


Makalenin Türü : Araştırma Makalesi
Geliş Tarihi : 16.11.2023
Kabul Tarihi : 18.03.2024



 <https://doi.org/10.29029/busbed.1391790>


BORSA ENDEKS YÖNÜNÜN AĞAÇ TABANLI TOPLULUK MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNİ: BİST-100 ÖRNEĞİ


Yasin BÜYÜKKÖR¹, Seyyide DOĞAN²

ÖZ

Borsa endeks yönünün tahmininde finansal verilerin karmaşık ve durağan olmayan yapısı nedeniyle etkin bir tahmin modelinin kurulması oldukça zordur. Bazı dışsal faktörlerin ve şokların etkilerinin daha derin gözlemlendiği gelişmekte olan ülke borsalarında, borsa endeksinin aşağı veya yukarı yönlü hareketini tahmin etmek gerek yatırımcılar, hükümetler, finansal kurumlar ve kreditorler gibi paydaşlar gerekse de araştırmacılar için önemli bir konudur. Bu çalışmanın amacı, Borsa İstanbul 100 (BİST-100) endeksinin borsa endeksinin yönünü ağaç tabanlı topluluk Makine Öğrenmesi (ML) yöntemleriyle tahmin etmektir. Üç yılın günlük Açılış, Kapanış, En Yüksek, En Düşük ve Hacim verilerine Üstel Düzgünleştirme uygulandıktan sonra hesaplanan Teknik Göstergeler modelin girdi değişkenleri olarak ele alınmıştır. Ayrıca Teknik Göstergelerin pencere uzunlukları artırılarak girdi değişkeni uzayı genişletilmiştir. Çalışmada Karar Ağaçlarına dayanan topluluk makine öğrenmesi yöntemlerinden Random Forest, XGBoost ve CatBoost kullanılmıştır. Modelin parametreleri Bayesyan Arama (Bayesian Search) yöntemi ile optimize edilmiştir. Çalışmanın bulgularına göre, tercih edilen bütün yöntemler %89,7 ile %90,4 aralığında doğruluk oranına sahipken ve diğer performans değerlendirme kriterleri de dikkate alındığında en iyi performansa sahip yöntemin XGBoost olduğu görülmektedir.


Anahtar Kelimeler: Topluluk Makine Öğrenmesi, Teknik Göstergeler, Borsa Yönü Tahmini, Rassel Orman, XGBoost, CatBoost

¹ Dr. Öğr. Üyesi, Karamanoğlu Mehmetbey Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, yasinbuyukkor@kmu.edu.tr,  <https://orcid.org/0000-0002-1006-0539>

² Dr. Öğr. Üyesi, Karamanoğlu Mehmetbey Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, dogans@kmu.edu.tr,  <https://orcid.org/0000-0001-7835-7905>

Article Type : Research Article
Date Received : 16.11.2023
Date Accepted : 18.03.2024



 <https://doi.org/10.29029/busbed.1391790>


PREDICTION OF STOCK MARKET INDEX DIRECTION WITH TREE-BASED ENSEMBLE MACHINE LEARNING METHODS: THE CASE OF BIST-100


Yasin BÜYÜKKÖR¹, Seyyide DOĞAN²

ABSTRACT

The establishment of an effective prediction model for the direction of stock market indices is quite challenging due to the complex and non-stationary nature of financial data. Predicting the upward or downward movements of the stock market index, especially in emerging market exchanges where the impacts of external factors and shocks are observed more deeply, is of significant importance to stakeholders such as investors, governments, financial institutions, and creditors, as well as researchers. The aim of this study is to predict the direction of the stock market index with tree-based ensemble Machine Learning (ML) methods. Technical Indicators calculated after applying Exponential Smoothing to daily Opening, Closing, Highest, Lowest and Volume data of three years were considered as input variables of the model. In addition, the input variable space was expanded by increasing the window lengths of the Technical Indicators. In the study, Random Forest, XGBoost and CatBoost methods, which based on Decision Trees, are used as ensemble ML methods. Bayesian Search was employed to determine the optimal parameters. According to the findings of the study, all selected methods demonstrated accuracy rates ranging from 89.7% to 90.4%, and considering other performance evaluation criterias, XGBoost was identified as the best prediction method.

Keywords: Ensemble Machine Learning, Technical Indicators, Stock Market Way Prediction, Random Forest, XGBoost, CatBoost

¹ Asst. Prof., Karamanoglu Mehmetbey University, Faculty of Economics and Administrative Sciences, yasinbuyukkor@kmu.edu.tr, 
<https://orcid.org/0000-0002-1006-0539>

² Asst. Prof., Karamanoglu Mehmetbey University, Faculty of Economics and Administrative Sciences, dogans@kmu.edu.tr, 
<https://orcid.org/0000-0001-7835-7905>

1. GİRİŞ

Finansal veriler ekonomik, politik, sosyal, doğal afetler ve ikame finansal ürünlerin fiyatlarındaki gelişmeler gibi pek çok dışsal faktörden etkilenebilmektedir. Finansal zaman serileri doğası gereği gürültü (modelde yer alamayan bilgi) içerir, durağan (dağılımın zaman içinde değişkenliği) değildir ve deterministik olarak kaotiktir (kısa vadeli rassal; uzun vadeli öngörülemeyen deterministik davranış) (Kumar & Thenmozhi, 2006; Yu vd., 2009; Nava vd., 2016; Cao vd., 2019). Bu sebeple finansal zaman serisi kullanarak gerek gelecek eğilimi gerekse de fiyat tahmininde bulunmak çok karmaşık ve zorlu bir iştir. Hem araştırmacılar hem de uygulayıcılar açısından güçlü bir tahmin modeli geliştirebilmek piyasa dinamiklerini doğru analiz edebilmek için değerlidir (Leung vd., 2000).

Son yıllarda, finansal piyasalardaki hareketlerin yönüne veya eğilimine olan ilgi artmaktadır; ancak böylesi önemli bir alan için az sayıda çalışma bulunmaktadır (Jiao ve Jakubowicz, 2017). Rassal Orman (RF), AdaBoost ve Extrem Gradyan Artırma (eXtreme Gradient Boosting- XGBoost) gibi topluluk (ensemble) öğrenme yöntemleri ile hisse senedinin gelecek eğilime odaklanmış finansal zaman serisi tahmin çalışmalarının önemli uygulamaları Kumar ve Thenmozhi (2006), Ballings vd. (2015) ve Patel vd. (2015) gibi araştırmacılar tarafından gösterilmiştir.

Literatürde gelişmekte olan piyasalarda borsa endeks hareketinin yönünü tahmin etmeye yönelik çok az sayıda araştırma bulunması sebebiyle bu çalışmada BİST-100 endeksi tercih edilmiştir. Bununla birlikte, Covid 19 küresel pandemisi, Şubat ayı içerisinde gerçekleşen deprem, hükümet seçimleri, dövizdeki hızlı ve önlenemez yükseliş gibi borsa üzerinde önemli etkilere sahip bazı dışsal faktörlerin ve şokların etkilerini gözlediğimiz bu endeks ile çalışmak, ilgili literatür açısından önemli olacaktır. BİST-100 endeksinin son zamanlardaki kırılmalı ve makroekonomik göstergelerin endeks üzerindeki etkileri bazı ekonometrik çalışmalarla da desteklenmiştir (Ünlü, 2023).

Bu çalışmada, RF, XGBoost ve gradyan artırma algoritmalarının son versiyonu olan CatBoost olmak üzere ağaç tabanlı topluluk makine öğrenmesi (ML) yöntemlerini kullanarak BİST-100 Endeksinin yönü tahmin edilmekte ve tercih edilen ML yöntemlerinin performansı karşılaştırılmaktadır. BİST-100 endeksindeki değişimlerinin büyüklüğü göz önünde bulundurulduğunda, günlük fiyat hareketlerinin yönünün iyi bir tahmini akademik çevre için son derece önemli olmasının yanı sıra uygulamacılar açısından risk yönetimi için de önemlidir. Literatürde izini sürdüğümüz çalışmaların sonuçlarından hareketle, tahmin performansında güçlü bir iyileşmenin gözlemlendiği tüm güncel yaklaşımlar bu çalışmada uygulanmış ve BİST 100 endeksinin yön tahmininin bir uygulaması gösterilmiştir. Dolayısıyla bu çalışmanın finansal zaman serisi tahmin çalışmalarına katkı sunacağı düşünülmektedir. Bu bağlamda bu çalışmada, makalenin geri kalan kısmı şöyle düzenlenmiştir: Takip eden bölümde, finansal zaman serileri kullanarak yapılan hisse senedi yön tahmini ile ilgili çeşitli ML yöntemlerinin literatür taraması verilmektedir. Tercih edilen ML yöntemlerinin teorik alt yapısı 3. bölümde tanıtılmaktadır. Deney tasarımı ve BİST-100 endeksinin açılış, kapanış, hacim vb. gibi özelliklerini ele aldığımız veri setinden edilen ampirik sonuçlar 4. bölümde rapor edilmektedir. Son olarak, 5. bölümde sonuçlar ve tartışma yer almaktadır.

2. Literatür Taraması

Hisse senedi fiyat ve trend tahmini, risk ve getiri analizleri ve portföy yönetimi borsa yatırımcılarının karşılaştığı önemli problemlerdendir. Hisse senedi piyasasında yatırımın getirilerinden faydalanma potansiyeli, piyasanın doğrusal olmayan kaotik yapısı nedeniyle zorluklar içermekte ve bu durum piyasanın değerlendirilmesine olan talebi artırmaktadır (Thakkar & Chaudhari, 2021).

Hisse senedi trend tahminlerinde, piyasa bilgisine dayanarak ve baskın faktörler incelenerek temel analiz; açılış fiyatı, kapanış fiyatı ve işlem hacmi olmak üzere farklı zaman serilerinden elde edilen bilgiler ise teknik analiz için kullanılabilir (Yu vd., 2009; Kumar vd., 2016; Yang vd., 2016) veya geleneksel olmayan uzmanlar olarak görülen Google, Twitter vs. gibi sosyal medya platformlarına ait verilerinin de dahil edildiği farklı stratejiler geliştirilebilir (Cavalcante vd., 2016; Chen vd., 2016; Hu vd., 2018).

Hisse senedinin gelecekteki fiyatı ve/veya eğilimi tahmin eden çalışmalar görece son yıllarda hız kazanmıştır. Hisse senedi getirilerini tahmin etmede açıklayıcı değişken olarak teknik göstergelerin kullanımı gibi hesaplamalı yaklaşımlar makine öğrenimi içerisinde son yıllarda önemli bir yer edinmiştir ve ekonomistlerin kullandıkları geleneksel yöntemlere göre daha iyi bir tahmin performansına sahip oldukları ileri sürülmüştür (Hsu vd., 2016). Bu yöntemler içerisinde ilk uygulamalardan bu yana oldukça sık tercih edilmiş bir yöntem Yapay Sinir Ağları (YSA)'dır (Tsaih, 1998; Dixit vd., 2013; Jhaveri, 2016; Weng, 2017). Çok sayıda başarılı uygulamasına karşılık, YSA'nın tutarsız sonuçlar sergilediği ve bazı kısıtlamalara sahip olduğu yönünde eleştirilerde bulunan yazarlar da bulunmaktadır (Hann vd., 1996). Bu noktada ana yönelim sıklıkla

performansının kıyaslandığı ve başarılı sonuçların elde edildiği Destek Vektör Makine (SVM)'ne olmuştur (Tay & Cao, 2001; Huang vd., 2005; Yu vd. 2005; Kara vd., 2011; Dai & Zhang, 2013; Pabuçcu, 2019).

Belirli sayıda karar ağacının eğitilmesinin esas alındığı topluluk ML yöntemlerinden RF'nin tek bir model ile tahmin yapan SVM'den daha iyi sonuçlar verdiği Patel vd. (2015) tarafından gösterilmiştir. Bu öncü çalışmanın ardından topluluk ML yöntemlerinin tekli ML yöntemleri ile kıyaslandığı ve daha iyi performans gösterdiği sonucu Ballings vd. (2015) ve Mehta vd. (2019) tarafından desteklenmiştir.

Jiao ve Jakubowicz (2017), topluluk ML yöntemlerinden RF, gradyan destekli ağaçlar ve tekli ML yöntemlerinden YSA ve Lojistik Regresyon Analizi (LRA)'nin S&P 500 endeksinin trend tahmin performansını karşılaştırmalı olarak araştırmışlardır. Bu çalışmanın sonuçları yukarıda bahsi geçen çalışmaların bulguları ile ters yöndedir. L2 düzleştirme ile LRA diğer yöntemlerden az bir farkla da olsa daha iyi tahminde bulunmuştur. Bununla birlikte çalışmanın ilgi çekici bir diğer sonucu, doğrulama setinin sıralı olarak yapıldığında çapraz doğrulama ve tekli doğrulamaya göre daha iyi tahmin sonucu elde edilmiş olmasıdır.

Hisse senedi fiyat tahmin literatüründe XGBoost algoritmasının tercih edildiği eser sayıda çalışma bulunmaktadır. Dey vd. (2016)'nin doğrusal olmadığını gösterdikleri APPL hisse senedi üzerinde 60 günlük ve 90 günlük tahminlerinde doğruluk oranı sırasıyla %87,79 ve %89,70 olarak elde etmişlerdir. Bu çalışmada uzun vadeli tahminlerde üstel düzleştirme ve doğrusal olmayan verilerde karar ağaçları tabanlı metrik olmayan yöntemlerin tercih edilmesinin önemi vurgulanmıştır. Ancak bu iddia, diğer yöntemler ile kıyaslamalı olarak desteklenmemiştir.

En güncel ve en kapsamlı olarak, ağaç tabanlı topluluk ML yöntemlerinin başarılı bir uygulaması Başak vd. (2019) tarafından gösterilmiştir. Bu çalışmada, yazılım (FB, TWTR), elektronik (AMS) ve otomobil (TATA) şirketlerine ait hisse sentlerinin eğilimlerini tahmin etmek için XGBoost ve RF algoritması tercih edilmiştir. Çalışmanın ilgi çekici tarafı kazançları ve kayıpları gösteren hisse senedinin n gün önceki fiyatına göre yukarı ve aşağı yönlü hareketinin bir regresyon modeli sonucu elde edilen eşik değeri ile belirlenmesidir. Geçmiş çalışmalar ile kıyaslandığında bu alan için iyi performanslardan birinin elde edildiği bu çalışmada doğruluk oranı %78 olarak elde edilmiş ve bu iyileşmenin önemli bir nedenin topluluk ML yöntemlerin tercih edilmiş olmasına bağlanmıştır.

Topluluk ML yöntemlerini en kapsamlı olarak araştıran çalışma Prasad ve Bakhshi (2022) tarafından yürütülmüştür. Bu çalışmada Hindistan VIX 'nin günlük hareketlerinin tahmini LRA ile beraber 11 ML yöntemi kullanılarak gerçekleştirilmiştir. Tahminler %63,33 ile %67,67 arasında değişen doğruluk oranları ile elde edilirken, CatBoost, Light Gradient Boosted Machine (LightGBM), XGBoost, oylama, istifleme (Stacking), torbalama (Bagging) ve RF sınıflandırıcıları, istatistiksel olarak benzer performanslara sahip en iyi yöntemler olduğu sonucuna varılmıştır.

3. YÖNTEM

Karar ağaçlarına dayanan ML yöntemleri yüksek doğruluk oranlarına sahip, kolay uygulanabilir ve modellerin nispeten yorumlanabilir olması nedeniyle sınıflama problemlerinde sıklıkla kullanılmaktadır. ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993) ve CART (Breiman vd., 1984) farklı amaçlar için geliştirilmiş olan karar ağacı yöntemleridir. Bu başlık altında karar ağaçlarına dayanan ML yöntemlerinden RF, XGBoost ve CatBoost yöntemlerinden bahsedilecektir.

3.1. Rassal Orman (Random Forest - RF)

RF (Breiman, 2001), karar ağaçlarını temel alan bir topluluk (ensemble) ML yöntemidir. Hem sınıflama (classification) hem de regresyon problemlerinde uygulama kolaylığı sağlaması ve yüksek performans göstermesi nedeniyle araştırmacı ve uygulamacılar arasında popüler bir yöntemdir. Karar ağaçlarının temel eğitim prensibi, öznitelik uzayının bir ağaç yapısı kullanarak sürekli olarak bölünmesidir. Bu yaklaşımda, her bir alt (child) düğüm saf (pure) düğümlere kadar bölünür ve böylece tek bir sınıfın örneklerini içeren düğümlere ulaşılan kadar bölünme devam ettirilir. Bu bölünme alt düğümlerin saflığını, ana (parent) düğümlerine göre maksimize etmeye çalışan bir kriter aracılığıyla yapılır. Alt düğümlerde maksimum saflık sağlandıktan sonra saf düğümlere ulaşılır. Elde edilen saf düğümler daha fazla bölünemez ve yaprak düğümlerini (leaf nodes) oluşturur. Sınıflama problemlerinde karar ağaçları kullanıldığında, ağacın yapısı yaprak düğümlerine kadar takip edilir ve yaprak düğümlerinin saf oldukları karar verilirse, test veri setinin sınıf etiketleri, eğitim örneklerinin yaprak düğümüne ait sınıf etiketleri baz alınarak atanır. Bu yöntemde bir test örneğini sınıflandırmak için büyük miktarda birbirleriyle korelasyonsuz karar ağacı bootstrap (yeniden örnekleme) yöntemi kullanarak büyütülür. Böylece test örneğinin hangi sınıfa ait olduğu özellik uzayının titizlikle incelenmesiyle karar verilir. Random Forest yöntemi aşırı uyumun (over-fitting) etkilerini azaltmak amacıyla birden fazla sayıda (ensemble) karar

ağacı kullanır ve çoğunluk oylamasına (majority voting) dayanarak nihai tahmin sonucunu elde eder (Basak vd., 2019).

Her düğümdeki bölünme performansını (saflığı) ölçmek amacıyla genellikle Gini Katsayısı kullanılır. N düğümündeki Gini Katsayısı Eşitlik 1’de gösterildiği gibi hesaplanır.

$$G(N) = 1 - (P_1)^2 - (P_{-1})^2 \quad (1)$$

Eşitlik 1’de P_i , i sınıf etiketine sahip popülasyon oranıdır. Bir düğümdeki en iyi bölünme kararı en düşük Gini Katsayısına sahip olan bölünme olarak seçilir. Başka bir ifadeyle, en iyi bölünme en yüksek bilgi kazancına (information gain) sahip olan bölünmedir (Basak vd., 2019)

RF, parametrik olmayan bir ML yöntemidir. Bu yöntemde öğrenme parametresi ve önsel (prior) dağılım varsayımı yoktur (Quinlan vd., 1986). Bu nedenle RF birçok araştırmacı için popüler bir yöntem olarak kullanılmaktadır.

3.2. Ekstrem Gradyan Artırma (eXtreme Gradient Boosting- XGBoost)

Chen ve Guestrin (2016) tarafından geliştirilen XGBoost bir topluluk ML yöntemidir. XGBoost’un en önemli özelliği daha az hafıza (memory) gerekliliği ve daha fazla hesaplama kapasitesine sahip olmasıdır. XGBoost algoritmasına ait kayıp fonksiyonu;

$$L(\phi) = \sum_i \ell(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (2)$$

ile gösterilir. Eşitlik 2’de birinci terim türevlenebilir ve konveks kayıp fonksiyonu iken ikinci terim düzeltme terimi olarak bilinir ve Eşitlik 3’te olduğu gibi gösterilir.

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (3)$$

Eşitlik 3’te karar ağacı içerisindeki yaprak sayısı T ile gösterilir. Ayrıca γ karmaşıklık parametresidir ve yaprak düğümlerinin sayısını gösterir. Buna ek olarak Eşitlik 3’ün ikinci teriminde w , yaprak düğümlerinin ağırlıklarını belirlerken λ aşırı öğrenmeyi önlemek için kullanılan parametredir. Buradan anlaşılacağı gibi XGBoost aşırı öğrenmeyi engellemek amacıyla bir düzenleme (regularization) terimine sahiptir.

XGBoost, veri setinde eksik veya hatalı gözlem olduğunda bu gözlemleri interpolasyon yöntemiyle hesaplayıp veri setine atar (sparsity-aware) ve veri setini tam ve eksiksiz olarak kullanılabilmesine olanak sağlar. Sınıflama problemlerinde ML yöntemlerinin büyük bir bölümü eşit ağırlıklı verilerde daha iyi çalışırken XGBoost dengesiz (unbalanced) veri setleri için de uygun bir ML yöntemidir.

3.3. Categorical Boosting (CatBoost)

CatBoost algoritması birden fazla sayıda karar ağaçlarını birleştiren topluluk (ensemble) öğrenme algoritmalarına gradyan artırma (boosting) uygulayan yeni bir ML algoritmadır. Dorogush vd. (2018) ve Prokhorenkova vd. (2018) tarafından önerilen bu algoritma, XGBoost (Chen vd., 2016) ve LigthGBM (Ke vd., 2017) gibi diğer gradyan artırma algoritmalarından farklı olarak kategorik özelliklerin daha iyi işlenmesini sağlayan sıralı güçlendirme algoritması kullanmakta ve aşırı uyumu önlemek için simetrik karar ağaçları oluşturmaktadır.

Modelin potansiyel doğruluğunu artırabilmek amacıyla gradyan artırmanın uygulanmasında karşılaşılan istatistiksel sorun “tahmin kayması”dır. Tahmin kayması, tüm eğitim örnekleri kullanıldığında (tüm eğitim etiketlerini öğrenmiş olduğundan), artırma adımları sırasında sapmalı model oluşmasına yol açar. Bu sapma tahmin modeliyle test örneklerinin dağılımında farklılık oluşmasına neden olur.

CatBoost algoritması bu sorunla başa çıkmak için sıralı artırmayı uygular. Kategorik özellikleri minimum bilgi kaybı ile işlemek için hedef istatistiği (Target Statistics-TS) değerine göre gruplama yapar. Kategorik özellikler için sıralı artırma işlemi veri kümesinde rastgele bir permütasyon gerçekleştirilmesi ve aynı kategori değerine sahip örnekler için ortalama bir etiket değerinin hesaplanması adımlarını takip eder (Prokhorenkova vd., 2018).

CatBoost algoritmasının kullandığı karar ağacı fonksiyonu ise Dorogush vd. (2018) aktardığı şekliyle;

$$H(x_i) = \sum_{j=1}^J c_j 1_{\{x \in R_j\}} \quad (4)$$

Eşitlik 4'te $H(x_i)$ açıklayıcı değişken x_i 'nin karar fonksiyonu ve R_j ağaç yapraklarına karşılık gelen ayrık bölgeyi tanımlamaktadır.

4. UYGULAMA

4.1. Veri ve Veri Ön Süreçler

Çalışmada kullanılan yöntemlerin performansını değerlendirmek amacıyla Borsa İstanbul 100 Endeksine (BİST-100) ait veriler günlük zaman serisi olarak kullanılmış ve veriler Yahoo Finance web sitesinden derlenmiştir. Veri setinin tamamı 27.07.2020'den 04.07.2023'e kadar günlük Açılış, Kapanış, En Yüksek, En Düşük, Düzeltilmiş Kapanış ve Hacim verilerini kapsamaktadır. Belirlenen tarihler arasında toplam 731 gözlem bulunmaktadır. Bu gözlemler kullanılarak hesaplanan teknik göstergelerden sonra kalan 680 gözlemden 544 gözlem (%80) eğitim ve 136 gözlem (%20) test veri seti olarak iki gruba ayrılmıştır. Veri setine ait özet değerler Tablo 1'de verilmiştir.

Tablo 1.

Veri Seti Artış ve Azalış Gün Sayıları

Veri Seti	Artış	Azalış	Artış/Azalış	Toplam
Eğitim	376	168	2,23	544
Test	94	42	2,23	136
Toplam	470	210	2,23	680

Tablo 1 incelendiğinde veri seti içerisinde artış ve azalışların oranlarının birbirlerinden önemli derecede farklı oldukları (unbalanced) görülmektedir. Bu farklılık ML öğrenmesi yöntemlerinin sayısı daha fazla olan sınıf lehine etkileyeceğinden (Aras, 2020) analizler aşamasında eğitim ve test veri setleri ayrılırken orijinal veri setindeki oranlar Tabakalı Grup Karışımı (Stratified Group Shuffle) kullanılarak korunmaya çalışılmıştır. Böylece Eğitim veri seti için 376 artış yönü ve 168 azalış yönü elde edilirken Test veri seti için 94 artış yönü ve 42 azalış yönü elde edilmiştir.

Çalışmada analizler Phyton 3.9 programı ile yapılmış ve RF yöntemi için *scikit-learn* paketi, XGBoost yöntemi için *xgboost* paketi ve CatBoost yöntemi için *catboost* paketi kullanılmıştır.

4.2. Üstel Düzgünleştirme (Exponential Smoothing)

Bir zaman serisi analizinde Üstel Düzgünleştirme, son gözlemlere daha büyük ve geçmiş gözlemlere daha küçük ağırlıklar vererek zaman serisini rassal değişkenlikten veya geçmiş verilerde var olan gürültüden arındırmaya yarayan bir istatistiksel yöntemdir (Basak vd., 2019; Khaidem vd., 2016). Üstel Düzgünleştirme yöntemi kullanılarak kurulacak modelin borsa endeks davranışındaki uzun dönemli değer eğilimi kolayca belirlenebilir. Bir Y zaman serisinin Üstel Düzgünleştirilmiş istatistiği ardışık olarak Eşitlik 5'teki gibi hesaplanır.

$$S_0 = Y_0$$

$$t > 0 \text{ için } S_t = \alpha Y_t + (1 - \alpha)S_{t-1} \quad (5)$$

Eşitlik 5'te α , düzgünleştirme faktörüdür ve 0 ile 1 arasında değer alır ($0 < \alpha < 1$). Eğer α değeri büyük ise düzgünleştirme düzeyi azalır ve $\alpha = 1$ olduğunda düzgünleştirilmiş istatistik gerçek gözleme eşit olur. Düzgünleştirilmiş istatistik S_t ancak ardışık gözlemler var olduğunda hesaplanabilir.

4.3. Özellik Çıkarma

4.3.1. Teknik Göstergeler

Endeks yönü tahminiyle ilgili geçmiş çalışmalar dikkate alındığında, gelecek dönem için yapılacak tahminlerde teknik göstergelerin kullanılmasının daha iyi tahmin sonuçları verdikleri görülmüştür (Saad vd., 1998). Finansal sektörde çalışan birçok yönetici ve yatırımcı teknik göstergelerin etkin birer araç olduklarını ve geleceği tahmin etmek için bir sinyal olarak kullanılabilecekleri konusunda hemfikirdirler. Bu çalışmada girdi değişkenleri olarak Tablo 2'de verilen teknik göstergeler kullanılmıştır (Kara vd., 2011; Patel vd., 2015).

Tablo 2.

Teknik Göstergeler ve Formülleri

İndikatör Adı	Formülü
Hareketli Ortalama (SMA)	$\frac{C_t + C_{t-1} + \dots + C_{t-n}}{n}$
Ağırlıklı Hareketli Ortalama (WMA)	$\frac{nC_t + (n-1)C_{t-1} + \dots + C_{t-n}}{n + (n-1) + \dots + 1}$

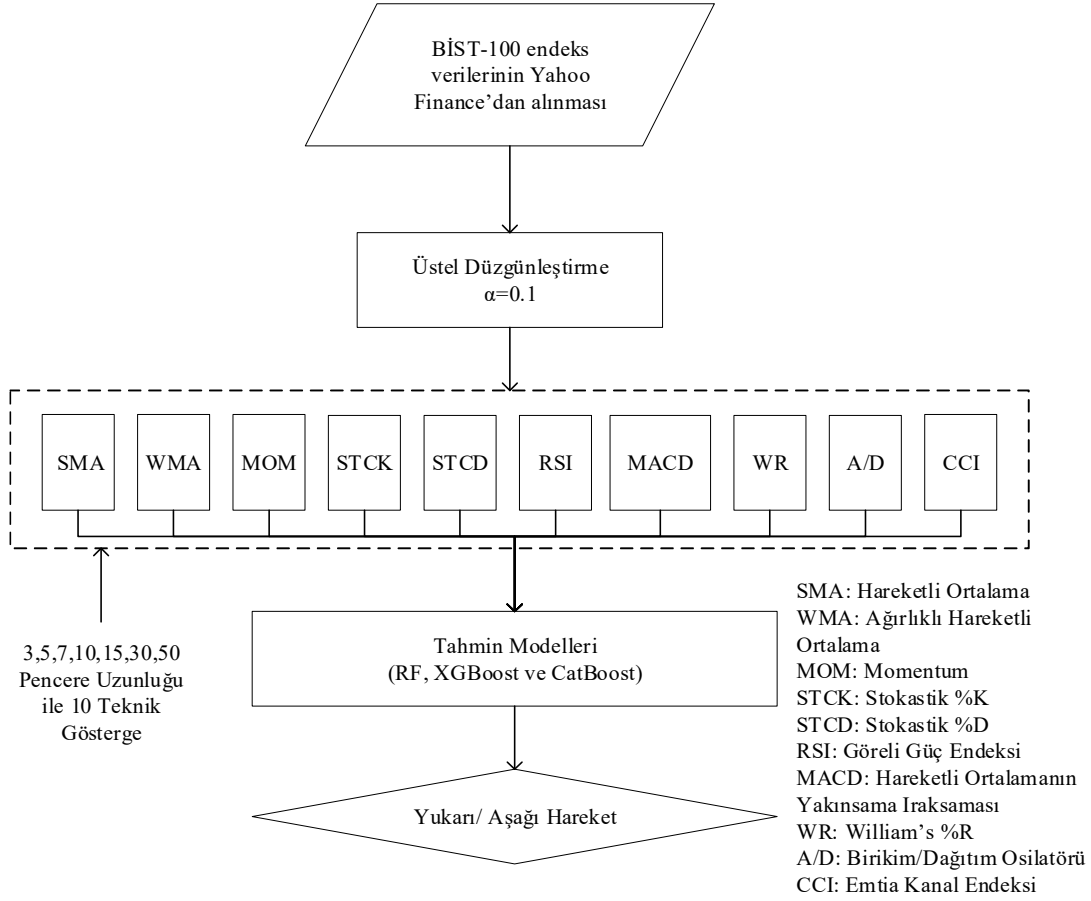
Momentum (M)	$C_t - C_{t-n}$
Stokastik %K	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100$
Stokastik %D	$\frac{\sum_{i=0}^{n-1} \%K_{t-i}}{n}$
Görelü Güç Endeksi (RSI)	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} Up_{t-i}/n) / (\sum_{i=0}^{n-1} Dw_{t-i}/n)}$
Hareketli Ortalamanın Yakınsama İraksaması (MACD)	$MACD(n)_{t-1} + 2/(n+1)(DIFF_t - MACD(n)_{t-1})$
Williams's %R	$\frac{H_t - C_t}{H_t - L_t} \times 100$
Birikim/Dağıtım Osilatörü (A/D)	$\frac{H_t - C_{t-1}}{H_t - L_t} \times 100$
Emtia Kanal Endeksi (CCI)	$\frac{M_t - SM_t}{0.015D_t} \times 100$

C_t : Kapanış Fiyatı, L_t : t zamanındaki En Düşük Fiyat, H_t : t zamanındaki En Yüksek Fiyat, $DIFF = EMA(12)_t - EMA(26)_t$ ve EMA Üstel Hareketli Ortalama, $EMA(k)_t = EMA(k)_{t-1} + \alpha(C_t - EMA(k)_{t-1})$ ve α düzleştirme sabiti ve $2/(k+1)$, k zaman periyodu için hesaplanan EMA , LL_t ve HH_t sırasıyla geçmiş t günde en küçüklerin en küçüğü ve en büyüklerin en büyüğü, $M_t = (H_t + L_t + C_t)/n$, $SM_t = (\sum_{i=1}^n M_{t-i+1})/n$, $D_t = (\sum_{i=1}^n |M_{t-i+1} - SM_t|)/n$, Up_t , t zamanındaki yukarı yönlü fiyat değişikliği ve Dw_t , t zamanındaki aşağı yönlü fiyat değişikliğini göstermektedir.

Teknik Göstergeler hesaplanırken pencere uzunlukları (window length) 3,5,7,10,15,30 ve 50 olarak alınmıştır. Böylece toplam girdi değişkeni sayısı 58 olarak belirlenmiştir. Pencere uzunluklarının belirlenmesinde hem önceki çalışmalar (Basak vd., 2019; Shynkevich vd., 2017) hem de analiz sonuçlarında doğruluk oranını artıran uzunluklar tercih edilmiştir. Bu çalışmada, daha sonra özellik matrisi olarak kullanılacak olan teknik göstergeler, veri seti düzleştirildikten sonra hesaplanmıştır. Ayrıca çalışmanın ana amacı olan borsanın yön tahmininde kullanılan hedef değer (target) Eşitlik 2'deki verildiği şekilde hesaplanmıştır.

$$hedef_i = \text{sign}((\text{Borsa Kapanış Endeksi})_{i+d} - (\text{Borsa Kapanış Endeksi})_i) \quad (6)$$

Eşitlik 6'da i , tahmin edilecek hedef değer gününü ve d değeri tahminin yapılacağı gün sayısını göstermektedir. Hedef değişkenin +1 değerini alması d gün sonra endekste yukarı bir hareket olduğu anlamına gelirken; -1 değerini alması d gün sonra endekste aşağı bir hareket olduğu anlamına gelir. Böylece hedef değişken, i . günde endeksin hangi yönde hareket ettiği hakkında bilgi verir (Basak vd., 2019). Şekil 1'de uygulamaya ait akış şeması verilmiştir.



Şekil 1. Uygulama akış şeması

4.3.2. Parametre Optimizasyonu

ML yöntemlerinde en uygun (optimal) parametrelerin belirlenmesi (hyper—parameter tuning) elde edilecek sonuçların güvenilirliğini ve performansını artıracığından önemli bir süreçtir. Bu çalışmada en uygun parametre setleri belirlenirken önceki çalışmalar ve verilere uygun olabilecek parametre setleri kullanılmıştır. Tablo 3'te kullanılan ML yöntemlerinin parametre setlerine ait aralıklar ve optimal parametreler verilmiştir. Optimal parametreler belirlenirken Bayesyan Arama (Bayesian Search) yöntemi kullanılmıştır (Xia vd., 2017).

Tablo 3.
Hyper-Parametre Aralıkları ve Optimal Parametreler

Yöntem	Parametreler	Parametre Aralığı	Optimal Parametre
XGBoost	Learning Rate	0.0001, 0.001, 0.01, 0.1, 0.2, 0.3	0.01
	Max Depth	3,5,6,10,50,100,1000	10
	Min Child Weight	1,2,5,15,50,100	2
	N Estimator	5,10,50,100,500,1000	1000
	Subsample	0.1,0.3,0.5,0.7,0.9	0.3
RF	N Estimator	200,300,..., 2000	500
	Max Features	Auto, Sqrt	Sqrt
	Max Depth	10,20,...,110	90
	Min Sample Split	2,5,10	5
	Min Samples Leaf	1,2,4,5,10	1
	Bootstrap	True, False	True
CatBoost	Depth	5,6,7,8,9,10	7
	Learning Rate	0.01,0.1,0.5,1	0.01
	L2 Leaf Reg	1,2,3,4,5,6,7,8,9,10,20,30,50,100,200	20

Optimal parametreler belirlendikten sonra aynı veri seti üzerinde her üç ML yöntemi ile analizler yapılmış ve yöntemlerin performansları karşılaştırılmıştır.

4.3.3. Performans Metrikleri

Sınıflama problemlerinde ML yöntemlerinin performanslarını değerlendirmek amacıyla en sık kullanılan ölçüm aracı karmaşıklık matrisidir (confusion matrix). Bu matris, gerçek değerler ile tahminlenmiş değerlerin yer aldığı bir tablodur. Tablo 4'te karmaşıklık matrisi verilmiştir.

Tablo 4.
Karmaşıklık Matrisi

		Gerçek	
		(+)	(-)
Tahmin	(+)	Doğru Pozitif (TP)	Yanlış Pozitif (FP)
	(-)	Yanlış Negatif (FN)	Doğru Negatif (TN)

Tablo 4 ele alınırsa Doğruluk oranı (Accuracy), Hassasiyet (Sensitivity-Recall), Specificity (Özgünlük), Kesinlik (Precision), Matthews Correlation Coefficient (MCC) ve F₁- Scoru Tablo 5'te verilmiştir.

Tablo 5.
Performans Metrikleri

Doğruluk Oranı (Accuracy)	$\frac{TP + TN}{TP + FP + FN + TN}$
Hassasiyet (Sensitivity- Recall)	$\frac{TP}{TP + FN}$
Özgünlük (Specificity- Selectivity)	$\frac{TN}{TN + FP}$
Kesinlik (Precision)	$\frac{TP}{TP + FP}$
Matthews Correlation Coefficient (MCC)	$\frac{TP + FN}{(TP * FN) - (FP * FN)}$
F1- Score	$\frac{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}{2TP}$

4. UYGULAMA SONUÇLARI

Bu çalışmada borsa endeksi yön tahmini ikili (binary) sınıflama problemi olarak ele alınmıştır. BİST-100 endeksi için yapılan yön tahmini için tercih edilen ağaç tabanlı topluluk ML yöntemlerinden RF, XGBoost ve nispeten yeni tanıtilen CatBoost'un tahmin sonuçları Tablo 6'da yer almaktadır.

Genel olarak tercih edilen tüm ML sınıflandırıcılar doğruluk oranları dikkate alındığında oldukça iyi sonuçlar vermişlerdir. Buna ilave olarak recall, specificity, precision, MCC ve F1-Score gibi diğer değerlendirme kriterlerinde de dengesiz bir veri seti ile çalışılmasına rağmen iyi sonuçlar elde edilmesi yöntemlerin sağlamlığı (robust) hakkında önemli kanıtlar sunmaktadır. Uygulanan ML yöntemlerinden en yüksek doğruluk oranına sahip yöntemin %90,4 ile XGBoost olduğu görülmektedir. Her üç yöntemin doğruluk oranları incelenirse her ne kadar XGBoost yönteminin doğruluk oranı yüksek olsa bile RF (%89,7) ve CatBoost (%89,7) yöntemlerinin doğruluk oranları birbirleri aynı ve XGBoost yöntemine oldukça yakındır.

Tablo 6.
Analiz Sonuçları

Yöntem	Accuracy	Recall	Specificity	Precision	MCC	F1- Score
RF	0,897	0,926	0,833	0,926	0,759	0,926
XGBoost	0,904	0,926	0,857	0,935	0,778	0,930
CatBoost	0,897	0,915	0,857	0,935	0,762	0,925

Bu yöntemlerin zorlu bir problem olan hisse yön tahmininin doğrusal olmayan, karmaşık ve kaotik yapısını yakalama konusundaki becerileri dikkate değerdir. SVM, ANN ve Karar Ağacı (DT) gibi tekli ML yöntemleri ile kıyaslandığında topluluk öğrenme yöntemlerinin bu alana önemli bir katkısı olduğu görülmektedir. Öncü çalışmalardan biri olan Kara vd. (2011), BİST veri seti üzerinde verilerin daha dengeli bir dağılımı elde etmek için verileri yeniden örnekleme ile ANN doğruluk oranını %65 ve SVM doğruluk oranını %84 olarak elde etmiştir. Patel vd. (2015), bir önceki çalışmada geçen sürekli verilerden oluşan teknik göstergelerin trendlerini de modele ekleyerek sonuçları daha iyileştirilmiştir. Söz konusunu bu çalışmanın tahmin doğruluk oranları: %86,69 ile ANN, %89,33 ile SVM, %89,98 ile RF ve %90,19 Naive Bayes (NB)'dir.

Mevcut literatür incelendiğinde iyi olarak kabul edilebilecek %90 ve üzeri doğruluk oranlarının elde edilmesi pek mümkün olmamıştır. Literatürde tahmin doğruluğunu arttırmak için önemli yaklaşımlar bulunmaktadır. Bu yaklaşımlardan bazılarını:

- i) Farklı ML ve derin öğrenme yöntemlerinin denenmesi,
- ii) Sosyal medya trendleri ya da makro ekonomik göstergelerin modele dahil edilmesi,
- iii) Örneklem döneminin çeşitlendirilmesi,
- iv) Teknik göstergelerin pencere uzunlularının artırılması örnek olarak verilebilir.

En kapsamlı topluluk ML yöntemlerinin tercih edildiği yakın tarihli Prasad ve Bakhshi (2022)'nin çalışmasında yöntemlerin doğruluk oranı %63,33 ile %67,37 aralığında kalmıştır. Google Trend verilerinin de dahil edildiği bir çalışmada Hu vd. (2018), BPNN ile en fazla %88,98'lik doğruluk oranını elde etmiştir. Shynkevich vd. (2017) teknik göstergelerinin tüm pencere boyutlarını modele dahil ederek en fazla %74,44 doğruluk oranına ulaşabilmiştir. Basak vd. (2019) tarafından önerilen yaklaşım ile RF doğruluk oranı %94,76 ve XGBoost doğruluk oranı %94,44 ulaşmıştır. Bu çalışmada elde edilen sonuçlara iki önemli veri işleme aşaması büyük katkı sağlamıştır. Bu aşamalardan birincisi, pencere boyutunun 90'güne kadar uzatılması ve ikincisi, rassal varyansı ortadan kaldırmak için ham verilere üstel düzleştirilmenin uygulanmasıdır.

Hisse senedi yön tahmininde topluluk öğrenmesi yöntemlerinin tercih edildiği bu çalışmada, mevcut çalışmaların önemli iyileştirmeleri dikkate alınmıştır. Bu iyileştirmeler, pencere boyutunun çeşitlendirilmesi ile girdi verilerinin sayısının artırılması ve ML literatüründe henüz veri ön işleme adımı olarak ele alınmayan üstel düzleştirilmenin çalışmaya adapte edilmesi olarak özetlenebilir.

5. SONUÇ

Bu çalışmada üstün tahmin yeteneğini göz önünde bulundurarak Karar Ağaçlarına dayanan topluluk ML yöntemlerinden RF, XGBoost ve CatBoost yöntemleri BİST-100 endeksinin yönünü tahmin etmek amacıyla kullanılmıştır. Çalışmada, bağımlı değişken -1 (azalış) ve +1 (artış) değerlerini alırken, özellik uzayında farklı pencere uzunluklarına sahip 10 teknik gösterge ele alınmıştır. Önceki çalışmalar ışığında, özellikle borsa endeks yönü ve hisse senedi fiyat tahminlerinde, teknik göstergeler veri setinin kendisine ait değerleri kullanılarak hesaplandığından oldukça kullanışlı ve veri setinin doğasını iyi bir şekilde yansıtmaktadırlar.

Çalışma sonuçları dikkate alındığında her üç ML yönteminin performanslarının birbirlerine çok yakın oldukları görülmüştür. Ancak XGBoost yöntemi dengesiz veri setlerinde de üstün performansı ile öne çıkmaktadır. RF ve CatBoost yöntemleri XGBoost yöntemine çok yakın sonuçlar vermekle birlikte birbirlerine doğruluk oranı açısından üstünlük sağlayamamışlardır. Buna ilave olarak, modelimizin sağlamlığı, kesinlik, özgüllük, Matthews korelasyon katsayısı ve F-skoru gibi çeşitli parametreler hesaplanarak değerlendirilmiştir. Çalışmada artış (376 artış yönlü hareket) ve azalışları (168 azalış yönlü hareket) temsil eden sınıfların gözlem sayıları dengesiz (unbalance) dağılmıştır. Bu gibi durumlarda yanlış pozitiflerin oranını (precision) ve yanlış negatiflerin oranını (recall) ortaya koyan iki ölçüm arasındaki dengeyi sağlayan F-1 score ve karmaşıklık matrisinin tüm elemanlarını dikkate alan Matthew korelasyon katsayısı model performansını değerlendirmek için daha değerli bir ölçümdür.

Bu çalışmada önerilen süreç karşılaştırmalı olarak incelendiğinde, literatürde yer alan tek modele dayalı ML yöntemlerinden daha iyi sonuçların elde edildiği görülmektedir. Bir dizi karar ağacı modelinin kurulması esasına dayanan topluluk ML yöntemleri finansal verilerin doğasında yer alan doğrusal olmayan karmaşık yapıyı kavramak açısından avantaj sağlamıştır. Bu çalışmada kullanılan topluluk ML yöntemleri, SVM, ANN, DT gibi yöntemler ile kıyaslandığında uygulama ve hesaplama açısından kolaylıkları sebebiyle teorisyenler ve uygulayıcılar için yararlı bir araç olmaktadır. Bununla beraber %90 ve üzeri doğruluk oranı ile tahmin yapabilen bu modeller, piyasa uygulayıcılarına ve borsa yatırımcılarına mevcut piyasa risklerinin azaltılması bakımından yeni bir alternatif olarak sunulabilirler.

KAYNAKÇA

- Aras, S. (2020). *Using technical indicators to predict stock price index movements by machine learning techniques*, E. Sarikaya (Ed.), In Theory and Research in Social, Human and Administrative Sciences II, (1. Baskı, s. 249-274). Gece Publishing.
- Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert systems with Applications*, 42(20), 7046-7056.
- Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, 552-567.
- Breiman, L., Friedman, J.H., Olshen, R., & Stone, A.C.G. (1984). *Classification and Regression Trees*, (1). Wadsworth International Group, Belmont, California, USA.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.

- Chen, H. Y., Lee, C. F., & Shih, W. K. (2016). Technical, fundamental, and combined information for separating winners from losers. *Pacific-Basin Finance Journal*, 39, 224-242. <https://doi.org/10.1016/j.pacfin.2016.06.008>
- Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29(4), 309-317. [http://doi.org/10.1016/S0305-0483\(01\)00026-3](http://doi.org/10.1016/S0305-0483(01)00026-3)
- Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and its Applications*, 519, 127-139. <https://doi.org/10.1016/j.physa.2018.11.061>
- Cavalcante, R. C., Brasileiro, R. C., Souza, V. L., Nobrega, J. P., & Oliveira, A. L. (2016). Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, 55, 194-211. <https://doi.org/10.1016/j.eswa.2016.02.006>
- Chen, T., & Guestrin, C. (2016, August, 13-17). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* [Oral presentation]. San Francisco, California, USA
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Li, Y. (2019). Xgboost: Extreme gradient boosting. R package version 0.90.0.2. *R Package Version 0.90.0.2*.
- Dai, Y., & Zhang, Y. (2013). Machine learning in stock price trend forecasting. *Stanford University*, <http://cs229.stanford.edu/proj2013/DaiZhang-MachineLearningInStockPriceTrendForecasting.pdf>.
- Dey, S., Kumar, Y., Saha, S., & Basak, S. (2016). Forecasting to classification: Predicting the direction of stock market price using Xtreme Gradient Boosting. PESIT South Campus, 1-10. [10.13140/RG.2.2.15294.48968](https://doi.org/10.13140/RG.2.2.15294.48968)
- Dixit, G., Roy, D., & Uppal, N. (2013). Predicting India volatility index: An application of artificial neural network. *International Journal of Computer Applications*, 70(4). Doi: [10.5120/11950-7768](https://doi.org/10.5120/11950-7768)
- Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: Gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*.
- Hann, T. H., & Steurer, E. (1996). Much ado about nothing? Exchange rate forecasting: Neural networks vs. Linear models using monthly and weekly data. *Neurocomputing*, 10(4), 323-339. [https://doi.org/10.1016/0925-2312\(95\)00137-9](https://doi.org/10.1016/0925-2312(95)00137-9)
- Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32(10), 2513-2522.
- Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google Trends. *Neurocomputing*, 285, 188-195.
- Hsu, M. W., Lessmann, S., Sung, M. C., Ma, T., & Johnson, J. E. (2016). Bridging the divide In financial market forecasting: Machine learners & financial economists. *Expert Systems With Applications*, 61, 215-234.
- Jhaveri, K., Shah, D., Bhanushali, S., & Johri, E. (2016, April). Financial market prediction using hybridized neural approach. In 2016 International Conference on Computation of Power, Energy Information and Commuication (ICCPEIC) (pp. 009-014). IEEE.
- Jiao, Y., & Jakubowicz, J. (2017, December). Predicting stock movement direction with machine learning: An extensive study on S&P 500 stocks. In *2017 IEEE International Conference on Big Data* (pp. 4705-4713). IEEE.
- Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert systems with Applications*, 38(5), 5311-5319.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017, December). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, LongBeach, CA, USA.
- Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*.
- Kumar, D., Meghwani, S. S., & Thakur, M. (2016). Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. *Journal of Computational Science*, 17, 1-13.
- Kumar, M., & Thenmozhi, M. (2006, January). Forecasting stock index movement: A comparison of support vector machines and random forest. In *Indian institute of capital markets 9th capital markets conference paper*. <http://dx.doi.org/10.2139/ssrn.876544>
- Leung, M. T., Daouk, H., & Chen, A. S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of forecasting*, 16(2), 173-190.
- Mehta, S., Rana, P., Singh, S., Sharma, A., & Agarwal, P. (2019, August). Ensemble learning approach for enhanced stock prediction. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-5). IEEE.
- Nava, N., Di Matteo, T., & Aste, T. (2016). Time-dependent scaling patterns in high frequency financial data. *The European Physical Journal Special Topics*, 225, 1997-2016.

- Ünlü, M. (2023). Borsa endeksi ve makroekonomik değişkenler arasındaki zamanla değişen nedensellik ilişkisi: Bist 100 endeksi üzerinden ampirik kanıtlar. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (60), 243-256.
- Pabuçcu, H. (2019). Borsa endeksi hareketlerinin tahmini: trend belirleyici veri. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 22(1), 246-256.
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems With Applications*, 42(1), 259-268.
- Prasad, A., & Bakhshi, P. (2022). Forecasting the direction of daily changes in the India VIX index using machine learning. *Journal of Risk and Financial Management*, 15(12), 552.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018, December 3-8). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31, Montréal, Canada.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1, 81-106.
- Quinlan, J. R. (1993). C4. 5: Programs for machine learning (Vol. 1). Morgan kaufmann.
- Saad, E. W., Prokhorov, D. V., & Wunsch, D. C. (1998). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Transactions on Neural Networks*, 9(6), 1456-1470.
- Shynkevich, Y., McGinnity, T. M., Coleman, S. A., Belatreche, A., & Li, Y. (2017). Forecasting price movements using technical indicators: Investigating the impact of varying input window length. *Neurocomputing*, 264, 71-88.
- Thakkar, A., & Chaudhari, K. (2021). Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions. *Information Fusion*, 65, 95-107.
- Tsaih, R., Hsu, Y., & Lai, C. C. (1998). Forecasting S&P 500 stock index futures with a hybrid AI system. *Decision support systems*, 23(2), 161-174.
- Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications*, 79, 153-163.
- Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert systems with applications*, 78, 225-241.
- Yang, J., Rao, R., Hong, P., & Ding, P. (2016, December,16-19). Ensemble model for stock price movement trend prediction on different investing periods. 12th International Conference on Computational Intelligence and Security (CIS), Wuxi, China
- Yu, L., Chen, H., Wang, S., & Lai, K. K. (2009, February,19). Evolving least squares support vector machines for stock market trend mining. *IEEE transactions on Evolutionary Computation*, 13(1), 87-102.
- Yu, L., Wang, S., & Lai, K. K. (2005, December). Mining stock market tendency using GA-based support vector machines. *International workshop on Internet and network economics*. Springer Berlin Heidelberg.

ARAŞTIRMACILARIN KATKI ORANI

1.yazarın araştırmaya katkı oranı %50, 2. yazarın araştırmaya katkı oranı %50'dir.

Yazar 1: Araştırmada Özet, Yöntemler, Uygulama, Kaynak Araştırılması ve Sonuç bölümlerine katkıda bulunmuştur.

Yazar 2: Araştırmada Özet, Giriş, Literatür Taraması, Yöntemler, Uygulama ve Sonuç bölümlerine katkıda bulunmuştur.

ÇATIŞMA BEYANI

Araştırmada herhangi bir kişi ya da kurum ile finansal ya da kişisel yönden bağlantı bulunmamaktadır. Araştırmada herhangi bir çıkar çatışması bulunmamaktadır.