



Pay Senedi Fiyat Yönünün Makine Öğrenmesi Yöntemleri ile Tahmini: Borsa İstanbul Örneği

Baris Aksoy^a

Öz: Bu çalışmada Borsa İstanbul 30 Endeksinde ve Kurumsal Yönetim Endeksinde bulunan beş imalat sanayi şirketinin 2010/3 ve 2020/3 dönemine ait verilerle üç ay sonraki ortalaması alınmış pay senedi fiyat yönü tahmin edilmiştir. Veri seti, örnek kapsamındaki şirketlerin üçer aylık dokuz mali tablo verisi ve üç aylık ortalaması alınmış beş makroekonomik değişkenden oluşmaktadır. Tahmin yöntemleri olarak Yapay Sinir Ağları, Sınıflandırma ve Regresyon Ağacı ile K- En Yakın Komşu Algoritması kullanılmıştır. Çalışmada kullanılan tüm yöntemlerde 10 katlı çapraz doğrulama yöntemi kullanılmıştır. Yapay Sinir Ağları ile Sınıflandırma ve Regresyon Ağacı analizlerinde en iyi parametrelerin ve performans kriterlerinin belirlenmesi fonksiyonu kullanılarak verilen parametre aralıkları doğrultusunda en iyi sonuç veren modeller elde edilmiştir. Analiz sonucuna göre pay senedi fiyat yönünü Yapay Sinir Ağları %98,05, Sınıflandırma ve Regresyon Ağacı %96,10 ve K- En Yakın Komşu Algoritması %92,20 oranında genel sınıflandırma doğruluğu elde etmiştir. Sınıflandırma ve Regresyon Ağacının oluşturulmasında verileri ikiye bölen önemli değişkenler olarak "Net Kâr Marjı", "Fiyat/Kazanç Oranı", "Hisse Başına Kâr Oranı", "CDS Primi (üç aylık ortalama)" ve "Tüketici Güven Endeksi" bulunmuştur. Bu bulgular, yatırımcıların bu çalışmada kullanılan modelleri kullanabileceklerini ve/veya hali hazırda kullanmış oldukları modellere dâhil edebileceklerini göstermektedir.

Anahtar Sözcükler: Pay Senedi Fiyat Yönü Tahmini, Borsa İstanbul, Yapay Sinir Ağları, K- En Yakın Komşu Algoritması, Sınıflandırma ve Regresyon Ağacı

JEL: C45, C51, E27, E44, G11

Geliş : 23 Temmuz 2020

Düzeltilme : 15 Ekim 2020

Kabul : 12 Kasım 2020

Tür : Araştırma

Predicting Direction of Stock Price Using Machine Learning Techniques: The Sample of Borsa İstanbul

Abstract: In this study, stock price with the calculated next three-month average of five manufacturing industry companies in the Borsa İstanbul 30 Index and the Corporate Governance Index was predicted with the data of the 2010/3 and 2020/3 periods. The dataset consisted of quarterly nine financial statements and five macroeconomic variables with a three-month average of the sample companies. Artificial Neural Networks, Classification and Regression Tree, and K- Nearest Neighbor Algorithm were used as prediction methods. A 10-fold cross-validation method was used in all methods in the study. In Artificial Neural Networks, Classification and Regression Tree analysis, the models that gave the best results in line with the given parameter ranges were obtained by using the determining the best parameters and performance criteria function. According to the results of the analysis, general classification accuracy was achieved 98.05% for Artificial Neural Networks, 96.10% for Classification and Regression Tree, and 92.20% K-Nearest Neighbor Algorithm. "Net Profit Margin", "Price/Earning", "Profit Per Share", "CDS Premium (3-month average)", "Consumer Confidence Index" were found as important variables that divided the data into two in the creation of the Classification and Regression Tree (CART) analysis. This result shows that the models used in this study can be incorporated into the models used by investors.

Keywords: Prediction of Stock Price Direction, Borsa İstanbul, Artificial Neural Networks, K- Nearest Neighbor Algorithm, Classification and Regression Tree

JEL: C45, C51, E27, E44, G11

Received : 23 July 2020

Revised : 15 October 2020

Accepted : 12 November 2020

Type : Research

Cite this article as: Aksoy, B. (2021). Pay senedi fiyat yönünün makine öğrenmesi yöntemleri ile tahmini: Borsa İstanbul örneği. *Business and Economics Research Journal*, 12(1), 89-110. <http://dx.doi.org/10.20409/berj.2021.312>

The current issue and archive of this Journal is available at: www.berjournal.com

^a Asst. Prof., PhD., Sivas Cumhuriyet University, Faculty of Economics and Administrative Sciences, Department of Banking and Finance, Sivas, Türkiye, baksoy@cumhuriyet.edu.tr (ORCID ID: 0000-0002-1090-5693)

1. Giriş

Yatırımcılar, ellerindeki fazla fonlarını finansal piyasalarda gelir getiren araçlarda değerlendirmektedirler. Yatırımcıların amacı, sahip oldukları atıl fonlarını bekledikleri düzeyde getiri sağlayacak finansal varlıklara yatırmaktır. Finansal piyasalarda alınan yanlış yatırım kararları, kaynakların verimli alanlara yönelmemesine neden olmaktadır. Bu nedenle pay senedi fiyatlarının doğru bir şekilde tahmin edilmesi önemlidir (Özçalıcı, 2017). Politik olaylar, ekonomik koşullar, alım/satım yapanların beklentileri ve pay senedi fiyatlarını etkileyebilecek diğer çevresel faktörler gibi çok fazla faktör bulunduğundan pay senedi fiyat yönünün tahmini oldukça karmaşık ve çok zor bir iştir. Ayrıca, pay senedi fiyat serileri genellikle oldukça gürültülü, dinamik, karmaşık olan ve parametrik olmayan bir yapıya sahiptir. Analiz yapılacak olan verilerde gürültülü verinin bulunması, finansal piyasaların gelecekteki ve geçmiş fiyatlar arasındaki bağımlılığı tam olarak yakalayamaması anlamına gelmektedir. Çalışmaların çoğu pay senedi fiyatının doğru tahminine odaklanmıştır. Ancak, farklı yatırımcılar farklı alım/satım stratejileri benimseyebilmektedir. Bu nedenle, gerçek değerler ile tahminler arasındaki hatanın en aza indirilmesine dayanan tahmin modeli her yatırımcı için uygun olmayabilir. Bunun yerine, pay senedi fiyatı hareket yönünün doğru tahmini, etkili pay senedi alım/satım stratejileri yapabilmeleri için çok önemlidir. Özellikle, yatırımcılar potansiyel piyasa riskine karşı etkili bir şekilde korunabilir ve spekülörler pay senedi fiyatı yönünün doğru tahminini elde edebildiklerinde kâr elde etme şansına sahip olabilirler (Imandoust ve Bolandraftar, 2014).

Öngörü, bir değişkenin belirli varsayımlar altında gelecek bir tarihte alabileceği değerlerin yaklaşık olarak önceden belirlenmesi şeklinde tanımlanabilir. Nicel ve nitel öngörü yöntemlerinin her ikisinin de çıkış noktasını, geçmiş ve bugünkü gözlem değerlerinden gelecek dönem gözlem değerlerinin belirli kurallar ışığında tahmin edilmesi oluşturmaktadır (Ataseven, 2013). Finansal piyasalar son derece oynaktır ve günlük bazda büyük miktarda veri üretmektedir. Borsa tahmini, her zaman araştırmadaki en sıcak konulardan biri olmasının yanı sıra karmaşık ve değişken yapısı nedeniyle zor bir çalışma alanıdır. Düşük fiyatla satın alma ve sonra yüksek fiyatla satış zamanlama yönünden önemli bir sorundur. Bu sorunu çözmek için pay senedi fiyatları eğiliminin öngörülmesi önemlidir (Anbalagan ve Maheswari, 2015). Pay senedi fiyat yönünün tahmini, hem borsada pay senedi alım satımı yaparak kâr etmek isteyenler hem de borsa verilerinde saklanan bilgileri ortaya çıkarmaya çalışan araştırmacılar için büyük ilgi alanı olmuştur. Veri madenciliği tekniklerinin borsa tahmini için uygulamaları, son zamanlarda büyük ilgi gören bir araştırma alanıdır (Nair, Mohandas ve Sakthivel, 2010). Borsa endeksi gibi finansal zaman serilerini tahmin etme konusunu ele alan makaleler olmasına rağmen, ampirik bulguların çoğu gelişmiş finansal piyasalarla (İngiltere, ABD ve Japonya) ilgilidir. Bununla birlikte, literatürde gelişmekte olan piyasalarda pay senedi hareketinin yönünü tahmin etmek için çok az araştırma bulunmaktadır (Imandoust ve Bolandraftar, 2014). Finansal zaman serilerinin geleneksel yöntemlerle analiz edilmesi yanında, yapay zekâ ve makine öğrenmesi algoritmaları ile analizi, son yıllarda oldukça popüler olmuştur (Papuçcu, 2019). Ayrıca fiziksel finans (ekonofizik) üzerine yapılan çalışmaların son yıllarda önemli düzeyde arttığı gözlenmektedir.

Bu çalışmanın birinci amacı; Borsa İstanbul (BIST) 30 Endeksinde ve BIST Kurumsal Yönetim Endeksinde bulunan BIST İmalat Sanayi sektöründe faaliyet gösteren beş işletmenin 2010/3-2020/3 döneminde üç aylık mali tablo verilerinden oluşan dokuz değişken ve üç aylık ortalaması alınmış ekonomik göstergelerden oluşan beş değişkenle üç ay sonraki fiyat ortalaması alınmış pay senedi fiyat hareketleri yönünün tahmin edilmesidir. Çalışmada kullanılan Yapay Sinir Ağı (ANN, Artificial Neural Network), Sınıflandırma ve Regresyon Ağacı (CART, Classification and Regression Tree) ve K-En Yakın Komşu Algoritması (KNN, K Nearest Neighbor) olmak üzere üç veri madenciliği yöntemi performansının, yöntemlerin ayırt edici özellikleri altında karşılaştırılması bu çalışmanın ikinci amacıdır. Çalışma dönemini içeren 10 yıllık sürede üç aylık mali tablo verileri ve ekonomik göstergelerin üç aylık ortalamaları kullanılarak bir sonraki üç aylık fiyat ortalamaları yönünün tahmin edilmesi üzerine bir çalışmaya literatürde rastlanmamıştır. Bu anlamda çalışmanın literatüre katkısı olacağı düşünülmektedir. Çalışma beş bölümden oluşmaktadır. Birinci bölümde giriş, ikinci bölümde literatür incelemesi, üçüncü bölümde yöntemler, dördüncü bölümde bulgular ve beşinci bölümde sonuç yer almaktadır.

2. Literatür İncelemesi

Zaman serisi analiz yöntemlerinde, serilerin yapısı ile ilgili varsayımlar ve pay senedi fiyatı ile değişkenler arasında doğrusallık varsayımı gibi metodolojik sorunlar vardır. Makine öğrenmesi yöntemleri gibi sınırlayıcı varsayımın az olduğu esnek yöntemlerin kullanımı gün geçtikçe armaktadır (Özdemir, Tolun ve Demirci, 2011). Aşağıda bu alandaki araştırmalardan genel hatları ile bahsedilecektir.

Huang, Nakamori ve Wang (2005) çalışmalarında SVM ile NIKKEI 225 endeksinin hareket yönünün öngörülebilirliğini araştırmışlardır. SVM'nin (Support Vector Machine; Destek Vektör Makineleri) tahmin yeteneğini değerlendirmek için, performansını Lineer Diskriminant Analizi ve Elman Geri Yayımlı Sinir Ağları ile karşılaştırmışlardır. Ampirik analizde NIKKEI 225 haftalık değişikliklerini incelemişlerdir. SVM'yi diğer sınıflandırma yöntemleriyle birleştirme modeli önermişlerdir. Analiz sonucunda SVM'nin diğer sınıflandırma yöntemlerinden daha iyi performans gösterdiği ve birleştirme modelinin, tüm öngörme yöntemleri arasında en iyi performansı gösterdiği belirtilmiştir.

Nair, Mohandas ve Sakthivel (2010) çalışmalarında, Bombay Menkul Kıymetler Borsası'ndaki (BSESENSEX) bir sonraki günün trendini tahmin etmek için hibrit bir karar ağacı temelli sistemin tasarım ve performans değerlendirmesini sunmuşlardır. Hibrit kaba küme tabanlı sistemin performansı, yapay sinir ağı tabanlı trend tahmin sistemi ve Naif Bayes tabanlı trend tahmincisi ile karşılaştırılmaktadır. Çalışmada 21 farklı teknik gösterge kullanmışlardır. Trend tahmin sistemlerinin performansı 10 kat çapraz doğrulama yöntemi kullanılarak değerlendirilmiştir. Önerilen hibrit karar ağacı kaba küme temelli eğilim tahmin sisteminin %90,22'lik bir doğruluk sağladığı görülmektedir. Tek başına kaba küme temelli eğilim tahmin sistemi, herhangi bir değişken seçimi olmadan %88,18 olarak, yapay sinir ağı tabanlı tahmin sistemi değişken seçimi olmadan, %77,66'lık bir doğruluk, Naive Bayes tabanlı trend tahmincisi ise %72,36'lık performans sağladığı bulunmuştur.

Ulusoy (2010) çalışmasında 1997-2000 yılları arasında İstanbul Menkul Kıymetler Borsası'nda (İMKB) gerçekleşen borsa endeks değerini tahmin etmiştir. Araştırmacı, İMKB'de gerçekleşen borsa endeks değerinin yönünün tahmin edilmesinde 13 değişkenli bir ANN sistemi kurmuş ve Hatayı Geriye Yayma Algoritması kullanmıştır. Hesaplamalar İMKB Nöral Ağ Simülatörü adı verilen bir programla yapılmıştır. Araştırmada aşırı artış ve azalış olmayan ve siyasi etkilerin olmadığı günlerde daha başarılı çıktı değerleri verdiği bulgusuna ulaşılmıştır. Simülasyon sonucu 1997 yılında %5,48, 1998 yılında %6,05, 1999 yılında %5,34, 2000 yılında ise %6,62 hata ortalamaları verdiği bulunmuştur. Genel hata sonuçlarının %0,5- %33 arasında salınım göstermesi bu simülasyonun kısmen de olsa başarıya ulaştığının bir göstergesi olarak değerlendirilmiştir. Oluşturulan ağ mimarisi ve simülasyonun Türk finans dünyası için kullanılabilirliğinin bir göstergesi olduğu belirtilmiştir. Borsada gerçekleşen fiyatların tahmininde kullanılan ANN algoritmalarının geliştirilerek sağlıklı sonuçlar alınmasının mümkün olduğu bulgusuna ulaşılmıştır.

Kara, Boyacıoğlu ve Baykan (2011) çalışmalarında iki model geliştirmişler ve günlük BIST Ulusal 100 Endeksinin hareket yönünü tahmin etme konusundaki performansları karşılaştırmışlardır. Söz konusu çalışmanın temel amacı ANN ve Destek Vektör Makineleri (SVM, Support Vector Machine) kullanarak günlük BIST Ulusal 100 Endeksinde hareket yönünü tahmin etmektir. Veri seti 2 Ocak 1997 ile 31 Aralık 2007 arasındaki dönemi kapsamaktadır. Önerilen modellerin girdisi olarak 10 teknik gösterge seçilmiştir. Tahmin performanslarını artırmak amacıyla her iki model için parametre belirleme deneyi yapılmıştır. Deney sonuçları, ANN modelinin ortalama performansının (%75,74), SVM modelinden (%71,52) önemli ölçüde daha iyi bulunduğunu göstermiştir.

Imandoust ve Bolandraftar (2014) çalışmalarında Karar Ağacı (DT), Rastgele Orman (RF) ve Naive Bayesian (NBC) olmak üzere üç model geliştirmişler ve günlük Tahran Menkul Kıymetler Borsası (TSE) endeksindeki hareket yönünü tahmin modellerinin performanslarını karşılaştırmışlardır. Önerilen modellerin girdisi olarak 10 mikro ekonomik değişken ve üç makroekonomik değişken seçilmiş ve performansları 2007-2012 yılları arasındaki günlük verilerle karşılaştırılmıştır. Araştırmada teknik ve temel değişkenler üç bölüme ayrılmıştır. İlk olarak, girdi değişkenleri olarak 10 teknik gösterge NBC'ye verilmiştir. Daha sonra araştırmada kullanılan üç temel değişken modele uygulanmıştır. Son olarak, modele 13 temel-teknik değişken uygulanmıştır. Deneysel sonuçlar, Karar Ağacı modelinin (%80,08) performansının Rastgele Orman (%78,81)

ve Naive Bayesian Sınıflandırıcı'dan (%73,84) daha iyi olduğunu göstermektedir. Karar Ağacı modelinin (hem teknik hem de teknik-temel değişkenlerde %80,08 doğrulukla) diğer tekniklerden daha iyi performans gösterdiği belirtilmiştir.

Patel, Shah, Thakkar ve Kotecha (2015) çalışmalarında Hindistan borsaları için pay senedi hareket yönünü ve pay senedi fiyat endeksini tahmin etmişlerdir. Söz konusu çalışmada, Hint borsalarından CNX Nifty, S&P BSE Sensex, Infosys Ltd. ve Reliance Industries'in 10 yıllık geçmiş verilerine (2003-2012) dayanılarak pay senedi ve pay senedi fiyat endeksi hareketini tahmin etmek için ANN, SVM, Random Forest ve Naive-Bayes algoritmalarının tahmin performansı karşılaştırılmıştır. Çalışmada 10 teknik parametre kullanılmıştır. Çalışmada sürekli değerli girdileri ayrık girişlere dönüştüren Trend Deterministik Veri Hazırlama Katmanı önerilmektedir. Sürekli değerli verilerle yapılan deneyler, Naive Bayes (Gauss işlemi) modelinin %73,30 doğrulukla en az performans ve Rastgele Orman'ın %83,56 doğrulukla en yüksek performans sergilediğini göstermektedir. Modellerin performansı, trend belirleyici veriler kullanıldığında önemli ölçüde iyileştirilmektedir. ANN %86,69, SVM %89,33, Random Forest %89,98 ve Naive Bayes %90,19'lük doğruluğa sahip olduğu belirtilmiştir.

Chen ve Hao (2017) çalışmalarında Şangay ve Shenzhen borsa endekslerini etkili bir şekilde tahmin edebilmek için, özellik ağırlıklı K-En Yakın Komşunun yanı sıra, özellik ağırlıklı Destek Vektör Makinesinin temel melezleştirilmiş bir çerçevesini önermektedir. Analiz sonucunda FWSVM (Feature Weighted SVM, Özellik Ağırlıklı SVM)-FWKNN (Feature Weighted KNN, Özellik Ağırlıklı KNN) pay senedi analiz algoritmasının, karşılaştırılan diğer modellere göre daha iyi tahmin yeteneğine sahip yöntem olduğu belirtilmiştir. FWSVM tahmin doğruluğu sırasıyla 1, 5, 10, 15, 20, 30; günlük tahminler sırasıyla 92,6, 93,6, 95,0, 94,0, 94,4, 97,4 olarak bulunmuştur. FWSVM tahmin doğruluğu sırasıyla 96,0, 96,8, 97,6, 95,6, 95,8 ve 98,4 olarak bulunduğu belirtilmiştir.

Yakut ve Gemici, (2017) çalışmalarında BIST 100 Endeksindeki kimya, kauçuk ve plastik imalat sanayinde 2009-2014 döneminde faaliyet gösteren 18 şirketin pay senedi getirilerini tahmin etmişlerdir. Çalışmada 20 finansal oran ve TÜFE değişkeni kullanılarak Lojistik Regresyon (LR), C5.0 algoritması, CART algoritması ve SVM yöntemi ile analizler gerçekleştirilmişlerdir. C5.0 algoritması ile 12 ve CART algoritması ile 15 kural üretmişlerdir. Analizler sonucunda LR %75, C5.0 algoritması %88, CART algoritması %89,8 ve SVM analizi %75,9 oranında doğru sınıflandırdığı belirtilmiştir. Pay senedi getirisinin pozitif ve negatif sınıflandırma tahmininde en önemli değişkenlerin "piyasa/defter değeri", "TÜFE değişkeni" ve "brüt kâr marjı" olduğu belirtilmiştir.

Papuçcu (2019) çalışmasında BIST 100 endeksi hareket yönünü tahmin etmiştir. Bu çalışma ile yapay sinir ağları, destek vektör makineleri ve Naive Bayes algoritmaları kullanılarak borsa endeksi hareketleri tahmin edilmiştir. Borsa endeksi tahminlerinde 10 adet teknik gösterge kullanmışlardır. Veri seti 2009-2018 dönemi arasındaki günlük kapanış değerlerinden oluşmuştur. Çalışmanın eğitim aşaması için 10 katlı çapraz doğrulama yöntemi kullanılmıştır. Tahmin sonuçlarının sırasıyla ANN (%99,10), SVM (%99,10) ve NB (%90,40) olarak bulunduğu belirtilmiştir.

Finansal piyasalardaki hareketlerin doğrusal olmaması nedeniyle gelecekteki pay senedi getirileri veya endeksleri hakkında daha iyi tahmin sonuçları elde etmek için, araştırmacılar ve finansal analistler tarafından makine öğrenmesi yöntemleri kullanılarak gerçekleştirilen analizler giderek yaygınlaşmaktadır (Enke ve Thawornwong, 2005).

3. Yöntem ve Veri

Makine öğrenmesi, matematik ve istatistik yöntemlerini kullanarak verilerden elde ettiği çıkarımlarla tahminlerde bulunan modelleme ve algoritmalarından oluşan yapay zekânın bir alt dalıdır. Makine öğrenmesinin amacı doğru kestirimler yapmaktır. Bunu yaparken kestirim fonksiyonlarını yorumlamak ve belirli olasılık modeli ile ilişkilendirmek zor olabilmektedir. Denetimli öğrenme, eğitim veri setine dayalı olarak makine öğrenme modelinin oluşturulması sürecidir. Denetimli öğrenmede verilerin büyük bir kısmından oluşan eğitim verisi kullanılarak algoritmanın öğrenmesi tamamlanmakta ve öğrenme aşaması, test verileri

kullanılarak denetlenmektedir. Denetimli makine öğrenmesi genel olarak tahmin ve kestirim problemleri üzerinde yoğunlaşmaktadır (Akay, 2018).

3.1. Veri Seti ve Araştırmanın Kısıtları

2010/3-2020/3 döneminde üçer aylık mali tablolarına düzenli olarak ulaşılabilen ve aynı sektörde faaliyet gösteren imalat sanayi şirketleri çalışmaya dâhil edilmiştir. Borsa İstanbul'da işlem gören perakende, finans şirketleri, yatırım şirketleri, leasing şirketleri, bankalar ve sigorta şirketlerinin bilanço yapıları farklılık gösterdiğinden bu sektörlerdeki şirketler çalışmaya dâhil edilmemiştir. Çalışmanın sınırlandırılması amacıyla BIST 30 Endeksi ve BIST Kurumsal Yönetim Endeksinde bulunan altı imalat sanayi işletmesi verilerinden Aselsan şirketinin 2010/3-2020/3 döneminin tamamına ait üç aylık mali tablo verilerine ulaşamadığından nihai olarak beş şirket verisi ile analizler gerçekleştirilmiştir. Örnek kapsamındaki beş şirketin 2010/3-2020/3 arasında 41 adet üçer aylık mali tablo verisi ve makro ekonomik değişkenlerle yürütülen bu çalışmada elde edilen bulguların (veri setinin kısıtlı olması nedeniyle) BIST'de işlem gören tüm şirketlere genellenmesi uygun değildir. Bu çalışmanın en önemli kısıtı örnek sayısının sınırlandırılmış olmasıdır. Tablo 1'de çalışma örneklemini oluşturan şirketler verilmektedir.

Tablo 1. Çalışma Örneklemini Oluşturan Şirketler

Kodu	Unvanı
ARCLK	ARÇELİK A.Ş.
EREGL	EREĞLİ DEMİR VE ÇELİK FABRİKALARI T.A.Ş.
SISE	TÜRKİYE ŞİŞE VE CAM FABRİKALARI A.Ş.
TOASO	TOFAŞ TÜRK OTOMOBİL FABRİKASI A.Ş.
TUPRS	TÜPRAŞ-TÜRKİYE PETROL RAFİNERİLERİ A.Ş.

Pay senedi fiyat yönü tahmini alanında yapılan literatür incelemesinde en çok kullanılan değişkenler bu çalışmaya dâhil edilmiştir. Araştırmada Tablo 2'de belirtilen değişkenlerin seçilmesinin diğer bir nedeni, belirlenen 14 girdi değerine Ulusoy (2010)'un çalışması takip edilerek araştırmacıların kolaylıkla ulaşabilmelerinin amaçlanmış olmasıdır. Bu çalışmada kullanılan değişkenler ile söz konusu değişkenleri çalışmalarında kullanan araştırmacılar Tablo 2'de verilmiştir.

Tablo 2. Çalışmada Kullanılan Değişkenler ile Bu Değişkenlerin Kullanıldığı Çalışmalar

No	Değişkenler	Hesaplanması	Bu Çalışmada Kullanılan Değişkenleri Çalışmalarında Kullanan Araştırmacılar
X1	Likidite Oranı	(Dönen Varlıklar-Stoklar)/KVYK	Oliveira, Nobre ve Zarate (2013), Barak ve Modarres (2015), Yakut ve Gemici (2017)
X2	Aktif Devir Hızı	Net Satış /Aktif Toplam	Barak ve Modarres (2015); Yakut ve Gemici (2017)
X3	Aktif Karlılık Oranı	Net Kar/Aktif Toplam	Akcan ve Kartal (2011), Barak ve Modarres (2015), Yakut ve Gemici (2017)
X4	Faaliyet Kâr Marjı	Faaliyet Karı /Net Satışlar	Barak ve Modarres (2015), Yakut ve Gemici (2017)
X5	Net Kâr Marjı	Net Kar/Net Satışlar	Oliveira vd. (2013), Barak ve Modarres (2015), Yakut ve Gemici (2017)

Tablo 2. Çalışmada Kullanılan Değişkenler ile Bu Değişkenlerin Kullanıldığı Çalışmalar (Devamı)

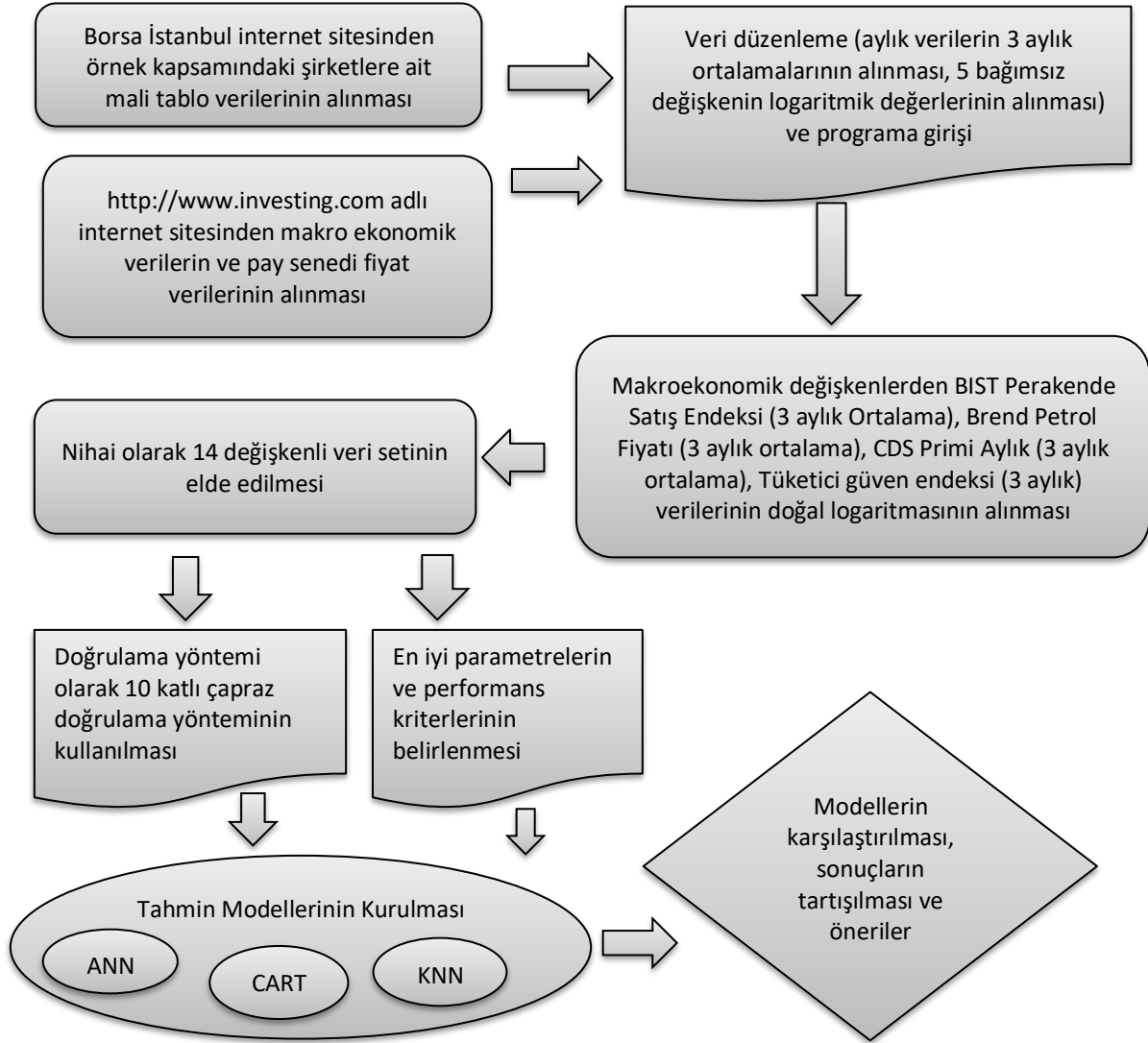
No	Değişkenler	Hesaplanması	Bu Çalışmada Kullanılan Değişkenleri Çalışmalarında Kullanan Araştırmacılar
X6	Kaldıraç Oranı	Toplam Borç/Aktif Toplam	Oliveira vd. (2013), Barak ve Modarres (2015), Yakut ve Gemici (2017)
X7	Fiyat/Kazanç	(Hisse Senedi Fiyatı x ödenmiş Sermaye)/(Net Dönem Karı x1000)	Akcan ve Kartal (2011), Oliveira vd. (2013), Yakut ve Gemici (2017)
X8	Piyasa/Defter Değeri	(Hisse Senedi Fiyatı x Ödenmiş Sermaye/1000)/Özsermaye	Akcan ve Kartal (2011), Yakut ve Gemici (2017)
X9	Hisse Başına Kâr Oranı	Net Kar/Ödenmiş Sermaye	Akcan ve Kartal (2011), Oliveira vd. (2013)
X10Ln	BIST Perakende Satış Endeksi (3 aylık Ortalama)	Aylık verilerin 3 aylık aritmetik ortalaması alınmıştır.	Oliveira vd. (2013)
X11Ln	Brend Petrol Fiyatı (3 aylık ortalama)	Aylık verilerin 3 aylık aritmetik ortalaması alınmıştır.	Oliveira vd. (2013)
X12Ln	CDS Primi Aylık (3 aylık ortalama)	Aylık verilerin 3 aylık aritmetik ortalaması alınmıştır.	Yazar tarafından eklenmiştir.
X13Ln	Tüketici güven endeksi (3 aylık)	Aylık verilerin 3 aylık aritmetik ortalaması alınmıştır.	Oliveira vd. (2013)
X14Ln	TÜFE (3 Aylık)	Aylık verilerin 3 aylık aritmetik ortalaması alınmıştır.	Akcan ve Kartal (2011), Oliveira vd. (2013), Yakut ve Gemici (2017)
X15	Bağımlı Değişken (Bir önceki döneme göre düşüş "0", bir önceki döneme göre yükseliş "1")		Yazar tarafından eklenmiştir.

3.2. Çalışma Metodolojisi

Bu çalışmada verilerin alındığı 2020 yılında BIST 30 endeksinde ve Kurumsal Yönetim Endeksinde bulunan beş imalat sanayi şirketinin 2010/3-2020/3 dönemindeki üç aylık fiyat ortalaması bir önceki üç ay ortalama fiyata göre yükseliş olduğunda "1", düşüş olduğunda ise "0" olarak kodlanmıştır. BIST 30 Endeksinde ve Kurumsal Yönetim Endeksi'nde bulunan beş imalat sanayi şirketine ait 41 adet üç aylık mali tablo verisi (dokuz değişken) ve makroekonomik (beş değişken) olmak üzere toplam 14 bağımsız değişken programa girilmiştir. 205 örneğin ve 14 bağımsız değişkenin kullanıldığı çalışmada eğitim verisinin sınırlı sayıda olması nedeniyle 10 katlı çapraz doğrulama yöntemi kullanılmıştır. 205 örneğin 136'sı bir sonraki üç aylık ortalama fiyatı yükselen 69'u ise bir sonraki üç aylık ortalama fiyatı düşen örneklerden oluşmaktadır.

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. Bu çalışmada ANN, CART ve KNN analizleri için RAPİDMİNER 9.7 programı kullanılmıştır. Şekil 1'de çalışmanın kavramsal yapısı verilmiştir.

Şekil 1. Çalışmanın Kavramsal Yapısı



Zaman serisi, zaman sırasına göre dizilmiş gözlem değerleri şeklinde tanımlanabilir. Zaman serisinde incelenen özellik bir değişkendir. Bu değişken zaman içerisinde çeşitli nedenlerle farklı değerler almaktadır. Dolayısıyla zaman serisi, gözlemlerin belirli bir dönem için gün, hafta, ay, üç ay, yıl gibi birbirini izleyen eşit aralıklarla yapılması ile elde edilen seriler zaman serileri olarak adlandırılmaktadır. Zaman serisi analizi, tahmin edilecek değişkene ait geçmiş zaman değerleri kullanılarak aynı değişkenin gelecekteki değerlerinin tahmin edilmesi sürecinde model geliştirmek için kullanılmaktadır. Model geliştirme, tahmin edilmesi beklenen değişkene ait zaman serisinin analiz edilmesi, serinin temel eğiliminin ve özelliklerinin belirlenmesine dayanmaktadır (Ataseven, 2013). Bu çalışmada kullanılan veri yapısının zaman serisi şeklinde değerlendirilmesi, analiz süreçlerinin yanlış anlaşılmasına neden olabilecektir. Bu nedenle analiz verisi hakkında kısa bir açıklama yapılması ihtiyacı doğmaktadır.

Çalışma metodolojisi başlığı altında belirtildiği üzere araştırma, 2010/3-2020/3 dönemindeki aritmetik ortalaması alınmış üç aylık fiyat değeri aynı şirketin pay senedine ait üç ay önceki aritmetik ortalaması alınmış bağımsız değişkenler ile tahmin edilmesine dayalıdır. Tablo 3'de veri yapısı verilmektedir. Kısaca her şirketin 41 dönemlik yatay kesit veri seti art arda sıralanmıştır. Zaman serisinde bir tek değişkenin günlük, haftalık, aylık, üç aylık gözlemleri zamana bağlı olarak sıralanmakta ve örneğin dönem sonu fiyatı tahmin edilmektedir. Bu çalışmada ise 5 farklı şirketin her birinin 41 dönemlik (3 aylık aritmetik ortalaması

alınmış fiyat) ortalama fiyatı 14 bağımsız değişken kullanılarak bir önceki 3 aylık fiyatına göre düşmüş ise 0, artmış ise 1 ile kodlanmış ve toplam 205 örneğe ulaşılmıştır. 3 farklı makine öğrenmesi yöntemi kullanılarak 3 ay sonraki pay senedi fiyat yönü tahmin edilmiştir. Eğitim esnasında doğrulama yöntemi olarak 10 katlı çapraz doğrulama yöntemi ve örnek seçim yöntemi olarak ise tabakalı örnek seçim yöntemleri kullanılmıştır. Dolayısı ile bağımlı değişken de her gruptan (fiyatı düşen 0, fiyatı yükselen 1 olarak) eşit sayıda örnek alınarak tüm veri seti 10 parçaya bölünmüş (veri yapısı şirket bazında değil 5 şirkete ait toplam 205 örneğin alt alta sıralandığı bir yatay kesit verisi olduğuna dikkat edilmeli) bir parçası test için ayrılarak diğer dokuz parça veri eğitim için kullanılmıştır. Örneğin eğitim verisi olarak yaklaşık 180 örnek tesadüfi olarak seçilmiş, kalan yaklaşık 20 örnek tesadüfi olarak test örneği olarak seçilerek birinci kat için test sonucu belirlenmiştir. Bu işlem 10 defa tesadüfi olarak farklı örnekler seçilerek gerçekleştirilmiş ve 10 farklı test sonucu elde edilerek ortalaması alınmış değer araştırmada paylaşılmıştır. Analiz aşamasının anlaşılması açısından örneğin eğitim için 1, 7, 12, 26, 34, 43, 48, 57, 58, 65, 70, 73, 76, 79, 84, 96, 100, 121, 134, 145, 167, 189, 194, 205... şeklinde örnekler seçildiği düşünülürken 1-34 arasındaki örnekler ARCLK şirketi pay senedi bilgilerini içermekte, 43-79 arasındaki örnekler EREGL şirketi pay senedi fiyat bilgilerini içermektedir. Aynı yöntem test için seçilen örnekler için de düşünülebilir. Dolayısıyla ne eğitim verisi seçilirken ne de test verisi seçilirken veri dağılımında zaman serisi yapısı bulunmamakta (her satırdaki örnekler ayrı ayrı değerlendirildiğinden ve bir önceki satır ile bir sonraki satır arasında eğitim ve test örneklerinin seçilmesinde bağlantı bulunmadığına dikkat edilmeli) olup bir yatay kesit veri analizi gerçekleştirilmektedir.

Tablo 3. Veri Yapısı

Şirket	Sıra	Dönem	Bağımsız Değişken 1	...	Bağımsız Değişken 14	Bağımlı Değişken 15
ARCLK	1	2010.3	Değer	Değer	Değer	1
ARCLK	2	2010.6	Değer	Değer	Değer	0
.	.	.	Değer	Değer	Değer	.
.	.	.	Değer	Değer	Değer	.
ARCLK	41	2020.3	Değer	Değer	Değer	.
EREGL	42	2010.3	Değer	Değer	Değer	.
EREGL	43	2010.6	Değer	Değer	Değer	.
.	.	.	Değer	Değer	Değer	.
.	.	.	Değer	Değer	Değer	.
EREGL	82	2020.3	Değer	Değer	Değer	.
ŞİSE	83	2010.3	Değer	Değer	Değer	.
ŞİSE	84	2010.6	Değer	Değer	Değer	.
.	.	.	Değer	Değer	Değer	.
.	.	.	Değer	Değer	Değer	.
ŞİSE	123	2020.3	Değer	Değer	Değer	.
TOASO	124	2010.3	Değer	Değer	Değer	.
TOASO	125	2010.6	Değer	Değer	Değer	.
.	.	.	Değer	Değer	Değer	.
.	.	.	Değer	Değer	Değer	.
TOASO	164	2020.3	Değer	Değer	Değer	.
TUPRS	165	2010.6	Değer	Değer	Değer	.
TUPRS	165	2010.9	Değer	Değer	Değer	.
.	.	.	Değer	Değer	Değer	.
.	.	.	Değer	Değer	Değer	.
TUPRS	205	2020.3	Değer	Değer	Değer	.

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). Enke ve Thawornwong (2005)'ün çalışması takip edilerek bu çalışmada kullanılan 10 katlı çapraz doğrulama yöntemi şu şekilde açıklanabilir: Veri örneği rastgele on eşit boyutlu kısma bölünür ve ağ 10 kez eğitilir. Eğitim geçişlerinin her birinde, 10 kısma ayrılan eğitim verilerinden bir kat atlanır ve ortaya çıkan model, bir doğrulama seti olarak da bilinen bu atlanmış kattaki verilerle doğrulanır. Tahminleri güvenilir bir şekilde değerlendirmek için, yalnızca tahmin doğruluğu değil, duyarlılık ve kesinlik de göz önünde bulundurulmuştur. Tabakalı örnekleme (stratified sampling) seçiminde her iki sınıftan (3 aylık ortalama fiyatı yükselen/3 aylık ortalama fiyatı düşen) eşit sayıda ancak rastgele örnekler alınmaktadır. Bağımlı değişkenin kategorik olduğu durumlarda başarılı sonuç vermektedir (Liang, Tsai ve Wu, 2015). Bu nedenle bu çalışmada örnekleme seçiminde tabakalı örnekleme seçimi (stratified sampling) kullanılmıştır. Gerçek yön ve tahmini yön değişkenleri aynı değeri aldığı anda doğru tahmin gerçekleştirilmiş anlamına gelmektedir.

3.3. Çalışmada Kullanılan Yöntemler

3.3.1. ANN Analizi

Yapay sinir ağı modeli, girdi katmanı, çıktı katmanı ve gizli katmandan oluşmakta ve her katmanda nöronlar bulunmaktadır. Gizli katman birden fazla katmandan oluşabilmektedir. Çıktı katmanındaki nöron sayısı kullanılan çıktı kadar bulunmaktadır. Nöronların girdi değeri önceki katman nöronlarının çıktı değerlerini kullanmaktadır. Gizli katman ve çıktı katmanında bulunan nöronlar kendilerine gelen sinyalleri bir aktivasyon fonksiyonu kullanarak işlemekte ve kendisinden sonra katmanın bulunması durumunda bu katmana iletmektedir (Özçalıcı, 2017). ANN'de bilgi, bağlantı ağırlıklarında depolanmaktadır. Yapay sinir ağına öğrenme, istenen bir işlevi yerine getirecek şekilde bu bağlantı ağırlıklarının değiştirilmesi ile gerçekleşmektedir. ANN'da öğrenme sınırlar arasındaki ağırlıkların değiştirilmesi ile gerçekleşir. Buna göre sınırlar arası bağlantılar üzerindeki ağırlıkların öğrenme kuralları uyarınca dinamik olarak değiştirilebilen ağırlar eğitilebilir (Elmas, 2018). Çok katmanlı ileri beslemeli sinir ağı, üzerinde eğitim yapmak için kullanılan en popüler sinir ağı algoritmasıdır. Ağırlıklar daha sonra ağı tahmini ile gerçek hedef arasındaki ortalama kare hatasını en aza indirecek şekilde değiştirilir. Genellikle, gizli katmanda çok fazla nöronun bulunması ve çok fazla bağlantının olması, verileri ezberleyen ve genelleme yeteneğinden yoksun bir sinir ağı oluşturmaktadır. Aşırı öğrenmeyi önlemek için kullanılacak bir yaklaşım n-kat çapraz geçerliliklidir (Enke ve Thawornwong, 2005).

Bu araştırmada üç katmanlı ileri beslemeli model kullanılmıştır. Geriye yayılım ağı, katmanlar arasında tam bağlantının olduğu "çok katmanlı", "ileri beslemeli" ve "denetimli" olarak eğitilen bir ANN modelidir. Bu modelde girdi katmanı, problemin yapısına göre bir gizli katman ve çıktı katmanı bulunmaktadır. Girdiyi oluşturan örnekler ağı ilk katmanındaki düğümlerden çıktı katmanına ulaşıncaya kadar veriler üzerinde işlemler gerçekleştirilmektedir. İşlem elemanlarından her biri diğer işlem elemanlarından gelen bilgileri birleştirerek tek bir çıkış değeri ortaya koymaktadır. İşlem elemanı tüm girişlerin ağırlıklı toplamını oluşturmakta ve bir eşik değeri ile karşılaştırmaktadır. Girişlerin önceden belirlenmiş bir eşik değerini aşması durumunda nöronun çıktısı yüksek bir değere diğer durumda ise düşük bir değere atanmaktadır (Ulusoy, 2010). Bu işlemler sonucunda modelin verdiği çıktı gerçek çıktı ile karşılaştırılmakta ve farkın olduğu durumda her çıktı düğümü için hata sinyali hesaplanmaktadır. Hesaplanan bu hata sinyalleri çıktı düğümlerine karşılık gelen gizli katmanda bulunan düğümlere aktarılmaktadır. Gizli katmandaki düğümler hatanın sadece hesaplanan kısmını içermekte ve bu hata sinyalleri girdi katmanına iletilmektedir. Hata sinyalleri göz önünde bulundurularak bağlantı ağırlıkları her düğümde yeniden değiştirilmektedir (Akpınar, 2014). Gizli katmandaki nöron sayısı (n), öğrenme hızı değeri (η), momentum sabiti (mc) ve yineleme sayısı (ep) etkin bir şekilde belirlenmesi gereken ANN modeli parametreleridir (Kara vd., 2011). Nöron, önceki aşamada (n) bir ağırlıkla çarpılan ve daha sonra bir eşik b ile toplanan birkaç düğümden oluşan bir giriş olan x_i 'den oluşur. Transfer fonksiyonu, bir nöron çıkışını belirleyen matematiksel bir fonksiyonla hesaplanır (Gaganis, 2009):

$$in = \sum_{t=1}^n w_i \cdot x_i + b \quad (1)$$

ileri sürümlü ANN'de kullanılan ağırlıklar her seferinde Δw kadar düzeltilerek yenilenir.

$$W_1^{yeni} = w_1^{eski} + \Delta w_1 \quad (2)$$

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 y ile gösterilirse, en küçük kareler yöntemiyle elde edilecek hata fonksiyonu E şu şekilde hesaplanabilir (Silahtaroglu, 2016):

$$E_r = 1/2 e^2 = 1/2 (g - y)^2 \quad (3)$$

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 fiyat yönü tespit modeli, belirli bir firma için, pay senedi fiyat yönünü 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).

$$Z = f(\sum_{i=1}^n w_{ij} x_i + b_j) \cdot (\sum_{j=1}^p w_j) + b \quad (4)$$

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 çıkış nöronunun ağırlığını göstermektedir (Jardin, 2016).

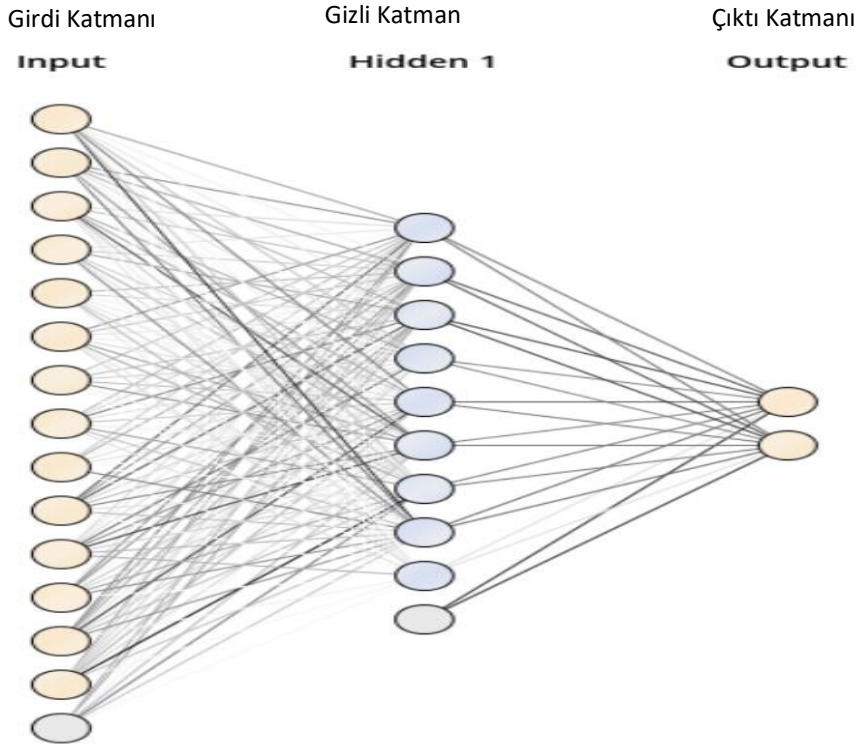
Ağ için girdiler, giriş katmanındaki 14 nöron tarafından temsil edilen finansal oranlar ve makro ekonomik göstergelerden oluşmaktadır. Ağın çıktısı, pay senedi fiyatı yönünün iki örneği bir önceki döneme göre fiyat düşüşü 0 veya bir önceki döneme göre fiyat yükselişi 1 ile kodlanmıştır. Üç katmanlı ileri beslemeli ANN'nin mimarisi Şekil 2'de gösterilmiştir. Bir katmanın nöronları, bağlantı katsayıları (ağırlıklar) ile komşu katmanların nöronları ile bağlantılıdır.

Momentum, bir önceki ağırlık değişimlerinin bir sonraki geçişte hesaplanan ağırlıkları etkileme düzeyini göstermektedir. ANN, finans alanındaki çalışmalarda Sigmoid fonksiyon sağlıklı sonuçlar verdiği için Ulusoy (2010)'un çalışması takip edilerek algoritmada Sigmoid fonksiyonun özelliklerinden yararlanılmıştır. Modelin çıktıları 0 ile 1 arasında değişecektir. Yukarı veya aşağı hareket tahminini belirlemek için 0,5 eşiği kullanılır. 0,5'ten büyük veya 0'a eşit çıktı değeri için, tahmin yukarı hareket, 0,5'ten küçük ise aşağı hareket olarak kabul edilir. Enke ve Thawornwong (2005)'un çalışması takip edilerek geri yayılım algoritmasını kullanan ileri beslemeli sinir ağı için, giriş değerleri üzerinde eşit bir dağılım oluşturmak üzere aktivasyon fonksiyonu olarak bir sigmoid hiperbolik tanjant fonksiyonu seçilmiştir. Sinir ağı modeli için finansal sınıflandırma ve tahmin için başarıyla kullanıldığından tek bir gizli katman seçilmiştir. Buna göre, ileri beslemeli sinir ağı, giriş katmanı, gizli katman ve çıkış katmanı olmak üzere üç katman ile inşa edilmiştir. Bu çalışmada, bağlantı ağırlıkları başlangıçta rastgele atanmış ve daha sonra geri yayılım eğitimi sürecinde belirlenmiştir. Eğitim esnasında hata değerinin daima azalması beklenemez. Hatanın artması çözümden uzaklaşılmasının göstergesidir. Hata yüzeyinde yerel bir en küçük noktaya ulaşıldığı halde eğitime devam edilmesi durumunda yerel en küçük noktadan kurtuluncaya kadar hata değeri belirli düzeyde artar. Başka bir en küçük noktaya ilerlendiğinde hata değeri yeniden azalır. Hata değerinin devamlı artması ilgili noktanın genel bir en küçük nokta olduğunun göstergesi sayılmaktadır. Çözüm esnasında hata yüzeyi hakkında bilgi bulunmadığından varılan noktanın yerel veya genel en küçük nokta olup olmadığı ayrımını yapmak zor olabilmektedir. Bu nedenle birçok çalışmada hata değerinin artmaya başladığı gözlemlendiğinde eğitime işlemi sonlandırılmaktadır. ANN eğitiminde öğrenme katsayısı önemli bir parametredir. Bu katsayı bağlantı ağırlıklarındaki değişme miktarını ayarlamaktadır. Öğrenme katsayısının pozitif ve birden küçük değerler alması gerekmektedir (Ataseven, 2013).

Çok katmanlı bir ağıın kullanılması durumunda üç katmanlı ağıın yeterli olup olmadığı, her katmanda kaç düğüm bulunması gerektiğı tecrübe sonucunda çözülecek problemin özelliğine göre belirlenecektir. Birçok problemin çözümünde üç katmanlı ağı yeterli olurken ağıın daha hızlı öğrenmesi için ilave katmanlar da eklenebilir. Girdi ve çıktı katmanında bulunacak düğümlerin sayısı çözülecek problemin yapısına bağlıdır. Ağıın öğrenme hızını belirleyen η (öğrenme oranı) değeri için genellikle 0,05 ile 0,25 arasında küçük bir değerin belirlenmesi tercih edilmektedir. Bu değerin gereğinden küçük belirlenmesi iterasyon sayısının artmasına büyük belirlenmesi ise gerçek bir minimum noktaya ulaşamamasına neden olmaktadır. Erişilen noktada hata değeri kabul edilebilir düzeyde ise bu noktanın lokal en düşük veya genel en düşük olmasının önemi olmayacaktır (Akpınar, 2014).

Bu çalışmada parametre değeri verinin ezberlenmesi sorunu ile karşılaşılmaması için en yüksek değeri öğrenme oranı 0,30, momentum 0,20 ve devir sayısı 500 ile sınırlandırılmıştır. ANN analizlerinde aktivasyon fonksiyonu, gizli katman sayısı, nöron sayıları ve diğeri parametre değeri problemin yapısı, örnek sayısı dikkate alınarak ve çok sayıda analiz yapılarak deneme yanılma yöntemi ile belirlenmiştir. Verilen parametre değeri içerisinde çok sayıda deneme ile algoritma en yüksek sınıflandırma ve tahmin sonucuna ulaşan parametre değeri belirlemektedir. Tablo 8’de verilen sonuçlar algoritma tarafından üretilmektedir. Çeşitli gizli katman nöronları, öğrenme algoritmaları ve öğrenme oranları ile yapılan birçok deneyden sonra, giriş katmanında 14 nöron, gizli katmanda 9 nöron, 0,30 öğrenme oranı, momentum 0,10 ve esnek bir geri yayılım eğitim algoritması kullanan ileri beslemeli sinir ağı on katlı çapraz doğrulama yöntemi ile en iyi ağı mimarisi bulunmuştur. Özdemir, Tolun ve Demirci (2011)’nin çalışması takip edilerek büyük sayısal değişkenlerin daha küçük sayısal değişkenlerin üzerinde baskınlık kurmasını engellemek için çalışmada kullanılan değişkenler [-1,1] aralığında ölçeklendirilmiştir. Şekil 2’de ANN mimarisi verilmiştir.

Şekil 2. ANN Mimarisi



Bu çalışmada kullanılan yöntemlerden en yüksek tahmin ve sınıflandırma sonucu veren modeli elde etmek amacıyla parametre değerlerine alt ve üst değerler atanmıştır. En iyi parametrelerin ve performans kriterlerinin belirlenmesi fonksiyonuyla atanan alt ve üst değerler farklı parametre kombinasyonu ile

denenerek en yüksek tahmin ve sınıflandırma sonucunu veren model belirlenmektedir. Tablo 4’de ANN parametreleri verilmiştir.

Tablo 4. ANN Parametreleri

Ağ Türü	Çok Katmanlı Perseptron			
Öğrenme Algoritması	Geri Yayılm			
Öğrenme Kuralı	Momentum			
Girdi Katmanındaki Düğüm Sayısı	14 Bağımsız Değişken			
Gizli Katman Sayısı	1			
Çıktı Katman Düğüm Sayısı	2 (Fiyatı düşen (0) / Fiyatı yükselen (1))			
Doğrulama Türü	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

3.3.2. CART Analizi

Karar ağacı, her iç düğümün (yapraksız düğüm) bir öznitelik üzerinde bir testi temsil ettiği, her dalın testin bir sonucunu temsil ettiği ve her bir yaprak düğümünün (veya terminal düğümünün) bir sınıf etiketi tuttuğu, ağaç yapısına benzer bir akış şemasıdır. Bir ağaçtaki en üstteki düğüm kök düğümdür. İlişkili sınıf etiketinin bilinmediği bir grup verildiğinde, grubun nitelik değerleri karar ağacına göre test edilir. Karar ağacı sınıflandırıcılarının inşası herhangi bir alan bilgisi veya parametre ayarı gerektirmez ve bu nedenle bilgi keşfi için uygundur. Karar ağaçları yüksek boyutlu verileri işleyebilir. Ağaç yapımı sırasında, öznitelik seçim ölçütleri ile en iyi sınıflara ayıran özniteliği seçmek için kullanılır. Karar ağaçları inşa edildiğinde, dalların çoğu eğitim verilerinde gürültü veya aykırı değerleri yansıtabilir. Ağaç budama işlevi bu tür dalları belirlemeye ve kaldırmaya çalışır. Bir öznitelik seçim ölçüsü (veya bir bölme kuralı), sınıf etiketli eğitim gruplarının belirli bir veri bölümünün en iyi şekilde ayıran bölme ölçütünü seçer. Verilen eğitim gruplarını tanımlayan her özellik için bir sıralama sağlar. Ölçüm için en iyi puana sahip olan öznitelik, ayırma özniteliği olarak seçilir (Nair, Mohandas ve Sakthivel, 2010).

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). Hangi düğümün kök düğüm olacağına karar vermenin dışında düğümün hangi noktadan ikiye ayrılması gerektiğini de hesaplar. CART, dallara ayırma kriterini hesaplar 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).

$$\Psi(s/t) = 2P_L P_R \sum_{j=1}^M |P(C_j|t_L) - P(C_j|t_R)| \quad (5)$$

t: Dalların yapacağı 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).

Gini Endeksi $g(t)$, p_i her bir kategorinin olasılığı ve c kategori sayısı olmak üzere 6 nolu eşitlik kullanılarak hesaplanır (Akpınar, 2014).

$$g(t) = 1 - \sum_{i=1}^c p_i^2 \quad (6)$$

CART karar ağacı algoritması analiz parametreleri en düşük ve en yüksek değerler dalların bölünme kriterleri, ağacın büyüklüğü, budanması gibi temel kriterler araştırmacı tarafından deneme yanılma ile programa verilmektedir. Verilen parametre değerleri içerisinde çok sayıda deneme ile algoritma en yüksek sınıflandırma ve tahmin sonucuna ulaşılan parametre değerlerini belirlemektedir. Tablo 10'da verilen sonuçlar algoritma tarafından üretilmektedir.

Tablo 5'te CART sınıflandırma algoritması analiz parametreleri verilmiştir.

Tablo 5. CART Sınıflandırma Algoritması Analiz Parametreleri

Analiz İçin Belirlenen Parametreler	Açıklama			
Doğrulama Türü	10 Katlı Çapraz Doğrulama			
Değişken Sayısı	14 Bağımsız Değişken			
Örneklem Seçimi	Tabakalı Örneklem Seçimi			
Bölünme Kriteri	Gini Index			
Analiz İçin Belirlenen Parametreler	En Düşük	En Yüksek	Adımlar	Ölçek
Bölünme İçin Minimum Boyut (Minimal Size For Split)	1,0	4,0	10	Doğrusal
En Düşük Dal Boyutu (Minimal Leaf Size)	1,0	2,0	10	Doğrusal
Minimum Kazanım (Minimal Gain)	1,0	10,0	10	Doğrusal
Maksimum Derinlik (Maximal Depth)	1,0	10,0	10	Doğrusal
Güven Düzeyi (confidence)	0,0	0,25	-	-
Ön Budama Sayısı	0,0	10	-	-

3.3.3. KNN Analizi

Bellek tabanlı yöntemler arasında sayılan K-En Yakın Komşu Algoritması gözlem değerlerinin arasındaki uzaklıklardan yararlanarak sınıflandırma işlemi yapmaktadır. Bu algoritma denetimli bir öğrenme yöntemi olup verilerin sınıflandırmasında seçilen bir özelliğin kendine en yakın özellikler arasındaki yakınlığı kullanmaktadır. K-En Yakın Komşu Algoritması (KNN) yöntemi hangi sınıfa girdiği bilinmekte olan bir örnek kümesindeki gözlem değerlerinden yararlanarak yeni katılacak olan bir gözlemin sınıflardan hangisine gireceğini belirlemek amacıyla kullanılmaktadır. K-En Yakın Komşu Algoritması örnek kümede bulunan gözlemlerin sonradan belirlenen bir gözlem değerine olan uzaklıkları hesaplandıktan sonra en yakın k sayıda gözlemin seçilmesine dayalı bir yöntemdir. Uzaklıkların hesaplanmasında i ve j noktaları için Öklid uzaklık formülü kullanılmaktadır (Özkan, 2016):

$$d(i, j) = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (7)$$

Algoritmada k değeri önceden belirlenmektedir. k değerinin yüksek olduğu durumda birbirine benzemeyen noktalar bir araya toplanmakta, düşük olduğu durumda ise birbirine benzediği (aynı sınıfa ait olduğu) halde bazı gözlemlerin farklı sınıflara konulmasına veya farklı sınıfların açılmasına neden olmaktadır (Silahtaroglu, 2016).

K en yakın komşu algoritması analiz aşamaları şu şekildedir (Özkan, 2016):

- Verilen bir noktaya en yakın komşuların sayısını gösteren k parametresi belirlenir.
- Verilen noktanın tüm gözlem noktalarına olan uzaklıkları hesaplanır.
- Hesaplama yapıldıktan sonra büyükten küçüğe doğru sıralanır ve bunlar arasından en küçük olan k tanesi seçilir.
- Seçilen satırların hangi sınıfa ait oldukları belirlenir ve en çok tekrarlanan sınıf değeri seçilir.
- Seçilen sınıf, tahmin edilmesi beklenen gözlem değeri sınıfı olarak alınır.

Tablo 6’da KNN analizi parametreleri verilmiştir.

Tablo 6. KNN Analizi Parametreleri

Değişken Seçimi	14 Değişkenli Veri Seti
Doğrulama Türü	10 Katlı Çapraz Doğrulama
Örneklem Seçimi	Tabakalı Örneklem Seçimi (Stratified sampling)
k sayısı	1-10
Ölçüm Tipi (Measure types)	Karışık ölçü (Mixed Measures)
Karışık Ölçü (Mixed Measure)	Karışık Öklid Mesafesi (Mixed Euclidean Distance)

KNN analizinde en yüksek sınıflandırma sonucu k=3 alındığında elde edildiğinden k değeri 3 olarak alınmıştır.

4.1. Bulgular

4.1. ANN Analizi Bulguları

ANN analizi karışıklık matrisi Tablo 7’de verilmiştir. Gözlem grubunda gerçekte fiyatı yükselen 136 örnekten 133’ü ANN tarafından doğru tahmin edilmiş, 3 örnek gerçek durumda fiyatı yükseldiği halde fiyatı düşen olarak yanlış tahmin edilmiştir. Gözlem grubunda gerçekte fiyatı düşen 69 örnekten 68’i ANN tarafından doğru tahmin edilmiş, bir örnek gerçek durumda fiyatı düştüğü halde fiyatı yükselen olarak yanlış tahmin edilmiştir.

Tablo 7. ANN Analizi Karışıklık Matrisi

		Tahmin Edilen Grup		
		Fiyatı Yükselen	Fiyatı Düşen	Toplam
Gözlemlenen Grup(Gerçek)	Fiyatı Yükselen	133	3	136
	Fiyatı Düşen	1	68	69
	Toplam	134	71	205

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 ölçüttür (Li vd., 2017). Çalışmada kullanılan modeller 6 kriter yönünden sıralanmıştır. Bunlar; i) doğruluk (accuracy), ii) kesinlik (precision), iii) duyarlılık (recall), iv) Kappa değeri, v) AUC ve vi) F ölçütüdür. En iyi modeller, söz konusu 6 kriter tarafından ölçülen yüksek sınıflandırma gücüne dayalı olarak seçilmiştir. Ek 1’de ANN ağırlıkları verilmiştir. Tablo 8’de en iyi parametrelerin ve performans kriterlerinin belirlenmesi ile en iyi performans gösteren ANN model sonucu verilmektedir.

Tablo 8. En İyi Parametrelerin ve Performans Kriterlerinin Belirlenmesi ile En İyi Performansı Gösteren Yapay Sinir Ağları Model Sonucu

Parametreler	Sonuçlar
Doğruluk	%98,05
Sınıflandırma Hatası	%1,95
Kappa	0,957
AUC	0,986
Kesinlik	%95,77
Duyarlılık	%98,55
F ölçütü	%97,14
Belirleyicilik	%97,79
Öğrenme Oranı	0,30
Momentum	0,10
Devir Sayısı	500

Tabloda 8’de yer alan kappa istatistiğinden kısaca bahsetmek gerekmektedir. Kappa, iki ya da daha fazla gözlemci arasındaki uyumun güvenilirliğini ölçen istatistik 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.

Kappa değeri (-)1 ile (+)1 arasında değer alabilmekte ve bulunan değer şu şekilde yorumlanmaktadır:

K = +1 ise iki gözlemcinin sonuçları tümüyle birbiri ile uyumludur.

K = 0 ise iki gözlemci arasındaki uyum sadece şansa bağlıdır.

K = -1 ise iki gözlemci tümüyle birbirinin tersini değerlendirmektedir.

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.

4.2. CART Analizi Bulguları

Tablo 9’da CART karar ağacı analizi karışıklık matrisi verilmektedir. Gözlem grubunda gerçekte fiyatı yükselen 136 örnekten 133’ü CART tarafından doğru tahmin edilmiş üç örnek gerçek durumda fiyatı yükseldiği halde fiyatı düşen olarak yanlış tahmin edilmiştir. Gözlem grubunda gerçekte fiyatı düşen 69 örnekten 64’ü CART tarafından doğru tahmin edilmiş, beş örnek gerçek durumda fiyatı düştüğü halde fiyatı yükselen olarak yanlış tahmin edilmiştir.

Tablo 9. CART Karar Ağacı Analizi Karışıklık Matrisi

		Tahmin Edilen Grup		
		Fiyatı Yükselen	Fiyatı Düşen	Toplam
Gözlemlenen Grup(Gerçek)	Fiyatı Yükselen	133	3	136
	Fiyatı Düşen	5	64	69
	Toplam	138	67	205

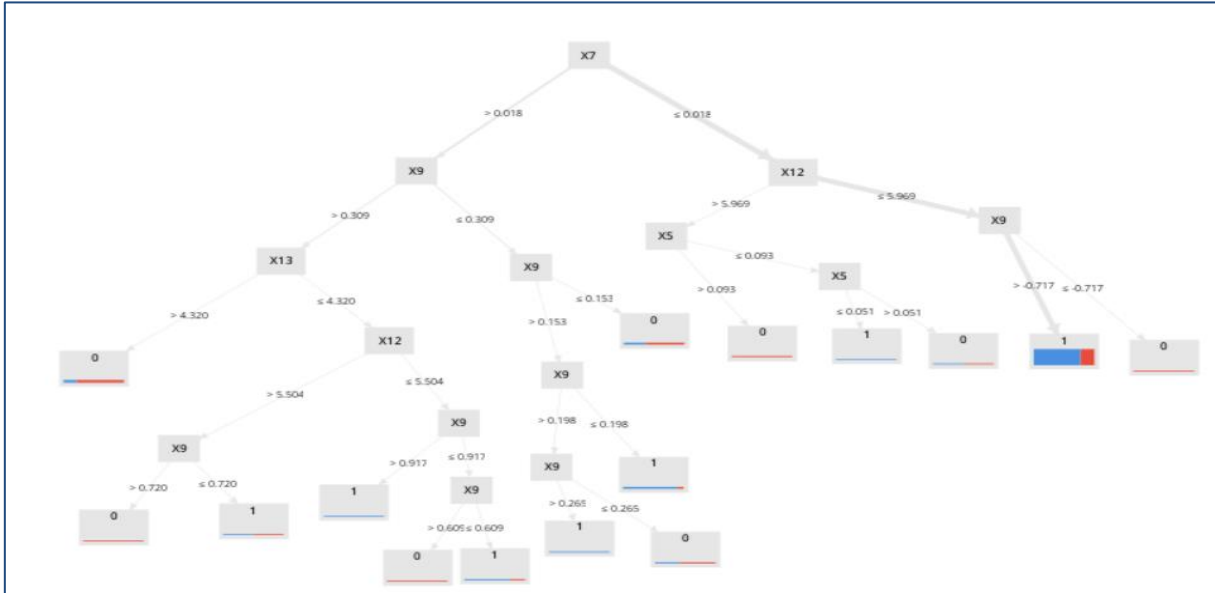
Tablo 10’da en iyi parametrelerin ve performans kriterlerinin belirlenmesi ile en iyi performans gösteren CART analizi performans sonuçları verilmektedir.

Tablo 10. En İyi Parametrelerin ve Performans Kriterlerinin Belirlenmesi ile En İyi Performansı Gösteren CART Analizi Performans Sonuçları

Parametreler	Sonuçlar
Doğruluk	%96,10
Sınıflandırma Hatası	%3,90
Kappa	0,912
AUC	0,989
Kesinlik	%95,52
Duyarlılık	%92,75
F Ölçütü	%94,12
Bölünme İçin Minimum Boyut	1
En Düşük Dal Boyutu	1
En düşük Kazanım	2,8
En Yüksek Derinlik	10

“X7, Net Kâr Marjı” ağacın kökünü oluşturan en önemli değişken olarak bulunmuştur. X7, Net Kâr Marjı” 0,018’den büyük olan örnekler için ağacın karar vermesinde ikinci önemli değişken “X9 Hisse Başına Kâr Oranı” değişkenidir. “X9 Hisse Başına Kâr Oranı” 0,309’dan büyük işletmeler için ağacın karar vermesinde üçüncü önemli değişken “X13Ln, Tüketici Güven Endeksi” değişkenidir. “X13Ln, Tüketici Güven Endeksi” 4,320’den büyük 4 örneğin fiyatı yükselen, 14 örneğin fiyatı düşen olarak çıkmıştır. Çalışma hacmini sınırlamak için tüm karar kuralları verilmemiş örnek olması açısından karar kuralının okunması yukarıda kısaca ele alınmıştır. Ek 2’de CART Analizi Karar Kuralı verilmiştir. Şekil 3’de CART karar ağacı görüntüsü verilmiştir.

Şekil 3. CART Karar Ağacı Görüntüsü



4.3. KNN Analizi Bulguları

Tablo 11’de KNN analizi karışıklık matrisi verilmiştir. Gözlem grubunda gerçekte fiyatı yükselen 136 örnekten 136’sı KNN tarafından doğru tahmin edilmiştir. Gözlem grubunda gerçekte fiyatı düşen 69 örnekten 53’ü KNN tarafından doğru tahmin edilmiş, 16 örnek gerçek durumda fiyatı düştüğü halde fiyatı yükselen olarak yanlış tahmin edilmiştir.

Tablo 11. KNN Analizi Karışıklık Matrisi

		Tahmin Edilen Grup		
		Fiyatı Yükselen	Fiyatı Düşen	Toplam
Gözlemlenen Grup (Gerçek)	Fiyatı Yükselen	136	0	136
	Fiyatı Düşen	16	53	69
	Toplam	152	53	205

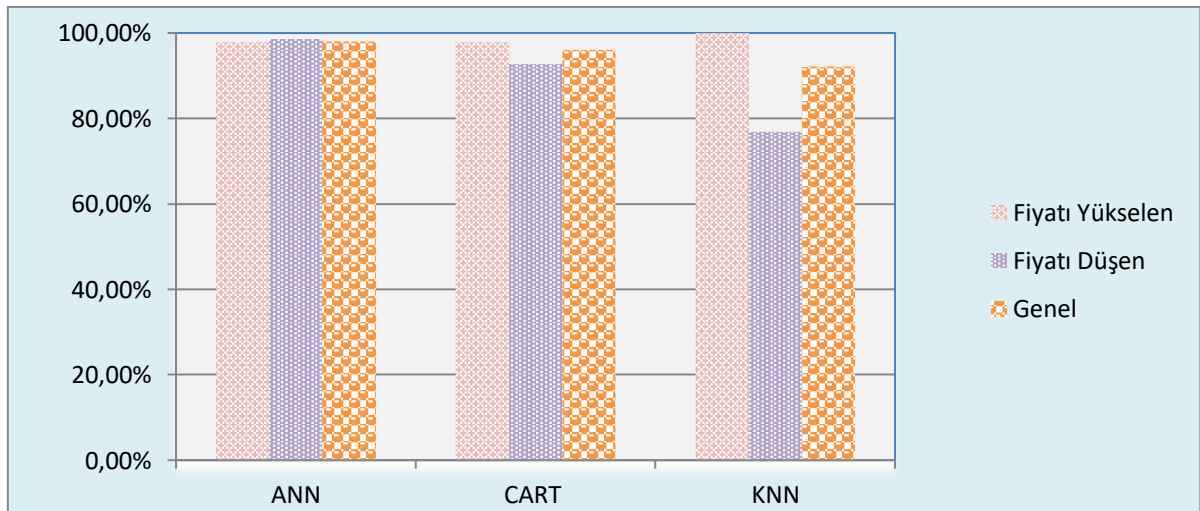
Tablo 12’de KNN analizi model sonucu verilmektedir.

Tablo 12. KNN Analizi Model Sonucu

Parametreler	Sonuçlar
Doğruluk	%92,20
Sınıflandırma Hatası	%7,80
Kappa	0,815
AUC	0,981
Kesinlik	%100,00
Duyarlılık	%76,81
F Ölçütü	%86,89

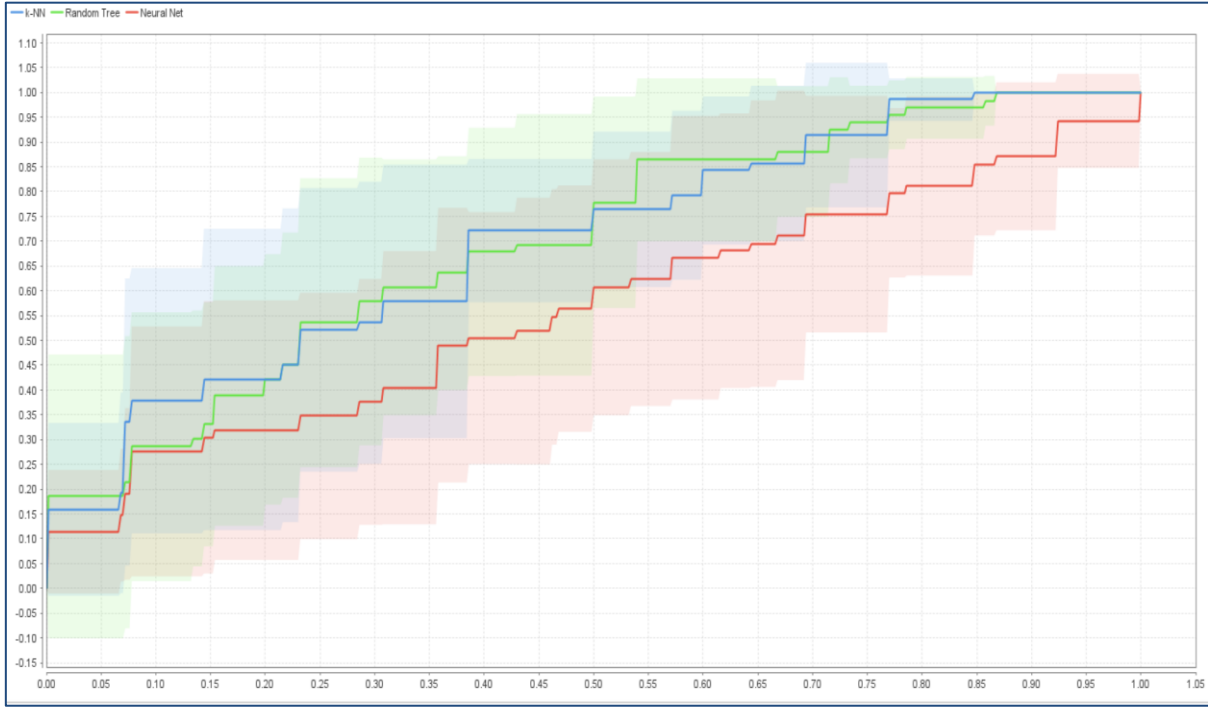
4.4. Bulguların Değerlendirilmesi

Bu çalışmada elde edilen bulgular modelin performansının veri setine, seçilen algoritmaya ve değişkenlere bağlı olduğunu ortaya koymuştur. Gaganis (2009)’in 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)’in sınıflandırmasına göre bu çalışmadaki ANN, CART ve KNN analiz sonuçları çok iyi, olarak değerlendirilebilir. Analiz sonucunda ANN, fiyatı yükselen örnekleri %97,79 olarak, fiyatı düşen örnekleri %98,55 olarak ve genel sınıflandırma sonucunda ise %98,05 doğruluk oranı ile tahmin etmiştir. CART, fiyatı yükselen örnekleri %97,79 olarak, fiyatı düşen örnekleri %92,75 olarak ve genel sınıflandırma sonucunda ise %96,10 doğruluk oranı ile tahmin etmiştir. KNN algoritması fiyatı yükselen örnekleri %100,00 olarak, fiyatı düşen örnekleri %76,81 olarak ve genel sınıflandırma sonucunda ise %92,20 doğruluk oranı ile tahmin etmiştir. Şekil 4’de çalışmada kullanılan yöntemlerin tahmin performansının karşılaştırılması verilmiştir.

Şekil 4. Çalışmada Kullanılan Yöntemlerin Tahmin Performansının Karşılaştırılması

ROC eğrisi, modelin dikey ekseninde “isabetleri” (yani gerçek pozitifler) yüzdesini ve yatay ekseninde “yanlış alarmların” 1-özgüllüğü veya yüzde oranlarını çizmektedir. Sonuç, 45° çizgisinden sol üst köşeye yükselen eğimli bir eğridir. Bükme keskinliği ve sol üst köşeye ne kadar yakınsa, modelin doğruluğu da o kadar yüksek olur. Şekil 5’te ROC eğrisi verilmektedir. Eğri altındaki alan (AUC), çeşitli kesme noktalarının tüm olası seçimlerine göre yanlış sınıflandırma oranlarının ortalaması olarak kabul edilebilir. Bu nedenle AUC, sınıflandırma modellerinin maliyeti veya ciddiyeti hakkında bir bilgi bulunmadığında farklı sınıflandırma modellerini karşılaştırmak için kullanılabilir (Gaganis, 2009).

Şekil 5. ROC Eğrisi



Literatürde pay senedi fiyat endeksi hareketini tahmin etmeyi amaçlayan birçok çalışma olmasına rağmen, bu çalışma sonuçlarını pay senedi fiyat yönü tahmini ve endeks yönü tahmini ile karşılaştırmak mantıklı olacaktır. Kara vd. (2011) çalışmalarında SVM ve ANN yöntemlerini kullanarak BIST Ulusal 100 endeksi hareket yönünü tahmin etmişlerdir. ANN modeli çalışmamızda olduğu gibi en yüksek sınıflandırma ve tahmin modeli olarak bulunmuştur. Imandoust ve Bolandraftar (2014) çalışmalarında Karar Ağacı (DT), Rastgele Orman (RF) ve Naïve Bayesian (NBC) olmak üzere üç model geliştirmişler ve günlük Tahran Menkul Kıymetler Borsası (TSE) endeksindeki hareket yönünü tahmin performanslarını karşılaştırmışlardır. Sonuç olarak karar ağacı modelinin (%80,08) performansının Rastgele Orman (%78,81) ve Naïve Bayesian Sınıflandırıcıdan (%73,84) daha iyi bulmuşlardır. Bu çalışmada karar ağacı algoritması olarak CART Sınıflandırma ve regresyon ağacı kullanılmıştır. CART modeli ANN’den sonra en iyi (%96,10) sınıflandırma sonucu elde etmiştir. Patel vd. (2015) çalışmalarında pay senedi ve pay senedi fiyat endeksi hareketini tahmin etmek için ANN, SVM, Random Forest ve Naive-Bayes algoritmalarının tahmin performansını karşılaştırmıştır. Analiz sonucunda ANN %86,69, SVM %89,33, Random Forest %89,98 ve Naive Bayes %90,19’luk doğruluğa sahip olduğu belirtilmiştir. Bu sonuç çalışmamızda elde edilen ANN’nin en yüksek sınıflandırma sonucuna sahip model olduğu bulgusuyla uyumlu değildir.

Yakut ve Gemici (2017) çalışmalarında, BIST 100 endeksindeki kimya, kauçuk ve plastik ürünleri sanayinde bulunan 18 şirketin pay senedi getirileri Lojistik Regresyon, C5.0, CART ve SVM yöntemlerini kullanarak pay senedi getirilerini tahmin etmişlerdir. Analizler sonucunda LR %75,00 C5.0 %88,00 CART %89,8 ve SVM %75,9’luk doğru sınıflandırma başarısı gerçekleştirdiği belirtilmiştir. CART ilgili çalışmada en yüksek doğrulukla tahmin gerçekleştiren model olarak görülmekte, bu çalışmada CART karar ağacı modeli ANN’den

sonra en iyi tahmin modeli olarak bulunmuştur. Ayrıca ilgili çalışmada pozitif ve negatif pay senedi getiri sınıflandırma tahminine etki eden en önemli değişkenlerin “piyasa/defter değeri değişkeni”, “TÜFE değişkeni” ve “brüt kâr marjı değişkeni” olduğu belirtilmiştir. Bu çalışmada CART karar ağacının oluşturulmasında verileri ikiye bölen önemli değişkenler olarak “Net Kâr Marjı”, “Fiyat/Kazanç Oranı”, “Hisse Başına Kâr Oranı”, “CDS Primi Aylık (3 aylık ortalama)”, “Tüketici Güven Endeksi” bulunmuştur. Papuçcu (2019) çalışmasında Yapay Sinir Ağları, Destek Vektör Makineleri ve Naive Bayes algoritmaları kullanarak BIST 100 endeksi hareket yönünü tahmin etmiştir. İlgili çalışma tahmin sonuçları sırasıyla ANN (%99,10), SVM (%99,10) ve NB (%90,40) olarak bulunmuş, bu sonuç bu çalışmada ANN'nin diğer iki yöntemden (CART, KNN) daha yüksek sınıflandırma ve tahmin sonuçlarıyla uyumludur.

5. Sonuç

Bu çalışmada Borsa İstanbul 30 Endeksinde ve Kurumsal Yönetim Endeksinde bulunan beş imalat sanayi şirketinin 2010/3 ve 2020/3 dönemine ait üçer aylık aritmetik ortalaması alınmış dokuz mali tablo verisi ve üç aylık ortalaması alınmış beş makro ekonomik değişken dâhil edilerek örnek kapsamındaki şirketlerin bir sonraki üç aylık ortalaması alınmış pay senedi fiyat yönü tahmin edilmiştir. Tahmin yöntemleri olarak Yapay Sinir Ağları, Sınıflandırma ve Regresyon Ağacı ve K-En Yakın Komşu Algoritması kullanılmıştır. Veri setinin kısıtlı olması nedeniyle çalışmada kullanılan tüm yöntemlerde 10 katlı çapraz doğrulama yöntemi kullanılmıştır. Yapay Sinir Ağları ile Sınıflandırma ve Regresyon Ağacı analizlerinde en iyi parametrelerin ve performans kriterlerinin belirlenmesi fonksiyonu kullanılarak verilen parametre aralıkları doğrultusunda en iyi sonucun alındığı modeller elde edilmiştir. Analiz sonucuna göre örnek kapsamındaki beş şirkete ait 41 dönem ve 205 örneğin üç ay sonraki ortalaması alınmıştır. 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)'in sınıflandırmasına göre bu çalışmadaki ANN, CART ve KNN analiz sonuçları çok iyi, olarak değerlendirilebilir.

Analiz sonucunda ANN, fiyatı yükselen örnekleri %97,79 olarak, fiyatı düşen örnekleri %98,55 olarak ve genel sınıflandırma sonucunda ise %98,05 doğruluk oranı ile tahmin etmiştir. CART, fiyatı yükselen örnekleri %97,79 olarak, fiyatı düşen örnekleri %92,75 olarak ve genel sınıflandırma sonucunda ise %96,10 doğruluk oranı ile tahmin etmiştir. KNN fiyatı yükselen örnekleri %100,00 olarak, fiyatı düşen örnekleri %76,81 olarak ve genel sınıflandırma sonucunda ise %92,20 doğruluk oranı ile tahmin etmiştir. Bu çalışmada CART karar ağacının oluşturulmasında verileri ikiye bölen önemli değişkenler olarak “X5, Net Kâr Marjı”, “X7, Fiyat/Kazanç Oranı”, “X9, Hisse Başına Kâr Oranı”, “X12Ln, CDS Primi Aylık (3 aylık ortalama)”, “X13Ln, Tüketici Güven Endeksi” bulunmuştur.

Bu araştırma, üç farklı makine öğrenmesi yöntemi kullanılarak belirlenmiş değişkenlerle modeller oluşturma ve her bir yöntemin ayırt edici özellikleri çerçevesinde sınıflandırma ve tahmin performanslarının karşılaştırılmasına dayanmaktadır. Araştırmada kullanılan yöntemlerden ANN ve CART için en iyi parametrelerin ve performans kriterlerinin belirlenmesi için problemin çözümü için verinin yapısı göz önünde bulundurularak literatürde sıklıkla kullanılan parametre değerleri ve aralıkları programa girilerek algoritmaların verilen parametre aralıklarında en yüksek sınıflandırma ve tahmin sonuçlarını veren parametre değerleri belirlenmiştir. Araştırmanın alanında literatüre en yüksek katkıyı sağlayacağı ya da Borsa İstanbul ile diğer yabancı ülke borsalarında genelleştirilebileceği iddia edilmemektedir. Kullanılan değişkenler ve parametre değerlerinde ilgili yöntemlerin pay senedi fiyat yönü hakkında belirli düzeyde bilgi verebileceği ve araştırmacıların, uygulayıcıların bu çalışmada elde edilen modelleri hali hazırda kullandıkları modellere dahil edebilecekleri belirtilmektedir. Uzun vadeli yatırımı tercih eden analistler, yatırımcılar, spekülâtörler ve araştırmacılar, kullandıkları modellerin içerisine bu çalışmada kurulan modelleri de alternatif olarak kullanabileceklerdir. Bu çalışmada elde edilen bulgular makine öğrenme yöntemlerinin pay senedi fiyat yönü tahmini konusunda umut verici olduğunu göstermektedir.

Makine öğrenmesi yöntemlerinin geliştirilmesi gereken yönleri (nedensel ilişkilerle ilgilenmemesi vb.) ve ekonometrik yöntemlerin geliştirilmesi gereken yönleri (büyük verinin analize dâhil edilmesi vb.)

birleştirerek daha doğru ve güvenilir sonuçlara ulaşılabilir. Bunun yanı sıra örnek sayısının artırılarak Borsa İstanbul'da daha fazla örneğin alınması ve dünya borsalarından örnekler alınarak ilgili borsalarda işlem gören pay senetleri fiyat yönü/fiyat tahmin çalışmaları literatüre değer katacaktır. Modellerin tahmin performanslarının iyileştirilebilmesi için diğer makine öğrenmesi veya geleneksel yöntemlerle birlikte kullanımlarının da faydalı olacağı düşünülmektedir. Ayrıca, veri setinin genişletilerek her yıl için ayrı bir tahmin yapıp sonuçların karşılaştırılması önemli sonuçları ortaya çıkaracaktır. Öngörülebilir risklerin modele dâhil edilmesi ile elde edilecek tahminlerin yatırımcılara yol göstermesi bakımından önemli olduğu düşünülmektedir. Öngörülemez risklerin de her zaman bulunabileceği göz önünde bulundurularak hatasız bir tahminin olamayacağı açıktır. Gelecekteki yapılacak çalışmalarda daha fazla temel ve teknik analiz değişkenlerinin pay senedi fiyat tahmin kalitesi üzerindeki etkisi belirlenebilir. Son olarak, gelecek araştırmalarda gerçekçi yatırım uygulamalarını artırmak için pay senedi temettüleri, işlem maliyetleri ve bireysel vergi dilimleri senaryoları altında alım/satım simülasyonunun dikkate alındığı durum sonuçlarının irdelenmesinin literatüre önemli katkılar sunacağı düşünülmektedir.

Beyan ve Açıklamalar (Disclosure Statements)

1. Bu çalışmanın yazarı, araştırma ve yayın etiği ilkelerine uyduğunu kabul etmektedir (The author of this article confirms that her work complies with the principles of research and publication ethics).
2. Yazar tarafından herhangi bir çıkar çatışması beyan edilmemiştir (No potential conflict of interest was reported by the author).
3. Bu çalışma, intihal tarama programı kullanılarak intihal taramasından geçirilmiştir (This article was screened for potential plagiarism using a plagiarism screening program).

Kaynaklar

- Akay, E. Ç. (2018). Ekonometride yeni bir ufuk: Büyük veri ve makine öğrenmesi. *Social Sciences Research Journal*, 7(2), 41-53.
- Akcan, A., & Kartal, C. (2011). İMKB sigorta endeksini oluşturan şirketlerin hisse senedi fiyatlarının yapay sinir ağları ile tahmini. *Muhasebe ve Finansman Dergisi*, (51), 27-40.
- Akpınar, H. (2014). *Data veri madenciliği veri analizi*. İstanbul: Papatya Yayınları.
- Anbalagan, T., & Maheswari, S. U. (2015). Classification and prediction of stock market index based on fuzzy metagraph. *Procedia Computer Science*, (47), 214-221.
- Ataseven, B. (2013). Yapay sinir ağları ile öngörü modellemesi. *Öneri Dergisi*, 10(39), 101-115.
- Barak, S., & Modarres, M. (2015). Developing an approach to evaluate stocks by forecasting effective features with data mining methods. *Expert Systems with Applications*, (42), 1325-1339.
- Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications*, (80), 340-355.
- Çelik, U., Akçetin, E., & Gök, M. (2017). *Rapidminer ile veri madenciliği*. İstanbul: Pusula Yayınları.
- Dhanalakshmi, S., & 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.
- Elmas, Ç. (2018). *Yapay zeka uygulamaları*. Ankara: Seçkin Yayıncılık.
- Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, (29), 927-940.
- 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.
- Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, (32), 2513-2522.

-
- Imandoust, S. B., & Bolandraftar, M. (2014). Forecasting the direction of stock market index movement using three data mining techniques: The case of Tehran Stock Exchange. *Int. Journal of Engineering Research and Applications*, 4(6), 106-117.
- Jardin, P. D. (2016). A Two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research*, (254), 236-252.
- Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, (38), 5311-5319.
- Kılıç, S., (2015). Kappa testi. *Journal of Mood Disorders*, 5(3), 142-144.
- Li, A., Wu, J., & Liu, Z. (2017). Market manipulation detection based on classification methods. *Procedia Computer Science*, (122), 788-795.
- Liang, D., Tsai, C. F. & Wu, H. T. (2015). The effect of feature selection on financial distress prediction. *Knowledge-Based Systems*, (73), 289-297
- Nair, B. B., Mohandas, V., & Sakthivel, N. R. (2010). A Decision tree- rough set hybrid system for stock market trend prediction. *International Journal of Computer Applications*, 6(9), 1-6.
- Oliveira, F. A., Nobre, C. N., & Zárate, L. E. (2013). Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, (40), 7596-7606.
- Özçalıcı, M. (2017). Aşırı öğrenme makineleri ile hisse senedi fiyat tahmini. *Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 35(1), 67-88.
- Özdemir, A. K., Tolun, S., & Demirci, E. (2011). Endeks getirisi yönünün ikili sınıflandırma yöntemiyle tahmin edilmesi: İMKB 100 endeksi örneği. *Niğde Üniversitesi İİBF Dergisi*, 4(2), 45-59.
- Özkan, Y. (2016). *Veri madenciliği yöntemleri*, İstanbul: Papatya Yayınları.
- Öztemel, E. (2012). *Yapay sinir ağları*. İstanbul: Papatya Yayınları.
- Papuçcu, H. (2019). Borsa endeksi hareketlerinin makine öğrenme algoritmaları ile tahmini. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (23), 179-190.
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, (42), 259-268.
- Silahtaroglu, G. (2016). *Veri madenciliği kavram ve algoritmaları*. İstanbul: Papatya Yayınları.
- Ulusoy, T. (2010). İMKB endeks öngörüsü için ileri beslemeli ağ mimarisine sahip yapay sinir ağı modellemesi. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (5), 21-40.
- Yakut, E., & Gemici, E. (2017). LR, C5.0, CART, SVM yöntemlerini kullanarak hisse senedi getiri sınıflandırma tahmini yapılması ve kullanılan yöntemlerin karşılaştırılması: Türkiye’de BIST’de bir uygulama. *Ege Akademik Bakış*, 17(4), 461-479.
- <https://tr.investing.com> (Erişim Tarihi, 25.06.2020).
- <https://www.kap.org.tr> (Erişim Tarihi, 26.06.2020).

Ekler

Ek 1. ANN Ağırlıkları

Girdi Katmanı	Gizli Katman								
	Düğüm 1	Düğüm 2	Düğüm 3	Düğüm 4	Düğüm 5	Düğüm 6	Düğüm 7	Düğüm 8	Düğüm 9
Bağımsız Değişken	Sigmoid								
X1	5,157	-0,692	-1,115	-3,194	-6,885	0,441	2,918	-10,840	-0,593
X2	-6,458	5,118	-4,012	1,626	0,469	5,727	-0,544	7,331	0,028
X3	2,033	3,471	-0,152	-4,428	-4,453	9,521	-2,687	-0,928	-1,449
X4	1,398	2,729	-5,039	0,005	2,571	-4,000	-1,652	6,525	0,568
X5	0,283	-3,925	-2,262	-0,608	-1,100	-1,317	2,322	3,788	1,119
X6	6,174	3,482	1,357	0,732	1,834	-5,298	4,675	1,288	1,992
X7	-3,411	0,739	3,589	2,605	-3,521	-6,373	-0,936	-0,525	0,024
X8	3,862	2,372	7,310	1,784	-3,078	0,817	5,371	-3,505	1,751
X9	-3,039	-4,389	-0,444	1,555	1,503	0,279	0,060	7,016	0,398
X10	8,899	5,392	9,330	2,269	-3,770	-7,089	-1,672	3,684	-0,143
X11	2,099	1,122	-1,825	-7,941	6,321	-10,931	-2,588	-4,032	-4,181
X12	1,654	-6,205	-5,766	0,414	-2,911	-2,265	-5,125	-6,366	-0,004
X13	6,599	-4,988	4,354	-1,295	-10,665	2,370	3,673	6,120	-0,650
X14	4,611	2,152	-2,212	2,624	3,953	2,021	-13,189	-3,389	0,487
Bias	0,186	4,177	-4,814	0,332	0,616	3,715	-7,117	-3,253	-0,418
Fiyatı Yükselen (1)	7,503	-9,747	-10,529	6,622	7,332	6,748	7,288	8,239	1,697
Fiyatı Düşen (0)	-7,503	9,746	10,529	-6,622	-7,332	-6,748	-7,288	-8,239	-1,698

Ek 2. CART Analizi Karar Kuralı

```

X7 > 0.018
  |
  | X9 > 0.309
  | |
  | | X13 > 4.320: 0 {1=4, 0=14}
  | | X13 ≤ 4.320
  | | |
  | | | X12 > 5.504
  | | | |
  | | | | X9 > 0.720: 0 {1=0, 0=2}
  | | | | X9 ≤ 0.720: 1 {1=1, 0=1}
  | | | |
  | | | | X12 ≤ 5.504
  | | | | |
  | | | | | X9 > 0.917: 1 {1=2, 0=0}
  | | | | | X9 ≤ 0.917
  | | | | | |
  | | | | | | X9 > 0.609: 0 {1=0, 0=2}
  | | | | | | X9 ≤ 0.609: 1 {1=3, 0=1}
  | | |
  | | | X9 ≤ 0.309
  | | | |
  | | | | X9 > 0.153
  | | | | |
  | | | | | X9 > 0.198
  | | | | | |
  | | | | | | X9 > 0.265: 1 {1=3, 0=0}
  | | | | | | X9 ≤ 0.265: 0 {1=2, 0=3}
  | | | | | |
  | | | | | | X9 ≤ 0.198: 1 {1=9, 0=1}
  | | | | |
  | | | | | X9 ≤ 0.153: 0 {1=4, 0=7}
  |
  | X7 ≤ 0.018
  | |
  | | X12 > 5.969
  | | |
  | | | X5 > 0.093: 0 {1=0, 0=4}
  | | | X5 ≤ 0.093
  | | | |
  | | | | X5 > 0.051: 0 {1=1, 0=1}
  | | | | X5 ≤ 0.051: 1 {1=2, 0=0}
  | | |
  | | | X12 ≤ 5.969
  | | | |
  | | | | X9 > -0.717: 1 {1=105, 0=31}
  | | | | X9 ≤ -0.717: 0 {1=0, 0=2}

```