T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DENETİMSİZ ÖĞRENME İLE KREDİ KARTI KOPYALAMA NOKTALARININ TESPİT EDİLMESİ

Fuat ÖGME

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Doç. Dr. M. Elif KARSLIGİL

Temmuz, 2020

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DENETİMSİZ ÖĞRENME İLE KREDİ KARTI KOPYALAMA NOKTALARININ TESPİT EDİLMESİ

Fuat ÖGME tarafından hazırlanan tez çalışması 21.07.2020 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. M. Elif KARSLIGİL
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. M. Elif KARSLIGİL, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Banu DİRİ, Üye
Yıldız Teknik Üniversitesi _____

Prof. Dr. Çiğdem EROĞLU ERDEM, Üye
Marmara Üniversitesi _____

Danışmanım Doç. Dr. M. Elif KARSLIGİL sorumluluğunda tarafımca hazırlanan Denetimsiz Öğrenme ile Kredi Kartı Kopyalama Noktalarının Tespit Edilmesi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Fuat ÖGME

İmza

Destek Alındıysa buraya

Aileme
ve sevdiklerime

TEŞEKKÜR

Tez çalışmamda yardımlarını esirgemeyen değerli hocam Doç. Dr. Mine Elif Karslıgil, hayatımın her anında bana yardımcı olan anne ve babama, varlıklarıyla teşekkür ederim.

Fuat ÖGME

İÇİNDEKİLER

SİMGE LİSTESİ

σ	Aktivasyon Fonksiyonu
χ^2	ki-kare istatistiği

KISALTMA LİSTESİ

BKM	Bankalararası Kart Merkezi
PCA	Principle Component Analysis
LLE	Locally Linear Embedding
ICA	Independent Component Analysis
SQL	Structured Query Language
MCC	Mathew Correlation Coefficient
ANN	Artificial Neural Network
RNN	Recurrent Neural Network
LSTM	Long-Short Term Memory
GRU	Gated Recurrent Unit
RBM	Restricted Boltzman Machine
UCI	University of California, Irvine
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
SGD	Stochastic Gradient Descent
ADAM	Adaptive Moment Estimation

ŞEKİL LİSTESİ

TABLO LİSTESİ

ÖZET

Denetimsiz Öğrenme ile Kredi Kartı Kopyalama Noktalarının Tespit Edilmesi

Fuat ÖGME

Bilgisayar Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Doç. Dr. M. Elif KARSLIGİL

Kredi kartı kullanımının yaygınlaşması ile birlikte, kredi kartı sahteciliği de zamanla artış göstermiştir. Kart dolandırıcıları, kart bilgilerini ele geçirmek için çeşitli yöntemler geliştirmektedir. Bu yöntemlerden birisi de, kullanıcılara ait kredi kartlarının, dolandırıcılar tarafından ATM'lerde ve POS cihazlarında oluşturulan düzeneklerle kopyalanarak kullanılmasıdır. Bu çalışmada, sahtekarlığa karıştığı tespit edilmiş kredi kartlarının kopyalandıkları iş yerlerini tespit etmeye yönelik bir yöntem önerilmiştir. Bunun için önce kredi kartı sahtekarlıklarında kopyalama vakaları incelenerek kart kopyalama problemi incelenmiştir. Daha sonra kart bilgilerinin suçlular tarafından ele geçirildği kart kopyalama noktalarını tespit edilmesini sağlayarak, kart kopyalama kaynaklı kredi kartı sahtekarlıklarını azaltmayı amaçlayan yeni bir yöntem önerilmiştir. Önerilen yöntem, kopyalanma sonucunda gerçekleşen benzer sahte işlemlerin kümeleme yöntemi ile gruplanması ve oluşturulan gruplardaki ortak özelliklere sahip kredi kartlarının geçmiş işlemlerinin incelenerek kopyalandıkları olası iş yerlerinin tespit edilmesini hedeflemektedir. Çalışmada 2016-2018 yılları arasında gerçekleşen ve sahte olduğu beyan edilmiş gerçek işlemler ve bankalar tarafından kopyalama merkezi olduğu tespit edilmiş iş yerlerine ait veriler kullanılmıştır. Önce veri kümesi üzerinde ön işlemler gerçekleştirilmiştir. Daha sonra karakteristik özelliklerin elde edilebilmesi için farklı özellik çıkarma yöntemleri ile testler yapılmış ve Özkodlayıcı (AutoEncoder) yöntemi ile özellik çıkarma işlemi gerçekleştirilmiştir. Sahte işlemler benzerliklerine göre K-ortalama kümeleme yöntemiyle gruplanmıştır. Kümeleme sonucunda oluşan her bir küme için, kredi kartlarının geçmiş işlemleri analiz edilmiştir. Analiz aşamasında kredi kartlarının

geçmişte kullanıldıkları iş yerleri, iş yerlerinde ortak kullanılan kart sayısına ve iş yeri skorlarına göre derecelendirilerek bu iş yerlerinden kopyalama merkezi adayları seçilmiştir. Yöntemin başarısı bir banka tarafından sağlanan, gerçek kart kopyalama noktaları ile değerlendirilmiştir.

Anahtar Kelimeler: Kart Kopyalama Noktası Tespiti, Özkodlayıcı, Derin Öğrenme

ABSTRACT

Point of Compromise Detection with Unsupervised Learning

Fuat ÖGME

Department of Computer Engineering
Master of Science Thesis

Advisor: Assoc. Prof. M. Elif KARSLIGİL

With the widespread use of credit cards, credit card fraud has also increased over time. Credit card fraudsters develop various methods to capture card information. One of these methods is to use the credit cards of the users by copying them with the mechanisms created by the fraudsters in ATMs and POS devices. In this study, a method is proposed to identify the merchants where credit cards, which are found to be involved in fraud, are copied. For this; initially, copying cases in credit card frauds were examined and the problem of card copying was examined. Then, a new method is proposed, which aims to reduce credit card frauds, by identifying merchants where the card copying operation took place and card information has been captured by criminals. The proposed method aims to group similar transactions that are reported as card copy frauds with the clustering method and to identify possible merchants that are involved in this scheme by finding matching merchants in the created groups with the examination of previous transactions of the credit cards. In the study, real transactions that have been reported as fraudulent between 2016-2018 and the data of the merchants marked as copying centers by banks were used. Research began with acquisition and preprocessing of the data. Later, in order to obtain the characteristic properties, tests were performed with the usage of different feature extraction methods and feature extraction has decided to be carried out by AutoEncoder method. Fraudulent transactions then, are grouped by K-mean clustering method according to their similarities. For each cluster formed as a result of clustering, the past transactions of credit cards has been analyzed. In the analysis stage, the merchants where credit cards were used in the past were graded in accordance with

the number of fradulent common cards used in the merchants. From this merchants scores, candidates for copy centers has been selected. The accuracy of the method has been evaluated with real card copy points provided by a bank.

Keywords: Point of Compromise Detection, Autoencoder, Deep Learning

1
GİRİŞ

Teknolojinin gelişmesi ile birlikte ticari ve finansal kurumlar daha çok müşteriye hitap etmek ve müşterilerine daha iyi hizmet verebilmek için çevrim içi çalışmaya dayanan e-ticaret siteleri gibi uygulamalar geliştirmiştir. Bununla birlikte, kredi kartı kullanıcılarının sayısı artarak tüketicilerin alışveriş alışkanlıkları bu doğrultuda değişmiştir.

Bankalar ve hizmet veren diğer şirketler, müşterilerine en iyi şekilde hizmet verebilmek için rekabet içinde uygulamalarını ve sistemlerini geliştirirken, bunun paralelinde suçlular bu sistemleri yasal olmayan yollarla alt etmek ve kredi kartı kullanıcılarının kart bilgilerini ele geçirmek için yeni yollar aramaktadırlar. Dolandırıcıların, bilgilerini ele geçirdikleri kredi kartlarıyla yaptığı yasal olmayan her işlem, sahte işlem olarak belirtilmektedir.

Sahte işlemler, kaybolan/çalınan kredi kartları, sahibine ulaşmamış kredi kartı, sahte kart başvurusu gibi farklı sebeplerden dolayı gerçekleşmektedir. Sahte işlemlerin gerçekleşmesine sebep olan bir diğer faktör ise kart kopyalamadır. Kart kopyalama, daha önce bilgileri açığa çıkmış veya kopyalanmış kredi kartlarının, belli bir zaman sonra kötü niyetli kişiler tarafından kullanılması ile gerçekleşmektedir [1]. Kredi kart kopyalama, fiziksel ve çevrim içi (online) olmak üzere iki farklı şekilde gerçekleştirilebilir. Fiziksel kredi kart, kopyalama cihazından (skimmer) geçirildiğinde kredi kartına ilişkin kart numarası, son kullanma tarihi ve kart sahibinin isim, soy isim bilgileri elde edilmektedir. çevrim içi kopyalamada ise kredi kartı bilgileri *oltalama (phishing)*, *SQL enjeksiyonu (SQL injection)*, *anahtar kaydedici (keylogger)* gibi siber saldırı yöntemleri ile kullanıcıdan veya e-ticaret sisteminin sunucularından ele geçirilmektedir. Kredi kart bilgilerinin fiziksel olarak çalındığı yerlere kart kopyalama noktaları denilmektedir. Bu kredi kart kopyalama noktalarının tespiti yapılabilirse, bu noktalarda kartların kopyalanmasını engellemek için gerekli tedbirlerin alınması mümkün olacaktır. Mevcut sistemlerde, kart kopyalama noktalarının tespit edilmesi, kart kullanıcıları tarafından yapılan bildirimlerin bu alanda çalışan banka personelleri tarafından incelenmesiyle gerçekleştirilmektedir. Kredi kartı sahteciliği ile ilgili her

ay yaklaşık 20.000 ile 40.000 arasında değişen sayılarda sahte işlem bildirilmektedir. Bilginin büyüklüğü ve banka personelleri tarafından yapılan kontrollerin zaman alıcı olmasından ötürü aylık 6 ile 18 arasında değişen kopyalama merkezleri tespit edilebilmektedir.

Bu çalışmada, öncelikle kredi kartı sahtecilikleri ve kart kopyalama problemi detaylı olarak incelenmiştir. Kart sahteciliklerini ve kart kopyalama vakalarını tespit etmeye yönelik yapılan çalışmalar için literatür taraması yapılmış, kart sahtecilikleriyle ilgili akademik çalışmalarda paylaşılan veri kümeleri incelenmiştir. Daha sonra, problemin çözümüne yönelik geliştirilecek olan sistemin ana adımları belirlenmiş ve bu adımlarda kullanılabilecek yöntemler incelenmiştir. Son olarak, kart kopyalama noktalarını tespit etmek için önerilen sistemin bileşenleri açıklanmış ve gerçek bir veri kümesi ile sistem performansı ölçülerek elde edilen sonuçlar değerlendirilmiştir.

1.1 Literatür Özeti

Kart kopyalama merkezlerinin tespit edilmesi, kredi kartı sahtekarlıkları problemleriyle ilişkilidir. Bu nedenle, önce kart sahtekarlıklarının tespiti ile ilgili yapılan çalışmalar incelenmiş daha sonra doğrudan kart kopyalama noktalarının tespitine yönelik yapılmış olan çalışmalar incelenmiştir.

1.1.1 Kredi Kartı Sahtekarlıklarının Tespiti

Dolandırıcıların kredi kartı sahtekarlığı yapması, banka ve diğer kurumlar için ortak bir sorundur. Kredi kartı sahtekarlığı tespitinde son yıllarda ses, görüntü işleme ve doğal dil işleme gibi birçok alanda yaygın olarak kullanılan derin öğrenme yöntemleri gibi daha modern yaklaşımları içeren çözümler de öne sürülmüştür. Kredi kartı sahtekarlığı tespiti için yapılan literatür taraması, geleneksel makine öğrenmesi yaklaşımlarını içeren ve derin öğrenme yaklaşımlarını içeren çalışmalar olmak üzere iki bölümde incelenmiştir.

1.1.1.1 Geleneksel Makine Öğrenmesi Yöntemleri ile Kredi Kartı Sahtekarlığı Tespiti

Maes vd. [2], kredi kartı sahteciliği tespiti için yaptıkları çalışmada, Bayes Ağı ve yapay sinir ağı modeli kullanmıştır. Europay International aracılığıyla elde ettikleri gerçek işlemlere ait bir veri kümesi kullanmıştır. Veri gizliliği nedeniyle, verilerdeki özellikler ve diğer detaylar paylaşılmamıştır. Başarı değerlendirmesi için ROC eğrisi kullanılmıştır. Yapay sinir ağı modelleri, daha hızlı tahmin üretse de, Bayesian Belief Network yaklaşık olarak %10 daha az yanlış olumlu sonuç üretmiştir.

A. Srivastava vd. [3], kart kullanıcılarının harcama alışkanlıklarındaki değişimden yola çıkarak kart sahtekarlığını tespit etmeye çalışan bir yöntem önermiştir. Bunun için Saklı Markov Model Yöntemi ile anomali tespiti yapan bir sistem kullanılmıştır. Önerdikleri sistemde kart kullanıcılarının harcama alışkanlıkları belirlemek için, K-ortalama kümeleme yöntemi ile benzer harcama alışkanlıklarına sahip kullanıcılar kümelenerek profiller oluşturulmuş ve her profil için bir Saklı Markov Model eğitilmiştir. Kullanılan veri kümesi ile ilgili detaylar veri gizliliği nedeniyle paylaşılmamıştır. Başarı değerlendirmesi için doğruluk oranı metriği kullanılmıştır. Elde edilen doğruluk oranı için, model başarısı %80 olarak belirtilmiştir.

Y. Sahin vd. [4], kart sahtekarlığı problemini bir sınıflandırma problemi olarak ele almış ve denetimli sınıflandırma yöntemleri olan Destek Vektör Makineleri (DVM) ve Karar Ağaçlarını kullanarak bir çalışma yapmıştır. Karar Ağacı yöntemleri için, c5.0, c&RT ve CHAID Karar Ağacı Modelleri, DVM için polinomsal, sigmoid doğrusal ve radyal tabanlı (radial basis) çekirdekli 3 farklı DVM modeli kullanılmıştır. Çalışmalarında gerçek işlemlere ait bir veri kümesi kullanmıştır. Veri kümesindeki özellikler, gizlilik esası gereği paylaşılmamış olsa da, özelliklerin işlem istatistikleri, bölgesel istatistikler, sektörel istatistikler, günlük harcama miktarı ve günlük işlem istatistikleri gibi detay bilgilerden oluştuğu paylaşılmıştır. Y. Sahin vd., veri kümesinde 978 sahte işlem bulurken 22 milyon normal işlem bulunmaktadır. Bu da veri kümesindeki sahte işlemlerin normal işlemlere oranının yaklaşık olarak 1:22500 olduğunu göstermektedir. Bu oran, DVM ve karar ağacı modellerinin eğitimleri için büyük bir problem olduğu için, oransal örnekleme yöntemi kullanarak bu oran 1:2, 1:4, 1:9 olacak şekilde farklı örneklemeler ile model eğitimleri gerçekleştirilmiştir. Başarı değerlendirmesi, doğruluk (accuracy) hesaplanarak yapılmıştır. Buna göre, elde edilen en iyi doğruluk skorları, DVM test kümesinde %83 iken, karar ağacı olan c5.0 karar ağacında ise %91 olarak elde edilmiştir.

Bahnsen vd. [5], yaptıkları çalışmada Karar Ağacı, Lojistik Regresyon ve Rassal Orman (Random Forest) sınıflandırma yöntemlerini kullanmıştır. Çalışmada, yararlanılan veri kümesi, 27 özellik içeren 120 milyon işleme ait veriden oluşmaktadır. Veri kümesindeki işlemlerde, sahte işlem olarak etiketlenmiş 40.000 işlem bulunmaktadır. Veri kümesindeki 27 özelliği doğrudan kullanmak yerine, işlemlerdeki harcama davranışlarından profil oluşturmak için, işlem verilerini birleştirerek 293 yeni özellik elde etmişlerdir. Oluşturulan bu profilleme ile elde edilen yeni veri kümesinde 236,735 işlem bulunurken bu işlemlerden %1.5 kadarı sahte işlem olarak belirtilmiştir. Başarı değerlendirmesi için veri kümesindeki gerçekleşen sahte işlemlerden kaynaklanan finansal kaybı da kullanarak, finansal maliyete duyarlı başarı metrikleri belirlemişlerdir. Çalışma sonucunda, harcama alışkanlıklarını gösteren yeni özelliklerin eklenmesi ile elde edilen sonuçların, 27 özellik kullanıldığında elde edilen

sonuçlara göre yaklaşık olarak %287 daha başarılı olduğu belirtilmiştir.

A. Mishra vd. [6], Karar Ağaçları, Lojistik Regresyon, Gradyan Artırma Ağacı (Gradient Boosted Tree) ve Rassal Orman gibi sınıflandırma yöntemler ile kart sahteciliği tespiti yapmıştır. Çalışmalarında, açık kaynak olarak erişilebilen European Credit Cards 2013 [7] veri kümesi kullanılmıştır. Bu veri kümesinde 284.315 yasal işlem, 492 tane ise sahte işlem bulunmaktadır. Bu veri kümesindeki özellikler içerisinde işlem zamanı ve harcama miktarı gibi özellikler bulunmazken, diğer özellikler ise PCA yöntemi ile maskelenerek veri gizliliğine uygun bir şekilde paylaşılmaktadır. A. Mishra vd, kullandıkları modelleri eğitirken sahte işlemlerin yasal işlemlere oranını dengelemek için alt örnekleme (undersampling) yöntemleri kullanmışlardır. Alt örnekleme ile yasal işlemlerden rasgele örnekler seçilerek, sahte işlem yasal işlem oranı 1:1 hale getirilmiştir. Başarı değerlendirmesi için, hassasiyet (precision) ve duyarlılık (recall) metrikleri kullanılmıştır. Buna göre, Rassal Orman Yöntemi ile her iki metrikte de göre diğer yöntemlere göre daha yüksek başarı elde edildiği belirtilmiştir.

K. Randhawa vd. [8], kart sahteciliği tespiti için yaptıkları çalışmada Naive Bayes, Karar Ağacı, Rassal Orman ve Gradyan Bazlı Ağaç (Gradient Based Tree) tekil sınıflandırıcılarını kullanarak, hem Adaboost hem de Oy Çokluğu (Majority Voting) Yöntemleri ile kollektif öğrenme modelleri uygulamıştır. Veri kümesi olarak Malezya'daki bir enstitüden temin edilen gerçek veri kümesi kullanılmıştır. Veri kümesi 2017 Şubat ve 2017 Nisan ayları arasında gerçekleşen, 102 tanesi sahte işlem olmak üzere toplam 287.2244 işlemden oluşmaktadır. Veri kümesindeki işlemlerde 11 özellik bulunmaktadır. Başarı değerlendirmesi olarak Matthews korelasyon katsayısı (Mathew Correlation Coefficient, MCC) kullanılmışlardır. Elde edilen sonuçlara göre, Adaboost ve Oy Çokluğu kullanarak oluşturulan kollektif öğrenme yöntemleri ile tekil sınıflandırıcılardan daha yüksek başarılara sahip yeni modeller oluşturulduğu belirtilmiştir.

1.1.1.2 Derin Öğrenme Yöntemleri ile Kredi Kartı Sahtekarlığı Tespiti

Z. Kazemi vd. [9], yaptıkları çalışmada Özkodlayıcı (AutoEncoder) modelinin son katmanına eklediği Softmax aktivasyon fonksiyonu ile, kredi kartı sahteciliği tespiti yapan bir sınıflandırma modeli önermiştir. Çalışmada, açık kaynaklı olarak paylaşılan ve yaygın olarak kullanılan UCI German Credit Data [10] veri kümesini kullanmışlardır. Bu veri kümesinde, 1000 kredi kartı işlemi ve bu işlemlere ait 20 özellik bulunmaktadır. Başarı değerlendirmesi için, doğruluk metriğinden faydalanılmıştır. Ayrıca önerdikleri yöntemi, K En Yakın Komşu Algoritması (K Nearest Neighborhood, KNN), Özdüzenyeci Haritalar (Self Organizing Map, SOM) gibi diğer yöntemlerin sınıflandırma başarısıyla karşılaştırmışlardır. Yapılan deneyin sonucunda,

doğruluk skoru 0.84 ile Özkodlayıcı ile diğer yöntemlere göre daha iyi sonuçlar elde edilebildiği gösterilmiştir.

A. Roy vd. [11], çalışmalarında derin öğrenme tabanlı sınıflandırma yöntemleri olan Yapay Sinir Ağları (Artificial Neural Network, ANN), Tekrarlayan Sinir Ağları (Recurrent Neural Network, RNN), Uzun Kısa Süreli Bellek Birimleri (Long-Short Term Memory, LSTM) ve Geçitli Tekrarlayan Birim (Gated Recurrent Unit, GRU) derin öğrenme yöntemlerini kullanmışlardır. Çalışmada, Bir bankadan temin ettikleri, 8 aylık zaman diliminde gerçekleşen kredi kartı işlemleri ile veri kümesi oluşturmuşlardır. Veri kümesinde işlem miktarı, iş yeri kategorisi, iş yeri tipi gibi işlemlere ve iş yerlerine ait detaylar yer almaktadır. Ayrıca veri kümesindeki işlemlerde, sahte işlemlerin yasal işlemlere oranı %0.14 olarak belirtilmiştir. Çalışmada sahte işlemlerin yasal işlemlere oranını dengelemek için alt örnekleme yöntemi kullanarak %0.14 olan sahte işlem oranını, %10 olacak şekilde dengelemişlerdir. Ayrıca sayısal özellikler kullanan derin öğrenme yöntemlerini gerçekleştirebilmek için, kategorik özellikleri One-Hot Encoding Yöntemi ile sayısal hale dönüştürmüşlerdir. Başarı değerlendirmesinde metrik olarak doğruluk oranı kullanılmış ve değerlendirme için 10 katlamalı çapraz doğrulama yapılmıştır. Deney sonuçlarına göre %91.6 doğruluğa sahip GRU yönteminin diğer yöntemlere göre daha iyi sonuç verdiği belirtilmiştir.

Pumsirirat vd. [12],kredi kartı sahteciliği problemini anomali tespiti problemi olarak ele alıp, Özkodlayıcı ve Kısıtlı Boltzmann Makinesi (Restricted Boltzman Machine,RBM) derin öğrenme yöntemlerini içeren bir çalışma yapmışlardır. Veri kümesi olarak açık kaynaklı olarak erişilebilen, UCI German Credit Data [10], UCI Australian Credit Approval [13] ve European Credit Cards [7] olmak üzere üç farklı veri kümesi kullanılmıştır. UCI German Credit Data veri kümesinde 1000 adet, UCI Australian Credit Approval veri kümesinde ise 690 adet işlem bulunmaktadır. Başarı değerlendirmesi için Alıcı İşletim Karakteristiği (Reciever Operator Characteristics, ROC Eğirisi) ve Eğri Altında Kalan Alan (Area Under Curve, AUC) skor kullanılmıştır. Çalışmada, elde edilen sonuçlara göre Özkodlayıcı ve RBM yöntemlerinin başarı sonuçlarının birbirine oldukça yakın olduğu, her iki yöntem için de European veri kümesinde AUC skoru %96'lara varan başarılar olduğu belirtilmiştir.

1.1.2 Kart Kopyalama Noktası Tespiti

Kredi kartı kopyalama noktalarını bulmaya yönelik yapılan çalışmalar, üzerinde çalışılabilecek herkese açık veri kümesi bulmak ve başarı değerlendirmesi yapmak için gerçek kopyalama noktalarının bilgisine erişmek zor olduğu için kısıtlıdır. Yaptığımız literatür taramasında doğrudan kart kopyalama noktalarını tespit etmeye yönelik

ancak bir adet bilimsel çalışma incelenebilmiştir.

M. Araujo vd. [14], yaptıkları çalışmada, kopyalama noktalarını bulma problemini, bilgisayar dosya sistemindeki zararlı yazılımları bulma problemine benzeterek, bu problemi çözebilmek için iki parçalı graf (bipartite graph) kullanmıştır. Ayrıca önerdikleri yöntemin paralel işlem yapabilmeye müsait olmasından dolayı, ölçeklenebilir olduğu vurgulanmıştır. Ancak, yapılan çalışmada kullanılan veri kümesine ait detaylar, veri gizliliği nedeniyle paylaşılmamıştır. Ayrıca önerilen yöntem ile hangi zaman aralığı için kaç adet kopyalama noktasının tespit edildiği ve başarı değerlendirmesi için referans bilgisi olarak kullanılacak gerçek kart kopyalama noktası sayısı da belirtilmemiştir. Çalışmalarında önerilen yöntemin bir diğer dezavantajı ise, yeni kopyalama noktalarını tespit edebilmek için, daha önceden tespit edilmiş başka kopyalama noktalarının varlığına da ihtiyaç duymasıdır.

1.2 Tezin Amacı

Tezin amacı, kredi kartı kopyalama noktaları olan iş yerlerini tespit edebilmektir. Bu sayede, kopyalama işlemini gerçekleştiren bu iş yerlerine müdahale edilebilecek ve daha önce bu iş yerlerinde kopyalanmış ancak henüz sahte işlem yapılmamış kart gruplarının, gerçekleşebilecek yeni sahte işlemlerde kullanılması engellenebilecektir.

Benzer sahte işlemlere sahip kredi kartlarının aynı iş yerlerinde kopyalanmış olmaları muhtemeldir. Bu yaklaşımla, kopyalanma kaynaklı gerçekleşen benzer sahte işlemlere sahip kredi kartları gruplandırılır. Tespit edilen kart gruplarının geçmiş işlemlerinde ortak olarak kullanıldıkları üye iş yerleri incelenerek potansiyel kart kopyalama noktalarının bulunması amaçlanmıştır.

1.3 Hipotez

Kredi kartı kopyalama işlemi, kart sahtekarlarının fiziksel veya çevrim içi ortamlarda, hedef seçtiği kullanıcıların kredi kart bilgilerini ele geçirmesi işlemidir. Kart sahtekarları, daha sonra bu ele geçirilen kredi kart bilgileri ile sahte işlemler gerçekleştirebilmektedir. Kart kopyalama esnasında genellikle kart sahtekarları, çok sayıda kullanıcının bilgisini ele geçirebilir. Ele geçirilen farklı kullanıcılara ait bu kart bilgilerinin genellikle aynı yerlerde kullanıldığı görülmüştür. Bir diğer deyişle, aynı kopyalama noktasında kopyalanan kredi kartlarının birlikte hareket etmesi durumu söz konusudur. Ayrıca burada bir diğer önemli nokta ise, kullanıcıların kartlarının kopyalanması esnasında gerçekleştirdiği kredi kartı işlemlerin genellikle yasal işlemler olmasıdır.

Bu tez çalışmasında önerilen yöntemde, ilk olarak benzer sahte işlemlere karışan kopyalanmış kredi kartlarının birlikte hareket ettiği varsayılmıştır. Bundan yola çıkarak, sahte kullanımı beyan edilen kredi kartlarının geçmiş harcamalarına yönelik sahte/yasal tüm kart işlemlerinde ortak olarak kullanıldıkları iş yerleri tespit edilmiştir. Tespit edilen iş yerlerinden kopyalama noktası adayları belirlenmiştir. Önerilen yöntemde, benzer işlemlere kullanılan her kart grubu için en az bir kopyalama noktası adayı belirlenmektedir.

2
KREDİ KARTI SAHTEKARLIKLARI VE KREDİ KARTI KOPYALAMA

Günümüzde, kredi kartı dolandırıcılığı en sık rastlanan suç kategorilerinden birisidir. Kart dolandırıcılıkları çeşitli yollarla gerçekleştirilebilir. Bunlardan bazıları; kart başvuru sahtecilikleri, kart hesabını ele geçirme, kayıp çalıntı ve kart kullanım sahteciliği başlıkları ile özetlenebilir. Kart kullanım sahteciliği başlığı altında incelenmiştir [1]. Bu tez çalışmasında, kart kopyalama sahtekarlığı kategorisinde gerçekleşen sahte işlemlerde sebebiyet veren kartın kopyalandığı merkez olarak nitelendirilen iş yerlerinin tespit edilmesi amaçlanmıştır.

2.1 Kart Başvuru Sahtecilikleri

Bankalar müşterilere kart talep etme işlemlerinde kolaylık sağlama amacı ile çeşitli yollar sunmaktadır. Bu yolların tümünde, kişinin kart talebinde bulunması için kimlik doğrulama adımlarını geçmesi gerekir. Kart başvuru sahteciliği, genellikle kişilerin kimlik bilgilerinin çalınması sonucunda gerçekleştirilir. Kart talebinde bulunma esnasında gereken tüm kişisel bilgilerin elde edilmesiyle, kimlik doğrulaması adımı başarıyla geçilir ve şifre oluşturma, kartın teslim alınması gibi süreçler kişinin haberi olmadan gerçekleştirilebilir. Bu kartların kullanımı şüphe uyandırmadan sürdürülebilir. Kart bilgilerinin çalınmasının yanı sıra yaşlı, akli sağlığı yerinde olmayan vatandaşlar gibi kandırılması veya zorlanması daha kolay kişiler kullanılarak sahte kart başvuruları gerçekleştirilebilmektedir.

2.2 Kart Hesabının Ele Geçirilmesi

Kart hesabının ele geçirilmesi kredi kartı limitini veya kart hesabında bulunan parayı kullanmak için gerçekleştirilir. Dolandırıcılar hesap bakiyesi kullanımının yanı sıra, ek kart çıkartma gibi işlemler de uygulamaktadırlar. Kart hesabının ele geçirilmesi, kredi kartı sahibinin kart kullanımı için gerekli olan temel kişisel bilgilerin, sahtekar

tarafından ele edilmesiyle gerçekleştirilebilen bir dolandırıcılık türüdür.

2.3 Kayıp veya Çalıntı Kart Kullanımı

Kayıp ve çalıntı kart kullanımı dolandırıcılık türü, kayıp veya çalıntı kartların fiziksel olarak dolandırıcılar tarafından kullanılmasıdır. Kişinin dikkatsizlik ile kartını kaybetmesi, alışveriş sırasında şifre ile birlikte dolandırıcı tarafından el çabukluğuyla çalınması, kartın zorla gasp edilmesi, kuryelerin kartı yanlış kişiye teslim etmesi veya kuryelerin kişisel bilgileri kötüye kullanması gibi çeşitli girişimler ile yapılan dolandırıcılık türüdür. Kartların temassız ödeme özelliği de bu tür dolandırıcılıklarda dezavantaj oluşturur. Fiziksel olarak elde edilen bir kartın şifresi bilinmese bile, temassız işlem ile kart kullanılabilir.

2.4 Kart Kopyalama Sahtekarlığı

Gerçek bir kartın bilgileri çalınarak orijinal kart bilgileri ile kopya kartlar oluşturulmasıdır. Orijinal kart bilgilerinin ele geçirilmesi için çeşitli yollar izlenmektedir. Bunlardan birisi ATM veya POS cihazlarına kopyalama cihazlarının yerleştirilmesidir. İkinci yol ise orijinal bilgilerin bulunduğu veri tabanlarından bilgilerin çalınmasıdır. Dolandırıcılar bu verileri geçerliliğini yitirmiş kartlara kopyalayarak sahte kartları oluştururlar. En baştan boş bir kartı gerçek karta benzer şekilde üretmek maliyetli ve zor olduğu için genellikle kullanılmayan kartlar üzerine yazma işlemini yaparlar. Beyaz plastik kart adı verilen gerçek kartlarda bulunan hologramları içermeyen basit manyetik kartlara da bu kopyalama işlemi uygulanmaktadır. Bu kartların kullanımı ATM'lerde veya kötü amaçlı iş yerlerinde gerçekleştirilir. Kart kopyalama sahtekarlıkları fiziksel ortamlarda sahtekarlık amacına yönelik tasarlanan cihazlar aracılığıyla gerçekleşen fiziksel kopyalama, sanal ortamda siber saldırılar ile gerçekleşen çevrim içi kopyalamalar olarak iki kategoride ele alınmıştır.

2.4.1 Fiziksel Kopyalama

Fiziksel kopyalamalar, ATM kaynaklı ve POS cihazı kaynaklı kart kopyalama olmak üzere iki farklı şekilde gerçekleşebilir.ATM kaynaklı kart kopyalama, kart kopyalama işlemi kopyalama cihazının (skimmer) ATM'lere yerleştirilmesi ile gerçekleştirilir. ATM yoluyla kart kopyalamalarında kart giriş yuvasına kopyalama cihazının yerleştirilmesi ile kart manyetiğinde bulunan bilgiler kopyalanır. Şekil 2.1'de ATM'lerde kullanılan sahte kart giriş yuvasına yerleştirilen cihaz görülmektedir. Kullanıcılar, normal işlemlerini yaparken karta ait bilgiler bu cihaz ile kopyalanmaktadır. Fakat bu

cihaz kart şifresini elde edemez. Şifreyi elde etmek için gizli kamera ile girilen şifrenin kaydedilmesi, dolandırıcıların yardımcı olma gibi bahanelerle kart sahiplerini kandırması veya sahte klavye (PINPAD) kullanılması gibi yollar izlenmektedir. Daha az rastlanan bir yol ise, sahte ATM'ler üzerinden kopyalama işleminin yapılmasıdır.

Şekil 2.1 Kart kopyalama amaçlı ATM'lerin kart giriş yuvasında kullanılan kart kopyalama aygıtı (skimmer) [15]

POS cihazı ile kopyalama işlemlerinde, iş yerleri kötü amaçlı davranmaktadır. Dolandırıcı iş yerlerinden yapılan alışverişlerde, ödemeyi alan şahıs kartı ikinci bir cihazdan geçirerek kopyalamayı gerçekleştirir. Kart şifresini de aynı şahıs almaktadır veya POS cihazının içerisine şifreleri kaydeden bir eklenti entegre edilmektedir. Bunun yanı sıra, Şekil 2.2'de görülen sahte POS cihazları da kullanılabilmektedir. Bu cihazlar POS cihazları görünümünde, sahte slip veren kopyalama cihazlarıdır. Genellikle taksiler, yemek siparişleri gibi en sık alışveriş yapılan işlemlerde kullanılırlar.

2.4.2 Çevrim İçi Kopyalama

İnternet kullanımının artmasıyla bankacılık sektörü de sanal ortamda müşterilerine hizmet sunmaktadır. Bu hizmet hayatımızı yüksek düzeyde kolaylaştırsa da diğer taraftan dolandırıcılara yeni kapılar açmaktadır. Bu bölümde çevrim içi kopyalama işlemlerinden yaygın olanları kısaca anlatılmaktadır.

2.4.2.1 Oltalama

Bu sahtekarlık türünde dolandırıcıların amacı sahte bir e-posta hazırlayarak, göndereni resmi bir kurum, marka veya kişi olarak gösterip, kurbanın bilgilerini ele geçirmektir. Dolandırıcı, e-posta ile kurbanı kart şifresini yenilemeye yönlendirebilir veya virüslü ekler göndererek bilgisayarını ele geçirebilir. Bu e-postalar, genellikle kart

Şekil 2.2 Kart kopyalama amaçlı kullanılan sahte POS cihazı ve gerçek POS cihazı [16]

kullanım şifresinin süresinin dolması gibi uyarılar, kampanya, hediye gibi korkutucu veya teşvik edici içeriklere sahiptir.

2.4.2.2 SQL Enjeksiyonu

Dolandırıcılar, siber suç yöntemlerinden biri olan SQL enjeksiyonu ile e-ticaret uygulamalarının veya finansal veri taşıyan diğer kuruluşların veritabanlarına yasal olmayan yollar ile erişip, veri sızıntısı gerçekleştirebilir. Başarılı bir SQL enjeksiyonu saldırısı gerçekleştiren siber suçlular, eriştikleri veri tabanındaki bütün kullanıcıların kredi kartı bilgilerini aynı anda ele geçirebilir.

2.4.2.3 Anahtar Kaydedici

Siber suçluların kredi kartı bilgilerini elde etmeleri için kullandıkları siber saldırı yöntemlerinden birisi de anahtar kaydedici yöntemidir. Bu yöntem genellikle doğrudan kart kullanıcılarını hedef alır. Anahtar kaydedici ile suçlu, kredi kartı kullanıcısının kullandığı klavye gibi donanımlarındaki hareketlilikleri kaydedip, taklit edebilir.

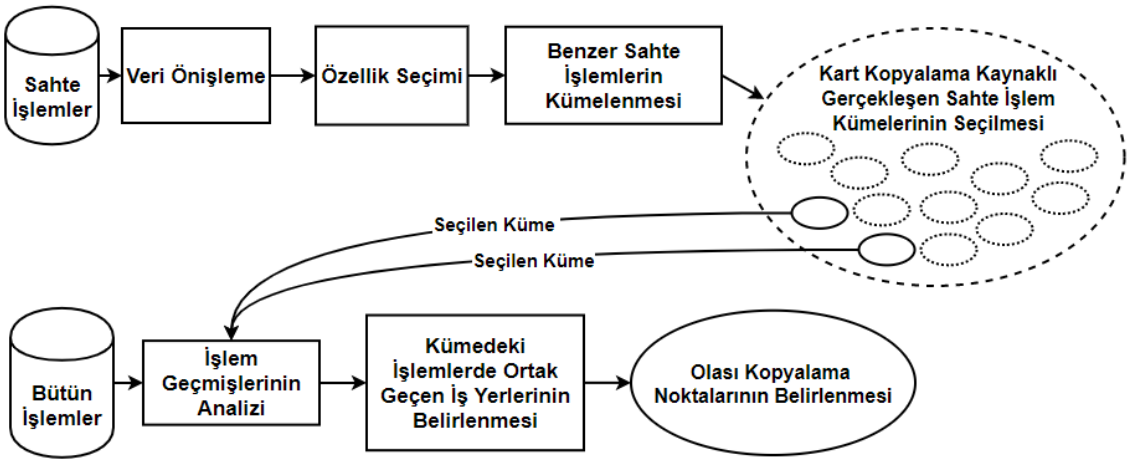
3
KREDİ KARTI KOPYALAMA MERKEZİ TESPİT SİSTEMİ TASARIMI

Kart kopyalama, kredi kartı sahtekârlarının yaygın olarak başvurduğu yöntemlerden biridir. Kart kopyalayan suçlular, genellikle aynı noktada birden fazla kredi kartı için kopyalama gerçekleştirirler. Kopyalanan kartlar daha sonra aynı veya farklı noktalarda harcamalarda kullanılabilir. Kopyalanan kartların aynı kişi tarafından kullanılması durumunda, bu harcamalarda ortak özellikler olacağı öngörülebilir. Bu nedenle, bu çalışmada sahte işlemlerden yola çıkarak kart kopyalama noktalarının tespit edilebilme durumu incelenmiştir.

Tespit edilen sahte işlemlerin kaynağının daha önce gerçekleşmiş bir kopyalama olduğunu belirlemek oldukça güçtür. Bankalar bu tarz durumlarda genellikle kullanıcılardan gelen bildirimler doğrultusunda harekete geçmektedir. Kredi kartı kullanıcıları, sahte işlem bildirisi yaptıklarında, genellikle sahte işleme neden olan dolandırıcılık yönteminin sebebini bilemezler. Bu nedenle sadece kullanıcılardan yapılan geri bildirimler ile kart kopyalama merkezlerini tespit etmek güçtür. Bankalar kart kopyalama merkezlerini operatörlerin yaptıkları geniş kapsamlı, uzun zaman alan incelemeler sonunda tespit edebilmektedirler.

Bu çalışma, aynı kopyalama merkezinde kopyalanan kredi kartlarının benzer sahte işlemlerde kullanılma ihtimalinin yüksek olması fikrinden yola çıkarak gerçekleştirilmiştir. Kopyalayan kişiler tarafından aynı yerlerde kullanılmasından dolayı karakteristik olarak birbirlerine benzeyen bu sahte işlemleri tespit edebilmek için, önce karakteristik özellikler belirlenmiştir. Daha sonra, denetimiz öğrenme ile benzer sahte işlemler ve dolayısıyla benzer sahte işleme karışmış kredi kartları kümelere ayrılmıştır. Elde edilen kümelerden, kart kopyalama kaynaklı sahte işlemlerin yoğun olduğu kümeler seçilmiş ve bu kümelerdeki kartların geçmiş harcamalarındaki ortak olarak kullanıldıkları noktalardan kopyalama noktalarını tespit etmek için kartların geçmiş harcamaları üzerinden geriye yönelik analiz yapılmıştır. Şekil 3.1'de bu çalışmada önerilen sistemin temel işlem adımlarını

gösteren blok diyagramı verilmiştir. Kullanılan yöntemler ve veri kümesi bu bölümde açıklanmıştır.



Şekil 3.1 Sistemin blok diyagramı

3.1 Veri Kümesi

BKM, bankalar arası gerçekleşen her türlü ödemeyi veya para transferini sağlayan veya destekleyen sistem, platform ve altyapıları oluşturan bir kuruluştur. Çalışmada kullanılan veri kümesine erişim BKM aracılığıyla gerçekleştirilmiştir. 2016-2018 yıllarında bankalar tarafından BKM'ye bildirilen, bankalar arasında gerçekleşen yurt içi işlemlerden oluşan 681.862 tane sahte işleme ait bir veri kümesi hazırlanmıştır. Veri kümesinin 2018 yılına ait Eylül-Aralık aylarında gerçekleşen sahte işlemler deneylerde kullanılmıştır. Yine bu doğrultuda yapılan deneylerde, 2018 Eylül - 2018 Aralık aylarında gerçekleşen yasal ve sahte bütün işlemlere ait yaklaşık 2.44 milyar işlem üzerinde, belirlenen kart kümelerine ait işlemler için geriye yönelik analiz yapılmıştır. Ayrıca bir banka tarafından 2018 Ekim ve 2018 Aralık ayları arasında gerçek kart kopyalama merkezi olarak belirlenmiş 17 satın alma noktası, yöntemin başarısını değerlendirmek için referans olarak kullanılmıştır.

Çalışmada, benzer sahte işlemlere konu olmuş kredi kart gruplarını bulmak için sahte işlemlerin olduğu, VISA TC40 [17] sahtekarlık bildirimi standartlarına uygun olarak düzenlenen işlemlerin olduğu veri kümesi, kopyalanan kredi kartlarını tespit etmek için kullanılmıştır. Gerçek kişi ve kurumlara ait verilerden oluşan bu veri kümelerine erişim, BKM ile yapılan gizlilik sözleşmesi çerçevesinde gerçekleştirilmiştir.

Veri kümesinde yer alan sahte işlemler, kart sahibi tarafından bankalar aracılığıyla BKM'ye iletilen finansal olmayan sahte işlem bildirilerinden oluşmaktadır. Bu veri kümesinde bulunan işlemler için iş yeri detayları, bankaya ait detaylar ve işlemin tipi

(temassız, manyetik vb.) işlem zamanı gibi işleme ait 44 özellik yer almaktadır. Tablo 3.1'de bu 44 özellik, özelliğe ait sayısal bilgiler, özellik tipleri verilmiştir. Tablonun özellik sütununda bir kart harcamasını tanımlayan bilgiler yer almaktadır. Bunlar kart tipi, işlem kanalı, iş yeri adı ve harcamaya ait detay bilgilerdir. Farklı değer sayısı, bir özelliğe ait farklı değerlerin sayısını ifade etmektedir. Boş değer sayısı, veri kümesinde bir özellik için girilmemiş/boş değerlerin sayısını ifade etmektedir. Ayrıca birlikte hareket ettiği tespit edilen kart gruplarından yola çıkarak yapılan geçmiş işlemlerin analizi için sahte/yasal olan bütün işlemlerin olduğu bir ikinci veri kümesi kullanılmıştır. İşlem geçmişi analizinde kullanılan bu ikinci veri kümesindeki özelliklere ait sayısal değerler ise Tablo 3.2'de verilmiştir.

Veri hassasiyeti nedeniyle kurumlar, kredi kartı bilgilerini açık olarak paylaşmamaktadır. Bu nedenle, bu alanda yapılan çalışmalarda kullanılan veri kümelerinin içeriği, gizlenmiş ya da sınırlı olarak yayınlanmıştır. Bu çalışmada kullanılan veri kümesinin yapısını literatürdeki mevcut çalışmalarda kullanılan veri kümeleri ile karşılaştırmak için farklı kredi kartı sahteciliği çalışmalarında kullanılan veri kümelerine örnekler Tablo 3.3'te verilmiştir. BKM veri kümesi, bankalar arası işlemleri içerdiği için verideki özellik sayısının ve toplam işlem sayısının diğer veri kümelerine göre oldukça fazla olduğu açıkça görülmektedir. Ayrıca paylaşılan veri kümeleri içerisinde sadece Machine Learning Group [7] veri kümesi erişilebilir durumdadır.

Kredi kartları genellikle yasal işlemler sırasında kopyalanır. Kopyalanmış bir kart sahte işlemde kullanılana kadar, kart sahipleri ve bankalar tarafından kartın kopyalanmış olduğundan haberdar olması güçtür. Bu nedenle, sahte işlemlere ek olarak benzer kopyalanma vakalarına karıştığı belirlenen kredi kartlarının geçmiş işlemlerini analiz edebilmek için yasal işlemlerin de olduğu ek bir veri kümesine de ihtiyaç duyulmuştur.

3.2 Veri Analizi

Veri kümesinin daha efektif değerlendirilebilmesi için önce çeşitli analizler yapılmıştır. Veri analizi aşamasında, sahte işlemler için Tablo 3.1'de, bütün işlemler için Tablo 3.2'de verilen bilgilerden faydalanılmıştır. Daha sonra, özellik ikililerinin birbirleriyle ilişkilerini değerlendirmek için, korelasyon matrisleri oluşturulmuştur. Korelasyon matrisleri oluşturulurken, özellik ikililerinin tiplerine göre farklı metrikler kullanılmıştır:

- Cramer's V metriği [21]: Korelasyon değeri için hesaplanacak her iki özellik de kategori tipinde ise, Cramer's V metriği kullanılmıştır. Eşitlikteki değişkenler; χ^2

Tablo 3.1 Sahte işlem veri kümesinde bulunan özellikler

Özellik	Farklı Değer Sayısı	Boş Değer Sayısı	Özellik Tipi
BKM Genel İşyeri No	160.512	-	Nominal Kategorik
Terminal No	134.528	1.300.185	Nominal Kategorik
Acquirer Üye No	31	-	Nominal Kategorik
Acquirer Üye Adı	31	-	Nominal Kategorik
Issuer Üye No	34	-	Nominal Kategorik
Issuer Üye Adı	34	-	Nominal Kategorik
İşlem Tarihi	5.878	-	Aralıklı Nümerik
İşlem Tutarı	186.976	-	Sürekli Nümerik
Ödeme Sistemi Kodu	5	15.775	Nominal Kategorik
Ödeme Sistemi	6	-	Nominal Kategorik
Kart Tipi Kodu	3	15.775	Nominal Kategorik
Kart Tipi	4	-	Nominal Kategorik
Alt Kart Tipi Kodu	7	15.775	Nominal Kategorik
Alt Kart Tipi	8	-	Nominal Kategorik
Kart Numarası	1.334.656	-	Nominal Kategorik
Kart No İlk 6 Rakam	662	-	Nominal Kategorik
Kart No Son 4 Rakam	9.876	-	Nominal Kategorik
Sahtecilik/Dolandırıcılık Tarihi	5.266	-	Aralıklı Nümerik
Sahtecilik/Dolandırıcılık Tipi Kodu	10	-	Nominal Kategorik
Sahtecilik/Dolandırıcılık Tipi	10	-	Nominal Kategorik
İşlem Kanal Kodu	10	-	Nominal Kategorik
İşlem Kanalı	10	-	Nominal Kategorik
İşyeri Kategori Kodu	682	-	Nominal Kategorik
İşyeri Kategori Grup Kodu	5	-	Nominal Kategorik
İşyeri Kategori Grup Kodu	5	-	Nominal Kategorik
Geri Ödeme Değer Kodu	24	-	Nominal Kategorik
Geri Ödeme Değeri	24	-	Nominal Kategorik
MOTO/E-Ticaret Tipi	40	466.997	Nominal Kategorik
MOTO/E-Ticaret Tipi Açıklaması	17	-	Nominal Kategorik
Kart Giriş Tipi Kodu	130	1.300.185	Nominal Kategorik
Kart Giriş Tipi	22	-	Nominal Kategorik
Acquirer Referans No	4.396.032	-	Nominal Kategorik
Yurtiçi/Yurtdışı	2	-	Nominal Kategorik
ATM İşlemi mi?	2	-	Nominal Kategorik
Çipli İşlem mi?	-	4.414.584	Nominal Kategorik
MOTO/E-Ticaret Tipi	2	-	Nominal Kategorik
TemassızİşlemMi?	2	-	Nominal Kategorik
İşyeri Adı	814.144	-	Nominal Kategorik
İşyerinin İli	96.832	-	Nominal Kategorik
İşyerinin Ülke Kodu	189	1.415	Nominal Kategorik
İşyeri Ülkesi	189	-	Nominal Kategorik
Terminal Özellik Kodu	15	1.300.185	Nominal Kategorik
Terminal Özellikleri	11	-	Nominal Kategorik
Kart SahibiBelirleme Metod Kodu	16	1.300.185	Nominal Kategorik
Kart Sahibi Belirleme Metodu	6	-	Nominal Kategorik

ki-kare istatistiğidir, N örnek sayısıdır, k ise ikili özelliklerden daha az kategoriye sahip olan özelliğin kategori sayısını ifade eder.

$$\text{Cramer's V} = \sqrt{\frac{\chi^2}{N(k-1)}} \tag{3.1}$$

- Pearson korelasyon metriği [22]: Korelasyon değeri için hesaplanacak her iki özellik de sayısal değere sahipse, Pearson korelasyon metriği kullanılmıştır. Eşitlikteki X ve Y veri kümesi değişkenlerini, $\overline{X}$ ve $\overline{Y}$ bu değişkenlerin

Tablo 3.2 İşlem Veri kümesinde bulunan özellikler

Özellik	Farklı Değer Sayısı	Boş Değer Sayısı	Özellik tipi
İşlem Tarihi	16	-	Aralıklı Nümerik
BKM Genel İşyeri Numarası	2.819.072	12.531.754	Nominal Kategorik
Terminal No	23.642.112	1.325.927.238	Nominal Kategorik
Acquirer Üye No	39	-	Nominal Kategorik
Acquirer Üye Adı	39	-	Nominal Kategorik
Issuer Üye No	41	-	Nominal Kategorik
Issuer Üye Adı	41	-	Nominal Kategorik
İşlem Tipi	16	-	Nominal Kategorik
İşlem Tipi Açıklaması	16	-	Nominal Kategorik
İşlem Saati	144.654.336	188	Aralıklı Nümerik
İşlem Tutarı	14.945.280	-	Sürekli Nümerik
Kur Bilgisi	3	-	Nominal Kategorik
Ödeme Sistemi Kodu	9	-	Nominal Kategorik
Ödeme Sistemi	9	-	Nominal Kategorik
Kart Tipi Kodu	4	-	Nominal Kategorik
Kart Tipi	4	-	Nominal Kategorik
Alt Kart Tipi Kodu	8	-	Nominal Kategorik
Alt Kart Tipi	8	-	Nominal Kategorik
Kart Numarası	267.485.184	-	Nominal Kategorik
Kart No İlk 6 Rakam	64.484	8.926.834	Nominal Kategorik
Kart No Son 4 Rakam	10.173	8.926.834	Nominal Kategorik
ARN	23.112.712.192	-	Nominal Kategorik
İşyeri Kategori Kodu	915	-	Nominal Kategorik
İşyeri Kategori Grup Kodu	29	-	Nominal Kategorik
İşyeri Kategori Grup Kodu	29	-	Nominal Kategorik
Geri Ödeme Değeri	30	188	Nominal Kategorik
Yurtdışı işlem mi?	2	-	Nominal Kategorik
Not OnUs işlem mi?	2	-	Nominal Kategorik
ATM işlem mi?	2	-	Nominal Kategorik
Chipli işlem mi?	2	-	Nominal Kategorik
Marka Paylaşım işlemi mi?	2	-	Nominal Kategorik
Ödeme Tipi	3	188	Nominal Kategorik
MOTO / E-Ticaret Tipi	21	-	Nominal Kategorik
İşyeri Adı	258.490.368	1.971.577.225	Nominal Kategorik
İşyeri Şehri	15.558.656	2.150.843.634	Nominal Kategorik
İşyeri Ülke Kodu	240	95.679.528	Nominal Kategorik
İşyeri Ülkesi	240	95.679.528	Nominal Kategorik
Kullanım Kodu	3	1	Nominal Kategorik
Neden Kodu	169	-	Nominal Kategorik
DWH Neden Kodu	644	-	Nominal Kategorik
Elektronik Terminal kodu	7	188	Nominal Kategorik
Terminal Özellikleri	20	-	Nominal Kategorik
Kart Sahibini Belirleme Metot Kodu	4	188	Nominal Kategorik
POS Kategori Kodu	5	233.620	Nominal Kategorik
POS Kategori Seviyesi	5	233.620	Nominal Kategorik
ATM Hesap Seçim Özelliği	10	188	Nominal Kategorik
Çip Durum Özellikleri	15	188	Nominal Kategorik
EMV İşlem Tipi	261	188	Nominal Kategorik
Kart Giriş Tipi	24	-	Nominal Kategorik
Kart Şifre Giriş Tipi	26	188	Nominal Kategorik
Terminal Ülkesi	646	188	Nominal Kategorik
Terminal Seri Numarası	5.149.184	14.220.350.896	Nominal Kategorik
Dosya Oluşturma Tarihi	-	-	Aralıklı Nümerik

ortalamasını, n ise toplam örnek sayısını ifade eder.

$$Pearson_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{3.2}$$

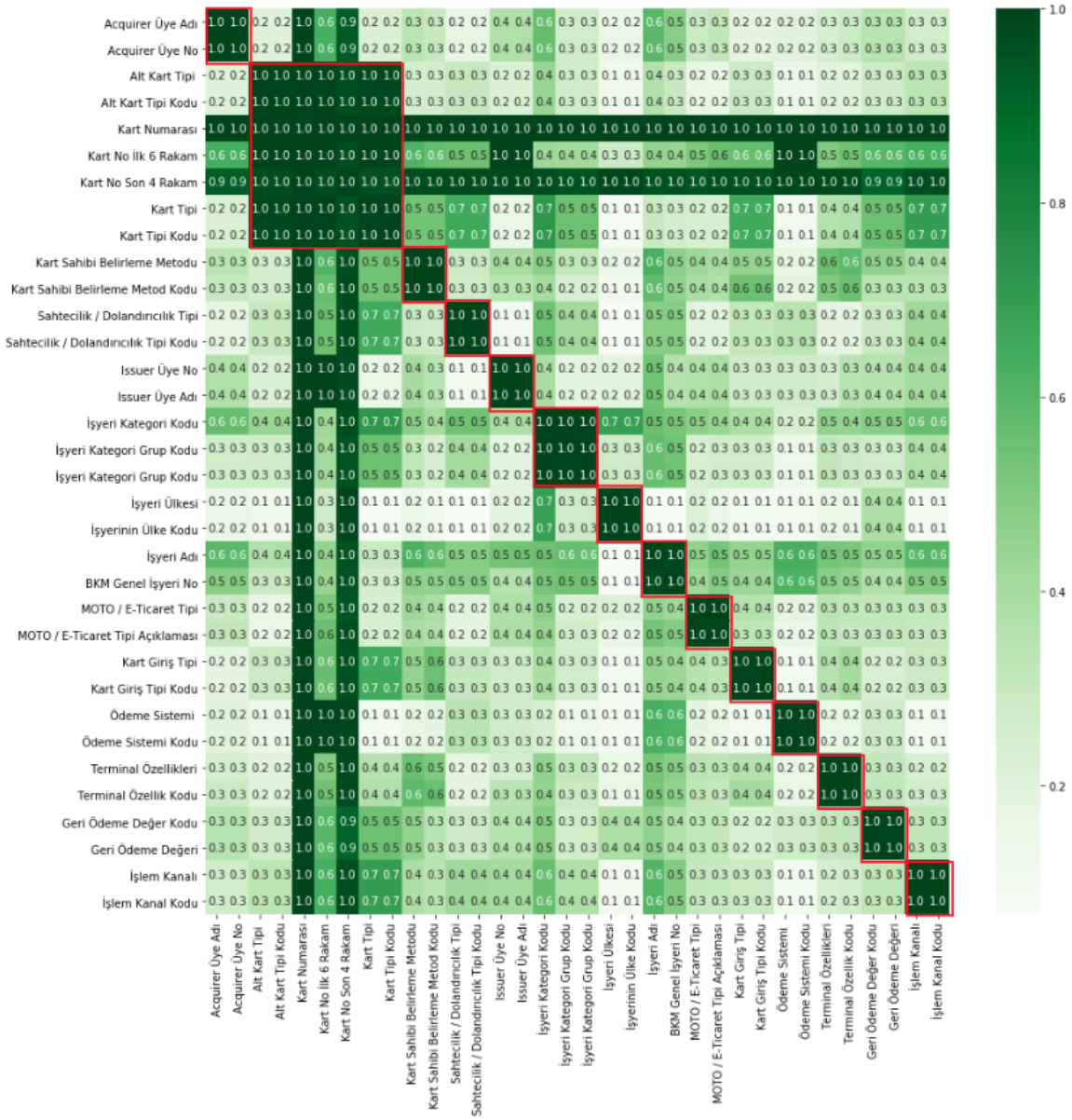
Tablo 3.3 Veri Kümelerinin Karşılaştırması

Kümeyi Sağlayan Kurum	Veri Kümesi Tipi	İçerik Zamanı	Özellik Sayısı	Toplam İşlem Sayısı	Toplam Sahte İşlem Sayısı
ATOS Worldline [18]	Gerçek	2015 Mart - 2015 Mayıs.	12	16,5 Milyon	117.250
A large European card processing company[19]	Gerçek	2012 Ocak - 2013 Haziran	27	120 Milyon	40
Machine Learning Group – ULB[7]	Çevrim içi	Eylül 2013	31	284.807	492
A commercial bank[20]	Gerçek	Bir Yıllık Veri	8	260 Milyon	4.000
National bank's credit card data ware house[4]	Gerçek	-	-	22,978 Mbilyon	978
BKM	Gerçek	2016 Ocak - 2019 Ocak	44	∼2.4 milyar	∼680.000

- Korelasyon oranı metriği [23]: Korelasyon değeri hesaplanacak özelliklerden birisi kategorik diğeri numerik ise, Korelasyon oranı metriği kullanılmıştır. Eşitlikte X ve Y rastsal değişkenlerdir. $Var(Y)$, Y'nin varyansıdır, $Var(Y|X)$ koşullu varyans, E de beklenen değer ya da diğer bir deyişle ortalamasıdır.

$$\eta^2_{Y|X} = 1 - \mathsf{E}\left[\frac{\mathsf{V}ar(Y \mid X)}{\mathsf{V}ar(Y)}\right] \tag{3.3}$$

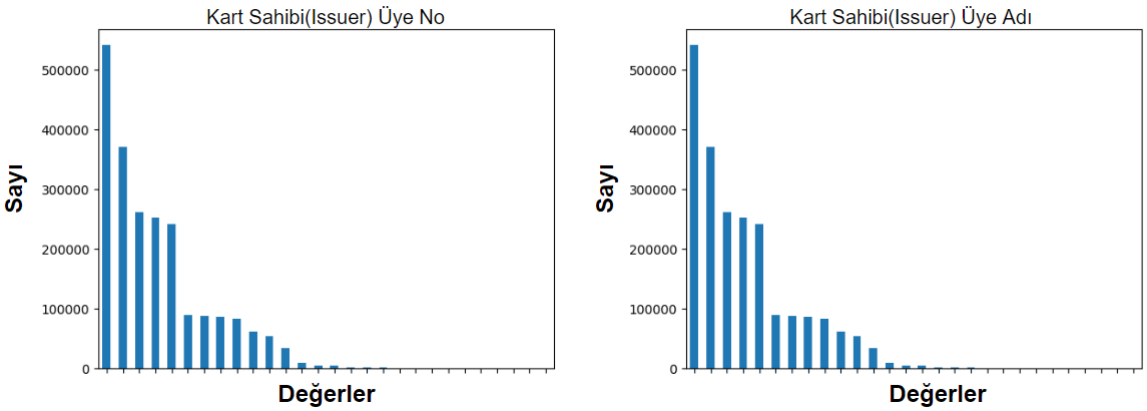
Elde edilen korelasyon matrisi, sahte işlem veri kümesinde bulunan 44 özellik için, 44x44 boyutunda büyük bir matris oluşturmaktadır. Korelasyon matrisini daha anlamlı ve görsel olarak daha kolay anlaşılır sunabilmek için, korelasyon matrisi iki görsel harita halinde organize edilmiştir. İlk korelasyon haritası, ikili korelasyonu 1 olan özellikleri içerecek şekilde, Şekil 3.2'de verilmiştir. Bu haritaya göre, kredi kartı numarasına ilişkin bilgileri taşıyan, "Kart Numarası", "Kart No İlk 6 Rakam", "Alt Kart Tipi" gibi özelliklerin korelasyon değerlerinin 1 olduğu görülmektedir. Ayrıca, "İş Yeri Kodu" ve "İş Yeri Adı" gibi özelliklerin de korelasyon değerlerinin 1 olduğu görülmektedir. Bu gibi korelasyon değerleri 1 olan özellik ikililerinin değer dağılımları da incelenmiştir. Bu dağılım incelemesine bir örnek, Şekil 3.3'te "Kart Sahibi Üye No" ve "Kart Sahibi Üye Adı" özellikleri için verilmiştir. Buna göre "Kart Sahibi Üye No" ve "Kart Sahibi Üye Adı" özellik ikilisinin aynı değer dağılımına sahip olduğu açıkça görülmektedir.



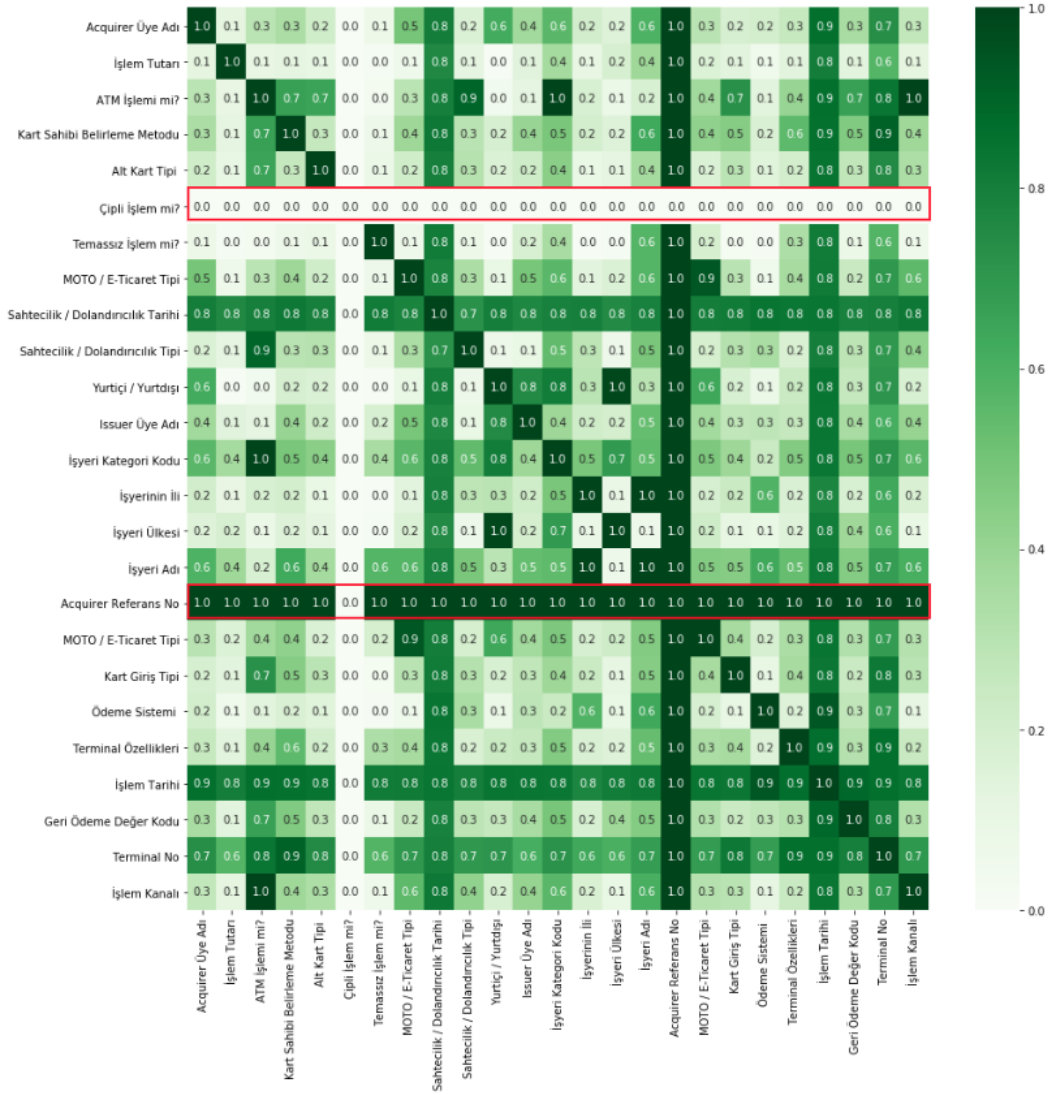
Şekil 3.2 İkili korelasyonu 1 olan özelliklerin grupları içeren korelasyon matrisi

Şekil 3.2'de verilen korelasyon değeri 1 olan özellik gruplarından sadece bir özelliği seçmek yeterlidir. Şekil 3.4'te verilen ikinci korelasyon haritası bu seçilen özelliklerin ve geri kalan bütün özelliklerin korelasyon haritasını içermektedir. Korelasyon matrisine göre "Çipli işlem mi" özelliğinin diğer bütün özelliklerle ilişkisiz olduğu gözlemlenmiştir. Bunun nedeni, Tablo 3.1'de gösterildiği üzere bu özelliğin çok fazla sayıda boş değer içermesidir. Ayrıca "Acquirer Referans No" özelliğinin ise bütün özelliklerle korelasyonunun 1 olduğu gözlemlenmiştir. Bunun nedeni ise, Tablo 3.1'de gösterildiği üzere, bu özelliğin çok fazla sayıda farklı değere sahip olmasıdır.

Veri analizi aşamasında elde edilen özelliklerin ikili ilişkileri, verilerdeki boş değer, farklı değer sayıları gibi veri istatistikleri, sonraki aşamalar olan veri temizliği ve veri



Şekil 3.3 Aynı dağılıma sahip olan Kart Sahibi Üye No ve Kart Sahibi Üye Adı özelliklerinin gösterimi



Şekil 3.4 İkili korelasyonu 1 olan özellik gruplarından seçilen özellikleri ve geri kalan diğer özellikleri içeren korelasyon matrisi

dönüşümleri aşamalarında kullanılmıştır.

3.3 Veri Ön Hazırlığı

Denetimli veya denetimsiz bütün makine öğrenme yöntemlerinin başarılı olması veri kalitesiyle doğrudan ilgilidir. Veri kümelerinde yanlış, eksik veya fazla girilen örnekler makine öğrenmesi yöntemlerini yanıltmaktadır. Bu yüzden, veri madenciliği, yapay zeka ve makine öğrenmesi çalışmalarında veri ön hazırlığı yapılması gereken bir aşamadır. Veri ön hazırlığı aşamasında veri temizliği ve veri dönüşümleri adımları gerçekleştirilmiştir.

3.3.1 Veri Temizliği

Kullanılan veri kümesi bankalar tarafından BKM'ye bildirilen işlem detaylarından oluşmaktadır. Bunun doğal bir sonucu olarak, farklı kaynaklardan bu verileri sağlayan bankalar verileri farklı biçimlerde iletilebilmektedir. Bu nedenle yapılan çalışmada, veri temizliğine duyulan ihtiyaç daha da önemli bir hale gelmiştir.

Özellik çıkarımı ve kümeleme yöntemlerinin başarılı sonuçlanması için, Tablo 3.1'de detayları verilen özellikler incelenerek aşağıda belirtilen veri temizliği işlemleri gerçekleştirilmiştir.

1. Tekrarlı veriler, veri kümesi kayıtlarında birden fazla aynı veya çok yakın değere sahip kaydın bulunması durumudur. Veri kümesinde yapılan inceleme sonucunda tekrarlı veri olmadığı görülmüştür.

2. Gereksiz veriler, aynı bilgiyi taşıyan veri özelliklerinin bir arada bulunması durumudur. Bu bağlamda, aralarında korelasyon olduğu tespit edilen özellikler veri kümesinden çıkarılmıştır. Bu özellik grupları, veri analizi bölümünde, Şekil 3.2'de işaretlenen "Kart Tipi Kodu"-"Kart Tipi", "Terminal Özellikleri" - "Terminal Özellik Kodu" gibi aynı bilgiyi taşıyan özellik gruplarıdır.

3. Yapısal hatalar, aynı bilgi için farklı kayıtların girilmiş olması durumudur. Genellikle büyük-küçük harf duyarlılığının olmadığı veya imla, söz dizimi kurallarının denetlenilmediği durumlarda gerçekleştiği görülmüştür. İşyerinin İli alanında, aynı il için farklı değerlerin olduğu gözlemlenmiştir. Örneğin, "İstanbul" ili için, farklı kaynaklardan veri alınmasından dolayı "ISTANBUL","İSTANBUL" gibi farklılıklar olduğu görülmüştür. Bu nedenle "İşyerinin İli", veri kümesinden çıkarılmıştır. Bu sorunu gidermek için bir harici lokalizasyon servisinden faydalanılmıştır. Bu servis üye iş yeri adı, açıklaması

gibi bilgileri kullanarak, bir arayüz üzerinden iş yerinin coğrafik konumuna erişir. Elde edilen coğrafik konum üzerinden, enlem, boylam, açık adres, posta kodu gibi çok daha fazla lokasyon bilgisine erişilebilmektedir.

4. Eksik/Boş veriler, verilerin düzgün kaydedilmemiş olması durumudur. Eksik/Boş veri özelliklerini gidermek için, "MOTO/E-ticaret Tipi", "Kart Giriş Tipi Kodu", "Ödeme Sistem Kodu" özellikleri için, her bir özelliğin veri kümesinde bulunan tüm değerlerin dağılımlarına bakılarak en sık gözlemlenen değer (özellik değerlerinin modu) ile boş değerler doldurularak töhmet (imputation) işlemi gerçekleştirilmiştir. Ayrıca "Terminal No" için çok fazla boş değere sahip veri kümesinden çıkarılmıştır. "Çipli İşlem mi ?" özelliği diğer özelliklerle ilişkisi olmayan ve farklı değer içermeyen bir özellik olduğu için bu bağlamda veri kümesinden çıkarılmıştır.

Yapılan incelemelerde, yurt dışı işlemler için üye iş yerlerine ait birçok kayıtta özelliklerin yanlış/eksik girildiği gözlemlenmiştir. Bu eksik/yanlış bilgiler yurtdışındaki iş yerinler ait iş yeri adı, iş yeri lokasyonu gibi birçok bilginin, BKM bünyesinde bir karşılığının olmamasından kaynaklanmaktadır. Bu nedenle yurt dışında gerçekleşen sahte işlemler veri kümesinden çıkarılmıştır. Dolayısıyla, bu çalışmada yapılan bütün deneyler ve sonuçları yurt içinde gerçekleşen işlemler için yapılmıştır.

3.3.2 Veri Dönüşümü

Veri dönüşümü aşamasında, özellik çıkarıcı modellerin eğitiminde kullanılacak olan verilerin uygun formata dönüştürülmesi sağlanmıştır. Ayrıca, veri dönüşümü ile bazı özellikler için örnekleme ve normalizasyon dönüşümleri gerçekleştirilmiştir.

Aralıklı nümerik özellik tipinde olan, "İşlem Tarihi", "Sahtecilik/Dolandırıcılık Tarihi" alanları için örnekleme yöntemi ile dönüşüm gerçekleştirilmiştir. Bu dönüşüm işlemi, tarih alanlarında verilen günün tam tarihi yerine, yılın haftasına dönüştürülmüştür. Örneğin, "13.12.2018" tarihi, yılın 50. haftası olduğu için, bu tarih için ilgili özelliğin değeri "50" olacak şekilde dönüştürülmüştür. Tarih alanlarının bu şekilde dönüşümü ile özel gün ve haftalarda yapılan harcama alışkanlıklarının ön plana çıkacağı öngörülmüştür.

Veri kümesindeki tek sürekli nümerik özellik tipinde olan "İşlem Miktarı" için, eşitlik 3.4'te formülü verilen Min-Max normalizasyonu uygulanmıştır. Eşitlikte mevcut değer x_i, en küçük değer $min(x)$, en yüksek değer $max(x)$ ve dönüştürülen yeni değer z_i

olarak ifade edilmektedir.

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{3.4}$$

Özellik çıkartma adımında kullanılan Özkodlayıcı model, sayısal değerler için eğitim gerçekleştirdiğinden dolayı, kategorik özelliklerin sayısal özelliklere dönüştürülmesi gerekmektedir. Bunu sağlamak için, veri ön hazırlığı aşamasında son olarak, etiket kodlama (*Label Encoding*) ve yapay kodlama (*dummy encoding*) uygulanmıştır.

3.3.2.1 Etiket Kodlama

Etiket kodlama, kategorik olan değerlerin sayısal olarak ifade etmek için kullanılan bir dönüşümdür. Şekil 3.5'te bir etiket kodlama örneği gösterilmiştir.

ID	Miktar	ATM	Ülke Kodu	...
1	119.99	Evet	TR	...
2	14.99	Hayır	US	...
3	0.99	Hayır	CH	...
4	100.00	Evet	JP	...
5	20.99	Hayır	TR	...
..	...	...	...	...

Etiket Kodlama

ID	Miktar	ATM	Ülke Kodu	...
1	119.99	0	0	...
2	14.99	1	1	...
3	0.99	1	2	...
4	100.00	0	3	...
5	20.99	1	0	...
..	...	...	...	...

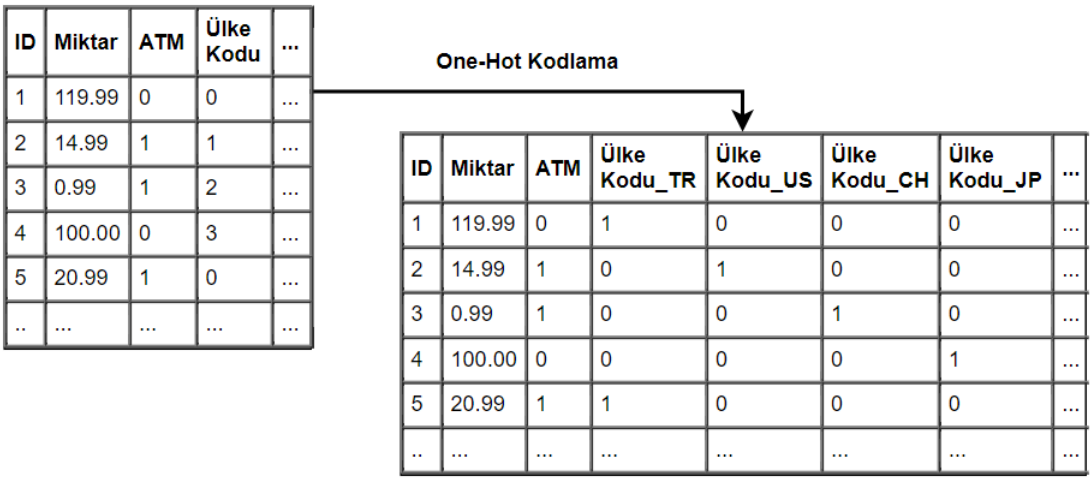
Şekil 3.5 Etiket Kodlama ile kategorik özelliklerin dönüştürülmesi

Etiket kodlama tekniği, kategorik bir değeri basitçe farklı bir skalar, sayısal değerle değiştirir. Etiket kodlama sonrasında bir özelliğin farklı kategorideki değerleri arasındaki sayısal mesafe farklı olduğundan sıralamasız (nominal) kategorik değerleri ifade etmek güçleşir. Bu durum, yapay sinir ağları gibi yöntemlerin eğitilmesinde model eğitiminde başarısız sonuçlara neden olacaktır. Bunun üstesinden gelebilmek için etiketleme kodlamadan sonra, yapay kodlama yapılması yaygın bir pratiktir.

3.3.2.2 Yapay kodlama

Yapay kodlama ya da diğer bir deyişle One-hot kodlama, kategorik olan değerlerin sayısal olarak ifade etmek için kullanılan bir dönüşümdür. Şekil 3.6'da bir yapay kodlama örneği gösterilmiştir.

Yapay kodlama, kategorik verilere sahip bir sütunu alır ve bu sütunda bulunan farklı kategori değerleri kadar sütun oluşturulur. Oluşturulan yeni sütunlar, kategorik



ID	Miktar	ATM	Ülke Kodu	...
1	119.99	0	0	...
2	14.99	1	1	...
3	0.99	1	2	...
4	100.00	0	3	...
5	20.99	1	0	...
..	...	...	...	...

ID	Miktar	ATM	Ülke Kodu_TR	Ülke Kodu_US	Ülke Kodu_CH	Ülke Kodu_JP	...
1	119.99	0	1	0	0	0	...
2	14.99	1	0	1	0	0	...
3	0.99	1	0	0	1	0	...
4	100.00	0	0	0	0	1	...
5	20.99	1	1	0	0	0	...
..	...	...	...	...	...	...	...

Şekil 3.6 Yapay Kodlama ile kategorik özelliklerin dönüştürülmesi

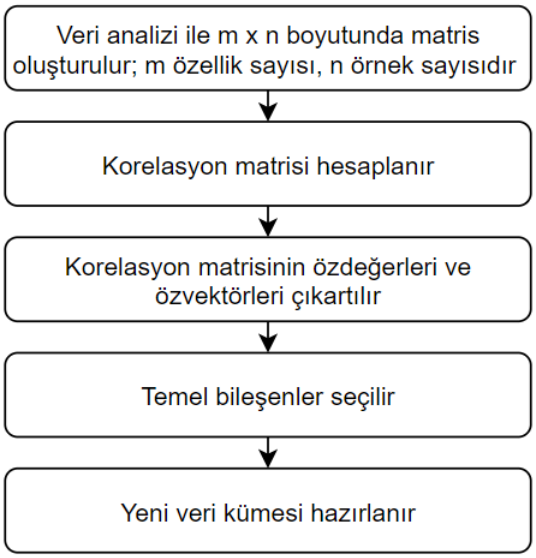
verilere göre ’1’ veya ’0’ değerlerini alır. Sırasıyla, kategorik veriye karşılık gelen ilgili sütun hücresine ’1’ değeri atanırken, aynı satırda bulunan diğer sütun hücrelerine ’0’ değeri atanır. Bu sayede farklı kategoriler arasındaki sayısal farklar eşit olur.

3.4 Özellik Çıkarımı

Aynı sınıftaki örnekler arasındaki benzerliğin fazla, farklı sınıflardaki örnekler arasındaki benzerliğin az olmasını sağlamak için karakteristik özelliklerin seçilmesi gerekir. Sahte kredi kartı işlemleri, birbiriyle ilişkili özellikler içermektedir. Özellik seçimi ve azaltma için, Temel Bileşen Analizi (PCA) ve Özkodlayıcı yöntemleri kullanılmıştır.

3.4.1 Temel Bileşen Analizi

Temel Bileşen Analizi [24], 1901 yılında Karl Pearson tarafından bulunmuştur ve Harold Hotelling tarafından 1930 yıllarında geliştirilmiştir. PCA, varyansı yüksek olan özellikleri ön plana çıkararak veri uzayını küçültmeyi sağlayan bir linear dönüşüm yöntemidir. PCA denetimsiz öğrenme yöntemidir, etiket verisine ihtiyaç duymaz. Ancak varyans değerlerine bağımlı olduğundan kategorik özellikler içeren veriler için uygun bir yöntem değildir. Temel bileşen analizinde gerçekleşen temel işlem adımları Şekil 3.7’de verilmiştir.



Şekil 3.7 Temel Bileşen Analizi adımları

Temel bileşen analizinde, çıkarılan özellikler popülasyon içerisinde sahip olduğu varyans ile doğrudan ilişkilidir. Bu nedenle, temel bileşen analizinde özellik seçimi için ihtiyaç duyulan özellik sayısını belirlemede özdeğer ve özdeğer vektörlerinin varyans değerleri incelenmektedir.

3.4.2 Özkodlayıcı

Özkodlayıcı [25], insan beynindeki nöronları model alarak geliştirilen perseptron adlı yapılardan oluşan çok katmanlı yapıya sahip bir yapay sinir ağ modeli türüdür. Yapay sinir ağları, denetimli öğrenme yöntemi olsa da özkodlayıcılarda sınıf bilgisi verinin kendisi olduğundan denetimsiz öğrenme yöntemi olarak da nitelendirilmektedir.

Özkodlayıcılarda etiket verisine ihtiyaç duyulmaması, eğitimin kolay olması gibi nedenlerden dolayı araştırmacıların dikkatini çekmiş ve görüntü işleme, doğal dil işleme, özellik çıkartma, anomali tespiti, gürültü temizleme gibi birçok alanda kullanım şekline özel farklı yapılarda autoencoder modelleri önerilmiştir [26–30].

Şekil 3.8'de görseli verilmiş olan özkodlayıcı model, kodlayıcı, darboğaz, çözümleyici ve tekrar oluşturma maliyet fonksiyonu olmak üzere 4 temel bileşenden oluşur.

1. Kodlayıcı: Giriş verisini (x) daha küçük boyuta sıkıştırır böylece verideki gürültüyü mümkün olduğu kadar azaltmaya çalışır. Bu bileşenin çıktısı genellikle darboğaz (z) olarak ifade edilir. Eşitlik 3.5'te darboğaz çıktısını

oluşturan aktivasyon fonksiyonu ifade edilmiştir.

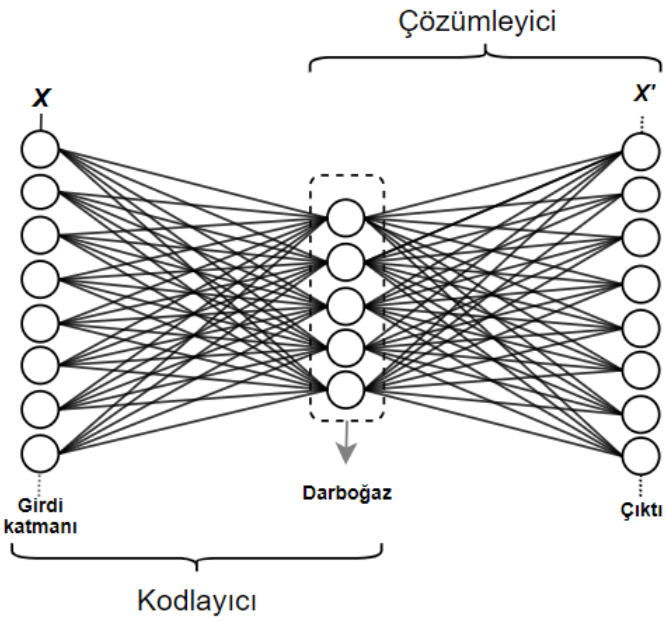
$$z = \sigma(Wx + b) \tag{3.5}$$

2. Darboğaz: Genellikle giriş katmanında verilen girdiden daha az sayıda özelliğe sahip olan bir ara katmandır. Bu katmanda eğitim esnasında oluşan gizli özellikler (latent features) çıkış verisini oluşturmada tekrar kullanılır.

3. Çözümleyici: Kodlanmış olan darboğaz çıktısını alır ve giriş verisine benzer bir veri oluşturmaya çalışan bileşendir. Bu bileşenin çıktısı ise tahmin çıktısı (x') olarak ifade edilir. Eşitlik 3.6'da darboğaz çıktısından giriş verisine yakın olan tahmin çıktısı olarak verilmiştir.

$$x' = \sigma(W'z + b) \tag{3.6}$$

4. Tekrar oluşturma maliyet fonksiyonu: bu fonksiyon, çözümleyici fonksiyonunun çıktısının giriş verisine ne kadar yakın olacağını belirleyen maliyet fonksiyonundur.



Şekil 3.8 Tek gizli katmana sahip özkodlayıcı model mimarisi

Maliyet fonksiyonu seçimi, ileri beslemeli yapay sinir ağ modellerinin eğitim başarısını etkileyen önemli parametrelerden birisidir. Verideki etiket bilgisi, model çıktısı ve veri kümesinin özellikleri maliyet fonksiyonu seçimini belirleyen önemli etkenlerdir. Model çıktısı olarak elde edilen değer $\hat{y}_i$ ve gerçek değer y_i olmak üzere yaygın olarak kullanılan maliyet fonksiyonları aşağıda verilmiştir:

- İkili Çapraz Entropi (Binary Cross-Entropy): Çıkış verisi 0 ve 1 gibi olasılıksal değer olan modeller için kullanılan bir metriktir. Modelin çıktısı olan tahmin ile gerçek değer arasındaki entropi hesaplamasıyla elde edilir. Binary Cross Entropy maliyet fonksiyonu, Denklem 3.7'de verilen formül ile hesaplanır.

$$\text{İkili Çapraz Entropi} = -(y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)) \tag{3.7}$$

- Ortalama Mutlak Hata (Mean Absolute Error): Model çıktısı olan tahmin değeri ile gerçek değer arasındaki mutlak farkların ortalaması ile hesaplanır. Maliyet fonksiyonu Denklem 3.8'de verilen formül ile hesaplanır.

$$\text{Ortalama Mutlak Hata} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{3.8}$$

- Ortalama Karesel Hata (Mean Squared Error): Model çıktısı olan tahmin değeri ile gerçek değer arasındaki farkların kareleri toplamının ortalaması ile hesaplanır. Maliyet fonksiyonu Denklem 3.9'da verilen formül ile hesaplanır.

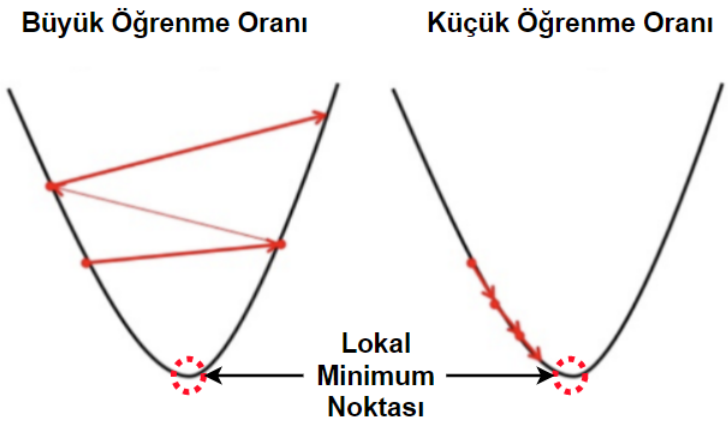
$$\text{Ortalama Karesel Hata} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} \tag{3.9}$$

- Ortalama Taraflılık Hatası (Mean Bias Error): Model çıktısı olan tahmin değeri ile gerçek değer arasındaki farkların, toplamının ortalaması ile hesaplanır. Maliyet fonksiyonu Denklem 3.10'da verilen formül ile hesaplanır.

$$\text{Ortalama Taraflılık Hatası} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{n} \tag{3.10}$$

Özkodlayıcı modellerde amaç, tahmin çıktısını giriş verisine yakınsatacak bir hipotez fonksiyonu eğitmektir. Bu amaçla, modelin kodlayıcı ve çözümleyici bileşenleri birlikte eğitilir. Eğitim, giriş (x) ve tahmin (x') değerleriyle oluşturulan eşitlik 3.5'te verilen maliyet fonksiyonunun minimalize edilmesi ile gerçekleşir. Bu maliyeti minimalize etme için eğitimde geri yayılım (back propagation) ve dereceli iniş yöntemleri [31] birlikte kullanılır. Dereceli iniş yöntemi, maliyet fonksiyonunun lokal minimum noktasına ulaşmayı hedefler. Bunu yaparken maliyet fonksiyonunun türevini alarak, minimum noktası bulunmaya çalışılır. Lokal minimum noktasına erişirken her geri yayılımda model ağırlıkları belli oranda güncellenir, bu güncelleme oranına öğrenme oranı denir. Öğrenme oranının büyüklüğüne göre Şekil 3.9'da verilen iki durum gerçekleşebilir:

Modelin başarılı bir şekilde eğitilmesini sağlayan bir diğer faktör ise, uygun optimizasyon yönteminin seçilmesidir. Optimizasyon yöntemleri, geri yayılım ile



Şekil 3.9 Öğrenme oranının çok büyük veya çok küçük seçilmesi durumunda oluşabilecek durumlar

öğrenme gerçekleştirilirken model parametrelerini güncelleme oranını (learning rate) belirlemek için kullanılır. Optimizasyon için Stokastik Gradyan İniş (Stochastic Gradient Descent,SGD) gibi lokal minimum noktasına yakınsamayı uzun sürede gerçekleştiren yöntemler yerine, Momentum yöntemleri ve ADAM, AdaGrad gibi öğrenme oranını adaptif hesaplayan yöntemler de önerilmiştir [32]. Bu çalışmada, yaygın olarak kullanılan ADAM [33] ve RMSProp [34] optimizasyon yöntemine yer verilmiştir:

- RMSProp: Adaptif optimizasyon yöntemlerinden birisi olan RMSProp, geçmiş gradyan bilgilerinin karelerinin üstel ağırlıklı ortalamalarını kullanır. Daha sonra, öğrenme oranını bu ortalamaya bölerek lokal minimum noktasına yakınsamayı hızlandırır. Denklem 3.11'de s_{dw} geçmiş gradyenlerin üstel ağırlıklı ortalaması, β belirlenmesi gereken bir hiperparametre, $\frac{\partial \mathscr{J}}{\partial W}$ ağırlıkları hesaplanan katmanın maliyet fonksiyonunun türevi, α öğrenme oranı, ε 0 ile bölmeyi engelleyecek çok küçük bir sayı, W katmandaki ağırlık olmak üzere RMSProp ile ağırlık güncelleme formülü verilmiştir.

$$\begin{aligned} s_{dW} &= \beta s_{dW} + (1-\beta)(\frac{\partial \mathscr{J}}{\partial W})^2 \\ W &= W - \alpha \frac{\frac{\partial \mathscr{J}}{\partial W}}{\sqrt{s_{dW}^{corrected}} + \varepsilon} \end{aligned} \tag{3.11}$$

- ADAM: Adaptive Moment Estimation (ADAM), RMSMProp ve Momentum yöntemlerini birlikte kullanarak daha hızlı yakınsamayı sağlayan bir yöntemdir. Buna göre, ilk olarak geçmiş gradyen bilgisinin üstel ağırlıklı ortalaması alınır. Daha sonra geçmiş gradyen bilgisinin karelesin üstel ağırlıklı ortalaması alınır.

Daha sonar her iki ortalama için de taraflılık doğrulama (bias correction) uygulanır. Son olarak hesaplaması yapılan katmanın ağırlıkları bu ortalamalar kullanılarak hesaplanır. Denklem 3.12'de v_{dw} geçmiş gradyenlerin üstel ağırlıklı ortalaması, s_{dw} geçmiş gradyenlerin karesinin üstel ağırlıklı ortalaması, β_1 ve β_2 hesaplanması gereken hiperparametreler, $\frac{\partial \mathcal{J}}{\partial W}$ ağırlıkları hesaplanan katmanın maliyet fonksiyonunun türevi, α öğrenme oranı, ε 0 ile bölmeyi engelleyecek çok küçük bir sayı, W katmandaki ağırlıklar olmak üzere ADAM ile ağırlık güncelleme formülü verilmiştir.

$$
\begin{aligned}
v_{dW} &= \beta_1 v_{dW} + (1-\beta_1)\frac{\partial \mathcal{J}}{\partial W} \\
s_{dW} &= \beta_2 s_{dW} + (1-\beta_2)(\frac{\partial \mathcal{J}}{\partial W})^2 \\
v_{dW}^{corrected} &= \frac{v_{dW}}{1-(\beta_1)^t} \\
s_{dW}^{corrected} &= \frac{s_{dW}}{1-(\beta_1)^t} \\
W &= W - \alpha \frac{v_{dW}^{corrected}}{\sqrt{s_{dW}^{corrected}} + \varepsilon}
\end{aligned}
\tag{3.12}
$$

Sahte işlem verisi üzerinde özellik çıkarma yöntemi olarak özkodlayıcı model seçilmiştir. Özkodlayıcı modeller, diğer özellik çıkarma yöntemlerinin aksine özellikler arasındaki linear olmayan bağıntıları da kullanarak yeni özellik uzayı oluşturabilir. Çalışmada kullanılan veri kümesinde, 681.682 sahte işlem bulunması özkodlayıcı modelin eğitimi için ihtiyaç duyulan veri miktarı için yeterli görülmüştür.

Eğitim sonrasında, kodlayıcı giriş verisini, daha az sayıda özellikle ifade edilebilecek bir darboğaz oluşturur. Çözümleyici ise bur darboğaz girdisi ile giriş verisini tekrar oluşturabilecek hale getirmeye çalışır. Nihayetinde kodlayıcı modelin darboğaz katmanında oluşturduğu özelliklerin kullanılarak giriş verisini tekrar oluşturulabilmesi nedeniyle giriş verisinin daha az boyutlu bir uzayda ifade edilmesi sağlanır.

3.5 Benzer Sahte İşlemlerin Kümelenmesi

Benzer sahte işlemleri gruplamak için kümeleme yönteminden faydalanılmıştır. Az hiperparametresi olması ve kolay eğitilebilmesi nedeniyle, sahte işlemleri kümelemek için K-ortalama [35] algoritması kullanılmıştır. K-ortalama veri etiketlerine ihtiyaç duymaz, bu yüzden etiket bilgisinin yetersiz olduğu problemlerde yaygın olarak kullanılan yöntemlerden biridir. K-ortalama kümeleme yöntemi, veri kümesini k adet kümeye bölen bir yöntemdir. K-ortalama yönteminde, her bir örnek x_j olmak üzere, $X = \{x_1, x_2, ..., x_n\}$ bir X veri kümesinde rastgele ilklendirilen k tane küme merkezi seçilir. Her bir küme merkezi v_i olmak üzere $V = \{v_1, v_2, ..., v_k\}$ ile verilen küme merkezleri V rastgele oluşturulur. K-ortalama yönteminde eğitim, eşitlik 3.13'te verilen maliyet fonksiyonunu minimalize eden küme merkezlerinin belirlenmesi ve örneklerin küme merkezlerine yakınlıklarına göre sınıflandırılmasıyla gerçekleşir.

$$J(V) = \sum_{i=1}^{k} \sum_{j=1}^{k_i} (\left\| x_j - v_i \right\|)^2 \tag{3.13}$$

Maliyet fonksiyonunu minimalize etmek için aşağıdaki işlemler, örneklerin atandığı kümelerde değişiklik olmayana kadar tekrarlanır:

1. Her bir veri örneği x_j, k küme merkezinden en yakın olan küme merkezine atanır.
2. Küme merkezleri için, küme içerisindeki örneklerin ortalaması baz alınarak yeni küme merkezi belirlenir. Eşitlik 3.14'te yeni küme merkezlerinin belirleyen denklem verilmiştir.

$$v_i = \frac{1}{k_i} \sum_{j=1}^{k_i} x_i \tag{3.14}$$

Algoritma 3.1'de k-ortalama yönteminin sözde kodu verilmiştir.

Algoritma 3.1 K-ortalama Algoritması

Girdiler:

Veri kümesi : $X = \{x_1, x_2, ..., x_n\}$,

Küme sayısı : k

Çıktılar:

Küme etiketleri : $L = \{l(x) \mid x \in X\}$

Algoritma:

1: **for each** $v_i \in V$ **do**
2: $v_i \leftarrow e_j \in E$ (Küme Merkezlerinin İlklendirilmesi)
3: **end for**
4: **for each** $v_i \in V$ **do**
5: $l(x_i) \leftarrow argminDistance(x_i, v_j) \in \{1...k\}$
6: **end for**
7: $changed \leftarrow false$
8: **repeat**
9: **for each** $v_i \in V$ **do**
10: $updateClusters(x_i)$
11: **end for**
12: **for each** $x_i \in X$ **do**
13: $minDist \leftarrow argminDistance(x_i, v_j) j \leftarrow \{1...k\}$
14: **if** $minDist \neq l(x_i)$ **then**
15: $l(x_i) \leftarrow minDist$
16: $changed \leftarrow true$
17: **end if**
18: **end for**
19: **until** changed = true

Veri örnekleri küme merkezine atanırken, numerik özellikler için *öklid mesafesi*, kategorik değerler için ise *Hamming uzaklığı* hesaplanır. Bu nedenle kategorik değerlerin yeni küme merkezini bulmak için ortalama yerine mod kullanılır. K-ortalama yöntemi numerik özellikler üzerinde ortalama alınarak hesaplanır. K-ortalama yönteminin diğer türevlerinden olan K-modes yöntemi sadece kategorik özellikler üzerinde çalışırken, bir diğer türevi olan K-Prototypes hem numerik, hem kategorik özellikler üzerinde çalışabilir [36].

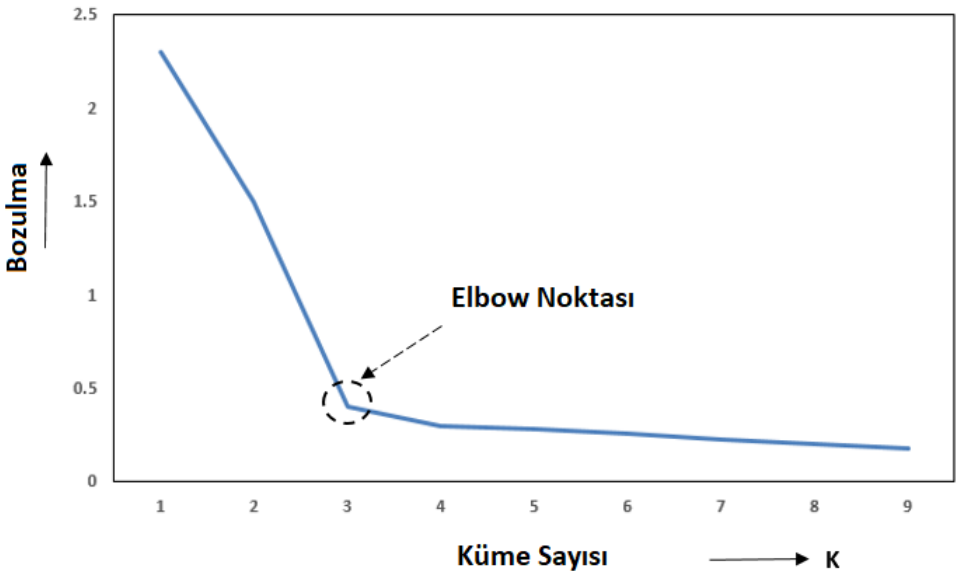
K-Means kümeleme yöntemi, denetimsiz öğrenme yöntemidir. Bu yüzden, kümeleme esnasında sınıf/küme bilgisi bulunmamaktadır. Verileri etkin bir k sayısı ile kümelemek için farklı k sayıları ile denemeler yapılarak kümeleme başarıları ölçülür. Bu başarı ölçümü yöntemlerinden birisi de *dirsek (elbow)* yöntemidir. Buna göre, En

uygun k sayısını bulabilmek için, Şekil 3.10'da bir örneği gösterilen *dirsek* yöntemi uygulanmıştır.

Dirsek yöntem ile, farklı küme sayıları için örneklerin küme merkezlerine olan uzaklıklarının karelerinin toplamı hesaplanarak bir bozulma (distortion) değeri elde edilir. Denklem 3.15'te bozulmanın hesaplanma şekli ifade edilmiştir. Denklemde; x veri noktalarını temsil ederken, c_i küme merkezlerini temsil eder.

Küme merkezi sayısının arttırılmasının doğal bir sonucu olarak bu bozulma değeri azalacaktır. Şekil 3.10'da farklı küme sayıları için bozulma oranlarını gösteren dirsek yöntemi grafiği gösterilmiştir. Buna göre en uygun küme sayısı, grafiğin kırılma noktasındadır.

$$\text{Bozulma} = \sum_{i=1}^{k} \sum_{x \epsilon C_i} \text{Mesafe}(c_i, x)^2 \tag{3.15}$$



Şekil 3.10 Dirsek (Elbow) yöntemi

Giriş bölümünde bahsedildiği gibi sahte işlemlerin sebepleri farklı olabilir. Sahte işlemler tespit edildiğinde, bu sahte işlemin kaynağını tespit etmek oldukça zordur. Kümeleme ile elde edilmek istenen bir diğer amaç ise, kart kopyalama kaynaklı sahte işlem olan ancak sahte işlem nedeni yanlış belirlenen sahte işlemleri de aynı küme içinde değerlendirmektir.

Kümeleme işlemi yapıldıktan sonra küme seçimi yapılmıştır. Kümelerdeki sahte işlemlerin nedenleri kopyalanmış kart, kayıp veya çalıntı kart, sahte başvuru, kart numarasının sahtekarlık amaçlı kullanılması gibi bir çok nedenden dolayı gerçekleşmiş olabilir. Sahte işlem tipi kart kopyalama olan sahte işlemlerin, diğer

sahte işlem tiplerinden dolayı oluşmuş gibi bildirildiği durumlar da mümkündür. Bu yüzden, kümeleme işlemi, bütün sahte işlemler üzerinde uygulanmakta ve kümeleme işleminden sonra kopyalama kaynaklı işlemlerin yoğun olduğu kümeler seçilmektedir. Burada hangi kümelerin seçileceğini belirlemek için kullanılacak olan kopyalama nedenli sahte işlem yoğunluk oranı (d) deneysel olarak belirlenmiştir.

3.6 İşlem Geçmişi Analizi

Kart kopyalayan suçlular, kopyaladıkları kartlarla sahtekarlık işlemleri gerçekleştirirler. Aynı kopyalama işlemine karışan kartlar, daha sonra sahte işlemlerde kullanılırlar. Bu sahte işlemler zaman, yer, miktar vb. gibi yönlerden benzerlik gösterebilir. Önceki adımlarda gerçekleştirilen özellik çıkarımı ile işlemlerin benzerliğini gösterebilecek olan önemli özellikler elde edilmiş, kümeleme yöntemi ile ise bu özelliklere göre benzerlik gösteren işlemler kümelenmiştir.

Kümeleme sonucunda elde edilen kart grupları, sadece kopyalama nedenli işlemleri değil, kart başvuru dolandırıcılığı, kayıp/çalıntı kart gibi diğer nedenlerden kaynaklanan sahte işlemleri de içermektedir. Dolayısıyla kopyalama sebebiyle gerçekleşmiş kart kümelerinin seçilmesi gerekmektedir. Bu yüzden küme seçimi işlemi yapılmalıdır.

Bir sahte işlem gerçekleştiğinde, kart kullanıcısının geri bildirimi olmadan, bu sahte işlemin nedenini tahmin etmek oldukça güçtür. Bankalar kopyalama merkezlerini tespit edebilmek için sahte işlemleri inceleyerek, kopyalama nedenli gerçekleşen sahte işlemlerin bir kısmını belirleyebilirler. Küme seçimi işleminde, bankaların "kopyalama nedenli gerçekleşen sahte işlem" olarak belirlediği işlemler küme seçiminde kullanılmıştır. Sahte işlemin nedenini belirleyen bu bilgi, veri kümesinde "Sahtecilik Tip Kodu" özelliğinde yer almaktadır. Sahtecilik Tip Kodu ile ilgili özellik değer şeması, Tablo 3.4'te verilmiştir. Tablo 3.4'te de gösterildiği üzere sahtecilik tip kodu "4" olan sahte işlemler, kopyalama nedenli gerçekleştiği bilinen sahte işlemleri göstermektedir.

Hangi kümedeki işlemlerin kopyalama kaynaklı gerçekleşen sahte işlemleri içerdiğini tespit edebilmek için, kümelerdeki kopyalama nedenli olduğu bilinen sahte işlemlerin küme içerisindeki yoğunluğunu dikkate alarak karar verilmiştir. Çünkü, nedeni kopyalama olarak belirlenmiş sahte işlemler, kart kullanıcılarının ve bankaların geri bildirimleri ile kesinleştiği için bu işlemlerin kopyalama sebebiyle gerçekleştiği bilgisi güvenilirdir. Öte yandan, kopyalama nedenli olup, diğer nedenli işaretlenmiş sahte işlemlerin de olmaması olasıdır. Kümeleme yönteminin bir diğer sağladığı avantaj ise kopyalanma sonucu olup yanlış etiketlenmiş bu sahte işlemlerin de dikkate alınmasını

Tablo 3.4 Sahticilik Tip Kodu ve Sahtecilik Tipleri

Sahtecilik Tip Kodu	**Sahtecilik Tipi**
0	Kayıp kart
1	Çalıntı Kart
2	Hamiline Ulaşmamış Kart
3	Sahte Başvuru
4	Sahte kart ve manyetik şerit kopyalama
6	Kart numarasının Sahtekarlık amaçlı kullanımı

sağlamaktır.

Kümeleme işlemi sonucunda, kopya sahte işlemler benzer işlemler oldukları için bazı kümelerde yoğunlaşacaktır. Kopya sahte işlem yoğunluğu yüksek olan bu kümelerdeki kredi kartlarının, geçmiş işlemlerinde, ortak bir iş yerinde kopyalanmış olma ihtimali yüksektir. İşlem yoğunluğu yüksek olan kümeleri seçmek için bir yoğunluk eşik değeri d deneysel olarak belirlenir. Kümedeki sahte işlem tipi kopyalama olan (sahtecilik tip kodu=4 olan) işlem sayısının (n_k), bütün işlem sayısına (n) oranı, belirlenen eşik değerinden (d) yüksek ise küme seçimi Eşitlik 3.16'da verildiği gibi gerçekleşir.

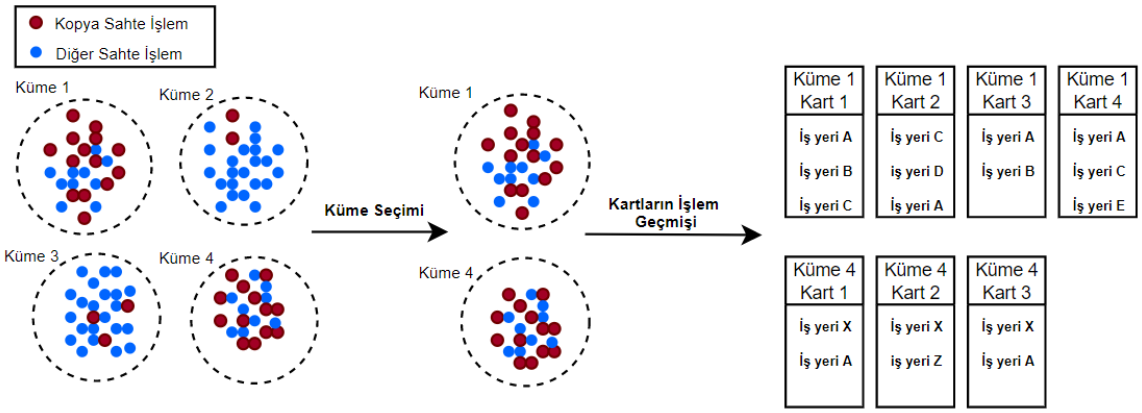
$$\text{Küme Seçimi } = \begin{cases} \text{Seçilir, } d < \frac{n_k}{n} \\ \text{Seçilmez, diğer} \end{cases} \tag{3.16}$$

Seçilen kümelerdeki sahte işlemlerde kullanılan kredi kartlarının kopyalandığı iş yerlerini bulabilmek için geçmiş işlemleri incelenmiştir. Kopyalanan kredi kartlarının kopyalandıktan ne kadar süre sonra sahte işlemlerde kullanılacağı kart bilgisini ele geçiren dolandırıcıya bağlıdır. Bu yüzden, kredi kartlarının geçmişini incelemek için ne kadar geçmişe gidileceğini belirlemek ancak deneysel olarak tahmin edilebilir.

Şekil 3.11'da, kümeleme işlemi sonrası küme seçimini ve kümelerdeki kartların geçmiş işlemlerinin belirlendiği bir örnek akış şeması verilmiştir. Buna göre kopya sahte işlem yoğunluğu, yoğunluk eşik değerinden yüksek olan "küme 1" ve "küme 4" seçilmiştir. Sonrasında bu iki kümedeki işlemlerin yapıldığı kredi kartlarının geçmiş işlemlerinin listelenmesi gösterilmiştir.

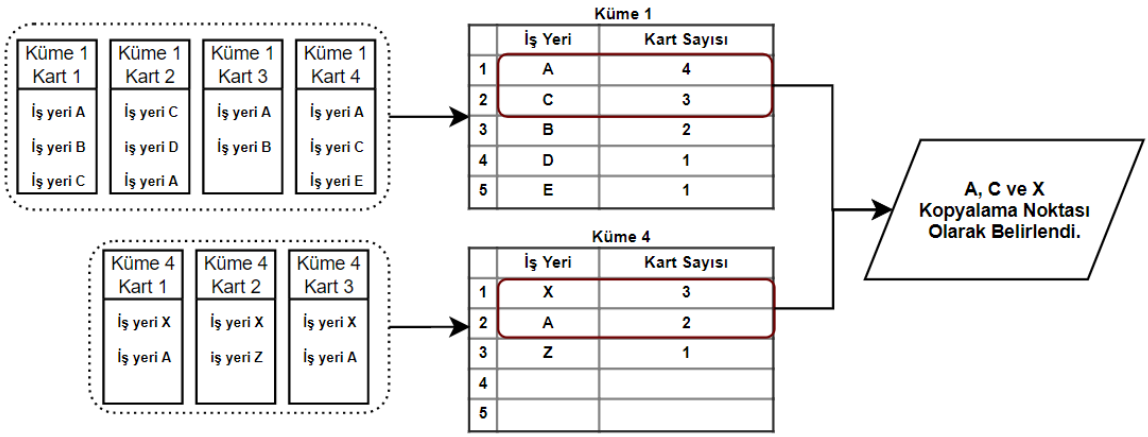
Seçilen her bir küme için kartların geçmiş harcamalarının yapıldığı iş yerleri çıkarılarak listelenmiştir. Küme bazında değerlendirilen bu iş yerlerini bazılarında, kartların ortak bir harcama yapmış ve bu harcama esnasında da kopyalama işlemi yapıldığı varsayılarak, kopyalama tahmini yapılması hedeflenmiştir.

Şekil 3.12'de küme bazında verilmiş olan kartların kullanıldığı iş yerleri listesinden,



Şekil 3.11 Küme seçimi ve kümelerdeki sahte işlemlerin gerçekleştiği kartların geçmiş işlemlerinin listelenmesi

iş yerinde geçen kartların sayısını veren listeye geçiş ve bu listeden seçilen kopyalama merkezi adaylarının seçilmesi işlemlerinin akış diyagramı verilmiştir. Buna göre, "Küme 1" için "A,C,B,D ve E" işyerleri için sırasıyla "4,3,2,1,1" adet kümedeki kartlara ait işlem gerçekleşmişken, "Küme 4" için bu değerler, "X,A,Z" iş yerleri için sırasıyla "3,2,1" olacak şekilde belirlenmiştir. Bir sonraki adımda, bu belirlenen iş yerleri için kümelerden kaç aday çıkarılacağıdır. Bu aday seçim sayısı (N), verilen şekildeki gösterime göre N=2 olarak ifade edilmiştir. Son olarak, her kümeden belirlenmiş olan adaylar programın çıktısını oluşturmaktadır.



Şekil 3.12 Geçmişi listelenmiş kartlardan üye iş yerlerinde geçen kart sayısının belirlenmesi ve olası kopya merkezi adaylarının tespit edilmesi

İşlem geçmişi analizi için geliştirilen Algoritma 3.2'de sözde kodu verilmiştir. Verilen sözde kodda, algoritmanın girdileri, parametreleri, çıktıları ve yorum satırları ile birlikte işlem adımları gösterilmiştir. Algoritmada kullanılan izlenen diğer metotlar Fonksiyon 3.1 ve Fonksiyon 3.2'de açıklanmıştır.

Küme seçimi için baz alınacak, kopyalama nedenli gerçekleşmiş sahte işlem yoğunluğunu 0 ile 1 aralığında ifade eden d parametresi ve iş yeri-kart sayısı sıralı listesinden seçilecek olan kopyalama merkezi aday sayısını ifade eden N parametresi önerdiğimiz yöntemin hiperparametreleridir. Bu hiperparametreler için deneylerde kullanılan değerler, deneysel sonuçlar bölümünde verilmiştir.

Algoritma 3.2 İşlem Geçmişi Analizi Algoritması

Girdiler:

Küme sayısı : k

Sahte İşlem kümeleri : $C = \{c_1, c_2, ..., c_k\}$

Çıktılar:

Kopyalama merkezi adayları : $candidates(list)$

Parametreler:

$t\_months(integer)$: geriye kaç ay gidileceğini belirleyen parametre

$d(float)$: sahte işlem tipi kopyalama olan işlemlerin (tip 4) yoğun olduğu kümeleri seçmek için kullanılan eşik değeri

$N(integer)$: Belirlenen sıralanmış işyeri listesinden ilk kaç adayın seçileceğini belirten parametre

Algoritma:

for each $c_i \in C$ **do**
    $copyFraudRatio \leftarrow getType4Ratios(c_i)$
    if $copyFraudRatio \geq d$ **then**
        $cardNos \leftarrow getDistinctCardNos(c_i)$
        $dict \leftarrow getCommonMerchantScores(cardNos, t\_months)$ ▷ Fonksiyon 3.1'de devamı.
        $merchantList \leftarrow sortAndSelectMerchants(dict, N)$ ▷ Fonksiyon 3.2'te devamı
        $candidateList.append(merchantList)$
    end if
end for
$candidates = getDistinctCandidates(candidateList)$

Fonksiyon 3.1 Ortak İş Yerleri Skorlarının Getirilmesi

Fonksiyon: getCommonMerchantScores($cardNos, t\_months$**)**

Girdiler:

Kümedeki farklı kart sayısı: n

Kümedeki kart numaraları: $k = \{k_1, k_2, ..., k_n\}$

İşlem geçmişi için seçilen zaman aralığı: t_months

Çıktılar:

İş Yeri Skorları: $dict$

Adım 1 – İş Yerlerinin Skorlarının Getirilmesi

```
1: function getMerchantScore(merchantNo)
2:     score = getMerchantScoreFromDB(merchantNo)        ▷ iş yeri skoru getirilir
3:     return - score        ▷ güvenilir olmayan iş yerlerinin skorları daha düşüktür
4: end function
```

Adım 2 – İşyerlerinde Ortak Geçen Kart Sayısının ve İş Yeri Skorlarının Oluşturulması

```
5: dict = dict()                                  ▷ sözlük değişkeni oluşturulmaktadır
6: for i to n do
7:     M_i = getMerchantListFromDB(k_i,t_months)        ▷ M_i, i. kartın son t_months ayda
   kullanıldığı iş yerleri listesi
8:     for each m in M_i do
9:         if m in merchantDict then
10:            dict[m].counter ← dict[m].count + 1
11:        else
12:            dict[m].counter ← 1
13:            dict[m].score ← getMerchantScore(m)
14:        end if        ▷ sözlük, ortak kart sayısını ve skor bilgisini tutmaktadır
15:    end for
16: end for
```

Fonksiyon 3.2 İş Yerlerinin Sıralanması ve Seçilmesi

Fonksiyon: sortAndSelectMerchants($dict, N$**)**

Girdiler:

İş yeri skorları: $dict$

Sıralanmış iş yerlerinden seçilecek olan aday sayısı: N

Çıktılar:

Seçilmiş İş Yerleri: $merchantList$

Adım 1 – İş Yerlerinin Sıralanması

```
1: function comparator(item1, item2)
2:     if item1.counter > item2.counter return true
3:     if item1.counter < item2.counter return false
4:     if item1.score > item2.score return true
5:     if item1.score < item2.score return false
6: end function
7: repeat
8:     swapped ← false
9:     for each i in 1 to length(dict) − 1 do
10:        if comparator(dict[i − 1], dict[i]) then
11:            swap(dict[i − 1], dict[i])
12:            swapped ← true
13:        end if
14:    end for
15: until swapped
```

Adım 2 – İlk "N" Kopyalama Adayının Seçilmesi

```
16: merchantList ← dict.first(N).keys    ▷ Sözlükteki Sıralanmış İlk N İş Yerinin Seçilmesi
```

4
DENEYSEL SONUÇLAR

Bu bölümde, deneylerin yapıldığı ortama ait teknik detaylar ve deney sonuçlarına yer verilmiştir. Deney kapsamında, öncelikle kart kopyalama noktalarını tespit edebilmek için, benzer sahte işlemler üzerinde kümeleme yapılmıştır. Sonra her kümedeki sahte işlemlerde geçen kartların geçmiş ortak işlemleri analiz edilmiştir. Bu analizin sonucunda, kart kopyalama noktası adayları çıktısı oluşturulmuştur. Bu amaçla, uygun kümeleme yöntemleri seçilmiş ve deneyler yapılmıştır.

4.1 Deney Ortamı

Kart kopyalama noktalarını tespit edebilmek için, donanımsal destekten faydalanılmıştır. Özellik çıkarımı için kullanılan özkodlayıcı modelin eğitilmesi için, yüksek işlemci gücüne ihtiyaç duyulmuştur. İşlem geçmişi analizini gerçekleştirebilmek için ise, milyarlarca kayıt içerisinde sorgu gerçekleştirebilmek için güçlü bir veri tabanı mimarisi ve donanımına ihtiyaç duyulmuştur.

İhtiyaç duyulan işlemci gücü desteğini sağlamak için IBM PowerAI kullanılmıştır. Genellikle büyük veriler üzerinde yapılan derin öğrenme çalışmalarında kullanılmak için tasarlanmış olan IBM PowerAI'a ait bazı önemli özellikler Tablo 4.1'de verilmiştir.

Tablo 4.1 Çalışmada kullanılan sistemin özellikleri

Sistem Özellikleri	
Çekirdek Sayısı	160 (5x32)
Çekirdek Frekansı	3.6 Ghz
Hafıza(RAM)	1 TB
Disk	500 GB
Grafik İşlemcisi(GPU)	NVIDIA Tesla V100
İşletim sistemi	Linux - power pc 64
Sanallaştırma Ortamı	Docker

Çalışma Docker sanallaştırma ortamındaki keras, scikit-learn gibi makine öğrenmesi kütüphaneleri kullanılarak gerçekleştirilmiştir. İhtiyaç duyulan veri depolama ve

sorgulama gücü, Oracle veri ambarı bulunan farklı bir sunucudan karşılanmıştır.

4.2 Değerlendirme Ölçütleri

Kullanılan veri kümeleri, bankalar arası gerçekleşen finansal işlemleri ve sahte işlem bildirimlerini içermektedir. Yapılan deney sonuçlarının değerlendirilebilmesi için, gerçek kart kopyalama noktası olduğu bilinen üye iş yerlerine ihtiyaç duyulmuştur. Bunun için, kart kopyalama noktası olduğu bilinen 17 kart kopyalama noktası bir özel bankadan tarafından temin edilmiştir. Bu bildirilen 17 kart kopyalama noktasının doğruluğu kesin olmakla birlikte, sadece bir banka tarafından bildirilen bu 17 kart kopyalama merkezinden çok daha fazlasının da olduğu göz önünde bulundurulmalıdır. Budan ötürü, yöntem başarılarını ifade edebilmek için farklı başarı ölçütleri kullanılmıştır.

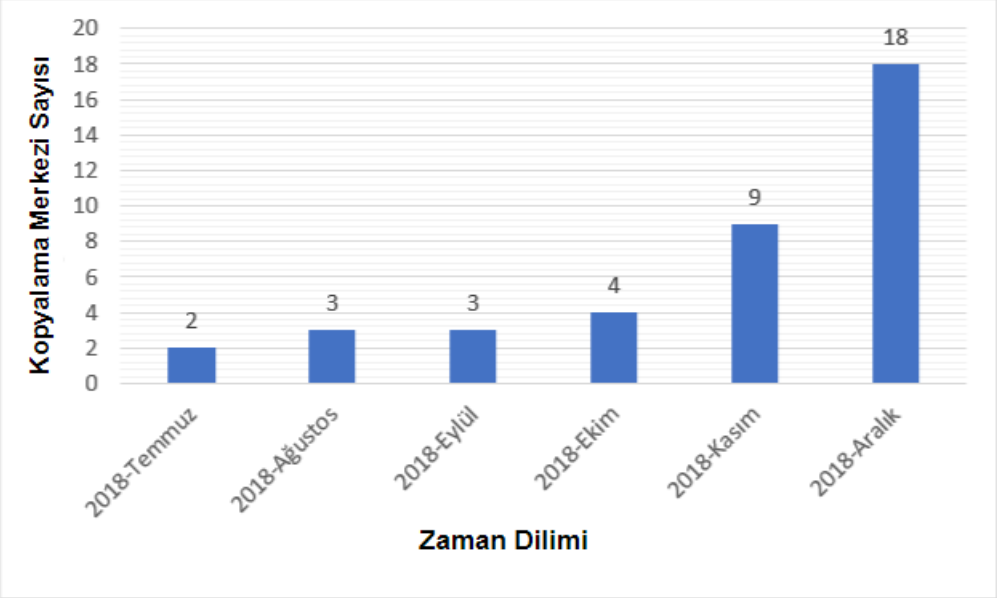
Önerilen yöntem, kart kopyalama noktası adayları önermektedir. Bu önerilen kart kopyalama noktalarından kaç tanesinin doğru olduğu ve gerçek kart kopyalama noktalarından kaç tanesinin doğru tahmin edilebildiği bilgileri önemlidir.

Yapılan deneylerde uygulanan farklı yöntem ve paramatreler için elde edilen sonuçlar değerlendirilirken, ulaşılabilen en fazla doğru tahmin sayısına bakılmıştır.

4.3 Deneyler

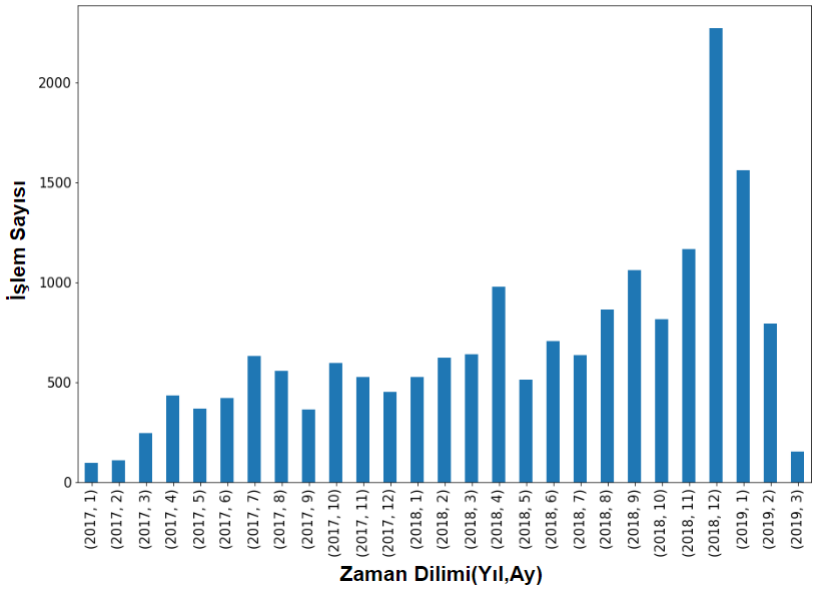
Önerilen modelin başarısını değerlendirmek için kart kopyalama noktası olduğu bilinen etiketlenmiş kopyalama merkezlerinin en yoğun olduğu aylar üzerinde çalışılmıştır. Bir özel banka tarafından bildirilmiş 2018 haziran ayından 2019 ocak ayına kadar farklı zamanlarda tespit edilmiş toplam 39 kart kopyalama noktası model başarısını ölçmek için kullanılmıştır. Şekil 4.1'de başarı ölçümü için verilen 39 kart kopyalama noktasının aylara göre dağılımı verilmiştir. Bu bildirilen kopyalama noktalarının bu aylardaki dağılımlarının düzensiz olması ve testlerin uzun sürmesi sebebiyle, yapılan deneyler kopyalama noktalarının en çok bildirildiği zaman dilimleri için gerçekleştirilmiştir. Bu amaçla, deney için kullanılacak olan zaman dilimi 2018 Aralık ayı olarak belirlenmiştir. Bu ay içerisinde gerçekleşen yurt içi sahte işlemlerden oluşan bir alt veri kümesi deneylerde kullanılmıştır. Bu veri kümesinde yaklaşık olarak 22.000 sahte işlem bulunmaktadır. Bu 22.000 sahte işlemde, birbirinden farklı yaklaşık 12.000 kredi kartı kullanılmıştır. Buna ek olarak, bu 12.000 kredi kartı Aralık ayından itibaren geçmiş 3 aylık zaman diliminde toplam 114.714 farklı üye iş yerinde kullanılmıştır.

Referans olarak kullanılan 39 kart kopyalama noktasında kopyalanan kredi kartlarının



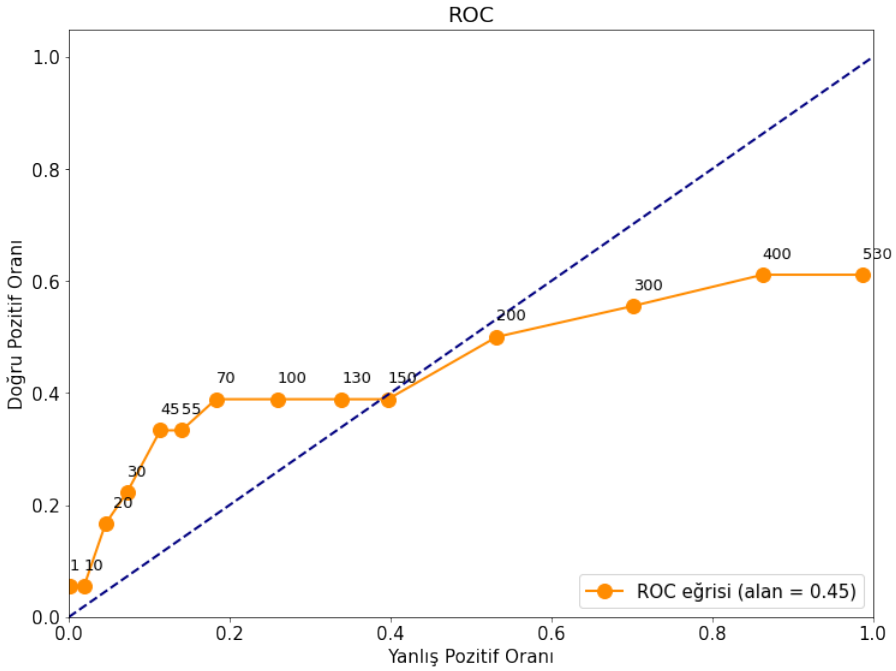
Şekil 4.1 Referans olarak kullanılan kopyalama noktalarının aylık dağılımı

kullanıldıkları sahte işlem sayılarının aylara göre dağılımı 4.2'de verilmiştir. Buna göre kopyalanan kartların, 10-13 ay boyunca kullanıldığı görülmektedir. Ayrıca sahte işlem sayılarının, kopyalama noktasının tespit edildiği zamandan öncesine gidildikçe azaldığı gözlemlenmiştir. Bu dağılım ve kullanılan işlemci gücü göz önüne alınarak, geçmişe yönelik yapılan işlem geçmişi analizi için kullanılacak olan zaman aralığı, 3 ay olarak belirlenmiştir.



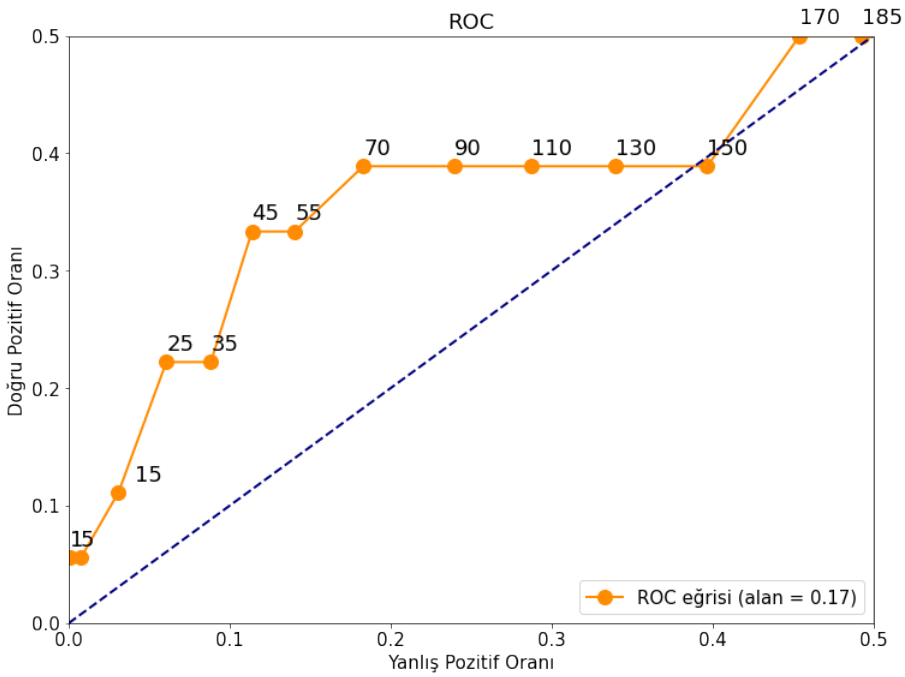
Şekil 4.2 2018 ekim - 2018 aralık aylarında, bankalar tarafından tespit edilen kart kopyalama noktalarından geçen kredi kartlarına ait bildirilen sahte işlem sayısının zamana göre dağılımı

Geriye yönelik analiz adımından sonra, kopyalama merkezleri adayları içerisinden seçilecek olan, maksimum aday sayısı olan n parametresi için en uygun belirleyebilmek için, ROC eğrisinden faydalanılmıştır. Şekil 4.3'te önerilen yöntem için elde edilen ROC eğrisine göre, n sayısı 150'yi geçtikten sonra yanlış pozitif oranı yükselmektedir.



Şekil 4.3 Maksimum aday sayısı parametresini belirlemek için kullanılan ROC eğrisi

Grafikte verilen ROC eğrisinde doğru pozitif oranının yanlış pozitif oranına göre en yüksek olduğu değer bize en uygun parametre değerini gösterecektir. Bu nedenle, ROC eğrisinin daha iyi sunabilmek için, Şekil 4.4'te ROC eğrisi için doğru pozitif oranının yüksek olduğu n sayısı 0-185 aralığında olan kesiti verilmiştir. Buna göre kümelerden seçilecek n sayısı için en uygun değerin 45 olduğu gözlemlenmiş ve deneylerde bu değer sabit olarak kullanılmıştır.



Şekil 4.4 Pozitif doğru oranının yüksek olduğu kesit için ROC eğrisi

4.3.1 Kategorik Özelliklerin Kümelenmesi

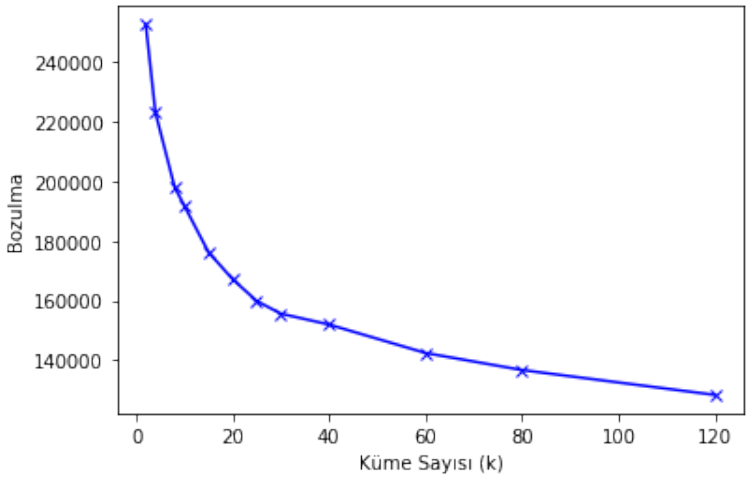
Çalışmada önerilen yöntem, benzer sahte işlemlerin tespit edilmesi için kümeleme yapılmasını ve daha sonra kümelenerek elde edilen benzer sahte işlemlerde kullanılan kartların geçmiş işlemleri analiz edilerek, kullanıldıkları ortak üye iş yerleri üzerinden kopyalama noktalarını tespit etmeyi amaçlamaktadır. Bu amaçla, ilk olarak, K-modes kümeleme yöntemi ile yapılan bir deney yapılmıştır. K-modes, kategorik veriler üzerinde kümeleme işlemi yapmaktadır. Sahte işlem veri kümesinde bulunan "Harcama Miktarı" sayısal bir özelliktir. Harcama miktarı özelliği için gerçekleştirilen kategorik dönüşüm için Tablo 4.2'de verilen dönüşüm uygulanmıştır. Harcama miktarları için kategori belirlenirken, logaritmik ölçek kullanılmıştır.

Tablo 4.2 Harcama miktarının kategorik dönüşüm tablosu

Harcama Miktarı(TL)	**Kategori**
0-10	A
10-100	B
100-1000	C
1000-10000	D
10000-...	E

Harcama miktarı özelliği, yapılan dönüşüm ile kategorik olduktan sonra, K-Modes yöntemi ile kümeleme işlemi gerçekleştirilmiştir. Kümeleme işlemi yapılırken, seçilebilecek uygun küme sayılarını (k) inceleyebilmek için dirsek yönteminden

faydalanılmıştır. Şekil 4.5'da K-Modes için elde edilen dirsek grafiği verilmiştir. Bu grafiğe göre, k sayısı 25 olduktan sonra küme elemanlarının küme merkezlerine yakınlaştığı görülmektedir.



Şekil 4.5 K-Modes yöntemi için elde edilen dirsek grafiği

Sahte işlemlerin olduğu kümeler içinden, kopya sahte işlemlerin küme içindeki yoğunluğu referans alınarak oluşturulan eşik değerler üzerinden kümeler seçilmiştir. Seçilen kümelerin de kopyalama merkezinde kopyalanmış olma ihtimali yüksek olan kart gruplarını oluşturduğu varsayılmıştır. Buna göre, seçilen bu kart kümelerinin geçmiş 3 aylık bütün işlemleri incelenerek, işlemlerin ortak olarak gerçekleştiği iş yerleri ortak işlem sayısına göre sıralanarak bir şüpheli üye iş yeri listesi oluşturulmuştur. Şüpheli üye işyeri listesinden, ilk 45 tanesi seçilmiş ve kart kopyalama noktası adayları belirlenmiştir. Farklı küme sayısı, farklı yoğunluk değeri d için Tablo 4.3'te verilen sonuçlar elde edilmiştir. Sonuçlarda verilen doğru tahmin değerleri, önerilen üye iş yerleri adaylarından 2018 Aralık ayı için bildirilen 17 kart kopyalama noktası ile eşleşen, doğru tahmin edilen üye iş yeri sayısını vermektedir. Elde edilen bu sonuçlara göre, K-modes yöntemiyle oluşturulan kümelerde kopyalama nedenli sahte işlemlerin yoğunluk eşik değerlerini geçemediği görülmüştür. Bunun yanı sıra, eşik değerini geçerek seçilen kümeler doğru tahmin oluşturmak için oldukça yetersiz gözükmektedir. Bu da K-Modes yöntemi ile oluşturulan kümelerin, benzer sahte işlemleri kümelemekte yetersiz olduğunu göstermektedir.

K-Modes yöntemi ile elde edilen sonuçlar için elde edilebilen en iyi sonuçta doğru tahmin sayısı 4'tür. Bu da, kopyalanmadan kaynaklanan benzer sahte işlemlerin kümelenmediğini göstermektedir. Bu nedenle, özellik çıkarımı yöntemleri ile elde edilen yeni özellik uzayı benzer işlemleri kümelemek için kullanılmasının daha başarılı olacağı öngörülmüştür.

Tablo 4.3 2018 Aralık ayı ve Kmodes yöntemi için, farklı küme sayısı ve küme seçim kriterlerinde elde edilen sonuçlar

Toplam Küme Sayısı	Yoğunluk Eşik Değeri	Seçilen Küme sayısı	Önerilen Aday Sayısı	Doğru Tahmin
4	0.2	1	45	0
10	0.2	1	45	3
25	0.2	2	75	4
30	0.2	1	45	3
40	0.2	2	80	3
4	0.4	4	79	0
10	0.4	1	45	3
25	0.4	2	74	4
30	0.4	1	45	3
40	0.4	1	45	3
4	0.8	4	79	0
10	0.8	1	45	3
25	**0.8**	**2**	**74**	**4**
30	0.8	1	45	3
40	0.8	1	45	3

4.3.2 Özellik Çıkarımı ile elde edilen özelliklerin Kümelenmesi

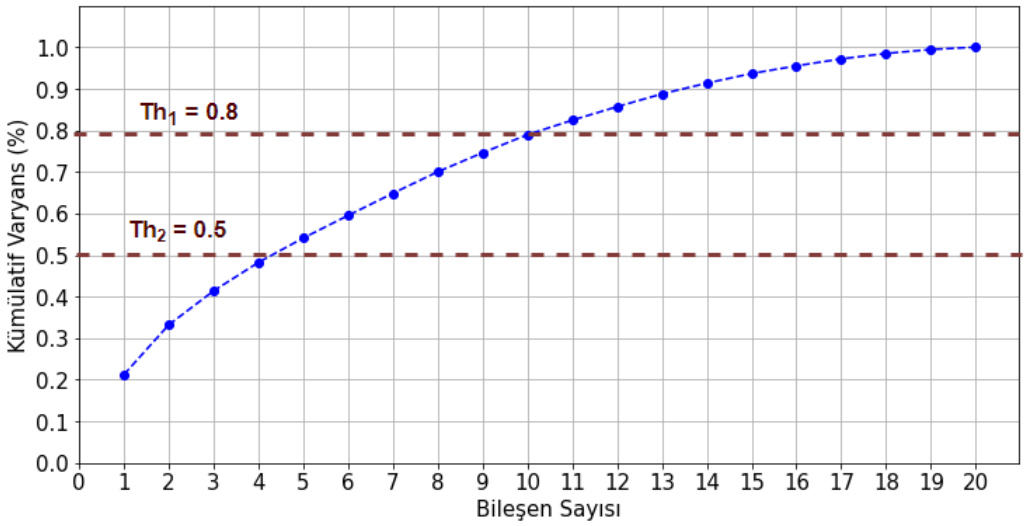
Özellik çıkarımı için PCA ve özkodlayıcı yöntemleri kullanılmıştır. Daha sonra bu yöntemler ile elde edilen özellikler ile kümeleme işlemi yapılıp, daha sonra işlem geçmişi analizleri yapılarak kopyalama noktaları tespit edilmiştir.

4.3.2.1 PCA ile çıkarılan özelliklerin kümelenmesi

Deneyde, bir özellik çıkarımı yöntemi olan PCA uygulanmış ve sonuçları verilmiştir. 2018 Aralık ayındaki sahte işlemler için önce, PCA ile özellik çıkartılmıştır. PCA ile çıkarılan özellikler sayısal özelliklerdir, bu nedenle K-ortalama yöntemi ile kümeleme işlemi yapılmıştır.

Şekil 4.6 PCA ile farklı bileşen/özellik sayıları ile oluşabilecek kümülatif varyans değerleri verilmiştir. Bu varyans değerlerine bakıldığında, bileşen sayısı 19'u geçtiğinde, varyans bakımından zengin özelliklerin elde edilemeyeceği görülmektedir. Bileşen sayısını belirlemek için kümülatif varyans değerleri incelenerek eşik değer belirlenmiştir. Eşik değerleri için, için $th_1 = 0.5$ ve $th_2 = 0.8$ olmak üzere iki farklı eşik değeri deneysel olarak seçilmiştir. Bu eşik değerlerine denk gelen bileşen sayısı 5 ve 10 olacak şekilde iki farklı özellik uzayı ile deney yapılmıştır.

PCA ile çıkarılan 5 özellik ile yapılan kümeleme işlemleri sonrasında, farklı küme sayısı, farklı yoğunluk değeri d için Tablo 4.4'te verilen sonuçlar elde edilmiştir. Sonuçlarda verilen doğru tahmin değerleri, önerilen üye iş yerleri adaylarından 2018



Şekil 4.6 PCA ile oluşturulabilecek farklı bileşen sayıları için kümülatif varyans değerleri

Aralık ayı için bildirilen 17 kart kopyalama noktası ile eşleşen, doğru tahmin edilen üye iş yeri sayısını vermektedir. Elde edilen başarı sonuçlarına göre PCA, 5 doğru tahmine ulaşmayı başarabilmiştir.

PCA ile 10 özellik çıkararak ikinci bir deney daha yapılmıştır. Deneyin sonuçları Tablo 4.5'te verilmiştir. Bu deney ile, çıkarılan özellik sayısının arttırılması ile doğru tahmin sayıları artmış olsa da, önerilen modeldeki özellik çıkarımı yöntemi olan özkodlayıcıdan daha iyi bir başarının elde edilemediği görülmektedir.

Tablo 4.4 2018 Aralık Ayı için, PCA ile çıkartılan 5 Özellik ile farklı küme sayısı ve küme seçim kriterleri için elde edilen sonuçlar

Toplam Küme Sayısı	Yoğunluk Eşik Değeri	Seçilen Küme sayısı	Önerilen Aday Sayısı	Doğru Tahmin
4	0.2	1	45	3
10	0.2	3	72	4
25	0.2	5	140	5
30	0.2	8	195	4
40	0.2	9	193	4
4	0.4	1	45	4
10	0.4	2	69	3
25	0.4	4	137	5
30	0.4	6	188	4
40	0.4	7	176	4
4	0.8	1	45	4
10	0.8	2	69	4
25	**0.8**	**3**	**107**	**5**
30	0.8	4	142	4
40	0.8	6	174	4

Tablo 4.5 2018 Aralık Ayı için, PCA ile çıkartılan 10 Özellik ile farklı küme sayısı ve küme seçim kriterleri için elde edilen sonuçlar

Toplam Küme Sayısı	Yoğunluk Eşik Değeri	Seçilen Küme sayısı	Önerilen Aday Sayısı	Doğru Tahmin
4	0.2	1	45	3
10	0.2	2	48	3
25	0.2	4	109	5
30	0.2	4	109	5
40	0.2	5	122	5
4	0.4	1	45	3
10	0.4	1	45	3
25	0.4	3	106	5
30	0.4	3	105	5
40	0.4	3	106	5
4	0.8	1	45	3
10	0.8	1	45	3
25	0.8	3	106	5
30	0.8	3	106	5
40	**0.8**	**3**	**106**	**5**

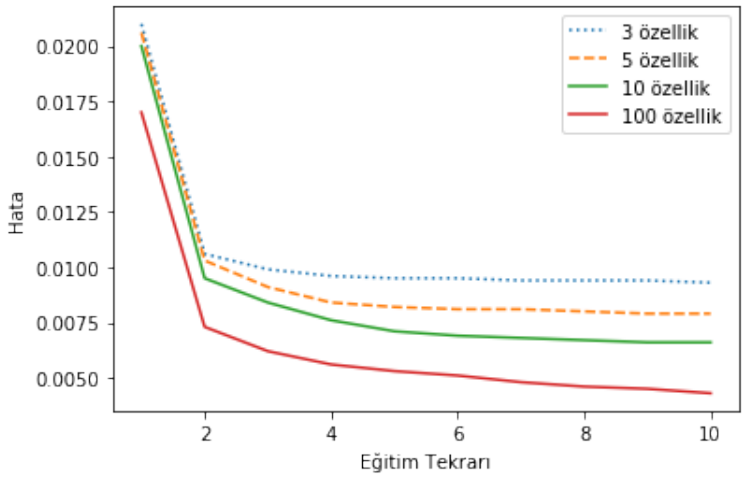
4.3.2.2 Özkodlayıcı ile çıkarılan özelliklerin kümelenmesi

Özkodlayıcı modelin eğitim için 2016-2018 yılları arasında gerçekleşen yaklaşık 680.00 sahte işlemden oluşan veri kümesi kullanılmıştır. Özkodlayıcı modelin mimarisi için tek gizli katmandan oluşan sade bir model mimarisi kullanılmıştır.

Özkodlayıcı model eğitimi için, uygun bir maliyet fonksiyonu seçmek önemlidir. Bu çalışmada, ikili çapraz entropi (binary cross entropy) fonksiyonu model eğitimi için deneysel olarak seçilmiştir. Ayrıca, model eğitiminin hızlı olması ve az iterasyonla yakınsamayı sağlamak için, AdaDelta Optimizasyonu [37] kullanılmıştır.

Şekil 4.7'de eğitilen 3, 5, 10 ve 100 gizli özellik sayısına sahip modellerin farklı eğitim tekrarlarındaki ortalama karesel hataları verilmiştir. Buna göre, farklı özellik sayılarında karesel hatalarda meydana gelen değişimlerin yok sayılacak değerlerde olduğu ve 5 eğitim tekrarından sonra modellerin daha fazla eğitim tekrarına gereksinim duymadığı görülmüştür. Özellik çıkarımı için gereksinim duyulan özkodlayıcı model, mümkün oldukça az özellik sayısına ve az hataya sahip olmalıdır. Bu nedenle özkodlayıcı model için, 5 özellik içeren tek bir gizli katmandan oluşan basit bir mimarinin yeterli olacağı belirlenmiştir.

Özkodlayıcı model ile çıkarılan 5 özellik, sayısal özelliklerdir. Bu nedenle K-ortalama yöntemi uygulanarak , 2018 Aralık ayındaki sahte işlemler üzerinde kümeleme yapılmıştır. K-ortalama yönteminde küme elemanlarının küme merkezlerine olan uzaklıklarını incelemek için dirsek yöntemi uygulanmıştır. Şekil 3.10'da, kümeleme için uygun küme sayısını (k) bulabilmek için kullanılan dirsek yöntemi verilmiştir.



Şekil 4.7 Farklı boyutlardaki özellik uzayları için, eğitim tekrarlarında gözlemlenen ortalama karesel hata değerlerinin grafiği

Buna göre, uygun küme sayısı 15 olduktan sonra verilerin küme merkezlerine yakınlaştığı görülmektedir.

Kümeler belirlendikten sonra, kopya sahte işlem yoğunluk eşik değerini geçen kümeler seçilmiş ve bu kümeler için 3 aylık geriye yönelik işlem geçmişi analizi yapılmıştır. Farklı küme sayısı, farklı yoğunluk eşik değeri d için Tablo 4.6'te verilen sonuçlar elde edilmiştir. Sonuçlarda verilen doğru tahmin değerleri, önerilen üye iş yerleri adaylarından 2018 Aralık ayı için bildirilen 17 kart kopyalama noktası ile eşleşen, doğru tahmin edilen üye iş yeri sayısını vermektedir. Elde edilen sonuçlara göre önerilen yöntem 2018 Aralık ayı için, parametreler olan küme sayısı 30 ve yoğunluk eşik değeri $d = 0.8$ için, kullanılan diğer özellik uzayları dikkate alındığında en fazla doğru tahmine ulaşmayı sağlamıştır.

Bir diğer deneyde, çıkarılan farklı özellik sayısı ile elde edilen bir başka sonuç değerlendirilmiştir. Bu deneyde, özkodlayıcı için 100 özellik oluşturan bir model eğitilmiş ve sonuçları Tablo 4.7'de verilmiştir. Sonuçlar incelendiğinde, özellik sayısı 5 olduğunda elde edilen sonuçlardan daha iyi sonuç elde edilemediği görülmektedir.

Tablo 4.6 2018 Aralık ayı ve 5 özellik çıkartan özkodlayıcı ve sonrasında kullanılan K-ortalama yöntemi için, farklı küme sayısı ve küme seçim kriterlerinde elde edilen sonuçlar

Toplam Küme Sayısı	**Yoğunluk Eşik Değeri**	**Seçilen Küme sayısı**	**Önerilen Aday Sayısı**	**Doğru Tahmin**
4	0.2	1	45	3
10	0.2	3	108	4
25	0.2	6	149	6
30	0.2	6	154	6
40	0.2	7	181	6
4	0.4	1	45	3
10	0.4	3	108	4
25	0.4	6	151	6
30	0.4	6	155	6
40	0.4	7	182	6
4	0.8	1	45	3
10	0.8	2	78	4
25	**0.8**	**3**	**114**	**6**
30	0.8	3	114	6
40	0.8	4	143	6

Tablo 4.7 2018 Aralık ayı ve 100 özellik çıkartan özkodlayıcı ve sonrasında kullanılan K-ortalama yöntemi için, farklı küme sayısı ve küme seçim kriterlerinde elde edilen sonuçlar

Toplam Küme Sayısı	**Yoğunluk Eşik Değeri**	**Seçilen Küme sayısı**	**Önerilen Aday Sayısı**	**Doğru Tahmin**
4	0.2	1	45	3
10	0.2	3	105	4
25	0.2	5	120	4
30	0.2	5	121	4
40	**0.2**	**7**	**175**	**6**
4	0.4	1	45	3
10	0.4	2	81	4
25	0.4	4	91	4
30	0.4	4	91	4
40	0.4	6	149	5
4	0.8	1	45	3
10	0.8	2	81	4
25	0.8	2	81	4
30	0.8	2	81	4
40	0.8	4	142	5

Çıkarılan özellik sayısı 5 olan özkodlayıcı ile gerçeklenen yöntem, önerilen yöntemdir. Bu yöntem, özkodlayıcı modelin olduğu diğer deneylerde baz alınarak başarı değerlendirmeleri gerçekleştirilmiştir. Diğer deneylerde, önerdiğimiz yöntemin diğer, parametreleri için elde edilen sonuçlar verilmiştir.

4.3.2.2.1 Farklı Geçmiş Zaman Periyotlarının Değerlendirilmesi Bu deneyde, önerilen yöntem olan özkodlayıcı ve kümeleme yöntemleri kullanılarak, geçmişe yönelik analiz aşamasında 3 ay geriye yönelik işlemlerin incelenmesi yerine 4, 5 ve 6 aylık geçmiş işlemlerin incelenerek başarının artıp artmadığı incelenmiştir. Sonuçlar, 4, 5 ve 6 aylık geçmiş işlemleri incelendiği durumlar için sırasıyla Tablo 4.8, Tablo 4.9 ve Tablo 4.10'de verilmiştir.

Tablo 4.8 Baz alınan model için, farkı küme sayısı ve küme seçim kriterlerine göre, geriye yönelik 4 ay analiz yapıldığında elde edilen sonuçlar

Toplam Küme Sayısı	**Yoğunluk Eşik Değeri**	**Seçilen Küme sayısı**	**Önerilen Aday Sayısı**	**Doğru Tahmin**
4	0.2	1	45	2
10	0.2	3	100	3
25	0.2	6	147	4
30	0.2	6	147	5
40	**0.2**	**7**	**183**	**6**
4	0.4	1	45	2
10	0.4	3	103	2
25	0.4	6	149	5
30	0.4	6	147	4
40	0.4	7	182	5
4	0.8	1	45	2
10	0.8	2	74	4
25	0.8	3	109	4
30	0.8	3	108	4
40	0.8	4	145	4

Tablo 4.9 Baz alınan model için, farkı küme sayısı ve küme seçim kriterlerine göre, geriye yönelik 5 ay analiz yapıldığında elde edilen sonuçlar

Toplam Küme Sayısı	**Yoğunluk Eşik Değeri**	**Seçilen Küme sayısı**	**Önerilen Aday Sayısı**	**Doğru Tahmin**
4	0.2	1	45	2
10	0.2	3	98	3
25	0.2	6	142	4
30	0.2	6	146	4
40	0.2	7	179	4
4	0.4	1	45	2
10	0.4	3	98	3
25	0.4	6	143	3
30	0.4	6	147	3
40	0.4	7	172	3
4	0.8	1	45	2
10	0.8	2	71	3
25	**0.8**	**3**	**106**	**4**
30	0.8	3	108	4
40	0.8	4	142	3

Tablo 4.10 Baz alınan model için, farkı küme sayısı ve küme seçim kriterlerine göre, geriye yönelik 6 ay analiz yapıldığında elde edilen sonuçlar

Toplam Küme Sayısı	Yoğunluk Eşik Değeri	Seçilen Küme sayısı	Önerilen Aday Sayısı	Doğru Tahmin
4	0.2	1	45	1
10	0.2	3	99	2
25	0.2	6	144	3
30	0.2	6	146	3
40	0.2	7	181	4
4	0.4	1	45	1
10	0.4	3	98	2
25	0.4	6	145	3
30	0.4	6	143	3
40	0.4	7	181	4
4	0.8	1	45	1
10	0.8	2	72	2
25	0.8	3	105	3
30	0.8	3	105	3
40	**0.8**	**4**	**143**	**4**

Sonuçlar incelendiğinde, geriye yönelik analizde fazla geçmişe gidildiğinde doğru tahmin sayısının azalması ve yanlış olumlu sonuçların artmasından dolayı başarının düştüğü açıkça görülmektedir.

4.3.2.2.2 Farklı Aylardaki Sahte İşlemlerin Kümelenmesi Bu deneyde, 9 kopyalama noktası rapor edilen 2018 Kasım ayı için, önerilen yöntemin başarıları test edilmiştir. Tablo 4.11'da 2018 Kasım ayı için çıkarılan özellikler üzerinden kümeleme işlemi gerçekleştirilip 2018 Kasım ayında rapor edilen 9 kopyalama noktası için test sonuçları verilmiştir. Sonuçlara göre, önerilen model, 2018 Kasım ayı için de, 2018 Aralık ayı için yakaladığı başarıya yakın başarılar elde etmiştir.

Kasım ayındaki sahte işlemlerin kümelenmesi ile yapılan bir ikinci deneyde ise, bu sefer bir sonraki ay olan 2018 Aralık ayında rapor edilen 17 kopyalama noktası tespit edilmeye çalışılmıştır. Ancak bu deneyde herhangi bir doğru tahmin elde edilememiştir.

Tablo 4.11 Baz alınan model kullanılarak 2018 kasım ayı için, kasım ayında bildirilen kopyalama noktaları referans alınarak, farklı küme sayısı ve küme seçim kriterleri için elde edilen sonuçlar

Toplam Küme Sayısı	Yoğunluk Eşik Değeri	Seçilen Küme sayısı	Önerilen Aday Sayısı	Doğru Tahmin
4	0.2	1	45	3
10	0.2	2	63	4
25	0.2	3	90	5
30	0.2	4	94	5
40	0.2	5	121	4
4	0.4	1	45	3
10	0.4	1	45	3
25	0.4	2	79	5
30	**0.4**	**3**	**81**	**5**
40	0.4	5	120	4
4	0.8	1	45	3
10	0.8	1	45	4
25	0.8	1	45	3
30	0.8	1	45	3
40	0.8	3	82	3

4.3.2.2.3 Sadece Kopyalama Kaynaklı İşlemlerin Kümelenmesi Bu deneyde, sadece sahte işlem nedeni kopyalama olarak bildirilen bir alt küme kullanılarak deney yapılmıştır. Veri kümesinde bulunan, 2018 Aralık ayında gerçekleşen yaklaşık 21.298 sahte işlemden yalnızca 1.384 tanesi kopyalama nedenli (Sahte İşlem Tipi = 4 olan) kaynaklandığı tespit edilmiştir. Tablo 4.12'da, bu 1.384 sahte işlem üzerinde baz alınan model kullanılarak elde edilen sonuçlar verilmiştir. Sonuçlara göre, 5 doğru tahmin sayısı ancak çok fazla yanlış olumlu cevabın oluşmasıyla elde edilmiştir. Bu da, bu alt veri kümesinin kullanılmasıyla elde edilebilecek sonuçların, önerilen yöntemde kullanılan bütün sahte işlem tiplerinin olduğu veri kümesinin kullanılmasıyla elde edilebilecek sonuçlardan daha iyi olmadığını göstermektedir.

Tablo 4.12 Baz Alınan model kullanılarak, 2018 Aralık Ayı için, Kopyalama Kaynaklı Bildirilen Sahte İşlemlerin kümelenmesi ile elde edilen sonuçlar

Toplam Küme Sayısı	Seçilen Küme sayısı	Önerilen Aday Sayısı	Doğru Tahmin
4	4	50	2
10	10	82	3
25	25	153	4
30	**30**	**182**	**5**
40	40	225	5

4.3.2.2.4 Referans Kopyalama Noktalarından Oluşturulan Alt Veri Kümesinin Değerlendirilmesi Sahte işlem veri kümesi, bankalar arası işlemleri içermekteyken başarı değerlendirmelerini yapabilmek için kullandığımız referans kopya noktaları bir banka aracılığıyla bildirilmiştir. Bu deneyde, kullanılan veri kümesi ve referans noktaların farklı kaynaklardan elde edilmiş olmasının deney sonuçlarına olan etkisi incelenmek istenmiştir.

Gerçekleştirilen deneylerde, veri kümesindeki tüm sahte işlemler kümeleme işlemine tabi tutulmuştur. Bu deneyde ise, referans olarak bildirilen 17 kopyalama noktası baz alınarak bir alt veri kümesi oluşturulmuştur. Amaç, sadece kopyalama nedenli gerçekleşen sahte işlemler üzerinde yöntemimizin başarısını ölçmektir. Bunun için, tüm veri noktalarını kümelemek yerine, 17 referans kopyalama noktalarının herhangi birinden geçtikten sonra sahte işleme karışan kartlara ait işlemler kümelenmiştir.

Başarı değerlendirmesinde kullanılan 17 referans noktasından geçtikten sonra farklı aylarda gerçekleşen yaklaşık 18.000 sahte işlem tespit edilmiştir. Bu 18.000 işlemden yaklaşık 3.000 işlem deneylerin yapıldığı 2018 Aralık ayında gerçekleşmiştir. 18.000 işlem üzerinde özkodlayıcı model eğitilmiştir. Daha sonra 3.000 işlem için özellik çıkarımı gerçekleştirilip, çıkarılan özellikler üzerinde K-ortalama yöntemi ile kümeleme yapılmıştır. Farklı küme sayısı, farklı yoğunluk değeri d için Tablo 4.13'de verilen sonuçlar elde edilmiştir. Sonuçlarda verilen doğru tahmin değerleri, önerilen üye iş yerleri adaylarından 2018 Aralık ayı için bildirilen 17 kart kopyalama noktası ile eşleşen, doğru tahmin edilen üye iş yeri sayısını vermektedir. Alt veri kümesinin kullanılmasıyla elde edilen doğru tahmin sayısı yönünden tüm veri kümesi kullanıldığında elde edilen başarıyı geçmiştir. Ancak bu deneyde, doğrudan referans bilgileri kullanılarak bir alt veri kümesi oluşturulduğu için daha yüksek başarıya ulaşılması beklenmekteydi. Ayrıca alt veri kümesi elde edilen doğru tahmin sayıları ile tüm veri kümesi üzerinde elde edilen doğru tahmin sayıları benzer olsa da, doğru tahmin edilen iş yerleri incelendiğinde farklılıklar olduğu gözlemlenmiştir. Bu da tüm veri kümesinde verilen sahte işlemler içerisinde yanlış/eksik bildirilen ve gerçekte kopyalama olan ancak kopyalama nedenli bildirilmemiş sahte işlemlerinde olduğunu göstermektedir.

Tablo 4.13 Alt Veri Kümesinde, 2018 Aralık ayı ve Özkodlayıcı + K-ortalama yöntemi için, farklı küme sayısı ve küme seçim kriterlerinde elde edilen sonuçlar

Toplam Küme Sayısı	Yoğunluk Eşik Değeri	Seçilen Küme sayısı	Önerilen Aday Sayısı	Doğru Tahmin
4	0.2	2	79	6
10	0.2	5	121	6
25	0.2	11	154	8
30	0.2	12	156	8
40	0.2	14	155	8
4	0.4	2	79	6
10	0.4	5	121	6
25	0.4	9	154	8
30	0.4	11	156	8
40	0.4	12	158	9
4	0.8	1	45	3
10	0.8	4	120	6
25	0.8	8	150	8
30	**0.8**	**10**	**155**	**9**
40	0.8	12	156	9

4.3.2.2.5 Kopyalama Merkezi Aday Seçimi için İş Yeri Skorlarının Kullanılması Baz alınan yöntemde, kopyalama merkezi adayları belirlenirken iş yerlerinde, kopyalandığı düşünülen kart sayısı dikkate alınarak sıralama yapılmaktadır. Ancak aynı sayıda kopyalama işlemi gerçekleştiği düşünülen, kopyalama merkezleri adaylarını kendi içlerinde sıralamak için herhangi bir ölçüt kullanılmamaktadır. Bu duruma çözüm sağlamak amacıyla yapılan bu deneyde, kopyalama merkezi adayında gerçekleşen ortak işlem sayısına ek olarak, yeni bir ölçüt olarak iş yerleri skorları kullanılmıştır. İş yeri skorları BKM bünyesinde gerçekleştirilen bir başka çalışmadan elde edilmiş, iş yerlerinin güvenilirliğini belirten bir veridir. Tablo 4.14'de verilen sonuçlara göre, iş yeri skorlarının kullanılması, baz modelde 6 olarak elde edilen maksimum doğru tahmin sayısını, 40 küme kullanıldığında 7'ye çıkarırken, diğer durumlarda da baz modele yakın yüksek başarılar elde edilmiştir. Bu da iş yeri skorlarının kopyalama merkezi adaylarını belirlemede etkin olduğunu göstermektedir.

Tablo 4.14 2018 aralık ayı ve baz alınan model için kopyalama merkezi adaylarını sıralamada iş yeri skorları da kullanıldığında elde edilen sonuçlar

Toplam Küme Sayısı	Yoğunluk Eşik Değeri	Seçilen Küme sayısı	Önerilen Aday Sayısı	Doğru Tahmin
4	0.2	1	45	3
10	0.2	3	105	4
25	0.2	6	152	6
30	0.2	6	151	6
40	0.2	7	177	7
4	0.4	1	45	3
10	0.4	3	105	4
25	0.4	6	152	6
30	0.4	6	151	6
40	0.4	7	177	7
4	0.8	1	45	3
10	0.8	2	77	4
25	0.8	3	113	6
30	0.8	3	113	6
40	**0.8**	**4**	**139**	**7**

4.3.2.2.6 Kredi Kartı Sahteciliği Tespiti İçin Seçilmiş Özelliklerin Kopyalama Merkezlerinin Tespitinde Kullanılması BKM bünyesinde kredi kartı sahteciliği tespiti için kullanılan bir başka çalışmada, çeşitli özellik seçimi yöntemleriyle belirlenen Tablo 4.15'de verilen özellikler kullanılmıştır. Bu deneyde, bu özellikler kopyalama merkezlerinin tespitinde benzer işlemleri gruplamak için kullanılmıştır. Bu deneyi gerçekleştirmek için öncelikle, sahte işlem veri tablosunda bulunan 2018 Aralık ayında gerçekleşen sahte işlemler, bütün işlemlerin olduğu tablodaki işlem kayıtlar ile eşleştirilmiştir. Bu eşleşme ile, 2018 aralık ayında gerçekleşen yaklaşık 22.000 sahte işlem için bütün işlemler tablosunda yaklaşık 12.000 adet işlem eşleştirilmiştir. Daha sonra, kredi kartı sahteciliği için çıkarılan özelliklerle test yapılmış ancak herhangi bir küme veya küme seçimi kriteri için doğru tahmin sayısı 2'yi geçmemiştir.

Tablo 4.15 Kart Sahteciliği Çalışmasında Kullanılan Özellikler

Özellik	**Özellik tipi**
İşlem Tipi	Nominal Kategorik
İşyeri Ülke Kodu	Nominal Kategorik
İşyeri Kategori Kodu	Nominal Kategorik
Kullanım Kodu	Nominal Kategorik
DWH Neden Kodu	Nominal Kategorik
Elektronik Terminal kodu	Nominal Kategorik
POS Kategori Kodu	Nominal Kategorik
Kart Giriş Tipi	Nominal Kategorik
Geri Ödeme Değeri	Nominal Kategorik
Kart Tipi Kodu	Nominal Kategorik
Alt Kart Tipi Kodu	Nominal Kategorik
Ara Kart Tipi Kodu	Nominal Kategorik
Issuer Üye No	Nominal Kategorik
Acquirer Üye No	Nominal Kategorik
MOTO / E-Ticaret Tipi	Nominal Kategorik
Ödeme Sistemi Kodu	Nominal Kategorik
EMV İşlem Tipi	Nominal Kategorik
Yurtdışı işlem mi?	Nominal Kategorik
Sistem Kaynağı	Nominal Kategorik
ATM işlem mi?	Nominal Kategorik
Chipli işlem mi?	Nominal Kategorik
Ödeme Tipi	Nominal Kategorik
Öneri Kodu	Nominal Kategorik
Marka Paylaşım işlemi mi?	Nominal Kategorik
İşlem Tutarı	Sürekli Nümerik

4.4 Deney Sonuçları

Yapılan deneylerde, 2018 aralık ayı üzerinde kümeleme işlemi yapılmıştır. Kümeleme işlemi sonucunda elde edilen her bir işlem kümesi için, kümelerdeki her kredi kartının 2018 Ekim ayından 2018 Aralık ayına kadar olan bütün işlemleri incelenerek ortak üye iş yerleri tespit edilmiştir. Her bir kümeden en fazla 45 tane kart kopyalama noktası çıkarılmasını sağlayacak şekilde bir eşik değeri seçilmiştir. Bu eşik değeri, önerilen model üzerinde ROC eğrisi yardımıyla belirlenmiştir.

Önerilen yöntem için, en iyi sonuçların bir aylık sahte işlem verisi için özkodlayıcı model ile özellik çıkararak, çıkarılan özellikler üzerinden K-ortalama yöntemi ile kümeleme yapılmasıyla elde edilmiştir. Öte yandan, özellik çıkarımı yöntemi kullanılmadan, doğrudan K-mod yöntemi ile elde edilen sonuçlar beklenen başarıya ulaşmamıştır.

Tablo 4.16'te yapılan deneylerde kullanılan yöntemler ve farklı parametreleri için elde edilen en iyi sonuçlar karşılaştırılmıştır. Özellik çıkarımı için özkodlayıcı model ile elde edilen sonuçlar diğer yöntemlere göre oldukça iyi sonuçlar vermiştir. Bütün yöntemler için küme sayısı 25'ten büyük olduğunda en iyi sonuçlar elde edilmektedir. Yöntemlerin büyük çoğunluğunda, yoğunluk eşik değeri için 60% iyi sonuca götüren küme seçim kriteri olarak değerlendirilebilir.

Kümeleme yöntemleri için uygun k sayılarını belirlemek için dirsek yöntemi

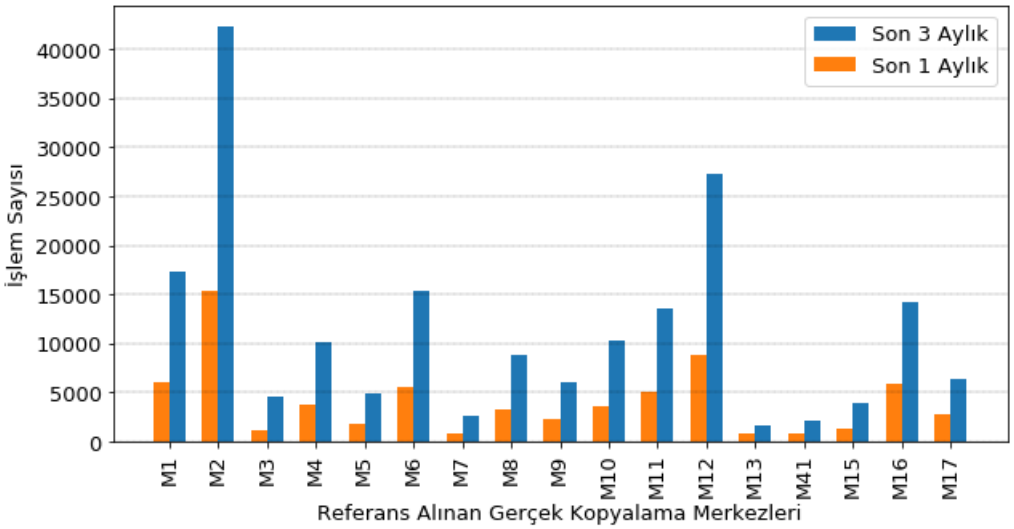
Tablo 4.16 Yöntemler ve parametreler için elde edilen en iyi sonuçlar

Kullanılan Yöntem	Parametre Açıklaması	Küme Sayısı(k) ve Yoğunluk Eşik Değeri(d)	Önerilen Aday Sayısı	Doğru Tahmin
Kategorik özellikler ve K-Mod	Orjinal özellikler	k=25, d=0.8	74	4
Özkodlayıcı ve K-ortalama	Çıkarılan 5 özellik, 3 aylık işlem geçmişi analizi	k=30, d=0.8	114	6
Özkodlayıcı ve K-ortalama	Çıkarılan 100 özellik, 3 aylık işlem geçmişi analizi	k=40, d=0.2	175	6
Özkodlayıcı ve K-ortalama	Çıkarılan 5 özellik, 4,5,6 aylık işlem geçmişi analizi	k=40, d=0.2	183	6
Özkodlayıcı ve K-ortalama	Çıkarılan 5 özellik, 3 aylık işlem geçmişi analizi, İş yeri skorlarından faydalanma	k=40, d=0.8	139	7
PCA ve K-ortalama	Çıkarılan 5 özellik, 3 aylık işlem geçmişi analizi	k=25, d=0.8	107	5
PCA ve K-ortalama	Çıkarılan 10 özellik, 3 aylık işlem geçmişi analizi	k=25, d=0.8	106	5

kullanılmış ancak yüksek başarıya ulaştıran küme sayısının, dirsek yöntemi ile elde edilen küme sayısından daha fazla olması gerektiği deneylerle gösterilmiştir.

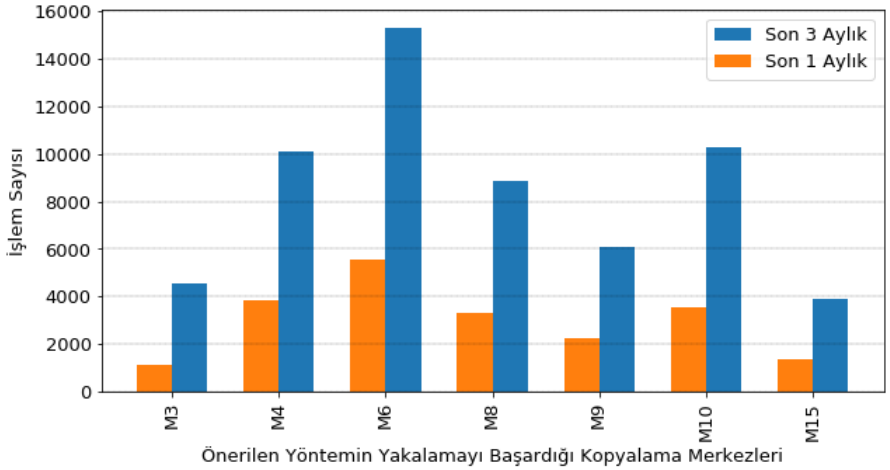
Yöntemimizin başarısını göstermek için kullandığımız 17 referans noktası ile, kullandığımız veri kümesinin arasındaki ilişkiyi gösterebilmek için bir alt veri kümesi oluşturduğumuz deney yapılmıştır. Deney sonucunda, 17 referans noktasında kullanıldıktan sonra gerçekleşen sahte işlemler tespit edilmiş ve bu sahte işlemler bir alt veri kümesi olarak kullanılmıştır. Bu alt veri kümesi üzerinde yapılan deney sonucunda, elde edilen başarı beklenilen başarıdan daha düşük olmuştur. Bunun nedeni, bankaların bütün sahte işlemleri dolayısıyla kart kopyalama noktalarını tespit edememiş olabilmesi ve/veya tespit edilen her bir kopya noktasının veri kümesinde karşılığının olmamasıdır.

Deneylerde kullanılan olan 2018 aralık ayındaki 17 referans kopyalama noktasında gerçekleşen bütün işlemler incelenmiştir. İnceleme sonunda Şekil 4.8'te verilen grafik elde edilmiştir. Grafikte verilen kopyalama merkezlerinin tespit edildiği tarihten itibaren geçmiş 1 ve 3 aylık zaman dilimlerinde gerçekleşen işlem sayılarını yansıtmaktadır. Ayrıca kopyalama merkezleri isimleri veri gizliliği nedeniyle gizlenmiş ve her bir iş yeri M1,M2..M17 olarak ifade edilmiştir. Buna göre, BKM bünyesinde kopyalama merkezlerinden M2 kopyalama merkezinin göreceli olarak diğer kopyalama merkezlerinden daha fazla işleme sahip olduğu görülmektedir.



Şekil 4.8 2018 aralık ayında bankalar tarafından tespit edilen 17 kart kopyalama noktasında gerçekleşen 1 ve 3 aylık kredi kartı işlem sayıları

Deneylerde gerçekleştirilen yöntemlerin doğru tespit ettiği kopyalama merkezleri incelenmiştir. İnceleme sonucunda farklı yöntemlerin aynı kopyalama merkezlerini tespit ettiği görülmüştür. Bu iş yerlerinin geçmiş 1 ve 3 aylık zaman dilimlerinde gerçekleşen işlem sayıları Şekil 4.9'da verilmiştir. Buna göre, işlem sayısı diğer iş yerlerine göre daha az olan kopyalama merkezlerinin de tespit edilebildiği görülmektedir.



Şekil 4.9 Deneylerde yöntemlerin ortak olarak tespit ettiği 7 kart kopyalama noktasında gerçekleşen 1 ve 3 aylık kredi kartı işlem sayıları

Bütün deney ve başarı değerlendirmeleri göz önünde bulundurulduğunda, özkodlayıcı modelle çıkarılan özelliklerden K-ortalama yöntemiyle kümeleme uygulandığında 17 kart kopyalama noktasından 7 tanesi bulunabilmiştir. Bulunan noktalar incelendiğinde, bu 7 noktadan 3 tanesinin orta ölçekli iş yerleri olduğu gözlemlenmiştir. Bir ay içerisinde on binlerce iş yerinde işlem gerçekleştiği

düşünüldüğünde, önerilen yöntemin sadece 100-150 kart kopyalama noktası adayı önererek zaten az sayıda olan 17 referans kopya noktasından 7'sini bulabilmesi başarı olarak değerlendirilmiştir.

5
SONUÇ VE ÖNERİLER

Bu tez çalışmasında, kredi kartı sahteciliklerinin yaşanmasında rol oynayan kredi kartı kopyalama sorununu çözmeye yönelik veri madenciliği ve denetimsiz makine öğrenmesi yöntemlerini de içeren yeni bir sistem önerilmiştir. Önerilen yöntem, kredi kartlarının kopyalandığı iş yerlerini tespit ederek ileride gerçekleşebilecek yeni kredi kartı sahtekarlıklarının önüne geçilmesini sağlamaktadır.

Önerilen yöntemde, sahte işlem verileri üzerinde özkodlayıcı ile özellik çıkarılmış, çıkarılan özellikler üzerinden kümeleme yapılarak benzer sahte kart işlemleri gruplanmış, daha sonra her bir grup için, işlemlerde geçen kredi kartlarının geçmiş işlemleri incelenerek kopyalandıkları noktalar tespit edilmiştir.

Veri kümesinde bulunan sahte işlemler için, sahte işlemin kaynağının birçok veride yanlış girildiği, model başarısını ölçerken kullanılan 17 gerçek kopya merkezinden daha fazlasının olabileceği göz önünde bulundurulduğunda, model tarafından 114.714 üye iş yerin içerisinden yaklaşık olarak sadece 100-150 kopyalama merkezi adayı önererek, 7 farklı kart kopyalama merkezinin bulunması iyi bir başarı olarak değerlendirilmiştir. Genel olarak, bankacılık sektöründe personellerin müşterilerden gelen bildirimler üzerinden manuel olarak yaptığı analizler sonucunda daha az sayıda kopyalama noktası tespit etmesi, önerdiğimiz sadece sahte işlem verisine ihtiyaç duyan modelin başarılı olduğunu göstermiştir.

Sonraki çalışmalarda, önerilen yöntemin geriye yönelik analiz aşamasında, iş yeri merkezlerinin sahte işlem, iş yeri skoruna ek olarak yeni güvenilirlik istatistikleri gibi verileri de kopyalama merkezi tespit etmede kullanarak başarı arttırılabilir. Ayrıca farklı aylarda kopyalama merkezi olduğu bilinen iş yerinin sayısının artmasıyla da yoğunluk bazlı kümeleme yöntemlerinin de başarıyı arttırabileceği düşünülmektedir.

KAYNAKÇA

[1] TBB, *Bankacılıkta Dolandırıcılık Eylemleri Tespit ve Önleme Yöntemleri*, https://www.tbb.org.tr/gec/KTPV14.pdf, Online; accessed 02 January 2020, 2018.

[2] S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, "Credit card fraud detection using bayesian and neural networks," in *Proceedings of the 1st international naiso congress on neuro fuzzy technologies*, 2002, pp. 261–270.

[3] A. Srivastava, A. Kundu, S. Sural, A. Majumdar, "Credit card fraud detection using hidden markov model," *IEEE Transactions on dependable and secure computing*, vol. 5, no. 1, pp. 37–48, 2008.

[4] Y. G. Şahin, E. Duman, "Detecting credit card fraud by decision trees and support vector machines," 2011.

[5] A. C. Bahnsen, D. Aouada, A. Stojanovic, B. Ottersten, "Detecting credit card fraud using periodic features," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2015, pp. 208–213.

[6] A. Mishra, C. Ghorpade, "Credit card fraud detection on the skewed data using various classification and ensemble techniques," in *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, IEEE, 2018, pp. 1–5.

[7] Machine Learning Group, *Credit Card Fraud Detection | Kaggle*, https://www.kaggle.com/mlg-ulb/creditcardfraud, Online; accessed 18 January 2020, 2020.

[8] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, A. K. Nandi, "Credit card fraud detection using adaboost and majority voting," *IEEE access*, vol. 6, pp. 14 277–14 284, 2018.

[9] Z. Kazemi, H. Zarrabi, "Using deep networks for fraud detection in the credit card transactions," in *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, IEEE, 2017, pp. 0630–0633.

[10] UCI, *UCI Machine Learning Repository*, https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data), Online; accessed 18 January 2020, 2020.

[11] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, P. Beling, "Deep learning detecting fraud in credit card transactions," in *2018 Systems and Information Engineering Design Symposium (SIEDS)*, IEEE, 2018, pp. 129–134.

[12] A. Pumsirirat, L. Yan, "Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine," *International Journal of advanced computer science and applications*, vol. 9, no. 1, pp. 18–25, 2018.

[13] UCI, *UCI Machine Learning Repository*, http://archive.ics.uci.edu/ml/datasets/statlog+(australian+credit+approval), Online; accessed 18 January 2020, 2020.

[14] M. Araujo, M. Almeida, J. Ferreira, L. Silva, P. Bizarro, "Breachradar: Automatic detection of points-of-compromise," in *Proceedings of the 2017 SIAM International Conference on Data Mining*, SIAM, 2017, pp. 561–569.

[15] Daniel Rodriguez, *How to spot an atm skimmer*, https://www.itstactical.com/digicom/security/how-to-spot-an-atm-skimmer/, Online; accessed 30 January 2020, 2018.

[16] allbankingalerts, *What is Credit Card Skimming*, https://https://allbankingalerts.com/what-is-credit-card-skimming-what-a-card-skimmer-looks-like/, Online; accessed 30 January 2020, 2019.

[17] VISA, *Visa Global Acquirer Risk Standards*, https://usa.visa.com/dam/VCOM/download/merchants/visa-global-acquirer-risk-standards.pdf, Online; accessed 29 January 2020, 2018.

[18] J. Jurgovsky, M. Granitzer, K. Ziegler, S. Calabretto, P.-E. Portier, L. He-Guelton, O. Caelen, "Sequence classification for credit-card fraud detection," *Expert Systems with Applications*, vol. 100, pp. 234–245, 2018.

[19] A. C. Bahnsen, D. Aouada, A. Stojanovic, B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, pp. 134–142, 2016.

[20] K. Fu, D. Cheng, Y. Tu, L. Zhang, "Credit card fraud detection using convolutional neural networks," in *International Conference on Neural Information Processing*, Springer, 2016, pp. 483–490.

[21] H. Cramér, *Mathematical methods of statistics*. Princeton university press, 1999, vol. 43.

[22] K. Pearson, "Vii. note on regression and inheritance in the case of two parents," *proceedings of the royal society of London*, vol. 58, no. 347-352, pp. 240–242, 1895.

[23] R. A. Fisher, "Statistical methods for research workers," in *Breakthroughs in statistics*, Springer, 1992, pp. 66–70.

[24] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[25] G. E. Hinton, R. S. Zemel, "Autoencoders, minimum description length and helmholtz free energy," in *Advances in neural information processing systems*, 1994, pp. 3–10.

[26] P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*, ACM, 2008, pp. 1096–1103.

[27] A. Ng *et al.*, "Sparse autoencoder," *CS294A Lecture notes*, vol. 72, no. 2011, pp. 1–19, 2011.

[28] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 689–696.

[29] M. Sakurada, T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, ACM, 2014, p. 4.

[30] C. Doersch, "Tutorial on variational autoencoders," *arXiv preprint arXiv:1606.05908*, 2016.

[31] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[32] X. Wang, "Optimization for training deep models," 2017.

[33] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[34] T. Tieleman, G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

[35] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651–666, 2010.

[36] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data mining and knowledge discovery*, vol. 2, no. 3, pp. 283–304, 1998.

[37] M. D. Zeiler, "Adadelta: An adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.

TEZDEN ÜRETİLMİŞ YAYINLAR

İletişim Bilgileri: fuato@yildiz.edu.tr

Konferans Bildirisi

1. F. Ögme, M. E. Karslıgil, A. G. Yavuz ve M. A. Güvensan, "Denetimsiz Öğrenme ile Kredi Kartı Kopyalama Noktalarının Tespit Edilmesi", Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU), 2020 (baskıda).