

Finansal Tablo Hilelerinin Makine Öğrenmesi Yöntemleri ve Lojistik Regresyon Kullanılarak Tahmin Edilmesi: Borsa İstanbul Örneği¹

Barış AKSOY²

Makale Gönderim Tarihi: 07 Mayıs 2020

Makale Kabul Tarihi: 05 Mart 2021

Öz

Bu çalışmada Borsa İstanbul’da 2000-2019 döneminde kesintisiz işlem gören 88 şirketin, finansal tablolarında sahtekârlık yapıp yapmadıklarını bir yıl öncesinden tahmin etmek için etkili bir model oluşturulması amaçlanmıştır. Bu amaçla makine öğrenmesi yöntemlerinden Yapay Sinir Ağları (ANN), Sınıflandırma ve Regresyon Ağaçları (CART), Destek Vektör Makinesi (SVM) ile Lojistik Regresyon (LR) kullanılarak mali tablo dolandırıcılığı tahmin edilmiştir. Analiz sonucunda ANN (%96,15), CART (%96,15), SVM (%80,77) ve LR (80,77) test örneği genel tahmin doğruluğu elde edilmiştir. ANN ve CART yöntemleri test örneğinde mali tablolarında sahtekârlık yapmış 13 şirketin tamamını (%100.00) doğru sınıflandırmıştır. Bu sonuç mali tablo sahtekârlığı tahmin çalışmalarında kullanılan yöntemlere, bu çalışmada elde edilen tüm modellerin dâhil edilebileceğini göstermektedir.

Anahtar Kelimeler: Finansal Tablo Hileleri, Makine Öğrenmesi Yöntemleri, Lojistik Regresyon, Borsa İstanbul,

Jel Sınıflandırması: C45, C63, G17, G32, M42

¹ Bu çalışma, 3. Uluslararası Ekonomi Araştırmaları ve Finansal Piyasalar Kongresinde (IERFM) sözlü olarak sunulan “Borsa İstanbul’da İşlem Gören Şirketlerde Makine Öğrenmesi Yöntemleri Kullanılarak Mali Tablo Manipülasyonunun Tahmin Edilmesi” adlı tebliğde alınan eleştiri ve katkılar doğrultusunda genişletilmiş ve son şekli verilen halidir.

² Dr. Öğr. Üyesi, Sivas Cumhuriyet Üniversitesi, İİBF, Finans ve Bankacılık Bölümü, Sivas, Türkiye, baksoy@cumhuriyet.edu.tr, Orcid: 0000-0002-1090-5693

Predicting Financial Statement Frauds Using Machine Learning Methods and Logistic Regression: The Case of Borsa Istanbul

Abstract

This study aims to create an effective model to predict one year before whether 88 firms, continuously traded at Borsa Istanbul between 2000-2019, commit fraud in their financial statements. For this purpose, financial statement fraud was predicted by using Artificial Neural Networks (ANN), Classification and Regression Trees (CART) and Support Vector Machine (SVM) and Logistic Regression (LR) methods among machine learning methods. As a result, the overall prediction accuracy of ANN (96.15%), CART (96.15%), SVM (80.77%) and LR (80.77%) test samples were obtained. ANN and CART classified correctly in test samples all 13 firms that fraudulent in their financial statements. This shows that all methods used in this study, can be used in studies to predict financial statement fraud.

Keywords: Financial Statement Fraud, Machine Learning Methods, Logistic Regression, Borsa Istanbul

Jel Classification: C45, C63, G17, G32, M42

1. Giriş

Şirketler, zorlayıcı rekabet ortamı ve küresel ekonominin belirsizlikleri arasında artan bir şekilde zorluklarla ve risk yönetimi sorunları ile karşı karşıya kalmaktadır. Sürekli büyüme ve zorlukların üstesinden gelme sürecinde şirketler, yasalara ve küresel değişimlere uyum sağlamak durumundadırlar. Piyasada rekabetin getirdiği iş zorlukları ve ayakta kalma çabaları şirket yöneticilerinin şirkete ait finansal tablolarda hile yapma konusundaki çabalarını arttırmaktadır. Finansal tablolar, bir şirketin finansal durumunu yansıtan temel belgelerdir. Finansal tablolar, yatırım yapan kamunun, alacaklıların, paydaşların ve diğer muhasebe bilgi kullanıcılarının karar vermeleri için temel dayanaktır. Son dönemde finansal tablo hileleri üzerine yapılan akademik çalışmaların yoğunluğunun literatürde geniş kapsamlı yer bulmasının altında finansal tabloların gerçeği ne kadar yansıttığı konusu yatmaktadır (Küçükkoçaoğlu & Küçüksözen, 2005: 161-163).

Genel olarak kabul edilmiş bir finansal dolandırıcılık tanımı olmasa da finansal dolandırıcılık “yetkisiz şekilde finansal fayda elde etme amaçlı yasa, yönetime veya politikaya aykırı kasıtlı hareket” olarak tanımlanabilir

(Ngai vd., 2011: 559). Yüksek itibara sahip şirketlerin çöküşleri sonucunda kurumsal yönetimin etkinliği, finansal raporların kalitesi ve denetim fonksiyonlarının güvenilirliği konusunda finansal bilgi kullanıcılarında soru işaretleri oluşmuştur (Chintalapati & G.Jyotsna, 2013: 717). Finansal tablolarındaki olağandışı işaretler genellikle finansal dolandırıcılık olmadan önce ortaya çıkmaktadır (Chen, 2016: 14). Mali tablo dolandırıcılığının varlığı bunun ortaya çıkmasından önceki birkaç çeyrek dönemden bir yıla kadar mali tablolarda bulunabilir. Finansal tablo dolandırıcılığının önlenmesi ve yatırımcıların aldatılmasının engellenmesi mali tablo sahtekârlığı olduktan sonra suçluların cezalandırılmasını sağlamaktan daha iyi ve daha düşük maliyete sahiptir. Bu nedenle, etkili bir mali tablo dolandırıcılık tespit modeli oluşturmak oldukça önemlidir (Chen, 2016: 15).

Geleneksel olarak, denetçiler mali tablo dolandırıcılığını tespit etmekten sorumludur. Denetim için harcanan zamanın ve çabanın artmasına rağmen, tespit edilen dolandırıcılıkların sayısı ve tespit oranı büyük ölçüde azalmıştır (Tangod & Kulkarni, 2015: 549-550). Sahtekârlık yapılmış mali tabloların tespitinde oluşabilecek bir başarısızlık, denetim mesleğinin güvenilirliğine ciddi şekilde zarar verebilir. Veri madenciliği gibi yeni teknikler ileri sınıflandırma ve tahmin kabiliyetine sahip bulunmakta ve dolandırıcılık tespit görevini başarıyla yerine getirmede denetçilerin rolünü kolaylaştırmak için kullanılabilir (Omidi vd., 2019: 4). Böyle bir modele duyulan ihtiyaç sonucunda finansal tablo hilelerinin tespit edilmesi ile birlikte önlenmesi çok önemlidir (Tangod & Kulkarni, 2015: 554).

Bu çalışmada Tangod ve Kulkarni (2015) çalışması takip edilerek mali tablo dolandırıcılığının tahmin edilmesi ve tespiti için veri madenciliği yöntemleri kullanılmaktadır. Çalışma, Jan (2018) çalışması takip edilerek yatırımcılara, denetçilere ve finansal piyasadaki tüm paydaşlara verilen zararları azaltmak amacıyla, mali tablo dolandırıcılığının erken dönem işaretlerine dikkat çekerek işletmelerin mali tablo hilelerini tespit etmek ve önleyebilmek için üstün bir model oluşturmayı amaçlamaktadır. 2000-2019 yılları arasındaki SPK haftalık bültenlerinden tek tek SPK denetimleri sonucu finansal tablo dolandırıcılığı yaptığı tespit edilen örnekler alınmıştır. Çalışmada Borsa İstanbul'da işlem görmüş mali tablo hileleri yapan- mali tablo hileleri yapmayan şirketlere ait yayımlanmış mali veriler ve nitel veri kullanılarak şirket yöneticilerinin mali tablo hilelerini tahmin etmede üç makine öğrenmesi yöntemi ve bir istatistiksel yöntem olan Lojistik Regresyonun tahmin performansı, yöntemlerin ve modellerin ayırt edici özellikleri altında karşılaştırılmıştır. Araştırma bulguları ile yatırımcılar, hissedarlar, muhasebeciler, şirket yöneticileri, kredi derecelendirme kuruluşları, denetçiler, menkul kıymet analistleri,

finansal piyasa kurumları yönetici ve danışmanları, mali düzenleyici otoriteler ve ilgili akademik uzmanlar için fayda sağlaması beklenmektedir. İlgili tarafların içinde özellikle denetçiler bu bulgular sayesinde daha etkili kanıtlar toplayabilecek ve denetim planı gerçekleştirebileceklerdir. Diğer taraftan denetçiler, kullandıkları yazılıma bu çalışmada öne çıkan değişkenleri ve sonuçları ekleyerek şirketlerin finansal tablolarını analiz edebilecek ve kırmızı bayrak olarak belirtilen risk faktörlerini belirleyebileceklerdir. Çalışmanın ikinci bölümünde mali tablo dolandırıcılığı, üçüncü bölümünde literatür incelemesi, dördüncü bölümde yöntem ve veri, beşinci bölümde sonuç yer almaktadır.

2. Mali Tablo Dolandırıcılığı

Mali tablo dolandırıcılığı, yatırımcıları ve alacaklıları aldatmak amacıyla bir şirketin hesap defterlerinde yönetim tarafından maddi gerçeklerin kasten yanlış beyan edilmesidir. Şirket yöneticileri tarafından gerçekleştirilen bu yasa dışı eylemler, yatırımcıların güvenini önemli ölçüde azalttığından dünya genelinde ekonomi üzerinde ciddi bir etkiye sahiptir. (Tangod & Kulkarni, 2015: 549). Mali tablo dolandırıcılığı nedeniyle, şirketlerin gerçek finansal durumlarının ve faaliyet sonuçlarının doğru bir şekilde sunulmaması, hem bu şirketlerin menkul kıymetlerine yatırım yapan yatırımcıların zarar etmelerine ve sisteme olan güvenlerinin azalmasına hem de yatırımcıların yanlış yatırım kararları nedeniyle kaynakların verimsiz alanlara tahsis edilmesine sebep olmaktadır (Küçükkoçaoğlu & Küçüksözen, 2005: 161-163).

Bir işletmede hileler, çalışanlar ve yöneticiler tarafından yapılan şekilde ikiye ayrılmaktadır. İşletme yöneticileri tarafından yapılan hilelerin en önemlisi, mali tablo hileleridir. Mali tablo dolandırıcılığı, işletme yöneticilerinin yatırımcılar ve kredi kuruluşları başta olmak üzere finansal tablo kullanıcılarını aldatmak ve yanlış yönlendirmek amacıyla kasıtlı olarak hileli finansal tablo hazırlamasıdır (Terzi, 2012: 52-53). Mali tablo dolandırıcılığının amaçları; pay senedi fiyatının değiştirilmesi, borç sözleşmelerindeki şartların uygun hale getirilmesi, halka açılırken ya da sermaye artırımını yaparken elde edilecek fon tutarını artırmak, politik ve hukuki risklerden kaçınmak, piyasaya olumlu sinyal göndermek, teşvik ve vergi avantajı sağlamak şeklinde sayılabilir. Mali tablo dolandırıcılığının sonuçları ise ekonomide kaynakların en verimli alanlara tahsisinin gerçekleşmemesi, pay senedi fiyatlarında değer düşüklüğü, borçlanma maliyetlerinin artışı, yatırımcıların zarar etmesi sonucu piyasalara güven duyulmaması, bağımsız denetim kuruluşlarının denetim işinden çekilmeleri, müşterilerini kaybetmeleri ya da yüksek tutarlarda tazminata muhatap olmaları olarak sayılabilir (Küçükkoçaoğlu & Küçüksözen, 2005: 163-164).

En iyi bilinen mali tablo dolandırıcılığı ABD’de Enron’un iflâs davasıdır. Enron davası, yatırımcıların finansal tablolara duyulan güveni kaybetmelerine ve şirketlerin bağımsız yöneticilerin başkanlık ettiği denetim komitelerini oluşturmalarını zorunlu kılan Sarbanes-Oxley Yasası’nın (1992) kurulmasını sağladı. Enron davası muhasebe standartlarının reformu ve düzenleyici mekanizmaların yeniden ele alınması ile sonuçlanmıştır (Chen, 2016: 14-15). ABD Kongresi, 2002’deki Sarbanes-Oxley Yasası ile finansal tablo hilelerini önlemek amacıyla yönetim ve denetçilere daha fazla sorumluluk yüklemeyi amaçlamıştır. Sarbanes-Oxley Yasası, düzenleyici gözetim, kurumsal yönetimin güçlendirilmesi ve denetçilerin bağımsızlığını içermektedir (Jan, 2018: 1-2). Günümüzde, SAS 82 gibi standartlar muhasebe ve denetim uzmanları tarafından denetçilerin finansal verilerdeki önemli yanlışlıkları tespit etmedeki performansını arttırmak amacıyla geliştirilmiştir (Omidi vd., 2019: 1). Bağımsız denetçi tarafından olumlu görüşe sahip mali tabloların Türkiye Muhasebe Standartlarına uygun olarak tüm önemli yönleri bakımından gerçeğe uygun biçimde sunulduğu belirtilmektedir. Ancak, olumlu denetçi görüşüne sahip olan finansal tablolar üzerinde hilelerin yapılması muhtemeldir (Fındık & Öztürk, 2016: 484).

Sertifikalı Dolandırıcılık Müfettişleri Birliğinin (ACFE, Association of Certified Fraud Examiners) 2018 yılı raporuna göre finansal dolandırıcılık türlerinden en yaygın olanı “varlıkların kötüye kullanılması” olup bunlar çalışmada meydana gelen olayların % 89’unu oluşturmakta ve en düşük maliyetli dolandırıcılık türü olarak ortalama 114.000 ABD Doları zarara neden olmuştur. “Haksız edinim” çalışmadaki vakaların % 38’ini oluşturmakta ve mağdur kuruluşlara ortalama 250.000 ABD Dolarlık bir zarar vermiştir. En az yaygın olan ancak en fazla zarara neden olan dolandırıcılık biçimi olan “mali tablo sahtekarlıkları” dolandırıcılık olayları içerisinde % 10 pay almakta ve ortalama olarak 800.000 ABD Doları zarara neden olmaktadır. Tablo 1’de finansal dolandırıcılık türlerinin gerçekleşme oranı ve ortalama zarar miktarı bilgisi verilmiştir.

Tablo 1. Finansal Dolandırıcılık Türlerinin Gerçekleşme oranı ve Ortalama Zarar Miktarı (2018 Yılı Raporu)

Hile Türleri	Gerçekleşme Yüzdesi %	Ortalama Zarar
Varlıkların kötüye kullanımı	%89	114.000 USD
Haksız edinim	%38	250.000 USD
Mali tablo sahtekarlıkları	%10	800.000 USD

Kaynak: <https://www.acfe.com/report-to-the-nations/2018/>Erişim Tarihi: 21.08.2019 (2018 Report to the Nations)

3. Literatür İncelemesi

Son yıllarda, bazı uzmanlar ve araştırmacılar tespit hatalarını azaltmak için veri madenciliği kullanmışlardır. Veri madenciliği teknikleri, bu araştırmanın hedefi olan sınıflandırma ve tahmin etmeyi yüksek oranda gerçekleştirmek için uygundur. Geniş miktarda veriyi işleyebilen makine öğrenimi ve veri madenciliği yüksek doğruluk oranları ile tahmin ve sınıflandırma sonuçları verebilmektedir. Yapay Sinir Ağları (ANN), Karar Ağaçları (DT), Destek Vektör Makineleri (SVM) ve Bayesian İnanç Ağları (BBN) muhasebe dolandırıcılığının tespitinde yaygın şekilde kullanılan araçlardır (Jan, 2018: 3).

Beneish (1997) çalışmasında aşırı yüksek finansal performans gösteren şirketlerin mali tablo hileleri yapıp yapmadığını tespit etmek için bir model önermiş ve önerdiği modeli diğer araştırmacıların modelleri ile karşılaştırmıştır. Çalışmada aşırı getiri elde edip etmeme yönünden mali tablo hilesi yapan şirketleri mali tablo hilesi yapmayan şirketlerle karşılaştırmıştır. Çalışmada önerilen modelin mali tablo hilelerini zamanında tespit ettiğini ve yüksek kâr açıklayan şirketlerin finansal tablolarında hile yapıp yapmadıklarını değerlendirmek için kullanılabilceği belirtilmiştir. Aşırı performans gösteren firmaları tespit etmek için yöntem arayan araştırmacıların, değiştirilmiş Jones modelini gecikmiş toplam tahakkuklarla ve geçmiş fiyat performansının bir ölçüsü ile birlikte değerlendirmesi gerektiğini belirtmiştir.

Küçükkocaoğlu ve Küçüksözen (2005) çalışmalarında pay senetleri Borsa İstanbul'da (BİST) işlem gören imalat sanayi şirketlerinin gerçeğe aykırı finansal tablolarını tespit edebilecek bir model geliştirmeyi amaçlamışlardır. Sermaye Piyasası Kurulu tarafından 1993-2002 yılları arasında en fazla mali tablo dolandırıcılığının 2001 yılında gözlemlendiği tespit edilmiştir. 2001 yılında finansal tablolarında gerçeğe aykırı finansal bilgi bulunan ve Sermaye Piyasası Kurulu tarafından finansal tablolarında düzeltme yaptırılan 23 şirketi finansal tablo hileleri yapan şirket olarak belirlemişlerdir. Mali tablo sahtekârlıkları yaptığına dair bulguya rastlanmayan 99 şirket de kontrol şirketi olarak seçmişlerdir. Çok değişkenli lojistik regresyon modelinin kullanıldığı bu çalışmada 12 adet finansal değişken potansiyel tahminci olarak belirlemişlerdir. Analiz sonucuna göre mali tablo dolandırıcılığı yapan şirketleri mali tablo dolandırıcılığı yapmayan şirketlerden ayıran en önemli değişkenlerin “net kârın toplam aktife oranı” ve “toplam finansman giderlerinin toplam faaliyet giderlerine oranı” olduğu bulgusuna ulaşmışlardır.

Kirkos vd. (2007) çalışmalarında veri madenciliği sınıflandırma yöntemlerinden Karar Ağaçları (ID3 Algoritması), Yapay Sinir Ağları, Bayesyan İnanç Ağlarının mali tablo dolandırıcılığı tespitindeki performansı ve bu per-

formansın elde edilmesinde önemli finansal faktörleri belirlemeyi amaçlamışlardır. Girdi verileri 76 Yunan imalat şirketinden elde edilen finansal tablolarından (bilanço ve gelir tabloları) oluşmaktadır. 27 finansal oran belirlenmiş ve boyutu azaltmak amacıyla, iki sınıf arasındaki farkın her değişken için anlamlı olup olmadığını test etmek için ANOVA kullanılmıştır. Performans açısından, 10 katlı çapraz doğrulama yöntemi kullanılan analizlerde Bayesian İnanç Ağı modeli, test örneğinin % 90,30'unu doğru şekilde sınıflandırma ile en iyi performansı göstermiştir. Yapay Sinir Ağı modeli %80,00 ve Karar Ağacı modeli doğruluk oranı ise %73,60 olarak bulunmuştur.

Kotsiantis vd. (2007) çalışmalarında finansal tablo hileleri yapan firmaları tespit etmek için 2001-2002 yılsonu mali tabloları ile Atina Menkul Kıymetler Borsası (ASE) imalat sanayinde listelenen 41 finansal tablo hileleri yapan firma, 123 finansal tablo sahtekârlığı yapmayan firma olmak üzere toplam 164 Yunan firmasının veri seti elde edilerek makine öğrenme algoritmaları kullanılmıştır. Çalışmada algoritmaları birleştiren hibrit bir karar destek sistemi uygulamışlardır. Analiz sonuçları en iyi sınıflandırıcıdan en düşüğe doğru sırasıyla; C4.5 (%91,20), Ripper algoritması (%86,80), kNN (%79,70), Sıralı Minimal Optimizasyon (%78,66), Lojistik Regresyon (%75,30), Bayes (%74,10), ANN (%73,40) doğru şekilde sınıflandırmıştır. Çalışmada az sayıda finansal oran kullanılarak yüksek düzeyde sınıflandırma doğruluğu elde edildiğini belirtmişlerdir. Performans açısından önerilen hibrit karar destek sisteminin incelenen basit ve topluluk yöntemlerinden daha iyi performans sağladığı bildirilmiştir.

Gaganis (2009) çalışmasında sahtekârlık yapılmış finansal tabloların tespiti Lojistik Regresyon Analizi, Diskriminant Analizi, Yapay Sinir Ağları, Olasılıksal Sinir Ağları, En Yakın Komşular, UTADIS ve MHDIS kullanarak 10 alternatif sınıflandırma modeli geliştirmiştir. Modeller finansal ve finansal olmayan veriler kullanılarak yarısı olumlu denetim görüşü almış Yunan firmalarından 398 mali tablo örneği kullanmıştır. İlk olarak 2001 ve 2004 arasında mali tablo hilelerinin yapıldığı belirlenen 199 firma hakkında bilgi elde edilmiştir. 199 mali tablo hileleri yapan şirketten her biri, aynı sektörden ve yaklaşık olarak eşit büyüklükteki mali tablo hileleri yapmayan şirketlerden yıllara göre rastgele eşleştirilmiştir. Analizlerde öncelikle sadece finansal veriler kullanılmıştır. Doğrulama örneğindeki ortalama sınıflandırma doğruluk oranı %76,83 ile %87,20 arasındadır. Finansal olmayan değişkenler eklenerek analiz yapıldığında, ortalama sınıflandırma doğruluğu %84,15 ve %90,24 olarak bulunmuştur.

Şen ve Terzi (2012) çalışmalarında Borsa İstanbul finans sektöründe işlem gören 113 finansal hizmet şirketinin finansal tablo sahtekârlığı yapıp yapmadığını belirlemek amacıyla ANN ve Karar Ağacı yöntemlerini kullanmışlardır. 2009-2010 dönemine ait şirketlerin bilançolarından ve gelir tablolarından elde edilen veriler 12'sinin finansal tablo hileleri yapmış olduğuna yönelik belirtiler içermekte, kalan 101 şirket için denetçilerin görüşlerini kullanarak finansal tablolarında hile yapmamış olarak belirlemişlerdir. ANN analizi sınıflandırma sonucu test seti için %90,00 ve Karar ağacının doğru sınıflandırmadaki başarı oranı %89,40 olarak gerçekleşmiştir. Çalışmada kullanılan 18 oranın 7'sinin hile yapılmış finansal tabloların belirlenmesinde iyi bir gösterge olduğu sonucuna ulaşılmıştır. Bu oranlar, “aktif kârlılığı”, “finansal kaldıraç oranı”, “nakit oranı”, “nakit ve eşdeğerinin toplam varlıklara oranı” ve “öz kaynak kârlılığı” ve “öz kaynağın toplam varlıklara oranı” olarak belirlenmiştir.

Tangod ve Kulkarni (2015) çalışmalarında finansal tablo sahtekârlığı yapan firmaları makine öğrenme teknikleri kullanarak tahmin etmeyi amaçlamışlardır. Bunun için şirketlerin finansal tablolarından oluşan veri seti elde edilerek Olasılıksal sinir ağı, K-Means Kümeleme Algoritması ve Çok Seviyeli İleri Besleme Ağı (MLFF) adlı üç veri madenciliği yöntemi kullanarak modellerin performansını karşılaştırmışlardır. Analiz sonuçlarına göre az sayıda finansal oranın sınıflandırma sonuçlarını büyük ölçüde belirlediği bulunmuştur. Çalışma bulgularına göre “Olasılıksal sinir ağı”, her iki algoritma ile karşılaştırıldığında en yüksek doğruluğu sağlamıştır. Sonuç olarak “denetçi devir hızı”, “toplam isteğe bağlı tahakkuklar”, “4 büyük denetçi tarafından denetlenip denetlenmediği”, “analist tahminleri” ve “beklenmeyen çalışan üretkenliği” gibi bazı değişkenlerin önemli tahmin edici olarak belirlendiğini bildirmişlerdir.

Chen (2016) çalışmasında 2002-2013 döneminde 44 hileli finansal tablo sunan şirketi belirlemiştir. Finansal tablo sahtekârlığı yapmış 44 şirket ve finansal tablo sahtekârlığı yapmamış 132 şirket dâhil olmak üzere toplam 176 şirket seçilmiştir. Analizin ilk aşamasında, sınıflandırma ve regresyon ağaçları (CART) ve Chi kare otomatik etkileşim dedektörü (CHAID) olmak üzere iki karar ağacı algoritması sınıflandırma ve tahminde önemli değişkenlerin seçiminde kullanılmıştır. İkinci aşama, hile yapılmış finansal tabloları saptamak için CART, CHAID, Bayesian inanç ağı (BBN), Destek Vektör Makinesi ve Yapay Sinir Ağı modelleri oluşturulmuştur. Elde edilen sonuçlara göre, CHAID-CART modelinin genel tespit performansı %87,97 (mali tablo sahtekârlığı yapan şirketi tespit doğruluğu %92,69'dır) bir doğrulukla en etkili model olarak bulunmuştur.

Jan (2018) çalışmasında, 2004–2014 döneminde Tayvan Menkul Kıymetler Borsasında mali tablo sahtekârlığı yaptığı tespit edilen 40 şirket ve 120 mali tablo sahtekârlığı yapmayan şirket olmak üzere toplam 160 şirketten oluşan örnek kullanmıştır. Mali tablo sahtekârlığı yapmış bir şirketi aynı yıl ve aynı sektördeki finansal tablo sahtekârlığı yapmamış üç normal şirketle eşleştirilmiştir. Çalışmada ilk aşamada, önemli değişkenleri belirlemek için ANN ve SVM kullanmıştır. İkinci aşamada, sınıflandırma için dört tip karar ağacı “sınıflandırma ve regresyon ağacı (CART)”, “ki-kare otomatik etkileşim dedektörü (CHAID)”, “C5.0” ve “(QUEST)” yöntemleri ile modeller oluşturmuştur. Ampirik bulgular, ANN kullanılarak değişken seçimi sonucu elde edilen az sayıda değişkenle yürütülen CART analizinin finansal tablo hileleri tespitinde %90,83 doğrulukla en iyi genel sınıflandırma sonuçlarını verdiğini göstermektedir. ANN + CART modelinin finansal tablo hilelerinin tespiti ve tahmininde en yüksek sınıflandırma doğruluğuna (%90,21) sahip olduğu bildirilmiştir.

Omidi vd. (2019) çalışmalarında kullanılan veri seti, Çin borsasında listelenen 2659 şirketten elde edilmiştir. Bu çalışma, çok yönlü olarak denetimli öğrenme yöntemlerinden çok katmanlı ileri beslemeli sinir ağı (MFFNN), olasılıksal sinir ağı (PNN), SVM, çok terimli log-lineer model (MLM) ve diskriminant analizinin (DA) finansal tablo hileleri tespit performansını incelemişlerdir. Ampirik sonuçlar, MFFNN’nin, finansal tablolarda sunulan gerçeğe aykırı verilerin tespitinde en iyi sınıflandırma sonuçlarını verdiğini gösterdiği belirtilmiştir.

Yukarıda belirtilen çalışmalardan bazıları (Küçükkocaoğlu ve Küçük-sözen (2005), Chen (2016)) finansal tablo dolandırıcılığını tespit etmek için yeni bir yonteme ait model sunmaya odaklanmıştır. Bazı çalışmalar (Küçükkocaoğlu ve Küçüksözen (2005), Kirkos vd. (2007), Gaganis (2009)) hileli finansal tablo tespit çalışması için yalnızca mali tablolardan elde edilen finansal oran değişkenlerini kullanmışlardır. Bazı çalışmalarda (Kirkos vd. (2007)) sadece 10 katlı çapraz doğrulama yöntemi kullanılmıştır. Bu çalışmayı literatürdeki diğer çalışmalardan ayıran yönleri; i) 2000-2019 yılları arasındaki SPK haftalık bültenlerinden tek tek SPK denetimleri sonucu finansal tablo dolandırıcılığı yaptığı tespit edilen örnekler alınmıştır. ii) İlgili örneklere ait ve bu şirketlerle aynı yıl, aynı sektörde olan ancak finansal tablo sahtekârlığı yapmamış aynı sayıda örnek belirlenmiştir. iii) Finansal tablo sahtekârlığı yapmış 44 ve Finansal tablo sahtekârlığı yapmamış 44 işletmeye ait finansal tablo dolandırıcılığı yılından bir yıl öncesi finansal tablolardan elde edilen finansal oranlar, şirket haber duyuruları, dipnotlar, yatırım danışmanlık şirketleri internet sayfaları ve gazetelerden elde edilen nicel değişkenler ve nitel

değişken bu çalışmada kullanılmıştır. iv) Makine öğrenmesi yöntemlerinde daha hassas bir tahmin sonucu elde etmek için doğrulama yöntemi olarak veri seti öncelikle %70 eğitim ve %30 test örneği olarak ikiye ayrılmıştır. Tüm veri setinin %70'ini oluşturan eğitim örneği üzerinde 10 katlı çapraz doğrulama yöntemi kullanılmış ve elde edilen modeli test etmek için algoritmanın eğitim esnasında daha önce hiç görmediği tüm veri setinin %30'unu oluşturan test örneği kullanılmıştır. v) Çalışma finansal tablo dolandırıcılığının tespit edildiği yıldan bir yıl önceki yılsonu finansal tablolardan elde edilen verilerle yürütülmüştür. Dolayısı ile çalışmada finansal tablo dolandırıcılığının varlığına yönelik bir tespit dışında bir yıl öncesinden finansal tablo hilesi olup olmadığı tahmin edilmiştir.

4. Yöntem ve Veri

Günümüzde sayıları hızla artan araştırmacılar, finansal tablo hilelerini tespit etmek için veri madenciliği yöntemlerini uygulamışlardır. Finansal tablo hilelerini tespit ve tahmin çalışmalarında finansal değişkenler genellikle normal dağılım gibi istatistiksel koşullara uymamaktadır. Bu nedenle, verilerin istatistiksel varsayım gerektirmeyen makine öğrenme yöntemi ortaya çıkmış ve araştırmacılar tarafından sınıflandırıcı olarak kullanılmaktadır. Ampirik sonuçlar, makine öğrenme yönteminin olumlu bir sınıflandırma etkisine sahip olduğunu göstermektedir (Chen, 2016: 3-4). Albashrawi (2016) çalışmasında finansal tablo sahtekârlığının tahmin ve tespit edilmesine yönelik olarak yapılan çalışmalarda en çok kullanılan yöntemlerden az kullanıma göre sırasıyla “Lojistik Regresyon”, Yapay Sinir Ağları”, “Karar Ağaçları”, “Diskriminant Analizi”, “Bayesian İnanç Ağı”, “Destek Vektör Makinesi” ve “En Yakın K Komşu” olduğu görülmüştür. Bu nedenle bu çalışmada finansal tablo sahtekârlığının tahmin edilmesinde Yapay Sinir Ağları (ANN, Artificial Neural Networks), Sınıflandırma ve Regresyon Ağaçları (CART, Classification and Regression Trees), Destek Vektör Makinesi (SVM, Support Vector Machine) ve Lojistik Regresyon (LR, Logistic Regression) yöntemleri kullanılmıştır.

4.1. Veri Seti ve Araştırmanın Kısıtları

Mali tablo sahtekârlıklarının tespit edilmesi ve önlenmesine yönelik kesin sonuç veren yöntem bulunmamaktadır. Bunun nedeni hile ve usulsüz gibi görünen işlemlerin kesin bir biçimde hileli olduğunun söylenememesidir. Bunun için hile denetiminde her işlem şüpheli görülür (Terzi, 2012: 53). Finansal sektörün mali tabloları diğer sektörlerle karşılaştırılabilir değildir ve finansal oranları genel sektörlerinkinden farklıdır. Bu nedenle literatürde birçok çalışmada finansal sektör çalışmaya dâhil edilmemiştir. Zaman, endüstri ve şirket büyüklüğü gibi pek çok dış çevresel faktörü kontrol etmek için eş-

leştirme yöntemi benimsenebilir (Chen, 2016: 9). Jan (2018) çalışmasında da zaman dilimleri, endüstriler ve firma büyüklükleri gibi dış çevre faktörlerini kontrol etmek için örnekleme dâhil olan şirketleri karşılaştırmak için eşleştirme yöntemi kullanılmıştır.

Araştırma ile ilgili veriler Varıcı ve Er (2013) çalışması takip edilerek BIST Kamuyu Aydınlatma Platformu (KAP) altında yer alan bilgi sorgulama kısmında şirketlerin finansal tablo verileri, dipnotlar ve faaliyet raporlarından yararlanılarak elde edilmiştir. Ayrıca çalışmanın yapıldığı dönemde özellikle mali tablo sahtekârlığı yapan şirketlerin büyük kısmının pay senetleri, çalışmanın yapıldığı 2019 yılında BIST’de alınıp satılmadığından dolayı BIST ve KAP’tan şirketlerle ilgili nicel ve nitel veriler elde edilememiştir. İlgili şirketlerin bazı nicel verileri (örneğin halka açılma tarihi, yönetim kurulu üye sayısı, halka açıklık oranı gibi) aracı kuruluş internet sayfalarından, gazete ve finans dergilerinin internet sayfalarından tek tek incelenerek ulaşılmıştır. Mali tablo dolandırıcılığı yapan 44 şirketin her biri ile aynı sektör ve döneme ait 44 finansal tablo sahtekârlığı yapmayan şirket örneği olmak üzere toplam 88 örnek belirlenmiştir. Örneklerin alındıkları sektörler; imalat sanayi (177 adet), toptan ve perakende ticaret, otel (32 adet), ulaştırma (10 adet), holding ve yatırım şirketleri (46 adet) olmak üzere toplam ana kütle 265 iken çalışma, verilerin elde edilebilirliği göz önüne alınarak 88 örnek ile yürütülmüştür. Bu çalışmada 44 finansal tablolarında hile yapan şirket ile eşleşmesi için aynı sayıda aynı sektörden finansal tablolarında hile yapmayan şirket örneği alınması yoluna gidildiğinden örnek sayısı 88 şirket ile sınırlıdır. Dolayısı ile çalışmanın en önemli kısıtı örnek sayısının düşüklüğü olarak belirtilebilir.

Halka açık şirketlerdeki finansal tablo sahtekârlıkları, SPK’nın denetim ve incelemeleri sonucunda belirlenmeye çalışılmaktadır. İşletmenin finansal tabloları üzerinde sahtekârlık yapıp yapmadıklarına yönelik kriter kararlaştırıldıktan sonra ana kütlede mali tablo dolandırıcılığı yapan ve mali tablo dolandırıcılığı yapmayan firmaların belirlenmesi gerekmektedir. En yaygın kullanılan örnekleme yöntemi, finansal tablo dolandırıcılığı yapan firmaların örnekleme belirlendikten sonra aynı sayıda ve aynı endüstride olan finansal tablo dolandırıcılığı yapmayan şirket örnekleri belirlemektir (Gallego & Quintana, 2012: 159-160). Bu çalışmada Küçükkocaoğlu ve Küçüksözen (2005) çalışması takip edilerek SPK’nın denetim ve incelemeleri sonucu finansal tablo dolandırıcılığı yaptığı tespit edilerek kamuya açıklanan şirketler finansal tablo dolandırıcılığı yapmış olan şirketler olarak alınmıştır. Bu şirketleri eşleştirmek amacıyla aynı sektörde ancak dolandırıcılık yaptığı yönünde bir bilgi bulunmayan aynı sayıda ve aynı yıllara ait mali tablolara dayalı olan ve mali tablolara dayalı olmayan bilgileri elde edilen şirketler finansal tablo dolandırıcılığı yapmamış şirket örneği olarak alınmıştır.

Tablo 2’de mali tablo hilesi yapan ve mali tablo hilesi yapmayan şirketlerin seçilme kriterleri verilmiştir.

Tablo 2. Mali Tablo Hilesi Yapan ve Mali Tablo Hilesi Yapmayan Şirketlerin Seçilme Kriterleri

No	SPK’nın Mali Tablo Hilesi Yapıldığına Dair Haftalık Bültenlerinde Belirttiği ve Suç Duyurusunda Bulunduğu Kriterler
1	X şirketinin yasal defterlerinin mevzuata uygun olarak tutulmaması tespiti
2	Kuruluma sunulan yasal defter ve kayıtlarının gerçeği tam ve doğru olarak yansıtmaktan uzak olması...
3	Şirketin muhasebe kayıtlarının gerçeği yansıtmaması ve işlemlerin süresi dahilinde muhasebe kayıtlarına intikal ettirilmemesi nedeniyle şirketin yasal defterleri ile ilgili olarak gerçeğe aykırı beyanda bulunarak yasal defterleri ibrazdan imtina etmesi nedeniyle...
4	Holding ve grup şirketlerine ait defter ve belgeleri Kurul uzmanlarına ibraz edilmemesi, defter ve kayıt tutulmaması, defter ve kayıtlarda gerçeğe aykırı kayıt tutulması, ortaklardan toplanan paraların büyük kısmının kayıtlara yansıtılmamış olması, şirkete ortak olmak amacıyla para vermiş olan şahısların yasal kayıtlarda ortak olarak gösterilmemiş olması nedeniyle...
5	Gerçeğe aykırı muhasebe kayıtları, sahte ödeme makbuzları ve mutabakat yazıları düzenlenmesi dolayısıyla...
6	Şirket muhasebe kayıtlarında verilen sipariş avansı olarak gösterilen tutarların gerçek durumu yansıtmamasına ilişkin olarak
7	Şirket’in 2005 ve 2006 yıllarına ilişkin kamuya açıklanan mali tablolarının yalan, yanlış, yanıltıcı olmasının ve gerçeği yansıtmamasının SPK’nın 47/1-A-3 maddesinde tanımlanan “Sermaye piyasası araçlarının değerini etkileyebilecek yalan, yanlış, yanıltıcı, mesnetsiz bilgi verme” suçu kapsamına girmesi nedeniyle...
8	Gerçeğe aykırı muhasebe kayıtlarıyla X’in Y’ye borçlandırılması, ilişkili taraflardan ve Z A.Ş.’den olan alacaklara ilişkin hesapların gerçeğe aykırı tutulması ve grup şirketlerine verilen kefaletler nedeniyle maruz kalınan dava ve takipler ile akabinde imzalanan protokoller kapsamında sorumlu olunan borçlara ilişkin olarak mali tablolarda gerekli karşılığın ayrılması
9	2005 yılı ve sonrasındaki dönemlerde şirketin ticari mallarının kayıt dışı satışı ve şirket varlıklarının satış tutarının resmi defterlere eksik intikal ettirilmesi suretiyle, hileli ve gerçeğe aykırı muhasebe kayıtları tesis edilmesi ve şirketin zarara uğratılması nedeniyle...
10	Şirket’in toplam 10,37 ton miktarındaki X alım satım kayıtlarını doğuran/içeren ve bu kayıtlara bağlı olarak gerçekleştirilen tüm muhasebe kayıtlarında muhasebe hilesi yapılması ve söz konusu hilelerle 31.03.2008 tarihli finansal tablolardan başlamak üzere Şirket tarafından 2008, 2009, 2010, 2011 ve 2012 yıllarında yayımlanan tüm finansal tabloların gerçeğe aykırı olması nedeniyle
11	Şirket’in stok hesaplarına gerçeğe aykırı kayıtlar yapılması ve şirket malvarlığının azalmasına yol açılması.
12	X A.Ş.’nin yönetim, denetim veya sermaye bakımından doğrudan veya dolaylı olarak ilişkide bulunduğu gerçek veya tüzel kişiler ile emsallerine uygunluk, piyasa teamülleri, ticari hayatın basiret ve dürüstlük ilkelerine aykırı şartlar içeren anlaşmalar veya ticari uygulamalarla karının veya malvarlığının azaltılması. X A.Ş.’nin yasal defterlerinde, muhasebe kayıtlarında ve finansal tablo ve raporlarında usulsüzlük yapılması. X A.Ş. yöneticilerince Kurulumuzca istenen bilgi ve belgelerin verilmemesi
13	Finansal tablolarda maddi duran varlıkların, bu varlıklarda gerçekleşen çıkış tutarlarının, değer artışlarının ve buna bağlı olarak özkaynakların olması gerekenden fazla gösterilmesi.
14	Toplam tutarı 11.800.000-TL ve 1.205.000-USD olan şirket çeklerine ilişkin bilgilere Şirket’in muhasebe kayıtlarında, dolayısıyla Şirket’in 31.12.2016 ve 31.03.2017 tarihli finansal tablolarında yer verilmemesi.

4.2. Değişken Seçimi

Mali tablo dolandırıcılığı ile kurumsal yönetim gibi finansal olmayan bilgiler arasında bir korelasyon bulunmaktadır. Bir şirket iyi bir finansal statüye sahip olduğunda ve iyi bir kurumsal yönetim mekanizmasına sahip olduğunda, mali tablo dolandırıcılığının rapor edilmesi muhtemeldir. Bu nedenle finansal tablo hileleri tespitinde finansal olmayan bilgileri de dikkate almak gerekir (Jan, 2018: 2). Jan (2018) çalışması takip edilerek finansal tablo dolandırıcılığını tespit etmek için tahmin doğruluğunu artırmak amacıyla bu çalışmada finansal tablolardan elde edilen nicel değişkenler ve nitel değişken analize dâhil edilmiştir. Finansal oran ile işletme durumu arasındaki ilişki dinamik olmakta ve oranlar farklı ülkelerde, farklı sanayilerde ve ekonomik döngülerin her aşamasında farklılık göstermektedir. Bu nedenle Lussier (1995) çalışmasında olduğu gibi çalışmamızda da elde edilen finansal oran setinin genelleştirilebileceği iddia edilmemektedir. Finansal oranlar kullanılarak, bir endüstri içindeki şirketler, sektörler arasında veya bir firma içinde karşılaştırmalar yapılabilir. Bu tür bir araç, aynı zamanda, farklı büyüklükteki şirketlerin göreceli performansını karşılaştırmak için de kullanılabilir (Delen: 2013: 3970). Bağımsız değişkenler arasındaki ilişkilerin yüksek olması aynı olguları ifade etme ihtimalinin yüksek olması demektir. Aynı olguları ifade eden değişkenlerin tamamının modele dâhil edilmesi bilgi kirliliği oluşturmaktadır. Sonuç olarak bağımsız değişkenler arasındaki güçlü korelasyon değişkenlerin modele tek tek katkısının ayırt edilmemesine neden olmaktadır (Varıcı & Er, 2013: 49). Bu nedenle bu çalışmada korelasyon analizi sonucunda “Faaliyet Kârı/Net Satışlar” ve Faaliyet Gideri/Net Satışlar” değişkenleri arasında %70’in üzerinde korelasyon bulunduğundan dolayı söz konusu iki değişken veri setinden çıkarılmıştır.

Tablo 3’de çalışmada kullanılan değişkenler verilmiştir. Ek 3’te bu çalışmada kullanılan değişkenleri çalışmalarında kullanan araştırmacılara ait bilgiler verilmiştir.

Tablo 3. Çalışmada Kullanılan Değişkenler

No	Değişkenler	No	Değişkenler
X1	Stoklar/ Aktif Toplam	X10	Vergi Öncesi Kâr / Net Satışlar
X2	Maddi Duran Varlık/Aktif Toplam	X11	Vergi, Faiz ve Amortisman Öncesi Kâr/Aktif Toplam
X3	Toplam Borç/Aktif Toplam	X12	Nitel Değişken: Dört büyük denetim şirket tarafından denetlenip denetlenmediği, denetlenmiyorsa 0, denetleniyorsa 1 ile kodlanmıştır)
X4	Net Kâr /Aktif Toplam	X13	Yönetim Kurulu Üye Sayısı
X5	Dönen Varlıklar/Kısa Vadeli Yabancı Kaynaklar	X14	Yönetim Kurulunda Bağımsız Üye Sayısı
X6	Brüt Kâr/Net Satışlar	X15	Halka Açıklık Oranı %
X7	Net Satışlar/Aktif Toplamı	X16	Halka Açıklık Süresi (Yıl)
X8	Brüt Kâr / Aktif Toplam	X17	Halka Açıklıkta Yabancı Payı (Sermayede Doğrudan %5 veya Daha Fazla Paya veya Oy Hakkına Sahip Gerçek ve Tüzel Kişiler)
X9	Satılan Malın Maliyeti/Ortalama Stoklar	X18	Bağımlı Değişken: Mali tabloları hileli işletme örneği 0 ile, mali tablo hileli olmayan işletme örneği 1 ile kodlanmıştır)

4.3. Çalışma Metodolojisi

Bir sınıflandırma algoritması eğitim seti aracılığıyla sonuç için tahminler sağlayacak nitelikler arasındaki ilişkileri bulmaya çalışır. Sonraki aşamada, öğrenilen ilişkiler sınıf özniteliği haricinde aynı nitelikleri içeren bir test seti kullanılarak tahminler oluşturulur. Son olarak, bu tahminler algoritmanın performansını analiz etmek için gerçek sınıf değerleri ile karşılaştırılır. Performans seviyeleri algoritma yapılarına ve parametre değerlerine bağlıdır. Performanslar doğruluk, hassaslık, F ölçütü, Kappa istatistiği vb. ile ölçülebilmektedir (Özdağoğlu vd., 2017: 70). Her bir sınıflandırıcı model oluşturulurken parametreler optimize edilmiştir. Parametre optimizasyonu, geniş aralıktaki bir dizi parametrenin en iyi olanlarını keşfetmek için tekrar tekrar test edilebildiği bir süreçtir. Parametrelerin optimize edilmiş değerleri ve seçilen sınıflandırıcıların performansı, veri setine, seçilen finansal göstergeler ve nitel değişkenlere bağlı olarak değişebilir.

K-katlı çapraz doğrulama yöntemi optimal veri dağılımı için literatürde sıklıkla kullanılan etkili bir yöntemdir. K-katlı çapraz doğrulamada veriler rastgele k sayıda eşit miktarda parçaya ayrılır. Sırasıyla bir parça test için, kalanlar eğitim için kullanılarak analiz yapılır. Sonra başka bir parça test, diğerleri eğitim için kullanılır. Her aşamada veri madenciliği analizi yapılır ve parçaların tümü test edildikten sonra genel performans elde edilir. Yapılan deneysel çalışmalarda, uzman görüşlerine göre k sayısı için en uygun değer 10 bulunmuştur (Çelik vd., 2017: 243). Bu çalışmada 10 katlı çapraz doğrulama

kullanılmadan önce veriler %70 eğitim ve %30 test örnekleme olarak ikiye ayrılmıştır. %30 test veri kümesi algoritmaya gösterilmeyerek %70 eğitimden oluşan veri seti 10 parçaya bölünmüş dokuz parçası eğitim geri kalan bir parça test verisi olarak analiz edilmiş, değişkenlere ağırlıklar atanmış ve değişkenlerin her birinin önem dereceleri belirlenerek modellerin öğrenmesi tamamlanmıştır. Her bir değişkene ağırlıkların ataması yapılarak öğrenme süreci tamamlandıktan sonra algoritmanın daha önce hiç görmediği tüm veri setinin %30'unu oluşturan test veri seti kullanılmış ve modellerin performansı test edilmiştir.

Veri Madenciliği uygulamalarını gerçekleştirmek için ticari ve açık kaynak olmak üzere birçok program mevcuttur. Bu programlar arasında RapidMiner (YALE), WEKA ve R programları en çok kullanılanlar arasındadır. RapidMiner özellikle WEKA ve R dâhil olmak üzere açık kaynak kodlu veri madenciliği programları arasında liderdir (Dener vd., 2009: 1-2). Bu nedenle bu çalışmada ANN, CART ve SVM analizleri için RAPİDMİNER 9.3 programı kullanılmıştır. Çalışma metodolojisi aşağıda verilmektedir.

- ✓ 2000-2019 yılları arasında Sermaye Piyasası Kurulu (SPK) tarafından yapılan denetim sonucunda finansal tablo dolandırıcılığı yaptığı tespit edilen şirket ana kütesinin belirlenmesi,
- ✓ SPK tarafından mali tablo sahtekârlığı tespit edilen ana kütlede bankalar, sigorta şirketleri ve finans sektöründe faaliyet gösteren diğer şirketlerin çıkarılması,
- ✓ SPK tarafından mali tablo sahtekârlığı tespit edilen ana kütledeki şirketlere ait nitel verinin ve nicel verinin BIST, KAP ve ilgili şirketlerin internet sitelerinden, yatırım ve finansal kuruluşlardan, gazetelerden elde edilmesi,
- ✓ Finansal tablo sahtekârlığı yapan şirketlerle aynı sektörde, aynı yıllara ait finansal tablo sahtekârlığı yapmayan şirketlerin mali tablo verileri ve nitel verinin elde edilmesi,
- ✓ Mali tablo dolandırıcılığı yapan ve mali tablo dolandırıcılığı yapmayan işletmelere ait nicel değişkenlerin ve nitel değişkenin programa girilmesi,
- ✓ ANN, CART ve SVM yöntemlerinde 17 değişkenli veri seti, LR analizinde diskriminant ileri adımlı analiz ile belirlenen 6 değişkenli veri setinin kullanılması,

- ✓ ANN, CART ve SVM analizlerinde veriler %70 eğitim ve %30 test verisi olarak ikiye ayrılmıştır. Eğitim verisi üzerinde 10 katlı çapraz doğrulama yöntemi kullanılarak oluşturulan modeller, algoritmaların daha önce hiç görmediği test verisi üzerinde test edilmiş ve elde edilen test sonuçları çalışmada paylaşılmıştır. LR analizinde veriler %70 eğitim ve %30 test seti olarak ikiye ayrılmış, eğitim seti üzerinde kurulan model test seti üzerinde doğrulama yapılmıştır.

Literatürde bazı araştırmacılar (Tangod ve Kulkarni (2015), Kotsiantis vd. (2007)) çeşitli yöntemlerle finansal oran sayısını azaltarak daha yüksek sınıflandırma sonucuna ulaştıklarını belirtmişlerdir. Bu sonuç çalışmamızda yalnızca LR için geçerlidir. 17 değişkenli tüm veri seti kullanıldığında %69,23 sınıflandırma performansı elde edilmişken diskriminant ileri adımli analiz ile belirlenen 6 değişkenle LR analizi yürütüldüğünde %80,77 genel sınıflandırma doğruluğuna ulaşılmıştır. Bu çalışmada kullanılan ANN, CART ve SVM yöntemlerinde diskriminant ileri adımli regresyon ile belirlenen 6 değişken ve CART analizi ile belirlenen 4 değişken kullanılarak ANN, CART ve SVM analizleri gerçekleştirilmiştir. Ancak diskriminant ileri adımli regresyon ile belirlenen 6 değişken ve CART analizi ile belirlenen 4 değişken kullanılarak yapılan analizlerde doğru tahmin oranları tüm değişkenlerin kullanıldığı analiz sonuçlarından daha düşük çıktığından 17 değişkenle yürütülen ANN, CART ve SVM analiz sonuçları bu çalışmada paylaşılmıştır.

4.4. Çalışmada Kullanılan Yöntemler

4.1. Yapay Sinir Ağları

Yapay Sinir Ağları (ANN, Artificial Neural Networks) biyolojik sinir ağının hesaplama yeteneklerini taklit eden bir sistem olup bir dizi birbirine bağlı düğüm kullanarak insan beyninin işlevselliğinden ilham alan doğrusal olmayan istatistiksel bir modelleme aracıdır (Chintalapati & G.Jyotsna, 2013: 721). Bir sinir ağı içindeki nöronlar genellikle katmanlar halinde düzenlenir. Yapay sinir ağının yapısı, sinirlerin girdi katmanı, gizli katman ve çıktı katmanı ile birbirine bağlandığı temelde üç katmandan oluşmaktadır. Girdi katmanı istatistikte bağımsız değişkenlere karşılık gelir. Çıktı katmanı istatistikte bağımlı değişkenlere karşılık gelmektedir. Gizli katmandaki nöronlar girdi katmanından gönderilen sinyalleri alırlar ve çıktı katmanına sinyaller gönderirler. Katmanların yanı sıra, en önemli unsurlardan biri nöronlar arasında veri iletimini sağlayan bağlantıdır (Şen & Terzi, 2012: 86). Her girdi değişkeni, bir bağlantı ağırlığına sahiptir. Sinir hücresi, tüm girdi değişkenlerini ağırlıklarına göre birleştirir. Sinir ağlarının evrimsel sürecinde, başlangıçtaki tahmin değişkeni için gereken ağırlığı belirlemek için eğitim modeline sahip bir para-

metreye ihtiyaç duyulmakta ve bu parametreler rastgele üretilmektedir (Chen, 2016: 8).

Katman sayısı, her tabakadaki nöron sayısı, öğrenme hızı, momentum değerleri bu tür ağların tasarımı için önemli parametrelerdir. Parametreler öğrenme hızlarının, momentumun ve diğer işlem elemanlarının belirlenmesinde önemlidir ve parametrelerin seçilmesi konusunda net bir kılavuz bulunmamaktadır (Özdağoğlu vd., 2017: 71). ANN'de işlem elemanı ağırlık değerlerinin belirlenmesinde (ağın eğitilmesi) ağırlıklar rastgele atanır. ANN'ler kendilerine örnekler gösterildikçe bu ağırlık değerleri değişmektedir. Örnekler ağa defalarca gösterilerek en doğru ağırlık değerleri bulunmaya çalışılır ve doğru ağırlık değerine ulaşıldığında olay hakkında genelleme yapılır ise ağın öğrendiği belirtilebilir. Ağın eğitimden sonra öğrenip öğrenmediğini test etme işlemi eğitim sırasındaki bağlantı ağırlıkları değiştirilmeden önce ağın görmediği örnekler için çıktı üretmesi ile olur. Test çıktı değeri ne kadar iyi olursa eğitimin performansı da o kadar iyi demektir (Öztemel, 2012: 55). Özdağoğlu vd. (2017) çalışması takip edilerek sınıflandırıcı olarak ANN analizi üç önemli parametrenin (öğrenme hızı, eğitim döngü sayısı ve momentum) değerleri esas alınarak yürütülmüştür. ANN algoritması, optimize edilmiş eğitim döngü sayısı, öğrenme oranı, momentum ve sigmoid fonksiyonu, tahminler için kullanılacak bir gizli katman ve normalleştirilmiş veri kümesinde yürütülmüştür.

İleri sürümlü YSA'da kullanılan ağırlıklar her seferinde Δw kadar düzeltilerek yenilenir.

$$w_1^{yeni} = w_1^{eski} + \Delta w_1 \quad (1)$$

Algoritmanın en hassas noktası Δw değerlerini bularak en uygun w ağırlıklarını elde etmektir. Bunun için her seferinde oluşan hatayı minimuma indirecek bir yapı kullanılır. Gerçekte var olan değer g ile; w ağırlıklarıyla elde edilen değer de y ile gösterilirse en küçük kareler yöntemiyle elde edilecek hata fonksiyonu E şu şekilde hesaplanabilir (Silahtaroglu, 2016: 124-125):

$$E_r = 1/2 e^2 = 1/2 (g - y)^2 \quad (2)$$

Her bağlantı, bir aktivasyon fonksiyonu, çoğunlukla bir lojistik fonksiyon veya hiperbolik teğet girişlerinin ağırlıklı toplamı kullanılarak, iki nöron arasındaki ve her bir nöron arasındaki ilişkinin gücünü temsil eden bir (sayısal) ağırlık ile temsil edilir. Sinir ağı kullanılarak tasarlanan bir finansal tablo sahtekârlığı tespit modeli, belirli bir firma için, finansal tablo sahtekârlığı ihtimalini temsil eden bir gizli katman, bir çıktı nöronu ve bir girdi katmanından oluşan ağ ile aşağıdaki şekilde ifade edilebilen bir Z skoru hesaplanmaktadır (Öztemel, 2012: 55).

$$Z = f \left(f \left(\sum_{i=1}^n w_{ij} x_i + b_j \right) \cdot \left(\sum_{j=1}^p w_j \right) + b \right) \quad (3)$$

Formülde f aktivasyon fonksiyonu, n değişken sayısı, p gizli nöron sayısı, x_i girdi katmanı nöronları, w_{ij} girdi katmanı ve gizli katman arasındaki ilişkileri temsil eden ağırlıkları, w_j gizli katman ve çıktı katmanı grupları arasındaki ağırlıkları, b_j gizli nöronların ağırlıkları ve b çıktı nöronunun ağırlığını göstermektedir (Jardin, 2016: 241).

4.4.2. CART Karar Ağacı

Sınıflandırma ve regresyon ağacı (Classification and Regression Tree, CART) veri kümesinin sınıflandırılmasında kullanılan karar ağacı tekniğidir ve hangi kayıtların belirlenmiş bir sonuca dâhil olacağını tahmin etmek için yeni (sınıflandırılmamış) bir veri kümesine uygulanabilecek bir dizi kural sunar. CART ikili bölmeler oluşturarak veri kümesini sınıflara ayırmaktadır (Dhanalakshmi ve Subramanian, 2014: 330). Örneklerin art arda bölünmesi sonucu büyük bir ağaç üretilmiş olabilir. Ağacın dallarından bazıları, yanlış değerler veya aykırı değerler gibi eğitim setindeki anomalileri yansıtabildiğinden ağacın budanması gerekmektedir. Ağaç budaması, bölme düğümlerinin, modelin doğruluk oranını önemli ölçüde etkilemeyecek şekilde çıkarılmasını gerektirir (Kirkos vd., 2007: 999).

Gini ölçütü bir frekans dağılımında değerler arasındaki eşitsizliğin ölçüsüdür. Bu ölçütte öznitelik değerlerinin sol ve sağda olmak üzere iki bölme ayrılması, her bir bölüm için ayrı ayrı Gini ölçütünün hesaplanması ve elde edilen sonuçların karşılaştırılması esasına dayanmaktadır. (Özkan, 2016:111). CART hangi düğümün kök düğüm olacağına karar vermenin dışında düğümün hangi noktadan ikiye ayrılması gerektiğini de hesaplar. CART, dallara ayırma kriterini hesaplarken kayıp verileri dikkate almaz. Hesaplanan $\Psi(s/t)$ değerleri içinden en büyük değere sahip nokta, düğüm olarak seçilir ve işlem tüm yapraklara kadar aynı şekilde devam ettirilir (Silahtaroglu, 2016:83).

$$\Psi(s/t) = 2P_L P_R \sum_{j=1}^M |P(C_j | t_L) - P(C_j | t_R)| \quad (4)$$

t : Dalların yapılacağı düğüm

c : Kriter

L : Ağacın sol tarafı

R : Ağacın sağ tarafı

P_L, P_R : Öğrenim kümesindeki bir kaydın sağda veya solda olma olasılığı

$|P(C_j | t_L) - P(C_j | t_R)|$: C sınıfındaki bir kaydın sağda veya solda olma olasılığı

4.4.3. Destek Vektör Makinesi

SVM, Vapnik (1995) tarafından istatistiksel öğrenme teorisi ve SRM (yapısal risk azaltma) temelli bir makine öğrenme yöntemidir. Öncelikle eğitim verileri kullanılarak iki veya daha fazla sınıfı öğrenme mekanizmasıyla ayırt edebilen optimal bir ayırıcı hiper düzlem bulmaktadır (Chen, 2016: 8). Destek vektör makineleri öğrenme teorisine dayanır ve veriler doğrusal olarak ayrıldığında çok iyi çalışır. Regresyon modellerinden farklı olarak, bu modeller genellikle değişkenler arasında etkileşim üretmeyi gerektirmez ve ANN'lere benzer şekilde, gürültüye ve aykırı değerlere karşı dayanıklıdır (Omidi vd., 2019: 4). SVM metodolojisi, iki veri sınıfını ayıran bir "marj" kavramı ile açıklanır. Marjı maksimize etmek ve böylelikle, ayırıcı hiper düzlemler arasında mümkün olan en büyük mesafeyi oluşturmak, beklenen genelleme hatasını azaltabilir (Apparao vd., 2009: 161). İki grup, iki boyutlu bir düzlem üzerinde gösterilmiştir. Bu düzlem ve boyutlar birer özellik olarak düşünülebilir. Yani basit anlamda sisteme giren her girdinin bir özellik çıkarımı yapılmış ve sonuçta bu iki boyutlu düzlemde her girdiyi gösteren farklı bir nokta elde edilmiştir. SVM çekirdek ağırlıkları, her bir örneğin bağımsız değişkenini tabloda paylaşılan değerler ile çarpılması ve toplanması sonucu ilgili örneğin hangi sınıfa ait olduğunun görülebilmesi için Ek 1'de verilmiştir.

4.4.4. Lojistik Regresyon

Lojistik regresyon analizi adını bağımlı değişkene uygulanan logit dönüşürmeden almaktadır. Lojistik regresyonda model oluşturulmasında en küçük kareler yöntemi yerine en çok olabilirlik yöntemi kullanılmaktadır. Lojistik regresyon olasılık, odds ve odds'un logaritmasına dayanmaktadır (Çokluk, Şekercioğlu ve Büyüköztürk, 2016: 62-63). Lojistik regresyonda odds, bir olayın meydana gelme olasılığının o olayın oluşmaması olasılığına bölünmesi şeklinde tanımlanmaktadır ve hesaplanması formül 5'de verilmektedir:

$$odds = \frac{p(x)}{1-p(x)} \quad (5)$$

$P(x)$ = Bir x olayının gerçekleşme olasılığı

$1-p(x)$ = X olayının gerçekleşmeme olasılığı

Lojistik regresyon analizi diskriminant analizi ve çok değişkenli regresyon analizinden farklı olarak bağımsız değişkenin dağılımı ile ilgili varsayımlar gerektirmemektedir. Lojistik regresyon analizinde bağımsız değişkenlerin normal dağılması, doğrusallık ve varyans-kovaryans matrislerinin eşitliği gibi varsayımların karşılamasına gerek bulunmamaktadır (Çokluk, Şekercioğlu ve

Büyüköztürk, 2016: 49). 17 değişkenle yürütülen lojistik regresyon analizi tahmin sonuçları düşük çıktığından diskriminant ileri adımlı analiz ile belirlenen 6 değişkenle lojistik regresyon analizi gerçekleştirilmiştir.

4.5. Makine Öğrenmesi Yöntemlerine Ait Bulgular ve Değerlendirme

Bu çalışmada elde edilen bulgular modelin performansının veri setine, seçilen algoritmaya ve değişkenlere bağlı olduğunu ortaya koymuştur. Gaganis (2009) çalışmasında model doğrulama türü olarak 10 katlı çapraz doğrulamanın tespit doğruluğunu artırmak için en iyi yöntemlerden biri olduğunu belirtmiştir ve %75'in üzerinde tespit doğruluğunun sosyal bilimler alanında iyi bir sonuç olduğunu belirtmiştir. Gaganis (2009) sınıflandırmasına göre bu çalışmadaki CART ve ANN sonuçları çok iyi, SVM ve LR sonucu ise iyi olarak değerlendirilebilir. Analiz sonucunda ANN (%96,15), CART (%96,15), SVM (%80,77) ve LR (%80,77) test örneği genel tahmin doğruluğu elde edilmiştir.

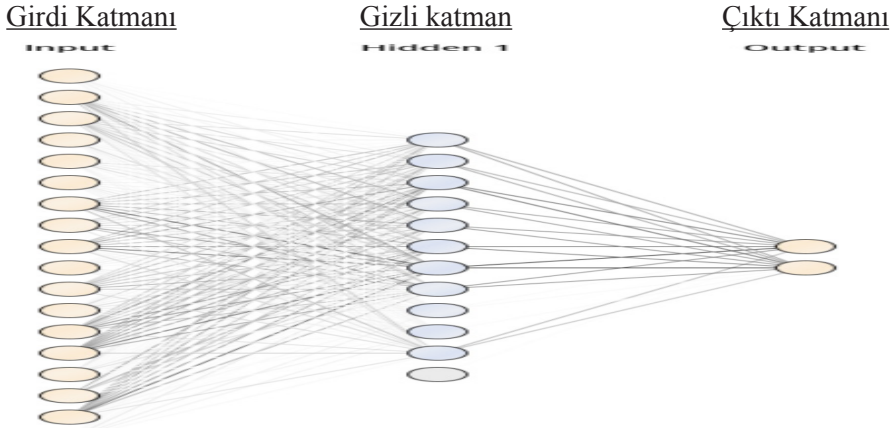
4.5.1. Yapay Sinir Ağları Analizi ve Bulguları

Yapay Sinir Ağları analizinde öğrenme oranı ve momentum %80, %60, %40 ve devir sayısı 1000, 5000 gibi farklı değerlerle analizler yürütülmüş ancak parametre optimizasyonu ile en iyi performansı gösteren yapay sinir ağları modeli sonucunda herhangi bir iyileşme olmadığından Tablo 4'de verilen parametreler programa girilmiştir. Tabakalı örneklem (stratified sampling) seçiminde her iki sınıftan (mali tabloları hileli, mali tabloları hileli olmayan) eşit sayıda ancak rastgele örnekler alınmıştır. Tabakalı örneklem seçimi bağımlı değişkenin kategorik olduğu durumlarda başarılı sonuç vermektedir (Liang vd., 2015: 291). ANN ağırlıkları, her bir örneğin bağımsız değişkenini verilen tabloda paylaşılan değerler ile çarpılması ve toplanması sonucu ilgili örneğin hangi sınıfa ait olduğunu görülebilmesi için Ek 2'de verilmiştir. Kısaca aynı veri seti ve aynı parametreler kullanılarak her bir bağımsız değişkenle ilgili düğüm ağırlıkları aynı verilir ise çalışmada elde edilen sonuçlara ulaşılabileceğinin belirtilmesi amacıyla düğüm ağırlıkları Ek 2'de paylaşılmıştır.

Tablo 4. Yapay Sinir Ağı Parametreleri

Ağ Türü	Çok katmanlı perseptron			
Öğrenme Algoritması	Geri yayılım			
Öğrenme Kuralı	Momentum			
Girdi Katmanındaki Düğüm Sayısı	17			
Gizli Katman Sayısı	1			
Gizli Katmandaki Düğüm Sayısı	11			
Çıktı Katman Düğüm Sayısı	2 (Mali tabloları hileli -Mali tabloları hileli olmayan)			
Değişken Seçimi	17 Değişkenli veri seti			
Doğrulama Türü	%70 Eğitim ve %30 test seti olarak ayrılmış, eğitim seti üzerinde 10 katlı çapraz doğrulama yöntemi kullanılmış, algoritmanın eğitim esnasında görmediği %30 test verisi ile modelin test edilmesi			
Örneklem Seçim Türü	Tabakalı örneklem seçimi (Stratified sampling)			
Aktivasyon Fonksiyonu	Sigmoid			
Öğrenme Oranı	En Düşük: 0,00	En Yüksek: 0,30	Adımlar: 10	Ölçek: Doğrusal
Momentum	En Düşük: 0,00	En Yüksek: 0,30	Adımlar: 10	Ölçek: Doğrusal
Eğitim Devir Sayısı	En Düşük: 1,00	En Yüksek: 500	Adımlar: 10	Ölçek: Doğrusal

Şekil 1’de Yapay sinir ağları model görüntüsü verilmektedir.

**Şekil 1. Yapay Sinir Ağları Model Görüntüsü**

Çalışmada kullanılan modellerin performansları 6 kriter yönünden sıralanmıştır. Bunlar: i) Doğruluk (accuracy) ii) Kesinlik (Precision), iii) Duyarlılık (Recall), iv) Kappa değeri, v) AUC, vi) F ölçütüdür. Son olarak en iyi modeller, söz konusu 6 kriter tarafından ölçülen yüksek sınıflandırma gücünün uyumluluğuna dayalı olarak seçilmiştir.

Tablo 5’de parametre optimizasyonu sonucu en iyi performansı gösteren yapay sinir ağıları model sonucu verilmektedir. Tablo 5’de yer alan Kappa istatistiğinden kısaca bahsetmek gerekmektedir. Kappa, iki ya da daha fazla gözlemci arasındaki uyumun güvenilirliğini ölçen istatistiksel bir yöntemdir. Uyumun değerlendirildiği değişken kategorik (nominal) olduğundan dolayı parametrik olmayan bir istatistiktir. “Cohen’in kappa katsayısı” iki gözlemci arasındaki uyumu incelerken, gözlemci sayısının ikiden fazla olduğu durumda “Fleiss’in kappa katsayısı” kullanılmaktadır. Fleiss tarafından yapılan sınıflamada, Kappa değeri 0,75 ve üzerinde olması durumunda mükemmel, 0,40-0,75 arasında olması orta-iyi, 0,40’ın altında ise zayıf bir uyumun varlığı şeklinde değerlendirilmektedir (Kılıç, 2015). Bu çalışmada elde edilen ANN modeline ait Kappa değeri 0,957 olduğundan iki gözlemcinin sonuçları mükemmel uyumlu olduğu görülmektedir. AUC, sınıflandırma maliyeti veya doğruluğu hakkında bir bilgi bulunmadığında farklı sınıflandırma modellerini karşılaştırmak için kullanılabilir (Gaganis, 2009: 222).

Tablo 5. Parametre Optimizasyonu Sonucu En İyi Performans Gösteren Yapay Sinir Ağları Model Sonucu

Parametreler	Sonuçlar
Doğruluk (accuracy)	%96,15
Sınıflandırma Hatası (classification error)	%3,85
Kappa	0,923
AUC	0,964
Kesinlik (precision)	%100,00
Duyarlılık (recall)	%92,31
F ölçütü	%96,00
Öğrenme Oranı (learning rate)	0,21
Momentum	0,0
Devir Sayısı (training cycles)	101

Gaganis (2009) çalışmasında model doğrulama türü olarak 10 katlı çapraz doğrulamanın tespit doğruluğunu artırmak için en iyi yöntemlerden biri olduğunu belirtmiştir ve %75’in üzerinde tespit doğruluğunun sosyal bilimlerde iyi bir sonuç olduğunu belirtmiştir. Gaganis (2009) sınıflandırmasına göre bu çalışmadaki ANN sonucu çok iyi, olarak değerlendirilebilir. Analiz sonucunda ANN modeli ile %96,15 test örneği genel tahmin doğruluğu elde edilmiştir. ANN modeli test örneğini oluşturan 13 mali tablolarında hile yapan şirketin tamamını %100,00 doğrulukla, 13 mali tablolarında hile yapmayan şirketin 12’sini %92,31 doğrulukla tahmin etmiş, genel tahmin doğruluk oranı olarak ise %96,15 bulunmuştur.

4.5.2. CART Karar Ağacı Analizi ve Bulguları

CART Karar Ağacı Analizinde farklı değerlerle analizler yürütülmüş ancak parametre optimizasyonu ile en iyi performansı gösteren CART modeli sonucunda herhangi bir iyileşme olmadığından Tablo 6’da verilen parametreler programa girilmiştir.

Tablo 6. CART Sınıflandırma Algoritması Analiz Parametreleri

Analiz İçin Belirlenen Parametreler	Açıklama			
Değişken Seçimi	17 Değişkenli Veri Seti			
Doğrulama Türü	%70 Eğitim ve %30 test seti olarak ayrılmış, eğitim seti üzerinde 10 katlı çapraz doğrulama yöntemi kullanılmış, algoritmanın eğitim esnasında görmediği %30 test verisi ile modelin test edilmesi			
Örneklem Seçimi	Tabakalı örneklem seçimi (Stratified sampling)			
Bölünme Kriteri	Gini Index			
Analiz İçin Belirlenen Parametreler	En Düşük	En Yüksek	Adımlar	Ölçek
Bölünme İçin Minimum Boyut (Minimal Size For Split)	1,0	4	10	Doğrusal
En Düşük Dal Boyutu (Minimal Leaf Size)	1,0	2	10	Doğrusal
Minimum Kazanım (Minimal Gain)	1,0	20	10	Doğrusal
Maksimum Derinlik (Maximal Depth)	1	20	10	Doğrusal
Güven Düzeyi (confidence)	0,0	0,25	10	
Ön Budama Sayısı	1,0	10	10	-

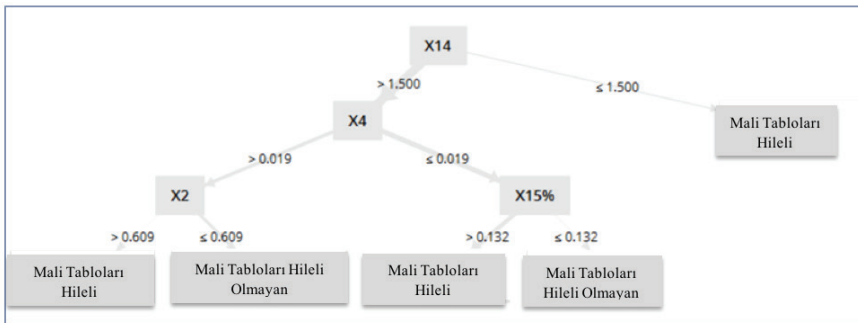
Tablo 7’de parametre optimizasyonu sonucu en iyi performans gösteren CART Sınıflandırma ve regresyon ağacı analizi performans sonuçları verilmektedir.

Tablo 7. Parametre Optimizasyonu Sonucu En İyi Performansı Gösteren CART Sınıflandırma ve Regresyon Ağacı Analizi Performans Sonuçları

Parametreler	Sonuçlar
Doğruluk (accuracy)	%96,15
Sınıflandırma Hatası (classification error)	%3,85
Kappa	0,923
AUC	0,962
Kesinlik (Precision)	%100,00
Duyarlılık (Recall)	%92,31
F Ölçütü	%96,00
Bölünme İçin Minimum Boyut (Minimal Size for Split)	2
En Düşük Dal Boyutu (Minimal Leaf Size)	1
En düşük Kazanım (Minimal Gain)	8,60
En Yüksek Derinlik	11

Gaganis (2009) sınıflandırmasına göre bu çalışmadaki CART sonucu çok iyi olarak değerlendirilebilir. Chen (2016) çalışmasında, CHAID-CART modelinin genel tespit performansı % 87,97 (finansal tablolarda hile yapan şirketi tespit doğruluğu % 92,69'dır) bir doğrulukla en etkili model olarak bulunmuştur. Bu çalışmada verilerin optimal bir şekilde dağılımı amacıyla doğrulama yöntemi olarak %70 eğitim-%30 test verisi ayrımı ve 10 katlı çapraz doğrulama yöntemi kullanılarak hassas bir tahmin modeli oluşturulmuştur. CART modeli test örneğini oluşturan 13 mali tablolarında hile yapan şirketin tamamını %100,00 doğrulukla, 13 mali tablolarında hile yapmayan şirketin 12'sini %92,31 doğrulukla tahmin etmiş, genel tahmin doğruluk oranı olarak ise %96,15 bulunmuştur.

Şekil 2'de CART Karar ağacı görüntüsü verilmiştir. CART karar ağacının oluşturulmasında ağacın kökünü "X14, Yönetim Kurulunda Bağımsız Üye Sayısı" değişkeni bulunmuştur. X14 değişken değeri 1,5'den küçük veya eşit olan 9 işletmenin tamamı "Mali tabloları hileli" olarak bulunmuştur. X14 değişken değeri 1,5'ten büyük olan işletmeler için ağacın karar vermesinde ikinci kırılımı sağlayan ve ilk dalı oluşturan "X4, Net Kâr/Aktif Toplam" değişkenidir. X4 değişkeni 0,019'dan büyük olan işletmeler için ağacın karar vermesinde ikinci dalı oluşturan "X2, Maddi Duran Varlık / Aktif Toplam" değişkenidir. X2 değişkeni 0,609'dan küçük veya eşit olan 20 işletmeden 18'i "Mali tabloları hileli olmayan" 2 işletme "Mali tabloları hileli" olarak bulunmuştur. X2 değişkeni 0,609'dan büyük olan 2 işletme "Mali tabloları hileli" olarak bulunmuştur. "X4, Net Kârın Aktif Toplama Oranı" değişkeni 0,019'dan küçük veya eşit olan işletmeler ile ilgili ağacın karar vermesinde üçüncü dalı oluşturan "X15, Halka Açıklık Oranı" değişkenidir. X15 değişkeni 0,132'den büyük olan 27 işletmeden 18'i "Mali tabloları hileli", 9'u "Mali tabloları hileli olmayan" olarak bulunmuştur. X15 değişkeni 0,132'den küçük veya eşit olan 4 işletme "mali tabloları hileli olmayan" olarak bulunmuştur.



Şekil 2. CART Karar Ağacı Görüntüsü

4.5.3. Destek Vektör Makinesi Analizi ve Bulguları

Destek Vektör Makinesi analizinde farklı değerlerle analizler yürütülmüş ancak parametre optimizasyonu ile en iyi performansı gösteren SVM modeli sonucunda herhangi bir iyileşme olmadığından Tablo 8’de verilen parametreler programa girilmiştir.

Tablo 8. SVM Analiz Parametreleri

Analiz İçin Belirlenen Parametreler	Açıklama
Veri Seti	17 Değişkenli Veri Seti
Doğrulama Türü	%70 Eğitim ve %30 test seti olarak ayrılmış, eğitim seti üzerinde 10 katlı çapraz doğrulama yöntemi kullanılmış, algoritmanın eğitim esnasında görmediği %30 test verisi ile modelin test edilmesi
Örnekleme Seçimi	Tabakalı Örnekleme Seçimi (Stratified sampling)
SVM.C	0.03125-0.125-0.5-2-8-32-128-512-2048-8192-32768-131072
SVM.gamma	0.00003052-0.00012207-0.000488-0.00195-0.0078125-0.03125-0.125-0.5-2-8
Önbellek Boyutu (Cache Size)	80
Çekirdek Tipi (Kernel Type)	rbf

Tablo 9’da SVM Analizi performans sonuçları verilmiştir.

Tablo 9. SVM Analizi Performans Sonuçları

Parametreler	Sonuçlar
Doğruluk (accuracy)	%80,77
Sınıflandırma Hatası (classification error)	%19,23
Kappa	0,615
AUC	0,781
Kesinlik (Precision)	%83,33
Duyarlılık (Recall)	%76,92
F ölçütü	%80,00
SVM.C	8192
SVM.gamma	0,00003052

Gaganis (2009) sınıflandırmasına göre bu çalışmadaki SVM sonucu iyi olarak değerlendirilebilir. SVM 13 mali tabloları hileli şirketin 11’ini %84,62 doğrulukla, 13 mali tabloları hileli olmayan şirketin 10’unu %76,92 doğrulukla tahmin etmiş, genel tahmin doğruluk oranı olarak ise %80,77’dir.

4.5.4. Lojistik Regresyon Analizi ve Bulguları

LR analizinde tüm değişkenlerle analiz yürütülmüş ancak diskriminant ileri adımlı analiz ile belirlenen 6 değişkenle yürütülen analiz sonucu daha yüksek tahmin performansı gösterdiğinden Tablo 10’da verilen parametreler programa girilmiştir.

Tablo 10. Lojistik Regresyon Analiz Parametreleri

Analiz İçin Belirlenen Parametreler	Açıklama
Veri Seti Ayrımı	%70 Eğitim - %30 Test Veri Seti Ayrımı
Değişkenler	X7-Net Satışlar/Aktif Toplamı X8-Brüt Kâr / Aktif Toplam X9-Satılan Malın Maliyeti/Ortalama Stoklar X13-Yönetim Kurulu Üye Sayısı X14-Yönetim Kurulunda Bağımsız Üye Sayısı X17-Halka Açıklıkta Yabancı Payı (Sermayede Doğrudan %5 veya Daha Fazla Paya veya Oy Hakkına Sahip Gerçek ve Tüzel Kişiler)

Tablo 11’de Lojistik Regresyon analizi performans sonuçları verilmektedir.

Tablo 11. Lojistik Regresyon Analizi Performans Sonuçları

Parametreler	Sonuçlar
Doğruluk (accuracy)	%80,77
Sınıflandırma Hatası (classification error)	%19,23
Kappa	0,615
Kesinlik (precision)	%90,00
Duyarlılık (recall)	%69,23
F ölçütü	%78,26

Gaganis (2009) sınıflandırmasına göre bu çalışmadaki LR analiz sonucu iyi olarak değerlendirilebilir. LR modeli test örneğindeki 13 mali tabloları hileli şirketin 12’sini %92,31 doğrulukla, test örneğindeki 13 mali tabloları hileli olmayan şirketin 9’unu %69,23 doğrulukla tahmin etmiştir. Analiz sonucunda LR modeli ile %80,77 test örneği genel tahmin doğruluğu elde edilmiştir.

5. Sonuç

Bu çalışmada şirketlerde mali tablo dolandırıcılığını bir yıl öncesinden tahmin ve tespit etmek için etkili bir model oluşturulması amaçlanmıştır. Bu amaçla Sermaye Piyasası Kurulu tarafından yürütülen denetimler sonucunda 2000-2019 yılları arasında pay senetleri Borsa İstanbul'da alınıp satılan ve mali tablolarında hile yaptığı tespit edilen 44 işletme örneği alınmıştır. Mali tablolarında hile yapan şirketle eşleşmesi amacıyla mali tabloları hileli şirket ile aynı sektörde ve aynı yıl mali tablolarında hile yapmayan şirket örneği alınarak toplam 88 örnek elde edilmiştir. Örneklem Borsa İstanbul sektörlerinden imalat (177 şirket), toptan ve perakende ticaret, lokanta otel (32 şirket), ulaştırma (10 şirket), holding ve yatırımdan (46 şirket) oluşan 265 şirketin olduğu ana kütlede 44 finansal tablolarında hile yapmış, 44 finansal tablolarında hile yapmamış şirket olmak üzere 88 şirket verilerinden oluşmaktadır. Örnek sayısının artırılması amacıyla 2000-2019 yılları arasında finansal tablolarında hile yapan şirketleri belirlemek için 988 haftalık Sermaye Piyasası haftalık bültenleri tek tek incelenmiş ancak sadece 44 finansal tablolarında hile yapan şirket örneğine ulaşılmıştır. Bu çalışmada 44 finansal tablolarında hile yapan şirket ile eşleşmesi için aynı sayıda aynı sektörden finansal tablolarında hile yapmayan şirket örneği alınması yoluna gidildiğinden örnek sayısı 88 şirket ile sınırlıdır. Ana kütleli oluşturan 265 şirketten elde edilen 88 örnek şirket verisi ile yürütülen bu çalışmanın en önemli sınırlılığı veri setinin sınırlı olmasıdır. Bu nedenle çalışma sonuçlarının ana kütleli oluşturan sektörlerdeki şirketlerin tümüne veya Borsa İstanbul'da işlem gören tüm şirketlere genelleme yapılması uygun değildir.

Mali tablo sahtekârlığı yılından bir yıl önceki yılsonu mali tablo verileri ve nitel veriler elde edilerek makine öğrenmesi yöntemlerinden Yapay Sinir Ağları (ANN), Sınıflandırma ve Regresyon Ağaçları (CART), Destek Vektör Makinesi (SVM) ve Lojistik Regresyon (LR) yöntemleri kullanılarak mali tablo sahtekârlığı yapan ve mali tablo sahtekârlığı yapmayan şirketler tahmin edilmiştir. Kurulan modelleri doğrulamak amacıyla tüm veri seti %70 eğitim ve %30 test verisine ayrılmış, %70 eğitim verisi üzerinde 10 katlı çapraz doğrulama yöntemi kullanılarak parametre optimizasyonu ile en iyi sonucu veren modeller elde edilmiştir. Kurulan modeller eğitim esnasında algoritmanın daha önce hiç görmediği veri setinin %30'unu oluşturan test seti üzerinde uygulanmış ve oluşturulan modeller test edilmiştir. Analiz sonucunda ANN (%96,15), CART (%96,15), SVM (%80,77) ve LR (%80,77) test örneği genel tahmin doğruluğu elde edilmiştir. ANN ve CART yöntemleri 88 örneklemden oluşan veri setinin %30'unu oluşturan 26 test örnekleminin (13 manipüle edilen, 13 manipüle edilmeyen) biri dışında tamamını doğru tahmin etmiştir. Her

iki yöntem mali tablo sahtekârlığı yapan işletmelerin tamamını doğru tahmin etmiş, mali tablo sahtekârlığı yapmayan bir işletmeyi manipüle edildi şeklinde yanlış sınıflandırmıştır. Bu sonuç finansal tablo sahtekârlığı tahmin çalışmalarında ilgililerin kullandıkları yöntemlerin arasına bu çalışmada kurulan tüm modelleri dâhil edebileceklerini göstermektedir. SVM 13 mali tabloları hileli şirketin 11'ini %84,62 doğrulukla, 13 mali tabloları hileli olmayan şirketin 10'unu %76,92 doğrulukla tahmin etmiş, genel tahmin doğruluk oranı olarak ise %80,77'dir. LR ise 13 mali tabloları hileli şirketin 12'sini %92,31 doğrulukla, 13 mali tabloları hileli olmayan şirketin 9'unu %69,23 doğrulukla tahmin etmiştir. Tablo 12'de ANN, CART, SVM ve LR analiz sonuçları verilmektedir.

Tablo 12. ANN, CART, SVM ve LR Analiz Sonuçları

Sınıflandırma Yöntemi	Mali Tabloları Hileli	Mali Tabloları Hileli Olmayan	Genel
ANN	100,00% 13	92,31% 12	96,15%
CART	100,00% 13	92,31% 12	96,15%
SVM	84,62% 11	76,92% 10	80,77%
LR	92,31% 12	69,23% 9	80,77%

Bu çalışma, finansal tablo dolandırıcılığı tespitinde makine öğrenme-yöntemlerine yönelik mevcut yaklaşımları optimize eder ve iyileştirir. Bu çalışmanın sonuçları çeşitli grupların, özellikle denetçilerin ve yatırımcıların, denetledikleri veya yatırım yaptıkları şirketleri dikkatlice seçmelerine ve karar vermelerine olanak sağlar. Diğer taraftan akademisyenler ve yatırımcıların borsa yatırımlarına ilişkin stratejik karar vermelerinde yarar sağlar. Ayrıca, finansal tabloların denetçileri, bu araştırmada önerilen teknikleri denetim prosedürlerinde uygulayabilirler. Diğer borsalardan bir veri seti alındığında da karşılaştırmalı deneysel çalışmalar yapılabilir. Yasal normlar ve denetim şartlarının yeterince katı olup olmadığı ve kurumsal yönetimin finansal tablolarda sahtekârlık yaparak şirket varlıklarını suiistimal, zimmete geçirme, kasıtlı olarak azaltma eylemlerini önleyip önleyemeyeceği de ele alınması gereken konulardandır.

Kaynakça

- Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: a decade review from 2004 to 2015. *Journal of Data Science*, 14, 553-570.
- Apparao, G., Singh, A., G.S.Rao, Bhavani, B., K.Eswar, & D.Rajani. (2009). Financial statement fraud detection by data mining. *Int. J. of Advanced Networking and Applications*, 1(3), 159-163.
- Beneish, M.D., (1997), "Detecting GAAP Violation: Implications for Assessing Earnings Management Among Firms with Extreme Financial Performance", *Journal of Accounting and Public Policy*, 16(3), 271-309.
- Chen, S. (2016). Detection of fraudulent financial statements using the hybrid data mining approach. *Chen SpringerPlus*, 5(89), 1-16.
- Chintalapati, S. S., & G.Jyotsna. (2013). Application of data mining techniques for financial accounting fraud detection scheme. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3 (11), 717-724.
- Çelik, U., Akçetin, E., & Gök, M. (2017). *Rapidminer ile Veri Madenciliği*. İstanbul: Pusula Yayınları.
- Çokluk, Ö., Şekercioğlu G. ve Büyüköztürk Ş. (2016). *Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve LISREL Uygulamaları*, Ankara: Pegem Akademi Yayını.
- Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*, 40, 3970-3983.
- Dener, M., Dörterler, M., & Orman, A. (2009). Açık kaynak kodlu veri madenciliği programları: weka'da örnek uygulama. *Conference: XI. Akademik Bilişim Konferansı, Şanlıurfa: 1-11.*
- Fındık, H., & Öztürk, E. (2016). Finansal bilgi manipülasyonunun beneish modeli yardımıyla ölçülmesi: BIST imalat sanayi üzerine bir araştırma. *İşletme Araştırmaları Dergisi*, 8(1), 483-499.
- Gaganis, C. (2009). Classification Techniques for the identification of falsified financial statements: a comparative analysis. *Intelligent Systems in Accounting, Finance and Management*, 16, 207-229.
- Gallego, A. G., & Quintana, M. J. (2012). Business failure prediction models: finding the connection between their result and the sampling method. *Preparation of Electronic Manuscripts for Publication*, 157-168.
- Jan, C.-I. (2018). An effective financial statements fraud detection model for the sustainable development of financial markets: evidence from Taiwan. *Sustainability*, 10(513), 1-14.
- Jardin, P. d. (2016). A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research*, 254, 236-252.
- Kılıç, S., (2015). Kappa testi, *Journal of Mood Disorders*, 5(3), 142-144.
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32, 995-1003.

- Kotsiantis, S., Koumanakos, E., Tzelepis, D., & Tampakas, V. (2007). Forecasting fraudulent financial statements using data mining. *International Scholarly and Scientific Research & Innovation*, 1(12), 844-849.
- Küçükkocaoğlu, G., & Küçüksözen, C. (2005). Gerçeğe aykırı finansal tabloların ortaya çıkarılması: İMKB şirketleri üzerine ampirik bir çalışma. *Muhasebe ve Finansman Dergisi*, 28, 160-171.
- Liang, D., Tsai, C.-F. & Wu, H.-T. (2015). The effect of feature selection on financial distress prediction. *Knowledge-Based Systems*, 73, 289-297
- Lussier, R. N. (1995). A nonfinancial business success versus failure prediction model for young firms . *Journal of Small Business Management*, 8-20.
- Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50, 559-569.
- Omidi, M., Min, Q., Moradinaftchali, V., & Piri, M. (2019). The efficacy of predictive methods in financial statement fraud. *Discrete Dynamics in Nature and Society*, 1-12.
- Özdağoğlu, G., Özdağoğlu, A., Gümüş, Y., & Kurt-Gümüş, G. (2017). The application of data mining techniques in manipulated financial statement classification: The case of Turkey. *Journal of AI and Data Mining*, 5(1), 67-77.
- Özkan, Y. (2016). *Veri Madenciliği Yöntemleri*, İstanbul: Papatya Yayınları.
- Öztemel, E. (2012). *Yapay Sinir Ağları*. İstanbul: Papatya Yayınları.
- Dhanalakshmi, S. & C.Subramanian. (2014). An analysis of data mining applications for fraud detection in securities market. *International Journal of Data Mining Techniques and Applications*, 3, 326-335.
- Silahtaroglu, G. (2016). *Veri Madenciliği Kavram ve Algoritmaları* . İstanbul: Papatya Yayınları.
- Şen, İ. K., & Terzi, S. (2012). Detecting falsified financial statements using data mining: empirical research on finance sector in Turkey. *Maliye Finans Yazıları* , 96, 76-94.
- Tangod, K., & Kulkarni, G. (2015). Detection of financial statement fraud using data mining technique and performance analysis. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(7), 549-555.
- Terzi, S. (2012). Hile ve usulsüzlüklerin tespitinde veri madenciliğinin kullanımı . *Muhasebe ve Finansman Dergisi*, Nisan, 51-64.
- Varıcı, İ., & Er, B. (2013). Muhasebe manipülasyonu ve firma performansı ilişkisi: İMKB uygulaması . *Ege Akademik Bakış*, 13(1), 43-52.
- 2018 Report to the Nations. (2019). 08 21, 2019 tarihinde <https://www.acfe.com/report-to-the-nations/2018/>:<https://s3-us-west-2.amazonaws.com/acfe-public/2018-report-to-the-nations.pdf> adresinden alındı.

EKLER

Ek 1. SVM Çekirdek Ağırlıkları

Değişkenler	Ağırlıklar	Değişkenler	Ağırlıklar
w(X1)	11709,105	w(X11)	10205,045
w(X2)	45405,639	w(X12)	34294,028
w(X3)	58559,513	w(X13)	651454,183
w(X4)	-5180,597	w(X14)	220965,944
w(X5)	158685,052	w(X15)	49348,611
w(X6)	25167,667	w(X16)	2027173,879
w(X7)	84546,122	w(X17)	3793,715
w(X8)	10047,579	Mali tabloları manipüle edilmeyen sınıf için destek vektörlerin sayısı	16
w(X9)	1101550,392	Mali tabloları manipüle edilen sınıf için destek vektörlerin sayısı	20
w(X10)	2252,607	Bias (offset)	-38,045

Ek2. ANN Ağırlıkları

Girdi Katman	Gizli Katman											Eşik
	Düğüm 1	Düğüm 2	Düğüm 3	Düğüm 4	Düğüm 5	Düğüm 6	Düğüm 7	Düğüm 8	Düğüm 9	Düğüm 10	Düğüm 11	
Bağımsız Değişken	Sigmoid											
X1	0,028	0,045	-0,188	-0,113	-0,053	-0,182	0,018	0,038	0,077	0,076	-0,070	
X2	-0,264	-0,245	0,561	0,860	0,892	0,469	1,017	-0,437	0,070	0,115	0,505	
X3	-0,486	-0,444	-0,636	0,391	0,466	-0,640	0,688	-0,577	0,127	0,129	0,341	
X4	-0,154	-0,077	0,064	0,115	0,151	0,084	0,449	-0,172	-0,107	-0,067	0,082	
X5	0,054	0,090	0,482	0,259	0,291	0,468	0,120	0,022	0,092	0,117	0,130	
X6	0,118	0,090	-0,194	-0,162	-0,207	-0,161	-0,288	0,183	-0,116	-0,088	-0,111	
X7	-0,895	-0,863	-0,969	1,036	1,159	-1,072	1,950	-1,223	0,164	0,064	0,715	
X8	0,633	0,604	0,329	-0,506	-0,625	0,387	-1,420	0,854	-0,002	0,007	-0,323	
X9	0,846	0,811	1,347	-0,793	-0,958	1,437	-1,911	1,008	0,095	0,121	-0,506	
X10	0,084	0,076	0,129	-0,213	-0,172	0,097	0,004	0,230	-0,195	-0,208	-0,175	
X11	0,495	0,585	0,884	-0,764	-0,831	0,907	-1,106	0,798	-0,063	-0,116	-0,493	
X12	0,080	0,101	0,507	-0,134	-0,150	0,470	0,350	0,236	0,152	0,125	0,025	
X13	0,647	0,581	1,376	-0,604	-0,716	1,358	-1,378	0,840	0,058	0,061	-0,450	
X14	0,835	0,920	0,981	-1,111	-1,181	1,042	-1,789	1,202	-0,054	-0,057	-0,723	
X15	-0,206	-0,251	-0,094	0,059	0,059	-0,090	-0,258	-0,384	-0,090	-0,025	0,013	
X16	-0,081	-0,071	0,984	-0,106	-0,078	0,903	0,608	-0,046	0,005	0,050	-0,016	
X17	0,720	0,700	1,023	-0,702	-0,762	1,052	-1,705	0,887	0,094	0,171	-0,405	
Bias	0,004	-0,029	-0,077	-0,064	-0,024	-0,059	0,249	0,020	-0,187	-0,157	-0,086	
Çıktı katmanı (Mali Tabloları Hıfeli) Sigmoid	-1,096	-1,044	-1,727	1,311	1,428	-1,805	2,504	-1,477	0,050	0,011	0,881	0,028
Çıktı katmanı (Mali Tabloları Hıfeli Olmayan) Sigmoid	1,055	1,098	1,763	-1,301	-1,466	1,745	-2,540	1,474	-0,065	-0,054	-0,813	-0,005

Ek3. Bu Çalışmada Kullanılan Değişkenleri Çalışmalarında Kullanan Araştırmacılar

No	Değişken	Değişkenleri Çalışmalarında Kullanan Araştırmacılar
X1	Stoklar/ Aktif Toplam	Kirkos, Spathis ve Manolopoulos (2007); Omid, Min, Moradinaftchali ve Piri (2019)
X2	Maddi Duran Varlık/Aktif Toplam	Chen (2016), Omid, Min, Moradinaftchali ve Piri (2019)
X3	Toplam Borç/Aktif Toplam	Chen (2016); Jan (2018); Kirkos, Spathis ve Manolopoulos (2007); Tangod ve Kulkarni (2015); Küçükkocaoğlu ve Küçüksözen (2005)
X4	Net Kâr /Aktif Toplam	Chen (2016); Jan (2018); Kirkos, Spathis ve Manolopoulos (2007); Omid, Min, Moradinaftchali ve Piri (2019); Tangod ve Kulkarni (2015); Küçükkocaoğlu ve Küçüksözen (2005)
X5	Dönen Varlıklar/Kısa Vadeli Yabancı Kaynaklar	Chen (2016); Omid, Min, Moradinaftchali ve Piri, 2019
X6	Brüt Kâr/Net Satışlar	Chen (2016), Jan (2018); Kirkos, Spathis ve Manolopoulos (2007)
X7	Net Satışlar/Aktif Toplamı	Chen (2016); Tangod ve Kulkarni (2015); Küçükkocaoğlu ve Küçüksözen (2005)
X8	Brüt Kâr / Aktif Toplam	Omid, Min, Moradinaftchali ve Piri (2019); Küçükkocaoğlu ve Küçüksözen (2005)
X9	Satılan Malın Maliyeti/Ortalama Stoklar	Chen (2016); Jan (2018); Kirkos, Spathis ve Manolopoulos (2007); Tangod ve Kulkarni (2015)
X10	Vergi öncesi kar / Net satışlar	Chen (2016); Jan (2018); Kirkos, Spathis ve Manolopoulos (2007)
X11	Vergi, Faiz ve Amortisman Öncesi Kâr/Toplam Varlıklar	Chen (2016); Tangod ve Kulkarni (2015)
X12	Dört büyük denetim şirket tarafından denetlenip denetlenmediği (Denetlenmiyorsa 0, denetleniyorsa 1 ile kodlanmıştır).	Yazar tarafından eklenmiştir.
X13	Yönetim Kurulu Üye Sayısı	Chen (2016)
X14	Yönetim Kurulunda Bağımsız Üye Sayısı	Yazar tarafından eklenmiştir.
X15	Halka Açıklık Oranı %	Yazar tarafından eklenmiştir.
X16	Halka Açıklık Süresi (Yıl)	Yazar tarafından eklenmiştir.
X17	Halka Açıklıkta Yabancı Payı (Sermayede Doğrudan %5 veya Daha Fazla Paya veya Oy Hakkına Sahip Gerçek ve Tüzel Kişiler)	Yazar tarafından eklenmiştir.