



FİNANS SEKTÖRÜNDE DOLANDIRICILIK TESPİTİ ÜZERİNE MELEZ SINIFLANDIRMA VE REGRESYON AĞACI UYGULAMASI

Ebru PEKEL ÖZMEN

İstanbul Üniversitesi Cerrahpaşa, Endüstri Mühendisliği Bölümü
pkl.ebru@gmail.com

Dr. Öğretim Üyesi Tuncay ÖZCAN

İstanbul Üniversitesi Cerrahpaşa, Endüstri Mühendisliği Bölümü
tuncay.ozcan@istanbul.edu.tr

ÖZET

Veri madenciliği günümüzde ekonomik faaliyetler neticesinden doğan nicel ve nitel verinin karar mekanizmalarında kullanılmak üzere farklı tekniklerle bir araya toplanıp saklı trendlerin ve anlamlı alakaların ortaya konulmasında faal bir biçimde kullanılmaktadır. Finansal illegal işlemlerin tespit edilmesinde yüksek derecede ciddi bir analitik yaklaşım olan veri madenciliği; matematik, istatistik, makine öğrenmesi, örüntü tanıma ve yapay zeka gibi melez tekniklerle geniş veritabanlarından yararlı veriyi elde edip tanımlandırma ve profil oluşturma işlemleri olarak açıklanmaktadır. Bu çalışmada ise finans sektöründe meşru olmayan davranışların (dolandırıcılık faaliyetlerinin) saptanmasına yönelik bir sınıflandırma çalışması yapılmıştır. Kaggle veri tabanında, PaySim adlı simülator kullanılarak üretilen bir sentetik veri seti üzerinde analizler yapılmıştır. Bu amaç doğrultusunda CART algoritmasını uygulanmış ve daha sonrasında CART algoritması Genetik Algoritma ile optimize edilmiştir. Önerilen GA-CART modeli klasik CART algoritmasına göre %37 daha iyi sonuç vermiştir.

Anahtar Kelimeler: CART, sınıflandırma, makine öğrenmesi, dolandırıcılık

HYBRID CLASSIFICATION AND REGRESSION TREE APPLICATION ON FRAUD DETECTION IN FINANCE SECTOR

ABSTRACT

Today, data mining is used actively in the determination of hidden trends and meaningful relevancy by gathering different techniques to be used in decision mechanisms of quantitative and qualitative data arising from economic activities. Data mining, a highly serious analytical approach to identifying financial illegal transactions; It is described as the process of identification and profiling by obtaining useful data from large databases through hybrid techniques such as mathematics, statistics, machine learning, pattern recognition and artificial intelligence. In this study, a classification study was carried out to determine the illegitimate behaviors in the financial sector (fraud activities). In the Kaggle database, analyzes were performed on a synthetic data set produced using a simulator called PaySim. For this purpose, CART algorithm was applied and then CART algorithm was optimized with Genetic Algorithm. The proposed GA-CART model performed a good classification performance. The proposed GA-CART model is 37% better than the traditional CART algorithm

Keywords: CART, classification, machine learning, fraud

1. GİRİŞ

Dolandırıcılık suçu, bireylerin malvarlığı güvenliğini tehdit eden en ağır ihlallerden biri olarak görülmektedir. Bu durum, insanları gündelik hayattaki sıradan işlemlerde dahi fazladan tedbirli olmaya sevk etmektedir. Bunun esas sebebi olarak dolandırıcılık suçunun dinamik yapısı gösterilebilir. Özellikle gelişen teknoloji, getirdiği daha rahat yaşam ile birlikte suçun icrasını kolaylaştıran bir başka etken olmaktadır. Böylelikle suç modern sanayi toplumunda daha büyük bir sorun teşkil etmektedir. Konuya dair literatürde de oldukça fazla çalışmalar bulunmaktadır. Ghosh ve Reilly çalışmalarında, bir kredi kartı düzenleyicisinden gelen geniş bir kredi kartı hesap işlem numunesi üzerinde eğitim vermek için, sinir ağı tabanlı sahtekârlık tespit sistemini kullandı (Ghosh ve Reilly, 1994). Sinir ağı, kayıp kartlar, çalınan kartlar, uygulama sahtekârlığı, sahte dolandırıcılık, posta sipariş dolandırıcılığı gibi dolandırıcılık örnekleri üzerine eğitilmiş ve sonraki iki aylık dönem boyunca tüm hesap etkinliklerini kapsayan gizli bir veri kümesinde test edilmiştir. Ağ, kural tabanlı dolandırıcılık saptama prosedürlerine kıyasla çok daha az sahte pozitif sahtekârlık hesapları tespit etmiştir. Hanagandi ve arkadaşları çalışmalarında yoğunluk tabanlı kümeleme yaklaşımı ile radyal bir temel işlev ağı (RBFN) kullanan bir sınıflandırma yöntemini tanımlamıştır (Hanagandi, Dhar ve ark., 1996). Bu model ile kredi kartıyla ilgili geçmiş işlemleri, dolandırıcılık puanı yaratarak ve daha sonrasında kredi kartı dolandırıcılığını tespit etmek amaçlanmıştır. Dorronsoro, Ginel, Sgncnez ve Cruz 1997 yılında yaptıkları çalışmalarında kredi kartı işlemlerinde sahtekârlık tespiti için sinirsel bir sınıflandırıcıya dayalı çevrimiçi bir sistem kurdular (Dorronsoro, Ginel ve ark., 1997). Dorronsoro ve ekibine göre kredi kartı trafiğinin temel özellikleri arasında, uygun ve hileli işlemlerdeki dengesizlik durumu ve her ikisi arasında büyük bir karıştırma vardır. Uygun model yapımını sağlamak için, dolandırıcılık operasyonlarını diğerlerinden daha iyi ayıran Fisher'in diskriminant analizinin doğrusal olmayan bir biçimi kullanılmıştır.

Fawcett ve Provost çalışmalarında dolandırıcılık tespiti amacıyla, bir dizi veri madenciliği tekniği kullanarak kullanıcı profillemeye yöntemlerinin otomatik tasarımını açıklamışlardır (Fawcett ve Provost, 1997). Özellikle, büyük müşteri işlemleri veri tabanından hileli davranış göstergelerini ortaya çıkarmak için bir kural öğrenme programı kullanılmış ardından, göstergeler, meşru müşteri davranışını sınıflandıran ve anormallikleri gösteren bir dizi monitör oluşturulmuştur. Monitörlerin çıktıları, yüksek güvenlikte alarmlar üretmek için kanıt birleştirmeyi öğrenen bir sistemdeki özellikler olarak kullanılır. Sistem, arama kayıtlarının olduğu bir veri tabanı üzerinde hücresel klonlama sahteciliğini tespit etme sorununa uygulanmıştır. Quah ve Sriganesh çalışmalarında gerçek zamanlı dolandırıcılık tespiti üzerinde durmuş ve potansiyel dolandırıcılık davalarını çözmek için harcama kalıplarını anlamada yeni ve yenilikçi bir yaklaşım önermişlerdir (Quah ve Sriganesh, 2008). Dolandırıcılık tespiti için müşteri davranışlarını çözmek, filtrelemek ve analiz etmek için öz düzenleme haritası (SOM) kullanmışlardır. Srivastava ve arkadaşları çalışmalarında kredi kartı dolandırıcılığını tespit etmek için Gizli Markov Modeli (HMM) kullanılmasını önerdi (Srivastava, Kundu ve ark., 2008). Kredi kartı işlem prosedüründeki işlemleri sıralayıp ve dolandırıcılık tespiti için nasıl kullanılabileceğini göstermişlerdir. Ma ve Li ise çalışmalarında kredi kartı dolandırıcılığının değerlendirilmesinde genetik algoritmanın bir uygulaması önermişlerdir (Ma ve Li, 2009). Ticari bir bankadaki kredi kartı müşterilerinin tarihsel verilerini inceleyerek genetik algoritmaların kredi kartı dolandırıcılığını önleme ve bu tür bir problemle uğraşmada bankaya destek verme imkânı olduğuna işaret etmektedirler. Ma ve Li çalışmalarında gerçek vaka analizinde, genetik algoritmanın tanımlama modelini kullanarak ideal doğruluk tanımlama sonucunu ürettiğini göstermişlerdir. Whitrow ve arkadaşları çalışmalarında işlem kayıtlarını toplamak için bir çerçeve geliştirerek verileri sahtekâr sınıflandırıcıya en iyi şekilde sunmanın genel problemi ele almaktadır (Whitrow, Hand ve ark., 2009). Bunun, algılama performansını nasıl etkilediğini araştırmış, çeşitli sınıflandırma yöntemleri ve gerçekçi bir maliyet temelli performans ölçütü kullanılarak işlem seviyesi algılamasına karşı değerlendirmişlerdir. Böylelikle, toplama süresi için optimum uzunluk bulmayı amaçlamışlardır. Duman ve Özçelik çalışmalarında, bir bankada kullanılan kredi kartı sahtekârlık tespit sistemini parametrelerin değerleriyle oynayarak

geliştiren bir yöntem geliştirmişlerdir(Duman ve Ozelik, 2011). Çalışmanın amacı, tipik sınıflandırma problemlerinden farklı olarak, yanlış sınıflandırma maliyeti değişkenliğidir. Standart veri madenciliği algoritmaları bu durumla iyi uyuşmadığından, meta-sezgisel algoritmalar kullanmaya karar vermişlerdir. Bu amaçla iki iyi bilinen yöntemi birleştirmişlerdir: genetik algoritma (GA) ve dağılık arama (SS). Bu çözüm ile her işlem puanlanır ve bu puanlara dayanarak işlemler hileli veya meşru olarak sınıflandırılır. Yöntem, gerçek verilere uygulanmış ve mevcut uygulama ile karşılaştırıldığında çok başarılı sonuçlar elde edilmiştir. Taklikar ve Kulkarni, GA ve Yapay Bağışıklık Sistemi (AIS) ile kullanıcı tabanlı modele dayalı bir kredi kartı dolandırıcılığı tespit sistemi üzerinde çalışmışlardır(Taklikar ve Kulkarni, 2015). Önerilen sistem, işlemin hakiki veya hileli işlem olduğunu tespit eden kendi kendini uyarlayan, dolandırıcılık tespit etme yeteneğini sürekli güncelleyen bir sistemdir. Model belirli bir işlemi algılamak için tarihsel ayrıntıları, açıklamasını ve görelî çözümünü normal veya sahte işlem olarak tutar ve herhangi bir hileli işlem, fatura işlemi yapılmadan önce uyarı sayesinde önceden fark edilir. Taklikar ve Kulkarni'ye göre bu durum, kart hamiline, kart düzenleyicisine veya bankalara yönelik büyük maddi kayıpların azaltılmasında yardımcı olacaktır. Doğrusal ayırt edici işlevler daha az karmaşık sınıflandırıcı olmasına ve kredi kartı sahtekârlığı algılama gibi yüksek boyutlu problemlerde çalışmasına rağmen bugüne kadar önemli bir ilgi görmemiştir (Mahmoudi ve Duman, 2015). Zeager ve arkadaşları çalışmalarında sahtekârlık motivasyonları ve bilgi tabanına ilişkin bilgileri uyarlanabilir sahtekârlık tespit sistemine dâhil etmeyi amaçlamaktadır(Zeager, Sridhar ve ark., 2017). Sahtekârların en iyi stratejisini modellemek ve dolandırıcılık tespit sistemini, gelecekteki dolandırıcılık işlemlerini daha iyi sınıflandırmak için önceden uyarlamak için bir oyun kuramsal düşman öğrenme yaklaşımı kullanmışlardır. Safa ve Ganga (2019) k-en yakın komşu algoritması, lojistik regresyon ve saf bayes algoritmalarını kullanarak kredi kartı sahtekarlığı verileri üzerindeki performansını araştırmışlardır. Tekniklerin performansı doğruluk, zaman süresi ve dengeli sınıflandırma oranına göre değerlendirmişlerdir. Sonuçlar saf bayes, lojistik regresyon ve k-en yakın komşu sınıflandırıcıları için optimal doğruluğun sırasıyla %83,00, %97,69 ve %54,86 olduğunu göstermektedir. Karşılaştırmalı sonuçlar, lojistik regresyonun, saf Bayes ve en yakın komşu tekniklerinden daha iyi performans gösterdiğini göstermişlerdir. (Kamusweke, Nyirenda ve ark., 2019) gizli desenleri keşfeden ve keşfedilen desenlere dayanarak dolandırıcılık durumunu tahmin edebilen bir model oluşturmak için veri madenciliğini kullanmışlardır. Çalışmalarında iki aşamada ilerlemişlerdir; öncelikle gizli desenleri ve verideki eğilimleri bulmuşlar, ikinci aşamada ise keşfedilen kalıplara dayanarak sahtekarlığı öngören bir model oluşturmuşlardır. Yazarlar, önerilen işlem modellerinin yüksek dolandırıcılık şansı sunmaya devam edebileceğini ve dolayısıyla dolandırıcılık tespiti için veri madenciliği konusunda daha fazla araştırmanın önemini ortaya koyduğunu göstermiştir.

Bu makale şu şekilde yapılandırılmıştır: ilk bölüm, çalışmanın kapsamı ve bu kapsamda yapılmış olan çalışmalar ile ilgili açıklamalara yer vermektedir. Bölüm II, önerilen modelin genel çerçevesi hakkında bilgiler sunarken, bölüm III önerilen modele ilişkin performans değerlendirme sürecinde faydalanılacak ölçütleri sunmaktadır. IV. bölümde analizde kullanılan verisetine dair bilgiler verilmiş ve bölüm V'de analiz sonucunda elde edilen değerler sunulmuştur. Son olarak bölüm VI çalışma sonuçlarını özetler şekilde açıklamalara yer vermektedir.

MATERYAL VE YÖNTEM

C&RT Algoritması

C&RT algoritması Breiman ve diğerleri tarafından (1984) önerilmiştir. "Classification and Regression Trees" olarak adlandırılan İngilizce ismin baş harflerinden yola çıkarak C&RT algoritması olarak adlandırılmıştır. C&RT algoritması kök düğümden başlayarak her düğüm için olası tüm ayırma şekillerini gözden geçirerek bunlardan en iyisini seçer. C&RT algoritmasında, her düğümde iki dal üretir. Yani bölünmeler (ya da ayrılmalar) ikilidir, nitel değişken üç seviyeli de olsa üç farklı bölünme şekli de incelenmelidir. Olası farklı bölünme şekillerine karar verildikten sonra

algoritma her adayın ne kadar başarılı bir ayırıcı olduğunu hesaplayarak en yüksek olan kök düğüm olarak seçilecektir. Bütün düğümler saflaşana kadar algoritmaya devam edilir(Breiman, 2017).

GA-C&RT Algoritması

Karar ağaçlarında tüm nitelikler ağaç yapısına dâhil edilirse; test aşamasında bazı örnekler için yanlış sonuçlar çıkabilmektedir. Bu nedenle karar ağacı oluşturulurken her nitelik bir dala açılmak zorunda değildir. Aksi takdirde aşırı öğrenme denilen durum meydana gelir ve modelin tahmin ediciye sunduğu doğruluk oranı düşük olacaktır. Genetik Algoritmali C&RT (GA-C&RT)'nda hesaplamalardaki belirginliği ve modelin sunacağı doğruluk oranını anlamlı bir şekilde artırabilmek için veriler işleme alınmadan önce ağırlıklandırılma yoluna gidilmektedir. GA-C&RT'in oluşturulabilmesi için veriler en optimal ağırlık değerleriyle güncellenir ve karar ağacı modeline dahil edilir. Klasik karar ağacı modellerinde mevcut data direkt olarak dallanma işlemine alınmaktadır. Önerilen Genetik Algoritma ile Karar Ağacı modelinde ise girdi değişkenlere belirli bir ağırlık yüklenerek ağaca olan etkileri, modelin sunacağı doğruluk oranını optimize edecek şekilde türetilmektedir. Klasik bir Karar Ağacı algoritmasında tüm girdi değişkenler eşit ağırlığa sahiptir. Genetik Algoritma ile ağırlıklar (denklem 1) türetildiğinde her girdinin çıktıya etkisi göz önünde bulundurulmakta ve elde edilen performans değeri artırılmaya çalışılmaktadır.

$$(x, Y) = (w_1x_1, w_2x_2, \dots, Y) \quad (1)$$

Algoritmanın sözde kodu aşağıda verildiği gibidir.

Başlangıç ağırlık değerleri **ata** (w_0)

$t \leftarrow 1$;

doğruluk_değeri₍₀₎ = 0;

durma koşulu sağlanana kadar **döngüye al**

verisetini ağırlıklar ile **güncelle**

ağaç yapısını **oluştur**

analiz *doğruluk* değerini **hesapla**

iyileşme \leftarrow *doğruluk*_(t) - *doğruluk*_(t-1)

ga ile ağırlık değerlerini **güncelle**

$t \leftarrow t+1$;

Döngü **bitir**

PERFORMANS ÖLÇÜTLERİ

Model başarımını değerlendirirken kullanılan temel kavramlar hata oranı, kesinlik, duyarlılık ve F-ölçütüdür. Modelin başarısı, doğru sınıfa atanan örnek sayısı ve yanlış sınıfa atılan örnek sayısı nicelikleriyle alakalıdır. Test sonucunda ulaşılan sonuçların başarımları karışıklık matrisi ile ifade edilebilir. Karışıklık matrisinde satırlar test kümesindeki örneklere ait gerçek sayıları, kolonlar ise modelin tahminlemesini ifade eder(Coşkun and Baykal 2011).

Tablo 1. Karışıklık Matrisi

Asıl Sınıf	C1	Doğru Pozitif (TP)	Yanlış Negatif (FN)
	C2	Yanlış Pozitif (FP)	Doğru Negatif(TN)

Doğruluk ve Hata Oranı: Model başarımının ölçülmesinde kullanılan en popüler ve basit yöntem, modele ait doğruluk oranıdır. Doğru sınıflandırılmış örnek sayısının, toplam örnek sayısına oranıdır. Literatürde en çok kullanılan yöntemdir. Hesaplanması oldukça kolaydır. Hata oranı ise bu değerlerin 1'e tamlayanıdır(Nizam ve Akın 2014). Denklem 2 ve 3'te formülleri gösterilmektedir.

$$\text{Doğruluk} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$\text{Hata Oranı} = \frac{FP + FN}{TP + FP + TN + FN} \quad (3)$$

Kesinlik: Kesinlik, sınıfı 1 olarak tahmin edilmiş True Pozitif (TP) örnek sayısının, sınıfı 1 olarak tahmin edilmiş tüm örnek sayısına oranıdır(Denklem 4)(Coşkun and Baykal 2011).

$$\text{Kesinlik} = \frac{TP}{TP + FP} \quad (4)$$

Anma: Doğru sınıflandırılmış pozitif örnek sayısının toplam pozitif örnek sayısına oranıdır(Nizam and Akın 2014).

$$\text{Anma} = \frac{TP}{TP + FN} \quad (5)$$

F- Ölçütü:Kesinlik ve anma ölçütleri tek başına anlamlı bir karşılaştırma sonucu çıkarmamıza yeterli değildir. Her iki ölçütü beraber değerlendirmek daha doğru sonuçlar verir. Bunun için f-ölçütü tanımlanmıştır. F-ölçütü, kesinlik ve anmanın harmonik ortalamasıdır(denklem 6)(Nizam ve Akın 2014).

$$F \text{ Ölçütü} = \frac{2 \times Anma \times Kesinlik}{Anma + Kesinlik} \quad (6)$$

VERİ

Finansal hizmetlerde ve özellikle yeni ortaya çıkan mobil para işlemleri alanında kamuya açık veri setlerinin eksikliği söz konusudur. Finansal veri setleri birçok araştırmacı için ve özellikle de dolandırıcılık tespiti alanında araştırma yapmak için önemlidir. Böyle bir soruna yaklaşım olarak Kaggle veri tabanında, PaySim adlı simülatör kullanılarak üretilen bir sentetik veri seti sunulmuştur. Orijinal veri seti 1048576x7 birimden oluşmakta ve 6 nitelik ve sınıf etiketi vardır. Veri setinde bulunan nitelikler Tablo 2’de özetlenmiştir.

Tablo 2I. Veri Setinin Yapısı

Nitelik	Nitelik Adı	Nitelik Tipi	Nitelik
1	İşlem Tipi	Nominal	1: Ödeme 2: Transfer 3: Nakit 4: Borç 5: Nakit
2	Tutar	Nümerik	
3	Başlangıç	Nümerik	
4	Yeni Bakiye	Nümerik	
5	Alıcı	Nümerik	
6	Alıcı Yeni	Nümerik	
7	Dolandırıcılık (Sınıf Etiketleri)	Nominal	0: 1:

BULGULAR

Analizler Matlab R2014a sürümünde gerçekleştirilmiştir. Önerilen algoritmanın etkisinin daha net görülebilmesi için klasik CART ve GA-CART algoritmalarına ilişkin performans ölçütleri de verilmiştir. Modeller 10-çapraz geçirme yöntemi ile analiz edilmiştir. Aşağıdaki Tablo 3 VE 4'de CART ve GA-CART algoritmalarına ilişkin karmaşıklık matrisi bulunmaktadır.

Sınıflandırıcının performansı için en önemli ölçüt hileli işlemin tahminindeki başarısıdır. Veri setinde sınıf etiketinin dağılımında hileli işlemlerin sayısı, hileli olmayan işlemlere göre çok azdır. Bu sebeple şekil 1 ve 2'de adı geçen doğruluk oranı kriteri, sınıflandırıcının hileli olmayan işlemi tespit edebilme becerisinden soyutlanarak hesaplanmıştır. Böylelikle sınıflandırma algoritmasının dolandırıcılık faaliyetini tespit edebilme becerisi net bir şekilde açıklanmış olacaktır. Tablo 1'deki karmaşıklık matrisinde bakıldığında veriseti üzerindeki doğruluk oranı $TN/(FP + TN + FN)$ olarak dikkate alınmıştır.

Tablo II. CART Karmaşıklık Matrisi

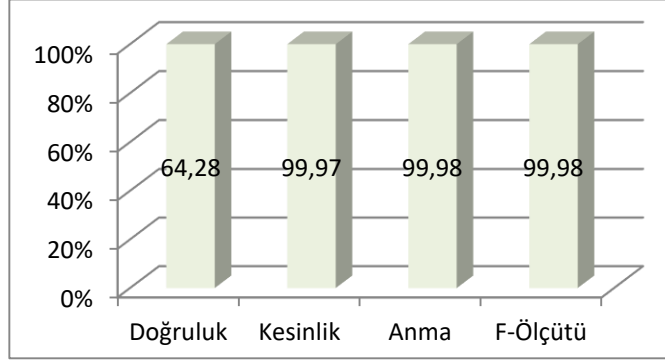
Karmaşıklık Matrisi	
1047209	224
264	878

Tablo 4III. GA-CART Karmaşıklık Matrisi

Karışıklık Matrisi	
1047388	4
98	1044

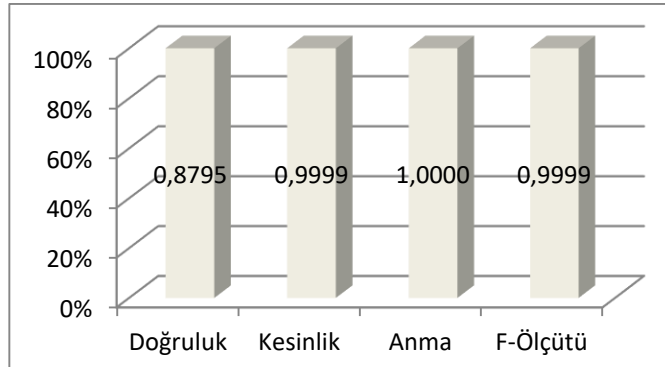
Aşağıdaki şekil 1 ve 2'de modellere ilişkin performans ölçütleri gösterilmektedir. Doğru pozitif değerlerin(TP) gözardı edilmesi ile klasik CART algoritması %64,28 değerini vermiştir. Diğer ölçümlerde TP değerleri hesapmalara katıldığı için oldukça yüksek performanslar ortaya çıkmıştır. Ancak analizleri net bir şekilde yorumlayabilmek adına karşılaştırma yaparken büyük ölçüde doğruluk oranı dikkate alınmıştır.

Şekil 1. Klasik CART Performans Ölçütleri



Önerilen GA-CART modelinin sınıflandırma performansı ise %87,95 olarak bulunmuştur. Klasik CART algoritmasına göre çok daha iyi sonuçlar vermiştir. Bu da göstermektedir ki optimizasyon eşliğinde veri madenciliğinin uygulanması ile probleme göre esnek bir hal alınması sağlanır ve daha efektif sonuçlar alınabilmektedir. Doğruluk kriteri dışındaki performans kriterleri genel olarak yüksek değerlerdir bunun sebebi olarak doğru tahminlenen değerlerin sayıca çokluğu gösterilebilir. Aynı zamanda verilerin sentetik olması, sınıflandırma için fazlasıyla uygun koşulların olması anlamına gelebilmektedir. GA-CART modelinde kesinlik, anma ve f-ölçütü değerlerinin 1 ve 1'e yakın değerler elde edilmesinin bu durumdan kaynaklı olabilir.

Şekil 2. GA-CART Performans Ölçütleri



TARTIŞMA ve SONUÇ

Kredi kartı dolandırıcılığının önlenmesi ve tespiti için birçok model ve yöntem üzerinde araştırma yapılmıştır. Finansal işlem sistemlerinin güvenliğini otomatik ve etkili bir şekilde geliştirmek için, doğru ve etkili bir kredi kartı dolandırıcılığı tespit sistemi kurmak, finansal kurumlar için en önemli görevlerden biridir. Dolayısıyla, dolandırıcılık tespitinin geliştirilmesi meşru işlemleri güvence altına almak için zorunlu hale gelmiştir. Bu doğrultuda yapılan çalışma da finans sektöründe meşru olmayan davranışların saptanmasına yöneliktir. Çalışmada klasik CART algoritmasının dolandırıcılık tespiti üzerindeki performansları değerlendirilmiştir. Aynı zamanda CART algoritması optimizasyon sürecine tabi tutularak GA ile melez bir model önerilmiştir, böyle daha efektif bir sınıflandırma yapılması amaçlanmıştır. Neticede ise önerilen algoritma sınıflandırma performansı açısından daha doğru sonuçlar vermiştir. Belirlenen doğruluk oranı kriteri bazında bakıldığında klasik CART yöntemi ortalama bir değer sunmuştur. Ancak veri seti GA ile bir optimizasyon sürecine dahil olarak güncellendiğinde daha etkin bir sonuç elde edilmiştir ve yaklaşık olarak sınıflandırma performansında daha iyi bir oran elde edilmiştir. Ayrıca klasik CART algoritmasının %64,28 ve GA-CART algoritmasının %87,95 doğruluk oranlarını ürettiği düşünülürse önerilen modelin %37 oranında daha iyi performans sergilediği söylenebilir.

Öte yandan bu veri setinde önerilen algoritmanın efektif olarak çalışıyor olması önerilen modelin tüm problemlerde efektif sonuçlar üreteceği garantisini vermemektedir. Önerilen algoritmaların farklı veri setlerindeki performansları ileriki çalışmalarda sunulacaktır.

KAYNAKÇA

- Breiman, L. (2017). Classification and regression trees, Routledge.
- Coşkun, C. and A. Baykal (2011). "Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması." Akademik Bilişim **2011**: 1-8.
- Dorrnsoro, J. R., F. Ginel, C. Sgnchez and C. S. Cruz (1997). "Neural fraud detection in credit card operations." IEEE transactions on neural networks **8**(4): 827-834.
- Duman, E. and M. H. Ozcelik (2011). "Detecting credit card fraud by genetic algorithm and scatter search." Expert Systems with Applications **38**(10): 13057-13063.
- Fawcett, T. and F. Provost (1997). "Adaptive fraud detection." Data mining and knowledge discovery **1**(3): 291-316.
- Ghosh, S. and D. L. Reilly (1994). Credit card fraud detection with a neural-network. System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on, IEEE.
- Hanagandi, V., A. Dhar and K. Buescher (1996). Density-based clustering and radial basis function modeling to generate credit card fraud scores. IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering (CIFER), IEEE.
- Kamusweke, K., M. Nyirenda and M. onde Kabemba (2019). "A Data Mining Model for Predicting and Forecasting Fraud in Banks."
- Ma, H. and X. Li (2009). Application of data mining in preventing credit card fraud. 2009 International Conference on Management and Service Science, IEEE.
- Mahmoudi, N. and E. Duman (2015). "Detecting credit card fraud by modified Fisher discriminant analysis." Expert Systems with Applications **42**(5): 2510-2516.
- Nizam, H. and S. S. Akın (2014). "Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması." XIX. Türkiye'de İnternet Konferansı.
- Quah, J. T. and M. Sriganesh (2008). "Real-time credit card fraud detection using computational intelligence." Expert systems with applications **35**(4): 1721-1732.
- Safa, M. U. and R. Ganga (2019). "Credit Card Fraud Detection Using Machine Learning."
- Srivastava, A., A. Kundu, S. Sural and A. Majumdar (2008). "Credit card fraud detection using hidden Markov model." IEEE Transactions on dependable and secure computing **5**(1): 37-48.
- Taklikar, S. H. and R. Kulkarni (2015). "Credit Card Fraud Detection System Based On User Based Model With GA And Artificial Immune System." J. Multidiscip. Eng. Sci. Technol. **2**.
- Whitrow, C., D. J. Hand, P. Juszczak, D. Weston and N. M. Adams (2009). "Transaction aggregation as a strategy for credit card fraud detection." Data mining and knowledge discovery **18**(1): 30-55.
- Zeager, M. F., A. Sridhar, N. Fogal, S. Adams, D. E. Brown and P. A. Beling (2017). Adversarial learning in credit card fraud detection. 2017 Systems and Information Engineering Design Symposium (SIEDS), IEEE.