



Makine Öğrenmesi Teknikleri ile Hisse Senedi Fiyat Tahmini

Nesrin Koç Ustalı¹ 

Nedret Tosun² 

Ömür Tosun³ 

Makine Öğrenmesi Teknikleri ile Hisse Senedi Fiyat Tahmini	Stock Price Forecasting Using Machine Learning Techniques
<p>Öz</p> <p>Bu çalışma, Borsa İstanbul Anonim Şirketi (BİST) 30 Endeksi'nde işlem gören firmaların hisse senetlerinin gelecek fiyatlarını tahmin etmeyi amaçlamaktadır. Bu amaçla öncelikle BİST 30 Endeksi firmalarının 2010-2019 yılları arasındaki üçer aylık finansal tabloları temin edilmiş daha sonra bu tablolar vasıtasıyla firmalara ait finansal oranlar hesaplanmıştır. Ayrıca firma hisse senetlerinin aylık kapanış fiyatlarına ulaşılmış ve firmalara ait finansal oranlarla denk olacak şekilde üçer aylık ortalamaları alınmıştır. Bu şekilde veriler temin edildikten sonra Yapay Sinir Ağları (YSA), Rastgele Orman (RO) algoritması ve XGBoost algoritması kullanılarak her bir firmaya ait hisse senedinin gelecek fiyatı tahmin edilmiştir. Daha sonra her bir yöntemle elde edilen tahmin sonuçları karşılaştırılmıştır. XGBoost ve Rastgele Orman algoritmaları birbirlerine yakın sonuçlar vermelerine rağmen XGBoost algoritması en iyi sonucu vermektedir. Ayrıca her iki modelin de YSA'ya göre daha yüksek performans gösterdiği tespit edilmiştir.</p>	<p>Abstract</p> <p>This study aims to estimate the future prices of stocks of firms listed in Borsa İstanbul Joint Stock Company (BIST) 30 Index. For this purpose, firstly, quarterly financial statements of BIST 30 Index companies between 2010-2019 have been provided and then financial ratios of firms have been calculated through these tables. In addition, monthly closing prices of company stocks were reached, and quarterly averages were taken in line with the financial ratios of firms. After obtaining the data, the future price of each company's stock was estimated by using Artificial Neural Networks (ANN), Random Forest (RF) algorithm and XGBoost algorithm. Then, the estimation results obtained according to each method were compared. It was determined that although XGBoost and Random Forest algorithms gave similar results, XGBoost has slightly better forecast results. Also, both models performed better than ANN.</p>
<p>Anahtar Kelimeler: BİST 30 Endeksi, Gelecek Fiyat Tahmini, Yapay Sinir Ağları, Rastgele Orman Algoritması, XGBoost Algoritması</p>	<p>Keywords: BIST 30 Index, Future Price Forecast, Artificial Neural Networks, Random Forest Algorithm, XGBoost Algorithm</p>
<p>JEL Kodları: G12, C45, G17</p>	<p>JEL Codes: G12, C45, G17</p>

<p>Araştırma ve Yayın Etiği Beyanı</p>	<p>Bu çalışma bilimsel araştırma ve yayın etiği kurallarına uygun olarak hazırlanmıştır.</p>
<p>Yazarların Makaleye Olan Katkıları</p>	<p>Yazar 1'in makaleye katkısı %40, Yazar 2'nin makaleye katkısı %30 ve Yazar 3'ün makaleye katkısı %30'dur.</p>
<p>Çıkar Beyanı</p>	<p>Yazarlar açısından ya da üçüncü taraflar açısından çalışmadan kaynaklı çıkar çatışması bulunmamaktadır.</p>

¹ Arş. Gör., Akdeniz Üniversitesi Uygulamalı Bilimler Fakültesi, Uluslararası Ticaret ve Lojistik Bölümü. nesrinkoc@akdeniz.edu.tr

² Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Ekonometri Bölümü. tosunn@baib.gov.tr

³ Doç. Dr. Akdeniz Üniversitesi Uygulamalı Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü. omurtosun@akdeniz.edu.tr

1. Giriş

Ülkeler arasındaki ekonomik sınırların yok olup sosyal, kültürel ve toplumsal değerlerin sürekli bir değişim içinde olduğu günümüz küresel dünyasında “karar vermek”, hem devletler ve firmalar hem de toplumun her kesiminden bireyler açısından oldukça zor bir süreci ifade etmektedir. Çünkü sürekli değişen koşulların ve muğlak verilerin olması, karar verme sürecini negatif yönde etkilemektedir. Bu süreç, yatırımcıların doğru karar vererek yatırım yapabilmelerini ya da başka bir ifade ile doğru bir portföy yönetimi yapabilmelerini etkilemektedir. Bu nedenle yatırımcılar için de çok önemli olan karar verme sürecinde, yatırımcıların yatırım yapacakları hisse senedi ile ilgili işlem hacmi, işlem miktarı, piyasa değeri gibi bilgilere ulaşabilmeleri ve hisse senedine ait gelecek fiyat beklentilerini tahmin edebilmeleri oldukça önemlidir. Ancak işlem hacmi, işlem miktarı, piyasa değeri gibi bilgiler kolay ulaşılabilir ve elde edilebilir olmasına rağmen gelecek fiyat beklentisi, modern iş dünyasında son zamanlarda ortaya çıkmış ve doğru tahmin edilmesi zor bir bilgi özelliği taşımaktadır. Çünkü hisse senedi fiyatı ekonomik ve ekonomik olmayan pek çok faktörden etkilenmektedir (Liu vd., 2012: 1348). Bu nedenle de doğru bir tahmin yapılabilmesi için pek çok faktör göz önünde bulundurulmalı ve doğru bir tahmin yöntemi seçilmelidir.

Portföy yönetimi için önemli olan hisse senedi gelecek fiyat tahmini, doğru bir şekilde yapıldığı zaman yatırımcıların mevcut karlarından daha yüksek karlar elde etmelerine ya da zararlarını düşürmelerine veya tamamen ortadan kaldırmalarına yardımcı olabilmektedir. Bu önemi nedeniyle hisse senedi gelecek fiyat tahmini konusunda pek çok çalışma yapılmış ve YSA, ARIMA, Destek Vektör Makineleri (DVM), Zaman Serileri, Doğrusal Regresyon, Karar Ağaçları, Markov Zincirleri gibi çok sayıda yöntem denenmiştir. Bu yöntemlerden bazıları başarısız bulunmuş, bazıları da başka yöntemlerle desteklenmiştir. Ancak YSA ile yapılan hisse senedi gelecek fiyat tahminlerinin Tektaş ve Karataş (2004), Toraman (2008), Hadavandi vd. (2010), Guresen vd. (2011), Adebıy vd. (2012) ve Zahedi ve Rounaghi (2015)’in çalışmalarında da görüldüğü üzere doğruluk oranının daha yüksek bulunduğu ortaya çıkmıştır. Bir başka ifadeyle, hisse senedi gelecek fiyat tahmini konusunda en geçerli sonuçlar YSA ile elde edilmiştir. Bu nedenle çalışmada tercih edilen yöntemlerden biri YSA olmuştur.

YSA, yapay zekâ çalışmalarıyla bağlantılı olarak insan beyninin yapısından esinlenilerek geliştirilmiş bir yöntemdir. Geliştirildiği ilk günden bu yana sürekli gelişim göstermiş ve günümüz bilgisayarlarının algoritmik yapısından çok farklı bir yapıya sahip olmuştur. Böylece doğrusal olmayan pek çok probleme çözüm üretmede çok tercih edilen modellerden olmuştur. Bu çalışmada da YSA’nın bu özelliklerinin çalışmanın sonuçlarının geçerliliği için önemli katkılar sağlayacağı düşünülmektedir. Çalışmada kullanılan ikinci yöntem RO algoritmasıdır. Bu algoritma karar ağaçları ve düğümlerden oluşan bir yapıya sahiptir. Düğümlerde yer alan ağırlıklara göre dalların aldığı değerler, nihai kararın verilmesine yardım etmektedir. Çalışmada kullanılan üçüncü yöntem ise Aşırı Gradyan Güçlendirme (XGBoost) algoritmasıdır. XGBoost algoritması, temelde Karar Ağaçları algoritmasına dayanmaktadır. Bu algoritma, Gradyan Güçlendirmeli Karar Ağaçları algoritmasının genişletilmiş ve geliştirilmiş şekli olarak ifade edilmekte ve pek çok tahmin probleminde çözüm üretmede etkili sonuçlar vermektedir.

Bu çalışmada, BİST 30 Endeksi'nde işlem gören ve finansal tablolarına ulaşılabilen 22 firmanın hisse senetlerinin gelecek fiyatları tahmin edilmiştir. Bunun için ilk olarak söz konusu firmaların 2010-2019 yılları arasındaki 3 aylık finansal tabloları temin edilmiş ardından bu finansal tablolar yardımıyla finansal oranları hesaplanmıştır. Ayrıca firmalara ait aylık geçmiş kapanış fiyatları temin edilmiş ve finansal oranlarla denk olacak şekilde üçer aylık ortalamaları alınmıştır. Söz konusu firmalara ait elde edilen finansal oranlar ve kapanış fiyatları sırasıyla girdi ve çıktı verisi olarak kullanılmış ve YSA, RO algoritması ve XGBoots algoritması ile fiyat tahmini yapılmıştır. Daha sonra yöntemlerden elde edilen sonuçlar karşılaştırılmıştır.

Çalışma, genel olarak altı bölümden oluşmaktadır. İlk bölüm olan giriş bölümünün ardından ikinci bölümde, hisse senedi gelecek fiyat tahmini konusunda daha önce yapılan çalışmalara yer verilmiştir. Üçüncü bölümde, çalışmada kullanılan YSA, RO algoritması ve XGBoost algoritması anlatılmıştır. Dördüncü bölümde, analizde kullanılan verilerin nasıl toplandığı ve analiz konusunda bilgi verilmiştir. Çalışmanın beşinci bölümünde, analiz sonucunda elde edilen bulgulara yer verilmiştir. Çalışmanın son bölümünde ise genel değerlendirmeler yapılmış ve gelecek çalışmalar için önerilerde bulunulmuştur.

2. Literatür İncelemesi

Bu bölümde hisse senedi gelecek fiyat tahmini konusunda daha önce yapılan çalışmalardan bazıları incelenmiş ve aşağıda yer aldığı şekilde özetlenmiştir.

Tektaş ve Karataş (2004)'te YSA'yı tanıtmak ve finans alanında katkıda bulunmak için Türkiye ile ilgili bir uygulama yapmak amaçlanmıştır. Bu doğrultuda çalışmada, İstanbul Menkul Kıymetler Borsası'nda işlem gören gıda ve çimento sektörlerinde faaliyet gösteren halka açık ve hisse senedi fiyat istatistiklerine ulaşılabilen yedi firmanın hisse senedi fiyatları tahmin edilmiştir. Analizde günlük ve haftalık olmak üzere iki farklı veri seti kullanılmıştır. Günlük analizlerde girdi değişkeni olarak İstanbul Menkul Kıymetler Borsası 100 Endeksi, İstanbul Menkul Kıymetler Borsası Sınai Endeksi ve Türkiye Cumhuriyeti Merkez Bankası (TCMB) Amerikan doları satış kuru kullanılmıştır. Haftalık analizlerde ise günlük analizlerde kullanılan girdi değişkenlerine ek olarak, aylık enflasyon endeksi, TCMB tarafından ağırlıklandırılmış bir aylık mevduat faiz oranı kullanılmıştır. Sonuç olarak, günlük veri kullanmanın modelin performansını arttırdığı görülmüştür. Ayrıca günlük verilerin kullanıldığı YSA ve Doğrusal Regresyon modelleri karşılaştırılmış ve YSA modelleri sonucuna göre oluşan korelasyon katsayılarının daha anlamlı olduğu görülmüştür.

Pai ve Lin (2005)'te hisse senedi fiyat tahmini yapmak için ARIMA ve DVM yöntemlerinden oluşan hibrid bir model önerilmiştir. Önerilen modelin performansını değerlendirmek için 10 hisse senedinin günlük kapanış fiyatlarından oluşan bir veri seti kullanılmıştır. Sonuçlar farklı tahmin modellerinin birbirini tamamlayabileceğini ortaya koymuştur.

Chen vd. (2007)'de Tayvan borsası hisse senedi tahmini için Bulanık Zaman Serileri modeli önerilmiştir. Söz konusu model, Tayvan'da Chen (1996), Yu (2004) ve Huarng (2006) tarafından hisse senedi gelecek fiyatının tahmin edildiği çalışmalar ile karşılaştırılmıştır. Analizlerde girdi değişkeni olarak, beş yıllık dönemdeki hisse senedi fiyatları ve 13 yıllık dönemdeki hisse senedi endeks verileri kullanılmıştır. Sonuç olarak önerilen modelin bu çalışmalardan daha yüksek doğruluk oranına sahip olduğu ortaya çıkmıştır.

Toraman (2008)'de İstanbul Menkul Kıymetler Borsası'nda işlem gören ve demir-çelik sektöründe faaliyet gösteren Erdemir A.Ş. ve Kardemir A.Ş. firmalarının hisse senedi fiyatları YSA ile tahmin edilmiştir. Bu doğrultuda iki farklı YSA modeli geliştirilmiştir. Analizlerde girdi değişkeni olarak ABD Merkez Bankası gösterge faiz oranı, TCMB üç aylık mevduat faiz oranı, enflasyon oranı, TCMB dolar efektif satış kuru, İstanbul Altın Borsası bir gram altın kapanış fiyatı, cari oran, sermaye yapısı oranı, duran varlıklar oranı, aktif karlılık oranı, fiyat-kazanç oranı, piyasa değeri-defter değeri oranı, hisse başına kar, kar dağıtım oranı, temettü verimi ve ekonomik verimlilik oranı kullanılmıştır. Sonuç olarak YSA'nın, oluşturulma sürecinde volatilitesi yüksek olan dönemlerin modellenmesinde başarılı olduğu görülmüştür.

Tsai ve Wang (2009)'da yatırımcıların hisse senedi seçiminde daha iyi karar verebilmeleri için hisse senetlerinin gelecekteki fiyatları tahmin edilmiştir. Çalışmada, YSA ve Karar Ağaçları yöntemleri kullanılmıştır. Yöntemleri test etmek için Tayvan elektronik sanayi firma hisse senedi verilerinden yararlanılmıştır. Analiz sonucunda, YSA ile Karar Ağaçlarının entegre edilerek uygulandığı modelin doğruluk oranının, YSA ve Karar Ağaçlarının ayrı ayrı uygulandığı modellerdeki doğruluk oranından daha yüksek olduğu ortaya çıkmıştır.

Hadavandi vd. (2010)'da hisse senedi piyasasında en başarılı tahmini, en basit şekilde verecek olan ve en az girdi bilgisi gerektiren yöntemin bulunması amaçlanmıştır. Bunun için, Bulanık Genetik Sistem ve YSA yöntemlerinin entegre edilmesiyle oluşan yeni bir model önerilmiştir. Önerilen model, havayolu ve teknoloji sektörlerinden elde edilen açılış fiyatı, kapanış fiyatı, en yüksek fiyat ve en düşük fiyat değişkenleri kullanılarak test edilmiştir. Sonuçlar, önerilen modelin hisse senedi gelecek fiyat tahmininde kullanılan tüm önceki modellerden daha yüksek performans sağladığını ortaya koymuştur.

Akcan ve Kartal (2011)'de İstanbul Menkul Kıymetler Borsası sigorta sektöründe faaliyet gösteren yedi firmanın hisse senedi fiyatları YSA ile tahmin edilmiştir. YSA modeli ile 15 gün, bir aylık, 1,5 aylık ve iki aylık dönemler için fiyat tahminleri yapılmıştır. Analizlerde girdi değişkeni olarak İstanbul Menkul Kıymetler Borsası 100 Endeksi günlük kapanış değeri, tüketici fiyat endeksi, TCMB dolar efektif satış kuru, Türkiye Cumhuriyeti Merkez Bankası günlük cumhuriyet altını fiyatı, fiyat-kazanç oranı, piyasa değeri-defter değeri oranı, aktif karlılığı oranı, alınan net prim büyüme hızı oranı, alınan primler-toplam aktifler oranı, öz sermaye karlılığı oranı, hisse başına kar tutarı ve net kar-alınan primler oranı kullanılmıştır. Sonuç olarak iki aya kadar yapılan tüm tahminlerin doğru olduğu ortaya çıkmıştır.

Guresen vd. (2011)'de hisse senedi endeks tahmininde YSA ve Çok Katmanlı Algılayıcı yöntemlerinin etkinliği değerlendirilmiştir. Analizde Nasdaq Menkul Kıymetler Borsası endeksinin günlük döviz kur değerleri kullanılmıştır. Sonuçlar klasik YSA modelinin daha güvenilir sonuçlar verdiğini ve oluşturulan hibrit modelin ise bu konuda başarısız olduğunu ortaya koymuştur.

Adebiyi vd. (2012)'de yatırımcıların hisse senedi seçimlerinde karar vermelerine yardımcı olmak için hisse senedi fiyat tahmininde kullanılmak üzere bir model önerilmiştir. Önerilen model, YSA ile piyasa göstergelerinin entegre edilmesiyle oluşturulmuştur. Analizlerde girdi değişkeni olarak açılış fiyatı, günlük en yüksek fiyat, günlük en düşük fiyat, kapanış fiyatı, işlem hacmi, yıllık kazanç, defter değeri gibi değişkenler kullanılmıştır. Sonuçlar önerilen modelin, sadece

teknik analiz ile elde edilen sonuçlardan daha kayda değer olduğunu ve yatırımcıların kararları için tatmin edici tahminler yaptığını ortaya koymuştur.

Liu vd. (2012)'de hisse senedi fiyat tahmini için Tip 2 Nötr Bulanık Modelleme yaklaşımı önerilmiştir. Önerilen modelde, hisse senetlerinin altı yıllık döneme ait günlük kapanış fiyatları değişken olarak kullanılmıştır. Analizler TAEX, Dow Jones ve Nasdaq endeksi veri setleri kullanılarak yapılmıştır. Sonuçlar hisse senedi fiyat tahmininde bu modelin etkin olduğunu göstermiştir.

Yang ve Panvada (2012)'de Avustralya Menkul Kıymetler Borsası'nda çeşitli sektör verileri kullanılarak gün içi hisse senedi fiyatlarının dağılımı incelenmiş ve bir sonraki fiyat, GARCH yöntemi ile tahmin edilmeye çalışılmıştır. Analizde girdi değişkeni olarak, her bir işlem için tarih, saat, boyut, fiyat, ticaret göstergesi, derinlik ve her hisse senedine karşılık gelen sektör endeksi kullanılmıştır. Sonuçlara göre, sistem %71 oranında bir sonraki fiyatı doğru tahmin etmiştir.

İlarslan (2014)'te hisse senedi fiyat hareketleri Markov Zincirleri yöntemi ile tahmin edilmeye çalışılmıştır. Çalışmada, İstanbul Menkul Kıymetler Borsası 10 banka endeksinde yer alan bankaların hisse senetlerine ait günlük kapanış verileri kullanılmıştır. Sonuç olarak bir banka dışındaki diğer tüm bankaların bir gün sonraki hisse senedi fiyat hareketliliği başarılı bir şekilde tahmin edilmiştir.

Zahedi ve Rounaghi (2015)'te Tahran'da 20 hisse senedinin gelecekteki fiyatları tahmin edilmiştir. Analizlerde girdi değişkeni olarak defter değeri, net hisse başına kar, öngörülen hisse başına kar, öz sermaye karlılığı, aktif karlılığı oranı, kar-zarar oranı, temettü yüzdesi, temettü-fiyat oranı, beta değeri, faaliyet karı/zararı, nakit kar, temettü büyümesi, derecelendirme likiditesi, firma büyüklüğü, defter değeri-piyasa değeri oranı, mali yıl sonunda toplam öz sermaye, net varlıklar, yıllık hisse senedi getirileri kullanılmıştır. YSA ve Temel Bileşen Analizinin kullanıldığı çalışmada, YSA ile doğruluk oranı daha yüksek tahminler yapılabileceği ortaya çıkmıştır.

3.Yöntem

Bu bölümde, çalışmada kullanılan makine öğrenmesi teknikleri olan YSA, RO algoritması ve XGBoost algoritması yöntemleri anlatılmıştır.

3.1. Yapay Sinir Ağları

Yapay Sinir Ağları (YSA)'nın temelleri insanoğlunun biyolojik beyin hücrelerinin nasıl çalıştığına ilişkin merakları üzerine yaptıkları çalışmalara dayanmaktadır. Bu çalışmalar sonucunda elde ettikleri bulguları bilgisayar sistemlerine entegre etmeleriyle ilk YSA geliştirilmeye başlanmıştır. YSA, ilk ortaya çıktığı dönemden bugüne kadar sürekli gelişim göstermiş ve günümüz bilgisayarlarının algoritmik hesaplama yöntemlerinden farklı bir yapıya sahip olmuştur. Genel olarak ağırlıklı bağlantılarla birbirine bağlanan ve kendi belleğine sahip elemanlardan oluşan paralel ve dağıtılmış bilgi işleme yapıları olarak tanımlanan (Elmas, 2003: 23) YSA, bu farklı yapıları sayesinde doğrusal olmayan birçok probleme çözüm üretmede faydalı sonuçlar sağlamıştır. Günümüzde pek çok alanda kullanılmakta olan YSA, özellikle endüstriyel, ulaştırma, havacılık, iletişim, tıp ve finans gibi alanlarda kullanılmaktadır.

YSA'nın tıpkı biyolojik sinir ağlarında olduğu gibi sinir hücreleri vardır. Bu yapay sinir hücrelerinin her biri 5 temel elemana sahiptir. Bunlar (Öztemel, 2006: 49-51):

Girdiler: Oluşturulan bir ağın öğrenmesi istenen örnekler tarafından belirlenen ve dış dünyadan olabileceği gibi ağ içerisinde başka hücrelerden veya kendi kendisinden de gelebilen bilgilerdir.

Ağırlıklar: Bir yapay sinir hücresine gelen girdilerin önemini ve hücre üzerindeki etkisini göstermektedir. Ağırlıkların büyük ya da küçük olması o girdinin önemini ifade etmemekle birlikte ağırlık değerleri artı, eksi ve sıfır değerlerini alabilmektedir. Artı değer etkinin pozitif olduğu, eksi değer etkinin negatif olduğu ve sıfır değeri etkinin olmadığı anlamına gelmektedir.

Toplama Fonksiyonu: Bir hücreye gelen net girdiyi hesaplayan fonksiyondur. Toplama fonksiyonu her yapay sinir hücresi için farklı olabileceği gibi bazı sinir hücreleri grup halinde aynı fonksiyona da sahip olabilmektedir.

Aktivasyon (etkinlik) Fonksiyonu: Bu fonksiyon hücreye gelen girdiyi işleyerek bu girdiyi karşılık gelecek çıktıyı üretmektedir. Ağdaki yapay sinir hücrelerinin her biri farklı aktivasyon formülüne sahip olabileceği gibi bazıları aynı diğer bazıları farklı aktivasyon formülüne sahip olabilir. Günümüz uygulamalarında yaygın olarak Çok Katmanlı Algılayıcı modelinde sigmoid fonksiyonu kullanılmaktadır.

Hücrenin Çıktısı: Aktivasyon fonksiyonu ile üretilen çıktı değeri hücre çıktısı olarak ifade edilmektedir. Bir sinir hücresinin birden fazla girdisi olmasına rağmen sadece bir çıktısı olmaktadır. Ayrıca hücre ürettiği çıktığı dış dünyaya sunabildiği gibi kendine girdi olarak da kullanabilmektedir.

Öğrenme yöntemi ve ağ yapısı gibi etmenlere bağlı olarak çeşitlilik gösteren YSA'lar temelde aynı felsefeye dayanmaktadır. Buna göre bir YSA 3 katmandan oluşmaktadır. Bunlar (Elmas, 2003: 45):

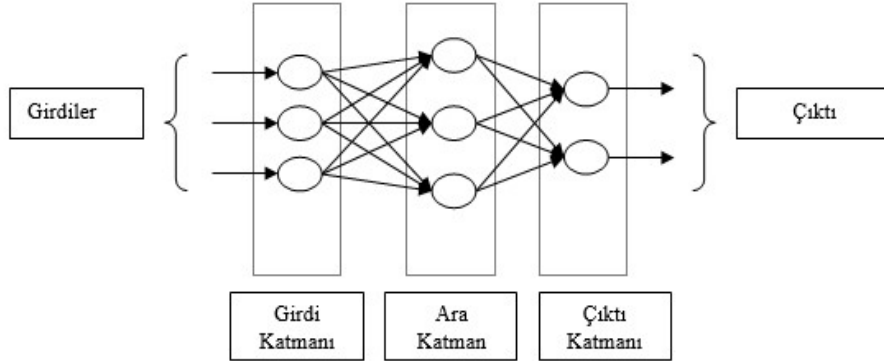
Girdi Katmanı: Bu katman dış çevreden gelen bilgileri ara katmanlara iletmekle sorumludur. Girdi katmanında herhangi bir bilgi işleme süreci gerçekleşmemektedir.

Ara Katmanlar: Girdi katmanından ara katmana gönderilen bilgilerin işlenerek çıktı katmanına gönderilmesi bu katmanda gerçekleşmektedir. Bir ağ birden fazla ara katmana sahip olabilmektedir. Bu katmandaki sinirler gerçek anlamda görülmediklerinden ara katmana gizli katman adı da verilmektedir.

Çıktı Katmanı: Ağın ara katmanlarından gelen verilerin işlenerek, girdi değerlerine karşılık gelen çıktının üretilmesi bu katmanda gerçekleşmektedir. Üretilen çıktı dış dünyaya ya da kullanıcıya gönderilmektedir.

Bu katmanlar aşağıdaki Şekil 1 yardımıyla gösterilmektedir.

Şekil 1: Yapay Sinir Ağı



Kaynak: Elmas, 2007: 56

YSA'da bu katmanlar arasında bağlantılar vardır ve bunların ağırlık değerleri vasıtasıyla bilgiler aktarılmaktadır. Bağlantıların ağırlık değerlerinin belirlenmesi işlemine "ağın eğitilmesi" denilmektedir. Başlangıçta rastgele atanan bu ağırlıklar, örnekler ağa gösterildikçe değiştirilmektedir. Burada amaç en iyi çözüme ulaşmak için ağa çok sayıda örnek göstererek, ağın doğru ağırlık değerini bulmasını sağlamaktır. Böylece ağ genelleştirme özelliğine kavuşmuş olur. Bu durum "ağın öğrenmesi" olarak adlandırılmaktadır (Öztemel, 2006: 55).

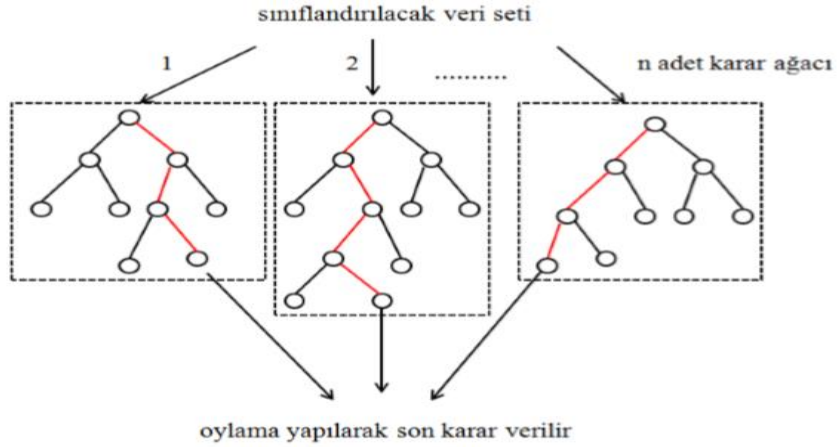
YSA'da öğrenme, danışmanlı ve danışmansız öğrenme olmak üzere temelde iki kategoriye ayrılmaktadır. Öğretmenli ya da eğitmenli öğrenme de denilen danışmanlı öğrenmede, YSA eğitim aşamasından sonra kullanılmaktadır. Eğitimde ağa veri setinin bir kısmı ile girdi ve çıktı bilgileri sunulmaktadır. Test aşamasında ise veri setinin kalan kısmı ile öğrenmenin gerçekleşip gerçekleşmediği kontrol edilmektedir (Elmas, 2003: 96). Öğretmensiz ya da eğitmensiz öğrenme olarak da adlandırılan danışmansız öğrenmede ise ağın doğru çıktı hakkında bilgisi yoktur, girdi bilgilerine göre kendisini örneklemektedir (Elmas, 2003: 149).

Ağın eğitimi tamamlandıktan sonra öğrenip öğrenmediğini yani performansını ölçmek için yapılan deneylere "ağın test edilmesi" denilmektedir. Test etme sürecinde ağın daha önce görmediği girdiler kullanılmakta ve ağırlık değerleri değiştirilmemektedir. Elde edilen çıktıların doğruluk derecesi ağın öğrenmedeki başarısı hakkında bilgi vermektedir (Öztemel, 2006: 56).

3.2. Rastgele Orman Algoritması

Rastgele Orman (RO) algoritması Breiman (2001) tarafından geliştirilmiş karar ağaçları ve düğümlerden oluşan bir yöntemdir. Bu yöntem ile regresyon ve sınıflandırma analizleri yapılabilmektedir. Yönteme göre düğümleri dallara ayırmak için düğümlerdeki rastgele değerlerden en iyi olanı seçilmekte ve oluşturulan karar ağaçlarına belirli ağırlıklar verilmektedir. Bu ağırlıklar karar ağaçlarının iç hatalarına göre belirlenmekte ve en düşük hataya sahip olan karar ağacına en yüksek, en yüksek hataya sahip karar ağacına en düşük ağırlık verilmektedir. Verilen bu ağırlıklar sınıf tahmininde oy verme işlemi için kullanılmaktadır. Daha sonra bu oylar toplanarak nihai karar verilmektedir. Örnek bir RO algoritması Şekil 2'de gösterilmektedir.

Şekil 2: RO Algoritması Genel Yapısı



Kaynak: Okumuş ve Aydemir, 2017: 3.

RO algoritması dört adımdan oluşmaktadır. Bunlar (Pal, 2005: 218; Watts vd. 2011: 70; Okumuş ve Aydemir, 2017: 3):

Adım 1: İlk adımda veri özelliklerine göre oluşturulmak istenen karar ağacı adedi (n) belirlenir.

Adım 2: Oluşturulan karar ağaçlarındaki her düğümde rastgele m adet değişken seçilir ve Gini indeksi ile hesaplanarak en iyi dal belirlenir.

Adım 3: Bir önceki aşamada belirlenen en iyi dal iki alt dala ayrılır. Bu işlem Gini indeksi sıfır oluncaya kadar başka bir ifadeyle her düğümde bir sınıf kalıncaya kadar devam ettirilir.

Adım 4: Son aşamada ise n tane karar ağacının ayrı ayrı yaptığı tahminler içinden en fazla oyu alan sınıf son karar tahmini olarak seçilir.

3.3. XGBoost Algoritması

Aşırı Gradyan Artırma (XGBoost) algoritması, karar ağaçları ve makine öğrenmesinde sıklıkla tercih edilen bir araç haline gelen bir uygulamadır. Geliştirilen bu yaklaşım sınıflandırma, regresyon ve sıralama görevlerinde yüksek performans sağlayan denetimli öğrenme alanında önemli bir araç olarak kabul edilmektedir (Mitchell ve Frank, 2017: 127). Denetimli öğrenme, bir dizi etiketli eğitim örneğinden tahmin edici bir model çıkarım görevi anlamına gelir. Bu öngörücü model ile bir üretim hatasının meydana gelip gelmediğini belirleme ya da belirli bir günde sıcaklık veya nem tahmini gibi sorunlara çözüm üretilebilmektedir. Bu bağlamda XGBoost algoritması, doğru modeller üretmek için artırma adı verilen bir işlemi uygulayan, denetimli bir öğrenme algoritması olarak ifade edilebilmektedir (Mitchell ve Frank, 2017: 129).

XGBoost algoritmasındaki artırılmış ağaçlar, regresyon ve sınıflandırma ağaçlarına bölünmüştür. Bu algoritmanın özü, amaç fonksiyonu değerinin optimize edilmesine dayanmaktadır (Zheng vd., 2017: 1171). XGBoost algoritmasının en önemli özelliği, tüm senaryolarda ölçeklenebilirliğidir. Sistem, tek bir makinedeki mevcut popüler çözümlerden 10 kattan daha hızlı çalışır ve dağıtılmış veya hafıza sınırlı ayarlarda milyarlarca örneğe ölçeklenir (Chen ve Guestrin, 2016: 785).

4. Verilerin Toplanması ve Analizi

Bu çalışma, hisse senedi gelecek fiyatını tahmin etmeyi amaçlamaktadır. Literatürde bu amaçla yapılan benzer çalışmalardan elde edilen bilgiler sonucunda, bu problem için en geçerli sonuçların YSA ile bulunduğu görülmüştür. Bu nedenle analizde öncelikle YSA ardından RO ve XGBoost algoritmaları kullanılarak gelecek fiyat tahmini yapılmıştır. Daha sonra farklı yöntemlere göre elde edilen sonuçlar karşılaştırılmıştır. Bu doğrultuda BİST 30 Endeksi'nde işlem gören ve finansal tablolarına ulaşılabilen 22 firma ele alınmıştır. BİST 30 Endeksi'nde işlem gören bankaların, finansal tabloları temel yapı itibarıyla diğer firmalardan farklı olduğundan dolayı araştırma dışında bırakılmıştır. Analizde ele alınan firmalar Tablo 1'de gösterilmektedir.

Tablo 1: Analizde Ele Alınan BİST 30 Endeksi Firmaları

Sıra	Firma Kodu	Firma Adı
1	ARCLK	Arçelik A.Ş.
2	ASELS	Aselsan Elektronik Sanayi ve Ticaret A.Ş.
3	BIMAS	Bim Birleşik Mağazalar A.Ş.
4	DOHOL	Doğan Şirketler Grubu Holding A.Ş.
5	EKGYO	Emlak Konut Gayrimenkul Yatırım Ortaklığı A.Ş.
6	EREGL	Ereğli Demir ve Çelik Fabrikaları T.A.Ş.
7	SAHOL	Hacı Ömer Sabancı Holding A.Ş.
8	KRDMD	Kardemir Karabük Demir Çelik Sanayi ve Ticaret A.Ş.
9	KCHOL	Koç Holding A.Ş.
10	KOZAL	Koza Altın İşletmeleri A.Ş.
11	KOZAA	Koza Anadolu Metal Madencilik İşletmeleri A.Ş.
12	PGSUS	Pegasus Hava Taşımacılığı A.Ş.
13	PETKM	Petkim Petrokimya Holding A.Ş.
14	SODA	Soda Sanayi A.Ş.
15	TAVHL	Tav Havalimanları Holding A.Ş.
16	TKFEN	Tekfen Holding A.Ş.
17	TOASO	Tofaş Türk Otomobil Fabrikası A.Ş.
18	TCELL	Turkcell İletişim Hizmetleri A.Ş.
19	TUPRS	Tüpraş-Türkiye Petrol Rafinerileri A.Ş.
20	THYAO	Türk Hava Yolları A.O.
21	TTKOM	Türk Telekomünikasyon A.Ş.
22	SISE	Türkiye Şişe ve Cam Fabrikaları A.Ş.

Kaynak: <https://www.kap.org.tr/tr/Endeksler>, erişim tarihi: 03.11.2018.

Bu firmalara ait hisse senetlerinin fiyat tahmini için firmaların 31.01.2010-31.12.2019 tarihleri arasındaki üç aylık finansal tabloları, Kamuyu Aydınlatma Platformu (KAP) web sayfasından temin edilmiştir. Finansal tablolar yardımıyla firmaların finansal oranları hesaplanmış ve elde edilen oranlar analizde girdi verisi olarak kullanılmıştır. Analizde kullanılan finansal oranlar Tablo 2'de gösterilmektedir. Analizde çıktı verisi olarak hisse senetlerinin aylık kapanış fiyatları kullanılmıştır. Çıktı verisi, Türkiye Hisse Senedi Piyasası-investing.com web sayfasından alınan 31.01.2010-31.12.2019 tarihleri arasındaki aylık kapanış fiyatlarının üçer aylık ortalaması alınarak elde edilmiştir.

Tablo 2: Analizde Kullanılan Finansal Oranlar

Oran Grubu	Oran Adı	Açıklama
Likidite Oranları	Cari Oran	Dönen Varlıklar / Kısa Vadeli Yabancı Kaynaklar
	Asit-Test Oranı	Dönen Varlıklar- Stoklar / Kısa Vadeli Yabancı Kaynakları
	Nakit Oran	Nakit ve Nakit Benzerleri / Kısa Vadeli Yabancı Kaynaklar
Faaliyet Oranı	Özsermaye Devir Hızı	Net Satışlar / Özsermaye
Karlılık Oranları	Özsermaye Karlılığı	Net Kar / Özsermaye
	Net Kar Marjı	Net Kar / Net Satışlar
Mali Yapı Oranları		Kısa Vadeli Yabancı Kaynaklar / Toplam Aktifler
		Uzun Vadeli Yabancı Kaynaklar / Toplam Aktifler
		Özsermaye / Toplam Aktifler

Kaynak: Okka, 2009: 109-123.

5. Bulgular

Makine öğrenmesi süreci veri setinin “eğitim verisi” ve “test verisi” olmak üzere ikiye ayrılması ile başlamaktadır. Literatürde herhangi bir kural olmamakla birlikte verinin %70’i rastgele seçilerek eğitim verisi olarak atanır. Eğitim verisi kullanmanın amacı kullanılan çeşitli makine öğrenmesi algoritmalarına ait kontrol parametrelerinin en iyi değerlerinin belirlenmesidir (öğrenme süreci). Optimal parametre değerleri seçilirken gerçek gözlem değeri ile modelin hesapladığı sonucun (tahmin çıktısı) arasındaki fark olarak adlandırabileceğimiz toplam hata değerinin en küçüklenmesi temel alınır. En küçük hatayı veren parametre değerleri seçilir.

Eğitim aşamasından sonra geliştirilen modelin gerçek hayattaki başarısını ifade eden genelleme gücü test verisi yardımı ile ölçülür. Sadece test verisinin yüksek başarıyı göstermesi yeterli değildir, aynı zamanda makine öğrenmesi modelinin eğitim ve test verilerinde gösterdiği performans değerlerinin de birbirine yakın olması beklenir.

Başarılı bir makine öğrenmesi algoritmasından istenen, eğitim verileri yardımıyla öğrendiği süreci o alanındaki tüm verilere genelleme yapabilmesidir. Bu, modelin daha önce hiç karşılaşmadığı veya ilk kez gördüğü veriler üzerinde başarılı tahminlerin yapılmasını sağlar. Eğer eğitim verisinde yüksek performans göstermiş fakat test verisinde gösterdiği başarı oranı düşük ise bu durumda aşırı öğrenme (ezberleme) olduğu söylenebilir. Aşırı öğrenme bir makine öğrenmesi algoritmasının eğitim verisinin çok iyi öğrenmesi fakat modelin ilk kez karşılaştığı bir veriyi (test verisi) kullanarak elde ettiği tahminlerin başarısız veya düşük performans oranına sahip olma durumudur.

Herhangi bir makine öğrenmesi modelinin performansı doğrudan doğruya seçilen kontrol parametrelerine (hiper parametre) bağlıdır. Doğru parametrenin seçilmesi modelin performansını arttırmaktadır. Her makine öğrenmesi modeli farklı sayıda kontrol parametrelerine sahiptir ve bunların eş zamanlı olarak en iyilerinin belirlenmesi gerekmektedir. Çalışmada kullanılacak farklı makine öğrenmesi modelleri için geliştirilen deney tasarımı Tablo 3’te yer almaktadır. Tabloda yer alan değerler, literatür analizi ve uzman görüşleri doğrultusunda belirlenmiştir.

Tablo 3: Parametre Optimizasyonu için Geliştirilen Deney Tasarımı

Model	Parametre	Değerler
YSA	Gizli Katman Sayısı	1, 2
	Gizli Katmandaki Nöron Sayısı	2, 4, 6, 8, 10
	Aktivasyon Fonksiyonu	Lojistik, Relu
XGBoost	Maksimum Ağaç Derinliği	4, 5, 6, 7, 8, 9
	Öğrenme Katsayısı	0.05, 0.01, 0.5, 0.1
	Tahminleyici (Ağaç) Sayısı	150, 200, 250, 300, 350
RO	Tahminleyici (Ağaç) Sayısı	150, 200, 250, 300, 350
	Maksimum Ağaç Derinliği	4, 5, 6, 7, 8, 9
	Maksimum Öznitelik Sayısı	4, 5, 6, 7, 8, 9

Tablo 3'te yer alan tüm olasılıklar için modeller çalıştırılmıştır. Örneğin YSA için (1, 4, Relu) deney tasarımında model; 1 gizli katman, gizli katmanda 4 nöron ve Relu aktivasyon fonksiyonu ile çalışacaktır. Uygulama için algoritmalar Python dili ile kodlanmıştır. Performans göstergesi olarak modelin uygunluğunun ölçümü için determinasyon katsayısı (R^2) kullanılmıştır. R^2 'nin hesaplanması için kullanılan formül eşitlik 1'de yer almaktadır (Terzi vd., 2011: 35).

$$R^2 = 1 - \frac{\sum_{i=1}^n (R_{i(gerçek)} - R_{i(model)})^2}{\sum_{i=1}^n (R_{i(gerçek)} - R_{i(ort)})^2} \quad (\text{Eşitlik 1})$$

Formüle yer alan indislerin açıklamaları aşağıda verilmiştir. Buna göre;

n : Toplam veri sayısı,

$R_{i(gerçek)}$: Ölçülen hisse senedi fiyatı,

$R_{i(model)}$: Geliştirilen modelin sonuçları,

$R_{i(ort)}$: Ortalama hisse senedi fiyatıdır.

Her bir deney için 10 tekrar yapılmış, en iyi sonuç ile 10 tekrarın ortalaması raporlanmıştır. Yapılan deney tasarımı sonucunda bulunan en iyi parametreler Tablo 4'te yer almaktadır.

Tablo 4: Kullanılan Modellere İlişkin En İyi Parametreler

Model	Parametre	En İyi Değerler
YSA	Gizli Katman Sayısı	2
	Gizli Katmandaki Nöron Sayısı	6 – 10
	Aktivasyon Fonksiyonu	Relu
XGBoost	Maksimum Ağaç Derinliği	6
	Öğrenme Katsayısı	0,1
	Tahminleyici (Ağaç) Sayısı	300
RO	Tahminleyici (Ağaç) Sayısı	200
	Maksimum Ağaç Derinliği	10
	Maksimum Öznitelik Sayısı	5

Uygulama için seçilen her üç model Tablo 4'te verilen optimal parametreler için kullanılmıştır. Seçilen parametreler için her bir model 10 kez çalıştırılmış ve bulunan en iyi değerler ile tekrarların ortalamaları Tablo 5'te verilmiştir.

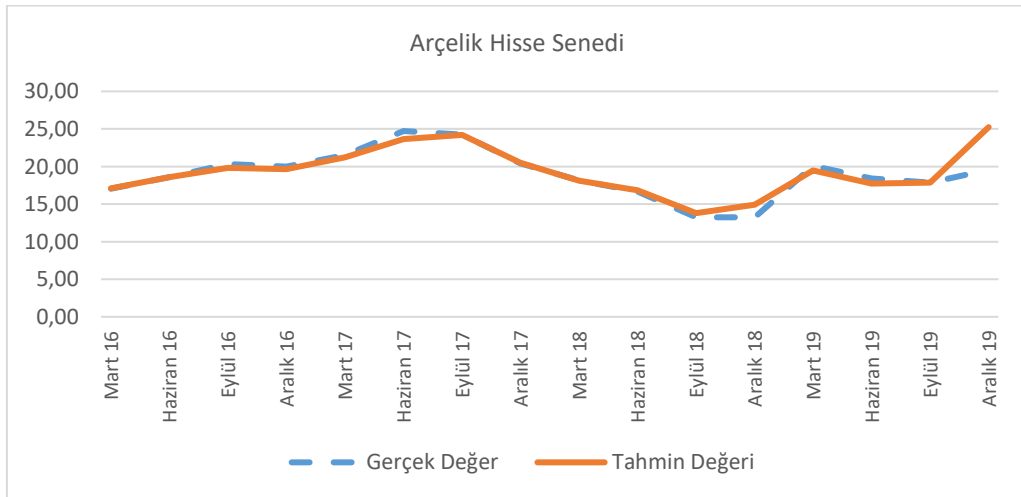
Tablo 5: Kullanılan Modellerin Performans Karşılaştırılması

	XGBoost		RO		YSA	
	En iyi Değer	Ortalama Değer	En iyi Değer	Ortalama Değer	En iyi Değer	Ortalama Değer
Eğitim Verisi Performansı	0,999	0,999	0,925	0,931	0,814	0,848
Test Verisi Performansı	0,758	0,722	0,702	0,695	0,708	0,686

Her üç modelinde başarılı tahminler verdiği Tablo 5'te görülebilmektedir. R^2 değeri bire ne kadar yakınsa modelin başarımı da yüksek olacaktır. Bir başka ifadeyle, tahmin değerleri ve gerçek gözlem değerleri arasındaki fark o kadar düşüktür. Sonuçlar incelendiğinde ele alınan veri seti için XGBoost ve RO modelleri birbirlerine yakın sonuçlar vermiştir. Ayrıca her iki model de YSA'dan daha iyi performans göstermiştir. Tek katmanlı yapay sinir ağı mimarisi kötü performans gösterdiği için iki gizli katmana sahip bir YSA tercih edilmiştir. Fakat iki katmanlı model bile kullanılan diğer iki makine öğrenmesi tekniğinin gerisinde kalmıştır. Eğitim seti performansları baz alındığında XGBoost en iyi performansı göstermiştir. Modellerin test verisi performanslarına bakarsak XGBoost algoritması %75,8, RO algoritması %70,2 ve YSA %70,8 oranında doğru tahmin vermiştir. Test veri setine ilişkin R^2 değerleri modellerin daha önce karşılaşmadığı gerçek hayat verileri karşısındaki genelleme gücünü ifade etmektedir. Bu değerler doğrultusunda YSA en düşük performansa sahiptir.

Tahmin modelinin performansına yönelik bir örnek Şekil 3'te verilmiştir. Şekilde XGBoost algoritması ile elde edilen Mart 2016 – Aralık 2019 dönemine ait Arçelik hisse senedi tahmin değerleri görülmektedir. XGBoost algoritmasının yüksek eğitim seti başarısından dolayı tahmin değerleri gerçek değerlere oldukça yakındır.

Şekil 3: Arçelik Hisse Senedi için Tahmin Değerleri



6. Sonuç

Bu çalışma, hisse senedi gelecek fiyatını doğru tahmin edebilmeyi sağlayacak yöntemi bulmayı amaçlamaktadır. Bu amaçla 3 farklı model analiz edilmiştir. Çalışmada, YSA, RO algoritması ve XGBoost algoritması kullanılarak analiz yapılmış ve sonuçlar karşılaştırılmıştır. Uygulama için algoritmalar Python dili ile kodlanmıştır. Performans göstergesi olarak modelin uygunluğunun ölçümü için determinasyon katsayısı (R^2) kullanılmıştır. Seçilen yöntemler için öncelikle bir deney tasarımı yapılmıştır. Her bir model için ele alınan kontrol parametreleri için belirlenen değerlerle yapılan deney tasarımı sonucunda en iyi parametre değerleri elde edilmiştir.

Optimal parametreler için her üç model uygulanmış ve her bir model için 10 kez çalıştırılmıştır. Test verisi performansları karşılaştırıldıklarında her ne kadar birbirlerine yakın da olsa XGBoost algoritması, RO algoritmasından daha iyi sonuç göstermiş, YSA ise bunların gerisinde kalmıştır. Bu değerler her üç modelin de başarılı tahminler verdiğini göstermektedir. Literatürde Tektaş ve Karataş (2004), Tsai ve Wang (2009), Guresen vd. (2011), Adebıy vd. (2012) ve Zahedi ve Rounaghi (2015) tarafından yapılan ve bir makine öğrenmesi tekniği olarak YSA ile birlikte başka bir analiz yönteminin kullanıldığı çalışmalarda ise her iki modelin de başarılı olduğu ancak YSA'nın daha yüksek performans sağladığı ortaya çıkmıştır. Bu çalışmada ise üç farklı makine öğrenmesi tekniği ile elde edilen sonuçlar karşılaştırılmıştır. Bu nedenle mevcut değişkenler ışığında yatırımcıların hisse senedi gelecek fiyat tahmini için XGBoost veya RO tekniklerinin herhangi birisini ile analizler yapmaları onları daha doğru sonuçlara ulaştıracaktır.

Gelecekteki çalışmalarda, BİST 30 endeksi üzerine yapılabilecek analizler ile süreç içerisindeki değişiklikler sonucunda nasıl bir tahmin sonucu elde edileceği araştırılabilir. Ayrıca, farklı endekslerle ve farklı algoritmalarla yapılan benzer analiz sonuçları karşılaştırma yapma imkânı sağlayacaktır.

Kaynaklar

- Adebiyi, Ayodele A.; Ayo, Charles K.; Adebiyi, Marion O.; Otokiti, Sunday O. (2012), "Stock Price Prediction Using Neural Network with Hybridized Market Indicators", *Journal of Emerging Trends in Computing and Information Sciences*, Vol. 3, No. 1: 1-9.
- Akcan, Ahmet; Kartal, Cem (2011), "İMKB Sigorta Endeksini Oluşturan Şirketlerin Hisse Senedi Fiyatlarının Yapay Sinir Ağları ile Tahmini", *Muhasebe ve Finansman Dergisi*, S. 51: 27-40.
- Breiman, Leo (2001), "Random Forests", *Machine Learning*, Vol. 45, No.1: 5-32.
- Chen, Tai L.; Cheng, Ching H.; Teoh, Hia J. (2007), "Fuzzy Time-Series Based on Fibonacci Sequence for Stock Price Forecasting", *Physica A: Statistical Mechanics and Its Applications*, Vol. 380: 377-390.
- Chen, Tianqi; Guestrin, Carlos (2016), "XGBoost: A Scalable Tree Boosting System", in *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, ACM, 785-794.
- Elmas, Çetin (2003), *Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama)*, Ankara: Seçkin Yayıncılık.
- Elmas, Çetin (2007), *Yapay Zekâ Uygulamaları*, Ankara: Seçkin Yayıncılık.
- Guresen, Erkam; Kayakutlu, Gulgun; Daim, Tugrul U. (2011), "Using Artificial Neural Network Models in Stock Market Index Prediction", *Expert Systems with Applications*, Vol. 38, No. 8: 10389-10397.
- Hadavandi, Esmail; Shavandi, Hassan; Ghanbari, Arash (2010), "Integration of Genetic Fuzzy Systems and Artificial Neural Networks for Stock Price Forecasting", *Knowledge-Based Systems*, Vol. 23, No. 8: 800-808.
<https://www.kap.org.tr/tr/Endeksler>, (Erişim: 03.11.2018).
- İlarslan, Kenan (2014), "Hisse Senedi Fiyat Hareketlerinin Tahmin Edilmesinde Markov Zincirlerinin Kullanılması: İMKB 10 Bankacılık Endeksi İşletmeleri Üzerine Ampirik Bir Çalışma", *Journal of Yaşar University*, C. 9, S. 35: 6158-6198.
- Liu, Chih F.; Yeh, Chi Y.; Lee, Shie J. (2012), "Application of Type-2 Neuro-Fuzzy Modeling in Stock Price Prediction", *Applied Soft Computing*, Vol. 12, No. 4: 1348-1358.
- Mitchell, Rory; Frank, Eibe (2017), "Accelerating the XGBoost Algorithm Using GPU Computing", *PeerJ Computer Science*, Vol.3: 127-164.
- Okka, Osman (2009), *Finansal Yönetim Teori ve Çözümlü Problemler*, Ankara: Nobel Yayıncılık.
- Okumuş, Hatice; Aydemir, Önder (2017), "Random Forest Classification for Brain Computer Interface Applications", In *Signal Processing and Communications Applications Conference (SIU)*, No.25: 1-4.
- Öztemel, Ercan (2006), *Yapay Sinir Ağları*, İstanbul: Papatya Yayıncılık.
- Pal, Mahesh (2005), "Random Forest Classifier for Remote Sensing Classification", *International Journal of Remote Sensing*, Vol. 26, No. 1: 217-222.
- Pai, Ping F.; Lin, Chih S. (2005), "A Hybrid Arıma and Support Vector Machines Model in Stock Price Forecasting". *Omega*, Vol. 33, No. 6: 497-505.
- Tektaş, Arzu; Karataş, Abdülmecit (2004), "Yapay Sinir Ağları ve Finans Alanına Uygulanması: Hisse Senedi Fiyat Tahminlemesi", *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, C. 18, S. 3-4: 337-349.
- Terzi, Özlem; Küçüksille, Ecir U.; Ergin, Gülşah; İlker, Ahmet (2011), "Veri Madenciliği Süreci Kullanılarak Güneş Işınımının Tahmini", *Uluslararası Teknolojik Bilimler Dergisi*, C. 3, S. 2: 29-37.
- Toraman, Cengiz (2008), "Demir-Çelik Sektöründe Yapay Sinir Ağları ile Hisse Senedi Fiyat Tahmini: Erdemir A.Ş. ve Kardemir A.Ş. Üzerine Bir Tahmin Uygulaması", *Muhasebe ve Finansman Dergisi*, S. 39: 44-57.
- Tsai, Chih F.; Wang, Sammy P. (2009), "Stock Price Forecasting by Hybrid Machine Learning Techniques", In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. 1, No. 755: 60-66.
- Watts, Jennifer D.; Powell, Scott L.; Lawrence, Rick L.; Hilker, Thomas (2011), "Improved Classification of Conservation Tillage Adoption Using High Temporal and Synthetic Satellite Imagery", *Remote Sensing of Environment*, Vol. 115, No. 1: 66-75.
- Yang, Joey W.; Parwada, Jerry (2012), "Predicting Stock Price Movements: An Ordered Probit Analysis on the Australian Securities Exchange", *Quantitative Finance*, Vol. 12, No. 5: 791-804.
- Zahedi, Javad; Rounaghi, Mohammad. M. (2015), "Application of Artificial Neural Network Models and Principal Component Analysis Method in Predicting Stock Prices on Tehran Stock Exchange", *Physica A: Statistical Mechanics and Its Applications*, Vol. 438: 178-187.
- Zheng, Huiting; Yuan, Jiabin; Chen, Long (2017), "Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation", *Energies*, Vol. 10, No. 8: 1168-1188.

Extended Summary

Stock Price Forecasting using Machine Learning Techniques

In today's global world, where economic boundaries between countries disappear and social, cultural and economic values are continually changing, "making decisions" represents a challenging process for both states, companies and individuals from all walks of society. This process affects investors' ability to invest by making the correct decision. In other words, it affects their ability to make the correct portfolio management. Investors need to have some knowledge of the stock when making the correct portfolio management. The future stock price forecasting from this information has emerged recently in the modern business world and has an information feature that is difficult to predict. Because the stock price is affected by many economic and non-economic factors (Liu et al., 2012: 1348), for this reason, many factors should be taken into account to make a correct prediction. Besides, an accurate forecasting method should be chosen.

Many studies have been done on the future stock price forecasting. Numerous methods such as Artificial Neural Networks (ANN), ARIMA, SVM, Time Series, Linear Regression, Decision Trees, Markov Chains have been tried in these studies. Some of these methods have been found unsuccessful and some have been supported by other methods. However, as can be seen in the studies conducted by Tektaş ve Karataş (2004), Toraman (2008), Hadavandi et al. (2010), Guresen et al. (2011), Adebisi et al. (2012) and Zahedi and Rounaghi (2015), it was revealed that the accuracy rate of the future stock price forecasting with ANN was higher than other methods. In other words, the most valid results regarding the future stock price forecasting were obtained by the ANN. For this reason, one of the preferred methods in the study was ANN.

This study aims to find the method that will enable to predict the future stock price correctly. In line with this purpose, the ANN, the Random Forest (RF) algorithm and the eXtreme Gradient Boosting (XGBoost) algorithm methods are discussed. The ANN has developed continuously since its first appearance. Ultimately, it has a different structure than the algorithmic calculation methods of today's computers. Thanks to these different structures, it provided useful results in producing solutions to many nonlinear problems. The RF algorithm is a method consisting of decision trees and nodes developed by Breiman (2001). Regression and classification analysis can be done with this method. The XGBoost algorithm is an application that has become a frequently preferred tool in decision trees and machine learning. This advanced method is considered as an essential tool in the field of supervised learning (Mitchell and Frank, 2017: 127).

In the analysis made in the study, firstly price prediction was made by using the ANN, then the RF and the XGBoost algorithms. Later the results obtained by different methods are compared. In this direction, 22 firms that are traded on the BIST 30 Index and whose financial statements are available are discussed. Quarterly financial statements of these companies between 01.01.2010 - 31.12.2019 were obtained from the Public Disclosure Platform (PDP) official website. Financial ratios (liquidity, activity, profitability and financial structure ratios) of firms were calculated with the help of financial statements. The rates obtained are used as input data in the analysis. In the analysis, monthly closing prices of stocks were used as output data. Output data were obtained from Turkey Stock Market-investing.com's official website. This data has been prepared by taking the quarterly average of closing prices of stocks between 01.01.2010 - 31.12.2019.

The machine learning process begins with the division of the dataset into two as "training data" and "test data". Although there are no rules in the literature, 70% of the data is randomly selected and assigned for educating the algorithms. Generalization power, which expresses the real-life success of the model developed after the training phase, is measured with the help of test data. It is not only that the test data shows high performance, but also the performance values of the machine learning model in the training and test data are expected to be close to each other.

The performance of any machine learning model depends directly on the control parameters (hyperparameter) selected. Choosing the correct parameter increases the performance of the model. Each machine learning model has a different number of control parameters and the best of them need to be determined simultaneously. The values for the experimental design developed for different machine learning models to be used in the study were determined by the literature analysis and expert opinions. Models were run for all possibilities in the experimental design. For example, the model in experiment design for the ANN (1, 4, Relu); it will work with 1 hidden layer, 4 neurons in the hidden layer and Relu activation function. Algorithms for implementation are coded in Python language. As a performance indicator, the coefficient of determination (R^2) was used to measure the suitability of the model. 10 runs were made

for each experiment, the best result and the average of these runs were reported. As a result of the experimental design, the best parameters are for the ANN; number of hidden layers 2, number of neurons in hidden layers are 6 and 10, Relu as activation function, for the XGBoost algorithm; maximum tree depth 6, learning coefficient 0,1, number of estimators 300, for the RF algorithm; the number of estimators 200, the maximum tree depth 10, the maximum number of attributes 5.

The optimal parameters found as a result of the experimental design were used for all three models selected for the application. It was determined that all three models gave successful predictions. When the results were examined for the training set, it was seen that the XGBoost and the RF models gave close results for the dataset used. Besides, both models outperformed the ANN. The ANN with two hidden layers was preferred because the single-layer artificial neural network architecture performed poorly. However, even the two-layer model lags behind the other two machine learning techniques used. Based on the training set performances, the XGBoost performed best. When looking at the test data performance of the models, they gave that the XGBoost algorithm 75.8%, the RF algorithm 70.2% and the ANN 70.8% accurate estimation. The R^2 values for the test dataset refer to the generalization power against the real-life data that the models have not encountered before. In line with these values, it was observed that the ANN had the lowest performance.

In the literature, studies conducted by Tektaş and Karataş (2004), Tsai and Wang (2009), Guresen et al. (2011), Adebisi et al. (2012) and Zahedi and Rounaghi (2015), in which another analysis method was used together with the ANN, both models were found to be successful. However, it was determined that ANN provides higher than other methods of performance. In this study, the results obtained with three different machine learning techniques were compared and it was found that the ANN performed lower. For this reason, in the light of existing variables, investors may be advised to analyze with any of the XGBoost or the RF techniques when making the future stock price forecasting because these techniques will lead them to more accurate results.