

Araştırma Makalesi / Research Article

CNN TABANLI DERİN ÖĞRENME VE MAKİNE ÖĞRENMESİ TEKNİKLERİNİN ENTEGRASYONU: İŞTEN AYRILMA TAHMİNLERİNDE YENİ BİR METODOLOJİ

Yunus Emre GÜR¹ , Cem AYDEN² 

ÖZET

İşgücü devri, organizasyonlar için önemli maliyet ve verimlilik kayıplarına yol açmaktadır. Bu çalışma, işten ayrılma tahminlerini geliştirmek amacıyla, geleneksel istatistiksel modellerin ötesine geçerek makine öğrenimi ve derin öğrenme tekniklerini entegre eden yenilikçi bir yaklaşım sunmaktadır. Çalışma, veri setindeki değişkenleri 2B karekod görüntülerine dönüştürmek suretiyle, CNN tabanlı derin öğrenme modellerinin bu görüntüler üzerinde sınıflandırma yapabilmesini sağlamıştır. Bu yenilikçi adım, derin öğrenme modellerinin görsel veri işleme yeteneklerini kullanarak daha karmaşık veri yapılarını analiz etme potansiyelini ortaya koymaktadır. Araştırma, çeşitli makine öğrenmesi modellerini değerlendirdikten sonra ResNet-18 modeli kullanılarak derin öğrenme tabanlı özellik çıkarımı gerçekleştirilmiştir. Daha sonra, RelieF algoritması kullanılarak seçilen en etkili 10 özelliğe dayanarak optimize edilmiş Hafif Gradyan Artırma (LightGBM) modeli, %100 doğruluk, %100 hassasiyet ve %100 F1-skoru gibi mükemmel performans metrikleri elde etmiştir. Bu sonuçlar, bu modelin işten ayrılma tahminlerinde yüksek etkinlik sergilediğini ve insan kaynakları yönetimi pratiğine önemli katkılarda bulunabileceğini göstermektedir.

Anahtar Kelimeler: CNN Tabanlı Derin Öğrenme, Makine Öğrenmesi, İşten Ayrılma Tahmini, RelieF Özellik Seçimi, 2B Karekod Dönüşümü.

JEL Sınıflandırması: C45, C53, J63.

INTEGRATION OF CNN BASED DEEP LEARNING AND MACHINE LEARNING TECHNIQUES: A NOVEL METHODOLOGY IN JOB SEPARATION PREDICTIONS

ABSTRACT

Labor turnover leads to significant cost and productivity losses for organizations. This paper presents an innovative approach that goes beyond traditional statistical models and integrates machine learning and deep learning techniques to improve turnover prediction. By transforming the variables in the dataset into 2D QR code images, the study enables CNN-based deep learning models to perform classification on these images. This innovative step demonstrates the potential of deep learning models to analyze more complex data structures using visual data processing capabilities. After evaluating various machine learning models, the researchers performed deep learning-based feature extraction using the ResNet-18 model. Then, based on the 10 most influential features selected using the RelieF algorithm, the

¹ Dr. Öğretim Üyesi, Fırat Üniversitesi, İİBF, Elazığ, Türkiye, yegur@firat.edu.tr.

² Doç. Dr., Fırat Üniversitesi, İİBF, Elazığ, Türkiye, cayden@firat.edu.tr.

optimized Light Gradient Boosting (LightGBM) model achieved excellent performance metrics of 100% accuracy, 100% precision, and 100% F1-score. These results show that this model exhibits high efficiency in turnover prediction and can make significant contributions to human resource management.

Keywords: CNN-Based Deep Learning, Machine Learning, Job Separation Estimation, Relief Feature Selection, 2D QR Transformation.

JEL Classification Codes: C45, C53, J63.

EXTENDED SUMMARY

Research Questions & Purpose

Employee turnover has been a persistent challenge for organizations, leading to financial and operational losses. Traditional statistical models have been widely used to predict turnover; however, their effectiveness diminishes when dealing with large and complex datasets. This study integrates machine learning and deep learning techniques to improve turnover prediction by leveraging Convolutional Neural Networks (CNNs) for feature extraction and machine learning models for classification. Unlike previous studies, which primarily rely on structured tabular data, this research transforms dataset variables into 2D QR code images, allowing CNN-based models to analyze the data more comprehensively. The research questions guiding this study are:

1. How do different machine learning methods impact turnover prediction accuracy?
2. What is the effect of transforming structured data into 2D QR code images on the performance of deep learning models?
3. To what extent does the combination of CNN-based feature extraction and machine learning models improve classification accuracy? By addressing these questions, this study aims to provide an innovative methodology that enhances turnover prediction accuracy, thereby contributing to the field of human resource management and predictive analytics.

Literature Review

The literature review was conducted using databases such as Scopus, Web of Science, and Google Scholar. The review identified two primary approaches to turnover prediction: traditional statistical methods (e.g., logistic regression, ANOVA) and machine learning techniques (e.g., random forests, support vector machines). While traditional models offer interpretability, they often fail to capture non-linear relationships in large datasets. Recent studies have explored the application of artificial intelligence (AI) techniques in turnover prediction, with an emphasis on ensemble learning and deep learning models. Several studies have implemented feature selection techniques, such as ReliefF, to improve prediction accuracy. However, limited research has explored the use of image-based transformations for structured data in turnover prediction. This study contributes to the literature by introducing a novel approach that utilizes 2D QR code transformation, allowing CNN models to extract intricate patterns from data. Furthermore, by integrating CNN-based feature extraction with optimized machine learning models, this research bridges a critical gap in predictive analytics for employee turnover.

Methodology

This study adopts an applied research approach with a descriptive and causal design. The dataset used in this study was sourced from Kaggle's "Employee Attrition Dataset," which contains employee demographic, job satisfaction, and organizational factors. The dataset comprises 74,498 observations, with a balanced distribution of employees who stayed (0) and left (1).

Research Model and Hypotheses

The study proposes the following hypotheses:

- H1: Different machine learning methods exhibit varying levels of accuracy in predicting turnover.
- H2: Transforming structured data into 2D QR code images enhances the performance of deep learning models in turnover prediction.
- H3: CNN-based feature extraction, when integrated with machine learning classifiers, significantly improves prediction accuracy.

Data Preprocessing & Analysis

The dataset was preprocessed by removing irrelevant features, applying one-hot encoding for categorical variables, and normalizing numerical attributes using Standard Scaler. It was then split into 80% training and 20% testing sets. Each observation was converted into QR codes using the qrcode and Pillow libraries, enabling CNN models (AlexNet, MobileNet, ResNet-18, SqueezeNet) to extract features. ResNet-18 was used for feature extraction, and the 10 most significant features were selected via ReliefF. These features were then fed into machine learning classifiers (Decision Trees, Gradient Boosting, KNN, LightGBM, Logistic Regression, Random Forest) for turnover prediction.

Ethical Considerations

Since the dataset was publicly available, no ethical approval was required.

Results and Conclusion

The findings indicate that machine learning models exhibit varying degrees of performance, with LightGBM achieving the highest accuracy (76%) and ROC-AUC (0.855). The QR code transformation significantly improved the CNN models' ability to extract meaningful features, enhancing classification accuracy. The integration of ResNet-18-based feature extraction with LightGBM resulted in optimal performance, achieving 100% accuracy, 100% precision, and 100% F1-score. These results demonstrate that the proposed methodology provides a highly effective approach for turnover prediction. The combination of deep learning feature extraction and machine learning classifiers offers a robust framework for handling complex employee datasets. This study contributes to the field by introducing an innovative approach that leverages QR code transformations for structured data, enhancing the predictive capabilities of AI-driven turnover models. Future research should explore the applicability of this methodology in different industries and expand the model's generalizability by incorporating additional organizational factors. Moreover, investigating interpretability techniques for CNN-extracted features could provide further insights into employee turnover drivers, aiding decision-makers in strategic workforce planning.

1. Giriş

Teknolojinin hızlı gelişimi ve global ekonomik koşullardaki dalgalanmalar, iş dünyasını sürekli bir dönüşüm sürecine itmektedir. Bu dönüşüm, işgücü piyasasındaki dinamikleri derinden etkilemekte ve işveren ile çalışan arasındaki ilişkilerde yeni trendlerin ortaya çıkmasına neden olmaktadır. Çalışanların işten ayrılması, özellikle bilgi yoğun sektörlerde, organizasyonlar için hem finansal hem de operasyonel büyük zorluklar yaratmaktadır (Marquez, 2023). İşten ayrılmanın neden olduğu bilgi kaybı, yetenek açığı ve eğitim yatırımlarının kaybı gibi direkt maliyetler, şirketlerin sürdürülebilirliğini ve büyüme potansiyelini doğrudan tehdit etmektedir (Kanuto, 2024). Ayrıca, işten ayrılma oranlarının yüksek olması, çalışan bağlılığını ve toplam iş verimliliğini olumsuz yönde etkileyebilmekte, bu da müşteri memnuniyeti ve şirket itibarına zarar verebilmektedir (Liao, 2023). Bu bağlamda, işten ayrılma eğilimlerini öngörebilen ve bu eğilimlere proaktif olarak müdahale edebilen tahmine dayalı modellerin geliştirilmesi, şirketler için stratejik bir gereklilik haline gelmiştir (Al Akasheh, 2024). Etkili çalışan yönetimi uygulamaları, şirketlerin bu tür zorluklarla başa çıkmalarını sağlayarak, rekabet avantajı elde etmelerine ve işgücü piyasasında üstün bir konum kazanmalarına yardımcı olmaktadır (Wijaya vd., 2021).

Bu geniş kapsamlı zorluklar çerçevesinde, işten ayrılma eğilimlerinin saptanması ve analizi için geleneksel yöntemlerin yanı sıra, yapay zekâ tekniklerinin devreye sokulması kaçınılmaz hale gelmiştir. Geleneksel yöntemler bazı temel iç görüler sağlasa da, iş dünyasının sürekli değişen ve karmaşık yapısını tam olarak yansıtamamaktadır. Bu nedenle, makine öğrenimi ve derin öğrenme gibi teknolojiler, işten ayrılma analizi ve yönetiminde yeni ufuklar açmaktadır. Geleneksel istatistiksel yöntemler incelendiğinde, işten ayrılma tahminlerinde uzun yıllardır kullanılmakta olan regresyon analizleri, lojistik regresyon ve ANOVA gibi tekniklerden faydalanılmıştır. Bu yöntemler, belirli koşullar altında ve sınırlı değişkenlerle etkili olabilmektedir (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, yapay zekâ teknikleri ise, büyük ve karmaşık veri kümelerinden öğrenme yeteneği sayesinde, geleneksel yöntemlerin aşamadığı birçok zorluğu aşmaktadır (Bae vd., 2021). Makine öğrenimi modelleri, veriler arasındaki gizli kalıpları ve ilişkileri tanıyarak, değişkenlerin etkileşimlerini ve bu etkileşimlerin işten ayrılma üzerindeki etkilerini çok daha başarılı bir şekilde modelleyebilmektedir (Grebovic vd., 2022). Derin öğrenme, özellikle büyük veri setleri üzerinde, insan müdahalesi olmaksızın özellik çıkarımı yapabileme kabiliyeti ile öne çıkmakta, bu sayede çok daha dinamik ve adapte olabilir modeller geliştirilmektedir (Wang ve Zhao, 2022). Genel olarak, işten ayrılma tahmininde yapay zekâ ve makine öğrenimine geçiş, kuruluşların işgücü yönetimi yaklaşımında bir paradigma değişikliğine işaret etmektedir. Şirketler, gelişmiş algoritmaların ve veriye dayalı iç görülerin gücünden yararlanarak, yetenekleri elde tutma ve stratejik işgücü planlamasında rekabet avantajı elde edebilir ve sonuçta daha iyi kurumsal performans ve istikrar sağlayabilir (Pakarinen vd., 2022).

Bu bilgiler ışığında, geleneksel yöntemler ve makine öğrenimi modelleri arasındaki farkların net bir şekilde görülmesi, işten ayrılma analizinde yeni ve daha etkin yöntemlerin araştırılmasını zorunlu kılmaktadır. Bu çalışma, mevcut modellerin sınırlarını zorlayarak, derin öğrenme tekniklerini kullanmanın potansiyel avantajlarını derinlemesine incelemektedir. Çalışmamız, makine öğrenimi ve derin öğrenme modellerini benzersiz bir şekilde entegre ederek işten ayrılma tahminlerinde farklılık yaratmayı hedeflemektedir. Özellikle, derin öğrenme

modelleri için özgün bir uygulama olan veri setindeki her bir satırın 2D barkod görüntülerine dönüştürülmesi, bu alanda bir ilk teşkil etmektedir. Bu dönüşüm, verilerin görsel olarak işlenmesi ve öğrenilmesi sürecinde derin öğrenme algoritmalarının doğal güçlerini kullanarak, geleneksel veri işleme tekniklerinin ötesine geçmeyi mümkün kılmaktadır. Bu gelişmiş yaklaşımların entegrasyonu, işten ayrılma tahminlerinin doğruluğunu artırarak işletmelerin insan kaynakları stratejilerini daha etkin bir şekilde yönlendirmelerine olanak tanıyacaktır. Yenilikçi metodoloji, işletmelerin çalışan bağlılığını artırma, işten ayrılma oranlarını azaltma ve genel iş performansını iyileştirme çabalarını destekleyecek şekilde tasarlanmıştır. Bu sayede, işgücü yönetiminde proaktif kararlar oluşturulabilir ve potansiyel ayrılma riskleri önceden tespit edilerek gerekli önlemler zamanında alınabilir. Aynı zamanda, bu çalışma akademik literatürde de önemli bir boşluğu doldurarak, derin öğrenme ve makine öğrenimi tekniklerinin çalışanların işten ayrılma tahmini gibi spesifik insan kaynakları uygulamalarına nasıl uyarlanabileceği konusunda yeni perspektifler sunmaktadır. Bu yenilikçi uygulamalar, sadece teorik açıdan değil, aynı zamanda pratik uygulamalarda da önemli katkılar sağlamaktadır. Çalışmanın sunduğu metodolojik ilerlemeler, işten ayrılma modellerinin gelecekte nasıl geliştirilebileceğine dair yol gösterici olabilir.

Çalışmanın bundan sonraki bölümlerinde, ilgili literatür incelemesi ile başlayarak, çalışmanın temelini oluşturan teorik ve pratik çalışmalar ele alınacaktır. Ardından, metodoloji bölümünde, makine öğrenimi ve derin öğrenme modellerinin entegrasyonunu detaylandırarak, veri setine ilişkin bilgiler, veri işleme süreçleri, model geliştirme ve özellik seçimi yöntemleri açıklanacaktır. Deneysel bulgular, elde edilen sonuçların analizi ile sunulacak olup, bu sonuçlar çeşitli metriklerle değerlendirilecek ve modelin performansı gözden geçirilecektir. Son olarak, sonuç ve tartışma bölümünde, elde edilen bulguların özetlenmesi, çalışmanın kısıtlılıkları, uygulamalı yönleri ve gelecek çalışmalar için öneriler ile çalışma tamamlanacaktır.

2. Literatür İncelemesi

Bu bölümde, işten ayrılma tahminleri konusunda geleneksel yöntemlerden yapay zekâ tekniklerine kadar uzanan geniş bir yelpazedeki mevcut araştırmaları ve bu çalışmaların nasıl bir çerçeve sunduğu derinlemesine incelenecektir. Özellikle, makine öğrenimi ve derin öğrenme modellerinin işten ayrılma tahminlerindeki etkinlikleri, bu alanın literatüründeki boşlukları belirlemeye ve araştırma sorularına yönelik metodolojik yaklaşımlar geliştirmeye olanak tanıyacaktır. İlgili literatür incelendiğinde, özellikle lojistik regresyon gibi geleneksel istatistiksel modeller, çalışan devrini tahmin etmek için yaygın olarak kullanılmaktadır (Zhu vd., 2016). Ancak bu modeller, özellikle büyük veri kümelerinde ve çok boyutlu yapılarla, işten ayrılma dinamiklerinin karmaşıklığını yakalamada sınırlamalara sahiptir (Wijaya vd., 2021). Son zamanlarda yapılan çalışmalar, lojistik regresyon gibi geleneksel istatistiksel yöntemlerin kapsamlı veri kümelerindeki karmaşık ilişkileri ve gizli kalıpları ortaya çıkarmada yeterli olmayabileceğini ve daha gelişmiş yaklaşımların kullanılmasını gerektirdiğini göstermiştir (Sperer, 2021). Geleneksel istatistiksel modellerin sınırlamaları nedeniyle, çalışan devrini tahmin etmek için makine öğrenimi tekniklerinin benimsenmesine doğru bir kayma olmuştur. Rastgele ormanlar ve sinir ağları da dahil olmak üzere makine öğrenimi modelleri, geleneksel istatistiksel yöntemlere kıyasla üstün performans göstermiştir (Adeusi vd., 2024). Kuruluşlar, makine öğrenimi algoritmalarını kullanarak, daha geniş bir değişken ve etkileşim yelpazesini dikkate alan daha doğru tahmin modelleri geliştirebilir ve bu da daha etkili işten ayrılma tahminlerine

yal açabilir (Lim vd., 2024). Geleneksel istatistiksel modeller, iş tatmini ve demografik özellikler gibi faktörlere dayalı işten ayrılma dinamiklerini anlamada faydalı olsa da (Xue vd., 2023), işten ayrılma davranışının incelikli ve çok yönlü doğasını yakalamada yetersiz kalabilmektedir (Feeley & Barnett, 1997). Buna karşılık, makine öğrenimi algoritmaları, daha geniş bir değişkenler dizisini ve bunların etkileşimlerini dikkate alarak işten ayrılma tahminine daha kapsamlı bir yaklaşım sunmaktadır (Wu, 2023).

Bununla birlikte, işten ayrılma durumunun tahmininde, değişkenlerin etkilerini inceleyen çalışmalara bakıldığında, literatürde birçok çalışma yer almaktadır. Çalışan devri, bir bireyin bir kuruluştan ayrılma kararını etkileyebilecek çeşitli faktörlerden etkilenmektedir. Bu faktörler, çalışma ortamının çeşitli yönlerini, iş tatminini, örgütsel bağlılığı ve dış koşulları kapsamaktadır. Çalışmalar, çalışanların işten ayrılma olasılığını şekillendirmede önemli bir rol oynayan birkaç temel belirleyicinin altını çizmiştir. Örgütsel bağlılık ve iş tatmini, işten ayrılma olasılığını etkileyen önemli faktörler olarak tanımlanmıştır (AlMohamed vd., 2022). Kurumlarına karşı güçlü bir bağlılık hisseden ve işlerinden memnun olan çalışanların işten ayrılmayı düşünme olasılığı daha düşüktür. Tersine, rekabetçi olmayan maaşlar, kariyer gelişimi için sınırlı fırsatlar ve çalışanların çabalarının takdir edilmemesi gibi faktörler daha yüksek işten ayrılma oranlarına katkıda bulunabilir (Samaşonok, 2024). Bununla birlikte, ücret oynaklığı ve ücret seviyeleri, işten ayrılma olasılığını tahmin etmek için etkileşime girerek, finansal istikrar ve tazminatın çalışanların elde tutulmasında çok önemli bir rol oynadığını göstermektedir (Conroy vd., 2021). Ayrıca, iş stresi, ücret ve iş tatmini gibi faktörler işten ayrılma niyetiyle ilişkilendirilmiştir ve işten ayrılma oranlarını azaltmak için bu hususların ele alınmasının önemini vurgulamaktadır (Ingsih vd., 2022). Öte yandan, liderlik kalitesi, kariyer ilerleme fırsatları ve kurum kültürü de işten ayrılma olasılığının kritik belirleyicileri olarak tanımlanmıştır (Stamolampros vd., 2019). Liderlik gelişimine öncelik veren, kariyer gelişimi için açık yollar sunan ve olumlu bir çalışma ortamını teşvik eden firmaların çalışanları elde tutma olasılığı daha yüksektir. Buna karşılık, değer çatışmaları, istismarcı denetim ve güçlendirme eksikliği işten ayrılma olasılığını artırabilir (Harrison & Gordon, 2014; Kim & Fernández, 2016; Tews & Stafford, 2020). Sonuç olarak, çalışanların işten ayrılma olasılığı, iş tatmini ve ücretten liderlik kalitesi ve örgüt kültürüne kadar uzanan karmaşık faktörler dizisinin etkileşiminin sonucunda ortaya çıkmaktadır. Kuruluşlar, bu belirleyicileri ele alarak ve çalışan bağlılığını, kariyer gelişimini ve çalışma ortamını iyileştirmeye yönelik stratejiler uygulayarak, işten ayrılma risklerini azaltabilir ve daha istikrarlı ve bağlı bir işgücünü teşvik edebilir.

İşten ayrılma durumunu tahmin etmek ve sınıflandırmak için makine öğrenimi tekniklerinin uygulanması son araştırmalarda büyük ilgi görmüştür. Çeşitli çalışmalar, işten ayrılma durumunu sınıflandırmak için öngörücü modeller oluşturmak üzere farklı makine öğrenimi algoritmaları kullanarak bu alanı araştırmıştır. Örneğin, Liao (2023), Destek Vektör Sınıflandırması, Karar Ağacı Sınıflandırıcısı, AdaBoost Sınıflandırıcısı, Rastgele Orman Sınıflandırıcısı, Ekstra Ağaçlar Sınıflandırıcısı, Lojistik Regresyon ve Gradyan Güçlendirme Sınıflandırıcısı gibi algoritmaları kullanarak 16 çalışan özelliğine dayalı olarak çalışan devir sınıflandırmasını araştırmıştır. Yapılan deneylerle, makine öğrenimi modellerinin etkinliği doğrulanmıştır. Kanuto (2024), çalışan devir modellerini tahmin etmek amacıyla makine öğrenimi yaklaşımlarını kullanarak önemli bulgular sunmuştur. Çalışmada, Kaggle'dan alınan 4653 geçerli katılımcı kaydı içeren bir veri seti analiz edilmiştir. Çalışma, K-En Yakın Komşu (KNN), Karar Ağacı ve Destek Vektör Makinesi (SVM) algoritmalarını kullanarak modellemiştir ve her bir modelin

doğruluk, hassasiyet, geri çağırma oranları gibi performans metrikleri değerlendirilmiştir. Karar Ağacı modeli %79 doğruluk oranı ile en iyi performansı sergilerken, KNN modeli %75.40, SVM modeli ise %61 doğruluk oranı ile izlemiştir. Çalışma, bu modellerin çalışan devir hızını tahmin etmedeki etkinliğini ve organizasyonların çalışan bağlılığını artırıcı stratejiler geliştirmek için bu bulguları nasıl kullanabileceğini tartışmaktadır. Ayrıca, Adibaji & Marleen (2022), bir çalışanın bir şirketten veya mevcut işinden ayrılma olasılığını ifade eden işten ayrılma niyetini tahmin etmek için K-En Yakın Komşu, Destek Vektör Makinesi ve Karar Ağacı yöntemlerini karşılaştırmaya odaklanmıştır. Kullanılan veri seti Kaggle'dan alınan 311 kayıtlı insan kaynakları verilerini içermekte olup, bu veriler 36 özelliğinden 24 tanesi kullanılarak analiz gerçekleştirilmiştir. Her üç yöntemin doğruluk, hassasiyet ve duyarlılık metrikleri karşılaştırılmış ve en yüksek ortalama yüzdeliğe sahip olan yöntem tahmin modeli olarak seçilmiştir. Sonuçlara göre, Karar Ağacı yöntemi %98 doğruluk, %100 hassasiyet ve %98 duyarlılık ile en iyi performansı sergileyerek diğer yöntemlerden üstün bulunmuştur. Bu bulgular, organizasyonların çalışan devir niyetini azaltmak ve uzun vadeli iş gücü sürdürülebilirliğini sağlamak için stratejiler geliştirmesine olanak tanıyabilir.

Derin öğrenme teknikleri kullanılarak çalışan devir hızı tahmini, son araştırmalarda önemli bir ilgi görmektedir. Pekel Özmen & Özcan (2022), çalışan işten ayrılma tahmini için evrimsel sinir ağlarına dayalı yeni bir derin öğrenme modeli sunarak derin öğrenmenin bu alandaki etkinliğini ortaya koymuştur. Çalışmada, evrimsel sinir ağları (CNN) tabanlı bir yaklaşım kullanarak, mevcut sınıflandırma algoritmalarının ötesine geçmeyi amaçlamıştır. Çalışmada öncelikle numerik veri seti üzerinde CNN modeli uygulanmış, daha sonra veri kaybını azaltmak için CNN algoritmasını geliştiren yeni bir hibrit genişletilmiş evrimsel karar ağacı modeli (ECDT) önerilmiştir. Son olarak, ECDT modelinin sınıflandırma doğruluğunu artırmak için ızgara arama optimizasyonu uygulanan bir model (ECDT-GRID) geliştirilmiştir. Sonuçlar, geliştirilen ECDT-GRID modelinin, CNN ve ECDT modelleri ile temel sınıflandırma algoritmalarına göre sınıflandırma doğruluğu açısından %79 doğruluk oranı ile daha üstün performans sergilediğini göstermiştir. Bu model, çalışan devir tahmininde etkili bir yöntem sağlamaktadır. Ayrıca, Srivastava & Eachempati (2021), gerçek zamanlı senaryolarda çalışan kaybı tahmini için Random Forest ve Gradient Boosting gibi topluluk makine öğrenimi tekniklerine kıyasla derin öğrenmenin üstün tahmin gücünü ortaya koymayı amaçlamıştır. Çalışmada, çalışan devir oranını azaltma ve şirket içi iş tatmini ve bağlılığı artırma yollarını belirlemek amacıyla kapsamlı bir makine öğrenimi ve çok kriterli karar verme modeli kullanılmıştır. Çalışan kaydı veri setinden yola çıkılarak, derin öğrenme ve ensemble makine öğrenimi teknikleri ile çalışan ayrılık oranlarının tahmin edilmesi hedeflenmiştir. Sonuçlar, derin sinir ağlarının (%91.6 doğruluk) random forest ve gradyan artırma algoritmalarına (%82.5 ve %85.4 doğruluk) göre daha etkili olduğunu göstermiştir. Bu bulgular, insan kaynakları yöneticilerine çalışanların iş yerinde daha uzun süre kalmasını sağlamak için stratejiler geliştirmeleri konusunda önemli iç görüler sunmaktadır. Çalışma aynı zamanda, regresyon modelleri ve çok kriterli bulanık analitik hiyerarşi süreci kullanılarak sonuçların doğrulanmasıyla da dikkat çekmiştir.

Bununla birlikte, özellik seçimi ve boyut azaltma teknikleri, çalışan devrine yönelik tahmin modellerinin performansını iyileştirmek için çok önemlidir. Bu teknikler, temel öngörücülerin belirlenmesine ve veri kümesi boyutluluğunun azaltılmasına yardımcı olarak işten ayrılma tahmin modellerinin doğruluğunu ve verimliliğini artırır. Pourkhodabakhsh vd. (2022), çalışan devrini etkileyen faktörleri belirlemek ve bilinçli kararlar almak için Gri Kurt Optimizer

ve Genetik Algoritma gibi meta sezgisel algoritmaların yanı sıra Özyinelemeli Özellik Eleme algoritması ve Karşılıklı Bilgi gibi özellik seçim yöntemlerinin önemini vurgulamıştır. Ayrıca Gao vd. (2019), özelliklerin önemini sıralamak ve boyutları azaltmak için rastgele orman algoritmasını kullanmış ve çalışan devrine yönelik tahmin sürecini kolaylaştırmada özellik seçiminin etkinliğini göstermiştir. Al Akasheh vd. (2024), çalışan devrinin tahminini geliştirmek için açıklanabilir yapay zekâ teknikleriyle birlikte özellik seçiminin önemini vurgulamıştır. Araştırmacılar, bu gelişmiş modellerde özellik seçimi yöntemlerinden yararlanarak, işten ayrılmaya neden olan en etkili faktörleri belirleyebilir ve daha sağlam tahmin çerçeveleri oluşturabilirler.

Öte yandan, çeşitli veri kümelerinden elde edilen görüntülerden özellik çıkarımı için derin öğrenme modellerinin kullanılması, farklı alanlardaki tahmin performansını artırmada önemli bir umut vaat etmektedir. Bunun gibi çalışmalar, özellik çıkarma ve sınıflandırma süreçlerini entegre ederek görüntü sınıflandırma görevlerinde derin öğrenme modellerinin etkinliğini göstermiş ve sınıflandırmada gelişmiş doğruluğa yol açmıştır (Liu & An, 2020). Benzer şekilde, Zhang (2024) tarafından yapılan araştırma, tek boyutlu hata örneklerini üç boyutlu özellikli görüntü örneklerine dönüştürmek için GAF dönüşüm yöntemini uygulamış, ardından derin öğrenme modeli girdisi için veri kümesini genişletmek üzere GAN tabanlı yöntemler kullanılarak veri iyileştirmesi yapmış ve görüntü dönüşümünün model performansı üzerindeki etkisini göstermiştir. Ayrıca, Park vd. (2021), derin öğrenme tabanlı bir izinsiz giriş tespit modeli geliştirmek için vektör verilerini görüntülere dönüştürme üzerine araştırma yapmış ve anomali tespiti için görüntü tabanlı veri temsiline önemini vurgulamıştır (Park vd., 2021). Bu çalışmalar, derin öğrenme modelleri kullanılarak daha sonra özellik çıkarımı için veri kümelerinin görüntülere dönüştürülmesinin dönüştürücü etkisinin altını toplu olarak çizmektedir. Genel olarak, derin öğrenme modelleri kullanılarak daha sonra özellik çıkarımı için veri kümelerinin görüntülere dönüştürülmesi, çeşitli alanlarda daha doğru sınıflandırma, teşhis ve analiz sağlayarak tahmin performansını artırma konusunda önemli bir umut vaat etmiştir. Araştırmacılar ve uygulayıcılar, görüntü verilerinden özellik çıkarma konusunda derin öğrenmenin gücünden yararlanarak değerli iç görülerin kilidini açabilir ve çeşitli alanlarda ilerlemeler sağlayabilir.

Bu çalışma, literatürde belirgin boşlukları doldurma potansiyeline sahiptir. Özellikle, makine öğrenimi ve derin öğrenme tekniklerinin işten ayrılma tahminlerindeki etkinliği üzerine yapılan çalışmalar, genellikle belirli algoritmaların performansına odaklanmaktadır. Ancak, bu çalışmalar sıklıkla büyük veri kümelerinde ve çok boyutlu yapılarda karmaşıklıkları yeterince modelleyememekte veya çeşitli veri türlerinden elde edilen öngörülerin birleştirilmesi konusunda eksik kalmaktadır. Bu çalışmada, veri setindeki geleneksel veri yapılarını 2D barkod görüntülerine dönüştürmek gibi yenilikçi bir yaklaşım kullanılmaktadır. Bu dönüşüm, derin öğrenme modellerinin görüntü işleme kabiliyetlerinden yararlanarak daha karmaşık özellik çıkarımları yapılmasına olanak tanıyacaktır. Öte yandan, derin öğrenme modellerinden elde edilen özellikler üzerine yapılan kapsamlı bir özellik seçimi ve optimizasyon süreci, işten ayrılma tahminlerini daha da rafine edebilir. Bu durum, bu çalışmada kullanılan veri setinde olduğu gibi özellikle büyük veri kümelerinde önemli bir avantaj sağlayacaktır. Bu bilgiler doğrultusunda aşağıda gösterilen araştırma soruları ve hipotezler oluşturulmuştur.

Araştırma Soruları:

1. Farklı makine öğrenimi yöntemleri, işten ayrılma tahminlerinin doğruluğunu nasıl etkilemektedir?
2. Veri setindeki geleneksel veri yapılarını 2D barkod görüntülerine dönüştürmek, derin öğrenme modellerinin işten ayrılma tahminlerindeki performansını nasıl etkilemektedir?
3. Derin öğrenme modelinden çıkarılan özelliklerin tekrardan makine öğrenmesi modellerine girdi olarak verilerek gerçekleştirilen sınıflandırma işlemi, işten ayrılma tahminlerini ne ölçüde iyileştirmektedir?

Hipotezler:

H1: Farklı makine öğrenimi yöntemlerinin kullanımı, işten ayrılma tahminlerinin doğruluğunu farklı şekillerde etkileyecektir; bazı yöntemler diğerlerine göre daha yüksek doğruluk sağlayacaktır.

H2: Geleneksel veri yapılarının 2D barkod görüntülerine dönüştürülmesi, derin öğrenme modellerinin özellik çıkarım kabiliyetini artırarak tahmin performansını önemli ölçüde iyileştirir.

H3: Derin öğrenme modelinden çıkarılan ve makine öğrenmesi modellerine girdi olarak verilen özellikler, işten ayrılma tahminlerinde doğruluğu önemli ölçüde artırır.

3. Metodoloji

Bu çalışmada kullanılan metodoloji, çalışanların işten ayrılma durumlarını sınıflandırma performansını iyileştirmek için farklı makine öğrenimi ve derin öğrenme modellerinin etkin bir şekilde nasıl entegre edilebileceğini göstermeyi amaçlamaktadır. Araştırmada, çeşitli veri ön işleme teknikleri, model eğitim süreçleri, özellik çıkarımı yöntemleri stratejileri kullanılmıştır. Özellikle, derin öğrenme modelinden elde edilen özelliklerin, makine öğrenimi modellerine nasıl entegre edildiği ve bu entegrasyonun sınıflandırma performansına etkileri ayrıntılı bir şekilde incelenmiştir. Aşağıdaki alt başlıklar altında, metodoloji kapsamında ele alınan ana bileşenler ve uygulanan süreçler detaylandırılmıştır.

3.1. Veri Seti

Bu çalışmada kullanılan veri seti, Zia (2024) tarafından paylaşılan “Çalışan Kaybı Veri Seti” olarak bilinen ve Kaggle platformu üzerinden erişilebilen bir veri kaynağından elde edilmiştir. Bu veri seti, çeşitli endüstrilerde çalışan bireylerin demografik bilgileri, iş tatmini, maaş düzeyleri, çalışma saatleri, işe bağlılık düzeyleri vd. gibi birçok değişken içermektedir. Çalışmamızda kullanılan veri seti, çalışanların işten ayrılma eğilimlerini sınıflandırmak için geniş bir veri yelpazesi sunmaktadır. Veri setinde kullanılan değişkenler, bunların açıklamaları ve veri türleri detaylı bir şekilde Tablo 1’de sunulmuştur.

Tablo 1: Çalışmada Kullanılan Veri Setine İlişkin Detaylar

Değişken Adı	Açıklama	Veri Türü
Çalışan Kimliği	Her çalışana atanan benzersiz bir tanımlayıcı.	int64
Yaş	Çalışanın yaşı 18 ila 60 arasında değişmektedir.	int64
Cinsiyet	Çalışanın cinsiyeti	object
Şirkette Geçirilen Yıllar	Çalışanın şirkette çalıştığı yıl sayısı.	int64
İş Rolü	Çalışanın çalıştığı departman veya rol; Finans, Sağlık, Teknoloji, Eğitim ve Medya gibi kategorilerde kodlanmıştır.	object
Aylık Gelir	Çalışanın dolar cinsinden aylık maaşı.	int64
İş-yaşam dengesi	Çalışanın iş ve özel hayatı arasında algıladığı denge, (Zayıf, Ortalamanın Altında, İyi, Mükemmel)	object
İş Tatmini	Çalışanın işinden memnuniyeti: (Çok Düşük, Düşük, Orta, Yüksek)	object
Performans Derecelendirmesi	Çalışanın performans derecesi: (Düşük, Ortalamanın Altında, Ortalama, Yüksek)	object
Promosyon Sayısı	Çalışanın aldığı toplam terfi sayısı.	int64
Eve Olan Mesafe	Çalışanın evi ile işyeri arasındaki mesafe, mil cinsinden.	int64
Eğitim Seviyesi	Çalışanın ulaştığı en yüksek eğitim seviyesi: (Lise, Ön Lisans, Lisans, Yüksek Lisans, Doktora)	object
Medeni Durum	Çalışanın medeni durumu: (Boşanmış, Evli, Bekâr)	object
İş Seviyesi	Çalışanın iş seviyesi: (Giriş, Orta, Kıdemli)	object
Şirket Büyüklüğü	Çalışanın çalıştığı şirketin büyüklüğü: (Küçük, Orta, Büyük)	object
Şirket Görev Süresi	Şirkette geçirilen toplam süre (ay)	int64
Uzaktan Çalışma	Çalışanın uzaktan çalışıp çalışmadığı: (Evet veya Hayır)	object
Liderlik Fırsatları	Çalışanın liderlik fırsatlarına sahip olup olmadığı: (Evet veya Hayır)	object
İnovasyon Fırsatları	Çalışanın inovasyon için fırsatlara sahip olup olmadığı: (Evet veya Hayır)	object
Şirket İtibarı	Çalışanların şirketin itibarına ilişkin algısı: (Çok Kötü, Kötü, İyi, Mükemmel)	object
Çalışan Takdiri	Çalışanın aldığı takdir düzeyi: (Çok Düşük, Düşük, Orta, Yüksek)	object
Çalışılan Şirket Sayısı	Kişinin çalıştığı şirket sayısı	int64
Bakmakla Yükümlü Olunan Kişi Sayısı	Çalışanın bakmakla yükümlü olduğu kişi sayısı	object
İşten Ayrılma Durumu (Hedef Değişken)	Çalışanın şirketten ayrılıp ayrılmadığı, 0 (kaldı) ve 1 (ayrıldı) olarak kodlanmıştır.	object

Kaynak: Zia (2024), Çalışan Yıpranma Sınıflandırması Veri Seti, Yıpranma Analizi ve Tahmini için Derinlemesine Bir Sentetik Simülasyon.

Veri setinde “İşten Ayrılma Durumu” sütunu için iki sınıfın dağılımı şu şekildedir. “Kaldı” (0 sınıfı): 39,128 girişten, “Ayrıldı” (1 sınıfı): 35,370 veri girişinden olmak üzere toplamda 74,498 veri girişi bulunmaktadır. Bu, veri setindeki sınıflar arasında nispeten dengeli bir dağılım olduğunu gösterir, bu da model eğitimi için uygun bir durum oluşturmaktadır.

3.2. 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 etkileyebilir. Başlangıçta, modellerin eğitilmesi sırasında gereksiz olan “Çalışan Kimliği” sütunu, veri setinden çıkarılmıştır. Bu sütun, modelin tahmin kabiliyetine katkıda bulunmayan bir tanımlayıcıdır. Daha sonra, İşten ayrılma durumunu gösteren sütun, hedef değişken olarak ayrılmış ve ana veri setinden çıkarılmıştır. Hedef değişken, “Ayrıldı” değeri için 1 ve “Kaldı” değeri için 0 olacak şekilde ikili formata dönüştürülmüştür.

Bununla birlikte, modelleme sürecinde kullanılacak olan kategorik değişkenlerin, makine öğrenimi algoritmaları tarafından işlenebilir hale getirilmesi gerekmektedir. Kategorik değişkenler için yaygın olarak kullanılan yöntemler arasında one-hot encoding (Hien vd., 2020) ve label encoding (Sihare, 2024) bulunmaktadır. Bu çalışmada one-hot encoding yöntemi ile dönüşüm işlemi uygulanmıştır.

Ardından, dönüştürülen veri seti, modelin eğitimi ve testi için eğitim (%80) ve test (%20) setlerine ayrılmıştır. Test setindeki veri dağılımı, 0 sınıfı için 7804, 1 sınıf için 7096 olmak üzere toplamda 14900 veri girişinden oluşmaktadır. Ayrıca, rastgele dağıtım için bir rastgelelik durumu (seed) belirlenmiştir. Son olarak, sayısal özellikler, Standard Scaler kullanılarak ölçeklendirilmiştir. Bu işlem, özellikler arasındaki ölçek farklılıklarını gidererek modelin daha etkili öğrenmesine yardımcı olur (Zelner & Fergus, 2014). Matematiksel olarak bu işlem Denklem 1’de gösterildiği gibi ifade edilmektedir:

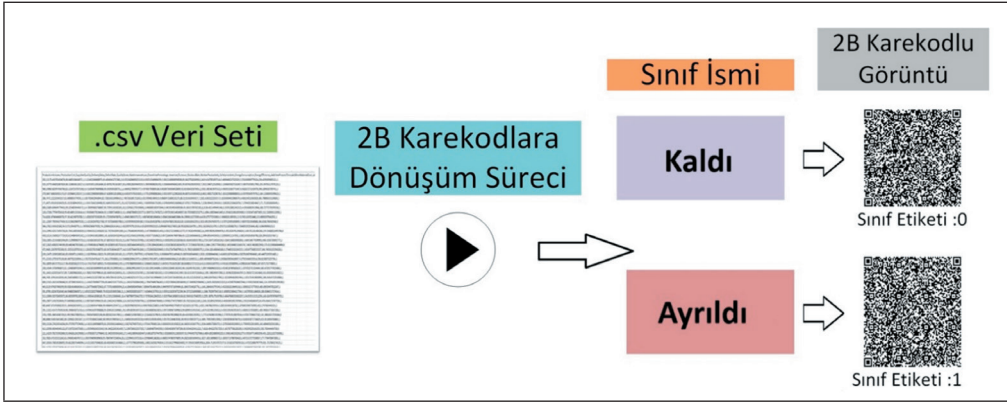
$$Z = \frac{(x - \mu)}{\sigma} \quad (1)$$

Burada, x ölçeklendirilecek olan özellik değerini, μ bu özelliğin ortalamasını, σ bu özelliğin standart sapmasını ve z ölçeklendirilmiş sonucu temsil etmektedir. Bununla birlikte, eğitim seti üzerinde fit ve transform işlemi uygulanırken, test setine sadece transform işlemi uygulanmıştır. Bu adımların tamamı, Jupyter Notebook ortamında Python v11 programlama dili kullanılarak gerçekleştirilmiştir. Her bir işlem, veri bilimi projelerinde yaygın olarak kullanılan pandas, numpy ve scikit-learn kütüphaneleri aracılığıyla uygulanmıştır. Bu süreç, modellerin daha doğru tahminler yapabilmesi için veri setinin temizlenmesi, dönüştürülmesi ve hazırlanması aşamalarını kapsamaktadır.

Çalışmada, CNN bazlı derin öğrenme modelleri ile sınıflandırma işlemlerini gerçekleştirmek için her bir veri girişi, Kare kodlar (QR Kod) oluşturularak görüntü formatında kaydedilmiştir. Bu görüntüler, derin öğrenme modelleri tarafından işlenecek şekilde düzenlenmiştir. Oluşturulan QR kodlu görüntüler yüklenmiş ve model girdisi olarak kullanılmak üzere uygun boyutlara ölçeklendirilmiştir. Hedef değişkenler, derin öğrenme modelleri için uygun olacak şekilde kategorik formatlara dönüştürülmüştür. Bu süreçler, veri setinin derin öğrenme model-

lerine uygun hale getirilmesi için kritik öneme sahiptir. QR kodlarının oluşturulması ve işlenmesi için kullanılan Python kütüphaneleri, Jupyter Notebook ortamında etkin bir şekilde entegre edilmiştir. Verileri QR kod formatına dönüştürmeyi ve bu kodları resim olarak kaydetmeyi sağlamak için “qrcode” kütüphanesi kullanılmıştır. Görüntü işleme kütüphanesi olan “Pillow”, QR kodlarını görüntü formatında işlemek ve düzenlemek için kullanılmıştır. Görüntü verileri üzerinde matris işlemleri gerçekleştirmek için “numpy” kütüphanesi tercih edilmiştir. Ayrıca “os” kütüphanesi, dizin yönetimi ve dosya işlemleri için kullanılmıştır. Ayrıca, “keras.utils” kütüphanesinin “to_categorical” fonksiyonu, hedef değişkenleri derin öğrenme modelleri için uygun kategorik formata dönüştürmek amacıyla kullanılmıştır. Bu kütüphaneler, veri setinin derin öğrenme modelleri tarafından işlenebilir hale getirilmesinde kritik roller oynamaktadır. QR Kod dönüşüm işlemi Şekil 1’de gösterilmiştir.

Şekil 1: QR Kod Dönüşüm Süreci



3.3. Makine Öğrenmesi ve Derin Öğrenme Modelleri

Bu bölümde sınıflandırma işlemi için kullanılan makine öğrenmesi ve derin öğrenme modelleri matematiksel formülleri ile birlikte açıklanmaktadır.

3.3.1. Karar Ağaçları (Decision Trees)

Karar ağaçları, veri setinden en uygun özellikleri ve eşik değerleri seçerek sınıflandırma ya da regresyon için kullanılır. İlk adım, veri setini iyi bir şekilde ayırabilecek en iyi özelliğin seçilmesidir. Bu seçim sırasında, her özelliğin veri setini nasıl böleceğine dair homojenlik veya saflık artışı gibi metrikler hesaplanır; bu metrikler arasında Gini impurity, entropi veya kazanç oranı bulunur (Ağlin vd., 2020). Seçilen özellik temelinde, veriler belirlenen bir eşik değere göre ikiye ayrılır ve bu süreç, her bir alt kümenin homojenliği yeterli düzeye ulaşana veya önceden belirlenen maksimum derinliğe erişene kadar devam eder. Bu işlem sonunda elde edilen homojen alt kümeler yaprak düğüm olarak adlandırılır ve her yaprak düğüm bir sınıf etiketi veya bir regresyon sonucu içerir. Bu yöntemle, karar ağaçları verileri etkili bir şekilde modelleyerek çeşitli tahmin ve sınıflandırma işlemleri gerçekleştirir (Hussain, 2014). Karar ağaçları için düğüm bölünmelerini belirlemede kullanılan önemli metriklerden biri olan Gini impurity'nin hesaplanması Denklem 2'de gösterildiği gibi yapılmaktadır:

$$Gini(p) = 1 - \sum_{i=1}^J p_i^2 \quad (2)$$

Burada p_i , i 'nci sınıfa ait örneklerin oranını ifade eder ve J , sınıf sayısını belirtir. Bu formül, bir düğümdeki karışıklığı (impurity) ölçer ve ideal olarak, bir düğümdeki tüm örneklerin aynı sınıftan olması durumunda sıfır değerini alır. Karar ağaçlarının amacı, Gini impurity'yi minimize edecek şekilde veri setini bölmektir.

3.3.2. Gradyan Artırma (Gradient Boosting)

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 genel bir modelin doğruluğunu artırır. Her yeni model, önceki modellerin hatalarını düzeltmeye odaklanır, böylece her adımda modelin performansı artar (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 & Knoll, 2013).

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

$$F_{m+1}(x) = F_m(x) + \gamma_m h_m(x) \quad (3)$$

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:

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_m(x_i) + \gamma h_m(x_i)) \quad (4)$$

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

3.3.3. K-En Yakın Komşu (KNN)

K-En Yakın Komşu (KNN), gözetimli öğrenme kategorisinde yer alan basit ve etkili bir sınıflandırma algoritmasıdır. Veri noktaları arasındaki benzerlikleri temel alarak çalışır ve bir noktanın sınıfını, en yakın komşularının sınıflarına göre belirlemektedir. KNN, modeli oluştururken "K" parametresine ihtiyaç duymaktadır. Bu parametre, bir noktanın sınıfını belirlerken dikkate alınacak en yakın komşu sayısını ifade etmektedir (Eldora vd., 2024). Her bir

sınıflandırma işlemi, bir test noktası ile mevcut veri noktaları arasındaki uzaklığı hesaplamayı gerektirir. Bu hesaplamada genellikle Öklid uzaklığı kullanılır, ancak Manhattan, Minkowski ve Hamming gibi diğer uzaklık ölçümleri de tercih edilebilir (Kaharuddin & Sholeha, 2021). Öklid Uzaklığı formülü Denklem 5'te gösterilmektedir.

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \quad (5)$$

Burada, p ve q iki noktayı, n ise özelliklerin (features) sayısını temsil etmektedir. Hesaplanan uzaklıklar kullanılarak, test noktasına en yakın “K” komşu belirlenir. Seçilen “K” komşunun sınıfları incelenir ve hangi sınıfın daha fazla temsil edildiği belirlenir. Test noktası, en çok rastlanan sınıfa atanır. Bu süreç, tüm test seti üzerinde tekrarlanır ve her bir test noktası için bir sınıf tahmini yapılır.

3.3.4. Hafif Gradyan Artırma Makinesi (LightGBM)

LightGBM, Microsoft tarafından geliştirilen, verimli ve etkili bir gradyan artırma çerçevesidir. Büyük veri kümeleri ve çok boyutlu veri üzerinde çalışabilen bu çerçeve, özellikle sınıflandırma ve regresyon görevlerinde kullanılır. Gelişmiş algoritma yapıları sayesinde hızlı ve hafif bir işlem sunar. LightGBM, sürekli özellik değerlerini sabit sayıda kutuya (bin) ayırarak histogram tabanlı bir algoritma kullanır. Bu yöntem, veri üzerinden geçiş sayısını ve hafıza kullanımını azaltarak hesaplama süresini düşürür. 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 işlemlerini azaltır (Nagassou vd., 2023).

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.

3.3.5. Lojistik Regresyon (Logistic Regression)

Lojistik Regresyon (Logistic Regression), iki veya daha fazla kategorik sonucu tahmin etmek için kullanılan istatistiksel bir sınıflandırma modelidir. Genellikle ikili sınıflandırma problemlerinde, yani sonuçların sadece iki olası kategoriden birine ait olabileceği durumlarda tercih edilir (Chivukula vd., 2021). Bu model, bağımsız değişkenlerin (özelliklerin) sonucu nasıl etkilediğini anlamak ve sonuçların olasılığını tahmin etmek için kullanılmaktadır. Model, bağımsız değişkenler ile bağımlı değişken (sınıflar) arasındaki ilişkiyi kurarak başlar. Bağımlı değişken genellikle 0 ve 1 olmak üzere iki sınıfa ayrılır. Lojistik regresyon, bağımsız değişkenlerin her bir kombinasyonu için bağımlı değişkenin 1 olma olasılığını (yani bir sınıfa ait

olma olasılığını tahmin eder. Model, bir logit dönüşüm kullanarak bağımsız değişkenlerin lineer kombinasyonunu sınıflandırma olasılığına dönüştürür (Zhang, 2024). Logit fonksiyonu Denklem 7’de gösterildiği gibidir:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) \quad (7)$$

Burada p bağımlı değişkenin 1 olma olasılığıdır.

3.3.6. Rastgele Orman (Random Forest)

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 & 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. Her bir karar ağacı bağımsız bir tahmin yapar ve sınıflandırma görevlerinde bu tahminlerin çoğunluğunun oyu (majority voting) veya olasılıkların ortalaması alınarak nihai tahmin elde edilir (Özen ve Bal, 2019). Sınıflandırma için tahmin yaparken genellikle Denklem 8’de gösterildiği gibi bir ifade kullanılır:

$$y = \text{mod}\{y_1, y_2, \dots, y_n\} \quad (8)$$

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

3.3.7. AlexNet

AlexNet, derin öğrenme ve bilgisayarlı görme alanında çığır açan bir sinir ağı modelidir. 2012 yılında Alex Krizhevsky, Ilya Sutskever ve Geoffrey Hinton tarafından geliştirilmiştir ve o yıl ImageNet yarışmasını kazanarak büyük ilgi toplamıştır. Bu model, büyük ve derin bir evrişimli sinir ağıdır (CNN - Convolutional Neural Network) ve görüntü tanıma görevlerinde kullanılmak üzere tasarlanmıştır (Cheng vd., 2019). AlexNet, toplamda beş evrişimli katman (convolutional layer), üç tam bağlantılı katman (fully connected layer) ve çıktı katmanından oluşur. Bu katmanlar arasında, bazı katmanlarda max pooling ve dropout işlemleri bulunur (Liu vd., 2020). Evrişimli katmanlar, girdi olarak aldığı görüntülerdeki özellikleri (features) çıkarmak için kullanılır. Her katman, farklı özellikleri (kenarlar, dokular vb.) belirlemek üzere çeşitli filtreler (kernels) uygular. AlexNet, ReLU (Rectified Linear Unit) aktivasyon fonksiyonunu kullanır. Bu fonksiyon, modelin eğitim sürecini hızlandırmak ve gradyan kaybolması problemini azaltmak için tercih edilir (Liu vd., 2020). ReLU formülü Denklem 9’daki gibidir.

$$f(x) = \max(0, x) \quad (9)$$

Max pooling işlemi, evrişimli katmanlar tarafından üretilen özellik haritalarının (feature maps) boyutunu azaltmak ve en belirgin özellikleri korumak için kullanılır. Tam bağlantılı

katmanlar, evrişimli katmanlar tarafından çıkarılan özellikleri kullanarak nihai sınıflandırmayı gerçekleştirir. Çıktı katmanı, softmax aktivasyon fonksiyonu kullanarak her bir sınıf için bir olasılık üretir (Pal vd., 2021). Softmax aktivasyon fonksiyonu Denklem 10'da gösterilmektedir.

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (10)$$

Burada z katmanın çıktılarıdır ve i indeksi sınıfı belirtir.

Dropout, aşırı uyum (overfitting) problemini önlemek için kullanılan bir yöntemdir. Eğitim sırasında rastgele seçilen nöronlar geçici olarak “devre dışı bırakılır”. Modelin eğitimi sırasında stokastik gradyan inişi (SGD) ve geri yayılım kullanılır. Kayıp fonksiyonu olarak çapraz entropi (cross-entropy) tercih edilir (Pal vd., 2021).

3.3.8. MobileNet

MobileNet, özellikle mobil ve gömülü cihazlar için tasarlanmış hafif bir derin öğrenme modelidir. Howard vd. (2017) tarafından geliştirilen bu model, yüksek verimlilik ile nispeten düşük hesaplama gücü gerektiren uygulamalarda kullanılmak üzere optimize edilmiştir. MobileNet, görüntü sınıflandırma, nesne algılama ve yüz tanıma gibi görevlerde kullanılabilir ve büyük, karmaşık modellere kıyasla daha az kaynak tüketir. MobileNet, standart evrişim işlemlerini derinlik sezgisel evrişimlere böler, bu sayede modelin boyutunu ve hesaplama maliyetini büyük ölçüde azaltır. Bu evrişimler, önce her bir giriş kanalı için ayrı ayrı evrişim (depthwise convolution) uygular, ardından 1×1 boyutunda evrişimlerle (pointwise convolution) kanallar arası bilgileri birleştirir (Jaderberg vd., 2014). MobileNet, ReLU6 adı verilen, sıfırdan küçük değerler için sıfır, altıdan büyük değerler için altı değerini alan bir ReLU türevi kullanır. Bu türev Denklem 11'de gösterilmektedir.

$$f(x) = \min(\max(0, x), 6) \quad (11)$$

Her evrişim katmanından sonra, model batch normalization işlemi uygular. Bu, modelin eğitim sürecini hızlandırır ve daha stabil hale getirir. Modelin sonunda, elde edilen özellikler tam bağlantılı katmanlar aracılığıyla işlenir ve çıktı olarak her sınıf için bir olasılık üretilir. Sınıflandırma için son katman, softmax aktivasyon fonksiyonunu kullanarak, her sınıfa ait olasılıkları hesaplar. Softmax aktivasyon fonksiyonu bir önceki bölümde Denklem 10'da gösterilmiştir.

3.3.9. ResNet-18

ResNet-18, derin sinir ağları içerisinde yer alan ve “Kalan Ağ” (Residual Network) olarak da bilinen bir modeldir. ResNet, derin öğrenme modellerinde karşılaşılan “kaybolan gradyan” problemine çözüm olarak ortaya çıkmıştır. ResNet-18, ResNet ailesinin 18 katmanlı bir versiyonudur ve genellikle görüntü sınıflandırma gibi görevlerde kullanılır (He vd., 2016).

ResNet-18, her bir katman çıkışına girişini ekleyerek çalışır. Bu, derin ağlarda sıkça karşılaşılan kaybolan gradyan problemine çözüm getirir. Her birim için, ana sinyal doğrudan sonraki katmanlara aktarılırken, öğrenilen ağırlıklar bu ana sinyale eklenir. Model, 18 katman boyunca temel bloklardan oluşur. Her blok, birkaç evrişimli katmandan ve ardından bir ka-

lan bağlantıdan oluşur. Bu bloklar, modelin daha derin öğrenme yeteneğini artırırken, eğitimi stabil hale getirir (Sarwinda vd., 2021). Modelde 3x3 boyutunda evrişimler kullanılır ve bu evrişimler, girdi ve çıktıdaki kanal sayısını kontrol etmek için ayarlanabilir. Her evrişimli katmandan sonra, ReLU (Rectified Linear Unit) aktivasyon fonksiyonu uygulanır. ReLU, negatif değerleri sıfıra eşitler ve pozitif değerleri değişmeden bırakır. Modelin sonunda, özellik haritalarının (feature maps) boyutunu küçülten bir küresel ortalama havuzlama katmanı bulunur. Bu, son tam bağlantılı katmana giriş olarak hizmet eder. Çıktı katmanı, sınıf sayısına göre ayarlanan bir softmax katmanıdır. Softmax, her sınıfa ait olasılığı hesaplar ve bu değerleri normalize eder (Guo & Du, 2019). ResNet-18 modelinin temel matematiği kalan bağlantılar üzerinden Denklem 12’de gösterildiği gibi açıklanabilir:

$$\text{output} = \text{ReLU}(F(x, \{W_i\}) + x) \quad (12)$$

Burada x bir katmanın girdisidir. $F(x, \{W_i\})$ girdi x üzerine uygulanan ağırlıklar ve işlemler sonucu elde edilen dönüşüm fonksiyonudur ve W_i ağırlıkları temsil eder. Bu formülde, x doğrudan sonraki katmana aktarılır (kalan bağlantı) ve $F(x, \{W_i\})$ ile elde edilen sonuçlara eklenir. Bu yapı, eğitim sırasında gradyanların daha derin katmanlara etkili bir şekilde ulaşmasını sağlar.

3.3.10. SqueezeNet

SqueezeNet, gömülü sistemler ve mobil cihazlar için optimize edilmiş, düşük parametrelili ama yüksek doğrulukta çalışan bir evrişimli sinir ağı (CNN) modelidir. SqueezeNet, AlexNet benzeri performans sunarken model boyutunu ve parametre sayısını önemli ölçüde azaltmayı başarır. Bu durum, hem depolama alanından tasarruf sağlar hem de modelin daha hızlı çalışmasını mümkün kılmaktadır (Siraj vd., 2024). SqueezeNet’in temel yapı taşları Fire modülleridir. Her Fire modülü, “squeeze” evrişim katmanı (1x1 evrişimler) ve “expand” katmanı (1x1 ve 3x3 evrişimlerin bir kombinasyonu) içerir. Squeeze katmanı, kanal sayısını azaltarak parametre sayısını düşürür, expand katmanı ise çeşitlilik sağlayarak daha zengin özellik haritaları oluşturur (Hassanpour ve Malek, 2020). Squeeze ve Expand katmanı sırasıyla Denklem 13 ve 14’te gösterilmektedir.

$$S(x) = f(W_s * x + b_s) \quad (13)$$

Burada W_s ve b_s , squeeze katmanının ağırlıkları ve biasları, f ise ReLU gibi bir aktivasyon fonksiyonudur.

$$E(x) = f(W_e * x + b_e) \quad (14)$$

Burada W_e ve b_e , expand katmanının ağırlıkları ve biaslarıdır.

Modelin son katmanında, her bir sınıf için bir çıktı üreten küresel ortalama havuzlama kullanılır. Bu, son evrişim katmanından elde edilen her özellik haritası için ortalama bir değer hesaplar ve bu değerler doğrudan son sınıflandırma katmanına beslenir (Siraj vd., 2024). Küresel ortalama havuzlama Denklem 15’te gösterilmektedir.

$$y_i = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W x_{i,h,w} \quad (15)$$

Burada $x_{i,h,w}$ i 'nci özellik haritasının (h,w) pozisyonundaki değeri, H ve W ise özellik haritasının yüksekliği ve genişliğidir.

3.4. Modellerin Hiperparametre Ayarları

Hiperparametre ayarlamaları, model performansının optimize edilmesinde kritik bir rol oynamaktadır. Bu hiperparametrelerin ayarlanması, modelin eğitim verilerine nasıl uyum sağladığını ve test verilerinde nasıl performans gösterdiğini önemli ölçüde etkiler. Makine öğrenmesi modelleri için hiperparametre optimizasyonu sürecinde, geniş bir parametre aralığını hızlı bir şekilde keşfetmek amacıyla Rastgele Arama (Random Search) yöntemi kullanıldı. Bu yöntem, parametrelerin geniş bir aralığında rastgele örnekler seçilerek modelin performansını değerlendirmek için kullanılır (Hasan vd., 2024). Veri setinin büyüklüğü ve çeşitliliği göz önünde bulundurularak, her bir model için uygun hiperparametrelerin belirlenmesi süreci detaylı bir şekilde planlanmıştır. Aşağıda Tablo 2'de makine öğrenmesi modellerinin hiperparametre ayarları gösterilmektedir.

Tablo 2: Makine Öğrenmesi Modellerinin Hiperparametre Ayarları

Model	Hiperparametre	Seçilen Ayar
Karar Ağaçları	max_depth	20
	min_samples_split	4
	min_samples_leaf	2
	Criterion	Gini
Gradyan Artırma	n_estimators	200
	Learning_rate	0,005
	max_depth	5
	Subsample	0,9
K-En Yakın Komşu	n_neighbors	15
	Weights	Distance
	Algorithm	Auto
Hafif Gradyan Artırma	num_leaves	50
	max_depth	20
	Learning_rate	0,05
	n_estimators	300
Lojistik Regresyon	C	0.1
	Solver	Liblinear
	Penalty	12
Rastgele Orman	n_estimators	300
	max_features	Auto
	max_depth	25
	min_samples_split	5

Bununla birlikte, derin öğrenme modelleri için hiperparametre optimizasyonunda İzgara Arama (Grid Search) metodunu kullanarak, model performansını en üst düzeye çıkarmak amacıyla geniş bir hiperparametre yelpazesini sistematik bir şekilde denenmektedir. İzgara Arama, belirli hiperparametre kombinasyonlarının her birini kapsamlı bir şekilde test eder, böylece en iyi performans gösteren kombinasyonları belirlemeye olanak tanımaktadır (Ogunsanya vd., 2023). Aşağıda Tablo 3'te, bu çalışmada kullanılan AlexNet, MobileNet, ResNet-18 ve SqueezeNet modelleri için belirlenen hiperparametre ayarları yer almaktadır.

Tablo 3: Derin Öğrenme Modellerinin Hiperparametre Ayarları

Model	Hiperparametre	Seçilen Ayar
AlexNet	Learning_rate	0,01
	Epochs	30
	Batch_size	128
	Activation	Softmax
	Optimizer	Adam
	Dropout_rate	0,5
MobileNet	Alpha	1,0
	Learning_rate	0,01
	Epochs	30
	Depth_multipler	1,0
	Batch_size	128
	Activation	Softmax
	Optimizer	Adam
Dropout_rate	0,3	
ResNet-18	Learning_rate	0,01
	Epochs	50
	Batch_size	64
	Activation	Softmax
	Optimizer	Adam
Dropout_rate	0,4	
SequuzeNet	Learning_rate	0,01
	Epochs	30
	Batch_size	128
	Activation	Softmax
	Optimizer	SGD
Dropout_rate	0,3	

3.5. Model Performans Değerlendirme Kriterleri

Modellerin performansını objektif bir şekilde değerlendirmek için çeşitli metrikler ve yöntemler kullanılmaktadır. Bu çalışmada, makine öğrenmesi ve derin öğrenme modellerinin performansları, Kesinlik (Precision), Duyarlılık (Recall), F1 Skoru ve Doğruluk (Accuracy) metrikleri ile ölçülmüştür. Bu metriklerin, matematiksel formülleri sırasıyla Denklem 16, 17, 18 ve 19'da gösterilmektedir. Denklemlerde, DP doğru pozitif, YP yanlış pozitif, DN doğru negatif ve YN yanlış negatiftir. Ayrıca, model tahminlerinin doğruluğunu görsel olarak değerlendirmek için karmaşıklık matrisleri de kullanılmıştır.

$$\text{Kesinlik (Precision)} = \frac{DP}{DP+YP} \quad (16)$$

$$\text{Duyarlılık (Recall)} = \frac{DP}{DP+YN} \quad (17)$$

$$\text{F1 Skoru} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (18)$$

$$\text{Doğruluk (Accuracy)} = \frac{DP+DN}{DP+DN+YP+YN} \quad (19)$$

3.6. Yazılım ve Donanım

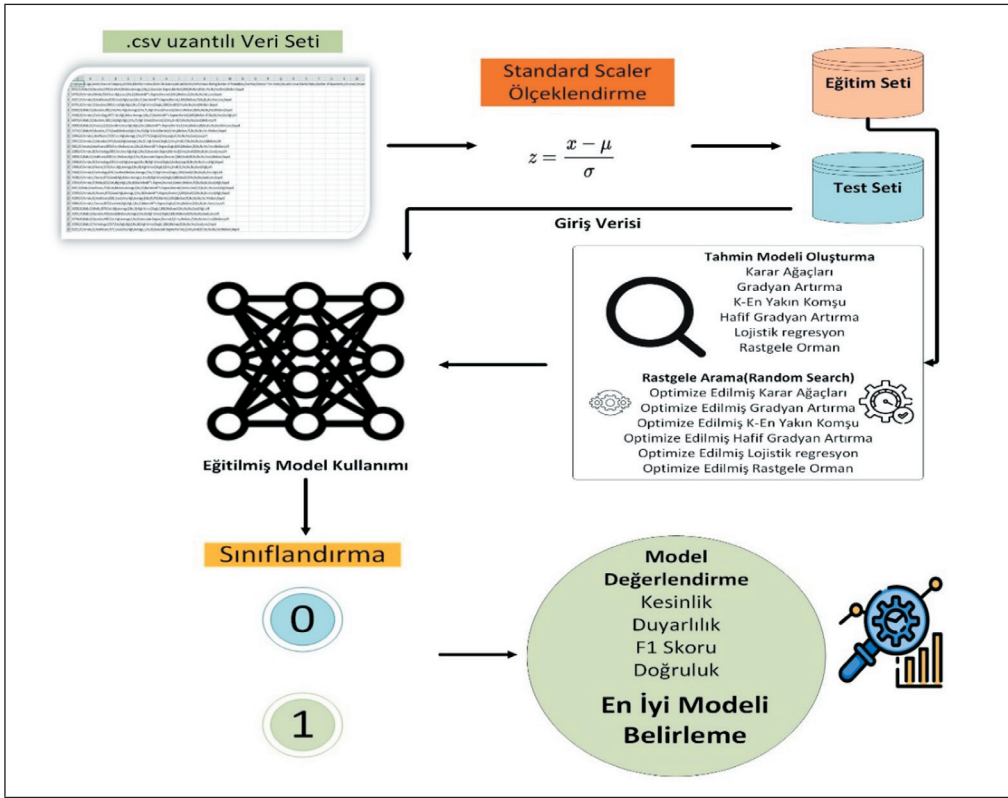
Bu çalışmada kullanılan yazılım araçları ve kullanılan bilgisayarın donanım özellikleri bu bölümde detaylandırılmıştır. Analizler, Python programlama dilinin güncel bir sürümü olan Python v11 kullanılarak Jupyter Notebook ortamında gerçekleştirilmiştir. Python, bilimsel hesaplamalar ve veri analizi için geniş kütüphanelere (NumPy, Pandas, Scikit-learn, TensorFlow, Keras) sahip olmasıyla tercih edilmiştir. Jupyter Notebook, kodları, görselleştirmeleri ve açıklamaları bir arada sunabilen interaktif bir ortam sağlayarak, analiz sürecinin anlaşılır ve tekrarlanabilir olmasını kolaylaştırır.

Kullanılan bilgisayar, son teknoloji ürünleri yansıtan yüksek performanslı özelliklere sahiptir. İşlemci olarak, çoklu görevler ve paralel işleme kapasitesi için uygun olan, 32 çekirdekli bir AMD Ryzen Threadripper bulunmaktadır. 64 GB DDR4 RAM, büyük veri setlerinin hafızada tutulması ve hızlı erişim için yeterli alan sağlamaktadır. Veri işleme ve model eğitimi hızını artırmak üzere, özellikle derin öğrenme modeli eğitimleri sırasında GPU desteği olarak NVIDIA RTX 4090 kullanılmıştır. Bu GPU, paralel hesaplama yetenekleri ile büyük ve karmaşık derin öğrenme algoritmalarının etkin bir şekilde eğitilmesine imkân sağlamıştır. Yazılım ve donanımın bu kombinasyonu, özellikle büyük veri setleri üzerinde karmaşık makine öğrenmesi ve derin öğrenme modellerinin hızlı ve etkili bir şekilde eğitilmesi için kritik öneme sahiptir. Böylece, model geliştirme süreci optimize edilirken, elde edilen sonuçların doğruluğu ve güvenilirliği de maksimize edilmiştir. Bu altyapı, araştırma sorularına yanıt ararken yüksek hesaplama gereksinimlerini karşılayarak, bilimsel keşiflerin sınırlarını zorlamamıza imkân tanımıştır.

3.7. Önerilen Metodoloji

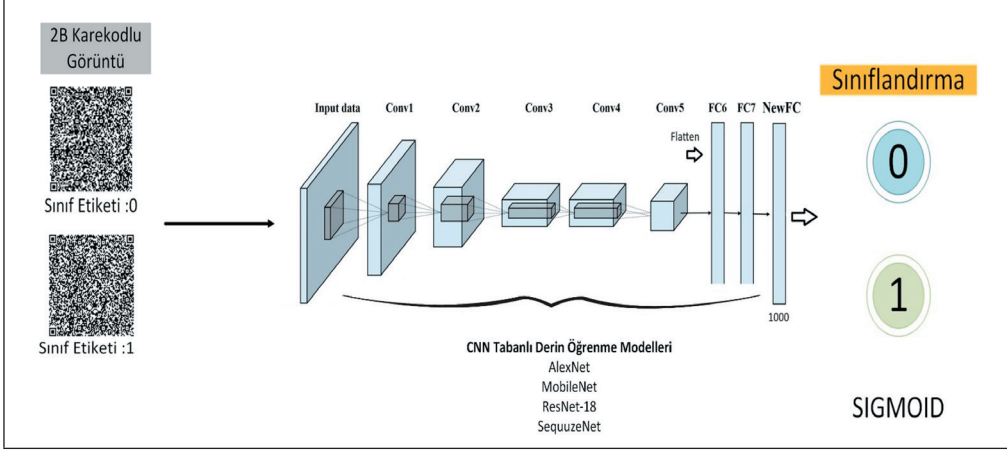
Bu çalışmada izlenen ve önerilen metodoloji, çeşitli aşamalardan oluşmakta olup, her biri belirli hedeflere yönelik tasarlanmıştır. İlk aşamada, makine öğrenmesi modelleri kullanılarak sınıflandırma işlemi gerçekleştirilmiştir. Bu süreçte, Karar Ağaçları, Gradient Boosting, KNN, LightGBM, Logistic Regression ve Random Forest gibi çeşitli algoritmalar üzerinde denemeler yapılmıştır. Bu modellerin her biri, veri seti üzerinde eğitilerek, performansları doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metriklerle değerlendirilir. Bu değerlendirme sonucunda, en iyi performans gösteren model belirlenmiştir. Birinci aşama Şekil 2’de gösterilmektedir.

Şekil 2: Önerilen Metodolojinin İlk Aşamaları



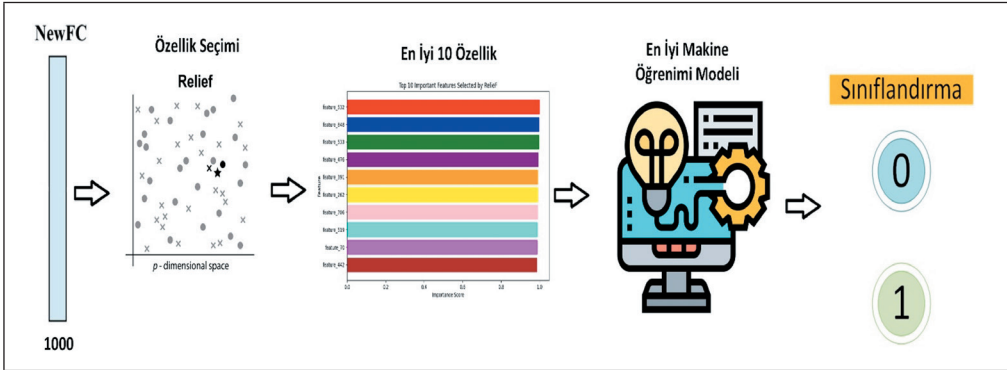
İkinci aşamada, derin öğrenme modelleri kullanılarak sınıflandırma işlemi yapılmıştır. Bu kapsamda AlexNet, MobileNet, ResNet-18 ve SqueezeNet gibi modeller, yine benzer performans metrikleri kullanılarak değerlendirilmiştir. Bu modellerden en iyi sonuçları verenin belirlenmesiyle, bu model üzerinden "NewFC" (Yeni Tam Bağantı Katmanı) katmanı eklenerek 1000 adet özellik çıkarımı yapılmıştır. Bu süreci gösteren şema ise Şekil 3’te gösterilmektedir.

Şekil 3: Önerilen Metodolojinin İkinci Aşaması



Üçüncü aşamada, çıkarılan bu 1000 özellik içerisinde en bilgilendirici 10 özelliği seçmek için Relief algoritması kullanılmıştır. Relief algoritması, özelliklerin sınıflandırma başarısına katkılarını analiz ederek, en etkili özellikleri belirler. Özellik sayısının düşürülmesi ile makine öğrenmesi modellerinin hızlı ve etkili bir şekilde eğitilmesi hedeflenerek zaman tasarrufu sağlanmaya çalışılmıştır. Son aşamada, seçilen bu 10 özellik, ilk aşamada en iyi performansı gösteren makine öğrenmesi modeline girdi olarak verilmiştir. Bu model, yeni özellik setiyle tekrar eğitilmiş ve sınıflandırma işlemi gerçekleştirilmiştir. Bu adımda, önerilen metodolojinin etkinliği ve özellik seçiminin model performansına olan etkisi değerlendirilmektedir. Bu süreç, Şekil 4’te gösterilmiştir.

Şekil 4: Önerilen Metodolojinin Üçüncü Aşaması



Bu metodoloji, hem makine öğrenmesi hem de derin öğrenme tekniklerini kapsamlı bir şekilde kullanarak, sınıflandırma problemleri için güçlü ve etkili çözümler sunmayı amaçlamaktadır. Her adım, modelin veri üzerindeki performansını artırmak ve daha doğru sınıflandırma sonuçları elde etmek için titizlikle planlanmıştır.

4. Deneysel Bulgular

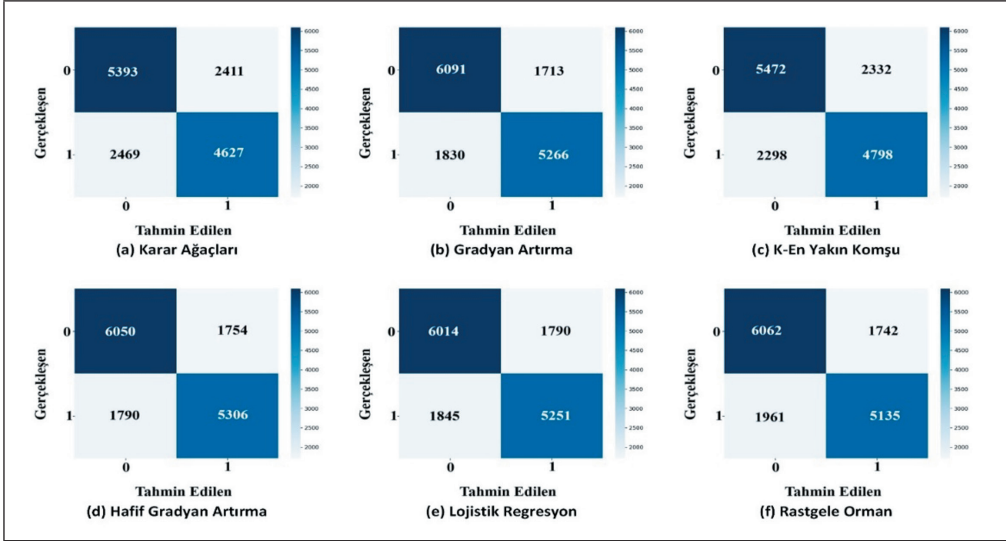
Bu bölümde, makine öğrenmesi ve derin öğrenme modellerinin sınıflandırma performanslarının yanı sıra, özellik çıkarımı ve seçimi ile ilgili elde edilen sonuçlar detaylı bir şekilde ele alınacaktır. Bu bölüm, model performanslarını değerlendirme ve araştırma hipotezlerini test etme süreçlerine dair kritik bilgiler sağlamaktadır. İlk olarak, makine öğrenmesi modellerinin test performansları, çeşitli istatistiksel metrikler kullanılarak değerlendirilmiştir. Her modelin doğruluk, kesinlik, duyarlılık, F1-Skoru ve ROC-AUC değerleri dikkatlice incelenmiştir. Modellerin test performanslarına ilişkin sınıflandırma raporu Tablo 4’te gösterilmektedir.

Tablo 4’te gösterilen sonuçlara göre, Lojistik Regresyon modeli %76 doğruluk ve 0,848 ROC-AUC skoru ile etkili bir sınıflandırma kapasitesine sahipken, Karar Ağacı modeli %67 doğruluk ve 0,672 ROC-AUC skoru ile aşırı uyuma eğilimli olduğunu göstermiştir. Rastgele Orman modeli, %75 doğruluk ve 0,841 ROC-AUC skoru ile iyi bir denge sunarken, Gradyan Artırma ve Hafif Gradyan Artırma modelleri sırasıyla %76 doğruluk ve 0,854, 0,855 ROC-AUC skorları ile veri yapılarındaki karmaşık kalıpları başarıyla modellemiştir. KNN modeli ise %69 doğruluk ve 0.747 ROC-AUC skoru ile göreceli olarak daha düşük performans sergilemiş ve gürültülü verilerden daha fazla etkilenmiş olabileceği düşünülmüştür. Makine öğrenmesi modelleri arasında değerlendirildiğinde, Hafif Gradyan Artırma ve Gradyan Artırma modelleri en yüksek performansı göstermişlerdir. Her iki model de %76 doğruluk oranına ulaşmış ve sırasıyla 0,855 ve 0,854 ROC-AUC skorları ile öne çıkmışlardır. Bu metrikler, her iki modelin de sınıflandırma işleminde yüksek derecede doğruluk ve dengeli bir performans sergilediğini göstermektedir. Ancak, Hafif Gradyan Artırma modeli biraz daha yüksek ROC-AUC skoru ile en başarılı model olarak kabul edilebilir. Bu model, karmaşık veri yapılarını etkili bir şekilde işleyebilir ve çeşitli sınıflandırma senaryolarında yüksek doğruluk ve dengeli sonuçlar sunar, bu da onu işten ayrılma tahminleri için özellikle uygun kılar. Bu bulgular, Hafif Gradyan Artırma modelinin makine öğrenmesi modelleri içinde en etkili model olduğunu ve araştırma sorularına cevap vermede kritik bir rol oynayabileceğini göstermektedir. Ek olarak modellerin test sonuçlarını gösteren karmaşıklık matrisleri Şekil 5’de gösterilmektedir.

Tablo 4: Makine Öğrenmesi Modellerinin Test Performanslarına İlişkin Sınıflandırma Raporu

Model	Sınıf Etiketi	Hassasiyet	Duyarlılık	F1-Skoru	Destek
Karar Ağacı	0	0,69	0,69	0,69	7804
	1	0,66	0,65	0,65	7096
	Doğruluk			0,67	14900
	Makro Ortalama	0,67	0,67	0,67	14900
	Ağırlıklı Ortalama	0,67	0,67	0,67	14900
	ROC-AUC			0,6715	14900
Gradyan Artırma	0	0,77	0,78	0,77	7804
	1	0,75	0,74	0,75	7096
	Doğruluk			0,76	14900
	Makro Ortalama	0,76	0,76	0,76	14900
	Ağırlıklı Ortalama	0,76	0,76	0,76	14900
	ROC-AUC			0,8544	14900
K-En Yakın Komşu	0	0,70	0,70	0,70	7804
	1	0,67	0,68	0,67	7096
	Doğruluk			0,69	14900
	Makro Ortalama	0,69	0,69	0,69	14900
	Ağırlıklı Ortalama	0,69	0,69	0,69	14900
Hafif Gradyan Artırma	0	0,77	0,78	0,77	7804
	1	0,75	0,75	0,75	7096
	Doğruluk			0,76	14900
	Makro Ortalama	0,76	0,76	0,76	14900
	Ağırlıklı Ortalama	0,76	0,76	0,76	14900
Lojistik Regresyon	0	0,70	0,70	0,70	7804
	1	0,67	0,68	0,67	7096
	Doğruluk			0,69	14900
	Makro Ortalama	0,69	0,69	0,69	14900
	Ağırlıklı Ortalama	0,69	0,69	0,69	14900
Rastgele Orman	0	0,76	0,78	0,77	7804
	1	0,75	0,72	0,73	7096
	Doğruluk			0,75	14900
	Makro Ortalama	0,75	0,75	0,75	14900
	Ağırlıklı Ortalama	0,75	0,75	0,75	14900
	ROC-AUC			0,8414	14900

Şekil 5: Makine Öğrenmesi Modellerinin Test Sonuçlarına İlişkin Karmaşıklık Matrisleri



Şekil 5'te gösterilen karmaşıklık matrisleri, makine öğrenmesi modellerinin test sonuçlarını somutlaştırmakta ve her modelin sınıflandırma başarısını detaylı bir şekilde göstermektedir. Karar Ağaçları, Gradyan Artırma, K-En Yakın Komşu (KNN), Hafif Gradyan Artırma (LightGBM), Lojistik Regresyon ve Rastgele Orman modelleri için oluşturulan bu matrisler, gerçek pozitif, gerçek negatif, yanlış pozitif ve yanlış negatif değerleri içermekte ve modellerin sınıflandırma performansını değerlendirmekte kullanılmıştır. Matris sonuçları, LightGBM modelinin, her iki sınıfı da dengeli bir şekilde sınıflandırdığını ve diğer modeller arasında en üstün performansı gösterdiğini ortaya koymaktadır. Özellikle, LightGBM modeli, doğru pozitif tahminlerde 5306 ve doğru negatif tahminlerde 6050 olarak gözlemlenmiştir, bu da modelin yüksek doğruluk ve güvenilirlikle çalıştığını gösterir. Diğer modellerle karşılaştırıldığında, LightGBM'in üstün performansı, işten ayrılma tahminleri için kritik olan karmaşık veri yapılarını ve örüntüleri etkin bir şekilde işleme yeteneğinden kaynaklanmaktadır. Bu analiz, LightGBM'in işten ayrılma durumlarını tahmin etmede en güvenilir model olarak tercih edilmesini desteklemektedir.

Bununla birlikte, çalışmada CNN tabanlı derin öğrenme modellerinin sınıflandırma performansına bakılacak olursa, modellerin test sonuçlarına ilişkin sınıflandırma raporu Tablo 5'te sunulmaktadır.

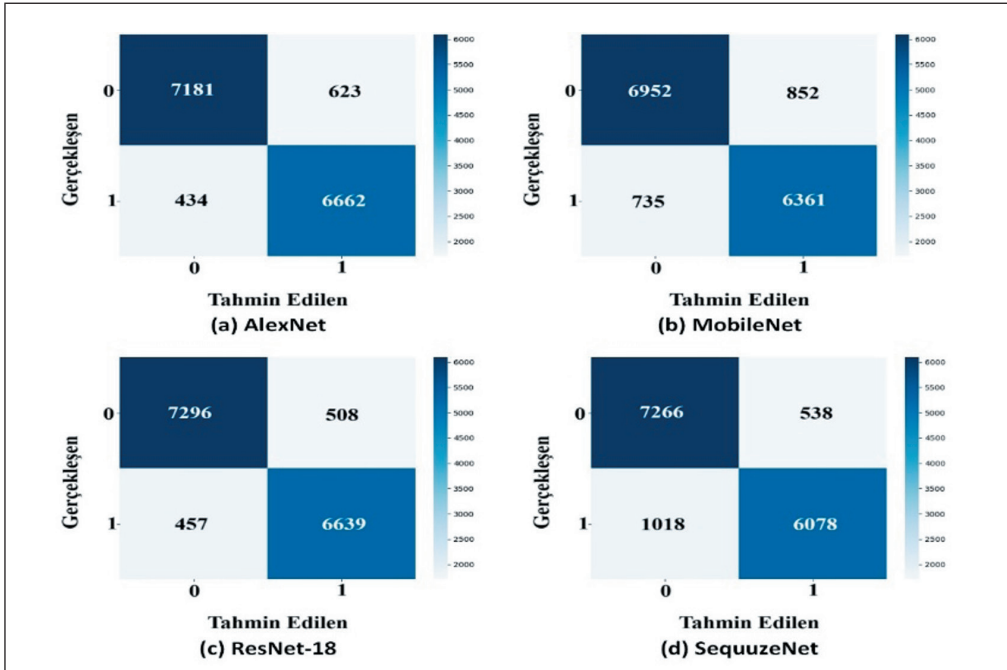
Tablo 5: Derin Öğrenme Modellerinin Test Performanslarına İlişkin Sınıflandırma Raporu

Model	Sınıf Etiketi	Hassasiyet	Duyarlılık	F1-Skoru	Destek
AlexNet	0	0,94	0,92	0,93	7804
	1	0,91	0,94	0,93	7096
	Doğruluk			0,93	14900
	Makro Ortalama	0,93	0,93	0,93	14900
	Ağırlıklı Ortalama	0,93	0,93	0,93	14900
	ROC-AUC			0,9852	14900
MobileNet	0	0,90	0,89	0,90	7804
	1	0,88	0,90	0,89	7096
	Doğruluk			0,89	14900
	Makro Ortalama	0,89	0,89	0,89	14900
	Ağırlıklı Ortalama	0,89	0,89	0,89	14900
	ROC-AUC			0,9646	14900
ResNet-18	0	0,94	0,93	0,94	7804
	1	0,93	0,94	0,93	7096
	Doğruluk			0,94	14900
	Makro Ortalama	0,93	0,94	0,94	14900
	Ağırlıklı Ortalama	0,94	0,94	0,94	14900
	ROC-AUC			0,9865	14900
SequuzeNet	0	0,88	0,93	0,90	7804
	1	0,92	0,86	0,89	7096
	Doğruluk			0,90	14900
	Makro Ortalama	0,90	0,89	0,89	14900
	Ağırlıklı Ortalama	0,90	0,90	0,90	14900
	ROC-AUC			0,9687	14900

Tablo 5’te gösterilen sonuçlar, derin öğrenme modellerinin işten ayrılma durumlarını sınıflandırmada gösterdiği başarıyı ortaya koymaktadır. AlexNet modeli %93 doğruluk oranı ile etkileyici bir performans sergilemiş, %94 kesinlik ve %92 duyarlılık ile sınıf 0 (kalmış) ve %91 kesinlik ve %94 duyarlılık ile sınıf 1 (ayrılmış) için yüksek değerler kaydetmiştir. Modelin ROC-AUC değeri 0,985, bu da modelin mükemmel bir tahmin yeteneğine sahip olduğunu gösterir. MobileNet modeli, %89 doğruluk oranı ile %90 kesinlik ve %89 duyarlılık ile sınıf 0 ve %88 kesinlik ve %90 duyarlılık ile sınıf 1 için dengeli performans sergilemiştir. ROC-AUC değeri 0,965, modelin genel olarak iyi tahmin yeteneğine sahip olduğunu göstermektedir. ResNet-18 ise %94 doğruluk ile AlexNet’i hafifçe aşarak en yüksek performansı göstermiştir. Bu model, %94 kesinlik ve %93 duyarlılık ile sınıf 0 ve %93 kesinlik ve %94 duyarlılık ile sınıf 1 için dengeli ve güçlü sonuçlar elde etmiştir. Modelin ROC-AUC değeri 0,987 olup, bu, sınıflandırma performansının son derece yüksek olduğunu belirtir. SqueezeNet, %90 doğruluk

oranı ile diğer modellere kıyasla biraz daha düşük performans göstermiş olmasına rağmen, %88 kesinlik ve %93 duyarlılık ile sınıf 0 ve %92 kesinlik ve %86 duyarlılık ile sınıf 1 için güçlü sonuçlar sunmuştur. ROC-AUC değeri 0,969 ile yine yüksek tahmin yeteneğine işaret eder. Bu sonuçlar, derin öğrenme modellerinin karmaşık veri yapılarını etkin bir şekilde işleyerek yüksek doğruluklu tahminler yapabildiğini göstermektedir. Özellikle, ResNet-18 modeli en yüksek genel performansı sunarak, işten ayrılma tahminleri için en uygun derin öğrenme modeli olarak öne çıkmıştır. Bununla birlikte, modellerin test sonuçlarına ilişkin karmaşıklık matrisleri Şekil 6'da gösterilmektedir.

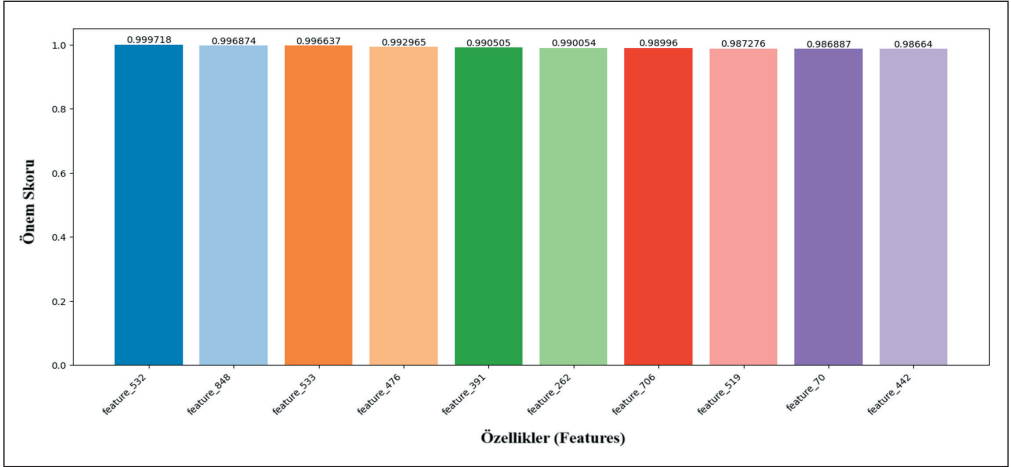
Şekil 6: CNN Tabanlı Derin Öğrenme Modellerinin Test Sonuçlarına İlişkin Karmaşıklık Matrisleri



Şekil 6'da gösterilen karmaşıklık matrisleri, derin öğrenme modellerinin test sonuçlarını detaylı bir şekilde göstermekte olup, bu modeller arasında ResNet-18 açık bir şekilde en başarılı model olarak öne çıkmaktadır. ResNet-18, 7296 doğru negatif (kalmış) ve 6639 doğru pozitif (ayrılmış) tahminle en yüksek doğru sınıflandırma sayılarına ulaşmıştır. Bu model, düşük yanlış pozitif (508) ve yanlış negatif (457) sayıları ile dengeli bir performans sergilemiş ve yüksek doğruluk oranları elde etmiştir. AlexNet modeli de 7181 doğru negatif ve 6662 doğru pozitif tahminle yüksek performans göstermiş, ancak yanlış negatif (434) ve yanlış pozitif (623) sayıları ResNet-18'e göre biraz daha yüksek kalmıştır. MobileNet ve SqueezeNet modelleri, sırasıyla 6952 ve 7266 doğru negatif tahminle iyi performans sergilemişlerdir; ancak SqueezeNet, 1018 yanlış pozitif ile diğer modellere göre daha fazla hata yapmıştır. Bu sonuçlar, ResNet-18'in sınıflandırma doğruluğu ve dengeli tahmin yeteneği açısından derin öğrenme modelleri arasında en etkili olduğunu göstermektedir, bu da onu işten ayrılma tahminleri için en uygun model haline getirmektedir.

CNN tabanlı derin öğrenme modelleri arasında en iyi sınıflandırma performansını sergileyen ResNet-18 modelinin son katmanına, “NewFC” adında yeni bir tam bağlantı katmanı eklenmiş ve 1000 adet özellik çıkarımı yapılmıştır. Daha sonra, çıkarılan bu 1000 özellik içerisinden en bilgilendirici 10 özelliği seçmek için Relief algoritması kullanılmıştır. Relief algoritması, özelliklerin sınıflandırma başarısına katkılarını analiz ederek, etkili özellikleri belirlemektedir. Özellik sayısının düşürülmesi ile makine öğrenmesi modellerinin hızlı ve etkili bir şekilde eğitilmesi hedeflenerek zaman tasarrufu sağlanmaya çalışılmıştır. Relief algoritması kullanılarak yapılan özellik seçimi sonuçları, belirli özelliklerin sınıflandırma performansına olan katkılarını açıkça ortaya koymaktadır. Bu süreçte, toplam 1000 özellik içerisinden sınıflandırma başarısına en yüksek katkıda bulunan 10 özellik seçilmiştir. Bu özelliklerin her birinin önem derecesi, algoritmanın özelliklerin sınıflandırma başarısı üzerindeki etkilerini analiz ederek hesaplanmıştır. Aşağıda Şekil 7’de, seçilen özelliklerin önem dereceleri gösterilmektedir.

Şekil 7: Relief Algoritması Kullanılarak Seçilen 10 Özelliğin Önem Dereceleri



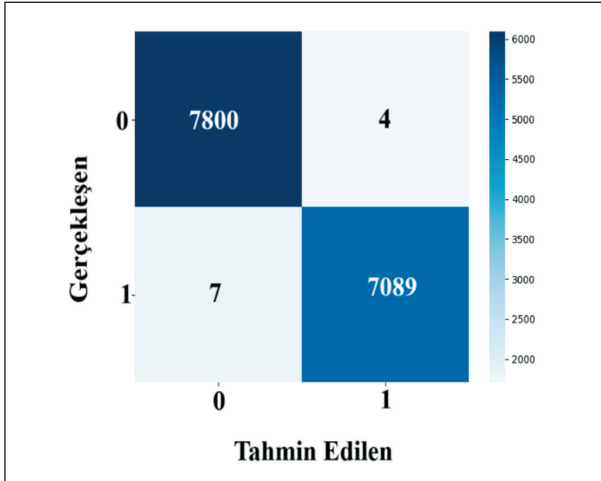
Bu aşamadan sonra, seçilen özellikler, makine öğrenmesi modelleri arasında en iyi sınıflandırma performansını gösteren, optimize edilmiş Hafif Gradyan Artırma (LightGBM) modeline girdi olarak verilmiş ve sınıflandırma işlemi yeniden gerçekleştirilmiştir. Modelin test sonuçlarına ilişkin sınıflandırma raporu Tablo 6’da gösterilmektedir.

Tablo 6: Özellik Seçimi Sonrası Hafif Gradyan Artırma Modelinin Test Performansına İlişkin Sınıflandırma Raporu

Model	Sınıf Etiketi	Hassasiyet	Duyarlılık	F1-Skoru	Destek
Hafif Gradyan Artırma	0	1,00	1,00	1,00	7804
	1	1,00	1,00	1,00	7096
	Doğruluk			1,00	14900
	Makro Ortalama	1,00	1,00	1,00	14900
	Ağırlıklı Ortalama	1,00	1,00	1,00	14900
	ROC-AUC			1,00	14900

RelieF algoritması ile seçilen en bilgilendirici 10 özelliğin kullanıldığı Hafif Gradyan Artırma (LightGBM) modelinin sınıflandırma performansı, sonuçlar üzerinde olağanüstü bir etki yaratmıştır. Seçilen özelliklerin modelin girdisi olarak verilmesi sonrasında gerçekleştirilen testler, modelin her iki sınıfı (0 ve 1) mükemmel bir şekilde tahmin ettiğini göstermiştir. Model, sınıf 0 ve sınıf 1 için hassasiyet, duyarlılık ve F1-skoru değerlerinde %100 başarıya ulaşmıştır. Toplam 14900 veri noktası üzerinden yapılan değerlendirmede, modelin doğruluk, makro ortalama ve ağırlıklı ortalama değerleri de %100 olarak kaydedilmiştir. ROC-AUC değeri de %100 olarak ölçülmüş, bu sonuç modelin sınıflandırma işleminde mutlak bir başarı sergilediğini ve veri setindeki tüm pozitif ve negatif sınıfları kesinlikle doğru bir şekilde ayırt ettiğini ortaya koymuştur. Bu sonuçlar, özellik seçiminin ve model optimizasyonunun, makine öğrenmesi modellerinin performansını nasıl radikal bir şekilde iyileştirebileceğini göstermektedir. Buna ek olarak, modelin test sonuçlarına ilişkin karmaşıklık matrisi Şekil 8'de gösterilmektedir.

Şekil 8: Hafif Gradyan Artırma Modelinin Test Sonuçlarına İlişkin Karmaşıklık Matrisi



Şekil 8'de gösterilen matristen görüldüğü üzere, model 7800 gerçek pozitif ve 7089 gerçek negatif tahminde bulunmuştur, bu da modelin genel olarak yüksek doğrulukla sınıflandırma yaptığını göstermektedir. Yanlış pozitif sayısı 4 iken, yanlış negatif sayısı 7'dir. Bu sonuçlar, modelin hem pozitif hem de negatif sınıfları yüksek hassasiyetle ayırt edebildiğini ve çok düşük hata oranları ile performans sergilediğini belirtmektedir. Bu matris, modelin genel performansını değerlendirmede kritik bir araç olarak kullanılmakta ve modelin gerçek dünya verilerine uygulanabilirliğini göstermektedir. Bununla birlikte, modelin sonuçlarını doğrulamak amacıyla 5 katlı çapraz doğrulama yöntemi uygulanmıştır. Bu, modelin genelleştirilebilirliğini ve kararlılığını test etmek için önemlidir. Bu yöntem, veri setinin rastgele beş eşit parçaya bölünmesini ve her bir parçanın sırayla test seti olarak kullanılmasını içerirken, kalan dört parça eğitim seti olarak kullanılır. Bu süreç, modelin farklı veri alt kümeleri üzerinde beş kez eğitilmesi ve test edilmesi anlamına gelmektedir. Çapraz doğrulama, modelin sadece bir kısmı üzerinde değil, veri setinin geneli üzerinde nasıl performans gösterdiğini anlamaya yardımcı

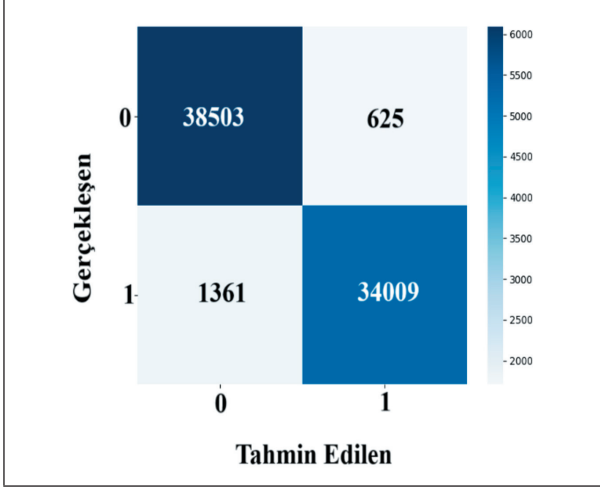
olmaktadır. Bu yöntemin uygulanması, modelin farklı veri kümelerindeki tutarlılığını ölçmekte ve aşırı uyum (overfitting) gibi potansiyel problemleri ortaya çıkarmaya yardımcı olmaktadır. Eğer model, çapraz doğrulama sürecinde her bir kat için benzer ve yüksek performans gösterirse, modelin verilerin farklı bölümlerine iyi genelleştirdiği ve dolayısıyla daha önce görmediği verilere karşı da iyi tahminler yapabileceği anlaşılır. Ayrıca, çapraz doğrulama modelin her bir sınıflandırma iterasyonundaki performansını ortaya koyarak, modelin belirli bir veri tipine veya örneğe özgü olmadığını, genel bir başarı sergilediğini doğrular. Dolayısıyla, bu yöntemle elde edilen sonuçlar, modelin gerçek dünya uygulamalarında nasıl performans göstereceğine dair güçlü ve güvenilir kanıtlar sağlamakla birlikte, model geliştiricilere ve karar vericilere modellerini değerlendirme ve iyileştirme konusunda değerli bilgiler sunacaktır. Modelin 5 katlı çapraz doğrulama sonuçlarına ilişkin sınıflandırma raporu Tablo 7’de gösterilmektedir.

Tablo 7: Hafif Gradyan Artırma Modelinin Çapraz Doğrulama Sonuçlarına İlişkin Sınıflandırma Raporu

Model	Sınıf Etiketi	Hassasiyet	Duyarlılık	F1-Skoru	Destek
Hafif Gradyan Artırma	0	0,97	0,98	0,97	39128
	1	0,98	0,96	0,97	35370
	Doğruluk			0,97	74498
	Makro Ortalama	0,97	0,97	0,97	74498
	Ağırlıklı Ortalama	0,97	0,97	0,97	74498

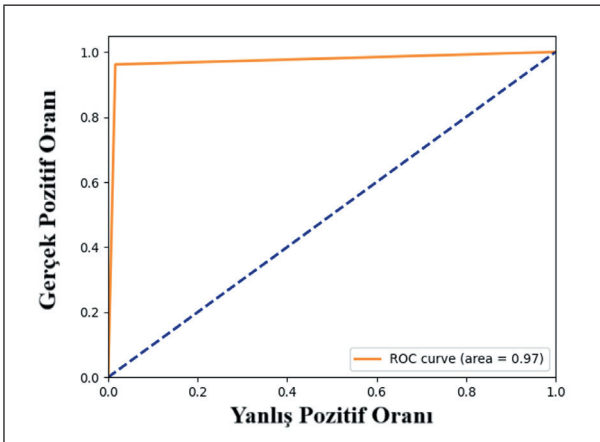
Tablo 7’deki sonuçlar, Hafif Gradyan Artırma (LightGBM) modelinin 5 katlı çapraz doğrulama sürecinde gösterdiği performansı detaylı bir şekilde ortaya koymaktadır. Model, hem sınıf 0 (kalmış çalışanlar) hem de sınıf 1 (ayrılmış çalışanlar) için %97 hassasiyet ve %97’lik bir F1-skoru ile dikkat çekici bir başarı sergilemiştir. Sınıf 0 için %98 duyarlılık ve sınıf 1 için %96 duyarlılık elde edilmiştir, bu oranlar modelin her iki sınıfı da yüksek doğrulukla tahmin edebildiğini göstermektedir. Toplam 74,498 veri noktası üzerinden yapılan değerlendirmede, modelin genel doğruluk oranı %97 olarak ölçülmüştür. Bu yüksek doğruluk oranı, modelin veri setinin genelinde güçlü ve tutarlı tahminler yapabildiğinin bir göstergesidir. Makro ve ağırlıklı ortalama değerlerinin her ikisi de %97 olarak belirlenmiş olması, modelin sınıflandırma performansının tüm sınıflar arasında dengeli ve tutarlı olduğunu belirtir. Bu sonuçlar, Hafif Gradyan Artırma modelinin, işten ayrılma durumlarını tahmin etmede son derece etkili ve güvenilir bir araç olduğunu doğrular. Çapraz doğrulama sonuçları, modelin genelleştirilebilirliğini ve farklı veri alt kümelerindeki performansını başarıyla test ederek, modelin pratik uygulamalarda nasıl performans göstereceğine dair güçlü kanıtlar sağlamıştır. Bu, modelin gerçek dünya senaryolarında güvenle kullanılabileceği anlamına gelir. Bununla birlikte, modelin çapraz doğrulama sonuçlarına ilişkin karmaşıklık matrisi Şekil 9’da gösterilmektedir.

Şekil 9: Modelin Çapraz Doğrulama Sonuçlarına İlişkin Karmaşıklık Matrisi



İlgili karmaşıklık matrisi sonuçları, modelin yüksek doğrulukla tahmin yaptığını göstermektedir. Toplamda, model 38,503 doğru negatif tahmin yapmış ve bu sınıfta sadece 625 yanlış pozitif, sınıf 1 için ise 34,009 doğru pozitif ve yalnızca 1,361 yanlış negatif tahminde bulunmuştur. Bu, modelin özellikle sınıf 0'daki duyarlılığının yüksek olduğunu ve sınıf 1'deki gerçek pozitifleri başarıyla tanıyabildiğini gösterir. Düşük yanlış pozitif ve negatif değerleri, modelin yanlış tahmin oranının az olduğunu ve güvenilir olduğunu belirtir. Karmaşıklık matrisi, modelin her iki sınıfı da yüksek doğrulukla sınıflandırabildiğini ve işten ayrılma durumlarını etkili bir şekilde tahmin edebildiğini kanıtlamaktadır. Ek olarak, modelin ROC eğrisi grafiği Şekil 10'da sunulmuştur.

Şekil 10: Modelin Çapraz Doğrulama Sonuçlarına İlişkin ROC Eğrisi



İlgili şekilde görüldüğü üzere, ROC-AUC değerinin %97 olması, modelin neredeyse hatasız bir sınıflandırma performansına sahip olduğunu ve yanlış pozitif ile yanlış negatif tahminlerde mükemmel yakın sınıflandırma yaptığını, bu da modelin pratik kullanım için ideal olduğunu gösterir. Özetle, çapraz doğrulama ile teyit edilen sonuçlar sayesinde, modelin sadece bir veri setine spesifik olmadığı, geniş veri setlerinde güvenilir tahminler yapabileceği doğrulanmış olur. Bu süreç aynı zamanda modelin daha önce görmediği veriler üzerinde de başarılı olabileceğine dair güven sağlar ve modelin gerçek dünya uygulamalarında nasıl performans göstereceği konusunda daha sağlam kanıtlar sunmaktadır. Bu, modeli gerçek dünya senaryolarında kullanmadan önce bir güven testi niteliği taşır ve modelin genel başarısını artırır.

5. Sonuç

Bu çalışma, işten ayrılma tahminleri yapmak amacıyla makine öğrenmesi ve derin öğrenme modellerinin etkinliğini ve optimizasyonunu detaylı bir şekilde incelenmiştir. Günümüzde işgücü devir hızının artması, şirketler için ciddi maliyetlere ve iş gücü verimliliğinde düşüşlere yol açmaktadır. Dolayısıyla, işten ayrılma olasılığını erken tahmin edebilmek, insan kaynakları yönetimi için stratejik bir öneme sahiptir. Bu çalışma, bilhassa özellik çıkarımı ve özellik seçimi konularında derinlemesine bir analiz yaparak, yüksek doğrulukta tahminler sağlamaktadır. Çalışmanın diğer benzer araştırmalardan ayrılan en önemli özelliği, çeşitli makine öğrenmesi ve derin öğrenme modellerini kapsamlı bir şekilde karşılaştırması ve en etkili özellikleri belirleyerek model performansını maksimize etme stratejisidir.

Çalışmada, ilk olarak, makine öğrenmesi modelleri arasında Hafif Gradyan Artırma (LightGBM) modeli, yüksek doğruluk oranları ve dengeli sınıflandırma performansı sayesinde en etkili model olarak belirlenmiştir. Ardından, derin öğrenme modelleri kullanılarak yapılan sınıflandırma işlemlerinde, ResNet-18 modeli öne çıkmıştır. Daha sonra, model performanslarının daha da artırılması amacıyla, ResNet-18 modeli üzerinden 1000 adet özellik çıkarılmış ve bu özellikler arasından RelieF algoritması kullanılarak en bilgilendirici 10 özellik seçilmiştir. Bu özellikler, daha sonra LightGBM modeline girdi olarak verilerek sınıflandırma işlemi tekrar gerçekleştirilmiş ve model, neredeyse mükemmel bir sınıflandırma başarısı göstermiştir (%100 doğruluk). Çalışmanın son aşamasında, modelin genelleştirilebilirliği ve kararlılığını değerlendirmek amacıyla 5 katlı çapraz doğrulama uygulanmış ve modelin farklı veri alt kümelerinde de yüksek doğruluk oranları elde ettiği teyit edilmiştir. Bu sonuçlar, modelin aşırı uyuma karşı dayanıklı olduğunu ve geniş veri setlerinde güvenilir sonuçlar üretebileceğini göstermiştir.

Bu çalışmada, işten ayrılma durumunun tahminleri için makine öğrenmesi ve derin öğrenme modellerinin entegrasyonunu kapsamlı bir şekilde ele alınmıştır. Literatürde benzer yöntemler genellikle ya sadece makine öğrenmesi modelleri veya sadece derin öğrenme modelleri kullanılarak yapılmaktadır. Örneğin, Zhu vd. (2016) tarafından yapılan çalışma, lojistik regresyonun işten ayrılma tahminleri üzerindeki etkisini incelerken, bu çalışma büyük veri setlerindeki karmaşıklıkları yeterince ele alamamıştır. Bu noktada, çalışmamız makine öğrenmesi ve derin öğrenme tekniklerini birleştirerek, her iki dünyanın en iyi yönlerini entegre etmekte ve daha karmaşık veri yapılarını etkili bir şekilde modellemektedir. Literatürdeki diğer bir örnek olan Kanuto (2024)'nin çalışması, K-En Yakın Komşu ve Destek Vektör Makineleri gibi algoritmalar kullanılarak işten ayrılma tahminleri gerçekleştirilmiş ve bu modeller %75 ile %79 arasında bir doğruluk oranı elde etmiştir. Buna karşın, bizim çalışmamızda, Hafif Gradyan Artırma (LightGBM) modeli özellik seçimi sonrası %100 doğruluk oranına ulaşarak

literatürdeki benzer çalışmalardan belirgin bir şekilde ayrılmaktadır. Pekel Özmen & Özcan (2022) tarafından geliştirilen evrişimsel sinir ağlarına dayalı modelin sınıflandırma doğruluğu %79 iken, bizim ResNet-18 tabanlı modelimiz %94 doğruluk elde etmiştir. Bu, derin öğrenme modellerinin özellikle görüntü tabanlı veri dönüşümünden yararlanılarak yapılan işten ayrılma tahminlerinde üstün performans sergileyebileceğini göstermektedir. Bununla birlikte, çalışmamız, işten ayrılma tahminlerinde, özellikle de derin öğrenme ve makine öğrenimi tekniklerinin entegrasyonu ile Adibaji & Marleen (2022) tarafından yapılan çalışmadan önemli ölçüde ayrılmaktadır. Adibaji & Marleen'in çalışması, daha geleneksel makine öğrenimi yöntemlerine odaklanmış ve sınırlı sayıda veri kaydı (311 kayıt) üzerinde gerçekleştirilmişken, bu çalışma büyük bir veri seti (74,498 kayıt) üzerinde geniş kapsamlı bir analiz sunmaktadır ve derin öğrenme tekniklerini de içermektedir. Çalışmamızda, son teknoloji derin öğrenme modellerinden çıkarılan özellikler makine öğrenmesi modeline entegre edilerek %100 doğruluk, hassasiyet ve duyarlılık değerleri elde etmiştir. Bu, Adibaji & Marleen'in çalışmasında elde edilen %98 doğruluk, %100 hassasiyet ve %98 duyarlılık sonuçlarını geçmektedir. Yüksek ROC-AUC değeri (1.00) ile çalışmamız, sınıflandırma modelinin mükemmel bir şekilde dengelenmiş olduğunu ve yanlış pozitif ile yanlış negatif tahminlerde üstün performans sergilediğini göstermektedir. Sonuç olarak, bu çalışma, karmaşık ve büyük veri kümelerinde işten ayrılma tahminleri yapabilme kapasitesini artırmış ve literatürdeki benzer çalışmalara kıyasla daha yüksek doğruluk oranları sağlayarak önemli bir katkı sunmuştur. Ayrıca, özellik seçimi ve optimizasyon süreçlerinin bu başarıdaki rolü, gelecekteki araştırmalara yol gösterici olabilir. Bu bağlamda, işten ayrılma tahminlerinde kullanılan yöntemlerin sürekli olarak güncellenmesi ve iyileştirilmesi, insan kaynakları yönetimi pratiğindeki etkinliği artıracaktır.

Bu çalışmanın kısıtlılıkları arasında, özellik seçimi sürecinin yalnızca bir algoritma ile sınırlı olması bulunmaktadır. Farklı özellik seçimi algoritmaları kullanılarak sonuçların nasıl etkilenebileceği üzerine daha fazla araştırma yapılabilir. Ayrıca, modelin farklı endüstrilerde ve coğrafi bölgelerdeki veri setleri üzerinde test edilmesi, genelleştirilebilirlik açısından faydalı olacaktır. Bununla birlikte, gelecekteki çalışmalar, farklı veri kaynaklarından elde edilen verileri entegre etmeye yönelik metodolojiler geliştirebilir. Örneğin, sosyal medya, performans değerlendirmeleri ve iş tatmini anketleri gibi çeşitli kaynaklardan toplanan verilerin birleştirilmesi, daha kapsamlı ve doğru tahminler yapılmasını sağlayabilir. Ayrıca, farklı makine öğrenmesi ve derin öğrenme modellerini hiyerarşik veya kademeli bir şekilde kullanarak, her bir modelin güçlü yönlerinden en iyi şekilde yararlanılabilir. Örneğin, birinci aşamada genel bir filtreleme için basit modeller kullanılabilirken, ikinci aşamada daha detaylı analizler için karmaşık modeller devreye sokulabilir. Bununla birlikte, modellerin gerçek zamanlı verilerle sürekli olarak güncellenmesi, işten ayrılma tahminlerinin zaman içindeki değişikliklere daha hızlı ve etkili bir şekilde adapte olmasını sağlayabilir. Bu, özellikle dinamik iş gücü piyasalarında ve hızlı değişen organizasyonel yapılar içinde değerlidir. Ek olarak, insan kaynakları uygulamalarında yapay zekâ etiğine özel bir önem verilmelidir. Modelleme süreçlerinin şeffaf olması, kullanılan verilerin ve alınan kararların adil olması gerekmektedir. Çalışanlar üzerinde potansiyel olarak büyük etkilere sahip olan bu tür sistemlerin etik standartlara uygun olarak geliştirilmesi ve yürütülmesi kritik öneme sahiptir. Ek olarak, organizasyonlar, insan kaynakları personelinin makine öğrenmesi ve derin öğrenme konularında eğiterek, bu teknolojilerin işten ayrılma tahminlerinde nasıl kullanılabileceği konusunda bilgi sahibi olmalarını sağlamalıdır. Bu, teknolojik araçların daha etkin kullanımını ve stratejik insan kaynakları yönetimi uygulamalarının geliştirilmesini teşvik edecektir. Bu çalışmanın bulguları, insan kaynakları strateji-

lerinin geliştirilmesine yönelik değerli bilgiler sunmaktadır. Özellikle, işten ayrılma oranlarını azaltmak ve çalışan bağlılığını artırmak için proaktif stratejiler geliştirilebilir. Bu stratejiler, çalışan memnuniyetini artırmaya yönelik girişimler, kariyer gelişim fırsatlarının sunulması ve çalışma ortamının iyileştirilmesini içerebilir.

Katkı Oranı Beyanı

Makale yazarları çalışmaya eşit oranda katkıda bulunmuşlardır.

Çıkar Çatışması Beyanı

Bu çalışmada herhangi bir potansiyel çıkar çatışması bulunmamaktadır.

Kaynakça

- Adeusi, K. B., Amajuoyi, P., & Benjami, L. B. (2024). Utilizing machine learning to predict employee turnover in high-stress sectors. *International Journal of Management & Entrepreneurship Research*, 6(5), 1702-1732. <https://doi.org/10.51594/ijmer.v6i5.1143>
- Adibaji, S. S., & Marleen, O. (2022). Comparative analysis of methods k-nearest neighbor, support vector machine and decision tree on prediction model of turnover intention. *Journal Research of Social Science, Economics, and Management*, 2(2). <https://doi.org/10.59141/jrssem.v2i02.241>
- Aglin, G., Nijssen, S., & Schaus, P. (2020). Pydl8.5: A library for learning optimal decision trees. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. <https://doi.org/10.24963/ijcai.2020/750>
- Al Akasheh, M., Hujran, O., Faisal Malik, E., & Zaki, N. (2024). Enhancing the prediction of employee turnover with knowledge graphs and explainable ai. *IEEE Access*, 12, 77041-77053. <https://doi.org/10.1109/access.2024.3404829>
- AlMohamed, M., AlAqeel, A., & Alkandari, K. (2022). Turnover and organizational commitment in the oil and gas industry in saudi arabia. *International Journal of Research in Human Resource Management*, 4(2), 01-06. <https://doi.org/10.33545/26633213.2022.v4.i2a.105>
- Bae, C. Y., Im, Y., Lee, J., Park, C., Kim, M., Kwon, H. U., ... 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., Kyrlyenko, O., Parfenyuk, Y., Krivtsov, S., Meniailov, I., Kuznietcova, V., ... 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>
- Cheng, L., Lin, C. H., Sun, C., & Wang, S. (2019). Evolutionary-fuzzy-integral-based convolutional neural networks for facial image classification. *Electronics*, 8(9), 997. <https://doi.org/10.3390/electronics8090997>
- Chivukula, R., Sajja, M. V., Lakshmi, T. J., & Harini, M. (2021). Empirical study on Microsoft malware classification. *International Journal of Advanced Computer Science and Applications*, 12(3). <https://doi.org/10.14569/ijacsa.2021.0120361>
- Conroy, S. A., Roumpi, D., Delery, J. E., & Gupta, N. (2021). Pay volatility and employee turnover in the trucking industry. *Journal of Management*, 48(3), 605-629. <https://doi.org/10.1177/01492063211019651>

- Eldora, K., Fernando, E., & Winanti, W. (2024). Comparative analysis of knn and decision tree classification algorithms for early stroke prediction: A machine learning approach. *Journal of Information Systems and Informatics*, 6(1), 313-338. <https://doi.org/10.51519/journalisi.v6i1.664>
- Feeley, T. H., & Barnett, G. A. (1997). Predicting employee turnover from communication networks. *Human Communication Research*, 23(3), 370-387. <https://doi.org/10.1111/j.1468-2958.1997.tb00401.x>
- Gao, X., Wen, J., & Zhang, C. (2019). An improved random forest algorithm for predicting employee turnover. *Mathematical Problems in Engineering*, 2019, 1-12. <https://doi.org/10.1155/2019/4140707>
- Grebovic, M., Filipović, L., Katnić, I., Vukotic, M., & Popović, T. (2022, November). Overcoming limitations of statistical methods with artificial neural networks. In *Proceedings of the 2022 International Arab Conference on Information Technology (ACIT)*, (1-6), IEEE. <https://doi.org/10.1109/acit57182.2022.9994218>
- Guo, M., & Du, Y. (2019, October). Classification of thyroid ultrasound standard plane images using ResNet-18 networks. In *Proceedings of the 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, (324-328), Xiamen, China. <https://doi.org/10.1109/ICASID.2019.8925267>
- Harrison, S., & Gordon, P. A. (2014). Misconceptions of employee turnover: evidence-based information for the retail grocery industry. *Journal of Business & Economics Research (JBER)*, 12(2), 145. <https://doi.org/10.19030/jber.v12i2.8528>
- Hasan, M. K., Sundarajan, E., Islam, S., Ahmed, F. R. A., Babiker, N. B. M., Alzahran, A. I., ... Khan, M. A. (2024). A novel segmented random search based batch scheduling algorithm in fog computing. *Computers in Human Behavior*, 158, 108269. <https://doi.org/10.1016/j.chb.2024.108269>
- Hassanpour, M., & Malek, H. (2020). Learning document image features with squeezenet convolutional neural network. *International Journal of Engineering*, 33(7). <https://doi.org/10.5829/ije.2020.33.07a.05>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016, October). Identity mappings in deep residual networks. In *Proceedings of the Computer Vision—ECCV 2016: 14th European Conference Proceedings Part IV 14*, (630-645), Amsterdam, The Netherlands. <https://doi.org/10.48550/arxiv.1603.05027>
- Hien, D. T. T., Thi, C., Kim, T., The, D., & Nguyen, C. (2020). Optimize the combination of categorical variable encoding and deep learning technique for the problem of prediction of vietnamese student academic performance. *International Journal of Advanced Computer Science and Applications*, 11(11). <https://doi.org/10.14569/ijacsa.2020.0111135>
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint*, arXiv:1704.04861.
- Hussain, S. (2014). Total path length and number of terminal nodes for decision trees. *Procedia Computer Science*, 35, 514-521. <https://doi.org/10.1016/j.procs.2014.08.132>
- Ingsih, K., Kadarningsih, A., & Rijati, N. (2022, February). Job stress, compensation, job dissatisfaction and turnover intention. In *Proceedings of the 2nd International Conference on Industry 4.0 and Artificial Intelligence (ICIAI 2021)*, (pp. 68-72). Atlantis Press. <https://doi.org/10.2991/aisr.k.220201.013>
- Jaderberg, M., Vedaldi, A., & Zisserman, A. (2014). Speeding up convolutional neural networks with low rank expansions. *arXiv preprint*, arXiv:1405.3866.
- 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>

- Kaharuddin, K., & Sholeha, E. W. (2021). Classification of fish species with image data using k-nearest neighbor. *International Journal of Computer and Information System (IJCIS)*, 2(2), 54-58. <https://doi.org/10.29040/ijcis.v2i2.33>
- Kanuto, A. E. (2024). Identifying patterns and predicting employee turnover using machine learning approaches. *International Journal of Science and Business*, 36(1), 20-35. <https://doi.org/10.58970/ijsb.2373>
- Kim, S. Y., & Fernández, S. (2016). Employee empowerment and turnover intention in the U.S. federal bureaucracy. *The American Review of Public Administration*, 47(1), 4-22. <https://doi.org/10.1177/0275074015583712>
- Liao, C. (2023, February). Employee turnover prediction using machine learning models. In *Proceedings of the International Conference on Mechatronics Engineering and Artificial Intelligence (MEAI 2022)*, (227-231), Changsha, China. <https://doi.org/10.1117/12.2672733>
- Lim, C. S., Malik, E. F., Khaw, K. W., Alnoor, A., Chew, X., Chong, Z. L., ... Al Akasheh, M. (2024). Hybrid GA–DeepAutoencoder–KNN Model for employee turnover prediction. *Statistics, Optimization & Information Computing*, 12(1), 75-90. <https://doi.org/10.19139/soic-2310-5070-1799>
- Liu, J. E., & An, F. P. (2020). Image classification algorithm based on deep learning-kernel function. *Scientific programming*, 2020(1), 1-14. <https://doi.org/10.1155/2020/7607612>
- Liu, Y., Dou, Y., & Qiao, P. (2020). Beyond top-n accuracy indicator: A comprehensive evaluation indicator of cnn models in image classification. *IET Computer Vision*, 14(6), 407-414. <https://doi.org/10.1049/iet-cvi.2018.5839>
- Marquez, B. Y., Realyvásquez-Vargas, A., Lopez-Esparza, N., & Ramos, C. E. (2023). Application of ordinary least squares regression and neural networks in predicting employee turnover in the industry. *Archives of Advanced Engineering Science*, 2(1), 30-36. <https://doi.org/10.47852/bonviewaaes32021326>
- Nagassou, M., Mwangi, R. W., & Nyarige, E. (2023). A hybrid ensemble learning approach utilizing light gradient boosting machine and category boosting model for lifestyle-based prediction of type-ii diabetes mellitus. *Journal of Data Analysis and Information Processing*, 11(04), 480-511. <https://doi.org/10.4236/jdaip.2023.114025>
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7. <https://doi.org/10.3389/fnbot.2013.00021>
- Oguine, O. C., & 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 & Reports*, 1-7. [https://doi.org/10.47363/jccsr/2021\(3\)182](https://doi.org/10.47363/jccsr/2021(3)182)
- Ogunsanya, M., Isichei, J., & Desai, S. (2023). Grid search hyperparameter tuning in additive manufacturing processes. *Manufacturing Letters*, 35, 1031-1042.
- Ou, R. (2020). Out-of-core gpu gradient boosting. *arXiv preprint*, <https://doi.org/10.48550/arxiv.2005.09148>
- Özen, H., & Bal, C. (2019). A study on missing data problem in random forest. *Osmangazi Journal of Medicine*, 42(1), 103-109. <https://doi.org/10.20515/otd.496524>
- Pakarinen, O., Karsikas, M., Reito, A., Lainiala, O., Neuvonen, P., & 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>
- Pal, S., Pramanik, A., Maiti, J., & Mitra, P. (2021). Deep learning in multi-object detection and tracking: State of the art. *Applied Intelligence*, 51(9), 6400-6429. <https://doi.org/10.1007/s10489-021-02293-7>

- Park, D., Kim, S. S., Kwon, H., Shin, D., & Shin, D. (2021). Host-based intrusion detection model using siamese network. *IEEE Access*, 9, 76614-76623. <https://doi.org/10.1109/access.2021.3082160>
- Pekel Ozmen, E., & Ozcan, T. (2022). A novel deep learning model based on convolutional neural networks for employee churn prediction. *Journal of Forecasting*, 41(3), 539-550. <https://doi.org/10.1002/for.2827>
- Pourkhodabakhsh, N., Mamoudan, M. M., & Bozorgi-Amiri, A. (2022). Effective machine learning, meta-heuristic algorithms and multi-criteria decision making to minimizing human resource turnover. *Applied Intelligence*, 53(12), 16309-16331. <https://doi.org/10.1007/s10489-022-04294-6>
- Samašonok, K. (2024). Employee turnover: Causes and retention strategies. *Entrepreneurship and Sustainability Issues*, 11(3), 134-148. [https://doi.org/10.9770/jesi.2024.11.3\(9\)](https://doi.org/10.9770/jesi.2024.11.3(9))
- Sarwinda, D., Paradisa, R. H., Bustamam, A., & Anggia, P. (2021). Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. *Procedia Computer Science*, 179, 423-431. <https://doi.org/10.1016/j.procs.2021.01.025>
- Sihare, M. (2024). Evaluation of machine learning methods for prediction student performance. *International Journal for Research in Applied Science and Engineering Technology*, 12(1), 534-544. <https://doi.org/10.22214/ijraset.2024.58001>
- Siraj, F. M., Ayon, S. T. K., Samad, M. A., Uddin, J., & Choi, K. (2024). Few-shot lightweight squeezenet architecture for induction motor fault diagnosis using limited thermal image dataset. *IEEE Access*, 12, 50986-50997. <https://doi.org/10.1109/access.2024.3385430>
- 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>
- Srivastava, P. R., & Eachempati, P. (2021). Intelligent employee retention system for attrition rate analysis and churn prediction. *Journal of Global Information Management*, 29(6), 1-29. <https://doi.org/10.4018/jgim.20211101.0a23>
- Stamolampros, P., Korfiatis, N., Chalvatzis, K., & Buhalis, D. (2019). Job satisfaction and employee turnover determinants in high contact services: Insights from employees'online reviews. *Tourism Management*, 75, 130-147. <https://doi.org/10.1016/j.tourman.2019.04.030>
- Tews, M. J., & Stafford, K. (2020). The impact of abusive supervision and constituent attachment on entry-level employee turnover. *Journal of Hospitality; Tourism Research*, 44(8), 1318-1334. <https://doi.org/10.1177/1096348020947139>
- Tong, S., Sun, W., Xu, J., & 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>
- Viswanadapalli, A. (2021). Efficient data mining model for employees churn prediction and safety measure. *Psychology and Education Journal*, 58(1), 1962-1982. <https://doi.org/10.17762/pae.v58i1.1049>
- Wang, L., & Zhao, L. (2022). Digital economy meets artificial intelligence: Forecasting economic conditions based on big data analytics. *Mobile Information Systems*, 2022, 1-9. <https://doi.org/10.1155/2022/7014874>
- Wijaya, D., Ds, J. H., Barus, S., Pasaribu, B., Sirbu, L. I., & Dharma, A. (2021). Uplift modeling vs conventional predictive model: A reliable machine learning model to solve employee turnover. *International Journal of Artificial Intelligence Research*, 5(1). <https://doi.org/10.29099/ijair.v4i2.169>
- 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>

- Xue, X., Sun, X., Wang, H., Zhang, H., & Feng, J. (2023). Neural network fusion with fine-grained adaptation learning for turnover prediction. *Complex & Intelligent Systems*, 9(3), 3355-3366. <https://doi.org/10.1007/s40747-022-00931-2>
- Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In D., Fleet, T., Pajdla, B., Schiele, T. Tuytelaars, (eds) *Computer Vision – ECCV 2014. Lecture Notes in Computer Science*, vol 8689. Springer, Cham, (818-833), Zurich, Switzerland. https://doi.org/10.1007/978-3-319-10590-1_53
- Zhang, C., Yang, Z., Fang, Z., Li, Y., Li, X., & Zhang, Y. (2024, May). Transformer fault diagnosis method based on gramian angular field and graph convolution network. In *Proceedings of the Eighth International Conference on Energy System, Electricity, and Power (ESEP 2023)*, (1627-1635). SPIE. <https://doi.org/10.1117/12.3024388>
- Zhang, J. (2024). Neuroscientific prediction model of mouse brain activity patterns. *Highlights in Science, Engineering and Technology*, 92, 295-301. <https://doi.org/10.54097/yecss238>
- Zhu, X., Sawhney, R., & Upreti, G. (2016). Determinates of employee voluntary turnover and forecasting in departments: A case study. *Studies in Engineering and Technology*, 3(1), 64. <https://doi.org/10.11114/set.v3i1.1635>
- Zia, U. (2024), Employee Attrition Classification Dataset, An In-Depth Synthetic Simulation for Attrition Analysis and Prediction, <https://www.kaggle.com/datasets/stealthtechnologies/employee-attribution-dataset/data> adresinden erişilmiştir. Erişim Tarihi: 25.07.2024.