Akdeniz Mühendislik Dergisi
Akdeniz Journal of Engineering
(AKUJE)
2024, 2(1), pp. 13-23
Araştırma Makalesi Research Article

GERÇEK ZAMANLI VERİ AKIŞLARINDA OTOMATİK ANORMALLİK TESPİTİ: FİNANSAL TEKNOLOJİLER ŞİRKETİNDE BİR UYGULAMA

AUTOMATED ANOMALY DETECTION IN REAL-TIME DATA STREAMS: AN APPLICATION AT FINANCIAL TECHNOLOGIES COMPANY

Dicle ASLAN[*1]

1 Token Finansal Teknolojiler, İTÜ Maslak Kampüsü Arı 8 Binası, İstanbul, Türkiye

ABSTRACT

Abnormalities are sample in data that do not fit the normal patterns. Various reasons such as malware, fraud, cyber-attack, terrorist activities, faults, system behavior changes, instrument and human error might generate abnormalities. Anomaly detection is a technique that provides unexpected situations or patterns to be found in the data. These unexpected situations or patterns are called as anomalies, outliers and unexpected cases in the literature that do not fit to the expected behavior of the data. Diverse research and applications have been carried out for the anomaly detection which is a critical task for the industries. The studies about the anomaly detection in the literature are mostly trying to build the system for security and error finding activities. Increasing IoT devices in the banking and finance sector, healthcare, manufacturing, IT and telecom, defense and government might drive the anomaly detection market. These industries regularly deal with the important data, enabling criminals to be prone to serious fraud, theft, and attacks that give them control over the firm's infrastructure. To predict the anomaly, there are numerous learning methods as supervised, semi-supervised and unsupervised. As a traditional method, anomalies might be alarmed according to fixed threshold-level. This approach might not be having satisfied and accurate outcome. In this scope, this paper proposes a novel concept as building automated anomaly detector system for business operation platform at Financial Technologies company which is a leader in payment systems industry in Turkey by using Isolation Forest algorithm developed in Python. Thanks to this system, abnormal data in the system might be detected in real time. In this study, to integrate the business operation platform, we have firstly examined the data of deleting banking applications on the EFT-POS devices and detected the anomaly. The detection helped the company to save more than 50% of the banking applications on the devices from deletion by contacting banks and customers instantly in the last quarter of 2023.

Keywords: Anomaly Detection, Ensemble Methods, Isolation Forest, Outlier Detection, Payment Systems, Unsupervised Learning, Financial Technologies

ÖZET

Anormallikler, normal desenlere uymayan veri örnekleridir. Kötü amaçlı yazılım, dolandırıcılık, siber saldırı, terörist faaliyetler, hatalar, sistem davranışı değişiklikleri, araç ve insan hataları gibi çeşitli nedenler anormallikler oluşturabilir. Anomali tespiti, verilerde beklenmeyen durumların veya desenlerin bulunmasını sağlayan bir tekniktir. Bu beklenmeyen durumlar veya desenler, literatürde verinin beklenen davranışına uymayan anormallikler, aykırı değerler ve beklenmeyen durumlar olarak adlandırılır. Endüstriler için kritik bir görev olan anormallik tespiti için çeşitli araştırmalar ve uygulamalar yapılmaktadır. Literatürdeki anormallik tespitiyle ilgili çalışmalar genellikle güvenlik ve hata bulma faaliyetleri için sistem oluşturmaya çalışmaktadır. Bankacılık ve finans sektöründe, sağlık, üretim, bilişim teknolojileri ve telekomünikasyon, savunma ve hükümet gibi IoT cihazlarının artması, anormallik tespiti pazarını artırabilir. Bu endüstriler düzenli olarak önemli verilerle uğraşır ve suçluların firma altyapısı üzerinde kontrol sahibi olmalarına neden olan ciddi dolandırıcılık, hırsızlık ve saldırılarına karşı savunmasız olabilirler. Anormalliği tahmin etmek için denetimli, yarı denetimli ve denetimsiz birçok öğrenme yöntemi bulunmaktadır. Geleneksel bir yöntem olarak, anormallikler sabit eşik seviyesine göre alarm verilebilir. Bu yaklaşım tatmin edici ve doğru bir sonuç sağlamayabilir. Bu kapsamda, bu çalışma, Türkiye'de ödeme sistemleri sektöründe lider olan bir firmanın işletme platformu için otomatik bir anormallik tespit sistemi oluşturma konseptini önermektedir. Python'da geliştirilen İzolasyon Ormanı algoritmasını kullanarak bu sistem sayesinde Finansal Teknolojiler şirketinde, sistemdeki anormal veriler anlık olarak tespit edilmektedir. Bu çalışmada, işletme platformunu entegre etmek için, EFT-POS cihazlarındaki bankacılık uygulamalarının silinme verilerini inceledik ve anormalliği tespit ettik. Bu tespit, Finansal Teknolojiler alanında yer alan firmanın 2023 yılının son çeyreğinde bankalar ve müşterilere hemen ulaşarak cihazlardaki bankacılık uygulamalarının %50'den fazlasının silinmesini engellemesine yardımcı oldu.

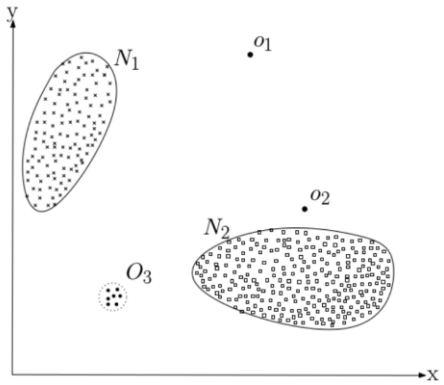
Anahtar Kelimeler: Anormallik Tespiti, Denetimsiz Öğrenme, Finansal Teknolojiler, İzolasyon Ormanı, Ödeme Sistemleri, Topluluk Öğrenmesi

*Corresponding Author (Sorumlu Yazar), e-mail: dicle.aslan@tokeninc.com

Submission Date Başvuru Tarihi	Revision Date Revizyon Tarihi	Accepted Date Kabul Tarihi	Published Date Yayım Tarihi
01.03.2024	15.05.2024	11.06.2024	30.06.2024

1. GİRİŞ

Anomali, verinin beklenen davranışıyla uyuşmayan durumları veya desenleri ifade edebilir [1]. Bu, diğer gözlemlerden önemli ölçüde farklı olan nadir gözlemlerdir. Şekil 1'de, N_1 ve N_2 olarak iki gözlem alanı bulunmaktadır. Bu alanlardaki noktalar normaldir; ancak, bu alanlardaki noktalar; o_1, o_2, o_3 anomali olarak kabul edilir.



Şekil 1. 2-boyutlu bir veri setindeki anormallikler [1]

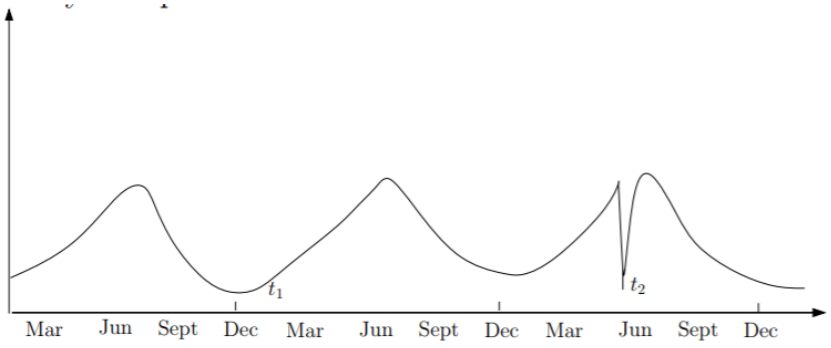
Şekil 1'de gösterilen yapıda normal ve anormal noktaların tespiti genellikle kolaydır. Ancak, genellikle normal ve anormal davranış durumlarının sınırı belirsizdir. Dolayısıyla, yakın bir alandaki anormal bir gözlem normal gibi görünebilir ve bu tanımlanması zor olabilir. Anormallik tespitinde belirli zorluklar vardır. Çoğu veri setinde, anormallik tespiti için normal ve anormal veriler genellikle açıkça küme oluşturmaz. Bir normal veri, anormallik verisi kümesine yakın olabilir ve bir anormallik verisi, normal veri kümesine yakın olabilir. Bu durumda, anormallik tespiti oldukça zorlaşır. Normal olarak adlandırdığımız davranışlar veya veriler zamanla değişebilir. Dolayısıyla, normal davranışları her zaman tanımlamak mümkün olmayabilir. Tüm alanlara belirli bir anormallik tespit tekniği uygulanamayabilir. Örneğin, tıbbi alandaki vücut sıcaklığındaki küçük bir dalgalanma anormallik davranışını gösterebilirken, hisse senetlerindeki küçük bir dalgalanma normal davranışı gösterebilir. Bu nedenle, tüm alanlara bir anormallik tespit yöntemi uygulamak mümkün değildir. Veri setlerindeki gürültü anormalliklerinin tespiti için gürültü giderme gereklidir. Ancak, gürültüyü ayırt etmek çok zor bir süreçtir [1].

Anormallik türleri, nokta anormallik, bağlamsal anormallik ve kolektif anormalliktir. Bir bireysel durumun diğer normal verilere uzak olması durumunda, bu bir nokta anormalliği verisidir çünkü anormallik tespiti bir özellikten bağımlıdır. Örneğin, kredi kartımızda harcanan tutar (harcama miktarı), anormallikleri tespit etmek için kullanılabilir. Şekil 1 nokta anormallikleri gösterir. Bir veri örneği her metinde anormal ise, bu bağlamsal bir anormallik olarak adlandırılır ve aynı zamanda koşullu bir anormallik olarak da adlandırılır. Bir bağlam, veri setindeki yapı tarafından indirgenir ve problem formülasyonunun bir parçası olarak belirtilmelidir. Bu bağlam, bazı durumlarda bazı verilerimizin anormalliğe işaret ettiği ve diğer durumlarda normal verilere işaret ettiği bir örnek olarak, bir bağlamda anormallik davranışı sergilerse bir anormallik örneğidir.

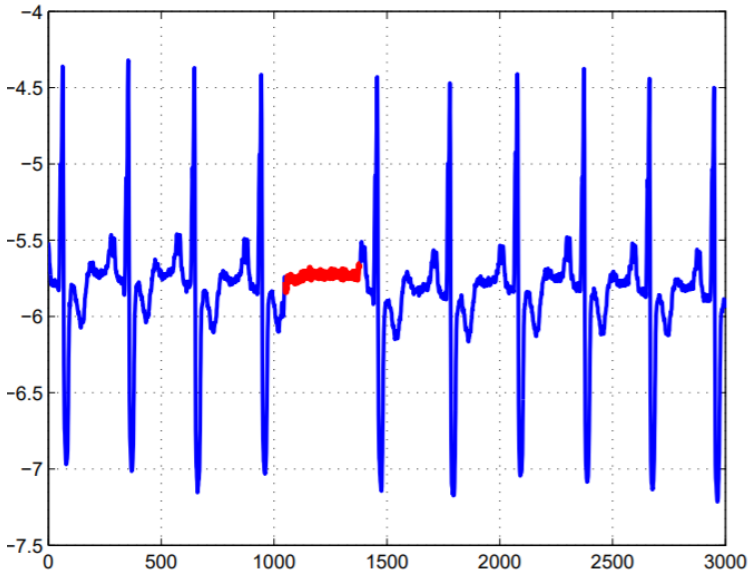
Şekil 2, belirli zamanlarda yılın belirli zamanlarında sıcaklık değişikliklerini gösteren bir zaman serisini sunar. Bu örnekte, t_1 zamanında düşük sıcaklık kışın normal davranışken, t_2 zamanında düşük sıcaklık yazın anormaldir.

Bu, verilerin birbiriyle ilişkilendirilmesi durumunda tüm veri setinde anormallik davranışı oluşturan kolektif bir anormallik örneğidir. Onunla ilişkilendirilen verilerden bazıları birlikte anormallik oluşturabilirken, bu veriler tek başına herhangi bir anormallik davranışı göstermeyebilir. Bu tür bir anormallik, bazı bilgisayar üretimi eylemlerin birlikte gerçekleştiğinde anormallik gösterdiğinde görülebilir.

Şekil 3, bir insanın elektrokardiyogramının çıktısını gösteren bir örneği açıklar. Kırmızı çizgi bölgesi, uzun süreli dönemde istikrarlı düşük değerin anormal olarak mevcut olduğu için anormallik gösterir.



Şekil 2. Sıcaklık zaman serisinde bağlamsal anormallik [1]



Şekil 3. İnsan elektrokardiyogramı çıktısında kolektif anormallik [1]

Uygulama Türkiye'de önde gelen ve %50 Pazar payına sahip bir Finansal Teknolojiler firmasında yapılmaktadır. İlgili firma, Türkiye'de 850.000 üzerinde terminale ve içerisinde de bankacılık, yemek kartı ve diğer uygulamalara sahiptir.

Literatürde, gerçek zamanlı veri akışları için otomatik anormallik tespit sistemi, bir Finansal Teknolojiler firması tarafından her EFT-POS (Noktadan Satışta Elektronik Fon Transferi) cihazı için uygulanan ilk deneyimdir. Çalışmanın geri kalanı şu şekilde yapılandırılmıştır: İkinci bölümde, anormallik tespitinde önceki çalışmalar tartışılmaktadır. Üçüncü bölüm, denetimsiz anormallik tespit algoritmaları hakkında bilgi sunmaktadır. Sonraki bölümde, yaklaşım, veri ve modelleme dahil olmak üzere çalışmanın metodolojisi incelenmektedir. İlerleyen bölüm, deneysel sonuçların tartışılmasını içermektedir. Sonuçlar ve ileri araştırmalar, makalenin son bölümünde önerilmektedir.

2. LİTERATÜR ARAŞTIRMASI

Anormallik tespiti ve aykırı değer tespiti konuları literatürde en yaygın olarak kullanılan konulardandır. Ayrıca, bu çalışmalar gelişen teknolojinin etkisiyle endüstriler tarafından gerçekleştirilmektedir. Çalışma, istatistiksel ve makine öğrenimi algoritmalarında geliştirilen aykırı değer tespiti tekniklerinin kapsamlı bir incelemesini sunmaktadır. Bu çalışmada, sınıflandırma, kümeleme, en yakın komşu istatistiksel temelli teknikler kullanılmıştır [2]. [3], hem sayısal hem de sembolik veriler için aykırı değer tespiti madenciliği tekniklerinin kapsamlı bir incelemesini sunmaktadır. Bu çalışmalar, sırasıyla istatistiksel ve sinir ağı tabanlı alanlarda yenilik tespiti sağlar [4-5]. [6], anormallik tespit sistemleri ve siber sızma tespit sistemlerinin geniş bir incelemesini sunmaktadır.

Son dönem çalışmaları genellikle Nesnelerin İnterneti (IoT) konusundaki anormalliklerle ilgilidir. Genişletilmiş IoT altyapısının kullanımıyla, çalışma, IoT sistemlerindeki saldırıları ve anormallikleri doğru bir şekilde tahmin etmek için çeşitli makine öğrenimi tekniklerini kullanır. Bu çalışmada, makine öğrenimi algoritmaları olarak lojistik regresyon (LR), karar ağacı (KA), destek vektör makinesi (DVM), yapay sinir ağı (YSA), rastgele orman (RO) kullanılmış ve bu yöntemler performans ölçütlerine göre karşılaştırılmıştır. Sistem, KA, RO ve YSA teknikleri kullanılarak %99,4 doğruluk elde etmiştir [7]. [8], gerçek dünya veri setinde sistem sağlık durumunu tahmin etmek için makine öğrenimi sınıflandırıcıları önermektedir. Sonuçlar, tekrarlayan sinir ağları tekniğinin sağlık sorunlarında anormallikleri tahmin etmede daha etkili olduğunu göstermektedir. Çalışma, anormallik davranışına dayalı bir sızma tespit sistemi geliştirir. Çalışmanın deneysel sonuçları, bilinen ve bilinmeyen sensör saldırıları için doğru tespit sağlar [9].

Çalışma, denetimsiz öğrenme paradigmasından türetilmiş yarı denetimli anormallik tespiti için matematiksel bir yöntem sunmaktadır. Bu, destek vektör veri açıklamasına (SVDD) dayanmaktadır [10]. [11], yüksek boyutlu veri setleri için hibrit yarı denetimli anormallik tespiti önermektedir. Veri, derin otokodlayıcı (DAE) ve ansambl k-en yakın komşu grafik tabanlı anormallik dedektöründen oluşur. [12], izolasyon ormanı algoritmasını kullanarak insansız hava aracında gerçek zamanlı anormallik tespiti sunar. [13], 10 farklı veri setinde kullanılan 19 farklı denetimsiz anormallik tespiti algoritmasının uygulanmasını önermektedir. Bu makale, denetimsiz öğrenme için yeni ve iyi finanse edilmiş bir inceleme olmayı amaçlamaktadır.

[14], sunuculardaki anormallikleri tespit etmek için kullanılan tekniklerin performanslarını incelemiştir. Çalışmada beş tekniğin (Yapay Sinir Ağları (YSA), Karar Ağacı, Rastgele Orman, K-En Yakın Komşu ve Ekstra Karar Ağacı) performansları karşılaştırılmış ve en iyi sonucun YSA algoritmasının verdiği tespit edilmiştir. [15], bilgisayar ağlarına yapılan saldırıları ağ trafiğindeki anormallikler kapsamında saldırı tespit sistemi geliştirilmiştir. İlgili sistemin geliştirilmesi kapsamında Rastgele Orman ve Ekstrem Gradyan Artırma algoritmaları en başarılı performans göstermiştir. [16], elektrik güç dağıtımında akıllı sayaç verilerindeki anormallikleri tespit etmek ve tahminleme yapmak için İzolasyon Ormanı, Yerel Aykırı Değer Faktörü ve FbProphet algoritmaları kullanılarak performanları karşılaştırılmıştır. Enerji tüketim ölçümlerinin hatalı yapılmasının önüne geçmek ve verimliliği ve tasarrufu artırmaya yönelik bir sistem geliştirmek istenmiştir. Çalışma sonucunda ise İzolasyon Ormanı yönteminin en başarılı sonuçlara sahip olduğu ve ilgili yöntemin kullanılarak sistemin geliştirilmesi gerektiği tespit edilmiştir. [17], IoT sistemlere yönelik siber saldırıların önceden tahmin edilmesine yönelik yapay zeka teknikleri ile tahminleme yapmışlardır. Klasik makine öğrenme tekniklerinden Destek Vektör Makineleri ve Naive Bayes algoritmaları ile derin öğrenme tekniklerinden Uzun Kısa Süreli Bellek (LSTM) algoritması çalışma kapsamında tercih edilmiş ve performans karşılaştırması yapılmıştır. LSTM algoritma sonuçlarının diğer klasik makine öğrenme tekniklerine göre daha etkili sonuç gösterdiği tespit edilmiştir.

Son yıllarda, araştırmalara göre, yarı denetimli ve denetimsiz öğrenme anormallik tespiti ve IoT konuları literatürde en çok incelenen konular arasındadır. Literatürde yer alan çalışmalarda da denetimsiz öğrenme tekniklerinden İzolasyon Ormanı tekniği ile pratik ve etkili sonuçlar elde edildiği görülmüştür. Bu çalışma, bir Finansal Teknolojiler firmasındaki ilk gerçek kullanım vakası olarak cihazlardaki bankacılık uygulamalarının İzolasyon Ormanı tekniği kullanılarak denetimsiz anormallik tespitini sağlamaktadır. Bankacılık uygulaması silinmesi, şirket için oldukça önemli bir konudur çünkü gelirin büyük bir kısmı uygulamalardan gelmektedir.

3. METODOLOJİ

Bu metodoloji bölümünde, yaklaşımı, veri toplama yöntemini, modeli ve veri etiketleme adımlarını açıklıyoruz. Bu çalışmada, veri toplama ve model oluşturma kısımları, diğerlerine kıyasla daha kritik ve zordur. Verilerin ve özniteliklerin hazırlanmasının ardından, anormallikleri tespit etmek için Python kullanarak denetimsiz öğrenmeyi kullandık.

3.1. Yaklaşım

Bu çalışmada, ilk olarak anormallik tespit modelini oluşturmak için ilgili verileri ve Finansal Teknolojiler firmasına ait EFT-POS cihazlarında bankacılık uygulamalarının silinmesine ilişkin verileri kullandık. Şirket için gerçek zamanlı veri akışlarında anormallikleri tespit etmek kritik bir konudur.

Verilerde herhangi bir eksik değer veya geçersiz ölçüm bulunmamaktadır. Bu, son 3 yıla ait haftalık uygulama silme verilerini içerir. Bu çalışmada, anormallik tespiti ikili bir sınıflandırma olarak tanımlanmıştır. Verileri "normal" ve "anormallik" sınıfları olarak etiketliyoruz. Ardından, durumumuza en iyi çözümü bulmak için Python'da İzolasyon Ormanı modelini oluşturup geliştiriyoruz. Anormallik tespiti için uygun olan ve en iyi çözümü elde etmek için denetimsiz öğrenme tekniğini kullandık. İlerleyen bölümlerde, bu adımları detaylı olarak açıklayacağız.

3.2. Veri Toplama

Firmaya ait işlem platformuna otomatik anormallik tespit aracını entegre ettik. Operasyonel anormallikleri tespit etmek için öncelikle şirket için hayati öneme sahip olan bankacılık uygulamalarının silinme verilerini aldık. Veri setimizde, son 3 yılı içeren 18 ortağın haftalık bankacılık uygulamalarının silinme verileri bulunmaktadır. Her ortağın 150 uygulama silme kaydı bulunmaktadır. Bankacılık uygulamalarının silinmesine ilişkin verilerde bağlamsal anormallikler bulunmaktadır. Bu bağlam, veri noktalarımızın bazı durumlarda normal ve anormalliğe işaret ettiği, ancak belirli bir bağlamda, örneğin farklı ortaklarda anormallik gösteren bir örnektir. Üretim, kalite, satış, kurulum uygulamaları gibi diğer işletme verileri bir sonraki aşamada birbirine bağlanmıştır.

3.3. Model

Modelimizde, veri setindeki anormallikleri tahmin etmek için Python'da derlenmiş bir denetimsiz öğrenme tekniği olarak İzolasyon Ormanı algoritmasını kullandık. Bu Python kodunu, ortaklara ait tek değişkenli veriler (uygulama silme) için anormallikleri ve veri setindeki anormallik puanını bulmak amacıyla uyguladık. İzolasyon Ormanı algoritmasının parametrelerini belirlemenin kritik bir nokta olduğunu belirtmek gerekir.

Klasik "parametre" kodu, İzolasyon Ormanı modelinde şu şekildedir:

“class sklearn.ensemble.IsolationForest(n_estimators=100, max_samples=’auto’, contamination=’legacy’, max_features=1.0, bootstrap=False, n_jobs=None, behaviour=’old’, random_state=None, verbose=0, warm_start=False)”

Kontaminasyon, veri setindeki aykırı değerlerin oranını ifade eder. Yani, kontaminasyon, veri kümesindeki toplam veri noktalarının ne kadarının anormal olduğunu ifade eder. Veri kümesinin anomalilerle nasıl kirlendiğinin belirlenmesi gerekmektedir. Modelimizde, kontaminasyon oranının veri görselleştirmemizden gelen sezgisel bir değere dayanarak %12 olduğunu belirledik. Genellikle literatürde bu değer %22'dir. Çalışmamızda kontaminasyon değerini daha düşük belirlememizin sebebi, verimizdeki aykırılıkları (kirliliği) daha düşük tutmak istememizdendir. Özellik sayısı 1'dir. Her nokta, İzolasyon Ormanı yöntemiyle diğer noktalardan rastgele olarak ayrılır. Model, her bir noktanın bir düğümü temsil ettiği bir ağaç oluşturur. Anomali veri noktalarının ağaç yollarının genellikle normal veri noktalarından çok daha kısa olması nedeniyle, izolasyon ormanındaki ağaçların geniş bir derinliğe sahip olmasına gerek yoktur, bu nedenle daha küçük bir *max_depth* kullanılabilir ve daha az bellek gereksinimi oluşur. Orman izolasyon yönteminde*, n_estimators* ve max_sample parametrelerine dayanarak bir ağaç ormanı oluşturulur ve skor çıkarılır.

3.4. Veri Etiketleme

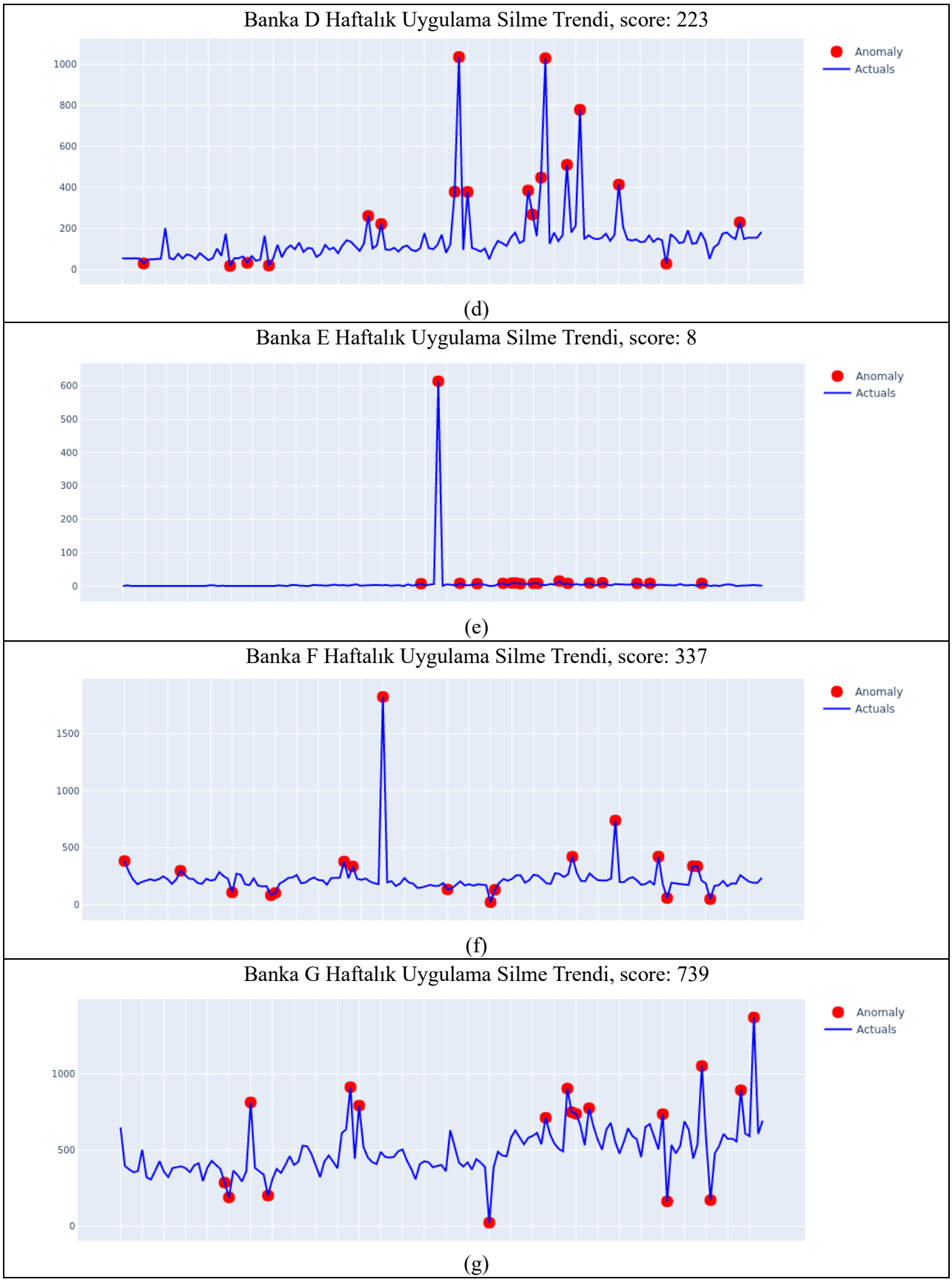
Anormallik tespit algoritmasının iki olası etiket sonucu vardır: anormal veya normal. Çalışma kapsamında her banka verisi için etiket ataması yaptık. Anormallik puanı veya güven değeri, anormalliğin derecesini gösterir. Anormallikleri tespit etmek için veri setindeki büyük miktardaki veriyi manuel olarak işlemek oldukça zor, zaman alıcı ve hata yapmaya açık bir faaliyettir. Ayrıca, geleneksel bir yöntem olarak, veri setindeki anormalliği belirlemek için genellikle sabit bir değer düşünülür. Ancak, bu yöntem tüm verilere ve özelliklere uygun olmadığından, bu şekilde doğru ve etkili sonuçlar sağlamaz. Bu görevi sürekli olarak daha doğru, daha kolay ve daha hızlı hale getirmek için işletme işlem platformuna entegre edilecek otomatik bir anormallik tespit sistemi geliştirdik. Bu sistem sayesinde, hem tek değişkenli hem de çok değişkenli durumlarda veri setindeki her bir öğe için anormallikler otomatik olarak tespit edilebilir.

3.4. Model Sonuçları

Bu bölümde, modelimizin sonuçlarını belirtiyoruz. Şekil 4, bankacılık uygulaması silme işleminin haftalık eğilimini ve gerçek zamanlı veri setinde tespit edilen anormallikleri göstermektedir. Ayrıca, sistemde uyarı yapmak için her ortak için anormallik puanını bulduk.

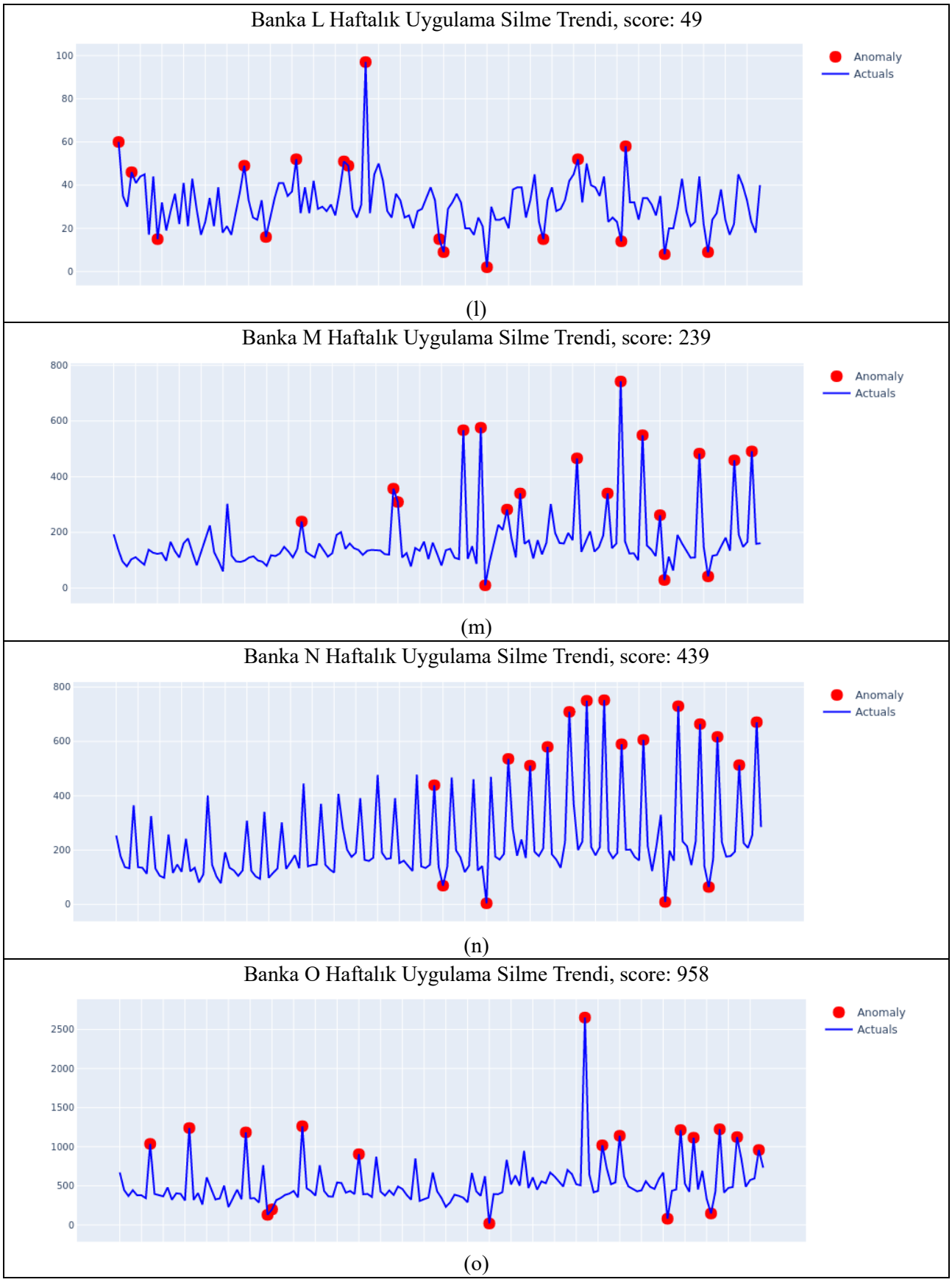
Şekil 4, her bir ortak için haftalık bankacılık uygulaması silme eğilimini, anormallikleri ve anormallik puanını göstermektedir. Anormallik puanı, sabit eşik seviyesi yerine kullanılan alarm seviyesidir. Banka A için, uygulama silme işlemi 490'a ulaştığında, bu kritik değer olduğu için sistem otomatik olarak kullanıcıları uyarır.

Şekil 4. (a) Banka A; **(b)** Banka B; **(c)** Banka C; **(d)** Banka D; **(e)** Banka E; **(f)** Banka F; **(g)** Banka G; **(h)** Banka H; **(i)** Banka I; **(j)** Banka J; **(k)** Banka K; **(l)** Banka L; **(m)** Banka M; **(n)** Banka N; **(o)** Banka O; **(p)** Banka P

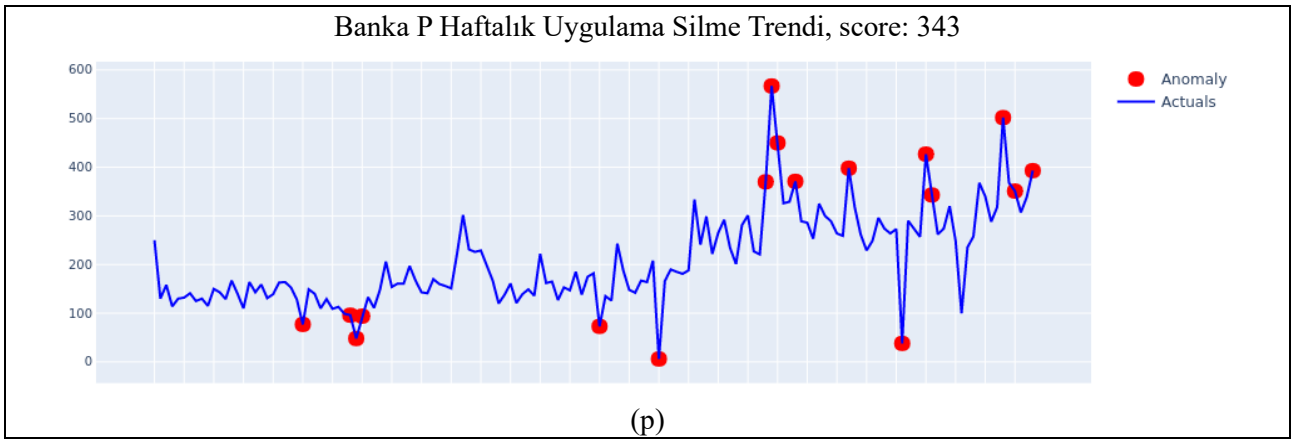


Şekil 4. (devam ediyor)

Şekil 4. (devam ediyor)



Şekil 4. (devam ediyor)



Şekil 4. (devam ediyor)

4. SONUÇLAR VE TARTIŞMA

Son yıllarda, sistemlerdeki anormal durumları bulmak, endüstriler için hayati bir konu haline gelmiştir. Anormallik tespit sistemleri sayesinde, beklenmedik durumlar ve istisnalar tespit edilebilir ve sorunu çözmek için gerekli adımlar atılabilir. Bu tespitlerin önceden yapılabilmesi sistem verimliliğini ve etkinliğini artırmakta aynı zamanda sistem tasarruf sağlayabilmektedir.

Bu çalışmanın literatüre katkılarından biri, Finansal Teknolojiler firmasında işlem platformuna entegre edilecek otomatik anormallik tespit sisteminin uygulanmasına yönelik ilk girişim olmasıdır. Diğer bir katkı ise, çalışmanın şirketin 18 iş ortağı için bankacılık uygulaması silme verilerini kapsaması ve verilerdeki anormallikleri tespit etmek ve her ortak için anormallik puanı bulmak olmuştur. Ortaklar kapsamında bu anormalliklerin önceden tespit edilmesi ortaklar ile iş birliği kapsamında aksiyon alınabilmesini sağlayabilmiştir. Çalışma kapsamında Python'da otomatik anormallik tespiti oluşturmak için İzolasyon Ormanı algoritmasını kullandık. Bu çalışma sayesinde, 2023'ün son çeyreğinde, firma, bankalar ve müşterilerle hemen iletişime geçerek cihazlardaki bankacılık uygulamalarının silinmesini %50'den fazla engelledi. Bu da gelir kaynağı uygulamalar olan firmanın gelirlerin dramatik düşüşünü önleyebilmiş ve uygulamaların silinme nedenleri iş ortakları ile görüşülmüştür. Aynı zamanda çalışma, firma bünyesinde bir tespit sistemi kurulmasını sağlamıştır.

Daha ileri araştırmalarda, işletme işlem platformunda otomatik çok değişkenli anormallik tespiti geliştirmek için üretim, kalite, satış, şikâyet gibi diğer ilgili veriler de dikkate alınacaktır. Mevcut çalışma kapsamında performansının ve etkinliğinin yüksek olduğu bilinen İzolasyon Ormanı yöntemi kullanılmıştır. Ancak ileriki çalışmalarda, yarı denetimli ve diğer denetimsiz teknikler kullanılabilir ve sonuçlar denetimsiz öğrenme tekniği olan İzolasyon Ormanı ile karşılaştırılabilir. Böylece firma bünyesinde kurulan anomali tespit sisteminin de performansı artırılabilecektir.

REFERANSLAR

[1] Chandola, V., Banerjee, A. & Kumar, V. Anomaly Detection: A Survey. ACM Computing Surveys. 15, 1-72, 2009.

[2] Hodge, V. J., &Austin, J. A Survey of Outlier Detection Methodologies. Artificial Intelligence Review. 13-18, 2004.

[3] Agyemang, M., Barker, K. & Alhajj, R. A comprehensive survey of numeric and symbolic outlier mining techniques. Intelligent Data Analysis. 10, 521-538, 2006.

[4] Markou, M. & Singh, Sameer. Novelty detection: a review-part 1: statistical approaches. Signal Processing. 83, 2481-2497, 2003.

[5] Markou, M. & Singh, Sameer. Novelty detection: a review-part 2: neural network based approaches. Signal Processing. 83, 2499-2521, 2003.

[6] Patcha, A. & Park, J-M. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks. 51, 3448-3470, 2007.

[7] Hasan, M., Islam, M., Zarif, I. I., & Hashem, M. M. A. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet of Things, 7, 1-14, 2019.

[8] Huch, F., Golagha, M., Petrovska, A., & Krauss, A. Machine Learning-Based Run-Time Anomaly Detection in Software Systems: An Industrial Evaluation. IEEE Workshop on Machine Learning Techniques for Software Quality Evaluation (MaLTeSQuE). 13-18, 2018.

[9] Pacheco, J. & Hariri, S. Anomaly behavior analysis for IoT sensors, 2016.

[10] Görnitz, N. & Kloft, M. Toward Supervised Anomaly Detection. Journal of Artificial Intelligence Research. 46, 235-262, 2013.

[11] Song, H., Jiang, Z., Men, A. & Yang, B. A hybrid semi-supervised anomaly detection model for high dimensional data. Computational Intelligence and Neuroscience. 1-9, 2017.

[12] Khan, S., Liew, C. F., Yairi, T. & McWilliam, R. Unsupervised anomaly detection in unmanned aerial vehicles. Applied Soft Computing Journal. 83, 1-15, 2019.

[13] Goldstein, M. & Uchida, S. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. Plos one. 1-31, 2016.

[14] Savran, M. F. & Müngen, A. A. Sunucuların Anomali Durumlarinin Yapay Zeka Metotları ile Tahmin Edilmesi. Journal of Computer Science. 8 (2), 57-65, 2023.

[15] Ekici, B. & Takcı, H. Bilgisayar Ağlarında Anomali Tespiti Yaklaşımı ile Saldırı Tespiti. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi. 22, 1016-1027, 2022.

[16] Yarat, S. & Orman, Z. Elektrik Güç Dağıtımında Akıllı Sayaç Verileri için Anomali Tespiti ve Tahminleme. Bilgisayar Bilimleri ve Teknolojileri Dergisi. 3 (2), 72-85, 2023.

[17] Gökdemir, A. & Çalhan, A. Deep learning and machine learning based anomaly detection in internet of things environments. Mühendislik Mimarlık Fakültesi Dergisi. 37 (4), 1945-1956, 2022.