



Makine Öğrenmesi Teknikleri ile İşten Ayrılacak Personelin Tahminlenmesi ve Tekniklerin Performanslarının Karşılaştırılması

Prediction of Staff Turnover by Machine Learning Techniques and Comparison of the Performances of Techniques

Batuhan Bilenler^{1*}, Sait Gül², Tamer Uçar³

¹ Bahçeşehir Üniversitesi, İşletme Mühendisliği Bölümü, batuhan.bilenler@bahcesehir.edu.tr

ORCID: <https://orcid.org/0000-0002-7925-4876>

² Bahçeşehir Üniversitesi, İşletme Mühendisliği Bölümü, sait.gul@bau.edu.tr

ORCID: <https://orcid.org/0000-0002-6011-0848>

³ Bahçeşehir Üniversitesi, Yazılım Mühendisliği Bölümü, tamer.ucar@bau.edu.tr

ORCID: <https://orcid.org/0000-0002-9397-6656>

MAKALE BİLGİLERİ

Makale Geçmişi:

Geliş 17 Temmuz 2024
Revizyon 10 Ekim 2024
Kabul 22 Ekim 2024
Online 23 Aralık 2024

Anahtar Kelimeler:

Personel kayıp tahmini, makine öğrenmesi, insan kaynakları analitiği, veri analizi.

ÖZ

Şirketlerin sürdürülebilir başarısı için yetenekli insan kaynağını şirkette tutundurmak oldukça önemlidir. Bu çalışmada, araştırma yapmak amacıyla açık kaynak olarak Kaggle platformunda HR_comma_sep.csv ismiyle yer alan 14.999 personelin verileri kullanılarak işten ayrılacak olan personelin tahminlenmesi amaçlanmıştır. Gradyan Artırma, Rastgele Orman ve Ekstrem Gradyan Artırma regresyon teknikleri ile tahminleme yapılmıştır. Değerlendirme metrikleri olan ortalama mutlak hata (mean absolute error), hata karelerinin ortalaması (mean squared error), R² skoru ve düzeltilmiş R² (Adjusted R²) değerleri karşılaştırılmış olup 3 modelin de anlamlı ölçüde tahminleme yapabilir olduğu sonucuna ulaşılmıştır. Ortalama hata skorları birbirine oldukça yakın olduğu için R² değeri 1'e en yakın olan Rastgele Orman tekniği üzerinden özellik önemi çıkarılmıştır. İncelenen öznitelikler arasında, işten ayrılmayı etkileyen en önemli öznitelğin personel tatmini olduğu görülmüştür. Makine öğrenmesi tekniklerinin insan kaynakları alanında kullanımının, şirket içerisindeki yetenekli personeli şirkete tutundurma stratejilerini belirleme anlamında insan kaynakları yöneticilerine oldukça faydalı çıktılar üretebileceği düşünülmektedir.

ARTICLE INFO

Article history:

Received 17 July 2024
Received in revised form 10 October 2024
Accepted 22 October 2024
Available online 23 December 2024

Keywords:

Staff turnover rate, machine learning, human resources analytics, data analysis.

ABSTRACT

Retaining talented human resources within a company is crucial for sustainable success. This study aims to predict staff turnover using a company's personnel data shared as open source for research purposes. Predictions were made using Gradient Boosting Tree, Random Forest Trees, and XGBoosting Regression techniques. The evaluation metrics used were mean absolute error, mean squared error, R² score and adjusted R² values. It was concluded that all three models could make significant predictions. Since the mean error scores of the three techniques were quite close to each other, feature importance was derived from the Random Forest technique, which had an R² value closest to 1. Among the features examined, employee satisfaction was found as the most important factor affecting staff turnover. It is believed that the usage of machine learning techniques in human resources management could produce highly beneficial insights for human resources managers in terms of determining strategies for retaining talented employees.

Doi: 10.24012/dumf.1517176

* Sorumlu Yazar

Giriş

Bir şirketin hedeflerine ulaşması için insan kaynağı yönetimine önem vermesi gerekmektedir. Yöneticilerin, hızla değişen piyasa koşullarına göre dinamik şekilde karar alabilmeleri ama aynı zamanda şirketi hedeflerine taşıyacak personelin mutluluğunu göz ardı etmemeleri gereklidir. Özellikle anahtar rolde bulunan personeli şirkete tutundurma amacıyla şirketlerin üst yönetimleri çeşitli yöntemler uygulayarak personellerin tatminini artırıcı hamleler yapmaktadırlar.

Gün geçtikçe farklılaşan sosyal, ekonomik ve teknolojik koşullar, her alanda olduğu gibi insan kaynakları alanında da yeni ihtiyaçlar doğurmakta ve bu ihtiyaçlar yeni kavramların, teorilerin ve uygulamaların ortaya çıkmasına neden olmaktadır [1]. İşletmeler için önemli değerlerden birisi de kayıpların önceden doğru şekilde tahmin edilebilmesidir [2]. Müşterileri elinde tutma, sadece en iyi ve faydalı müşterileri rakiplerden korumak değil, aynı zamanda onlar arasındaki ilişki ve bağlantıları sürdürmek ve böylece hizmet kullanımını ve buna bağlı geliri korumaktır [3].

İnsan Kaynağı (İK) analitiği, insan kaynakları süreçlerinin her biri hakkında daha iyi içgörü sağlamak, ilgili verileri toplamak ve daha sonra bu süreçlerin nasıl iyileştirileceğine dair objektif kararlar almak amacıyla kullanılan süreç, hesaplama veya modellerdir. İnsan kaynakları analitiği alanında literatür taraması gerçekleştirildiğinde analitik çalışmaların, personel seçim problemi ve mevcut personelin ayrılıp ayrılmaması üzerine yoğunlaştığı görülmektedir. Personelin işten ayrılma (*staff turnover*) tahmini probleminin ele alınış yöntemi olarak farklı farklı yaklaşımlara başvurulduğu görülmektedir. Shankar ve diğ. [4] veri madenciliği teknikleri kullanarak personelin işten ayrılma problemini istatistiksel bakış açısı ile tahminlemeye çalışmıştır. Alao ve diğ. [5], karar ağacı yöntemini kullanarak geçmiş verileri analiz ederek personelin ayrılma tahmininin yapılabileceğini göstermiştir. Abdulsalam ve diğ. [6], rastgele orman (random forest) ve convolution neural network algoritmaları kullanarak performans analizi gerçekleştirmiş ve trend olan iki yöntemi derinlemesine inceleyerek performans karşılaştırması yapmıştır. Gao ve diğ. [7] ise 2009 yılında rastgele orman yönteminin diğer makine öğrenmesi yöntemlerinden daha başarılı sonuçlar elde ettiğini gösteren bir karşılaştırmalı analiz gerçekleştirmişlerdir.

Tablo 1'de insan kaynağı analitiği alanında çözülen problemler ve bu problemleri çözme yöntemleri sınıflandırılmıştır. Görüldüğü gibi yazında çoğunlukla, personelin işten ayrılması ve personelin işe alımı konularında yoğunlaşmış olup, yöntemsel açıdan makine öğrenmesi tekniklerinin tercih edildiği görülmektedir.

Tablo 1. Makine Öğrenmesi ile İK Analitiğinde Yapılan Sınıflandırma ve Regresyon Çalışmaları

Çalışma-Yılı	Çalışmanın Türü	Değerlendirme Metrikleri	Çalışma Yapılan Veri seti
[8]- 2018	Sınıflandırma	Doğruluk oranı, Keskinlik, Duyarlılık, F1 değeri	Çalışmada kullanılan veri seti IBM HR Analytics Employee Attrition and Performance veri seti olarak belirtilmiştir.
[9]- 2019	Sınıflandırma	Doğruluk oranı, ROC Eğrisi altındaki alan, F1 değeri	Çin'de görev yapan bir Telekom şirketinin personel veri seti kullanılmıştır.
[10]- 2019	Sınıflandırma	Doğruluk oranı, F1 değeri, ROC Eğrisi altındaki alan	Çalışmada kullanılan veri seti IBM HR Analytics Employee Attrition and Performance veri seti olarak belirtilmiştir.
[11]- 2021	Sınıflandırma	Doğruluk oranı, ROC Eğrisi altındaki alan, F1 değeri	IBM HR Analytics Employee Attrition and Performance veri setini kullanmışlardır.
[12]- 2018	Regresyon	Kök Ortalama Kare Hatası, Ortalama Mutlak Hata	Çalışmada kullanılan veri seti Kaggle üzerinde mevcut olan IBM HR Analytics Employee Attrition and Performance'dır
[13]- 2023	Sınıflandırma	Doğruluk oranı, F1 değeri, ROC Eğrisi altındaki alan	Çalışmada belirtilen veri seti Kaggle'daki IBM HR Analytics Employee Attrition and Performance veri setidir.
[14]- 2018	Sınıflandırma	Doğruluk oranı, Keskinlik, Duyarlılık	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
[15]- 2022	Sınıflandırma	Doğruluk oranı, Keskinlik, Duyarlılık, F1 değeri	Bir telekomünikasyon şirketine ait gerçek veriler kullanılmıştır, ancak şirket adı belirtilmemiştir.
[16]- 2013	Sınıflandırma	Doğruluk oranı, F1 değeri	Nijerya Yüksek Enstitüsünün 1978-2006 arasında çalışan personel listesi
[17]- 2020	Sınıflandırma	Doğruluk oranı, ROC Eğrisi altındaki alan	Kaggle platformundaki IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
[18]- 2018	Sınıflandırma	Doğruluk oranı, Keskinlik, Duyarlılık, F1 değeri	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
[19]- 2021	Sınıflandırma	Doğruluk oranı, F1 değeri, ROC Eğrisi altındaki alan	IBM HR Analytics Employee Attrition veri seti kullanılmıştır.
[20]- 2020	Sınıflandırma	Doğruluk oranı, Keskinlik, F1 değeri	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
[21]- 2023	Sınıflandırma	Doğruluk oranı, F1 değeri, Eğrinin altında kalan alan	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
[22]- 2022	Regresyon	Kök Ortalama Kare Hatası, Ortalama Mutlak Hata	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
[23]- 2021	Sınıflandırma	Doğruluk oranı, F1 değeri, ROC Eğrisi altındaki alan	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.

Çoğu çalışma IBM HR Analytics Employee Attrition and Performance veri setini kullanmıştır. Bu veri seti, Kaggle'da halka açık olarak bulunan ve yaygın olarak kullanılan bir veri

setidir. Özel veri kaynakları kullanılan çalışmalarda ise şirket ismi genellikle belirtilmemiştir.

Regresyon tabanlı yaklaşım ile personelin işten ayrılma tahmini ile ilgili yapılan bir çalışmada, ortalama hata karelerinin kökü, destek vektör regresyon yöntemi ile 0,671; karar ağacı regresyon yönteminde 0,624; rastgele orman yöntemiyle 0,614 bulunmuştur [22]. Bu üç teknik arasında en başarılı sonucu rastgele orman tekniği vermiştir. Yapılan başka bir çalışmada ise, rastgele orman tekniğinin doğruluk oranı 0,985; Adaboosting tekniğinin ki ise 0,984 bulunmuştur. F1 değerleri ise 0,982 olarak aynı bulunmuştur. Bu da sınıflandırma tabanlı işten ayrılma tahmininin her iki yöntemin de başarılı şekilde yapabileceğini göstermiştir [13].

Personelin işten ayrılma tahmini üzerine yöntemleri karşılaştıran bir çalışmada ise, en iyi sınıflandırma modelinin %97,5 doğruluk oranı ile rastgele orman olduğunu önermektedir. İkinci en iyi yöntem, %96,6 doğruluk oranı ile naïve bayes'tir ve en düşük doğruluk oranına sahip sınıflandırma modeli ise %88,7 ile karar ağacı olarak çıkmıştır [14]. Çalışma bu makine öğrenmesi teknikleri içerisinde, çalışan devrini tahmin etmek için en güvenilir ve doğru sınıflandırma modelinin rastgele orman algoritması olduğunu tespit etmiştir.

Başka bir çalışmada ise personelin ayrılma tahmini kNN ve rastgele orman tekniği ile ayrı ayrı yapılmış ve performanslar karşılaştırılmıştır. Rastgele orman %80 oranında doğruluk oranı verirken kNN %84 performans göstermiştir. ROC altında kalan alanlar karşılaştırıldığında ise rastgele orman tekniğinin daha çok alan kapladığını göstermiştir.

Literatürde yapılan çalışmalara bakıldığında zaman sınıflandırma problemi için doğruluk değeri, F1 skoru, duyarlılık değerlerine ağırlıklı olarak bakıldığını görmekteyiz. Regresyon probleminde ise, Ortalama Mutlak Hata, hata karelerinin ortalamasının karekökü, gerektiğinde R-kare değerine bakıldığı görülmektedir. Yapılan bu çalışma regresyon problemi olduğu için ortalama hata, mutlak hata, hata karelerinin ortalamasının karekökü ve R-kare değerlerine bakılmaktadır.

Personelin işten ayrılma tahmini üzerine yapılan ampirik çalışmalar, insan kaynakları yönetiminde büyük önem taşımaktadır. Bu çalışmalarda genellikle makine öğrenimi algoritmaları kullanılarak, çalışanların işten ayrılma eğilimlerini belirlemeye yönelik modeller geliştirilir. IBM HR Analytics gibi popüler veri setleri üzerinde çalışan araştırmacılar, çalışanların yaş, kıdem, maaş, departman, iş memnuniyeti ve eğitim gibi çeşitli özelliklerini analiz ederek ayrılma olasılığını tahmin etmeye çalışmışlardır.

Personelin İşten Ayrılma Tahmin Problemi

Literatürdeki çoğu çalışma, çalışan devrini iki farklı kategori ve ikisinin de altında ikiye farklı tür olmak üzere dört açıdan incelemektedir [24-25].

- *İşten Çıkarma Devri:* Bu tür devir, organizasyon tarafından başlatılır ve bireysel bir çalışanı hedefler. Sebepler ise, çalışanın disiplin ve/veya performans sorunlarıyla ilgili olabilir.

- *Küçülme Devri:* Bu tür devir, organizasyon tarafından başlatılır ve organizasyonel yeniden yapılanmanın bir parçası olarak meydana gelir (kaybedilen fon, iş gereksinimlerinin değişmesi, yeniden yapılanma gibi).
- *Kaçınılabılır Devir:* Bu tür devir, çalışan tarafından başlatılır ve organizasyon tarafından muhtemelen önlenebilir.
- *Kaçınılamaz Devir:* Bu tür devir, çalışan tarafından başlatılır ve emeklilik, ölüm gibi kaçınılmaz durumlarda meydana gelir.

Personelin İşten Ayrılmasını Etkileyen Parametreler ve Yapılmış Benzer Çalışmalar

Bir isteğe bağlı devir çalışmasının meta-analiz yaklaşımı ile gösterdiği gibi, bir çalışanın yaş, cinsiyet gibi demografik özellikleri ile maaş, kıdem, iş tatmini gibi birçok çalışma ortamı özelliklerinin devir tahmininde birlikte rol oynadığı görülmektedir [26]. Yaş, ekonomik faaliyet, kıdem, mevcut pozisyondaki çalışma süresi ve eğitim gibi değişkenlerin devir üzerindeki en güçlü göstergeler olduğunu belirten başka çalışmalar da vardır [27]. İş tatmininin çalışan devrinde anahtar rol oynadığını gösteren pek çok çalışma söz konusudur [27]-[33]. Çalışmada göz önüne alınan öznitelikler Tablo 2'de özellikleriyle birlikte verilmektedir.

İş bağlılığı kavramını kullanarak çalışanların işten ayrılma niyetlerini tahmin etmeye yönelik çoklu yöntemler kullanılır. Hangi özniteliklerin işten ayrılma kararını etkilediğini görebilmek amacıyla yapılan bir çalışmada yaş, cinsiyet, eğitim seviyesi gibi demografik faktörlerin iş bağlılığı ve işten ayrılma niyeti üzerindeki etkileri incelenmiştir. Çalışma sonucunda, çalışanların organizasyona ve topluluğa olan bağlılıklarının işten ayrılma niyetlerini önemli ölçüde azalttığı gözlemlenmiştir [34]. Başka bir çalışmada ise, çalışanların iş tatmini ve işten ayrılma niyeti arasındaki ilişkiye psikolojik güçlendirme faktörünün etkisi incelenmiştir. Psikolojik güçlendirme, iş tatmininin işten ayrılma niyeti üzerindeki etkisini moderatör olarak değerlendirmiştir. İş tatminini artırıcı önerilerde bulunulmuştur [35].

Sağlık sektöründe gerçekleştirilen bir başka çalışmada, sağlık çalışanlarında iş-aile çatışmasının işten ayrılma niyeti üzerindeki etkileri araştırılmıştır. İş-aile çatışmasının iş tatmini ve genel iş memnuniyeti üzerindeki olumsuz etkilerini değerlendirip, bu çatışmanın işten ayrılma niyeti ile olan ilişkisi irdelenmiştir. İş-aile memnuniyeti azaldıkça işten ayrılmanın arttığı sonucuna ulaşılmıştır [36].

Makine öğrenimi tekniklerinin işten ayrılma tahmini için nasıl kullanılabileceğine dair bir çalışmada ise, veri madenciliği ve makine öğrenimi algoritmalarını kullanarak çalışanların işten ayrılma niyetlerini tahmin etmeye yönelik çeşitli modeller sunulmuştur. Rastgele orman ve Gradyan artırma tekniklerinin diğer tekniklerden daha başarılı sonuç verdiği çalışmanın sonucunda ortaya konulmuştur [37].

Çalışan direncinin, işten ayrılma niyetini nasıl etkileyebileceği ve finans sektöründeki örneklerle bu etkinin nasıl değiştiğini araştıran başka bir çalışmada, anket üzerinden alınan verilerle yapısal eşitlik modeli kurulmuş ve yüksek çalışan dayanıklılığının, işten ayrılma niyetini

azaltıcı bir etkisi olduğu bulunmuştur. Dayanıklı çalışanlar, zorluklar karşısında daha dirençli oldukları için işlerinden ayrılma niyetleri daha düşük olmaktadır [38].

İş ortamında bireylerin mutlu olmasının, iş bağlılığı, iş tatmini ve duygusal organizasyonel bağlılığı etkilediği fikriyle yola çıkılmış olan başka bir çalışmada tüm organizasyonel seviyelerde, mutluluğun bireyler ve organizasyonlar için önemli sonuçları olduğu ve işe bağlılık ile iş tatmini arasında anlamlı bir ilişki olduğu ortaya çıkarılmıştır [39]. Başka bir çalışmada ise, 245 itfaiyeci ve onların süpervizörleri ile çalışan bağlılığını etkileyen özniteliklerin tespit edilmesi hedeflenmiştir. Bağlılığın, algılanan örgütsel destek ve temel öz değerlendirmeler görev performansı ve örgütsel vatandaşlık davranışı arasındaki ilişkilere aracılık ettiğine dair sonuçlar ortaya çıkmıştır [40].

Tükenmişlik, motivasyon ve performans kavramlarının iş ortamındaki kaynak miktarındaki artış sağlandığı takdirde nasıl etkileneceği sorusuna yanıt bulmak için yapılan bir çalışmada, çalışma ortamında iş kaynaklarındaki artışın, çalışanların bağlılık ve performansı üzerinde olumlu bir etkisi olduğu bulunmuştur. Yüksek iş tatmini, çalışanların işlerine karşı daha olumlu bir tutum sergilemesine ve bağlılıklarının artmasına neden olabilir sonucuna ulaşılmıştır. Bu durum, hem bireysel hem de örgütsel düzeyde verimliliği artırabilir öngörüsü desteklenmiş ve stresle başa çıkma duygusunun kuvvetlendirildiği, uzun çalışma sürelerinin azalacağına dair inancın artırıldığına ve böylece de işten ayrılma duygusunun azaldığı sonucuna ulaşılmıştır [41]. Esnek çalışma programlarının iş-yaşam dengesi, çalışan memnuniyeti ve genel iş performansı üzerindeki etkilerini incelemek üzere yapılan başka bir çalışmada, esnek çalışma düzenlerinin çalışanların iş stresini azaltma, iş tatmini artırma ve işe olan bağlılıklarını güçlendirme potansiyeline sahip olduğunu gösterilmiştir [42].

Bu çalışma kapsamında, Gradyan artırma, rastgele orman ve ekstrem Gradyan artırma regresyon teknikleri ile personelin işten ayrılma kararı alıp almayacağını ayrı ayrı tahminlenmesi ve kullanılan üç tekniğin performanslarının karşılaştırılması yapılmaktadır. Ayrıca personeli ayrılma kararına götüren en önemli özniteliklerin belirlenmesi de çalışmanın amaçlarından biridir.

Materyal ve Metot

Bu çalışmada personelin işten ayrılma tahmini için kullanılacak modeller oluşturulmadan önce veri analiz edilmiş olup birtakım veri ön işleme süreçleri işletilmiştir. Veri analizi ve modelleri Knime Analytic Tool aracı kullanılarak gerçekleştirilmiştir. Knime, veri analizi ve makine öğrenmesi yöntemlerinin kullanımına olanak sağlayan kullanıcı dostu bir analitik araçtır. İş parçacıklarını yapmakla görevli düğümleri hazır olarak sunması önemli bir özelliğidir. Düğümleri uç uca ekleme mantığı ile çalışmaktadır.

Bu yöntemlerin tercih edilmesindeki temel motivasyon ağaç yöntemlerinin ileri versiyonu olarak kabul edilen torbalama (bagging) ve artırmalı ağaç (boosting) yöntemlerini birlikte kullanan ve performanslarını karşılaştıran bir karar destek sistemi oluşturmaktır. Zaman içerisinde yeni personel

verileri eklendikçe veri setinin varyansı değişecektir. Bu nedenle, düşük ve yüksek varyans durumlarının her ikisinde de başarılı çalışabilecek ağaç modelleri tercih edilmiştir. Herhangi bir t anında veri setinin güncel durumuna bağlı olarak yöntem değişiklik gösterebilecektir. Bu sayede karar destek sisteminde yöntem bazlı bir esneklik yaratması hedeflenmiştir.

Kullanılan Veri Seti

Bu çalışmada, araştırma yapmak amacıyla açık kaynak olarak Kaggle platformunda bulunan HR_comma_sep.csv isimli veri seti kullanılmıştır [49]. Veri seti 14.999 personel verisinden oluşmaktadır. Tablo 2’de veri setini oluşturan özniteliklerin yapısı hakkında bilgi verilmektedir.

Tablo 2. Veri setini oluşturan özniteliklerin yapısı

Sütun Adı	Veri Tipi	Alt Sınır	Üst Sınır
Emp_Id	Metin	-	-
Satisfaction_level	Ondalıklı sayı	0,09	1
Last_evaluation	Ondalıklı sayı	0,36	1
Number_project	Tamsayı	2	7
average_monthly_hours	Tamsayı	96	310
time_spend_company	Tamsayı	2	10
Work_accident	Tamsayı	0	1
left	Tamsayı	0	1
promotion_last_5years	Tamsayı	0	1
Department	Metin	-	-
salary	Metin	-	-

Paylaşılan veri setinde bu verilerin nasıl elde edildiği veya hesaplama metodu hakkında bilgi mevcut değildir. Veri setinin datacard bölümünde bir not veya bilgi görülememiştir. Çalışma, verilerin bu hali kullanılarak yapılacaktır. Sütun isimleri ve veri yapısı tutulan veri hakkında kolaylık sağlamaktadır. Verilerin temel istatistiksel bilgileri Tablo 3’te gösterilmiştir.

- Emp_Id: Personelin numarası.
- Satisfaction_level: Personelin tatmin seviyesi.
- Last_evaluation: Son değerlendirme puanı.
- Number_project: Yer aldığı toplam proje sayısı.
- average_monthly_hours: Ortalama aylık kaç saat çalıştığı.
- time_spend_company: Şirkette kaç yıldır bulunduğu.
- Work_accident: İş kazası geçirip geçirmediği.
- Left: İşten ayrıldı mı? (Bağımlı değişken)
- promotion_last_5years: Şirkette son 5 yılda terfi aldı mı?
- Department: Çalıştığı bölüm.
- Salary: Maaş seviyesi (Düşük/Orta/Yüksek)

Tablo 3. Verilerin temel istatistikleri

	En küçük	En büyük	Ortalama	Standard Sapma	Varyans	Kayıp Veri
satisfaction_level	0,09	1	0,612	0,248	0,061	0
last_evaluation	0,36	1	0,716	0,171	0,029	0
number_project	2	7	3,803	1,232	1,5192	0
average_monthly_hours	96	310	201,05	49,943	2494,313	0

time_spend_company	2	10	3,498	1,46	2,131	0
Work_accident	0	1	0,144	0,351	0,123	0
left	0	1	0,238	0,425	0,181	0
promotion_last_5years	0	1	0,021	0,144	0,02	0

Veri Ön İşleme

Veri setinde personelin departman ve maaş bilgileri kategorik değişken şeklindedir. Maaş bilgisi düşük, orta ve yüksek olarak firma tarafından etiketli olarak paylaşılmıştır. Bu bilgileri kullanabilmek için “One-to-Many” düğümü yardımıyla veri manipülasyona (0-1) uğratılmıştır. Dönüşüm gerçekleştikten sonra modele eski halinin girdi olarak verilmemesi için veri setinden kaldırılmıştır.

Satırlar üzerinde yapılan kontrolle boş verinin bulunmadığı görülmüştür. Herhangi bir anomali değerinin olmadığı, hayatın olağan akışına ters gelecek bir bilginin bulunmadığı ve modellemeye uygun hale bir veri setinin elde bulunduğu anlaşılmıştır.

Veri manipülasyonu sonrası verisetindeki Satisfaction_level, Last_evaluation, Number_project, average_monthly_hours, time_spend_company, Work_accident, left, promotion_last_5years, Department, salary öznitelikleri modellere girdi olarak verilmiştir.

Makine Öğrenmesi

Yapay zekâ alanının çatısı altında yer alan makine öğrenmesi kavramı, insan beyninin hesaplayamayacağı kadar karmaşık model ve süreçler oluşturarak analitik çıktılar üretmeyi sağlayan bir teknikler topluluğudur. Karar alma sürecinde geçmiş verileri kullanarak tahmin sonuçları üretilir. Bu çalışma kapsamında regresyon algoritmalarından olan Gradient Boosting Tree, Random Forest Trees, XGBoosting Regresyon teknikleri kullanılmış olup ağaç teknikleri üzerinde performans karşılaştırması da gerçekleştirilmiştir.

Rastgele Orman Yöntemi: Bu algoritma, rastgele bir şekilde öznitelikleri düğümlere atar. Başarılı sonuç verdiği görülen bir modelde özniteliklerin hangisinin kullanıldığı ve kaçınıcı seviyeden modelde başarılı sonuç verdiği o parametrenin önem düzeyini de gösterir. Rastgele orman yöntemi, veri madenciliği ve makine öğrenimi uygulamalarında kullanılan bir topluluk (ensemble) yöntemidir ve birçok karar ağacını bir araya getirerek regresyon analizi yapar. İlk adım, veri toplamak ve kullanmak istenilen bağımlı (hedef) değişkeni ve bağımsız (girdi) değişkenleri tanımlamaktır. Örneğin, ev fiyatları tahmin edilecekse, odaların sayısı, banyo sayısı, evin yaşı vb. bağımsız değişkenler kullanılabilir.

Herhangi bir karar ağacının amacı en iyi özelliği belirleyip, aşağı doğru dallanarak mimari yapıyı oluşturmaktır. Rastgele orman algoritmasında ise özelliğin iyi olup olmadığına bakmadan, oldukça yüksek sayıda ağaç oluşturup her birinden farklı boyutlarda kazanımlar sağlanmaktadır. Bu özelliği sayesinde aşırı öğrenme (overfitting) oldukça azalacaktır. Bu kazanım, rastgele orman yönteminin en büyük kazanımlarından biri olarak gösterilebilmektedir. Rastgele orman modeli bir problemi çözmek için hem veri setinden

hem de öznitelik setinden rastgele 10'larca ya da 100'lerce farklı alt-setler seçer ve bunları eğitir [43].

Gradyan Artırma Yöntemi (Gradient Boosting): Boosting yöntemleri genel olarak zayıf olan öğrencileri güçlü hale getirmek için kullanılan bir tekniktir. Gradyan artırma yöntemi bir önceki ağacın hatalarını en aza indirmek amacıyla bir sonraki ağacı oluşturur. Başlangıç olması için bir adet ağaç oluşturulur. İlk ağacın tahmin ettiği hatanın negatif gradyanını tahmin etmek için bu algoritma kullanılır [44]. Yeni ağaç bir önceki ağacın hatalarını tahmin eder. Bu süreç yinelemeli olarak ilerler ve sonuçlar giderek daha da iyileşmeye başlar. Bu tekniğin ana amacı hataları en düşük seviyeye çekerek ağaç modelini kurmaktır. Veri sayısının fazla olduğu modellerde avantajlıdır. Kirli veri üzerinde başarılı sonuç verdiği de bilinmektedir. Hem regresyon hem de sınıflandırma problemlerinde kullanılabilir.

Genellikle karar ağaçları gibi basit modeller kullanılır. Gradyan artırmanın temel mantığı, hata fonksiyonunun gradyanına (eğim) dayalı olarak tahmin hatalarını minimize etmeye yönelik adımlar atmaktır. Gradyan artırma, bir hata fonksiyonu tanımlar ve her adımda bu fonksiyonun gradyanına göre modelin performansını iyileştirir. Özellikle, veri seti üzerinde ardışık olarak oluşturulan her bir karar ağacı, önceki modelin hatalarını azaltacak şekilde eğitilir. Bu sayede, model her adımda daha iyi tahminler yapar.

Gradyan artırmanın önemli avantajlarından biri, doğruluğu yüksek tahminler yapabilmesidir. Küçük öğrenme oranları kullanarak, modelin yavaş yavaş öğrenmesini ve daha genel bir tahmin yapmasını sağlayabilir. Ayrıca, erken durdurma (early stopping) gibi tekniklerle modelin gereksiz yere çok fazla öğrenmesi engellenebilir. Bu özellikler, yöntemin tercih edilmesinde etkili olabilmektedir.

Gradyan artırma, karmaşık veri yapıları ve büyük veri setlerinde oldukça etkili bir yöntemdir. Özellikle, XGBoost ve LightGBM gibi modern versiyonları, büyük veri setlerinde hız ve doğruluk açısından önemli iyileştirmeler sağlar. Bu algoritmalar, yüksek performanslı tahminler yapabilme yeteneğine sahiptir. Gradyan artırmanın bu esnek ve güçlü yapısı, finans, sağlık, pazarlama ve daha birçok alanda yaygın olarak kullanılabilir.

Ekstrem Gradyan Artırma Yöntemi (XGBoosting): Ekstrem gradyan artırma algoritması, gradyan artırma algoritmasının optimize edilmiş ileri bir versiyonu sayılabilir. Önceki versiyonlara göre sağladığı avantajları Ekstrem gradyan artırma kullanımının yaygınlaşmasındaki en önemli nedendir [45]. Ağaç oluşturulurken fazla dallanma ve derinliğe ulaştığı takdirde ağaç budanmaya başlar. Aşırı öğrenmeyi engellediği için oldukça fazla kullanılan bir tekniktir.

Çalışmada, bağımlı değişken olan “left” değeri ikili (binary) olarak tutulmaktadır. Personeller için oluşturulan ayrılma tahmin skorlarının hangi eşik değere göre değerlendirileceğine dair insan kaynakları yönetiminin tercih kullanabileceği düşünülerek problem regresyon bakış açısıyla irdelenmektedir.

Ekstrem gradyan artırma, zayıf öğrencileri ardışık olarak birleştirerek hataları minimize ederken, Rastgele orman çok

sayıda karar ağacını bir araya getirip aşırı öğrenmeyi azaltır ve öznelik önemini belirlemede etkilidir. Ekstrem gradyan artırma ise, optimize edilmiş yapısı ile hızlı ve verimli modeller üretir. Bu yöntemler yüksek boyutlu veri setlerinde etkili sonuçlar vermektedir. Literatürdeki çalışmalarda da ağaç tabanlı yöntemlerin kullanıldığı ve başarılı sonuçlar elde edildiği gözlemlenmiştir

Regresyon Modelinin Değerlendirme Ölçütleri

R² Skoru: verilerin yerleştirilmiş regresyon hattına olan mesafesinin istatistiksel ölçüsüdür. Regresyon noktasına isabet edemeyen her noktaya artık adı verilmektedir. Bu artıkların her birinin karelerinin toplamına Artıkların Karelerinin Toplamı (AKT) ve her bir noktanın ortalamaya uzaklığının karesini de Ortalamaya Uzaklığın Kareler Toplamı (OUKT) şeklinde adlandırırsak R² değeri, bu iki istatistik kullanılarak Denklem 1'deki gibi hesaplanır.

$$R^2 \text{ Skoru} = 1 - (\text{AKT}/\text{OUKT})$$

R² değeri 1'e ne kadar yakın ise regresyon modelinin o kadar doğru olduğu anlaşılır. Bu metrikte dikkat edilmesi gereken husus R² değeri eğer 1 çıkarsa modelin ezberlemiş olabileceğidir. 1 değerinden uzaklaşmasıyla modelin tahminleme gücünde azalma olduğu anlaşılmaktadır.

Ortalama Mutlak Hata (Mean Absolute Error): Mutlak hata, tahmin edilen değerler ile gerçek değerler arasındaki farktır. Tam olarak ifade etmek gerekirse, veri kümesinin tüm örneğinde, o örnek için gerçek değer ile öngörülen değer arasındaki her bir farkın mutlak değerinin ortalamasına ortalama mutlak hata (mean absolute error) denir. Düşük bir değer olması modelin az hata yaptığını, yani iyi tahminlemede bulunduğunu gösterir.

Düzeltilmiş R-Kare (Adjusted R Square-R²): Modeldeki değişkenlerin sayısı için ayarlanmış olan R² değiştirilmiş bir versiyondur. Düzeltilmiş R² değeri, herhangi bir yeni değişken modele eklendiğinde artar. Bağımsız değişken eklendikçe model karmaşılaşır, model karmaşılaşınca aşırı ezberleme gibi problemler ortaya çıkabilir.

Ortalama Kare Hatası (Mean Squared Error): En sık kullanılan regresyon kaybı fonksiyonudur. Ortalama Kare Hatası, tüm veri kümesinde örnek başına ortalama hata kareleri kaybıdır. Bu metriğin 0'a yakın çıkması modelin daha doğru tahmin ürettiğini gösterir.

Uygulama Sonuçları

3 farklı makine öğrenmesi algoritması eldeki veri setine uygulanmış ve sonuçları karşılaştırılmıştır. İnsan kaynakları analitiği problemindeki özneliklerin önem seviyelerinin belirlenmesinden önce, modellerin performans göstergeleri Tablo 4'te paylaşılmaktadır.

Tablo 4. Yöntemlerin performans sonuçları

	Rastgele Orman Yöntemi (Ağaç sayısı=100, tree deep=10)	Gradyan Artırma Yöntemi (Ağaç sayısı=10, tree deep=10)	Ekstrem Gradyan Artırma Yöntemi (tree deep=10)
R² Skoru	0,921	0,854	0,870
Ortalama Mutlak Hata	0,014	0,026	0,022
Hata Kareleri Ortalamasının Karekökü	0,044	0,031	0,023
Düzeltilmiş R²	0,906	0,853	0,869

Her 3 modelin de oldukça başarılı tahminleme yapabildiği görülmektedir. Ortalama hata skorları birbirine oldukça yakın çıktığı için yöntem seçiminde R² değeri 1'e en yakın çıkan rastgele orman modeli üzerinden öznelik önemi belirlenmiştir.

R² skorlarını incelediğimizde:

R² Skoru, bir modelin açıklayıcı gücünü gösterir. 1'e ne kadar yakınsa, modelin hedef değişkenin varyansını o kadar iyi açıkladığını belirtir.

Rastgele Orman (R² skoru = 0,921) ile en yüksek değeri sağlıyor. Bu, modelin hedef değişkenin %92.1' ini açıkladığı anlamına gelir. Gradyan Artırma (R² skoru = 0,854) değeri ile daha düşük ve hedef değişkenin yaklaşık %85.4' ünü açıklamaktadır. Ekstrem Gradyan Artırma (R² skoru = 0,870) ile Gradyan Artırma yönteminden biraz daha iyi sonuç verdiği görülmektedir. Sonuç olarak, R² açısından Rastgele Orman Yöntemi en iyi performansı göstermiştir.

Ortalama mutlak hata skorlarını incelediğimizde:

Ortalama mutlak hata, modelin tahmin ettiği değerler ile gerçek değerler arasındaki ortalama farkı ölçer. Daha düşük bir ortalama mutlak hata daha başarılıdır.

Rastgele Orman (ortalama mutlak hata = 0,014), en düşük hata oranına sahip olduğu görülmüştür. Gradyan Artırma (ortalama mutlak hata = 0,026) ile en yüksek hata oranına sahiptir. Ekstrem Gradyan Artırma (ortalama mutlak hata = 0,022), Gradyan Artırma yönteminden daha iyi ancak Rastgele Orman kadar iyi değildir. Sonuç olarak, hata analizi açısından da Rastgele Orman yöntemi en iyi performansı sergilemiştir.

Düzeltilmiş R² değerlerini incelediğimizde:

Düzeltilmiş R² değerlerine göre de sonuçlar yorumlanabilir. Rastgele orman (Düzeltilmiş R²=0,906) yüksek bir değer ile verileri oldukça iyi açıklamaktadır. Gradyan Artırma (Düzeltilmiş R²=0,853) daha düşük bir düzeltilmiş R² skoru ile en az iyi performans gösteren modeldir. Ekstrem Gradyan Artırma (Düzeltilmiş R²=0,869), Gradyan artırmadan daha iyi performans sergilemekte ancak rastgele ormandan daha düşük performans göstermektedir.

Hata Kareleri Ortalamasının Karekökü değerlerini incelediğimizde:

Hata kareleri ortalamasının karekökü, hataların karekökünü olarak ortalamayı hesaplar ve büyük hatalara daha fazla ağırlık verir. Rastgele Orman (Hata kareleri ortalamasının karekökü = 0,044) değeri ile en yüksek değere sahip yöntem olduğu görülmüştür. Bu, büyük hatalar konusunda diğer iki modele göre daha zayıf olduğunu gösterir. Gradyan Artırma (Hata kareleri ortalamasının karekökü = 0,031), ise biraz daha iyi bir sonuç verdiği görülmüştür. Ekstrem Gradyan Artırma (Hata kareleri ortalamasının karekökü = 0,023) ise, 3 yöntem arasında en düşük değeri sunarak büyük hataları en iyi şekilde minimize edebildiğini göstermiştir. Ekstrem Gradyan Artırma yönteminde hata büyüklüklerinin küçük ve tutarlı olması ortalama mutlak hata ile ortalama kare hata değerlerinin birbirine yakın olmasını açıklamaktadır. Sonuç olarak, Hata kareleri ortalamasının karekökü değerleri baz alındığında büyük hatalar açısından Ekstrem Gradyan Artırma Yöntemi en iyi performansı sergiliyor.

Tüm metrikleri genel olarak değerlendirecek olursak:

Rastgele Orman Yöntemi, R^2 Skoru, ortalama mutlak hata ve düzeltilmiş R^2 açısından en yüksek performansı göstermektedir. Ancak hata kareleri ortalamasının karekökü değeri, diğer modellerden biraz daha yüksektir. Bu da modelin büyük hatalara daha duyarlı olduğunu göstermektedir.

Ekstrem Gradyan Artırma Yöntemi, en düşük hata kareleri ortalamasının karekökü değerine sahip olup, büyük hataları minimize etme konusunda başarılı olarak değerlendirilebilir. Ancak genel performansı (R^2 ve düzeltilmiş R^2) açısından Rastgele Orman'ın gerisinde kalmıştır.

Gradyan Artırma Yöntemi, genel olarak en düşük performansa sahip model olarak gözükmekte olup özellikle hem R^2 hem de ortalama mutlak hata açısından daha zayıf kalmıştır.

Bu analiz sonucunda, **Rastgele Orman Yöntemi**, veri setinin yapısına en uygun model olarak öne çıkmıştır.

Ağaç algoritmalarında, önce kök düğüm, sonrasında ise ilk hangi nitelikten bölüneceği ve kurallarının tespit edilmesi gerekmektedir. En iyi ayırım noktalarının tespit edilmesi için ise gini indeks yöntemi kullanılabilir. Gini indeks değeri, bir sınıfın içindeki izafi olarak sıklığı ifade eder. Her bir özellik için ayrı ayrı hesaplanan Gini indeks değerleri arasından en küçük olanı seçilir. Bölünme işlemi bu değere göre yapılmaktadır. Bu işlemler kalan veriler için de tekrar edilir ve diğer bölünmeler için hesaplanır. Sonuç olarak, bir özneliğin gini indeks değeri ne kadar düşük ise önemi ve dolayısıyla ağırlığı o kadar yüksek olacaktır. Gini hesaplaması aşağıdaki şekilde formüle edilmiştir. Sklearn kütüphanesinin içerisinde yer alan Feature Importance fonksiyonu, özellik önemlerini model üzerinde otomatik olarak hesaplayabilmektedir.

$$\text{Gini} = 1 - \left(\sum_{i=0}^m p(i)^2 \right)$$

Burada $p(i)$, sınıf i 'nin toplam nüfus içerisindeki oranını temsil eder. m ise, toplam sınıf sayısıdır.

S	Öznitelikler	S	Gini-based feature importance
	promotion_last_5years		0.000147
	product_mng		0.000403
	RandD		0.000587
	hr		0.000634
	management		0.000877
	marketing		0.001181
	high		0.001249
	accounting		0.001326
	Work_accident		0.001541
	IT		0.002037
	sales		0.002061
	medium		0.002492
	support		0.002744
	low		0.003191
	technical		0.003720
	average_monthly_hours		0.102119
	number_project		0.105349
	last_evaluation		0.124229
	time_spend_company		0.143736
	satisfaction_level		0.500376

Şekil 1. Özniteliklerin özellik önemlerinin gösterimi

Tüm özneliklerin işten ayrılmaya olan etkileri Şekil 1'de gösterilmiştir. Öznelik önemleri en düşükten en yükseğe doğru sıralanmış olup en önemli özneliğin açık ara farkla personel tatmini olduğu açığa çıkmıştır. Bu da yazındaki çalışmaları destekler niteliktedir. Önem skorları incelendiğinde 5 öznelik ile işten ayrılma kararının büyük ölçüde açıklanabileceği sonucuna ulaşmak mümkündür: çalışan tatmini (%50,03), şirkette bulunulan yıl (%14,37), değerlendirme puanı (%12,42), proje sayısı (%10,54) ve ortalama aylık çalışma süresi (%10,21). Bu 5 özneliğin toplam ağırlığı %97,57'dir.

Sonuç ve Değerlendirme

İnsan kaynağı analitiği çalışmaları son yıllarda giderek trend olan çalışma alanlarından biridir. Personellerin işten ayrılma tahminini önceden yapabilmek şirket için önemli bir değer olabilecektir. İşten ayrılma kararını öznel veriler yerine, nesnel veriler üzerinden ortaya koyabilen makine öğrenmesi modelleri bu çalışmada kullanılmıştır. Şirketin geçmiş dönem verileri, kurum kültürünü yansıtmada anlamında objektif bir perspektif sunabilmektedir. Veriden anlam çıkarma tekniğine ve devamlı öğrenme mantığına dayanan farklı makine öğrenmesi yaklaşımları ile tahminleme problemi ele alınmıştır. Ortalama hata skoru birbirine yakın çıkan tekniklerin yanında R^2 skoruna da bakılmasında fayda vardır.

Çalışan tatmini, çalışanların işten ayrılmalarında en önemli etken olarak öne çıkmıştır. Bu öznelik toplam ağırlığın yarısından fazlasını (%50,03) oluşturmaktadır. Çalışanlar, iş yerinde memnun olmadıklarında işten ayrılma eğilimi göstermektedirler. Tatmin düzeyi düşük olan çalışanlar, başka bir iş arayışına girebilir veya mevcut pozisyonlarında uzun süre kalma konusunda motivasyonlarını kaybedebilirler. Bu durum iş yükü, çalışma ortamı, maaş memnuniyeti, kariyer olanakları gibi faktörlerden etkilenebilir. Analiz, çalışan memnuniyetinin yüksek olduğu durumlarda işten ayrılma olasılığının önemli ölçüde azaldığını vurgulamaktadır.

Çalışanların şirkette ne kadar süre buldukları, işten ayrılma riskini etkileyen ikinci en önemli faktör olarak belirlenmiştir (%14,37). Genel olarak, çalışanlar şirkette ne kadar uzun süre kalırsa, kurumsal bağlılıkları o kadar artar, ancak bir noktadan sonra kariyer gelişimi, monotonluk ya da yeni

fırsatlar arama isteği işten ayrılma olasılığını artırabilir. Yeni çalışanların ayrılma riski yüksek olabilirken, uzun süreli çalışanlar için bu risk değişkenlik gösterebilir. Bu nedenle, şirkette geçirilen yılın işten ayrılma olasılığı üzerindeki etkisi çift yönlü değerlendirilebilir.

Çalışanların performans değerlendirmeleri de önemli bir etken olarak görülmektedir (%12,42). Değerlendirme puanları, çalışanların kariyer ilerlemelerini, ödüllendirme sistemlerini ve terfi fırsatlarını etkileyebilmektedir. Düşük performans puanlarına sahip çalışanlar, işten çıkarılma riskiyle karşı karşıya kalabilir ya da düşük motivasyon nedeniyle gönüllü olarak ayrılmayı tercih edebilirler. Buna karşılık, yüksek performans gösteren çalışanlar, kendilerine yeterli ödül ya da kariyer fırsatları sunulmazsa ayrılma eğilimi gösterebilirler.

Çalışanların üzerinde çalıştıkları proje sayısı, işten ayrılma kararında %10,54 oranında etkilidir. Fazla sayıda proje üstlenen çalışanlar, iş yükünün artması nedeniyle stres veya tükenmişlik hissedebilirler, bu da işten ayrılma olasılığını artırabilir. Öte yandan, projelerde yeterince görev almayan çalışanlar ise işlerinde gelişim fırsatı bulamayıp motivasyonlarını kaybedebilirler. Bu yüzden, proje sayısının işten ayrılma üzerindeki etkisi, hem iş yükünün dengesizliği hem de kariyer gelişimi eksikliğine bağlı olarak değişebilir.

Ortalama aylık çalışma süresi de işten ayrılma riskini etkileyen önemli bir faktör olarak ortaya çıkmaktadır (%10,21). Çalışanlar, fazla mesai yapmaya zorlandıklarında ya da uzun çalışma saatleriyle karşı karşıya kaldıklarında, iş-yaşam dengesini kurmakta zorlanabilirler ve bu durum, işten ayrılma eğilimini artırabilir.

Çalışan tatmini, şirkette kalınan süre, değerlendirme puanı, proje sayısı ve çalışma süresinden oluşan 5 özneliğin toplam ağırlığının %97,57 olması, işten ayrılma kararının büyük ölçüde bu faktörlere dayandığını açıkça göstermektedir. Çalışan tatmini en baskın faktör olup, çalışanların işlerinde ne kadar memnun olduklarının, işte kalma eğilimleri üzerinde çok büyük bir etkiye sahip olduğunu ortaya koymaktadır. Diğer dört öznelik ise daha spesifik durumlara işaret eder: şirkette kalınan süre, çalışanların bağlılığını ve kariyer gelişimini; değerlendirme puanı, performansla ilgili motivasyonu; proje sayısı ve çalışma süresi ise iş yükü ve iş-yaşam dengesi ile ilgili zorlukları temsil etmektedir. Bu özneliklerin bir arada değerlendirilmesi, çalışan bağlılığını artırmaya ve işten ayrılma oranlarını azaltmaya yönelik stratejilerin geliştirilmesinde kritik rol oynayabilir.

Bu çalışmada yapıldığı gibi rastgele orman yöntemi kullanıldıktan sonra özellik önemi veya özellik azaltma amaçlarıyla yapılmış çalışmalar incelenmiş olup, Gini indeksi yardımıyla başarılı şekilde sonuçlar elde edilebilmiştir. [46], yapmış olduğu çalışmada rastgele orman yöntemini kullanarak özellik önemi ve özellik eleme işlemi için gini değerini kullanmıştır. [47] ise, rastgele ormanın gini önemi, spektral verilere dayalı özelliklerin önemini ölçmede üstün bir yöntem sağlarken, optimal bir özellik alt kümesinde, düzenlenmiş sınıflayıcıların rastgele orman sınıflayıcısından daha tercih edilir olabileceği belirtilmiştir. Bununla birlikte, düzenlenmiş sınıflayıcılar yalnızca doğrusal bağımlılıkları modelleme sınırlamasına sahiptir. Gini önemine dayanan bir

özellik seçimi, optimal özellik alt kümesini belirlemek ve sınıflandırma görevindeki gürültüyü ortadan kaldırarak hem boyut azaltma hem de gürültü eliminasyonu sağlamak için düzenlenmiş doğrusal sınıflandırmadan önce uygulanabileceğini çalışmalarında göstermiştir. Başka bir çalışmada ise, rastgele orman yöntemi kullanılırken özellik öneminin gini değeri üzerinden başarılı şekilde değerlendirilebileceği sonucuna ulaşılmıştır [48].

Bu çalışma karşılaştırma tabanlı karar destek sisteminin bir altyapısı niteliği taşımaktadır. İlerleyen zamanlarda yeni verilerle besleyerek veri setindeki değişimlere bağlı olarak algoritma seçiminde değişiklik gösterebilecektir. Bu nedenle, yeni veriler geldikçe farklı teknikleri karşılaştıran bir karar destek sistemine dönüşebilecektir.

Yazında oldukça büyük yer kaplayan personelin işten ayrılma kararını tahminleme problemi için başkaca makine öğrenmesi teknikleri de denenebilir. Ayrıca başka personel veri setlerinde bu algoritmalarla modellerin nasıl performans göstereceği irdelenebilir. Şirketlerin önerilerini daha iyi görebilmeleri için ellerindeki personelleri daha stratejik bir açıdan değerlendirmeleri ve insan kaynakları politikalarını da bu kapsamda güncellemeleri iyi olabilir.

Kaynaklar

- [1] S. Reşitoğlu, "Yetkinlik Bazlı Performans Değerlendirme ve Çalışan Memnuniyeti- Bir Uygulama," Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, İzmir, 2011.
- [2] C. Yalçın, "Müşteri Kayıp Analizi (Customer Churn Analysis)," YBS Ansiklopedi, vol. 7, 2019.
- [3] V. Gülpınar, "Yapay Sinir Ağları ve Sosyal Ağ Analizi Yardımı ile Türk Telekomünikasyon Piyasasında Müşteri Kaybı Analizi," Marmara Üniversitesi İktisadi İdari Bilimler Dergisi, vol. 34, no. 1, pp. 331-350, 2013.
- [4] R. S. Shankar, J. Rajanikanth, V. V. Sivaramaraju, and K. V. S. S. R. Murthy, "Prediction of Employee Attrition Using Data Mining," in IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2018, doi: 10.1109/ICSCAN.2018.8541242.
- [5] D. Alao and A. B. Adeyemo, "Analyzing Employee Attrition Using Decision Tree Algorithms," Computing, Information Systems, Development Informatics and Allied Research Journal, vol. 4, no. 1, pp. 17-28, 2013.
- [6] S. O. Abdulsalam, J. F. Ajao, B. F. Balogun, and M. O. Arowolo, "A Churn Prediction System for Telecommunication Company Using Random Forest and Convolution Neural Network Algorithms," 2022.
- [7] X. Gao, J. Wen, and C. Zhang, "An Improved Random Forest Algorithm for Predicting Employee Turnover," Mathematical Problems in Engineering, vol. 2019, no. 1, 2019.

- [8] S. S. Alduayj and K. Rajpoot, "Predicting Employee Attrition using Machine Learning," in 2018 International Conference on Innovations in Information Technology (IIT), pp. 93-98.
- [9] X. Gao, J. Wen, and C. Zhang, "An Improved Random Forest Algorithm for Predicting Employee Turnover," *Mathematical Problems in Engineering*, vol. 2019, no. 1, 2019.
- [10] S. F. Sari and K. M. Lhaksmana, "Employee Attrition Prediction Using Feature Selection with Information Gain and Random Forest Classification," *Journal of Computer System and Informatics (JoSYC)*, 2022.
- [11] R. Joseph, S. Udupa, S. Jangale, K. Kotkar, and P. Pawar, "Employee Attrition Using Machine Learning And Depression Analysis," in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1000-1005.
- [12] M. M. Alam, K. Mohiuddin, M. K. Islam, M. Hassan, M. A. Hoque, and S. M. Allayear, "A Machine Learning Approach to Analyze and Reduce Features to a Significant Number for Employee's Turn Over Prediction Model," *Advances in Intelligent Systems and Computing*, 2018.
- [13] Y. Chen, X. Lin, and K. Zhan, "Employee Attrition Classification Model Based on Stacking Algorithm," *Psychology Research*, vol. 13, no. 6, pp. 279-285, 2023.
- [14] A. Alamsyah and N. Salma, "A Comparative Study of Employee Churn Prediction Model," in 2018 4th International Conference on Science and Technology (ICST), pp. 1-4.
- [15] S. O. Abdulsalam, J. F. Ajao, B. F. Balogun, and M. O. Arowolo, "A Churn Prediction System for Telecommunication Company Using Random Forest and Convolution Neural Network Algorithms," *EAI Endorsed Transactions on Mobile Communications and Applications*, vol. 7, no. 21, e4, 2022. doi: 10.4108/eetmca.v6i21.2181.
- [16] D. O. Alao and A. B. Adeyemo, "Analyzing Employee Attrition Using Decision Tree Algorithms," 2013.
- [17] M. Subhashini and R. Gopinath, "Employee Attrition Prediction in Industry Using Machine Learning Techniques," *International Journal of Advanced Research in Engineering and Technology (IJARET)*, vol. 11, no. 12, 2020.
- [18] S. S. Reddy, J. Rajanikanth, V. V. Sivaramaraju, and K. V. S. S. R. Murthy, "Prediction of Employee Attrition Using Data Mining," in 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA), pp. 1-8.
- [19] A. Qutub, A. Al-Mehmadi, M. Al-Hssan, R. Aljohani, and H. S. Alghamdi, "Prediction of Employee Attrition Using Machine Learning and Ensemble Methods," *International Journal of Machine Learning*, vol. 11, no. 2, 2021.
- [20] S. Dutta and S. K. Bandyopadhyay, "Employee Attrition Prediction Using Neural Network Cross Validation Method," *International Journal of Commerce and Management*, vol. 6, no. 3, pp. 80-85, 2020.
- [21] S. Agarwal, C. Bhardwaj, G. Gatkamani, R. Gururaj, N. Darapaneni, and A. R. Paduri, "AI Based Employee Attrition Prediction Tool," *Lecture Notes in Computer Science*, pp. 580-588, 2023. doi: 10.1007/978-3-031-36402-0_54.
- [22] M. Atef, D. Elzanfaly, and S. Ouf, "Early Prediction of Employee Turnover Using Machine Learning Algorithms," *International Journal of Electrical and Computer Engineering Systems*, vol. 13, no. 2, pp. 135-144, 2022.
- [23] T. S. Poornappriya and R. Gopinath, "Employee Attrition In Human Resource Using Machine Learning Techniques," *Webology*, vol. 18, no. 6.
- [24] M. A. Abelson, "Examination of Avoidable and Unavoidable Turnover," *Journal of Applied Psychology*, vol. 72, no. 3, pp. 382-386, 1987.
- [25] H. G. Heneman, T. Judge, and J. D. Kammeyer-Mueller, *Staffing Organizations*, 7th ed. Mendota House, 2012.
- [26] J. L. Cotton and J. M. Tuttle, "Employee Turnover: A Meta-Analysis and Review with Implications for Research," *Academy of Management Review*, vol. 11, no. 1, 1986.
- [27] J. R. Terborg and T. W. Lee, "A Predictive Study of Organizational Turnover Rates," *Academy of Management Journal*, vol. 27, no. 4, pp. 793-810, 1984.
- [28] S. M. Carraher, "Turnover Prediction Using Attitudes Towards Benefits, Pay, and Pay Satisfaction Among Employees and Entrepreneurs in Estonia, Latvia, and Lithuania," *Baltic Journal of Management*, vol. 6, no. 1, pp. 25-52, 2011.
- [29] A. Erosa, L. Fuster, and D. Restuccia, "Fertility Decisions and Gender Differences in Labor Turnover, Employment, and Wages," *Review of Economic Dynamics*, vol. 5, no. 4, pp. 856-891, 2002.
- [30] E. M. Ineson, E. Benke, and J. László, "Employee Loyalty in Hungarian Hotels," *International Journal of Hospitality Management*, vol. 32, pp. 31-39, 2013.
- [31] A. I. Kraut, "Predicting Turnover of Employees from Measured Job Attitudes," *Organizational Behavior and Human Performance*, vol. 13, no. 2, pp. 233-243, 1975.
- [32] D. C. Maynard, T. A. Joseph, and A. M. Maynard, "Underemployment, Job Attitudes, and Turnover Intentions,"

Journal of Organizational Behavior, vol. 27, no. 4, pp. 509-536, 2006.

[33] S. Sellgren, G. Ekvall, and G. Tomson, "Nursing Staff Turnover: Does Leadership Matter?", *Leadership in Health Services*, vol. 20, no. 3

[34] D. G. Allen and R. W. Griffeth, "Job embeddedness and turnover intention: A multi-method study," *Journal of Applied Psychology*, vol. 102, no. 4, pp. 620-633, 2017. doi: 10.1037/apl0000181.

[35] J. Liu, Z. Wang, and C. Lu, "Employee turnover intention and job satisfaction: The moderating role of psychological empowerment," *Frontiers in Psychology*, vol. 8, p. 1018, 2017. doi: 10.3389/fpsyg.2017.01018.

[36] Y. Chang and C. Chen, "The relationship between work-family conflict and turnover intention: A study of healthcare professionals," *Journal of Nursing Management*, vol. 26, no. 5, pp. 587-596, 2018.

[37] X. Zhai and S. Xu, "Turnover prediction using machine learning techniques: An empirical study," *Journal of Business Research*, vol. 120, pp. 19-28, 2020.

[38] P. Chen and L. Zhang, "The role of employee resilience in mitigating turnover intention: Evidence from the financial sector," *International Journal of Human Resource Management*, vol. 30, no. 8, pp. 1239-1260, 2019.

[39] F. C. D. Fisher, "Happiness at work," *International Journal of Management Reviews*, vol. 12, no. 4, pp. 384-412, 2010.

[40] B. L. Rich, J. A. LePine, and E. R. Crawford, "Job engagement: Antecedents and effects on job performance," *Academy of Management Journal*, vol. 53, no. 3, pp. 617-635, 2010.

[41] M. Tims, A. B. Bakker, and D. Derks, "The impact of job crafting on job demands, job resources, and well-being," *Journal of Occupational Health Psychology*, vol. 18, no. 2, pp. 230-240, 2013.

[42] E. E. Kossek and J. S. Michel, "Flexible work schedules," in *APA Handbook of Industrial and Organizational Psychology*, vol. 1, S. Zedeck, Ed. Washington, DC, USA: American Psychological Association, 2011, pp. 535-572.

[43] H. K. Şimşek, "Makine Öğrenmesi Dersleri 5b: Random Forest (Regresyon)," *Medium*, Dec. 12, 2022. [Online]. Available: <https://medium.com/data-science-tr/makine-ogrenmesi-dersleri-5a-random-forest-regresyon-2a91715a8b66>. [Erişim Tarihi: 05.04.2024].

[44] E. Güler, "Gradient Boosting Nedir?" *Medium*, Mar. 14, 2024. [Online]. Available: <https://efecanxrd.medium.com/gradient-boosting-nedir-2ba518700777>. [Erişim Tarihi: 05.04.2024].

[45] A. C. Kelle ve H. Yüce, "MQTT Trafikinde DoS Saldırılarının Makine Öğrenmesi ile Sınıflandırılması ve Modelin SHAP ile Yorumlanması," *J. Mater. Mechat. A*, vol. 3, no. 1, pp. 50-62, 2022, doi: 10.55546/jmm.995091.

[46] Saarela, M., Jauhiainen, S. Comparison of feature importance measures as explanations for classification models. *SN Appl. Sci.* 3, 272 (2021). <https://doi.org/10.1007/s42452-021-04148-9>

[47] Menze, B.H., Kelm, B.M., Masuch, R. et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 10, 213 (2009).

[48] Kristin K. Nicodemus, Letter to the Editor: On the stability and ranking of predictors from random forest variable importance measures, *Briefings in Bioinformatics*, Volume 12, Issue 4, 2011, Pages 369-373, <https://doi.org/10.1093/bib/bbr016>

[49] Kaggle. HR_comma_sep.csv. <https://www.kaggle.com/datasets/liujiqi/hr-comma-sepcsv>. Erişim Tarihi: 04.03.2024