T.C.
İSTANBUL GEDİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

KREDİ KARTI DOLANDIRICILIK TESPİTİNDE MAKİNE ÖĞRENME ALGORİTMALARININ KARŞILAŞTIRMALI ANALİZİ

YÜKSEK LİSANS TEZİ

Kemal ÇİLBURUNOĞLU

Yapay Zekâ Mühendisliği Anabilim Dalı

Yapay Zekâ Mühendisliği Tezli Yüksek Lisans Programı

EYLÜL 2023
İSTABUL

T.C.
İSTANBUL GEDİK UNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

KREDİ KARTI DOLANDIRICILIK TESPİTİNDE MAKİNE ÖĞRENME ALGORİTMALARININ KARŞILAŞTIRMALI ANALİZİ

YÜKSEK LİSANS TEZİ

Kemal ÇİLBURUNOĞLU
210039008
0009-0003-7419-2041

Yapay Zekâ Mühendisliği Anabilim Dalı

Yapay Zekâ Mühendisliği Tezli Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Şerife Esra DİNÇER

İstanbul 2023

T.C.
İSTANBUL GEDİK ÜNİVERSİTESİ
Lisansüstü Eğitim Enstitüsü Müdürlüğü

Jüri Tez Onay Formu

07.09.2023

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜ

Bu çalışma 07.09 2023 tarihinde aşağıdaki jüri tarafından Yapay Zeka Mühendisliği Anabilim Dalı, Yapay Zeka Mühendisliği (Tezli Yüksek Lisans) Programı Yüksek Lisans Tezi olarak kabul edilmiştir.

TEZ JÜRİSİ

Dr. Öğr. Üyesi Şerife Esra DİNÇER
Danışman
İstanbul Gedik Üniversitesi

Dr. Öğr. Üyesi Feridun Cemal ÖZÇAKIR.
Üye (İmza)
İstanbul Gedik Üniversitesi

Dr. Öğr. Üyesi Ümit ÖZTÜRK.
Üye (İmza)
İstanbul Beykent Üniversitesi

YEMİN METNİ

Yüksek Lisans tezi olarak sunduğum “Kredi Kartı Dolandırıcılık Tespitinde Makine Öğrenme Algoritmalarının Karşılaştırmalı Analizi” adlı çalışmanın, bilimsel ahlak ve geleneklere uygun şekilde tarafımdan yazıldığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, yararlandığım eserlerin tamamının kaynaklarda gösterildiğini ve çalışmamın içinde kullanıldıkları her yerde bunlara atıf yapıldığını, patent ve telif haklarını ihlal edici bir davranışımın olmadığını belirtir ve bunu onurumla doğrularım (07/09/2023).

Kemal ÇİLBURUNOĞLU

ÖNSÖZ

Finans sektörünün hızlı büyümesi ve teknolojik gelişmeler, dolandırıcılık faaliyetlerinin artmasına yol açmıştır. Bu nedenle, dolandırıcılık tespit yöntemlerinin geliştirilmesi, güvenli ve sürdürülebilir finansal sistemlerin temelini oluşturur. "Kredi Kartı Dolandırıcılık Tespitinde Makine Öğrenme Algoritmalarının Karşılaştırmalı Analizi" başlıklı bu tez, dolandırıcılık tespitinde kullanılan çeşitli makine öğrenme algoritmalarını incelemekte ve performanslarını karşılaştırmaktadır.

Bu çalışmanın temel amacı, dolandırıcılık tespitinde kullanılan mevcut makine öğrenme algoritmalarını değerlendirmek, en etkili yöntemleri belirlemek ve bu alandaki bilgi tabanını genişletmektir. Tez, teorik çerçeveyi araştıracak ve gerçek veri seti ile uygulamalı analizleri yapacaktır.

Bu süreç boyunca, dolandırıcılık tespitinde kullanılan makine öğrenme algoritmaları hakkında içgörü kazanmak için başlangıçta geniş bir literatür incelemesi yapılacaktır. Daha sonra, seçilen algoritmaların performansı karşılaştırılacak ve bulgulara dayanarak bu alana ilişkin öneriler sunulacaktır.

Bu tez boyunca bana yol gösteren ve destek olan değerli danışmanım Dr. Şerife Esra Dinçer'e teşekkürlerimi sunarım. Ayrıca, sabrı ve desteği için sevgili eşim Habibe Çilburunoğlu'na ve 1.5 yaşındaki kızım Ece Çilburunoğlu'na gönülden teşekkürlerimi sunarım.

Umarım bu tez, dolandırıcılık tespitinde makine öğrenme algoritmaları alanındaki bilgilere katkıda bulunur ve gelecekteki araştırmalara ilham verir.

Eylül 2023 Kemal ÇİLBURUNOĞLU

İÇİNDEKİLER

Sayfa

KISALTMALAR

ADASYN : Adaptive Synthetic
AI : Artificial Intelligence
ANN : Artificial Neural Networks
ARIMA : AutoRegressive Integrated Moving Average
AUC : Area Under the Curve
CNN : Convolutional Neural Networks
GAN : Generative Adversarial Networks
GBM : Gradient Boosting Machines
GloVe : Global Vectors for Word Representation
k-NN : k-Nearest Neighbors
LSTM : Long Short-Term Memory
MLP : Multilayer Perceptrons
NCR : Tomek Lin*k*s and Neighborhood Cleaning Rule
PCA : Principal Component Analysis
RFE : Recursive Feature Elimination
RFE : Recursive Feature Elimination
RNN : Recurrent Neural Networks
ROS : Random Over-Sampling
RUS : Random Under-Sampling
SARIMA : Seasonal AutoRegressive Integrated Moving Average
SMOTE : Synthetic Minority Over-sampling Technique
SVM : Support Vector Machines

ÇİZELGE LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

KREDİ KARTI DOLANDIRICILIK TESPİTİNDE MAKİNE ÖĞRENME ALGORİTMALARININ KARŞILAŞTIRMALI ANALİZİ

ÖZET

Finansal sektörün hızla büyümesi ve dijital işlemlerin yaygınlaşması dolandırıcılık faaliyetlerinin artmasına yol açmış ve dolayısıyla dolandırıcılığı tespit etme yöntemlerinin geliştirilmesi gerekliliğini ortaya çıkarmıştır. Bu tez kredi kartı dolandırıcılığını tespit etmek için çeşitli makine öğrenimi ve derin öğrenme tekniklerinin etkinliğini incelemeyi amaçlamaktadır. Lojistik Regresyon, Destek Vektör Makineleri (SVM), Karar Ağaçları, Rastgele Orman ve Yapay Sinir Ağları (ANN) yöntemleri kullanılmıştır.

Veri kümesi Worldline ve Brüksel Libre Üniversitesi (ULB) işbirliği ile toplanmıştır. 284,807 işlem içeren veri kümesinde dolandırıcılık işlemleri toplam işlemlerin %0.172'sine denk gelmektedir. Bu dengesizlik, veri kümesinin yapısını vurgulamaktadır.

Lojistik Regresyon ve Destek Vektör Makineleri %95 doğruluk oranları elde etmiştir. Karar Ağaçları ve Rastgele Orman %98 ve %97 doğruluk oranları ile yüksek performans sergilemektedir. Yapay Sinir Ağları (ANN) %100 doğruluk oranı ile öne çıkmaktadır.

Modellerin avantajları ve dezavantajları bulunmaktadır. Eğitim süresi, karmaşıklık, anlaşılabilirlik gibi faktörler göz önünde bulundurulmalıdır. Yanlış pozitif ve negatif oranlarının dengelemesi de modelin pratik kullanımını ve maliyetini etkileyebilmektedir.

Sonuç olarak bu çalışma kredi kartı dolandırıcılığını tespit etmek için makine öğrenimi ve derin öğrenme modellerini incelemiş olup Karar Ağaçları ve Random Forest gibi modellerin yüksek doğruluk oranı ve dengeli sonuçlar elde etmesiyle literatüre yeni bir katkı sunmaktadır. Farklı algoritmaların farklı performanslar sergilediği gözlemlenmiş ve model seçiminin algoritmanın özelliklerine ve veri kümesine bağlı olarak yapılması gerektiği vurgulanmaktadır. Bu çalışma finansal kurumların maliyet tasarrufu sağlamasına ve güvenli alışveriş deneyimleri sunmasına katkı sağlayacaktır.

Anahtar Kelimeler: *Dolandırıcılık Tespiti, Kredi Kartı Dolandırıcılığı, Makine Öğrenimi*

COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR CREDIT CARD FRAUD DETECTION

ABSTRACT

As the financial sector continues to grow rapidly and digital transactions become more widespread, the need for effective fraud detection methods is paramount. This thesis aims to examine the effectiveness of various machine learning and deep learning techniques in detecting credit card fraud. Methods such as Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forest, and Artificial Neural Networks (ANN) will be explored within this scope.

The dataset has been collected through a collaboration between Worldline and Universite Libre de Bruxelles (ULB). Within the dataset containing 284,807 transactions, fraudulent transactions account for 0.172% of the total transactions, underscoring the dataset's inherent imbalance.

Logistic Regression and SVM have yielded significant results with accuracy rates of 95%. Decision Trees and Random Forest have demonstrated high performance with accuracy rates of 98% and 97%, respectively. Notably, Artificial Neural Networks (ANN) emerged as the most effective method, achieving a perfect accuracy rate of 100%.

Each model has its own advantages and disadvantages, with factors such as training time, complexity, interpretability needing consideration. Balancing the false positive and false negative rates can also impact the practical usage and cost of the model.

In conclusion, this study has examined machine learning and deep learning models to detect credit card fraud, contributing a novel insight to the literature on the high accuracy and balanced outcomes of models such as Decision Trees and Random Forest in this context. Observations reveal different algorithms exhibit varying performances, underscoring the necessity of adapting the model selection to the algorithm's characteristics and the dataset. This research could contribute to cost savings for financial institutions and the provision of secure shopping experiences.

Keywords: *Fraud Deteciton, Credit Card Fraud, Machine Learning*

1. GİRİŞ

Dünya genelindeki dolandırıcılık faaliyetleri, toplumları ve ekonomileri ciddi şekilde tehdit eden bir sorundur. Bu tür suçlar sadece maddi kayıplara yol açmakla kalmamakla birlikte aynı zamanda insanların güven duygusunu da sarsmaktadır. Dolandırıcılık bireylerin ve kurumların en büyük endişelerinden biri haline gelmiştir. Dolandırıcılar her geçen gün daha farklı yöntemler geliştirerek bireylerin ve işletmelerin zarar görmesine neden olmaktadır. Finansal dolandırıcılık, bu tür suçların en yaygın ve etkili formlarından biridir. Dolandırıcılık faaliyetleri haksız kazanç elde etmek için gerçekleştirilen dolandırıcılık eylemleri olarak tanımlanabilir ve finansal sektör başta olmak üzere birçok sektörü olumsuz yönde etkilemektedir. Bu alandaki gelişmeleri anlamak ve etkili çözümler üretmek, finansal kurumlar ve toplumlar için hayati bir önem taşımaktadır (Hussaini, Bakar, & Yusuf, 2019).

Dolandırıcılık faaliyetleri farklı yöntemler ile tüm finansal sektöre yayılmakta olup Çizelge 1.1'de yer alan çalışma veriler ile sektörel durumu göstermektedir. Dolandırıcılık türlerini ve tespit kavramını anlamak finansal hizmetler önemli bir adımdır. Güncel ve etkili dolandırıcılık tespit ve önleme yöntemleri kullanarak, finansal kurumlar müşteri bilgilerinin gizliliğini ve güvenliğini koruyabilmekte, işlemleri ve hizmetleri güvence altına alabilmekte ve dolandırıcılığın önlenmesine ve sınırlandırılmasına yardımcı olabilmektedir. Bu nedenle finansal kurumlar için dolandırıcılık tespiti ve önleme stratejileri geliştirmek ve uygulamak büyük önem taşımaktadır (Johnson & Smith, 2019).

Çizelge 1.1: Sektörel Dolandırıcılık Üzerine Çalışma

Sektör	Adet	Fatura	Nakit Dolandırıcılığı	Elden Nakit Ödeme	Çek ile Ödeme	Yolsuzluk - Rüşvet	Masraf Geri Ödemeleri	Mali Tablo Dolandırıcılığı	Nakit Dışı İşlemler	Maaş Bordrosu	Masraf Kayıtları	Kart Kopyalama
Bankacılık ve Finansal Hizmetler	351	10%	11%	14%	14%	46%	8%	11%	11%	4%	2%	10%
Kamu Yönetimi ve İdaresi	198	21%	8%	7%	9%	57%	12%	8%	16%	16%	3%	8%
Üretim	194	26%	5%	9%	7%	59%	10%	12%	23%	10%	4%	8%
Sağlık	130	20%	6%	8%	8%	50%	11%	9%	18%	12%	2%	9%
Enerji	97	24%	9%	6%	8%	64%	16%	8%	13%	6%	3%	2%
Perakende	91	19%	10%	9%	9%	43%	7%	4%	24%	5%	7%	14%
Sigorta	88	15%	9%	8%	10%	40%	9%	5%	8%	10%	2%	11%
Teknoloji	84	21%	6%	10%	6%	54%	14%	8%	30%	5%	1%	1%
Ulaşım ve Depolama	82	20%	9%	15%	4%	59%	11%	7%	22%	9%	4%	11%
Yapı ve İnşaat	78	24%	8%	10%	14%	56%	17%	18%	24%	24%	3%	9%
Eğitim	69	26%	9%	12%	12%	49%	12%	12%	19%	14%	4%	12%
Bilgi ve Hizmet	60	15%	5%	5%	8%	58%	12%	12%	33%	7%	2%	7%
Yemek ve Konaklama	52	19%	10%	21%	17%	54%	13%	13%	29%	19%	10%	17%

Kaynak: (Association of Certified Fraud Examiners, 2020)

Yapay zekâ ve makine öğrenimi büyük veriyi hızlı ve doğru bir şekilde analiz ederek dolandırıcılık faaliyetlerini gerçek zamanlı olarak tespit edebilmekte ve otomatik aksiyonlar alabilmektedir. Ayrıca yapay zekâ ve makine öğrenimi her işlemden öğrenerek kendini geliştirebilir ve değişen dolandırıcılık yöntemlerine adapte olabilmektedir. Böylece dolandırıcılık tespit sistemleri hem operasyonel verimliliği hem de maliyet tasarrufunu artırabilmektedir (Phua, Lee, Smith, & Gayler, 2010).

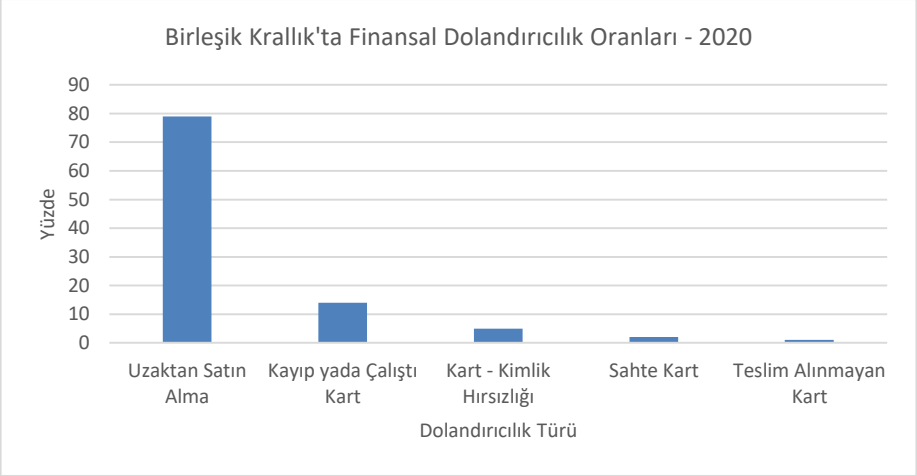
Bu tezde yapay zekâ ve makine öğrenimi ile dolandırıcılık tespiti konusunu detaylı bir şekilde incelenecektir. Yapay zekâ ve makine öğrenimi kavramlarını tanımlayacak ve temel prensipleri açıklanacaktır. Dolandırıcılık tespiti konusunun kapsamını, çeşitlerini ve zorlukları ele alınacaktır. Yapay zekâ ve makine öğrenimi ile dolandırıcılık tespiti arasındaki ilişkiyi kurulacak ve bu alanda kullanılan yöntemler, modeller ve uygulamalar gösterilecektir. Yapay zekâ ve makine öğrenimi ile dolandırıcılık tespiti konusunda Lojistik Regresyon, Destek Vektör Makineleri

(SVM), Karar Ağaçları, Rastgele Orman ve Yapay Sinir Ağları (ANN) modelleri kullanılacak olup avantajlar ve dezavantajlar tartışılacak ve gelecek perspektifler sunulacaktır.

2. GENEL BİLGİLER

2.1 Bankacılık ve Ödeme Sistemlerinde Dolandırıcılık Türleri

Finansal hizmetler ve ödeme sistemleri alanında, dolandırıcılık; sahte kimlik kullanma, kredi kartı dolandırıcılığı, ATM dolandırıcılığı, internet ve mobil bankacılık dolandırıcılığı, havale/EFT dolandırıcılığı, çek dolandırıcılığı ve kötü amaçlı yazılım veya sosyal mühendislik kullanımı gibi çeşitli türlerde gerçekleşebilmektedir (Yero, 2018). Dolandırıcılık türleri çok çeşitli olup Şekil 1.1'de Birleşik Krallık'taki Finansal Dolandırıcılık verisi bu çeşitliliği göstermektedir.



Şekil 2.1: Türe göre 2020'de Birleşik Krallık'ta Finansal Dolandırıcılık Oranları

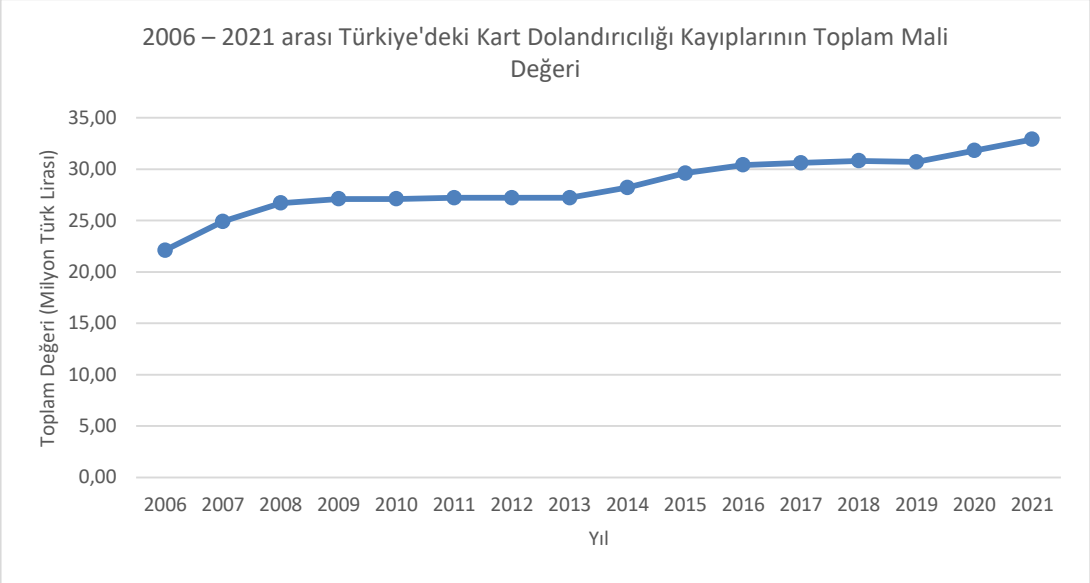
Kaynak: (UK - Finans, 2021)

2.1.1 Sahte kimlik kullanımı

Dolandırıcılar başkasının kimlik bilgilerini çalarak veya sahte belgelerle banka hesapları açabilir ve işlemler gerçekleştirebilmektedir. Kredi başvuruları, kredi kartı başvuruları ve diğer bankacılık işlemleri için kullanılabilmektedir. Sahte kimlik kullanma, kimlik avı ve kimlik hırsızlığı gibi çeşitli tekniklerle gerçekleştirebilmektedir. Ayrıca başkalarının banka hesaplarına erişmek ve bu hesapları kötü amaçlı kullanmak için sahte kimlik bilgileri kullanılabilmektedir (McNally, 2007).

2.1.2 Kredi kartı dolandırıcılığı

Kredi kartı dolandırıcılığı, dolandırıcıların çalınan veya kopyalanan kredi kartı bilgilerini kullanarak sahte işlemler gerçekleştirdiği bir dolandırıcılık türüdür. Bu sahte alışverişler yapmak, çevrimiçi hizmetlere üye olmak veya başkalarının hesaplarına para transfer etmek gibi işlemleri içermektedir. Kredi kartı dolandırıcılığı kart sahteciliği (kart kopyalama), kayıp veya çalıntı kart kullanımı gibi çeşitli şekillerde gerçekleşebilmektedir (Yero, 2018). Kart dolandırıcılığı finansal kurumlara ve kişilere maddi zarar vermektedir. Şekil 1.2'de 2006 – 2021 yılları arasında Türkiye'deki kart dolandırıcılığı kayıplarının toplam mali değeri belirtilmiştir.



Şekil 2.2: 2006 – 2021 arası Türkiye'deki Kart Dolandırıcılığı Kayıplarının Toplam Mali Değeri

Kaynak: (Euromonitor International, 2023)

2.1.3 ATM dolandırıcılığı

ATM dolandırıcılığı, dolandırıcıların ATM'lerde kart kopyalama (skimming) cihazları kullanarak kart bilgilerini çalarak sahte işlemler gerçekleştirdiği bir dolandırıcılık türüdür. Bu cihazlar kartın manyetik şeridini okuyarak kart bilgilerini kopyalamak için tasarlanmıştır. Ayrıca, dolandırıcılar ATM kameralarına benzer cihazlar kullanarak kullanıcıların kart şifrelerini çalabilmekte ve bu bilgileri kötü amaçlı kullanabilmektedirler (Adeel & Hussain, 2017).

2.1.4 İnternet bankacılığı dolandırıcılığı

İnternet bankacılığı dolandırıcılığı, kullanıcıların şifrelerini ele geçirerek veya oltalama saldırıları düzenleyerek internet bankacılığı üzerinden sahte işlemler gerçekleştiren bir dolandırıcılık türüdür. Dolandırıcılar bankaların web sitelerine benzeyen sahte siteler oluşturarak kullanıcıları yanıltır ve kimlik bilgilerini çalmaktadır. Bu bilgilerle dolandırıcılar kullanıcıların hesaplarına erişebilmekte ve para transferleri veya diğer kötü amaçlı işlemler gerçekleştirebilmektedirler (Cheng & Liu, 2019).

2.1.5 Havale/EFT dolandırıcılığı

Dolandırıcılar, başkalarının banka hesaplarına izinsiz erişim sağlayarak sahte havale veya EFT işlemleri gerçekleştirdikleri bir dolandırıcılık türüdür. Bu tür dolandırıcılık e-posta yoluyla yapılan kimlik avı, oltalama saldırıları ve şifre çalma yöntemleri ile gerçekleştirilebilir. Dolandırıcılar, başkalarının hesaplarına para transfer etmek, fatura ödemelerini durdurmak veya değiştirmek gibi kötü amaçlı işlemler gerçekleştirebilirler (Dunn, 2018).

2.1.6 Mobil bankacılık dolandırıcılığı

Mobil bankacılık dolandırıcılığı, dolandırıcıların sahte uygulamalar, kimlik avı, oltalama SMS'leri ve zararlı yazılımlar kullanarak kullanıcıların mobil bankacılık hesaplarına erişmeye çalıştığı bir dolandırıcılık türüdür. Mobil bankacılık dolandırıcılığında, dolandırıcılar genellikle kullanıcılardan şifre veya kimlik bilgilerini çalmak amacıyla sahte mesajlar veya uygulamalar yoluyla hedef alınmaktadır. Bu bilgilerle, dolandırıcılar kullanıcıların hesaplarını ele geçirebilmekte ve kötü amaçlı işlemler gerçekleştirebilmektedir (Cheng & Liu, 2019).

2.1.7 Çek dolandırıcılığı

Çek dolandırıcılığı; sahte, çalıntı veya boş çeklerle işlem yaparak gerçekleştirilen bir dolandırıcılık türüdür. Dolandırıcılar, başkalarının çeklerini çalarak veya sahte çekler oluşturarak para çekmeye veya ödeme yapmaya çalışmaktadır. Ayrıca dolandırıcılar kara para aklama amacıyla da sahte çekler kullanabilmektedir (Powell, 2008).

2.2 Dolandırıcılık Tespit Kavramı

Dolandırıcılık tespiti finansal hizmetler ve ödeme sistemleri alanında gerçekleşen şüpheli veya anormal işlemleri, davranışları ve aktiviteleri belirlemeye yönelik süreçler ve tekniklerin uygulanmasıdır. Dolandırıcılık tespiti kavramları hem teknolojik çözümleri hem de insan faktörünü içeren stratejileri kapsamaktadır.

2.2.1 Veri analizi ve yapay zekâ

Veri analizi ve yapay zekâ, dolandırıcılık tespitinde önemli bir rol oynamaktadır. Finansal kurumlar ve ödeme sistemleri, işlemleri ve müşteri davranışlarını analiz etmek ve şüpheli veya anormal işlemleri tespit etmek için büyük veri analizi, makine öğrenimi ve yapay zekâ teknolojilerinden yararlanmaktadır. Bu yöntemler gerçek zamanlı veya yakın gerçek zamanlı olarak gerçekleşen işlemleri, davranışları incelemeye ve potansiyel dolandırıcılığı belirlemeye olanak tanımaktadır (Smith & Johnson, 2020).

2.2.2 Davranış analizi

Davranış analizi; müşterilerin normal işlem ve etkinliklerini analiz ederek şüpheli veya anormal davranışları tespit etmeye çalışan bir dolandırıcılık tespit yöntemidir. Bu yöntem işlem miktarları, saatleri, türleri ve yerleri gibi faktörleri dikkate alarak dolandırıcılık riskini değerlendirmeye yardımcı olmaktadır (Kaur & Rani, 2021).

2.2.3 Güvenlik protokolleri ve şifreleme

Güvenlik protokolleri ve şifreleme dolandırıcılık tespitinde önemli bir rol oynamaktadır. Finansal kurumlar ve ödeme sistemleri, işlem verilerini ve müşteri bilgilerini korumak için güçlü şifreleme algoritmaları ve güvenlik protokolleri kullanmaktadır. Bu önlemler, dolandırıcıların bilgilere erişmesini ve kötü amaçlı kullanmasını zorlaştırmaktadır (Nadeem & Zawoad, 2021).

2.2.4 Çok faktörlü kimlik doğrulama

Çok faktörlü kimlik doğrulama, müşterilerin kimliğini doğrulamak için birden fazla doğrulama yöntemi kullanarak dolandırıcılığı önlemede ve tespit etmede etkili bir yöntemdir. Bu yöntemler, şifreler, biyometrik veriler (parmak izi, yüz tanıma vb.) ve tek kullanımlık şifreler (OTP) gibi bir dizi faktörü içermektedir. Çok faktörlü

kimlik doğrulama, dolandırıcıların bir müşterinin hesabına erişmesini ve kötü amaçlı işlemler gerçekleştirmesini zorlaştırmaktadır (Jin, Gao, & Liu, 2021).

2.2.5 İzleme ve uyarı sistemleri

İzleme ve uyarı sistemleri finansal kurumların ve ödeme sistemlerinin şüpheli işlemleri ve aktiviteleri tespit etmelerine ve müşterilere zamanında bildirimler göndermesine yardımcı olmaktadır. Bu sistemler, işlem verileri, hesap hareketleri ve müşteri davranışları üzerinde sürekli gözlem yaparak, dolandırıcılığa işaret edebilecek anormallikleri tespit etmekte ve uyarılar üretmektedir. Müşterilere gönderilen uyarılar, dolandırıcılığın önlenmesine ve sınırlandırılmasına yardımcı olmaktadır (Shen & Hsiao, 2021).

2.2.6 Eğitim ve farkındalık programları

Eğitim ve farkındalık programları müşterilerin ve finansal hizmet sağlayıcıların personelinin dolandırıcılık türlerini, risklerini ve dolandırıcılığı önleme ve tespit etme yöntemlerini anlamalarına yardımcı olmaktadır. Bu programlar müşterilere ve çalışanlara dolandırıcılık türleri ve riskleri hakkında bilgi sağlamaktadır. En iyi güvenlik uygulamalarını ve politikalarını öğreterek aynı zamanda dolandırıcılığın önlenmesi ve tespiti için kullanılacak araçlar ve teknikler hakkında eğitim vermektedir (Rejeb & Abdul-Rahman, 2021).

Dolandırıcılık tespit kavramlarını kullanarak finansal kurumlar ve ödeme sistemleri dolandırıcılık riskini azaltabilir, müşteri bilgilerini ve işlemlerini koruyabilir ve finansal hizmetlerin güvenliğini ve bütünlüğünü sağlayabilmektedir. Dolandırıcılık yöntemleri sürekli olarak geliştiği için, dolandırıcılık tespit stratejilerinin, teknolojilerinin güncel ve etkili olması önem arz etmektedir. Bu nedenle finansal kurumlar ve ödeme sistemleri için dolandırıcılık tespiti ve önleme stratejileri geliştirmek ve uygulamak büyük önem taşımaktadır (Rejeb & Abdul-Rahman, 2021).

2.2.7 İşlem izleme ve risk değerlendirme

İşlem izleme ve risk değerlendirmesi finansal kurumların ve ödeme sistemlerinin müşteri işlemleri ve davranışları üzerinde sürekli gözlem yaparak dolandırıcılık tespitinde kullanılan etkili bir yöntemdir. Bu süreç işlem türleri, miktarlar, saatler ve yerler gibi faktörleri dikkate alarak risk seviyelerini

değerlendirmeye yardımcı olmaktadır. Yüksek riskli işlemler ve davranışlar dolandırıcılık tespit sistemleri tarafından daha yakından incelenir ve gerektiğinde müşterilere ve ilgili kurumlara uyarılar gönderilebilmektedir (Singh & Patel, 2020).

2.2.8 Sosyal mühendislik ve kötü amaçlı yazılım tespiti

Sosyal mühendislik ve kötü amaçlı yazılım tespiti finansal kurumların ve ödeme sistemlerinin, dolandırıcıların insanları kandırmak ve sistemlere sızmak için kullandığı yöntemlerin belirlenmesine yardımcı olmaktadır. Sosyal mühendislik saldırıları, dolandırıcıların hedeflerine güvendikleri bilgileri vermeye ikna etmeye çalıştığı telefon görüşmeleri, e-postalar ve mesajlaşma uygulamaları gibi çeşitli iletişim kanalları üzerinden gerçekleşebilmektedir. Kötü amaçlı yazılımlar ise kullanıcıların cihazlarına ve sistemlerine sızarak bilgi çalmaya, işlemleri yönlendirmeye veya kötü amaçlı işlemler gerçekleştirmeye çalışmaktadır. Dolandırıcılık tespit sistemleri sosyal mühendislik saldırılarını ve kötü amaçlı yazılımları belirlemeye ve önlemeye yönelik stratejiler ve teknolojiler içermektedir (Tharaka Kaushalya, 2018).

2.2.9 İş birliği ve paylaşılan istihbarat

İş birliği ve paylaşılan istihbarat finansal kurumların, ödeme sistemlerinin, düzenleyicilerin ve kolluk kuvvetlerinin, dolandırıcılık türlerini, yöntemlerini ve tespit stratejilerini paylaşarak dolandırıcılık ile daha etkili bir şekilde mücadele etmelerine yardımcı olmaktadır. Bu iş birliği dolandırıcılıkla mücadelede bilgi ve deneyimlerin paylaşılmasını teşvik etmektedir. Sektör genelinde güvenlik ve dolandırıcılık tespiti konularında daha iyi anlayış ve uyum sağlamaktadır (Crossler, 2013).

Tüm bu dolandırıcılık tespit kavramlarını bir arada kullanarak, finansal kurumlar ve ödeme sistemleri dolandırıcılık riskini daha etkili bir şekilde yönetebilmekte, müşteri bilgilerini koruyabilmekte, finansal hizmetlerin güvenliğini ve bütünlüğünü sağlayabilmektedir. Dolandırıcılık yöntemleri sürekli olarak geliştiği için dolandırıcılık tespit stratejilerinin ve teknolojilerinin güncel ve etkili olması önem arz etmektedir. Bu nedenle finansal kurumlar ve ödeme sistemleri için dolandırıcılık tespiti ve önleme stratejileri geliştirmek ve bu konuda iş birliği uygulamak büyük önem taşımaktadır (Crossler, 2013).

2.2.10 Proaktif yaklaşımlar ve sürekli iyileştirme

Dolandırıcılık tespiti sadece mevcut tehditlere karşı mücadele etmekle kalmaz, aynı zamanda gelecekteki tehditleri öngörmeye ve bu tehditlere karşı önlem almak için proaktif yaklaşımlar benimsemeye de odaklanmaktadır. Finansal kurumlar ve ödeme sistemleri, yeni dolandırıcılık yöntemlerini ve teknolojilerini analiz etmeli, güvenlik ve tespit sistemlerini sürekli olarak güncellemeli, iyileştirmeli, yeni tehditler ve zayıf noktalar hakkında bilgi toplamalıdır (Nurse, 2011).

2.2.11 Yasal düzenlemeler ve uyumluluk

Dolandırıcılık tespiti yasal düzenlemeler ve uyumluluk gereklilikleriyle de yakından ilgilenmektedir. Finansal kurumlar ve ödeme sistemleri dolandırıcılık tespiti ve önleme yöntemleri geliştirirken ve uygularken yerel ve uluslararası yasal düzenlemelere ve standartlara uyum sağlamalıdır. Uyumluluk müşteri bilgilerinin korunması, işlem izleme ve raporlama, risk değerlendirmesi ve yönetimi ve diğer güvenlik önlemleri gibi alanları içermektedir (Santos & Maynard, 2018).

Dolandırıcılık tespit kavramları teknolojik çözümler ve insan faktörü arasında uyumlu bir denge sağlayarak finansal hizmetler ve ödeme sistemlerinin dolandırıcılık riskini yönetmeye ve azaltmaya yardımcı olmaktadır. Bu süreç sürekli gelişim, adaptasyon ve inovasyon gerektirmektedir, çünkü dolandırıcılar ve dolandırıcılık yöntemleri sürekli olarak gelişmekte ve değişmektedir. Bu nedenle finansal kurumlar ve ödeme sistemleri için yasal düzenlemeleri uygulamak ve dolandırıcılık tespiti önleme stratejilerine uyumluluk göstermek, güvenli ve güvenilir finansal hizmetler sunmak için kritik bir öneme sahiptir (Santos & Maynard, 2018).

2.2.12 Küresel iş birliği ve koordinasyon

Dolandırıcılık, küresel bir sorun olduğu için, ülkeler ve uluslararası kuruluşlar arasındaki iş birliği ve koordinasyon, dolandırıcılık tespitinde kritik bir faktördür. Ülkeler ve uluslararası kuruluşlar dolandırıcılıkla mücadelede bilgi ve istihbarat paylaşımı, uyumlu düzenlemeler ve standartlar geliştirme ve ortak operasyonlar yoluyla iş birliği yapmaktadır. Bu küresel finansal sistemlerin güvenliğini ve bütünlüğünü korumaya ve dolandırıcılık riskini azaltmaya yardımcı olmaktadır (Crossler, 2013).

2.2.13 Etkin iç kontrol sistemleri

Etkin iç kontrol sistemleri finansal kurumların ve ödeme sistemlerinin dolandırıcılığı tespit etmeye ve önlemeye yönelik süreçlerini ve politikalarını düzenli olarak gözden geçirmelerini ve değerlendirmelerini sağlamaktadır. İç kontrol sistemleri operasyonel süreçler, bilgi teknoloji altyapısı, çalışan eğitimi ve politika uygulamaları gibi alanlarda sürekli iyileştirmeler yapılmasını teşvik etmektedir (Arnone, 2008).

2.2.14 Olay yönetimi ve müdahale

Dolandırıcılık tespitinde olay yönetimi ve müdahale, şüpheli veya doğrulanmış dolandırıcılık vakalarının etkin bir şekilde ele alınması ve çözülmesi için önemlidir. Finansal kurumlar ve ödeme sistemleri dolandırıcılık vakalarının hızlı ve etkili bir şekilde tespit edilmesini, analiz edilmesini ve çözülmesini sağlamak için önceden belirlenmiş prosedürler ve protokoller geliştirmekte ve uygulamaktadır. Bu dolandırıcılığın önlenmesine ve sınırlandırılmasına yardımcı olmakta, müşteri güvenini ve finansal hizmetlerin güvenliğini korumaktadır (Shen & Hsiao, 2021).

2.2.15 Sürekli teknolojik gelişim ve yenilik

Dolandırıcılık tespiti ve önleme stratejileri, sürekli teknolojik gelişim ve yeniliklere dayanmaktadır. Finansal kurumlar ve ödeme sistemleri yapay zekâ, makine öğrenimi, büyük veri analitiği ve blok zinciri gibi yeni teknolojileri dolandırıcılık tespiti ve önleme yöntemlerine entegre etmektedir. Bu teknolojiler dolandırıcılığın tespit edilmesini ve engellenmesini daha hızlı ve etkili hale getirmekte, müşteri verilerinin korunmasını sağlamakta ve finansal hizmetlerin güvenliğini ve bütünlüğünü desteklemektedir (Nurse, 2011).

2.2.16 Müşteri deneyimi ve güvenliği dengelemek

Dolandırıcık tespiti ve önleme stratejilerinin etkili olabilmesi için finansal kurumlar ve ödeme sistemleri müşteri deneyimini ve güvenliği dikkate alarak dengelemektedir. Kullanıcı dostu ve güvenli çözümler sunarak, müşterilerin finansal hizmetlere erişimini kolaylaştırmak ve aynı zamanda dolandırıcılık riskini yönetmek ve azaltmak mümkündür. Bu denge, müşteri güvenini ve sadakatini artırmakta ve finansal hizmetlerin kullanımını teşvik etmektedir (Singh & Patel, 2020).

Sonuç olarak, dolandırıcılık tespit kavramları ve stratejileri, finansal kurumlar ve ödeme sistemleri için büyük öneme sahiptir. Dolandırıcılık yöntemleri sürekli olarak geliştiği için kurumlar güncel ve etkili dolandırıcılık tespit stratejileri ve teknolojileri benimseyerek müşteri bilgilerini ve işlemlerini koruyabilir aynı zamanda finansal hizmetlerin güvenliğini ve bütünlüğünü sağlayabilirler.

3. YAPAY ZEKÂ İLE DOLANDIRICILIK TESPİTİ: GELİŞİM, KULLANILAN YÖNTEMLER VE LİTERATÜR TARAMASI

3.1 Yapay Zekâ ile Dolandırıcılık Tespitinin Yıllara Göre Gelişimi

Finansal dolandırıcılık tespitinde teknolojilerin yıllara göre gelişimi incelendiğinde 2010 yılında, Özellik Seçimi ve Kümeleme teknikleri ön planda kullanılan yöntemlermiş. Bu dönemde algoritmalarda önemli özellikleri belirleyerek ve benzer veri gruplarını kümeleyerek dolandırıcılık işlemlerini tespit etmek amaçlanıyormuş (Bolton, 2002).

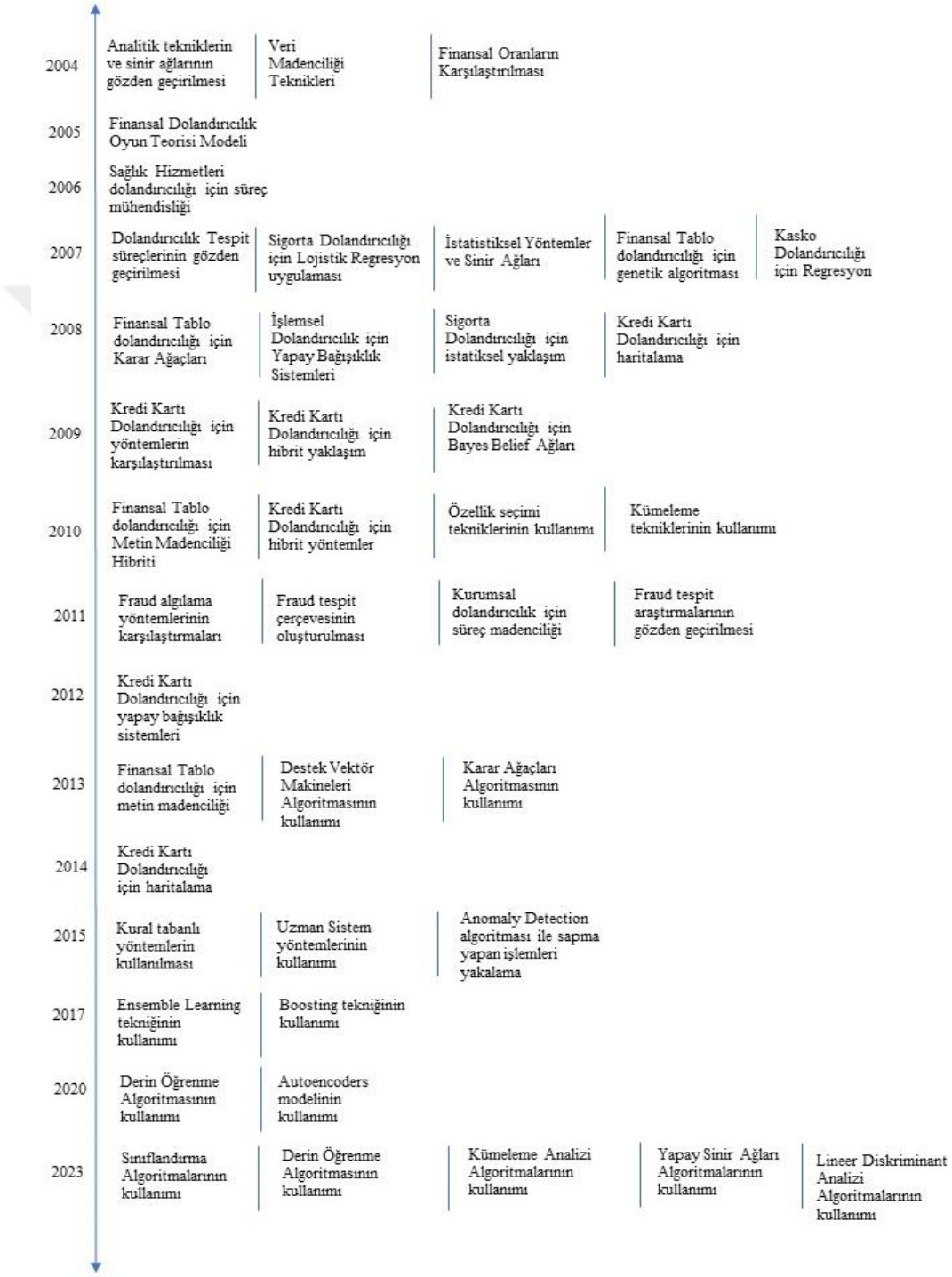
2015 yılına gelindiğinde Kural Tabanlı (Rule-based) ve Uzman Sistemler (Expert Systems) yöntemleri yaygınlaşmaya başlanmış. Bu sistemlerde, önceden belirlenen kurallar ve uzman bilgisi kullanılarak potansiyel dolandırıcılık faaliyetlerini belirlemek hedefleniyormuş. Ayrıca bu dönemde, Anomaly Detection yöntemleri de kullanılmaya başlanarak, normalden sapma gösteren işlemler tespit edilip, dolandırıcılık ihtimali yüksek olanlar işaretleniyormuş.

2015-2020 yılları arasında Toplu Öğrenme (Ensemble Learning) ve Artırma (Boosting) teknikleri daha yaygın hale gelmiş. Bu teknikler birden fazla zayıf işlemi bir araya getirerek güçlü bir sınıflandırıcı oluşturmayı amaçlar ve genellikle daha yüksek tespit doğruluğu elde etmektedir (Ahmad & Raza, 2021).

2020 yılında ise, Derin Öğrenme (Deep Learning) ve Otomatik Kodlayıcılar (Autoencoders) gibi daha gelişmiş yöntemler finansal dolandırıcılık tespitinde kullanılmaya başlanmış. Bu dönemde büyük veri setlerini işleyebilme yeteneği ve özellik öğrenme kabiliyeti sayesinde daha karmaşık ve gelişmiş dolandırıcılık faaliyetleri tespit edilebilir hale gelmiş.

Sonuç olarak, dolandırıcılık tespiti alanında teknolojiler ve yöntemler zamanla önemli ölçüde gelişmektedir. Yeni yaklaşımlar ve algoritmalar, dolandırıcılık faaliyetlerini daha etkin ve hızlı bir şekilde tespit etmeye ve önlemeye yardımcı olmaktadır (Ahmad & Raza, 2021).

Şekil 3.1'de yer alan bilgiler belirtilen kaynaklar dışında elektronik arama motoru olan Google Scholar, Scinapse ve Semantic Scholar internet sayfalarında yapay zekânın gelişimi, kredi kartı dolandırıcılık tespiti, dolandırıcılık tespitinde kullanılan yapay zekâ algoritmaları gibi kelimelerin kombinasyonları ile arama stratejisi oluşturularak çıkan sonucun yıllara göre gruplandırılması ile ulaşılmıştır.



Şekil 3.1: Yıllara Göre Kronolojik Finansal Dolandırıcılık Gelişmeleri

Kaynak: (Bolton, 2002; Ahmad & Raza, 2021)

3.2 Yapay Zekâ ile Dolandırıcılık Tespitinde Kullanılan Yöntemler

Yapay zekâ kredi kartı dolandırıcılığı gibi finansal dolandırıcılık tespitinde kullanılan yöntemlerin önemli bir bileşenidir. Bu alandaki araştırmalar dolandırıcılığı etkili bir şekilde tespit etmek ve önlemek için gelişmiş algoritmalar ve teknikler kullanmaktadır (Bolton, 2002). Bu bölümde yapay zekâ ile dolandırıcılık tespitinde kullanılan yöntemler hakkında bilgiler sunulmaktadır.

3.2.1 Sınıflandırma algoritmaları

Dolandırıcılık tespitinde kullanılan temel Yapay Zekâ yöntemlerinden biri sınıflandırma algoritmalarıdır. Bu algoritmalar veri setindeki işlemleri dolandırıcılık ve dolandırıcılık olmayan olarak ikiye ayırarak tespit sağlamaktadır. Kullanılan yaygın sınıflandırma algoritmaları aşağıda yer almaktadır (Ahmad & Saini, 2019):

- Lojistik Regresyon
- Destek Vektör Makineleri (SVM)
- Karar Ağaçları
- Rastgele Ormanlar
- Yapay Sinir Ağları (ANN)
- Naif Bayes
- k-En Yakın Komşu (k-NN)

3.2.2 Kümeleme ve anomali tespiti

Dolandırıcılık tespitinde kullanılan diğer bir yöntem, kümeleme ve anomali tespiti algoritmalarıdır. Bu yöntemler, veri setindeki benzer işlemleri gruplayarak normal ve anormal işlemleri ayırt etmeyi amaçlamaktadır. Anomali tespiti algoritmaları, özellikle yüksek boyutlu veri setlerinde ve dengesiz veri kümesi durumlarında etkili olabilmektedir. Yaygın kullanılan kümeleme ve anormallik (anomali) tespiti algoritmaları şunlardır (Wu, Zhu, & Ding, 2009):

- k-Ortalama Kümeleme
- Yoğunluk Tabanlı Uzaklıkla Nokta Gruplaması
- Hiyerarşik Kümeleme
- İzolasyon Ormanı
- Yerel Aykırı Gözlem Faktörü

- Tek Sınıf Destek Vektör Makineleri

3.2.3 Derin öğrenme

Derin öğrenme, yapay sinir ağlarına dayalı gelişmiş bir yapay zekâ teknolojisidir ve dolandırıcılık tespitinde kullanılan yöntemlerin geliştirilmesinde önemli bir rol oynamaktadır. Derin öğrenme modelleri, büyük veri setlerinde karmaşık özellik ilişkilerini ve yapıları öğrenebilir, dolandırıcılığı daha etkili bir şekilde tespit edebilmektedir (Ribeiro, 2019). Derin öğrenme algoritmaları aşağıda yer almaktadır (LeCun, Bengio, & Hinton, 2015):

- Çok Katmanlı Algılayıcılar
- Evrişimli Sinir Ağları
- Tekrarlayan Sinir Ağları
- Uzun Kısa Dönemli Bellek
- Dikkat Tabanlı Uzun Kısa Dönemli Bellek
- Otomatik Kodlayıcılar
- Varyasyonel Otomatik Kodlayıcılar
- Üretken Çekişmeli Ağlar

3.2.4 Toplu öğrenme ve hiperparametre optimizasyonu

Dolandırıcılık tespitinde, birden fazla modelin veya algoritmanın kombinasyonunu kullanarak daha güçlü ve kararlı sonuçlar elde etmeyi amaçlayan Toplu Öğrenme (Ensemble Learning) yöntemleri kullanılabilmektedirr. Toplu Öğrenme (Ensemble Learning) algoritmaları şunlardır (Chen & Guestrin, 2016):

- Uyarlamalı Arttırma
- Eğim Arttırma Makineleri
- Ektrem Eğim Arttırma
- Hafif Eğim Arttırma
- Kategorik Gradyan Arttırma

Ayrıca hiperparametre optimizasyonu, dolandırıcılık tespiti algoritmalarının performansını iyileştirmek için önemli bir rol oynamaktadır. Hiperparametre optimizasyonu için kullanılan yöntemleri aşağıda yer almaktadır (Bergstra & Bengio, 2012):

- Izgara Arama
- Rastgele Arama
- Bayes Optimizasyon
- Genetik Algoritmalar
- Parçacık Sürü Optimizasyonu

3.2.5 Özellik seçimi ve özellik mühendisliği

Yapay zekâ ile dolandırıcılık tespitinde kullanılan yöntemlerin başarısı, veri setindeki özelliklerin kalitesi ve seçimi ile de ilişkilidir. Özellik seçimi ve özellik mühendisliği, model performansını artırarak daha doğru ve etkili tespit sağlamaktadır. Özellik seçimi ve mühendisliği yöntemleri aşağıda yer almaktadır (Guyon, 2003):

- Yinelemeli Öznitelik Elemesi
- Tek Değişkenli Öznitelik Seçimi
- En Küçük Mutlak Büzülme ve Seçim Operatörü Düzenlemesi
- Temel Bileşen Analizi
- Öznitelik Önemi

3.2.6 Aktarım öğrenmesi ve gömme teknikleri

Dolandırıcılık tespitinde farklı alanlardan öğrenilen bilgi ve modellerin kullanılması da etkili olabilmektedir. Aktarım öğrenme ve gömme teknikleri önceden eğitilmiş modellerin ve özellik vektörlerinin dolandırıcılık tespiti için kullanılmasına olanak tanımaktadır. Bu teknikler özellikle veri eksikliği yaşanan durumlarda ve daha iyi özellik temsilleri elde etmek için önem arz etmektedir. Aktarım öğrenme ve gömme teknikleri şunlardır (Mikolov, Chen, Corrado, & Dean, 2013; Devlin, Chang, Lee, & Toutanova, 2019):

- Kelime Gömme ve Belge Gömme
- Global Vektörler
- Doğal Dil İşleme Modeli
- Görüntü Ağı

3.2.7 Zaman serisi analizi ve tahmine dayalı yöntemler

Dolandırıcılık tespiti zaman serisi verileri ile de ilişkili olabilmektedir. Özellikle işlem verilerinin zamanla değişkenlik gösterdiği durumlar önemli bir örnektir. Zaman serisi analizi ve tahmine dayalı yöntemler geçmiş verilere dayalı olarak gelecekteki dolandırıcılık olaylarını tahmin etmeye çalışmaktadır. Bu yöntemler şunlardır (Taylor & Letham, 2018):

- Otomatik Regresyon Entegre Hareketli Ortalama
- Mevsimsel Otomatik Regresyon Entegre Hareketli Ortalama
- Uzun Kısa Dönem Bellek ve Tekrarlayan Sinir Ağları

3.2.8 Veri dengesizliği ve örneklem yöntemleri

Dolandırıcılık tespitinde veri dengesizliği önemli bir sorundur ve model performansını olumsuz etkileyebilmektedir. Veri dengesizliği ile başa çıkmak için örneklem yöntemleri kullanılabilmektedir. Bu yöntemler dengesiz sınıflar arasındaki dağılımı dengelemeye ve modelin dolandırıcılık tespiti performansını artırmaya yardımcı olmaktadır. Örneklem yöntemleri şunlardır (Haixiang, Yijing, Shang, Mingyun, & Yuanyue, 2017):

- Rastgele Azaltma Örnekleme
- Rastgele Artırma Örnekleme
- Sentetik Azınlık Artırma Tekniği
- Uyarlamalı Sentetik
- Tomek Bağlantıları ve Komşuluk Temizleme Kuralı

3.2.9 İşbirlikçi filtreleme ve oylama tabanlı yöntemler

Dolandırıcılık tespitinde işbirlikçi filtreleme ve oylama tabanlı yöntemler, kullanıcılar ve işlemler arasındaki ilişkileri analiz ederek dolandırıcılık faaliyetlerini tespit etmeye yardımcı olmaktadır. Bu yöntemler kullanıcı davranışlarının ve işlem özelliklerinin benzerliklerine dayalı olarak dolandırıcılık örüntülerini belirlemeye çalışmaktadır. İşbirlikçi filtreleme ve oylama tabanlı yöntemler şunlardır (Su, Khoshgoftaar, & Zhu, 2009):

- Kullanıcı Tabanlı İşbirlikçi Filtreleme
- İşlem Tabanlı İşbirlikçi Filtreleme

- Oylama Tabanlı Sınıflandırma (Oylama Sınıflandırıcısı)

3.2.10 Graf tabanlı ve ağ analizi yöntemleri

Graf tabanlı ve ağ analizi yöntemleri dolandırıcılık tespitinde kullanıcılar, işlemler ve diğer bağlantılı öğeler arasındaki ilişkileri analiz etmektedir. Bu yöntemler karmaşık ağ yapılarını ve örüntülerini inceleyerek dolandırıcılık faaliyetlerinin tespitine ve önlenmesine katkıda bulunmaktadır. Graf tabanlı ve ağ analizi yöntemleri şunlardır (Gülpınar, Özlük, & Öztaş, 2015):

- Graf Teorisi ve Ağ Analizi
- Bağlantı Analizi Temelli Algoritmalar
- Bağlantı Analizi

Yapay zekâ ve makine öğrenimi teknolojilerinin hızlı gelişimiyle birlikte dolandırıcılık tespitinde kullanılan yöntemler ve teknikler de sürekli evrim geçirmektedir. Bu literatür taraması dolandırıcılık tespiti alandaki temel yöntemler ve teknikler hakkında bilgi sunmakta, dolandırıcılık tespiti ve önleme alanındaki çalışmaların daha etkili ve güçlü hale getirilmesine katkıda bulunmayı amaçlamaktadır. Bu teknolojilerin geliştirilmesi ve adapte edilmesi, finansal dolandırıcılıkla mücadelede daha başarılı ve etkili çözümlerin ortaya çıkmasını sağlayacaktır.

3.3 Yapay Zekâ ile Dolandırıcılık Tespiti: Literatür Taraması ve Kullanılan Yöntemler

Aşağıda Yapay Zekâ ile dolandırıcılık tespiti için kullanılan yöntemlerin literatür taraması sonucu bulunmaktadır. Çizelge 3.1’de ve Çizelge 3.2’de yer alan bilgiler elektronik arama motoru olan Google Scholar, Scinapse ve Semantic Scholar internet sayfalarında yapay zekâ ile kredi kartı dolandırıcılığı, kredi kartı dolandırıcılık tespiti, dolandırıcılık tespitinde kullanılan yapay zekâ algoritmaları gibi kelimelerin kombinasyonları ile arama stratejisi oluşturularak sonuca ulaşılmıştır.

Çizelge 3.1: Akademik Çalışmalarda Dolandırıcılık Tespiti için Kullanılan Yöntemler

Yöntem	Makale	Tez	Akademik Çalışma	Toplam
Sınıflandırma Algoritmaları	42	26	18	86
Derin Öğrenme	35	16	10	61
Karar ağaçları	28	18	13	59
Destek Vektör Makineleri	23	12	6	51
Kümeleme Analizi Algoritmaları	8	5	3	16
Lineer Ayrım Analizi	6	3	2	11
Çok Çıkışlı Yapay Sinir Ağları	4	2	1	7

Çizelge 3.1'de sunulan veriler bir literatür taramasının sonucudur. Sınıflandırma algoritmaları, Derin Öğrenme, Yapay Sinir Ağları, Kümeleme Analizi, Linear Discriminant analizi, Karar Ağaçları, Destek Vektör Makineleri ve Çok Çıkışlı Yapay Sinir Ağları yöntemleri akademik çalışmalarda dolandırıcılık tespiti için yaygın olarak kullanılan yaklaşımlardır.

Çizelge 3.2: Dolandırıcılık Tespitinde Kullanılan Algoritmaların Sıklığı

Algoritma	Kullanım Sıklığı	Çalışma Sayısı
Lojistik Regresyon	Yüksek	80
Yapay Sinir Ağları	Yüksek	66
Karar Ağaçları	Yüksek	59
Destek Vektör Makineleri	Yüksek	51
Rastgele Orman	Yüksek	50
Naif Bayes	Orta	28
k-En Yakın Komşu	Orta	24
k-Ortalama Kümeleme	Orta	16
Tek Sınıf Destek Vektör Makineleri	Orta	16
İzolasyon Ormanı	Orta	10
Hiyerarşik Kümeleme	Orta	7
Evrişimli Sinir Ağları	Orta	7
Eğim Artırma Makineleri	Orta	7
Izgara Arama	Orta	7
Yerel Aykırı Gözlem Faktörü	Orta	6
Genetik Algoritmalar	Orta	6
Çok Katmanlı Algılayıcılar	Düşük	5
Tekrarlayan Sinir Ağları	Düşük	5
Uzun Kısa Dönem Bellek	Düşük	5
Dikkat Tabanlı Uzun Kısa Dönem Bellek	Düşük	5

Çizelge 3.2: (Devamı) Dolandırıcılık Tespitinde Kullanılan Algoritmaların Sıklığı

Algoritma	Kullanım Sıklığı	Çalışma Sayısı
Otokodlayıcılar	Düşük	5
Aşırı Gradyan Artırma	Düşük	5
Bayes Eniyileme	Düşük	5
Tek Değişkenli Özellik Seçimi	Düşük	5
Temel Bileşen Analizi	Düşük	4
Yoğunluk Tabanlı Spatial Kümelenme Uygulaması	Düşük	3
Değişkenlik Tabanlı Otokodlayıcılar	Düşük	3
Üretici Karşıt Ağlar	Düşük	3
Uyarlamalı Artırma	Düşük	3
Rastgele Arama	Düşük	3
Yinelemeli Özellik Elemesi	Düşük	3
Hafif Eğim Artırma Makineleri	Düşük	2
Kategorik Eğim Artırma Makineleri	Düşük	2
Parçacık Sürü Optimizasyonu	Düşük	2
Kelime Gömme ve Belge Gömme	Düşük	2
Genel Vektörler	Düşük	2
Uzun Kısa Süreli Bellek ve Tekrarlayan Sinir Ağları	Düşük	2
En Küçük Mutlak Büzülme ve Seçim Operatörü	Düşük	1
Özellik Önemi	Düşük	1
Doğal Dil İşleme Modeli	Düşük	1
Görüntü Ağı	Düşük	1
Zaman Serisi Analizi Modeli	Düşük	1

3.4 Uygulanacak Algoritmalar ve Özellikleri

Literatür taraması sonucunda sık kullanılan algoritmalar aşağıda listelenmektedir. Kredi Kartı Dolandırıcılık Tespit uygulamasında seçilen bu algoritmalar kullanılacaktır.

- Lojistik Regresyon
- Destek Vektör Makineleri
- Karar Ağaçları
- Random Forest
- Yapay Sinir Ağları

3.5 Lojistik Regresyon (Logistic Regression) Algoritması

Lojistik regresyon, istatistik ve makine öğrenimi alanlarında sıklıkla kullanılan bir sınıflandırma yöntemidir. Genellikle bağımlı değişkenin ikili olduğu durumlarda kullanılmaktadır. Bu nedenle, lojistik regresyon, kredi kartı dolandırıcılığı tespiti gibi ikili sınıflandırma problemlerinde uygulanabilir bir yöntemdir (Udjianto, 2006). Lojistik regresyon algoritması aşağıdaki gibi genel olarak anlatılabilir:

Lojistik regresyon, bağımlı değişkenin olasılığını tahmin etmek için bağımsız değişkenlerin doğrusal bir kombinasyonunu kullanılmaktadır. Modelin matematiksel ifadesi şu şekildedir: (3.1) (D. W. Hosmer Jr, 2013)

$$P(Y = 1|X) = 1 / (1 + exp(-(\beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn)) \quad \textbf{(3.1)}$$

$P(Y = 1|X)$: Y'nin 1 olma olasılığı, X verilen bağımsız değişkenler vektörüdür.

$\beta 0, \beta 1, \ldots, \beta n$: Modelin katsayılarıdır.

$X1, X2, \ldots, Xn$: Bağımsız değişkenlerdir.

$exp(x)$: $e\hat{}x$, e'nin x kuvvetidir.

Lojistik regresyon modelini eğitmek için maksimum olabilirlik tahmini kulanılmaktadır. Bu yöntem, modelin katsayılarını, gerçek verilere en uygun olan değerlere ayarlamak için kulanılmaktadır. Maksimum olabilirlik tahmini, gerçek verilere en uygun modeli bulmak için bir maliyet fonksiyonunu minimize etmektedir: (3.2) (D. W. Hosmer Jr, 2013)

$$L(\beta) = \sum[yi(\beta 0 + \beta 1xi1 + \ldots + \beta nxin) - log(1 + exp(\beta 0 + \beta 1xi1 + \beta nxin))] \quad \textbf{(3.2)}$$

$L(\beta)$: Maliyet fonksiyonudur.

yi : İlgili gözlemin bağımlı değişkenin değeri (0 ya da 1).

$xi1, xi2, \ldots, xin$: İlgili gözlemin bağımsız değişkenlerinin değerleri.

Maliyet fonksiyonunun minimum değeri için katsayıları β0, β1, ..., βn optimize ederek modeli eğitiriz.

Lojistik regresyon modelinin performansını ölçmek için çeşitli metrikler kullanılabilir, örneğin doğruluk, kesinlik, duyarlılık ve F1 skoru. Ayrıca, modelin performansını görselleştirmek için ROC eğrisi ve AUC (Eğri Altındaki Alan) değeri kullanılabilmektedir (Bkz., 5.1. Bölüm, s48-49) (Dal Pozzolo, Caelen, Johnson, & Bontempi, 2015).

Eksik değerlerin doldurulması, kategorik değişkenlerin kodlanması ve özelliklerin ölçeklendirilmesi gibi ön işleme adımları gerçekleştirilmektedir. Veri dengesizliği, dolandırıcılık veya dolandırıcılık olmayan işlem sayısı arasındaki fark nedeniyle, makine veri çoğaltma yöntemlerinden olan Sentetik Azınlık Yüksek Örnekleme Tekniği (SMOTE) ile veri kümesi üretimi yapılabilmektedir (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

Modelin performansını artırmak için önemli özelliklerin seçilmesi önem arz etmektedir. Özellik seçimi için çeşitli yöntemler kullanılmaktadır. Örneğin LASSO, RFE veya istatistiksel testler (ör. kikare testi) seçilebilecek yöntemler arasındadır (Demsar, 2006).

Ön işleme ve özellik seçimi tamamlandığında, veri kümesi eğitim ve test setlerine ayrılır. Lojistik regresyon modeli eğitim seti üzerinde eğitilir ve test seti üzerinde değerlendirilir. Modelin performansını ölçmek için doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metrikler kullanılabilmektedir (Dal Pozzolo, Caelen, Johnson, & Bontempi, 2015).

Lojistik regresyon modelinin hiperparametreleri (ör. düzenlileştirme parametresi) optimize edilerek modelin performansı artırılabilmektedir. K-fold çapraz doğrulama tekniği kullanılarak bu hiperparametrelerin farklı kombinasyonları denetilebilmektedir. K-katlı çapraz doğrulama, veri kümesinin rastgele alt kümelerine ayrılmasını ve modelin bu alt kümeler üzerinde eğitilip değerlendirilmesini içeren bir doğrulama sürecidir. Her bir alt küme, sırayla test verisi olarak kullanılırken diğerleri eğitim verisi olarak kullanılır. Bu işlem K kez tekrarlanır, her alt küme sırayla test kümesi olarak kullanıldığı için her gözlem verisi hem eğitim hem de test setinde yer amaktadır. Bu sayede modelin genel performansı daha güvenilir bir şekilde değerlendirilir ve hiperparametre kombinasyonları arasında karşılaştırma yapılır (Friedman, Hastie, & Tibshirani, 2009). Bu yöntem sayesinde en iyi sonuç veren hiperparametreler seçilir.

Lojistik regresyon modeli, kredi kartı dolandırıcılık tespiti için başarılı bir şekilde uygulanabilmektedir. Modelin performansı yeterli düzeyde olduğunda, gerçek finansal işlemlerde dolandırıcılık tespiti için kullanılabilmektedir. Model sürekli olarak yeni verilerle güncellenerek ve performansı takip edilerek daha iyi sonuçlar elde etmeye devam edebilmektedir.

Algoritmanın genel mantığı verilere en uygun parametreleri öğrenerek bağımlı değişkenin olasılığını tahmin etmektir. Modelin performansı çeşitli metrikler ve görselleştirmeler kullanılarak değerlendirilebilmektedir.

Kredi kartı dolandırıcılık tespiti uygulamasında lojistik regresyon modeli veri ön işleme, özellik seçimi, model eğitimi ve değerlendirme, model ayarlanması ve sonuçların uygulanması gibi adımları içermektedir. Bu süreçte veri dengesizliği gibi sorunlarla başa çıkmak ve modelin performansını artırmak için çeşitli teknikler kullanılabilmektedir.

Lojistik regresyonun başarılı bir şekilde uygulanması kredi kartı dolandırıcılık tespitinde önemli bir rol oynayarak hem tüketicilerin hem de bankaların mali kayıplarını azaltmaya yardımcı olabilmektedir. Ayrıca, lojistik regresyon modeli, sürekli olarak yeni verilerle güncellenerek ve performansı takip edilerek daha iyi sonuçlar elde etmeye devam edebilmektedir.

Sonuç olarak, lojistik regresyon, kredi kartı dolandırıcılık tespiti alanında güçlü ve kullanışlı bir yöntemdir. Bu tür problemlerde kullanılması, etkili ve hızlı bir şekilde dolandırıcılık tespiti sağlayarak, finansal sistemlerin güvenliğini ve istikrarını korumaya katkıda bulunacaktır.

3.6 Destek Vektör Makineleri (Support Vector Machine) Algoritması

Destek Vektör Makineleri (SVM), sınıflandırma ve regresyon problemleri için kullanılan güçlü ve esnek bir makine öğrenimi yöntemidir. İkili sınıflandırma problemlerinde, SVM algoritması iki sınıf arasındaki en geniş marjı bulmaya çalışır ve bu marjın sınırlarındaki veri noktalarına "destek vektörleri" denir. SVM, kredi kartı dolandırıcılığı tespiti gibi ikili sınıflandırma problemleri için uygulanabilir bir yöntemdir (Bhattacharyya, 2011). SVM algoritması aşağıdaki gibi genel olarak anlatılabilir:

SVM, iki sınıfı ayıran bir hiperdüzlem bulmaya çalışır. Hiperdüzlem şu şekilde tanımlanabilir: : (Aaron Hertzmann, 2015) (3.3)

$$w \cdot x + b = 0 \quad \textbf{(3.3)}$$

w : Ağırlık vektörü

x : Özellik vektörü

b : Bias terimi

SVM iki sınıf arasındaki marjı en büyük hale getirecek şekilde w ve b değerlerini öğrenir. Bunu ise aşağıdaki optimizasyon problemini çözerek gerçekleştirir: (Aaron Hertzmann, 2015) (3.4)

$$min\ ||w||^2 \,/\, 2\ s.t.\ yi(w \cdot xi + b) \geq 1, i = 1, \ldots, n \quad \textbf{(3.4)}$$

yi :İlgili gözlemin sınıf etiketi (-1 veya 1)

xi : İlgili gözlemin özellik vektörü

n : Veri noktalarının sayısı

Doğrusal olarak ayrılabilir olmayan problemler için, SVM, özellik alanını dönüştüren çekirdek yöntemini kullanılmaktadır. Bu, özellik vektörlerini daha yüksek boyutlu bir alana eşlemek için bir çekirdek fonksiyonu (ör. RBF, polinom) kullanır. Bu sayede, doğrusal olarak ayrılabilir olmayan problemler için bile SVM başarılı bir şekilde kullanılabilmektedir (Aaron Hertzmann, 2015).

Eksik değerlerin doldurulması, kategorik değişkenlerin kodlanması ve özelliklerin ölçeklendirilmesi gibi ön işleme adımları gerçekleştirilmelidir. Veri dengesizliği, dolandırıcılık veya dolandırıcılık olmayan işlemlerin sayısı arasındaki fark nedeniyle, makine veri çoğaltma yöntemlerinden olan Sentetik Azınlık Yüksek Örnekleme Tekniği (SMOTE) ile veri kümesi üretimi yapılabilmektedir (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

Modelin performansını artırmak için önemli özelliklerin seçilmesi önemlidir. Özellik seçimi için çeşitli yöntemler kullanılabilmektedir. Örneğin LASSO, RFE veya istatistiksel testler (ör. kikare testi) seçilebilecek yöntemler arasında yer almaktadır (Demsar, 2006).

Ön işleme ve özellik seçimi tamamlandığında veri kümesi eğitim ve test setlerine ayrılır. SVM modeli eğitim seti üzerinde eğitilir ve test seti üzerinde

değerlendirilir. Modelin performansını ölçmek için doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metrikler kullanılabilmektedir (Dal Pozzolo, Caelen, Johnson, & Bontempi, 2015).

SVM modelinin hiperparametreleri (ör. C, gamma) optimize edilerek modelin performansı artırılabilir. K-katlı çapraz doğrulama kullanarak farklı hiperparametre kombinasyonları denenebilir ve en iyi sonucu veren hiperparametreler seçilebilir (Friedman, Hastie, & Tibshirani, 2009).

SVM modeli, kredi kartı dolandırıcılık tespiti için başarılı bir şekilde uygulanabilmektedir. Modelin performansı yeterli düzeyde olduğunda, gerçek dünya işlemlerinde dolandırıcılık tespiti için kullanılabilmektedir. Model, sürekli olarak yeni verilerle güncellenerek ve performansı takip edilerek daha iyi sonuçlar elde etmeye devam edebilmektedir.

Destek Vektör Makineleri, kredi kartı dolandırıcılık tespiti gibi ikili sınıflandırma problemlerinde güçlü ve kullanışlı yöntemler arasında yer almaktadır. Bu tür problemlerde kullanılması, etkili ve hızlı bir şekilde dolandırıcılık tespiti sağlayarak, finansal sistemlerin güvenliğini ve istikrarını korumaya katkıda bulunabilir.

3.6.1 Karar ağaçları (Decision trees) algoritması

Karar ağaçları sınıflandırma ve regresyon problemleri için kullanılan yaygın ve sezgisel bir makine öğrenimi yöntemidir. Karar ağacı modeli veri kümesini daha küçük alt kümelerine bölerek ve bu süreç boyunca öğrenilen kurallarla bir ağaç yapısı oluşturarak çalışır. Bu algoritma, kredi kartı dolandırıcılığı tespiti gibi sınıflandırma problemleri için uygulanabilir.

Karar ağaçlarının temel yapı taşları şunlardır (Han, 2011):

- Düğüm (Node): Karar ağacının her bir öğesini temsil eder ve bir özellik üzerinde bir test gerçekleştirir.
- Kök Düğüm (Root Node): Ağacın en üstündeki düğüm, tüm örnekleri içerir ve ilk bölünmeyi gerçekleştirir.
- İç Düğüm (Internal Node): Alt dallara sahip düğümlerdir ve veri kümesini daha küçük alt kümelerine bölmektedir.

- Yaprak Düğüm (Leaf Node): Ağacın altında kalan düğümlerdir ve sınıflandırma sonucunu temsil etmektedir.

Karar ağaçlarında öğrenme süreci sırasında ağacın dallanma kriteri olarak kullanılacak özellikler ve dallanma noktaları seçilmelidir. Bu seçimler, sınıflandırma hatalarını en aza indiren özellikler ve eşik değerleri temel alınarak gerçekleştirilir (Chen, 2005). Aşağıda, bir karar ağacının dallanma kriterlerini belirlemek için kullanılan iki popüler ölçü sunulmaktadır:

Gini Impurity: Bir düğümdeki örneklerin karışıklığını ölçer ve sınıflandırma hatalarını en aza indiren özellikler ve eşik değerleri temel alınarak dallanma kriteri olarak kullanılmaktadır (Han, 2011).(3.5)

$$Gini(D) = 1 - \sum [p(i|D)^2] \quad \textbf{(3.5)}$$

Burada, $p(i|D)$ özellik i'nin D düğümündeki olasılığıdır.

Information Gain: Bir düğümdeki veri kümesinin entropisi (rastgelelik) azalma miktarını ölçer ve özellikler ve eşik değerleri seçilerek entropi azaltılır (Han, 2011). (3.6)(3.7)

$$Entropy(D) = -\sum [p(i|D) * log_2(p(i|D))] \quad \textbf{(3.6)}$$

$$Information\_Gain(D, A) = Entropy(D) - \sum [(|D\_v| / |D|)$$

$$Entropy(D\_v\ ) \quad \textbf{(3.7)}$$

Burada D mevcut düğüm ve A dallanma kriteri olarak kullanılacak özelliktir;

D_v ise özellik A üzerindeki değerlere göre oluşturulan alt kümelerdir.

Karar ağaçlarının temel avantajlarından bazıları, sezgisel yapısı, hızlı eğitim süresi ve anlaşılabilir sonuçlardır. Kredi kartı dolandırıcılık tespiti için aşağıda, karar ağaçlarının nasıl kullanılacağına dair adımlar sunulmaktadır.

Veri ön işleme adımları, eksik değerlerin doldurulması, kategorik değişkenlerin kodlanması ve özelliklerin ölçeklendirilmesi gibi işlemleri içermektedir. Ayrıca, veri dengesizliği nedeniyle, makine veri çoğaltma yöntemlerinden olan Sentetik Azınlık Yüksek Örnekleme Tekniği (SMOTE) ile veri kümesi üretimi yapılabilmektedir (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

Karar ağaçlarının performansını artırmak için önemli özelliklerin seçilmesi önemlidir. Özellik seçimi için çeşitli yöntemler kullanılabilir. Örneğin Gini impurity,

information gain, kikare testi veya başka özellik seçimi yöntemleri seçilebilecek yöntemler arasında yer almaktadır (Demsar, 2006).

Ön işleme ve özellik seçimi tamamlandığında, veri kümesi eğitim ve test setlerine ayrılır. Karar ağacı modeli eğitim seti üzerinde eğitilir ve test seti üzerinde değerlendirilmektedir. Modelin performansını ölçmek için doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metrikler kullanılabilir. Ayrıca, modelin performansını görselleştirmek için ROC eğrisi ve AUC değeri kullanılabilir (Dal Pozzolo, Caelen, Johnson, & Bontempi, 2015).

Karar ağacının hiperparametreleri (ör. ağaç derinliği, minimum örnek sayısı) optimize edilerek modelin performansı artırılabilir. Çapraz doğrulama kullanarak farklı hiperparametre kombinasyonları denenebilir ve en iyi sonucu veren hiperparametreler seçilebilir (Friedman, Hastie, & Tibshirani, 2009).

Karar ağaçları kredi kartı dolandırıcılık tespiti gibi sınıflandırma problemlerinde sezgisel ve anlaşılır bir yöntemdir. Modelin başarılı uygulanması için uygun veri ön işleme, özellik seçimi ve model ayarlaması adımlarının gerçekleştirilmesi önemlidir.

Karar ağaçları, diğer makine öğrenimi algoritmalarına göre çeşitli avantajlara ve dezavantajlara sahiptir. Karar ağaçlarının avantajları arasında sezgisel yapısı, kolay anlaşılabilirliği ve hızlı eğitim süreleri bulunmaktadır. Ancak, karar ağaçlarının aşırı uyum (overfitting) eğilimi ve yapılarının her veriye göre değişkenlik göstermesi (yani, modelin kararlı olmaması) gibi dezavantajları da bulunmaktadır.

Bu bağlamda, kredi kartı dolandırıcılık tespiti için birleştirilmiş yaklaşımlar (ör. Rastgele Orman veya Uyarlamalı Artırma gibi toplu öğrenme yöntemleri) kullanılarak karar ağaçlarından daha güçlü ve kararlı modeller elde etmek mümkündür.

3.6.2 Rastgele orman (Random forest) algoritması

Random Forest, birden fazla karar ağacını bir araya getirerek sınıflandırma ve regresyon problemleri için kullanılan bir topluluk makine öğrenimi yöntemidir. Rastgele Orman algoritması, temel olarak karar ağaçlarının güçlerini kullanarak aşırı uyum ve kararsızlık gibi dezavantajları gidermektedir.

Rastgele Orman algoritmasının temel adımları şunlardır (Louppe, 2015):

- Yeniden örnekleme: Veri kümesinden tekrarlı örnekleme ile birden fazla alt küme oluşturulur. Her alt küme, ayrı bir karar ağacı eğitmek için kullanılmaktadır.
- Rastgele özellik seçimi: Her düğümde, rastgele özellikler seçilir ve en iyi dallanma kriteri belirlenir.
- Karar ağaçlarının eğitimi: Her alt küme için bir karar ağacı eğitilir ve rastgele özellik seçimi ile dallanma kriterleri belirlenir.
- Topluluk sonuçları: Tüm karar ağaçlarının tahminleri bir araya getirilerek, oylama yöntemi ile nihai sınıflandırma sonucu elde edilmektedir.

Rastgele Orman algoritması için matematiksel olarak, B adet karar ağacı kullanılmakta ve her bir ağaç için yeniden örnekleme uygulanmaktadır (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) (3.8).

T_i : i'nci karar ağacı

X : Veri kümesi

N : Örnek sayısı

B : Bootstrap örnek sayısı

Bootstrap örnekleme: B adet bootstrap örnek oluşturulur.

$$X_b = \{x_1, x_2, \ldots, x_n\}_b, b = 1, 2, \ldots, B \quad \textbf{(3.8)}$$

Rastgele özellik seçimi: Her düğümde, rastgele özellikler seçilir ve en iyi dallanma kriteri belirlenir.

Karar ağaçlarının eğitimi: Her alt küme için bir karar ağacı eğitilir ve rastgele özellik seçimi ile dallanma kriterleri belirlenir. (Chawla, Bowyer, Hall, & Kegelmeyer, 2002)(3.9)

$$T\_i = T\_i(X\_b), i = 1, 2, \ldots, B \quad \textbf{(3.9)}$$

Topluluk Sonuçları: Tüm karar ağaçlarının tahminleri bir araya getirilerek, oylama yöntemi ile nihai sınıflandırma sonucu elde edilir. (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) (3.10)

$$y\_pred = majority\_vote(\{T\_1(x), T\_2(x), \ldots, T\_B(x)\}) \quad \textbf{(3.10)}$$

Kredi kartı dolandırıcılık tespiti için Rastgele Orman algoritması aşağıdaki adımları içermektedir.

Veri ön işleme adımları, eksik değerlerin doldurulması, kategorik değişkenlerin kodlanması ve özelliklerin ölçeklendirilmesi gibi işlemleri içermektedir. Ayrıca, veri dengesizliği nedeniyle, makine veri çoğaltma yöntemlerinden olan Sentetik Azınlık Yüksek Örnekleme Tekniği (SMOTE) ile veri kümesi üretimi yapılabilmektedir (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

Ön işleme tamamlandığında, veri kümesi eğitim ve test setlerine ayrılır. Rastgele Orman modeli eğitim seti üzerinde eğitilir ve test seti üzerinde değerlendirilir. Modelin performansını ölçmek için doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metrikler kullanılabilir (Dal Pozzolo, Caelen, Johnson, & Bontempi, 2015).

Rastgele Orman hiperparametreleri (ör. ağaç sayısı, ağaç derinliği, minimum örnek sayısı) optimize edilerek modelin performansı artırılabilmektedir. K-katlı çapraz doğrulama kullanarak farklı hiperparametre kombinasyonları denenebilir ve en iyi sonucu veren hiperparametreler seçilebilir (Friedman, Hastie, & Tibshirani, 2009).

Rastgele Orman modeli, kredi kartı dolandırıcılık tespiti için başarılı bir şekilde uygulanabilmektedir. Modelin performansı yeterli düzeyde olduğunda gerçek finansal işlemlerde dolandırıcılık tespiti için kullanılabilir. Model, sürekli olarak yeni verilerle güncellenerek ve performansı takip edilerek daha iyi sonuçlar elde etmeye devam edebilir.

Özetle Rastgele Orman, kredi kartı dolandırıcılık tespiti gibi sınıflandırma problemlerinde güçlü ve kararlı bir yöntemdir. Karar ağaçlarının aşırı uyum ve kararsızlık gibi dezavantajlarını gidererek, daha iyi performanslı modeller elde edilebilir. Başarılı bir uygulama için, uygun veri ön işleme, model ayarlaması ve performans değerlendirmesi adımları gerçekleştirilmelidir.

3.6.3 Yapay sinir ağları (Artificial neural networks) algoritması

Yapay Sinir Ağları, insan beynindeki biyolojik sinir ağlarının çalışma prensiplerinden esinlenerek geliştirilmiş, karmaşık ve doğrusal olmayan problemler

için kullanılan güçlü bir makine öğrenimi yöntemidir. ANN, sınıflandırma ve regresyon problemleri gibi çeşitli alanlarda başarıyla uygulanabilmektedir.

Yapay sinir ağlarının temel yapı taşı, "nöron" adı verilen basit işlem birimleridir. Nöronlar, ağırlıklı girdilerin toplanarak aktivasyon fonksiyonu kullanarak çıktı üretir. Yapay sinir ağları, bu nöronların katmanlar halinde bir araya getirilmesi ile oluşturulur (Goodfellow, 2016).

Bir Yapay Sinir Ağının temel matematiksel ifadesi şu şekildedir (Nath & Rokonuzzaman, 2018):

Giriş katmanı : $x = (x\_1, x\_2, \ldots, x\_n)$

Ağırlıklar: w_ij, i: girdi indeksi, j: nöron indeksi

Toplam ağırlıklı girdisi: (3.11)

$$z\_j = \sum w\_ij * x\_i \quad \textbf{(3.11)}$$

Aktivasyon fonksiyonu: (3.12)

$$f(z) \quad \textbf{(3.12)}$$

Çıktı: (3.13)

$$y\_j = f(z\_j) \quad \textbf{(3.13)}$$

Kredi kartı dolandırıcılık tespiti için Yapay Sinir Ağları algoritması aşağıdaki adımları içermektedir.

Veri ön işleme adımları, eksik değerlerin doldurulması, kategorik değişkenlerin kodlanması ve özelliklerin ölçeklendirilmesi gibi işlemleri içermektedir. Ayrıca, veri dengesizliği nedeniyle, makine veri çoğaltma yöntemlerinden olan Sentetik Azınlık Yüksek Örnekleme Tekniği (SMOTE) ile veri kümesi üretimi yapılabilmektedir (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

Ön işleme tamamlandığında, veri kümesi eğitim ve test setlerine ayrılır. ANN modeli eğitim seti üzerinde eğitilir ve test seti üzerinde değerlendirilir. Modelin performansını ölçmek için doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metrikler kullanılabilir. Ayrıca, modelin performansını görselleştirmek için ROC eğrisi ve AUC değeri kullanılabilir (Dal Pozzolo, Caelen, Johnson, & Bontempi, 2015).

Yapay Sinir Ağları hiperparametreleri (ör. katman sayısı, nöron sayısı, aktivasyon fonksiyonları, öğrenme oranı) optimize edilerek modelin performansı artırılabilir. K-katlı çapraz doğrulama kullanarak farklı hiper parametre kombinasyonları denenebilir ve en iyi sonucu veren hiperparametreler seçilebilir (Friedman, Hastie, & Tibshirani, 2009).

Yapay Sinir Ağları modeli kredi kartı dolandırıcılık tespiti için başarılı bir şekilde uygulanabilir. Modelin performansı yeterli düzeyde olduğunda, olduğunda gerçek finansal işlemlerde dolandırıcılık tespiti için kullanılabilir. Model sürekli olarak yeni verilerle güncellenerek ve performansı takip edilerek daha iyi sonuçlar elde etmeye devam edebilir.

Özetle, Yapay Sinir Ağları, kredi kartı dolandırıcılık tespiti gibi sınıflandırma problemlerinde güçlü ve esnek bir yöntemdir. Karmaşık ve doğrusal olmayan ilişkileri öğrenme yeteneği sayesinde Yapay Sinir Ağları algoritması başarılı sonuçlar elde etmeye yardımcı olmaktadır. Başarılı bir uygulama için, uygun veri ön işleme, model ayarlama ve performans değerlendirmesi adımları gerçekleştirilmelidir.

4. VERİ HAKKINDA

4.1 Veri Seti İncelemesi ve Analizi

Kredi kartı dolandırıcılığı, finansal sektörde önemli ve sürekli büyüyen bir sorundur. Bu tezde kullanılan veri seti Worldline ve ULB'nin (Brüksel Libre Universitesi) Makine Öğrenimi Grubu'nun büyük veri madenciliği ve dolandırıcılık tespiti konusundaki bir araştırma iş birliği sırasında toplanmış ve analiz edilmiştir. Avrupalı kredi kartı sahipleri tarafından Eylül 2013'te gerçekleştirilen işlemlere dayanan anonimleştirilmiş bir veri kümesidir. Bu veri seti, kredi kartı dolandırıcılık tespitinde kullanılacak modellerin geliştirilmesi ve değerlendirilmesi amacıyla toplanmıştır (ULB, 2022).

4.2 Veri Setinin Yapısı ve Özellikleri

Veri seti 284.807 işlemi kapsamakta olup bunların 492'si dolandırıcılık olarak sınıflandırılmaktadır. Dolandırıcılık işlemlerinin toplam işlemlere oranı %0.172'dir, bu da veri setinin önemli ölçüde dengesiz olduğunu göstermektedir. Çizelge 4.1'de görüldüğü gibi veri setinde 30 adet özellik bulunmaktadır. Bunlar (ULB, 2022):

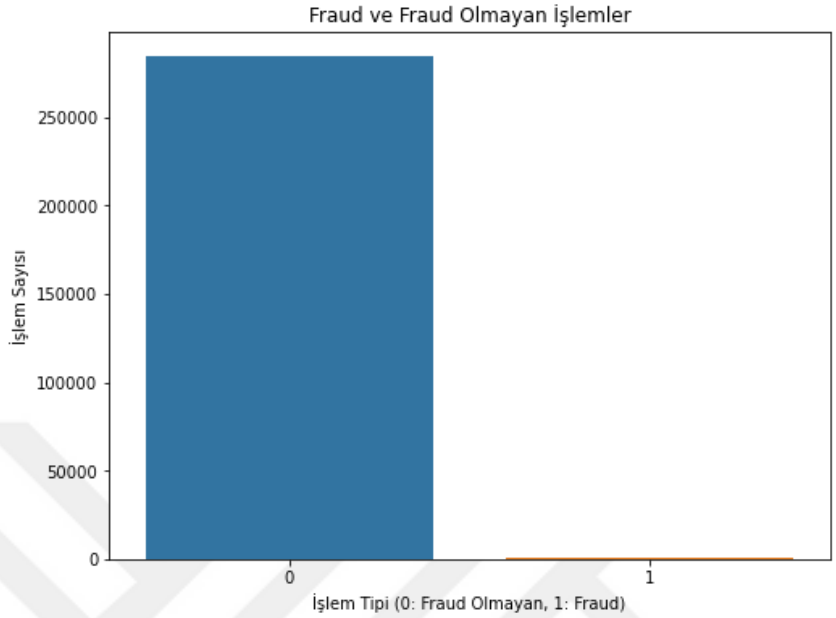
Çizelge 4.1: Veri Kümesinin Yapısı ve Özellikleri

Değişken	Tür	Açıklama
Sınıf	int	Sonuç Değişkeni (1 = Dolandırıcılık ve 0 = dolandırıcılık Olmayan)
Zaman	int	Her işlem arasındaki süre (sn)
Tutar	num	Tutar
V1, V2, ..., V28	num	Bilinmeyen bilgilere sahip özellik değişkeni

4.3 Veri Setinin Ön İşlemesi ve Temizlenmesi

Bu veri seti ön işleme ve temizleme gerektiren eksik değer veya hatalı veri içermemektedir. Bununla birlikte veri setinin dengesizliği nedeniyle, model eğitimi ve değerlendirmesi sırasında örnekleme veya makine veri üretme yöntemlerinin (ör. SMOTE, ADASYN veya Random Under/Over Sampling) kullanılması

gerekmektedir (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). Ayrıca 'Tutar' özelliği, diğer özelliklerle benzer ölçekte olmadığı için ölçeklendirme veya normalleştirme uygulanmalıdır.



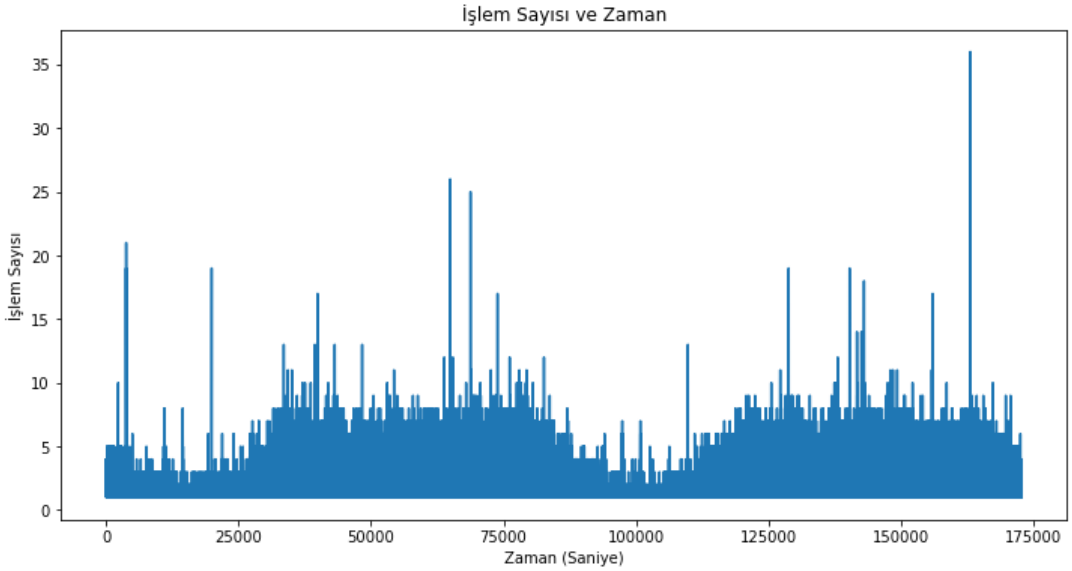
Şekil 4.1: Miktar Özelliğine Göre Sınıf Dağılımları

```
# Dolandırıcılık ve dolandırıılık olmayan işlem sayılarını
hesaplama
fraud_counts = data['Class'].value_counts()

# Grafik oluşturma
plt.figure(figsize=(8, 6))
sns.barplot(x=fraud_counts.index, y=fraud_counts.values)
plt.title('Fraud ve Fraud Olmayan İşlemler')
plt.xlabel('İşlem Tipi (0: Fraud Olmayan, 1: Fraud)')
plt.ylabel('İşlem Sayısı')
plt.show()
```

Şekil 4.2: Miktar Özelliğine Göre Sınıf Dağılımları Kodsal Gösterimi

Şekil 4.1 'de görüldüğü üzere, normal işlemlerin sayısı, sahtekarlık işlemlerine kıyasla daha fazladır. İşlemlerin büyük çoğunluğu dolandırıcılık içermeyen işlemler olduğundan veri kümesi ciddi derecede dengesizdir. Bu durum, yanlış aşırı uyum ve korelasyonlara sebep olabilmektedir (Japkowicz & Stephen, 2002).



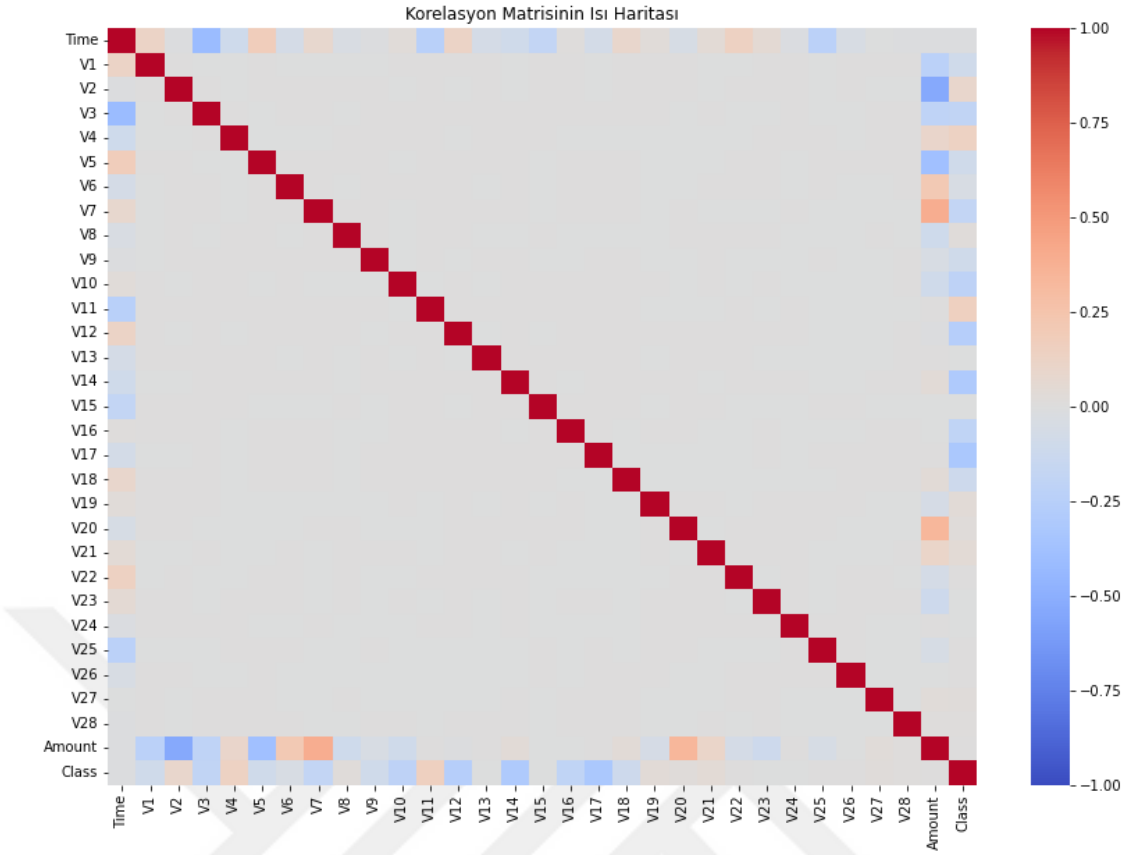
Şekil 4.3: Zaman Serisi Grafiği

Şekil 4.3'de görülebileceği gibi, açık bir döngüsel davranış bulunmaktadır. Ancak genel olarak dağılım herhangi aykırı değeri ortaya çıkaramayacak kadar yoğunluktadır. Belirgin bir zaman modeli bulunmamaktadır (Chatfield, 2003).

```
# Zaman ve işlem sayısını hesaplama
transaction_count = data['Time'].value_counts().sort_index()

# Zaman serisi grafiğini çizme
plt.figure(figsize=(12, 6))
transaction_count.plot()
plt.xlabel('Zaman (Saniye)')
plt.ylabel('İşlem Sayısı')
plt.title('İşlem Sayısı ve Zaman')
plt.show()
```

Şekil 4.4: Zaman Serisi Grafiği Kodsal Gösterimi



Şekil 4.5: Korelasyon Matrisi

```
# Korelasyon matrisi hesaplama
correlation_matrix = data.corr()

# Isı haritası grafiğini çizme
plt.figure(figsize=(14, 10))
sns.heatmap(correlation_matrix, annot=False, cmap='coolwarm', vmin=-
1, vmax=1)
plt.title('Heatmap of Correlation Matrix')
plt.show()
```

Şekil 4.6: Korelasyon Matrisi Kodsal Gösterimi

Şekil 4.5'de görülebileceği gibi her bir değişken arasındaki ilişkiyi gösteren bir korelasyon matrisini sunmaktadır. Bu korelasyon matrisi, V1'den V28'e kadar olan ana bileşenlerin birbiriyle herhangi bir ilişki göstermediğini ortaya koymaktadır. Ancak, daha detaylı incelediğimizde, 'Sınıf' yanıt değişkeninin ana bileşenlerle hem pozitif hem de negatif ilişkileri bulunurken, 'Zaman' ve 'Miktar' değişkenleri ile anlamlı bir ilişki sergilemediğini görülebilmektedir (Raschka & Mirjalili', 2020).

4.4 Veri Setinin Kullanımı

Kredi kartı dolandırıcılık veri seti dolandırıcılık tespiti için sınıflandırma modellerinin geliştirilmesi ve değerlendirilmesi amacıyla kullanılabilir. Bu veri seti ile çalışırken sınıflandırma algoritmaları (Lojistik Regresyon, Destek Vektör Makineleri, Karar Ağaçları, Rastgele Orman Ve Yapay Sinir Ağları) kullanarak dolandırıcılık işlemlerini tespit etmeye yönelik modeller oluşturulabilir ve değerlendirilebilir (Bhattacharya, Das, & Koner, 2019).

4.5 Değerlendirme Metrikleri ve Performans

Dolandırıcılık tespiti için geliştirilen modellerin performansını değerlendirmek amacıyla, dengesiz veri kümesi nedeniyle doğruluk metriği yerine kesinlik, duyarlılık ve F1 skoru gibi metrikler kullanılmalıdır. Ayrıca, ROC eğrisi ve AUC değeri de model performansını ölçmede kullanılabilmektedir (Dal Pozzolo, Caelen, Johnson, & Bontempi, 2015).

4.6 Model Seçimi ve Hiperparametre Ayarı

Farklı sınıflandırma algoritmalarının performansını karşılaştırmak ve en uygun modeli seçmek için çapraz doğrulama yöntemleri kullanılabilmektedir. Ayrıca, modellerin hiperparametrelerinin ayarlanması için Izgara Arama ve Rastgele Arama gibi yöntemler kullanılabilmektedir (Friedman, Hastie, & Tibshirani, 2009).

5. UYGULAMA

Bu çalışmada bir dizi makine öğrenmesi ve derin öğrenme algoritması kullanılarak kredi kartı dolandırıcılık tespiti yapılmıştır. Seçilen algoritmalar, Lojistik Regresyon, Destek Vektör Makineleri (SVM), Karar Ağaçları, Rastgele Orman ve Yapay Sinir Ağları'dır.

5.1 Algoritmaların Uygulanması

Bu bölüm, kredi kartı dolandırıcılık işlemlerinin analizi için Lojistik Regresyon, Destek Vektör Makineleri (SVM), Karar Ağaçları, Rastgele Orman ve Yapay Sinir Ağları (ANN) algoritmalarının uygulanma süreçlerini ayrıntılı olarak ele alınacaktır.

Algoritmaların performansını değerlendirmek için aşağıda yer alan performans ölçütleri ve metrikler kullanılacaktır. Bu ölçütler, algoritmaların etkinliğini ve performansını çeşitli açılardan değerlendirmemize yardımcı olmaktadır.

Doğruluk (Accuracy): Doğruluk, doğru tahmin edilen örneklerin toplam örnek sayısına oranını temsil etmektedir. Yüksek doğruluk, modelin genel olarak doğru tahminlerde bulunma yeteneğini göstermektedir. Ancak sınıf dengesizliği durumlarında yanıltıcı olabilmektedir. Örneğin nadir görülen dolandırıcılık işlemleri nedeniyle algoritma tüm işlemleri normal olarak sınıflandırıp yüksek doğruluk elde edebilir (Powers, 2011).

Kesinlik (Precision): Kesinlik, pozitif olarak tahmin edilen örneklerin gerçekten pozitif olan örneklerin oranını ifade etmektedir. Düşük yanlış pozitif oranı hedeflendiğinde kesinlik önemlidir. Yani, algoritmanın yanlış yere dolandırıcılık etiketi yapıştırmaması amaçlanmaktadır. Düşük kesinlik değeri yanlış pozitiflerin yüksek olduğunu ve modelin yanlışlıkla normal işlemleri dolandırıcılık olarak sınıflandırdığını gösterebilmektedir (Davis & Goadrich, 2006).

Duyarlılık (Recall): Duyarlılık, gerçek pozitif olarak tahmin edilen örneklerin gerçekten pozitif olan örneklerin oranını ifade etmektedir. Düşük yanlış negatif oranı hedeflendiğinde duyarlılık önemlidir. Algoritmanın gerçek dolandırıcılık işlemlerini yakalayabilmesi amaçlanmaktadır. Düşük duyarlılık değeri modelin gerçek dolandırıcılık işlemlerini tespit edemediğini ve yanlışlıkla normal işlemleri yanlış pozitif olarak sınıflandırdığını gösterebilmektedir (Flach, 2015).

F1 Skoru: F1 skoru, kesinlik ve duyarlılığın harmonik ortalamasını temsil etmektedir. Dengeli bir performans ölçümü sunar ve modelin hem yanlış pozitifleri hem de yanlış negatifleri minimize etme yeteneğini yansıtmaktadır. Eğer hem yanlış pozitiflerin hem de yanlış negatiflerin önemli olduğu bir senaryoda çalışılıyorsa, F1 skoru iyi bir ölçüdür (Saito & Rehmsmeier, 2015).

ROC AUC Skoru (Receiver Operating Characteristic Area Under Curve): ROC AUC skoru, algoritmaların sınıflandırma yeteneklerini değerlendirmek için yaygın olarak kullanılan bir metriktir. ROC eğrisi, farklı kesme noktalarında duyarlılık (gerçek pozitif oran) ile 1 eksi özgüllük (gerçek negatif oran) arasındaki dengeyi göstermektedir. ROC AUC, bu eğrinin altındaki alanı hesaplayarak algoritmanın farklı duyarlılık ve özgüllük düzeylerindeki performansını toplu olarak ölçmektedir. Yüksek ROC AUC skoru, modelin sınıflandırmada iyi bir ayırma yeteneği olduğunu göstermektedir (Powers, 2011).

Bu performans ölçütleri, algoritmaların farklı yönlerini analiz etmemize yardımcı olmaktadır. Yüksek doğruluk genel performansı yansıtırken, yüksek kesinlik yanlış pozitifleri minimize etme yeteneğini, yüksek duyarlılık yanlış negatifleri minimize etme yeteneğini, ve yüksek F1 skoru dengeleyici bir performansı ifade etmektedir. Bu ölçütler, algoritmaların dolandırıcılık tespitindeki etkinliğini anlamamızı ve farklı senaryolara göre değerlendirmemizi sağlamaktadır (Powers, 2011).

5.1.1 Lojistik regresyon uygulaması

Lojistik regresyon, bağımlı değişkenin kategorik olduğu durumlarda kullanılan istatistiksel bir modelleme tekniğidir. Bu özellik, dolandırıcılık tespiti gibi ikili sınıflandırma problemlerinde özellikle kullanılmaktadır. Bu çalışmada, lojistik regresyon kullanarak kredi kartı dolandırıcılık tespiti gerçekleştirilmiştir (Udjianto, 2006).

Lojistik regresyon verilerin dağılımını analiz ederek bir olasılık tahmini oluşturmaktadır. Bu tahmin, belirli bir eşik değerine göre sınıflandırma yapmak için kullanılmaktadır. Bu yaklaşım, özelliklerin etkileşimini ve dolandırıcılık durumunun karmaşıklığını modellememizi sağlamaktadır (Udjianto, 2006).

Öncelikle kullanılacak veri setini pandas kütüphanesi ile okuduk. Veri seti, kredi kartı işlemleri ile ilgili özellikleri ve işlemlerin dolandırıcılık olup olmadığını belirten bir etiket içermektedir.

Veri setinden özellikler ve etiketleri ayırıldı. Bu veri seti özelliklerin farklı ölçeklerde olması nedeniyle ölçeklendirildi. Bu lojistik regresyon gibi birçok makine öğrenmesi modeli için önemli bir adımdır, çünkü ölçeklendirme tüm özelliklerin model tarafından eşit derecede dikkate alınmasını sağlamaktadır.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report
from imblearn.over_sampling import SMOTE

# Veri kümesini okuma
df = pd.read_csv("creditcard.csv")

# Özellikler ve etiketleri ayırma
X = df.drop("Class", axis=1)
y = df["Class"]

# Özellikleri ölçeklendirme
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Şekil 5.1: Lojistik Regresyon Özelliklere Ayırma ve Ölçeklendirme Kodsal Görünümü

Veri seti dolandırıcılık olmayan işlemlerin sayısının dolandırıcılık işlemlerinden çok daha fazla olduğu dengesiz bir veri setidir. Bu modelin çoğunluk

sınıfını öğrenmesine ve azınlık sınıfını göz ardı etmesine neden olabilmektedir. Bu problemi çözmek için, azınlık sınıfını çoğunluk sınıfının boyutuna çıkarmak için makine veri çoğaltma yöntemlerinden olan Sentetik Azınlık Yüksek Örnekleme Tekniği (SMOTE) ile veri kümesi üretim tekniğini kullanmaktadır.

```
# Veri dengesizliği ile başa çıkmak için veri üretim tekniği
(SMOTE) kullanılmıştır.
sm = SMOTE(sampling_strategy='minority')
X_sm, y_sm = sm.fit_resample(X_scaled, y)
```

Şekil 5.2: Lojistik Regresyon Sentetik Azınlık Yüksek Örneklem Tekniği Kodsal Gösterimi

Veri kümesi içerisindeki özellikler (X) ve hedef sınıfı (y) ayırıldı. Daha sonra, tüm özelliklerin aynı ölçeğe sahip olmasını sağlamak için ölçeklendirme işlemi gerçekleştirildi. Bu, algoritmanın tüm özellikleri adil bir şekilde değerlendirmesine yardımcı olmaktadır.

```
# Eğitim ve test setlerinin ayrılması
X_train, X_test, y_train, y_test = train_test_split(X_sm, y_sm,
test_size=0.20, random_state=42)
```

Şekil 5.3: Lojistik Regresyon Eğitim ve Test Parçaları Ayrımının Kodsal Görünümü

Veri kümesindeki dengesizlik, makine öğrenmesi modellerinin performansını etkileyebilecek önemli bir sorun olabilir. Bu durum, çoğunluk sınıfının örnekleri, azınlık sınıfının örneklerini büyük ölçüde aştığında meydana gelmektedir. Bu çalışmada, veri dengesizliği ile başa çıkmak için kullanılmaktadır.

Veriyi ölçeklendirdikten ve yeniden örnekledikten sonra, veri setini eğitim ve test setlerine bölünmesi gerekmektedir. Daha sonra, GridSearchCV kullanarak lojistik regresyon modelimizin hiperparametrelerinin ayarlanması gerekmektedir. Bu işlem, modelin veriye en iyi şekilde uymasını sağlar.

```
# Lojistik Regresyon modeli için parametre aralığını tanımlama
param_grid = {'C': [0.01, 0.1, 1, 10, 100], 'max_iter': [100, 200, 500, 1000]}

# GridSearchCV ile modeli optimize etme
lr = LogisticRegression()
grid_search = GridSearchCV(lr, param_grid, cv=5, scoring='f1_macro')
grid_search.fit(X_train, y_train)

# En iyi parametrelerle Lojistik Regresyon modelini eğitme
best_lr = grid_search.best_estimator_
```

Şekil 5.4: Lojistik Regresyon Parametre Tanımı ve Model Optimizasyonu Kodsal Görünümü

Modeli eğittikten ve en iyi hiperparametrelerin bulunması ile modeli test veri kümesindeki verileri sınıflandırmak için kullanılmaktadır. Modelin performansını değerlendirmek için, test veri setindeki tahminlerin gerçek etiketlerle karşılaştırılması gerekmektedir. Bu amaçla, sınıflandırma raporu adı verilen bir metrik kullandık.

Bu rapor, modelin dolandırıcılık ve dolandırıcılık olmayan işlemleri ne kadar iyi tahmin ettiğini gösterir. Özellikle, kesinlik, duyarlılık ve F1 skoru gibi metrikler, modelin performansını özetler.

```
# Modelin performansını değerlendirme
y_pred = best_lr.predict(X_test)
print(classification_report(y_test, y_pred))
```

Şekil 5.5: Lojistik Regresyon Model Performans Ölçümü Kodsal Görünümü

Şekil 5.5'de görülebileceği üzere Lojistik Regresyon modelinin ROC AUC skoru 0.989 olarak bulunmaktadır. Bu, modelin kredi kartı dolandırıcılık tespit etme konusunda iyi bir yeteneğe sahip olduğunu göstermektedir.

Sonuç olarak lojistik regresyon modelimiz dolandırıcılık ve dolandırıcılık olmayan işlemleri oldukça iyi bir şekilde tespit etmiştir. Kesinlik, geri çağırma ve F1 skoru metriklerine göre modelin genel performansı yüksektir. Bu sonuçlar lojistik regresyonun kredi kartı dolandırıcılık tespiti gibi karmaşık sınıflandırma problemlerini çözmek için etkili bir araç olabileceğini göstermektedir.

Logistic Regression Uygulamasının Sonucu:

	Kesinlik	Duyarlılık	f1-skoru	Destek
0	0.93	0.98	0.95	56750
1	0.97	0.92	0.95	56976
Doğruluk			0.95	113726
Makro ort	0.95	0.95	0.95	113726
Ağırlıklı ort	0.95	0.95	0.95	113726

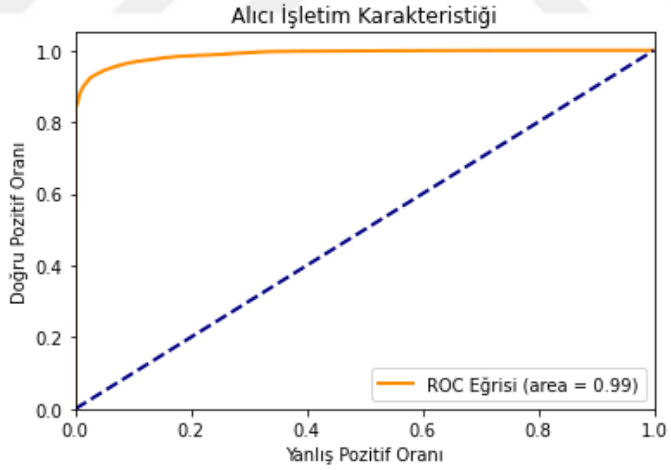
ROC AUC Skoru :0.9892609246091096

Doğruluk :0.9482704043050841

Kesinlik :0.9741809744779583

Duyarlılık :0.9211597865768043

F1 Skoru :0.946928760227693



Şekil 5.6: Lojistik Regresyon Modeli için ROC Eğrisi

5.1.2 Destek vektör makineleri uygulaması

Kredi kartı dolandırıcılık tespitinde bir başka algoritma da Destek Vektör Makineleri (SVM) algoritmasıdır. SVM, yüksek boyutlu veri setlerinde karmaşık sınıflandırma görevlerini gerçekleştirebilme yeteneği ile bilinmektedir. Bu algoritma, ayrılabilir hale getirmek için veri noktalarını bir kenara yerleştirerek, iki sınıf arasında optimum bir hiper düzlem bulmayı hedefler (Bhattacharyya, 2011).

Bu çalışmada hedef değişkenimiz olan Sınıf'ı tahmin etmek için SVC modeli kullanılmıştır. SVM iki sınıf arasındaki en geniş marjı bulmayı amaçlayan bir sınırlayıcı olarak çalışır ve bu modelin genelleme yeteneğini artırmaktadır. Bu fonksiyon modelin karmaşıklığını kontrol etmek için C hiperparametresini ve modelin maksimum iterasyon sayısını tanımlamak için max_iter hiperparametresini kabul etmektedir. Ayrıca dengesiz veri setlerinde kullanılmak üzere class_weight hiperparametresini de kabul etmektedir (Aaron Hertzmann, 2015).

Modeli eğitmek ve değerlendirmek için standart bir süreç izlenmektedir. Öncelikle, hedef değişken Sınıf'ı tahmin etmek için kullanılacak olan özellikler (X) ve hedef değişkeni (y) ayırılmaktadır. Ardından dengesiz bir sınıf dağılımına sahip veri kümesini dengelemek azınlık sınıfını çoğunluk sınıfının boyutuna çıkarmak için makine veri çoğaltma yöntemlerinden olan Sentetik Azınlık Yüksek Örnekleme Tekniği (SMOTE) ile veri kümesi üretim tekniğini kullanılmaktadır. Model eğitimi ve testi için veri seti %80 eğitim ve %20 test verisi olacak şekilde bölünmektedir.

Veri kümesini ve hedef değişkeni tanımlanmaktadır. Bu durumda hedef değişkenimiz "Sınıf" sütunu olacaktır ve bu bir işlemin dolandırıcılık olup olmadığını belirtecektir.

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import classification_report, roc_curve, auc
from imblearn.over_sampling import SMOTE
import matplotlib.pyplot as plt

# Veri kümesini okuma
df = pd.read_csv("creditcard.csv")

# Özellikler ve etiketleri ayırma
X = df.drop("Class", axis=1)
y = df["Class"]

# Özellikleri ölçeklendirme
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Şekil 5.7: SVM - Özelliklere Ayırma ve Ölçeklendirme Kodsal Görünümü

Veriyi özellikler ve hedef değişken olarak ayırılmaktadır. Özellikler ölçeklendirilir ve dengesizlik giderilmektedir. Veri seti daha sonra eğitim ve test setlerine ayrılmaktadır.

Bu çalışmada SVM modelini eğitmek için Scikit-learn kütüphanesinin SVC sınıflandırıcı fonksiyonu kullanılmaktadır. Modelin eğitildikten sonra test veri seti üzerinde tahminler yapmak için kullanılmaktadır.

Dolandırıcılık tespiti genellikle dengesiz bir veri kümesi ile karşı karşıya olduğumuz bir problemdir çünkü dolandırıcılık işlemlerinin sayısı genellikle normal işlemlerin çok altındadır. Bu durumda veri kümesi üretim teknikleri kullanılarak azınlık sınıfını (dolandırıcılık işlemleri) veri kümesi arttırımı yapılmaktadır. Bu modelin dolandırıcılık işlemlerini daha iyi tanımasına ve dolayısıyla daha doğru tahminler yapmasına yardımcı olmaktadır.

```
# Veri dengesizliği ile başa çıkmak için veri üretim tekniği kullanılarak yeniden örnekleme
sm = SMOTE(sampling_strategy='minority')
X_sm, y_sm = sm.fit_resample(X_scaled, y)
```

Şekil 5.8: SVM - Sentetik Azınlık Yüksek Örneklem Tekniği Kodsal Gösterimi

Verinin önceden işlenmesinden ve hazırlanmasından sonra, SVM modelini eğitmek için SVC fonksiyonunu kullanılmaktadır. Modeli eğittikten sonra, tahminler yapmak ve modelin performansını değerlendirmek için test setini kullanılmaktadır.

```
# Eğitim ve test setlerini ayırma
X_train, X_test, y_train, y_test = train_test_split(X_sm, y_sm, test_size=0.20, random_state=42)
```

Şekil 5.9: SVM - Eğitim ve Test Parçaları Ayrımının Kodsal Görünümü

```
# Modelin performansını değerlendirme
y_pred = svc.predict(X_test)
print(classification_report(y_test, y_pred))
```

Şekil 5.10: SVC Model Oluşturulmasının ve Eğitiminin Kodsal Görünümü

```
# SVC modeli oluşturma ve eğitme
svc = SVC(C=1, kernel='linear', probability=True,
class_weight='balanced', max_iter=1000)
svc.fit(X_train, y_train)
```

Şekil 5.11: Destek Vektör Makineler Model Performans Ölçümü Kodsal Görünümü

Kesinlik, duyarlılık ve f1-skoru metriklerine göre modelimizin hem dolandırıcılık işlemlerini hem de normal işlemleri tahmin etmede %95'lik bir doğruluk oranına sahip olduğunu görülmektedir. Bu SVM'nin bu tür bir görev için güçlü bir model olduğunu göstermektedir. Ancak modelin performansını daha da iyileştirilmesi için çeşitli yöntemler uygulanabilir. Örneğin daha farklı çekirdek fonksiyonları (RBF, polinom vb.) kullanarak SVM modelinin daha karmaşık veri yapılarıyla daha iyi değerlendirmesini sağlayabilmektedir. Diğer bir alternatif hiperparametre ayarlaması yapmaktır. C ve max_iter gibi SVM parametrelerini ayarlamak modelin genelleme kapasitesini ve sonuçta toplam performansını iyileştirebilmektedir.

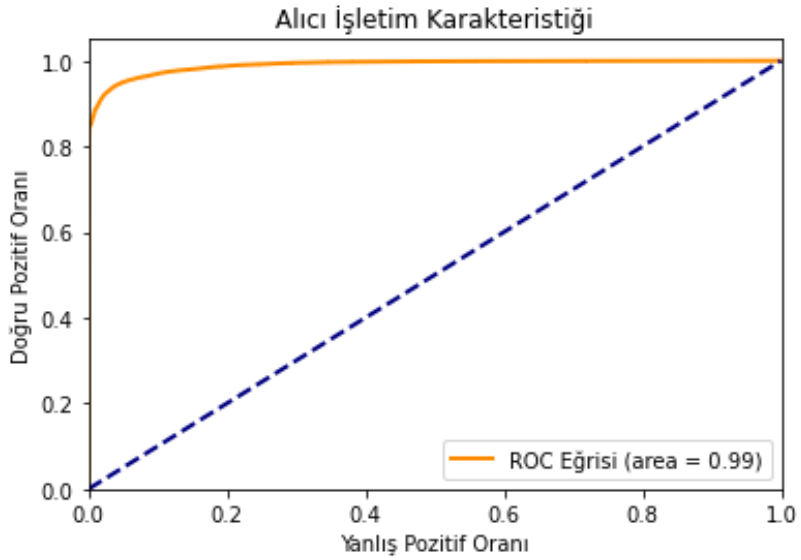
Şekil 5.12'de görüldüğü gibi Destek Vektör Makineleri modelinin ROC AUC skoru 0.989 olarak bulunmuştur. Modelin dolandırıcılık tespit etme konusunda performansını göstermektedir.

Sonuç olarak SVM modelimiz dolandırıcılık tespitinde etkili bir performans sergilemektedir. Model dolandırıcılık ve normal işlemleri ayırt etme konusunda yüksek bir doğruluk oranına sahip olduğunu göstermektedir.

Destek Vektör Makineleri (SVM) Uygulamasının Sonucu:

	Kesinlik	Duyarlılık	f1-skoru	Destek
0	0.93	0.98	0.95	56750
1	0.97	0.92	0.95	56976
Doğruluk			0.95	113726
Makro ort	0.95	0.95	0.95	113726
Ağırlıklı ort	0.95	0.95	0.95	113726

```
ROC AUC Skoru  :0.9900143794991507
Doğruluk       :0.9493871234370329
Kesinlik       :0.9742065696404103
Duyarlılık     :0.923423897781522
F1 Skoru       :0.9481357337225857
```



Şekil 5.12: Destek Vektör Makineleri Modeli için ROC Eğrisi

5.1.3 Karar ağaçları uygulaması

Kredi kartı dolandırıcılığı finans sektöründe önemli bir sorundur ve etkili bir çözüm bulunması hem tüketicilerin hem de bankaların çıkarına olacaktır. Karar ağaçları özellikle anlaşılabilirliği ve yorumlanabilirliği nedeniyle birçok uygulama alanında kullanılmaktadır. Bu analiz bir karar ağacı sınıflandırıcısının, kredi kartı dolandırıcılığını tespit etmek için nasıl uygulanabileceğini göstermektedir.

Bu uygulamada Karar Ağacı Sınıflandırıcı kullanılarak kredi kartı dolandırıcılığı tespit edilmektedir. Karar ağaçları veriyi bir dizi karar kurallarına göre bölerek çalışan sınıflandırma algoritmasıdır. Her bir düğümde bir özellik değerlendirilir ve bir karar verilmektedir. Bu süreç, hedef değişkenin tahmin edildiği bir yaprak düğümüne ulaşılıncaya kadar devam etmektedir (Han, 2011).

Veriyi özellikler ve hedef değişken olarak ayırılmaktadır. Özellikler ölçeklendirilir ve dengesizlik giderilmektedir. Veri seti daha sonra eğitim ve test setlerine ayrılmaktadır.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, roc_curve,
roc_auc_score
from imblearn.over_sampling import SMOTE

# Veri kümesini okuma
df = pd.read_csv("creditcard.csv")

# Özellikler ve etiketleri ayırma
X = df.drop("Class", axis=1)
y = df["Class"]
```

Şekil 5.13: Özelliklere Ayırma ve Ölçeklendirme Kodsal Görünümü

```
# Veri dengesizliği ile başa çıkmak için veri üretme tekniği
(SMOTE) kullanılmıştır.
sm = SMOTE(sampling_strategy='minority')
X_sm, y_sm = sm.fit_resample(X, y)

# Eğitim ve test setlerini ayırma
X_train, X_test, y_train, y_test = train_test_split(X_sm, y_sm,
test_size=0.20, random_state=42)
```

Şekil 5.14: Karar Ağaçları Sentetik Azınlık Yüksek Örneklem Tekniği Kodsal Gösterimi

Veri setindeki sınıf dengesizliği veri üretim tekniği kullanılarak giderilmiştir. Bu teknik, azınlık sınıfını yapay olarak artırarak veri dengesizliğini düzeltmektedir. Daha sonra, veri seti eğitim ve test setlerine bölünmektedir.

Karar ağacı algoritması sklearn kütüphanesinin DecisionTreeClassifier fonksiyonu ile uygulanmaktadır. Algoritmanın iki önemli parametresi vardır: max_depth ve criterion. max_depth parametresi ağacın ne kadar derine inebileceğini belirlerken, criterion parametresi ağaç bölünmelerinin kalitesini değerlendirmek için kullanılan işlevi belirlemektedir. Bu uygulamada, entropy kriteri kullanılmıştır, bu da her bölünmenin bilgi kazancını maksimize etmeye çalışmaktadır.

```
# DecisionTreeClassifier modeli oluşturma ve eğitme
clf = DecisionTreeClassifier(max_depth=6, criterion="entropy")
clf.fit(X_train, y_train)
```

Şekil 5.15: Karar Ağaçları Model Oluşturulmasının ve Eğitiminin Kodsal Görünümü

```
# Modelin performansını değerlendirme
y_pred = clf.predict(X_test)
print(classification_report(y_test, y_pred))
```

Şekil 5.16: Karar Ağaçları Model Performans Ölçümü Kodsal Görünümü

Karar ağacı modeli eğitim verisi üzerinde eğitilmiş ve test verisi üzerinde tahminlerde bulunmaktadır. Sonuçlar, modelin yüksek bir doğrulukla dolandırıcılık durumlarını tespit edebildiğini göstermektedir.

Sonuçları değerlendirdiğimizde Karar Ağacı Sınıflandırıcısı modelinin kredi kartı dolandırıcılık tespitinde oldukça etkili olduğunu görünmektedir. Model hem dolandırıcılık durumlarını hem de dolandırıcılık olmayan durumları yüksek bir doğrulukla tahmin edebilmiştir.

Kesinlik tahmin edilen pozitif durumların gerçekten kaçının pozitif olduğunu ölçmektedir. Kesinlik değeri modelin dolandırıcılık durumlarını tespit ederken ne kadar kesin olduğunu göstermektedir. Bu durumda dolandırıcılık durumlarına ilişkin tahminlerin %98'inin doğru olduğunu görünmektedir. Diğer yandan dolandırıcılık olmayan durumlar için de kesinlik %98'dir. Bu, modelin dolandırıcılık olmayan durumları tespit ederken de oldukça kesin olduğunu göstermektedir.

Duyarlılık gerçek pozitif durumların ne kadarının doğru bir şekilde tespit edildiğini göstermektedir. Bu durumda dolandırıcılık durumlarının %98'si doğru bir şekilde tespit edilmiştir. Dolandırıcılık olmayan durumlar için duyarlılık %98'dir. Bu, modelin dolandırıcılık olmayan durumları tespit etme konusunda oldukça duyarlı olduğunu göstermektedir.

F1-Skoru, Kesinlik ve Duyarlılık'ın harmonik ortalamasıdır ve modelin genel performansını ölçmektedir. Dolandırıcılık ve dolandırıcılık olmayan durumlar için F1-Score'un her ikisi de %98'dir, bu da modelin genel olarak mükemmel bir performansa sahip olduğunu göstermektedir.

Şekil 5.17’de görüldüğü gibi ROC AUC skoru 0.997 olarak elde edilmiş bu da modelin yüksek bir ayrım gücüne sahip olduğunu ve sağlam bir performans sergilediğini göstermektedir.

Sonuç olarak Karar Ağacı Sınıflandırıcısı Kredi kartı dolandırıcılık tespitinde etkili bir araç olduğu sonucuna varılmaktadır. Modelin yüksek doğruluk, kesinlik, duyarlılık ve F1-Score değerleri, bu algoritmanın kredi kartı dolandırıcılık tespitinde güçlü bir aday olduğunu göstermektedir.

Karar Ağacı Sınıflandırıcısı Uygulamasının Sonucu:

	Kesinlik	Duyarlılık	f1-skoru	Destek
0	0.98	0.98	0.98	56750
1	0.98	0.98	0.98	56976
Doğruluk			0.98	113726
Makro ort	0.98	0.98	0.98	113726
Ağırlıklı ort	0.98	0.98	0.98	113726

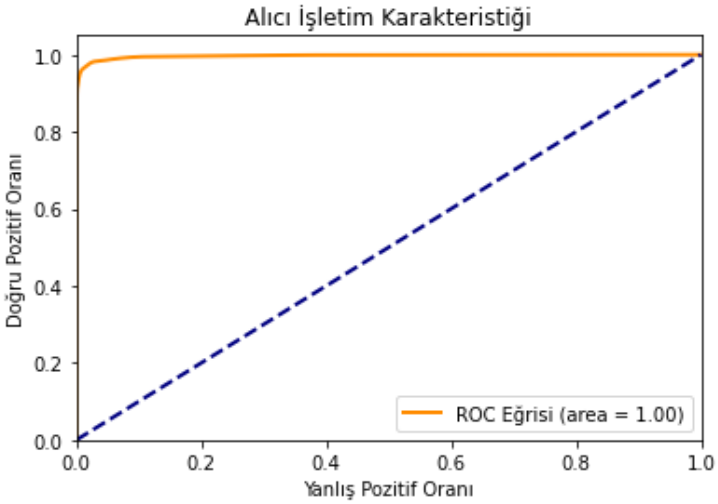
ROC AUC Skoru :0.9969362393254381

Doğruluk :0.9778766508977719

Kesinlik :0.9784916004779645

Duyarlılık :0.9773237854535243

F1 Skoru :0.9779073443152682



Şekil 5.17: Karar Ağaçları Modeli için ROC Eğrisi

5.1.4 Rastgele orman uygulaması

Rastgele Orman birçok karar ağacının birleştirilmesiyle oluşturulan bir topluluk öğrenme algoritmasıdır. Her bir ağaç veri setinin bir alt kümesi üzerinde bağımsız olarak eğitilmekte ve her bir ağacın çıktısı, son tahminin oluşturulması için birleştirilmektedir. Bu yaklaşım, modelin genel hata oranını azaltır ve riskini en aza indirmektedir (Louppe, 2015).

Rastgele Orman algoritması birçok karar ağacının tahminlerini birleştirerek daha doğru ve sağlam bir tahmin elde etme prensibine dayanmaktadır. Her bir ağaç veri setinin bir alt kümesi üzerinde bağımsız olarak eğitilmekte ve her bir ağacın çıktısı, son tahminin oluşturulması için birleştirilmektedir. Bu yaklaşım, modelin genel hata oranını azaltır ve riskini en aza indirmektedir (Louppe, 2015).

Veri setimizi, dolandırıcılık olan ve olmayan işlemleri içeren etiketlere (y) ve bu etiketlere karşılık gelen özelliklere (X) ayrılmaktadır.

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report
from imblearn.over_sampling import SMOTE
from sklearn.metrics import roc_curve, auc, roc_auc_score,
accuracy_score, precision_score, recall_score, f1_score
import matplotlib.pyplot as plt

# Veri kümesini okuma
df = pd.read_csv("creditcard.csv")

# Özellikler ve etiketleri ayırma
X = df.drop("Class", axis=1)
y = df["Class"]
```

Şekil 5.18: Rastgele Orman Özellik ve Etiketlere Ayırma Kodsal Görünümü

Veri üretme tekniği kullanarak azınlık sınıfın (bu durumda dolandırıcılık olan işlemler) yeniden üretilmektedir. Bu teknik azınlık sınıfının örneklerini sentetik olarak artırarak veri dengesizliğini giderir.

```
# Veri dengesizliği ile başa çıkmak için veri üretme tekniği
kullanılmıştır.
sm = SMOTE(sampling_strategy='minority')
X_sm, y_sm = sm.fit_resample(X, y)

# Eğitim ve test setlerini ayırma
X_train, X_test, y_train, y_test = train_test_split(X_sm,
y_sm, test_size=0.20, random_state=42)
```

Şekil 5.19: Rastgele Orman - Sentetik Azınlık Yüksek Örneklem Tekniği Kodsal Gösterimi

Rastgele Orman algoritması, scikit-learn kütüphanesi içinde RandomForestClassifier adlı bir sınıf olarak mevcuttur. Bu sınıf, bir Rastgele Orman modelini tanımlamak ve eğitmek için kullanılır. Bu örnekte, 100 ağaç (veya "tahminci") kullandık ve her bir ağacın maksimum derinliği 6 olarak ayarlandı.

```
# RandomForestClassifier modeli oluşturma ve eğitme
clf = RandomForestClassifier(n_estimators=100, max_depth=6,
criterion="entropy", random_state=42)
clf.fit(X_train, y_train)
```

Şekil 5.20: Rastgele Orman Model Oluşturma ve Eğitme Kodsal Görünümü

Veri setini önce eğitim ve test setlerine ayrılmaktadır. Eğitim seti, modelin öğrenmesi ve kendini ayarlaması için kullanılırken, test seti modelin performansını değerlendirmek için kullanılmaktadır. Model eğitildi ve predict metodu ile test seti üzerinde tahminler yapılmaktadır.

```
# Modelin performansını değerlendirme
y_pred = clf.predict(X_test)
print(classification_report(y_test, y_pred))

# Tahmin olasılıklarını alma
y_pred_proba = clf.predict_proba(X_test)[::,1]
```

Şekil 5.21: Rastgele Orman Model Performans Ölçümü Kodsal Görünümü

Rastgele Orman modeli, dolandırıcılık durumlarını tespit etmede etkili olduğunu göstermektedir. Model, dolandırıcılık durumlarına ilişkin tahminlerin %100'ünün ve dolandırıcılık olmayan durumlar için de tahminlerin %95'inin doğru olduğunu göstermektedir. Dolandırıcılık durumları için yapılan tahminlerin kesinlik ve duyarlılık oranı çok yüksek, bu da modelin bu tür durumları tespit etme yeteneğinin oldukça güçlü olduğunu göstermektedir.

F1 skoru, bir modelin kesinlik ve duyarlılık oranlarının harmonik ortalamasını alır ve bu durumda da modelin başarısını göstermektedir. Dolandırıcılık durumları için F1 skoru %97, dolandırıcılık olmayan durumlar için ise %97'dir. Bu, modelin hem dolandırıcılık durumlarını hem de dolandırıcılık olmayan durumları tespit etme konusunda dengeli bir performans gösterdiği görülmektedir.

Şekil 5.22'de görüldüğü gibi AUC skoru ise %99.8'dir. Bu değer, ROC eğrisinin altında kalan alanın yüzdesini temsil eder ve genellikle 0.5 ile 1.0 arasında bir değer alır. 1.0'e yakın bir AUC skoru, modelin sınıfları mükemmel bir şekilde ayırt ettiğini gösterirken, 0.5 değeri, modelin sınıfları ayırt etme yeteneğinin

tamamen rastgele olduğunu göstermektedir. Burada, AUC skoru %99.8 olarak çok yüksek sonuca ulaşması modelin mükemmel bir sınıflandırma performansına sahip olduğunu göstermektedir.

Sonuç olarak Rastgele Orman algoritması dolandırıcılık tespiti konusunda etkili bir araçtır. Dengesiz veri setleri üzerinde bile yüksek performans gösterir ve genellenebilir sonuçlar sağlamaktadır. Bu çalışmada gösterildiği gibi Rastgele Orman algoritması ile dolandırıcılık tespiti kredi kartı işlemlerinde dolandırıcılık tespit etmek ve önlemek için güçlü bir araç olmaktadır.

Rastgele Orman Uygulamasının Sonucu:

	Kesinlik	Duyarlılık	f1-skoru	Destek
0	0.95	1.00	0.97	56750
1	1.00	0.95	0.97	56976
Doğruluk			0.97	113726
Makro ort	0.97	0.97	0.97	113726
Ağırlıklı ort	0.97	0.97	0.97	113726

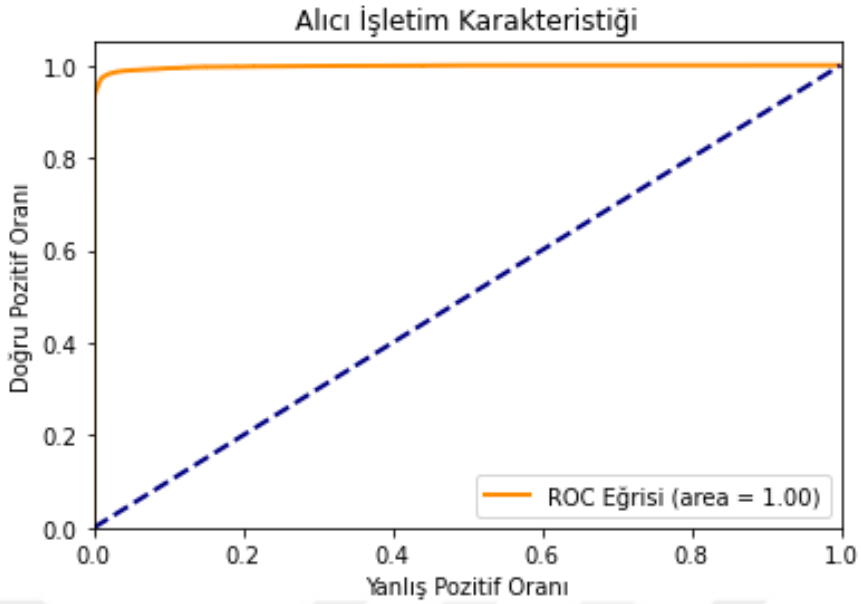
ROC AUC Skoru :0.9975892992118482

Doğruluk :0.972178745405624

Kesinlik :0.9958900070035756

Duyarlılık :0.9483817747823645

F1 Skoru :0.9715554596616143



Şekil 5.22: Rastgele Orman Modeli için ROC Eğrisi

5.1.5 Yapay sinir ağları uygulaması

Yapay Sinir Ağları (ANN), özellikle karmaşık veri yapılarını anlamada ve örüntüleri tanımada başarılı bir makine öğrenmesi algoritmasıdır. İnsan beynindeki biyolojik sinir ağlarının çalışma şeklini taklit eder ve büyük miktarda veriyi işleyebilme yeteneğiyle bilinmektedir. Bu durum, ANN'leri dolandırıcılık tespiti gibi karmaşık problemler için ideal bir araç haline getirmektedir. Bu çalışmada, kredi kartı dolandırıcılık tespit etmek için ANN modeli kullanılmaktadır (Goodfellow, 2016).

Veri setimiz dolandırıcılık olan ve olmayan işlemleri içeren etiketlere (y) ve bu etiketlere karşılık gelen özelliklere (X) ayrılmaktadır. StandardScaler kullanılarak veri seti ölçeklendirilmektedir. Bu modelin eğitim sürecini hızlandırır ve performansını iyileştirir.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from imblearn.over_sampling import SMOTE
from keras.models import Sequential
from keras.layers import Dense
from keras.callbacks import EarlyStopping
from sklearn.metrics import classification_report, roc_curve, auc,
accuracy_score, precision_score, recall_score, f1_score
import matplotlib.pyplot as plt

# veriyi yükle
df = pd.read_csv('creditcard.csv')

# Özellikler ve hedef değişken
X = df.drop('Class', axis=1)
y = df['Class']

# Veriyi ölçeklendir
scaler = StandardScaler()
X = scaler.fit_transform(X)
```

Şekil 5.23: ANN - Özellik ve Etiketlere Ayırma Kodsal Görünümü

Model dolandırıcılık sınıfının azınlıkta olduğu bir veri setinde kullanılmaktadır. Veri setinin dengesiz olması, modelin performansını etkileyebilmektedir. Bu sorunu çözmek için, Sentetik Azınlık Yüksek Örnekleme Tekniği kullanılmaktadır. Sentetik Azınlık Yüksek Örnekleme Tekniği azınlık sınıfından sentetik veriler oluşturarak veri setini dengeler (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

```
# Veri dengesizliği ile başa çıkmak için veri üretme tekniği
kullanılmıştır.
sm = SMOTE(random_state=42)
X_res, y_res = sm.fit_resample(X, y)

# Eğitim ve test setlerini ayır
X_train, X_test, y_train, y_test = train_test_split(X_res,
y_res, test_size=0.2, random_state=42)
```

Şekil 5.24: ANN - Sentetik Azınlık Yüksek Örneklem Tekniği Kodsal Gösterimi

Bu çalışmada kullanılan ANN modeli, üç katmandan oluşmaktadır. Bunlar; giriş katmanı, gizli katman ve çıkış katmanıdır. Model, binary_crossentropy loss fonksiyonunu ve adam optimizer'ını kullanarak derlenmektedir. Modelin eğitim süreci boyunca, modelin doğruluk metriği takip edilmektedir (Nath & Rokonuzzaman, 2018).

```
# Modelin ve katmanların oluşturulması
model = Sequential()
model.add(Dense(30,                 input_dim=X_train.shape[1],
activation='relu'))  # Giriş katmanı
model.add(Dense(50, activation='relu'))  # Gizli katman
model.add(Dense(1, activation='sigmoid'))  # Çıkış katmanı

# Modelin derlenmesi
model.compile(loss='binary_crossentropy',     optimizer='adam',
metrics=['accuracy'])
```

Şekil 5.25: ANN – Modelin Katmanlarının Oluşturulması ve Derlenmesi Kodsal Görünümü

Model, eğitim veri seti üzerinde 100 epoch boyunca eğitilmiştir. Her bir epoch'ta, modelin loss ve doğruluk değerleri hesaplanmış ve veri setindeki performansı değerlendirilmektedir. Erken durdurma mekanizması modelin aşırı uyumu (overfitting) önlemek için kullanılmaktadır.

```
# Modelin eğitilmesi
es     =     EarlyStopping(monitor='val_loss',     mode='min',
verbose=1, patience=3)  # Erken durma
history    =    model.fit(X_train,    y_train,    epochs=100,
batch_size=1000,     verbose=1,     validation_data=(X_test,
y_test), callbacks=[es])
```

Şekil 5.26: ANN – Modelin Eğitilmesi Kodsal Görünümü

Modelin performansı test veri seti üzerinde yapılan tahminler kullanılarak değerlendirilmektedir.

```
# Tahminlerde bulun
y_pred = model.predict(X_test)
y_pred_binary = (y_pred > 0.5).astype("int32")
```

Şekil 5.27: ANN - Model Performans Ölçümü Kodsal Görünümü

Model dolandırıcılık durumlarını %100 doğrulukla tespit etmiştir. Bu modelin bu tür durumları tespit etme yeteneğinin, kullanılan özelliklerin ve algoritmanın doğru bir şekilde ayarlanmasına bağlı olduğunu göstermektedir.

Şekli 5.28'de görüldüğü gibi ANN (Artificial Neural Network) modeli, mükemmel bir performans sergileyerek ROC AUC skorunda 0.9999 gibi yüksek bir değer elde etmiştir. Bu, modelin pozitif ve negatif sınıfları ayırt etme yeteneğinin son derece iyi olduğunu, dolayısıyla sahtecilik tespiti konusunda güvenilir ve etkili bir araç olduğunu göstermektedir.

Ancak tüm modellerde olduğu gibi, bu modelin de hatalar yapabileceğini unutmamak önemlidir. Özellikle, model hala yanlış pozitifler (dolandırıcılık olmayan işlemlerin dolandırıcılık olarak sınıflandırılması) ve yanlış negatifler (dolandırıcılık işlemlerinin dolandırıcılık olmayan olarak sınıflandırılması) üretebilmektedir. Bu nedenle, modelin uygulamada kullanılması durumunda, bu potansiyel hataların farkında olmak ve onları minimize etmek için uygun önlemler almak önemlidir.

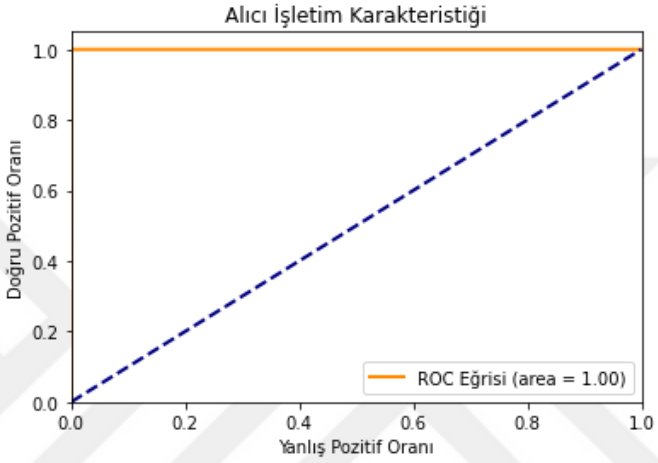
Sonuç olarak Yapay Sinir Ağları, kredi kartı dolandırıcılığı gibi karmaşık problemlerin çözümünde oldukça etkili bir araçtır. Ancak modelin etkinliği kullanılan veri setinin kalitesine, modelin doğru şekilde ayarlanmasına ve eğitilmesine bağlıdır. Bu çalışmada modelin performansı, doğru veri ön işleme ve dengeli bir veri seti ile önemli ölçüde iyileştirilmiştir.

Yapay Sinir Ağları Uygulamasının Sonucu:

	Kesinlik	Duyarlılık	f1-skoru	Destek
0	1.00	1.00	1.00	56750
1	1.00	1.00	1.00	56976
Doğruluk			1.00	113726
Makro ort	1.00	1.00	1.00	113726

```
Ağırlıklı ort      1.00      1.00      1.00    113726
```

```
ROC AUC Skoru  :  0.9999376494562359
Doğruluk       :  0.9994196577739479
Kesinlik       :  0.9988604488078542
Duyarlılık     :  0.999982448750351
F1 Skoru       :  0.9994211338759472
```



Şekil 5.28: ANN Modeli için ROC Eğrisi

6. BULGULAR

Bu çalışma, kredi kartı dolandırıcılık tespiti için çeşitli makine öğrenme ve derin öğrenme algoritmalarını karşılaştırmaktadır. Bu algoritmaların performansı gerçek finansal verileri olan bir kredi kartı dolandırıcılık veri seti üzerinde karşılaştırılmıştır. Lojistik Regresyon, Destek Vektör Makineleri (SVM), Karar Ağaçları, Rastgele Orman ve Yapay Sinir Ağları (ANN) gibi çeşitli algoritmaların performansını ve etkinliklerini incelenmiştir. Bu algoritmaların hem benzerliklerini hem de farklılıkları aşağıda belirtilmiştir.

6.1 Algoritmaların Benzerlikleri

Lojistik Regresyon, Destek Vektör Makineleri, Karar Ağaçları ve Random Forest, her biri bir denetimli öğrenme algoritmasıdır ve sınıflandırma problemlerinde yaygın olarak kullanılmaktadır. Her biri özellikler ve hedef sınıflar arasındaki ilişkiyi anlamak ve yeni örneklere genellemek için veri setinden öğrenmektedir. Her bir algoritmanın amacı modelin sahtekarlık işlemlerini gerçek işlemlerden ayırt etme yeteneğini en üst düzeye çıkarmaktır.

6.2 Algoritmaların Farklılıkları

Algoritmaların genel amacı benzer olsa da kullanılan yöntemler ve performans açısından önemli farklılıklar bulunmaktadır:

Lojistik Regresyon, veri noktalarını iki sınıfa ayıran doğrusal bir sınırlayıcı kullanarak sınıflandırma yapmaktadır. Model karmaşıklığı düşük olduğu için daha hızlı eğitilir ve daha az hesaplama gücü gerektirmektedir. Bununla birlikte doğrusal sınırlayıcılar karmaşık ve doğrusal olmayan ilişkileri modellemekte zorlanabilmektedir.

Destek Vektör Makineleri veri noktalarını iki sınıfa ayıran en iyi ayrım hiperdüzlemi (sınırlayıcı) bulmaya çalışmaktadır. Destek Vektör Makineleri doğrusal

ve doğrusal olmayan sınırlayıcılar kullanabilir, ancak eğitim süresi ve hesaplama gücü daha fazladır.

Karar Ağaçları özellikler üzerinde ardışık kararlar alarak veri noktalarını sınıflandırmaktadır. Karar ağaçları verilerin doğrusal olmayan ilişkilerini yakalayabilmektedir. Ancak aşırı uyum (overfitting) eğilimindedir.

Rastgele Orman birden fazla karar ağacını bir araya getirerek daha güçlü ve daha genelleştirilebilir bir model oluşturmaktadır. Bu sayede, aşırı uyum riski azalır ve performans artmaktadır.

Yapay Sinir Ağları (ANN), biyolojik sinir sistemlerinden esinlenerek, karmaşık ve doğrusal olmayan ilişkileri modelleme yeteneğine sahiptir. Yapay Sinir Ağları en yüksek doğruluk oranına sahip olsa da eğitim süresi ve hesaplama gücü açısından daha pahalıdır ve açıklanabilirlik eksikliği nedeniyle daha zor anlaşılabilmektedir.

6.3 Karşılaştırmalı Analiz Sonucu

Çeşitli makine öğrenme ve derin öğrenme algoritmalarının kredi kartı dolandırıcılık tespiti için karşılaştırmalı bir analizini sunmaktadır. Bu algoritmalar arasında Lojistik Regresyon, Destek Vektör Makineleri (SVM), Karar Ağaçları, Rastgele Orman ve Yapay Sinir Ağları (ANN) bulunmaktadır.

Lojistik Regresyon bağımlı değişkenin iki sonuçlu olduğu durumlarda kullanılan bir regresyon modelidir. İstatistiksel bir yaklaşım sunar ve sonuçları olasılıklar olarak sağlamaktadır. Modellerin yorumlanabilirliği ve hesaplama maliyetinin düşük olması gibi avantajları bulunmaktadır. Ancak karmaşık ve doğrusal olmayan veri kümeleri ile etkin bir şekilde başa çıkma konusunda sınırlıdır.

Destek Vektör Makineleri (SVM) doğrusal ve doğrusal olmayan sınıflandırma problemlerine uygulanabilen güçlü bir algoritmadır. Destek Vektör Makineleri veri noktalarını bir hiperdüzlem ile ayrılarak çalışır ve en büyük marjı bulmayı hedeflemektedir. Destek Vektör Makineleri yüksek boyutlu verilerle iyi çalışabilmesi ve karmaşık karar sınırları çizebilmesiyle bilinir fakat model eğitimi zaman alıcı olabilir ve büyük veri kümeleriyle ölçeklendirme konusunda zorluklar yaşayabilmektedir.

Karar Ağaçları bir dizi karar kuralına dayalı olarak verileri sınıflandırır veya tahminlerde bulunmaktadır. Veri özelliklerini ve hedef değişken arasındaki ilişkileri açık bir şekilde modellediği için son derece açıklanabilir bir modeldir. Ancak, karar ağaçları genellikle aşırı öğrenme eğilimindedir ve gürültülü verilere karşı duyarlıdır.

Rastgele Orman birden çok karar ağacını bir araya getirerek daha sağlam bir model oluşturan bir topluluk öğrenme tekniğidir. Her bir ağacın ayrı ayrı eğitilmesi ve sonuçların birleştirilmesi modelin aşırı öğrenmeyi önlemesine ve genel performansını iyileştirmesine yardımcı olmaktadır. Ancak Rastgele Orman algoritması karar ağaçlarına kıyasla daha karmaşıktır ve daha fazla hesaplama kaynağı gerektirmektedir.

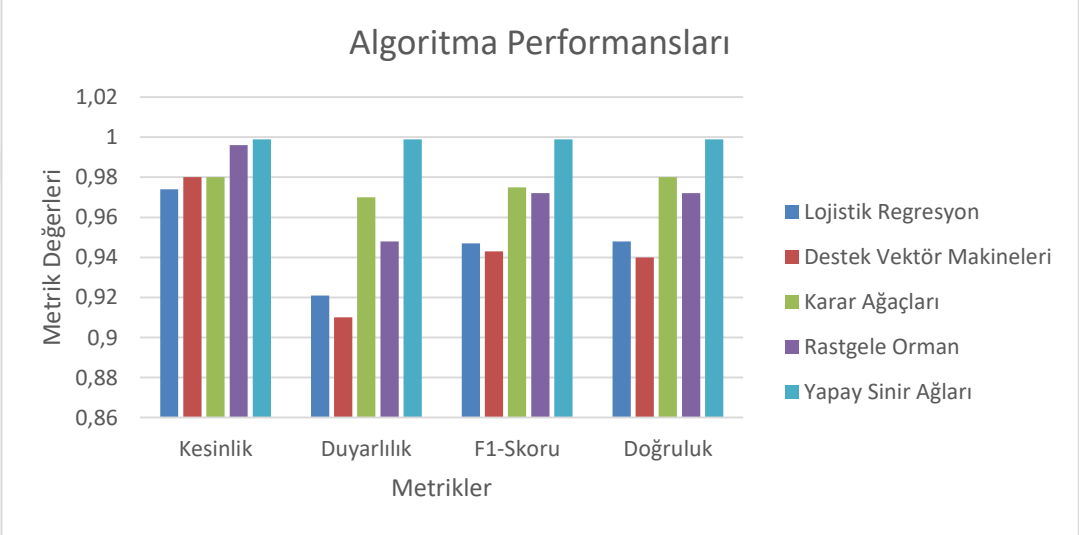
Çizelge 6.1 ve Şekil 6.1'de görüldüğü gibi Yapay Sinir Ağları (ANN) ise tüm algoritmalar arasında en yüksek doğruluk oranını sağlamış olup bu da derin öğrenme tekniklerinin karmaşık veri kümelerini modelleme ve sınıflandırma yeteneğini göstermektedir. Ancak, Yapay Sinir Ağları mükemmel performansına rağmen, eğitim süresi, hesaplama gücü ve modelin açıklanabilirliği gibi faktörler de dikkate alınmalıdır. Yapay Sinir Ağları genellikle yoğun hesaplama kaynakları gerektirir ve kapalı kutu olarak kabul edilirler yani çıktılarının nasıl elde edildiğini tam olarak açıklamak zordur. Bu finansal düzenleyiciler ve kurumlar tarafından sıklıkla istenen bir özellik olan 'açıklanabilirlik' noktasında bir zorluk oluşturabilmektedir.

Şekil 6.2'de görüldüğü gibi tüm modellerin AUC değerlerinin 1'e oldukça yakın olması, bu modellerin kredi kartı dolandırıcılık tespitinde yüksek bir yetenek ve performansa sahip olduğunu göstermektedir. AUC değeri, bir sınıflandırma modelinin yanlış pozitifler ile doğru pozitifleri dengeleme yeteneğini ölçer, bu yüzden yüksek bir AUC skoru, modelin hem hassasiyeti hem de hatırlamayı iyi bir şekilde yönettiğini göstermektedir.

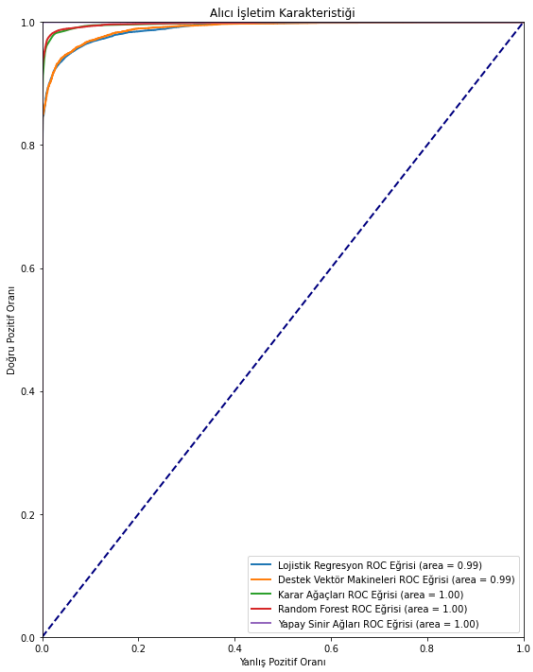
Sonuç olarak her modelin kendi güçlü ve zayıf yönleri bulunmaktadır ve en uygun model seçimi uygulamanın gereksinimlerine, kullanılabilir kaynaklara, veri setinin özelliklerine ve spesifik hedeflere bağlıdır. Bu nedenle model seçimi süreci ayrıntılı bir analiz ve değerlendirme gerektirir ve seçilen modelin performansını sürekli olarak izlemek ve gerekli düzenlemeleri yapmak önem arz etmektedir.

Çizelge 6.1: Algoritma Perfomansları

Yöntem	Kesinlik	Duyarlılık	F1-Skoru	Doğruluk	AUC Skoru
Lojistik Regresyon	0.974	0.921	0.947	0.948	0.989
Destek Vektör Makineleri	0.973	0.919	0.945	0.947	0.989
Karar Ağaçları	0.980	0.970	0.975	0.980	0.997
Rastgele Orman	0.996	0.948	0.972	0.972	0.998
Yapay Sinir Ağları	0.999	0.999	0.999	0.999	0.999



Şekil 6.1: Algoritma Performansları



Şekil 6.2: Algoritmaların ROC Eğrisi

7. SONUÇ VE ÖNERİLER

Bu çalışmada kredi kartı dolandırıcılık işlemlerini tespit edebilmek için makine ve derin öğrenme tekniklerinin etkinliğini değerlendirmiştir. Uygulanan yöntemler arasında Lojistik Regresyon, Destek Vektör Makineleri, Karar Ağaçları, Rastgele Orman ve Yapay Sinir Ağları bulunmaktadır.

Lojistik Regresyon ve SVM modelleri, 0.95 doğruluk oranı elde edilmiştir. Ayrıca bu iki modelin ROC AUC skorları sırasıyla 0.989 ve 0.990 olmuştur ki bu da modellerin dengeli bir performans sergilediğini göstermektedir. Fakat bu iki model Kesinlik ve Duyarlılık değerleri açısından bir miktar farklılık göstermektedir. Lojistik regresyon modeli, 0.974 Kesinlik ve 0.921 Duyarlılık değeri elde ederken, Destek Vektör Makineleri modeli, 0.974 Kesinlik ve 0.923 Duyarlılık değeri elde etmiştir. Bu sonuç Destek Vektör Makineleri modelinin biraz daha dengeli bir performans gösterdiğini göstermektedir.

Karar Ağaçları ve Rastgele Orman modelleri daha yüksek bir doğruluk oranı sağlamıştır. Karar Ağaçları modeli %98 doğruluk oranı ile önemli bir performans sergilemiştir. ROC AUC skoru 0.997, Kesinlik değeri 0.978 ve Duyarlılık değeri 0.977 olan Karar Ağaçları, dengeli ve etkileyici sonuçlar vermiştir. Rastgele Orman modeli ise %97 doğruluk oranı, 0.998 ROC AUC skoru, 0.996 Kesinlik değeri ve 0.948 Duyarlılık değeri ile oldukça iyi sonuçlar elde etmiştir.

Yapay Sinir Ağları (ANN) modeli, %99.94 doğruluk oranı ile yüksek bir performans sergilemiştir. ROC AUC skoru 0.9999 olan bu model, 0.9989 Kesinlik ve 0.9999 Duyarlılık değerleri ile tüm test verilerini doğru bir şekilde sınıflandırabilmiştir.

Buna rağmen, her modelin avantajları ve dezavantajları vardır ve en uygun model seçimi, uygulamanın gereksinimlerine ve kısıtlamalarına bağlıdır. Örneğin eğitim süresi, modelin karmaşıklığı, anlaşılabilirlik ve açıklanabilirlik gibi faktörler de göz önünde bulundurulmalıdır. Bu ayrıntılı inceleme, bankacılık ve ödeme

sistemleri üzerindeki dolandırıcılık türlerinin çeşitliliğini ve karmaşıklığını göstermektedir.

Bu çalışmada Yapay Sinir Ağları (ANN), Karar Ağaçları ve Rastgele Orman olmak üzere en yüksek performansa sahip üç algoritmayı ayrıca incelediğimizde üç temel nokta öne çıkmaktadır. İlk olarak özellik önem sıralaması açısından incelediğimizde Karar Ağaçları ve Rastgele Orman algoritmalarının veri setindeki hangi özelliklerin modelin karar vermesi için belirleyici özellik olduğunu belirleyebilmek ve önem sıralaması sağlamak karar verilmesi noktasında etkili bir araç olarak öne çıkmaktadır (Breiman, 2001) (Liaw, 2002). Veri setindeki kritik faktörlerin belirlenmesinde önemli katkılar sağlamaktadır. Ayrıca, hızlı eğitim ve tahmin yetenekleri sayesinde anlık işlemleri hızlı bir şekilde değerlendirmek için ideal bir yapı sunmaktadırlar (Awwalu, Ghazvini, & Bakar, 2014). Bu nedenle özellikle hızlı karar verme gerekliliği olan durumlarda Rastgele Orman ve Karar Ağaçları algoritmaları tercih edilebilmektedir. Bu modeller ayrıca anlaşılabilirlik ve açıklanabilirlik açısından Yapay Sinir Ağlarına göre karar mekanizmalarını daha şeffaf bir şekilde gösterirler (Adadi & Berrada, 2019). Bu çalışma, kredi kartı dolandırıcılığı tespiti alanında daha etkili ve özelleştirilmiş yaklaşımların geliştirilmesine yönelik önemli bir adım olarak değerlendirilebilir.

Sonuç olarak, bu çalışma kredi kartı dolandırıcılığı tespiti alanında diğer çalışmalardan farklı bir yaklaşım sunarak çeşitli makine öğrenme ve derin öğrenme modellerinin etkinliğini değerlendirmektedir. Farklı algoritmaların dolandırıcılık tespiti konusunda farklı performanslar sergilediği ve her bir algoritmanın özellikleri, veri kümesinin özellikleri ve performans gereksinimleri gibi faktörler dikkate alınarak seçilmesi gerektiği görülmektedir. Yapay Sinir Ağlarının yanında özellikle Karar Ağaçları ve Rastgele Orman modelleri yüksek doğruluk oranları ve dengeli sonuçlar elde edilmiştir. Bu farklı perspektif, dolandırıcılık tespiti konusundaki daha etkili ve özelleştirilmiş yaklaşımların geliştirilmesine yol açabilir. Bu nedenle farklı algoritmaların bir arada kullanıldığı ve birbirleriyle karşılaştırıldığı yeni çalışmaların yapılması önerilmektedir. Gelecekteki araştırmalar, bu algoritmaların daha da iyileştirilmesi ve özelleştirilmesi üzerinde durabilir. Böylece dolandırıcılık tespitinde daha etkili ve etkin bir yaklaşım sağlanabilir. Bu çalışma hem finansal kurumlar için maliyet tasarrufu sağlar hem de tüketicilere daha güvenli bir alışveriş deneyimi sunabilir.

KAYNAKLAR

Aaron Hertzmann, D. J. (2015). Machine Learning and Data Mining. *Machine Learning and Data Mining* (s. 115-117).

Adadi, A., & Berrada, M. (2019). Peeking Inside the Black-Box: A Survey on. *IEEE Access*, 52138-52160.

Adeel, M., & Hussain, S. (2017). ATM Skimming and Its Effects on Banking Industry: A Case Study of Pakistan. *International Journal of Academic Research in Accounting, Finance, and Management Sciences*, 2-5.

Ahmad, A., & Saini, D. (2019). Machine Learning Based Approach for Credit Card Fraud Detection. *International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (s. 1-6). IEEE.

Ahmad, S., & Raza, S. (2021). A Comprehensive Review on Ensemble Learning in Fraud Detection. *International Conference on Engineering Sciences and Management* (s. 623-628). Singapore: Springer.

AICPA. (2021). *Fraud Risk Frameworks.* Forensic and Litigation Services (FLS) Fraud Task Force.

Arnone, M. (2008). Anti-money laundering by international institutions: A preliminary assessment. *European Journal of Law and Economics*, s. 26.

Association of Certified Fraud Examiners. (2020). *Report to the Nations.* Association of Certified Fraud Examiners.

Awwalu, J., Ghazvini, A., & Bakar, A. A. (2014). Performance Comparison of Various Data Mining Algorithms: A Review. *International Journal of Computer Trends and Technology*, 78-82.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 281-305.

Bhattacharya, A., Das, S., & Koner, S. (2019). Credit Card Fraud Detection Using Machine Learning: A Comparative Study. *Journal of Data Science and Engineering*, 4-18.

Bhattacharyya, S. J. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, s. 602-613.

Bolton, R. J. (2002). Statistical fraud detection: A review. *Statistical Science*, s. 235-249.

Breiman, L. (2001). Random forests. *Machine Learning*, 5-32.

Chatfield, C. (2003). The Analysis of Time Series: An Introduction. *Chapman and Hall*, 15.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 321-357.

Chen, R. C. (2005). Detecting credit card fraud by decision trees and support vector machines. *International Conference on Cyberworlds*, s. 83-88.

Chen, X., & Guestrin, C. (2016). A Scalable Tree Boosting System. *International Conference on Knowledge Discovery and Data Mining* (s. 785-794). IEEE.

Cheng, F. F., & Liu, L. (2019). *Fraudulent activities in mobile payment systems.* 72-92: A comprehensive survey.

Crossler, R. E. (2013). Future directions for behavioral information security research. *Computers & Security*, s. 91-92.

D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Logistic Regression, Applied logistic regression s. 31-33. London: Wiley.

Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating Probability with Undersampling for Unbalanced Classification. *Computational Intelligence and Data Mining*, 3-5.

Davis, J., & Goadrich, M. (2006). The Relationship Between Precision-Recall and ROC Curves. *In Proceedings of the 23rd International Conference on Machine Learning* (s. 233-240.). ICML.

Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 1-30.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Learning Representations*, 2-5.

Dunn, J. E. (2018). The Role of Phishing and Malware Attacks in EFT and Wire Fraud. *Dark Reading*, 7-10.

Euromonitor International. (2023, 02 15). *www.fico.com*. FICO: https://www.fico.com/europeanfraud/turkey adresinden alındı

Flach, P. (2015). Precision-Recall-Gain Curves: PR Analysis Done Right. *In Advances in Neural Information Processing Systems*, 838-846.

Friedman, J., Hastie, T., & Tibshirani, R. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer Science & Business Media*, 10-15.

Goodfellow, I. B. (2016). Deep Learning. *MIT Press* (s. 109-110).

Gülpınar, N., Özlük, Ö., & Öztaş, N. (2015). A graph-based approach for identifying organized frauds in financial crimes. *Expert Systems with Applications*, s. 6-18.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, s. 1157-1182.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., & Yuanyue, H. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, s. 220-239.

Han, J. P. (2011). *Data Mining: Concepts and Techniques (3rd ed.)* s. 336,337-341-342. Morgan Kaufmann.

Hussaini, U., Bakar, A. A., & Yusuf, M.-B. O. (2019). he effect of fraud risk management, risk culture and performance of banking sector: A conceptual framework. *International Journal of Multidisciplinary Research and Development*, s. 71-80.

Hyndman, R. J. (2018). Forecasting: principles and practice. *OTexts*, s. 15-17.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, s. 429-449.

Jin, M., Gao, L., & Liu, Y. (2021). Multi-factor Authentication for Financial Applications: A Survey. *Journal of Network and Computer Applications*, s. 187-189.

Johnson, R. E., & Smith, M. A. (2019). Fraud Prevention Strategies: A Comprehensive Review. *International Journal of Risk and Contingency Management*, s. 23-37.

Kaur, K., & Rani, R. (2021). Behavioral Analytics in Fraud Detection: A Review. *Procedia Computer Science*, s. 467-474.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 436-444.

Liaw, A., & Wiener, M. (2002). Classification and regression. *R news*, 18-22.

Louppe, G. (2015). Understanding Random Forests: From Theory to Practice.

McNally, G. R. (2007). Identity theft: A Research Review. *National Institute of Justice*, s. pp. 1-6.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *Learning Representations*, 3-7.

Nadeem, A., & Zawoad, S. (2021). Data Encryption Techniques in Financial Sector. *International Journal of Computer Science and Information Security*, 19-20.

Nath, B. D., & Rokonuzzaman, M. (2018). Artificial Neural Network Based Pullout Capacity Prediction. *Proceedings of the 4th International Conference on Civil Engineering for Sustainable Development* (s. 2-10).

Bangladesh: International Conference on Civil Engineering for Sustainable Development.

Nurse, J. R. (2011). Guidelines for usable cybersecurity: Past and present. *Cyberspace Safety and Security*, s. 21-26.

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A Comprehensive Survey of Data Mining-based Fraud Detection Research. *The Computer Journal*, 2-6.

Powell, S. E. (2008). Check Fraud: A Guide to Avoiding Losses. *The Journal of Corporate Accounting & Finance*, 6-19.

Powers, D. M. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 37-63.

Raschka, S., & Mirjalili', V. (2020). *Python Machine Learning, 2nd Edition.* Packt Publishing: Birmingham.

Rejeb, A. B., & Abdul-Rahman, A. (2021). The Importance of Financial Education in Preventing Fraud. *International Journal of Economics, Commerce, and Management*, 1-7.

Ribeiro, A. H. (2019). A survey on ensemble learning for data stream classification. *ACM Computing Surveys (CSUR)*, s. 236-239.

Sahin, C., & Duman, E. (2011). Detecting credit card fraud by genetic algorithm and scatter search. Expert Systems with Applications. Expert Systems with Applications, s. 65-67.

Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE*, 10.

Santos, A., & Maynard, S. (2018). Financial Fraud Prevention and Anti-Money Laundering Compliance: The Role of Artificial Intelligence. *European Business Organization Law Review*, 29-56.

Shen, J., & Hsiao, K. F. (2021). Fraud Detection and Prevention with Big Data Analytics and Machine Learning. *Information & Management*, 58.

Sherman, R. (2000). The check is not in the mail. *Security Management*, s. 52-58.

Singh, A., & Patel, M. (2020). Real-time Fraud Detection and Prevention using Machine Learning Techniques. *International Journal of Recent Technology and Engineering*, 83*87.

Smith, J., & Johnson, R. (2020). The Role of Data Analysis and Artificial Intelligence in Fraud Detection. *International Journal of Finance and Banking Studies*, 32-46.

Su, X., Khoshgoftaar, T., & Zhu, X. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 4.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *PeerJ Preprints*, 6.

Tharaka, K. S. L. (2018). An Overview of Social Engineering in the Context of Information Security. *International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, s. 3-4.

Udjianto, A. W. (2006). Statistical methods for credit card fraud detection. *Section on Quality and Productivity*, 34-39.

UK - Finans. (2021). *Fraud - The Facts 2021*. London: UK - Finans.

Ulb, M. L. (2022, 10 10). *kaggle.com*. kaggle: ttps://www.kaggle.com/datasets/mlg-ulb/creditcardfraud adresinden alındı

Wu, G., Zhu, X., & Ding, W. (2009). Data Mining with Big Data. *Transactions on Knowledge and Data Engineering*, 97-107.

Yero, S. I. (2018). Types and Patterns of Fraud in the Banking Industry. *International Journal of Scientific and Research Publications*, 8-10.

ÖZGEÇMİŞ

ÖĞRENİM DURUMU

- Lisans : 2017, Düzce Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği
- Yüksek Lisans : 2023, (Devam) İstanbul Gedik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Yapay Zekâ Mühendisliği Anabilim Dalı, Yapay Zekâ Mühendisliği Tezli Yüksek Lisans Programı

MESLEKİ DENEYİM

- **2022** Bankalararası Kart Merkezi AŞ. Sistem Analistliği
- **2020** Vakıf Katılım Bankası AŞ. Sistem Analistliği
- **2019** Etiya Bilgi Teknolojileri AŞ. Sistem Analistliği