

Makine öğrenmesi yöntemleri ile hisse senedi fiyat tahmini: kâğıt firması örneği

Stock price prediction with machine learning methods: the case of paper company

Selahattin BARDAK¹  Nadir ERSEN^{2*}  Kinyas POLAT³  Kadri Cemil AKYÜZ⁴ 

¹Sinop Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Sinop

²Artvin Çoruh Üniversitesi, Artvin Meslek Yüksekokulu, Ormancılık ve Orman Ürünleri Programı, Artvin

³Sinop Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Metalürji ve Malzeme Mühendisliği Bölümü, Sinop

⁴Karadeniz Teknik Üniversitesi, Orman Fakültesi, Orman Endüstri Mühendisliği Bölümü, Trabzon

Eser Bilgisi / Article Info

Araştırma makalesi / Research article

DOI: 10.17474/artvinofd.1500569

Sorumlu yazar / Corresponding author

Nadir ERSEN

e-mail: nadirersen20@artvin.edu.tr

Geliş tarihi / Received

13.06.2024

Düzeltilme tarihi / Received in revised form

23.07.2024

Kabul Tarihi / Accepted

06.08.2024

Elektronik erişim / Online available

15.10.2024

Anahtar kelimeler:

Hisse senedi fiyatı

Gradyan güçlendirme

Rastgele orman

Doğrusal regresyon

Yapay sinir ağları

Tahmin

Keywords:

Stock price

Gradient boosting

Random forest

Linear regression

Artificial neural network

Prediction

Özet

Bir finansal formül kullanarak hisse senedi fiyatlarını tahmin etmek zordur. Hisse senetleri fiyatları, siyasi gelişmeler, küresel ekonomi, beklenmedik olaylar, piyasa anormallikleri ve ilgili şirketlerin özellikleri gibi çok sayıda faktörden etkilenir. Hisse senedi fiyatlarına ilişkin daha doğru tahminler yapmak için bilgisayar biliminin gelişmesiyle birlikte birçok bilgisayar bilimi yöntemi kullanılmaktadır. Bu çalışmada, Kartonsan şirketinin hisse senedi fiyatını tahmin etmek için doğrusal regresyon (LR) algoritmaları, rastgele orman (RF), gradyan güçlendirme makinesi (GBM) ve yapay sinir ağları (YSA) gibi makine öğrenmesi teknikleri kullanılmıştır. Daha sonra kullanılan algoritmaların sonuçları karşılaştırılmıştır. Hisse senedi fiyatı tahmini için ilk olarak BIST (Borsa İstanbul)'te işlem gören Kartonsan firmasının 2011-2022 yılları arasındaki üçer aylık finansal çizelgeler kullanılarak firmaya ait finansal oran hesaplanmıştır ve bu oranlar girdi olarak kullanılmıştır. Çıktı olarak kullanılan firmanın hisse senedi fiyatlarının ise üçer aylık ortalamaları alınmıştır. GBM ve RF algoritmaları başarılı tahmin sonuçlarına sahip olmasına rağmen GBM algoritması en başarılı sonucu vermiştir. RF algoritmasının ise LR ve YSA'ya göre daha iyi performans gösterdiği bulunmuştur. YSA'nın hisse senedi fiyat tahmininde en kötü performansa sahip teknik olduğu belirlenmiştir.

Abstract

It is difficult to predict stock prices using a financial formula. Stock prices are affected by numerous factors, including political developments, the global economy, unexpected events, market anomalies and the characteristics of the companies involved. With the development of computer science, many computer science methods are being used to make more accurate predictions of stock prices. In this study, machine learning techniques such as linear regression (LR) algorithms, random forest (RF), gradient boosting machine (GBM) and artificial neural Networks (ANN) were used to predict the stock price of Kartonsan company. Then, the results of the algorithms used were compared. For stock price prediction, firstly, the company's financial ratio was calculated using the quarterly financial statements of Kartonsan company, which is traded on BIST (Borsa İstanbul), between 2011 and 2022, and these ratios were used as input. Quarterly averages of the company's stock prices used as output were taken. Although GBM and RF algorithms have successful prediction results, the GBM algorithm gave the most successful results. It has been found that the RF algorithm performs better than LR and ANN. ANN has been identified as the technique with the worst performance in stock price prediction.

GİRİŞ

Gelişmekte olan ülkelerin büyümesinde borsa çok önemli bir rol oynamaktadır. Borsa yükselirse, ülkelerin ekonomik büyümesi yüksek olur. Borsa düşerse, o zaman ülkelerin ekonomik büyümesi de düşer. Başka bir deyişle, ülke büyümesinin borsanın performansına sıkı sıkıya bağlı olduğunu söyleyebiliriz (Sharma ve ark. 2017).

Borsa (hisse senedi piyasası), hisse senetlerinin, stokların, devlet tahvillerinin, teminatsız tahvillerin ve diğer onaylı menkul kıymetlerin alınıp satıldığı bir piyasadır. Borsa tahmini fikri ise yeni değildir, iş adamları genellikle

ekonomik göstergeler, kamuoyu ve mevcut siyasi iklim gibi dış parametreleri yorumlayarak piyasayı tahmin etmeye çalışmaktadırlar (Idowu ve ark. 2012). Borsanın nasıl performans göstereceğini tahmin etmek ise yapılması en zor şeylerden biridir. Tahmini etkileyen birçok faktör (fiziksel, fizyolojik, rasyonel ve irrasyonel davranışlar vb.) bulunmaktadır. Tüm bu unsurlar bir araya gelerek hisse fiyatlarını değişken hale getirmekte ve yüksek derecede doğrulukla tahmin edilmesi çok zor olmaktadır (Ravikumar ve Saraf 2020). Bu faktörlerin borsa analistleri tarafından dikkatli bir şekilde analiz edilmesi ve yatırımcıların bu analistler tarafından verilen tavsiyelere uyması

gerekmektedir. Hisse senedi piyasasındaki sürekli istikrarsızlık, yatırımcıların yanlış zamanda hisselerini satmalarının ve çoğu zaman kazanç elde edememelerinin ana nedenidir (Cavalcante ve ark. 2016). Borsa hareketini tam doğrulukla tahmin etmek mümkün olmasa da, hisse senedinin yanlış zamanda satılmasından kaynaklanan kayıplar ve etkileri, geçmiş verilere dayalı analize dayalı borsa hareketi tahmini kullanılarak büyük ölçüde azaltılabilir (Sakhare ve Imambi 2019).

Yapay zekâ alanındaki gelişmeler, akıllı özellikleriyle hayatın her alanında pek çok şeyi desteklemektedir. Yapay zekâda gelecekteki tahminlerde rol oynayan çeşitli algoritmalar mevcuttur. Makine öğrenmesi, hisse senedi fiyatlarının değerlerini tahmin etmede oyun değiştirici olarak kullanılan bir yapay zekâ alanıdır (Ghani ve ark. 2019, Ravikumar ve Saraf 2020). Makine öğrenmesi algoritmalarına dayalı olarak hisse senedi tahminlerinin gelecekteki fiyat hareketi eğiliminin araştırılması da giderek daha fazla ilgi görmeye başlamıştır (Hu ve ark. 2022). Makine öğrenmesi yöntemleri ile hisse senedi fiyatlarına ilişkin daha iyi tahminler yapılabilir ve bu yöntemler hisse senedi yatırımcılarına yatırım kararı vermede yardımcı olabilir (Cao 2021). Hisse senedi piyasası hareketinin tahmin edilmesi için klasik makine öğrenmesi (doğrusal regresyon, LASSO (Least Absolute Shrinkage and Selection Operator), Ridge, vb.), hem sınıflandırma hem de regresyon problemleri için kullanılan makine öğrenmesi (yapay sinir ağları, karar ağaçları, k-en yakın komşu, destek vektör makineleri, destek vektör regresyonu, vb.) ve topluluk makine öğrenmesi (torbalama, güçlendirme-hızlandırma, gradyan güçlendirme, rastgele orman, vb.) yöntemleri kullanılabilir (Ceylan 2018). Borsa hareketini yüksek hassasiyetle tahmin edebilecek en iyi algoritma diye bir şey yoktur. Bununla birlikte, makine öğrenmesi hisse senedi fiyat tahmini yaklaşımlarının çoğu, bireysel algoritmaların etkinliğinin değerlendirilmesine veya iyileştirilmesine odaklanırken, algoritmaların karşılaştırmalı çalışmasına daha az önem verilmektedir. Bu nedenle, hisse senedi hareketinin en optimal ve kesin tahminini sağlayan en iyi makine öğrenmesi algoritmasına ulaşmak ve böylece hisse senedi piyasasına yatırım yapan yatırımcıların mali kayıplarını en aza indirmek için farklı makine öğrenmesi algoritmalarının performans analizine ihtiyaç duyulmaktadır (Sakhare ve Imambi 2019, Vijh ve ark. 2020, Hu ve ark. 2022). Bunun için bu çalışmada, seçilen hisse senedinin fiyat tahmini için birden fazla makine öğrenmesi algoritması kullanılacaktır. Bunun, yatırımcılara algoritma seçme, risk-getiri tartma ve daha rasyonel yatırımlar yapma konusunda rehberlik etme konusunda önemli pratik sonuçları bulunmaktadır. Çalışmada kullanılan

makine öğrenmesi yöntemleri belirlerken klasik makine öğrenmesi, hem sınıflandırma hem de regresyon amacıyla kullanılan makine öğrenmesi ve topluluk makine öğrenmesi yöntemlerinden her birinden en az bir tane yöntem olmasına dikkat edilmiştir.

Bu çalışmada, makine öğrenmesi tekniklerinden olan yapay sinir ağı, gradyan güçlendirme makine, rastgele orman, doğrusal regresyon teknikleri kullanılarak Kartonsan firmasının hisse senedi fiyatı tahmin edilmeye çalışılmıştır ve kullanılan teknikler karşılaştırılmıştır. Çıktı değişkeni olarak firmanın son on iki yıla ait hisse senedi aylık kapanış fiyatlarının üç aylık ortalaması alınarak kullanılmıştır. Girdi değişkeni olarak ise firmaya ait dokuz finansal oran kullanılmıştır. Çalışmanın ilk bölümünde hisse senetleri ve makine öğrenme yöntemleri ile ilgili bilgiler verilmiştir. İkinci bölümde makine öğrenme yöntemleri ile hisse senedi tahminine yönelik çalışmalardan bahsedilmiş, üçüncü bölümde çalışmada kullanılan veriler ve teknikler hakkında bilgiler sunulmuştur. Dördüncü bölümde analiz sonucunda elde edilen sonuçlara yer verilmiştir ve çıkan sonuçlar ile benzer çalışmaların sonuçları karşılaştırılmıştır. Son bölümde ise çalışma sonuçları genel olarak değerlendirilmiş olup, gelecekte yapılabilecek araştırmalar konusunda bilgilerde verilmiştir.

Literatür Taraması

Uluslararası ve ulusal literatürde makine öğrenmesi yöntemleri hisse senedi fiyat ve fiyat yönü tahminine yönelik çalışmalar bulunmaktadır.

Yıldırım ve ark. (2011), yapay sinir ağları (YSA) ve çoklu doğrusal regresyon (MLR) kullanarak Türkiye'deki kâğıt sektörünün finansal getirisini tahmin etmişlerdir. Çalışma sonucunda her iki yöntemin de oldukça başarılı sonuçlar vermesine rağmen YSA'nın MLR'den önemli ölçüde daha iyi performans gösterdiği gözlemlenmiştir.

Manojlovic ve Stajduhar (2015), Zagreb Borsa'sındaki dört hisse senedinin 5 gün ileri ve 10 gün ileri yönlerini tahmin etmek için rastgele orman algoritmasını kullanarak bir model oluşturmuşlardır. Elde edilen sonuçlar hisse senedi piyasası eğilimlerinin yönünü tahmin etmeye yönelik tahmine dayalı modeller oluşturmak için rastgele orman algoritmasının başarıyla kullanılabileceğini göstermektedir.

Tekin ve Çanakoğlu (2018), tarafından yapılan çalışmada, çeşitli makroekonomik veriler ve şirketlere ait temel finansal değerleri girdi ve Çok Katmanlı Algılayıcı, Rastgele Orman, Rastgele Ağaç, Lojistik Regresyon, Naive Bayes ve Destek Vektör Makineleri yöntem olarak kullanılarak Borsa

İstanbul'da önde gelen 25 şirketin hisse senedi getirilerinin tahmini araştırılmıştır. Rastgele Orman algoritmasının en iyi sonuç veren algoritma olduğu sonucuna ulaşılmıştır.

Bhardwaj ve Akil Ansari (2019), tarafından piyasa davranışını analiz etmek amacıyla gelecekte hisse senedi fiyatlarını değerlendirmek için makine öğrenmesi destekli algoritmalarından olan lojistik regresyon, rastgele orman, K-en yakın komşular ve otoregresif entegre hareketli ortalama (ARIMA) algoritmaları karşılaştırılmıştır. Burada % 74 doğrulukla Lojistik Regresyonun en iyi performansı gösterdiğine ulaşılmış olup, bunu sırasıyla rastgele orman, K-en yakın komşular ve ARIMA takip etmektedir.

Diğer çalışmada, karar ağacı, çoklu regresyon ve rassal orman yöntemlerini kullanarak beş bankanın hisse senedi fiyatları tahmin edilmiştir. Girdi olarak ilk önce 46 teknik gösterge ve 4 temel gösterge olmak üzere 50 gösterge ve daha sonra teknik gösterge 29'a indirgenerek 33 gösterge ile analiz yapılmıştır. Analiz sonucunda kullanılan yöntemlerin banka hisse senetlerinin bir gün sonraki kapanış değerlerini tahmininde başarılı olduğu tespit edilmiştir (Akşehir ve Kılıç 219).

Vijh ve ark. (2020) tarafından JP Morgan, Nike, Johnson ve Johnson, Goldman Sachs ve Pfizer olmak üzere beş farklı sektör şirketinin gelecekteki hisse senedi fiyatlarının tahmini yapılmıştır. Bu amaç doğrultusunda YSA ve RF teknikleri kullanılmış ve iki teknik arasında bir karşılaştırma yapılmıştır. Araştırma sonucunda, YSA'nın RF'ye kıyasla hisse senedi fiyatlarına ilişkin daha iyi tahmin sağladığı bulunmuştur.

Öztürkmen ve Eren (2020), tarafından yapılan çalışmada, on beş farklı yapay öğrenme algoritmaları kullanarak BIST100 endeksinin fiyat yükseliş ve düşük yönünü tahmin edilmeye çalışılmıştır. En başarılı performans gösteren algoritmanın Hafif Gradyan Arttırma Makineleri (LightGBM) algoritması olduğu tespit edilmiştir.

Apple hisse senedi veri kümesinin hisse senedi fiyatı ve momentumunun tahmin edildiği çalışmada, hisse senedi fiyatlarının tahmininde en iyi sonuç veren % 99.57 doğrulukla rastgele orman modeli ve sınıflandırmada ise % 68.622 doğrulukla lojistik regresyon modelinin olduğu sonucuna varılmıştır (Ravikumar ve Saraf 2020).

Roy ve ark. (2020) yaptıkları çalışmada, Kore Bileşik Hisse Senedi Fiyat Endeksi'nden (KOSPI) işlem gören şirketlerin hisse senet fiyatları tahmin edilmiştir. Bu amaç için derin öğrenme, rastgele orman ve gradyan destekli makineler kullanılmıştır. Araştırma sonucunda, tüm yöntemlerin

tatmin edici bir performans sağladığı bulunmuştur. Ancak, derin öğrenme diğer iki yönteme göre ikinci planda çıkmıştır.

Koç Ustalı ve ark. (2021), tarafından yapılan çalışmada, yapay sinir ağları, rastgele orman algoritması ve ekstrem gradyan arttırma (XGBoost) algoritması gibi makine öğrenmesi yöntemlerinin BİST-30 endeksinde yer alan firmaların hisse senetlerinin fiyatlarının tahmininde başarılı sonuçlar verdiğini belirtilmiştir.

Ereğli Demir ve Çelik Fabrikaları T. A. Ş.'nin hisse senedi fiyatları Arslankaya ve Toprak (2021) tarafından tahmin edilmeye çalışılmıştır. Bu amaç doğrultusunda makine öğrenmesi yöntemlerinden Polinom Regresyon (PR) ve Random Forest Regresyon (RFR), derin öğrenme yöntemlerinden Tekrarlayan Sinir Ağları (RNN) ve Uzun Kısa Süreli Bellek (LSTM) algoritmaları kullanılmıştır. Hata değerlerine göre en başarılı modeller sırasıyla RFR modeli, LSTM modeli, RNN modeli ve PR modeli çıkmıştır.

Çalışmada, iki borsa endeksi ve İstanbul borsasının yüksek hacimli iki hisse senedinin gelecek değerleri çeşitli makine öğrenmesi yöntemleri (Destek vektör regresyonu, rastgele orman, ekstrem gradyan arttırma) ile tahmin edilmeye çalışılmıştır. Çalışmada tercih edilen endeksler New York Borsası (NYSE) ve İstanbul Borsasına (BİST 100) ait birleşik endeksler iken, hisse senetleri ise Borsa İstanbul'da işlem gören Ereğli Demir ve Çelik ile Aselsan Şirketlerine ait hisse senetleridir. Makine öğrenmesi yöntemlerinin hem yatırımcılar hem de araştırmacılar için kullanılabilecek faydalı bir araç olduğu sonucuna varılmıştır (Doğan ve Büyükkör 2022).

Hisse senedi fiyat tahminine yönelik yapılan başka bir çalışmada, gradyan arttırma makine (GBM) ile destek vektör makineleri (SVM) karşılaştırılmıştır. Karşılaştırma sonucunda, GBM yönteminin SVM yöntemine göre daha başarılı olduğu sonucuna ulaşılmıştır (Reddy ve Kumar 2022).

Hu ark. (2022), tarafından sonuçların tutarlılığını doğrulamak için üç hisse senedi üzerinde Hafif Gradyan Arttırma Makineleri, Rastgele Orman ve Lojistik Regresyonun bir dizi tahmin penceresindeki yukarı veya aşağı sınıflandırma performansları karşılaştırılmıştır. Göreceli güç endeksi, momentum endeksi, denge hacmi, dalgalanma (volatilite) oranı, stokastik gösterge gibi 12 teknik gösterge modelleri eğitmek için girdiler olarak seçilmiştir. Modellerin tahmin performansının kısa ve uzun zaman aralığında önemli ölçüde farklı olduğu sonucunda ulaşılmıştır.

Verma ve Soni (2022), gradyan artırmaya dayalı bir yaklaşım kullanarak hisse senedi fiyatları açısından hisse senedi piyasası hareketini tahmin etmeye yönelik yaptığı çalışmada, GB yaklaşımının dört kıyaslama veri seti için %89.75, %90.40, %87.48 ve %88.78 doğruluk değerlerine sahip olduğu saptanmıştır.

2022 yılında yapılan başka bir çalışmada ise Çin borsasındaki çeşitli makine öğrenmesi yöntemlerinin tahmin gücü araştırılmıştır. Çalışmada en az mutlak küçültme ve seçim operatörü (LASSO), sıradan en küçük kareler (OLS) regresyonu, kısmi en küçük kareler regresyonu (PLS), gradyan destekli regresyon ağaçları (GBRT), elastik ağ (Enet), rastgele orman (RF), değişken alt örnekleme tahmin toplama (VASA) ve sinir ağları yöntemleri kullanılmıştır. Çalışma sonucunda, makine öğrenmesi yöntemlerinin ABD pazarından tamamen farklı özelliklere sahip pazarlara başarıyla uygulanabileceği belirlenmiştir (Leippold ve ark. 2022).

Nakagawa ve Yoshida (2022) yaptığı araştırmada, hisse senedi fiyat tahmininde zaman serisi gradyan artırma ağacının hem kârlılık hem de doğruluk açısından önceki araştırma yöntemlerinden üstün olduğunu doğrulamıştır.

BIST 30 Endeksi'nde işlem gören holding, beyaz eşya, petrokimya, demir çelik, ulaştırma ve iletişim sektörlerindeki şirketlerin hisse senetlerinin gelecekteki fiyatlarının tahmin edilmesinin amaçlandığı bu çalışmada, makine öğrenmesi algoritmaları olarak XGBoost, SVM, KNN ve RF kullanılmıştır. Analiz sonucunda XGBoost algoritmasının hem sektörler hem de şirketler bazında en başarılı algoritma iken, en kötü performans ise SVM algoritması çıkmıştır. Ayrıca, RF algoritmasının sektör ve şirket bazında en iyi ikinci performansa sahip algoritma olduğu sonucuna ulaşılmıştır (Kocaoğlu ve ark. 2022).

Xu (2022) tarafından araştırılan çalışmada, Netflix hisselerinin 2002'den 2021'e kadar olan fiyat verilerine dayanarak, Netflix hisselerinin son yedi günlük hisse senedi getirisine dayalı olarak sekizinci günlük hisse senedi getirisini tahmin edilmiştir. Çalışmada model olarak doğrusal regresyon, karar ağacı ve gradyan güçlendirme olmak üzere üç model kullanılmıştır. Çalışma sonucunda, doğrusal regresyon modelinin Netflix tipi hisse senetlerinin uzun vadeli getirisini tahmin etmek için en iyi model olduğu belirlenmiştir.

Diğer bir taraftan, 2022 yılında Shrivastav tarafından yapılan bir çalışmada, Derin Öğrenme (DL), Gradient Boosting Machine (GBM) ve dağıtılmış Rastgele Orman (DRF) veri analitiği teknikleri kullanılarak borsa veri tahmini

için bir topluluk modeli önerilmiştir. Ayrıca, Topluluk modelinin performans sonuçları, DP, GBM ve DRF gibi bireysel yöntemlerin her biriyle karşılaştırılmıştır. Topluluk modelinin Derin Öğrenme ve GBM'den çok daha iyi sonuç verdiği ve topluluk modeli ve DRF kullanılarak elde edilen tahmin sonuçlarının ise hisse senedi fiyatlarının orijinal değerlerine oldukça yakın olduğu tespit edilmiştir.

Singh (2022) yaptığı çalışmada, adaptif güçlendirme, k-en yakın komşular, doğrusal regresyon, yapay sinir ağları, rastgele orman, stokastik gradyan iniş, destek vektör makinesi ve karar ağaçları olmak üzere sekiz denetimli makine öğrenmesi modeli kullanarak Hindistan Hisse Senedi Piyasası Nifty 50 Endeksi'ni tahmin etmeye odaklanmıştır. Analiz sonucunda, Destek Vektör Makinesinin, diğer modeller arasında daha iyi performans gösterdiği ancak veri kümesinin boyutundaki artışla Stokastik Gradyan İniş, Destek Vektör Makinesinden daha iyi performans gösterdiği belirlenmiştir. Veri seti boyutunun artmasıyla birlikte Adaptif Boost, k-En Yakın Komşular, Rastgele Orman ve Karar Ağaçlarının yetersiz performans gösterdiği anlaşılmıştır. Bununla beraber, Doğrusal Regresyon ve Yapay Sinir Ağları, farklı boyuttaki verilerin tüm segmentlerinde neredeyse eşit performans sergilemiştir.

Akyol Özcan (2023) gerçekleştirdiği çalışmada, borsa endekslerinin artış ve azalışları makine öğrenmesi algoritmaları ile tahmin etmeye çalışılmış ve algoritmalar karşılaştırılmıştır. Bu amaç doğrultusunda on değişkenin 1. ve 2. gecikmeli değerleri alınmıştır ve dokuz makine öğrenmesi algoritması kullanılmıştır. Doğruluk oranı, F-1 skoru ve Matthew korelasyon katsayısı kullanılarak makine öğrenmesi algoritmalarının performansları değerlendirilmiştir. Değerlendirme sonucunda, uzun ve orta vadede lojistik regresyon ve lineer diskriminant ve kısa vadede polinomial çekirdek fonksiyonu ile destek vektör makinelerinin performansı diğer algoritmalara göre daha yüksek çıkmıştır. Aynı zamanda rastgele orman algoritması eğitim verisinde yüksek, test verisinde düşük sonuçlar verdiği saptanmıştır.

Hartanto ve ark. (2023), yüksek doğruluk ve verimliliğe sahip olan LightGBM algoritmasını kullanarak hisse senedi fiyat tahminleri oluşturmuşlardır. Aynı zamanda LightGBM algoritması diğer algoritmalarla karşılaştırılarak hisse senedi fiyat verilerinin tahmininde hangi modelin optimal olduğu gözlemlenmiştir. Analiz sonuçları, GBM'nin versiyonu olan LightGBM modelinin yine GBM'nin versiyonları olan XGBoost, AdaBoost ve CatBoost gibi güçlendirme tabanlı tahmin modelleriyle rekabet

edebileceğini ve onlardan daha iyi performans sergileyebileceğini göstermiştir.

Kesharwani ve ark. (2023) tarafından yapılan bir çalışmada ise hisse senedi varlıklarının fiyatlarını tahmin etmek için bir makine öğrenmesi modeli sunulmuştur. Gradient Boosted Regressor modeli %98.28 test puanı ile bu çalışmada önerilen dört model arasında en doğru model olarak belirlenmiştir.

Reddy ve Kumar (2023), hisse senedi değerlerine ilişkin hisse senedi fiyatları tahmin edilmeye çalışılmış ve bunun için rastgele orman ve gradyan güçlendirme makinesi algoritmalarından yararlanılmışlardır. Rastgele orman algoritmasının doğruluğu ile karşılaştırıldığında gradyan güçlendirme makinesi algoritmasının doğruluğu çok daha yüksek (%92.3) çıkmıştır.

Toprak ve ark. (2023) tarafından yapılan bir çalışmada ise BİST'teki Petkim Petrokimya Holding A.Ş.'nin hisse senedi fiyatları rastgele orman regresyonu, uzun-kısa süreli bellek ve evrimsel sinir ağları ile tahmin edilmiştir. Analiz sonucunda, çalışmada kullanılan algoritmaların determinasyon katsayılarına (R^2) göre hisse senedi fiyatlarının tahmininde kullanılabileceği görülmüştür.

MATERYAL VE YÖNTEM

Materyal

Kartonsan firmasının hisse senedinin fiyatının tahmini için çıktı olarak kullanılan firmanın 2011-2022 tarihleri arasındaki hisse senedi aylık kapanış fiyatları Investing.com web sayfasından temin edilmiş olup (URL-1 2023), verilerin üçer aylık ortalaması alınmıştır. Çalışmada girdi olarak kullanılan finansal oranların hesaplanmasında yararlanılan firmanın çalışma kapsamındaki döneme ait üç aylık finansal çizelgeler Kamu Aydınlatma Platformu (KAP) web sayfasından temin edilmiştir (URL-2 2023). Çalışmada hesaplanan finansal oranlar ise şunlardır: Cari oran, asit-test oranı, nakit oranı, özsermaye devir hızı, özsermaye kârlılığı, net kâr marjı, uzun vadeli yabancı kaynaklar / toplam aktifler, kısa vadeli yabancı kaynaklar / toplam aktifler, özsermaye / toplam aktifler'dir.

Yöntem

Yapay Sinir Ağları

Yapay sinir ağları insan beyni bilgilerinin toplanmasını ve işlenmesini teşvik etmek için tasarlanmış bilgisayar sisteminin bir bileşenidir (Patel ve ark. 2021). YSA, birbirine

bağlı birçok düğümden (yapay sinirlerden) oluşur. Sinir ağı, nöron adı verilen birkaç tek işlem elemanından oluşan, biyo-esinli bir sistemdir. Nöronlar, bir dizi atanmış ağırlıktan oluşan eklem mekanizmasıyla birbirine bağlanır. Nöron, giriş parametrelerinin değerlerini alır, bunları atanan ağırlıklara göre toplar ve bir sapma değeri ekler. Transfer fonksiyonu uygulanarak çıktıların değeri belirlenir. Giriş katmanındaki nöronların sayısı, giriş parametrelerinin sayısına eşittir (Moghaddam ve ark. 2016). Matematiksel açıdan p nöronunun performansı şu şekilde tanımlanabilir:

$$U_p = \sum_{i=1}^n w_{pi} x_i \quad (2.2.1)$$

$$y_p = \varphi(u_p + b_p) \quad (2.2.2)$$

Burada, $x_1, \dots, x_n$, giriş parametreleridir; $w_{p1} \dots w_{pn}$, P nöronunun bağlantı ağırlıklarıdır; u_p , giriş birleştiricisidir; b_p , eşik değeridir; φ , aktivasyon fonksiyonudur ve y_p , nöronun çıktısıdır (Moghaddam ve ark. 2016).

YSA'ların önemli bir uygulama alanı tahmindir ve araştırmacıların yanı sıra uygulayıcılar için de çekici bir alternatif araç sağlar. YSA'ların aşağıdaki özellikleri onları bir tahmin görevi için değerli ve çekici kılmaktadır.

i) Geleneksel model tabanlı yöntemlerin aksine, örneklerden öğrenir ve altta yatan ilişkiler bilinmese veya tanımlanması zor olsa bile veriler arasındaki incelikli işlevsel ilişkileri yakalar. Deneyimlerden öğrenme yeteneği verir ve gerçek dünya sorunlarını çözmek için pratik ve uygulanabilir bir yol sağlar.

ii) YSA'lar, örnek veriler en az prensipte gürültülü bilgiler içerse bile, verilerin görünmeyen kısmını genelleyebilir ve doğru bir şekilde çıkarım yapabilir.

iii) YSA'lar evrensel işlevsel yaklaşım araçlarıdır ve istenen doğruluk için sürekli bir işlev sağlar.

iv) YSA'lar, girdi ve çıktı değişkenleri arasındaki ilişki hakkında önceden bilgi sahibi olmadan doğrusal olmayan modelleme yapma kapasitesine sahiptir. Bu nedenle YSA'lar tahmin için daha genel ve esnek bir modelleme aracıdır (Devadoss ve Ligori 2013).

Rastgele Orman

Rastgele orman, bir dizi regresyon karar ağacından elde edilen sonuçları birleştirerek çıktı tahminleri yapan ve sınıflandırma ve regresyon analizi için kullanılan bir

topluluk makine öğrenmesi tekniğidir. RF, mevcut bir veri kümesinin yerine yeni bir veri kümesi oluşturma yöntemi olan torbalama (veya önyükleme toplama) tekniğini uygulamaktadır. Bu yöntemde, her ağaç bağımsız olarak oluşturulur ve ormandaki tüm ağaçlar aynı dağılıma sahip olacak şekilde girdi verilerinden örneklenen rastgele bir vektöre bağlıdır. Ormanlardan gelen tahminlerin önyükleme toplama ve rastgele özellik seçimi kullanılarak ortalaması alınır. RF, hem küçük veri setleri hem de yüksek veri setleri ile verimli bir şekilde çalışmaktadır (Biau ve Scornet 2016, Umoh ve ark. 2022). RF, eğitim işleminin diğer algoritmalara göre daha hızlı olması ve sınıflandırma doğruluğunun yüksek olması nedeniyle de oldukça kullanışlıdır (Srivastava ve ark. 2023). Ancak, diğer makine öğrenmesi algoritmalarıyla karşılaştırıldığında RF daha karmaşık ve zaman alıcıdır (Wang ve Gao 2022). RF regresyon tahmincisi ise aşağıdaki gibi ifade edilebilir (Ahmad ve ark. 2018).

$$f_{RF}^c(x) = \frac{1}{C} \sum_{i=1}^C T_i(x) \quad (2.2.3)$$

Burada, x vektörlenmiş giriş değişkeni, C ağaç sayısı ve $T_i(x)$ girdi değişkenlerinin bir alt kümesine ve önyüklemeli örneklerle dayalı olarak oluşturulan tek bir regresyon ağacıdır.

Doğrusal Regresyon

Doğrusal regresyon, bir öğrenme modeli oluşturmayı amaçlayan ve girdi verilerine en iyi şekilde doğrusal bir çizgi uydurarak iki değişken arasında ilişki kurmaya çalışan denetimli bir makine öğrenmesi tekniğidir (Sakhare ve Imambi 2019). Doğrusal regresyon modelinin genel matematiksel ifadesi ise aşağıda verilmiştir.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \epsilon \quad (2.2.4)$$

Burada, Y bağımlı değişken, β_0 regresyon sabiti, $\beta_1, \beta_2, \dots, \beta_n$ bağımsız değişkenlerin ağırlıkları, $X_1, X_2, \dots, X_n$ bağımsız değişkenler ve ϵ ise regresyon hatasıdır (Su ve ark. 2012). Y tahmin edilecek bir değişkendir ve kriter değişkeni olarak da bilinir. Y 'nin tahmini x 'in değerlerine dayandığından x 'e açıklayıcı değişken adı verilir. Açıklayıcı değişken bir veya daha fazla olabilir. Yalnızca bir açıklayıcı değişken olduğunda tahmin yöntemine basit regresyon, birden fazla açıklayıcı değişken olduğunda tahmin yöntemine çoklu regresyon adı verilir (Sharma ve ark. 2017, Sakhare ve Imambi 2019).

Tahmin için doğrusal regresyon modelini kullanmadan önce, tahmin edilecek (bağımlı) değişken ile girdi verileri (bağımsız değişkenler) arasında bir ilişki olup olmadığı belirlenmelidir.

Bu iki değişken arasındaki ilişkinin gücü dağılım grafiği kullanılarak belirlenebilir (Rubio ve ark. 2016).

Doğrusal regresyon, giriş veri noktaları boyunca en uygun düz çizgiyi aramaktadır. En uygun çizgiye regresyon çizgisi denir. Regresyon çizgisi bir veri grubu için modellendikten sonra çizgiden uzakta bulunan veri değerlerine ise aykırı değer adı verilir. Aykırı değerler hatalı verileri temsil eder ve uygun olmayan bir regresyon çizgisine işaret edebilir. Bu tür hatalı veri değerleri ile tahminin doğruluğu büyük ölçüde azalabilmektedir. Doğrusal regresyon her zaman çıktı değişkeni ile girdi değişkeni arasında doğrusal bir ilişkinin olduğu varsayımına dayanır ve bu her zaman doğru değildir. Doğrusal regresyon, doğrusal olmayan veriler için en iyi uyumu sağlamaz (Sakhare ve Imambi 2019). Son olarak, doğrusal regresyon modelleri genellikle en küçük kareler yaklaşımı kullanılarak uyarlanmaktadır (Cao 2021).

Gradyan Güçlendirme Makinesi

Gradyan güçlendirme makinesi, makine öğrenmesinde sıklıkla kullanılan ve doğrusal olmayan karmaşık ilişkileri modelleyebilen bir algoritmadır. Temel olarak GBM algoritması birçok basit karar ağaçlarını bir araya getirerek güçlü bir model oluşturur. Her biri adım adım modelin performansını artırır. GBM modeli bir başlangıç model ile tahmin oluşturur ve bu model matematiksel olarak aşağıdaki gibi ifade edilebilir (Touzani ve ark. 2018, Lu ve Mazumder 2020, Siswoyo 2021, Acharya ve ark. 2021, Konstantinov ve Utkin 2021);

$$F^0(x) = \argmin_{\gamma} \sum_{i=1}^n L(y_i, \gamma) \quad (2.2.5)$$

Burada $F^0(x)$ başlangıç modelidir. Model tahmini tüm veri seti üzerinde sabit bir değer (γ) ile başlar ve bu değer hedef değişkenin ortalaması veya medyanı olabilir. Bu değer kayıp fonksiyonunu (L) en küçük değere getirir. Eşitlikteki y_i ise veri setindeki gerçek değerleri temsil etmektedir. GBM, her adımda modelin önceki adımlardaki hatalarını düzeltmeye çalışan yeni bir karar ağacı ekler. Bu süreç, modelin eğitim verilerindeki ilişkileri daha iyi kavramasına yardımcı olur. Her yeni ağacın katkısı, önceki modelin tahminlerine eklenir. Matematiksel olarak bu durum aşağıdaki eşitlik ile ifade edilebilir;

$$F_m(x) = F_{m-1}(x) + v \cdot h_m(x) \quad (2.2.6)$$

Burada $F_{m-1}(x)$ bir önceki tahmini, $h_m(x)$ m inci tahmini, v ise öğrenme oranını temsil etmektedir. Her ağacın tahmini ($h_m(x)$), kayıp fonksiyonunun gradyanını (türevini)

azaltacak şekilde eğitilir. Bu süreçte, aşağıdaki eşitlikte belirtildiği gibi her adımdaki ağacın tahmini, önceki adımların toplam tahminlerine eklenerek, kayıp fonksiyonu minimize edilmeye çalışılır.

$$h_m = \operatorname{argmin}_h \sum_{i=1}^n L(y_i, F_m^{-1}(x_i) + h(x_i)) \quad (2.2.7)$$

Belirlenen adım sayısı (iterasyon) veya başka bir durdurma kriteri karşılanana kadar bu süreç devam eder. Modelin son hali, başlangıç modeli ve tüm iterasyonlardaki ağaçların toplamı olarak aşağıdaki gibi gösterilebilir.

$$F(x) = F^0(x) + v \cdot \sum_{m=1}^M h_m(x) \quad (2.2.8)$$

Modelin başarısı zayıf karar ağaçlarının bir araya getirilerek her adımda kayıp fonksiyonun minimize edilmesine dayalı optimizasyondan kaynaklanmaktadır (Touzani ve ark. 2018, Lu ve Mazumder 2020, Siswoyo 2021, Acharya ve ark. 2021, Konstantinov ve Utkin 2021).

Modellerin Değerlendirilmesi

Yöntemlerin performans kriteri olarak MSE, MAE ve R^2 değerleri belirlenerek karşılaştırma yapılmış ve modellerin eğitim ve test verileri üzerindeki başarısı değerlendirilmiştir. Performans ölçütleri olarak belirlenen MSE, MAE ve R^2 değerlerinin hesaplanması için kullanılan matematiksel eşitlikler aşağıda verilmiştir (Botchkarev 2018, Correa-Jullian ve ark. 2020).

Ortalama Karesel Hata (MSE): Tahmin edilen değerlerin gerçekten ne kadar uzakta olduğunu göstermektedir.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.2.9)$$

Ortalama Mutlak Hata (MAE): Hataların mutlak değerlerinin ortalamasını göstermektedir.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2.2.10)$$

Determinasyon Katsayısı (R^2): Modelin veriler arasındaki sapmayı ne ölçüde açıklayabildiğini gösterir.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (2.2.11)$$

Burada y_i gerçek değeri, $\hat{y}_i$ tahmin edilen değeri, N toplam örnek sayısını temsil etmektedir.

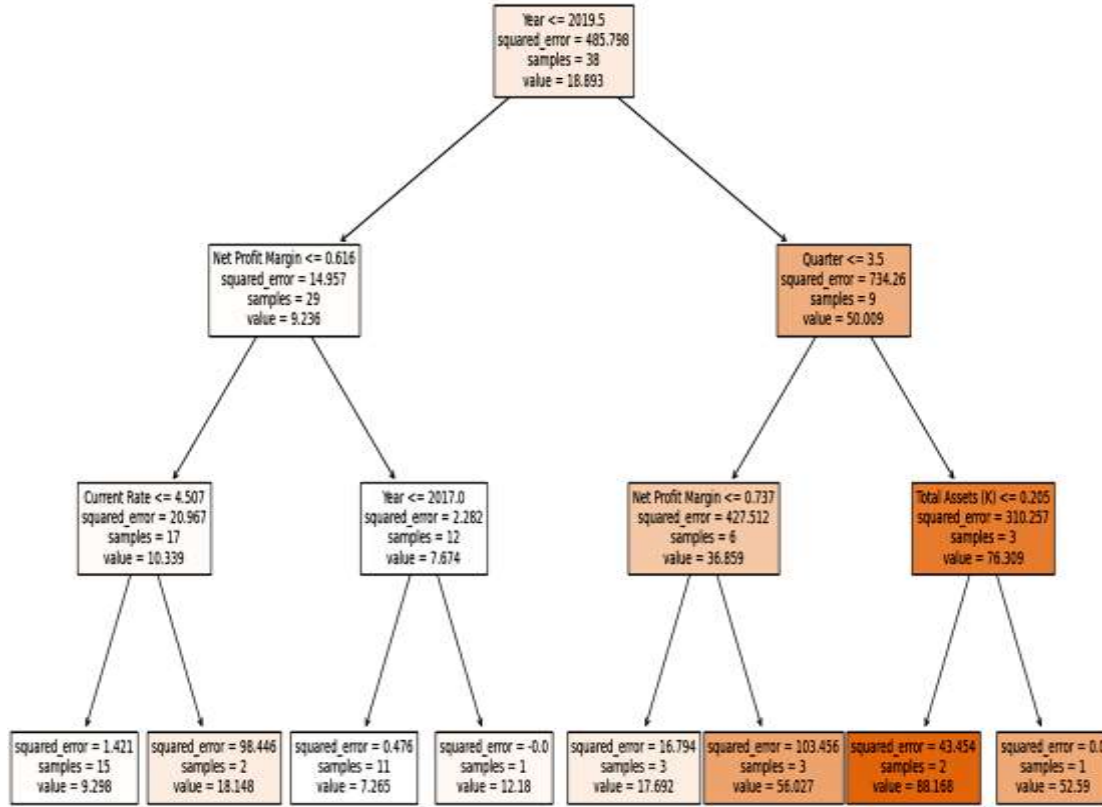
BULGULAR VE TARTIŞMA

Makine öğrenmesi python programlama dili kullanılarak gerçekleştirilmiştir. Bu amaçla yapay sinir ağları modeli (Dike ve ark. 2018) için programının pandas, torch, sklearn, matplotlib ve numpy kütüphaneleri yüklenmiştir. Veri setleri içinden yıl ve çeyrek bilgileri çıkarılmıştır. Hisse senedi fiyat değişkeni bağımlı değişken (çıktı) olarak atanırken, dokuz finansal oran bağımsız değişkenler (girdi) olarak atanmıştır.

Veriler %80 eğitim verisi ve %20 test verisi olarak gruplandırılmıştır. Veriler pytorch kütüphanesi yardımıyla iki boyutlu tensörlere dönüştürülmüştür. Daha sonra girdi katmanı 9 nöron, gizli katmanı 64 nöron ve çıktı katmanı bir nöron olan bir yapay sinir ağları modeli oluşturularak eğitilmiştir. Aktivasyon fonksiyonu olarak düzeltilmiş doğrusal birim (ReLU) kullanılmıştır (He ve ark. 2015). Model MSE (ortalama kare hata) kayıp fonksiyonu ve ADAM optimizasyon algoritması kullanılarak eğitilmiştir (Attrapadung ve ark. 2022). ADAM algoritması uyarlanabilir öğrenmeyi sağlayan momentum ve RMSprop olarak isimlendirilen iki bileşenden meydana gelmektedir. Momentum bileşeni önceki gradyanların birikimidir ve modelin minimumlardan kurtulmasını sağlamaktadır. RMSprop ise farklı parametreler için öğrenme oranını ayarlamaktadır. Modelde hesaplama için “numpy” kütüphanesi kullanılmıştır. Modelin eğitimi için oluşturulan algoritmanın uygulanması “scikit-learn” kütüphanesi ile gerçekleştirilmiştir. Veri düzenleme için “pandas” kütüphanesi, gerçek ve tahmini değerlerin iki boyutlu regresyon grafiğine dönüştürülmesi için “matplotlib” kütüphanesi kullanılmıştır.

Lineer regresyon modelinde python programının “pandas”, “scikit-learn” ve “matplotlib” kütüphanelerinden faydalanılmıştır.

Temel kütüphanelere ek olarak “Random Forest” modeli (Hu ve Szymczak 2023) için ise tipik regresyon problemlerinin dallanma kriterlerini kullanan “RandomForestRegressor” sınıfı tatbik edilmiştir. Buradaki dallanma kriteri “Mean Squared Error”(MSE) ve “Mean Absolute Error”(MAE) hata ölçütlerini kullanmaktadır. Random forest modeli için maksimum karar ağacı sayısı 100 olarak belirlenmiştir. Karar ağacı, her kutucukta, kararın alındığı özellik (feature), eşik değeri, örnek sayısı, değeri ve düğümün ortalamasını gösteren akış şeması Şekil 1’de verilmiştir.

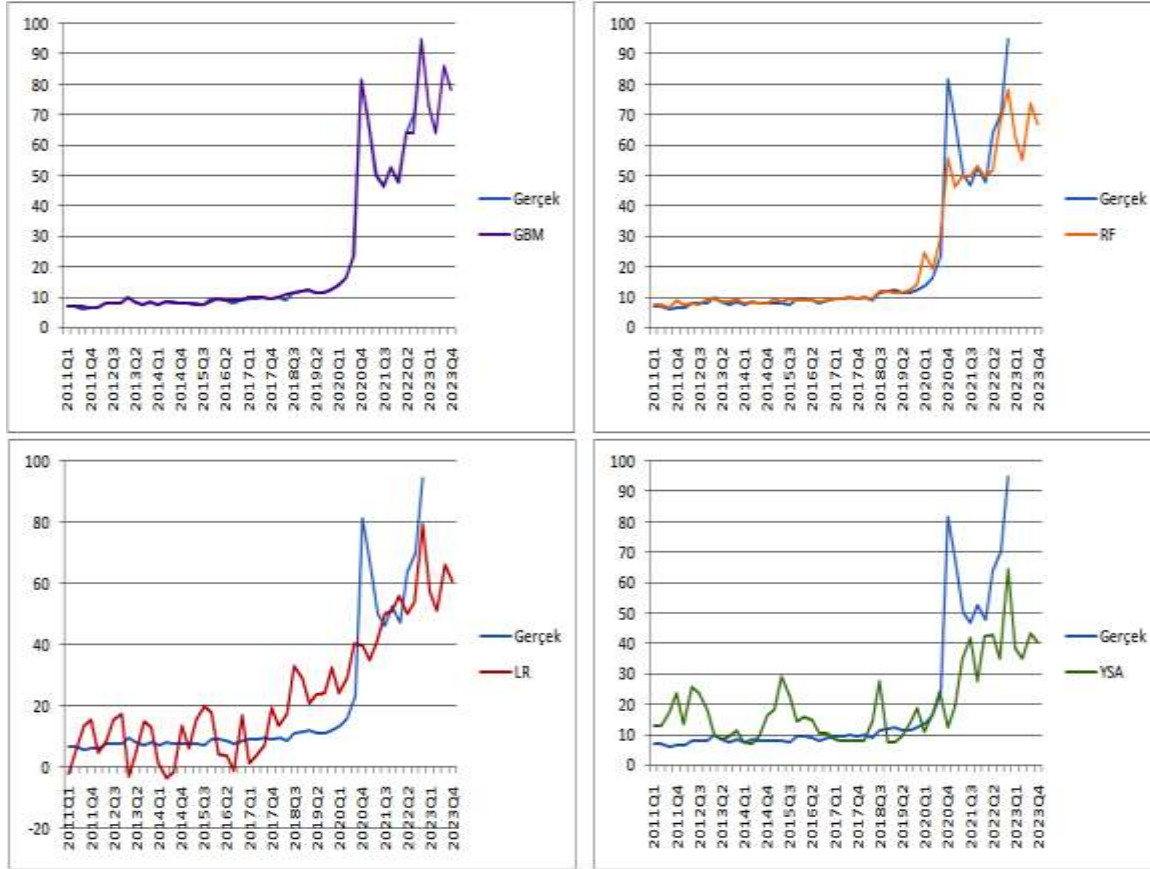


Şekil 1. Random forest karar ağacı

GBM modeli, Python programlama dilinde mevcut bulunan “scikit-learn” kütüphanesinin “GradientBoostingRegressor” fonksiyonu kullanılarak eğitilmiştir. Eğitim ve test verilerinin belirlenmesinde kullanılan “randomstate” parametresi modelin R^2 değerini en yüksek çıkaracak şekilde 0 ile 50 arasında bir değerden seçilmiştir. Modelin performansı eğitim ve test verilerinin ayrı ayrı MSE, MAE ve R^2 değerleri hesaplanarak belirlenmiştir. Eğitim ve test verilerinin yüzdelik oranı 0.8 ile 0.2 olarak ayarlanmıştır.

Çizelge 1'deki modellerin performansları incelendiğinde, GBM'nin eğitim aşaması için MAE değeri 0.055 ve test aşaması için 1.32, RF'nin eğitim aşaması için MAE değeri 2.14 ve test aşaması için 4.48, LR ile elde edilen MAE değerleri eğitim için 9.21 ve test için 13.16 ve YSA'nın eğitim aşaması için MAE değeri 8.85 ve test aşaması için 14.91 olarak bulunmuştur. Modelin geçerliliğini değerlendirmek için kullanılan önemli bir kriter de, deney ve tahmin sonuçları arasındaki korelasyon katsayısıdır (R^2). R^2 değeri 0 ile 1 arasında bir değer almaktadır. Bu değer 1'e yaklaşıyorsa model verilerle oldukça uyumludur (Özşahin 2012). Modellerin R^2 değerleri incelendiğinde, GBM'nin eğitim ve test veri setlerinde R^2 değerleri sırasıyla 0.99 ve 0.98, RF'nin eğitim ve test veri setlerinde R^2 değerleri sırasıyla 0.94 ve 0.89, LR'nin eğitim ve test

verilerinde R^2 değerleri sırasıyla 0.73 ve 0.55 ve YSA'nın eğitim ve test veri setlerinde R^2 değerleri ise sırasıyla 0.49 ve 0.32 çıkmıştır. Şekil 2'ye bakıldığında da GBM ile elde edilen tahmin değerlerinin gerçek değerlere daha yakın olduğu görülmektedir. GBM modeli ile gerçek değerlere en yakın değerler elde edilse de, RF ile elde edilen değerlerin de gerçek değerlere yakın olduğu gözlenmiştir. Özellikle RF sonuçlarının 2020 yılına kadar gerçek değerlere oldukça yakın olup, bu yıldan sonra yüksek standart sapmaların meydana gelmesi hisse senetlerinin dalgalı bir seyir izlemesinden kaynaklanabilir. Bu yüksek standart sapmalara ise 2019 yılında başlayan ve ülkemizde ve dünyada etkili olan Covid-19 pandemisi nedeniyle ekonomi ve ticarete yaşanan darboğazlar neden olabilir. Bununla beraber, 2020 yılından itibaren meydana gelen hisse senetlerindeki ani yükseliş ve azalışlara karşı GBM modelinin iyi performans sergilediği söylenebilir. LR ve YSA modelleri ile elde edilen sonuçlar ile ölçülen değerler arasında oldukça yüksek farkın olduğu görülmektedir. Yani LR ve YSA sonuçlarına göre GBM ve RF sonuçlarının ölçülen değerlerden daha düşük sapma değerlerine sahip olduğu görülmektedir. Çalışmada hisse senedi tahminde en iyi sonuç veren model GBM model olsa da, RF modelinin de hisse senedi tahmininde başarılı bir sonuç verdiğini söyleyebiliriz. YSA'nın R^2 değeri hem eğitim hem de test aşamasında 0.50'nin altında çıkmıştır.



Şekil 2. Gerçek ve tahmin değerleri

YSA tabanlı bir yöntemin başarılı bir şekilde uygulanması, çok sayıda eğitim veri seti gerektirmektedir (Qi ve Zhang 2001). GBM yönteminin diğer yöntemlere göre daha başarılı olduğunu gösteren yani bu çalışma ile paralellik gösteren çalışmalar bulunmakta olup, Reddy ve Kumar (2022), hisse senedi fiyat tahmininde GBM yönteminin %92.3 doğrulukla destek vektör makinesi algoritmasından önemli ölçüde daha iyi olduğunu saptamıştır. Verma ve Soni (2022) de araştırmasında, gradyan artırma yaklaşımının yüksek doğruluk derecesinde hisse senedi fiyatlarını tahmin ettiğini tespit etmiştir. Başka bir çalışmada gradyan artırılmış regresyon modelinin destek vektör makinesi, doğrusal regresyon, rastgele orman modellerinden daha başarılı performansa sahip model olarak öne çıkmıştır (Kesharwani ve ark. 2023). Yine, 2023 yılında Reddy ve Kumar tarafından yapılan başka bir çalışmada, GBM ve RF ile hisse senedi fiyat tahmini yapılmıştır. Çalışma sonucunda GBM algortimasının doğruluğu RF algoritmasına göre daha yüksek çıkmıştır. Hisse senedi fiyatı tahmininde rastgele orman yönteminin başarılı performans gösterdiğini sunan başka çalışmalar da (Tekin ve Çanakoğlu 2018, Ravikumar ve Saraf 2020, Roy ve ark. 2020, Kaynar ve Yiğit 2021) bulunmaktadır. Koç Ustalı ve ark. (2021) tarafından BIST 30 indekisinde işlem gören işletmelerin hisse senetlerinin tahmine

yönelik yapılan çalışmada, YSA'nın en kötü performansa sahip olduğu tespit edilmiştir. Aynı zamanda literatürde hisse senedi tahmininde YSA'nın başarılı bir performansa sahip olduğunu gösteren araştırmalarda mevcuttur (Yıldırım ve ark. 2011, Zahedi ve Rounaghi 2015, Sorayaei ve ark. 2016, Vijh ve ark. 2020, Ali ve ark. 2021). Ayrıca, bu çalışmanın aksine Abd Elminaam ve ark. (2022) hisse senedinin ertesi gün kapanış fiyatının tahmininde LR yönteminin RF, GB ve KNN yöntemlerinden daha başarılı performans gösterdiğini tespit etmiştir. Modellerin performans değerleri Çizelge 1'de verilmiştir.

Çizelge 1. Modellerin performanslarının karşılaştırılması

Modeller	Eğitim			Test		
	MSE	MAE	R ²	MSE	MAE	R ²
GBM	0.0044	0.055	0.99	4.853	1.32	0.98
RF	29.9	2.14	0.94	63.7	4.48	0.89
LR	133.24	9.21	0.73	249.61	13.16	0.55
YSA	250.05	8.85	0.49	382.5	14.91	0.32

GBM: Gradyan Güçlendirme Makineleri, RF: Rasgele Orman, LR: Doğrusal Regresyon, YSA: Yapay Sinir Ağları, MSE: Ortalama Karesel Hata, MAE: Ortalama Mutlak Hata, R²: Determinasyon Katsayısı

SONUÇ VE ÖNERİLER

Türkiye'de kâğıt ve kâğıt ürünleri sektörü, Türkiye'nin ihracatında önemli bir yere sahip olan tarım ürünlerinin taşınması başta olmak üzere ambalaj ürünlerinin bütün sektörlerle hitap etmesiyle beraber sektördeki teknolojik gelişmelerden dolayı da yatırım harcamalarındaki artışlar nedeni ile stratejik bir öneme sahiptir. Dahası, bu sektör ihracat değeri, firma sayısı ve istihdam sayısı ile önemli bir imalat endüstrisidir (Anonim 2022). Çalışmada hisse senetleri tahmini yapılan Kartonsan firması ise ilk “özel kuşeli karton üreticisi olup, piyasayı yerli kuşeli karton ile tanıştıran bir firmadır. Üretim kapasitesi açısından Türkiye'nin lider Avrupa'nın ise önde gelen kuşeli karton üreticilerinden biridir (URL-3 2024).

Bu çalışmada Borsa İstanbul'da işlem gören Kartonsan firmasına yatırım yapanlara veya yatırım yapmayı düşünenlere maksimum kar elde etmek için hisse senedi satın alma konusunda yardımcı olacak hisse senedi fiyat eğilimi makine öğrenme teknikleri ile analiz edilmesi amaçlanmıştır. Makine öğrenmesi yöntemlerinin karşılaştırılması ise MSE, MAE ve R^2 değerlendirme ölçütlerinden faydalanarak yapılmıştır. Sonuçlar incelendiğinde, makine öğrenmesi yöntemlerinden gradyan güçlendirme ve rastgele orman yöntemleri ile oluşturulan modellerin BIST'te işlem gören ve kâğıt sektöründe faaliyet gösteren Kartonsan firmasının hisse senetlerinin bir gün sonraki kapanış değerlerinin tahmininde başarılı olduğu görülmektedir. Gradyan güçlendirme ile rastgele orman yöntemleri karşılaştırıldığında gradyan güçlendirme yöntemi ile daha iyi sonuçlar elde edilmiştir. Yapay sinir ağlarının ise en kötü performansa sahip olduğu bulunmuştur. Yapay sinir ağları yönteminin kötü performans sergilemesi kullanılan veri sayısının az olmasından kaynaklanabilir. Her dört modelde eğitim kümelerinin değerlendirme ölçütlerinin değerleri test kümelerinin değerlerine göre daha düşük olduğu görülmektedir. Bu sonuç eğitim aşamasında kullanılan veri sayısının test aşamasında kullanılan veri sayısından çok daha fazla olmasından kaynaklanabilir. Çalışmada girdi olarak sadece firmaya ait dokuz finansal oran ve dört makine öğrenmesi yönteminin kullanılması da çalışmanın bir kısıtlayıcısıdır. Bu çalışmada sadece çeyrek dönemlik hisse senedi kapanış fiyatları ve incelenen işletmeye ait finansal oranlar seçilmiştir, ancak makroekonomik ve mali politikalar gibi hisse senedi ile ilgili diğer ekonomik faktörler dikkate alınmamıştır. Bu nedenle hisse senedi fiyat eğilimini yargılamak için hisse senedi fiyat dalgalanmasını etkileyen diğer faktörlerin bağımsız değişken olarak seçilmesi gelecekte üzerinde

çalışılacak konulardan biridir. Farklı makine öğrenmesi yöntemleri kullanılarak ve makine öğrenmesi yöntemleri birbiriyle entegre edilerek farklı çalışmalarda yapılabilir. Çalışmadaki veri sayısı ve girdi sayısı artırıldığında özellikle yapay sinir ağları ile daha başarılı sonuçlar elde edilebilir. Özellikle kâğıt ve kâğıt ürünleri sektörü ve sektördeki işletmelerin finansal getirilerinin makine öğrenme yöntemleri ile tahminine yönelik sınırlı sayıda çalışma bulunmakta olup, bu bakımdan mevcut çalışma literatüre katkı sağlayacaktır.

KAYNAKLAR

- Abd Elminaam DS, El Tanany A, Abdul Salam M, Abd El Fattah M (2022) Closing price prediction of stock market using machine learning models. 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, pp. 251-255.
- Acharya S, Pustokhina IV, Pustokhin DA, Geetha BT, Joshi GP, Nebhen J, Seo C (2021) An improved gradient boosting tree algorithm for financial risk management. Knowledge Management Research & Practice, 1-12.
- Ahmad M, Reynolds J, Rezgui Y (2018) Predictive modelling for solar thermal energy systems: a comparison of support vector regression, random forest, extra trees and regression trees. J Clean Prod, 203: 810–821.
- Akşehir ZD, Kılıç E (2019) Makine öğrenmesi teknikleri ile banka hisse senetlerinin fiyat tahmini. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 12(2): 30-39.
- Akyol Özcan K (2023) Borsa endeksi yönünün makine öğrenmesi yöntemleri ile tahmini: BIST 100 örneği. Gümüşhane Üniversitesi Sosyal Bilimler Dergisi, 14(3): 1001-1018.
- Ali M, Khan DM, Aamir M, Ali A (2021) Predicting the direction movement of financial time series using artificial neural network and support vector machine. Complexity, Special Issue, 1-13.
- Anonim (2022) Kağıt Sektör Raporu 2021. T.C. Sanayi ve Teknoloji Bakanlığı Sanayi Genel Müdürlüğü.
- Arslankaya S, Toprak Ş (2021) Makine öğrenmesi ve derin öğrenme algoritmalarını kullanarak hisse senedi fiyat tahmini. Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi, 13(1): 178-192.
- Attrapadung N, Hamada K, Ikarashi D, Kikuchi R, Matsuda T, Mishina I, Morita H, Schuldt JCN (2022) Adam in private: secure and fast training of deep neural networks with adaptive moment estimation. Proceedings on Privacy Enhancing Technologies, 4: 746-767.
- Bhardwaj N, Akil Ansari MD (2019) Prediction of stock market using machine learning algorithms. International Research Journal of Engineering and Technology, 6(5): 5994-6005.
- Biau G, Scornet EA (2016) Random forest guided tour. Test, 25: 197–227.
- Botchkarev A (2018) Performance metrics (error measures) in machine learning regression, forecasting and prognostics: properties and typology.
- Cavalcante RC, Brasileiro RC, Souza VLF, NobregaJP, Oliveira AI (2016) Computational intelligence and financial markets: a survey and future directions. Expert Systems with Applications, 55: 194-211.
- Cao M (2021) Predicting the link between stock prices and indices with machine learning in R programming language. Journal of Mathematics, 2021: 1275637.
- Ceylan T (2018) Perakende sektöründe makine öğrenmesine dayalı yaklaşımlar. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi, İstanbul.

- Correa-Jullian C, Cardemil JM, Droguett EL, Behzad M (2020) Assessment of deep learning techniques for prognosis of solar thermal systems. *Renewable Energy*, 145: 2178-2191.
- Devadoss AV, Ligor TAA (2013) Stock prediction using artificial neural networks. *International Journal of Data Mining Techniques and Applications*, 2: 283-291.
- Dike HU, Zhou Y, Deveerasetty KK, Wu Q (2018) Unsupervised learning based on artificial neural network: a review. 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS), Shenzhen, China, pp. 322-327.
- Doğan S, Büyükkör Y (2022) Makine öğrenmesi ile finansal zaman serisi tahminleme. *Ankara Hacı Bayram Veli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 24(3): 1205-1230.
- Ghani MU, Awais M, Muzammul M (2019) Stock market prediction using machine learning (ML) algorithms. *Advances in Distributed Computing and Artificial Intelligence Journal*, 8(4): 97-116.
- Hartanto A, Kholik YN, Pristyanto Y (2023) Stock price time series data forecasting using the light gradient boosting machine (LightGBM) model. *International Journal on Informatics Visualization*, 7(4): 2270-2279.
- He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 1026-1034.
- Hu Q, Qin S, Zhang S (2022) Comparison of stock price prediction based on different machine learning approaches. *Proceedings of the 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022)*, Wuhan, Hubei province, China, pp. 215-231.
- Hu J, Szymczak S (2023) A review on longitudinal data analysis with random forest. *Briefings in Bioinformatics*, 24(2): 1-11.
- Idowu PA, Osakwe C, Kayode AA, Adagunodo ER (2012) Prediction of stock market in Nigeria using artificial neural network. *I.J. Intelligent Systems and Applications*, 11: 68-74.
- Kaynar T, Yiğit ÖE (2021) Öznitelik mühendisliği ile makine öğrenmesi yöntemleri kullanılarak BIST 100 endeksi değişiminin tahminine yönelik bir yaklaşım. *Journal of Yaşar University* 16(64): 1741-1762.
- Kesharwani H, Mallick T, Nakarmi A, Raj G (2023) Evaluation of stock prices prediction using recent machine learning algorithms. In B. Unhelkar, H. M. Pandey, A. P. Agrawal, A. Choudhary (Eds.), *Advances and Applications of Artificial Intelligence & Machine Learning. Lecture Notes in Electrical Engineering*, Singapore, Springer, pp. 557-566.
- Kocaoglu D, Turgut K, Konyar MZ (2022) Sector-based stock price prediction with machine learning models. *Sakarya University Journal of Computer and Information Sciences*, 5(3): 415-426.
- Konstantinov AV, Utkin LV (2021) Interpretable machine learning with an ensemble of gradient boosting machines. *Knowledge-Based Systems*, 222: 106993.
- Koç Ustalı N, Tosun N, Tosun Ö (2021) Makine öğrenmesi teknikleri ile hisse senedi fiyat tahmini. *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, 16(1): 1-16.
- Leippold M, Wang Q, Zhou W (2022) Machine learning in the Chinese stock market. *Journal of Financial Economics*, 145: 64-82.
- Lu H, Mazumder R (2020) Randomized gradient boosting machine. *SIAM Journal on Optimization*, 30(4): 2780-2808.
- Manojlović T, Štajduhar I (2015) Predicting stock market trends using random forests: a sample of the Zagreb stock exchange. 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, pp. 1189-1193.
- Moghaddam AH, Moghaddam MH, Esfandyari M (2016) Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science*, 21: 89-93.
- Nakagawa K, Yoshida K (2022) Time-series gradient boosting tree for stock price prediction. *Data, Mining, Modelling and Management*, 14(2): 110-125.
- Özşahin Ş (2012) The use of an artificial neural network for modelling the moisture absorption and thickness swelling of oriented strand board. *BioResources*, 7(1): 1053-1067.
- Öztürkmen M, Eren E (2020) BIST 100 endeksi için yükseliş-düşüş tahmini için bir yapay öğrenme modeli. *Efil Journal*, 3(12): 42-52.
- Patel R, Choudhary V, Saxena D, Singh AK (2021) Review of stock prediction using machine learning techniques. *Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, TamilNadu, India, pp. 840-846.
- Qi M, Zhang G (2001) An investigation of model selection criteria for neural network time series forecasting. *European J. Oper. Res.*, 13 (3): 666-680.
- Ravikumar S, Saraf P (2020) Prediction of stock prices using machine learning (regression, classification) algorithms. 2020 International Conference for Emerging Technology (INCET), Belgaum, India, pp. 1-5.
- Reddy PV, Kumar SM (2022) A method for determining the accuracy of stock prices using gradient boosting and the support vector machines algorithm. *Proceedings of the Third International Conference on Smart Electronics and Communication (ICOS)* Trichy, India, pp. 1596-1599.
- Reddy PV, Kumar SM (2023) A novel approach to improve accuracy in stock price prediction using gradient boosting machines algorithm compared with random forest algorithm. *AIP Conference Proceedings*, 2822(1): 1-8.
- Roy SS, Chopra R, Lee KC, Spampinato C, Mohammadi-ivatlood B (2020) Random forest, gradient boosted machines and deep neural network for stock price forecasting: a comparative analysis on South Korean companies. *International Journal of Ad Hoc and Ubiquitous Computing*, 33(1): 62-71.
- Rubio A, Bermudez JD, Vercher E (2016) Forecasting portfolio returns using weighted fuzzy time series methods. *International Journal of Approximate Reasoning*, 75: 1-12.
- Sakhare NN, Imambi SS (2019) Performance analysis of regression based machine learning techniques for prediction of stock market movement. *International Journal of Recent Technology and Engineering*, 7(6S): 655-662.
- Sharma A, Bhuriya D, Singh U (2017) Survey of stock market prediction using machine learning approach. *International Conference on Electronics, Communication and Aerospace Technology* Coimbatore, India: 506-509.
- Shrivastav LK (2022) An ensemble of random forest gradient boosting machine and deep learning methods for stock price prediction. *Journal of Information Technology Research*, 15(1): 1-19.
- Singh G (2022) Machine learning models in stock market prediction. *International Journal of Innovative Technology and Exploring Engineering*, 11(3): 18-28.
- Siswoyo B (2021) Ensemble learning gradient boosting in improving classification and prediction in machine learning. *Turkish Journal of Computer and Mathematics Education*, 12(8): 1997-2002.
- Sorayaei A, Atf Z, Gholami M (2016) Prediction stock price using artificial neural network (case study: chemical industry firms accepted in Tehran stock exchange). *Bulletin de la Societe Royale des Sciences*, 85: 991-998.

- Srivastava R, Kumar S, Kumar B (2023) Classification model of machine learning for medical data analysis. In T. Goswami, G. R. Sinha (Eds.), Statistical Modeling in Machine Learning: Concepts and Applications, Cambridge, Massachusetts, ABD Academic Press.
- Su X, Yan X, Tsai CL (2012) Linear regression. Wiley Interdisciplinary Reviews: Computational Statistics, 4(3): 275-294
- Tekin S, Çanakoğlu E (2018) Borsa İstanbul hisse senedi getirilerinin makine öğrenmesi yöntemleriyle tahmini. 26th Signal Processing and Communications Applications Conference, İzmir, Türkiye, pp. 1-4.
- Toprak Ş, Çağıl G, Kökçam AH (2023) Stock closing price prediction with machine learning algorithms: PETKM stock example in BIST. Düzce University Journal of Science & Technology, 11: 958-976.
- Touzani S, Granderson J, Fernandes S (2018) Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy and Buildings, 158: 1533-1543.
- Umoh UA, Eyoh IJ, Murugesan VS, Nyoho EE (2022) Fuzzy-machine learning models for the prediction of fire outbreaks: a comparative analysis. In R. Pandey, S. K. Khatri, N. K. Singh, P. Verma (Eds.), Artificial Intelligence and Machine Learning for EDGE Computing, Cambridge, Massachusetts, ABD: Academic Press, pp. 207-233.
- URL-1. (2023). <https://tr.investing.com/equities/kartonsan>. Erişim tarihi: 15.10.2023
- URL-2. (2023). <https://www.kap.org.tr/tr/sirket-bilgileri/ozet/997-kartonsan-karton-sanayi-ve-ticaret-a-s>. Erişim tarihi: 15.10.2023
- URL-3. (2024). <https://www.kartonsan.com.tr/tr/#>. Erişim tarihi: 20.01.2024
- Verma A, Soni R (2022) Predicting stock market movement employing data filtration and gradient boost algorithm. In A. Abraham, P. Siarry, V. Piuri, N. Gandhi, G. Casalino, O. Castillo, P. Hung (Eds.), Hybrid Intelligent Systems. HIS 2021. Lecture Notes in Networks and Systems Cham, Springer, pp. 215-222.
- Vijh M, Chandola D, Tikkiwal VA, Kumar A (2020) Stock closing price prediction using machine learning techniques. Procedia Computer Science, 167: 599–606.
- Wang J, Gao RX (2022) Innovative smart scheduling and predictive maintenance techniques. In D. Mourtzis (Ed.), Design and Operation of Production Networks for Mass Personalization in the Era of Cloud Technology ABD: Elsevier, pp.181-207.
- Xu X (2022) Research on the stock price forecasting of netflix based on linear regression, decision tree, and gradient boosting models. Proceedings of the 2022 2nd International Conference on Business Administration and Data Science (BADs 2022) Kashgar, Xinjiang, China, pp. 1232-1242.
- Yıldırım İ, Özşahin Ş, Akyüz KC (2011) Prediction of the financial return of the paper sector with artificial neural networks. BioResources, 6(4): 4076-4091.
- Zahedi J, Rounaghi MM (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, 438: 177-187.