



## Gelişen Teknoloji ile Bankacılık Sektöründe Veri Analitiği: Müşteri Kaybı Tahmini İçin Makine Öğrenmesi Yaklaşımları

Büşra ÖZCAN<sup>1</sup> (0009-0001-4455-5505)

Kübra KAYAPINAR<sup>1</sup> (0009-0007-0433-2554)

Kemal ADEM<sup>2\*</sup> (0000-0002-3752-7354)

<sup>1</sup>Sivas Bilim ve Teknoloji Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Savunma Teknolojileri Bölümü, 5800, Sivas, Türkiye

<sup>2</sup>Sivas Bilim ve Teknoloji Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği Bölümü, 5800, Sivas, Türkiye

\*Sorumlu yazar (Corresponding author): kemaladem@sivas.edu.tr

**Geliş Tarihi (Received):** 15.06.2023

**Kabul Tarihi (Accepted):** 28.07.2023

### Özet

Günümüzde gelişen teknoloji ile birlikte banka sektörü daha da gelişmiş ve verilerden çokça faydalanır duruma gelmişlerdir. Artan verileri gelişen teknoloji ile birlikte değerlendirmeleri ve kendi faydalarına kullanmaları elzemdir. Bu makalede makine öğrenmesi yöntemleri ile banka müşterilerine ait olan bir veri seti değerlendirilmiştir. Veri seti üzerinde bankanın müşteriyi kayıp edip etmeyeceği tahmin edilmiştir. Bu çalışma kapsamında kullanılan yöntemler sırasıyla Rastgele Orman, Karar Ağacı, Gauss, K-En Yakın Komşu, Adaboost ve Lojistik Regresyon algoritmalarıdır. Modeller karşılaştırılmadan önce en iyi sonuçları alabilmek ve eşit bir karşılaştırma sağlayabilmek için modeller üzerinde hiperparametre optimizasyonu uygulanmıştır. Bu çalışmada Rastgele Orman algoritması karşılaştırılan diğer algoritmalarından daha iyi sonuçlar elde etmiştir. Karşılaştırılan diğer algoritmaların başarı sonuçları ise sırasıyla K-En Yakın Komşu, Karar Ağacı, Adaboost, Gauss ve Lojistik Regresyon şeklinde belirlenmiştir.

**Anahtar Kelimeler:** Banka, müşteri, yapay zekâ, ağaç algoritmaları

### Data Analytics in the Banking Sector with Emerging Technology: Machine Learning Approaches for Customer Loss Prediction

#### Abstract

With the advancement of technology, the banking sector has experienced significant development and has become highly reliant on data. It has become crucial for banks to assess and utilize the increasing volume of data with the aid of advancing technology. In this article, a dataset belonging to bank customers was evaluated using machine learning methods. The objective was to predict whether the bank would lose a customer based on the given dataset. The methods employed in this study were Random Forest, Decision Tree, Gaussian Naive

Bayes, K-Nearest Neighbors, Adaboost and Logistic Regression algorithms respectively. Prior to comparing the models, hyperparameter optimization was applied to ensure obtaining the best possible results and enable a fair comparison among the models. Among the algorithms compared in this study, the Random Forest algorithm yielded superior results. The success rates of the other compared algorithms were determined as follows: K-Nearest Neighbors, Decision Tree, Adaboost, Gaussian Naiva Bayes and Logistic Regression.

**Keywords:** Bank, customer, artificial intelligence, tree algorithms

---

## GİRİŞ

Şu anda, dünya çapında çok sayıda hizmet sağlayıcının, özellikle de bankaların mevcudiyeti nedeniyle pazar dinamik ve oldukça rekabetçidir (AL-Najjar ve ark., 2022). Bankalar günümüzde, insanların paralarını saklayabilmesi için en güvenilir ortamlar haline gelmiştir. Rekabetin arttığı günümüz dünyasında bir bankanın güçlü bir şekilde hizmet verebilmesi ve piyasada kalabilmesi için müşteri çok önemlidir. Uzun vadeli müşteriler doğrudan kar üretimine bağlıdır; dolayısıyla bankalar müşteri kaybetmekten kaçınmalıdır (Kim ve ark., 2005). Harvard Business Review, müşteri sayısındaki %5'lik bir sapmanın, firmalar için %25 ile %85 arasında bir kâr artışına yol açabileceğine inanıyor (Keramati ve ark., 2016). Müşteri kaybı, bir müşterinin bir rakibe karşı kaybedilmesi ve bu da kârda kayıplara yol açması olarak tanımlanabilir. Rakip bir bankaya geçmesi muhtemel olan müşterileri önceden belirlemek önemlidir (Domingos ve ark., 2021). Nüfus yoğunluğunun ivmelenerek arttığı bu dönemde nüfusun artması demek müşteri sayısının artması demektir. Günümüzdeki bu rekabetçi ortamda müşterilere hızlı bir şekilde hizmet verebilmek, tüm müşterilere ulaşabilmek ve var olan müşteriyi kaybetmemek için müşteriyi iyi tanımalı aynı zamanda hızlı ve güvenilir olunmalıdır. Gelişen teknoloji ile birlikte bankalar da bu sürece ayak uydurmuş ve elde ettikleri verileri saklamaya başlamışlardır. Sürekli artan veri ve bu verileri analiz etme ihtiyacı her geçen gün artmaktadır. Bu nedenle sürekli gelişmekte olan yapay zeka algoritmaları, artan veri yoğunluğu sebebiyle bu verileri hızlı ve istatistiksel olarak güvenilir bir şekilde analiz edebilmek için tercih edilmektedir (Keramati ve ark., 2016).

Bu çalışmada bankaların ihtiyaç duyduğu analizler doğrultusunda müşteri kaybı tahmini yapmak için makine öğrenmesi yöntemleri kullanılması hedeflenmiştir. Yapılan literatür çalışmaları sonucunda ağaç algoritmalarının daha iyi başarı oranı verdiği gözlemlenmiştir. Bu çalışmada ağaç algoritmaları ve diğer makine öğrenmesi yöntemleri değerlendirilecektir.

### Literatür taraması

Gerçekleştirilen literatür taraması sonucunda Caigny ve ark. (2018) tarafından yapılan çalışmada bir sınıflandırma algoritması olan Logit Yaprak Modeli algoritmasını önermişlerdir. Yapılan çalışmada 14 farklı veri seti kullanılmıştır. Finans alanında kullanılan veri setleri en az 100,000 veriden oluşmaktadır. Çalışmada Logit Yaprak Model algoritması denenmiş ve bilinen diğer yöntemlerle karşılaştırılmıştır. Performans ve doğruluk açısından Rastgele Orman algoritması ile yakın sonuçlar elde etmiştir. Rastgele Orman algoritması AUC değeri 0.644 iken Logit Yaprak Model algoritması AUC değeri 0.630 olarak elde edilmiştir. Amuda ve Adeyemo (2020) tarafından yapılan çalışmada kullanılan veri setinde Çin bankasına ait 50,000 müşteriden toplanmış 46,406 veri bulunmaktadır. Çalışmada Lojistik Regresyon, Karar Ağacı, K-En Yakın

Komşu ve Rastgele Orman gibi farklı algoritmalar kullanılmış ve karşılaştırılmıştır. Rastgele Orman algoritması %97.53 sonucunu vermiştir. Rahman ve Kumar (2020) tarafından yapılan çalışmada K-En Yakın Komşu, Destek Vektör Makinesi, Karar Ağacı ve Rastgele Orman sınıflandırma algoritmaları kullanılmıştır. Kullanılan veri seti Çin bankasına ait 50,000 farklı müşteriye ait verileri içermektedir. Veri seti üzerinde yapılan ön çalışmalar sonunda modele dahil edilen 46,406 veri kalmıştır. Yüksek hız ile birlikte %95.74 oranı ile en yüksek sonucu Rastgele Orman algoritması vermiştir. Kaur ve Kaur (2020) tarafından yapılan çalışmada Lojistik Regresyon, Karar Ağacı, K-En Yakın Komşu ve Rastgele Orman algoritmaları karşılaştırılmıştır. Çalışmaya dahil edilen veri seti 28,382 adet veri içermektedir. Yapılan çalışma sonucunda kullanılan algoritmalar arasında en yüksek doğruluk oranını %85.22 doğruluk oranı ile Rastgele Orman algoritması vermiştir. Guliyev ve Tatoğlu (2021) tarafından yapılan çalışmada düşük örnekleme yöntemleri ve ağaç tabanlı algoritmalar önerilmiştir. Yapılan çalışma için kullanılan veri seti 274,542 veriden oluşmaktadır. Karar Ağacı, Rastgele Orman ve XgBoost algoritmaları uygulanmış ve karşılaştırılmıştır. XgBoost %96.97 ROC eğrisi altındaki alan değeri ile her metrik için en iyi performansı göstermiştir. (Zaky ve ark., 2022) yaptığı çalışmada müşteri kaybını tahmin etmek için YSA (Yapay Sınır Ağları) algoritması kullanılmıştır. Çalışma sonucunda %87 doğruluk oranı elde edilmiştir. (Karvana Manik Gde Ketut ve ark., n.d.) tarafından yapılan çalışmada müşteri kaybı tahmini için makine öğrenimi modeli üreten veri madenciliğinden bir sınıflandırma tekniği kullanılmıştır. Çalışmada 5 farklı sınıflandırma yöntemi test edilmiştir. Bu yöntemler Karar Ağacı, Sınır Ağları, Destek Vektör Makinesi, Hassas Bayes ve Lojistik Regresyondur. Deneyler birden fazla kez gerçekleştirilmiş ve Destek Vektör Makinesi %73,57 doğruluk oranı ile kayıp müşteri tahmininde en iyi model olarak belirtilmiştir. Haddadi ve ark. (2022) yapmış oldukları çalışmada İran'da bulunan banka müşterilerinin verilerinin toplandığı bir veri seti kullanılmıştır. Önerdikleri model, diğer bilinen modeller ile karşılaştırılmıştır. Karşılaştırılan diğer modeller Karar Ağacı algoritması, Hassas Bayes ve Lojistik Resresyon algoritmalarıdır. Önermiş oldukları Bi-LSTM modeli %84 doğruluk oranı ile diğer modellerden daha yüksek bir doğruluk oranı elde etmiştir. Deng ve ark. (2021) tarafından yapılan çalışmada bankaların kullanıcı verilerini analiz etmek ve doğruluğu yükseltmek için Catboost, Lightgbm ve Rastgele Orman gibi algoritmalar kullanılarak bir model oluşturulmuştur. Oluşturulan model ile yapılan deneyler sonucunda Rastgele Orman modeli doğruluk oranı %91 ile en iyi model olarak belirlenmiştir.

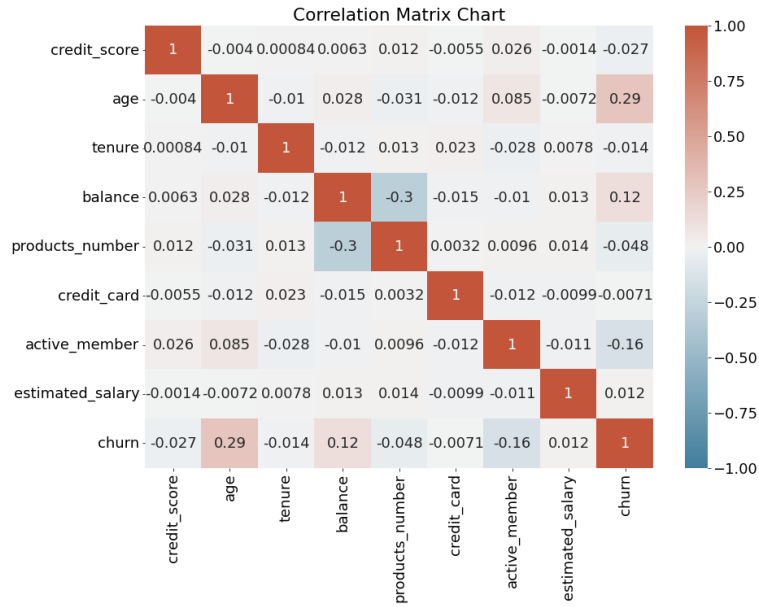
## MATERYAL VE YÖNTEM

Bu çalışmada Kaggle'dan alınan 'Bank Customer Churn Dataset' veri seti kullanılacaktır. Bu veri seti 10000 adet veri içermektedir ve 12 adet öznelikten oluşmaktadır. Tablo 1'de veri setinde bulunan özelliklerin detaylı tablosu gösterilmektedir.

**Tablo 1.** Veri setinde bulunan ilk 5 veriye ait bilgiler

Customer Id	Credit Score	Country	Gender	Age	Tenure	Balance	Products Number	Credit Card	Active Member	Estimated Salary	Churn
15634602	619	France	Female	42	2	0.00	1	1	1	101348.88	1
15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
15619304	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
15701354	699	France	Female	39	1	0.00	2	0	0	93826.63	0
15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0

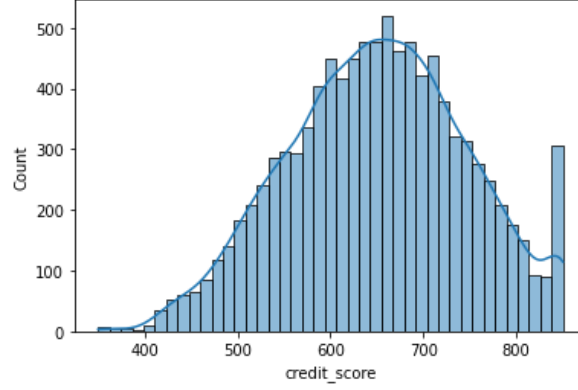
Özneliklerin birbirleri ile olan ilişkisini analiz etmek için kullanılan korelasyon matrisi grafiği Şekil 1'de verilmiştir. Burada kırmızı renkler pozitif korelasyonu temsil ederken; mavi renkler negatif korelasyonu temsil eder.



**Şekil 1.** Korelasyon matrisi grafiği

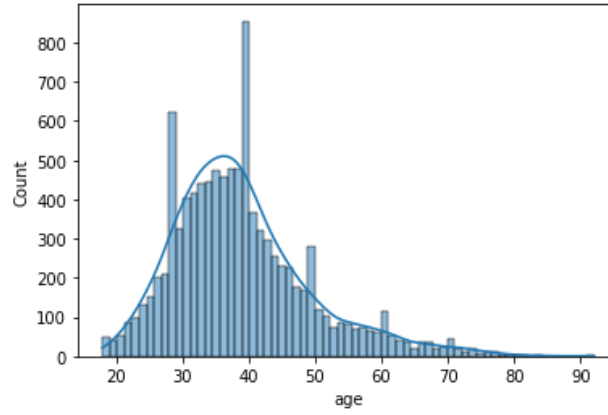
Şekil 2.'de ki görsel incelendiğinde 'credit\_score' özelliğinin dağılımı çan eğrisi şeklinde olduğu gözlemlenmektedir. Yapay zeka algoritmalarında verinin çan eğrisi şeklinde olması,

algoritmanın daha iyi öğrenme ve tahmin yapma performansı göstermesine sebep olur. Bunun nedeni çan eğrisi şeklindeki verilerin simetrik olmasıdır. Simetrik veri algoritmanın öğrenme sürecinde daha kolay işlenir ve anlaşılır. Bu sayede daha kolay tahminler yapılır.



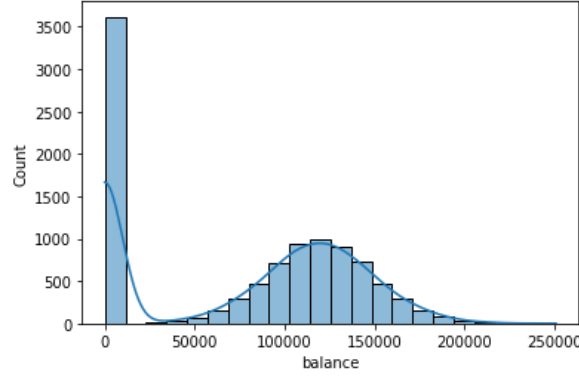
**Şekil 2.** Credit Score özelliği grafiksel gösterimi

Şekil 3 incelendiğinde 'age' özelliğinin de 'credit\_score' özelliği gibi çan eğrisi dağılımına sahip olduğu gözlemlenmektedir.



**Şekil 3.** Age özelliği grafiksel gösterimi

Şekil 4'te ki görsel de çan eğrisi dağılımına sahiptir. Bu dağılıma sahip olması verinin simetrik olmasının yanında merkezi eğiliminin de güçlü olmasına anlamına gelmektedir. Bu, verilerin ortalamaya daha yakın olduğunu ifade etmektedir. Bu algoritmanın daha kolay bir şekilde ortalama değeri tahmin etmesine yardımcı olur.



Şekil 4. Balance özelliği grafiksel gösterimi

Çalışmada kullanılacak veri setindeki özellikler incelendiğinde önce gereksiz özellikler sonucu kötü etkilememesi için kaldırılmıştır. Veri setindeki özellikler şu şekildedir; ‘customer\_id’, ‘credit\_score’, ‘country’, ‘gender’, ‘age’, ‘tenure’, ‘balance’, ‘products\_number’, ‘credit\_card’, ‘active\_member’, ‘estimated\_salary’ ve ‘churn’. Bu özelliklerden ‘churn’ özelliği tahmin etmeye çalıştığımız özelliştir. Geriye kalanlar dışında ‘customer\_id’ bir sıra numarası olduğu ve anlam ifade etmediği için listeden silinmiştir. Kalan on adet özellik tahmin işlemine dahil edilmiştir. Çalışma kapsamında veri seti üzerinde Rastgele Orman, Karar Ağacı, Lojistik Regresyon, Gauss ve K-En Yakın Komşu algoritmaları uygulanmıştır.

Rastgele Orman modeli birden çok karar ağacının bir araya getirilmesiyle oluşan bir modeldir (Biau ve Scornet, 2016). Karar Ağacı algoritması verileri karar ağacı şeklinde temsil eden bir algoritmadır. Her düğümde bir özellik ve o özellik için belirli bir değer aralığı kontrol edilir ve düğümler arası geçiş yapılır (Canete-Sifuentes ve ark., 2021). Lojistik Regresyon algoritması verilen değişkenlere bağlı olarak bir sınıfın olasılığını hesaplar ve en yüksek olasılık değerine sahip olan sınıfı tahmin eder (Connelly). Gauss algoritması bayes teoremine ve özelliklerin bağımsız olduğu varsayımına dayanan bir gauss dağılımı izleyen naive bayes algoritmalarından biridir (Schulz ve ark., 2018). K-En Yakın Komşu algoritması veri noktalarının k adet en yakın komşusunun sınıfını tahmin etmek için kullanılır (Zhang ve ark., 2018).

## BULGULAR VE TARTIŞMA

Deneyler python programlama dili ile gerçekleştirilmiştir. Veri sayısını arttırmak için hazır bir fonksiyon olan RandomOverSampler fonksiyonundan faydalanılmıştır. Bu aşırı örnekleme olarak adlandırılmaktadır. Kullanılacak veri seti tahmin yapılacak sınıf dağılımı incelendiğinde orantısız olduğu gözlemlenmiştir. Churn değişkeni ‘yes’ (bankayı terk edecek müşteri) sayısı veri arttırma işlemi öncesinde 1648, ‘no’ (bankayı terk etmeyecek müşteri) sayısı 6352’dir. Çıkacak sonuçların doğru olabilmesi için ‘yes’ ve ‘no’ sayılarının dengeli olması gerekmektedir. Bu sebeple veri arttırma işlemi uygulanmıştır. Bu uygulama sonucunda churn

değişkeni ‘yes’ veri sayısı 1648 ‘den 6352’ye çıkarılmıştır ve ‘no’ sayısı 6352 olarak kalmıştır. Verilerin %20’si test için ayrılmıştır. Çalışmada kullanılacak algoritmalar sırasıyla Rastgele Orman, Karar Ağacı, Lojistik Regresyon, Gauss, K-En Yakın Komşu ve Adaboost algoritmalarıdır. Bu modeller uygulanırken, model parametreleri en iyi duruma optimize edilmişlerdir.

Rastgele Orman modeli uygulanmadan önce  $n\_estimators$ ,  $max\_depth$ ,  $min\_samples\_split$ ,  $min\_samples\_leaf$  özelliklerine göre optimize edilmiştir. Optimizasyon sonucunda  $max\_depth$  değeri None  $min\_samples\_leaf$  değeri 1,  $min\_samples\_split$  değeri 2 ve  $n\_estimatros$  değeri 100 olarak seçilmiştir. Çalışmada Karar Ağacı algoritması  $max\_depth$ ,  $min\_samples\_split$  ve  $min\_samples\_leaf$  parametrelerine göre optimize edilmiştir. Modelin optimize edilmiş durumuna göre  $max\_depth$  değişkeninin değeri None,  $min\_samples\_leaf$  değeri 1,  $min\_samples\_split$  değeri 2 olarak elde edilmiştir. Lojistik Regresyon modeli  $penalty$ ,  $C$ ,  $solver$  ve  $max\_iter$  değişkenlerine göre optimize edilmiştir. Modelin optimize edilmiş durumdaki hiperparametreleri  $penalty$  için none,  $C$  için 0.001,  $solver$  için ‘sag’ ve  $max\_iter$  parametresi için 500’dür. Gauss algoritması optimizasyonu için  $var\_smoothing$  değişkeni ele alınmış ve değeri 1e-09 olarak elde edilmiştir. K-En Yakın Komşu modeli için hiper parametre optimizasyonu yapılmıştır ve  $n\_neighbors$ ,  $weights$ ,  $metric$ ,  $algorithm$  parametreleri optimize edilmiştir. Elde edilen optimum model için parametreler ve değişkenleri  $n\_neighbors$  1,  $metric$  manhattan,  $weights$  uniform,  $algorithm$  brute şeklinde elde edilmiştir.

**Tablo 3.** Modellere ait hata metrikleri

Model	Precision		Recall		F1-Score		Accuracy
	0	1	0	1	0	1	
LogisticRegression	0.90	0.39	0.72	0.70	0.80	0.50	0.71
GaussNB	0.91	0.44	0.77	0.69	0.83	0.54	0.76
KNeighborsClassifier	0.87	0.54	0.89	0.51	0.88	0.53	0.81
RandomForestClassifier	0.88	0.65	0.92	0.54	0.90	0.59	0.84
DecisionTreeClassifier	0.86	0.51	0.89	0.45	0.87	0.48	0.80
AdaBoostClassifier	0.91	0.51	0.82	0.71	0.86	0.59	0.80



Tablo incelendiğinde müşterilerin bankayı bırakıp bırakmayacaklarını tahmin etmede en iyi doğruluk oranını %84 ile Rastgele Orman algoritması vermiştir. Diğerleri sırası ile %81 doğruluk oranı ile K-En Yakın Komşu, %80 doğruluk oranı ile karar ağacı algoritması ve Adaboost algoritması, %76 oranı ile Gauss algoritması ve en kötü sonucu %71 doğruluk oranı ile veren Lojistik Regresyon algoritması olmuştur. Müşterilerin bankayı terk etmesi durumunu sınıflandırırken hassasiyet, f1-skor ve duyarlılık oranları bankayı terk etmeyecek müşteri durumuna göre düşük gözlemlenmektedir. Bunun nedeni veri setinde Churn sütununda ‘yes’ olarak belirtilmiş gerçek veri sayısının az olmasıdır. Rastgele Orman algoritması, ensemble (birleşik) bir öğrenme yöntemi kullanmasından dolayı güçlü bir algoritmadır. Bu algortmada, birden çok karar ağacı birleştirilerek tahminlerde bulunur ve sonuçları birleştirir. Bu süreçte farklı ağaçların güçlü yönleri bir araya getirilirken zayıf yönlerini telafi eder bu sayede daha doğru tahminler yapar. Bu nedenle kullanılan veri seti üzerinde diğer algoritmalara kıyasla daha yüksek doğruluk elde etmiştir.

## **SONUÇ**

Bu çalışmada yapay zekâ uygulamalarından olan makine öğrenmesi yöntemleri ile banka müşterilerine ait olan bir veri seti değerlendirilmiştir. Veri seti üzerinde bankanın müşteriye kayıp edip etmeyeceği tahmin edilmiştir. Bu çalışma kapsamında kullanılan yöntemler sırasıyla Rastgele Orman, Karar Ağacı, Gauss, K-En Yakın Komşu ve Adaboost algoritmalarıdır. Modeller karşılaştırılmadan önce en iyi sonuçları alabilmek ve eşit bir karşılaştırma sağlayabilmek için modeller üzerinde hiperparametre optimizasyonu uygulanmıştır. Yapılan literatür çalışmalarında da görüldüğü üzere Rastgele Orman algoritması en yüksek doğruluk değerini veren algoritma olmuştur. Bu çalışma kapsamında yapılan deneyler sonucunda da Rastgele Orman algoritması %84 oranı ile en yüksek doğruluk oranını veren algoritma olmuştur. Diğer algoritmaların başarıları sırasıyla K-En Yakın Komşu, Karar Ağacı, Adaboost, Gauss ve Lojistik Regresyon şeklindedir. Yapılan literatür çalışmaları ve deneyler sonucunda ağaç algoritmalarının kullandığımız veri seti üzerinde daha başarılı olduğu gözlemlenmektedir. Yapılan bu çalışma için veri setinde en iyi sonucu verecek algoritma olarak Rastgele Orman algoritması önerilmektedir. Elde edilen doğruluk oranlarını yükseltmek için gelecek çalışmalarda veri seti üzerindeki dengesizlikleri gidermek için farklı yaklaşımlar uygulanacak ve sonuçlar karşılaştırılacaktır.

## **Teşekkür**

Bu çalışma için kullanmış olduğumuz veri setini Kaggle platformunda paylaşan Gurav Topre'ye teşekkür ederiz.

**KAYNAKÇA**

AL-Najjar, D., Al-Rousan, N., AL-Najjar, H. 2022. Machine Learning to Develop Credit Card Customer Churn Prediction. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(4): 1529–1542.

Amuda, K.A., Adeyemo, A.B. 2020. Customers Churn Prediction in Financial Institution Using Artificial Neural Network. <https://arxiv.org/abs/1912.11346>

Biau, G., Scornet, E. 2016. A random forest guided tour. *Test*, 25(2): 197–227.

Canete-Sifuentes, L., Monroy, R., Medina-Perez, M.A. 2021. A Review and Experimental Comparison of Multivariate Decision Trees. *IEEE Access*, 9: 110451–110479.

Connelly, L. (n.d.). *Logistic Regression* (Vol. 29, Issue 5).

De Caigny, A., Coussement, K., De Bock, K.W. 2018. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2): 760–772.

Deng, Y., Li, D., Yang, L., Tang, J., Zhao, J. 2021. Analysis and prediction of bank user churn based on ensemble learning algorithm. *Proceedings of 2021 IEEE International Conference on Power Electronics, Computer Applications, ICPECA 2021*, 288–291.

Domingos, E., Ojeme, B., Daramola, O. 2021. Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. *Computation*, 9(3).

Guliyev, H., Yerdelen Tatoğlu, F. 2021. Customer churn analysis in banking sector: Evidence from explainable machine learning models. *Journal of Applied Microeconometrics*, 1(2), 85–99.

Haddadi, S.J., Mohammadi, M.O., Bahrami, M., Khoeini, E., Beygi, M., Khoshkar, M. H. 2022. Customer Churn Prediction in the Iranian Banking Sector. *2022 International Conference on Applied Artificial Intelligence, ICAPAI 2022*.

Karvana, K.G.M., Yazid, S., Syalim, A., Mursanto, P. 2019. Customer churn analysis and prediction using data mining models in banking industry. In *2019 international workshop on big data and information security (IWBIS)* pp. 33-38. IEEE.

Kaur, I., Kaur, J. 2020. Customer churn analysis and prediction in banking industry using machine learning. *PDGC 2020 - 2020 6th International Conference on Parallel, Distributed and Grid Computing*, 434–437.

Keramati, A., Ghaneei, H., Mirmohammadi, S.M. 2016. Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, 2(1).

Kim, S., Shin, K.S., Park, K. 2005. An Application of Support Vector Machines for Customer Churn Analysis: Credit Card Case. In *LNCS* (Vol. 3611).

Rahman, M., Kumar, V. 2020. Machine Learning Based Customer Churn Prediction in Banking. Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020, 1196–1201.

Schulz, E., Speekenbrink, M., Krause, A. 2018. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85, 1–16.

Zaky, A., Ouf, S., Roushdy, M. 2022. Predicting Banking Customer Churn based on Artificial Neural Network. 5th International Conference on Computing and Informatics, ICCI 2022, 132–139.

Zhang, S., Cheng, D., Deng, Z., Zong, M., Deng, X. 2018. A novel kNN algorithm with data-driven k parameter computation. *Pattern Recognition Letters*, 109, 44–54.