

## Research Article (Special Issue) | Araştırma Makalesi (Özel Sayı)

# Çalışan yıpranmasını tahmin etmede analitik bir yaklaşım: Topluluk öğrenme yöntemi

Mustafa Yurtsever | Dr. Öğr. Üyesi, İzmir Demokrasi Üniversitesi, [mustafa.yurtsever@idu.edu.tr](mailto:mustafa.yurtsever@idu.edu.tr), [0000-0003-2232-0542](https://orcid.org/0000-0003-2232-0542)

Corresponding author/Sorumlu yazar: Mustafa Yurtsever ✉ [mustafa.yurtsever@idu.edu.tr](mailto:mustafa.yurtsever@idu.edu.tr)

## Öz

Şirketler, profesyonel çalışanlarının ayrılmasını engelleyerek işe alım ve eğitim maliyetlerini azaltmak için çeşitli önlemler ararlar. Belirli bir çalışanın ayrılıp ayrılmayacağını önceden tahmin etmek, şirketin bu tür kayıpları minimize etmek için gereken adımları atmasını sağlar. Bu nedenle, çalışanların ayrılma olasılığını önceden tahmin etmek, işverenlere stratejik kararlar almalarında yardımcı olabilir. Çalışan yıpranması bu noktada çalışanların işten ayrılma niyetlerini anlama konusunda belirleyici olabilmektedir. Topluluk öğrenme modelleri, birden fazla algoritmanın çeşitli görüşlerini birleştirerek daha doğru ve güvenilir sonuçlar elde etme potansiyeline sahiptir. Bu çalışmada, 1.470 kayıttan oluşan IBM şirketi Watson Analytics tarafından hazırlanan çalışanların yıpranma durumunu gösteren veri seti kullanılmıştır. Bu amaçla, Rastgele Orman, Destek Vektör Makineleri, Çok Katmanlı Algılayıcı modeli ve iki farklı topluluk öğrenme modeli kullanılarak performansları değerlendirilmiştir. Sınıf dengesizliğini aşmak için adaptif sentetik veri üretimi (ADASYN) yaklaşımı kullanılmıştır. Sonuç olarak, ADASYN kullanılarak oluşturulan dengeli bir veri kümesi ile eğitilen yığın topluluk öğrenme modeli ile 0.96 doğruluk oranı elde etmiştir.

**Anahtar Kelimeler:** Çalışan Yıpranması, Tahminleme, Topluluk Öğrenmesi Yöntemleri **JEL Kodları:** J21, M51

## An analytical approach to predicting employee attrition: Ensemble learning method

### Abstract

Companies seek various measures to prevent the departure of professional employees and thereby reduce recruitment and training costs. Predicting whether a specific employee will leave or not enables the company to take the necessary steps to minimize such losses. Therefore, predicting the likelihood of employee attrition in advance can assist employers in making strategic decisions. Employee attrition plays a crucial role in understanding employees' intentions to leave their jobs. Ensemble learning models have the potential to achieve more accurate and reliable results by combining various perspectives of multiple algorithms. In this study, a dataset indicating the attrition status of employees, prepared by IBM Watson Analytics and consisting of 1,470 records, was utilized. For this purpose, the performances were evaluated using Random Forest, Support Vector Machines, Multi-Layer Perceptron model, and two different ensemble learning models. To overcome class imbalance, the adaptive synthetic data generation (ADASYN) approach was used. As a result, a stacked ensemble learning model trained on a balanced dataset created using ADASYN achieved an accuracy rate of 0.96.

**Keywords:** Employee Attrition, Prediction, Ensemble Learning Methods **JEL Codes:** J21, M51

### Extended Summary

Employee attrition can be defined as a gradual decrease in the number of personnel due to increased resignations (typically of competent, high-performing, talented, and team-compatible employees) and the failure to replace them with new hires. To address high turnover rates, organizations should take measures to understand the underlying reasons and implement strategies to mitigate them. These measures may include conducting exit interviews to gather feedback from departing employees, identifying areas for improvement in the workplace environment, offering competitive salaries and benefits packages, providing opportunities for development, and fostering a supportive organizational culture. By addressing the root causes of attrition and prioritizing employee engagement and satisfaction, organizations can work towards reducing turnover rates and promoting long-term stability and success.

### How to cite this article / Bu makaleye atıf vermek için:

Yurtsever, M. (2024). Çalışan yıpranmasını tahmin etmede analitik bir yaklaşım: Topluluk öğrenme yöntemi. *KOCATEPEİİBFD*, 26(Özel Sayı), 150-160.

<https://doi.org/10.33707/akuiibfd.1462567>

In recent years, with the advancement of artificial intelligence (AI) technology, it has been observed that businesses have started to implement this technology in various fields. AI is being used in areas such as production management, marketing, procurement, distribution, and human resources management to increase efficiency or improve financial performance.

Traditionally, organizations would react and take steps to mitigate losses after employee attrition incidents occurred. However, with the use of technologies like artificial intelligence (AI) today, it has become possible to predict the likelihood of attrition in advance and take preventive measures. This proactive approach presents a significant opportunity for organizations to enhance their workforce planning and management processes.

In this study, two types of ensemble models were created using Random Forest, Multilayer Perceptron (MLP), and Support Vector Machines (SVM), as well as individual models, to predict employee attrition. The aim is to identify the best classifier for the analyzed problem. To address the limitations of traditional techniques used in predicting employee attrition, a series of ensemble learning methods were employed.

Ensemble learning involves training multiple models simultaneously and combining their outputs to generate predictions. This technique leverages the different features and strengths of each model to yield more accurate and reliable predictions. The approach outlines the classification models used in the initial step and their corresponding training processes. Ensemble learning entails the simultaneous use of multiple classification models. In this approach, various classification models are trained, and their outputs are later combined through a higher-level meta-model to produce the final prediction value. The final prediction is generated by the meta-model through the combination of predictions from the trained models. Particularly, the aggregation process performed using logistic regression at the meta-stage aims to combine the predictions of the base models to achieve higher accuracy. Another prominent ensemble method utilized for making robust predictions is the Voting Classifier. The Voting Classifier is an ensemble learning method that combines predictions from multiple base predictors (machine learning models) and predicts the class label by taking a vote. It can be applied to both classification and regression problems. In hard voting, each base predictor (model) predicts the class label, and the final prediction is determined by a majority vote. The class label with the most votes becomes the final prediction. In soft voting, each base predictor estimates the probability distribution across all classes. The final prediction is determined by taking the average of these probabilities and then selecting the class with the highest average probability.

After analyzing the predictive performance of the models, the ensemble learning model constructed with employee attrition, Random Forest (RO), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) achieved the best performance across all metrics such as accuracy (0.96), F1 score (0.96), precision (0.96), and recall (0.96). The Voting Classifier ensemble model also demonstrated successful performance with an accuracy rate of 0.95. The model with the lowest performance was the Support Vector Machine (SVM) model, with an accuracy rate of 0.91. It is observed that ensemble learning models outperform individual machine learning models.

## Giriş

İnsan kaynakları, iş faaliyetlerini yürütmek ve iş hedeflerine ulaşmak için işletmeler için son derece önemlidir. Hızla değişen iş ortamı ve bir organizasyon içinde ortaya çıkan çeşitli insan sorunları nedeniyle, organizasyonlar insan kaynakları yönetimiyle ilgili çok sayıda sorunla karşı karşıyadır (Chung vd., 2023). Bu sorunlardan bir tanesi de işten ayrılmalardır (Douaidi ve Kheddouci, 2022). Ayrılma, bir çalışanın bir kuruluştan gönüllü veya gönülsüz istifa etmesi olarak tanımlanabilir (Srivastava ve Nair, 2018).

Çalışan memnuniyetinin düşük olması bir çalışanın şirketi terk etmesi için başlıca nedenlerden biridir. Çalışma saatlerinin belirsizliği, ücret politikası, çevresel koşullar, iş yükü gibi nedenler çalışan memnuniyetini etkileyen faktörlerdir. Oshagbemi (2003), genel iş memnuniyetinin işin çeşitli yönlerine bağlı olduğunu ve çoğu zaman işin doğasının birincil neden olabileceğini kanıtıyor. Oshagbemi ayrıca, iş tatminini etkileyebilecek iş arkadaşlarının kişi üzerindeki etkisi veya işteki denetim, teklif edilen ücret, çalışma ortamı, organizasyonda izlenen politikalar, prosedürler ve organizasyondaki büyüme fırsatları gibi birçok nedenin iş tatminini etkileyebileceğini ve bunun da bir çalışanın organizasyondan ayrılma kararını etkileyebileceğini açıklıyor.

Bir diğer çalışma işten ayrılmasının başlıca nedenleri olarak yöneticiyle ilişki, daha iyi kariyer beklentileri, tazminat, ödüller, takdir ve iş içeriği gibi insan kaynakları değişkenleriyle ilgili olduğunu ortaya koymaktadır (Barpanda ve Athira, 2022). Sriram vd. (2019) araştırması, örgüt kültürünün ve çalışanlara izin verilmesinin çalışan yıpranmasını etkileyen iki önemli faktör olduğunu ortaya koydu. Tüm bu faktörler çalışanların iş tatmin seviyesini etkilemektedir. İş tatmini düşük olan birinin gelecekte işini bırakma olasılığı, iş tatmini yüksek olan aynı kişiye göre çok daha yüksek olduğu bilinmektedir (Clark, 2001).

Yüksek işten ayrılma oranı, herhangi bir organizasyon için hem finansal hem de insan kaynakları açısından zararlı olabilir (Jain vd., 2020). Yalnızca yeni çalışanları işe almak ve onları eğitmek için değil, aynı zamanda mevcut çalışanların moralini düşürmesi ve geriye kalan çalışanların iş yükünü artırması açısından işletmeleri etkiler (Oruç, 2020). Ayrıca ayrılan çalışanların yerine geçecek

doğru kişiyi bulmakta işletme açısından zorlayıcıdır (Yadav vd., 2018). Bir diğer nokta ise temel teknolojilerin ve teknik bilgilerin sızma riski gibi çeşitli nedenlerden dolayı kuruluşta önemli kayıplara neden olmasıdır (Chung vd., 2023). Müşterilerin hizmet markasına ilişkin imajı üzerinde olumsuz bir etkiye sahip olduğu da bilinmektedir (Subramony ve Holtom, 2012). Projelerden kazanılan değerli bilgi ve deneyimlerin kaybına neden olması da diğer bir etkisidir (Kumar ve Yakhlef, 2016). Özellikle ileri teknoloji firmaları işten ayrılma oranlarını azaltmak ile yakından ilgilenmektedir (Hoffman ve Tadelis, 2021). Bu nedenle, çalışanların işten ayrılmasına katkıda bulunan faktörleri anlamak ve önleyici önlemler almak gerçekten önemlidir.

Çalışan yıpranması, işten ayrılmaların (genellikle yetkin, performansı yüksek, yetenekli ve takım uyumuna sahip çalışanların) artması ve yerlerine yeni çalışanların getirilememesi durumunda personel sayısında kademeli bir azalma olarak tanımlanabilir (Oruç, 2020). Yüksek işten ayrılma oranlarını ele almak için, kuruluşlar altta yatan nedenleri anlamak ve bunları azaltacak stratejiler uygulamak için önlemler almalıdır (El-Rayes vd., 2020). Bu önlemler; ayrılan çalışanlardan geri bildirim almak için çıkış görüşmeleri yapmayı, işyeri ortamında iyileştirme alanlarını belirlemeyi, rekabetçi ücret vermeyi, yan hak paketleri sunmayı, gelişim için fırsatlar sağlamayı ve destekleyici bir kurum kültürünü teşvik etmeyi içerebilir. Kuruluşlar, işten ayrılmanın temel nedenlerini ele alarak çalışan bağlılığı ve memnuniyetine öncelik vererek işten ayrılma oranını azaltmaya ve uzun vadeli istikrar ve başarıyı teşvik etmeye çalışabilir.

Geleneksel olarak, işten ayrılma durumları meydana geldikten sonra kurumlar tepki verir ve kayıpları azaltmaya yönelik adımlar atardı. Ancak, günümüzde YZ gibi teknolojilerin kullanımıyla, işten ayrılma olasılığını önceden tahmin etmek ve bu duruma önlem almak mümkün hale gelmiştir. Bu önleyici yaklaşım, kurumların işgücü planlaması ve yönetimi süreçlerini geliştirmek için önemli bir fırsat sunar.

Son yıllarda yapay zeka (YZ) teknolojisinin gelişmesiyle birlikte, işletmelerin çeşitli alanlarda bu teknolojiyi uygulamaya başladığı gözlemlenmektedir. Üretim yönetimi, pazarlama, satın alma, dağıtım ve insan kaynakları yönetimi gibi alanlarda verimliliği artırmak veya finansal performansı iyileştirmek amacıyla YZ kullanılmaktadır (Buntak vd., 2021; Enholm vd., 2022) İnsan kaynakları yönetimi alanında ise, YZ özellikle işe alım, performans yönetimi (Garg vd., 2022) ve insan kaynakları otomasyonu gibi alanlarda uygulanmaya başlamıştır ve bu kullanım alanı giderek genişlemektedir (Aldulaimi vd., 2021). Özellikle makine öğrenimi gibi veri analizi tekniklerinin çalışan devirlerini önleme yöntemi olarak kullanılması dikkat çekmektedir (Avrahami vd., 2022).

Bu doğrultuda, literatürde çalışan yıpranmasını önceden tahmin edilmesini analiz eden çeşitli çalışmalar bulunmaktadır. Makine öğrenmesi yöntemleri kullanılarak yapılan çalışmalarda Gaussian Naive Bayes (Fallucchi vd., 2020), Ragele Orman (Kaya ve Korkmaz, 2021), K-en Yakın Komşu (Alduayj ve Rajpoot, 2018) gibi makine öğrenmesi yöntemleri kullanılmıştır. Bu çalışmaların tek bir modele bağlı olduğu ve yöntemlerin sonuçlarının karşılaştırılmasına dayandığı görülmektedir.

Bu çalışmanın temel amacı, çalışan yıpranmasını ve bunun sonucunda çalışanların işten ayrılma olasılıklarını daha doğru bir şekilde tahmin etmek için çeşitli makine öğrenimi modellerini karşılaştırmak ve topluluk öğrenme modellerinin bu süreçteki etkisini incelemektir. Bu çalışma, çalışan yıpranmasını tahmin etme alanında birkaç yeniliği barındırmaktadır. İlk olarak, IBM Watson Analytics tarafından sağlanan 1.470 kayıttan oluşan veri seti üzerinde farklı makine öğrenimi algoritmaları test edilmiştir. İkinci olarak, sınıf dengesizliği gibi yaygın bir sorunla başa çıkmak için ADASYN (Adaptif Sentetik Veri Üretimi) yöntemi kullanılmıştır. Bu yöntem, sınıf dengesizliğini azaltarak daha dengeli bir veri kümesi oluşturmakta ve modellerin performansını artırmaktadır. Son olarak, topluluk öğrenme yöntemlerinin uygulanmasıyla elde edilen sonuçlar, bu tür yaklaşımların çalışan yıpranması tahmininde ne kadar etkili olabileceğini göstermektedir.

Çalışmanın katkıları arasında, çalışan yıpranmasını tahmin etmeye yönelik olarak kullanılan farklı makine öğrenimi algoritmalarının performanslarının karşılaştırılması ve topluluk öğrenme yöntemlerinin bu alandaki etkisinin değerlendirilmesi bulunmaktadır. Ayrıca, ADASYN gibi veri işleme tekniklerinin model performansı üzerindeki etkisi de incelenmiştir. Elde edilen bulgular, işverenlere, çalışanların işten ayrılma olasılıklarını daha iyi tahmin edebilecek ve bu kayıpları minimize edebilecek stratejik kararlar almaları için yol gösterici niteliktedir.

Bu çalışmada cevap aranan temel araştırma soruları şunlardır:

- Farklı makine öğrenimi modelleri, çalışan yıpranmasını tahmin etmede ne kadar başarılıdır?
- Topluluk öğrenme yöntemleri, çalışan yıpranması tahmininde doğruluğu artırmak için nasıl kullanılabilir?
- Sınıf dengesizliği gibi sorunlarla başa çıkmak için veri işleme teknikleri etkili sonuçlar verir mi?

Çoğu şirketin personel verisi bulunmakla birlikte, çalışan yıpranmasının çeşitli faktörlere bağlı olması daha sofistike yöntemlerin kullanılmasını gerektirir. Çalışmada kullanılan veri setinin dengesiz olması tahmin yapmayı daha da zorlaştırmaktadır. Bu çalışma, çalışan yıpranması tahminlerinin doğruluğunu artırmak için topluluk öğrenme modellerinin performanslarının karşılaştırılması amaçlanmaktadır. Bu makalenin 1. bölümü, çalışan yıpranması ile ilgili YZ alanındaki literatür incelemesine ele almıştır. 2. bölüm

araştırma yöntemini sunmakta ve 3. bölüm analiz sonuçlarını sunmaktadır. Son olarak, bu çalışmanın sonuçlarını, etkilerini, sınırlılıklarını ve ileriye dönük araştırma konularını sunmuştur.

## 1. Literatür Taraması

Hızla değişen iş ortamı ve bir şirkette ortaya çıkan çeşitli insani sorunlar nedeniyle kuruluşlar, insan kaynaklarıyla ilgili birçok sorunla karşı karşıya kalmaktadırlar (Rutherford vd., 2003). Çalışanların işten ayrılması veya işten ayrılma oranı, bir işyerindeki çalışanların belirli bir dönem içinde işlerini bırakma veya ayrılma durumunu ifade eder. Bu durum, çalışanların kendi istekleriyle ayrılması, emekli olması, işten çıkarılması veya başka nedenlerle işi terk etmesi gibi çeşitli sebeplerle gerçekleşebilir.

İşverenler için çalışanların işten ayrılması, iş gücü planlaması, maliyetler, verimlilik gibi birçok faktörü etkileyebilir (Yahia vd., 2021; Alsheref vd., 2022). Bu nedenle, işletmeler sıklıkla çalışanların işten ayrılma oranını takip eder ve bu durumu tahmin etmek ister. Ayrıca çalışan yıpranmasının iş birimlerinin performansı üzerinde önemli olumsuz etkileri olduğunu göstermektedir (Subramony ve Holtom, 2012). Yeni bir çalışanın yetiştirilmesi uzun ve maliyetli bir süreçtir ve çalışanların yıpranma oranını kontrol etmek ve azaltmak şirketin çıkarıdır (Fallucchi vd., 2020).

Bu nedenle bir kuruluşun ticari faaliyetlerini sürdürebilmesi ve yüksek rekabet avantajı elde edebilmesi için çalışan kaybının en aza indirilmesi gerekmektedir (Chung vd., 2023) Çalışanların neden ayrıldığına dair bir anlayış edinmek, İnsan Kaynakları departmanlarının gelişmiş planlama ve müdahale yoluyla çalışanlarını elde tutmalarına destek olur (Frye vd., 2018).

Literatürde araştırmacılar bu konuya farklı yöntemler ile yaklaşmışlardır. Alao ve Adeyemo (2013) beş farklı karar ağacı modeli ve iki kural kümesi ürettikleri çalışmalarında yeni çalışan kaybı durumlarını tahmin etmek için bir tahmin modeli geliştirmişlerdir. Khare vd. (2011), tarafından yapılan çalışmada lojistik regresyon yöntemi, ayrılmış ve mevcut çalışanların demografik verilerine dayalı olarak çalışan kaybını tahmin etmek için bir risk denklemi geliştirmek üzere önerilmiştir.

Alduayj ve Rajpoot (2018) Rastgele Orman (RO), k-en yakın komşu ve Destek Vektörü Makineleri (DVM) yöntemlerine kullandıkları çalışmada IBM İK veri setinin kaynak olarak kullanmışlardır. Veri setinin dengesizliğinden dolayı veri kümesini orijinal, sentetik olarak aşırı örneklenmiş ve az örneklenmiş olarak 3 farklı şekilde kullanmışlardır. Sentetik veri setine sahip sistemleri yüksek doğruluk gösterse de orijinal veri setiyle çıkan sonuçların doğruluğu yeterli olmamıştır. Bhatta vd. (2022), farklı makine öğrenmesi yöntemleri ve oylama sınıflandırıcı kullandıkları çalışmada, oylama sınıflandırıcısı yöntemi ile çalışan yıpranmasını 0,89'lük bir AUC (ROC Eğrisi Altındaki Alan) puanına ulaşmışlardır. Bir diğer araştırmada üç makine öğrenme algoritması, Karar Ağacı, Rastgele Orman ve İkili Lojistik Regresyon ile çalışan kaybını tahmin etmek için seçilmiştir. Lojistik Regresyon modeli, test edilen algoritmalar arasında %87,44'lük en yüksek doğruluğu sağlamıştır (Alsubaie ve Aldoukhi, 2024).

Fallucchi vd. (2020), IBM İK verilerini kullanarak Gaussian Naive Bayes, K-En Yakın Komşu, Karar Ağacı, RO, DVM ile modellerini oluşturdu. Tahmin doğruluğu açısından en iyi performansa sahip model doğrusal DVM oldu. Al-Darraji vd. (2021) yılında aynı veri setini derin sinir ağları kullanarak modellemişlerdir. Veri setinin dengesizliğini gidermek için sentetik veri ile genişletmişlerdir. Orijinal veri seti ile tahmin doğruluğu yaklaşık %91 iken sentetik veri seti kullanıldığında yaklaşık %94 doğruluk sonucuna ulaşılar.

Kaya ve Korkmaz (2021), aynı veri seti için dengesizlik problemini gidermek için bootstrap tekniğini kullanmışlardır ve sınıfların örnek sayılarını dengelemişlerdir. Yedi farklı makine öğrenmesi yöntemi kullandıkları çalışmada doğruluk oranı olarak %80 seviyelerine ulaşmışlardır. Celik (2019), ise karar ağacı ve destek vektör makinesi yöntemlerini kullandıkları çalışmalarında DVM ile %91 doğruluk oranı elde etmişlerdir.

## 2. Metodoloji

### 2.1. Rasgele Orman

RO algoritması, makine öğrenimi alanında sınıflandırma ve regresyon problemleri için kullanılan bir topluluk öğrenme algoritmasıdır. İlk olarak Breiman (2001) tarafından tanımlanan RO algoritması, daha güçlü bir model oluşturmak için birden fazla karar ağacının kombinasyonundan yararlanan bir topluluk yaklaşımıdır. Her bir karar ağacı rastgele bir alt örnek üzerinde eğitilir ve bireysel tahminler yapar. RO, aşırı uyum eğilimine karşı dirençlidir ve genellikle yüksek boyutlu veri setlerinde ve yüksek boyutlu özellik uzaylarında etkilidir. Oluşturulan tüm karar ağaçları kullanılarak sınıflandırma veya regresyon işlemi gerçekleştirilir. Sınıflandırma durumunda, her ağaç tek bir sınıf tahmini yapar ve ardından en çok oy alan sınıf tahmini sonuç olarak seçilir. Daha fazla optimizasyon elde etmek için, ağaç sayısı, maksimum derinlik ve minimum bölüm boyutu dahil olmak üzere hiperparametreler ayarlanarak RO'ya ince ayar yapılabilir. Sınıflandırma verilerine dayalı RO gerçekleştirirken Gini endeksinin kullanılmaktadır. Gini endeksinin azalması sınıf homojenliğini artırır. Bölünme işlemlerinde homejen sınıf dağılımına sahip düğümler tercih edilmektedir. Gini endeksinin formülü aşağıdaki gibidir.

$$Gini = 1 - \sum_{i=1}^C (p_i)^2$$

## 2.2. Destek Vektör Makineleri

DVM'nin temel amacı, veri noktalarını sınıflar arasında bir karar sınırı ile ayırmaktır. Bu karar sınırı, veri noktalarının en iyi şekilde ayrılmasını sağlamak için optimize edilir. DVM, sınıflar arasındaki marjı (mesafe) maksimize etmeye çalışarak, veri noktalarını ayıran en iyi karar sınırını bulur. DVM'nin ana avantajlarından biri, yüksek boyutlu veri kümeleri üzerinde iyi performans göstermesi (Suthaharan ve Suthaharan, 2016) ve ayrık, sürekli veya karmaşık veri türlerini işleyebilmesidir. Model, çekirdek fonksiyonu seçimi ve düşük marjinal hata için tolerans parametresi gibi çeşitli hiperparametreler ayarlanarak optimize edilebilir. İki boyutlu uzayda, optimal hiper-düzlem, uzayı iki parçaya bölen bir çizgi olarak düşünülebilir (Hearst vd., 1998). Uzayın bir kısmı bir sınıfa ait veri noktalarını içerirken, uzayın diğer kısmı diğer sınıfa ait veri noktalarını içerir. Bir sınıflandırıcı olarak hareket eden çizgiler kavramı, yalnızca veri noktaları doğrusal olarak ayrılabilirse geçerlidir. DVM, doğrusal olarak ayrılamayan veri noktalarını sınıflandırmak için kullanılabilir en uygun eğriyi bulmak için de kullanılabilir. Bu sınıflandırmayı yapabilmek için iki sınıfı ayıran bir doğru çizilir. Bu doğrunun arasında kalan bölgeye Margin adı verilir. Margin ne kadar geniş ise sınıflandırma o kadar başarılı olur. Aşağıda buna ilişkin formül verilmiştir.

$$y = \begin{cases} 0 & \text{if } w^T * x + b < 0, \\ 1 & \text{if } w^T * x + b \geq 0 \end{cases}$$

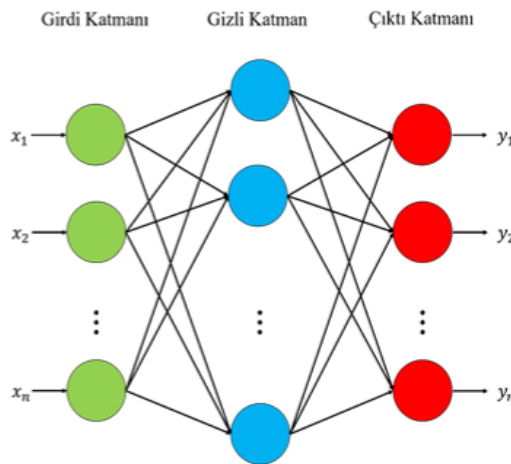
Formülde; w ağırlık vektörü, x girdi vektörü, b ise sapmayı ifade etmektedir.

## 2.3. Çok Katmanlı Algılayıcı (ÇKA)

Yapay sinir ağlarının bir türü olan ve en azından bir gizli katman içeren bir mimariyi ifade eden model girdi katmanı, en az bir gizli katman ve çıktı katmanından oluşur. Gizli katmanlardaki her bir düğüm, girdi katmanındaki girdilerin birleşik bir ağırlık toplamına dayanarak bir aktivasyon fonksiyonu tarafından işlenmesiyle oluşturulur (Gardner ve Dorling, 1998). ÇKA, özellikle sınıflandırma ve regresyon problemleri için yaygın olarak kullanılan bir modeldir. ÇKA öncelikle geriye yayılma algoritması yoluyla öğrenir ve gizli katman sayısı, nöron sayısı, aktivasyon fonksiyonlarının seçimi ve daha fazlası dahil olmak üzere hiper parametreler ayarlanarak daha da optimize edilebilir (Wilson ve Tufts, 1994).

ÇKA modelinde, her bir gizli katmanın çıktısı, bir önceki katmanın çıktısı girdi olarak kullanılarak elde edilir. Son çıktı katmanı, tahmin sonucunu üretmek için bir aktivasyon fonksiyonu uygular. Regresyon problemleri için genellikle doğrusal aktivasyon fonksiyonları kullanılırken, sınıflandırma problemleri için sigmoid veya softmax gibi uygun aktivasyon fonksiyonları kullanılır. ÇKA mimarisi Şekil-1 de gösterilmektedir. Her katman  $y = f(Wx + b)$  olarak temsil edilir. Burada f, aktivasyon fonksiyonudur. W, katmandaki parametre veya ağırlıklar kümesidir, x, aynı zamanda önceki katmanın çıkışı da olabilen giriş vektörüdür ve b, bias vektörüdür. ÇKA'nın katmanları, tamamen bağlı birkaç katmandan oluşur çünkü bir katmandaki her birim, bir önceki katmandaki tüm birimlere bağlıdır.

Şekil 1. ÇKA Mimarisi



Kaynak: Yazarlar.

## 2.4. Veri Seti

Bu çalışmada kullanılan veri seti, IBM Analytics tarafından oluşturulmuştur. Kaggle'ın yarışmalarında yayınlanan 'IBM HR Analytics Çalışan Yıpranması ve Performans' verileridir (IBM HR Analytics Employee, 2017). 1470 çalışan için 35 özellik içermektedir. "Yıpranma" özelliği, çalışanın şirkette kalma veya ayrılma kararını temsil eder. Tüm özellikler çalışanların çalışma hayatı ve kişisel özellikleri ile ilgilidir. Tablo 1'de veri setinde bulunan özellikler listelenmiştir.

**Tablo 1.** Veri Seti Özellikleri

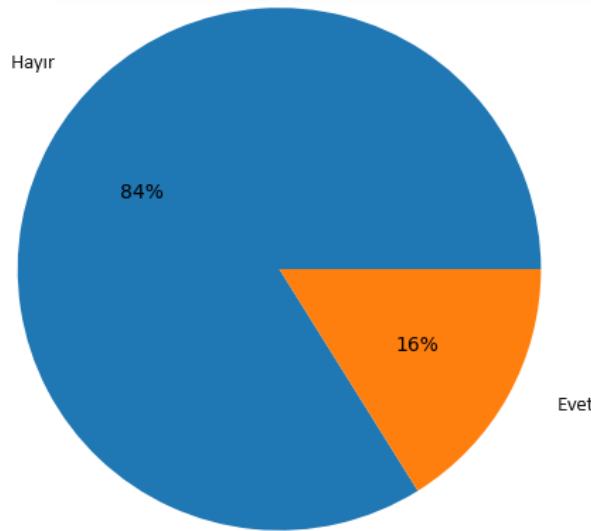
Yaş	Çalışan Yıpranması	Seyahat Durumu
Günlük Ücret	Departman	Evden uzaklık
Eğitim	Eğitim Alanı	İşçi Sayısı
Çalışan Numarası	Çevre Memnuniyeti	Cinsiyet Eğitim Süreleri
Saatlik Ücret	İşe Katılım	İş Seviyesi
İş Rolü	İş Tatmini	Medeni Durum
Aylık Gelir	Aylık Ücret	Önceki İşveren Sayısı
18 Yaş Üstü	Fazla Mesai	Yüzde Maaş Artışı
Performans Derecesi	İlişki Memnuniyeti	Standart Saatler
Hisse Senedi Opsiyonu Seviyesi	Toplam Çalışma Yılı	Son sene
İş Yaşam Dengesi	Şirkette Geçirdiği Yıl	Mevcut Görevdeki Yıl
Son Promosyon Zamanı	Mevcut Yönetici ile Kaç Yıl	

**Kaynak:** Yazarlar.

Veri setinde yapılan basit bir inceleme, çalışan sayısı, 18 yaş üstü ve standart çalışma saatleri gibi bazı özelliklerin tüm çalışanlar için aynı olduğunu ortaya koyuyor, bu nedenle bu özellikler bu aşamada ihmal edilmiştir. Ayrıca, çalışan numarası özelliği de ihmal edilmiştir çünkü değerleri sınıflandırma problemimizle ilgili değildir.

Veri seti incelendiğinde bazı özellikler sayılar yerine kategorik değerlerdir. Makine öğrenmesi yöntemlerinde kullanmak için kategorik veriler sayılara dönüştürülmüştür. Orijinal veri seti, iş seyahati, departman, eğitim alanı, cinsiyet, iş rolü, medeni durumu ve mesai gibi birçok kategorik özelliği içerir. Öncelikle benzersiz değerler ve sayıları belirlenir. Daha sonra her değer için bir tek etiketli ikili vektör atanır. İki kategorili kategorik değişkenlerde, sayılar 0 ve 1'e eşlenmiştir. Üç veya daha fazla kategoriye sahip kategorik değişkenlerde ise, değişken bağımsızlığını korurken kategorik değişkenleri sayısal verilere dönüştürmek için one-hot dönüşümü kullanılmıştır. Makine öğrenmesi modelleri, eğitim sürecindeki özellikler arasındaki göreceli değerlerin büyüklüğüne karşı duyarlıdır. Bir özelliğin diğerlerine üstün gelmesini önlemek için özellik ölçeklendirmesi kullanılmıştır. Bu çalışmada özellik değerleri [0, 1] aralığına yeniden ölçeklendirilmiştir.

Çalışan yıpranması oranları dengesiz olduğundan ADASYN algoritması uygulanarak yeni örnekler oluşturulmuştur. Şekil 2'de bu dengesizliğin grafik üzerinde gösterimi yapılmıştır. ADASYN algoritması yoğunluk dağılımı, her bir azınlık sınıfı örneği için oluşturulması gereken sentetik örneklerin sayısına otomatik olarak karar vermek için kullanılır. Algoritmanın temel işleyişi, her örnek için farklı miktarlarda sentetik veri üretmek amacıyla farklı azınlık sınıfı örneklerine ağırlıklar vermesidir (Gosaion ve Sardana, 2017). ADASYN algoritması kullanımı sonrası veri setinin büyüklüğü 2406 örneğe çıkmıştır. Yeni veri setinde çalışan yıpranması dağılımı (Evet:1173, Hayır:1233) şeklinde dengeli bir şekilde oluşmuştur.

**Şekil 2.** Çalışan Yıpranması Dağılımı

**Kaynak:** Yazarlar.

## 2.5. Model Tasarımı

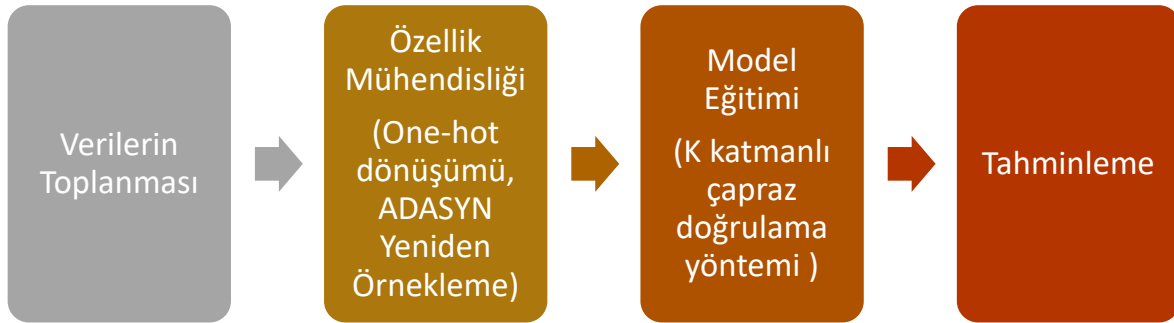
Bu çalışmada RO, ÇKA, DVM, ve bu bireysel modeller kullanılarak oluşturulan 2 tip topluluk modeli, çalışanların yıpranmasını tahmin etmek için oluşturulmuştur. Amaç, analiz edilen problem için en iyi sınıflandırıcıyı belirlemektir. Bu çalışmada, çalışan yıpranması tahmininde kullanılan geleneksel tekniklerin sınırlamalarını gidermek için bir dizi topluluk öğrenmesi yöntemi kullanılmıştır. RO, DVM ve ÇKA gibi algoritmalar, genellikle sınıflandırma problemlerinde güçlü performans gösterirler ve özellik uzayındaki karmaşıklığı iyi yönetirler. Çalışmamızda KNN ve Naif Bayes yöntemlerinin kullanılmamasının sebebi çalışma öncesi bu modellerinin denenip performans sonuçlarının diğer algoritmalara göre düşük olmasıdır.

Yığın topluluğu öğrenimi yaklaşımı, birden fazla modeli aynı anda eğitmeyi ve tahminler oluşturmak için çıktıların birleştirmeyi içerir. Bu teknik, her modelin farklı özelliklerinden ve güçlü yönlerinden yararlanarak daha doğru ve güvenilir tahminlere yol açar. İlk adımda kullanılan sınıflandırma modellerini ve bunlara karşılık gelen eğitim sürecini ortaya koymaktadır. Yığın topluluğu yöntemi birden fazla sınıflandırma modelinin aynı anda kullanılmasını içerir. Bu yaklaşımda çeşitli sınıflandırma modelleri eğitilir. Her modelden elde edilen daha sonra daha üst düzey bir meta model aracılığıyla birleştirilir ve nihai tahmin değeri elde edilir. Nihai tahmin, eğitilen modellerden gelen tahminlerin birleşimi yoluyla meta model tarafından üretilir. Özellikle meta aşamada lojistik regresyon kullanılarak yapılan toplama işlemi, temel modellerin tahminlerini birleştirerek daha yüksek bir doğruluk sağlaması amaçlanmıştır.

Öne çıkan bir diğer topluluk yöntemi, sağlam tahminler yapmak için kullanılan Oylama Sınıflandırıcısıdır. Oylama Sınıflandırıcısı, birden fazla temel tahmincinin (makine öğrenimi modelleri) tahminlerini birleştiren ve oy alarak sınıf etiketini tahmin eden bir topluluk öğrenme yöntemidir. Hem sınıflandırma hem de regresyon problemlerine uygulanabilir. Zorunlu oylamada, her temel tahminci (model) sınıf etiketini tahmin eder ve nihai tahmin çoğunluk oyu ile belirlenir. En çok oyu alan sınıf etiketi son tahmin olur. Yumuşak oylamada her temel tahminci, tüm sınıflar üzerindeki olasılık dağılımını tahmin eder. Nihai tahmin, bu olasılıkların ortalaması alınarak ve ardından en yüksek ortalama olasılığa sahip sınıfın seçilmesiyle belirlenir.

Tahmin modelinin performansını değerlendirmek için veri seti eğitim ve test veri seti olmak üzere iki bölüme ayrılmıştır. Eğitim veri seti %70'lik kısımdan test veri seti %30'luk kısımdan oluşturuldu. Eğitim seti modelin verilerde gizli olan ilişkileri öğrenmesine olanak sağlamak amacıyla ayrıldı. Test verisi ise modeli doğrulamak için kullanıldı. Aşırı uyum sorunlarını önlemek ve modeli basitleştirmek için k katmanlı çapraz doğrulama yöntemi kullanılmıştır. Eğitim seti rastgele beş parçaya bölündükten sonra, biri doğrulama seti olarak, diğer k-1'ler eğitim seti olarak kullanılmıştır. Bu işlem k kere gerçekleştirilmiştir. Her yinelemede farklı bir parça doğrulama seti olarak alınmıştır. Finalde k kerelik uygulamaların doğruluk değerlerinin ortalaması kullanılmıştır. Şekil 3'te tahmin modelinin grafiği gösterilmektedir.

Şekil 3. Önerilen Model



Kaynak: Yazarlar.

### 3. Analiz Sonuçları

Makalede sunulan modellerin performansı çeşitli ölçümler kullanılarak değerlendirildi. Performans değerlendirmesinde her bir algoritmanın başarısını ölçmek amacıyla doğruluk, kesinlik, duyarlılık ve F1 skorları kullanılmıştır. Doğruluk, modelin doğru tahmin ettiği örneklerin oranıdır. En yüksek değeri 1 olabilir. Kesinlik ise pozitif olarak sınıflandırılan örneklerin ne kadarının gerçekten pozitif olduğunu ölçer. Duyarlılık ise, gerçek pozitif örneklerin ne kadarının pozitif olarak tahmin edildiğini gösterir. F1-skoru ise sınıflandırma modelinin kesinlik ile duyarlılık değerlerinin harmonik ortalamasıdır. Bu metriklerin formülleri aşağıdaki gibidir.

$$\text{Doğruluk} = \frac{(TP+TN)}{TP+FP+TN+FN}$$

$$\text{Kesinlik} = \frac{TP}{TP+FP}$$

$$\text{Duyarlılık} = \frac{TP}{TP+FN}$$

$$\text{F1 Skor} = 2 * \frac{\text{Duyarlılık} * \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}}$$

Formüllerdeki;

**TP (True Positive):** doğru pozitif sayısı

**TN (True Negative):** doğru negatif sayısı

**FP (False Positive):** yanlış pozitif sayısı

**FN (False Negative):** yanlış negatif sayısıdır.

Topluluk öğrenme modelleri; RO, DVM ve ÇKA gibi tekli modellerinden daha iyi performans gösterdi. Bu, farklı modellerin birleştirilmesinden elde edilecek sinerjinin, bireysel modellerin performansından daha büyük olabileceğini göstermektedir. Tekli modeller arasında RO ve ÇKA'nın en iyi doğruluğa sahip olduğu görüldü. Çeşitli model kombinasyonlarından oluşan topluluk modeli, tekli modele göre daha iyi performans gösterse de, belirli algoritmaların kombinasyonuna bağlı olarak topluluk modelinin performansı farklılık göstermektedir.

Tablo 2, sentetik veriler ile genişletilen veri setinin test verileri aracılığıyla yapılan yıpranma tahmin değerleri ile ilgili olarak doğruluk, kesinlik, duyarlılık ve F1 skorları sonuçlarını göstermektedir. Tablo 3 ise orijinal veri seti ile eğitilen modelin test verileri aracılığıyla yapılan yıpranma tahmin değerleri ile ilgili olarak doğruluk, kesinlik, duyarlılık ve F1 skorları sonuçlarını göstermektedir. Ayrıca sentetik olarak üretilen veri ile orijinal veri kullanıldığında çıkan sonuç farklılıkları bu iki tablo incelendiğinde net olarak gözükmemektedir. Sentetik olarak üretilen veri ile sınıf dengesizliğinin ortadan kaldırılması sonuçları olumlu yönde etkilemiştir.

**Tablo 2.** Model Performansları (Sentetik Veri)

Model	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RO	0.93	0.94	0.93	0.93
DVM	0.91	0.91	0.91	0.91
ÇKA	0.94	0.95	0.94	0.94
Yığın Topluluk Öğrenme Modeli	0.96	0.96	0.96	0.96
Oylama Sınıflandırıcı Topluluk Modeli	0.95	0.95	0.95	0.95

**Kaynak:** Yazarlar.

**Tablo 3.** Model Performansları (Orijinal Veri)

Model	Doğruluk	Kesinlik	Duyarlılık	F1 Skor
RO	0.87	0.72	0.55	0.55
DVM	0.89	0.89	0.60	0.64
ÇKA	0.85	0.65	0.62	0.63
Yığın Topluluk Öğrenme Modeli	0.88	0.77	0.61	0.64
Oylama Sınıflandırıcı Topluluk Modeli	0.86	0.68	0.62	0.64

**Kaynak:** Yazarlar.

Model tahmin performansının analiz edilmesi sonucunda çalışan yıpranması, RO, DVM ve ÇKA ile oluşturulan yığın topluluk öğrenme modeli, doğruluk (0.96), F1 puanı (0.96), kesinlik (0.96) ve duyarlılık (0.96) gibi tüm göstergelerde en iyi performansı göstermiştir. Oylama sınıflandırıcı topluluk modeli de 0.95 doğruluk oranı ile başarılı bir performans sergilemiştir. En düşük performans gösteren model ise 0.91 doğruluk oranı ile DVM modelidir. Topluluk öğrenme modellerinin tek başına makine öğrenmesi modellerine göre daha yüksek performans gösterdiği görülmektedir.

Orijinal veri ile yapılan çalışan yıpranması modellerinin performansı sentetik veri ile genişletilen veri setine göre çok daha düşüktür. Bunun ilk sebebinin sınıflar arasındaki dengesizlikten kaynaklandığı söylenebilir. Orijinal veri modellerinde en iyi sonucu 0.89 doğruluk oranı ile DVM modeli sergilemiştir. Topluluk modellerinden yığın model (0.88) ve oylama sınıflandırıcı model (0.86) doğruluk oranı yakalamıştır.

## Tartışma ve Sonuç

Önerilen çalışma insan kaynakları departmanı için bir çalışanın işten ayrılma niyeti hakkında işletmelere önemli bilgiler sunabilir. Çalışan yıpranmasını tahmin eden çalışma, bu amaçla çeşitli makine öğrenmesi yöntemlerini ve bu yöntemler kullanılarak oluşturulan topluluk öğrenme yöntemlerini kullanmıştır. Genel değerlendirme için doğruluk, kesinlik, duyarlılık ve F1 skoru kullanılmıştır. Yığın topluluk öğrenme modeli 0.96 doğruluk oranı ile en iyi sonucu vermiştir. Tek modellerin sınırlamalarının topluluk modelleri ile avantaja dönüştürüldüğü ve sonuçlara yansıtıldığı görülmektedir.

Aynı veri setini kullanan diğer çalışmalar ile karşılaştırıldığında modelimiz kayda değer bir performans göstermiştir. Tablo 4'te aynı veri setini kullanan diğer çalışmalar listelenmiştir.

**Tablo 4.** Aynı Veri Setini Kullanan Diğer Çalışmalar

Referans	Model	Doğruluk Oranı
Alduayj ve Rajpoot, 2018	RO	0.91
Celik, 2019	DVM	0.91
Fallucchi vd., 2020	Doğrusal DVM	0.87
Al-Darraji vd., 2021	Derin Sinir Ağları	0.94
Kaya ve Korkmaz, 2021	Radyal Tabanlı Fonksiyon Ağı	0.80
Raza vd., 2022	Ekstra Ağaçlar	0.93
Alshiddy ve Aljaber, 2023	İç İçe Topluluk Öğrenme	0.94
Alsubaie ve Aldoukhi (2024)	Lojistik Regresyon	0.87

**Kaynak:** Yazarlar.



Tablo 4 aynı veri setinin kullanan diğer çalışmalar arasındaki karşılaştırmayı göstermektedir. Her çalışmada kullanılan özellik mühendisliği yöntemleri, veri ön işleme ve örnekleme yöntemleri farklı olsa da doğruluk oranları karşılaştırıldığında modelin başarılı olduğu söylenebilir. Tablodaki çalışmalara bakıldığında Alshiddy ve Aljaber (2023)'in çalışması dışında diğer çalışmalarda tek bir yöntemin kullanıldığı görülmektedir. Bu çalışmaların performanslarına bakıldığında bu çalışmadaki 0.96'lık doğruluk oranının altında kaldığı görülmektedir. Derin Sinir ağlarında birden fazla gizli katmanın kullanılması öğrenme performansını arttırdığı bilinmektedir (Sze vd., 2017). Bu doğrultuda Al-Darraji vd. (2021) çalışmasının sonucu bunu desteklemektedir. Topluluk öğrenme modellerinin ana amacı bireysel modelleri birleştirerek performans artırmaktır (Sagi ve Rokach, 2018). Çalışmada topluluk öğrenme modelinin bu avantajı kullandığı görülmektedir. Kullandığımız modelin, çalışanların yıpranma riskini doğru bir şekilde tahmin etmede karşılaştırma yaptığımız çalışmalardan daha iyi performans göstermiştir.

Bu çalışmada, çalışanların işten ayrılma olasılığını tahmin etmek amacıyla çeşitli makine öğrenmesi algoritmaları ve topluluk öğrenme modelleri kullanılarak IBM Watson Analytics tarafından sağlanan bir veri seti üzerinde analizler yapılmıştır. Giriş bölümünde belirtildiği üzere, çalışanların ayrılma olasılığını önceden tahmin etmek, şirketlerin işe alım ve eğitim maliyetlerini azaltarak verimliliği artırmalarına yardımcı olabilir. Çalışmanın sonuçları, bu problemin çözümüne önemli katkılar sağlamaktadır.

İlk olarak, kullanılan tahmin modelleri ile çalışanların işten ayrılma olasılıklarının yüksek doğrulukla tahmin edilebildiği gösterilmiştir. RO, DVM, ÇKA ve topluluk öğrenme modelleri ile yapılan değerlendirmeler sonucunda, ADASYN yöntemi ile dengelenmiş veri seti üzerinde eğitilen yığın topluluk öğrenme modelinin %96 doğruluk oranı ile en iyi performansı sergilediği tespit edilmiştir. Bu yüksek doğruluk oranı, insan kaynakları departmanlarının, çalışanların memnuniyet ve bağlılığını artırmak için proaktif ve hedefe yönelik stratejiler geliştirmesine olanak tanır. Dolayısıyla, araştırmamızda kullanılan modeller, çalışan bağlılığını artırmak ve ayrılma oranlarını düşürmek isteyen şirketler için pratik bir çözüm sunmaktadır.

İkinci olarak, çalışmamızda kullanılan tahmin yöntemlerinin güçlü yönleri, özellikle veri dengesizliği gibi yaygın bir problemi ele almadaki etkinliği, çözüm sürecine önemli katkılar sağlamıştır. ADASYN yöntemi, dengesiz veri kümelerinde azınlık sınıfın daha iyi temsil edilmesine yardımcı olmuş ve model performansını artırmıştır. Bu yaklaşım, veri dengesizliği problemini çözmek için yenilikçi bir çözüm sunmakta ve bu alandaki diğer araştırmacılar için de değerli bir örnek teşkil etmektedir. Ayrıca, topluluk öğrenme modellerinin çeşitli algoritmaların avantajlarını birleştirerek elde ettiği yüksek doğruluk oranı, bu modellerin karmaşık iş gücü veri setlerinde kullanımı için güçlü bir argüman sunmaktadır.

Son olarak, bu çalışma, çalışan yıpranması ve işten ayrılma tahminine yönelik literatüre anlamlı katkılar sağlamaktadır. Giriş bölümünde öngörülen hipotezlerin doğrulanmış olması, literatürdeki boşlukların doldurulmasına yardımcı olmuştur. Çalışma, topluluk öğrenme modellerinin ve ADASYN yönteminin bu alandaki uygulamaları konusunda literatüre önemli bir katkı sağlamış ve gelecekteki araştırmalar için yeni bir yol haritası çizmiştir.

Gelecekteki çalışmalar, burada elde edilen bulguları genişleterek farklı veri kümeleri veya endüstrilerde kullanılarak modellerin performansını değerlendirebilir. Ayrıca, farklı veri dengeleme yöntemleri veya yeni makine öğrenmesi algoritmalarının kullanılması, model doğruluğunu ve genelleme yeteneğini daha da artırabilir.



This research article has been licensed with Creative Commons Attribution - Non-Commercial 4.0 International License. Bu araştırma makalesi, Creative Commons Atıf - Gayri Ticari 4.0 Uluslararası Lisansı ile lisanslanmıştır.

#### Yazar Katkıları

Yazar çalışmayı tek başına hazırladığını beyan etmiştir.

#### Teşekkür Beyanı

Yazar teşekkür beyanında bulunmamıştır.

#### Destek Beyanı

Yazar destek beyanında bulunmamıştır.

#### Çıkar Çatışması

Yazar herhangi bir çıkar çatışması beyan etmemiştir.

#### Etik Beyanı

Yazar çalışma için Etik Kurul Onayı alınmasına gerek olmadığını beyan etmiştir.

#### Sorumlu Özel Sayı Editörleri

Doç. Dr. Beyza Mina Ordu-Akkaya, Ankara Sosyal Bilimler Üniversitesi

Doç. Dr. Görkem Ataman, Yaşar Üniversitesi

Arş. Gör. Yunus Yıldırım, Afyon Kocatepe Üniversitesi

## Kaynakça/References

- Alao, D. A. B. A., & Adeyemo, A. B. (2013). Analyzing employee attrition using decision tree algorithms. *Computing, Information Systems, Development Informatics and Allied Research Journal*, 4(1), 17–28.
- Al-Darraj, S., Honi, D. G., Fallucchi, F., Abdulsada, A. I., Giuliano, R., & Abdulmalik, H. A. (2021). Employee attrition prediction using deep neural networks. *Computers*, 10(11), 141. <https://doi.org/10.3390/computers10110141>
- Alduayj, S. S., & Rajpoot, K. (2018). Predicting employee attrition using machine learning. In *2018 International Conference on Innovations in Information Technology (IIT)* (pp. 93–98). IEEE. <https://doi.org/10.1109/INNOVATIONS.2018.8605976>
- Aldulaimi, S. H., Abdeldayem, M. M., Mowafak, B. M., & Abdulaziz, M. M. (2021). Experimental perspective of artificial intelligence technology in human resources management. In Hamdan, A., Hassanien, A. E., Khamis, R., Alareeni, B., Razzaque, A., & Awwad, B. (Eds.), *Applications of artificial intelligence in business, education and healthcare* (Vol. 954, pp. 605–619). Springer. [https://doi.org/10.1007/978-3-030-72080-3\\_26](https://doi.org/10.1007/978-3-030-72080-3_26)
- Alsheref, F. K., Fattoh, I. E., & M. Ead, W. (2022). Automated prediction of employee attrition using ensemble model based on machine learning algorithms. *Computational Intelligence and Neuroscience*, 2022(1), 7728668. <https://doi.org/10.1155/2022/7728668>
- Alshiddy, M. S., & Aljaber, B. N. (2023). Employee attrition prediction using nested ensemble learning techniques. *International Journal of Advanced Computer Science and Applications*, 14(7), 932–938. <https://doi.org/10.14569/IJACSA.2023.01400712>
- Alsubaie, F., & Aldoukhi, M. (2024). Using machine learning algorithms with improved accuracy to analyze and predict employee attrition. *Decision Science Letters*, 13(1), 1–18. <https://doi.org/10.5267/j.dsl.2023.12.006>
- Avrahami, D., Pessach, D., Singer, G., & Chalutz Ben-Gal, H. (2022). A human resources analytics and machine-learning examination of turnover: Implications for theory and practice. *International Journal of Manpower*, 43(6), 1405–1424. <https://doi.org/10.1108/IJM-12-2020-0548>
- Barpanda, S., & Athira, S. (2022). Cause of attrition in an information technology-enabled services company: A triangulation approach. *International Journal of Human Capital and Information Technology Professionals (IJHCITP)*, 13(1), 1–22. <https://doi.org/10.4018/IJHCITP.20220101.0a1>
- Bhatta, S., Zaman, I. U., Raisa, N., Fahim, S. I., & Momen, S. (2022, April). Machine learning approach to predicting attrition among employees at work. In *Computer Science On-line Conference* (pp. 285–294). Springer International Publishing. [https://doi.org/10.1007/978-3-030-93515-3\\_25](https://doi.org/10.1007/978-3-030-93515-3_25)
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(5), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Buntak, K., Kovačić, M., & Mutavdžija, M. (2021). Application of artificial intelligence in the business. *International Journal for Quality Research*, 15(2), 403. <https://doi.org/10.24874/IJQR15.02-03>
- Çelik, U. (2019). Veri madenciliği yöntemleri ile iş yaşam dengesinde yıpranma durumu tahmini. *Journal of Management and Economics Research*, 17(1), 63–76. <https://doi.org/10.11611/yead.519923>
- Chung, D., Yun, J., Lee, J., & Jeon, Y. (2023). Predictive model of employee attrition based on stacking ensemble learning. *Expert Systems with Applications*, 215, 119364. <https://doi.org/10.1016/j.eswa.2022.119364>
- Clark, A. E. (2001). What really matters in a job? Hedonic measurement using quit data. *Labour Economics*, 8(2), 223–242. [https://doi.org/10.1016/S0927-5371\(00\)00041-9](https://doi.org/10.1016/S0927-5371(00)00041-9)
- Douaidi, L., & Kheddouci, H. (2022, September). A new approach for employee attrition prediction. In *International Conference on Conceptual Structures* (pp. 115–128). Springer International Publishing. [https://doi.org/10.1007/978-3-030-76294-0\\_9](https://doi.org/10.1007/978-3-030-76294-0_9)
- El-Rayes, N., Fang, M., Smith, M., & Taylor, S. M. (2020). Predicting employee attrition using tree-based models. *International Journal of Organizational Analysis*, 28(6), 1273–1291. <https://doi.org/10.1108/IJOA-10-2019-1903>
- Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers*, 24(5), 1709–1734. <https://doi.org/10.1007/s10796-021-10186-w>
- Fallucchi, F., Coladangelo, M., Giuliano, R., & De Luca, E. W. (2020). Predicting employee attrition using machine learning techniques. *Computers*, 9(4), 86. <https://doi.org/10.3390/computers9040086>
- Frye, A., Boomhower, C., Smith, M., Vitovsky, L., & Fabricant, S. (2018). Employee attrition: What makes an employee quit?. *SMU Data Science Review*, 1(1), 9.
- Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14–15), 2627–2636. [https://doi.org/10.1016/S1352-2310\(97\)00447-0](https://doi.org/10.1016/S1352-2310(97)00447-0)

- Garg, S., Sinha, S., Kar, A. K., & Mani, M. (2022). A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management*, 71(5), 1590–1610. <https://doi.org/10.1108/IJPPM-08-2020-0427>
- Gosain, A., & Sardana, S. (2017). Handling class imbalance problem using oversampling techniques: A review. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 79–85). IEEE. <https://doi.org/10.1109/ICACCI.2017.8125865>
- Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications*, 13(4), 18–28. <https://doi.org/10.1109/5254.708428>
- Hoffman, M., & Tadelis, S. (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy*, 129(1), 243–285. <https://doi.org/10.1086/712436>
- IBM HR Analytics Employee. (2017). *IBM HR Analytics Employee Dataset*. Kaggle. <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>
- Jain, P. K., Jain, M., & Pamula, R. (2020). Explaining and predicting employees' attrition: A machine learning approach. *SN Applied Sciences*, 2(4), 757. <https://doi.org/10.1007/s42452-020-2519-4>
- Kaya, İ. E., & Korkmaz, O. (2021). Machine learning approach for predicting employee attrition and factors leading to attrition. *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi*, 36(4), 913–928. <https://doi.org/10.21605/cukurovaumfd.1040487>
- Khalid, S. M., Rashid, S., & Ullah, R. (2022). Employee retention prediction using machine learning techniques: A review of the state of the art. *International Journal of Electrical and Computer Engineering*, 12(4), 4498–4506. <https://doi.org/10.11591/ijece.v12i4.9440>
- Khare, R., Kaloya, D., Choudhary, C. K., & Gupta, G. (2011, January). Employee attrition risk assessment using logistic regression analysis. In *Proceedings of the International Conference Advanced Data Analytics Business Analytics Intelligence* (pp. 1–33).
- Kumar, N., & Yakhlef, A. (2016). Managing business-to-business relationships under conditions of employee attrition: A transparency approach. *Industrial Marketing Management*, 56, 143–155. <https://doi.org/10.1016/j.indmarman.2016.01.002>
- Oruç, E. (2020). Örgütsel yıpranma üzerine kuramsal bir inceleme. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 66, 319–334.
- Oshagbemi, T. (2003). Personal correlates of job satisfaction: Empirical evidence from UK universities. *International Journal of Social Economics*, 30(12), 1210–1232. <https://doi.org/10.1108/03068290310504380>
- Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting employee attrition using machine learning approaches. *Applied Sciences*, 12(13), 6424. <https://doi.org/10.3390/app12136424>
- Rutherford, M. W., Buller, P. F., & McMullen, P. R. (2003). Human resource management problems over the life cycle of small to medium-sized firms. *Human Resource Management*, 42(4), 321–335. <https://doi.org/10.1002/hrm.10093>
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249. <https://doi.org/10.1002/widm.1249>
- Sriram, K. V., Joseph, J., Mathew, A. O., & Rai, A. S. (2019). Factors affecting high employee attrition in manufacturing firms—A case study. *Calitatea*, 20(169), 23–28.
- Srivastava, D. K., & Nair, P. (2018). Employee attrition analysis using predictive techniques. In *Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 1* (pp. 293–300). Springer International Publishing. [https://doi.org/10.1007/978-3-319-98516-7\\_32](https://doi.org/10.1007/978-3-319-98516-7_32)
- Subramony, M., & Holtom, B. C. (2012). The long-term influence of service employee attrition on customer outcomes and profits. *Journal of Service Research*, 15(4), 460–473. <https://doi.org/10.1177/1094670512452792>
- Suthaharan, S. (2016). Support vector machine. In *Machine Learning Models and Algorithms for Big Data Classification* (Vol. 36, pp. 393–405). Springer. [https://doi.org/10.1007/978-1-4899-7641-3\\_9](https://doi.org/10.1007/978-1-4899-7641-3_9)
- Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12), 2295–2329. <https://doi.org/10.1109/JPROC.2017.2763201>
- Wilson, E., & Tufts, D. W. (1994). Multilayer perceptron design algorithm. In *Proceedings of IEEE Workshop on Neural Networks for Signal Processing* (pp. 61–68). IEEE. <https://doi.org/10.1109/NNSP.1994.336191>
- Yadav, S., Jain, A., & Singh, D. (2018, December). Early prediction of employee attrition using data mining techniques. In *2018 IEEE 8th International Advance Computing Conference (IACC)* (pp. 349–354). IEEE. <https://doi.org/10.1109/IACC.2018.8710222>
- Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. *IEEE Access*, 9, 60447–60458. <https://doi.org/10.1109/ACCESS.2021.3074559>