
FİNANSAL PİYASALARDA PAY SENEDİ MANİPÜLASYONUNUN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE TAHMİNİ: BORSA İSTANBUL ÖRNEĞİ¹

Bariş AKSOY²

Öz

Bu çalışmada Borsa İstanbul İmalat Sanayi sektöründe faaliyet gösteren 51 manipüle edilmiş ve 51 manipüle edilmemiş toplam 102 işletmeye ait pay senedinin 3-9 ay öncesinden piyasa manipülasyonuna maruz kalıp kalmama durumu tahmin edilmiştir. İlgili şirketlerin mali tablo ve nitel verileri elde edilerek sınıflandırma için Yapay Sinir Ağları (ANN), Sınıflandırma ve Regresyon Ağacı (CART), C5.0 Karar Kuralı Türetme Algoritması ve Lojistik Regresyon (LR) yöntemleri kullanılmıştır. Genel sınıflandırma doğruluğu sırasıyla ANN (%86,67), CART (%80,00), C5.0 (%76,67) ve LR (%70,00) olarak bulunmuştur. C5.0 algoritması test örneğinde yer alan 15 manipüle edilen pay senetlerinin 14'ünü doğru tahmin ederek %93,33 sınıflandırma başarısı göstermiştir. ANN test örneğinde bulunan 15 manipüle edilmemiş pay senedinin 14'ünü doğru tahmin ederek %93,33 sınıflandırma başarısı göstermiştir. Bu çalışmada tüm yöntemler %70'in üzerinde tahmin performansı gösterdiğinden, ilgililerin yapacakları pay senedi manipülasyon tahmini analizlerine bu modeller de dahil edilebilecektir.

Anahtar Kelimeler: Menkul Kıymet Borsaları, Manipülasyon Tahmini, Piyasa Dolandırıcılığı Tahmini, Veri Madenciliği, Borsa İstanbul

JEL Sınıflandırması: C38, C45, C51, C52, D53

FORECASTING STOCK PRICE MANIPULATION IN FINANCIAL MARKETS USING DATA MINING METHODS: THE CASE OF BORSA ISTANBUL

Abstract

In this study, 102 businesses, 51 manipulated and 51 non-manipulated, operating in the Borsa İstanbul Manufacturing Industry have been predicted 3-9 months before whether their stock price will be manipulated. Artificial Neural Networks (ANN), Classification and Regression Tree (CART), C5.0 decision tree algorithm and Logistic Regression (LR) methods were used for classification by obtaining financial statements and qualitative data of related businesses. Overall classification accuracy was found as ANN (86.67%), CART (80.00%), C5.0 (76.67%) and LR (70.00%), respectively. The C5.0 algorithm predicted 14 of the 15 manipulated stocks and achieved 93.33% classification accuracy. ANN predicted 14 of the 15 non-manipulated stocks and achieved 93.33% classification accuracy. In this study, since all methods have over 70% prediction performance, related models can also be included in stock manipulation prediction analyzes.

Keywords: Stock Exchanges, Manipulation Prediction, Market Fraud Crime, Data Mining Methods, Borsa İstanbul

JEL Classification: C38, C45, C51, C52, D53

¹ Bu çalışma 23. Finans Sempozyumunda sözlü olarak sunulan "Pay Senedi Manipülasyonunun Makine Öğrenmesi Yöntemleri Kullanılarak Tahmin Edilmesi: Borsa İstanbul Örneği" adlı tebliğde alınan eleştiri ve katkılar doğrultusunda genişletilmiş ve son şekli verilen halidir.

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

DOI: 10.18092/ulikidince.582919

Makalenin Geliş Tarihi (Received Date): 26-06-2019

Yayına Kabul Tarihi (Acceptance Date): 11-05-2020

1. Giriş

Ekonomik suçlar çoğu ülkede en tehlikeli yasadışı eylemleri oluşturmaktadır. Devletin piyasada işlem yapanlara eşit muamele ilkesi nedeniyle yatırımcılar, kendi bilgisine ve çalışmasına bağlı olarak gelir elde edebileceği bir piyasaya kavuşmuş olurlar. Ekonomik suçların işlenmesiyle bu tarafsız piyasa koşullarına saldırılmakta, devlet ile vatandaşları arasındaki temel düzen piyasa bozucu davranışlar ve manipülasyonlar ile baltalanmaktadır (Tellechea, 2008:214). Piyasayı düzenleme ve denetlemekle görevli kurumların gerekli düzenleme ve denetlemeyi yapmadığı bir piyasada diğer bireylerin zarar etmesi ve piyasadan çekilmesi, piyasa bozucu faaliyette bulunan azınlığın haksız kazancını artırırken nihai olarak piyasaların kendisinden beklenen işlevleri yerine getirememesine neden olmaktadır.

Manipülasyon, gelişmiş ve gelişmekte olan borsalar için önemli bir konudur. Pay senetlerinin fiyatları piyasa tarafından herhangi bir müdahale olmadan belirlenmelidir. Yatırımcılar tüm bilgileri göz önünde bulundurarak pay senetlerine değer verdikleri için, fiyatlama sürecini bozan herhangi bir etken olmadan belirlenen piyasa fiyatları, yatırımcıların pay senetlerinin değeri hakkındaki ortak değerlendirmesini yansıtmaktadır. Bu şekilde üzerinde hemfikir olunan fiyatın adil olduğu kabul edilmektedir. Pay senedi fiyatlarındaki en önemli sapma manipülasyondan kaynaklanmaktadır. Bu bağlamda pay senedi manipülasyonu makul bir fiyatın geçerli olmasını önleyen piyasa mekanizmasına çeşitli yollarla müdahale ederek pay senedinin fiyatını yapay olarak etkilemeye yönelik eylem olarak tanımlanabilir (Öğüt vd. 2009:11944).

Manipülatörlerin yasadışı kâr elde etmelerinden dolayı çok sayıda yatırımcı, menkul kıymetlere yatırdıkları fonları kaybetmektedirler. Bu nedenle manipülasyon eylemini gerçekleşmeden önce veya gerçekleşikten sonra tespit etmek çok önemlidir. Menkul kıymet piyasa manipülasyonunun tespitinde veri madenciliği teknikleri konusunda araştırmalar artmaktadır. Çok sayıda bilim insanının piyasa manipülasyonunu tespit etme çalışması teorik nitelikte kalmakta ve manipülasyon kalıplarını tanımlamaktan ibarettir. Bu da menkul kıymet piyasa manipülasyonunu doğru ve hızlı bir şekilde saptamanın hâlâ güç olduğunu ve konunun çalışılmaya ne kadar muhtaç olduğunu göstermektedir (Li vd. 2017:788-789). Düzenleyiciler açısından manipülasyon eylemi gerçekleşmeden önce manipülasyon olup olmayacağını tespit etmek daha da zordur (İmişiker ve Taş, 2013:127-128).

Bu çalışmanın gerek piyasa düzenleyici ve denetleyici kurumlar gerekse manipülasyona maruz kalması muhtemel işletme yöneticileri ve yatırımcıların yeni bakış açısına sahip olabilecekleri bazı sonuçları bulunmaktadır. Çalışmada manipülasyona uğramayan işletmelerle karşılaştırıldığında bazı işletmelerin piyasa manipülasyonuna daha duyarlı hale gelmesine neden olan firmaya özgü faktörler araştırılmaktadır. Bu çalışmada Sermaye Piyasası Kurulu'nun manipülasyon yapıldığını tespit ettiği işletmelere ait mali tablo verileri ve mali tablolara dayalı olmayan veriler elde edilerek aynı sayıda manipülasyona uğramayan işletmelere ait veriler kullanılarak manipülasyona neden olan faktörleri belirlemek amaçlanmıştır. Böylece manipüle edilmeye neden olan faktörler manipülasyon tarihinden 3-9 ay öncesine kadar belirlenerek işletme yöneticileri, ilgili menkul kıymetin alınıp satıldığı piyasa, piyasa düzenleyici ve denetleyici kurumlar, politika yapıcılar, ilgili kamu otoriteleri ve yatırımcılar yararına manipülasyon daha yapılmadan gerekli önlemin alınması sağlanabilir. Piyasa düzenleyici ve denetleyiciler, borsalar halka açık şirketlerin kotasyon standartlarına karar vermektedir. Bu çalışmada belirtilen faktörleri kullanarak, manipülasyon yapma olasılıklarına göre ilgili şirketlere ait pay senetlerini farklı bölümlere ayırabilirler. Sürekli açık artırma yerine çağrı açık artırmanın kullanılması gibi pazarın bazı bölümleri için farklı alım satım kuralları uygulanabilir.

Piyasa manipülasyonu tahmin çalışmasında, mali tablo verilerinin yanı sıra nitel değişkenlerin kullanılması, çalışmada kullanılan parametrelerin seçimi, parametre optimizasyonu ile en doğru sonuçları veren parametrelerin bulunması, mali tablo verilerinden elde edilen bağımsız değişkenlerin yanı sıra Borsa İstanbul şirket haber ve duyuruları, bilanço dipnotlarından elde edilen nitel bağımsız değişkenlerin kullanılmasının modelin tahmin başarısında önemli olduğu

görülmektedir. Çalışmada tüm veri seti %70 eğitim %30 test seti olarak ayrıldıktan sonra test verileri eğitim esnasında modelin oluşturulmasında algoritmaya gösterilmemektedir. Eğitim verisi ile değişkenlere ağırlıklar atandıktan ve en iyi model parametre optimizasyonu ile belirlendikten sonra modelin daha önce hiç görmediği test verisi algoritmaya gösterilmekte ve kurulan modelin tahmin performansı test edilmektedir. Bu yöntem oluşturulan modelde algoritmanın verileri ezberlemesinin önüne geçmekte ve daha güvenilir sonuç elde edilmesini sağlamaktadır. Pay senetleri manipüle edilen ve manipüle edilmeyen işletmelerin 3-9 ay öncesinden tahmin edilmesine ilişkin mali tablo verileri ve mali tablolara dayanmayan nitel değişkenlerin alındığı bu çalışmada veri seti %70 eğitim ve %30 test verisi olarak ayrıldıktan sonra doğrulama türü olarak 10 katlı çapraz doğrulama tekniği kullanılarak elde edilen ANN ve CART yöntemlerine ait modeller %80'in üzerinde başarılı tahmin performansı göstermişlerdir. Veri seti %70 eğitim ve %30 test verisi olarak ayrılarak yürütülen C5.0 ve LR yöntemlerine ait modeller %70 ve üzeri başarılı tahmin performansı göstermişlerdir. Bu sonuçlar piyasa manipülasyonu tahmini ile ilgilenen kişilerin manipülasyon tahmininde kullandıkları diğer yöntemlerin arasında bu çalışmada kurulan modelleri de kullanılabileceklerini göstermektedir. Bu nedenle çalışmanın literatüre değer katabileceği düşünülmektedir. Çalışmada birinci bölümde giriş, ikinci bölümde manipülasyon (piyasa dolandırıcılığı) suçu, üçüncü bölümde literatür incelemesi, dördüncü bölümde Borsa İstanbul'da pay senedi manipülasyonunun veri madenciliği yöntemleriyle tahmini ve beşinci bölümde sonuç ele alınmıştır.

2. Piyasa Dolandırıcılığı (Manipülasyon) Suçu

İyi işleyen bir sermaye piyasasının temelinde güven ve fırsat eşitliği ilkeleri yer almaktadır. Ancak bazen piyasalarda bu ilkelere ters düşen işlem ve davranışlar olduğu gözlemlenmektedir. Bazı yatırımcıların rasyonel olmayan davranışlarla pay senedi fiyatlarında dayanağı olmayan, beklenmedik dalgalanmalar yaratarak yatırımcıları yanlış yönlendirmeleri ve bundan haksız kazanç elde etmeye çalışmaları manipülasyon olarak nitelendirilmektedir (Şensoy, 2013:371). Genel olarak piyasa manipülasyonu, menkul kıymet fiyatı ile ilgili yapay, yanlış veya yanıltıcı görünümler yaratabilen piyasanın serbest ve dürüst işleyişine müdahalede bulunma girişimi olarak tanımlanabilir (Li vd., 2017:788).

Pay senedi fiyatları üzerinde gerçekleştirilen manipülasyon eylemleri, azalan işlem hacmi sonucunda fiyat keşfini bozmakta ve fiyatları doğal seviyelerinden aşağı veya yukarı yönde oluşmasını sağlamaktadır. Bunun piyasada fiyatların tahsis rolü üzerinde zararlı bir etkisi vardır. Uzun vadede piyasa manipülasyonu yatırımcıların piyasaya katılımını engellemekte ve yatırımcıların alternatif pazarlarda işlem yapmalarına neden olmaktadır. Bunun menkul kıymet piyasalarında likidite üzerine olumsuz etkisi vardır ve aynı zamanda piyasadaki diğer yatırımcıların işlem maliyetini arttırmaktadır. Sonuç olarak manipülasyon sermaye maliyetini artırma potansiyeline sahiptir ve bu da şirketlerin sermaye piyasalarında paylarını halka arz etmekte daha isteksiz olmalarına neden olmaktadır (Comerton-Forde ve Putnins, 2011:136).

Bilgiye dayalı manipülasyonda manipülatörler, pay senedi fiyatlarını etkilemek için söylenti ve yanlış bilgi yaymaktadırlar. İşleme dayalı manipülasyonda ise manipülatörler, aktif bir pazar görüntüsü oluşturmak amacıyla ilgili pay senedi için sahte alım satım gerçekleştirirler. Manipülatörler, diğer yatırımcıları pay senedi almak üzere çekmek için bir pay senedinin fiyatını ve hacmini yapay olarak artırma şeklinde kendi aralarında alım-satım yaparlar. Bilgiye dayalı manipülasyon ve işleme dayalı manipülasyon günümüzde en yaygın gerçekleşen manipülasyon türleridir. İçeriden öğrenenlerin ticareti ve finansal bilgi manipülasyonu, manipülasyon olarak kabul edilmekte ve yasa dışı işlem olarak değerlendirilmektedir. İçeriden öğrenenlerin ticaretinde ihraççının çalışanları ve ilişkili olduğu diğer kişi/kurumlar henüz halka açıklanmayan özel bilgileri kullanarak kişisel çıkar elde etmektedir. Finansal bilgi manipülasyonunda ise firmanın yöneticileri, işletmenin finansal durumu ve firmanın performansı hakkında yanlış bir izlenim vermek için finansal tablolarda sunulan bilgileri bozmaktadır. Manipülatif işlemler her borsada yakından takip

edilmekte ve tespit edildiğinde manipülatörlere belirli cezalar verilmektedir (Öğüt vd. 2009:11944-11945).

Borsanın “kumar” ve “tüyo” ifadeleri ile özdeşleştirildiği Türkiye’de tasarruf sahiplerinin piyasaya olumlu olarak bakması sağlanmalıdır. Bu olumsuz algının değişmesi yatırımcının korunması ilkesi çerçevesinde piyasada işlenecek suçların önlenmesi, işlenen suçların ise caydırıcı ve etkili bir yaptırım sistemi ile cezalandırılması gerekir. Finansal piyasalarda piyasa bozucu eylemleri-manipülasyonları teşhis etmek ve bunlara yaptırım uygulamak için düzenleme/denetleme mekanizmasının sağlıklı bir şekilde işlemesi ve konusunda uzman hızlı bir yargı süreci gereklidir (Gündoğdu, 2017:23). Sermaye Piyasası Kurulu Türk sermaye piyasasında işlenebilecek suçları önleyebilmek amacı ile 21 Ocak 2014 tarih ve 28889 sayılı Resmi Gazetede yayımlanarak yürürlüğe giren “Bilgi Suiistimali veya Piyasa Dolandırıcılığı Suçları Hakkında Bildirim Yükümlülüğü Tebliği (V-102.1)” yayınlamıştır (Gündoğdu, 2017:15).

6362 sayılı SPKn’nun 107/1 maddesinde yer alan piyasa dolandırıcılığı suçunun işlenmiş sayılması için sermaye piyasası araçlarının fiyatlarına, arz ve taleplerine ilişkin olarak yanlış veya yanıltıcı izlenim uyandırmak için alım veya satım yapılması, emir verilmesi, emir iptal edilmesi, emir değiştirilmesi gerekmektedir. Suç unsurlarının belirlendiği durumlarda, ikinci aşama yapılan işlemlerde piyasa dolandırıcılığı suçunun işlenmesine yönelik kastın meydana gelip gelmediği incelenmektedir. Bu nedenle yalnızca suça yönelik kastın bulunduğu durumlarda Cumhuriyet Başsavcılığına suç duyurusunda bulunulabilmektedir (Kesici ve Nacakçı, 2014:6-7). SPKn’nun 107/1 hükmünde belirtilen eylemi bir “tehlike suçu” olarak düzenleyerek sonucun gerçekleşmesini zorunlu tutmamıştır. SPKn’nun 107/2 maddesinde bilgiye dayalı piyasa dolandırıcılığı sayılan eylemler “sermaye piyasası araçlarının fiyatlarını yatırımcıların kararlarını etkilemek amacıyla yalan, yanlış veya yanıltıcı bilgi verilmesi, söylenti çıkarılması, haber verilmesi, yorum yapılması veya rapor hazırlanması ya da bunların yayılması” şeklinde sayılmıştır. Bu hükme göre maddi unsurun oluşabilmesi için bilgi, söylenti, haber, yorum ve raporun yalan, yanlış veya yanıltıcı olması gerekmektedir. Diğer taraftan işlemin büyüklüğü konusunda alt ve üst sınır söz konusu değildir. İlgili payın piyasasında kontrolü elde ederek fiyat hakimiyeti sağlamak için gerekli miktar fazla olabileceği gibi, payın kapanış fiyatının belirlenmesi için bir adetlik alış işlemleri de piyasanın dolandırılmasına neden olabilecek niteliktedir (Çalışkan, 2015:383-384).

Piyasa dolandırıcılığı suçu farklı şekillerde işlenebilir. Uluslararası Menkul Kıymetler Komisyonları Örgütü (IOSCO) (2000), manipülatörler tarafından yaygın olarak kullanılan manipülasyon yöntemlerini belirlemiştir (Gemici vd., 2017:370). IOSCO’nun bulguları dünyadaki birçok sermaye piyasası düzenleyicisinin ortak çabalarından oluşmaktadır (İmişiker ve Taş, 2013:120). Manipülasyonun engellenmesinde 2009 yılından önce Borsa İstanbul (BIST) ve Vadeli İşlem ve Opsiyon Borsası (VİOP) manipülatif işlemlerin gözetim ve denetimini yürütmekte, SPK da bu kurumların gözetim ve denetimini yürütmekteydi. 2009 yılından sonra SPK nezdinde “Piyasa Gözetim ve Denetim Dairesi” kurulmuş ve gözetim denetim faaliyetleri tek çatı altında toplansa da BIST ve VİOP ilgili denetim birimleri görevlerini eskisi gibi yapmaya devam etmektedir (Çalışkan, 2015:386). SPK’nın yaptığı denetimler sonucu pay senedi manipülasyonu sonucunda işlem yasakları ve suç duyuruları yıllardan beri verilmekte ancak manipülasyonla mücadelede fazla etkili bir yöntem olmadığı görülmektedir. Borsa İstanbul’da işlem yapan yatırımcı, düzenleyici ve denetleyici kurumların manipülasyon sonucu uğrayabileceği zarar ortaya çıkmadan önce tedbir alması gerekmektedir. Bu bağlamda her bir yatırımcı ve ilgili denetleyici kurumların kolayca ulaşılabilmesi mali oran ve az sayıda değişken ile manipülasyon sonucu oluşacak zararlara yönelik tedbir alınması mümkün olabilecektir (Çalışkan, 2015: 390).

3. Literatür İncelemesi

Borsa İstanbul gelişmekte olan bir borsa olduğundan manipülasyon işlemlerini makul süre önceden tespit etmek ve önlemek önemli bir araştırma konusudur. Manipülasyonla ilgili literatürde teorik ve ampirik çalışmalar olmasına rağmen, çok nadir olarak manipülasyonu öngörmek veya saptamak için model geliştirme üzerine odaklanan çalışmalar bulunmaktadır

(Öğüt vd., 2009:11945). Literatüre bakıldığında manipülasyon konusunun önemine rağmen ampirik yapılan çalışmaların yetersiz kaldığı görülmektedir. Bunun ana sebebi konuyu ampirik olarak incelemek için gerekli olan işlemlere ait verilerin gizli bilgi olarak görülmesi ve akademisyenlerin bu verilere sınırlı olarak ulaşabilmeleridir. Kısıtlı veri kullanılarak yapılan çalışmaların farklı manipülasyon türlerinin varlığını tespit ederek bunların önlenmesi ve piyasaların daha güvenilir olması için yeni önerilerin getirildiği görülmektedir (Özcan, 2012:20).

Öğüt vd. (2009) çalışmalarında Borsa İstanbul'a kayıtlı pay senetleri ile endeksin ortalama günlük getirisi, işlem hacmindeki ortalama günlük değişim ve ortalama günlük volatilité arasındaki farkı açıklayıcı değişken olarak kullanmışlardır. Hedeflenen pay senetlerinin fiyatını, hacmini ve volatilitésini gözlemleyerek olası pay senedi fiyat manipülasyonlarını tespit etmek için modeller geliştirmeye çalışmışlardır. Manipülasyon sonrası ve manipülasyon öncesi dönemlerdeki veriler manipüle edilmemiş pay senetleri olarak kullanılırken manipülasyon dönemindeki verileri manipüle edilmiş pay senetleri olarak almışlardır. Yapay Sinir Ağları (ANN), Destek Vektör Makinesi (SVM), Diskriminant ve Lojistik Regresyon (Logit) yöntemlerini sınıflandırma doğruluğu, duyarlılık ve özgüllük istatistiklerinin test performansı yönünden karşılaştırmışlardır. Analizlerinde veri madenciliği tekniklerinin (ANN ve SVM) pay senedi fiyat manipülasyonunu tespit etmek için toplam sınıflandırma doğruluğu ve hassasiyet istatistikleri bakımından performansları çok değişkenli istatistiksel tekniklerden (diskriminant, lojistik regresyon) daha iyi olduğu bulgusuna ulaşmışlardır.

Diaz vd. (2011) çalışmasında manipülasyonların incelenmesi için kapanış fiyatlarının yanı sıra gün içi işlem fiyatlarının manipüle edilip edilmediğini tespit etmeye çalışmışlardır. Analiz için karar ağaçlarına dayalı QUEST, C5.0 ve CART gibi farklı denetimli sınıflandırma algoritmaları kullanmışlardır. Ampirik bulgulara dayalı olarak likiditesi düşük olan pay senetlerinin manipüle edilmesinin muhtemel olduğunu ve manipülasyonların pay senedi fiyatında volatilitéyi arttırdığını belirlemişlerdir. Ayrıca, günlük pay senedi fiyatlarının manipülasyon süresi boyunca arttığını ve ardından manipülasyon sonrası dönemde düştüğünü, manipülatörün sattığı zaman fiyatların ve likiditenin, manipülatörün satın aldığı zamandan daha yüksek olduğu bulgusuna ulaşmışlardır. Tüm pay senetleri için piyasanın kapanma saatlerinde likiditenin daha yüksek, ancak bu etkinin manipüle edilen pay senedinde daha fazla olduğunu belirtmişlerdir.

Comerton-Forde ve Putnins (2011) çalışmalarında kapanış fiyatı manipülasyonunun ABD ve Kanada borsalarında işlem gören pay senetlerinin alım-satım özellikleri üzerindeki etkilerini araştırmışlardır. Ayrıca kapanış fiyatının manipüle edilme olasılığını ölçmek için bir endeks geliştirmişlerdir. Manipülasyon endeksi, hali hazırda mevcut olan alım satım ve teklif verileri kullanılarak, herhangi bir pay senedinin kapanış fiyatının manipüle edilme olasılığını tahmin etmektedir. Önceki çalışmaların aksine, kapanış fiyatı manipülasyonunun etkisini ilişkisiz gün sonu ve mevsimsel etkilerden ayırmışlardır. Çalışmada manipülasyonun fiyat doğruluğu üzerinde önemli derecede zararlı bir etkisi olduğunu belirlemişlerdir. Fiyat bozulmaları genellikle sadece kapanıştan kısa bir süre önce mevcut olsa da, kapanış fiyatlarının yaygın bir şekilde kullanılması nedeniyle etkilerinin çok önemli olduğu sonucuna ulaşmışlardır. Muhtemel manipülasyon örneklerini tanımlayarak bu endeksin, kovuşturma verilerinin hazır olmadığı çok sayıda piyasada ve zaman dilimlerinde manipülasyonu incelemek için kullanılabileceğini belirtmişlerdir.

Golmohammadi ve Zaiane (2012) çalışmalarında Bangladeş borsasında menkul kıymet dolandırıcılığının sık meydana geldiğini belirtmiş ve Bangladeş menkul kıymetler piyasasında dolandırıcılık tespitini incelemişlerdir. Veri seti üç aylık verilerden oluşmaktadır. Veriler, sabit zaman dilimlerinde alım satım emirlerinin istatistiksel değişkenleri (ortalama ve varyans) kullanılarak belirlenmiştir. Daha sonra, aynı grup üyelerinin grup merkezine olan mesafesi t-istatistiği kullanılarak hesaplanmış ve gruptan önemli ölçüde sapan örnekleri almışlardır. Bu pay senetleri ile ilişkilendirilen yatırımcı hesaplarını, ani olarak diğer işlem yapanlardan farklı davranan şüpheli yatırımcılar olarak belirlemişlerdir. Sonuç olarak veri madenciliği yöntemlerinin,

menkul kıymetler piyasasında farklı dolandırıcılık türlerini tespit etmede etkili olduğu sonucuna varmışlardır.

İmişiker ve Taş (2013) çalışmalarında Borsa İstanbul'da (BIST, İMKB) 1996 ve 2006 yılları arasında manipülasyon vakalarından oluşan bir veri seti elde edilmiş ve bu manipülasyonları açıklamak için firmaya özgü değişkenler kullanılmıştır. Probit regresyon sonuçlarına göre küçük firmaların, düşük Nakit Akım Oranı ve yüksek Kaldıraç Oranına sahip olan firmaların pay senedi fiyat manipülasyonuna maruz kalmaya daha yatkın olduğu bulgusuna ulaşmışlardır. Dinamik probit analizi ile daha önce manipüle edilmiş pay senetleri için ilgili pay senedinin yeniden manipülasyona uğrama olasılığının yüksek olduğu sonucuna varmışlardır. Çalışmada nispeten büyük firmaların manipüle edilme ihtimalinin daha düşük olduğu ve halka açıklık oranı yüksek olan firmaların daha az manipüle edilmiş olduğunu tespit etmişlerdir.

Li vd. (2017) çalışmalarında Çin Menkul Kıymetler Düzenleme Komisyonu (CSRC) tarafından yayınlanan manipüle edilmiş bilgilere dayanarak, günlük manipülasyonu tespit etmek ve manipüle edilmiş pay senetlerinden alım-satım verilerini belirlemek için denetimli ve denetimsiz makine öğrenme modelleri kullanmışlardır. CSRC web sitesinde yayınlanan 64 adet manipüle edilmiş hissenin günlük pay senedi fiyat verileri ve manipüle edilmemiş pay senedi fiyat verileri seçilmiştir. Model oluşturmak için denetimli makine öğrenimi modellerinden en yakın k komşuları (k Nearest Neighbor, KNN), karar ağacı sınıflandırıcısını (Decision Tree Classifier, DTC), Yapay Sinir Ağları (Artificial Neural Networks, ANN) ve Destek Vektör Makinesi (Support Vector Machine, SVM) kullanmışlardır. İstatistiksel yöntemlerden ise Doğrusal Diskriminant Analizi (Linear Discriminant Analysis, LDA), Kuadratik Diskriminant Analizi (Quadratic Discriminant Analysis, QDA), Lojistik Regresyon (Logistic Regression, LR) analizlerini gerçekleştirmişlerdir. Denetimli makine öğrenme modelleri, günlük verilerden manipülasyonları tespit etmek için mükemmel bir performans sergilerken, istatistiksel yöntemlerin alım-satım verileri üzerindeki manipülasyon tespit performansı zayıf bulunmuştur. Kullanılan makine öğrenme modelleri arasında, KNN ve DTC doğruluk, hassasiyet, özgüllük ve AUC yönlerinden tüm yöntemlerle kıyaslandığında %99,91 ile en iyi tahmin performansı gösterdiği bulunmuştur. Yöntemlerin tahmin performansı en yüksekte düşüğe doğru sırasıyla KNN (%99,91)=DTC (%99,91)>ANN (%98,00)>QDA (%93,99)>SVM (%91,81)>LDA (%90,46)>LR (%89,11) olarak bulunmuştur.

Gemici vd. (2017) çalışmalarında 2001-2014 döneminde Borsa İstanbul'a (BIST) kayıtlı olan ve manipülasyona maruz kaldığı SPK tarafından tespit edilen 273 manipülasyon durumunu SPK haftalık bültenlerden tek tek incelemişlerdir. Manipülasyona yönelik günlük olarak getiri, işlem hacmi, volatilité ve pay senedi devir hızı değişkenlerini kullanmışlardır. Manipüle edilmiş aynı işletmelere ait manipülasyon dönemi bir sınıfa ayrılmış, manipülasyon öncesi ve manipülasyon sonrası ise ayrı birer sınıfa ayrılmış olarak üçlü bağımlı değişken sınıfları oluşturulmuştur. Çoklu lojistik regresyon analizi sonucunda günlük getiri ve volatilité değişkeninin manipülasyon üzerinde daha fazla etkiye sahip olduğunu bulmuşlardır. Elde edilen bulgulara göre manipülasyon döneminde pay senetlerinin getirisinin, volatilitésinin, işlem hacminin ve devir hızının manipülasyon öncesi döneme göre belirgin bir şekilde artış göstermesi literatürde daha önce yapılmış çalışmaları destekler nitelikte olduğu belirtilmiştir. Özellikle volatilitéyi, manipülatif eylemlerin etkisiyle bir şirketin manipülasyon dönemine girme olasılığını etkileyen en önemli değişken olarak bulmuşlardır.

Yukarıda incelendiği gibi, menkul kıymet borsalarında pay senedi manipülasyonu tespit literatürü, manipülasyon süresince ilgili payın fiyat hareketleri, getiri, işlem hacmi ve volatilité incelemesi ile gerçekleştirilmektedir. Ancak manipülasyonu belirli bir süre öncesinden tespit edebilecek ve manipülasyona uğrayan işletmelerden kaynaklı faktörler üzerine yoğunlaşan bir çalışmaya İmişiker ve Taş (2013) çalışması dışında rastlanmamıştır. Bu çalışmanın literatüre olan ilk katkısı; mali tablo verilerinden oluşan nicel değişkenlerin yanında nitel değişkenlerin kullanılması sonucunda bir pay senedinin manipülasyona uğrayıp uğramama durumunu belirleyen faktörleri ampirik olarak analiz etmesidir. Çalışmanın literatüre ikinci katkısı; pay senedi

manipülasyonuna uğramış şirketler SPK haftalık bültenlerinden elde edilen örneklerden oluşmuş ve eşleşmesi için pay senedi manipüle edilen her bir işletmeyle aynı sektörde ve aynı sayıda manipülasyona uğramamış işletmeler tesadüfi olarak seçilmiştir. Belirlenen nitel ve nicel değişkenler kullanılarak makine öğrenmesi yöntemlerinden ANN, CART, C5.0 ve LR modellerinin tahmin performanslarının karşılaştırılması ile ilgili bir çalışmaya literatür incelemesinde rastlanmamıştır. Literatüre olan üçüncü katkı ise kullanılan doğrulama yönteminin sınıflandırma performansı üzerindeki etkisinin incelenmesi amacıyla veri seti %70 eğitim ve %30 test seti olarak ayrıldıktan sonra test verileri algoritmaya gösterilmemiştir. Eğitim verisi ile analiz gerçekleştirilerek değişkenlere ağırlıklar atanmaktadır. En iyi model, parametre optimizasyonu ile belirlendikten sonra modelin daha önce hiç görmediği test verisi algoritmaya gösterilmekte ve kurulan modelin tahmin performansı test edilmektedir. Bu yöntemle oluşturulan modelde algoritmanın verileri ezberlemesinin önüne geçilmekte ve daha güvenilir sonuç elde edilmesi sağlanmaktadır. Bu nedenle çalışmanın literatüre değer katabileceği düşünülmektedir.

4. Borsa İstanbul'da Pay Senedi Manipülasyonunun Veri Madenciliği Yöntemleriyle Tahmini

Düzenleyici kurumlar, finansal suiistimali ortadan kaldırmak ve piyasaların bütünlüğünü korumak amacıyla teknolojik çözümlere yönelmiştir. Hesaplama gücü, hız ve depolama kapasitesindeki gelişmelerin yanı sıra, yapay zeka ve veri görselleştirme araçları gibi yeni analitik yöntemlerin kullanımıyla düzenleyiciler artık piyasanın derinliklerinde daha fazla araştırma yapabilmektedirler. Bu teknolojilerin spesifik uygulamaları sayesinde piyasa duyarlılığındaki değişiklikler izlenebilmekte, içeriden öğrenenlerin ticareti ve manipülatif alım ve satım işlemleri araştırılmaktadır (Williams, 2013:544).

4.1. Çalışmanın Kısıtları ve Veri Seti

Pay senedi manipülasyonunun gerçekleştiği yıldan bir yıl öncesi verileriyle çalışılması, manipülasyon işlemlerinin dikkate alınmaması sonucunu doğurabilmekte ve bu da potansiyel sonuçları çarpıtabilmektedir (Diaz vd., 2011:12768). Bu çalışmada ana kütleli oluşturan işletmeler farklı yıllarda ve yılın farklı aylarında piyasa dolandırıcılığına (manipülasyon) maruz kalmışlardır. Finansal piyasalarda oluşan hareketlerin hızla değişmesi, piyasa dolandırıcılarının işlemlerinde etkili olan faktörleri incelememizi sınırlayabilmektedir. Böyle bir tehlikeden dolayı Diaz vd. (2011) çalışması takip edilerek manipülasyon tarihinden önceki en yakın mali tabloların kullanılması yoluna gidilmiştir. Pay senedi manipülasyonuna maruz kalan bazı işletmelerin 3'er aylık mali tablolarına düzenli olarak ulaşılamadığından manipülasyon tarihinden önceki 9 aya kadar olan mali tablolar alınmıştır. Bu nedenle manipülasyondan önceki 3 aydan 9 aya kadar olan ve pay senetleri BIST'de işlem gören işletmeler tarafından Kamuyu Aydınlatma Platformuna (KAP) gönderilmiş ara dönem mali tablolar çalışmada kullanılmıştır.

Çalışmada manipülasyona maruz kalan ve kalmayan işletmelerin mali tablo verileri kullanıldığından farklı sektörlerde ait mali tablo verilerinin kullanılması neticesinde oluşabilecek sakıncayı ortadan kaldırmak için ana kütlede aynı sektör işletmelerinin alınması yoluna gidilmiştir. Kısaca belirtmek gerekirse her bir manipülasyona maruz kalmış 51 şirket karşılığında eşleşmesi için manipülasyona uğramış şirketlerle aynı sektörden tesadüfi olarak seçilen 51 şirkete ait aynı yılın aynı dönemindeki mali tablo verileri elde edilerek 102 imalat sektörü işletmesi örneği alınmıştır. Ana kütlede "ulaştırma, haberleşme ve depolama", "inşaat, bayındırlık", "mali kuruluşlar", "teknoloji", "toptan perakende ticaret, otel ve lokanta", "idari ve destek hizmet faaliyetleri", "holding" işletmeleri çıkarılmış ve nihai olarak örnekleme ulaşılmıştır. Çalışmada "Faaliyet Kâr Marjı" ve "Net Kâr Marjı" bağımsız değişkenleri arasında yüksek korelasyon olduğundan "Faaliyet Kâr Marjı" değişkeni veri setinden çıkarılmıştır. "Yönetim Kurulunda Bağımsız Üye Sayısı" bağımsız değişkeni sınıflandırmadaki gücünün zayıflığı nedeniyle veri setinden çıkarılmıştır.

Tablo 1: Çalışmada Pay Senedi Manipüle Edilen Şirketlerin Belirlenmesine Yönelik Kriterler

No	Pay Senedinin Manipüle Edilme Kriterleri	Açıklama
1	SPK VI.104.1 Piyasa Bozucu Eylemler Tebliğinin 5'inci maddesinin 1'inci fıkrasının a, b ve f bentleri	... pay piyasasının arz ve talebi ile fiyatı hakkında yanıltıcı izlenim uyandırmak
2	6362 sayılı SPKn'nun 101.1 maddesi ve V101.1 sayılı Bilgi Suistimali ve Piyasa Dolandırıcılığı İncelemelerinde Uygulanacak Tedbirler Tebliğinin 18'inci maddesinin birinci fıkrasının (g) bendi ve üçüncü fıkrası	
3	6362 sayılı SPKn'nun 107.1 maddesi hükümleri kapsamında işlem yapmak üzere 6362 sayılı SPK'nın 115'inci maddesi uyarınca Cumhuriyet Başsavcılığına suç duyurusunda bulunulması	
4	6362 sayılı SPKn'nun 101/1 maddesi ve V.101.1 sayılı Tebliğin Bilgi Suistimali ve Piyasa Dolandırıcılığı İncelemelerinde Uygulanacak Tedbirler Tebliğinin 5/1 ve 6/1 maddeleri uyarınca 6 ay süreyle ... hakkında geçici işlem yasağı	
5	VI.104.1 Piyasa Bozucu Eylemler Tebliğinin 4. maddesinin üçüncü fıkrası	Mali tabloların kamuya duyurulmasına ilişkin işlem yapılamaz dönemde içsel ve sürekli bilgi sahiplerince işlem gerçekleştirilmesi
6	VI.104.1 Piyasa Bozucu Eylemler Tebliğinin 5. maddesinin birinci fıkrasının a ve f bentleri	... pay piyasasının arz ve talebi ile fiyatı hakkında yanıltıcı izlenim uyandırmak
7	6362 sayılı SPKn'nun 107.1 maddesinde sayılan fillerin işlendiğine dair makul şüphe bulunması nedeniyle 6362 sayılı SPKn'nun 128.1(a) ve 101'inci maddesine dayanarak yapay fiyat oluşturmaya yönelik işlemler	

Çin Menkul Kıymetler Düzenleme Komisyonu (CSRC) yetkilileri, herhangi birinin borsada manipülasyon yaptığını tespit ettiğinde, manipülatörleri cezalandırıp bu durumu düzenli olarak web sitesinde yayınlamaktadır (Li vd., 2017:791). SPK ve BIST de manipülasyonları izlemek ve tespit etmek için önemli veri oluşturan köklü düzenleyici ve denetleyici kurumlardır. Sermaye Piyasası Kurulunun bültenleri, manipüle edilmiş payları belirlemek için bize güvenilir bir kaynak sağlamaktadır (İmişiker ve Taş, 2013:121). Bu çalışma İmişiker ve Taş (2013) çalışmasını takip ederek manipülasyona maruz kalma olasılığı daha yüksek olan pay senetlerinin benzer özelliklerini tanımlamak için kullanılabilir firmaya özgü faktörleri araştırmaktadır. Sermaye Piyasası Kurulu (SPK) internet sayfasında haftalık bültenler yayınlanmaktadır. Haftalık bültenlerde SPK tarafından tespit edilen manipülasyon durumları, manipülasyona uğrayan pay senedi, manipülasyon dönemi, manipülasyon yapan kişi/kişiler, bu kişilere verilen idari para cezaları, bu kişiler hakkında savcılıklara yapılan suç duyuruları ve ilgili kişiler hakkında işlem yasakları bilgileri düzenli olarak paylaşılmaktadır. Bu çalışmanın dönem aralığı olarak SPKn V-101.1 nolu "Bilgi Suistimali ve Piyasa Dolandırıcılığı İncelemelerinde Uygulanacak Tedbirler Tebliği" yayınlanma tarihi olan 21.04.2014 tarihinden alınması planlanmış ancak 2014 yılındaki yaptırımların karşılığı işlenen suç tarihi 2010 yılına kadar geriye gittiğinden çalışma dönemi 13.04.2010'den itibaren

başlamıştır. Böylece çalışma dönemi Nisan 2010 ve Mart 2019 arasında belirlenmiştir. Bu çalışmada pay senedi manipüle edilen şirketlerin belirlenmesine yönelik kriterler Tablo 1'de verilmiştir.

4.2. Araştırmanın Yöntemi

Veri madenciliği, veri tabanı sistemlerindeki yöntemleri kullanarak büyük veri kümelerindeki kalıpları bulma sürecidir. Veri madenciliği işleminin genel amacı, bir veri setinden bilgi çıkarmak ve bilgiyi daha verimli alanlarda kullanmak amacıyla anlaşılabilir bir yapıya dönüştürmektir. Veri madenciliği analiz süreci ilginç ve karmaşık verileri veri ön işleme, görselleştirme ve çevrimiçi güncellemeyi içermektedir (Dhanalakshmi ve Subramanian, 2014:329). Veri madenciliği yöntemlerinin sermaye piyasalarında manipülasyon tespitinde analitik ve stratejik karar destek amaçlı olarak katkı sağlayacağı düşünülmektedir. Sadece sermaye piyasaları açısından değil tüm mali sistem açısından politika üretimi bakımından veri madenciliğinden etkin biçimde yararlanılabilmektedir (Koyuncugil, 2007: 16). Günümüzde, veri madenciliği, menkul kıymetler piyasası dâhil olmak üzere çeşitli sektörlerde yaygın olarak kullanılmakta ancak borsada aykırı değerlerin tespit edilmesi hâlâ keşif aşamasındadır (Li vd., 2016:246). Denetimli makine öğrenmesi, algoritmaya dışarıdan gösterilen örneklerden genel yargıya ulaşmakta ve daha sonra algoritmaya gösterilen örneklerin hangi sınıfa dâhil edilmesi gerektiğini tahmin edebilecek algoritmalara dayalı yöntemlerdir. Denetimli öğrenmenin amacı, sınıf etiketlerinin yordayıcı özellikleri açısından dağılımın kuralını veya modelini oluşturmaktır. Sonuçta ortaya çıkan sınıflandırıcı, yordayıcı özelliklerinin değerlerinin bilindiği teste sınıf etiketleri atamak için kullanılır. Sınıflandırma, temel olarak sınıflandırma ve regresyondan oluşur ve bir veri ögesini önceden tanımlanmış birkaç sınıftan birine eşleştiren (sınıflandıran) bir fonksiyon öğrenme işlemidir. Benzer şekilde regresyon, bir veri ögesini gerçek değerli tahmin değişkenine eşleyen bir fonksiyon öğrenme işlemidir (Li vd., 2017:790).

Bu çalışmada kullanılan veri seti SPK'nın manipülasyon tespiti üzerindeki uygulama gücü ile sınırlıdır ve SPK tarafından tespit edilmeyen başka manipülasyon faaliyetleri olabilir. Ancak düzenleyici kurumların son teknolojiye sahip algılama ve uygulama teknikleri göz önüne alındığında, gerçekleşen manipülasyonların bazı ortak özelliklerini belirlemek için bu verileri kullanmak oldukça mantıklıdır. BIST ve SPK, manipülasyonları tespit etmek için alım-satım emir ve fiyat bilgilerini kullanmaktadır. BIST ve SPK uzmanları manipülasyon olup olmadığı ile ilgili yaptıkları takipte tüm pay senetleri ve zaman dilimleri hem elektronik hem de manuel olarak incelemektedirler. Bu nedenle tahmin edilen katsayıların ve hata terimlerinin tarafsız olduğunu iddia edebilirken manipülasyonun tespiti konusunda bir önyargı olması ise düşük bir ihtimal olarak düşünülebilir (İmişiker & Taş, 2013, s. 125-126). Manipüle edilmiş pay senetlerinin BIST ve SPK tarafından tespit edilmesi için kullanılan değişkenler bu çalışmada incelenen değişkenlerden farklıdır. Tablo 2'de çalışmada kullanılan değişkenler verilmiştir.

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. 2007 yılında Veri Madenciliği uzmanlarının ziyaret ettiği sitede yapılan ankete göre WEKA, RapidMiner'e göre daha fazla indirilmiş olmasına rağmen uzmanlar arasında RapidMiner daha çok rağbet görmektedir (Dener vd., 2009:1-2). Bu nedenle bu çalışmada ANN, CART, C5.0 ve LR analizleri için RAPİDMİNER 9.2 programı kullanılmıştır. Çalışma metodolojisi sırasıyla aşağıda verilmektedir:

Tablo 2: Çalışmada Kullanılan Değişkenler

No	Bağımsız Değişkenler
X1	Stokların Toplam Varlıklara Oranı
X2	Finansal Kaldıraç Oranı
X3	Maddi Duran Varlık Devir Hızı
X4	Net Kâr Marjı
X5	Ekonomik Kârlılık Oranı (ROA)
X6	Özsermaye Kârlılığı (ROE)
X7	Dört Büyük Denetim Şirketi Tarafından Denetlenip Denetlenmediği, (Pricewaterhousecoopers-Deloitte Touche Tohmatsu-KPMG-Ernst and Young)
X8	Halka Açıklık Oranı %
X9	Sermayede doğrudan %5 veya daha fazla paya sahip gerçek ve tüzel kişiler -Halka açık olmayan paylarda yabancı sermaye payı %
X10	BIST Kurumsal Yönetim Endeksinde Olup Olmadığı (Endekste değilse 0, endekste ise 1)
X11	2010 yılından sonra manipüle edilip edilmediği (Manipüle edilmiş ise 0, manipüle edilmemiş ise 1)
X12	Ödenmiş / Çıkarılmış Sermaye

Kaynak: Literatür incelemesi sonucu yazar tarafından oluşturulmuştur.

- ✓ 2010-2019 Yılları arası Sermaye Piyasası Kurulu haftalık bültenlerinden manipülasyon yapılan şirketler ve manipülasyon tarihlerinin alınması
- ✓ Manipülasyon yapılan şirketlere ait mali tablo verilerinin Kamuyu Aydınlatma Platformundan elde edilmesi
- ✓ Borsa İstanbul İnternet sayfasından örnekleme oluşturan manipüle edilmiş ve manipüle edilmemiş şirketlere ait nitel değişkenlerin elde edilmesi
- ✓ Mali oranlar hesaplandıktan sonra nitel değişken verileri ile birlikte programa girilmesi
- ✓ 12 Değişkenli veri seti (ANN, CART, C5.0)
- ✓ 3 Değişkenli veri seti (LR)
- ✓ Veri seti %70 eğitim ve %30 test setine bölünmüştür. Doğrulama yöntemi olarak 10 katlı çapraz doğrulama kullanılmıştır. (ANN, CART)
- ✓ Veri seti %70 eğitim ve %30 test setine bölünmüştür. Test seti üzerinden model doğrulaması yapılmıştır. (C5.0, LR)
- ✓ Analizde kullanılan program: Rapidminer 9.2
- ✓ Analiz Sonuçlarının Tartışılması ve Öneriler

K-katlı çapraz doğrulama yöntemi uygun 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). K-katlı çapraz doğrulamada test için kullanılan örnekler, eğitim kümesi içerisinde modelin oluşturulması esnasında algoritmanın daha önce gördüğü örnekler olduğundan verinin ezberlenmesi sorunu ortaya çıkabilmektedir. Bu nedenle bu çalışmada ANN ve CART yöntemleri için 10 katlı çapraz doğrulama yöntemi 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 parametre optimizasyonu ile değişkenlerin her birinin önem dereceleri belirlenerek modelin öğrenmesi tamamlanmıştır. Ağırlıklar belirlenerek model öğrendikten sonra algoritmanın daha önce hiç görmediği tüm veri setinin %30'unu oluşturan test veri seti kullanılmış modelin performansı doğrulanmıştır. C5.0 ve LR yöntemlerinde veri seti %70 eğitim ve %30 test verisi olarak ikiye ayrılmış eğitim veri seti ile model kurulmuş ve test seti verileri ile kurulan modelin doğrulaması yapılmıştır.

4.3. Çalışmada Kullanılan Yöntemlere Ait Analizler ve Bulguları

Aralık 2006'da Veri Madenciliği Konferansı (ICDM) tarafından en iyi 10 veri madenciliği algoritması belirlenmiştir. Buna göre veri madenciliğinde kullanılan en iyi 10 algoritma arasında Sınıflandırma ve Regresyon Ağacı (CART) yer almaktadır (Li vd., 2010:5895). Bu çalışmada en iyi 10 veri madenciliği algoritmalarından CART'ın yanı sıra diğer alanlarda olduğu gibi finansal tahminleme ve sınıflandırma alanında da önemli başarılar elde eden ANN ve C5.0 karar ağacı yöntemi kullanılmıştır. İstatistiksel yöntemlerden ise karşılaştırma amacıyla LR kullanılmıştır. Çalışmada kullanılan 4 yöntemin sınıflandırma performansı, modellerin ayırt edici özellikleri belirtilerek karşılaştırılmıştır.

4.3.1. Yapay Sinir Ağları (ANN) Analizi ve Bulguları

Yapay Sinir Ağları (Artificial Neural Network, ANN) eğitim yoluyla öğrenen ve biyolojik sinir ağlarına benzeyen yapısıyla doğrusal olmayan kestirimci modellerdendir. ANN tahmin edicilerin değişkenler arasındaki olası tüm etkileşimleri saptama yeteneğine sahiptir. ANN aynı zamanda bağımlı ve bağımsız değişkenler arasındaki karmaşık doğrusal olmayan ilişkilerde bile şüphe duymadan tam bir algılama yeteneğine sahiptir. Bu nedenle ANN tekniği en iyi tahmin yöntemlerinden biri olarak seçilmiştir (Li, Wu ve Liu, 2017: 790).

ANN girdi katmanı, bir veya daha fazla gizli katman ve çıktı katmanı olarak üç katmandan oluşmaktadır. Her katmanda nöronlar bulunmakta ve bir katmandaki nöronlar bir sonraki katmandaki nöronlara bağlıdır. Her bağlantının bir ağırlığı vardır ve bağlı nöronun çıktısının ağırlıklı kombinasyonu bir sonraki nöronun girdisine gönderilir. Dönüşüm fonksiyonu girişi alır ve nöronun çıkış değerine dönüştürür. Son hesaplanan çıktı, çıkışın gerçek değeri ile karşılaştırılır ve ağırlığın yeniden hesaplanması için farklı yöntemlerle fark belirlenir. Bunlar arasında, geri yayılım algoritması en çok kullanılanıdır, gerçek ve hesaplanmış çıktı arasındaki farka dayanarak çıktı düğümlerinden iç düğümlere kadar olan ağırlıkları ayarlamaktadır (Öğüt vd., 2009:11946).

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 yapılırsa 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).

İleri sürümlü ANN'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 d 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 gösteren bir ağırlık ile temsil edilir. Sinir ağı kullanılarak tasarlanan bir pay senedi manipülasyon tespit modeli, belirli bir firma için, pay senedi manipülasyon ihtimalini temsil eden bir gizli katman, bir çıktı nöronu ve bir girdi katmanından oluşan ağ ile aşağıdaki şekilde ifade edilebilen bir Z skoru hesaplamaktadır (Öztemel, 2012:55).

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

Formülde f aktivasyon fonksiyonu, n değişken sayısı, p gizli nöron sayısı, xi girdi katmanı nöronları, wij girdi katmanı ve gizli katman arasındaki ilişkileri temsil eden ağırlıkları, wj gizli katman ve çıktı katmanı grupları arasındaki ağırlıkları, bj gizli nöronların ağırlıkları ve b çıkış nöronunun ağırlığını göstermektedir (Jardin, 2016:241). Tablo 3’de ANN parametreleri verilmiştir.

Tablo 3: ANN Parametreleri

Ağ Türü	Çok Katmanlı Perseptron			
Öğrenme Algoritması	Geri Yayılım			
Öğrenme Kuralı	Momentum			
Girdi Katmanındaki Düğüm Sayısı	12			
Gizli Katman Sayısı	1			
Çıktı Katman Düğüm Sayısı	2 (Manipüle Edilmiş/Manipüle Edilmemiş)			
Değişken Seçimi	12 Bağımsız Değişken			
Doğrulama Türü	%70 Eğitim ve %30 Test Veri Seti Ayrımı 10 Katlı Çapraz Doğrulama			
Örnekleme Seçim Türü	Tabakalı Örnekleme 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: 0,20	Adımlar: 10	Ölçek: Doğrusal
Eğitim Devir Sayısı	En Düşük : 1,00	En Yüksek: 500	Adımlar: 10	Ölçek: Doğrusal

Çalışmada tüm veri setinin %70’ini oluşturan (pay senedi manipüle edilen 51 ve pay senedi manipüle edilmeyen 51 olmak üzere toplam 102 örneğin yaklaşık 72’si) eğitim verisi olarak alınmış, veri setinin %30’unu oluşturan (pay senedi manipüle edilen 51 ve pay senedi manipüle edilmeyen 51 olmak üzere toplam 102 örneğin 30’u) test verisi olarak alınmış (30 test verisinden 15 şirket pay senedi manipüle edilen ve 15 şirket pay senedi manipüle edilmeyen olarak ayrılmış) ve test verisinden elde edilen sonuçlar verilmiştir. Tablo 4’de ANN analizi karışıklık matrisi verilmiştir. Manipüle edilmiş 15 pay senedinin 12’si doğru tahmin edilmiş 3 örnek gerçekte manipüle edildiği halde yanlış sınıflandırmayla “manipüle edilmemiş” olarak tahmin edilmiştir. Manipüle edilmemiş 15 pay senedinin 14’ü doğru tahmin edilmiş, 1 pay senedi gerçekte manipüle edilmemiş olduğu halde “manipüle edilmiş” olarak yanlış tahmin edilmiştir.

Tablo 4: %70 - %30 Test Veri Seti Ayrımı ve 10 Katlı Çapraz Doğrulama Yapılmış ANN Analizi Karışıklık Matrisi

		Tahmin Edilen Grup		
		Manipüle Edilmiş	Manipüle Edilmemiş	Toplam
Gözlemlenen Grup(Gerçek)	Manipüle Edilmiş	12	3	15
	Manipüle Edilmemiş	1	14	15
	Toplam	13	17	30

Sınıflandırma performansının kapsamlı olarak değerlendirmesi ROC tarafından gerçekleştirilir. Belirli bir özgülük değeri için daha yüksek bir hassasiyet değeri daha iyi performans olduğunu gösterir. ROC eğrisinin altındaki alan (AUC), sınıflandırıcı performansını değerlendirmek için yaygın olarak kullanılan bir ölçümdür (Li, Wu ve Liu, 2017: 792). Çalışmada kullanılan modeller 6 kriter yönünden sıralanmıştır. Bunlar: i) Doğruluk (accuracy) ii) Kesinlik (Precision), iii) Duyarlılık (Recall), iv) Kappa değeri, v) AUC, vi) F ölçütüdür. Tablo 5’de parametre optimizasyonu sonucu en iyi performans gösteren ANN model sonucu verilmektedir.

Tablo 5: ANN Analizi Performans Sonuçları

Parametreler	%70 - %30 Test Veri Seti Ayrımı ve 10 Katlı Çapraz Doğrulama Sonuçları
Doğruluk (accuracy)	86,67%
Sınıflandırma Hatası (classification error)	13,33%
Kappa	0,733
AUC	0,911
Duyarlılık (recall)	%93,33
Kesinlik (precision)	%82,35
F ölçütü	%87,50
Öğrenme Oranı (learning rate)	0,27
Momentum	0,08
Devir Sayısı (training cycles)	1

4.3.2. CART Karar Ağacı Analizi ve Bulguları

Sınıflandırma ve regresyon ağacı (Classification and Regression Tree, CART) veri kümesinin sınıflandırılmasında kullanılan karar ağacı tekniğidir ve hangi kayıtların belirlenmiş bir sonuç 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). 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

Gini ölçütü bir frekans dağılımında değerler arasındaki eşitsizliğin ölçüsüdür. Bu ölçütte öznitelik değerlerinin sol ve sağda olmak üzere iki bölüme ayrılması, her bir bölüm için ayrı ayrı Gini ölçütünün hesaplanması ve elde edilen sonuçların karşılaştırılması esasına dayanmaktadır. (Özkan, 2016:111). Çalışmada kullanılan modeller 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 ve karar ağaçlarının dallarında belirtilen yorumlama basitliğine dayalı olarak seçilmiştir. Tablo 6'de CART sınıflandırma algoritması analiz parametreleri verilmektedir.

Tablo 6: CART Sınıflandırma Algoritması Analiz Parametreleri

Analiz İçin Belirlenen Parametreler	Açıklama			
Veri Seti Ayrımı	%70 Eğitim - %30 Test Veri Seti Ayrımı			
Doğrulama Türü	10 Katlı Çapraz Doğrulama			
Değişken Sayısı	12 Bağımsız Değişken			
Örnekleme Seçimi	Tabakalı Örnekleme Seçimi			
Bölünme Kriteri	Gini Index			
Analiz İçin Belirlenen Parametreler	En Düşük	En Yüksek	Adımlar	Ölçek
Bölünme İçin Minimum Boyut (Minimal Size For Split)	1	4	10	Doğrusal
En Düşük Dal Boyutu (Minimal Leaf Size)	1	2	10	Doğrusal
Minimum Kazanım (Minimal Gain)	1	20	10	Doğrusal
Maksimum Derinlik (Maximal Depth)	1	20	10	Doğrusal
Güven Düzeyi (confidence)	0	0,25		
Ön Budama Sayısı	1	10	-	-

Tablo 7'de CART karar ağacı analizi karışıklık matrisi verilmektedir.

Tablo 7: %70 - %30 Test Veri Seti Ayrımı ve 10 Katlı Çapraz Doğrulama Yapılmış CART Karar Ağacı Analizi Karışıklık Matrisi

		Tahmin Edilen Grup		
		Manipüle Edilmiş	Manipüle Edilmemiş	Toplam
Gözlemlenen Grup(Gerçek)	Manipüle Edilmiş	13	2	15
	Manipüle Edilmemiş	4	11	15
	Toplam	17	13	30

Tablo 8'de parametre optimizasyonu sonucu en iyi performans gösteren CART Sınıflandırma ve regresyon ağacı analizi performans sonuçları verilmektedir.

Tablo 8: CART Sınıflandırma ve Regresyon Ağacı Analizi Performans Sonuçları

Parametreler	%70 - %30 Test Veri Seti Ayrımı ve 10 Katlı Çapraz Doğrulama
Doğruluk (accuracy)	80,00%
Sınıflandırma Hatası (classification error)	20,00%
Kappa	0,600
AUC	0,749
Kesinlik (Precision)	84,62%
Duyarlılık (Recall)	73,33%
F ölçütü	78,57%
Bölünme İçin Minimum Boyut (Minimal Size For Split)	4
En Düşük Dal Boyutu (Minimal Leaf Size)	1
En düşük Kazanım (Minimal Gain)	14,30
En Yüksek Derinlik	12

4.3.3. C5.0 Karar Kuralı Türetme Algoritması Analizi ve Bulguları

C4.5 algoritması 1993 yılında Ross Quinlan tarafından geliştirilen bir karar ağacı algoritmasıdır. CLS (Concept Learning System) birçok karar ağacı yönteminin temel taşıdır. CLS ailesinin özelliklerine sahip olan C4.5 en iyi bölme öznitelik seçiminde entropi ve enformasyon kazancı tekniklerini kullanır. C4.5 algoritmasında karar ağacı oluşturulurken kayıp veriler dikkate alınmamaktadır. Kazanım oranı hesaplanırken yalnızca eksik verisi olmayan diğer kayıtlar kullanılmaktadır. C4.5 algoritması kayıp verileri diğer veri ve değişkenlerle tahmin ederek kazanım oranının hesaplanmasında kullanılmaktadır. Bu şekilde daha duyarlı ve anlamlı karar ağaçları üretilebilmektedir (Silahtaroglu, 2016: 80). C5.0 Karar Kuralı Türetme Algoritması da Quinlan tarafından geliştirilmiştir. C4,5'e göre daha hızlı olması, belleği daha etkin kullanması, faydasız özniteliklerin elimine edilmesi gibi gelişmiş özelliklere sahiptir (Akpınar, 2014: 221-222).

Tablo 9'da C5.0 Karar Kuralı Türetme analiz parametreleri verilmektedir.

Tablo 9: C5.0 Karar Kuralı Türetme Algoritması Analiz Parametreleri

Analiz İçin Belirlenen Parametreler	Açıklama			
Veri Seti Ayrımı	%70 Eğitim - %30 Test Veri Seti Ayrımı			
Değişken Sayısı	12 Bağımsız Değişken			
Örnekleme Seçimi	Tabakalı Örnekleme Seçimi			
Bölünme Kriteri	Bilgi Kazancı (Entropi)			
Analiz İçin Belirlenen Parametreler	En Düşük	En Yüksek	Adımlar	Ölçek
Bölünme İçin Minimum Boyut (Minimal Size For Split)	1	4	10	Doğrusal
En Düşük Dal Boyutu (Minimal Leaf Size)	1	2	10	Doğrusal
Minimum Kazanım (Minimal Gain)	0	20	10	Doğrusal
Maksimum Derinlik (Maximal Depth)	-1	20	10	Doğrusal
Güven Düzeyi (confidence)	0	0,25		

Tablo 10’da parametre optimizasyonu sonucu en iyi performansı gösteren C5.0 Karar kuralı türetme algoritması analizi performans sonuçları verilmektedir.

Tablo 10: C5.0 Karar Kuralı Türetme Algoritması Analizi Performans Sonuçları

Parametreler	%70 - %30 Test Veri Seti Ayrımı
Doğruluk (accuracy)	76,67%
Sınıflandırma Hatası (classification error)	23,33%
Kappa	0,533
AUC	0,693
Kesinlik (Precision)	90,00%
Duyarlılık (Recall)	60,00%
F Ölçüsü	72,00%
Bölünme İçin Minimum Boyut (Minimal Size For Split)	2
En Düşük Dal Boyutu (Minimal Leaf Size)	2
En düşük Kazanım (Minimal Gain)	0,00
En Yüksek Derinlik	12

Tablo 11’de C5.0 Karar Kuralı Türetme Algoritması Analizi Karışıklık Matrisi verilmiştir.

Tablo 11: C5.0 Karar Kuralı Türetme Algoritması Analizi Karışıklık Matrisi

		Tahmin Edilen Grup		
		Manipüle Edilmiş	Manipüle Edilmemiş	Toplam
Gözlemlenen Grup(Gerçek)	Manipüle Edilmiş	14	1	15
	Manipüle Edilmemiş	6	9	15
	Toplam	20	10	30

4.3.4. Lojistik Regresyon Analizi ve Bulguları

Lojistik regresyon analizi adını bağımlı değişkene uygulanan logit dönüştürmeden 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). 12 değişkenle yürütülen lojistik regresyon analizi tahmin sonuçları düşük çıktığından diskriminant ileri adımlı analiz ile belirlenen 3 değişkenle lojistik regresyon analizi gerçekleştirilmiştir. Tablo 12’de LR analiz parametreleri verilmektedir.

Tablo 12: Lojistik Regresyon Analiz Parametreleri

Analiz İçin Belirlenen Parametreler	Açıklama
Veri Seti Ayrımı	%70 Eğitim - %30 Test Veri Seti Ayrımı
Değişken Sayısı	3 Bağımsız Değişken X3- Maddi Duran Varlık Devir Hızı X8- Halka Açıklık Oranı % X12- Ödenmiş / Çıkarılmış Sermaye
Örnekleme Seçimi	Tabakalı Örnekleme Seçimi

Tablo 13’de LR analizi performans sonuçları verilmektedir.

Tablo 13: Lojistik Regresyon Analizi Performans Sonuçları

Parametreler	%70 - %30 Test Veri Seti Ayrımı
Doğruluk (accuracy)	%70,00
Sınıflandırma Hatası (classification error)	%30,00
Kappa	0,40
AUC	0,769
Keskinlik (Precision)	71,43
Duyarlılık (Recall)	66,67
F ölçütü	68,97

Tablo 14’de LR analizi karışıklık matrisi verilmiştir.

Tablo 14: Lojistik Regresyon Analizi Karışıklık Matrisi

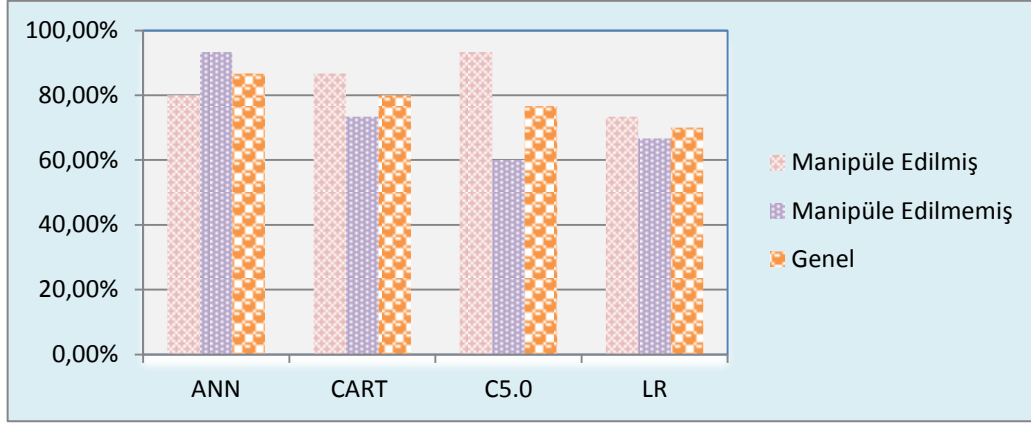
		Tahmin Edilen Grup		
		Manipüle Edilmiş	Manipüle Edilmemiş	Toplam
Gözlemlenen Grup(Gerçek)	Manipüle Edilmiş	11	4	15
	Manipüle Edilmemiş	5	10	15
	Toplam	16	14	30

4.4. Veri Madenciliği Yöntemlerine Ait Bulguların Değerlendirilmesi

Çalışmada Borsa İstanbul’da işlem görmüş 102 işletmeye ait pay senetleri üzerinde piyasa manipülasyonu yapıp yapılmayacağı işletmelere ait nicel ve nitel veriler kullanılarak 3-9 ay öncesinden tahmin edilmiştir. Çalışmada veri madenciliği yöntemlerinden ANN, CART, C5.0 ve LR yöntemleri kullanılarak ilgili yöntemlerin performansı modellerin ayırt edici özellikleri altında karşılaştırılmıştır. Analizler sonucunda genel sınıflandırma doğruluğu sırasıyla ANN (%86,67), CART (%80,00), C5.0 (%76,67) ve LR (%70,00) olarak bulunmuştur. C5.0 algoritması 15 manipüle edilen pay senetlerinin 14’ünü doğru tahmin ederek %93,33 sınıflandırma başarıları göstermiştir. ANN 15 manipüle edilmemiş pay senedinin 14’ünü doğru tahmin ederek %93,33 sınıflandırma başarıları göstermiştir. Bu çalışmada tüm yöntemler %70’in üzerinde tahmin performansı gösterdiğinden pay senedi manipülasyon tahmini analizlerine ilgili modeller de dahil edilebilir.

- Şekil 1’de çalışmada kullanılan yöntemlere ilişkin sınıflandırma sonuçları verilmiştir.

Şekil 1: Çalışmada Kullanılan Yöntemlere İlişkin Sınıflandırma Sonuçları



Golmohammadi ve Zaiane (2012) veri madenciliği yöntemlerinin, menkul kıymetler piyasasında farklı dolandırıcılık türlerini tespit etmede etkili olduğu sonucuna varmışlardır. Bu çalışmada da Golmohammadi ve Zaiane (2012) çalışması ile uyumlu olarak veri madenciliği yöntemlerinin, menkul kıymetler piyasasında farklı dolandırıcılık türlerini tespit etmede etkili olduğu bulgusuna ulaşılmıştır.

Öğüt vd. (2009) çalışması pay senedi manipülasyonunu tahmin etme üzerine değil endeksin ortalama günlük getirisi, işlem hacmindeki ortalama günlük değişim ve ortalama günlük volatilité arasındaki farkı açıklayıcı değişken olarak kullanarak manipülasyon yapıldığı esnada tespit etmeye yönelmiştir. Manipülasyon tespitinde ANN ve SVM'nin toplam sınıflandırma doğruluğu ve hassasiyet istatistikleri bakımından performansları Diskriminant ve LR'ye göre daha iyi olduğu bulgusuna ulaşmışlardır. Bu çalışmada ilgili makaleyle ortak değişkenler ANN ve LR'dir. Çalışmada veri madenciliği içerisinde makine öğrenimi yöntemleri olan ANN, CART ve C5.0 veri madenciliğinin istatistiksel yöntemlerinden LR'den daha yüksek tahmin performansı göstermiştir. Bu sonuç Öğüt vd. (2009) çalışmasının sınıflandırma doğruluğu yönünden ANN'nin LR'den daha iyi olduğu bulgusuyla uyumludur.

Liv vd. (2017) çalışmasında pay senedi alım/satım verilerinden faydalanarak pay senetlerinin manipüle edilip edilmeyeceğini tahmin etmek için yaptığı analizlerde KNN ve DTC'nin doğruluk, hassasiyet, özgüllük ve AUC yönlerinden tüm yöntemlerle kıyaslandığında %99,91 ile en iyi tahmin performansı gösterdiği bulunmuştur. Yöntemlerin tahmin performansı en yüksekten düşüğe doğru sırasıyla KNN (%99,91)=DTC (%99,91)>ANN (%98,00)> QDA (%93,99)> SVM (%91,81)> LDA (%90,46)> LR (%89,11) olarak bulunmuştur. Bu çalışmada en yüksek sınıflandırma doğruluğunu ANN elde etmiştir. Sırasıyla CART, C5.0 ve LR gelmektedir. Liv vd. (2017) çalışma bulgularından ANN'nin LR'dan daha yüksek sınıflandırma performansı göstermesi ve LR'nin en düşük performans elde etmesi bu çalışmanın bulgularıyla da uyumludur.

İmişiker ve Taş (2013) çalışmalarında yüksek kaldıraç oranına sahip olan firmaların hisse senedi fiyat manipülasyonuna maruz kalmaya daha yatkın olduğunu belirtmişlerdir. Bu sonuç bu çalışma bulgularıyla tutarlıdır. Doğrulama yöntemi olarak 10 katlı çapraz doğrulama kullanılarak gerçekleştirilen CART karar ağacının kökünü "X7- Dört Büyük Denetim Şirketi Tarafından Denetlenip Denetlenmediği" bağımsız değişkeni oluşturmuştur. Ağacın oluşturulmasında ikinci önemli değişken olarak "X2- Finansal Kaldıraç Oranı" bulunmuştur.

İmişiker ve Taş (2013) çalışmalarında bir diğer önemli sonuç olarak, daha önce manipüle edilmiş pay senetleri için ilgili pay senedinin yeniden manipülasyona uğrama olasılığının yüksek olduğu sonucuna varmışlardır. İlgili değişken bu çalışmada da kullanılmıştır. Ancak karar ağaçlarının oluşturulmasında önemli değişkenler arasında yer almamaktadır. İmişiker ve Taş (2013) çalışmalarında büyük firmaların manipüle edilme ihtimalinin düşük olduğu görüşünü desteklediği bulgusuna ulaşmışlardır. İlgili çalışmada büyük firma ayrımı pay senedi fiyatı ile pay senedi sayısı çarpımı ile elde edilmiştir. Bu çalışmada ödenmiş-çıkarılmış sermayesi yüksek olan firmaların daha az manipülasyona uğradığı tespit edilmiştir. Bu sonuç CART Karar Ağacının %70 eğitim %30 test veri seti ayrımı ile 10 katlı çapraz doğrulamanın kullanıldığı ağacın oluşturulmasında kökü oluşturan en önemli değişken olarak belirlenmiştir.

İmişiker ve Taş (2013) çalışmalarında halka açıklık oranı daha yüksek olan firmaların daha az manipüle edilmiş görüldüğünü tespit etmişlerdir. İlgili sonuç bu çalışma bulgularıyla tutarlıdır. Bu çalışmada gerçekleştirilen Diskriminant İleri Adımlı Analiz sonuçlarına göre halka açıklık oranı manipüle edilen ve manipüle edilmeyen firma ayrımında önemli bir değişken olarak bulunmuştur. İmişiker ve Taş (2013) çalışmalarında kârlılık, kaldıraç oranı ve firmanın cari oranını bir firmanın finansal performansının önemli göstergeleri olarak kabul edilmektedir. İlgili sonuç bu çalışma ile uyumludur. Bu çalışmada CART Karar Ağacının %70 eğitim %30 test veri seti ayrımı ile 10 katlı çapraz doğrulamanın kullanıldığı ağacın oluşturulmasında ikinci önemli değişken olarak "X5- Ekonomik Kârlılık Oranı" bulunmuştur. Sadece 10 katlı çapraz doğrulamanın kullanıldığı CART modelinde "X7- Dört Büyük Denetim Şirketi Tarafından Denetlenip Denetlenmediği", "X2- Finansal Kaldıraç Oranı" ve "X3- Maddi Duran Varlık Devir Hızı" değişkenleri ağacın oluşturulmasında önemli değişkenler olarak bulunmuştur.

5. Sonuç

Düzenleyici kurumlar ve borsalar manipülasyon yapan piyasa katılımcılarına hızlı ve katı yaptırımlar uygulamasına rağmen, manipülasyona maruz kalmak yatırımcıların en büyük endişesi olmaya günümüzde de devam etmektedir. Türk sermaye piyasalarının gelişiminin önündeki en önemli engellerden biri yatırımcıların finansal piyasalara olan güvenlerinin zayıflamasıdır. Sermaye piyasasında olması gereken "güven" unsurunu sarsan bu gibi yatırım ortamı yalnızca yatırımların artmasını engellemez aynı zamanda uzun vadede ülke ekonomisine de zarar vermektedir.

Bu çalışmada Borsa İstanbul İmalat Sanayi sektöründe faaliyet gösteren 51 manipüle edilmiş ve 51 manipüle edilmemiş toplam 102 işletmenin mali tablo verileri ve şirketlere ait nitel veriler kullanılmıştır. Manipülasyon tarihinden 3 ay öncesinden 9 ay öncesine kadar manipülasyona maruz kalıp kalmayacakları konusunda Yapay Sinir Ağları (ANN) ve Sınıflandırma ve Regresyon Ağacı (CART), C5.0 Karar Kuralı Türetme Algoritması ve Lojistik Regresyon (LR) yöntemleri kullanılarak tahmin yapılmıştır. Çalışmada ANN, CART ve C5.0 analizleri için parametre optimizasyonu ile en yüksek performansa sahip modeller belirlenmiştir. Analiz sonucunda veri seti %70 eğitim ve %30 test seti olarak ayrılmış ve 10 katlı çapraz doğrulama yapılmış olarak gerçekleştirilen analizde ANN modeli test örneğinde 15 manipüle edilmiş işletmenin 12'sini doğru tahmin etmiş ve 15 manipüle edilmemiş işletmenin 14'ünü doğru tahmin etmiştir. ANN'nın genel sınıflandırma doğruluğu %86,67 olarak bulunmuştur. CART karar ağacı modeli test örneğinde 15 manipüle edilmiş işletmenin 13'ünü doğru tahmin etmiş ve 15 manipüle edilmemiş işletmenin 11'ini doğru tahmin etmiştir. CART Karar Ağacının ortalama sınıflandırma doğruluğu %80 olarak bulunmuştur. C5.0 Karar Kuralı Türetme Algoritması Analizi Manipüle edilmiş 15 pay senedinin 14'ü doğru tahmin edilmiş, bir örnek gerçekte manipüle edildiği halde yanlış sınıflandırmayla manipüle edilmemiş olarak tahmin edilmiştir. Manipüle edilmemiş 15 pay senedinin 9'u doğru tahmin edilmiş, 6 pay senedi gerçekte manipüle edilmemiş olduğu halde manipüle edilmiş olarak yanlış tahmin edilmiştir. C5.0'in genel sınıflandırma doğruluğu %76,67 olarak bulunmuştur. LR analizi manipüle edilmiş 15 pay senedinin 11'i doğru tahmin edilmiş, 4 örnek gerçekte manipüle edildiği halde yanlış sınıflandırmayla manipüle edilmemiş olarak tahmin edilmiştir. Manipüle

edilmemiş 15 pay senedinin 10'u doğru tahmin edilmiş, 5 pay senedi gerçekte manipüle edilmemiş olduğu halde manipüle edilmiş olarak yanlış tahmin edilmiştir. LR'nin genel sınıflandırma doğruluğu %70,00 olarak bulunmuştur.

Sonuçta, finans, muhasebe ve denetim teknolojileri üzerine karşılıklı diyalog ve zenginleştirme için çeşitli fırsatlar vardır gelecekte bu tartışmaları sürdürmek için araştırmaların devam etmesi gereği açıktır. SPKn'nun 107/1 maddesindeki işlem kalıplarını içerecek şekilde içeriden öğrenenlerin ticareti ve piyasa dolandırıcılığı suçu işleyenlerin e-postaları ve sosyal ağlarının analizi ile ilişkilendirme üzerinde yapılacak sonraki çalışmalar literatüre değer katacaktır.

Kaynakça

- Akpınar, H. (2014). *Data Veri Madenciliği Veri Analizi*, İstanbul: Papatya Yayınları.
- Comerton-Forde, C. ve Putnins, T. J. (2011). Measuring Closing Price Manipulation. *J. Finan. Intermediation*, (20), 135-158.
- Çalışkan, M.M. Tuncer. (2015). Borsa İstanbul'da Manipülasyon: 2001 – 2013 Yılları Arasında İşlem Yasaklıların Profili. *Yönetim ve Ekonomi Araştırmaları Dergisi*. 13 (1), 381-391.
- Çelik, U., Akçetin, E. ve 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.
- Dener, M., Dörterler, M. ve Orman, A. (2009). Açık Kaynak Kodlu Veri Madenciliği Programları: Weka'da Örnek Uygulama. *Conference: XI. Akademik Bilişim Konferansı* (s. 1-11). Şanlıurfa: Conference: XI. Akademik Bilişim Konferansı.
- Diaz, D., Theodoulidis, B. ve Sampaio, P. (2011). Analysis of Stock Market Manipulations using Knowledge Discovery Techniques Applied to Intraday Trade Prices. *Expert Systems with Applications*, (38), 12757–12771.
- Dhanalakshmi, S. ve Subramanian, C. (2014). An Analysis of Data Mining Applications for Fraud Detection in Securities Market. *International Journal of Data Mining Techniques and Applications*, (3), 326-335.
- Gemici, E., Cihangir, M. ve Yakut, E. (2017). İşlem Bazlı Manipülasyon: Türkiye Örneği. *Ege Akademik Bakış*, 17 (3), 369-380.
- Golmohammadi, K., ve Zaiane, O. R. (2012). Data Mining Applications for Fraud Detection in Securities Market. *Conference: Intelligence and Security Informatics Conference (EISIC), 2012 European*, (s. 1-8). Edmonton, Canada: <https://www.researchgate.net/publication>.
- Gündoğdu, A. (2017). Türk Sermaye Piyasasında Bilgi Suistimali ve Piyasa Dolandırıcılığı Suçu Düzenlemelerinin İrdelenmesi. *Maliye Finans Yazıları*, (108), 9-26.
- İmişiker, S. ve Taş, B. K. (2013). Which Firms are More Prone to Stock Market Manipulation? *Emerging Markets Review*, (16), 119-130.
- Jardin, P.D. (2016). A Two-Stage Classification Technique for Bankruptcy Prediction, *European Journal of Operational Research*, (254), 236-252.
- Kesici, E. ve Nacakçı, M. B. (2014). Piyasa Bozucu Eylemler . *Capital Markets Journal*, (15), 1-12.
- Kılıç, S. (2015). Kappa Testi, *Journal of Mood Disorders*, 5 (3), 142-144.
- Koyuncugil, A. S. (2007). Veri Madenciliği ve Sermaye Piyasalarına Uygulanması. Ankara : Sermaye Piyasası Kurulu Araştırma Dairesi Araştırma Raporu.

- Li, H., Sun, J. ve Wu, J. (2010). Predicting Business Failure using Classification and Regression Tree: An Empirical Comparison with Popular Classical Statistical Methods and Top Classification Mining Methods, *Expert Systems with Applications*, (37), 5895-5904.
- Li, A., Feng, M., Li, Y. ve Liu, Z. (2016). Application of Outlier Mining in Insider Identification Based on Boxplot Method. *Procedia Computer Science*, (91), 245-251.
- Li, A., Wu, J. ve Liu, Z. (2017). Market Manipulation Detection Based on Classification Methods. *Procedia Computer Science*, (122), 788-795.
- Öğüt, H., Doğanay, M. M. ve Aktaş, R. (2009). Detecting Stock-Price Manipulation in an Emerging Market: The Case of Turkey. *Expert Systems with Applications*, (36), 11944-11949.
- Özcan, R. (2012). Hisse Senedi Piyasalarında Manipülasyon Stratejileri. *İMKB Dergisi*, 13(49), 20-41.
- Özkan, Y. (2016). *Veri Madenciliği Yöntemleri*. İstanbul : Papatya Yayıncılık.
- Öztemel, E. (2012). *Yapay Sinir Ağları*. İstanbul: Papatya Yayınları.
- Silahtaroglu, G. (2016). *Veri Madenciliği Kavram ve Algoritmaları* . İstanbul: Papatya Yayınları.
- Şensoy, D. (2013). Manipülasyon; Piyasa Dolandırıcılığı Suçu, Uygulanacak Tedbirler ve Yaptırımlar. *Ankara Barosu Dergisi*, 2013 (3), 369-399.
- Tellechea, A. F. (2008). Economic Crimes in The Capital Markets. *Journal of Financial Crime*, 15(2), 214-222.
- Williams, J. W. (2013). Regulatory technologies, risky subjects, and financial boundaries: Governing 'fraud' in the financial markets. *Accounting, Organizations and Society*, (38), 544-558.

Ek:1 ANN Ağırlıkları

Girdi Katmanı	Gizli Katman								Eşik
Bağımsız Değişken	Düğüm 1	Düğüm 2	Düğüm 3	Düğüm 4	Düğüm 5	Düğüm 6	Düğüm 7	Düğüm 8	Threshold
	Sigmoid								
X1	-0,020	-0,032	-0,043	-0,002	-0,024	0,004	0,030	-0,022	
X2	-0,035	0,000	0,006	0,047	0,009	-0,041	-0,049	-0,048	
X3	-0,031	0,035	0,036	0,037	-0,006	-0,008	-0,028	-0,026	
X4	-0,005	-0,008	-0,019	0,039	0,008	-0,014	0,021	0,007	
X5	-0,027	-0,006	0,017	-0,028	0,006	-0,001	-0,050	0,017	
X6	-0,034	-0,014	0,008	0,032	-0,015	0,013	-0,042	-0,029	
X7	-0,014	-0,007	-0,042	0,020	0,000	-0,037	0,023	-0,035	
X8	0,020	-0,007	-0,020	-0,032	-0,019	-0,003	-0,014	0,028	
X9	-0,012	-0,022	0,001	0,008	-0,026	0,044	-0,027	-0,006	
X10	-0,031	0,042	-0,029	-0,024	0,046	-0,015	0,001	0,027	
X11	0,032	-0,034	0,004	0,030	0,017	0,055	0,037	0,019	
X12	0,020	0,043	-0,026	-0,011	-0,031	-0,019	0,005	0,016	
Bias	-0,035	-0,003	0,029	-0,005	-0,058	0,029	0,020	-0,025	
Çıktı katmanı (Manipüle Edilmiş) Sigmoid	-0,022	0,026	0,040	-0,010	-0,014	0,014	-0,009	-0,011	-0,008
Çıktı katmanı (Manipüle Edilmemiş) Sigmoid	-0,014	0,008	-0,022	0,014	0,002	-0,031	-0,001	-0,007	0,027

FORECASTING STOCK PRICE MANIPULATION IN FINANCIAL MARKETS USING MACHINE LEARNING METHODS: THE CASE OF BORSA ISTANBUL

Extended Abstract

Aim: This study was conducted to determine the factors that cause manipulation by using financial table data and data that is not based on financial statements of the businesses that the Capital Markets Board determined that manipulation was made and the same number of businesses that were not manipulated. Thus, the factors that cause manipulation can be determined 3-9 months before the date of manipulation, so that the necessary measures can be taken without further manipulation for the benefit of business managers, market, regulatory and supervisory authorities, policy makers, relevant public authorities and investors.

Methods: This study investigates firm-specific factors that can be used to identify similar characteristics of businesses that are more likely to be manipulated. Weekly bulletins are published on the Capital Markets Board (CMB) website. In the weekly bulletins, the information on the manipulation cases determined by the CMB, the shares that have been manipulated, the period of manipulation, the person/persons who have been manipulating, the administrative fines imposed on these persons, the criminal announcements made to the public prosecutor's office and the prohibition of transactions about these persons are regularly shared. The manipulations and the dates of manipulations were taken from the weekly bulletins of the Capital Markets Board between 2010-2019. Financial statements and non-financial data of the manipulated companies were obtained from the Public Disclosure Platform. Data set with 12 variables was used as the independent variable.

In the study, data was divided into 70% training data and 30% test data before using 10-fold cross validation for ANN and CART methods. Data set consisting of 70% of the training was divided into 10 parts by not showing 30% test set to algorithm, 9 parts were analyzed as training data and 1 part as test data. Weights were assigned to the variables and the importance of each variable was determined and the learning of the model was completed. After determining the weights and the completion of the learning, 30% test set was used which was never seen before by the algorithm and the performance of the models was validated. Data set was divided into 70% training data and 30% test set with 12 variables of C5.0 algorithm, and the model with the best performance was determined as a result of parameter optimization. Since the prediction results of logistic regression analysis conducted with 12 variables of LR analysis were low, the logistic regression analysis was performed with 3 variables determined by forward stepwise discriminant analysis.

Finding: The highest performance models with parameter optimization for ANN, CART and C5.0 analyzes were determined in the study. As a result of the analysis, the data set was divided into 70% training and 30% test sets. In the analysis carried out with 10-fold cross validation, ANN model correctly predicted 12 of the 15 manipulated businesses and 14 of the 15 non-manipulated businesses in the test sample. ANN's overall classification accuracy was found to be 86.67%. The CART decision tree model correctly predicted 13 of the 15 manipulated businesses in the test sample and correctly predicted 11 of the 15 non-manipulated businesses. The average classification accuracy of the CART Decision Tree was 80%. C5.0 Decision Tree Algorithm correctly predicted 14 of the 15 manipulated stocks, while a sample was actually manipulated but was found to be non-manipulated by the wrong classification. 9 out of 15 non-manipulated stocks were predicted correctly and 6 stocks were incorrectly predicted as manipulated even though they were not actually manipulated. The overall classification accuracy of C5.0 was found to be 76.67%. The LR analysis correctly predicted 11 out of 15 manipulated stocks, while 4 samples were actually manipulated but found to be non-manipulated by incorrect classification. 10 of the 15 non-manipulated stocks were predicted correctly, and 5 stocks were incorrectly predicted as

manipulated even though they were not actually manipulated. The overall classification accuracy of LR was found to be 70.00%.

Conclusions: In this study, it has been predicted whether or not market manipulation will be performed on the stock shares of 102 businesses traded on Borsa Istanbul prior to 3-9 months using quantitative and qualitative data of businesses. The performance of the related methods were compared by using data mining methods such as ANN, CART, C5.0, LR. As a result of the analyzes, the overall classification accuracy was found as ANN (86.67%), CART (80.00%), C5.0 (76.67%) and LR (70.00%), respectively. The C5.0 algorithm predicted 14 of the 15 manipulated stocks correctly and achieved 93.33% classification accuracy. ANN predicted 14 of the 15 non-manipulated stocks correctly and achieved 93.33% classification accuracy. In this study, since all methods have over 70% prediction performance, related models can also be included in stock manipulation prediction analyzes.

As a result, there is a variety of opportunities for mutual dialogue and enrichment on finance, accounting and auditing technologies, and the need to continue research to continue these discussions in the future is evident. Subsequent studies on linking insider trading and e-mails and analysis of social networks, including transaction patterns in Article 107/1 of the CMB, will add value to the literature.