

## HİBRİT REGRESYON MODELLERİ İLE BİST'E ETKİ EDEN G20 ENDEKSLERİNİN BELİRLENMESİ\*

### DETERMINATION OF G20 INDICES AFFECTING BIST WITH HYBRID REGRESSION MODELS

Dr. Ahmet SEL<sup>1</sup>

#### ÖZ

Yatırımcılar için borsa endekslerinin yönelim tahmini endeksin etkilendiği çok fazla değişken olmasından dolayı zor olduğu kadar ihtiyaç duyulan bir konudur. Küresel olaylar, politika yapıcıların davranışları, ekonomik faktörler ile özellikle son dönemde görülen salgın hastalıklar endekslerin fazlasıyla değişkenlik göstermesine sebep olmuştur. Bu çalışmada parametrik olmayan regresyon modelleri ve hibrit yaklaşımları kullanarak BİST 100 endeksine etki eden G20 endeksleri incelenmiştir. Veri setinde 2010/01/01 ve 2019/12/31 tarihleri arasında endeks günlerinin açık olduğu ortak günler dikkate alınarak 2432 günlük kapanış değerleri kullanılmıştır. Bağımsız değişkenler Bovespa, İtalya40, KOSPI, Nikkei 225, BMVIIPC ve Shanghai endeksleri seçilmiştir. Çalışmada veri madenciliği uygulamaları Knime programı yardımıyla 30 bölütleme için en yüksek  $R^2=0,9918$  için en düşük MAE=0,0650, MSE=0,0082 ve RMSE=0,0903 değerleri bulunmuştur. Analiz sonuçlarına göre Bovespa, İtalya40, KOSPI, Nikkei 225 ve BMVIIPC endekslerinin BİST-100 endeksini pozitif yönlü olarak etkilediği, Shanghai endeksinin ise negatif yönlü etkilediği belirlenmiştir.

**Anahtar Kelimeler:** Veri Madenciliği, Büyük Veri, Hibrit Makine Öğrenmesi, Regresyon.

**JEL Sınıflandırma Kodları:** C13, C53, E27.


#### ABSTRACT

For investors, the prediction of the trend of the stock market indices is not only a difficult but needed issue due to various variables that affect the index. Global events, policymakers' actions, economic variables and, in particular, recent epidemics have caused indices to respond more than once. Using hybrid regression models, the study aims to predict the relationship between BİST 100 index and G20 index. The closing values of 2,432 days are taken as the data set by considering the common days when the index days between 2010/01/01 and 2019/12/31 are open. Independent variables, Bovespa, Italy40, KOSPI, Nikkei 225, BMVIIPC, Shanghai and Tadawul, are obtained by taking into account the correlation values of indices which are expected to have an effect on the BİST index in the G20 countries, with an absolute value below 0.80. In the study, data mining applications have found the highest  $R^2 = 0.9918$  for 30 segmentation with the Knime program, the lowest MAE = 0.0650, MSE = 0.0082 and RMSE = 0.0903. According to the results of the analysis, it is determined that Bovespa, Italy40, KOSPI, Nikkei 225 and BMVIIPC indices affect BIST-100 index positively, and Shanghai index negatively.

**Keywords:** Data Mining, Big Data, Hybrid Machine Learning, Regression.

**JEL Classification Codes:** C13, C53, E27.

\* Bu çalışma 1-4 Ekim 2020 tarihlerinde 1. Uluslararası Uygulamalı İstatistik Kongresi'nde sunulan ve bildiri özet kitabında "Determination of G20 Indices Affecting BIST with Hybrid Regression Models" başlığıyla basılan bildirinin revize edilmiş halidir.

<sup>1</sup>  T. C. Milli Eğitim Bakanlığı, selahmet43@gmail.com

## EXTENDED SUMMARY

### **Purpose and Scope:**

Investors can invest in any product they want and sell it with ease. As a result, the positive or negative consequences of any case have an effect on markets in other countries. Financial markets have emerged from a position that can be considered separate from one another, particularly in light of recent events. Data mining methods go beyond routine statistical methods in terms of big data analysis and obtaining different relationships and patterns on the basis of data. The advancement of technology and the increase in storage areas have made big data analysis a necessity. The increase in the programs and applications used in the field of data mining has been beneficial for the analysis to be done more easily. Data mining, which has many application areas, is preferred in the application of main methods such as association, clustering, causality and classification. The impact of the G20 countries' indices on the BIST 100 index was examined in this report. Data mining approaches have been favoured over big data analysis methods since Index data spans a decade. The Knime software was used to complete the applications. Machine learning models and hybrid methods were used in the study to increase the performance values of regression models. With the aid of the evaluations obtained prior to the start of the pandemic, it was attempted to evaluate the impact and direction of the indices to which the BIST 100 index is associated. The evolution of the BIST index in the normalization phase after the pandemic can thus be seen with the aid of similar indices using the results obtained. Furthermore, the findings are expected to aid in the direction of investments and portfolio diversification.

### **Design/methodology/approach:**

The increase in the amount of data and the creation of the appropriate resources for analyzing this data has increased interest in analyzing the collected data in recent years, thanks to advancements in information and communication technologies. Data mining activities have become commonplace in both the academic and business worlds as a result of the innovations. Since the random forest method, support vector regression analysis, and machine learning approaches were combined in the research, a hybrid approach was used. Furthermore, data mining techniques have been studied using the Knime software due to the large number of data and the preference for hybrid methods. The study's variables were found and normalized to find standardized values. Using the random forest regression method under the segment node, a model with the best output was found by repeating operations up to 50 segments and checking the number of segments in multiples of 5.

### **Findings:**

The results of random forest, support vector, and hybrid regression models, as well as their combinations, were compared in this analysis. Between January 1, 2010 and December 31, 2019, the independent variables Bovespa, Italy40, KOSPI, Nikkei 225, BMVIIPC, and Shanghai indices were chosen for the dependent variable BIST 100. Variables with an absolute correlation value of less than 0.80 were used as a precaution against the multiple association issue since the regression model was used in this collection. Using data mining applications, the Knime software and python plugin were used to construct the applications. For the results, 30 segmentation had the highest  $R^2 = 0.9918$  and the lowest MAE = 0.0650, MSE = 0.0082, and RMSE = 0.0903. In addition, the "rbf" kernel structure was chosen in the support vector regression model, and the data was divided into training (66%) and test (33%) data for review.

### **Conclusion and Discussion:**

$R^2$ : 0.940 was calibrated for the support vector regression model, while this value was 95.8% in the hybrid process, according to the results of the study. The applied method has been found to be more efficient, and the random forest method can also be used to solve the visualization problem. The BIST-100 index was found to be positively affected by the Brazil-Bovespa, Italy-Italy40, South Korea-KOSPI, Japan-Nikkei 225, and Mexico-BMVIIPC indices. The countries whose indices contributed the most to the BIST-100 index were Japan, Mexico, Brazil, and South Korea, according to the findings. The China-Shanghai index was found to have a negative effect. BIST-100 is unaffected by the Saudi Arabia-Tadawul index. The impact of increases and decreases in related country indices on the BIST index was investigated in this report. The findings are expected to aid decision-makers in their investment decisions. Different machine learning approaches and models are thought to be contributed to the literature in future research by creating different hybrid methods.

## 1. GİRİŞ

Dünyanın herhangi bir yerinde piyasa ve fiyatlar üzerinde etkili olabilecek bir haber anında ulaşmakta ve borsalar reaksiyon gösterebilmektedir. Sermaye hareketlerinin önündeki engellerin de kalkmasıyla birlikte bütün piyasalar aynı bilgiyi aynı anda elde ettiği için hisse senetleri arasında fiyat farklılıkları ortadan kalkmış ve piyasalar birlikte hareket etmeye başlamıştır (Yıldız ve Aksoy, 2014:22). Yatırımcılar istedikleri ürüne yatırım yapabildikleri gibi satış konusunda da sıkıntı yaşamamaktadırlar. Bu nedenle meydana gelen herhangi bir olayın olumlu ya da olumsuz etkileri diğer ülke piyasalarını da etkilemektedir. Özellikle son gelişmelerle finansal piyasalar birbirinden ayrı olarak düşünülebilecek bir konumdan çıkmıştır.

Yatırım argümanları günümüzde fazlasıyla olmasına rağmen hisse senetleri bu konuda temel olarak nitelendirilebilir. Hisse senetleri getirileri fazla olabilmesine karşın risk seviyeleri de bir o kadar yüksektir. Buna rağmen modellerle yardımcıyla riskler analiz edilebilmektedir. Bunların bir tanesi de hisse senetlerinin çeşitlendirilmesidir. Çeşitlendirme yoluyla yatırım sepetine dahil olan yatırım araçları riskin azaltılmasında etkili olabilmektedir. Çeşitlendirme de ana yöntem ilişki oranları az olan argümanların sepete dahil edilmesidir. Çeşitlendirme yönteminde hisse senedi bazında olduğu gibi borsaların birbirleriyle olan ilişkileri de büyük önem taşımaktadır. İlgili borsa altında bulunan hisseler borsa genel endeksinin ani hareketlerine genellikle aynı yönlü tepki vermektedirler. Aralık 2019'da ortaya çıkan koronavirüs (Covid-19) pandemisi piyasalarda ve yatırım araçlarının çoğunda olağan dışı dalgalanmalara sebep olmuştur. Teknolojinin ilerlemesi ile dünya piyasalarının tamamına ulaşabilen yatırımcılar için farklı borsalara yatırım imkânı sağlaması çeşitlendirmeye imkân tanımıştır. Özellikle çeşitlendirme konusunda tahminlerin doğruluğu başarılı kararlar alınmasını sağlar ve yatırımcıların fayda maksimizasyonuna ulaşmasına imkân tanır.

Veri madenciliği yöntemleri büyük veri analizi ve verinin temelinde olan farklı ilişki ve örüntülerin elde edilmesi yönünden rutin istatistiksel yöntemlerin dışına çıkmaktadır. Teknolojinin ilerlemesi ve depolama alanlarının artması büyük veri analizini bir gereklilik haline getirmiştir. Veri madenciliği alanında kullanılan programların ve uygulamaların artması analizlerin daha rahat yapılabilmesi konusunda faydalı olmuştur. Birçok uygulama alanı olan veri madenciliği ilişkilendirme, kümeleme, nedensellik ve sınıflama gibi ana yöntemlerin uygulanmasında tercih edilmektedir.

Bu çalışmada, G20 ülkeleri endekslerinin BİST 100 endeksine olan etkisi araştırılmıştır. Uygulamada G-20 ülkelerinin seçilme nedeni, Dünya ekonomisinin yaklaşık %85'ini, ticaretinin %75'ini ve nüfusunun 2/3'ünü temsil etmekte olan G20, farklı kıtalardan gelişmiş ve yükselen ekonomileri aynı platformda bir araya getiren, kapsayıcı yapısıyla uluslararası ekonomik düzlemde profili giderek yükselen bir oluşum olarak kendisini göstermesidir (Dışişleri Bakanlığı, 2020). Endeks verileri on yılı kapsamasından dolayı büyük veri analizi yöntemlerinden veri madenciliği yöntemleri tercih edilmiştir. Uygulamalar Ktime programı yardımcıyla gerçekleştirilmiştir. Çalışmada makine öğrenme modelleri ve hibrit yaklaşımlar kullanılarak regresyon modellerinin başarı değerleri artırılmıştır. Elde edilen değerlendirmeler yardımcıyla pandemi süreci başlamadan önce BİST 100 endeksinin ilişkili olduğu endekslerin etkisi ve yönü belirlenmeye çalışılmıştır. Böylelikle elde edilen sonuçlar yardımcıyla pandemi sonrası normalleşme sürecinde BİST endeksinin gelişimi ilişkili endeksler yardımcıyla görülebilecektir. Ayrıca sonuçların yatırımların yönlendirilmesi ve portföy çeşitlendirilmesinde yardımcı olması beklenmektedir.

## 2. LİTERATÜR TARAMASI

BİST endeksinin tahmini konusunda literatürde genellikle Yapay Sinir Ağları (YSA) ve makine öğrenme yöntemleri kullanılarak yapılan çalışmalar incelendiğinde; Diler (2003), çalışmasında bir gün gecikmeli veriler ile BİST 100 endeksini tahmin etmiştir. Sonuç olarak %60,81 açıklama oranı ile fiyat yönü tahmini gerçekleştirmiştir. Karaatlı, Güngör, Demir ve Kalaycı (2005), çalışmalarında BİST endeksi kapanış fiyatını YSA ve regresyon yöntemi ile tahmin etmiştir. Bağımsız değişkenler olarak faiz, tasarruf mevduatı ve enflasyon oranı, altın fiyatı, sanayi üretim endeksi ve Dolar kurunu kullanarak aylık tahminlerde bulunmuştur. Modeller arasında YSA modeli performansını daha yüksek elde edilmiştir. Altay ve Satman (2005), çalışmalarında BİST endeksi için yön ve nokta tahmininde bulunmuşlardır. YSA ve regresyon modeli kullanılan analizlerde nokta tahmini için her iki model ile yaklaşık değerler elde edilirken, yön tahmininde YSA modeli regresyon modeline göre daha iyi performans göstermiştir. Yıldız, Yalama ve Coşkun (2008), çalışmalarında YSA yardımcıyla bir gün gecikmeli veriler ile BİST 100 endeksini tahmin etmişlerdir. Bağımsız değişkenler BİST endeksinin en düşük, en yüksek, kapanış fiyatları ve ayrıca TL-Dolar kurudur. Sonuç olarak elde edilen model ile %54,37 performans elde edilmiştir. Kara,

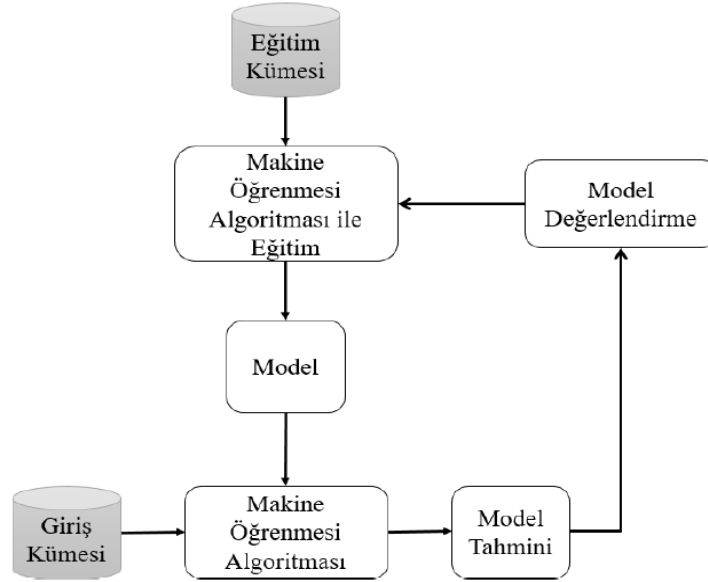
Boyacıoğlu ve Baykan (2011), çalışmalarında BİST 100 endeks değerini destek vektör makineleri ve YSA modeli ile tahmin etmişlerdir. Bağımsız değişken olarak 10 adet teknik gösterge kullanılmıştır. Sonuç olarak YSA modeli %75,74 ve destek vektör makineleri %71,52 oranında doğru tahmin etmiştir. Aygören, Sarıtaş ve Morali (2012), çalışmalarında BİST 100 endeks değeri tahmini için YSA ve Newton nümerik arama modelleri denenmiş ve performansları karşılaştırılmıştır. Sonuç olarak YSA modelinin performansı daha yüksek olduğu görülmüştür. Özçalıcı ve Ayrıçay (2016), çalışmalarında teknik göstergeler yardımıyla genetik algoritma tabanlı uzman bir sistem tasarlamıştır. Sonuç olarak 53 adet teknik gösterge için, kapanış fiyatlarının 6 günlük üçgensel hareketinin ortalamasının daha etkin olduğu görülmüştür. Filiz, Karaboğa ve Akoğul (2017), çalışmalarında BİST-50 endeksini makine öğrenmesi metotları ve YSA kullanılarak incelemişlerdir. Bağımlı değişken BİST-50 için bağımsız değişkenler; DAX, S&P 500, BİSTSINAİ, BİSTMALİ, FTSE, BİSTBANK, Ham Petrol Fiyatı, Euro/Dolar Paritesi, Faiz oranları ve GOLDINDEX kullanılmıştır. Analizde k-NN, Naive Bayes, C4.5 ve YSA modelleri kullanılmıştır. Sonuç olarak C4.5 algoritması %92,71 başarı oranı ile en iyi performansı göstermiştir. Özer, Sarı ve Başakın (2017), çalışmalarında BİST 100 endeksi değeri tahminini bağımsız değişkenler; Almanya, Hindistan, Meksika, İstanbul, ABD, Çin, İngiltere ve Fransa endeksleri kullanılarak YSA ve bulanık mantık yöntemleriyle gerçekleştirmiştir. Sonuç olarak bulanık mantık tekniğinin diğerlerine göre daha iyi bir performans sergilediği görülmüştür. Kara (2019), çalışmasında bağımsız değişken olarak teknik göstergeler kullanarak borsa endeksi hareket yönünü tahmin etmeye çalışmıştır. Makine öğrenmesi yöntemlerinden Naive Bayes, Destek Vektör Makineleri, YSA, Karar Ağaçları, k-En Yakın Komşuluk yöntemleri ile Doğrusal Diskriminant Analizi ve Lojistik Regresyon kullanılarak denemeler gerçekleştirilmiştir. Sonuç olarak %83,83 ile YSA'nın yüksek performans gösterdiği belirlenmiştir. Pabuççu (2019), çalışmasında YSA, destek vektör makineleri ve Naive Bayes yöntemleri ile bağımlı değişken olan Borsa endeksini bağımsız değişkenler olan on teknik gösterge ile tahmin etmiştir. Analiz sonucunda her üç modelinde başarılı olmasına rağmen YSA modelinin performansının daha fazla olduğu görülmüştür.

Ayrıca BİST endeksine etki eden değişkenler için regresyon ve zaman serileri yöntemleri kullanılarak yapılan çalışmalar incelenmiştir. Özmerdivanlı (2014), çalışmasında BİST 100 endeksi ile petrol fiyatları ilişkisini Granger nedensellik ve eş bütünleşme yöntemi kullanarak incelemiştir. Eş bütünleşme testi sonucunda petrol fiyatları ile BİST 100 endeksi için uzun dönem, nedensellik analizi sonucunda ise tek yönlü nedensellik belirlenmiştir. Yıldız ve Aksoy (2014), çalışmalarında gelişmekte olan endeksler olan Morgan Stanley ile BİST arasında hata düzeltme modeli, eşbütünleşme ve Engle- Granger yöntemi kullanılarak ilişki olup olmadığı incelenmiştir. Eşbütünleşme analizi sonucuna göre serilerin uzun dönemde birlikte hareket etmekte olduğu ve portföy çeşitlendirmesine uygun oldukları görülmüştür. Şimşek (2016), çalışmasında BRICS ülkeleri ve BİST hisse senedi piyasalarının etkileşimini ARCH ve GARCH modelleri kullanarak araştırmıştır. Sonuç olarak; BRICS ile BİST ülkelerinin ilişkili olduğu ve bu ilişki en fazla Güney Afrika ve Hindistan ülkeleri arasında olduğu tespit edilmiştir. Tekin ve Hatipoğlu (2017), çalışmalarında kantil regresyon yöntemi kullanarak BİST endeksine dolar kurunun, petrol fiyatlarının ve VIX endeksinin etkisini araştırmıştır. Analiz sonucunda; önemli ölçüde volatilité endeksinden etkilendiği, Dolar kurunun ise yüksek kantillerde BİST üzerinde etkili olduğu ve Petrol ile BİST arasında asimetric bir ilişki olmadığı görülmüştür. Caba (2017), çalışmasında firma büyüklüğü ve finansal kaldıraçın işletme finansal performansında etkisinin yönü araştırılmıştır. BİST Sınai Endeksi üzerinde yapılan çalışmada çoklu regresyon yöntemi kullanılmıştır. Sonuç olarak firma büyüklüğü ve finansal kaldıraçların işletmelerin finansal performansı üzerinde etkili olduğu görülmüştür. Güney ve Iğın (2019), çalışmalarında BİST 100 endeksi bağımlı değişken iken altın fiyatları, döviz kuru ve faiz oranlarının bağımsız değişken olarak alınmıştır. Çalışmada etki analizleri Johansen eşbütünleşme testi, Granger Nedensellik Analizi ve VAR modeli ile incelenmiştir. Sonuç olarak, BİST100 ile faiz ve altın arasında çift yönlü; döviz, altın ve faiz arasında ise tek yönlü nedensellik olduğu görülmüştür. Varyans ayrıştırması ve etki tepki analizi sonucunda ise BİST100 endeksine karşı alternatif yatırımların sırasıyla faiz oranları, döviz kurları ve altın fiyatları olduğu tespit edilmiştir. Güngör ve Polat (2020), çalışmalarında bağımlı değişken BİST'teki sektörler olmak üzere bağımsız değişkenler; döviz kuru, altın fiyatı ve faiz oranı kullanılmıştır. Sonuç olarak yatırım araçlarının hisse senedi fiyatını negatif yönde etkilediği görülmüştür. Altın fiyatı etkisiz bulunurken anlamlı olan döviz kurunun, hisse senedi fiyatını faiz oranından daha çok etkilediği belirlenmiştir.

### 3. YÖNTEM

Son yıllarda, bilgi ve iletişim teknolojilerinde meydana gelen gelişmeler yardımıyla veri miktarının artması ve bu verilerin analizi için gerekli araçların gelişmesi, toplanan verilerin analiz edilmesine yönelik ilgiyi de arttırmıştır.

Meydana gelen gelişmelere paralel olarak gerek akademik gerekse iş dünyasında veri madenciliği uygulamalarının yaygınlaştığı görülmektedir. Veri madenciliği, toplanan verilerden anlamlı bilgiler çıkarmak, veri içerisinde gizli olan birtakım örüntüleri ve eğilimleri tespit etmek ve çeşitli değişkenler arasında ilişkiler bulmak ve böylece karar vermeye yardımcı olmak amacıyla uygulanan bir yaklaşımdır. Veri madenciliği uygulaması temelinde istatistik, yapay zekâ, makine öğrenmesi gibi farklı alanlarda geliştirilmiş birçok teknik ve yöntem kullanılmaktadır (Rygielski, Wang ve Yen, 2002:483). Hibrit kelimesi Türk Dil Kurumuna (TDK) göre iki farklı güç kaynağının bir arada bulunması veya melez anlamında kullanılmaktadır (Türk Dil Kurumu, 2021). Çalışmada rasgele orman yöntemi, destek vektör regresyon analizi ve makine öğrenme yöntemleri bir arada kullanıldığından dolayı hibrit bir yaklaşım benimsenmiştir. Ayrıca veri sayısının büyüklüğü ve de hibrit yöntemlerin tercih edilmesinden dolayı, veri madenciliği yöntemleri Knime programı kullanılarak analiz edilmiştir. Uygulamanın temelinde yatan makine öğrenme süreci Şekil 1’de gösterilmiştir.



Şekil 1. Temel Makine Öğrenmesi Süreci

Makine öğrenmesi yardımıyla modelin belirlenmesi ve değerlendirilmesi eğitim ve test verileri yardımıyla gerçekleşmektedir. Eğitim kümesi ile oluşturulan modelin test kümesi yardımıyla performansı değerlendirilmektedir. Eğer istenilen düzeylerde performans elde edilemezse döngü yeniden oluşturulmakta ve süreç istenilen düzey yakalanana kadar devam etmektedir.

Standart sapma; verilerin ortalamadan sapmalarını gösteren bir dağılım ölçüsüdür. Uygulamalarda çoğu zaman veriler farklı olduğu için ortalama ve varyans farklı çıkar. Normalleştirme ile farklı birimlerdeki değişkenlerin karşılaştırılmaları mümkün olur. Bu yüzden veri setinde yer alan değişkenlerin normalleştirilmesi gerekir. Normalleştirme yapılmasının avantajları (Tunç ve Ülger, 2016);

- Standartlaştırılmış sonuçlara ulaşmayı sağlar.
- Standartlaştırılmış değerler ile değişkenler arası karşılaştırma daha kolaydır.
- Karmaşıklığı ortadan kaldırır.

Çalışmada kullanılan değişkenlerin standartize edilmiş değerleri bulunarak normalleştirme işlemine tabi tutulmuştur.  $Z'$  standardize edilmiş değer olmak üzere her bir veri  $X$  için  $\mu$ : ortalama ve  $\sigma$ : standart sapma değeri ile Denklem (1) ile normalleştirme işlemine tabi tutulmuştur.

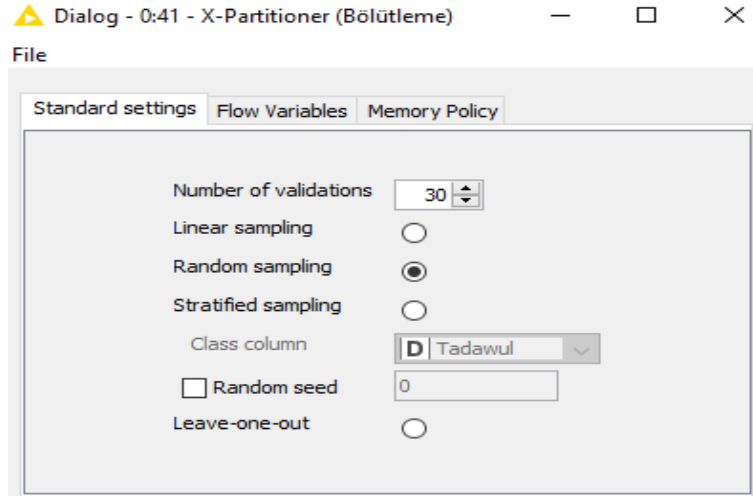
$$Z' = \frac{X - \mu}{\sigma} \quad (1)$$

Normalleştirme işlemi sonucunda elde edilen veriler için sırasıyla bölütleme, rasgele orman regresyon yöntemi, birleştirme, destek vektör regresyon analizi ve performans değerlendirme adımları izlenmiştir.

### 3.1. Bölütleme ve Birleştirme

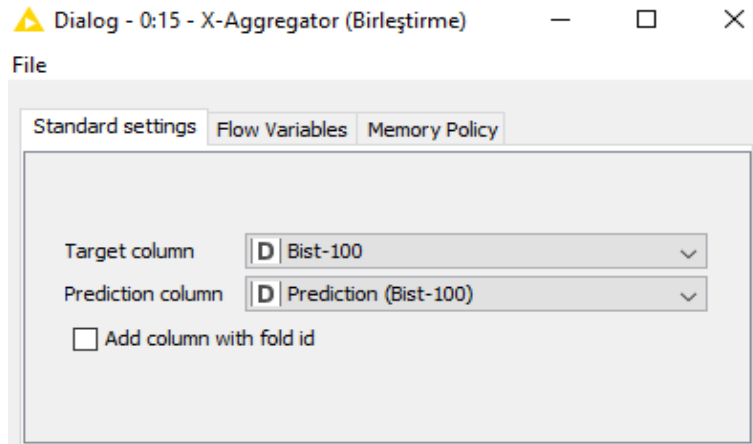
Bölütleme; veri setinin “n” parçaya ayrılması ve “n-1” parçasının eğitim, geriye kalan bir parçanın da test olarak kullanıldığı çapraz geçerlilik uygulamasıdır. Modelde bulunan tüm veriler her seferinde eğitim ve test kümesine ayrılarak “n” parçanın tamamı çalışmaya dahil olmaktadır. Örneğin, bölütleme işlemi ile n=5 parçaya ayrılan verinin n-1=4 parçası eğitim için kullanılırken bir parçası test için kullanılır. “n” kere tekrarlanan bu işlem sonucunun her biri için hesaplanan hata değerlerinin ortalaması alınarak genel hata değeri elde edilmektedir. Bölütleme işlemi ile veri madenciliği uygulamalarında verinin tamamının eğitim ve test olarak sürece katılması sağlanarak olası tek seçimden elde edilecek olan sapmalar azaltılmaktadır.

Knime programı altında X-Partitioner (X-Bölütleme) düğümü çapraz doğrulama döngüsünde ilk düğümdür. Bu adımda veri istenilen sayıda parçaya ayrılarak her seferinde yapılan işlemlerin tekrarı ile doğrulama sağlanır. Bu adımda birden fazla doğrulama yöntemi olmasına rağmen, çapraz doğrulama yineleme sayısı değiştirilerek yüksek açıklayıcılık katsayısı olan  $R^2$  aranmıştır. Knime programı bölütleme işlemi düğümü Şekil 2’de gösterilmiştir.



Şekil 2. Knime Programı Bölütleme Düğümü

Döngünün sonunda, her yinelemeden sonuçları toplamak için bir X-Aggregator (X-Toplayıcısı) programa yerleştirilmesi gerekmektedir. Sonucu bir tahmin düğümünden toplar ve tahmin sınıfı ile gerçek sınıfı karşılaştırır. Bu düğüm ile tüm satırlar ve yineleme istatistikleri için tahminler elde edilir. Bu iki düğüm arasındaki tüm düğümler, yinelemelerin gerçekleştirilmesi gerektiği kadar yürütülür. Knime programı birleştirme işlemi düğümü Şekil 3’te gösterilmiştir.



Şekil 3. Knime Programı Birleştirme Düğümü

Bölütleme ve birleştirme döngüsü için doğrulama işlemi BİST endeksinin gerçek değerleri ile tahmin değerleri arasında gerçekleştirilmiştir.

### 3.2. Rasgele Orman Yöntemi Regresyon Analizi

Rasgele orman yönteminin oluşturduğu yapı Karar Ağaçları (KA) yapısıyla meydana gelmektedir. KA ilgili uygulamada bulunan verilerin cinsine göre ağaç yapısı yardımıyla sınıflandırma ve regresyon uygulaması yapılabilmektedir. KA modelleri için ağaç yapısında dalların oluşumu programlama dilinde yer alan “eğer” yapısını kullanmaktadır. Bu yapıyla birlikte uygulamanın başında karmaşık halde bulunan veriler birbirini takip eden katmanlı bir yapıya sahip olur.

Ağaç yapısını oluşturan dalların oluşumunda en önemli adım meydana gelen dallanmaların hangi kriter ya da öznelik değerinin dikkate alınacağına belirlenmesidir. Kaynaklar incelendiğinde bu problemin çözümünde kullanılan çeşitli yaklaşımlar olduğu görülmektedir. Bunlardan en önemli olarak kabul edilenler ise Gini indeksi, Twoing kuralı, bilgi kazancı ve bilgi kazanç oranı ile Ki-Kare olasılık tablo istatistiği yaklaşımlarıdır. ID3 algoritması tek değişkenli karar ağaçlarından olmak üzere bilgi kazancı yaklaşımını kullanırken, C4.5 algoritması bilgi kazancı ve bölünme bilgisi kavramından yararlanmaktadır. Regresyon ağacı ve sınıflandırma yöntemi olarak bilinen CART algoritması ise Twoing kuralını kullanmaktadır (Breiman, Friedman, Olshen ve Stone, 1984:28).

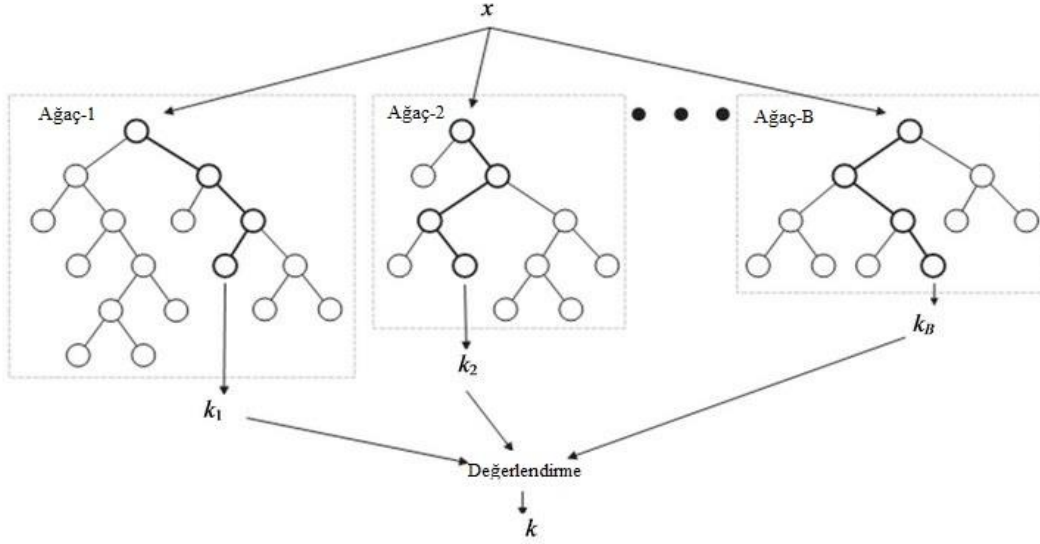
Bagging yöntemi, birçok sınıflama modeline uygulanabilmekle birlikte daha çok karar ağaçları için kullanılmaktadır. Bagging yöntemi veri setinden sınıf yapısını bozmayacak şekilde rastgele örnekler seçilerek (bootstrap) oluşturulan çok sayıda karar ağacının yaptığı sınıf tahminleri oylanarak en çok oy toplayan sınıfı asıl sınıf tahmini olacak şekilde belirleyen öğrenme yöntemidir. Bagging yönteminde art arda oluşturulan ağaçlar önceden oluşturulan ağaçlara bağımlı değildirler ve ağaçlar orijinal veri setinden bootstrap örnekleme yapılarak oluşturulmaktadır.

Rasgele Orman (Random Forests) yönteminde sonradan gelen veriye ait tahmin yapılmasının yanında, değişkenlerin önem derecesi de hesaplanmaktadır. Veri setinde çok sayıda değişken varsa, değişken önem derecesinin hesaplanması model indirgemesi açısından oldukça kullanışlıdır. Örneğin binlerce değişkenin bulunduğu veri setinde, Rasgele Orman yöntemiyle elde edilen ve önem derecesi yüksek olan değişkenler yardımıyla daha etkin tahminler yapılabilir.

Rastgele orman regresyon modelinde ağaç sayısının fazla olmasının her zaman daha yüksek performans elde etmeyi garanti etmediği söylenebilir (Breiman, 2017). Bundan dolayı veri sayısı da dikkate alınarak ağaç sayıları denenerek elde edilen modellerin performans değerlerinin karşılaştırılması uygun görülmektedir. Elde edilen modeller için görsel çıktı eksikliği ve işlem adımlarının karmaşıklığı dezavantaj olarak görülmektedir (Akman, Genç ve Ankaralı, 2011:38). Çalışmada bu dezavantajın ortadan kaldırılması adına parametrik olmayan destek vektör regresyon modelleri sürece eklenerek hibrit yöntemler elde edilmiştir.

Rastgele orman modelinde karar ağaçlarından oluşan bir topluluk sayesinde öğrenme gerçekleşir. Regresyon ağacı modellerinin her biri, farklı satırlarda bulunan örnekler ve farklı sütunlarda yer alan öznelikler üzerinde öğrenme gerçekleştirir. Ayrıca sütunlarda yer alan özellikler çalışmanın niteliğine göre değişkenlik gösterebilir. Çıktı modeli regresyon ağacı modellerinden oluşmaktadır. Her bir model ile elde edilen modeller için tek tek tahminlerin basit bir ortalaması kullanılarak ilgili tahmin düğümüne uygulanır.

Rasgele orman yönteminin yapısı karar ağaçları topluluğundan meydana gelen bir orman yapısını andırmaktadır. Şekil 4’te rasgele orman modelinin yapısı görülebilir.

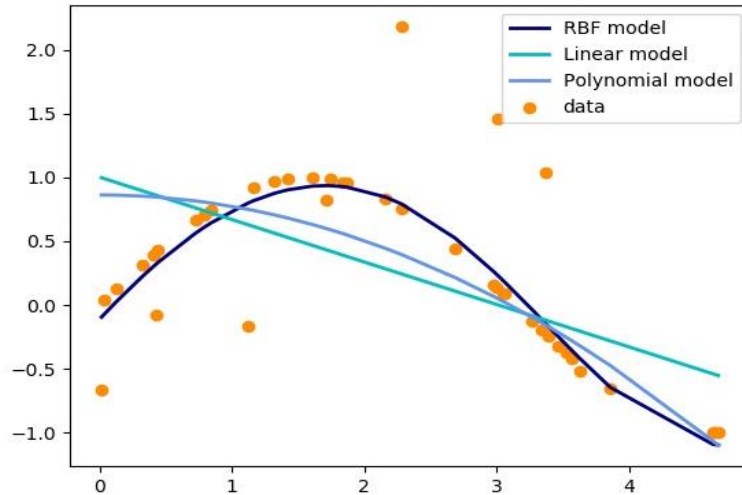


Şekil 4. Rasgele Orman Modeli Yapısı

Karar ağaçlarının her birinden elde edilen sonuçlar ( $k_1, k_2, k_3, \dots, k_B$ ) için ortalamaları tahmin değerini ( $k$ ) vermektedir. Ancak karar ağaçlarında olduğu gibi rasgele orman yönteminde elde edilen modelin görselleştirilmesi çoğu zaman sorun olabilmektedir.

### 3.3. Destek Vektör Makineleri

1963 yılında Vladimir Vapnik ve Alexey Chervonenkis tarafından temelleri atılan Destek Vektör Makineleri (Support Vector Machine) (SVM) istatistiksel öğrenme teorisine dayalı bir gözetimli öğrenme algoritmasıdır. Her ne kadar temelleri 60'lı yıllara dayansada 1995 yılında Vladir Vapnik, Bernhard Boser ve Isabelle Guyon tarafından geliştirilmiştir (Akpınar, 2017). SVM modelleri aşırı öğrenmeyi azaltan yapısı ile sınıflandırma ve regresyon problemlerinin çözümünde uygulanan ve daha iyi sonuçlar veren güdümlü bir öğrenme algoritmasıdır (Panigrahi ve Mantri, 2015). SVM algoritmasında, regresyon için Destek Vektör Bağlayıcısı (Support Vector Regressor-SVR) adı verilen bir yapı bulunmaktadır. SVR, regresyon hatalarını en aza indirmek için deneysel riskleri ölçer ve bunun için de bir maliyet fonksiyonu kullanır (Yu, Chen ve Chang 2016). SVM algoritması, doğrusal olan ve doğrusal olmayan türlere sahiptir. Hiper düzlemlerdeki en optimal noktayı bulabilmek için kernel adı verilen çekirdek yapıları kullanılır (Lin ve Wang, 2002). Şekil 5'te SVR modelleri için Radial Basis Function (RBF) metodu kernel yapısına göre doğrusal ve polinomal eğrilere göre daha başarılı sonuç verdiği görülmektedir.



Şekil 5. SVR (rbf)-Doğrusal-Polinomal Tahminler



Destek Vektör Makinesi yönteminde doğrusal olarak sınıflanabilen verileri birbirinden ayırt etmek için olası pek çok doğrusal fonksiyon arasından, marjini en büyük olan belirlenmektedir. Sınıflandırılacak örnekler doğrusal bir düzlemle ayrıştırılabilecek düzeyde olmadığına, bu yöntemde belirlenen bir Kernel fonksiyonu yardımı ile daha yüksek boyutlu bir uzaya aktarılması mümkün olmaktadır. Bu şekilde marjini en yüksek olan hiper düzlemler bulunur. Sonuç olarak veriler bu ayırt edici hiper düzleme göre sınıflara atanır (Coşgun ve Karağaoğlu 2011). Python programlama dili altında bulunan SVR modeli kernel yapıları {'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'} seçenek olarak verilmektedir. Verinin yapısı ve model başarısının yüksekliğine göre bu yapılardan uygun olanı modele uygulanır.

### 3.4. Doğrulama

Uygulama sonuçlarının elde edilmesi sonucunda performans değerlendirme işlemleri için Açıklayıcılık Katsayısı ( $R^2$ ) Denklem (2)'de, Ortalama Hata Kare (Mean Square Error-MSE) Denklem (3)'te, Ortalama Hata Kare Kökü (Root Mean Square Error-RMSE) Denklem (4)'te ve Ortalama Mutlak Hata (Mean Absolute Error-MAE) Denklem (5)'de olduğu gibi hesaplanmıştır.

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

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

$$MAE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (5)$$

Burada  $R^2$ 'nin 1'e en yakın değeri alması beklenirken hataların değerlendirilmesinde kullanılan MSE, RMSE ve MAE için en düşük değeri veren tahmin başarılı sayılacaktır.

## 4. UYGULAMA

### 4.1. Veri ve Değişkenler

Pandemi öncesi yapılan uygulama için G20 ülkelerinin verileri 01/01/2010-31/12/2019 tarihleri arasında olmak üzere (investing.com) adresinden alınmıştır. Veriler alınırken tüm ülkeler için işleme açık olduğu günler için kapanış değerleri dikkate alınmıştır. Çalışmada regresyon modelleri kullanıldığından dolayı çoklu doğrusallık sorunu olmaması adına bağımsız değişkenler arasında mutlak değerce 0,80 ve daha büyük olan değişkenler çikartılmıştır. Buna göre çikartılan endeksler ve korelasyon matrisi Tablo 1'de gösterilmiştir.

**Tablo 1.** G20 Endeksleri Korelasyon Matrisi

Endeksler	Nifty50	CAC40	DAX	DowJones	Ftse100	IDX	MOEX	NASDAQ	S&P500	S&PMerval	S&PTSX	S.Africa	
Nifty50	1,00												
CAC40	<b>0,93</b>	1,00											
DAX	<b>0,93</b>	<b>0,96</b>	1,00										
DowJones	<b>0,97</b>	<b>0,92</b>	<b>0,93</b>	1,00									
Ftse100	<b>0,87</b>	<b>0,91</b>	<b>0,94</b>	<b>0,90</b>	1,00								
IDX	<b>0,92</b>	<b>0,84</b>	<b>0,91</b>	<b>0,94</b>	<b>0,91</b>	1,00							
MOEX	<b>0,89</b>	<b>0,82</b>	0,76	<b>0,89</b>	0,72	0,78	1,00						
NASDAQ	<b>0,98</b>	<b>0,93</b>	<b>0,93</b>	1,00	<b>0,89</b>	<b>0,93</b>	<b>0,90</b>	1,00					
S&P500	<b>0,93</b>	<b>0,96</b>	<b>0,92</b>	<b>0,93</b>	<b>0,92</b>	<b>0,87</b>	<b>0,81</b>	<b>0,93</b>	1,00				
S&PMerval	<b>0,98</b>	<b>0,93</b>	<b>0,95</b>	0,99	<b>0,91</b>	<b>0,94</b>	<b>0,87</b>	0,99	<b>0,94</b>	1,00			
S&PTSX	<b>0,96</b>	<b>0,88</b>	<b>0,86</b>	<b>0,96</b>	<b>0,82</b>	<b>0,88</b>	<b>0,93</b>	<b>0,97</b>	<b>0,87</b>	<b>0,95</b>	1,00		
S.Africa	<b>0,92</b>	<b>0,92</b>	<b>0,91</b>	<b>0,90</b>	<b>0,92</b>	<b>0,88</b>	<b>0,81</b>	<b>0,91</b>	<b>0,94</b>	<b>0,92</b>	<b>0,86</b>	1,00	
	<b>0,89</b>	<b>0,89</b>	<b>0,96</b>	<b>0,90</b>	<b>0,90</b>	<b>0,91</b>	0,69	<b>0,90</b>	<b>0,88</b>	<b>0,93</b>	<b>0,81</b>	<b>0,87</b>	1,00

Tablo 1 incelendiğinde Nifty50 (Hindistan), CAC 40 (Fransa), DAX (Almanya), Dow Jones (Amerika), Ftse100 (İngiltere), IDX (Endonezya), MOEX (Rusya), NASDAQ (Amerika), S&P\_ASX 200 (Avustralya), S&P500 (Amerika), S&P Merval (Arjantin), S&P\_TSX (Kanada) ve SouthAfrica (Güney Afrika) endeksleri korelasyon değerleri 0,80'den büyük olduğundan dolayı uygulamaya alınmamıştır. Sonuç olarak elde edilen 7 bağımsız değişken için endeksler Tablo 2'de gösterilmiştir.

**Tablo 2.** Bağımsız değişkenler

Borsa Adı	Ülke
Bovespa	Brezilya
İtaly40	İtalya
KOSPI	Güney Kore
Nikkei 225	Japonya
BMVIIPC	Meksika
Shanghai	Çin
Tadawul	Suudi Arabistan

Tablo 2'de yer alan bağımsız değişkenler için korelasyon matrisi Tablo 3'te gösterilmiştir.

**Tablo 3.** Korelasyon Matrisi

Ülkeler	Endeks Adı	Bovespa	İtaly40	KOSPI	Nikkei 225	BMVIIPC	Shanghai	Tadawul
Brezilya	Bovespa	1,00	0,38	0,42	0,46	0,14	0,16	0,10
İtalya	İtaly40	0,38	1,00	0,44	0,56	0,21	0,60	0,38
G. Kore	KOSPI	0,42	0,44	1,00	0,69	0,75	0,39	0,23
Japonya	Nikkei 225	0,46	0,56	0,69	1,00	0,77	0,54	0,45
Meksika	BMVIIPC	0,14	0,21	0,75	0,77	1,00	0,36	0,31
Çin	Shanghai	0,16	0,60	0,39	0,54	0,36	1,00	0,02
S. Arabistan	Tadawul	0,10	0,38	0,23	0,45	0,31	0,02	1,00

Tablo 3 incelendiğinde bağımsız değişkenlerin aralarındaki korelasyon değerleri 0.80'den düşük olduğundan dolayı analize dahil edilmiştir.

#### 4.2. Model

Knime programı kullanılarak elde edilen hibrit regresyon ve makine öğrenme modeli Şekil 6'da gösterilmiştir.

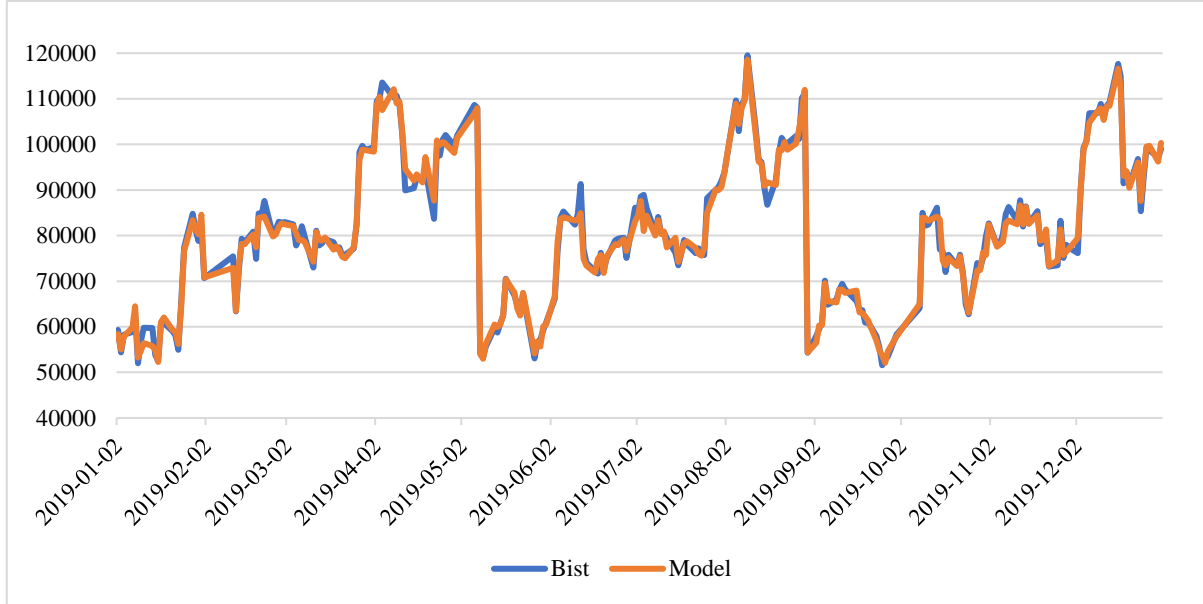


Tablo 3. incelendiğinde 30 bölüt için maksimum R<sup>2</sup> değeri ile minimum MAE, MSE ve RMSE değerleri elde edilmiştir. Elde edilen rasgele orman modelinin katsayılarının belirlendiği hibrit regresyon modeli için destek vektör regresyon modeli sonuçları Tablo 4’te gösterilmiştir.

**Tablo 4.** Hibrit Yöntem Destek Vektör Regresyon Modeli Sonuçları

Değişkenler	Katsayı	Standart Hata	t-değeri	P.
Bovespa	0.2501	0.013	19.843	<b>0.000</b>
İtaly40	0.0343	0.015	2.318	<b>0.021</b>
KOSPI	0.2450	0.017	14.411	<b>0.000</b>
Nikkei	0.3594	0.023	15.700	<b>0.000</b>
BMVIIPC	0.3130	0.023	13.464	<b>0.000</b>
Shanghai	-0.0711	0.013	-5.604	<b>0.000</b>
R-kare: 0.958		Düzeltilmiş R-kare: 0.958		
P. (F-statistic):0.00		Durbin-Watson:1.958		

Destek vektör regresyon modelinde Tadawul endeksi anlamlı olmadığından dolayı modelden çıkartılmıştır. Düzeltilmiş R<sup>2</sup> değeri %95,8 için bağımsız değişkenlerin BİST endeksine etkileri; Bovespa (0,2501), İtaly40 (0,0343), KOSPI (0,2450), Nikkei 225 (0,3594) ve BMVIIPC (0,3130) pozitif yönlü iken Shanghai (-0,0711) endeksinin ise negatif yönlü etkilediği belirlenmiştir. Hibrit yöntem kullanılmadan sadece destek vektör regresyon modelinin sonuçları ise; P. (F-statistic): 0.00 için anlamlı olmasıyla birlikte R-kare: 0.941, düzeltilmiş R-kare: 0.940 bulunmuştur. Elde edilen hibrit model kullanılarak 2019 yılı için BİST 100 ve model değeri Şekil 7’de gösterilmiştir.



**Şekil 7.** BİST 100 ve Tahmin Değerleri Grafiği

Sonuçlar incelendiğinde hibrit yapı kullanılarak bölütleme, rasgele orman regresyon yönteminin tek destek vektör modelinden daha iyi sonuçlar verdiği görülmüştür. Ayrıca rasgele orman yöntemi çıktılarının görselleştirilmesi sorunu destek vektör regresyon modeli kullanılarak ortadan kaldırılmıştır.

## 5. SONUÇ

Yatırımcıların yeni varlık alımına karar vermesi veya satın alınıp yönetilen varlıklardan çıkılması aşamasında belirleyici kıstaslardan biri, ilgilenilen varlıkların etkisi altında olduğu makroekonomik değişkenlerdir. Borsa

endeksinin etkilendiği çok fazla değişken olmasından dolayı endeksi etkileyen faktörlerin tespit edilebilmesi, zor olduğu kadar ihtiyaç duyulan bir konudur. Yatırımcılar risklerin azaltılması adına yatırımlarını yönlendirirken çeşitlendirme yolunu tercih ederler. Gelişen teknoloji yardımıyla çeşitlendirme sadece yerel piyasa ve yatırımlarla kalmayıp uluslararası düzeye taşınmıştır. Bu durum yatırımcılar için avantaj olarak görülse de pozitif etkilerin yanı sıra negatif etkilerin de tüm piyasalara kısa sürede yayılmasına da sebep olmaktadır. Ancak uzun dönemli analizler yatırımların yönlendirilmesi ve değerlendirilmesi açısından anlık etkilerden sıyrılarak etkinliğin artmasına sebep olabilir. Bundan dolayı çalışmada uzun dönemli etkiler araştırılarak anlık etkiler dışında BİST 100 endeksine etki eden ilişkili G20 endeksleri incelenmiştir.

Çalışmada rasgele orman, destek vektör ve bunların birleşim olan hibrit regresyon modelleri kullanılmış ve sonuçları karşılaştırılmıştır. 01/01/2010-31/12/2019 tarihleri arasında bağımlı değişken BİST 100 için bağımsız değişkenler Bovespa, İtalya40, KOSPI, Nikkei 225, BMVIIPC ve Shanghai endeksleri seçilmiştir. Bu seçimde regresyon modeli kullanıldığından çoklu bağlantı sorununa önlem olarak mutlak değerce korelasyon değeri 0,80 altı olan değişkenler kullanılmıştır. Uygulamalar veri madenciliği uygulamaları kullanılarak Knime programı ve python eklentisiyle gerçekleştirilmiştir. Veriler için 30 bölütleme için en yüksek  $R^2=0,9918$  ve en düşük  $MAE=0,0650$ ,  $MSE=0,0082$  ve  $RMSE=0,0903$  değerleri bulunmuştur. Ayrıca destek vektör regresyon modelinde “rbf” kernel yapısı seçilerek eğitim (%66) ve test (%33) verilerine ayrılarak analiz gerçekleştirilmiştir.

Analiz sonuçlarına göre destek vektör regresyon modeli için düzeltilmiş  $R^2$ : 0.940 iken bu değer hibrit yöntemde %95,8 bulunmuştur. Uygulanan yöntemin daha başarılı olduğu ve rasgele orman yönteminin görselleştirme sorununa da çözüm olabileceği görülmüştür. Brezilya-Bovespa, İtalya-İtalya40, Güney Kore-KOSPI, Japonya-Nikkei 225 ve Meksika-BMVIIPC endekslerinin BİST-100 endeksini pozitif yönlü olarak etkilediği belirlenmiştir. Elde edilen sonuçlar yardımıyla endeksleri pozitif yönlü olarak BİST-100 endeksine katkısı fazla olan ülkelerin sırasıyla Japonya, Meksika, Brezilya, Güney Kore olduğu görülmüştür. Çin-Shanghai endeksinin ise negatif yönlü etkilediği belirlenmiştir. Suudi Arabistan-Tadawul endeksinin ise BİST-100 üzerinde herhangi bir etkisinin olmadığı görülmüştür.

Yapılan çalışmada ilişkili ülkelerin endekslerinde meydana gelen artış ve azalışların BİST endeksine olan etkisi belirlenmiştir. Elde edilen sonuçların yatırımların yönlendirilmesinde karar vericilere yardımcı olacağı düşünülmektedir. Gelecek çalışmalarda farklı makine öğrenme yöntemleri ve modeller geliştirilerek oluşturulan farklı hibrit yöntemlerle literatüre katkı sağlanabileceği düşünülmektedir.

## YAZARIN BEYANI

**Katkı Oranı Beyanı:** Yazar, çalışmanın tümüne tek başına katkı sağlamıştır.

**Destek ve Teşekkür Beyanı:** Çalışmada herhangi bir kurum ya da kuruluştan destek alınmamıştır.

**Çatışma Beyanı:** Çalışmada herhangi bir potansiyel çıkar çatışması söz konusu değildir

## KAYNAKÇA

- Akman, M., Genç, Y. ve Ankaralı, H. (2011). Random forests yöntemi ve sağlık alanında bir uygulama. *Türkiye Klinikleri Biyoistatistik Dergisi*, 3(1), 36-48.
- Akpınar, H. (2014). *Data: Veri madenciliği veri analizi*. Papatya Yayıncılık Eğitim.
- Altay, E., & Satman, M. H. (2005). Stock market forecasting: artificial neural network and linear regression comparison in an emerging market. *Journal of Financial Management & Analysis*, 18(2), 18.
- Aygören, H., Sarıtaş, H. ve Morali, T. (2012). İMKB 100 endeksinin yapay sinir ağları ve Newton nümerik arama modelleri ile tahmini. *Journal of Alanya Faculty of Business/Alanya İisletme Fakültesi Dergisi*, 4(1), 73-88
- Breiman L. (2017). *Manual on setting up, using, and understanding random forests*. Erişim adresi: <http://www.stat.berkeley.edu/~breiman> (12.05.2020).
- Breiman, L., Friedman, J. H., Olshen, R. A. ve Stone, C. J. (1984). *Classification and regression trees*. Monterey, Calif., USA: Wadsworth.

- Caba, N. (2017). Finansal kaldıraç ve firma büyüklüğünün finansal performans üzerine etkisi: BİST sınai endeksinde işlem gören işletmeler üzerine bir uygulama. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 13(5), 796-811.
- Cosgun, E. ve Karaağaoğlu, E. (2011). Veri madenciliği yöntemleriyle mikrodizilim gen ifade analizi. *Hacettepe Tıp Dergisi*, 42, 180-189.
- Diler, A. İ. (2003). İMKB ulusal-100 endeksinin yönünün yapay sinir ağları hata geriye yayma yöntemi ile tahmin edilmesi. *İstanbul Menkul Kıymetler Borsası Dergisi*, 7(25-26), 65-82.
- Filiz, E., Karaboğa, H. A. ve Akoğul, S. (2017). BİST-50 endeksi değişim değerlerinin sınıflandırılmasında makine öğrenmesi yöntemleri ve yapay sinir ağları kullanımı. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 26(1), 231-241.
- Güney, S. ve Ilgin, K. S. (2019). Yatırım araçlarının BİST-100 endeksi üzerindeki etkisinin değerlendirilmesi. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (53), 226-245.
- Güngör, B. ve Polat, A. (2020). Geleneksel yatırım araçlarının hisse senedi fiyatlarına etkisi: BİST'te sektörel bazda karşılaştırmalı bir analiz. *Bingöl Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(1), 79-105.
- Investing.com. (2020). *Kripto paralar*. Erişim adresi: <https://tr.investing.com/crypto/currencies>, (06.07.2020).
- Kara, İ. (2019). *Borsa endeksi hareket yönünün tahmininde sınıflandırma yöntemlerinin performanslarının karşılaştırılması: BİST 100 örneği*. Doktora Tezi, Afyon Kocatepe Üniversitesi Sosyal Bilimler Enstitüsü, Afyonkarahisar.
- Kara, Y., Boyacioglu, M. A. ve 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.
- Karaatlı, M., Güngör, İ., Demir, Y. ve Kalaycı, Ş. (2005). Hisse senedi fiyat hareketlerinin yapay sinir ağları yöntemi ile tahmin edilmesi. *Yönetim ve Ekonomi Araştırmaları Dergisi*, 3(3), 38-48.
- Lin, C. F. ve Wang, S. D. (2002). Fuzzy support vector machines. *IEEE transactions on neural networks*, 13(2), 464-471.
- Özer A., Sarı, S. ve Başakın, E. E. (2017). Bulanık mantık ve yapay sinir ağları ile borsa endeksi tahmini: gelişmiş ve gelişmekte olan ülkeler örneği. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(1), 99-124.
- Özmerdivanlı, A. (2014). Petrol fiyatları ile BİST 100 endeksi kapanış fiyatları arasındaki ilişki. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi*, (43).
- Pabuçcu, H. (2019). Borsa endeksi hareketlerinin makine öğrenme algoritmaları ile tahmini. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (23), 179-190.
- Panigrahi, S. S. ve Mantri, J. K. (2015). Epsilon-SVR and decision tree for stock market forecasting. *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, IEEE, 761-766.
- Rygielski, C., Wang, J. C. ve Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in society*, 24(4), 483-502.
- Şimşek, M. (2016). Borsa İstanbul (BİST) ve Brics ülkelerinin hisse senedi piyasalarının ilişkisi üzerine bir inceleme. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, 5(3), 520-536.
- Tekin, B. ve Hatipoğlu, M. (2017). VIX endeksi, döviz kuru ve petrol fiyatlarının BİST 100 endeksi üzerindeki etkileri: Bir kuantil regresyon yaklaşımı. *ODÜ Sosyal Bilimler Araştırmaları Dergisi (ODÜSOBİAD)*, 7(3), 627-634.
- Tunç, A. ve Ülger, İ. (2016). Veri madenciliği uygulamalarında özellik seçimi için finansal değerlere binning ve five number summary metotları ile normalizasyon işleminin uygulanması. *XVIII. Akademik Bilişim Konferansı, Adnan Menderes Üniversitesi*, 30.

Türk Dil Kurumu. (2021). *Sözlükler*. Erişim adresi: <https://sozluk.gov.tr/>, (22.02.2021)

Yıldız, A. ve Aksoy, E. (2014). Morgan stanley gelişmekte olan borsa endeksi ile BİST endeksi arasındaki eşbütünleşme ilişkisinin analiz edilmesi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 28(1), 1-23.

Yıldız, B., Yalama, A. ve Coşkun, M. (2008). Forecasting the İstanbul stock exchange national 100 index using an artificial neural network. *An International Journal of Science, Engineering and Technology*, (46), 36-39.

Yu, P. S., Chen, S. T. ve Chang, I. F. (2006). Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, 328(3-4), 704-716.