DOKUZ EYLÜL ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

ANOMALİ BELİRLEME YAKLAŞIMLARI İLE SAHTEKARLIK TESPİTİ

Berkay ERBAYAT

Eylül, 2024

İZMİR

ANOMALİ BELİRLEME YAKLAŞIMLARI İLE SAHTEKARLIK TESPİTİ

Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü

Yüksek Lisans Tezi

İstatistik Anabilim Dalı, Veri Bilimi Programı

Berkay ERBAYAT

Eylül, 2024

İZMİR

YÜKSEK LİSANS TEZİ SINAV SONUÇ FORMU

BERKAY ERBAYAT tarafından **DOÇ. DR. ENGİN YILDIZTEPE** yönetiminde hazırlanan "**ANOMALİ BELİRLEME YAKLAŞIMLARI İLE SAHTEKARLIK TESPİTİ**" başlıklı tez tarafımızdan okunmuş, kapsamı ve niteliği açısından bir Yüksek Lisans tezi olarak kabul edilmiştir.

.....

Doç. Dr. Engin YILDIZTEPE

Danışman

.....	.....
Doç. Dr. Cemal Murat ÖZKUT	Prof. Dr. Neslihan DEMİREL
Jüri Üyesi	Jüri Üyesi

Prof.Dr. Okan FISTIKOĞLU
Müdür
Fen Bilimleri Enstitüsü

TEŞEKKÜR

Öncelikle, tez danışmanım Doç. Dr. Engin YILDIZTEPE'ye minnettarlığımı ifade etmek isterim. Kendisinin bilgi ve deneyimleri, yol göstericiliği ve sürekli desteği, bu tezin başarılı bir şekilde tamamlanmasında büyük katkı sağlamıştır. Ayrıca bana her zaman sevgi ve desteklerini sunan sevgili annem Fatma ERBAYAT ve babam Baki ERBAYAT'a sonsuz teşekkürlerimi sunuyorum. Onların varlığı ve desteği olmadan bu noktaya gelebilmem mümkün olmazdı. Bu süreçte bana sevgisi ve desteği ile güç veren Naz BULUKLU'ya da teşekkürlerimi sunmak istiyorum.

Berkay ERBAYAT

ANOMALİ BELİRLEME YAKLAŞIMLARI İLE SAHTEKARLIK TESPİTİ

ÖZ

E-Ticarette kredi kartı sahtekarlığı, bireyler ve kurumlar için büyük bir sorun teşkil etmektedir. Bu çalışma, kredi kartı işlemlerinde sahtekarlığın tespit edilmesi amacıyla anomali belirleme yaklaşımlarının kullanılmasını incelemektedir. Tezde, kredi kartı işlemlerinde sahtekarlığın tespiti için "FindOut" adı verilen yeni bir yaklaşım geliştirilmiştir. Önerilen yaklaşım, müşterilerin geçmiş harcama profillerini analiz ederek şüpheli işlemleri tespit etmeyi hedeflemektedir. Araştırma kapsamında kullanılan anomali belirleme yöntemleri arasında izolasyon ormanı, eliptik zarf, yerel aykırı değer ve tek sınıflı destek vektör makineleri yer almaktadır. Bu yöntemlerin etkinliği, farklı veriler ve sahtekarlık senaryoları üzerinde değerlendirilmiştir. Sonuçlar, geliştirilen modellerin sahtekarlık tespitinde yüksek doğruluk ve düşük yanlış pozitif oranlarına ulaşabildiğini göstermektedir.

Anahtar kelimeler: Finansal sahtekarlık, anomali belirleme, kredi kartı sahtekarlığı, izolasyon ormanı, tek sınıflı destek vektör makineleri

ANOMALY DETECTION APPROACHES TO IDENTIFY FRAUD

ABSTRACT

Credit card fraud in e-commerce poses a significant problem for individuals and organisations. This study examines the use of anomaly detection approaches to identify fraud in credit card transactions. In this thesis, a new approach called "FindOut" has been developed to detect fraud in credit card transactions. The approach aims to identify suspicious transactions analyzing customers' past spending profiles. The anomaly detection methods used in this research include isolation forest, elliptic envelope, local outlier factor, and one class support vector machine. The effectiveness of these methods has been evaluated on various datasets and fraud scenarios. The result shows that the developed models can achieve high accuracy and low false positive rates in fraud detection.

Keywords: Financial fraud, anomaly detection, credit card fraud, isolation forest, one-class support vector machine

İÇİNDEKİLER

Sayfa

YÜKSEK LİSANS TEZİ SINAV SONUÇ FORMU ii
TEŞEKKÜR iii
ÖZ iv
ABSTRACT v
ŞEKİLLER LİSTESİ viii
TABLOLAR LİSTESİ ix

BÖLÜM BİR – GİRİŞ 1

BÖLÜM İKİ – ANOMALİ BELİRLEME 6

2.1 Anomali Türleri 8
2.1.1 Nokta Anomali 9
2.1.2 Bağlamsal Anomali 9
2.1.3 Kolektif Anomali 10
2.2 Anomali Belirleme Yaklaşımları 10
2.2.1 Denetimsiz Anomali Belirleme 11
2.2.2 Denetimli Anomali Belirleme 11
2.2.3 Yarı-Denetimli Anomali Belirleme 12

BÖLÜM ÜÇ – FİNANSAL İŞLEMLERDE SAHTEKARLIK 14

3.1 Sahtekarlık Tanımı 14
3.2 Kredi Kartı Sahtekarlığı 15

BÖLÜM DÖRT – YÖNTEM 19

4.1 FindOut Yaklaşımı 20
4.1.1 Profilleme Aşaması 21
4.1.2 Değerlendirme Aşaması 22

4.2 Kullanılan Anomali Tespit Yöntemleri 23
4.2.1 İzolasyon Ormanı 23
4.2.2 Yerel Aykırı Değer 25
4.2.3 Destek Vektör Makineleri 27
4.2.4 Tek Sınıflı Destek Vektör Makineleri 28
4.2.5 Eliptik Zarf 29
4.3 Performans Değerlendirme 30

BÖLÜM BEŞ – UYGULAMA 33

5.1 Veri 33
5.2 Uygulama Şeması 36
5.3 Model Oluşturma 38
5.3.1 Profilleme Aşaması 38
5.3.2 Değerlendirme Aşaması 38
5.4 Profil Değerlendirme 41
5.5 Model Çıktıları 41

BÖLÜM ALTI – SONUÇ VE TARTIŞMA 50

KAYNAKLAR 52

ŞEKİLLER LİSTESİ

Sayfa

Şekil 1.1 Finansal sahtekarlık tipleri 1

Şekil 2.1 Anomali Örneği 7

Şekil 2.2 Nokta Anomali 8

Şekil 2.3 Bağlamsal Anomali 9

Şekil 2.4 Kolektif Anomali 10

Şekil 2.5 Anomali belirleme yöntemleri 12

Şekil 4.1 FindOut Yaklaşımı 21

Şekil 4.2 IF ile gözlemin izole edilmesi 24

Şekil 5.1 Müşterilerin 30 günlük harcamalarının gün içerisindeki dağılımı 34

Şekil 5.2 Sektörlere göre harcama miktarı 35

Şekil 5.3 Anomali gözlemlerin eklenmesi 39

Şekil 5.4 İşlem dakikası değerlerinin histogramı 40

Şekil 5.5 Özel profil F_1 puanları 48

Şekil 5.6 Genel profil F_1 puanları 49

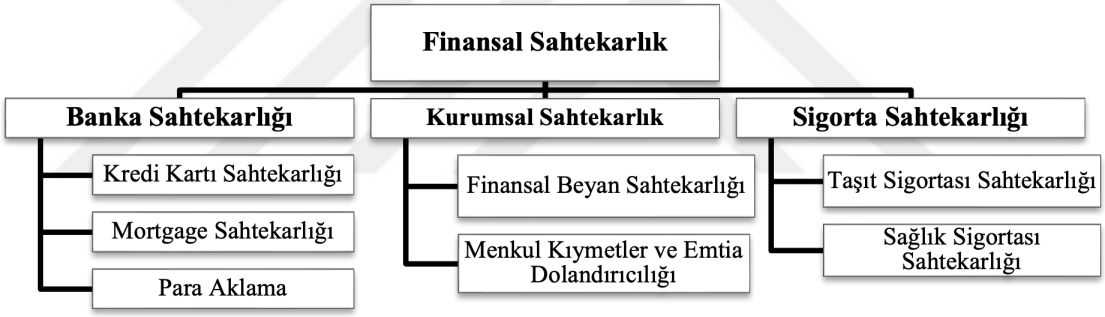
TABLOLAR LİSTESİ

Sayfa

Tablo 4.1 Kullanılan terimler 19
Tablo 4.2 Yanılgı Matrisi 30
Tablo 5.1 Değişken Tanımları 33
Tablo 5.2 Tutar Değişkeninin Tanımlayıcı İstatistikleri 38
Tablo 5.3 Müşteri 1 için sonuçlar 42
Tablo 5.4 Müşteri 2 için sonuçlar 42
Tablo 5.5 Müşteri 3 için sonuçlar 43
Tablo 5.6 Müşteri 4 için sonuçlar 44
Tablo 5.7 Müşteri 5 için sonuçlar 44
Tablo 5.8 Müşteri 6 için sonuçlar 45
Tablo 5.9 Müşteri 7 için sonuçlar 46
Tablo 5.10 Müşteri 8 için sonuçlar 47

BÖLÜM BİR
GİRİŞ

Günümüzün dijital dünyasında, finansal işlemlerin çoğu çevrimiçi olarak gerçekleşmektedir. Ancak geride bırakılan dijital ayak izi yetkisiz erişime karşı savunmasızdır. Siber suçlular, kredi kartı numaraları ve kullanıcı kimlikleri de dahil olmak üzere hassas bilgileri çalmak için bu güvenlik açığından yararlanabilir. Daha sonra, kullanıcının bilgisi olmadan hileli finansal işlemler gerçekleştirebilirler. "İnsanları kandırarak para elde etme suçu" olarak tanımlanabilen sahtekarlık, insanlık kadar eskidir ve sonsuz çeşitlilikte farklı biçimler alabilir (Bolton ve Hand, 2002). Bu biçimlerden biri de finansal sahtekarlıktır. Finansal sahtekarlık üç türe ayrılabilir. Şekil 1.1 yaygın finansal sahtekarlık türlerini göstermektedir. Bu çalışma kredi kartı sahtekarlığına odaklanmaktadır.



Şekil 1.1 Finansal sahtekarlık tipleri

Kredi kartı sahtekarlığı bireyler, işletmeler ve küresel finans ortamı için ciddi sonuçlar doğurabilir. Ayrıca bireyler ve işletmeler için mali kayıplara, itibar ve güvenin zedelenmesine yol açabilir. 2021 yılında Federal Ticaret Komisyonu (FTC), tüketicilerden 2.8 milyon adet dolandırıcılık raporu almıştır. Bu raporlar sonucunda bildirilen toplam kayıplar bir önceki yıla oranla %70 artışla 5.8 milyar dolara ulaşmıştır. En çok bildirilen dolandırıcılık türleri arasında sahte kimliklerle yapılan dolandırıcılıklar ve çevrimiçi alışveriş dolandırıcılıkları bulunmaktadır (Commission, 2022). Suçluların çoğu suç davranışlarını geçmişteki sahtekarlık tespit yöntemlerine göre şekillendirmektedir. Bu nedenle, sahtekarlık tespit stratejilerine yönelik yenilikçi yaklaşımlara ihtiyaç duyulmaktadır. Avrupa Merkez Bankası verilerine göre ödeme

kartlarıyla yapılan her 100 Euro'luk harcamanın 3.6 Cent'i sahtekarlık nedeniyle kaybedilmiştir. Sahtekarlık işlemlerinin finansal ekosistemde bir anomali olması sebebiyle, sahtekarlık tespitine bir anomali belirleme problemi olarak yaklaşmak gerekmektedir.

Anomali tespiti, çeşitli endüstri ve uygulama alanlarında, özellikle siber güvenlik, sağlık bilimleri, finans ve üretim gibi sektörlerde, önemli bir araştırma konusudur. Anomali tespiti, normalden sapma gösteren olayları veya gözlemleri tanımlama sürecidir. Bu süreç, veri madenciliği, istatistik, makine öğrenimi tekniklerini kapsayabilir. Literatürde, anomali tespiti üzerine yapılan çalışmalar, genellikle yöntemlerin geliştirilmesi, yöntemlerin karşılaştırılması ve çeşitli alanlardaki uygulamaları üzerine yoğunlaşmıştır.

Anomali tespitinde en yaygın kullanılan yöntemlerden biri, izolasyon ormanı (Isolation Forest - IF) yöntemidir. Bu yöntem, Liu ve arkadaşları tarafından önerilmiş olup, özellikle yüksek boyutlu verilerde iyi bir şekilde çalışmaktadır. İzolasyon ormanı, veri noktalarını rastgele seçilen bir öznitelik ve değer üzerinden ayrıştırarak anomalileri tespit eder. Anomaliler, normal veri noktalarına göre daha hızlı izole edilir, bu da izolasyon ormanı yönteminin etkinliğini artırır (Liu vd., 2008). Bir diğer popüler metot ise, yerel anomali skoru (Local Outlier Factor - LOF) yöntemidir. LOF, Breunig ve arkadaşları tarafından geliştirilmiştir ve bir veri noktasının yerel yoğunluğunu, komşu noktaların yerel yoğunlukları ile karşılaştırarak aykırı değerleri tespit eder. Bu yöntem, özellikle veri dağılımının homojen olmadığı durumlar için uygundur (Breunig vd., 2000). Derin öğrenme tabanlı anomali tespiti yöntemleri de son yıllarda popülerlik kazanmıştır. Özellikle, otokodlayıcı (autoencoders), verilerdeki normal davranışı öğrenmek ve anomali tespiti için kullanılmak üzere etkili bir şekilde kullanılmıştır (Sakurada ve Yairi, 2014). Otokodlayıcılar, giriş verilerini bir dizi sıkıştırma ve genişletme işleminden geçirerek bir çıktı üretir ve bu süreç sırasında verinin temel özelliklerini öğrenir. Anomali tespiti, otokodlayıcı tarafından üretilen çıktının giriş verisi ile arasındaki farkın ölçülmesiyle gerçekleştirilir (Sakurada ve Yairi, 2014).

Her bir yöntemin benzersiz avantajları ve kısıtlamaları vardır, bu yüzden uygulanacak duruma en uygun metodun seçilmesi önemlidir. Örneğin, derin öğrenme tabanlı yöntemler, özellikle büyük verilerde ve karmaşık veri yapılarında etkili olabilirken, yüksek hesaplama gücü gereksinimi ve modelin açıklanabilirliğinin düşük olması gibi dezavantajlara sahiptir. Öte yandan, geleneksel istatistiksel yöntemler ve makine öğrenimi tabanlı yöntemler, daha az hesaplama kaynağı gerektirir ve modelin nasıl çalıştığına dair daha fazla anlayış sunar, ancak bu yöntemler yüksek boyutlu veya karmaşık veri yapıları ile başa çıkmakta zorlanabilir.

Finansal sahtekarlık, dünya çapında büyük bir sorun teşkil ederken, bu suçla mücadelede kullanılan yöntemlerin etkinliği ve verimliliği büyük önem taşımaktadır. sahtekarlıkla mücadelede kullanılan geleneksel yöntemler, çoğu zaman yetersiz kalabilmekte ve dolandırıcıların sürekli evrilen tekniklerine ayak uyduramamaktadır. Bu nedenle, hem istatistiksel yöntemler hem de derin öğrenme tabanlı teknikler gibi modern yaklaşımlar büyük bir değer taşımaktadır.

Finansal sahtekarlık, modern ekonomiler için ciddi bir tehdit oluşturmakta ve hem işletmeler hem de tüketiciler üzerinde derin etkiler bırakmaktadır. Bu tür sahtekarlık, çeşitli yöntemlerle gerçekleştirilebilir ve genellikle kredi kartı sahtekarlığı, sigorta sahtekarlığı ve finansal beyan sahtekarlığı gibi alanları kapsar (Ngai vd., 2011). Kredi kartı sahtekarlığı, genellikle çalıntı kartların kullanılması veya kart bilgilerinin çevrimiçi olarak ele geçirilmesi şeklinde gerçekleşir. Bu tür sahtekarlık vakaları, yıllık milyarlarca dolarlık kayıplara neden olmakta ve finans kuruluşları için büyük bir sorun teşkil etmektedir (Asha ve Suresh Kumar, 2021).

Sigorta sahtekarlığı, genellikle müşterilerin veya sağlayıcıların poliçe şartlarını manipüle etmeleriyle gerçekleşir. Bu sahtekarlık türü, abartılmış zarar talepleri veya tamamen uydurma olaylar yoluyla sigorta şirketlerini maddi zarara uğratır (West ve Bhattacharya, 2016). Finansal beyan sahtekarlığı ise, şirketlerin hisse senedi fiyatlarını manipüle etmek, kredi şartlarını iyileştirmek veya vergi yükümlülüklerini azaltmak amacıyla mali durumlarını daha iyi göstermek için finansal tabloları kasıtlı olarak değiştirmeleri anlamına gelir (Asha ve Suresh Kumar, 2021).

Bu tür finansal sahtekarlık faaliyetleri, sadece mali kayıplara neden olmakla kalmaz, aynı zamanda tüketici güvenini sarsar ve piyasalarda istikrarsızlığa yol açabilir. Dolayısıyla, finansal sahtekarlığın tespiti ve önlenmesi, hem ekonomik hem de toplumsal açıdan büyük önem taşır. Geleneksel yöntemler, genellikle manuel incelemeler ve denetimler şeklinde yapılmakta olup, sahtekarlığı tespit etmek için yeterince hızlı veya etkili değildir. Bu nedenle, finans kuruluşları sahtekarlığı tespit etmek için makine öğrenimi gibi modern çözümlere yönelmektedir (Asha ve Suresh Kumar, 2021).

Makine öğrenimi yöntemleri, büyük veri kümeleri üzerinde karmaşık analizler yapabilme kapasitesine sahiptir ve bu sayede finansal işlemlerdeki düzensizlikleri hızla tespit edebilir. Örneğin, yapay sinir ağları, destek vektör makineleri ve karar ağaçları gibi yöntemler, sahtekarlık şüphesi taşıyan işlemleri belirlemek için kullanılabilir. Bu sistemler, müşterilerin alışkanlıklarını öğrenerek ve her işlemi bu alışkanlıklara göre değerlendirerek anormal aktiviteleri saptayabilir (West ve Bhattacharya, 2016).

Finansal sahtekarlıkla mücadelede teknolojinin rolü giderek artmaktadır. Finans sektörü, sahtekarlık tespit süreçlerini otomatize etmek ve bu tehditlere karşı daha etkin bir savunma geliştirmek için makine öğrenimi gibi gelişmiş araçlardan yararlanmaktadır. Bu teknolojiler, finansal sahtekarlıkla mücadelede yeni bir dönem başlatmakta ve kuruluşların bu tür suistimalleri daha etkin bir şekilde önlemelerine olanak tanımaktadır (Karpoff, 2021). Bu tez çalışmasında, kredi kartı işlemlerindeki sahtekarlık anomali olarak ele alınmış ve bu işlemleri tespit etmek için yeni bir yaklaşım önerilmiştir. "FindOut" adı verilen yaklaşım ile geçmiş kredi kartı işlemlerinden elde edilen profillerden yola çıkarak müşterinin olağan harcama davranışlarına uymayan işlemleri tespit etmek amaçlanmıştır. Yaklaşımın en büyük avantajı, yalnızca tutar ve işlem zamanı verilerine ihtiyaç duyması sayesinde hızlı tahminler yapabilmesidir. Böylece, gerçek dünya problemlerine karşı etkili bir çözüm sunabileceği düşünülmektedir.

Bu tez çalışmasında, kredi kartı işlemlerinde sahtekarlığın tespiti için anomali

belirleme yaklaşımlarının kullanımı incelenmiştir. Giriş bölümünde, literatürde bu alanda yürütülen çalışmalar özetlenerek, bu tez çalışmasının motivasyonu belirtilmiştir. Tezin ikinci bölümünde, anomali tespiti incelenmiştir. Üçüncü bölümde finansal sahtekarlık kavramı açıklanmıştır. Dördüncü bölümde, araştırma tasarımı ve kullanılan anomali tespit yöntemleri, izolasyon ormanı (isolation forest - IF), eliptik zarf (elliptic envelope - EE), yerel anomali skoru (local outlier factor - LOF) ve tek sınıflı destek vektör makineleri (one class support vector machine - OC-SVM) açıklanmıştır. Önerilen yöntem olan "FindOut" modeli tanıtılmış ve bu yöntemin kısıtlamaları ele alınmıştır. Beşinci bölümde, veri toplama, veri ön işleme, model çıktıları ve bu çıktılar üzerine yapılan karşılaştırmalar sunulmuştur. Son bölümde ise çıktılar değerlendirilmiş, araştırmanın kısıtlamaları tartışılmıştır.

BÖLÜM İKİ
ANOMALİ BELİRLEME

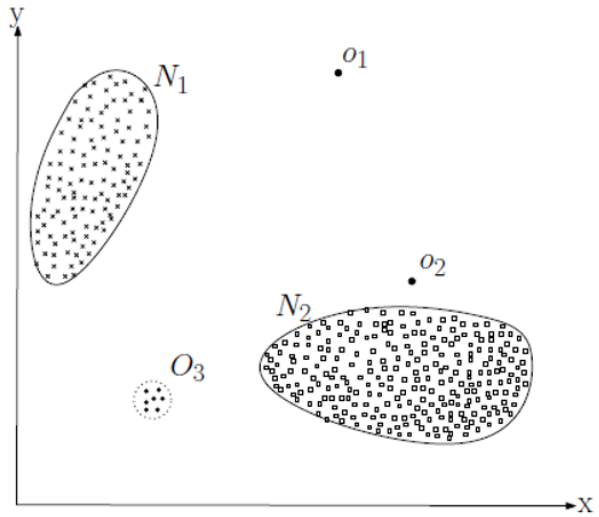
Anomali belirleme, verilerde beklenmedik öğelerin veya olayların tanımlanmasını ifade eder. Bu öğeler, normal olarak kabul edilen verilere göre farklılık gösterir (Goldstein ve Uchida, 2016). Anomali belirlemenin temelinde, verilerdeki anormal durumların veya değerlerin tespit edilmesi yatar. Özellikle etiketlenmemiş veriler üzerinde çalışıldığında, bu süreç daha da karmaşık bir hale bürünebilir çünkü bu durumda yalnızca verinin sağladığı özellikler dikkate alınır (Chandola vd., 2009). Bu bağlamda, çeşitli yöntemler geliştirilmiş olup, bu yöntemlerin genellikle uzaklık veya yoğunluk gibi ölçütlere dayandığı görülmektedir (Zimek vd., 2012). Araştırmalarda kullanılan yaygın bir yöntem olan denetimsiz anomali belirleme, etiketli veri gerektirmez ve verinin özniteliklerini analiz eder. (Goldstein ve Uchida, 2016).

Birçok uygulama alanında, anomali belirleme teknikleri yaygın olarak kullanılmaktadır. Örneğin, ağ güvenlik sistemlerinde, sahtekarlığın tespit edilmesinde, finansal işlemlerin izlenmesinde ve medikal görüntü analizlerinde bu tekniklerden faydalanılmaktadır (Chandola vd., 2009). Bu teknikler, genellikle belirli yöntemlerin güçlü ve zayıf yönlerini belirlemek için çeşitli veriler üzerinde değerlendirilir. Ayrıca, bu yöntemlerin performansı genellikle çeşitli parametre ayarlarının etkileri ile birlikte incelenir ve böylece daha genel bir değerlendirme yapılabilir (Goldstein ve Uchida, 2016).

Anomali belirleme, hem teorik hem de pratik açıdan oldukça önemli bir araştırma alanıdır. Bu alandaki çalışmalar, özellikle büyük verileri ve çeşitli uygulama alanları göz önüne alındığında, veri bilimi ve yapay zeka teknolojilerinin ilerlemesi için kritik öneme sahiptir. Bu nedenle, alandaki yeni gelişmeler sürekli takip edilmeli ve yöntemlerde yapılan iyileştirmeler doğrultusunda sistemler güncellenmelidir.

Anomali, genellikle bir verideki diğer öğelerden önemli ölçüde farklı olan ve bu nedenle beklenmedik olarak kabul edilen öğeleri veya olayları tanımlamak için kullanılan bir terimdir (Goldstein ve Uchida, 2016). Veri analizi ve makine öğrenimi bağlamında, anomali tespiti, verilerindeki sapmaları, hataları veya istisnaları

belirleme sürecidir. Bu, özellikle büyük verilerdenin analiz edildiği ve verilerin otomatik olarak işlenmesi gerektiği durumlarda kritik bir öneme sahiptir. Şekil 2.1'de, N_1 ve N_2 kümelerinden sapma gösteren veri noktalarını (O_1, O_2, ve O_3) gösterilmektedir.



Şekil 2.1 Anomali Örneği (Chandola vd., 2009)

Anomali tespiti, denetimsiz (unsupervised), yarı denetimli (semi-supervised) ve denetimli (supervised) olmak üzere üç ana yaklaşım altında incelenebilir. Denetimsiz anomali tespiti, etiketlenmemiş veri üzerinde çalışır ve verinin içsel yapılarından yararlanarak anormallikleri belirler. Bu yaklaşım, verideki normal davranışın modelini oluşturur ve bu modele uymayan veri noktalarını anomali olarak sınıflandırır. Yaygın kullanılan yöntemler arasında LOF, IF ve histogram tabanlı aykırı değer puanı (Histogram Based Outlier Score - HBOS) bulunmaktadır (Chandola vd., 2009).

Yarı denetimli öğrenme, etiketsiz verilerin kullanıldığı bir makine öğrenimi yöntemidir. Bu yaklaşım, genellikle büyük miktarda etiketsiz verinin mevcut olduğu durumlarda kullanılır. (Chapelle vd., 2006)

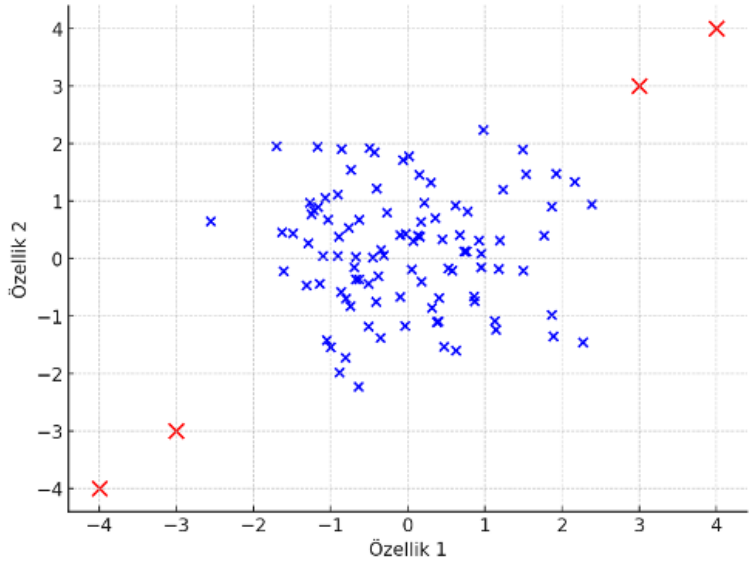
Denetimli anomali tespiti ise, etiketli veriler kullanarak çalışır. Bu yöntem, normal ve anormal veri örneklerinin önceden tanımlandığı durumlarda uygundur. Yöntem, verilen etiketlerden yararlanarak bir model eğitir ve bu modeli yeni verileri

sınıflandırmak için kullanır. Ancak gerçek hayat verilerinde çok sayıda gözlemin önceden etiketlenmesi pek olası değildir.

Anomali tespiti, birçok farklı uygulama alanında kullanılmaktadır. Örneğin, ağ güvenliği alanında veya normal dışı trafik akışlarını tespit ederek olası siber saldırıları önlemede kritik bir rol oynayabilir. Finans sektöründe, alışılmadık işlem modellerini belirleyerek sahtekarlık faaliyetlerinin önüne geçmede kullanılabilir. Sağlık bilimlerinde, hasta verilerindeki anormallikleri tespit ederek erken teşhis konulmasına yardımcı olabilir (Zimek vd., 2012). Tüm bu uygulamalar, anomali tespitinin çeşitli sektörlerde nasıl faydalı olabileceğine örnek olarak gösterilmektedir.

2.1 Anomali Türleri

Anomali tespitinde, tespit edilmeye çalışılan anormal olaylar veya öğelerin türüne göre farklı anomali kategorileri tanımlanmıştır. Bu anomali türleri, özellikle verinin yapısına ve uygulama alanına göre sınıflandırılır. Anomali tespitinde sıklıkla karşılaşılan üç ana anomali türü bulunmaktadır: nokta anomalileri, bağlamsal anomaliler ve kolektif anomaliler (Chandola vd., 2009).



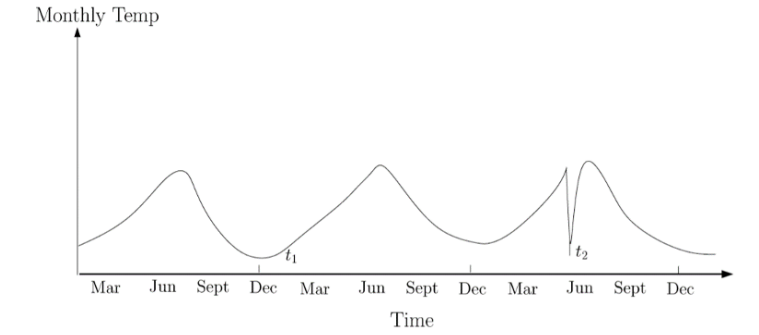
Şekil 2.2 Nokta Anomali

2.1.1 Nokta Anomali

Nokta anomalisi, bir veri noktasının diğer tüm verilere kıyasla önemli ölçüde farklı olduğu durumlarda ortaya çıkar (Chandola vd., 2009). Bu tür anomaliler, verideki diğer öğelerden izole bir şekilde sapma gösterir. Örneğin, bir kredi kartı işlem kaydında, alışılmışın dışında büyük miktarda bir harcama, nokta anomali olarak değerlendirilebilir. Bu tür anomaliler, genellikle finansal sahtekarlık tespiti veya hata tespit sistemlerinde karşımıza çıkar. Şekil 2.2'de kümeden önemli ölçüde sapan 4 anomali değer gösterilmiştir. Bu tekil gözlemler nokta anomali olarak sınıflandırılır.

2.1.2 Bağlamsal Anomali

Bağlamsal veya durumsal anomaliler, belirli bir bağlam veya durum içerisinde anormal olarak kabul edilen, ancak diğer bağlamlar içinde normal kabul edilebilecek veri noktalarıdır (Song vd., 2007). Örneğin, mevsimsel sıcaklık verilerinde, yaz aylarında görülen yüksek sıcaklıklar normaldir, ancak aynı sıcaklık değerleri kış ayları için anomali olarak değerlendirilir. Bu tür anomaliler, özellikle bağlamsal verinin önemli olduğu hava durumu izleme veya kişisel sağlık izleme sistemlerinde önem taşır.

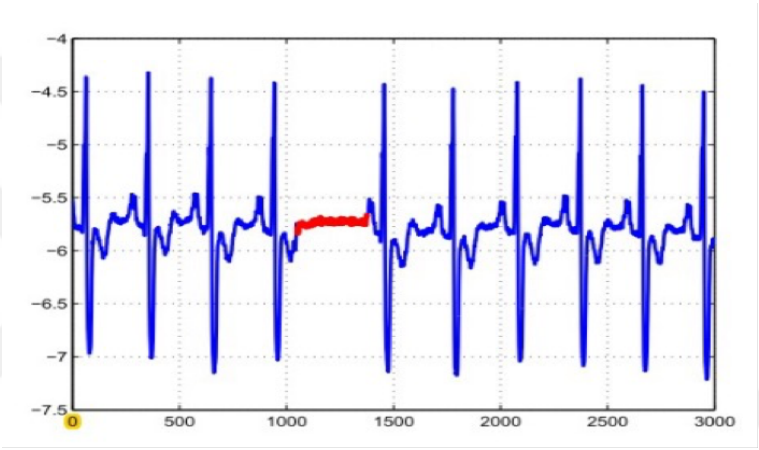


Şekil 2.3 Bağlamsal Anomali (Chandola vd., 2009)

Şekil 2.3'te t_1 ve t_2 sıcaklık değerleri birbirine eşit olmasına rağmen, içinde bulundukları mevsim normalinde t_2 değeri anomali olarak gözlenmektedir. Bu koşullarda t_1 değeri olağan olarak görülürken, t_2 değeri bağlamsal anomali olarak belirtilmektedir.

2.1.3 Kolektif Anomali

Kolektif anomali, bir grup veri noktasının, bireysel olarak anormal olmasalar bile birlikte anormal bir davranış sergilemesi durumunda ortaya çıkar (Song vd., 2007). Bu tür anomaliler genellikle zaman serisi verilerinde veya uzamsal verilerde gözlemlenir. Örnek olarak, bir hastanın kalp ritimlerinin belirli bir dönem içindeki düzensizlikleri kolektif bir anomali oluşturabilir. Benzer şekilde, bir ağdaki trafik akışlarının ani ve eşzamanlı değişiklikleri de kolektif bir anomaliye işaret edebilir. Şekil 2.4'te EKG verisi üzerinde tespit edilen kollektif anomali gösterilmektedir.



Şekil 2.4 Kolektif Anomali (Chandola vd., 2009)

Her bir anomali türünün doğru bir şekilde tanımlanması ve tespit edilmesi, uygulanan anomali tespit yönteminin doğruluğunu ve verimliliğini büyük ölçüde etkiler. Bu nedenle, yöntemleri seçerken ve tasarlarken verinin özelliklerini ve ihtiyaç duyulan tespit türünü dikkate almak önemlidir. Ayrıca, bu türlerin her biri için geliştirilen özelleşmiş algılama teknikleri, genel tespit performansını artırmada ve yanlış pozitifleri minimize etmede kritik rol oynar (Goldstein ve Uchida, 2016).

2.2 Anomali Belirleme Yaklaşımları

Anomali belirleme, modern veri analizi ve sistem izleme uygulamalarında önemli bir yere sahiptir. Bu süreç, genel olarak üç ana yaklaşım çerçevesinde ele alınabilir: denetimli, yarı denetimli ve denetimsiz anomali belirleme. Her bir yaklaşımın kendine

özgü metodolojileri, avantajları ve zorlukları vardır.

2.2.1 Denetimsiz Anomali Belirleme

Denetimsiz anomali belirleme (unsupervised anomaly detection), etiketlenmemiş verilerde uygulanır ve çoğu gerçek dünya uygulaması için geçerli olan senaryodur. Bu yaklaşım, verilerin normal dağılımını modellemeye çalışır ve bu modele uymayan veri noktalarını anomali olarak işaretler. Denetimsiz yöntemler, veri içerisindeki doğal yapıları ve ilişkileri analiz eder, dolayısıyla önceden belirlenmiş bir etiket setine ihtiyaç duymaz.

Yaygın denetimsiz anomali belirleme yöntemleri arasında IF, LOF ve HBOS gibi yöntemler bulunmaktadır. Bu yöntemler, veriler arasındaki uzaklık veya yoğunluk farklılıklarını kullanarak anormal davranışları tespit etmeye çalışır (Goldstein ve Uchida, 2016).

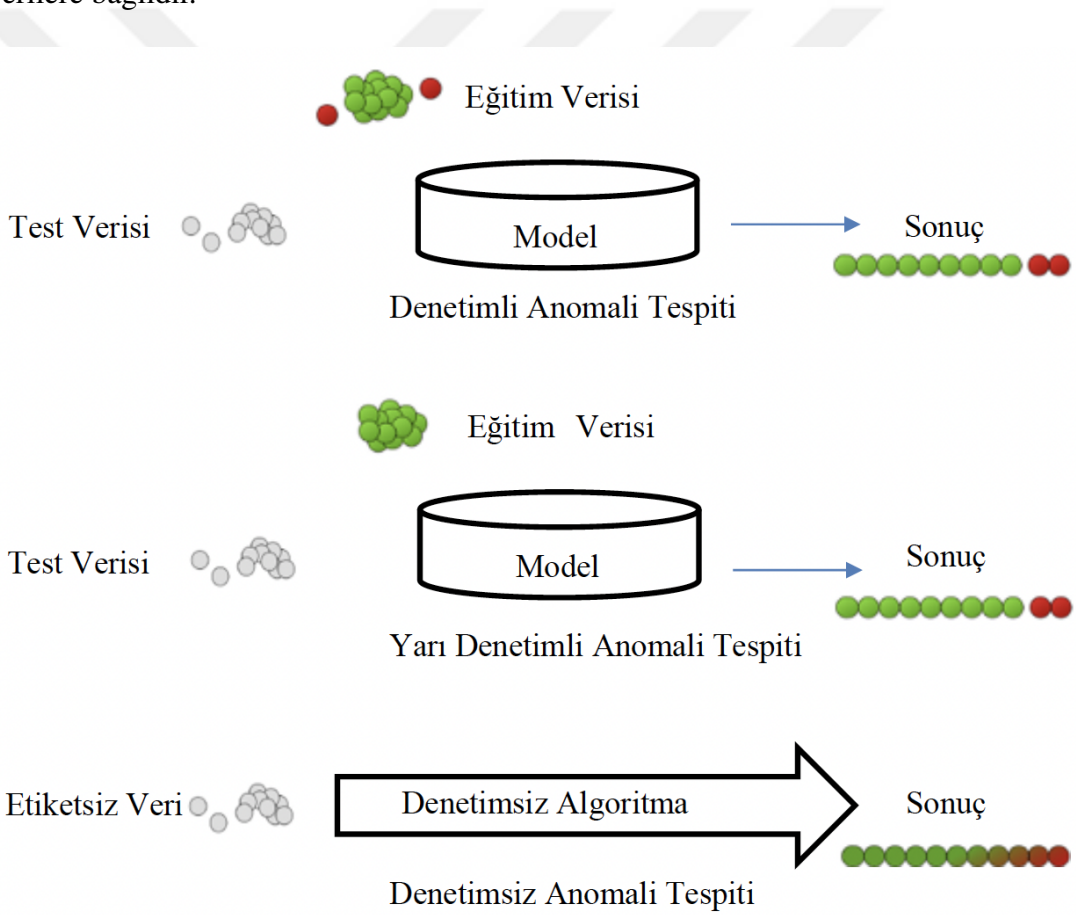
2.2.2 Denetimli Anomali Belirleme

Denetimli anomali belirleme (supervised anomaly detection), etiketlenmiş verilerde kullanarak gerçekleştirilir. Bu yöntemde, normal ve anormal veriler önceden belirlenmiş ve etiketlenmiştir. Eğitim sürecinde, bu etiketler kullanılarak bir model oluşturulur ve bu model, yeni verileri sınıflandırmak için kullanılır. Denetimli yaklaşımlar genellikle daha yüksek doğruluk oranları sunar çünkü model, anormallikleri tanımlamak üzere özel olarak eğitilmiştir. Bu yaklaşımın sıklıkla kullanıldığı teknikler arasında SVM, yapay sinir ağları ve karar ağaçları yer alır. Bu yöntemler çoğu durumda denetimsiz yöntemlerden daha iyi sonuçlar üretir (Chandola vd., 2009).

2.2.3 Yarı-Denetimli Anomali Belirleme

Yarı-denetimli anomali belirleme (semi-supervised anomaly detection), olağan gözlemleri kullanır. Bu teknikler, sadece olağan veri örneklerini kullanarak model oluşturur ve bu modele uymayan yeni verileri anomali olarak belirler. (Goldstein ve Uchida, 2016).

Anomali belirleme yaklaşımları, uygulama gereksinimlerine ve kullanılabilir veri türlerine göre çeşitlenmektedir. Her bir yaklaşımın benzersiz avantajları ve zorlukları vardır ve belirli bir durum için en uygun yöntemin seçimi, sorunun doğasına ve eldeki verilere bağlıdır.



Şekil 2.5 Anomali belirleme yöntemleri (Goldstein ve Uchida, 2016)

Şekil 2.5, farklı anomali belirleme yöntemlerini görsel olarak özetlemektedir. Üst kısımda yer alan "Denetimli Anomali Tespiti" bölümünde, modelin eğitim verisi kullanılarak oluşturulduğu ve bu modelin test verisi üzerinde anomali tespiti yaptığı

gösterilmektedir. Eğitim verisi hem olağan hem de anormal örnekler içermektedir. Orta kısımda, "Yarı Denetimli Anomali Tespiti" yöntemi yer almakta olup, burada model yalnızca olağan örnekleri içeren eğitim verisiyle eğitilmekte ve ardından test verisi üzerinde anomali tespiti yapılmaktadır. En alt kısımda ise, "Denetimsiz Anomali Tespiti" yöntemi gösterilmekte olup, bu yöntemde etiketsiz veri kullanılarak denetimsiz bir yöntem yardımıyla anomalilerin tespit edildiği açıklanmaktadır . Bu şekil, farklı anomali tespit yöntemlerinin eğitim ve test aşamalarındaki farklılıklarını net bir şekilde ortaya koymaktadır.

Anomali tespiti, sürekli gelişen bir alan olup, yeni yöntemlerin ve teknolojilerin ortaya çıkışı ile evrimleşmeye devam etmektedir. Araştırmacılar ve uygulayıcılar için, veriye özgü zorlukları ele alabilen ve spesifik uygulama gereksinimlerini karşılayabilen en uygun yaklaşımı seçmek önemlidir. Bu süreçte, çeşitli sektörlerden gelen gerçek dünya uygulamalarından elde edilen bilgi ve deneyimler, anomali tespiti alanındaki yeniliklerin ve ilerlemelerin anahtarını oluşturur.

BÖLÜM ÜÇ
FİNANSAL İŞLEMLERDE SAHTEKARLIK

3.1 Sahtekarlık Tanımı

Sahtekarlık, genel olarak 'bir kişinin diğerine karşı dürüst olmayan bir avantaj elde etmeyi amaçladığı herhangi bir davranış' olarak tanımlanabilir. Başka bir deyişle, sahtekarlık, bir kişiye haksız kazanç, diğerine ise haksız kayıp sağlamayı amaçlayan bir eylem veya ihmaldir (Chakrabarty, 2013). Finansal sahtekarlık ise, genellikle maddiyat ve itibar kaybına yol açan, yasal yaptırımlara konu olabilen ve birçok sektörde farklı türleri görülebilen bir terimdir. Sanal suçlar finansal sahtekarlığın önemli bir parçasını oluşturmaktadır. Çalışmanın ana konusu olan e-sahtekarlık ise sanal suçlar altında incelenmektedir. Sanal suçlar literatürünü incelediğimizde, e-sahtekarlıkla ilgili birden fazla tanım yapılmıştır. Literatürde sıklıkla atıfta bulunulan Button ve Cross (2017), e-sahtekarlığı "internet ve diğer bilgi iletişim teknolojilerini kullanarak yapılan, kurbanları aldatma amacı güden, maddi veya manevi zararlara yol açan sahtekarlık türü " olarak tanımlamıştır. Amerika Adalet Bakanlığı, e-sahtekarlığı "sahte talepler sunmak için İnternet'in bir veya daha fazla bileşenini - sohbet odaları, e-posta, mesaj panoları veya Web siteleri gibi - kullanan bir sahtekarlık şeması" olarak tanımlamaktadır (Akintoye ve Araoye, 2011). Graham (2002) ise e-sahtekarlığı daha geniş bir kavram olarak ele alarak "birinin dürüst olmayan bir avantaj elde etmeyi amaçladığı bilgisayarlaşma ile bağlantılı sahtekarlık davranışı" olarak tanımlamıştır. E-sahtekarlığın tanımı, gözlemlendiği zemine göre değişkenlik göstermektedir. Genel olarak sahtekarlık, elektronik ödeme sistemlerinde gerçekleşen yetkisiz faaliyetlerdir; bu faaliyetler yasalarca yasaklanmalı ve yasadışı olarak kabul edilmelidir (RamaKalyani ve UmaDevi, 2012). Sahtekarlık, finansal sistemler, telekomünikasyon, kamu ve özel hizmetler gibi çeşitli alanlarda ortaya çıkabilir.

3.2 Kredi Kartı Sahtekarlığı

Kredi kartı sahtekarlığı, modern finansal dünyada giderek artan bir endişe kaynağıdır. E-ticaretin ve küresel finansal işlemlerin yükselişi, kredi kartı sahtekarlığının yaygınlaşmasına neden olmuştur. Bu sorun, hem bireyler hem de finansal kurumlar için önemli mali kayıplara yol açmaktadır. Bu durum, sahtekarlık faaliyetlerini etkili bir şekilde tespit etmek ve önlemek için gelişmiş sahtekarlık tespit sistemlerinin geliştirilmesi ve uygulanmasını zorunlu kılmaktadır (Ghosh ve Reilly, 1994).

Finansal işlemlerde sahtecilik ise başkalarına ait kredi kartı ile herhangi bir işyeri veya ATM'den işlem yaparak haksız kazanç sağlamak ve bu sayede kartın gerçek sahibini, işyerini ya da bankayı zarara uğratmaktır.

Temel olarak iki tip sahtecilik vardır;

- Fiziksel ortamda yapılan sahtecilikler
- Sanal ortamda yapılan sahtecilikler

Fiziksel ortamda yapılan sahteciliklerde kart hamilinin rızası ve izni olmadan, kartın temassız özelliğini kullanarak veya ele geçirilen şifrenin girilmesi yoluyla işlem yapılır. Bu tür sahtecilikler kayıp, çalıntı ya da kopyalanmış (sahte) kartlar ile çoğunlukla şifresiz olarak yapılmaktadır. Bu sahtecilik en çok kuyumcu, akaryakıt istasyonu ve market gibi işletmelerde görülmektedir. Temassız ödeme özelliği olan kartların yaygınlaşması bu tarz sahtekarlıkların artmasına yol açmıştır. Bankalar bu durumlar için, belli sayıda temassız işlemden sonra şifre girilmesini istemek, temassız işlem limitini kısıtlamak gibi önlemler alırlar. Sanal ortamda gerçekleştirilen sahteciliklerde dolandırıcı, fiziken işyerinde bulunmamaktadır. Dolandırıcı, işlemleri gerçekleştirmesini sağlayacak hassas bilgilere erişmiş ve bu bilgileri kullanarak haksız kazanç elde etmeye çalışmaktadır (Akintoye ve Araoye, 2011). Bu tür sahteciliğin en yaygın görüldüğü sektörler, uygulama indirme mağazaları, bilet satışı yapan web siteleri ve elektronik ürün satışı yapan firmalardır.

Bu tip sahtecilik işlemlerinin önlenmesi için en etkili yöntem, güçlü bir sahtecilik önleme yazılımı ve deneyimli bir sahtekarlık önleme ekibidir. Şirketler, bu ekipleri kendi bünyelerinde barındırabileceği gibi dış firmalardan da destek alarak bu hizmeti sağlayabilirler. Bankalar ve sanal pos hizmeti sunan firmalar tarafından genellikle tercih edilen profesyonel sahtecilik önleme yazılımları, kural bazlı ve gelişmiş filtreleme sistemleri ile donatılmıştır.

Türkiye Bankalar Birliği, Ekim 2015'te yayımladığı raporda e-sahtekarlığı dört ana başlıkta incelemiştir. Bu başlıklar sırasıyla şöyledir:

1. Kart başvuru sahtecilikleri
2. Kartın hesabını ele geçirme
3. Kayıp çalıntı kart
4. Sahte kart

Kart başvuru sahteciliklerinde, dolandırıcı illegal yollarla ele geçirdiği kimlik bilgilerini kullanarak internet bankacılığı üzerinden kart talebinde bulunur. Bu tür sahtekarlık, bankalar için ciddi mali kayıplara yol açabilir. Kart hesabını ele geçirme, yabancı literatürde "Account Takeover" olarak adlandırılır ve dolandırıcının amacı, kredi kartı limitini veya karta bağlı hesap bakiyesini boşaltmaktır (Doerfler vd., 2019). Bu sahtekarlık türü, kartın kendisinin veya kartla işlem yapılabilecek asgari bilgilerin ele geçirilmesiyle gerçekleştirilir. Bu bilgiler sayesinde dolandırıcılar, ek kart çıkarma gibi işlemleri yaparak kart başvuru sahteciliğinde de bulunabilirler.

Kayıp veya çalıntı kart durumlarında, kart bilgilerinin fiziksel veya dijital olarak ele geçirilmesi söz konusudur. Alışveriş yapılan sitenin veri tabanlarından elde edilebilecek maskeli kart bilgilerinin çalınması sonucu dolandırıcı, kart bilgilerini internet üzerinden kullanarak maddi fayda sağlar (Jones, 2019). Günümüzde sıkça karşılaşılan kart sahteciliğinde, dolandırıcı gerçek ve geçerli kart bilgileri ile sahte bir kart oluşturur ve bu kart ile çeşitli işlemler yapar. Sahte kart oluşturmanın birinci koşulu, orijinal kart bilgisini ele geçirmektir. Bu bilginin ele geçirilmesi iki yöntemle

gerçekleşebilir. Birincisi, kart verisinin ATM veya POS cihazları üzerine yerleştirilen özel cihazlar ile kopyalanmasıdır. İkincisi, kart bilgilerinin depolandığı veri tabanından çalınarak elde edilen kart numaralarının sahte kartlara yüklenmesidir (Adeoti, 2011) .

Türkiye Bankalar Birliği'nin "Bankacılıkta Dolandırıcılık Eylemleri Tespit ve Önleme Yöntemleri" raporunda sıkça bahsedilen veri tabanı sızıntılarının üç ana sebebi vardır (Türkiye Bankalar Birliği, 2015). Birincisi, herhangi bir sebeple veri güvenliği politikalarının ihlal edilmesidir. Bu durum veri sızıntılarının başlıca sebeplerinden biridir. İkincisi, kaybolan bilgisayarlar ve e-posta adreslerinin önemli veri kaybı riski yaratmasıdır. Üçüncüsü ise, iş süreçleri ve yöntemleri gibi hassas verilerin kötü niyetli kişilerce öğrenilmesi, ele geçirilmesi veya saldırılara açık halde bulunmasıdır. Özellikle eskimiş veya izinsiz iş süreçlerinin korunmaması veya güncellenmemesi, sızıntı riskini artırmaktadır.

Son yıllarda hızla artan internet kullanımı, alışveriş alışkanıklarımızında değişmesine sebep oldu. İnternet üzerinden harcamalarımızın artmasıyla birlikte internet sahtekarlıkları da hızla artışa geçti (Yazdanifard vd., 2011). İnternetteki hacim arttıkça dolandırıcılar e-sahtekarlık yapmak için etkili ve daha karmaşık yollar geliştirmişlerdir (Fernandes, 2013). Avrupa komisyonu 2008 yılında yayınladığı raporunda sahtekarlık sorunu için "Ödeme araçlarına karşı sahtekarlık (e-sahtekarlık), ödemeler için iç pazarın başarısı için bir tehdit olmaya devam ediyor. Ödeme sahtekarlığı, tüketicinin nakit dışı ödeme araçlarına olan güvenini ve nihayetinde reel ekonomiyi etkiliyor" demiştir.

E-sahtekarlık süreci her vaka için çeşitlilik göstermekle birlikte belli bir şema içerisinde ilerler. Bu genel şema sayesinde müdahale edilebilecek noktalar daha doğru tespit edilebilir.

Kredi kartı sahtekarlığını tespit etmek için kullanılan çeşitli yöntemler vardır. Kural tabanlı sistemler, anomali tespiti için önceden belirlenmiş kuralları kullanır, ancak bu sistemler genellikle yüksek yanlış pozitif oranlarına sahiptir (Ghosh ve Reilly, 1994). Yapay sinir ağları (ANN), sahtekarlık davranışındaki yeni desenleri

öğrenme ve uyum sağlama yetenekleri nedeniyle yaygın olarak kullanılır ve tarihsel verilere dayanarak işlemleri sahte veya sahte olmayan olarak sınıflandırabilirler. ANN'ler, yüksek doğruluk ve düşük yanlış pozitif oranlarına sahip olmakla birlikte, önemli hesaplama kaynakları ve büyük veriler gerektirir. Bayes ağları, olasılıksal modeller kullanarak bir işlemin sahte olma olasılığını tahmin eder ve yeni verileri kademeli olarak ekleyebilirler. Ancak, bu ağlar da önemli hesaplama gücü ve olasılıksal modelleme uzmanlığı gerektirir (Raj ve Portia, 2011).

Etkili kredi kartı sahtekarlığı tespit sistemleri, mali kayıpları en aza indirmek ve finansal işlemlere olan güveni sürdürmek için hayati öneme sahiptir.

BÖLÜM DÖRT
YÖNTEM

Özellikle kredi kartı sahtekarlığı gibi finansal suçların tespitinde, müşteri harcama kalıplarının belirlenmesi ve kullanılması oldukça önemlidir. Bu kalıpların analizi, sahtekarlık tespit sistemlerinin hassasiyetini ve doğruluğunu artırabilir (Chung ve Lee, 2023). Bu doğrultuda önerilen yaklaşım olağan müşteri davranışlarını belirleme üzerine kurgulanmıştır.

Bu bölümde, şüpheli harcamaların tespiti için önerilen iki aşamalı yaklaşım açıklanmıştır. Çalışmada kullanılan terimler ve açıklamalarına Tablo 4.1'de yer verilmiştir.

Tablo 4.1 Kullanılan terimler

TERİM	**AÇIKLAMA**
MÜŞTERİ	Çevrim içi satış yapan işletme
SEKTÖR	Müşterinin faaliyet gösterdiği alan
İŞLEM	Kredi kartı ile gerçekleştirilen harcama
TUTAR	İşlem tutarı
FİNANS KURULUŞU	Çevrim içi kredi kartı ödeme hizmetini sağlayan bankalar veya ödeme hizmetleri ve elektronik para kuruluşları

İnternet bankacılığı sahtekarlıkları giderek daha karmaşık hale gelmekte ve geleneksel tek aşamalı tespit yöntemleri bu tür karmaşık sahtekarlıkları yakalamakta yetersiz kalabilmektedir. Bu nedenle, iki aşamalı yaklaşım, hem özel hem de genel kullanıcı profillemesi kullanarak sahtekarlık tespitinde daha etkili olmayı hedeflemektedir (Erbayat ve Yıldıztepe, 2023). Benzer bir iki aşamalı sahtekarlık tespiti "Banksealer" 2015 yılında Carminati ve diğerleri tarafından önerilmiştir (Carminati vd., 2015). Banksealer yönteminin ilk aşamasında, bireysel kullanıcıların geçmiş işlem davranışlarını modelleyerek, her yeni işlemi bu profillere karşı anomali açısından değerlendirir. İkinci aşaması ise, benzer harcama alışkanlıklarına sahip

kullanıcıların genel kümelerini belirleyerek, bireysel profillerin yetersiz kaldığı durumlarda destek sağlar. Carminati ve diğerlerine (2015) göre bu iki aşamalı yapı, sahtekarlık tespitinde iki önemli fayda sunar:

1. Daha Yüksek Tespit Oranları: Yerel profilleme, kullanıcıların alışılmış davranışlarından sapmaları tespit ederken, küresel profilleme benzer kullanıcı gruplarını analiz ederek yanlış pozitifleri azaltır ve tespit doğruluğunu artırır.

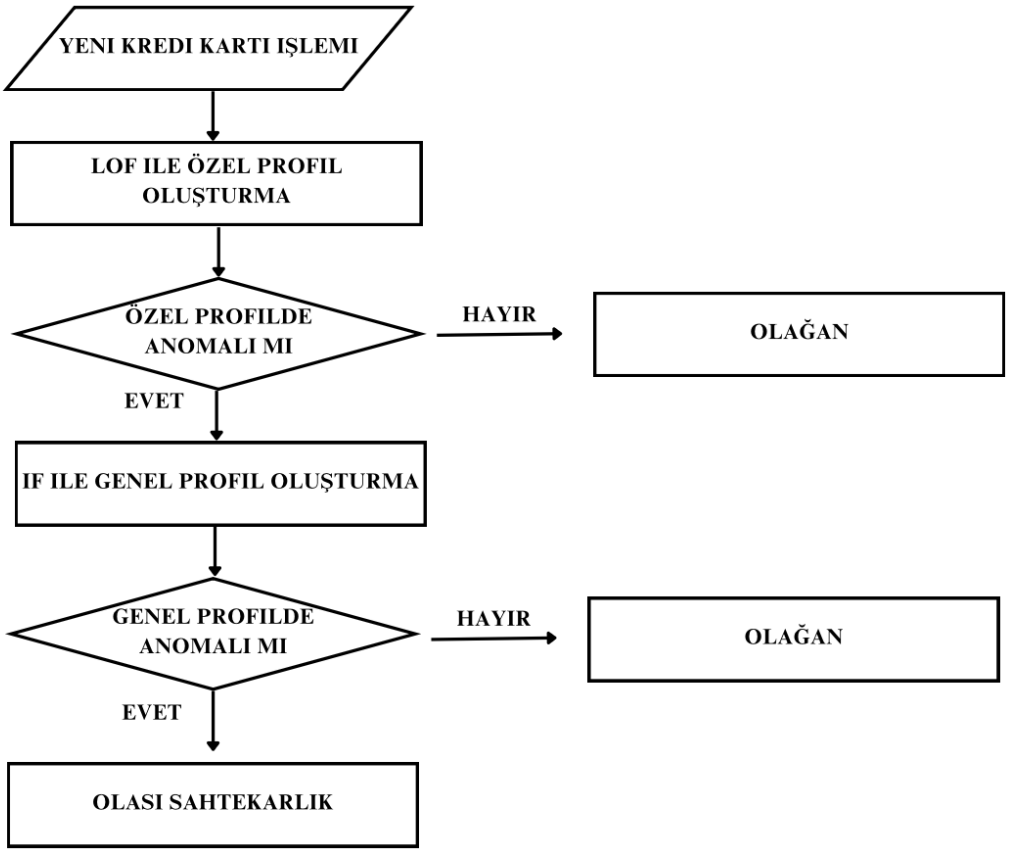
2. Uyarlanabilirlik ve Esneklik: Kullanıcı davranışlarındaki değişikliklere uyum sağlama kapasitesi, iki aşamalı yaklaşımın zaman içinde etkinliğini korumasını sağlar. Bu yöntem, değişen harcama alışkanlıklarını ve yeni sahtekarlık tekniklerini daha hızlı ve etkili bir şekilde tespit edebilir.

BankSealer, bu iki aşamalı yaklaşımı başarılı bir şekilde uygulayarak çevrimiçi bankacılık sahtekarlıklarını analiz etmek ve araştırmak için etkili bir karar destek sistemi sunmaktadır (Carminati vd., 2015).

Bu tezde önerilen "FindOut" yaklaşımında benzer şekilde, iki aşamalı profilleme benimsenerek, müşterilerin güvenliğini artırmak ve sahtekarlık tespitinde daha yüksek doğruluk oranları elde etmek amaçlanmıştır.

4.1 FindOut Yaklaşımı

“FindOut” adı verilen yaklaşım, özel ve genel profillerin belirlenmesi (profilleme) ve şüpheli işlem tespiti (değerlendirme) aşamalarından oluşmaktadır. Profilleme aşaması, derleme ve profil oluşturma süreçleri ile gerçekleştirilir. Derleme aşamasında ilk olarak, müşteri bilgisi, sektör bilgisi, harcama tutarı ve işlem zamanı değişkenlerinden oluşan 30 günlük kredi kartı harcama verileri elde edilir. Yaklaşımda son 30 günlük veriler kullanılarak enflasyon ve mevsimsellik etkisi en aza indirilmiştir. Tutar değişkeni Min-Max tekniği ile normalleştirilmelidir. Normalizasyon, veri özelliklerinin ölçeklenmesi sayesinde, anomali tespit yöntemlerinin performansını artırır (Seliya vd., 2021). Derleme aşaması her bir



Şekil 4.1 FindOut Yaklaşımı

müşteri için yapılır. Böylece faaliyet gösterilen sektöre göre de ayrım yapılmış olur. Şekil 4.1'de FindOut yaklaşımının mimarisi gösterilmiştir.

4.1.1 Profilleme Aşaması

Özel profil, her bir müşteri için işlem davranışlarını temsil ederken, genel profil daha geniş bir veriden türetilen ve her bir sektör için geçerli olan işlem davranışlarını yansıtır. Literatürde daha önce de kullanılan, bu iki düzeyli profil oluşturma yaklaşımı, yöntemin daha etkin olmasını amaçlar (Carminati vd., 2015). Finansal sahtekarlık tespitinde hızlı yanıt süreleri, potansiyel kayıpların minimize edilmesi açısından kritik öneme sahiptir . Bu nedenle, FindOut yaklaşımının özel profil oluşturulma aşamasında müşterinin tüm geçmiş verilerini kullanmak yerine son 30 günlük işlemler seçilerek yaklaşım hızlandırılır.

Çoğu finans kuruluşu her sektörden müşteri ile çalışmaktadır. Bu nedenle bu kuruluşların elinde yeni gelen müşterilerin harcama geçmişi olmamasına karşın müşterinin içinde bulunduğu sektörle ilgili geniş bir harcama verisi bulunmaktadır. Müşterilerin işlemleri içindeki bulundukları sektörün harcama davranışlarıyla benzerlik göstermektedir. Bu nedenle yaklaşımdaki genel profil müşterilerin ait oldukları sektörün 30 günlük harcama verileri ile oluşturulur.

4.1.2 Değerlendirme Aşaması

Bu çalışmada anomali tespitinde dört farklı yöntem kullanılmıştır. Verinin yapısı, performans talepleri gibi değişkenlere göre her profilde kullanılacak yöntem değişmektedir. Model kurulum aşamasında eldeki verinin ve kullanım alanının gerekliliğine göre yöntem seçilip modele önceden parametre olarak girilmelidir. Kullanılan yöntemler aşağıda listelenmiştir ;

- Yerel aykırı değer ,
- İzolasyon ormanı,
- Tek sınıflı destek vektör makineleri ,
- Eliptik zarf

Her yöntemin avantajları ve dezavantajları bulunmaktadır. Bu yöntemler detaylı olarak bölüm 4.2'de incelenmiştir. Model eğitimi aşamasında sahtekarlık sınaması yapılacak gözlemin müşteri ve sektör bilgileri alınmaktadır. İlk olarak müşterinin özel profili ile seçilen yöntem kullanılarak model oluşturulur. İncelenen işlem son olarak içinde bulunduğu sektörün işlemlerini temsil eden genel profil ile değerlendirilir.

Modellerin oluşturulmasından sonra yeni yapılan işlem özel profil kullanılarak anomali sınamasına girer, bu aşamada işlem geçmiş müşteri harcama alışkanlığıyla kıyaslanmaktadır. Eğer işlem, bu ilk değerlendirme aşamasında şüpheli olarak sınıflandırılırsa, ikinci aşamada genel profil kullanılarak tekrar sınanır. İşlemin her iki

profil tarafından da şüpheli olarak değerlendirilmesi durumunda "olası sahtekarlık" (potential fraud) olarak etiketlenir. Bu süreç şüpheli harcamaların belirlenmesini sağlar.

Özetle, FindOut yaklaşımı, hem bireysel kullanıcı harcama alışkanlıklarını hem de sektördeki harcama alışkanlıklarını kullanarak sahtekarlık tespit süreçleri için bir çerçeve sunar. Bu çift katmanlı yaklaşım, sahtekarlık tespit yöntemlerinin hem gerçek pozitif oranını hem de gerçek negatif oranını artırarak, finansal kurumların sahtekarlık kayıplarını azaltmasına ve yanlış pozitif oranlarını düşürmesine katkıda bulunabilir.

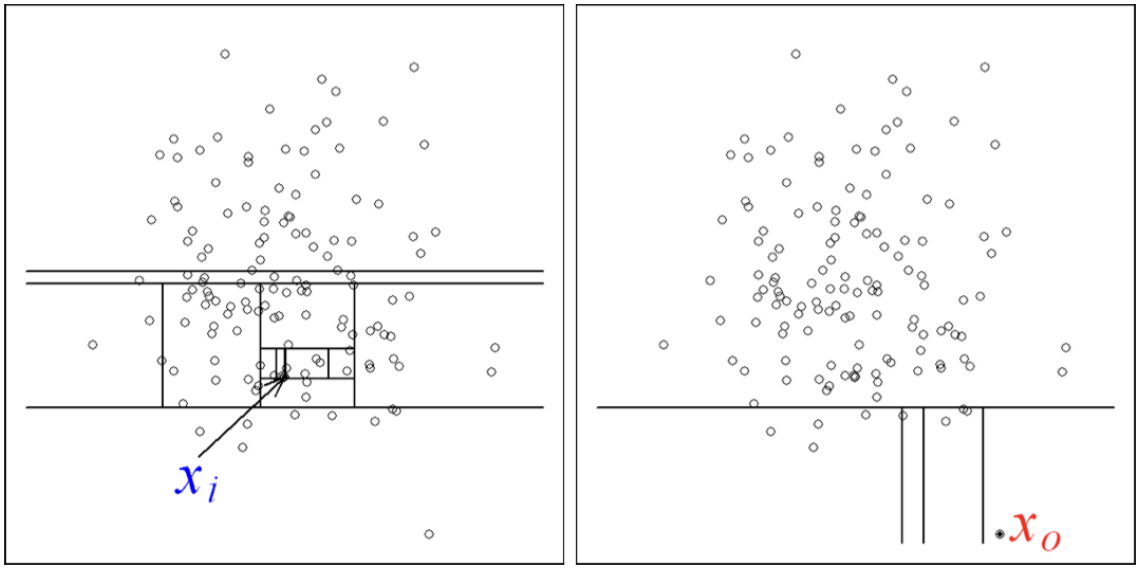
4.2 Kullanılan Anomali Tespit Yöntemleri

Bu çalışmada kullanılan dört yöntem, verinin yapısına, anomali tespitinin gerektirdiği performans taleplerine ve gözlemlerin sınırlı sayıda olduğu durumlarda en yüksek doğruluğu sağlayabilecek tekniklere dayalı olarak seçilmiştir. Bu yöntemlerin kombinasyonu, sahtekarlık tespitinde hem özel hem de genel profiller üzerinden yapılan sınamaların doğruluğunu artırmak amacıyla tercih edilmiştir.

4.2.1 İzolasyon Ormanı

IF yöntemi, özellikle anomali tespiti ve aykırı değerlerin belirlenmesi için kullanılan güçlü bir yöntemdir. Liu vd. (2008) tarafından geliştirilen bu yöntem, veri noktalarının rastgele alt kümelerini oluşturarak ve bu alt kümeler içinde rastgele bölünmeler yaparak çalışır. Yöntem, aykırı değerlerin diğer verilere kıyasla daha kolay izole edilebileceği prensibine dayanır. İzolasyon süresi kısa olan noktalar aykırı değer olarak değerlendirilir.

IF yöntemi, karar ağaçları kullanarak veriyi rastgele böler ve her bir veri noktası için izolasyon derinliği hesaplar. Aykırı değerlerin izolasyon derinliği daha düşüktür, çünkü bu noktalar daha az bölünme gerektirir. Bu sayede yöntem, aykırı değerleri daha hızlı bir şekilde belirler (Liu vd., 2008).



Şekil 4.2 IF ile gözlemin izole edilmesi (Liu vd., 2008)

Şekil 4.1'de, veri noktalarının izolasyon sürecini gösteren bir grafik bulunmaktadır. Grafikte, x_0 noktası diğer veri noktalarından farklı bir konumda yer aldığı için aykırı değer olarak izole edilmiştir. Grafiğin sol tarafında yer alan x_i noktası, verilerin yoğun olduğu bir bölümde yer almakta bu nedenle olağan bir gözlem olarak değerlendirilmektedir.

İzolasyon Ağacı (iTree) İnşası: Verilen veri $X = \{x_1, x_2, \ldots, x_n\}$ rastgele alt örnekleme yöntemi kullanılarak alt kümelere bölünür ve her bir alt küme üzerinde bir izolasyon ağacı inşa edilir.

İzolasyon Derinliği $h(x)$: Bir veri noktası x için izolasyon derinliği, x'in bulunduğu yaprak düğüme kadar olan yolun uzunluğudur. Bu, x'in izole edilmesi için gereken bölünme sayısıdır.

$C(n)$ Fonksiyonu: Bir veriden rastgele seçilen bir noktanın izolasyon derinliğinin beklenen değeri, $C(n)$ ile ifade edilir. Bu fonksiyon, bir tam sayı n için Harmonik Sayı kullanılarak yaklaşık olarak hesaplanabilir:

$$C(n) \approx 2H(n-1) - \frac{2(n-1)}{n} \tag{4.1}$$

Burada $H(i)$ Harmonik sayı olup, $H(i) = \sum_{k=1}^{i} \frac{1}{k}$ şeklinde tanımlanır.

Ortalama İzolasyon Derinliği: Bir veri noktası x için ortalama izolasyon derinliği $E(h(x))$ tüm iTree'ler üzerindeki izolasyon derinliklerinin ortalaması alınarak hesaplanır.

Anomali Skoru: Anomali skoru $s(x,n)$, x'in ortalama izolasyon derinliği ile normalleştirilmiş izolasyon derinliği arasındaki farkın bir fonksiyonudur:

$$s(x,n) = 2^{-\frac{E(h(x))}{C(n)}} \tag{4.2}$$

Burada, $s(x,n)$ skoru 0 ile 1 arasında bir değer alır. Skorun 1'e yakın olması, x'in bir anomali olduğunu gösterir, skorun 0'a yakın olması ise x'in normal bir veri noktası olduğunu gösterir.

IF yöntemi, iki temel parametreye dayanır; oluşturulacak ağaç sayısı ve alt örnekleme boyutu. Liu vd. (2008), bu yöntemin yüksek tespit performansı elde etmek için yalnızca küçük bir alt örnekleme boyutuna ihtiyaç duyduğunu belirtmişlerdir. Ek olarak, Hariri vd. (2019) IF yönteminin ağaçların dallanma yapısından kaynaklanan bir yanlılık problemi olduğunu ifade etmişlerdir.

4.2.2 Yerel Aykırı Değer

LOF yöntemi, veri madenciliğinde anomali tespiti için kullanılan bir tekniktir. Bu yöntem, bir veri noktasının, çevresindeki diğer noktalara kıyasla ne kadar izole olduğunu ölçerek anomalileri belirler (Breunig vd., 2000). LOF, bir veri noktasının yerel yoğunluğunu hesaplar ve bu yoğunluğu, komşularının yoğunlukları ile karşılaştırır. Yerel yoğunluk, veri noktası ile komşuları arasındaki uzaklık ölçümlerine dayanır. LOF skoru, bir veri noktasının anomali olup olmadığını belirlemek için kullanılır; yüksek LOF skorları genellikle anomalileri işaret eder. Yöntemin adımları aşağıdaki gibi listelenebilir.

Komşuların Belirlenmesi (k-nearest neighbors, k-NN) : Her veri noktası için k sayıda en yakın komşu (k-nearest neighbors) belirlenir. Bu adım, veri noktasının yerel yakınlığını hesaplamak için gereklidir (Breunig vd., 2000).

Erişim Uzaklığı (Reachability Distance) : Erişim uzaklığı, bir veri noktasının komşusuna olan uzaklığıdır, ancak bu uzaklık en azından komşunun k-komşularındaki en uzak veri noktasına olan uzaklığı kadar olmalıdır. Erişim uzaklığı $d_{\text{erisim}}(A, B)$ k-komşusu B olan veri noktası A için şu şekilde tanımlanır:

$$d_{\text{erisim}}(A, B) = maks(d(A, B), k - \text{uzaklik}(B)) \tag{4.3}$$

Yerel Erişim Uzaklığı Yoğunluğu (Local Reachability Density, LRD) : Bir veri noktasının yerel erişim uzaklığı yoğunluğu, komşularının ortalama erişim uzaklığına dayanır. Bu yoğunluk, $LRD(A)$ ile ifade edilir ve şu şekilde hesaplanır:

$$LRD(A) = \left(\frac{\sum_{B \in kNN(A)} d_{erisim}(A, B)}{|kNN(A)|} \right)^{-1} \tag{4.4}$$

Burada, $kNN(A)$ A veri noktasının k-komşularının kümesidir.

Yerel Aykırı Değer : LOF skoru, bir veri noktasının yerel erişim uzaklığı yoğunluğunun komşularının yoğunluklarına oranıdır. LOF skoru şu şekilde tanımlanır:

$$LOF(A) = \frac{\sum_{B \in kNN(A)} \frac{LRD(B)}{LRD(A)}}{|kNN(A)|} \tag{4.5}$$

Bu formül, veri noktası A'nın yerel yoğunluğunun, komşularının yoğunluklarına kıyasla nasıl olduğunu ölçer. LOF skoru 1'den büyükse, veri noktası anormal olarak kabul edilir.

LOF yöntemi yerel yapıyı dikkate alarak anomali tespiti yapar, bu nedenle kümeleme yapılarına duyarlıdır (Breunig vd., 2000). Verinin yoğunluk değişikliklerini dikkate alarak daha hassas tespitler yapabilir. Ancak, LOF'un

hesaplama maliyeti yüksektir, özellikle büyük veri kümelerinde performans sorunları yaşanabilir. Ek olarak k değeri seçimi, sonuçların doğruluğunu önemli ölçüde etkileyebilir ve bu parametrenin belirlenmesi genellikle deneysel olarak yapılır (Breunig vd., 2000).

LOF yöntemi, birçok farklı alanda anomali tespiti için başarıyla uygulanmıştır. Örneğin, kredi kartı sahtekarlığı tespiti, ağ güvenliği, tıbbi teşhis ve sanayi süreçlerini izleme gibi çeşitli uygulamalarda kullanılmıştır (Chandola vd., 2009). LOF yöntemi, veri madenciliğinde anomali tespiti için güçlü ve esnek bir araçtır. Yerel yoğunluk farklarını kullanarak anomali tespiti yapması, bu yöntemi birçok farklı veri türünde etkili kılar. Ancak, yüksek hesaplama maliyeti ve parametre seçiminin zorluğu gibi dezavantajlar göz önünde bulundurulmalıdır.

4.2.3 Destek Vektör Makineleri

SVM, sınıflandırma ve regresyon analizinde yaygın olarak kullanılan güçlü bir denetimli öğrenme yöntemidir. SVM, verilerde iki sınıf arasındaki (anomali ve olağan) en iyi ayrımı sağlayan hiper düzlemi bulmayı amaçlar. SVM, iki sınıfı ayıran doğrusal bir hiper düzlem oluşturur. Bu hiper düzlem, sınıflar arasındaki mesafeyi en büyük hale getirecek şekilde yerleştirilir. Hiper düzlem, veri noktalarının sınıflarını en iyi şekilde ayıran sınırdır. SVM ayrıca doğrusal olmayan ayırma problemleri için de çekirdek fonksiyonları kullanarak yüksek boyutlu uzaylarda çalışabilir (Cortes ve Vapnik, 1995).

SVM yöntemi yüksek boyutlu verilerle iyi performans gösterir. Ancak, SVM yönteminin bazı dezavantajları mevcuttur. Büyük verilerde hesaplama maliyeti yüksektir. Buna ek olarak çekirdek ve parametre seçimleri sonuçların doğruluğunu etkileyebilir (Schölkopf ve Smola, 2002).

4.2.4 Tek Sınıflı Destek Vektör Makineleri

OC-SVM , geleneksel SVM yönteminin bir uzantısıdır. OC-SVM, anomali tespiti için kullanıldığında belirlenen sınıfın dışındaki veriler anomali olarak nitelendirilir (Schölkopf vd., 2001). Bu yöntem, verilerin çoğunluğunu kapsayan bir hiper düzlem oluşturur ve bu düzlemin dışındaki verileri anomali olarak değerlendirir. OC-SVM, özellikle çok sayıda pozitif örneğe sahip verilerde etkilidir. OC-SVM yönteminin adımları aşağıda listelenmiştir.

Veri Dönüştürme: Veriler, bir çekirdek fonksiyonu kullanılarak yüksek boyutlu bir uzaya dönüştürülür. En yaygın kullanılan çekirdek fonksiyonlar, doğrusal (linear), polinom (polynomial) ve radyal bazlı fonksiyonlardır (radial basis function).

Hiper Düzlem Oluşturma: OC-SVM, verileri kapsayan bir hiper düzlem oluşturur. Bu hiper düzlem, verilerin yoğun olduğu bölgeyi tanımlar ve bu bölgenin dışındaki veriler anomali olarak kabul edilir. Hiper düzlemin denklemi şu şekildedir:

$$f(x) = w \cdot \phi(x) - \rho \tag{4.6}$$

Burada, w ağırlık vektörü, $\phi(x)$ çekirdek fonksiyonu ile dönüştürülmüş veri noktası ve ρ ise modelin bir parametresidir (Schölkopf vd., 2001).

Optimizasyon Problemi: OC-SVM, aşağıdaki optimizasyon problemini çözerek en iyi hiper düzlemi bulur:

$$\min_{w,p,\xi} \frac{1}{2}\|w\|^2 + \frac{1}{vn}\sum_{i=1}^{n}\xi_i - p \tag{4.7}$$

Buna bağlı sınırlamalar ise ; $(w \cdot \phi(x_i)) \geq p - \xi_i, \xi_i \geq 0$

Anomali Tespiti: Eğitim aşamasından sonra, yeni bir veri noktasının anomali olup olmadığına karar vermek için tekrar $f(x)$ fonksiyonu kullanılır. Eğer $f(x) \leq 0$ ise, veri noktası anomali olarak kabul edilir.

OC-SVM, anomali tespiti için güçlü ve esnek bir yöntem olduğundan, makine arızası tespiti, kredi kartı sahtekarlığı, biyometrik tanıma ve ağ güvenliği gibi birçok alanda anomali tespiti için başarıyla kullanılmıştır (Tax ve Duin, 2004). OC-SVM yöntemi yüksek boyutlu verilerle etkili bir şekilde çalışabilir. Bunun yanı sıra, sınırlı sayıda pozitif örnekle bile iyi performans gösterebilir (Schölkopf vd., 2001). Bu nedenle önerilen yaklaşımda SVM yerine OC-SVM kullanılmıştır.

Yüksek boyutlu verilerle etkili bir şekilde çalışabilmesi ve sınırlı pozitif örneklerle bile iyi performans göstermesi, bu yöntemi çeşitli uygulamalarda tercih edilen bir araç haline getirir. Bununla birlikte, OC-SVM kullanımının bazı kısıtlamaları bulunmaktadır. Çekirdek fonksiyonunun ve parametrelerin seçimi, sonuçların doğruluğunu önemli ölçüde etkilemektedir.

4.2.5 Eliptik Zarf

Elliptik zarf yöntemi, Mahalanobis uzaklığını kullanarak veri noktalarının çok değişkenli normal dağılıma ne kadar uyduğunu ölçer. Bu yöntem, veri içindeki merkezi ve kovaryans matrisini hesaplar ve bu bilgilere dayanarak bir elips sınır oluşturur. Verinin bu elips sınırının dışında kalan kısımları, aykırı değer olarak kabul edilir (Rousseeuw ve Driessen, 1999).

Yöntem özellikle finansal analiz, biyomedikal veriler ve mühendislik uygulamaları gibi çeşitli alanlarda anomali tespiti amacıyla kullanılmıştır. Aynı zamanda, Filzmoser vd. (2005) çalışmalarında bu yöntem, verilerdeki anomali tespiti için kullanılmış ve yüksek performans sergilemiştir.

EE yöntemi, elipsin şeklini ve boyutunu tahmin etmek için hızlı-minimum kovaryans belirleyicisini (FAST-MCD) (Rousseeuw ve Driessen, 1999) kullanır ve bu tahminden elde edilen Mahalanobis uzaklığı, aykırı değerleri veya anormallikleri belirlemek için eşiği tanımlamak için kullanılır. FAST-MCD , özellikle verideki aykırı değerlerin etkisini minimize etmek için geliştirilmiştir ve büyük veriyle başa çıkabilecek şekilde optimize edilmiştir. FAST-MCD, başlangıçta verilerin rastgele bir

alt kümesini seçerek başlar ve bu alt küme üzerinden determinantı minimize eder. Bu süreç, aykırı değerlerin belirlenmesini ve daha sonra sağlam bir merkezi eğilim ve yayılım ölçüsü sağlamak için kullanılır.

Eliptik zarf verileri Gauss dağılımına göre modeller ve verilerin çoğunu içeren bir elips tanımlar. Elipsin dışındaki bir veri noktası anomalidir. Elipsin boyutu ve şekli, kovaryans matrisinin determinantı yakınsayana kadar verilerden alınan alt örneklerin Mahalanobis uzaklığı (bir veri noktasının ortalamadan kaç standart sapma uzakta olduğunun bir ölçüsü) iteratif olarak hesaplayan FAST-MCD algoritması (Rousseeuw ve Driessen, 1999) tarafından belirlenir.

4.3 Performans Değerlendirme

FindOut yaklaşımı, sadece olağan işlemleri kullanarak çalıştığından yarı denetimli anomali belirleme olarak değerlendirilmektedir. Yaklaşımın sahtekarlık tespit başarısını belirlemek için yanılgı matrisinden (confusion matrix) yararlanarak , duyarlılık (recall), kesinlik (precision) ile F-Beta puanı kullanılmıştır.

Tablo 4.2 Yanılgı Matrisi

Yanılgı Matrisi		**Gerçek Durum**	
		Pozitif	**Negatif**
Tahmin	**Pozitif**	Doğru Pozitif	Yanlış Pozitif
	Negatif	Yanlış Negatif	Doğru Negatif

Sahtekarlık tespiti bağlamında, veri sınıflandırması tipik olarak düzenli, hileli olmayan işlemlerin 'Negatif' olarak etiketlenmesini ve şüpheli, potansiyel olarak hileli işlemlerin 'Pozitif' olarak etiketlenmesini amaçlar. Bu çerçevede, bir yöntemin etkinliği, bu etiketleri doğru bir şekilde tanımlamadaki doğruluğuna göre değerlendirilir.

Doğru Pozitifler (DP): Yöntemin, bir işlemi gerçekten hileli olduğunda doğru bir

şekilde hileli (Pozitif) olarak tanımladığı durumlardır. Bu doğruluk, yöntemin hileli faaliyetleri doğru bir şekilde işaretleme yeteneğini temsil ettiği için sahtekarlık tespitinde çok önemlidir.

Doğru Negatifler (DN): Yöntemin, tipik bir işlem olduğunda bir işlemi hileli olmayan (Negatif) olarak doğru bir şekilde tanımladığında ortaya çıkar. Bu, meşru işlemlerin yanlış bir şekilde işaretlenmemesini sağlayarak kullanıcı güvenini ve sistem verimliliğini koruduğu için aynı derecede önemlidir.

Yanlış Pozitifler (YP): Bu terim, yöntemin meşru bir işlemi (Negatif) yanlışlıkla hileli (Pozitif) olarak işaretlediği durumları ifade eder. Bu genellikle 'Tip I hata' olarak adlandırılır. Bu durum gereksiz soruşturmalara, zaman kaybına ve işlemleri meşru olan kullanıcılarda rahatsızlığa neden olacağından sorunlu olabilir.

Yanlış Negatifler (YN): Yöntemin hileli bir işlemi tanımlayamadığı, işlem hileli olduğu halde (Pozitif) hileli olmayan (Negatif) olarak etiketlediği durumlardır. 'Tip II hata' olarak bilinen bu durum, sahtekarlık tespitinde kritik bir hatadır çünkü sahtekarlık faaliyetlerinin tespit edilememesi anlamına gelir ve önemli mali kayıplara neden olabilir.

Bu ölçütlerden yola çıkarak hesaplanan duyarlılık gerçekte pozitif olan bir gözlemin doğru tahmin edilme oranını verir. Duyarlılık sınıflandırmanın pozitifleri doğru tespit etme kabiliyetini gösterir.

$$\text{Duyarlılık} = \frac{\text{Doğru Pozitif}}{\text{Doğru Pozitif} + \text{Yanlış Negatif}} \tag{4.8}$$

Kesinlik değeri ise pozitif olarak tahmin edilen gözlemlerin gerçekte ne kadarının pozitif olduğu ölçüsüdür.

$$\text{Kesinlik} = \frac{\text{Doğru Pozitif}}{\text{Doğru Pozitif} + \text{Yanlış Pozitif}} \tag{4.9}$$

F-Beta Puanı

F-beta puanı, kesinlik ve duyarlılığın göreceli önemini ayarlamak için beta parametresi ile ağırlıklandırılmış bir ölçüttür. Beta parametresi, duyarlılığın kesinliğe oranla ne kadar önemli olduğunu belirler. F-beta puanı şu şekilde hesaplanır:

$$F_\beta = (1+\beta^2)\frac{\text{Kesinlik} \times \text{Duyarlılık}}{\beta^2 \times \text{Kesinlik} + \text{Duyarlılık}} \tag{4.10}$$

$\beta > 1$ duyarlılığın daha önemli olduğu durumlarda, $\beta < 1$ ise kesinliğin daha önemli olduğu durumlarda kullanılır. $\beta = 1$ durumunda F_1 puanı elde edilir. F_1 puanı, sınıflandırma modellerinin performansını değerlendirmek için yaygın olarak kullanılan bir ölçüttür. F_1 puanı, kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır.

$$F_1 = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \tag{4.11}$$

Sahtekarlık tespitinde F puanında betanın 1 olarak seçilmesi, modelin duyarlılık ve kesinlik arasında denge kurmasına olanak tanır. Bu denge, hem yanlış pozitiflerin hem de yanlış negatiflerin maliyetlerinin yüksek olduğu senaryolarda kritik önem taşır. Yanlış pozitifler, tipik işlemin yanlışlıkla sahtekarlık olarak işaretlenmesine neden olurken, yanlış negatifler gerçek sahtekarlık vakalarının gözden kaçmasına ve finansal kayıplara yol açar. Bu nedenle, F_1 puanı, modelin genel performansını değerlendirirken dengeli bir bakış sağlar (Alfaiz ve Fati, 2022).

BÖLÜM BEŞ
UYGULAMA

Bu bölümde, önerilen 'FindOut' yaklaşımı sekiz farklı veride uygulanmış ve sonuçları paylaşılmıştır.

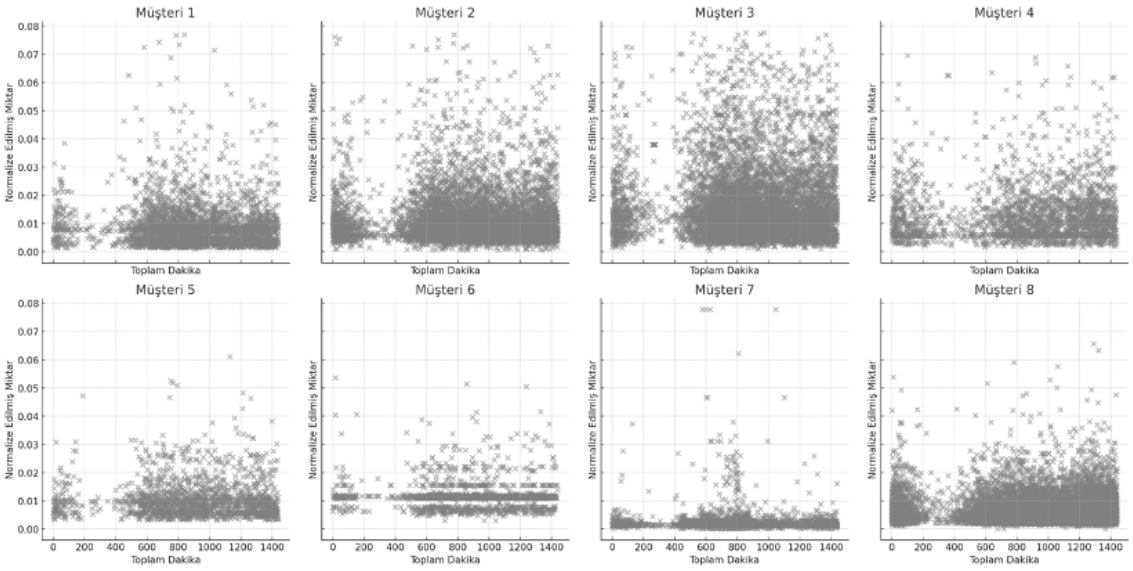
5.1 Veri

Uygulamaya konu olan veri, iki farklı sektörde faaliyet gösteren sekiz mağazadan yapılan kredi kartı alışveriş harcamalarını içermektedir. Analizler kredi kartı harcamasının gerçekleştiği mağaza için yapılmıştır. Kullanılan veriler, gerçek harcama verilerine uygun olarak türetilmiştir. Veri türetme işlemi sırasında, verinin dağılımına ve karakteristiğine sadık kalınmıştır. Veride sahtekarlık sayılabilecek işlemler yer almamakta olup, tüm gözlemler olağan harcama değerlerini ifade etmektedir. Harcamanın yapıldığı mağazalar bu bölümde müşteri olacak anılacaktır. Veriler, her bir işlem için tarih, müşteri kimliği, harcama miktarı , sektör ve işlem dakikası değişkenlerinden oluşmaktadır. Tablo 5.1'de veride yer alan değişkenler ve açıklamaları yer almaktadır.

Tablo 5.1 Değişken Tanımları

Değişken İsmi	**Açıklama**	**Veri Tipi**
İşlem Zamanı	İşlemin gerçekleştiği tarih ve saat bilgisi.	Tarih&Saat
Müşteri Kimliği	Her bir müşteriyi benzersiz şekilde tanımlar.	Kategorik
Harcama Miktarı	Her bir işlem için harcanan tutar.	Numerik
Sektör	İşlemin gerçekleştiği sektör.	Kategorik
İşlem Dakikası	İşlemin gerçekleştiği zamanın dakika karşılığı; 00:00 0. dakika, 23:59 1439. dakika	Kategorik

Müşteri 1'den Müşteri 8'e kadar olan müşterilerin harcama alışkanlıkları



Şekil 5.1 Müşterilerin 30 günlük harcamalarının gün içerisindeki dağılımı

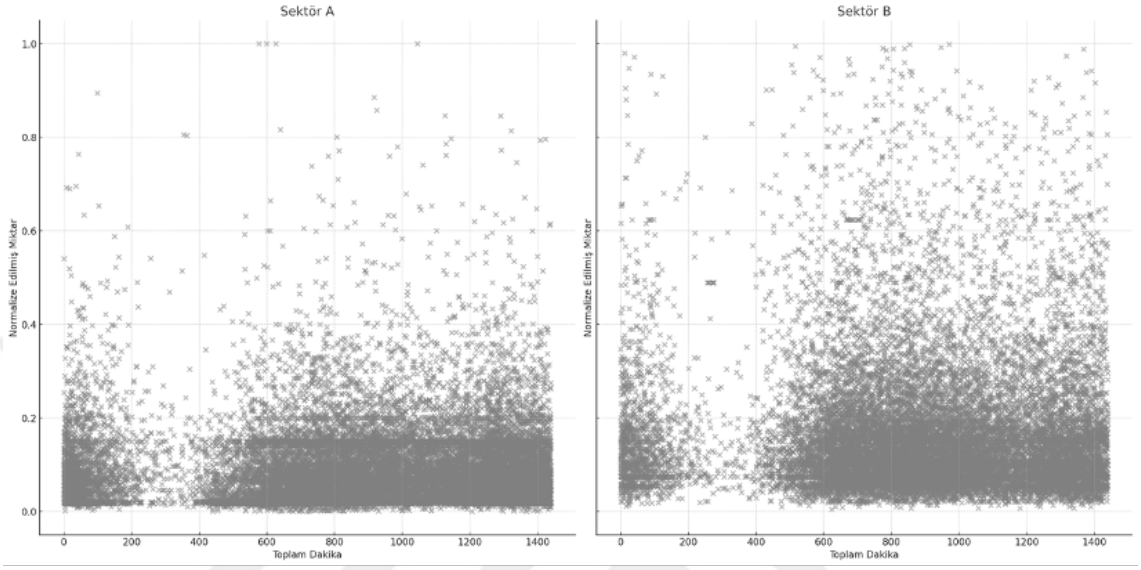
incelendiğinde, her bir müşterinin farklı harcama profilleri sergilediği gözlemlenmiştir. Bazı müşteriler daha sık ve düşük tutarlı işlemler gerçekleştirirken, diğerleri daha nadir fakat yüksek tutarlı işlemler yapmaktadır. Bu farklılıklar, müşteri davranışlarının ve harcama stratejilerinin çeşitliliğini ortaya koymaktadır.

Şekil 5.1'de gösterilen veriler, sekiz farklı müşteri için normalleştirilmiş harcama tutarlarının işlemin yapıldığı zamana göre dağılımını göstermektedir. Dikey eksende, harcamaların normalleştirilmiş değerleri yer alırken, yatay eksen gün içindeki zamanı dakika olarak temsil etmektedir. Yatay eksende 0 değeri saat 00:00'ı, 1439 değeri ise saat 23:59'u ifade etmektedir.

Müşterilerin harcama davranışları genel olarak incelendiğinde, farklı zaman dilimlerinde değişen yoğunluklar gözlenmektedir. Genel bir eğilim olarak, harcamaların günün erken saatlerinde daha az, öğleden sonra ve akşam saatlerinde ise daha fazla yoğunlaştığı görülmektedir. Özellikle müşteri 2 ve müşteri 3 grafiklerinde, günün ilerleyen saatlerinde harcama miktarlarının arttığı dikkat çekmektedir. Bu durum, uygulamadaki verilere göre bireylerin genellikle çevrimiçi alışveriş için akşam saatlerini tercih ettiğini göstermektedir.

Bazı müşteriler, örneğin müşteri 7, belirli zaman dilimlerinde ani harcama artışları

göstermektedir. Bu, mağazadan özel bir etkinlik, kampanya veya ihtiyaç nedeniyle belirli bir zaman diliminde yoğun harcama yapılmış olabileceğini gösterebilir. Diğer müşterilerde ise harcamalar gün içine daha dengeli bir şekilde dağılmış durumdadır, bu da düzenli alışveriş alışkanlıklarını yansıtabilir.



Şekil 5.2 Sektörlere göre harcama miktarı

Anonim olarak kullanılan verilerdeki sektörler A ve B olarak kodlanmıştır. A sektörü ve B sektörü arasındaki harcama alışkanlıkları ufak farklılıklar göstermektedir. A sektöründe yapılan harcamalar birbirine daha yakın ve düşük tutarlarda gerçekleşmektedir. Bu durum, sektördeki ürünlerin fiyat aralıklarının daha dar olması ve alışverişlerin daha sık tekrarlanması ile açıklanabilir. B sektöründe yapılan harcamalar ise değişken ve yüksek tutarlarda gerçekleşmektedir. Bu sektörde, müşteri harcamalarının geniş bir aralıkta olması, ürün çeşitliliğinin ve yüksek fiyatlı ürünlerin bulunmasından kaynaklanmaktadır.

Bu veriler, müşteri ve sektör bazında harcama alışkanlıklarının analiz edilmesi açısından önemli bilgiler sunmaktadır. Özellikle, sahtekarlık tespiti gibi uygulamalar için model geliştirilmesi açısından değerlidir. Müşteri harcamalarının sektöre göre dağılımlarının detaylı analizi, anomali tespiti ve risk yönetimi gibi konularda kritik bilgiler sağlayarak, daha etkili ve güvenli finansal sistemlerin geliştirilmesine katkı sunabilir.

5.2 Uygulama Şeması

Çoğu uygulamada anomali tespit modellerinin performansını değerlendirmek amacıyla etiketli veriler kullanılmıştır. Etiketli veriler sayesinde, modelin anormal ve normal verileri ne kadar doğru sınıflandırdığı ölçülebilir. Ancak, kredi kartı işlemleri gibi finansal verilerde genellikle işlemler etiketsiz olarak bulunur. Uygulamada ele alınacak veri de benzer bir şekilde etiketsiz yapıda bulunmaktadır. Buna karşılık veride sahtekarlık işleminin bulunmadığı bilinmektedir.

Önerilen yaklaşımın uygulanacağı veri, sahtekarlık işlemleri içermediği için tüm işlemler olağan (normal) olarak sınıflandırılmıştır. Modelin performansını değerlendirmek amacıyla, veriye kontrollü bir şekilde yapay anomaliler eklenmiş ve böylece veriler etiketlenmiştir. Anomali gözlemler oluştururken, her bir müşterinin harcama dağılımına ve tanımlayıcı istatistiklerine dikkat edilmiştir. Yapay anomaliler, her bir müşterinin harcama alışkanlıklarına uygun olarak türetilmiş ve sahtekarlık olarak işaretlenmiştir. Bu, modelin gerçek dünyada karşılaşabileceği sahtekarlık işlemlerini tanıma yeteneğininin değerlendirilmesini sağlamaktadır.

Ödeme işlemi sırasında, sahtekarlık tespit sistemlerinin düşük gecikme süreleri ile hızlı karar verebilmesi, hem kullanıcı deneyimi açısından hem de güvenlik açısından büyük önem taşımaktadır (Brown ve Green, 2019). Bu bağlamda, yaklaşımda kullanılan her yöntemin eldeki veriler için optimize edilmiş parametrelerle çalışması gerekmektedir. Parametre optimizasyonu, makine öğrenmesi algoritmalarının performansını önemli ölçüde artırarak, eğitim sürelerini kısaltabilir ve daha hızlı sonuçlar elde edilmesine olanak tanıyabilir (Brownlee, 2016).

Makine öğrenmesi ve anomali tespitinde parametre seçimi, model performansını doğrudan etkileyen kritik bir faktördür. Modellerin doğruluğu ve genelleme kapasitesi, seçilen parametrelerin uygunluğuna bağlıdır. Doğru parametrelerin belirlenmesi, modelin hem eğitim hem de test aşamalarında başarılı olmasını sağlar.

Hiperparametre seçimi, modelin eğitilmesi sırasında optimize edilen ve modelin karmaşıklığını kontrol eden parametrelerden biridir (Goodfellow vd., 2016). Örneğin,

SVM'de doğru C ve gamma değerlerinin seçilmesi, modelin doğruluğunu önemli ölçüde artırabilir. C parametresi modelin karar sınırının ne kadar sert veya esnek olacağını kontrol eder. Yüksek bir C, hatalara karşı daha az toleranslı olup eğitim verisindeki sınıflandırma hatalarını azaltmayı hedeflerken; düşük bir C, daha genel bir model oluşturarak hatalara daha fazla tolerans gösterir. Gamma parametresi ise bir veri noktasının karar sınırını nasıl etkilediğini belirler. Yüksek bir gamma, yalnızca çok yakın komşularını dikkate alarak modelin daha hassas olmasına yol açarken, düşük bir gamma tüm veri genelinde daha geniş etkiler yaratarak daha genel bir model oluşturur. Hiperparametre optimizasyonunda yaygın olarak kullanılan grid arama ve rastgele arama yöntemleri, hiperparametre uzayını tarayarak en iyi parametre kombinasyonunu bulmayı amaçlar (Bergstra ve Bengio, 2012).

Anomali tespitinde, parametre seçimi daha da kritik hale gelir, çünkü bu modeller genellikle dengesiz verileri ile çalışır ve yanlış pozitif ile yanlış negatif oranlarını dengelemeye özen gösterir (Chandola vd., 2009).

Modelin performansı, eldeki veriler için en uygun parametrelerin kullanılmasıyla optimize edilmiştir. Bu bağlamda, parametre ayarlamaları (tuning) süreci, IF, LOF , OC-SVM ve EE yöntemleri için yapılmıştır. Her bir yöntem için gerçekleştirilen parametre optimizasyonu, modelin doğruluk ve etkinliğini artırmayı amaçlamıştır. IF için ağaç sayısı (n_estimators) ve kontaminasyon oranı (contamination), LOF için komşu sayısı (n_neighbors) ve kontaminasyon oranı, OC-SVM için çekirdek tipi ve *nv* parametresi, EE için ise kontaminasyon oranı ve destek noktası sayısı gibi hiperparametreler optimize edilmiştir. Kontaminasyon oranı ("contamination rate"), modelin anormal (anomali) olarak kabul edeceği veri noktalarının oranını belirtir. Yani, kontaminasyon oranı, veride beklenen anomali yüzdesini temsil eder. Bu oran, modelin anomali eşik değerini ayarlamasında yardımcı olur. Parametre seçimi, modelin performansını ve doğruluğunu belirleyen çok önemli bir faktördür. Uygulamada, doğru parametreler seçilerek, modellerin genel performansı iyileştirilmiş ve anomali tespiti yeteneği artırılmıştır.

5.3 Model Oluşturma

5.3.1 Profilleme Aşaması

İki aşamalı yaklaşımın başarı ölçütleri için, gerçek dünya senaryosuna uygun bir simülasyon hazırlanmıştır. İlk olarak, yaklaşımın gerektirdiği Özel ve Genel profiller oluşturuldu. Sekiz müşteri için, müşterilerin önceki harcama yapıları göz önüne alınarak 8 ayrı profil oluşturuldu. Özel profiller, her müşterinin harcama yapılarını ve davranışlarını yansıtırken, Genel profiller o sektördeki tüm müşterilerin genel harcama eğilimlerini temsil etmektedir. Özel ve Genel profillerin oluşturulması, yöntemlerin farklı seviyelerdeki anomalileri tespit edebilmesi için önemlidir. Genel profil kısmında, 8 müşterinin oluşturduğu 2 sektör üzerinden harcama verileri toplanmıştır. Özel profildeki verilere göre daha fazla verinin bulunduğu genel profillemede, her sektördeki işlemlerin %70'i rastgele seçilerek kullanılmıştır.

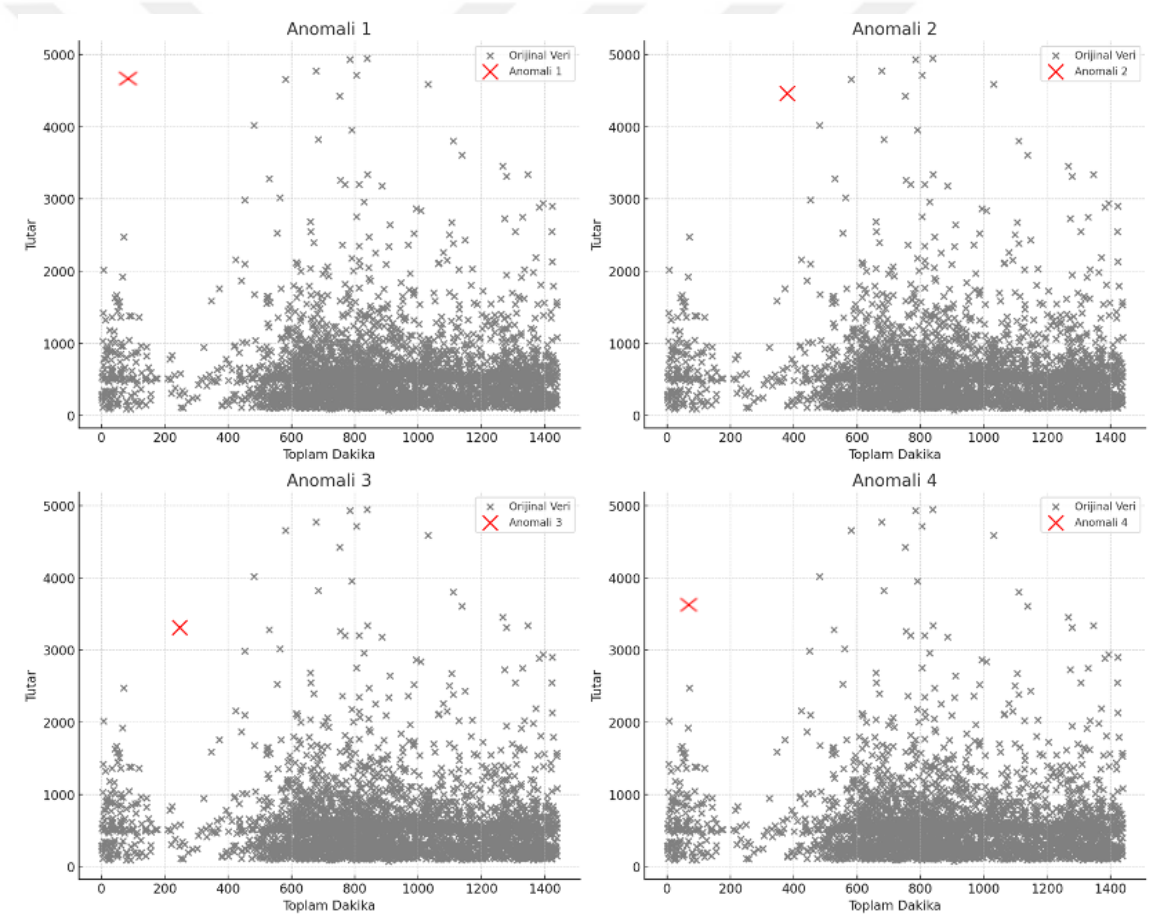
5.3.2 Değerlendirme Aşaması

Önceki bölümde oluşturulan profiller, kullanılan yöntemlerin eğitim kısmında kullanıldı. Her bir harcama bilgisi önem arz ettiğinden, yöntemler eğitilirken profillerin tüm işlemleri kullanıldı. FindOut yaklaşımının başarısını ölçmek için

Tablo 5.2 Tutar Değişkeninin Tanımlayıcı İstatistikleri

No	İşlem Sayısı	Ortalama	Std. Sapma	Min.	Ortanca	Mak.
1	3243	582,84	500,89	70,00	505,00	4947,00
2	8036	693,00	563,65	35,00	506,18	4949,00
3	7565	1077,89	816,31	22,52	838,00	4993,50
4	2347	842,67	624,89	100,00	665,00	4472,00
5	1751	672,55	422,60	194,90	574,80	3926,85
6	1743	762,66	325,15	194,06	749,90	3448,68
7	5813	129,23	174,05	1,00	99,90	4000,00
8	10442	447,82	338,11	1,01	341,35	4225,82

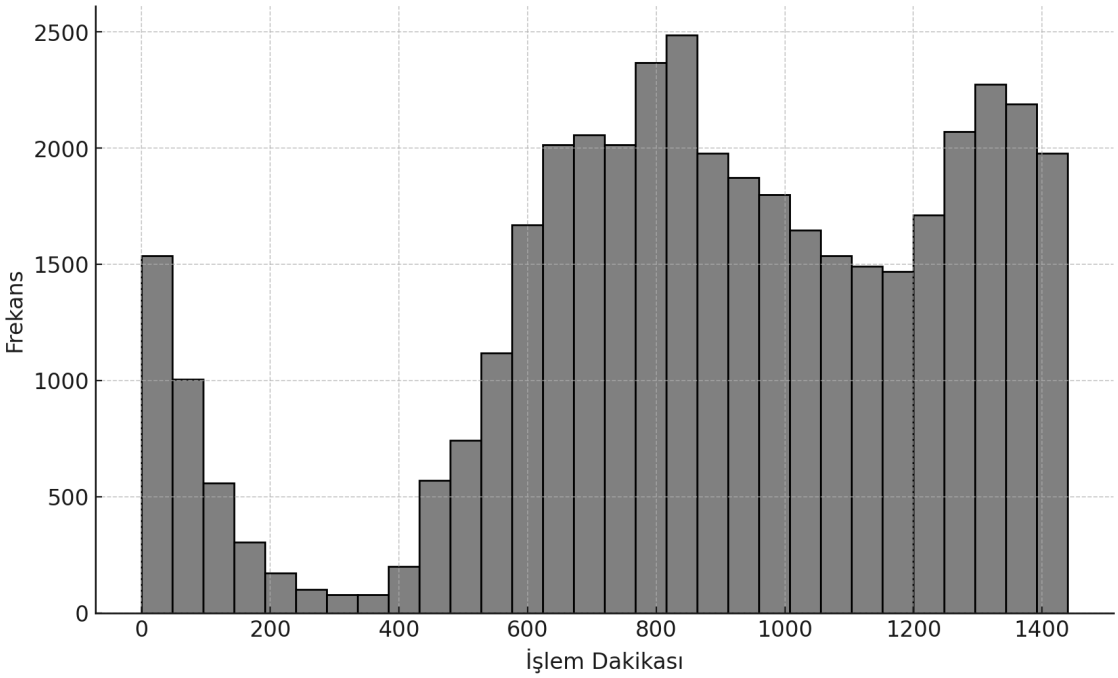
modele kontrollü bir biçimde olağan ve anomali veriler eklendi. Müşterilere göre anomali ve olağan işlem aralıklarının belirlenmesi için, ilk olarak her müşterinin ortalama ve standart sapma değerleri bulundu. Şekil 5.2'de müşterilerin "Harcama Miktarı" değişkeninin ortalama ve standart sapma değerleri gösterilmektedir. Olağan değerlerin alt ve üst sınırları, müşterinin ortalama harcamasına bir standart sapma eklenip çıkarılarak belirlenmiştir. Her deneme için "Harcama Miktarı" değeri, bu belirlenen alt ve üst sınırlar içerisinde rastgele olarak oluşturulmuştur. "Harcama Miktarı" değişkeni için anomali değerler ise, ortalamanın bir standart sapma üstündeki değerden başlayarak, ortalamanın üç standart sapma üstündeki değere kadar olan aralıktan rastgele olarak türetilmiştir.



Şekil 5.3 Anomali gözlemlerin eklenmesi

Şekil 5.3, anomali veri ekleme sürecini görselleştirmektedir. Her bir grafikte gri renkte gösterilen orijinal veri noktaları ile kırmızı renkte gösterilen anomali harcama verileri yer almaktadır. Anomali ekleme sürecinde, her bir anomali için belirlenen

kurallar çerçevesinde veri noktaları orijinal veriye eklenmiştir. Anomaliler, müşterinin ortalama harcama ve süre değerlerinden önemli ölçüde sapma gösteren veri noktaları olarak tanımlanmıştır.



Şekil 5.4 İşlem dakikası değerlerinin histogramı

Diğer bir değişken olan "işlem dakikası" için olağan değerler, histogramda sıklık değeri en yüksek olan iki aralık içinde rastgele olarak oluşturulmuştur. Şekil 5.4'te müşteri-2'nin "işlem dakikası" değerlerinin histogramı gösterilmektedir. Anomali değerler ise, sıklık değeri en az olan aralıkta rastgele seçilmiştir. Sahtekarlık tespit sistemleri, yapısı gereği, son gelen işlemin sahtekarlık olup olmadığını tespit etmeye çalışır. Model, elindeki eğitim verisini yeni gelen ödeme isteği ile sınayarak son işlem için bir sahtekarlık skoru belirler. Bu çalışmada, eğitim verilerine teker teker türetilen işaretli veriler eklenmiş, FindOut yaklaşımının eklenen veri üzerinde verdiği karara göre başarı ölçütleri belirlenmiştir.

5.4 Profil Değerlendirme

Türetilen her yeni işlem, önce özel profil kullanılarak değerlendirilmiştir. Eğer işlem özel profil tarafından olağandışı olarak işaretlenirse, bu işlem genel profil ile tekrar incelenmiştir. Genel profil değerlendirmesi, işlemin sektörel normlara göre de olağandışı olup olmadığını belirler. Bu çift aşamalı profil değerlendirme süreci, yanlış pozitifleri azaltmak ve daha doğru anomali tespiti sağlamak amacıyla uygulanmıştır. Yeni işlem verisi türetme süreci, hem olağan hem de anomali işlem senaryolarını oluşturmak için kullanılmıştır. Bu süreç, her müşteri için 50,000 kez tekrarlanmış ve bu denemelerin %90'ı olağan işlem, %10'u ise anomali işlem olacak şekilde yapılandırılmıştır.

Her müşteri için oluşturulan hem özel hem de genel profiller, türetilmiş veriler üzerinde anomali tespiti yapmak için kullanılmıştır. Bu yöntem, modelin hem müşteri bazında hem de sektörel düzeyde anomali tespit kabiliyetini test etmek ve doğrulamak için geniş bir veri sağlamıştır.

5.5 Model Çıktıları

Uygulama Apple Silicon M1 işlemcili 8 GB RAM'e sahip bir bilgisayarda yapılmıştır. Bir yeni işlemin tüm model içerisinde değerlendirme süresi ortalama 240 milisaniye olmuştur.

Tablo 5.3, Müşteri 1'in harcama verilerinde anomali tespitinde yöntemlerin "Özel" ve "Genel" profiller için performansını karşılaştırmaktadır. Performans değerlendirmesi, Kesinlik, Duyarlılık ve F_1 puanı ölçütleri üzerinden yapılmıştır. Bu ölçütler, yöntemlerin doğruluk ve güvenilirliğini değerlendirmede kritik rol oynamaktadır.

En iyi sonuçları veren yöntem, LOF olarak gözlemlenmiştir. LOF yöntemi, hem "Özel" hem de "Genel" profillerde yüksek kesinlik ve F_1 puanı değerleri sunmuştur. Bu durum, LOF yönteminin anomali tespitinde güçlü bir performans sergilediğini ve

Tablo 5.3 Müşteri 1 için sonuçlar

Profil	**Özel**			**Genel**		
	Kesinlik	**Duyarlılık**	F_1	**Kesinlik**	**Duyarlılık**	F_1
IF	0,9299	0,6486	0,7869	0,8643	0,6882	0,7663
EE	0,8349	0,4235	0,5951	0,8625	0,5732	0,6887
LOF	0,9965	0,8464	0,9153	0,8887	0,8860	0,8873
OC-SVM	0,9082	0,2398	0,3790	0,8968	0,3245	0,4766

çoğunlukla doğru anomalileri tespit ederken, yanlış pozitif oranının oldukça düşük olduğunu göstermektedir. Özellikle, LOF'un "Genel" profil altında da benzer başarıyı sürdürmesi, yöntemin genellenebilirliğinin ve farklı veri profilleri üzerinde tutarlı performans sergileyebilme kapasitesinin yüksek olduğunu ortaya koymaktadır.

Öte yandan, en düşük performansı gösteren yöntem OC-SVM olmuştur. OC-SVM'nin düşük duyarlılık ve F_1 puanı, yöntemin anomali tespit yeteneğinin sınırlı olduğunu göstermektedir. OC-SVM'nin, doğru anomali sayısını tespit etmede düşük başarı sergilemesi, özellikle nadir anomalileri yakalamada yetersiz kaldığını göstermektedir. Bu durum, yöntemin genel performansını olumsuz etkilemiş ve "Genel" profil kullanımıyla da iyileştirilememiştir. OC-SVM'nin yüksek kesinlik değerine rağmen, düşük duyarlılık ve F_1 puanı, modelin çoğunlukla normal örnekleri doğru tespit ettiğini ancak anomali tespitinde başarısız olduğunu göstermektedir.

Tablo 5.4 Müşteri 2 için sonuçlar

Profil	**Özel**			**Genel**		
	Kesinlik	**Duyarlılık**	F_1	**Kesinlik**	**Duyarlılık**	F_1
IF	0,9063	0,6164	0,7338	0,8572	0,6660	0,7496
EE	0,8371	0,5563	0,6684	0,8592	0,3878	0,5344
LOF	0,8410	0,8459	0,8434	0,8897	0,8954	0,8926
OC-SVM	0,8771	0,1997	0,3250	0,9279	0,2897	0,4415

Tablo 5.4'teki sonuçlar, Müşteri 2'nin harcama verilerinde anomali tespiti için yöntemlerin etkinliğini değerlendirmektedir. LOF yöntemi, yüksek duyarlılık ve F_1

puanlarıyla anomali tespitinde en başarılı yöntem olarak öne çıkmaktadır. Bu, LOF'un anomali tespitinde tutarlı ve güvenilir bir yöntem olduğunu göstermektedir.

Buna karşılık, OC-SVM yöntemi, düşük duyarlılık ve F_1 puanlarıyla beklenen anomalileri belirlemede yetersiz kalmıştır. Bu durum, OC-SVM'nin bu veride anomali tespiti için uygun bir seçenek olmadığını ortaya koymaktadır. Diğer yöntemler, performans açısından LOF ile karşılaştırıldığında daha sınırlı başarı göstermiştir, bu da LOF'un genel olarak daha etkin bir yaklaşım olduğunu işaret etmektedir.

Tablo 5.5 Müşteri 3 için sonuçlar

Profil	**Özel**			**Genel**		
	Kesinlik	**Duyarlılık**	F_1	**Kesinlik**	**Duyarlılık**	F_1
IF	0,8458	0,3281	0,4728	0,8270	0,5306	0,6464
EE	0,9467	0,1854	0,3101	0,9398	0,3689	0,5298
LOF	0,8693	0,7419	0,8006	0,8949	0,9444	0,9190
OC-SVM	0,9376	0,0077	0,0153	0,8580	0,1385	0,2385

Tablo 5.5, Müşteri 3'ün harcama verilerinde kullanılan çeşitli anomali tespit yöntemlerinin performansını göstermektedir. LOF yöntemi, hem "Özel" hem de "Genel" profillerde yüksek performans sergileyerek, özellikle duyarlılık ve F_1 puanında üstünlük sağlamıştır. Bu sonuçlar, LOF yönteminin anomali tespitinde yüksek doğruluk ve güvenilirlikle çalıştığını ortaya koymaktadır.

Diğer yandan, IF yöntemi, makul bir kesinlik sağlamasına rağmen, duyarlılık ve F_1 puanında düşük kalmış ve bu da anomali tespit yeteneğinde sınırlı bir başarıya işaret etmektedir. En düşük performansı ise OC-SVM ve EE yöntemleri göstermiştir. OC-SVM ve EE'nin düşük duyarlılık ve F_1 puanları, bu yöntemlerin anomali tespitinde yetersiz olduğunu göstermektedir.

Genel olarak, LOF yöntemi, anomali tespiti için en etkili yöntem olarak öne çıkmakta ve diğer yöntemlerle karşılaştırıldığında daha güvenilir ve tutarlı sonuçlar sunmaktadır.

Tablo 5.6 Müşteri 4 için sonuçlar

Profil	**Özel**			**Genel**		
	Kesinlik	**Duyarlılık**	F_1	**Kesinlik**	**Duyarlılık**	F_1
IF	0,9299	0,6486	0,7642	0,8643	0,6882	0,7663
EE	0,8349	0,3754	0,5179	0,8625	0,5478	0,6700
LOF	0,9965	0,8464	0,9153	0,8887	0,8860	0,8873
OC-SVM	0,9082	0,2198	0,3539	0,8968	0,1274	0,2231

Tablo 5.6, Müşteri 4'ün harcama verilerinde uygulanan çeşitli anomali tespit yöntemlerinin etkinliğini göstermektedir. LOF yöntemi, yüksek kesinlik ve F_1 puanıyla hem "Özel" hem de "Genel" profillerde üstün performans sergilemiştir. Bu durum, LOF'un anomali tespitinde güvenilir ve tutarlı bir yöntem olduğunu göstermektedir.

IF yöntemi de makul bir performans sergilemiş, özellikle kesinlik açısından iyi sonuçlar vermiştir. Ancak, duyarlılık oranının diğer yöntemlere göre düşük olması, IF'nin bazı anomalileri kaçırma olasılığını artırmaktadır.

OC-SVM ve EE yöntemleri, düşük duyarlılık ve F_1 puanları ile anomali tespitinde yetersiz kalmıştır. OC-SVM'nin özellikle düşük performansı, bu yöntemin Müşteri 4'ün verilerinde anomali tespitinde etkisiz olduğunu ortaya koymaktadır. Genel olarak, LOF yöntemi, diğer yöntemlere kıyasla daha yüksek doğruluk ve güvenilirlikle anomalileri tespit etmiştir.

Tablo 5.7 Müşteri 5 için sonuçlar

Profil	**Özel**			**Genel**		
	Kesinlik	**Duyarlılık**	F_1	**Kesinlik**	**Duyarlılık**	F_1
IF	0,9293	0,4546	0,6105	0,8366	0,5682	0,6768
EE	0,8862	0,6542	0,7527	0,9168	0,7458	0,8225
LOF	0,8860	0,8017	0,8417	0,8919	0,9153	0,9034
OC-SVM	0,8362	0,4905	0,6183	0,8201	0,2304	0,3597

Tablo 5.7, Müşteri 5'in harcama verilerinde kullanılan çeşitli anomali tespit

yöntemlerinin performansını karşılaştırmaktadır. LOF yöntemi, yüksek duyarlılık ve F_1 puanıyla öne çıkmakta, hem "Özel" hem de "Genel" profillerde başarılı sonuçlar vermektedir. Bu durum, LOF'un farklı profillerde tutarlı bir şekilde yüksek performans sergileyebildiğini göstermektedir.

IF yöntemi, özellikle kesinlik açısından iyi sonuçlar sunmuş, ancak duyarlılık değerlerinin görece düşük olması nedeniyle bazı anomalilerin tespit edilememesi riskini taşımaktadır. EE yöntemi, kesinlikte yüksek olmasına karşın, düşük duyarlılık ve F_1 puanıyla, anomalilerin büyük bir kısmını gözden kaçırma eğilimindedir.

OC-SVM yöntemi, hem "Özel" hem de "Genel" profillerde düşük performans göstermiştir. Bu, OC-SVM'nin Müşteri 5'in verilerinde anomali tespiti konusunda etkisiz kaldığını ortaya koymaktadır. Genel olarak, LOF yöntemi en başarılı sonuçları verirken, diğer yöntemler özellikle duyarlılık açısından sınırlı performans sergilemiştir.

Tablo 5.8 Müşteri 6 için sonuçlar

Profil	**Özel**			**Genel**		
	Kesinlik	**Duyarlılık**	F_1	**Kesinlik**	**Duyarlılık**	F_1
IF	0,8981	0,4630	0,6110	0,8391	0,5785	0,6848
EE	0,8680	0,2345	0,3692	0,8790	0,4597	0,6037
LOF	0,8426	0,8042	0,8230	0,8923	0,9197	0,9058
OC-SVM	0,9292	0,3123	0,4675	0,8877	0,3355	0,4870

Tablo 5.8, Müşteri 6'nın harcama verileri üzerinden uygulanan çeşitli anomali tespit yöntemlerinin etkinliğini ortaya koymaktadır. LOF yöntemi, özellikle duyarlılık ve F_1 puanı açısından hem "Özel" hem de "Genel" profillerde güçlü bir performans sergilemiştir. Bu, LOF'un anomali tespitinde güvenilir ve yüksek doğrulukla çalıştığını göstermektedir.

IF yöntemi, kesinlik açısından iyi sonuçlar verirken, duyarlılık değerleri düşük kalmış ve bu da bazı anomalilerin kaçırılmasına yol açabilecek bir durum yaratmıştır. EE yöntemi, özellikle düşük duyarlılık ve F_1 puanı ile dikkat çekmektedir, bu da

yöntemin anomali tespitinde sınırlı bir etkiye sahip olduğunu göstermektedir.

OC-SVM yöntemi, yine düşük duyarlılık ve F_1 puanı ile performans açısından zayıf kalmıştır. Bu, OC-SVM'nin Müşteri 6'nın verilerinde anomali tespiti için yeterince etkili olmadığını ortaya koymaktadır. Genel olarak, LOF yöntemi en başarılı sonuçları sunmuş, diğer yöntemler ise özellikle duyarlılık ve F_1 puanlarında düşük performans sergilemiştir.

Tablo 5.9 Müşteri 7 için sonuçlar

Profil	**Özel**			**Genel**		
	Kesinlik	**Duyarlılık**	F_1	**Kesinlik**	**Duyarlılık**	F_1
IF	0,8318	0,8126	0,8221	0,8953	0,8283	0,8605
EE	0,8396	0,6107	0,7071	0,8937	0,5345	0,6689
LOF	0,9830	0,8140	0,8906	0,8821	0,8298	0,8552
OC-SVM	0,8222	0,5543	0,6622	0,9568	0,2777	0,4305

Tablo 5.9, Müşteri 7'nin harcama verilerinde kullanılan farklı anomali tespit yöntemlerinin performansını değerlendirmektedir. IF yöntemi, hem "Özel" hem de "Genel" profillerde yüksek kesinlik ve F_1 puanı ile dikkat çekmekte, böylece anomali tespitinde etkili olduğunu göstermektedir. EE yöntemi, makul bir kesinlik sağlamasına rağmen, duyarlılık açısından diğer yöntemlere göre daha düşük performans sergilemiştir.

LOF yöntemi, "Özel" profilde yüksek duyarlılık ve F_1 puanı ile öne çıkarken, "Genel" profilde diğer yöntemlerle benzer bir performans sergilemiştir. Bu, LOF'un anomali tespitinde genel olarak güvenilir bir seçenek olduğunu ortaya koymaktadır.

OC-SVM yöntemi, özellikle düşük duyarlılık ve F_1 puanı ile hem "Özel" hem de "Genel" profillerde en düşük performansı göstermiştir. Bu, OC-SVM'nin Müşteri 7'nin verilerinde anomali tespiti için yeterince etkili olmadığını göstermektedir. Genel olarak, IF ve LOF yöntemleri, anomali tespitinde en başarılı sonuçları sunmuştur.

Tablo 5.10, Müşteri 8'in harcama verilerinde uygulanan çeşitli anomali tespit

Tablo 5.10 Müşteri 8 için sonuçlar

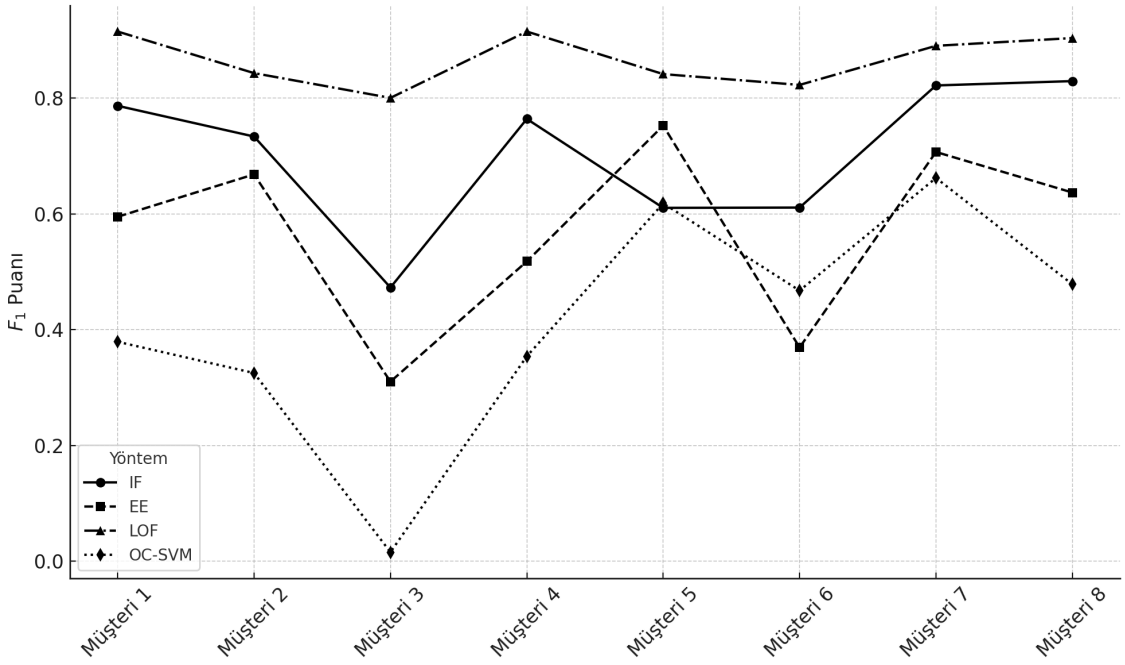
Profil	**Özel**			**Genel**		
	Kesinlik	**Duyarlılık**	F_1	**Kesinlik**	**Duyarlılık**	F_1
IF	0,9241	0,7527	0,8296	0,8889	0,7785	0,8300
EE	0,8216	0,5201	0,6370	0,8821	0,6984	0,7796
LOF	0,9839	0,8361	0,9040	0,8859	0,8619	0,8737
OC-SVM	0,8272	0,3365	0,4784	0,8217	0,3122	0,4525

yöntemlerinin performansını değerlendirmektedir. IF yöntemi, yüksek kesinlik ve F_1 puanı ile hem "Özel" hem de "Genel" profillerde etkin bir performans sergilemiştir. Bu, IF'nin anomali tespitinde güvenilir bir yöntem olduğunu göstermektedir.

Uygulamada 2 farklı sektörden 8 ayrı müşteri için önerilen yaklaşım denenmiştir. Müşterilerden elde edilen sonuçlara göre, LOF yöntemi, özellikle yüksek duyarlılık ve F_1 puanı ile öne çıkmakta ve anomali tespitinde yüksek doğruluk sağlamaktadır. LOF'un bu başarısı, yöntemin anomali tespiti için oldukça etkili olduğunu ortaya koymaktadır.

EE yöntemi, diğer yöntemlere göre daha düşük duyarlılık ve F_1 puanı ile sınırlı bir performans göstermiştir. OC-SVM yöntemi ise, özellikle düşük duyarlılık ve F_1 puanı ile en düşük performansı sergileyen yöntem olmuştur. Bu sonuçlar, kullanılan senaryolarda, OC-SVM'nin anomali tespiti için uygun bir seçenek olmadığını ve LOF ile IF'nin bu görevde daha başarılı olduğunu göstermektedir.

Tüm tablolarda yer alan sonuçlar, farklı anomali tespit yöntemlerinin çeşitli müşteri verilerinde nasıl performans gösterdiğini kapsamlı bir şekilde değerlendirmektedir. Genel olarak, LOF yöntemi, en yüksek kesinlik, duyarlılık ve F_1 puanları ile diğer yöntemlere kıyasla üstün bir performans sergilemiştir. LOF'un özellikle "Özel" profillerde yüksek başarı elde etmesi, bu yöntemin veriye özgü anomalileri yakalamada etkili olduğunu ve aynı zamanda "Genel" profillerde de tutarlı bir performans sağladığını göstermektedir.



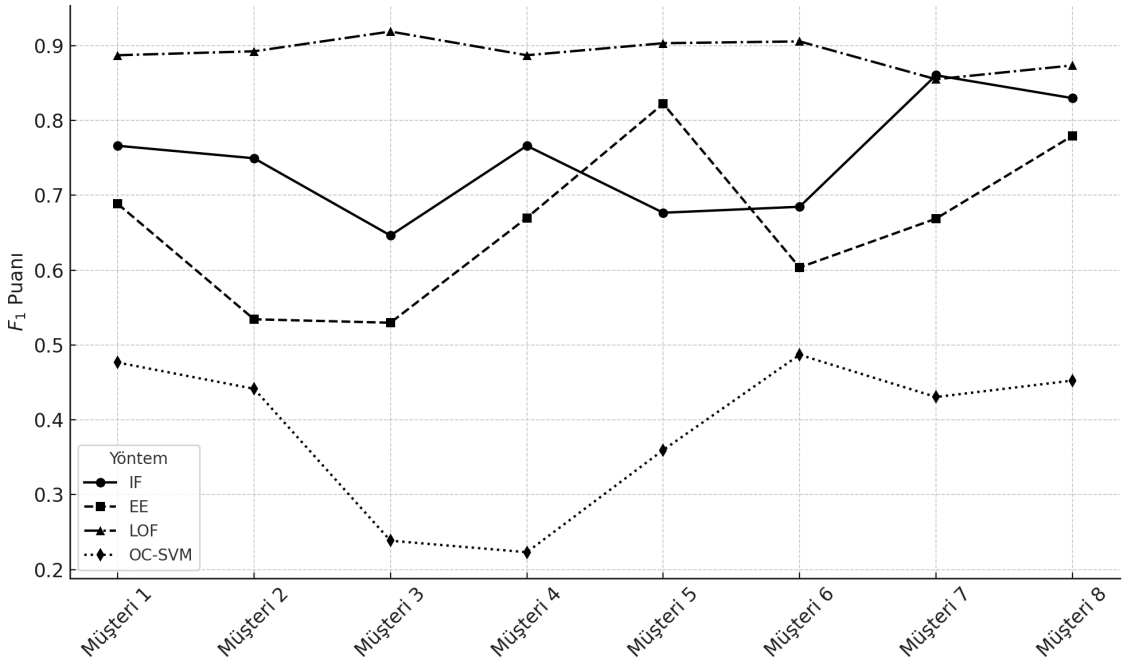
Şekil 5.5 Özel profil F_1 puanları

IF yöntemi, genellikle makul düzeyde kesinlik ve duyarlılık sunmuş, ancak bazı durumlarda duyarlılık değerlerinde düşüşler gözlenmiştir. Bu durum, IF'nin belirli anomalileri kaçırma olasılığını artırmakla birlikte, genel olarak güvenilir bir yöntem olduğunu göstermektedir. IF'nin güçlü yönleri, veriyi dışlama mekanizmasıyla verilerin büyük çoğunluğunu doğru bir şekilde tanımlayabilmesidir.

EE yöntemi, kesinlik açısından çoğu durumda kabul edilebilir düzeyde performans göstermiştir. Ancak duyarlılık ve F_1 puanlarındaki düşüklük, bu yöntemin anomali tespitinde yetersiz kalabileceğini ortaya koymaktadır. EE'nin düşük duyarlılığı, özellikle nadir veya heterojen anomalilerin tespitinde sınırlı bir etkiye sahip olduğunu göstermektedir.

OC-SVM yöntemi ise, genel olarak en düşük performansı sergilemiş, hem duyarlılık hem de F_1 puanlarında yetersiz kalmıştır. Bu durum, OC-SVM'nin ele alınan veride anomali tespiti için uygun bir yöntem olmadığını göstermektedir. OC-SVM'nin düşük performansı, çoğunlukla anomali sayısının az olduğu verilerde doğru anomalileri yakalamada yetersiz kalmasından kaynaklanmaktadır.

Şekil 5.6'da LOF ve IF değerlerinin birbirine yakınlaştığı görülmektedir. Özellikle



Şekil 5.6 Genel profil F_1 puanları

Müşteri 7 ve 8'de IF ve LOF birbirine yakın seyretmiştir.

Genel olarak, LOF yöntemi tüm değerlendirmelerde en güvenilir ve etkili yöntem olarak öne çıkmıştır. IF ise bazı durumlarda yüksek doğruluk sunarken, genel güvenilirlikte LOF'un gerisinde kalmıştır. EE ve OC-SVM ise, özellikle duyarlılık açısından yeterli performans gösterememiştir. Bu bulgular, anomali tespiti yaparken verinin doğası, anomali oranı ve verinin özellikleri göz önünde bulundurularak uygun yöntemlerin seçilmesinin önemini vurgulamaktadır. Anomali tespiti için en uygun yöntemi belirlerken, kesinlik, duyarlılık ve F_1 puanları gibi performans ölçütlerinin kapsamlı bir şekilde değerlendirilmesi kritik öneme sahiptir.

BÖLÜM ALTI
SONUÇ VE TARTIŞMA

Günümüz finans dünyasında kredi kartı sahtekarlığı, büyük mali kayıplara neden olabilen kritik bir sorundur. Bu çalışma, sahte işlemlerin erken tespit edilmesi için bir çözüm önerisi sunmaktadır. E-ticaret harcamalarının her geçen gün artan hacmi nedeniyle kredi kartı işlemleri bireylerin hayatının vazgeçilmez bir parçası olmuştur. Bu nedenle, sahtekarlık tehditlerini en aza indirmek için çeşitli gözetim teknikleri ve analizler geliştirilmiştir. Bu alanda kullanılan en önemli tekniklerden biri de anomali tespitidir. Kredi kartı işlemlerinde anomali tespiti, müşterilerin olağan dışı davranışlarını belirlemeyi amaçlayan bir çalışma alanıdır. Kredi kartı işlemlerinde olağan ve olağan dışı davranışları tanımlamak için genellikle harcama verileri kullanılmaktadır.

Bu çalışmada, sahtekarlık tespiti için önerilen iki aşamalı profilleme yöntemi kullanılmıştır. Önerilen model, müşterilerin geçmiş harcama verilerini kullanarak her yeni işlemi anomali açısından değerlendirmektedir. Bu model, hem özel profilleme hem de genel profilleme aşamalarından oluşmaktadır. Özel profilleme, müşterinin geçmiş işlemlerine dayanarak oluşturulan profildir. Genel profilleme ise, benzer harcama alışkanlıklarına sahip müşterilerin kümelerini belirleyerek oluşturulmaktadır. Uygulamada, normal işlem verileri ile anomali verileri arasındaki farklılıkları belirlemek için çeşitli veri analiz yöntemleri ve modeller kullanılmıştır. Elde edilen sonuçlar, önerilen yöntemin kredi kartı sahtekarlığının tespitinde hızlı ve etkili olduğunu göstermektedir.

FindOut modeli, bu iki aşamalı yapı sayesinde sahtekarlık tespitinde iki önemli fayda sunmaktadır: daha yüksek tespit oranları ve uyarlanabilirlik. Özel profilleme, kullanıcıların alışılmış davranışlarından sapmaları tespit ederken, genel profilleme benzer kullanıcı gruplarını analiz ederek yanlış pozitifleri azaltır ve tespit doğruluğunu artırır. Kullanıcı davranışlarındaki değişikliklere uyum sağlama kapasitesi, iki aşamalı yaklaşımın zaman içinde etkinliğini korumasını sağlar. Bu yöntem, değişen harcama alışkanlıklarını ve yeni sahtekarlık tekniklerini daha hızlı ve

etkili bir şekilde tespit edebilir.

Önerilen yaklaşım, finansal kurumlarda gerçek zamanlı sahtekarlık tespiti için uygulanabilirdir. Ancak, modelin başarımını artırmak için daha fazla sektörden ve müşteriden sağlanan veriler ve farklı yöntemler kullanılarak iyileştirmeler yapılabilir. Gelecek çalışmalarda, farklı anomali tespit yöntemlerinin kullanılması planlanabilir. Ayrıca, sistemin güvenliğini artırmak ve sahtecilik girişimlerini daha erken aşamalarda tespit etmek için çok katmanlı bir güvenlik yaklaşımı benimsenebilir.

Sonuç olarak, bu tez çalışması kredi kartı işlemlerindeki anomalilerin belirlenmesi yoluyla sahtekarlık tespiti alanına katkı sağlamayı amaçlamaktadır. Önerilen FindOut yaklaşımı, kredi kartı sahtekarlığı tespitinde hızlı işlem süresi, yüksek doğruluk ve düşük yanlış pozitif oranlarına ulaşabilmesi ile finansal sahtekarlıkla mücadelede yeni bir yaklaşım sunmaktadır.

KAYNAKLAR

Adeoti, J. O. (2011). Automated teller machine (atm) frauds in nigeria: The way out. *Journal of Social Sciences*, *27*(1), 53–58.

Akintoye, K., ve Araoye, O. (2011). Combating e-fraud on electronic payment system. *International Journal of Computer Applications*, *25*(8), 48–53.

Alfaiz, N. S., ve Fati, S. M. (2022). Enhanced credit card fraud detection model using machine learning. *Electronics*, *11*(4), 662.

Asha, R., ve Suresh Kumar, K. (2021). Credit card fraud detection using artificial neural network. *Global Transitions Proceedings*, *2*(1), 35–41. 1st International Conference on Advances in Information, Computing and Trends in Data Engineering (AICDE - 2020).

Bergstra, J., ve Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2).

Bolton, R. J., ve Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical science*, *17*(3), 235–255.

Breunig, M. M., Kriegel, H.-P., Ng, R. T., ve Sander, J. (2000). Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 93–104.

Brown, J., ve Green, E. (2019). *Fraud Detection in Payment Systems: Techniques and Technologies*. New York, NY: TechPress.

Brownlee, J. (2016). *Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras*.

Button, M., ve Cross, C. (2017). *Cyber frauds, scams and their victims*.

Carminati, M., Caron, R., Maggi, F., Epifani, I., ve Zanero, S. (2015). Banksealer: A decision support system for online banking fraud analysis and investigation. *computers & security*, *53*, 175–186.

Chakrabarty, K. C. (2013). Fraud in the banking sector–causes, concerns and cures. In *National Conference on Financial Fraud organised by ASSOCHAM, New Delhi*, vol. 26.

Chandola, V., Banerjee, A., ve Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), 1–58.

Chapelle, O., Scholkopf, B., ve Zien, A. (2006). *Semi-Supervised Learning*.

Chung, J., ve Lee, K. (2023). Credit card fraud detection: An improved strategy for high recall using knn, lda, and linear regression. *Sensors*, *23*(18), 7788.

Commission, F. T. (2022). *New data shows ftc received 2.8 million fraud reports from consumers in 2021*. from https://www.ftc.gov/news-events/press-releases/2022/02/new-data-shows-ftc-received-28-million-fraud-reports-consumers-2021.

Cortes, C., ve Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273–297.

Doerfler, P., Thomas, K., Marincenko, M., Ranieri, J., Jiang, Y., Moscicki, A., ve McCoy, D. (2019). Evaluating login challenges as adefense against account takeover. In *The World Wide Web Conference*, 372–382.

Erbayat, B., ve Yıldıztepe, E. (2023). An approach to identify credit card fraud with anomaly detection methods. In *Proceedings of the DEUISGR23 Symposium*, Turkey: DEUISGR.

Fernandes, L. (2013). Fraud in electronic payment transactions: Threats and countermeasures. *Asia Pacific Journal of Marketing & Management Review, ISSN*, *2319*, 2836.

Filzmoser, P., Garrett, R. G., ve Reimann, C. (2005). Multivariate outlier detection in exploration geochemistry. *Computers & geosciences*, *31*(5), 579–587.

Ghosh, S., ve Reilly, D. L. (1994). Credit card fraud detection with a neural-network. In *System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on*, vol. 3, *IEEE*, 621–630.

Goldstein, M., ve Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS one*, *11*(4), e0152173.

Goodfellow, I., Bengio, Y., ve Courville, A. (2016). *Deep learning*.

Hariri, S., Kind, M. C., ve Brunner, R. J. (2019). Extended isolation forest. *IEEE transactions on knowledge and data engineering*, *33*(4), 1479–1489.

Jones, A. S. (2019). Fraud risk management and corporate performance of deposit money banks (dmbs) in nigeria.

Karpoff, J. M. (2021). The future of financial fraud. *Journal of Corporate Finance*, *66*, 101694.

Liu, F. T., Ting, K. M., ve Zhou, Z.-H. (2008). Isolation forest. In *2008 eighth ieee international conference on data mining*, *IEEE*, 413–422.

Ngai, E. W., Chau, D. C., ve Chan, T. (2011). Information technology, operational, and management competencies for supply chain agility: Findings from case studies. *The Journal of Strategic Information Systems*, *20*(3), 232–249.

Raj, S. B. E., ve Portia, A. A. (2011). Analysis on credit card fraud detection methods. In *2011 International Conference on Computer, Communication and Electrical Technology (ICCCET)*, *IEEE*, 152–156.

RamaKalyani, K., ve UmaDevi, D. (2012). Fraud detection of credit card payment system by genetic algorithm. *International Journal of Scientific & Engineering Research*, *3*(7), 1–6.

Rousseeuw, P. J., ve Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, *41*(3), 212–223.

Sakurada, M., ve Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis*, 4–11.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., ve Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural computation*, *13*(7), 1443–1471.

Schölkopf, B., ve Smola, A. J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*.

Seliya, N., Abdollah Zadeh, A., ve Khoshgoftaar, T. M. (2021). A literature review on one-class classification and its potential applications in big data. *Journal of Big Data*, *8*, 1–31.

Song, X., Wu, M., Jermaine, C., ve Ranka, S. (2007). Conditional anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, *19*(5), 631–645.

Tax, D. M., ve Duin, R. P. (2004). Support vector data description. *Machine learning*, *54*, 45–66.

TBB. *(Türkiye Bankalar Birliği) (2015). Bankacılıkta Dolandırıcılık Eylemleri Tespit ve Önleme Yöntemleri*.

West, J., ve Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security*, *57*, 47–66.

Yazdanifard, R., WanYusoff, W. F., Behora, A. C., ve Sade, A. B. (2011). Electronic banking fraud: The need to enhance security and customer trust in online banking. *Advances in Information Sciences and Service Sciences*, *3*(10), 505–509.

Zimek, A., Schubert, E., ve Kriegel, H.-P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *5*(5), 363–387.