DOĞUŞ ÜNİVERSİTESİ DERGİSİ
DOGUS UNIVERSITY JOURNAL
e-ISSN: 1308-6979
https://dergipark.org.tr/tr/pub/doujournal

ÇALIŞAN DEVİR HIZINI TAHMİN ETMEDE CNN VE XGBOOST ALGORİTMALARI İLE GELİŞTİRİLEN HİBRİT MODEL ÖNERİSİ (*)

A HYBRIT PREDICTION MODEL DEVELOPED WITH CNN AND XGBOOST ALGORITMS FOR PREDICTING EMPLOYEE TURNOVER RATE

Alp ALPSOY[(1)], Dilek TÜKEL[(2)]

Öz: İşletmelerde yüksek çalışan devir hızı hem maliyetleri artırmakta hem de performansı olumsuz etkilemektedir. Yüksek devir hızı, işletmenin sürekliliğini bozmakta, çalışanların odaklanmasını zorlaştırmakta ve deneyimli personelin kaybına yol açmaktadır. Yeni personel bulma ve adaptasyon süreci zaman alıcı ve maliyetlidir, bu süreçte işletme kültürüne uyum sağlanması verimliliği düşürür. Bu durum, işletme içindeki güven ve istikrarı olumsuz etkileyerek müşteri ve iş ortakları arasında güven kaybına yol açar. Bu çalışmada, çalışanların işlen ayrılma nedenlerini anlamak için CNN ve XGBoost algoritmaları kullanılarak İK verileri analiz edilmiştir. Araştırmanın amacı, çalışan devir hızını tahmin eden bir modeli geliştirmektir. Araştırmanın yönteminde, Kaggle platformundan elde edilen sentetik bir İK veri seti üzerinde CNN ve XGBoost algoritmaları kullanılarak hibrit bir model oluşturulmuş, modelin performansı doğruluk, hassasiyet ve F1 skorları gibi metriklerle değerlendirilmiştir. Bulgular, hibrit modelin %99 doğruluk oranı ile çalışan devir hızı tahmininde yüksek performans sergilediğini göstermiştir. Sonuç olarak, geliştirilen modelin işgücü planlamasında ve çalışan memnuniyetini artırmada etkili bir aracı olabileceği belirtilmiştir. Ancak, bu tür modellerin kullanımı, etik ve yönetim açısından dikkatli değerlendirilmelidir. İşletmelerin, bu tahmin modellerini yalnızca iş gücü planlamasında değil, aynı zamanda çalışan memnuniyetini artırmaya yönelik stratejiler geliştirmede kullanmaları önemlidir.

Anahtar Kelimeler: Makine Öğrenimi, Çalışan Devir Hızı Tahmini, XGBoost, CNN.

Abstract: *High employee turnover in businesses not only increases costs but also negatively impacts performance. Elevated turnover disrupts business continuity, makes it difficult for employees to maintain focus, and leads to the loss of experienced personnel. The process of recruiting and onboarding new staff is both time-consuming and costly, while the adaptation period required for aligning with the organizational culture reduces overall efficiency. This situation adversely affects trust and stability within the organization, potentially resulting in a loss of confidence among customers and business partners. In this study, HR data were analyzed using CNN and XGBoost algorithms to understand the reasons behind employee attrition. The aim of the*

(*) Bu çalışma, birinci yazarın ikinci yazar danışmanlığında hazırladığı yüksek lisans projesinden üretilmiştir.

(1) Doğuş Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar ve Enformasyon Bilimleri Bölümü; alp.alpsoy@yahoo.com ORCID: 0009-0002-3544-531X
(2) Doğuş Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü; dtukel@dogus.edu.tr ORCID: 0000-0001-5288-5996

Geliş/Received: 10-09-2024; Kabul/Accepted: 10-04-2025

Atıf bilgisi: Alpsoy, A., Tükel, D. (2026). Çalışan devir hızını tahmin etmede CNN ve XGboost algoritmaları ile geliştirilen hibrit model önerisi. *Doğuş Üniversitesi Dergisi*, *27*(1), 1-18, DOI: 10.31671/doujournal.1547333

research is to develop a model that predicts employee turnover rate. In the methodology, a hybrid model was created by applying CNN and XGBoost algorithms to a synthetic HR dataset obtained from the Kaggle platform. The model's performance was evaluated using metrics such as accuracy, precision, and F1score. The findings demonstrated that the hybrid model exhibited high performance in predicting employee turnover, achieving an accuracy rate of 97%. Consequently, the developed model is suggested to be an effective tool for workforce planning and enhancing employee satisfaction. However, the use of such models should be carefully considered from both ethical and managerial perspectives. It is crucial for businesses to utilize these predictive models not only for workforce planning but also to develop strategies aimed at improving employee satisfaction.

Keywords: *Machine Learning, Employee Turnover Prediction, XGBoost, CNN.*

JEL: *C88, L86, L39.*

1. Giriş

Veri odaklı karar alma süreçleri, endüstride işletme performansını arttırmak için kullanılmaktadır. İşletmeler, büyük veri analizi ve yapay zekâ teknikleriyle elde edilen içgörüler sayesinde rekabet avantajı sağlamayı hedeflemektedir. İnsan kaynakları yönetimi bu tekniklerden yararlanarak personel performansını artırmayı, çalışan bağlılığını güçlendirecek uygulamalar geliştirmeyi hedeflemektedir.

Geleneksel insan kaynakları süreçleri, teknoloji ve yapay zekâ uygulamaları karşısında yetersiz kalmaktadır (Misican, 2020). Yapay zekâ destekli sistemler, işe alım süreçlerinde büyük veri kümelerini hızlı bir şekilde analiz ederek, adayların beceri ve deneyimlerini iş gereklilikleriyle eşleştirerek insan kaynakları departmanlarına zaman ve verimlilik kazandırarak süreci hızlandırır (Tonbil ve Aksakal, 2024). Yapay zekâ, aday değerlendirmelerinde önyargıları en aza indirerek daha nesnel sonuçlar elde edilmesine olanak sağlar (Dündar ve Ağaçayak, 2023). Bu teknolojiler sayesinde, yoğun emek isteyen işler otomatikleştirilirken, insan faktörü daha anlamlı işlere odaklanma fırsatı bulmaktadır. Yapay zeka tabanlı sistemler, adayların özgeçmişlerini tarayarak en uygun adayları belirleme ve eğitim programlarını kişiselleştirme gibi işlevleri yerine getirebilmektedir. Bu tür teknolojiler, geleneksel yöntemlerin aksine, büyük veri analitiği kullanarak daha isabetli ve objektif kararlar alınmasına olanak tanımaktadır (Kambur, 2021). Bu doğrultuda, yapay zekâ tabanlı sistemlerin yalnızca işe alım süreçlerinde değil, çalışan yönetimi ve organizasyonel planlamada da kritik roller üstlendiği ifade edilmektedir (Yurtsever, 2023). Özellikle, çalışanların işten ayrılma eğilimlerini tahmin etmek, insan kaynakları yönetimi için kritik bir konudur ve işletmelerin operasyonel süreçlerine etki eden birçok faktörü kapsar. Bu tahminler, çalışanların ayrılma olasılığını belirlemeye yönelik modellerin geliştirilmesiyle, personel devrini azaltma, yetenekleri koruma ve iş gücü planlamasını iyileştirmede stratejik kararlara yardımcı olur. Veri bilimi ve makine öğrenimi teknikleri, bu tahminlerde güçlü bir araç olarak kullanılarak, büyük veri analizi yoluyla çalışanların ayrılma eğilimlerini belirleme ve yönetme imkanı sunar.

Teknolojik yeniliklerle birlikte yaşanan hızlı değişimlere işletmelerin ve iş gücünün hızla adapte olması oldukça önemlidir. Alan taraması yapıldığında insan kaynakları yönetiminde teknolojik yenilikleri konu alan çalışmalar (Zhong, Fang, Li, Sun, 2008; Tunçer, 2012; Xu, Ji, 2013; Del Prado, Rosellon, 2017; Türkel, 2018; Aksu ve

Sürgevil, 2019; Çiftçioğlu vd., 2019; Pelenk, 2020; Şahin ve Yılmaz, 2021) bulunmaktadır ancak, iş gücü devir hızı ya da işten ayrılma tahminleme yöntemleri ile ilgili teknolojinin kullanımına yönelik yapılan çalışmaların daha sınırlı olduğu söylenebilir (Mahlasela ve Chinyamurindi, 2020; Tharani ve Raj, 2020; Bahadır, vd., 2021; Yıldız, 2023; Adeusi vd., 2024; Bilenler vd., 2024). Araştırmanın bu yönüyle özgün nitelik taşıdığı düşünülmektedir. Buradan hareketle çalışmanın amacı, çalışanların işten ayrılma olasılığını tahmin etmek için makine öğrenimi tekniklerini kullanarak bir model geliştirmektir. Çalışma kapsamında, Kaggle platformundan elde edilen sentetik İK verileri üzerinde çalışan devir hızı tahmini için CNN ve XGBoost algoritmaları kullanılarak hibrit bir model oluşturulmuştur. Çalışma farklı endüstrilerde faaliyet gösteren ve insan kaynakları yönetimi süreçlerine sahip olan işletmeleri kapsarken, insan kaynakları verilerinin analiz edilerek çalışanların işten ayrılma olasılıklarını tahmin eden makine öğrenimi algoritmalarının uygulanmasını içermektedir.

2. Kavramsal Çerçeve

İnsan kaynakları yönetimi, işletmelerin stratejik hedeflere ulaşılabilmesinde önemli bir role sahiptir. Çalışan devir hızının etkin bir şekilde yönetilmesi, işletmelerin sürdürülebilirlik ve rekabet gücünü artırması açısından önemlidir. Teknolojik gelişmelerin yaygınlaşması, iş gücü yönetiminde geleneksel yöntemlerin yerini makine öğrenimi ve yapay zeka tabanlı modellerin almasına yol açmaktadır. Bu da insan kaynakları yönetim şekillerinde değişiklikleri beraberinde getirmiştir. Değişiklikle değerlendirildiğinde bu bölümde çalışma kapsamında kullanılan bazı kavramlar (veri bilimi, makine öğrenmesi, XGBoost algoritması, evrişimli sinir ağları, işten ayrılma tahmini) literatür çerçevesinde ele alınacaktır. Konuya ilişkin ayrıntılı açıklamalar, aşağıdaki bölümlerde yer almaktadır.

2.1. Veri Bilimi

Veri bilimi, istatistik, makine öğrenimi ve bilgisayar bilimi gibi disiplinleri birleştirerek, veri analizi, modelleme ve tahmin yoluyla veriden anlamlı bilgiler çıkarma sürecidir (Ersöz ve Çınar, 2021). Veri bilimi, işletmelerin verilerini toplamasına, temizlemesine, analiz etmesine ve bu analizlerden elde edilen içgörülerle stratejik kararlar almasına yardımcı olur. Bu disiplin, finans, sağlık, perakende, e-ticaret ve birçok sektörde önemli uygulamalara sahiptir. Örneğin, sağlık kuruluşları hasta verilerini analiz ederek hastalık risklerini belirlemektedir (Fonseca, 2021)

2.1.1. Veri Bilimi Süreçleri ve Sınıflandırma

Veri bilimi süreçlerinin temel aşamaları şunlardır; veri temizleme ve ön işleme gerçekleştirilir (Şeker, 2018), bu aşamada eksik veya hatalı veriler düzeltilir ve veri setleri, veri normalleştirme, özellik mühendisliği ve boyut azaltma teknikleri ile hazırlanır (Oğuzlar, 2003). Veri görselleştirme, veri setinin yapısal özelliklerini ve ilişkilerini anlamak için grafikler, tablolar ve dağılım grafikleri gibi araçları kullanır (Aaslt, Damiani, 2015).

Veri sınıflandırma, bu alanda yaygın olarak kullanılan bir tekniktir ve veri noktalarını belirli sınıflara ayırmak için kullanılır. Sınıflandırma, çeşitli algoritmalarla gerçekleştirilmektedir (Coşkun ve Baykal, 2011). Karar ağaçları, veri setini ağaç benzeri bir yapıyla basit kararlar ve kural setleriyle sınıflandırır. Destek Vektör Makineleri (SVM), sınıfları ayıran optimal bir hiper düzlem seçmeye çalışır. K-En

Yakın Komşu (KNN) algoritması, bir veri noktasını en yakın komşularının çoğunluğuna göre sınıflandırır (Wang, Huang, Cao, 2023)

2.1.2. Makine Öğrenmesi

Makine öğrenmesi, yapay zekânın bir bileşeni olarak, bilgisayar sistemlerinin büyük veri kümelerini analiz ederek istenen davranışları geliştirmesini sağlayan algoritmalar bütünüdür (Nacar ve Erdebilli, 2021: 308). Yapay zekâ, bilgisayarların insan benzeri bir şekilde kendilerine verilen bilgiyi alıp karar vermelerine olanak tanıyan programlamaları içerir. Makine öğrenmesinin başarılı olabilmesi için yüksek miktarda ve nitelikli veri girişi gereklidir; veri ne kadar fazlaysa, öğrenme o kadar verimli ve doğru sonuçlar üretir (Gökalp, 2022: 2-3).

Makine öğrenmesi, denetimli ve denetimsiz öğrenme olmak üzere iki ana kategoriye ayrılır. Denetimli öğrenme, önceden tanımlanmış doğru sınıflandırma ile veri örneklerinin eğitilmesini içerirken, denetimsiz öğrenme, etiketlenmemiş verilerdeki gizli kalıpları belirlemek amacıyla kendi kendini organize eden sinir ağları aracılığıyla gerçekleştirilir (Sathya ve Abraham, 2013: 35). Denetimli öğrenme, sınıflandırma ve regresyon olarak ikiye ayrılırken, veri setinin içeriğine bağlı olarak farklı algoritmalar kullanılır (Nacar ve Erdebilli, 2021: 310).

2.2. XGBoost Algoritması Tahmin Esasları

XGBoost (Extreme Gradient Boosting), makine öğreniminde yaygın olarak kullanılan ve yüksek performans sağlayan bir algoritmadır (Tokmak, 2023). Özellikle sınıflandırma ve regresyon problemlerinde etkili olan XGBoost, her bir karar ağacının önceki ağaçların hatalarını düzeltmesine odaklanarak performansı artırır (Altun, Altun, 2025). Ayrıca, aşırı öğrenmeyi önlemek amacıyla karmaşıklığı kontrol için düğüm kesme, minimum ağırlık ve lambda parametreleri gibi regülerleştirme parametrelerini kullanarak modelin genelleme yeteneğini korur (Büyükkör, Doğan, 2024). Eğitim sürecinden sonra, yeni veriler üzerinde tahmin yapmak için tüm ağaçları birleştirerek final tahminler elde edilir (Ben, Stef, Carmona, 2023).

2.3. Evrişimli Sinir Ağları (Convolutional Neural Networks- CNN)

Evrişimli sinir ağları (CNN) görüntü tanıma, nesne tespiti ve görüntü sınıflandırma gibi görsel işleme görevlerinde başarı elde etmiştir (Gümüş, Eyüpoğlu, 2023). CNN'lerin temel bileşeni olan evrişim katmanları, girdi verisini işleyerek özellik haritaları üretir. Bu katmanlar, girdi verisindeki yerel ilişkileri öğrenir ve belirli özellikleri (kenarlar, köşeler, desenler) belirlemek için filtreler (kernel) kullanır (Fırat, Hanbay, 2022). Havuzlama katmanları, evrişim katmanlarının çıktısını küçültmek ve özetlemek için kullanılır; en yaygın kullanılan işlem maksimum havuzlamadır. Bu işlem, modelin öğrenme sürecini hızlandırırken aşırı uyum riskini de azaltır. Tam bağlantılı katmanlar, elde edilen öznitelikleri kullanarak sınıflandırma veya regresyon sonuçları üretir ve genellikle softmax aktivasyon fonksiyonu ile sınıflara dönüştürme işlemi gerçekleştirir. CNN'lerde yaygın olarak kullanılan aktivasyon fonksiyonları arasında ReLU, sigmoid ve tanh bulunur; ReLU, hızlı hesaplamalar ve eğitim süresince daha iyi performans sağlaması nedeniyle en çok tercih edilendir (Özmen, Özcan, 2021).

2.4. Çalışan Devir Hızı Tahmini ve İşletmeler İçin Önemi

Çalışanlar, bir şirketin en değerli unsurlarından biridir ve onların beklenmedik şekilde işten ayrılması ciddi maliyetlere ve zorluklara yol açmaktadır (Turan, 2017). Bu noktada çalışanların devir hızı tahmini oldukça önemlidir. Çalışan devir hızı tahmini, işletmelere çalışanların ayrılma olasılıklarını değerlendirerek personel yönetimi stratejileri geliştirme fırsatı sunar. Bu tahminler, işletmelere hangi departmanlarda ve pozisyonlarda ayrılma riskinin yüksek olduğunu belirleme konusunda rehberlik ederek yetenek kaybını önler ve çalışan bağlılığını artırır. Çalışan memnuniyeti ve bağlılığını artırmak için stratejiler geliştirilmesine de olanak tanır. İşletmeler, çalışanların ayrılma riskini azaltarak işe alım, eğitim maliyetleri ve üretkenlik kaybı gibi maliyetleri düşürür (Park, Feng, Jeong, 2024).

Şirketler, müşteri kayıplarını tahminleyen modellere yatırım yaparak, müşteri davranışlarını analiz eder ve potansiyel kayıp risklerini belirler. Hem müşteri hem de çalışan kayıplarını önlemek için veri odaklı yaklaşımlar, şirketlerin verimliliğini artırırken maliyetlerini de azaltır. Özellikle bilgi teknolojileri sektöründe %12-15 çalışan işten ayrılma oranının yüksek olması, şirketlere olumsuz etkiler yaratmaktadır. Bu durum, yeni çalışanların bulunması ve uyum sürecinin uzunluğu nedeniyle zaman, efor ve kaynak kaybına yol açmaktadır (Dolatabadi ve Keynia, 2017: 74; Saradhi ve Palshikar, 2011: 2005). Çalışanların işten ayrılma olasılığını tahmin etmeye yönelik analizler, şirketlerin iş gücü planlaması ve çalışan memnuniyeti stratejilerini geliştirmelerine yardımcı olabilmektedir (Holtom vd., 2008). Bu analizler, genellikle veri bilimi ve makine öğrenimi teknikleriyle gerçekleştirilir ve çalışma süresi, performans, maaş düzeyi gibi faktörlere dayanarak tahminler yapılır.

3. Metodoloji

Çalışmanın metodolojisi, sentetik bir veri kullanılarak işten ayrılma tahmini için geliştirilmiştir. Çapraz doğrulama yöntemi ile işten ayrılma yöntemi için gradient boosting, karar ağaçları ve destek vektör makineleri kullanılmıştır.

3.1. Araştırmanın Amacı ve Önemi

Bu çalışmanın amacı İnsan kaynakları alanında çalışanların işten ayrılma olasılığını tahmin etmek için veri bilimi ve makine öğrenimi tekniklerini kullanarak bir model geliştirmektir. Geliştirilen model, çalışan veri setleri üzerinde eğitilerek, işletmelere çalışanların ayrılma olasılıklarını tahmin etme konusunda yardımcı olabilecek bir çerçeve sunmayı amaçlamaktadır. Bu model, işletmelerin stratejik insan kaynakları planlamasında bilgiye dayalı kararlar almasına ve insan kaynakları yönetim süreçlerini optimize etmesine katkı sağlayacağı düşünülmektedir.

3.2. Araştırmanın Yöntemi

Bu çalışmanın evrenini, çeşitli endüstrilerde faaliyet gösteren ve insan kaynakları yönetimi süreçlerine sahip olan işletmeler oluşturmaktadır. Ancak çalışmada kullanılan veri seti, gerçek işletme verilerinden değil, bir çevrimiçi veri platformu olan Kaggle platformundan elde edilen sentetik bir insan kaynakları veri setinden oluşmaktadır. Kaggle, veri bilimi ve makine öğrenmesi alanında çalışan uzmanlar ve meraklılar için 2010 yılında ortaya çıkmış, 2017 yılında da Google tarafından satın alınmıştır. Kaggle, problem sahiplerinin veri setlerini ve veri setlerine yönelik problemlerini paylaştıkları, yarışmacılarında teknik bilgilerini kullanarak problemi

çözmeye çalıştıkları bir platformdur. Bu bağlamda yarışmacılar, problem sahibi tarafından sunulan para ödülü veya iş teklifleri ile ödüllendirilebilmekte ve ayrıca yeni veriler ve buna yönelik problemler ile teknik becerilerini geliştirebilmektedir. Bu çalışmada, Kaggle kaynağından elde edilen veri seti şu adresten edinilmiştir: https://www.kaggle.com/datasets/kadirduran/hr-dataset/data. Bu veri seti son değerlendirme, proje sayısı, ortalama aylık saat, zaman harcaması, iş kazası, departmanlar, maaş, promosyon, memnuniyet olmak üzere on öznitelik içermektedir.

Bu bağlamda, araştırma kapsamında kullanılan veri seti, gerçek bir işletmeye ait olmadığından, belirli bir ülke veya sektör için genelleme yapmaktan ziyade, makine öğrenimi tekniklerinin işten ayrılma oranını tahmin etmedeki etkinliğini test etmeyi amaçlamaktadır.

Örneklem seçiminde, geleneksel örnekleme yöntemleri (basit rastgele, tabakalı, küme örneklemesi vb.) yerine, veri biliminde kullanılan ve işten ayrılmayı tahmin etmeye yönelik bir veri kümesi kullanılmıştır. Buradan hareketle veri ön işleme ve makine öğrenimi modelleri kullanılarak oluşturulmuş bir veri seti tercih edilmiştir.

Veri ön işleme dört adımdan oluşmaktadır. (1) Veri setinde eksik gözlemler olup olmadığı kontrol edilmiş ve eksik veriler uygun istatistiksel yöntemlerle tamamlanmıştır (eksik verilerin analizi). (2) Departman, maaş gibi kategorik değişkenler, one-hot encoding veya label encoding teknikleriyle sayısal değerlere dönüştürülmüştür (kategorik değişkenlerin dönüştürülmesi). (3) Modelin daha iyi performans göstermesi için sürekli değişkenler standart ölçeklendirme (StandardScaler) yöntemi ile normalize edilmiştir (veri normalizasyonu). Son olarak (4) İşten ayrılan ve ayrılmayan çalışanların dengeli bir şekilde dağılması için veri seti üzerinde uygun dengeleme işlemleri yapılmıştır (örneklem dengesinin sağlanması).

Farklı makine öğrenimi modelleri değerlendrilmiş ve en yüksek performans gösteren modeller çalışma için seçilmiştir. Değerlendirilen modeller arasında; karar ağaçları, destek vektör makineleri, lojistik regresyon ve gradient boosting, XGBoost ve Evrişimli Sinir Ağları (Convolutional Neural Networks - CNN) modelleri bulunmaktadır. Modellerin, çapraz doğrulama sonuçları karşılaştırılmış ve performansları değerlendirilmiştir. En iyi performans gösteren model, daha fazla ayar ve optimizasyon adımlarıyla geliştirilmiş ve nihai model olarak seçilmiştir. Geliştirilen modelin performansı, doğruluk, hassasiyet, geri çağırma, F1 puanı ve ROC eğrisi gibi farklı değerlendirme metrikleri kullanılarak değerlendirilmiştir.

Ayrıca, modelin karmaşıklık matrisi üzerindeki performansı da incelenmiştir. Model, eğitim veri seti üzerinde ve ayrılmış test veri seti üzerinde değerlendirilmiştir. Model eğitimi ve değerlendirme sürecinde, çalışanların işten ayrılma nedenlerini daha iyi anlamak için veri görselleştirme teknikleri kullanılmıştır.

Tüm bu aşamalarda çalışmaya dahil edilerek kullanılan kütüphanelerin listesini şu şekilde sıralamak mümkündür: import pandas as pd, from sklearn.model_selection import train_test_split, from sklearn.linear_model import LogisticRegression, import numpy as np, from sklearn.metrics import accuracy_score, precision_score, recall_score, import pickle, from pandas.api.types import CategoricalDtype, from xgboost import XGBClassifier, from ydata_profiling import ProfileReport, import seaborn as sns, from sklearn.model_selection import cross_val_score, from sklearn.metrics import roc_curve, roc_auc_score, import matplotlib.pyplot as plt, from sklearn.model_selection import cross_validate, import warnings, from

sklearn.preprocessing import LabelEncoder, StandardScaler, OneHotEncoder, OrdinalEncoder, from sklearn.model_selection import cross_val_predict, from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier, VotingClassifier, from sklearn import set_config, from sklearn.metrics import classification_report, warnings.filterwarnings ("ignore"), from sklearn.compose import ColumnTransformer, from sklearn.neighbors import KNeighborsClassifier. Söz konusu import edilmiş kütüphaneler, işleri hızlı ve kolay yapabilmesi için eklenen hazır araçlardır. Geliştiriciler bunları kullanarak veri işleme, tahmin yapma, grafik çizme gibi işleri kolayca yapabilir.

Çalışmanın geliştirme ortamı Jupyter Notebook üzerinde tamamlanmıştır. Jupyter Notebook; özellikle veri analizi, makine öğrenimi ve bilimsel hesaplamalar için kullanılan interaktif bir kod geliştirme ortamıdır. Bu ortamın tercih edilme sebebini, satır satır kodlama yapılabilmesi, ilgili kodun hemen altında görüntülenebilmesi, sade bir arayüze sahip olması şeklinde açıklamak mümkündür.

4. Araştırmanın Bulguları

Çalışmanın bulguları, Kaggle'dan elde edilmiş insan kaynakları veri seti üzerinden elde edilen sonuçlar oluşmaktadır. Aşağıda bulunan başlıklarda bu çalışmaya ait bulgular detaylı bir şekilde ifade edilmiştir.

4.1. Veri Setine İlişkin Bulgular

Kaggle'dan elde edilen veri seti çeşitli öznitelikleri kapsamaktadır. İnsan kaynakları veri seti ile ilgili öznitelikler ile ilgili bilgiler Tablo 1'de açıklamalarıyla birlikte gösterilmektedir.

Tablo 1. İnsan Kaynakları Verisetinin Öznitelikleri

satisfaction_level	0-1 arasında değişen çalışan memnuniyet oranı
last_evaluation	0-1 arasında değişen çalışan performans ölçüm puanı
number_project	Çalışanın üzerinde bulunan proje sayısı
average_monthly_hours	Ortalama aylık çalışma sayısı
time_spend_company	Çalışanın işletmedeki kıdem yılı
work_accident	Çalışanın işletmede iş kazası geçirip geçirmediği
promotion_last_5years	Çalışanın son 5 yılda aldığı terfi
departments	İşletmedeki departmanlar
salary	Kademeli olarak ifade edilmiş ordinal değişken
left	Çalışanın işten ayrılıp ayrılmama durumu

Tablo 1 de açıklanan verilere göre insan kaynakları veri seti; satisfaction_level (çalışan memnuniyeti), last_evaluation (son performans değerlendirmesi), number_project (proje sayısı), average_monthly_hours (ortalama aylık çalışma zamanı), time_spend_company (kıdem yılı), work_accident (iş kazası), promotion_last_5years (son beş yılda aldığı terfi), departments (departmanlar), salary (değişken) ve left (ayrılma durumu) olmak üzere on öznitelik içermektedir.

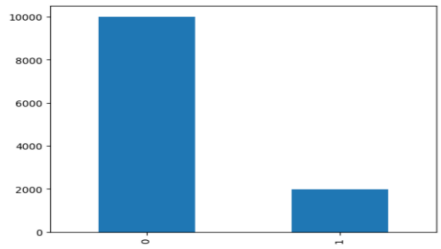
İnsan kaynakları veri setinin istatistiksel olarak betimlenmesi ile ilgili bilgiler Tablo 2'de gösterilmektedir.

Tablo 2. İnsan Kaynakları Veri Setinin İstatistiksel Betimlenmesi

	mean	std	min	%25	%50	%75	max
satisfaction_level	0.630	0.241	0.090	0.480	0.660	0.820	1.000
last_evaluation	0.717	0.168	0.360	0.570	0.720	0.860	1.000
number_project	3.803	1.163	2.000	3.000	4.000	5.000	7.000
average_monthly_hours	200.474	48.728	96.000	157.000	200.000	243.000	310.000
time_spend_company	3.365	1.330	2.000	3.000	3.000	4.000	10.000
work_accident	0.154	0.361	0.000	0.000	0.000	0.000	1.000
left	0.166	0.372	0.000	0.000	0.000	0.000	1.000
promotion_last_5years	0.017	0.129	0.000	0.000	0.000	0.000	1.000

Tablo 2'de veri dağılımının normal bir dağılımda olduğu gözlemlenmektedir. Veri setinde tekrar eden 3008 gözlem, daha sağlıklı bir öğrenmenin gerçekleşebilmesi için veri setinden silinmiştir. Veri seti içinse Object veri türü olarak bulunan veri tipleri için Category veri tipi olarak değiştirilmiş buna göre ordinal kategorik olarak dönüşümü yapılmıştır.

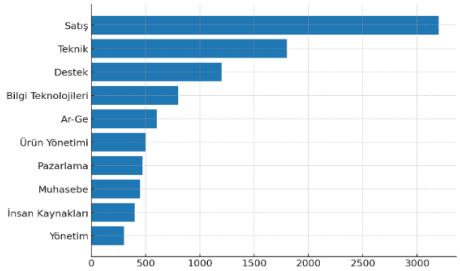
Çalışanlar ve işten ayrılanlar ile ilgili dağılım, şekil 1'de gösterilmektedir.



Şekil 1. Çalışan ve İşten Ayrılanların Grafik Gösterimi

Şekil 1'de, işletmede yaklaşık 10.000 kişinin çalışmakta olduğu, 1.991 kişinin işten ayrıldığı tespit edilmiştir. İşletmenin işten ayrılma oranı %16,604 olarak belirlenmiştir. Genellikle, düşük bir işten ayrılma oranı beklenir çünkü yüksek oranlar işletme için maliyetli olur ya da işletmenin istikrar eksikliğini gösterir. Çoğu sektörde, %10'un altındaki işten ayrılma oranları tercih edilirken, bazı sektörlerde daha yüksek oranlar kabul edilir. Örneğin, hızla büyüyen şirketlerde işten ayrılma oranları daha yüksektir, çünkü çalışanlar yeni fırsatlar aramakta, daha iyi teklifler almaktadır.

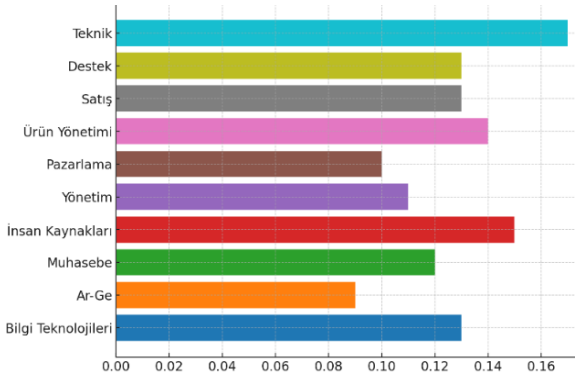
Departmanlara göre çalışan sayıları ile ilgili dağılım şekil 2'de gösterilmektedir.



Şekil 2. Departmanlarına Göre Çalışan Sayıları

Şekil 2'de görüldüğü gibi, 3.000 kişiden fazla çalışanı ile en fazla çalışan sayısına sahip departmanın satış bölümünde olduğu; 500'den az çalışan sayısı ile de en az çalışan departmanın yönetim bölümünde olduğu tespit edilmiştir.

Departmanlar kırılımında işten ayrılanlar ilgili dağılım Şekil 3'de yer almaktadır.



Şekil 3. Departmanlar Kırılımında İşten Ayrılanlar

Şekil 3'te 0.16'dan daha fazla bir oranla teknik departmanda çalışan kişilerin işten ayrılma oranı yüksek gözlemlenmiştir. Ar-Ge ve bilgi teknolojileri departmanındaki çalışanların ise işten ayrılma oranları daha düşüktür. Sayısal değişkenlerin hedef değişkenle olan korelasyonuna bakılmış ve işten ayrılma ile sonuçlanan durumların veri setindeki diğer değişkenlerle arasında bulunan pozitif ve negatif korelasyon incelenmiştir. Bu sayede modelde optimal sayıdaki değişkenlerin kullanılmasına olanak sağlanmıştır.

4.2. Model Geliştirme Süreci

Bu bölüm model geliştirme süreci aşamasını içermektedir. Model eğitiminde hedef değişkenimiz olan ve işten ayrılma durumunu ifade eden "left" değişkeniyle ilişkisi bulunan diğer veriler belirlenerek düzenlenmiştir. Veriler standart hale getirildikten sonra train test split metoduyla veri setini ikiye bölerek %20 lik kısmı test, %80'lik kısmı eğitim seti olarak kullanılmıştır. Ayrıca modelin doğruluğu kontrol etmek için 10 kez test edilerek sonuçlar incelenmiştir.

Çalışmada kullanılan XGBoost, makine öğrenimi alanında sınıflandırma ve regresyon problemleri için tercih edilen güçlü bir algoritmadır. XGBoost, gradient boosting yöntemiyle, modelin genel performansını optimize etmek amacıyla önemlilik sırasına göre seçilen değişkenlerle etkili bir model geliştirilmiştir. Bu yöntem, zayıf öğrencileri bir araya getirerek daha güçlü bir tahmin modeli oluşturur. Gradient boosting, hata fonksiyonunun gradyanını kullanarak her adımda yeni bir öğrenici eğitir ve bu öğreniciler, önceki hataları düzeltmeye yönelik eklenir. Model, overfitting problemini önlemek amacıyla L1 ve L2 regularizasyon tekniklerini uygularken, paralel işlemciler ve GPU'ları kullanarak hesaplama hızını artırır. Bu özellikler, XGBoost'un büyük ölçekli veri setleri üzerinde hızlı ve etkili çalışmasını sağlamaktadır.

XGBoost, GBM (Gradient Boosting Machine) algoritmasının optimize edilmiş bir versiyonu olarak tanımlanır ve yüksek tahmin gücü, aşırı öğrenmeyi önleme, boş verileri yönetme ve hızlı çalışma gibi avantajlarıyla öne çıkar. Chen ve Guestrin

(2016: 785) tarafından geliştirilen bu algoritmanın, diğer popüler algoritmalardan daha hızlı çalıştığı belirtilmiştir. XGBoost sınıflandırma raporu şekil 4'te gösterilmektedir.

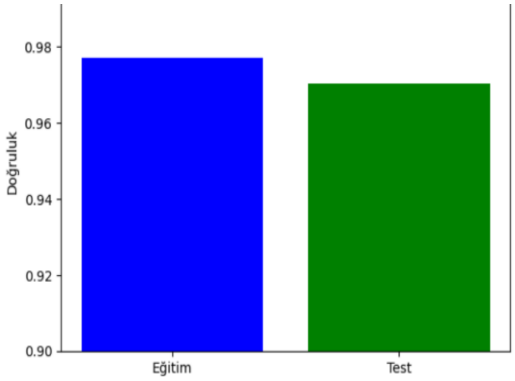
Sınıflandırma Raporu:

	precision	recall	f1-score	support
0	0.98	0.98	0.98	1998
1	0.92	0.90	0.91	401
accuracy			0.97	2399
macro avg	0.95	0.94	0.95	2399
weighted avg	0.97	0.97	0.97	2399

Şekil 4. Sınıflandırma Raporu

Şekil 4'te model incelendiğinde, doğruluk (accuracy) değeri %97 olarak belirlenmiştir. Sınıf 0 için precision, recall ve F1-score %98 olarak hesaplanmış, bu da modelin sınıf 0'ı doğru şekilde sınıflandırma yeteneğinin yüksek olduğunu göstermektedir. Sınıf 1 için ise precision %92, recall %90 ve F1-score %91 elde edilmiştir, bu da sınıf 1'in sınıf 0'a kıyasla biraz daha düşük bir performans sergilediğini ortaya koymaktadır. Bu sonuçlar, modelin genel olarak dengeli ve yüksek performans sunduğunu göstermektedir.

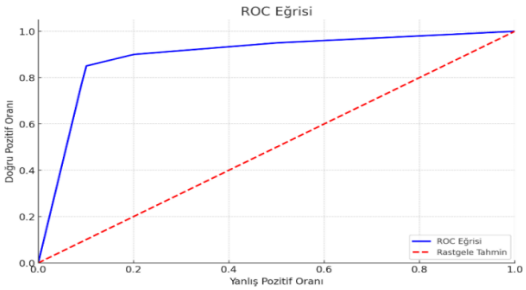
Özellik önemlilik skorlarına göre, modelin içindeki özelliklerin hangi değer aralığında olduğu ve model tahmininde hangi özelliğin kaç oranında etkilediği incelenmiştir. Şekil 5, eğitim ve test doğruluk dağımını göstermektedir.



Şekil 5. Eğitim ve Test Doğruluğu

Şekil 5'te görüldüğü üzere, modelin eğitim ve test seti arasındaki doğruluk dağılımı iyi durumdadır.

Sınıflandırma modelinin performansını ölçmek için kullanılan AUC değeri, ROC eğrisinin altında kalan alanı ifade etmektedir. Bu değer ile ilgili oranlar, Şekil 6'da gösterilmektedir.



Şekil 6. ROC Eğrisi

Şekil 6'ya göre, AUC değeri 1'e ne kadar yakınsa modelin performansı o kadar iyidir.

AUC değeri şu şekilde yorumlanır:

1.0: Mükemmel bir sınıflandırma modeli

0.9 - 0.99: Çok iyi

0.8 - 0.89: İyi

0.7 - 0.79: Orta

0.6 - 0.69: Zayıf

0.5 - 0.59: Rastgele tahmin etmekle aynı

Model geliştirme sürecinde, her bir algoritma için on katlı çapraz doğrulama yöntemi uygulanarak model başarımı değerlendirilmiş ve çapraz doğrulama skorları ile ortalama doğruluk oranları belirlenmiştir. Sonuçlara göre, en yüksek performansı gösteren algoritma seçilmiş ve bu algoritma ile model kurulmuştur. Tablo 3'te her bir algoritmanın model başarım sonuçları yer almaktadır.

Tablo 3. Karşılaştırma Skorları

XGBoostClassifier	
Çapraz Doğrulama	[0.975, 0.98020833, 0.96767466, 0.96976017, 0.97914494, 0.97184567, 0.97288843, 0.98123045, 0.96871741, 0.97705944]
Ortalama Doğruluk	0.9743529501216545
Random Forest Classifier	
Çapraz Doğrulama	[0.97291667, 0.97708333, 0.96558916, 0.9645464, 0.97393118, 0.96767466, 0.96663191, 0.97705944, 0.96663191, 0.97601668]
Ortalama Doğruluk	0.970808133472367
Gradient Boosting Classifier	
Çapraz Doğrulama	[0.97395833, 0.97916667, 0.96976017, 0.96767466, 0.97810219, 0.96976017, 0.97080292, 0.9801877, 0.96871741, 0.97914494]
Ortalama Doğruluk	0.9737275156412929
K Neighbors Classifier	
Çapraz Doğrulama	[0.97291667, 0.978125, 0.9645464, 0.96976017, 0.97705944, 0.97184567, 0.96976017, 0.97810219, 0.96871741, 0.97810219]
Ortalama Doğruluk	0.9728935305874176
Logistic Regression	
Çapraz Doğrulama	[0.81666667, 0.79375, 0.81438999, 0.8164755, 0.81334724, 0.82377477, 0.80813347, 0.80500521, 0.79353493, 0.81751825]
Ortalama Doğruluk	0.810259602015989

Tablo 3'e göre, XGBoost algoritması, %97.43 ortalama doğruluk oranı ile en yüksek performansı göstermiştir ve modelin tahmin gücünün oldukça yüksek olduğunu ortaya koymaktadır. Gradient Boosting (%97.37) ve K-Nearest Neighbors (%97.28) algoritmalarının da yüksek doğruluk değerlerini sağladığı görülse de, XGBoost'un küçük bir farkla daha iyi bir tahmin gücü sergilediğini söylemek mümkündür. Diğer yandan, Logistic Regression algoritması %81.02 ortalama doğruluk oranı ile diğerlerine daha göre düşük performans göstermiştir.

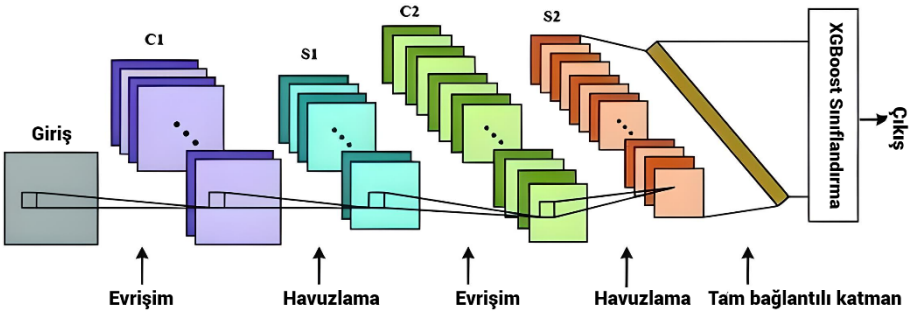
4.3. Önerilen Hibrit Model Yaklaşım

Veri işleme sürecinin çoğu, veriye ilk dokunuştan model eğitimine kadar olan aşamada gerçekleşir. Bu süreçte geliştirici, veriyi anlamaya, işlemeye ve model eğitimi için uygun hale getirmeye çalışır. Veri bilimi ve makine öğrenimi alanındaki yenilikler, karmaşık veri setlerinden anlamlı bilgiler çıkarmak için çeşitli algoritmaların birleştirilmesini sağlar.

CNN, bir resmi küçük parçalara ayırarak temel özellikleri bulur; bu işlem evrişim olarak adlandırılır ve resim çeşitli filtrelerle taranır. İlk katmanlar basit özellikleri çıkarırken, derin katmanlar daha karmaşık özellikleri tanır. Akıllı telefonlar, güvenlik kameraları ve otomatik etiketleme sistemleri gibi teknolojilerde yaygın olarak kullanılır.

XGBoost, verileri analiz etmek ve tahminler için kullanılan yapay zeka algoritmasıdır. Algoritma, verilerdeki desenleri ve ilişkileri öğrenerek tahminler yapar. XGBoost, kredi risklerinin tahmini, dolandırıcılık tespiti, piyasa trend analizi, hastalıkların erken teşhisi, müşteri davranış analizi, hedef reklamlar, satış tahminleri ve spor analitiği gibi birçok alanda kullanılmaktadır.

CNN ve XGBoost gibi yöntemleri bir araya getirmek, tahminleri yüksek oranda doğru hale getirir ve daha etkili modeller oluşturmaya yardımcı olur. Bu iki yöntemi birleştirerek, CNN modelinin resimlerden bilgi çıkarma yeteneğinden ve XGBoost algoritmasının güçlü tahmin yaparak doğru tahminler elde edilmesine yardımcı olur. Şekil 7, CNN – XGBoost Mimarisini göstermektedir.



Şekil 7. CNN – XGBoost Mimarisi

Şekil 7'ye göre, CNN'in özellik çıkarımı kabiliyeti ve XGBoost'un yüksek doğruluklu tahmin yapma yeteneği, hibrit modelin başarılı bir şekilde çalışmasını sağlayacaktır. Bunun sağlayacağı yüksek doğruluk oranı, modelin sınıflandırma ve tahmin süreçlerinde etkin bir şekilde kullanılmasını desteklemektedir.

CNN algoritması için import edilen kütüphaneler şunlardır: import numpy as np, from sklearn.compose import ColumnTransformer, import pandas as pd, from sklearn.metrics import classification_report, confusion_matrix, import xgboost as xgb, from tensorflow.keras.models import Sequential, from sklearn.model_selection import train_test_split, from tensorflow.keras.layers import Conv1D, MaxPooling1D, Flatten, from sklearn.preprocessing import StandardScaler, OneHotEncoder.

İnsan kaynakları veri seti üzerinde bir CNN ve XGBoost tabanlı hibrit model oluşturulmuştur. Veri seti yüklendikten sonra, özellikler (X) ve hedef değişken (Y) ayrılmış ve veri %80 eğitim, %20 test olarak bölünmüştür. Sayısal özellikler standartlaştırılırken kategorik değişkenler one-hot-enconding ile dönüştürülüp bir araya getirilmiştir. CNN modeli 3 boyutlu girdi beklediğinden veri seti 3 boyutlu bir tensöre dönüştürülmüştür. CNN modeli için evrişim katmanı, maksimum havuzlama katmanı ve düzleşme katmanından oluşmaktadır. Eğitilen model, test verisi üzerinde tahminlerde bulunarak ve performansı, sınıflandırma raporu ve karışıklık matrisi ile değerlendirilmiştir. Şekil 8, sınıflandırma raporunu göstermektedir.

Classification Report:

	precision	recall	f1-score	support
0	0.99	0.99	0.99	2294
1	0.98	0.96	0.97	706
accuracy			0.99	3000
macro avg	0.98	0.98	0.98	3000
weighted avg	0.99	0.99	0.99	3000

Şekil 8. Sınıflandırma Raporu

Şekil 8'e göre model, yüksek oranda bir doğruluk ve yüksek oranda bir performans sergilediğini göstermektedir. Hem sınıf 0 hem de sınıf 1 için yüksek keskinlik ve F1 skorları, modelin pozitif sınıflandırıldığını gösterir.

Şekil 9, hata matrisini göstermektedir. Hata matrisi, bir sınıflandırma modelinin performansını değerlendirmek için kullanılır ve iki sınıf arasındaki tahminlerin doğruluğunu gösterir.

```
Confusion Matrix:
[[2279   15]
 [  28  678]]
```

Şekil 9. Hata Matrisi

Şekil 9'daki hata matrisi raporuna göre, gerçek negatiflerin sayısı 2279, yanlış pozitiflerin sayısı 15, yanlış negatiflerin sayısı 28, gerçek pozitiflerin sayısı 678 hesaplanmıştır. Bu matrise göre modelin performansı iyi seviyededir. Gerçek negatifler ve pozitifler yüksek, yanlış negatif ve pozitifler iyi düzeydedir. Veri setine ait iki bine yakın gözlem 1 sınıfından oluştuğu göze alındığında son derece yüksek ve kabul edilir bir performans gözlemlenmiştir. Diğer taraftan geliştirilen bu hibrit model ile XGBoost algoritma sonuçları kıyaslandığında Accuracy değerinde %2.06 artış oranı gözlemlenmiştir. Bu da geliştirilen hibrit model ile geleneksel algoritmaların tahmin başarıları önemli ölçüde arttırmaktadır.

5. Sonuç, Tartışma ve Öneriler

Bu çalışmada, çalışan devir hızını tahmin etmek için CNN ve XGBoost algoritmalarını birleştiren hibrit bir model önerilmiştir. Model, Kaggle platformundan elde edilen sentetik bir insan kaynakları veri seti üzerinde eğitilmiş ve test edilmiştir. Elde edilen bulgular, hibrit modelin %99 doğruluk oranı ile çalışan devir hızını tahmin etmede yüksek performans sergilediğini göstermiştir. Ayrıca, modelin hassasiyet, geri çağırma ve F1 skoru gibi metriklerde de başarılı sonuçlar verdiği görülmüştür. Bu sonuçlar, hibrit modelin çalışan devir hızını tahmin etmede etkili bir araç olabileceğini ortaya koymaktadır.

Çalışan devir hızı, işletmeler için önemli bir sorun olmaya devam etmekte ve bu sorunun çözümüne yönelik çeşitli araştırmalar yapılmaktadır (Yılmaz ve Halıcı, 2010; Kaptanoğlu, 2020). Bu çalışma, özellikle makine öğrenimi tekniklerinin çalışan devir hızını tahmin etmedeki etkinliğini inceleyen önceki çalışmalarla benzer sonuçları ortaya çıkarmıştır. Saradhi ve Palshikar (2011), çalışan devir hızını tahmin etmek için makine öğrenimi algoritmalarını kullanmış ve mevcut çalışma ile benzer şekilde yüksek doğruluk oranları elde etmiştir. Dolatabadi ve Keynia (2017) çalışmasında, çalışan devir hızını tahmin etmek için veri madenciliği yöntemlerini kullanmış ve bu yöntemlerin etkinliğini vurgulamıştır.

Bu çalışmada önerilen hibrit model, özellikle CNN ve XGBoost algoritmalarının birleştirilmesiyle daha yüksek doğruluk oranları elde etmiştir. Bu sonuçlar, özellikle derin öğrenme ve gradient boosting tekniklerinin birleştirilmesinin, çalışan devir hızı tahmininde daha etkili sonuçlar verebileceğini göstermektedir. Özmen ve Özcan (2021) tarafından yapılan bir çalışmada CNN tabanlı bir modelin çalışan devir hızını tahmin etmede başarılı olduğu belirtilmiştir ve bu çalışmanın sonuçlarıyla benzer olduğu tespit edilmiştir. Fakat bu çalışma ile olan farkı, onların çalışmasında elde edilen hibrit modelin, daha yüksek doğruluk oranlara sahip olduğudur.

Ayrıca, bu çalışmanın bulguları, işletmelerin çalışan devir hızını tahmin etmek için veri odaklı yaklaşımları benimsemelerinin önemli olduğunu vurgulamaktadır. Özellikle, çalışan memnuniyeti, performans değerlendirmeleri ve kıdem gibi faktörlerin çalışan devir hızını tahmin etmede önemli değişkenler olduğu görülmüştür. Söz konusu bulgular, Holtom vd., (2008)'nin yaptığı çalışma ile uyumludur. Çalışma, çalışan memnuniyetinin, çalışanların işten ayrılma eğilimlerinde etkili olduğunu ortaya koymuştur.

Bu çalışmanın bulguları doğrultusunda, çalışan devir hızını azaltarak, işletmelerdeki insan kaynakları yönetimini iyileştirmeye yardımcı olabilecek öneriler sunmak mümkündür. İşletmeler, çalışan devir hızını tahmin etmek ve yönetmek için makine öğrenimi ve teknoloji tabanlı özellikle yapay zeka modelleri kullanarak, veri odaklı insan kaynakları yönetimini benimseyebilir. Bu da çalışanların işten ayrılma eğilimlerinin önceden tahmin edilmesine olanak sağlayarak, gerekli önlemlerin alınmasına yardımcı olabilir. İşletmeler, potansiyel ayrılma riski taşıyan çalışanları önceden tespit ederek, bu çalışanlarla ilgili özel stratejiler geliştirme ve iş gücü maliyetlerini optimize etmeyi dikkate almalıdır. Bu da işgücü verimliliğini artırarak operasyonel verimliliği ve iş memnuniyetini sağlamada kritik bir rol oynar. Ayrıca, doğru bir çalışan devir hızı tahmini modeli, işletmelere rekabet avantajı sağlar. Model, veri odaklı kararlar almayı destekleyerek, işletmelerin operasyonel maliyetlerini düşürmelerine, verimliliği artırmalarına ve çalışan memnuniyetini sağlamalarına yardımcı olur. Çalışanların memnuniyetlerini arttırmak için, düzenli anketler ve geri

bildirimlerin kullanılması önerilmektedir. Ayrıca, çalışanların kariyer gelişimlerini destekleyecek eğitim programları gerçekleştirilmektedir. Bu şekilde, işletmeler rekabet gücünü artırarak gelecekteki başarılarını desteklerler. Bu bakımdan geleneksel tahmin modelleri ile verinin işleme süreçleri oldukça uzun sürmekteyken, hibrit modeller ile veri işleme süreçleri kısaltılabilmektedir. Modellerin şeffaf ve adil ve şekilde kullanılmasının sağlanmalı, etik riskler dikkate alınmalıdır. Statik veri setlerinden ziyade, gerçek ve güncellenebilir veri akışı sağlayan modellerin kullanılması önerilir. Böylece işletmeler gerçek zamanlı verilerle modellerini eğitirken, modelin de uyumluluğunu geliştirebilir.

Gelecekte yapılacak çalışmalarda, farklı endüstrilerden elde edilen gerçek veri setleri üzerinde bu tür hibrit modellerin test edilmesi ve modellerin performansının daha detaylı bir şekilde incelenmesi önerilmektedir. Araştırmacılar, farklı sektörlerdeki çalışan devir hızını analiz etmek için daha geniş çaplı çalışmalar gerçekleştirebilir. Çalışan devir hızını etkileyen çalışma koşulları ve iş yaşam dengesi gibi faktörlerin modele dahil edilmesiyle daha kapsamlı bir analiz yapılması önerilebilir. Gerçekleştirilecek bu tür çalışmalar, işletmelerin çalışan devir hızını daha etkin bir şekilde yönetmelerine ve sürdürülebilir rekabet avantajı elde etmelerine yardımcı olabilir.

Referanslar

Aalst, W., Damiani, E. (2015). Processes meet big data: connecting data science with process science. *IEEE Transactions on Services* Computing, *8*(6), 810-819. http://dx.doi.org/10.1109/TSC.2015.2493732

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.

Altun, M. G., Altun, A. H. (2025). Yüksek performanslı betonun basınç dayanımının farklı makine öğrenimi algoritmaları ile tahmin edilmesi. *Journal of Innovative Engineering and Natural Science*, *5*(1), 347-361. https://doi.org/10.61112/jiens.1555284

Aksu, S. G., Sürgevil, O. (2019). Dijital çağın yetkinlikleri: Çalışanlar, insan kaynakları uzmanları ve yöneticiler çerçevesinden bakış. *Journal of Business in the Digital Age*, 2(2), 54-68.

Bahadır, M. B., Bayrak, A. T., Yücetürk, G., Ergun, P. (2021, June). A comparative study for employee churn prediction. In *2021 29th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

Ben Jabeur, S., Stef N., Carmona P., (2023). Bankruptcy prediction using the XGboost algorithm and variable importance feature engineering. *Computational Economics, 61*(2), 715-741. http://dx.doi.org/10.1007/s10614-021-10227-1

Bilenler, B., Gül, S., Uçar, T. (2024). Makine öğrenmesi teknikleri ile işten ayrılacak personelin tahminlenmesi ve tekniklerin performanslarının karşılaştırılması. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, *15*(4), 807-816.

Büyükkör, Y., Doğan, S. (2024). Borsa endeks yönünün ağaç tabanlı topluluk makine öğrenmesi yöntemleri ile tahmini: bist-100 örneği. *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,* (27), 324-335. https://doi.org/10.29029/busbed.1391790

Chen T., Guestrin C. (2016). XGBoost: A scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. http://dx.doi.org/10.1145/2939672.2939785

Coşkun, C., Baykal, A. (2011). Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması. *Akademik Bilişim*, *11*, 51-58.

Çiftçioğlu, B. A., Mutlu, M., & Katırcıoğlu, S. (2019). Endüstri 4.0 ve insan kaynakları yönetiminin ilişkisi. *Bandırma Onyedi Eylül Üniversitesi Sosyal Bilimler Araştırmaları Dergisi*, *2*(1), 31-53.

Del Prado, FLE., Roselloni MAD. (2017). Developing technological capability through human resource management: case study from the Philippines. *Asian Journal of Technology Innovation.* http://dx.doi.org/10.1080/19761597.2017.1385973

Dolatabadi S. H., Keynia F. (2017). Designing of customer and employee churn prediction model based on data mining method and neural predictor. *The 2nd International Conference on Computer and Communication Systems.* http://dx.doi.org/10.1109/CCOMS.2017.8075270

Dündar, K., Ağaçayak, A. (2023*).* Yapay zekâ ve makine öğrenmesi ile insan ilişkileri analizi. S. Neşeli, H. Terzioğlu (Ed.) *Mühendislikte Yenilikçi Yaklaşımlar-2* içinde (ss.123-144), İstanbul: Eğitim Yayınevi.

Ersöz, F., Çınar, Y. (2021). Veri madenciliği ve makine öğrenimi yaklaşımlarının karşılaştırılması: Tekstil sektöründe bir uygulama. *Avrupa Bilim ve Teknoloji Dergisi*, (*29*), 397-414.

Fırat, H., Hanbay, D. (2022). Comparison of 3D CNN based deep learning architectures using hyperspectral images. *Journal of the Faculty of Engineering and Architecture of Gazi University*, *38*(1).

Fonseca, F. (2021). Data objects for knowing. *AI & Society*. http://dx.doi.org/10.1007/s00146-021-01150-y

Gökalp, Ö. M. (2020). Makine öğrenmesi. Erişim adresi: https://www.researchgate.net/publication/357974983_Makine_Ogrenmesi_-_Machine_Learning.

Gümüş, H. T., Eyüpoğlu, C. (2023). Grafik sinir ağlarına genel bir bakış. *EMO Bilimsel Dergi*, *13*(2), 39-56.

Holtom, B. C., Mitchell, T. R., Lee, T. W., Eberly, M. B. (2008). Turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future. *The Academy of Management Annals*, 2(1), 231 274. https://doi.org/10.1080/19416520802211552

Kambur, E. (2022). Yapay zeka çağında insan kaynakları yönetimi konusunda yazılmış Türkçe makaleler üzerine bir araştırma. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* (*48*), 139-152. https://doi.org/10.30794/pausbed.872606

Kaptanoglu, R. Ö. (2020). İşten ayrılma niyeti ve toksik liderliğin etkisi. *IBAD Sosyal Bilimler Dergisi*, (6), 161-173. https://doi.org/10.21733/ibad.621500

Mahlasela, S., Chinyamurindi, WT (2020). Teknolojiyle ilgili faktörler ve işten ayrılma niyetleri üzerindeki etkileri: Güney Afrika'daki devlet çalışanları

örneği. *Gelişmekte Olan Ülkelerde Bilgi Sistemleri Elektronik Dergisi*, *86*(3), e12126.

Meade, J. G. (2003). The human resources software handbook: Evaluating technology solutions for your organization. *San Francisco: JosseyBass/Pfeiffer.*

Misican, D.Ö. (2020). İnsan kaynakları profesyonellerinin perspektifinden dijitalleşen çalışma hayatında yapay zekâ. *Journal of Academic Value Studies, 6*(2) 152-175. http://dx.doi.org/10.29228/javs.42120.

Nacar, E. N., Erdebilli (b.d.rouyendegh), B. (2021). Makine öğrenmesi algoritmaları ile satış tahmini. *Endüstri Mühendisliği*, *32*(2), 307-320. https://doi.org/10.46465/endustrimuhendisligi.811183

Oğuzlar, A. (2003). Veri ön işleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (21).

Özmen EP., Özcan T., (2021). A novel deep learning model based on convolutional neural networks for employee churn prediction. *Journal of Forecasting* 41(3), 539-550. http://dx.doi.org/10.1002/for.2827

Park J., Feng YT., Jeong SP., (2024). Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques. http://dx.doi.org/10.1038/s41598-023-50593-4

Pelenk, S.E. (2020). İnovatif (yenilikçi) insan kaynakları uygulamalarının yenilik kültürüne etkisi. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, (*Özel Ek*), 19 Temmuz 2020 237-261.

Saradhi, V.V. ve Palshikar, G.K. (2011). Employee churn prediction. *Expert Systems with Applications, 38,* 19-30. http://dx.doi.org/10.1016/j.eswa.2010.07.134

Sathya, R., Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. International *Journal of Advanced Research in Artificial Intelligence*, *2*(2), 34–38, http://dx.doi.org/10.14569/IJARAI.2013.020206

Şahin, D., Yılmaz, S. (2021). Endüstri 4.0 uygulamalarının sağlık kurumlarında insan kaynakları yönetimine etkilerinin değerlendirilmesi. *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi*, 7(1), 142-155.

Şeker, Ş. E. (2018). CRISP-DM: Endüstriler arası standart işleme–veri madenciliği için (cross ındustry standard processing–data mining). *YBS Ansiklopedi*, 5(2).

Tharani, S. M., Raj, S. V. (2020, October). Predicting employee turnover intention in IT&ITeS industry using machine learning algorithms. *In 2020 Fourth International Conference on I-SMAC* (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 508-513). IEEE.

Tokmak, M. (2023). XGBoost algoritması ile ikili parçacık sürü optimizasyonu öznitelik seçme tabanlı jar kötü amaçlı yazılımlarının tespiti. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, *10*(1), 140-152.

Tonbil, D. D., Aksakal, N. Y. (2024). İşe alım, temin-seçim süreçlerinde yapay zekâ ve teknolojilerinin kullanımı: nitel bir araştırma. *İstanbul Ticaret Üniversitesi Girişimcilik Dergis*i, *7*(15), 38-56. https://doi.org/10.55830/tje.1514895

Tunçer, P. (2012). Değişen insan kaynakları yönetimi anlayışında kariyer yönetimi. *Ondokuz Mayis University Journal of Education Faculty*, *31*(1). https://doi.org/10.7822/egt91

Turan, G. (2017). *Muhasebe meslek mensuplarında iş yaşamı kalitesi ve işten ayrılma niyeti ilişkisi*. Yüksek Lisans Tezi, Gaziantep Üniversitesi.

Türkel, S., Bozagaç, F. (2018). Endüstri 4.0' in insan kaynakları yönetimine etkileri. *Toros Üniversitesi İİSBF Sosyal Bilimler Dergisi*, 5(9), 419-441.

Walker, A. J. (2001). *Web-based human resources.* New York: McGraw-Hill.

Wang SY., Huang Y., Cao GQ., (2023). Review on functional data classification. *Wiley Interdisciplinary Reviews – Computainal Statics 16*(1). http://dx.doi.org/10.1002/wics.1638

Xu, XX., Ji, Y. (2013). Study on technology innovation and human resources reallocation. *Twldth Wuhan International Conference on E-Business*. 677-682.

Yıldız, K. (2023). *İstatistiksel yöntemler ile çalışanın işten ayrılmasının tahminlenmesi ve ücret belirlenmesi ile işten ayrılmanın optimizasyonu*. Doktora Tezi, Marmara Üniversitesi.

Yılmaz, B., Halıcı, A. (2010). İşgücü devir hızını etkileyen etmenler: Sekreterlik mesleğinde bir araştırma. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (4).

Zhong, HP., Fang, RS., Li, XY., Sun, XQ. (2008). Human resource slack and technological innovation: Evidence from Henan province in China. *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing,* Vols 1-31, 7065-7068.