**İŞLETME ARAŞTIRMALARI DERGİSİ**
**JOURNAL OF BUSINESS RESEARCH-TURK**
2022, 14(4), 3053-3065
https://doi.org/10.20491/isarder.2022.1549

# Hile Tespitinde Makine Öğrenmesi Yöntemlerinin Kullanılması ve Model Performanslarının Değerlendirilmesi*

Using Machine Learning Methods to Detect Fraud and Evaluation of Model Performances

**Önder GÜR** [a] **Banu TARHAN MENGİ** [b]

[a] Kırklareli Üniversitesi, Babaeski Meslek Yüksekokulu, Finans, Bankacılık ve Sigortacılık Bölümü, Kırklareli, Türkiye, ondergur@klu.edu.tr

[b] Marmara Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, İstanbul, Türkiye, btarhan@marmara.edu.tr

**MAKALE BİLGİSİ**

**Anahtar Kelimeler:**
Hile Tespiti
Makine Öğrenmesi Yöntemleri
Karar Ağacı

Gönderilme Tarihi 23 Temmuz 2022
Revizyon Tarihi 13 Aralık 2022
Kabul Tarihi 20 Aralık 2022

**Makale Kategorisi:**
Araştırma Makalesi

**ÖZET**

**Amaç –** Çalışmada, hilenin verdiği zararın azaltılmasına yönelik teknoloji temelli araçlarla çözüm üretilmeye çalışılmıştır. İşletmelerde sıklıkla karşılaşılan hileli ödemelerin tespiti için makine öğrenimi yöntemleriyle bir model oluşturulması amaçlanmaktadır.

**Yöntem –** Çalışmada, bir bankaya ait finansal ve finansal olmayanlar bilgilerle oluşturulan 594.643 adetlik yapay veri setinden yararlanılmıştır. Veri seti kullanılarak makine öğrenmesinin Karar Ağacı, Destek Vektör Makinesi, Lojistik Regresyon ve Yapay Sinir Ağları yöntemleriyle tahmin yapılmıştır. Yöntemlerin algoritmaları veri setinin %69'u ile önce eğitilmiş, sonra veri setinin %31'i ile tahmin yapması sağlanmıştır.

**Bulgular -** Yöntemlerin oluşturduğu değerlere bakıldığında doğruluk metriği sırasıyla Karar Ağacı %99,42, Destek Vektör Makinesi %99,11, Lojistik Regresyon %98,95 ve Yapay Sinir Ağları %99,35 hesaplandığı görülmüştür. 1.620 tane hileli işlemi doğru tahmin ederek en fazla doğru tahmin yapan ve bu süreçleri en hızlı (1,32sn) gerçekleştiren yöntem karar ağacı olmuştur. Sonuçlara göre ortalama olarak en başarılı, en hızlı ve en çok doğru tahmin yapan modeli karar ağacının oluşturduğu tespit edilmiştir.

**Tartışma -** Uygulamada kullanılan yöntemlere bağlı olarak veri setiyle yapılan model denemelerinde makine öğrenimi yöntemleri yanlış pozitif ve yanlış negatif değerleri üretmiştir. Her ne kadar bu değerleri sıfıra indirmek mümkün olmasa da azaltılması yönünde geliştirilmeye açıktır.

**ARTICLE INFO**

**Keywords:**
Detect Fraud
Machine Learning
Decision Tree

Received 23 July 2022
Revised 13 Aralık 2022
Accepted 20 Aralık 2022

**Article Classification:**
Research Article

**ABSTRACT**

**Purpose -** It has been tried to produce a solution with technology-based tools to reduce the damage caused by fraud in this study. It is aimed to create a model with machine learning methods for the detection of fraudulent payments, which are frequently encountered in businesses.

**Design/methodology/approach -** In the study, an artificial data set of 594.643 created with financial and non-financial information belonging to a bank was used. Using the data set, predictions were made with the Decision Tree, Support Vector Machine, Logistic Regression and Artificial Neural Networks methods of machine learning. Algorithms of the methods were first trained with 69% of the data set and then made predictions with 31% of the data set.

**Findings -** Considering the values created by the methods, it was seen that the accuracy metric was 99.42% for Decision Tree, 99.11% for Support Vector Machine, 98.95% for Logistic Regression and 99.35% for Artificial Neural Networks. Decision tree was the method that predicted 1.620 fraudulent transactions correctly and made the most correct predictions and performed these processes in the fastest (1.32 seconds). According to the results, it has been determined that the decision tree is the model that makes the most successful, fastest and most accurate predictions on average.

**Discussion -** Depending on the methods used in practice, machine learning methods produced false positive and false negative values in model trials with the data set. Although it is not possible to reduce these values to zero, it is open to development to reduce them.

*Bu çalışma, "Makine Öğrenmesi Destekli Hile Tespiti ve Bir Uygulama" adlı doktora tez çalışmasından üretilmiştir.

**Önerilen Atıf / Suggested Citation**

Gür, Ö., Tarhan Mengi, B. (2022). Hile Tespitinde Makine Öğrenmesi Yöntemlerinin Kullanılması ve Model Performanslarının Değerlendirilmesi, *İşletme Araştırmaları Dergisi*, 14 (4), 3053-3065.

## 1. GİRİŞ

21. yüzyılda büyük ölçekli hile vakalarının yaşanması finansal piyasalara olan güveni derinden sarsmış ve paydaşlarını zarara uğratmıştır. Sertifikalı Hile Denetçileri Birliği (ACFE)'nin hazırlamış olduğu raporlar doğrultusunda hilenin yapılış biçimiyle birlikte vermiş olduğu zarar gözler önüne serilmiştir. Uluslararası ölçekteki otoriteler, gelecekte oluşabilecek hile vakalarına karşı daha ihtiyatlı davranarak düzenlemeler yapmıştır. Ancak hileden korunmak için sadece standartlarla ve düzenlemelerin yeterli olmayacağının farkına varılması sonucunda işletmelerin kendi süreçlerini sürekli denetleyerek hileyi önlemeye yönelik yaptığı çalışmalar da hızlanmıştır.

Günümüzde hile birçok şekilde tanımlansa da hilenin ekonomilere zarar verdiği kesin olarak bilinmektedir. Özellikle son yıllarda yaşanan büyük kayıplar, hilekârların finansal raporlama sistemlerinin açıklarını kullanılmasıyla ortaya çıkmıştır (Jackson, 2015:30). Hileli işlemler işletmenin olağan akışı dışındaki nedenlerden, faaliyetlerin yürütülmesindeki eksikliklerden, bozuk etik değerlerinden ve kültüründen kaynaklanabilmektedir. Önemli görülen noktalarda oluşturulan kontrol noktalarıyla hileli işlemlerin var olduğuna ya da olabileceğine ilişkin işaretler elde edilebilmektedir (Pehlivanlı, 2011:36). Hile soruşturması genellikle zaman alır ve alanındaki uzman olan muhasebeciler, hile denetçileri ve avukatlar tarafından yürütülür. Bu süreçte, önleme, ortaya çıkartma ve araştırma gibi üç temel faaliyet yürütülür. Hem en zoru hem de en önemlisi hilenin ortaya çıkartılmasıdır. Etkili kontroller ve diğer önleyici yaklaşımlar büyük oranda hilekârlığa karşı bir caydırıcılığı olsa da bazıları yine de devam etmektedir. Yani hile yapmayı kafasına koymuş kişiler kasıtlı olarak sistemin içinde boşluk arayacaktır. Diğer yandan, hilenin araştırılabilmesi için hilenin varlığına yönelik önemli bulgulara ulaşılması gerekir (Albrecht ve Albrecht, 2007:49.3).

Hilenin ortaya çıkartılmasının ilk adımı, hile şüphesinin ortaya çıkmış olması veya anormal bir durumun fark edilmesiyle başlamaktadır. Bu tür şüphe; pasif yapının ürünü olan fısıltı hatlarına telefon, e-posta ve diğer web tabanlı araçlarla gelen ihbar ile şikâyetlerin değerlendirilmesi veya tesadüfi olarak karşılaşılmasıyla ya da aktif analizlerin yapılmasıyla ortaya çıkabilir. Kısacası, hile belirtilerinin tespit edilmesi pasif veya aktif yapılar sayesinde gerçekleşir. İki yapı arasındaki fark, pasif yapıların belirtileri ortaya koymak için özel önlemleri gerektirmemesidir. Aktif yapının doğası gereği kritik süreçlerin denetlenmesiyle, iç denetim bulgularıyla ve analizi yapılan yönetim muhasebesi raporlarıyla anormallikler ortaya çıkartılabilmektedir (Stamler vd., 2014:205).

Geçmişte denetçiler, hile belirtilerini tespit etmek için onlarca yıl manuel kontrol yöntemlerine göre uyarlanmış analiz tekniklerini uygulayarak verileri analiz etmişlerdir. Ancak teknoloji temelli araçların gelişmesi hile denetçilerine, büyük veri setlerini analiz etmek için tamamına erişilebilme sağlamıştır (Albrecht vd, 2011:168). Ancak, hile denetçilerinin hile tespit çalışmalarında kullanacağı büyük veri setlerinin elde edilmesinde birtakım zorluklar bulunmaktadır. Bu zorlukların belki de en önemlisi mevcut gerçek hile kayıtlarına ulaşmaktır. Bu zorluğun nedeni, şirketlerin kendi alanlarındaki dolandırıcılık kayıtlarını genellikle açıklamaktan hoşlanmamalarıdır. Çünkü bu durum, hissedar güveninin yanı sıra tüketici güveni üzerinde de negatif bir etkiye sahiptir (Lu, 2006:348-349). Ayrıca kişisel verilerin korunması kanunu ile kişi ya da kurumlara ait verilerin açık rızası olmadan işleyenlere 5.000 TL'den 1.000.000 TL'ye kadar büyük yaptırımların uygulanması (RG[1], 2016) nedeniyle veri setlerine ulaşmayı oldukça zorlaştırmıştır. Bu nedenlerle, kaynak veri olarak kamuya açık olan bir veri seti kullanılmıştır . Kaynak veri, kamuya açık makine öğrenimi ve veri bilimi topluluğu olan **kaggle.com** adlı internet sitesinden sağlanmıştır(Kaggle, 2017). Bu çalışmanın literatüre katkısı denetçilere ve finansal piyasadaki tüm paydaşlara hileli ödemelerin verdiği zararların azaltmasını sağlamak, işletmelerin veri setlerini kullanarak hilelerin tespit edilebilmesini ve önlenebilmesini sağlamak ile makine öğrenmesinin oluşturduğu modellerin performanslarını değerlendirerek denetim alanında yenilikçi yaklaşımların kullanılmasına yardımcı olmaktır. Çalışmaya literatür incelemesiyle başlanmış, hileli ödemeler hakkında bilgi verildikten sonra araştırmanın yöntemine ve verilerine değinilmiştir.

[1]Resmî Gazete, Sayısı: 29677

## 2. LİTERATÜR İNCELEMESİ

Hileli faaliyetlerin tespitiyle ilgili literatürde yapılmış birçok çalışma bulunmaktadır. Aynı zamanda bu çalışmalarda birçok yöntemin kullanıldığı görülmüştür. Ancak ağırlıklı olarak teknoloji temelli yöntemlerin kullanıldığı anlaşılmıştır. Çalışmada veri madenciliğinin bir kolu olan makine öğrenmesine dayalı yöntemler kullanacağı için bu yöntemleri konu alan çalışmalar aşağıda özetlenmiştir.

Son yıllarda, veri madenciliğine dayalı istatistik, makine öğrenimi, yapay zekâ, örüntü tanıma gibi çeşitli disiplinlerden kaynaklanan farklı türde tahmine dayalı analitik teknikler geliştirilmiştir. Hile tespit çalışmalarında hem sınıflandırma hem de regresyon modelleri aynı anda kullanılmaktadır. Örneğin, bir hile tespitinde regresyon analizi dolandırıcılık miktarını tahmin etmeye çalışırken sınıflandırma tekniği bir hile tespitinde dolandırıcılığın varlığına ya da yokluğuna ilişkin tespit yapmaya yardımcı olmaktadır (Baesens, 2015:122). Özellikle çalışan hilelerinden ve hileli finansal raporlamalardan kaynaklı işletmeler büyük tutarlarda zarar etmektedir. Bununla birlikte günümüzde artan iş hacimleri göz önüne alındığında oluşan büyük veriler içinden hileli işlemlerin tespit edilmesi kolay olmamaktadır. Büyük finansal veriler içindeki hileli faaliyetlerin tespit edilmesine yönelik yazılımlar geliştirilmiştir. Veri madenciliği araçlarının bu büyük finansal verilerde kullanılması, hilelerin tespiti ve önlenmesi bakımından işletmelere büyük avantajlar sağlayacaktır (Terzi, 2012:54).

Günümüzde uzmanlar ve araştırmacılar hileleri tespit için veri madenciliğini kullanmaya başlamıştır. Veri madenciliği teknikleri, veri setlerindeki değişkenleri kullanılarak yapılan sınıflandırma ve tahmin için oldukça uygundur. Veri madenciliğinin kolu olan makine öğrenimi, büyük miktardaki veriyi analiz etmede diğer yöntemlere göre daha başarılıdır. Yöntemlerden yapay sinir ağları (ANN), karar ağaçları (DT), destek vektör makineleri (SVM) ve bayes inanç ağları (BBN) hilenin tespiti için popüler araçlardır (Jan, 2018:1-14).

Lægreid (2007:287) çalışmasında, lojistik regresyon (LR) yönteminin oluşturduğu modeli kullanarak hayat dışı sigortalardaki hile faaliyet modellerini otomatik tanıma sistemi geliştirmiştir. Çalışmadaki ana fikir, geçmiş müşteri bilgileriyle hile riskini olasılıksal olarak değerlendirerek ekspertizlere yardımcı olmaktır. Diğer yandan veri madenciliği araçları bankacılıkta da yaygın olarak kullanılmaktadır. Çünkü kredi kartı dolandırıcılığı, bankalar için son derece ciddi bir sorun olmaktadır. Örneğin, Visa ve MasterCard, dolandırıcılıktan 1 yılda 700 milyon USD fazla zarar ederken Capital One tarafından geliştirilen yapay sinir ağı tabanlı kredi kartı dolandırıcılık tespit sistemi, şirketlerin kayıplarını %50'den fazla azaltmayı başarmıştır (Kantardzic, 2020:590).

Liou (2008:650-662) çalışmasında, hileli finansal raporlama ile iş hatası tahmini arasındaki farkları, benzerlikleri ve ikili arasındaki etkili faktörleri araştırmıştır. Taiwan Economic Journal veri bankasından ve Taiwan Stock Exchange Corporation'dan elde ettiği verilerden 52 finansal göstergeyi önemli etken olarak tanımlamıştır. Çalışmasında LR, ANN ve DT algoritmalarını kullanmıştır. Oluşturulan modeller hem hileli finansal raporlamayı tespit etmede hem de iş başarısızlıklarını tahmin etmede başarılı olmuştur. Genel doğruluk açısından DT, ortalama %98'lik başarı göstererek diğer iki algoritmadan daha iyi performans sergilemiştir.

Kirkos ve vd. (2007:1002) çalışmalarında denetim ve muhasebe alanına yardımcı olmak amacıyla DT, ANN ve BNN'yi kullanarak mali tablo dolandırıcılığını tespit etmeye çalışmışlardır. Deneylerden elde edilen sonuçlara göre, 76 adet Yunan imalat şirketine ait mali tablo verilerinin tahrif edildiğine dair göstergeleri olan veriler üzerinden yapılan araştırma hileli olan kesin sonuçlarla örtüşmüştür. Modellerin performanslarına bakıldığında, 10 katlı çapraz doğrulama yöntemi kullanılan analizlerde Bayes inanç ağı modeli %90,3'ünü doğru bir şekilde sınıflandırmayla en iyi performansı elde etmiştir. Yapay Sinir Ağı modelinin ve Karar Ağacı modelinin doğruluk oranları sırasıyla %80 ve %73,6'dır.

Kırlıoğlu ve Ceyhan (2014:13) çalışmalarında, kaliteli bir denetimin gerçekleştirilebilmesi adına araştırmacının kısa sürede şirket hakkında genel yargıya varılması gerektiğini savunmuşlardır. Çalışmalarında denetçilerin faydalandığı finansal tablo bilgilerini veri madenciliği algoritmalarıyla analiz ederek finansal olarak başarılı ya da başarısız olarak sınıflandırılmayı hedeflemişlerdir. Bu sayede denetçilerin kısa sürede işletmeler hakkında analitik bilgi elde edebileceğini ispatlamıştır. Çalışmasında k-en yakın komşu algoritmasıyla birlikte 10 katlı çapraz doğrulama tekniğini kullanmış ve oluşturduğu model %95'lik doğru finansal sınıflama tahmini sağlamıştır.

Dutta ve vd. (2017:374-393) yapmış oldukları çalışmada, hileli faaliyetlerde bulunduğu tespit edilen şirketlerin yeniden yaptıkları beyanları incelemişlerdir. 2001-2014 yılları arasındaki 3.513 vakayı içeren gerçek kapsamlı bir veri seti kullanarak ANN, DT, BBN ve destek vektör makinesi (SVM) yöntemleriyle modeller geliştirmişlerdir. Modellerin performansını değerlendirme aşamasında doğruluk, duyarlılık, kesinlik, özgüllük ve ROC eğrisi (AUC) ölçütlerini kullanmışlardır. 10 katlı çapraz doğrulamayla yapılan analiz sonucu ANN modellerinin yaklaşık %83'lük değeriyle diğer yöntemlere göre daha yüksek bir performans gösterdiğini bulmuşlardır.

Lakshmi & Kavila (2018:16823) çalışmalarında kredi kartlarında yapılan dolandırıcılığı tespit etmek için LR, DT ve Rastgele Orman (RF) gibi makine öğrenme yöntemlerini kullanmışlardır. Yöntemlerin performansını değerlendirmek için duyarlılık, özgüllük, doğruluk ve hata oranı metrikleri kullanılmıştır. Lojistik regresyon %90, karar ağacı %94,3 ve rastgele orman sınıflandırıcısı %95,5 olmak üzere performans sergilemiştir. Her üç yöntemi de karşılaştırdıklarında, rastgele orman sınıflandırıcısının LR ve DT'den daha iyi olduğunu bulmuşlardır.

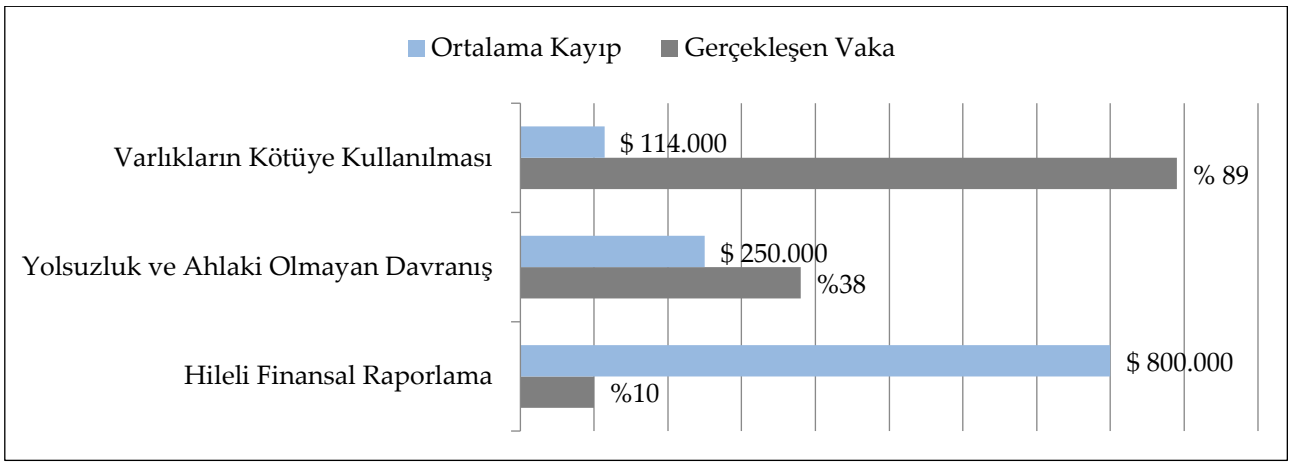
Aksoy (2021:29) çalışmasında Borsa İstanbul'da 2000-2019 yılları arasında işlem gören 88 şirketin, finansal tablolarına dayalı model geliştirerek hile faaliyetleri 1 yıl önceden tespit etmeyi amaçlamıştır. Çalışmasında veri madenciliğinin bir kolu olan makine öğrenmesi yöntemlerinden ANN, sınıflandırıcı regresyon ağacı (CART), SVM ile LR yöntemlerini kullanarak mali tablo dolandırıcılığını tahmin etmeye çalışmıştır. Yapılan çalışmanın sonucunda ANN %96,15, CART %96,15, SVM %80,77 ve LR %80,77 oranında başarı göstermiştir.

Chen (2016:15) çalışmasında, 2002 ve 2013 yılları arasında hem hileli hem de hileli olmayan mali tablolara sahip şirketlerin verilerini kullanarak bir mali tablo tespit modeli oluşturmayı amaçlamıştır. Çalışmasının ilk aşamasında CART ve ki-kare etkileşim detektörünü (CHAID) kullanarak ana değişkenleri belirlemiştir. İkinci aşamasında hile yapılmış tabloları belirlemek için CART, CHAID, BBN, SVM ve ANN tekniklerini birleştirilerek model oluşturmuştur. Araştırma sonuçlarına göre, CHAID–CART modelinin algılama performansı %87,97 genel doğrulukla en etkili olan ikili olarak belirlenmiştir. Jans vd. (2007:19-24) çalışmalarında veri madenciliği tekniklerinin gerçek dünya varsayımları altında dolandırıcılık tespiti için bir araç olarak verimli olup olmadığını araştırmışlardır. Çalışmalarında satın alma siparişlerine, mal girişlerine ve faturalara ilişkin bir dizi değerleri (sipariş adedi, satın alan, satıcı, fiyat, tarih gibi) kullanmışlardır. Daha sonra veri madenciliğinin denetimli öğrenme ve denetimsiz öğrenme tekniklerini kullanarak aykırı değerlere sahip bir küme tespiti yapmışlardır. Tespit edilen bu aykırı değerlerin normalde küçük olması gerekirken, ortalamaların yüksek olması tespitte önemli bir etken olarak bulunmuştur. Böylelikle işlemelerin bu aykırı değerlere yönelik kontrol noktalarının oluşturabileceğini ve dolandırıcılığın önlenebileceğini ifade etmişlerdir.

Bu çalışmanın literatürdeki diğer çalışmalardan farkı; 1) Yürütülecek hile tespit çalışmasında kaynak veri, bir bankanın bir dönemine ait işlemlerinden üretilen kayıtlar üzerinde yapılmasıdır. 2) 594.643 adet satır ve 9 adet öznitelik sütunun bulunduğu büyük sayılabilecek bir veri seti üzerinde gerçekleştirilmesidir. 3) Makine öğrenmesi yöntemleriyle hassas bir tahmin yapılabilmesi için veri seti %69'u eğitime, %31'i teste ayrılmış olmasıdır. 4) Hile tespit çalışmalarında zaman önemli olduğu için makine öğrenmesi yöntemlerinin eğitim ve tahmin için harcadığı süreleri değerlendirmede bir kriter olarak ele almasıdır. 5) Son olarak, literatürdeki makine öğrenimine dayalı hileli finansal raporlama tespit çalışmalarının aksine varlıkların kötüye kullanılması amacıyla yapılan hileli ödemeleri tespit etmeye yönelik olmasıdır.

## 2.1.Varlıkların Kötüye Kullanılması

Günümüze kadar hile ile ilgili pek çok çalışma yapıldığı bilinmektedir. Hile üzerine yapılan çalışmalara bakıldığında hilekârlığın üç ana kategoriye ayrıldığı konusunda uzmanlar birleşmiştir. Buna göre işletmelerde karşılaşılan hile türleri; a)varlıkların kötüye kullanımı, b)yolsuzluk ve c)hileli finansal raporlama şeklinde sınıflandırılmıştır. (Padgett, 2015:23). Çalışmaya uygun olması nedeniyle bu hile sınıflandırmasına göre ilerlenmiştir. ACFE'nin son raporundaki hileli işlemlere ait gerçekleşme sıklıkları ve ortalama kayıp tutarları Şekil.1'de gösterilmiştir.

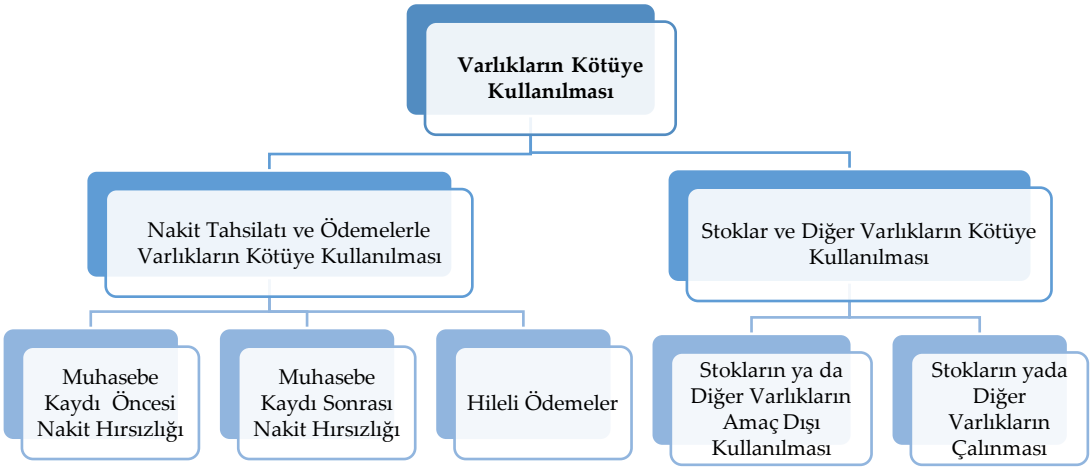



**Şekil 1:** Hileli Faaliyet Kategorileri ile Ortalama Kayıp (ACFE:2018,10)

İşletmelerde, hile riskleri değerlendirilirken ve hile tespit kontrolleri tasarlanırken hilekârların planları doğrultusunda ortaya çıkan her türlü fırsatı değerlendirebileceği unutulmamalıdır. Çünkü ACFE'nin raporuna göre 2.690 vakada, birçok hileli faaliyetin birden fazla mesleki hile türünü içinde barındırabileceği vurgulanmaktadır (ACFE, 2018:12). Varlıkların kötüye kullanılması, nakit veya envanter hırsızlığı, hasılatı eksik gösterme ve zimmete para geçirme gibi türlerini içerir. Varlıkların kötüye kullanımında en sık karşılaşılan fatura, bordro, gider ve çek gibi hileli ödemelerdir. Bazen çalışanlar, işletmeye mal satan satıcılarla ya da rekabet ettiği şirketlerle iş birliğine girerek işletmeyi zarara uğratabilmektedirler. Bununla birlikte işletmenin varlıklılarının kendi çıkarları için kullanılması, yanlış raporlama ve kendi adına yaptığı harcamalar da bu tür hileli faaliyetin içinde değerlendirilmektedir (Golden vd., 2011:5).

Varlıkların kötüye kullanılması, faturaların değiştirilmesi, malların çalınması veya alınmamış mal veya hizmet bedellerinin ödemesine neden olmak gibi çeşitli yollarla gerçekleştirilmektedir. Bu hile türü, muhtemel kontrolleri atlatarak yaratılan yanlış yahut yanıltıcı kayıtlar veya belgeler eşliğinde yapılabilmektedir (SAS[2], 2002:1721). Varlıkların kötüye kullanılması, hırsızlık veya zimmete geçirmekten daha fazlasını içermektedir. Buna herhangi bir şirket varlığının kişisel kazanç için kötüye kullanılması da dâhil edilebilir. Örneğin; çalışanların şirket bilgisayarını saatlerce kendi çıkarları için kullanması da bu kapsamda değerlendirilebilir (Wells, 2017:46).

Hiledeki üç temel kategoriden, varlıkların kötüye kullanımı en yaygın olanıdır, ACFE raporuna göre çalışmada yer alan 2.690 vakanın %89'unda meydana gelmiştir (ACFE, 2018:10). Çalışmada ana konudan ayrılmamak için aşağıdaki Şekil 2'de gösterilen nakit tahsilatı ve ödemelerle varlıkların kötüye kullanılmasındaki alt başlık olan hileli ödemeler türü ele alınacaktır.



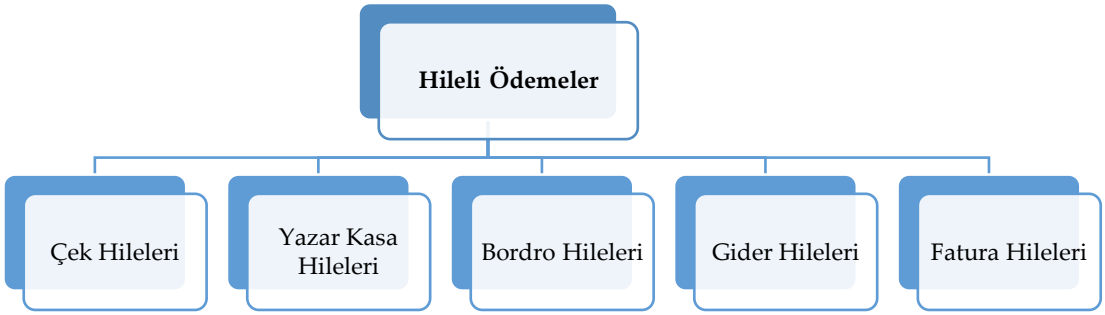

**Şekil 2:** Varlıkların Kötüye Kullanılma Türleri (ACFE:2018,11)

[2] Statement on Auditing Standards, No. 99.

### *2.1.1. Hileli Ödemeler*

Hileli ödemeler, işletme varlıklarının çalışanlar tarafından kötü amaçlı kullanımı sıkça karşılaşılan hile türlerindendir. Bu hile türü, bir işletme çalışanının görevini kötüye kullanarak, ödemeleri yasal olmayan kanallara yönlendirmesiyle oluşabilmektedir. Hileli ödemeler kendi içinde 5 ana başlık altında toplanmıştır. Bunlar; 1)Fatura hilesi, 2)Bordro hilesi, 3)Gider hilesi, 4) Çek hilesi, 5)Yazar kasa hilesi (Bozkurt, 2009:212,215).

Çek hileleri, bir çalışanın kuruluşun hesabına düzenlenmiş bir çeki kendi çıkarları için kullanması veya kuruluş tarafından üçüncü bir şahsa düzenlenmiş gibi göstererek el koyması yoluyla meydana gelir. Yazar kasada dolandırıcılığında ise, paranın kasadan gizlice çıkarılması vardır. Bordro hilesi genellikle sahte zaman çizelgesiyle işletmenin çalışanlardan birine borcu varmış gibi gösterilerek yapılan bir türdür. Gider hilesi, şirketin çalışanları tarafından yanlış gider raporlamasına dayandırılarak yapılan bir dolandırıcılıktır (Wells, 2017:110,145,188). Fatura hileleri, oldukça sık karşılaşılan ve işletmeye oldukça fazla zarar veren bir hileli ödeme türüdür. İşletmenin özellikle satın alma işlevine karşı yapılan operasyonlarda ortaya çıkmaktadır. Bu hile türü, işletmeye gelen nakitlere dokunulmadan bir nedenle işletme dışına çıkarılmasına dayanmaktadır (Bozkurt, 2009:215).




**Şekil 3:** Hileli Ödeme Hile Türleri(ACFE:2018,11)

Çalışmada kullanılan kaynak veri bir bankanın üye işyerleri aracılığıyla yapılan işlemlere dayalı kayıtlar oluşturmaktadır. Çalışmanın özelinde üye işyerlerini bankanın çalışanları olduğu düşünülürse, işletme çalışanları tarafından işletme kaynaklarının kötü amaçlı kullanımı sıkça rastlanılan "**hileli ödemeler**" türüne benzetilebilir. Daha önce de belirtildiği üzere, bu hile türü bir işletme çalışanının işletmedeki konumunu kullanarak, ödemeleri yasal olmayan kanallara yönlendirmesi biçiminde gerçekleştirilmektedir. Böylelikle işletmelere çoğunlukla var olmayan ya da olduğundan yüksek tutarlı ödemeler yaptırılarak zarara uğratılmıştır.

## 3.VERİ VE YÖNTEM

Jans vd. (2007:19), hilenin tespitinde faturayı düzenleyenin ve ödeyenin kim olduğu, faturanın ödenme şekli ve ödenme tarihi, toplam fatura sayısı, fatura açıklamaları ve ortalama tutar gibi bilgilerin kullanılması gerektiğini belirtmiştir. Hileli işlemlerle ile işletmenin finansal olmayan bilgileri arasında bir korelasyon bulunabilmektedir. Bu nedenle finansal tablo hilelerinin tespitinde finansal olmayan bilgileri de dikkate almak gerekir (Jan, 2018:2). Bununla hile tespit çalışmalarının büyük problemlerinin en önemlisi mevcut hile kayıtlarının olmayışıdır. Birçok araştırmacı, gerçek hile kayıtları olmadan, yöntemlerini test etmek için yapay olarak oluşturulmuş hile kayıtlarını kullanmaktadır. Test amaçlı üretilen yapay kayıtlar, gerçek hileli kayıtları gerçekten temsil ettikleri ölçüde faydalıdır (Lu vd., 2006:49). Bu alanda yapılan çalışmalar, hileli mali tabloları tespit etmek için veri madenciliği tekniklerinin kullanılmasının, doğruluk açısından geleneksel analiz tekniklerini benimsemekten daha üstün olduğuna işaret etmektedir (Chen,2016:4). Kirkos ve vd. (2007:998) DT, ANN ve BNN'yi, Aksoy (2021:38) ANN, CART, SVM ve LR'yi, Liou (2008:653), LR, ANN ve DT'yi, Dutta ve vd. (2017:374-393) ANN, DT, BBN ve destek vektör makinesi (SVM) yöntemini belirlemiş, performans ölçütü olarak doğruluk, duyarlılık, kesinlik, özgüllük ve AUC eğrisi ölçütlerini kullanmışlardır. Çalışmada, en çok tercih edilen ve başarılı uygulamalarda kullanılan karar ağaçları (DT), destek vektör makinesi (SVM), lojistik regresyon (LR) ve yapay sinir ağları (ANN) yöntemi kullanılmıştır.

### 3.1.Veri

Çalışmada kullanılan veri, ödemelerde kullanılan bir bankanın bir dönemine ait işlemlerini yansıtmaktadır. Normal ödemeler ile hileli ödemelerin yer aldığı veri seti, gerçeği yansıtacak şekilde hile araştırmaları için yapay olarak üretilmiştir. Bu veri seti, gerçeği yansıtan herhangi bir kişisel veya işlem bilgisi içermemektedir. Veri seti 594.643 adet satır, 9 adet değişken/öznitelik ve 1 adet sınıflandırıcı sütundan (hilesiz = 0, hileli=1) oluşmaktadır. Veri setindeki 587.443 tanesi normal işlemken, 7,200 tanesi hileli işlemdir (Kaggle, 2017). Veri setindeki dolandırıcılık işlemleri tüm işlemlerin yaklaşık %01,21'sini oluşturmaktadır. Çalışmada, yapılan işlemlere ilişkin finansal ve finansal olmayan veriler kullanılarak hileli kayıtların makine öğrenimi yöntemleri tarafından tespit edilmesi sağlanmıştır.

**Tablo.1 Veri Setindeki Değişkenler**

| Değişkenin Adı | Veri Tipi | Tanım |
|---|---|---|
| **STEP (İşlem Adımı)** | Integer | Yapılan İşlem Adımı |
| **CUSTOMER (Müşteri Numarası)** | String | Müşteri Numarası |
| **AGE (Yaş)** | Integer | Müşteri Yaş Grubu |
| **GENDER (Cinsiyet)** | String | Müşteri Cinsiyeti |
| **ZIP_CODE_ORI (Müşteri Posta Kodu)** | Integer | Müşteri Posta Kodu |
| **MERCHANT (Üye İş Yeri Numarası)** | String | Satıcı (Üye İşyeri) Numarası |
| **ZIP_MERCHANT (Üye İş Yeri Posta Kodu)** | Integer | Satıcı (Üye İşyeri) Posta Kodu |
| **CATEGORY (Harcama Kategorisi)** | String | Harcama Alanı |
| **AMOUNT (Harcama Alanı)** | Float | Harcama Tutarı |
| **FRAUD (İşlem Sınıfı)** | Integer | Sınıf Etiketi |

Teoride veri setlerinin doğru olduğu varsayılsa da veritabanlarında bulunan ham verilerin çoğu önceden işlenmemiş, gürültülü ya da tutarsız olabilmektedir. Verilerin kalitesi, eksik değerlerden, aykırı değerlerden, eskimiş veya değersizleşmiş verilerden kaynaklı olarak bozulabilmektedir. Veri madenciliğinin amaçları doğrultusunda faydalı olabilmesi için, veri tabanlarındaki verilerin ayıklanması ve temizlenmesi ya da birden çok kaynaktan gelen verilerdeki birleştirmeden dolayı oluşan tutarsızlıkların kaldırılması için bir ön işleme tabi tutulması gerekmektedir. Buradaki temel hedef ise, veri setindeki geçerliliğini kaybetmiş (çöp) verilerin varlığını asgari düzeye indirmektir (Larose, 2005:27-28). Veri seti içinde az da olsa değişkenlerin bazıları hatalı olarak adlandırılan uygun olmayan değerler bulunduğu anlaşılmıştır. Ön işlem süreci için veri setindeki eksiklikler Microsoft Excel programı kullanılarak tespit edilmiştir. Özellikle veri setindeki tanımlama ya da belirtme amacıyla kullanılan karakterlerden temizlenmesi ve gerekli dönüşüm işlemleri yapılmalıdır. Bu işlemler, verileri manipüle etmek için değil, seçilen sınıflandırıcı algoritmalara gereken değerlerin oluşması için yapılmalıdır.

### 3.2. Yöntem

Sınıflandırma algoritmaları ilk olarak, önceden sınıflandırılmış değerleri içeren bir eğitim kümesi ile eğitilir. Eğitim veri setinden sağlanan veri kullanılarak geçici bir veri madenciliği modeli oluşturulur. Sonra modelin bir test kümesi üzerinde nasıl performans gösterdiği incelenmektedir. Test veri kümesinde, hedef değişkenin değerleri geçici olarak geçici modelden gizlenmekte, daha sonra eğitim setinden öğrendiği model ve yapıya göre sınıflandırma yapması istenmektedir. Sınıflandırmaların etkinliği daha sonra hedef değişkenin gerçek değerleriyle karşılaştırılarak değerlendirilmektedir (Larose, 2005:91-92). Çalışmada, kaynak veri %69 ile %31 şeklinde ikiye ayrılmıştır. Algoritmalar, önce %69'lik kısım üzerinden eğitilecek, %31'lik kısım üzerinden test edilecektir.

Dutta vd. (2017: 374-393) çalışmasında yaptığı gibi, modellerin performans ölçüsünü belirlemek adına doğruluk, duyarlılık, kesinlik ve f1-skor metrikleriyle belirlenecektir. Sınıflandırıcının doğruluk ölçütü, toplamda doğru sınıflandırılmış pozitiflerin ve negatiflerin, toplam örnek sayısına bölünmesiyle; sınıflandırıcının kesinlik ölçütü, doğru sınıflandırılmış pozitif örneklerin, toplam pozitif tahmin edilmiş örneklere bölünmesiyle; sınıflandırıcının duyarlılık ölçütü, doğru sınıflandırılmış pozitiflerin, toplam gerçek pozitif sınıfa bölünmesiyle; sınıflandırıcının f1-skoru ölçütü, kesinlik ve duyarlılık değerlerinin harmonik

ortalaması alınarak hesaplanmaktadır(Memiş vd., 2019:3). Ayrıca çalışmada modellerin oluşturulma ve tahmin yapma süreleri de karşılaştırmada başarı kriteri olarak kullanılacaktır.

Günümüzde kullanılan tüm veri madenciliği uygulamaları çeşitli yazılımlar ile yapılmaktadır. Bu araçlardan en bilinenleri ve açık kaynak kodlu olanları DataLab, DBMiner, Knime, RapidMiner, Weka, R ve Orange sıklıkla kullanılmaktadır (Kantardzic, 2020:574-576). Ancak uygulayıcılara programlama açısından çok fazla esneklik sağlaması ve farklı görevleri bir araya getirmek için çok sayıda modüle sahip olması nedeniyle veri madenciliğinde Python programlama dilinin kullanımı da hızla artmaktadır. Python, programlama dilinin kolaylığı sayesinde veri madenciliği araçlarını daha geniş alanda aktif olarak kullanma imkânı vermektedir (Layton, 2015:1). Çalışmada verinin ön hazırlığıyla ve model oluşturma işlemlerinin bir arada yapılması için Python programlama diline sahip bir uygulama sistemi tasarlanmıştır. Bu uygulama sistemi Google Colab üzerinde çalıştırılarak daha hızlı işlem yapma imkânı elde edilmiştir. Kaynak verinin gerekli temizleme ve dönüştürme işlemleri yapıldıktan sonra sınıflandırma Weka programı ile de yapılabilmektedir.

### *3.2.1. Kullanılan Makine Öğrenmesi Yöntemleri*

**Karar Ağaçları:** Sınıflandırma ağaçları olarak da bilinen karar ağaçları en sezgisel ve en sık kullanılan veri madenciliği tekniklerinden biridir. Bir analist açısından, kullanımı ve yorumlanması kolaydır. Sınıflandırma ağaçları, adından da anlaşılacağı gibi, bir veri kümesini yanıt değişkenine ait sınıflara ayırmak için kullanılır. Genellikle yanıt değişkeninin iki sınıfı vardır: Evet veya Hayır, 1 veya 0 (Kotu ve Deshpande, 2019:66). Bir hilekarlığın tespitinde karar ağacı algoritmaları tanımlama yapmak için kullanılabilir. Basit bir örnekle, üst düğümde (kökte) bir işlem miktarına sınırlama getirerek, test koşulunu sağlayan ve diğer öznitelikleri (hilekarlık geçmişi, iş durumu) göz önüne alarak alt düğümler oluşturulabilir. Son olarak, ağacın uç kısımlarında "hile" veya "hile(li) değil" sınıfına atama yapılabilir (Baesens vd., 2015:1).

**Destek Vektör Makineleri:** Hem doğrusal olarak hem de doğrusal olmayan verileri sınıflandırmak için benzersiz bir yaklaşım sağlayan bir veri madenciliği algoritmasıdır. Diğer tekniklerin aksine sınıflandırması zor örneklere bile optimal çözüm önerileri sunabilmektedir. Günümüzde hem sınıflandırma hem de sayısal tahmin için kullanılsa da, daha çok ikili sınıflandırma modelleri tercih edilmektedir (Roiger, 2017:324-327). Doğrusal sınıflandırıcı denklem açısından önemli olan ayarlamalar, sahtekarlığa ilişkin oluşturulacak bir modelde kullanılabilirliğini arttırmaktadır. Oluşturulmak istenen model, hilenin varlığına $\mathbf{A_1}$ ya da hilenin yokluğuna $\mathbf{A_2}$ ilişkin sınıflandırılma işleminde, değişkenlerin $\mathbf{x_1, x_2,..}$ hileli işlem tutarının $\mathbf{w_1}$ ve hileli işlem sayısı $\mathbf{w_2}$ varsayılırsa, bu durumlarda birden fazla kıstas ile çarpılarak yapılabilmektedir. Ayrıca değişkenler arasında bir önemlilik seviyesi farkı varsa bu durum hileli işlem tutarı büyüktür hileli işlem sayısı $\mathbf{w_1>w_2}$ şeklinde ayarlanabilir. Oluşturulan modelin denklemine hata payı $\boldsymbol{b}$ eklenerek daha da hassaslaştırılabilir (Baesens vd., 2015:155-156).

**Lojistik Regresyon:** Sınıf etiketinin (yanıt değişkeni) kategorik olduğu bir regresyon analizi şeklidir. İlgilenilen bir olayın meydana gelme olasılığını tahmin etmek için yaygın olarak kullanılan veri madenciliği algoritmasıdır. Lojistik regresyon, dolandırıcılık, kötü kredi durumu, kayıp, satın alma eğilimi ve diğer birçok ikili hedef sonucunu tahmin etmek için kullanılabilir (Kudyba, 2014:8283). Lojistik regresyon modellerinin birden fazla tahmin değişkenini barındırması hile denetiminde başarılı performans sergilemesine yardımcı olmaktadır. Hedef değişken **Y**'nin tahmininin yapılabilmesi için $\mathbf{X_1}$, $\mathbf{X_2}$, $\mathbf{X_3}$**….**$\mathbf{X_n}$ öngörü sağlayan değişkenlerinin kullanılması gerektirir (Kantardzic,2020:178). Örneğin, araba sigortasındaki hile tespitine yönelik bir model oluşturmak istenirse, hile tutarına (**Y**) tespit etmek için şüphelinin yaşı ($\mathbf{X_1}$), talep edilen tutar ($\mathbf{X_2}$), kazanın ciddiyeti ($\mathbf{X_3}$) ve diğer öngörü sağlayan değişkenler ($\mathbf{X_n}$) eklenerek aşağıdaki gibi bir denklem oluşturulabilir (Baesens vd., 2015:125).

**Yapay Sinir Ağları:** Yapay bir ağ oluşturarak en iyi tahmini elde etmek için tekrarlanan denemelerde öğrenerek yeniden modellemeler yapan bir veri madenciliği aracıdır. Ağ, belirli bir sorunu çözmek için paralel olarak çalışan çok sayıda birbirine yüksek düzeyde bağlı işlem öğesinden (nöron) oluşmaktadır. Yapay sinir ağı kullanılarak oluşturulan modeller, büyük verileri ve birçok tahmin değişkeni ile karmaşık etkileşimleri çözümlemede kullanılmaktadır (Fernandez, 2003: 7:4) Yapay sinir ağlarının kendi kendine öğrenme yapısının oluşu gizli katmanlarında nasıl bir reaksiyon gösterdiği tam olarak da bilinmemektedir. Ancak yapay sinir ağlarını verilerdeki çok karmaşık kalıpları ve karar sınırlarını modelleyebilmesi onu güçlü kılmaktadır. Yapay sinir ağları, sahtekarlık olup olmadığını ve sahtekarlık miktarını aynı anda tahmin etmek için programlanabilir (Baesens vd., 2015:149).

## 4.BULGULAR

### 4.1.Karar Ağacı

Karar ağacının yaptığı tahminlere ilişkin istatistik sonuçları incelendiğinde, modelin doğruluk metriği %99,42, f1-skoru %75, kesinlik metriği %74 ve duyarlılık metriği %76 hesaplanmıştır. 1.620 tane hileli işlemi doğru tahmin etmiştir. Karar ağacı algoritması, eğitimini ve tahminini 1,32 saniyede tamamlamıştır.

**Tablo.2 Karar Ağacının Model İstatistikleri**

| Karar Ağacı | | | | | |
|---|---|---|---|---|---|
| **Doğruluk** | **:** | 0.994 | **Karışıklık Matrisi** | | |
| **f1-skoru** | **:** | 0.754 | | **Gerçek Pozitifler** | **Gerçek Negatifler** |
| **Kesinlik** | **:** | 0.748 | **Tah. Ed. Pozitifler** | 181.353 | 552 |
| **Duyarlılık** | **:** | 0.762 | **Tah. Ed. Negatifler** | 505 | 1.620 |
| **Süre** | **:** | 1,32 Sn. | | | |

### 4.2. Destek Vektör Makinesi

Destek vektör makinesinin yaptığı tahminlere ilişkin istatistik sonuçları incelendiğinde, modelin doğruluk metriği %99,11, f1-skoru %39, kesinlik metriği %95 ve duyarlılık metriği %24 hesaplanmıştır. 522 tane hileli işlemi doğru tahmin etmiştir. Destek vektör makinesi algoritması, eğitimini ve tahminini 22,14 saniyede tamamlamıştır.

**Tablo.3 Destek Vektör Makinesinin Model İstatistikleri**

| Destek Vektör Makinesi | | | | | |
|---|---|---|---|---|---|
| **Doğruluk** | **:** | 0.991 | **Karışıklık Matrisi** | | |
| **f1-skoru** | **:** | 0.390 | | **Gerçek Pozitifler** | **Gerçek Negatifler** |
| **Kesinlik** | **:** | 0.950 | **Tah. Ed. Pozitifler** | 181.878 | 27 |
| **Duyarlılık** | **:** | 0.245 | **Tah. Ed. Negatifler** | 1.603 | 522 |
| **Süre** | **:** | 22,14 Sn | | | |

### 4.3. Lojistik Regresyon

Lojistik regresyonun yaptığı tahminlere ilişkin istatistik sonuçları incelendiğinde, modelin doğruluk metriği %98,95, f1-skoru %48, kesinlik metriği %55 ve duyarlılık metriği %43 hesaplanmıştır. 920 tane hileli işlemi doğru tahmin etmiştir. Lojistik regresyon algoritması, eğitimini ve tahminini 9,51 saniyede tamamlamıştır.

**Tablo.4 Lojistik Regresyonun Model İstatistikleri**

| Lojistik Regresyon | | | | | |
|---|---|---|---|---|---|
| **Doğruluk** | **:** | 0.989 | **Karışıklık Matrisi** | | |
| **f1-skoru** | **:** | 0.487 | | **Gerçek Pozitifler** | **Gerçek Negatifler** |
| **Kesinlik** | **:** | 0.558 | **Tah. Ed. Pozitifler** | 181.179 | 726 |
| **Duyarlılık** | **:** | 0.432 | **Tah. Ed. Negatifler** | 1.205 | 920 |
| **Süre** | **:** | 9,51 Sn. | | | |

### 4.4. Yapay Sinir Ağları

Yapay sinir ağlarının yaptığı tahminlere ilişkin istatistik sonuçları incelendiğinde, modelin doğruluk metriği %99,35, f1-skoru %62, kesinlik metriği %93 ve duyarlılık metriği %47 hesaplanmıştır. 1.000 tane hileli işlemi doğru tahmin etmiştir. Yapay sinir ağı algoritması, eğitimini ve tahminini 51,44 saniyede tamamlamıştır.

**Tablo.5 Yapay Sinir Ağlarının Model İstatistikleri**

| Yapay Sinir Ağları | | | | |
|---|---|---|---|---|
| **Doğruluk** | : 0.993 | **Karışıklık Matrisi** | | |
| **f1-skoru** | : 0.626 | | **Gerçek Pozitifler** | **Gerçek Negatifler** |
| **Kesinlik** | : 0.936 | **Tah. Ed. Pozitifler** | 181.837 | 68 |
| **Duyarlılık** | : 0.470 | **Tah. Ed. Negatifler** | 1.125 | 1.000 |
| **Süre** | : 51,44 Sn | | | |

### 4.5 Yöntemlerin Karşılaştırılması

Yöntemler sırasıyla çalıştırılmış ve modeller oluşturulmuştur. Modellerin başarısı karşılaştırılırken üretmiş olduğu doğruluk, f1-skoru, duyarlılık, kesinlik metriklerinin yanı sıra eğitim ve tahmin yapma süresi de dikkate alınmıştır. Ancak modellerin istatistikleri, uygulamanın yapıldığı bilgisayarın donanımına ve internet hızına göre değişkenlik gösterebileceği unutulmamalıdır. Modellerin istatistikleri karşılaştırılmalı olarak gösterilmiştir.

**Tablo.6 Modellerin İstatistikleri**

| | **Doğruluk (Accuracy)** | **f1-skoru (f1-score)** | **Kesinlik (Precision)** | **Duyarlılık (Recall)** | **Süre (Duration)** |
|---|---|---|---|---|---|
| **Karar Ağacı** | %99,42 | %75 | %74 | %76 | 1,32 sn |
| **Destek Vektör Makinesi** | %99,11 | %39 | %95 | %24 | 22,14 sn |
| **Lojistik Regresyon** | %98,95 | %48 | %55 | %43 | 9,51 sn |
| **Yapay Sinir Ağı** | %99,35 | %62 | %93 | %47 | 51,44 sn |

Literatür incelemesi yer verilen Lægreid (2007), Kirkos ve vd. (2007), Liou (2008), Kırlıoğlu ve Ceyhan (2014), Chen (2016), Dutta ve vd. (2017), Aksoy (2021) çalışmalarla karşılaştırıldığında metriklerinin bazıları farklılık gösterse de genel itibari başarılı model istatistiği elde edilmiştir. Diğer yandan Tablo.6'da gösterildiği üzere, hilenin ortaya çıkarılmasındaki zamanın önemini vurgulamak adına kriter olarak belirlenen makine öğrenimi algoritmalarının öğrenme ve tahmin süreleri birbirlerinden farklı çıkmıştır. Kurulan modeller içerisinde de karar ağacının, destek vektör makinesi, lojistik regresyon ve yapay sinir ağlarına göre daha üstün bir performans sergilemiştir.

## 5. SONUÇ VE TARTIŞMA

Bu çalışmada şirketlerde sıkılıkla karşılaşılan hileli ödemelerin tespitine yönelik etkili bir model oluşturmak amaçlanmıştır. Çalışmada bir bankanın verilerine dayanılarak üretilen 594.643 adetlik veri seti kullanılmıştır. Veri setinde, hileli ödemeleri tespit etmek için yapılan işlemlerin adım sayısı, müşteri yaş grubu, müşteri cinsiyeti, müşteri posta kodu, satıcı (üye işyeri) numarası, satıcı (üye işyeri) posta kodu, harcamaların yapıldığı sektörler ve harcama tutarı gibi değişkenler kullanılmıştır. Böylelikle hem finansal hem finansal olmayan verileri bir arada kullanarak tespit yapılabileceği gösterilmiştir. Tahmin için kurulan modeller önce veri setinin %69'u ile eğitilmiş sonra %31'i ile test edilmiştir.

Yöntemlerin oluşturduğu değerlere bakıldığında doğruluk metriği sırasıyla karar ağacında %99,42, destek vektör makinesinde %99,11, lojistik regresyonda %98,95 ve yapay sinir ağlarında %99,35 hesaplandığı görülmüştür. Doğruluk metriğinde rakamlar birbirine yakın olsa da en anlamlı (%99,42) değeri karar ağacı algoritması üretmiştir. F1-skor metriğine bakıldığında karar ağacında %75, destek vektör makinesinde %39, lojistik regresyonda %48 ve yapay sinir ağlarında %62 hesaplandığı görülmüştür. F1-skor metriğinde en anlamlı değeri (%75) karar ağacı değeri üretmiştir. Kesinlik metriğine bakıldığında karar ağacında %74 , destek vektör makinesinde %95, lojistik regresyonda %55 ve yapay sinir ağlarında %93 hesaplandığı görülmüştür. Kesinlik metriğinde en anlamlı değeri (%95) destek vektör makinesi üretmiştir. Duyarlılık metriğine bakıldığında karar ağacında %76, destek vektör makinesinde %24, lojistik regresyonda %43 ve yapay sinir ağlarında %47 hesaplandığı görülmüştür. Duyarlılık metriğinde en anlamlı değeri (%76) karar ağacı üretmiştir. Yöntemlerin model ve tahmin için harcadığı sürelere bakıldığında, karar ağacı 1,32 saniyede, destek vektör makinesi 22,14 saniyede, lojistik regresyon 9,51 saniyede ve yapay sinir ağları 51,44 saniyede işlemlerini tamamlamıştır. Hileli kayıtları doğru tahmin etme sayılarına bakıldığında, karar ağacı 1.620 tane, destek vektör makinesi 522 tane, lojistik regresyon 920 tane ve yapay sinir ağları 1.000 tane hile kaydı doğru tahmin etmiştir. Yöntemlerin oluşturduğu istatistikler incelendiğinde, ortalama olarak en başarılı, en hızlı ve hileli kaydı en fazla doğru tahmin eden modeli karar ağacının oluşturduğu tespit edilmiştir.

Çalışma özünde, hileli ödemelerin tespitinde makine öğrenimi yöntemlerinin kullanılmasıyla alanında öncü çalışmalardan biridir. Çünkü literatür incelemesinde yer verilen pek çok çalışma hileli finansal raporlama üzerine kurulmuştur. Ayrıca bilindiği üzere, hilenin ortaya çıkarılma süresi uzadığında işletmelere verdiği zarar artmaktadır. Bu sebeple, makine öğrenimi yöntemlerinin eğitim ve test süresi de önem kazanmaktadır. Bu yönüyle hilenin hızlı ve doğru tespit edilmesinin vurgulandığı bir çalışma olmuştur.

Çalışmada kullanılan kaynak veri seti yapay olması sebebiyle uygulamanın gerçek bir veri seti üzerinde test edilme ihtiyacı bulunmaktadır. Diğer yandan makine öğrenimi yöntemlerinde yaşanabilecek gelişmeyle daha iyi sonuçlar elde edilebilir. Uygulamada kullanılan yöntemlere bağlı olarak veri setiyle yapılan model denemelerinde makine yanlış pozitif ve yanlış negatif değerleri üretmiştir. Her ne kadar bu değerleri sıfıra indirmek mümkün olmasa da azaltılması yönünde geliştirilmeye açıktır. Bu nedenle makine öğreniminin hata sayısını düşürmeye yönelik teknikler ve yaklaşımlar kullanılarak çalışmalar yapılabilir.

Sonuç olarak, günümüzde hileli faaliyetlerin artmasıyla beraber denetim alanında gelişmelerin yaşandığı bilinmektedir. Daha önce vurgulandığı üzere farklı alanlarda kullanılan veri madenciliği bağlı makine öğrenimi yöntemlerini denetim alanında kullanarak gelecekte hızla yaygınlaşacağı düşünülen teknoloji temelli denetime destek olmaya ve bilimsel katkı sağlamaya çalışılmıştır.

## KAYNAKÇA


Albrecht, W. Steve ve Conan C. Albrecht. "Chapter 49 Detecting Fraud" , D. R. Carmichael (Ed.), O. Ray Whittington (Ed.), Lynford Graham (Ed.), Accountant's Handbook Volume Two: Special Industries and Special Topics", New Jersey: John Wiley & Sons, 2007.

Albrecht W. Steve, Chad Albrecht ve Conan C. Albrecht. "Current Trends in Fraud and its Detection", Information Security Journal: A Global Perspective. Vol.17, No.1, March 2008.

Aksoy,Barış. "Finansal Tablo Hilelerinin Makine Öğrenmesi Yöntemleri ve Lojistik Regresyon Kullanılarak Tahmin Edilmesi: Borsa İstanbul Örneği", Maliye ve Finans Yazıları. S.115, 2021, s.29-60.

Association of Certified Fraud Examiners (ACFE), "2018 Global Study on Occupational Fraud and Abuse", Report to The Nation, USA: 2018, s.10.

Baesens, Bart, Véronique Van Vlasselaer ve Wouter Verbeke, Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection. First Edition. New Jersey: John Wiley & Sons, 2015.

Bozkurt, Nejat. İşletmelerin Kara Deliği Hile. 3.Baskı. İstanbul: Alfa Yayınları, 2009.

Chen, Suduan. "Detection of fraudulent financial statements using the hybrid data mining approach", SpringerPlus. 5(89), 2016, pp. 1-16.

Dutta,Ila, Shantanu Dutta ve Bijan Raahemi. "Detecting Financial Restatements Using Data Mining Techniques", Expert Systems With Applications. V. 90, 2017, pp.374-393.

Fernandez, George, Data Mining Using SAS Applications, USA: Chapman & Hall/ CRC Press, 2003.

Fletcher Lu, J. Efrim Boritz ve Dominic Covvey. "Adaptive Fraud Detection Using Benford's Law", Conference: Advances in Artificial Intelligence, 19th Conference of the Canadian Society for Computational Studies of Intelligence. Québec, Canada: Springer-Verlag, 7-9 June 2006, s.348-349.

Golden, Thomas W., Steven L. Skalak, Mona M. Clayton ve Jessica S. Pill. A Guide to Forensic Accounting Investigation. Second Edition. New Jersey: John Wiley & Sons, 2011.

I, Lægreid. "Automatic Fraud Detection — Does it Work?", Annals of Actuarial Science. Volume 2, Issue 02, September 2007, pp 271 – 288.

Jackson, Cecil W. Detecting Accounting Fraud: Analysis and Ethics. Global Edition. England: Pearson Education Limited, 2015.

Jan, Chyan-long. "An effective financial statements fraud detection model for the sustainable development of financial markets: evidence from Taiwan", Sustainability. 10(513), 2018,1-14.

Jans, Mieke, Nadine Lybaert ve Koen Vanhoof, "Data Mining for Fraud Detection: Toward an Improvement on Internal Control Systems?", 30. Eurapean Accounting Association Congress. Lisbon, 2007,pp.,1-27. https://documentserver.uhasselt.be/bitstream/1942/7872/1/EAA%20lisbon.pdf, [17.06.2022].

Kantardzic, Mehmed. Data Mining: Concepts, Models, Methods, and Algorithms. Third Edition, New Jersey, John Wiley & Sons,2020.

Kaggle. Synthetic Data From A Financial Payment System, 2017, https://www.kaggle.com/ealaxi/banksim1 [16Şubat 2022].

Kırlıoğlu, Hilmi ve İsmail Fatih Ceyhan. "Mali Tablo Denetiminde Ön Analitik İnceleme Tekniği Olarak Veri Madenciliğinin Kullanımı: Borsa İstanbul Uygulaması", Akademik Yaklaşımlar Dergisi, 5(1), 2014, s.13-36.

Kirkos, Efstathios, Charalambos Spathis & Yannis Manolopoulos. "Data mining techniques for the detection of fraudulent financial statements", Expert Systems with Applications. V. 32, 2007,pp, 995-1003.

Kotu, Vijay ve Bala Deshpande, Data Science Concepts and Practice, Second Edition, United States, Morgan Kaufmann Publishers, 2019.

Kudyba, Stephan. Big Data, Mining, and Analytics: Components of Strategic Decision Making. Florida: CRC Press, Taylor & Francis Group. 2014.

Larose, Daniel T. Discovering Knowledge In Data: An Introduction to Data Mining. New Jersey: John Wiley & Sons, 2005.

Layton, Robert. Learning Data Mining with Python, First Edition, Birmingham, UK, Packt Publishing Ltd,2015.

Liou, F. M. Fraudulent Financial Reporting Detection and Business Failure Prediction Models: A Comparison, Managerial Auditing Journal. 23(7), 2008.650-662.

Memiş, S, Enginoğlu ve S, Erkan, U. A Data Classification Method in Machine Learning Based on Normalised Hamming Pseudo-Similarity of Fuzzy Parameterized Fuzzy Soft Matrices. Bilge International Journal of Science and Technology Research. Özel Sayı (3) 2019,1-8.

Padgett, Simon. Profiling the Fraudster: Removing The Mask To Prevent And Detect Fraud. First Edition. New Jersey: John Wiley & Sons, 2015.

Pehlivanlı, Davut. Hile Denetimi Metodoloji ve Raporlama. İstanbul: Beta Yayınları, 2011.

Resmî Gazete, "Kişisel verilerin korunması kanunu" 07 Nisan 2016, Sayısı: 29677.

Roiger, Richard J. Data Mining: A Tutorial-Based Primer. USA: CRC Press, Taylor & Francis Group. 2017.

Stamler, Rodney T., Hans J. Marschdorf ve Mario Possamai. Fraud Prevention and Detection: Warning Signs and the Red Flag System. USA: Taylor & Francis Group, 2014.

Statement on Auditing Standards (SAS) No. 99, Consideration of Fraud in a Financial Statement Audit, American Institute of Certified Public Accountants, 2002, s.1721.

S V S, Lakshmi ve Selvani Deepthi Kavila. "Machine Learning For Credit Card Fraud Detection System", International Journal of Applied Engineering Research. V.13/4, 2018, pp.16819-16824.

Terzi, Serkan. "Hile ve Usulsüzlüklerin Tespitinde Veri Madenciliğinin Kullanımı", Muhasebe ve Finansman Dergisi (MUFAD). Sayı.54, Nisan, 2012, s.51-64.

Wells, Joseph T. Corporate Fraud Handbook: Prevention and Detection, Fifth Edition, New Jersey: John Wiley & Sons,2017.

## EKLER

**Veri Seti Erişim Linki**

https://www.kaggle.com/ealaxi/banksim1

**Veri Seti Alternatif Erişim Linki**

**https://drive.google.com/file/d/1MDYGhEVd8OzK9HhOT13BzhTpBpeZtPGq/view?usp=sharing**