Elektronik Sosyal Bilimler Dergisi www.esosder.org
Electronic Journal of Social Sciences e-ISSN:1304-0278
Nisan/April(2025) - Cilt/Volume:24 - Sayı/Issue:2 (946-964)

Makine Öğrenimi Modelleri ile Yetişkin Eğitimi Analizi: Modellerin Karşılaştırmalı Performansı

Analysing Adult Education with Machine Learning Models: The Comparative Performance of Models

Kamil Abdullah EŞİDİR[1]

Öz

Çalışmada, Türkiye İstatistik Kurumu'nun (TÜİK) 2022 Yetişkin Eğitimi Araştırması mikro veri seti kullanılarak, bireylerin örgün eğitime katılım durumu ve bu durumu etkileyen faktörler analiz edilmiştir. Analiz edilen veri seti, 14 değişken ve 24.462 gözlemden meydana gelmektedir. Veri ön işleme adımlarından sonra veri seti, eğitim ve test verileri olarak dengeli şekilde dağıtılmıştır. Sayısal veriler Standard Scaler yöntemi ile ölçeklendirilmiştir. Model performansları kesinlik, duyarlılık, F1 skoru ve doğruluk gibi metrikler ile ölçülmüş ve karmaşıklık matrisleri ile görselleştirilmiştir. Örgün eğitime katılım durumu bağımlı değişken olarak belirlenmiştir. Analizler sonucunda Gradient Boosting, XGBoost ve Cat Boost modelleri test doğruluğunda %94 başarı göstermiştir. Gradient Boosting ve LightGBM modelleri 0,92 AUC değeri ile ön plana çıkarken, Cat Boost ve XGBoost modelleri ise 0,91 AUC değerine ulaşmıştır. F1 skorunda ise Gradient Boosting ve LightGBM modelleri 0,75 ile en yüksek değerleri elde etmiştir. Örgün eğitime katılımda "yaş", "eğitim seviyesi" ve "medeni durum" faktörleri belirleyici olmuştur. Cat Boost modeli; dengeli performansı, yüksek doğruluk oranı, kesinlik ve AUC değerleri ile en uygun model olarak önerilmiştir. Çalışma sonucunda makine öğrenimi modellerinin, geleneksel istatistiksel yöntemlere kıyasla karmaşık veri yapılarındaki ilişkileri tespit etme kapasitesi anlaşılmıştır. Çalışma, Türkiye'deki eğitim politikalarının şekillendirilmesinde makine öğrenimi modellerinin kullanımı hakkında önemli bilgiler barındırmaktadır. Veriye dayalı analizler sayesinde eğitim politikaları daha etkin hale getirilebilir.

Anahtar Kelimeler: Makine Öğrenimi, LightGBM, Random Forest, Gradient Boosting, Cat Boost.

Abstract

In this study, the participation status of individuals in formal education and the factors affecting this status are analysed using the micro data set of the 2022 Adult Education Survey of the Turkish Statistical Institute (TurkStat). The analysed dataset consists of 14 variables and 24,462 observations. After the data preprocessing steps, the dataset was evenly distributed as training and test data. Numerical data were scaled with the Standard Scaler method. Model performances were measured with metrics such as precision, sensitivity, F1 score, and accuracy and visualised with complexity matrices. Formal education participation status is defined as the dependent variable. As a result of the analysis, Gradient Boosting, XGBoost and Cat Boost models achieved 94% success in test accuracy. Gradient Boosting and LightGBM models stand out with an AUC value of 0.92, while Cat Boost and XGBoost models achieved an AUC value of 0.91. In the F1 score, Gradient Boosting and LightGBM models obtained the highest values with 0.75. "Age", "education level" and "marital status" factors were determinants for participation in formal education. Cat Boost model was recommended as the most appropriate model with its balanced performance, high accuracy, precision and AUC values. The study revealed the capacity of machine learning models to detect relationships in complex data structures compared to traditional statistical methods. The study concludes that machine learning models can be used in shaping education policies in Türkiye.

Keywords: Machine Learning, LightGBM, Random Forest, Gradient Boosting, Cat Boost.

[1] Dr., Firat Development Agency, abdullahesidir@yahoo.com, https://orcid.org/0000-0002-8106-1758

Makale Türü/Article Type: Araştırma Makalesi/Research Article
Geliş Tarihi/Received Date: 27.11.2024 – Kabul Tarihi/Accepted Date: 14.02.2025
Atıf İçin/For Cite: EŞİDİR K.A., "Makine Öğrenimi Modelleri ile Yetişkin Eğitimi Analizi: Modellerin Karşılaştırmalı Performansı", *Elektronik Sosyal Bilimler Dergisi, 2025;24(2):946-964*
https://doi.org/10.17755/esosder.1589887
License: CC BY-NC 4.0

Giriş

Türkiye'de 18 yaş ve üzeri bireylerin mesleki ve kişisel bilgi becerilerini artırmak amacıyla katıldıkları eğitim faaliyetlerine ilişkin kapsamlı veri sunan Yetişkin Eğitimi Araştırması, ilk olarak 2007 yılında gerçekleştirilmiş ve 2012, 2016 ile 2022 yıllarında ise güncellenerek, eğitim politikalarına yön vermede değerli bir kaynak haline gelmiştir. Türkiye İstatistik Kurumu'nun (TÜİK) gerçekleştirdiği bu araştırma, bireylerin eğitim alma motivasyonlarını ve karşılaştıkları engelleri tespit etmeyi hedeflemektedir. Avrupa Birliği'nin (AB) hayat boyu öğrenme yaklaşımı ile uyumlu biçimde yürütülen bu çalışma, ülke genelindeki eğitim ihtiyaçlarını belirleyerek, eğitim programlarının geliştirilmesine ve revize edilmesine katkı sağlamaktadır. Özellikle bilgi ekonomisine geçiş sürecinde, ülkelerin sürdürülebilir kalkınma hedeflerine ulaşmasında yetişkin eğitimi kritik bir role sahiptir. Türkiye'de yetişkin eğitime katılım oranları, Avrupa Birliği ortalamalarının oldukça altındadır ve bu durum, eğitim politikalarının etkinliğini artırmak için yeni yaklaşımlara duyulan ihtiyacı ortaya koymaktadır (TÜİK Yetişkin Eğitimi Araştırması, 2022).

Bu çalışma, Türkiye'deki yetişkin eğitiminin analizine yönelik makine öğrenimi yöntemlerini kullanmaktadır. Geleneksel istatistiksel yöntemler genellikle doğrusal ilişkiler üzerine kurulu olup, karmaşık veri yapılarındaki çok boyutlu ilişkileri tam anlamıyla modelleyememektedir (Eşidir ve Gür, 2024). Oysa makine öğrenimi teknikleri, büyük ve karmaşık veri setlerinden öğrenerek değişkenler arasındaki gizli kalıpları tespit edebilme yeteneğine sahiptir (Tosunoğlu vd., 2021). Bu yönüyle, çalışmamız yetişkin eğitimi faaliyetlerine katılımı etkileyen faktörleri derinlemesine analiz etmeyi ve de eğitim politikalarına yönelik daha doğru öngörüler sunmayı hedeflemektedir.

Geleneksel istatistiksel yöntemler incelendiğinde; regresyon analizleri, lojistik regresyon ve ANOVA gibi tekniklerin, belirli koşullar altında ve sınırlı değişkenlerle etkili olabildiği görülmektedir (Zhu vd., 2016). Ancak, büyük veri setleri ve çok boyutlu veri yapıları söz konusu olduğunda, bu yöntemler yetersiz kalmakta ve karmaşık ilişkileri açığa çıkarmada başarısız olmaktadır (Ji, 2023). Buna karşın, makine öğrenmesi teknikleri ise, büyük ve karmaşık veri kümelerinden öğrenme yeteneği sayesinde, geleneksel yöntemler ile aşılamayan birçok zorluğu kolaylıkla aşabilmektedir (Bae vd., 2021). Makine öğrenimi modellerinin seçilmesinin temel nedeni, verinin karmaşıklığını ve çeşitliliğini daha iyi yönetebilme yeteneğidir. Örneğin, RandomForest ve LightGBM gibi modeller, büyük veri setlerinde ve çok boyutlu veri yapılarında etkin sonuçlar üreterek değişkenlerin etkileşimlerini yakalayabilmektedir (Pakarinen vd., 2022). Ayrıca, bu türden modeller eksik veri problemlerine karşı dayanıklılık gösterebilmekte ve aşırı uyum ve ezberlemeye (overfitting) karşı önlemler içermektedir. Bu nedenlerden ötürü, çalışmada makine öğrenimi modelleri tercih edilmiştir; böylelikle daha yüksek doğruluk oranlarına sahip tahminler elde edilmesi hedeflenmiştir.

1. Literatür Taraması

Son zamanlarda yapılan çalışmalar, geleneksel istatistiksel yöntemlerin kapsamlı veri kümelerindeki karmaşık ilişkileri ve gizli kalıpları ortaya çıkarmada yeterli olmayabileceğini ve daha gelişmiş teknik ve modellerin kullanılmasını gerektirdiğini göstermiştir (Speer, 2021). Makine öğrenimi algoritmaları kullanılarak, daha geniş bir değişken ve etkileşim yelpazesini dikkate alan daha doğru tahmin modelleri geliştirilebilir (Lim vd., 2024). Makine öğrenimi algoritmaları, daha geniş bir değişkenler dizisini ve bunların etkileşimlerini dikkate alarak daha kapsamlı bir yaklaşımlar sunmaktadır (Wu, 2023).

Yetişkin eğitimi, Avrupa Birliği tarafından hayat boyu öğrenme kapsamında 1980'lerden bu yana desteklenmektedir. Bu bağlamda, Türkiye'de gerçekleştirilen çalışmalar,

AB ülkelerindeki uygulamalara paralel olarak yaygın ve örgün eğitim faaliyetlerinin önemini ortaya koymaktadır

Güleç (2012), yaşam boyu öğrenmenin sosyal, ekonomik ve kültürel dönüşümlerle birlikte bir gereklilik haline geldiğini vurgulamaktadır. Çalışma, yaşam boyu öğrenmenin bireylerin mesleki, kişisel ve sosyal gelişimi için kritik bir öneme sahip olduğunu ve eğitim politikalarının, bilgi toplumunun gerekliliklerine uyum sağlamak için yeniden yapılandırılması gerektiğini ifade etmektedir.

Kaya (2014) çalışmasında, küreselleşmenin etkisiyle yetişkin eğitiminin giderek piyasa odaklı hale geldiğini ve sosyal devlet anlayışının bu alandaki etkisinin azaldığını vurgulamaktadır. Kaya'ya göre, yeni liberal ideolojinin yaygınlaşmasıyla, eğitim dezavantajlı grupların erişiminden uzaklaşarak, ekonomik gerekliliklere hizmet eden bir yapıya dönüşmüştür. Kaya, yetişkin eğitiminin insancıl değerlerden uzaklaştırılarak ekonomik bir araç haline getirilmesine eleştirel bir yaklaşım sunmaktadır.

Komşu (2017) çalışmasında, şiirin yetişkin eğitiminde kullanımı üzerine yoğunlaşmaktadır. Antik çağlardan bu yana eğitim materyali olarak kullanılan şiirin, metaforlar ve benzetmeler yoluyla öğrenme sürecine katkıda bulunduğu ve bireylerin empati kurma yetilerini geliştirdiği vurgulanmaktadır.

Polatoğlu (2022), Türkiye'de yetişkin eğitimine yönelik politika belgelerini 11. Kalkınma Planı ve 2019-2023 Milli Eğitim Bakanlığı stratejik planı çerçevesinde incelemiştir. Çalışma, yetişkin eğitiminin OECD ortalamalarına göre eksiklikler taşıdığını ve COVID-19'un katılım oranlarını olumsuz etkilediğini vurgulamaktadır. Türkiye'nin toplumsal ve ekonomik kalkınma hedefleriyle uyumlu ancak uygulamada eksiklikler içeren hayat boyu öğrenme sistemini güçlendirmek için yeni stratejiler geliştirmesi gerektiği belirtilmiştir.

Zilyas ve Yılmaz (2023) çalışmasında, ortaokul öğrencilerinin eğitim başarısını tahmin etmek ve başarıyı etkileyen faktörleri belirlemek için makine öğrenimi yöntemlerini kullanmıştır. Rastgele Orman algoritması, %88 doğruluk ve 0.98 R-Kare değerleri ile en yüksek performansı sergilemiştir. Çalışma, öğrencinin Türkçe dersindeki başarısını en çok etkileyen faktörlerin aile geliri ve ders çalışma süresi olduğunu ortaya koymuştur. Elde edilen bulgular, makine öğrenimi yöntemlerinin eğitim alanında öğrenci başarısını analiz etmek ve etkili politikalar geliştirmek için güçlü bir araç olduğunu göstermektedir.

Sinap (2024) çalışmasında, 2002-2022 yılları arasında eğitimde makine öğrenmesi üzerine yapılan 2.851 bilimsel yayını analiz ederek, bu alandaki gelişimi ve eğilimleri değerlendirmiştir. Çalışmada; "eğitsel veri madenciliği" ve "öğrenme analitiği" en fazla çalışılan konular olurken, ABD ve Çin'in ise en üretken ülkeler olduğu vurgulanmıştır. Ayrıca, eğitimde makine öğrenmesi araştırmalarının son yıllarda epeyce artış gösterdiği ve disiplinler arasında bir etki oluşturarak çeşitli alanlara katkı sağladığı anlatılmıştır.

2. Metodoloji

Büyük veri, çeşitlilik, hız ve hacimden ötürü çeşitli zorluklar barındırmaktadır (Suthaharan, 2014). Makine öğrenmesi veriye dayalı problemlerin çözümünde, uygun algoritmalarla modelleme yapmayı hedefleyen bir çeşit hesaplama disiplinidir (El Naqa ve Murphy, 2015). Analizlerde kullanılan mikro veri setinde eksik veri bulunmamaktadır. Veri seti, analizler öncesinde eğitim (%80) ve test (%20) olarak dengeli bir biçimde ikiye ayrılmıştır. Sayısal değişkenler Standard Scaler yöntemiyle ölçeklendirilmiştir. Çalışmada kullanılan mikro veri setinin çok boyutlu yapıda olması ve doğrusal olmayan ilişkileri modelleyebilme kapasitesinden ötürü yapılan analizlerde makine öğrenmesi modelleri tercih edilmiştir. Analizlerde LightGBM, Random Forest, Cat Boost, Gradient Boosting ve XGBoost makine

öğrenimi modelleri kullanılmıştır. Model performansları, kesinlik, duyarlılık, F1 skoru ve doğruluk gibi metriklerle değerlendirilmiş ve karmaşıklık matrisleri kullanılarak görselleştirilmiştir. Model performanslarının iyileştirilmesi için GridSearchCV yöntemiyle hiperparametre optimizasyonu gerçekleştirilmiştir. Çalışma, veriye dayalı eğitim politikalarının oluşturulması açısından önemli bir katkı sunmakta olup, gelecekte daha geniş veri setleri ve farklı modelleme yaklaşımlarıyla desteklenebilir.

2.1. Veri Seti

Çalışmada, TÜİK'in 2022 Yetişkin Eğitimi Araştırması mikro veri seti kullanılmıştır. Bu veri seti, bireylerin eğitim faaliyetlerini "Örgün Eğitim" ve "Yaygın Eğitim" olmak üzere iki ana kategoriye ayırmaktadır. Örgün eğitim, ilköğretim, ortaöğretim ve yükseköğretim gibi yapılandırılmış programları kapsarken; yaygın eğitim, kurslar, seminerler ve diğer gayri resmi öğrenme faaliyetlerini içermektedir. Söz konusu veri seti, bireylerin eğitim katılım dinamiklerini anlamak ve eğitim politikalarına yönelik öneriler geliştirmek amacıyla kapsamlı bir temel sunmaktadır. Mikro veri setinde eksik verileri azaltmak ve de analiz aşamalarını kolaylaştırmak adına 14 değişken kullanılmıştır.

Analizi gerçekleştirilen veri setinde toplamda 24.462 gözlem ve 14 değişken bulunmaktadır. Değişkenler, katılımcıların demografik özellikleri, eğitim katılım durumları ve diğer eğitimle ilişkili faktörleri kapsamaktadır. Bu değişkenler ve açıklamaları detaylı olarak Tablo 1'de sunulmuştur. Bu kapsamlı veri seti, analizlerin metodolojik doğruluğunu sağlamak ve eğitim faaliyetlerinin çok boyutlu bir şekilde değerlendirilmesine olanak tanımaktadır. Örgün eğitim faaliyetlerine katılım durumu ORGUN01 değişkeni bağımlı değişken olarak seçilmiştir. Diğer 13 değişken ise bağımsız değişkenler olarak analizlerde kullanılmıştır.

Tablo 1. Çalışmada Kullanılan Veri Seti, Değişkenler ve Açıklamaları

Değişken Adı	**Açıklama**	**Seçenekler**
HANE_GELIR	Hane halkı gelir dilimi	1 = 1. dilim (En düşük) 2 = 2. dilim 3= 3. dilim 4 = 4. dilim 5 = 5. dilim (En yüksek)
CINSIYET	Katılımcının cinsiyeti	1 = Erkek 2 = Kadın
YAS	Katılımcının yaşı	18-69 70+
MEDENI_DURUM	Katılımcının medeni durumu	1= Hiç evlenmedi 2= Evli 3= Boşandı 4= Eşi öldü
OKUL_BITEN	Bitirilen en yüksek eğitim seviyesi	000= Bir okul bitirmedi 100= İlkokul 200= Ortaokul, ilköğretim 344= Lise 550= 2 veya 3 yıllık yüksekokul 600= 4 yıllık yüksekokul veya fakülte 700= Yükseklisans 800= Doktora

Tablo 1. (devam)

FAAL_DURUM	Katılımcının iş durumu	11= Tam zamanlı çalışan 12= Yarı zamanlı çalışan 20= İşsiz/İş arıyor 31= Eğitim/öğretime devam ediyor (çırak, stajer dahil) 32= Emekli veya işten ayrılmış 33= Engelli 35= Ev işleri ile meşgul 36= Diğer -1= Cevap yok
EGITIM_ARASTIRMA	Eğitime katılım	1= Evet 2= Hayır
YAYGIN01	Kurslara katılım	1= Evet 2= Hayır
YAYGIN02	Seminer ve çalıştaylara katılım	1= Evet 2= Hayır
YAYGIN03	İşbaşında rehberli temel eğitime katılım	1= Evet 2= Hayır
YAYGIN04	Özel derslere katılım	1= Evet 2= Hayır
GENEL_SAGLIK	Genel sağlık durumu	1= Çok iyi 2= İyi 3= Ne iyi ne kötü, Orta 4= Kötü 5= Çok kötü -1= Cevap yok
SAGLIK_KISITLAMA	Sağlık nedeniyle sınırlanma durumu	1= Kısıtlanıyor 2= Kısıtlanıyor ama çok değil 3= Kısıtlanmıyor -1= Cevap yok
ORGUN01-Bağımlı Değişken	Örgün eğitim faaliyetlerine katılım	1= Evet 2= Hayır

Veri seti, bireylerin demografik, sosyo-ekonomik özellikleriyle (gelir, cinsiyet, yaş, medeni durum, eğitim seviyesi, iş durumu), sağlık durumları ve yaygın/örgün eğitime katılım durumlarını içermektedir.

2.2. Veri Ön İşleme

Veri ön işleme, veri setinin analiz ve modelleme için hazır hale getirilmesi sürecidir. Bu aşamada gerçekleştirilen işlemler, modellerin doğruluğunu ve genel performansını önemli ölçüde etkilemektedir (Gür, 2024a). Modellerin eğitilmesi sırasında ihtiyaç duyulmayan çok fazla eksik veriler içeren sütunlar, veri setinden çıkarılmıştır Hedef (bağımlı) değişken ORGUN01, "Evet" değeri için 0 ve "Hayır" değeri için 1 olacak şekilde ikili formata dönüştürülmüştür. Kategorik değişkenler, makine öğrenmesi algoritmalarıyla uyumlu hale getirilmek için sayısal değerlere dönüştürülmüştür. Ardından, dönüştürülen veri seti, eğitim (%80) ve test (%20) setlerine ayrılmıştır. Hedef değişken sınıflarının eğitim ve test veri setinde dengeli şekilde dağılmasını sağlamak amacıyla Stratify Parametresi kullanılmıştır. Sayısal veriler, Standard Scaler kullanılarak ölçeklendirilmiştir. Bu işlem, özellikler arasındaki ölçek farklılıklarını gidererek modelin daha etkili öğrenmesine yardımcı olmaktadır (Zeiler ve Fergus, 2014). Matematiksel olarak bu işlem Denklem 1'de gösterildiği gibi ifade edilmektedir:

$$z = \frac{(x-\mu)}{\sigma} \quad (1)$$

Burada, x ölçeklendirilecek olan özellik değerini, μ bu özelliğin ortalamasını, σ bu özelliğin standart sapmasını ve z ölçeklendirilmiş sonucu temsil etmektedir.

2.3. Model Performans Değerlendirme Kriterleri

Makine öğrenmesinde analizi gerçekleştirilen modellerin performansını objektif bir şekilde değerlendirmek için çeşitli metrikler ve yöntemler kullanılmaktadır (Gür, 2024c). Bu çalışmada, makine öğrenmesi modellerinin performansları, Kesinlik (Precision), Duyarlılık (Recall), F1 Skoru ve Doğruluk (Accuracy) metrikleri ile ölçülmüştür. Bu metriklerin, matematiksel formülleri sırasıyla Denklem 2, 3, 4 ve 5'de gösterilmiştir. Denklemlerde, DP doğru pozitif, YP yanlış pozitif, DN doğru negatif ve YN yanlış negatiftir. Ayrıca, tahminlerin doğruluğunu görsel olarak değerlendirebilmek için karmaşıklık matrisleri de kullanılmıştır.

$$Kesinlik\ (Precision) = \frac{DP}{DP+YP} \quad (2)$$

$$Duyarlılık\ (Recall) = \frac{DP}{DP+YN} \quad (3)$$

$$F1\ Skoru = 2 \times \frac{Kesinlik \times Duyarlılık}{Kesinlik + Duyarlılık} \quad (4)$$

$$Doğruluk\ (Accuracy) = \frac{DP+DN}{DP+DN+YP+YN} \quad (5)$$

Makine öğrenmesi modellerinin performansını değerlendirmek için kullanılan Kesinlik, Duyarlılık, F1 Skoru ve Doğruluk metrikleri, farklı senaryolarda modelin gücünü ve zayıflıklarının belirlenmesi açısından önemlidir. Özellikle Kesinlik, modelin pozitif tahminlerinde ne kadar seçici olduğunu ölçerken, Duyarlılık ise gerçek pozitifleri bulma konusundaki yeteneğini ifade etmektedir (Gür, 2024c). Karmaşıklık matrisleri, bu metrikleri görsel olarak temsil eden önemli bir araçtır. Karmaşıklık matrisleri sayesinde modelin hangi durumlarda hata yaptığını, yanlış pozitif ve yanlış negatif tahminleri detaylı bir şekilde görülüp anlaşılabilmektedir.

3. Makine Öğrenmesi Modelleri

Makine öğrenmesi, verilerden öğrenerek ve bu öğrendiklerini kullanarak tahminlerde bulunma veya karar verme üzerine kurulu bir yapay zeka alt dalıdır. Makine öğrenmesi modelleri, veri türüne ve problemin karmaşıklığına göre seçilmektedir. Kullanılan veri setinde, bireylerin yaş, cinsiyet, eğitim seviyesi gibi birden fazla değişkeni içerdiği göz önünde alındığında, makine öğrenmesi modellerinin bu tür karmaşık ilişkileri modellemedeki gücü açıkça anlaşılmaktadır. Bu bölümde sınıflandırma işlemi için kullanılan makine öğrenmesi modelleri matematiksel formülleri ile birlikte açıklanmaktadır. Makine öğrenmesi, bilgisayarların öğrenme yeteneği kazanmasını sağlamak amacıyla çeşitli algoritma, modelleme ve tekniklerin geliştirildiği bir bilim alanıdır. Çözüm süreçlerinde farklı matematiksel ve istatistiksel yöntemler kullanılmakla birlikte, yöntemlerin seçimi veri kümesinin özelliklerine göre belirlenmektedir (Akbulut ve Adem, 2023).

3.1. LightGBM (Hafif Gradyan Artırma Makinesi)

LightGBM, Microsoft firması tarafından geliştirilen, özellikle büyük veri setleri ve çok boyutlu özellikler üzerinde hızlı ve etkili sonuçlar sunabilen bir gradyan artırma algoritmasıdır. Yaprak bazlı bölme stratejisi ve histogram tabanlı yaklaşımı ile yüksek performans ve düşük hesaplama maliyeti sağlamaktadır.

Algoritma, daha büyük gradyanlara sahip veri örneklerine odaklanır ve bu örnekler üzerinde daha fazla bölünme gerçekleştirir. Bu, algoritmanın hızını ve verimliliğini artırır (Tong vd., 2024). LightGBM, seviye-dengeleme yerine yaprak-dengeleme stratejisi kullanır. Yani, her seferinde en iyi bölünmeyi sağlayan yaprağı seçer, bu da modelin daha az yaprakla daha iyi performans göstermesini sağlar. Ek olarak, kategorik özellikleri doğrudan işleyebilir ve bu özellikler için optimal bölünmeleri bulabilir. Bu, özellik mühendisliği sürecini basitleştirir ve veri ön işlemesini azaltır (Nagassou vd., 2023).

LightGBM, histogram tabanlı bir algoritma olup kategorik öznitelikleri doğrudan işleyebilir. Karar ağaçlarının eğitim süresi, yapılan hesaplamalar ve bölünme sayısıyla doğru orantılıdır; bu da kısa eğitim süresi ve düşük kaynak kullanımı sağlar. Yaprak bazlı bölme stratejisi sayesinde, diğer algoritmalara göre kaybı daha fazla azaltarak daha yüksek doğruluk değerlerine ulaşır (Kuş vd., 2021).

LightGBM'in temelinde, aşağıda Denklem 6'da gösterilen formülü yatmaktadır:

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (6)$$

Burada, $L^{(t)}$, t adımındaki toplam kaybı, l kayıp fonksiyonunu, y_i, i'nci gerçek değeri, $\hat{y}_i^{(t-1)}$, $t-1$ adımındaki tahmini, $f_t(x_i)$, t adımında model tarafından eklenen yeni fonksiyonu ve $\Omega(f_t)$, yeni fonksiyonun karmaşıklık cezasını temsil etmektedir.

3.2. Random Forest (Rastgele Orman)

RandomForest, farklı alt veri setlerinde birden fazla karar ağacının birleşiminden oluşur. Rastgele Orman (Random Forest), sınıflandırma ve regresyon görevleri için kullanılan, topluluk öğrenme metoduna dayanan bir makine öğrenimi algoritmasıdır. Bu algoritma, birden fazla karar ağacını bir araya getirerek her bir ağacın zayıflıklarını dengeler ve genel olarak daha güçlü ve dengeli bir model oluşturur (Oguine ve Oguine, 2021).

Rastgele Orman, eğitim veri setinden birçok bootstrap örneği çeker (her örnek, orijinal veri setinin rastgele seçilmiş örneklerini içerir). Her bir bootstrap örneği için bağımsız bir karar ağacı kurulur. Her ağaç, veri setinin farklı bir alt kümesi üzerinde eğitilir. Her düğümdeki bölünme işlemi, tüm özellikler yerine rastgele seçilen bir alt küme kullanılarak yapılır. Bu, modelin varyansını azaltmaya yardımcı olur ve ağaçlar arası korelasyonu düşürür. Sınıflandırma için tahmin yaparken genellikle Denklem 8'de gösterildiği gibi bir ifade kullanılır:

$$y = \text{mod}\{y_1, y_2, \ldots, y_n\} \quad (7)$$

Burada, $y_1, y_2, \ldots, y_n$ her bir karar ağacının tahmin ettiği sınıf etiketleridir ve mod fonksiyonu en sık rastlanan etiketi belirler.

3.3. Cat Boost (Categorical Boosting)

Cat Boost Yandex firması tarafından geliştirilen, özellikle kategorik veriler ile iyi şekilde çalışan bir boosting algoritmasıdır. Adını "Categorical Boosting" kelimelerinin birleşiminden alır ve kategorik değişkenlerin dönüştürülmesi sırasında oluşan bilgi kaybını minimize etmek için tasarlanmıştır. Diğer boosting algoritmalarına kıyasla daha az ön işleme gerektirir, bu da kullanıcıya zaman kazandırır ve modelin performansını artırır. Cat Boost'un en öenmli avantajlarından birisi de ezberlemeyi (overfitting) önleyebilme kapasitesidir.

Cat Boost, kategorik öznitelikleri etkin bir şekilde işlemek üzere tasarlanmış bir Gradyan Arttırma Karar Ağacı (GBDT) algoritmasıdır. Derin öğrenme modellerinin aksine, büyük veri gereksinimi olmadan, az veriyle de yüksek doğruluk sağlayabilir. Performanslı, kullanıcı dostu ve kategorik verileri otomatik işleme yeteneğiyle öne çıkar. Geleneksel GBDT algoritmalarından farklı olarak, önişleme aşaması yerine eğitim süreci boyunca kategorik öznitelikleri ele alır. Bu yaklaşımı sayesinde, kategorik verilerin işlenmesinde sıkça görülen tahmin kaymalarını önemli ölçüde azaltır ve daha güvenilir sonuçlar elde edilmesini sağlar (Kuş vd., 2021).

3.4. XGBoost (Extreme Gradient Boosting)

XGBoost, gradient boosting algoritmalarının geliştirilmiş bir versiyonudur ve yüksek performans gösteren modellerden biridir. Makine öğrenmesinde sınıflandırma ve regresyon

problemlerinde sıkça kullanılmaktadır. Hızlı ve yüksek doğruluk oranları sunması, onu veri bilimi yarışmaları ve büyük ölçekli projelerde popüler hale getirmiştir (Eşidir, 2025). XGBoost, son yıllarda makine öğrenmesi alanında en başarılı ve yaygın kullanılan yöntemlerden biri haline gelmiştir (Ma vd., 2018). Gereksiz e-postaların sınıflandırılması, reklam eşleştirme, dolandırıcılık tespiti ve fiziksel olaylardaki anomali tespiti gibi birçok alanda etkili sonuçlar sunmaktadır (Chen ve Guestrin, 2016). XGBoost, doğru tahminler üretmede etkili olduğunu ispatlamıştır ve birçok alternatif makine öğrenimi yaklaşımına göre hesaplama avantajları sunmaktadır (Gür, 2024b).

Paralel hesaplama yeteneği, L1 ve L2 regülarizasyon desteği ve eksik verilerle etkili çalışması, XGBoost'u büyük veri setleri için ideal bir seçim haline getirir. Özellikle tree pruning (ağaç budama) gibi optimizasyon özellikleri, modeli hem hızlı hem de verimli hale getirir. Finans, sağlık, pazarlama ve e-ticaret gibi birçok alanda yaygın olarak kullanılan bu algoritma, büyük veri setlerinde karmaşık ilişkileri analiz etmekte oldukça başarılıdır.

XGBoost, karar ağacı temelli bir makine öğrenmesi algoritması olup, sınıflama ve tahmin problemleri için kullanılan Gradyan Artırma (Gradient Boosting - GB) yöntemine benzer bir şekilde çalışır. Klasik yöntemlere kıyasla daha fazla eğitim süresi gerektirse de, yüksek doğruluk ve başarılı tahminler sağlamasıyla öne çıkar. Özellikle büyük veri setlerinde ve karmaşık problemlerde sağladığı etkili performans, XGBoost'u popüler bir tercih haline getirmektedir (Abar, 2020).

3.5. Gradient Boosting (Gradyan Artırma)

Gradient Boosting, sınıflandırma ve regresyon gibi gözetimli öğrenme görevleri için kullanılan güçlü bir makine öğrenimi tekniğidir. Bu yöntem, zayıf tahmin modellerini (genellikle karar ağaçları) ardışık bir şekilde geliştirerek modelin doğruluğunu artırır. Geliştirilen her yeni model, önceki modellerin hatalarını düzeltmeye odaklanmaktadır, böylece her adımda performansın artması sağlanmaktadır (Ou, 2020). Gradient Boosting, veri setindeki hedef değişkenin ortalama değeriyle başlayarak basit bir model kurar ve bu model regresyon için doğrudan ortalamayı, sınıflandırma için ise log (odds) değerini kullanır (Bazilevych vd., 2023). İlk modelin hataları (artıkları), sonraki modellerin iyileştirmeye çalışacağı hedef olarak belirlenir. Her adımda, önceki modelin hatalarını en iyi düzeltecek yeni bir zayıf öğrenici (genellikle basit karar ağaçları) eklenir. Her zayıf öğrenicinin katkısı, gradient descent benzeri bir optimizasyon tekniğiyle ayarlanır, böylece modelin hata fonksiyonu azaltılır. Ağaçların çıktıları belirli bir öğrenme hızıyla ağırlıklandırılarak toplanır ve bu süreç, model yeterince iyi performans gösterene veya belirlenen iterasyon sayısına ulaşana kadar devam eder. Bu yöntem, modelin hızla ve etkin bir şekilde öğrenmesini sağlar ve genellikle düşük bir öğrenme hızı tercih edilir (Natekin ve Knoll, 2013).

Her bir adımda $F_m(x)$ modeli için güncellenme formülü Denklem 3'te gösterildiği gibidir:

$$F_{m+1}(x) = F_m(x) + \gamma_m h_m(x) \quad (8)$$

Bu formül, mevcut modelin $F_m(x)$, bir sonraki adımda eklenen zayıf öğrenici $h_m(x)$ ve bu öğrenicinin model üzerindeki etkisini belirleyen ağırlık γ_m ile nasıl güncellendiğini ifade etmektedir. γ_m değeri, hata fonksiyonunu minimize edecek şekilde Denklem 4'te gösterildiği gibi seçilmektedir:

$$\gamma_m = \arg\ min_\gamma \sum_{i=1}^{n} L(y_i, F_m(x_i) + \gamma h_m(x_i)) \quad (9)$$

Burada, L kayıp fonksiyonudur ve y_i, x_i sırasıyla i'nci gerçek değer ve özelliklerdir.

4. Yazılım ve Donanım

Gerçekleştirilen analizlerde, Python programlama dilinin güncel bir sürümü olan Python 3.12.7 kullanılmış ve analizler Jupyter Notebook 7.2.2 ortamında gerçekleştirilmiştir. Python, bilimsel hesaplamalar ve veri analizi için geniş kütüphaneler (NumPy, Pandas, Scikit-learn, TensorFlow, Keras vb.) barındırması sebebi ile tercih edilmiştir. Jupyter Notebook ise; kodları, grafikleri ve açıklamaları bir arada sunabilen bir ortam sağlaması açısından, işlemlerin kolaylıkla yapılabilmesini ve süreçlerin anlaşılır olmasını sağlamaktadır.

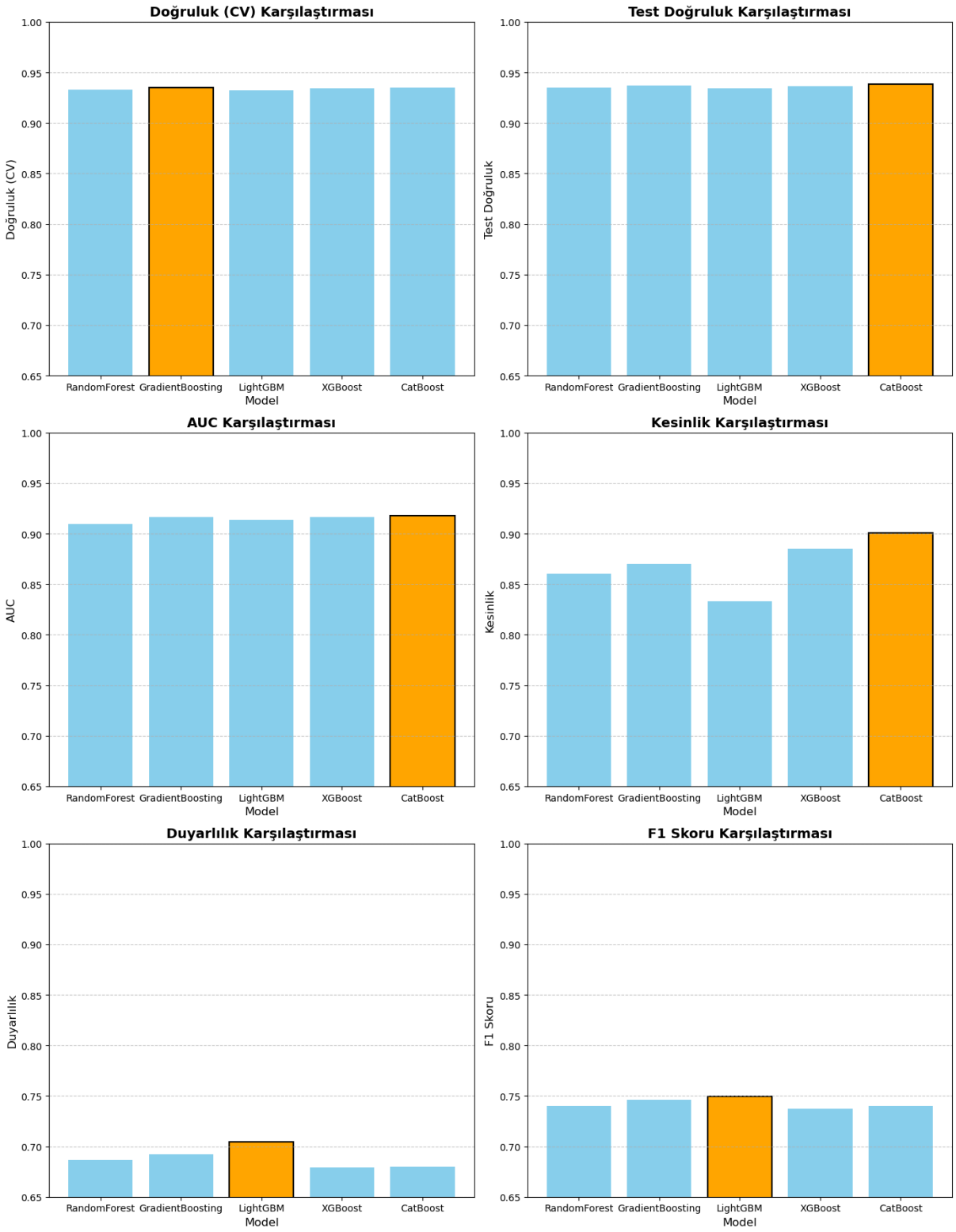
Çalışmada kullanılan bilgisayar, yüksek performans gerektiren veri analizi süreçlerini destekleyebilecek niteliklere sahiptir. Bilgisayarın işletim sistemi Windows 11 Pro tabanlı olup 10.0.22631-SP0 sürümünü çalıştırmaktadır. Cihaz, 64-bit mimariye sahip olup, Intel64 Family 6 Model 142 Stepping 10 türünde, güçlü ve modern bir işlemci barındırmaktadır. Cihaz, toplamda 16 GB RAM kapasitesine sahip olup, geniş veri kümelerini işlemek ve makine öğrenimi modellerini eğitmek için yeterli bellek alanına sahiptir. Kullanılan donanım özellikleri, makine öğrenimi modellerini eğitim ve test aşamalarını etkin bir şekilde desteklemiştir.

5. Bulgular

Tablo 2'de makine öğrenimi modellerinin performansları karşılaştırılmıştır. Gradient Boosting, XGBoost ve Cat Boost modelleri test doğruluğu açısından benzer performanslar sergilemektedir ve 0,94 değeri ile en iyi sonuçları vermektedir. RandomForest ve LightGBM bu değerin biraz gerisinde kalmasına rağmen, özellikle Gradient Boosting ve LightGBM modelleri F1 skoru açısından en iyi sonuçları sağlıyor. AUC değerlerine bakıldığında ise Cat Boost ve XGBoost modelleri, eğri altındaki alanı en geniş tutarak sınıflandırma performansında dikkat çekmektedir.

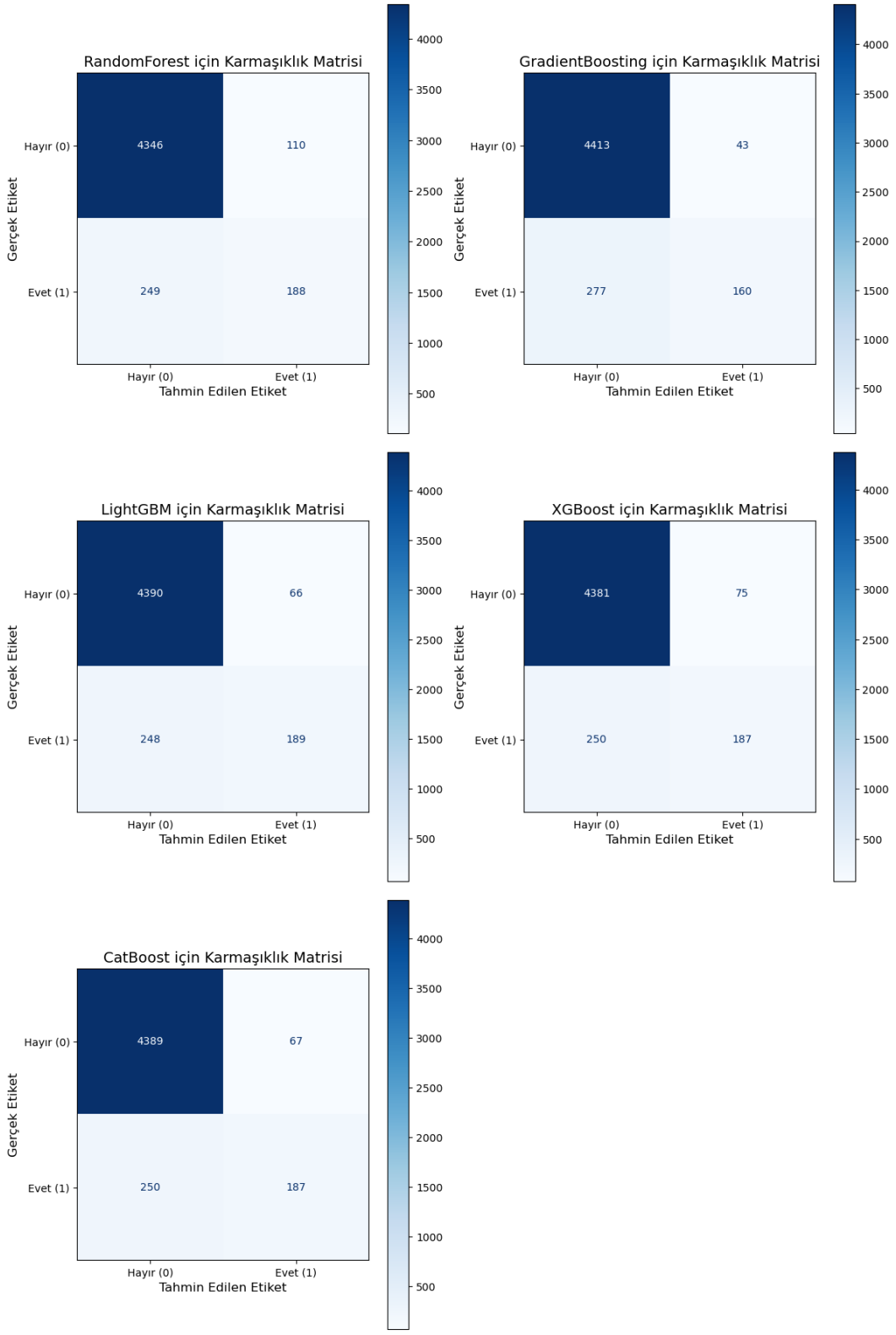
Tablo 2. Makine Öğrenimi Modellerinin Performans Ölçütlerinin Karşılaştırması

Modeller	Doğruluk (CV)	Test Doğruluk	AUC	Kesinlik	Duyarlılık	F1 Skoru
RandomForest	0,93	0,94	0,91	0,86	0,69	0,74
Gradient Boosting	0,94	0,94	0,92	0,87	0,69	0,75
LightGBM	0,93	0,93	0,91	0,83	0,70	0,75
XGBoost	0,93	0,94	0,92	0,89	0,68	0,74
CatBoost	0,93	0,94	0,92	0,90	0,68	0,74



Şekil 1. Makine Öğrenimi Modellerinin Performans Metrikleri Karşılaştırması

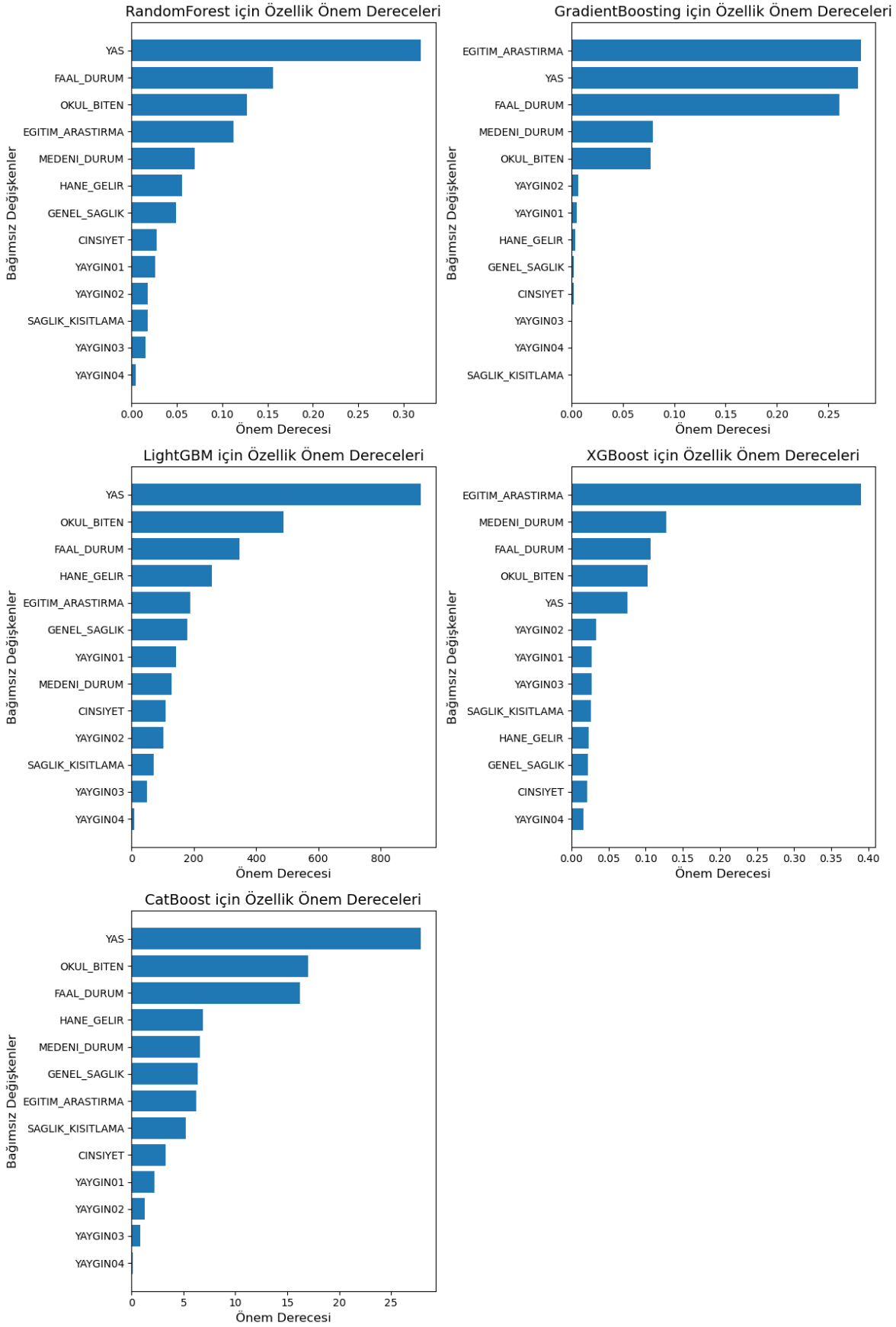
Şekil 1'de modellerinin performans metrikleri karşılaştırması grafik ortamında verilmiştir. Grafiklere göre, Cat Boost modeli öne çıkan performans değerlerine sahiptir. Test doğruluğu, kesinlik ve AUC değerlerinde CatBoost'un diğer modellerden üstün olduğu görülmektedir.



Şekil 2. Makine Öğrenimi Modelleri için Karmaşıklık Matrisleri (Evet-Hayır Tahminleri)

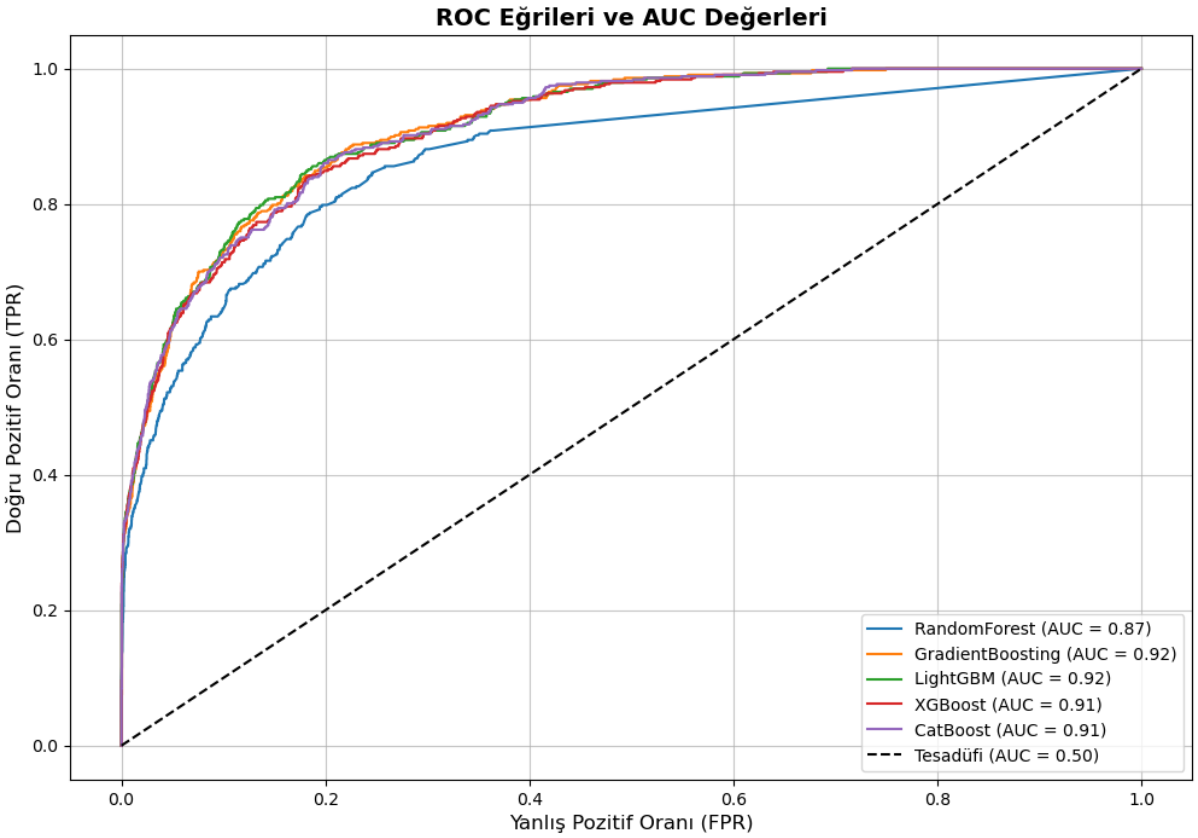
Şekil 2'de modellerin karmaşıklık matrisleri gösterilmiştir. Karmaşıklık matrisleri incelendiğinde, tüm modellerin "Hayır" sınıfını (True Negative) oldukça iyi tahmin ettiği görülmektedir, fakat Gradient Boosting bu sınıfta en düşük hatalı tahmin sayısına sahiptir. Gradient Boosting "Hayır" sınıfında en yüksek başarıya ulaşmıştır.

Öte yandan, "Evet" sınıfı (True Positive) açısından bakıldığında, Gradient Boosting modelinin genel performansı diğer modellere benzer olsa da, hatalı pozitif ve negatif tahminlerin dağılımı açısından diğer modellere kıyasla bir avantajı bulunmaktadır. Bu nedenle, pozitif sınıfın tespitinde daha dengeli sonuçlar elde etmek isteyen kullanıcılar için Cat Boost veya XGBoost da değerlendirilmesi gereken alternatifler olabilir.



Şekil 3. Makine Öğrenimi Modelleri için Önem Derecelerinin Karşılaştırılması

Şekil 3'te makine öğrenimi modellerinin önem derecelerinin karşılaştırılması grafik ortamında göterilmiştir. Gradient Boosting ve XGBoost modellerinde "EĞİTİM_ARAŞTIRMA" özniteliği öne çıkarken, RandomForest, LightGBM ve Cat Boost modelleri "YAŞ" değişkenini en önemli öznitelik olarak belirlemiştir. Ayrıca, "FAAL_DURUM" ve "OKUL_BITEN" öznitelikleri genel olarak tüm modellerde ilk sıralarda yer almaktadır. Ancak "CİNSİYET" ve "SAĞLIK_KISITLAMA" gibi değişkenlerin önem derecesi modeller arasında farklılık göstermektedir ve bazı modellerde oldukça düşük bir etkiye sahiptir.



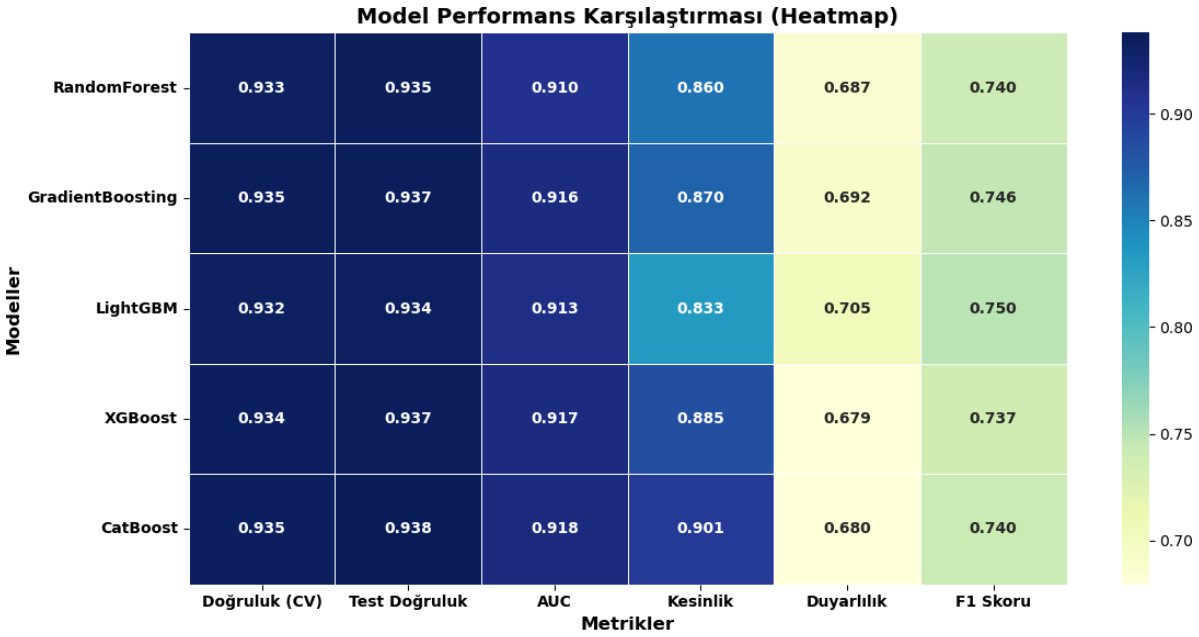
Şekil 4. Modellerin ROC Eğrileri ve AUC Değerleri

Şekil 4'te modellerin ROC eğrileri ve AUC değerleri gösterilmiştir. ROC eğrisi (Receiver Operating Characteristic Curve), modellerin sınıflandırma performansını değerlendirmede kullanılan bir metriktir ve eğrinin altında kalan alan (AUC) değeri, modelin genel başarı oranını özetlemektedir. Gradient Boosting ve LightGBM modellerinde AUC değerleri 0,92 ile en yüksek seviyededir. Bu modeller hem doğru pozitif oranını artırmada hem de yanlış pozitif oranını düşük tutmada diğer modellere göre üstün performans sergilemektedir.

XGBoost ve Cat Boost modelleri ise 0.91 AUC değeri ile Gradient Boosting ve LightGBM modellerini yakından takip etmektedir. Bu modeller, sınıflandırma performansı açısından benzer bir başarı göstermekte ve yanlış pozitif oranlarına karşı doğru pozitif oranlarını dengeli bir şekilde optimize etmektedir. Ancak RandomForest modeli, 0.87 AUC değeri ile diğerlerinden daha düşük bir performans sergileyerek yanlış pozitif oranını kontrol etme konusunda biraz daha zayıf kalmaktadır.

ROC eğrisi ayrıca, modellerin farklı eşik değerlerinde nasıl bir performans gösterdiğini de görselleştirmektedir. Gradient Boosting ve LightGBM'in eğrilerinin diğer modellere göre daha yukarıda ve sola yakın olması, bu modellerin doğru pozitif oranlarını daha hızlı artırırken

yanlış pozitif oranlarını düşük seviyede tutabildiğini göstermektedir. Bu sonuçlar, Gradient Boosting ve LightGBM'in bu problem için daha uygun modeller olduğunu işaret etmektedir.



Şekil 5. Isı Haritasına (Heatmap) Göre Modellerin Performansları

Şekil 5'te ısı haritasına (Heatmap) göre modellerin performansları gösterilmiştir. Heatmap grafiğine göre, Cat Boost ve Gradient Boosting modelleri hem doğruluk (CV ve test) hem de AUC açısından en yüksek değerlere sahiptir. Kesinlik metriklerinde XGBoost ve Cat Boost öne çıkarken, Duyarlılık açısından LightGBM en iyi performansı sergilemiştir. F1 skoru açısından LightGBM ve Gradient Boosting modelleri diğerlerinden daha başarılıdır.

6. Sonuç, Tartışma ve Öneriler

Çalışmada, TÜİK'in 2022 Yetişkin Eğitimi Araştırması mikro veri seti kullanılarak bireylerin örgün eğitime katılım durumları makine öğrenimi modelleri aracılığıyla analiz edilmiştir. Çalışmada kullanılan Gradient Boosting, XGBoost, Cat Boost, LightGBM ve Random Forest modellerinin performansları karşılaştırılmıştır. Analizler sonucunda, yaş, eğitim seviyesi ve medeni durum değişkenlerinin örgün eğitime katılım üzerinde en etkili faktörler olduğu tespit edilmiştir. Cat Boost modeli, yüksek doğruluk oranı, kesinlik ve AUC değerleri ile en uygun model olarak ön plana çıkmıştır.

Türkiye'de yetişkin eğitimi, bireylerin yaşam boyu öğrenme süreçlerinde önemli bir yer tutmaktadır. Yetişkin eğitimi, bireylerin mesleki ve kişisel gelişimleri açısından önemlidir. Araştırmada kullanılan modellerin performans değerlendirmesi, makine öğrenimi yöntemlerinin doğruluk ve öngörü gücünün geleneksel istatistiksel yöntemlere göre üstün olduğunu ortaya koymaktadır. Çalışmanın en önemli katkılarından biri, makine öğrenimi tekniklerini kullanarak yetişkin eğitimi alanında veri temelli karar alma süreçlerini desteklemektir. Eğitim politikalarını yönlendirmek için gelecekteki araştırmalar, daha geniş veri setleri ve farklı modelleme yaklaşımlarıyla bu tür analizleri daha kapsamlı hale getirebilir. Bu sonuçlar, literatürde sıklıkla vurgulanan makine öğrenimi tekniklerinin veri yoğun eğitim alanlarında kullanımına yönelik argümanları desteklemektedir (Chen ve Guestrin, 2016). Özellikle Türkiye bağlamında, eğitim politikalarını şekillendirme potansiyeli taşıyan veriye dayalı analizlerin sınırlı olduğu göz önüne alındığında, bu çalışmanın önemi daha da belirgin hale gelmektedir. Eğitim politikalarının etkinliğini artırmak için, dezavantajlı grupların eğitime katılımını teşvik edecek politikaların geliştirilmesi gerekmektedir (Güleç ve Çelik, 2013). Bu

bağlamda, eğitim politikalarının yalnızca akademik planlamaya değil, ekonomik ve sosyal faktörlere de odaklanması gerektiği söylenebilir. Çalışmanın bir diğer önemli sonucu, makine öğrenimi modellerinin eğitim verileri üzerindeki analiz kapasitesinin geleneksel istatistiksel yöntemlere kıyasla daha güçlü olmasıdır. Regresyon modelleri genellikle doğrusal ilişkileri analiz edebilmekte, ancak çok boyutlu ve karmaşık veri yapıları içerisinde yeterli performansı sergileyememektedir (Bae vd., 2021). Bu nedenden ötürü, makine öğrenimi modellerinin eğitim verileri ile daha yaygın bir şekilde kullanılması, eğitim politikalarının geliştirilmesine yönelik yeni bir perspektif sunacaktır. Öte yandan, çalışmada kullanılan veri seti Türkiye genelini kapsamakla birlikte, bölgesel bazda farklılıklar göz ardı edilmiştir. Türkiye'de farklı bölgelerde yaşayan bireylerin eğitime katılım oranları arasında belirgin farklılıklar bulunabilmektedir (Polatoğlu, 2022). Gelecekte yapılacak benzer çalışmalarda, bölgesel farklılıkları analiz eden makine öğrenimi modellerinin kullanılması, eğitim politikalarının daha etkin hale getirilmesini sağlayacaktır. Çalışma sonuçları, düşük gelirli ve düşük eğitim seviyesine sahip bireylerin eğitime katılım oranlarının daha düşük olduğunu göstermiştir. Bu nedenle; burs, eğitim kredileri ve teşvik programları gibi ekonomik destek mekanizmalarının artırılması önerilmektedir. Özellikle çalışan bireyler için çevrim içi eğitim programlarının genişletilmesi, eğitimde fırsat eşitliğine katkıda bulunacaktır. Dijital eğitim platformları, bireylerin zamandan ve mekândan bağımsız olarak eğitime erişimlerini basitleştirecektir.

Gelecekte yapılacak benzer çalışmalarda, farklı coğrafi bölgeler ve daha çeşitli demografik gruplar için veri toplanarak uluslararası karşılaştırmalı analizlerin yapılabilir. Özellikle Avrupa Birliğinin hayat boyu öğrenme politikaları dikkate alınarak Türkiye'nin stratejileri yeniden dizayn edilebilir. Dijital öğrenme platformlarının yaygınlaştırılması, ekonomik teşvikler ve dezavantajlı gruplara yönelik özel bir takım programların oluşturulması gibi uygulamalar, makine öğrenimi modelleriyle belirlenen hedef grupların eğitime katılımını artırmada önemli rol oynayacaktır. Eğitim politikalarının, makine öğrenimi modelleriyle desteklenmiş bu tür veriye dayalı yaklaşımlarla şekillendirilmesi, hem ulusal hem de uluslararası düzeyde daha etkili ve sürdürülebilir sonuçlar elde edilmesini sağlayacaktır.

Çıkar Çatışması Beyanı/ *Conflict of Interest Declaration:*
Bu çalışmada herhangi bir potansiyel çıkar çatışması bulunmamaktadır.
There are no potential conflicts of interest in this study

Etik/ *Ethic*:
Çalışmada bilimsel, etik ilkelere uyulmuş ve yararlanılan tüm çalışmalar kaynakçada belirtilmiştir.
In the study, scientific and ethical principles were followed and all the studies used were indicated in the bibliography.

Kaynakça

Abar, H. (2020). XGBoost ve Mars yöntemleriyle altın fiyatlarının kestirimi. *EKEV Akademi Dergisi*, 83, 427-446.

Akbulut, S., ve Adem, K. (2023). Derin öğrenme ve makine öğrenmesi yöntemleri kullanılarak gelişmekte olan ülkelerin finansal enstrümanlarının etkileşimi ile BIST 100 tahmini. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 12(1), 52-63. https://doi.org/10.28948/ngumuh.1131191

Bae, C. Y., Im, Y., Lee, J., Park, C., Kim, M., Kwon, H. U., and Kim, J. (2021). Comparison of biological age prediction models using clinical biomarkers commonly measured in clinical practice settings: AI techniques vs. traditional statistical methods. *Frontiers in Analytical Science*, 1. https://doi.org/10.3389/frans.2021.709589

Bazilevych, K., Kyrylenko, O., Parfenyuk, Y., Krivtsov, S., Meniailov, I., Kuznietcova, V., and Chumachenko, D. (2023). Comparative analysis of the machine learning models determining COVID-19 patient risk levels. *Radioelectronic and Computer Systems*, (3), 5-17. https://doi.org/10.32620/reks.2023.3.01

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939785

El Naqa, I., and Murphy, M. J. (2015). What is machine learning? In *Machine Learning in Radiation Oncology* (pp. 3-11). Springer. https://doi.org/10.1007/978-3-319-18305-3_1

Eşidir, K. A. (2025). Türkiye'nin Kimyasal Madde İthalatının Gelecek Tahmini: Makine Öğrenmesi ve Topluluk Öğrenme Yöntemleri Performans Analizi. *Firat University Journal of Social Sciences,* 35(1), 261-278. https://doi.org/10.18069/firatsbed.1580620

Eşidir, K. A., ve Gür, Y. E. (2024). Forecasting Türkiye's Paper and Paper Products Sector Import Using Artificial Neural Networks. *Hitit Sosyal Bilimler Dergisi,* 17(2), 206-224. https://doi.org/10.17218/hititsbd.1327799

Güleç, İ., Çelik, S., ve Demirhan, B. (2013). Yaşam boyu öğrenme nedir? Kavram ve kapsamı üzerine bir değerlendirme. *Sakarya University Journal of Education*, 2(3), 34-48. https://doi.org/10.19126/suje.27105

Gür, Y. E. (2024a). Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach. *Data Science in Finance and Economics*, 4(4), 469-513. https://doi.org/10.3934/DSFE.2024020

Gür, Y. E. (2024b). Forecasting the euro exchange rate using deep learning algorithms and machine learning algorithms. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, 23(49), 1435-1456. https://doi.org/10.46928/iticusbe.1379268

Gür, Y. E. (2024c). Stock price forecasting using machine learning and deep learning algorithms: A case study for the aviation industry. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 36(1), 25-34. https://doi.org/10.35234/fumbd.1357613

Ji, H. (2023). Robustness analysis on stock market prediction method. *Highlights in Business, Economics and Management*, 21, 791-801. https://doi.org/10.54097/hbem.v21i.14763

Kaya, H. (2014). Küreselleşme sürecinde yaşam boyu öğrenme ve yetişkin eğitimi gerçeği. *Akademik İncelemeler Dergisi*, 9(2), 91-111. https://doi.org/10.17550/aid.94961

Komşu, U. C. (2017). Yetişkin eğitiminde bir öğretim aracı olarak şiirin rolü: Literatür taraması örneği. *HAYEF Journal of Education*, 14(1), 1-20.

Kuş, İ., Bozkurt Keser, S., ve Yolaçan, E. N. (2021). Saldırı tespit sistemlerinde topluluk öğrenme yöntemlerinin kıyaslanması. *Avrupa Bilim ve Teknoloji Dergisi*, (31), 725-734.

Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., and Niu, X. (2018). Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGBoost algorithms according to different high dimensional data cleaning. *Electronic Commerce Research and Applications*, 31, 24-39.

Natekin, A., and Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7. https://doi.org/10.3389/fnbot.2013.00021

Oguine, O. C., and Oguine, M. B. (2021). Comparative analysis and forecasting on the death rate of COVID-19 patients in Nigeria using random forest and multinomial Bayesian epidemiological models. *Journal of Clinical Case Studies, Reviews and Reports*, 1-7. https://doi.org/10.47363/jccsr/2021(3)182

Ou, R. (2020). Out-of-core GPU gradient boosting. https://doi.org/10.48550/arxiv.2005.09148

Pakarinen, O., Karsikas, M., Reito, A., Lainiala, O., Neuvonen, P., and Eskelinen, A. (2022). Prediction model for an early revision for dislocation after primary total hip arthroplasty. *PLOS ONE*, 17(9), e0274384. https://doi.org/10.1371/journal.pone.0274384

Polatoğlu, Y. (2022). Ulusal politika belgelerinde yetişkin eğitimi. *Studies in Educational Research and Development*, 6(1), 52-72.

Sinap, V. (2024). Eğitimde makine öğrenmesi: Bir bilim haritalama çalışması. *Başkent University Journal of Education*, 11(1), 10-25.

Speer, A. B. (2021). Empirical attrition modelling and discrimination: Balancing validity and group differences. *Human Resource Management Journal*, 34(1), 1-19. https://doi.org/10.1111/1748-8583.12355

Suthaharan, S. (2014). Big data classification: Problems and challenges in network intrusion prediction with machine learning. *ACM SIGMETRICS Performance Evaluation Review*, 41(4), 70-73.

Tong, S., Sun, W., Xu, J., and Li, H. (2024). Robustness analysis and prediction of topological edge states in topological elastic waveguides. *Physica Scripta*, 99(7), 075402. https://doi.org/10.1088/1402-4896/ad504f

Tosunoğlu, E., Yılmaz, R., Özeren, E., ve Sağlam, Z. (2021). Eğitimde makine öğrenmesi: Araştırmalardaki güncel eğilimler üzerine inceleme. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, 3(2), 178-199.

Wu, Y. (2023). Job embeddedness review: Presentation, measurement and development. *Advances in Economics, Management and Political Sciences*, 47(1), 169-174. https://doi.org/10.54254/2754-1169/47/20230393

Yetişkin Eğitimi Araştırması Mikro Veri Seti. (2022). Yayın No: 4725, Yayım Tarihi: Ocak 2024, Türkiye İstatistik Kurumu Bilgi Dağıtım Grup Başkanlığı. ISBN: 978-625-8368-63-5.

Zeiler, M. D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. *Computer Vision – ECCV 2014*, 818-833. https://doi.org/10.1007/978-3-319-10590-1_53

Zhu, X., Sawhney, R., and Upreti, G. (2016). Determinates of employee voluntary turnover and forecasting in departments: A case study. *Studies in Engineering and Technology*, 3(1), 64-73. https://doi.org/10.11114/set.v3i1.1635

Zilyas, D., ve Yılmaz, A. (2023). Makine öğrenmesi yöntemleri ile eğitim başarısının tahmini modeli. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 14(3), 437-447. https://doi.org/10.24012/dumf.1322273