

Çalışan İşten Ayrılma Tahmini

Su Eda YILDIZ

Bilgisayar Mühendisliği Departmanı
Tobb Economy and Technology University
Ankara, Türkiye

suedayildiz@etu.edu.tr

Abstract – Çalışan kaybı dünya çapında bir problem haline gelmiştir. Şirketlerin başarıları çalışanlarına bağlıyken işten ayrılmalar önemli kayıplara sebep olur. Bunların başında bilgi kaybı, güvenlik sıkıntıları ve sürdürülebilirlik gelir. Dahası, yeni çalışanları işe almak mevcut çalışanları işte tutmaya kıyasla çok daha pahalı, yorucu ve zaman alıcıdır. Çünkü, yeni bir çalışanı işe alırken ve süreç boyunca verilen eğitimler ekstra masraflı olabilir. Ayrıca, çalışanın şirket kültürüne, çalışma ortamına, çalışma arkadaşlarına alışması aylar sürebilir. İnsan kaynakları çalışanların şirketten ayrılma riskini tahmin edebilirse riskleri belirleyebilir, çalışanları şirkette tutmak için gerekli desteği sağlayabilir veya organizasyon üzerindeki etkiyi en aza indirmek için proaktif işe alım yapabilir. Bu makalenin amacı, çalışanların işten ayrılma olasılığını tahmin etmek için makine öğrenimi tekniklerini kullanarak çalışan kaybını ve performansını analiz etmektir. Bu analiz yapılırken 'IBM HR Analytics Employee Attrition Performance' veri seti kullanılacak ve farklı model eğitimleriyle sonuçlar kıyaslanacaktır.

Keywords- Makine Öğrenmesi , Sınıflandırma, İşten Ayrılma, .

I. GİRİŞ

Çalışanların istifası ya da emekliliği anlamına gelen çalışan kaybı, kuruluşlar için önemli bir endişe kaynağıdır. Özellikle yüksek becerili çalışanların aşırı kaybı, şirketin başarısını doğrudan etkileyebilir. Kayıp iki türe ayrılabilir: İstemsiz ve Gönüllü. İstemsiz kayıp, üst düzey yetkililerin tatmin edici olmayan performans, verimlilik sorunları veya uygunsuz davranış nedeniyle çalışanları işten çıkarması durumunda meydana gelir. Öte yandan, gönüllü kayıp, çalışanların kendi istekleriyle ayrılması durumunda meydana gelir. Gönüllü kaybın yaygın nedenleri arasında düşük ücret, büyüme fırsatlarının olmaması, monoton çalışma rutinleri, üstler veya yönetimle zayıf etkileşim ve diğer şirketler tarafından sunulan daha cazip teşvikler yer alır.[1]

Yüksek düzeyde kayıp işletme üzerinde zararlı etkilere sahip olabilir. Bunların başında bilgi birikiminin kaybolması ve bunun sonucunda nitelikli işlerin çıkmaması gelir. Şirket, eskisi kadar iyi ürün üretemezse veya geliştiremezse satışlarda düşüş gözükülebilir. Ayrıca, nitelikli yeni çalışan bulmak, eğitmek ve çalışanın alışma süreci zaman alıcı ve maliyetlidir. Bu nedenle, çalışanların işten ayrılmasına sebep olan faktörleri anlamak, şirketlerin daha hızlı hareket etmesini, şartları düzeltmesini ve acil durum planları geliştirmesini sağlamak için önemlidir. Bu tahmin genellikle İnsan Kaynakları departmanının elinde bulunan temel çalışan verilerini analiz ederek yapılır.

Etkili bir elde tutma stratejisi oluşturmak için şirketlerin maaş, terfi fırsatları, işe gidiş mesafesi, meslektaşlarla ilişkiler ve genel iş memnuniyeti gibi çeşitli faktörleri göz önünde bulundurması gerekir.[1] Şirketler, bu faktörleri değerlendirerek, çalışanların istifa etmesini engellemek ve onlara kalmaları için güçlü teşvikler sağlamak üzere stratejiler geliştirebilir. Çalışan cirosunu tahmin etmek ayrıca, kayıpla ilişkili finansal kayıpları ve üretim düşüşlerini vurgulamaya yardımcı olur.

Emeklilik veya başka şehirlere taşınma gibi bazı kayıp türleri önlenemezken, belirli seviyelerde kayıp, bir şirketin mali sağlığını ve kurum kültürünü ciddi şekilde etkileyebilir. Bu nedenle kaliteli çalışanları elde tutmak kritik bir zorluktur.

Yapay Zeka (AI) ve Makine Öğreniminin (ML) gelişimi ile, şirketlerin çalışan kaybı da dahil olmak üzere gerçek zamanlı sorunları ele alma biçimi değişti. Bu teknolojiler, kuruluşların kayıplarının temel nedenlerinin belirlenmesini ve çalışma ortamını iyileştirmek için proaktif adımlar atılmasını sağladı. Lojistik Regresyon, Karar Ağaçları ve Destek Vektör Makineleri (SVM) gibi makine öğrenimi algoritmaları kayıp tahmininde sıkça uygulandı.

2. kısımda daha önceden yapılan çalışmaları literatür araştırması başlığı altında inceleyip yorumladım. 3. Kısımda veri seti hakkında genel bilgilere değindim. 4. Kısımda veri işleme, öznitelik seçimi ve benzeri konular için kullandığım metodolojiler, 5. Kısımda işten ayrılma tahmini için eğittiğim modeller, 6. Kısımda ise tüm bu sonuçların karşılaştırılmasını ekledim. 7. Kısımda proje hakkında genel değerlendirme var. En son kısımda ise referans aldığım kaynakları bulabilirsiniz.

II. LİTERATÜR ARAŞTIRMASI

Şuana kadar, çalışan işten ayrılma tahmini konusunda çok farklı çalışmalar yürütülmüştür. Bunlardan bazıları veri setini incelemek, görselleştirmek ve öznitelikler arasındaki ilişkileri çözmeye yönelik çalışmalardı. Diğer kısmı ise farklı makine öğrenmesi modelleri ile çalışan işten ayrılma tahmini yapıyordu. Gelişen teknolojilerle birlikte, yeni projeler daha çok derin öğrenme, yapay sinir ağları ve füzyon modellerle gerçekleştirilmektedir. İlerleyen kısımda bu projelerde alınan sonuçlar tartışılmaktadır.

[2] tarafından yapılan bir analiz, işletmelerin büyük veri kümelerini analiz etmek ve daha iyi sonuçlar elde etmek için AI ve makine öğrenimini nasıl kullanabileceğini göstermektedir. Bu araştırmanın temel amacı, ML ve AL'nin

çalışan davranışını ve üretkenliğini nasıl iyileştirebileceğini anlamaktır.

[3] tarafından yapılan araştırmada, çalışan özellikleri ile tahmini kayıp oranı arasındaki ilişki araştırıldı. Veriler, çalışanların iş ile ilgili nitelikleri ve demografik profillerinden oluşuyordu ve araştırma sonuçları, çalışan kaybı olasılığını tahmin etmek için tahmini bir model oluşturmak için kullanıldı.

[4] aynı IBM veri kümesinde SVM, Karar Ağaçları, Rastgele Orman ve K en yakın komşu gibi çeşitli modeller kullanmıştır. Attrition'a karşı Gender, BusinessTravel, OverTime vb. Özniteliklerin grafiklerini çizerek değerlendirir. Verilen değerlere dayanarak, diğer sınıflandırıcılara kıyasla %88.43 doğrulukla Rastgele Orman'da en iyi tahminde bulunur.

[5]'de, makale beş temel modelin doğruluğunu tahmin eder ve daha sonra bunları birleştirerek genellikle topluluk öğrenmesi olarak bilinen daha güçlü bir tahmin modeli elde eder. Bunlar arasında AdaBoost ve Rastgele Orman, SVM ve gradient boosting, Karar Ağaçları ve Linear Regresyon birleşimi olan modeller bulunur. Burada Karar Ağacı ve Linear Regresyon topluluk modeli %86.39 başarımla elde eder.

[6]'te ise daha üst düzey olan CNN ve Kapsül Ağlar birleştirildi. Metodoloji, CNN'leri kullanarak özellik çıkarmayı ve ardından Kapsül Ağları ile hiyerarşik ilişkileri yakalamayı içeriyordu.

III. VERİ SETİ

Yapılan çalışmada 'IBM HR Analytics Employee Attrition Performance' veri seti kullanılmıştır. Bu veri seti, IBM veri bilimcileri tarafından oluşturulmuş yapay bir veri setidir. Toplamda 35 öznitelik ve 1470 satır bulunur. Label olarak 'Attrition' sütünü kullanılır. Bu sütün, çalışanın işten ayrılıp ayrılmadığını gösterir. 35 öznitelik arasında 8 tanesi sayısal olmayan, 27 tanesi sayısal özniteliktir.

Kategorik verilerin üç tanesi (Attrition, Gender, OverTime) iki sınıflıdır ve sonucu yüksek oranda etkileyen özelliklerdir. Beş tanesi (BusinessTravel, Department, EducationField, JobRole, MaritalStatus) ise çok sınıflıdır.

Sayısal verilerin dokuz tanesi (Education, JobLevel, EnvironmentSatisfaction, JobInvolvement, JobSatisfaction, PerformanceRating, RelationshipSatisfaction, StockOptionLevel, WorkLifeBalance) ordinalken kalan on sekiz tanesi (DailyRate, DistanceFromHome, EmployeeCount, EmployeeNumber, MonthlyIncome, MonthlyRate, NumCompaniesWorked, PercentSalaryHike, StandardHours, TotalWorkingYears, TrainingTimesLastYear, Age, YearsAtCompany, HourlyRate, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager) ise nominaldır.

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	3
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	4
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	4
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	1

Fig. 1 Veri Seti

Fig. 1'de veri seti hakkında genel bilgi oluşması için ilk 5 satırı gösterilmiştir.

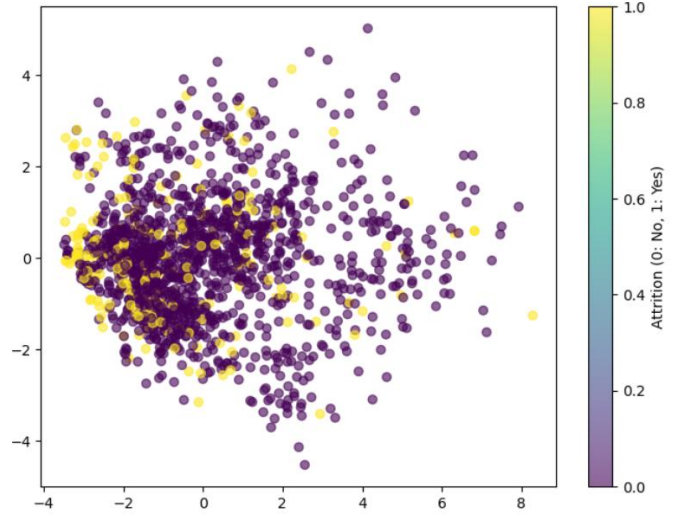


Fig. 2 Çıktının Uzayda Dağılımı

Veri setinde sonucu etkilemeyen, alakasız ve ilişkili (correlated) öznitelikler vardı. Bunlar hakkında yapılanlar bir sonraki bölümde ele alınmıştır.

IV. METODOLOJİ

A. VERİ ÖNİŞLEME

Veri ön işleme, bir algoritma veya modelin performansında önemli bir rol oynar. Algoritmanın verimliliği ve hızı, veri ön işlemenin düzgün bir şekilde yürütülüp yürütülmediğine önemli ölçüde etkilenir. Doğru şekilde yapıldığında, ön işleme algoritmasının hızını artırabilir ve büyük veri kümelerini daha verimli bir şekilde işlemesini sağlayabilir.

Veri ön işlemesine önce veri setinde eksik değer olup olmadığına bakarak başlıyorum. Veri setinde herhangi bir eksiklik yok. Bundan dolayı veri tamamlamaya dair bir adım yok.

Sonrasında aynı satırdan birden fazla (duplicate) var mı diye bakıyorum. Bu işlem, modelin başarımını etkilemese de aynı veriyi birden fazla eğiterek kaybedilecek zaman açısından önemlidir. Veri setinde duplicate değer olmadığından satır silme gibi bir işleme gerek yok.

Makine öğrenmesi modellerine girdi olarak sayısal değerler verilebilir. Bu yüzden veri setinde bulunan tüm

sayısal olmayan özniteliklerin değişimine ihtiyaç var. Bunun için ikili sınıflarda 0 ve 1'e atama kullanılmıştır. Çoklu sınıflarda ise label encoding yapılmıştır.

Sonucu aynı etkilemesi gereken özniteliklerde eğer bazılarının değiştiği aralık diğerlerine göre çok büyükse model yanlış ağırlaklandırmayla öğrenebilir. Bunun önüne geçebilmek için verilerin standartlaştırılması gerekir. Bunu yapabilmek için MinMaxScaler kullanılmıştır. Sadece ortalaması yüksek olan özniteliklere uygulanmıştır.

Veriler incelendiğinde 0-1 aralığına eşleştirmek gerekmediğinden normalizasyon adımları atlandı.

B. ÖZİNTELİK SEÇİMİ

Veri seçimi de veri önışleme kadar önemlidir. Alakasız verilerin modele verilmesi modelin yanlış öğrenmesine ve düşük başarımla sonuçlanabilir.

İlk olarak özniteliklerin her biri için eşsiz (unique) değerler bulundu. Her çalışan için değeri aynı olan öznitelikleri çıkarıldı.

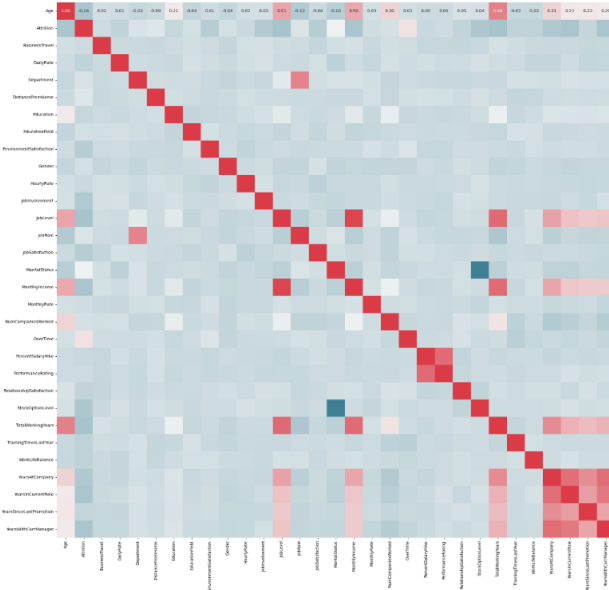


Fig. 3 Başlangıçtaki Korelasyon Matrisi

Veri seçimi yapabilmek ve aslında pair-plot gösterimini yapabilmek için önce PCA uygulanmıştır. Burada 5 öznitelik bulunmuş ve sonrasında Korelasyon matrisi yardımıyla korelasyonu fazla olan (belirli bir değer üzerinde olan) öznitelikler gruplanmıştır. Ayrıca diğer özniteliklerle hiç korelasyonu olmayan öznitelikler listelenmiştir. Korelasyon matrisinin ilk hali Fig. 2'deki gibidir.

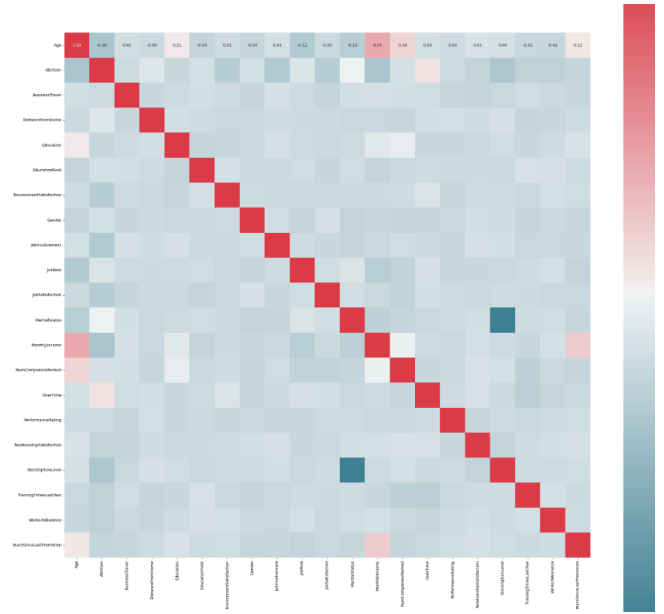


Fig. 4 Öznitelik Seçimi Sonrası Korelasyon Matrisi

Fig. 4 öznitelik seçimi sonrası korelasyon matrisini göstermektedir. Öznitelikler temizlenmiş ve son halini almıştır.

C. VERİ SETİNİ DENGELEME

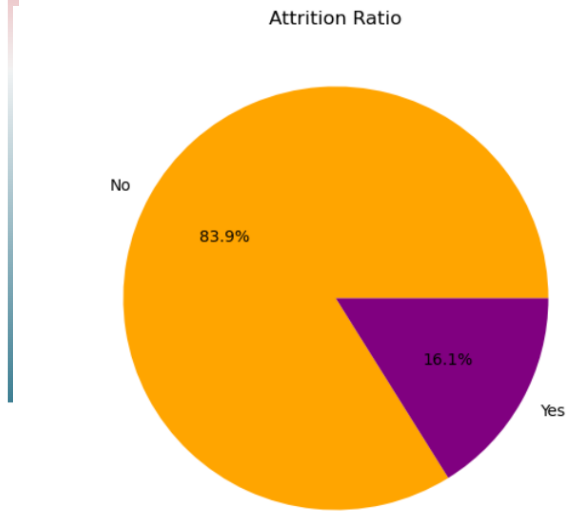


Fig. 5 Label Dağılımı

Veri setinde işten ayrılma olasılığı %16.12 oranına sahip. Veri setinde sınıfsal dengesizlik (imbalance) var. Modelin daha iyi sınıflandırma/tahmin yapabilmesi için çıktılarının dengelenmesi gerekiyor. Dengeleme için farklı yöntemler arasından SMOTE (Syntetic Minority Over-Sampling Technique)'u seçtim. SMOTE veri setinden özellikler seçerek en iyi olası örnekleri bulmak için örnekler arasında bir çizgi çeker. Over-sampling yönteminin kapsamlı bir versiyonudur

ve sentetik verilerin mevcut bir modelin tam bir kopyası olmadığından emin olur. Ayrıca sonucunda eksik değerler oluşturmaz.

V. KULLANILAN MODELLER

Bu projede, çalışanların işten ayrılma tahmini için farklı makine öğrenimi modelleri kullanılmıştır. Aşağıda kullanılan modellerin tanımları, metodolojileri ve neden bu modellerin seçildiği açıklanmıştır:

Veri seti, eğitim ve test olmak üzere iki parçaya ayrılmıştır. Veri setinin %80'i eğitim için kullanılmış ve %20'si test için ayrılmıştır. K en yakın komşu harici tüm modellerde çarpaz doğrulama yapılmış ve hyperparametre ayarlanmıştır. Çarpaz doğrulamada, 4 fold kullanıldığından veriler %60 eğitim %20 test ve %20 validasyon haline gelmiştir. Sonuçlar en iyi modele göre çıkarılmıştır.

A. SVM

SVM, veriyi iki sınıf arasında en iyi ayıran bir hiper düzlem bulmayı amaçlar. İkili sınıflandırma problemleri için kullanılır ve ayırma işlemi sırasında destek vektörlerini dikkate alır. Hiper düzlem, sınıflar arasındaki en büyük marjini sağlar. Doğrusal ve doğrusal olmayan verileri sınıflandırabilir. Kernel olarak linear ve rbf seçilmiştir. Belirtildiği gibi modelde hyper parametre tuning yapılmıştır. SVM tercih etmemin nedeni yüksek boyutlu veri setimin olmasıdır.

B. LOJİSTİK REGRESYON

Lojistik regresyon, bir bağımlı değişkenin olasılığını tahmin etmek için kullanılır. Çıktı olarak 0 ile 1 arasında bir olasılık değeri döner ve bir eşiğe göre sınıflandırma yapılır. Modelde bağımlı değişken işten ayrılmıştır. Sigmoid fonksiyonu kullanılarak tahmin edilen değer olasılık aralığına sıkıştırılmıştır. Yorumlama kolaylığından dolayı seçilmiştir.

C. K-EN YAKIN KOMŞU

K en yakın komşu algoritması, yeni bir veri noktasını sınıflandırmak için en yakın k komşusunun sınıflarına bakar. Temelde sınıflandırma problemlerinde kullanılır ve eğitim aşamasında bir model oluşturmaz. Öklidyen mesafesi ölçümü yapar.

$$d(i,j) = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2}$$

Model eğitilirken farklı komşu sayıları için eğitim sağlanmış ve en çok başarımlı aldığı komşu sayısı seçilmiştir. Kümeleme yaparak veri setini incelemek istediğim için bu model seçilmiştir.

D. ÇOK KATMANLI PERCEPTRON

Çok katmanlı perceptron, yapay sinir ağı modellerinden biridir. Bir veya daha fazla gizli katmana sahip olabilir ve her nöron diğer katmandaki nöronlara bağlıdır. MLP, geri yayılım (backpropagation) algoritması ile eğitilir. Doğrusal olmayan fonksiyonları öğrenmek için aktivasyon fonksiyonları ve ağırlıkların güncellenmesi için optimizasyon teknikleri kullanır. Bu projede, 2 katmanlı yapay sinir ağı ve aktivasyon fonksiyonu olarak tanh ve relu kullanılmıştır. Bu modelin seçilmesinin sebebi işten ayrılmalarındaki karmaşık ilişkilerin daha iyi öğrenilebileceğinin düşünülmesidir.

E. RASTGELE ORMAN

Rastgele orman, bir ensemble yöntemidir ve birden fazla karar ağacının bir araya gelmesiyle oluşturulur. Her ağaç, farklı veri alt kümeleri ve özelliklerle eğitilir. Son tahmin, tüm ağaçların ortalama sonucudur. Aynı zamanda hem doğrusal hem de doğrusal olmayan ilişkileri modelleyebilir. Her ağaç farklı ve random değerlerle eğitildiğinden overfit imkanı olmaz. Ama ne yazık ki fazla hesaplama gücü gerektirir. Daha hızlı çalışmasını sağlamak için projede paralelleştirme ve erken durdurma yöntemi kullanılmıştır. Ayrıca ağaçlarda entropy ve gini saflığı hesaplanmış ve bu metriklere göre ağaçlar oluşturulmuştur.

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2(p_i)^{1)}$$

$$Gini(t) = 1 - \sum_{i=1}^j P(i|t)^2$$

Bu modeli seçme sebebim, literatür araştırmalarında derin öğrenme harici en iyi sonuç veren model olmasıydı.

VI. SONUÇLARIN YORUMLANMASI

	Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
0	SVM (Train)	0.997465	0.997970	0.996957	0.997463	0.997465
1	Logistic Regression (Train)	0.816430	0.807087	0.831643	0.819181	0.816430
2	K-Nearest Neighbour (Train)	0.994422	1.000000	0.988844	0.994391	0.994422
3	Multi Layer Perceptron (Train)	0.919878	0.913174	0.927992	0.920523	0.919878
4	Random Forest (Train)	1.000000	1.000000	1.000000	1.000000	1.000000

Fig. 6 Eğitim Seti Metrik Sonuçları

	Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
0	SVM (Test)	0.896761	0.885827	0.910931	0.898204	0.896761
1	Logistic Regression (Test)	0.813765	0.811245	0.817814	0.814516	0.813765
2	K-Nearest Neighbour (Test)	0.917004	0.893130	0.947368	0.919450	0.917004
3	Multi Layer Perceptron (Test)	0.864372	0.860000	0.870445	0.865191	0.864372
4	Random Forest (Test)	0.890688	0.900415	0.878543	0.889344	0.890688

Fig. 7 Test Seti Metrik Sonuçları

Modeller için değerlendirme metriği olarak doğruluk (accuracy), hassasiyet (precision), geri çağırma (recall), fl puanı (f1 score), ROC eğrisi ve auc kullanılmıştır.

Fig.6'da eğitim verilerinin metrik sonuçları gösterilmiştir. Figürde de görüleceği üzere Rastgele Orman modeli eğitim setini tamamen doğru sınıflandırmıştır. Sonra, aslında tam eğitim denemeyen k en yakın komşu modeli ve SVM modeli %99 ile en iyi ikinci başarıyı elde etmiştir. Çok katmanlı perceptron modeli görece düşük başarımdadır. Yapay sinir ağı modeli olduğundan daha yüksek performans beklenen bu modelin bu kadar düşük performans göstermesinin nedeni veri setinin az sayıda satır içermesi olabilir. En son da lojistik regresyon modeli başarıyı sağlamıştır.

Fig.7'de test verilerinin metrik sonuçları gösterilmiştir. Figürde de görüleceği üzere k en yakın komşu modeli en yüksek başarıyı göstermiştir. Rastgele Orman modeli ve SVM modeli %89-91 ile en iyi ikinci başarıyı elde etmiştir. 4. olarak %86 ile çok katmanlı perceptron modeli bulunur. En son da yine lojistik regresyon modeli başarıyı sağlamıştır.

Yukarıda bahsettiğim başarımdan aslında sadece accuracy değeri değildir. Başarımdan bahsederken diğer metrikler de göz önüne alınmıştır.

Son olarak modelleri eğitmek için geçen süre hesaplandı. En uzun süre çok katmanlı perceptron eğitmek için geçmişken en az süre k en yakın komşu eğitimi için geçmiştir.

Ancak çok katmanlı perceptron ve rastgele orman modellerini eğitirken donanım sorunlarıyla karşılaştım. Bundan dolayı, değişkenleri azaltmak, maksimum iterasyon değerini azaltmak gibi yöntemlere başvurdum. Bu da hyper parametrelerin istenilen değerlerde ayarlanamamasıyla sonuçlanmış olabilir. Daha iyi bir bilgisayarla bu modellerde daha iyi başarımlar alınabilir.

VII. CONCLUSIONS

Bu projede, çalışan işten ayrılma tahmini gerçekleştirilmiştir. Bu tahminde, IBM veri analistlerinin oluşturduğu veri seti kullanılmıştır. Bu veri seti için öncelikle derinlemesine inceleme yapılmış, öznitelikler hakkında bilgi toplanılmıştır. Veri ön işleme, öznitelik seçimi gibi farklı yöntemlerle modellerin başarımları geliştirilmiştir. Bunlar yapılırken PCA, EDA, korelasyon matrisi gibi farklı yöntemler kullanılmıştır. Projede dengesiz veri seti konusunda neler yapılabileceği öğrenilmiştir ve SMOTE kullanılarak veriler dengelenmiştir. Klasik makine öğrenmesi modellerinden olan SVM ve lojistik regresyon modellerinin yanı sıra yapay sinir ağı modellerinden çok katmanlı perceptron eğitilmiştir. Ayrıca topluluk modeli

olan rastgele orman modeli de eğitilmiştir. Tüm bu modellerin eğitimi sırasında hyper parametre tuning hakkında bilgi sahibi olunmuştur. Ayrıca k en yakın komşu modeli de çalıştırılıp sonuçları karşılaştırılmıştır. Gelecek çalışmalarda klasik makine öğrenmesi modellerinden uzaklaşıp derin öğrenme tabanlı modeller incelenebilir.

REFERENCES

- [1] V. Mehta and S. Modi, "Employee Attrition System Using Tree Based Ensemble Method," *2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)*, Bangalore, India, 2021
- [2] K. K. Ramachandran, A. Apsara Saleth Mary, S. Hawladar, D. Asokk, B. Bhaskar, and J. R. Pitroda, "Machine learning and role of artificial intelligence in optimizing work performance and employee behavior," *Mater. Today Proc.*, vol. 51, pp. 2327–2331, 2022, doi: 10.1016/j.matpr.2021.11.544.
- [3] Valle, M.A., Varas, S., Ruz, G.A.: Job performance prediction in a call center using a naïve Bayes classifier. *Expert Syst. Appl.* 39, 9939–9945 (2012)
- [4] Patel, Adarsh, et al. "Employee attrition predictive model using machine learning." *International Research Journal of Engineering and Technology (IRJET)* 7.5 (2020).
- [5] Qutub, Aseel, et al. "Prediction of Employee Attrition Using Machine Learning and Ensemble Methods." *Int. J. Mach. Learn. Comput* 11 (2021).
- [6] R. Praveen, P. Pabitha, V. Sakthi, S. Madhavi and R. V. Sai, "Convolutional Neural Network and Capsule Network Fusion for Effective Attrition Classification," *2023 12th International Conference on Advanced Computing (ICoAC)*, Chennai, India, 2023, pp. 1-6, doi: 10.1109/ICoAC59537.2023.10249466.
- [7] V. Mehta and S. Modi, "Employee Attrition System Using Tree Based Ensemble Method," *2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)*, Bangalore, India, 2021
- [8] Qutub, Aseel, Asmaa Al-Mehmadi, Munirah Al-Hssan, Ruyan Aljohani and Hanan S. Alghamdi. "Prediction of Employee Attrition Using Machine Learning and Ensemble Methods." *International Journal of Machine Learning and Computing* 11 (2021): 110-114.
- [9] Poornapriya, T. & R., Dr. GOPINATH. (2021). Employee Attrition In Human Resource Using Machine Learning Techniques. *Webology*. 18. 2844-2856.
- [10] S. George, K. A. Lakshmi and K. T. Thomas, "Predicting Employee Attrition Using Machine Learning Algorithms," *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Greater Noida, India, 2022
- [11] S. R. Ahmed, A. K. Ahmed and S. J. Jwmaa, "Analyzing The Employee Turnover by Using Decision Tree Algorithm," *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Istanbul, Türkiye, 2023, pp. 1-4, doi: 10.1109/HORA58378.2023.10156709.
- [12] M. John, M. Alharbi, N. Bin Daham and S. Abu Theeb, "Analyzing Employee Attrition: A Visual Perspective," *2024 Seventh International Women in Data Science Conference at Prince Sultan University (WiDS PSU)*, Riyadh, Saudi Arabia, 2024, pp. 28-31, doi: 10.1109/WiDS-PSU61003.2024.00020.
- [13] A. I. Al-Alawi and Y. A. Ghanem, "Predicting Employee Attrition Using Machine Learning: A Systematic Literature Review," *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETIS)*, Manama, Bahrain, 2024, pp. 526-530, doi: 10.1109/ICETIS61505.2024.10459451.