin, İrlanda. <https://www.researchgate.net> sayfasından erişilmiştir.
- Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, 103, 1-15. <https://doi.org/10.1016/j.compedu.2016.09.005>

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Möller, A., Ruhlmann-Kleider, V., Leloup, C., Neveu, J., Palanque-Delabrouille, N., Rich, J., ..., & Pritchett, C. (2016). Photometric classification of type Ia supernovae in the SuperNova Legacy Survey with supervised learning. *Journal of Cosmology and Astroparticle Physics*, 2016(12), 008. <https://doi.org/10.1088/1475-7516/2016/12/008>
- Murat, G. (2017). Makine öğrenmesi yöntemleri ile akademik başarının tahmin edilmesi. *Gazi University Journal of Science Part C: Design and Technology*, 5(3), 139-148.
- Nadar, N. (2023). Enhancing student performance prediction through stream-based analysis dataset using modified xgboost algorithm. *International Journal on Information Technologies & Security*, 15(2), 75-87.
- Özdamar, K. (1999). *Paket programlar ile istatistiksel veri analizi 1*. Kaan.
- Özer, H. & Sarı, A. (2016). Kovaryans analizi modelleriyle üniversite öğrencilerinin başarılarını etkileyen faktörlerin belirlenmesi: Atatürk Üniversitesi İİBF öğrencileri için bir uygulama. *Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, 24(2), 105-126.
- Rençber, B. A. (2012). Üniversite öğrencilerinin akademik başarılarını etkileyen faktörler. *Çankırı Karatekin Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 3(1), 191-198.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804-818. <https://doi.org/10.1016/j.oregeorev.2015.01.001>
- Samson, P. J., Czarnik, A., & Gross, M. (2017). *Relationships between digital measures of student engagement and exam scores: is the LMS enough?* 7th International Learning Analytics & Knowledge Conference (LAK17): Practitioner Track'da sunulmuş bildiri, Mart 2017, Vancouver, Kanada. <https://www.researchgate.net> sayfasından erişilmiştir.
- Sandoval, A., Gonzalez, C., Alarcon, R., Pichara, K., & Montenegro, M. (2018). Centralized student performance prediction in large courses based on low-cost variables in an institutional context. *The Internet and Higher Education*, 37, 76-89. <https://doi.org/10.1016/j.iheduc.2018.02.002>
- Selvi, A. (2020). *Bilecik ilinde ilköğretimden liseye geçiş sınavlarında makine öğrenmesi yöntemleri ile öğrenci başarısının tahmini* (Yüksek Lisans Tezi). <http://tez.yok.gov.tr> sayfasından erişilmiştir.
- Sletten, M. A., Tøge, A. G., & Malmberg-Heimonen, I. (2023). Effects of an early warning system on student absence and completion in Norwegian upper secondary schools: a cluster-randomised

- study. *Scandinavian Journal of Educational Research*, 67(7), 1151-1165. <https://doi.org/10.1080/00313831.2022.2116481>
- Soman, K., Loganathan, R., & Ajay, V. (2009). *Machine learning with SVM and other kernel methods*. PHI Learning.
- Stapel, M., Zheng, Z., & Pinkwart, N. (2016). *An ensemble method to predict student performance in an online math learning environment*. 9th International Conference on Educational Data Mining'da sunulmuş bildiri, Haziran 2016, Raleigh, NC. <https://eric.ed.gov/?id=ED592647> sayfasından erişilmiştir.
- Su, W., Jiang, F., Shi, C., Wu, D., Liu, L., Li, S., ..., & Shi, J. (2023). An xgboost-based knowledge tracing model. *International Journal of Computational Intelligence Systems*, 16(1), 1-9. <https://doi.org/10.1007/s44196-023-00192-y>
- Tamayo, D., Silburt, A., Valencia, D., Menou, K., Ali-Dib, M., Petrovich, C., ..., & Paradise, A. (2016). A machine learns to predict the stability of tightly packed planetary systems. *The Astrophysical Journal Letters*, 832(2), L22. <https://doi.org/10.3847/2041-8205/832/2/L22>
- Tosunoğlu, E., Yılmaz, R., Özeren, E., & Sağlam, Z. (2021). Eğitimde makine öğrenmesi: Araştırmalardaki güncel eğilimler üzerine inceleme. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, 3(2), 178-199.
- Waddington, R. J., Nam, S., Lonn, S., & Teasley, S. D. (2016). Improving early warning systems with categorized course resource usage. *Journal of Learning Analytics*, 3(3), 263-290.
- Wei, J., Chu, X., Sun, X. Y., Xu, K., Deng, H. X., Chen, J., ..., & Lei, M. (2019). Machine learning in materials science. *InfoMat*, 1(3), 338-358. <https://doi.org/10.1002/inf2.12028>
- Yao, D. & Deng, X. (2020). *A learning situation early warning method based on linear regression*. International Conference on Big Data and Informatization Education'da sunulmuş bildiri, Nisan 2020, Zhangjiajie, Çin. <https://ieeexplore.ieee.org/abstract/document/9150254> sayfasından erişilmiştir.
- Yavuzarslan, M. & Çiğdem, E. (2022). Öğrenme yönetim sistemi log kayıtlarının akademik başarı tahmininde kullanılması. *Bilişim Teknolojileri Dergisi*, 15(2), 199-207. <https://doi.org/10.17671/gazibtd.837884>
- Yılmaz, E. & Aktürk, A. (2021). *Z kuşağı bir nesli anlamak*. Palet.

Extended Summary

Despite substantial investments in education in recent years, Türkiye has yet to achieve the desired levels of success. This issue becomes evident when examining students' performance in both domestic and international exams. According to Murat (2017), factors influencing success have not

been clearly defined. In this context, predicting student success in advance holds crucial significance for educational institutions.

Various parameters affecting student performance have been explored in literature studies. Selvi (2020) conducted a study evaluating students' demographic, social, and emotional conditions through a questionnaire with closed-ended questions. The data obtained were predicted using machine learning methods, with the random forest method identified as the most effective. Another study by Adak and Duralioğlu (2023) demonstrated the k-Nearest Neighbors method's effectiveness in predicting future performances based on current exam data.

In a study by Güvenç et al. (2022), student achievements were predicted and deficiencies addressed using computer knowledge levels and end-of-term grades. Aydoğan and Karıcı (2018) analyzed factors influencing student success among Bingöl University students through machine learning methods based on survey data. Kaya (2022) utilized classification models, such as random forest, extreme gradient boosting, and support vector machines, to determine factors influencing student success using two distinct datasets.

With the increasing digitization in education, the use of Learning Management Systems (LMS) is rapidly becoming widespread. Students generate significant data through these systems, enabling the development of models using artificial intelligence algorithms to understand the learning process (Yavuzarslan and Çiğdem, 2022). Some studies in this field focus on early warning systems using LMS (Kaban and Bilen, 2021).

The use of machine learning methods in educational studies is on the rise. According to the Web of Science database, decision trees were preferred in 22.5% of studies conducted between 2015 and 2020, support vector machines in 17.8%, and naive Bayes machine learning methods in 14.2%. Examining sample profiles in these studies reveals that undergraduate students constitute the highest percentage at 35.3%. Moreover, the majority of samples, at 29.7%, fall within the range of 101-1000 participants (Tosunoğlu et al., 2021).

Artificial intelligence, programmed to think like humans, learn from data, derive meaning from data, generate decisions, and solve problems, is a software category. Machine learning, a subtype of artificial intelligence, imparts learning and prediction abilities to computers using relevant data. Training is conducted in machine learning using data, and training models are generated. During model training, data is used as training, testing, and validation data. The success of models can be measured through various metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Model success depends on the representation capability, quantity, and methods used for the data. However,

continuous optimization and feeding of models with new data are crucial for enhancing their success in real-world scenarios (Jariwala et al., 2023).

In this study, a survey was conducted to identify factors influencing university placement success, and machine learning methods were employed to predict the most impactful factors in order of importance. The data collection process included survey design, participant selection, and survey implementation. During the survey design phase, a comprehensive questionnaire covering various topics such as student profiles, educational backgrounds, socioeconomic status, and learning habits was created. Survey questions were formulated with input from guidance and counseling experts, educators, and experts in exam preparation, and their validity was ensured through expert review. Subsequently, the survey was administered to university students selected according to predefined criteria, resulting in 309 students from different universities and departments participating.

The survey results were analyzed using machine learning methods, including random forest (RF), support vector machines (SVM), extreme gradient boosting (XGB), and linear regression (LR) algorithms. Additionally, analyses were conducted to predict student success.

The analysis of the collected data using machine learning methods unveiled the most influential factors impacting student success. Noteworthy among these factors were the student's educational background, exam preparation period, socioeconomic status, and learning habits. The results emphasized the significance of the student's high school achievement score and exam preparation period as the most influential factors.

Furthermore, machine learning studies were instrumental in predicting student success, offering a hierarchical order of factors influencing student outcomes. Evaluation based on metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) revealed that the random forest method excelled in MSE and RMSE metrics, the XGBoost method yielded the best results in MAE metric, and the linear regression method outperformed in MAPE metric.

Upon closer examination of the importance assigned by machine learning, intriguing insights were gained. While some traditionally expected factors aligned with the anticipated impact on success, such as high school entrance exam rankings and the total number of solved questions, unexpected outcomes emerged. Factors traditionally considered to have minimal impact, such as the father's educational status, family monthly income, and changes in the primary school teacher, were found to be influential. Conversely, some factors traditionally believed to have a significant impact, including repeating a grade in high school, receiving education in a state or private high school, and taking private lessons, exhibited a contrasting influence.

The findings of this study have significant implications for educational institutions, students, and families. Recommendations include placing greater emphasis on students' exam preparation processes, providing socioeconomic support, and instilling appropriate learning habits. The study's outcomes offer valuable insights for enhancing educational practices, aligning strategies with identified success factors.

In conclusion, this study serves as an exemplary application of machine learning methods in education, offering a detailed exploration of factors influencing student success in the Turkish context. The integration of technology and data-driven approaches has the potential to revolutionize educational practices, providing tailored support and interventions for students. However, the study also underscores the need for continued research and development efforts to ensure the sustainability and generalizability of such models.

Educational institutions must strategically plan the adoption of technological methods, continually updating programs to enhance the knowledge and skills of educators and administrators in educational technologies. This approach ensures the effective use of technology to drive positive outcomes in education. The study opens avenues for future research, encouraging further exploration of machine learning applications in diverse learning environments and student populations to gain a more comprehensive understanding of its impact.

Arařtırmacıların Katkı Oranı Beyanı

Birinci arařtırmacı makalenin yazılması, yapay zekâ algoritmalarının belirlenmesi ve uygulanması, gerekli yazılımların kodlanması, yöntem belirlenmesi ve uygulanması, alanyazının genişletilmesi, ikinci arařtırmacı anket hazırlanması ve uygulanması, alanyazın taraması, verilerin toplanması ve işlenmesinde katkı sağlamıştır.

Destek ve Teşekkür Beyanı

Bu arařtırmada herhangi bir kurum, kuruluş ya da kişiden destek alınmamıştır.

Çatışma Beyanı

Arařtırmacıların arařtırma ile ilgili diđer kişi ve kurumlarla herhangi bir kişisel ve finansal çıkar çatışması yoktur.

Etik Kurul Beyanı

Bu arařtırma, Hacettepe Üniversitesi Rektörlüğü Sosyal ve Beşeri Bilimler Arařtırma Etik Kurulunun 06.09.2023 tarih ve E-66777842-900-00003055772 sayılı onayı ile yürütülmüştür.