

---

## BORSA ENDEKSİ HAREKETLERİNİN MAKİNE ÖĞRENME ALGORİTMALARI İLE TAHMİNİ<sup>1</sup>

---

Hakan PABUÇCU<sup>2</sup>

### Öz

Finansal zaman serilerinin barındırdığı belirsizlik, kaotik hareketler yanında doğrusal olmayan dinamik yapı, tahminleri oldukça güçleştirmektedir. Borsa endekslerinin politik değişimler, ekonominin genel görünümü, yatırımcıların beklenti ve yatırım tercihleri ve diğer endekslerin hareketleri gibi birçok makroekonomik faktörden etkilenmeleri, endeks tahminlerini oldukça zor ancak bir o kadar da çekici kılmaktadır. Borsa endeksi hareketleri ve geleceğe dönük tahminler üretmede makine öğrenme algoritmalarının başarılı oldukları bilinmektedir. Bu çalışmada BIST 100 endeksi hareketlerinin yönünün tahmin edilmesi problemi ele alınmıştır. Üç farklı makine öğrenme algoritması olan yapay sinir ağları, destek vektör makineleri ve naive Bayes sınıflandırıcı algoritması kullanılmış ve performansları karşılaştırılmıştır. Borsa endeksi tahminleri için kullanılan on teknik gösterge modeller için girdi olarak kullanılmıştır. Veri seti 2009-2018 periyodunu kapsayan günlük kapanış değerlerini içermektedir. Analiz sonuçları, her üç modelin de borsa endeks hareketlerini yakalamada kullanılabilir olduğunu, yapay sinir ağı algoritmasının ise daha iyi bir sınıflandırıcı olduğunu göstermiştir.

**Anahtar Kelimeler:** BIST100, yapay sinir ağları, destek vektör makineleri, naive Bayes, makine öğrenme

**JEL Sınıflandırması:** C44, C45

---

## PREDICTING STOCK MARKET MOVEMENT BY USING MACHINELEARNING ALGORITHM

---

### Abstract

In addition to the uncertainty and chaotic movements of the financial time series, the nonlinear dynamic structure makes the forecasts very difficult. The fact that the stock market index are affected by the political changes, the general outlook of the economy, the investors' expectations and investment preferences, and the movements of other indexes, make the index estimates quite difficult but attractive. It is known that the machine learning algorithms are successful in estimating stock index movements and their future values. In this study, the problem of forecasting the direction of BIST 100 index movements is discussed. Three different machine learning algorithms, artificial neural networks, support vector machines and naive Bayes classifier were used and their performances were compared. Ten technical indicators were used as inputs for the models. The data set consists of ten-year daily closing price values covering the 2009-2018 period. Analysis results show that the models can be used to capture stock market index movements, whereas artificial neural network algorithm is a better classifier.

**Keywords:** BIST 100, artificial neural network, support vector machine, naive Bayes classifier, machine learning

**JEL Classification:** C44, C45

---

<sup>1</sup> Bu çalışma 19. Ekonometri, Yöneylem Araştırması ve İstatistik sempozyumunda sunulan bildirinin gözden geçirilmiş genişletilmiş halidir

<sup>2</sup> Bayburt Üniversitesi İİBF İşletme Bölümü, Bayburt/Türkiye hpabuccu@bayburt.edu.tr, orcid.org/0000-0003-2267-5175.

## 1. Giriş

Zaman serileri analizlerinde engellerle karşılaşılan, en çok zorlanılan uygulamaların başında borsa endeksi ve endeks hareketlerinin tahmin edilmesi gelmektedir. Yatırımcıların hisse senetlerine yatırım yapmadan önce faydalandıkları temel ve teknik analiz olmak üzere iki tür analiz vardır. Temel analiz ile yatırımcılar hisse senedinin gerçek değeri, politik iklim, sanayi ve ekonominin performansı gibi göstergeleri dikkate alarak yatırım yapıp yapmayacaklarına karar verirler. Teknik analizde ise, geçmiş değerler ve işlem hacimleri gibi piyasa hareketleri ile oluşturulan istatistiklerin kullanılarak hisse senetlerinin değerlendirilmesi sağlanır. Teknik analistler, bir hisse senedinin gerçek değerini ölçmeyi denemezler, bunun yerine gelecekte nasıl davranacağını gösterebilecek kalıpları ve eğilimleri tanımlamak için hisse senedi grafiklerini kullanırlar. Malkiel ve Fama (1970) tarafından ileri sürülen etkin piyasalar hipotezine göre, hisse senedi fiyatlarının bilgi açısından verimli olduğu, bunun da alım satım verilerine dayanarak hisse senedi fiyatlarının tahmin edilebileceği anlamına gelmektedir. Buradan hareketle borsa endeksi hareketleri için güvenli ve geçerli tahminlerin üretilmesi etkin pazar stratejilerinin geliştirilmesi için ciddi önem arz etmektedir (Leung vd., 2000). Böylece yatırımcılar potansiyel piyasa risklerine ve spekülörlere karşı korunma fırsatına sahip olurlar.

Finansal zaman serilerinin dinamik, doğrusal olmayan, karmaşık ve kaotik yapısı doğru tahminler üretme noktasında engeller oluşturmaktadır (Abu-Mostafa ve Atiya, 1996). Ayrıca, borsa endeksleri politik değişimler, ekonominin genel görünümü, yatırımcıların beklenti ve yatırım tercihleri ve diğer endekslerin hareketleri gibi birçok makroekonomik faktörden etkilenmektedir (Tan vd., 2007). Finansal zaman serilerinin geleneksel yöntemlerle analiz edilmesi yanında, yapay zekâ ve makine öğrenme algoritmaları ile analizi, son yıllarda oldukça popüler olmuştur. Buradan hareketle, BIST 100 endeksinin hareketlerinin yönünün tahmin edilmesi ve kullanılacak olan makine öğrenme algoritmalarının performanslarının karşılaştırılması bu çalışmanın temel hedefidir. Tahmin için kullanılacak ve performansları karşılaştırılacak olan modeller makine öğrenme algoritmalarından yapay sinir ağları (YSA), destek vektör makineleri (DVM) ve naive Bayes (NB) sınıflandırıcı algoritmasıdır. Modeller için girdi olarak on teknik gösterge seçilmiştir.

Finansal piyasalarda doğrusal olmayan ilişkinin var olduğu ve bu ilişkiyi ortaya çıkarmak için sinir ağları gibi makine öğrenme yöntemlerinin etkin bir şekilde kullanılabilirdiği birçok çalışma tarafından geniş çapta kabul görmüştür. Bu çalışmanın da makine öğrenme teknikleri kullanılarak finansal zaman serileri tahminlerine katkı sağlayacağı değerli sonuçlar üreteceği düşünülmektedir. Finansal zaman serileri tahminleri için gerçekleştirilmiş bazı seçilmiş çalışmaları ikinci bölümde aşağıdaki şekilde özetlemek mümkündür.

Çalışma beş bölümden oluşmaktadır. İlk bölümde teorik olarak ana temanın incelendiği giriş, ikinci bölümde ilgili literatür araştırması, üçüncü bölümde veri seti ve kullanılan yöntem sunulmaktadır. Dördüncü bölüm araştırmanın bulgularını beşinci bölüm ise sonuç ve önerileri içermektedir.

## 2. Literatür Özeti

Borsa tahmini için literatürde çok fazla araştırma bulunmakla birlikte, bunların önemli bir bölümü gelişmiş ülkeler için yapılan çalışmalardan oluşmaktadır. Bu çalışmalarda, hisse senedi getirilerinin yanında borsa endeksi hareketleri yönünün tahmini çok çeşitli yöntemlerle tahmin edilmeye çalışılmıştır. Bu yöntemlerin başında YSA ve DVM gelmektedir. YSA, Kara vd. (2011); Kimoto vd. (1990); Olson ve Mossman (2003) gibi çalışmalarla, DVM ise Hsu vd. (2009); Huang vd. (2005); Kumar ve Thenmozhi (2005) gibi çalışmalarla borsa tahminlerinde güvenilir sonuçlar üretme noktasında başarısını ispatlamıştır. Borsa hareketlerinin tahmini için literatürdeki seçilmiş çalışmaları şu şekilde özetlemek mümkündür:

Finansal zaman serilerinin tahmin edilmesi için önemli kaynaklardan biri olan Cao ve Tay (2001) destek vektör makineleri ve geri beslemeli sinir ağları tahminlerinin performanslarını karşılaştırmaktadır. Çalışma DVM'ni finansal zaman serilerinde uygulayan ilk çalışmalardan olması

nedeniyle önemlidir. DVM'nin YSA modeline kıyasla, gerek parametre seçimi gerekse tahmin hata oranları bakımından daha uygulanabilir ve iyi sonuçlar ürettiği belirtilmiştir. Tay ve Cao (2001)'de DVM ile elde edilen finansal zaman serileri tahminlerinin gücünün artırılması için öz düzenleyici haritalar yardımıyla iki aşamalı bir yöntem önerilmiş ve tek başına DVM tahminlerine oranla daha güçlü tahminler üretilmiştir. Chen vd. (2003)'de gelişmekte olan Asya ülkelerinin en hızlı büyüyen finansal borsalarından biri olan Tayvan Borsası'nın piyasa endeksindeki getiri yönünü tahmin etmişlerdir. Fiyat hareketinin yönüne dair tahminlerin yönlendirdiği ticaret stratejilerinin daha etkili olabileceği ve daha yüksek karlara yol açabileceği fikrinden hareketle yapılan bu çalışmada, olasılıksal sinir ağı tarihsel verilerle eğitildikten sonra endeks hareket yönünü tahmin etmek için kullanılmıştır. Sinir ağı tahminlerinin istatistiksel performansı ölçülmüştür. Ampirik sonuçlar, olasılıksal sinir ağı tabanlı yatırım stratejilerinin, geleneksel yatırım stratejilerinden daha yüksek getiri elde ettiğini göstermiştir. Kim (2003)'de ise DVM, YSA ve durum temelli muhakeme (case-based reasoning) yöntemlerinin finansal zaman serileri tahminleri için karşılaştırmalı analizi yapılmıştır. DVM'nin en iyi tahmin yöntemi olduğu belirtilmiştir. Thawornwong ve Enke (2004) çalışmalarında borsa endekslerinin yönünü farklı tahmin modelleri ve farklı değişken kombinasyonları deneyerek araştırmışlardır. Tahmin modellerine adım adım eklenen değişkenlerin tahminler üzerinde iyileştirme yapılıp yapılmadıklarını izlemişlerdir. Sonuçlar sinir ağı modellerinin geleneksel tahmin yöntemlerine göre daha düşük riskli ve yüksek getirili tahminler ürettiğini ortaya koymuştur. Enke ve Thawornwong (2005) çok sayıda finansal ve ekonomik değişkenin yordama kapasitelerini değerlendirmek için veri madenciliği ve makine öğrenme yöntemlerini kullanmışlardır. Sınıflandırma için sinir ağı modelleri, gelecekteki değerlerin etkili bir tahminini sağlama yetenekleri açısından incelenmiştir. Çeşitli modellerin genelleme yeteneğini geliştirmek için çapraz doğrulama tekniği kullanılmıştır. Sonuçlar, sınıflandırma modelleri tarafından yönlendirilen ticaret stratejilerinin, satın alma ve tutma stratejisinden daha yüksek riskli karlar sağladığını göstermiştir. Kara vd. (2011)'de finansal zaman serilerinin tahmininde karşılaşılan zorluklardan bahsedilmiş, YSA ve DVM tahmin performansları karşılaştırılmıştır. Örneklem olarak İstanbul borsası ulusal 100 endeksi verileri kullanılmıştır. Deneysel sonuçlar ise YSA performansının %75,7 DVM performansının ise %71,52 olduğunu göstermiştir. Patel vd. (2015a) Hindistan borsası günlük kapanış verileri ve bu verileri kullanarak kategorik hale getirdikleri yeni veri setinin davranışını DVM, YSA, NB ve random forest algoritmalarını kullanarak araştırmışlardır. Random forest algoritması en iyi tahmin modeli olarak belirlenmiştir. Yine, Hint borsalarına ait iki endeksin on yıllık veri seti kullanılarak Patel vd. (2015b) tarafından yapılan çalışmada, iki aşamalı bir endeks tahmin modeli önerilmiştir. İlk aşamada destek vektör regresyonu ikinci aşamada ise yapay sinir ağı ve random forest algoritmaları kullanılmış ve sonuçlar tartışılmıştır. Zhong ve Enke (2017)'de Standard & Poors 500 endeksinin günlük hareketlerinin açıklanması ve tahmin edilebilmesi için 60 ekonomik ve finansal gösterge kullanmışlardır. Temel bileşenler analizi gibi üç farklı yöntem ile göstergeleri azaltıp boyutlara indirgemişler ve YSA ile tahminler yapmışlardır. Farklı yöntem kombinasyonları ile elde ettikleri sonuçları karşılaştırmışlardır. DVM ve veri zarflama analizi ile Çin'de faaliyet gösteren bankalara ait finansal veriler kullanılarak banka etkinliklerinin değerlendirildiği Chen vd. (2018)'de bankaların önemli bir kısmının istenilen etkinlik seviyesine ulaşamadığı belirlenmiştir.

Literatürde çok sayıda çalışma makine öğrenme yöntemlerinin finansal zaman serileri tahminlerinde kullanıldığını göstermektedir. Bu çalışmalarla ilgili kapsamlı incelemeler Atsalakis ve Valavanis (2009); Chun ve Kim (2004) ve Vanstone ve Finnie (2009) çalışmalarında yer almaktadır.

Literatür araştırması ile Türkiye için güncel veri setleri ve makine öğrenme algoritmaları kullanılarak borsa tahminine ilişkin Kara vd. (2011) ve diğer birkaç çalışma dışında önemli sayıda çalışmaya rastlanmamıştır. Ayrıca; YSA, DVM ve NB sınıflandırma algoritmalarının birlikte kullanılması ile performansları karşılaştırma imkânı doğmaktadır. Bu nedenlerden dolayı, bu çalışmanın amacı Türkiye için en güncel borsa verisi kullanılarak hem yatırımcıya yol göstermek hem de makine öğrenme yöntemlerinin finansal zaman serisi tahmin performanslarını değerlendirmektir.

### 3. Veri Seti ve Yöntem

Çalışmada BIST 100 endeksinin negatif ve pozitif yönlü hareketlerinin tahmin edilebilmesi için üç farklı makine öğrenme algoritması kullanılmıştır. Bu algoritmalara girdi olarak Kara vd. (2011)'de kullanılan on teknik gösterge girdi olarak seçilmiştir. Teknik göstergeler endeks kapanış fiyatları kullanılarak yazar tarafından ayrı ayrı hesaplanmıştır. Borsa endeksi verisi ise 2009-2018 Aralığını kapsayan günlük kapanış fiyatlarından oluşmaktadır. Veri seti T.C. Merkez Bankasından alınmıştır. Veri analizinde ilişkilerin öğrenildiği eğitim aşaması için 10 katlı çapraz doğrulama (Ten-fold cross validation) yöntemi kullanılmıştır. Model seçimi ve performansların karşılaştırılabilmesi için çok sayıda istatistik kullanılmıştır. Yöntemlerin daha iyi tahminler üretebilmesi için eğitim sürecinde her bir yöntemin kendine özgü parametreleri değiştirilerek en iyi tahmin modeli araştırılmıştır. Kullanılan teknik göstergeler için şu tanımlamaları yapmak mümkündür.

**Hareketli ortalamalar (MA-WMA):** Hareketli ortalama ve ağırlıklı hareketli ortalama ilgili ürünün belli zaman aralığındaki fiyat değişimlerinin ortalamasını ifade eder. Hareketli ortalamalar eğilimin genel yönünü gösterir. Teknik analizde çok sık kullanılan en basit göstergedir. Bu çalışmada 14 günlük hareketli ortalamalar kullanılmıştır.

**Momentum:** Fiyat değişimlerinin ne kadar güçlü olduğu ve var olan trendin ne kadar devam edeceği hakkında bilgi veren göstergedir.

**Stokastik %K-%D ve Larry Williams %R:** Bu göstergeler trend içermeyen stokastik göstergelerdir. Seçilen bir dönem için kapanış fiyatlarını günlük en yüksek ve en düşük seviyelere göre ölçen, fiyatların dönüm noktalarını belirleyen göstergelerdir.

**Göreceli Güç Endeksi (RSI):** Bir momentum çizgisinin belirlenmesindeki en temel sorun, belirli bir zaman dilimindeki bazı değerlerde görülen keskin artış ya da düşüşlerdir. Bu keskin artış ya da düşüşler momentum çizgisinin ani değişimine neden olabilir. RSI bu değişimleri en aza indirecek düzleştirme işlemini gerçekleştirir.

Tablo 1: Seçilen Teknik Göstergeler ve Formülleri

Göstergeler	Formül
Basit 14 günlük hareketli ortalama (MA)	$C_t + C_{t-1} + \dots + C_{t-14} / 14$
Ağırlıklı 14 günlük hareketli ortalama (WMA)	$\frac{((n) * C_t + (n-1) * C_{t-1} + \dots + C_{t-14})}{(n + (n-1) + \dots + 1)}$
Momentum (Mom)	$C_t - C_{t-n}$
Stokastik K% (K%)	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} * 100$
Stokastik D% (D%)	$\frac{\sum_{i=0}^{n-1} K_{t-i} \% / n}{100}$
Göreceli güç indeksi (RSI)	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} Up_{t-i} / n) / (\sum_{i=0}^{n-1} Dw_{t-i} / n)}$
MACD(Hareketli Ortalamaların Uyumu/Uyumsuzluğu)	$MACD(n)_{t-1} + \frac{2}{n+1} * (DIFF_t - MACD(n)_{t-1})$
Larry William's R% (LW)	$\frac{H_n - C_t}{H_n - L_n} * 100$
A/D (Birikim/Dağılım göstergesi)	$\frac{H_t - C_{t-1}}{H_t - L_t}$
CCI (Emtia kanal endeksi)	$\frac{M_t - SM_t}{0.015D_t}$
Kaynak: Kara vd. (2011).	

\* $C_t$  kapanış fiyatı,  $L_t$  en düşük fiyat ve  $H_t$  en yüksek fiyat.  $DIFF_t$ :  $EMA(12)_t - EMA(26)_t$ .  $EMA$  Üstel hareketli ortalama,  $EMA(k)_t$ :  $EMA(k)_{t-1} + \alpha * (C_t - EMA(k)_{t-1})$ ,  $\alpha$  düzeltme faktörü.  $LL_t$  son t gün için minimumların minimumu,  $HH_t$  son t gün için maksimumların maksimum değeri.  $M_t = (H_t + L_t + C_t) / 3$ ,

$SM_t = (\sum_{i=0}^n M_{t-i+1}/n)$ ,  $D_t = (\sum_{i=1}^n |M_{t-i+1} - SM_t|/n)$ ,  $Up_t$  ve  $Dw_t$  sırasıyla  $t$  zamanındaki yukarı ve aşağı yönlü fiyat değişimlerini ifade etmektedir.

Hareketli Ortalamaların Uyumu/Uyumsuzluğu (MACD): İki farklı hareketli ortalama arasındaki ilişkiyi gösteren ve piyasa katılımcılarının sık kullandığı orta vadeli göstergedir. Aşırı alım satım bölgelerinin belirlenmesini sağlar.

Birikim/Dağılım göstergesi (A/D endeksi): Fiyatlardaki değişimin ne kadar yüksek işlem hacmiyle gerçekleşirse o kadar güçlü olacağından hareketle bir trendin var olup olmadığını ve var olan trendin sürüp sürmeyeceğini araştırır.

Emtia kanal endeksi (CCI): Fiyatların istatistiksel ortalamadan ne ölçüde saptığının tespiti için kullanılan bir göstergedir. Tablo 1’de on teknik gösterge ve formülleri yer almaktadır.

### 3.1. Yapay Sinir Ağları

Yapay sinir ağları biyolojik sinir hücresinin bir simülasyonudur. Girdi ve çıktı birimleri arasındaki ilişkilerin tespiti ve tahmini için bazı aktivasyon fonksiyonlarını kullanır. YSA, finansal modellemede sıklıkla kullanılan başarılı bir tahmin modelidir. Borsa endeksi hareket yönlerinin tahmin edilmesi için bu çalışmada üç katmanlı ileri beslemeli model kullanılmıştır. Girdi katmanında on sinir hücresi, gizli katmanda altı sinir hücresi ve çıktı katmanında iki sinir hücresi kullanılmıştır. Girdi katmanındaki her sinir hücresi bir teknik göstergiyi temsil etmektedir. Çıktı katmanında ise iki adet sinir hücresi  $\{0-1\}$  ile etiketlenmiştir. YSA parametreleri ve mimarisi deneysel olarak belirlenmiştir. Tablo 2’de YSA için kullanılan parametreler ve kullanım düzeyleri sunulmaktadır. YSA ile ilgili finansal tahminler konusunda ayrıntılı bilgi için aşağıdaki kaynaktan<sup>1</sup> faydalanılabilir.

Tablo 1: Eğitim sürecinde kullanılan parametre ayarları

Parametre	Kullanılan düzey
Gizli katman nöron sayısı (n)	5,...,50
İterasyon (ep)	250, 500,...,2000
Momentum sabiti (mc)	0.1, 0.2,...,0.9
Öğrenme oranı (lr)	0.1, 0.2, 0.3

### 3.2. Destek Vektör Makineleri

Sınıflandırma, karakter tanıma, regresyon ve zaman serileri gibi uygulama alanları olan bir algoritmadır. Sınıflandırmada yapısal risk minimizasyonuna dayalı olarak Vapnik (1995) tarafından geliştirilmiş çok güçlü bir algoritmadır. Temel fikri en uzun sınıflara sahip bir hiper düzlem tanımlamak, pozitif ve negatif örnekler arasındaki ayrımın marjının maksimize edilmesidir. DVM, Vapnik, (1995)’de ikili sınıflandırmalarda yapısal risk minimizasyonu ilkesinin bir uygulaması olarak ortaya çıkmıştır. DVM algoritması istatistiksel öğrenme kuramında genellemenin ve yapısal karmaşıklık ile deneysel risk arasındaki değiş tokuşun nasıl yapılacağı çalışmalarından doğmuştur. DVM, noktaları desen uzayında ya da daha yüksek boyutlu bir uzayda iki ayrık yarım uzaya atayarak sınıflandırır (Khemchandani vd., 2009). DVM’nin ana fikri, pozitif ve negatif örnekler arasındaki ayrım marjının maksimize edileceği şekilde karar yüzeyi olarak bir hiper düzlem oluşturmaktır (Xu vd., 2009). Örneklerin bir eğitim kümesi için,  $x_i \in R^d$  girdi vektörleri ve karşılığı olan  $y_i \in \{1, -1\}$  görüntü kümesi ile nesnelere iki sınıfa dâhil ederek nasıl sınıflandıracağını öğrenir. İki sınıflı bir sınıflandırma problemi için;  $x_i \in R^d (i = 1, 2, \dots, N)$  girdi vektörleri ve  $y_i \in \{1, -1\} (i = 1, 2, \dots, N)$  karşılık kümesi olduğu varsayalım. Burada (1) ve (-1) iki sınıfı ifade etsin. Burada amaç ikili bir sınıflandırıcı oluşturmak veya mevcut örneklerden, daha önce görülmemiş bir örneklerdeki birimleri yanlış sınıflandırma olasılığı düşük olan bir karar fonksiyonu türetmektir. DVM, giriş vektörlerini  $x_i \in R^d$  yüksek boyutlu bir özellik alanına  $\Phi(x_i) \in H$  haritalar ve en uygun ayırma

<sup>1</sup> Patel, J., Shah, S., Thakkar, P., ve Kotecha, K. (2015a). Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259–268.

hiper düzlemini oluşturur. Hiper düzlem ile  $H$  uzayındaki her bir sınıfın en yakın veri noktaları arasındaki mesafe maksimize edilir.  $\Phi$  Haritalaması  $H$  iç çarpım uzayında tanımlanan  $K(x_i, x_j)$  çekirdek fonksiyonları tarafından gerçekleştirilir. Elde edilen sınıflandırıcı  $\alpha_i$  katsayılarını belirlemek için eşitlik 1 de verilen karar fonksiyonu ve kuadratik programlama problemine dayalıdır (Hua ve Sun, 2001).

$$f(x) = \text{sgn}(\sum_{i=1}^N y_i \alpha_i \cdot K(x, x_i) + b) \quad (1)$$

Kuadratik problem ise Eşitlik 2'deki gibi tanımlanır:

$$\begin{aligned} \text{Maks. } \sum_{i=1}^N \alpha_i - 1/2 \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \cdot y_i y_j \cdot K(x, x_i), 0 \leq \alpha_i \leq c \\ \sum_{i=1}^N \alpha_i y_i = 0 \quad i = 1, 2, \dots, N \end{aligned} \quad (2)$$

Burada  $c$ ; marjin ile sınıflandırma hataları arasındaki dengeyi kontrol eden düzenleme parametresidir. DVM'de yaygın kullanılan polinomial ve radyal tabanlı olmak üzere iki adet çekirdek fonksiyon vardır.

$$\text{Polinomial fonksiyon: } K(x_i, x_j) = (x_i \blacksquare x_j + 1)^d \quad (3)$$

$$\text{Radyal tabanlı fonksiyon: } K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (4)$$

Tahminin başarısı için çekirdek fonksiyonu seçimi oldukça önemlidir. İlgili fonksiyonların parametrelerinin belirlenmesi belli bir yöntem olmamakla birlikte çok sayıda deneme yapılarak en iyi parametre setleri belirlenmektedir. Bu çalışmada kullanılan DVM modeline ilişkin çekirdek fonksiyonu ve parametre kümeleri Tablo 3'te sunulmaktadır.

Tablo 2: Eğitim sürecinde kullanılan parametre ayarları

Parametre	Düzyey (polinomial)	Düzyey (radial tabanlı)
Çekirdek fonksiyonu derecesi ( $d$ )	1, 2, 3, 4	
Çekirdek fonksiyonu Gamma Katsayısı ( $\gamma$ )		0, 0.1, 0.2, ..., 5.0
Düzenleme parametresi ( $c$ )	1, 10, 100	1, 10, 100

### 3.3. Naive Bayes

Naive-Bayes sınıflandırma algoritması şartlı olasılık kuralları üzerine inşa edilmiş makine öğrenme algoritmalarından bir tanesidir. Hesaplama ve kullanım kolaylığı ve dolayısı ile diğer makine öğrenme yöntemlerine göre uygulama için daha az zamana ihtiyaç duyması yönünden diğer yöntemlere göre üstünlük gösterir. Eğitim için ayrılan veri seti üzerinden örneklerin hangi sınıflara dâhil olduğunu öğrenir ve test verisi içindeki gözlemler için sınıf tahminlerini gerçekleştirir. Bu süreçte olasılıkların tahmini için Bayes teoremi kullanılır. Bayes teoremi sonsal (posterior)  $P(C|X)$  olasılıkların,  $P(C)$ ,  $P(X|C)$  ve  $P(X)$  olasılıkları kullanılarak hesaplanmasını sağlar. Bayes teoremi şu şekilde ifade edilebilir (Eşitlik 5).

$$P(C|X) = \frac{P(C)P(X|C)}{P(X)} \quad (5)$$

Burada;  $P(C|X)$ ,  $X$  olayı gerçekleştiğinde  $C$  olayının meydana gelmesi için hesaplanan sonsal olasılıktır. Bu çalışmada  $P(C)$ , borsa hareketleri (yükseliş/düşüş) sınıfına dâhil olma olasılığı ve  $X$  ise test verisini ifade eder.  $P(X|C)$   $C$  olayının gerçekleşmesi durumunda  $X$  olayının gerçekleşmesinin koşullu olasılığını gösterir. Şartlı olasılık eğitim verisi kullanılarak tahmin edilir. Naive Bayes sınıflandırıcı algoritmasının işleyişi şu şekilde özetlenebilir (Patel et al., 2015a):

$C_1, C_2, \dots, C_m$  olmak üzere  $m$  sınıf ve test verisinin gerçekleşme olayı  $X$  verilsin. Bayes sınıflandırıcı, Bayes teoremini (Eşitlik 6) kullanarak test verilerini en yüksek olasılıklı bir sınıfa sınıflandırır.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (6)$$

$A_1, A_2, \dots, A_m$  gibi çok özellikli bir veri seti için  $P(X|C_i)$  olasılıklarını hesaplamak oldukça güç ve zahmetli bir iş olacaktır. Hesaplamaları kolaylaştırmak için sınıf koşullu bağımsızlık varsayımı (Naive assumption) yapılır. Bu varsayım, özelliklerin değerlerinin diğerlerinden koşullu olarak bağımsız olduklarını ifade eder. Yani, özellikler arasında bağımlılık ilişkisi yoktur (Eşitlik 7).

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) = P(x_1|C_i) * P(x_2|C_i) * \dots * P(x_n|C_i) \quad (7)$$

Burada,  $x_k$   $X$  olayı için  $A_k$  özelliğinin aldığı değerdir.  $P(x_k|C_i)$ 'nin hesaplanması kategorik veya sürekli olmasına göre değişiklik gösterir. Eğer  $A_k$  kategorik ise,  $P(x_k|C_i)$ ,  $A_k$  özelliği için eğitim veri setinde  $x_k$  değerini alan  $C_i$  sınıfının gözlem sayısının, eğitim veri setindeki  $C_i$  sınıfının tüm gözlem sayısına bölümüdür. Eğer  $A_k$  sürekli ise, Gauss dağılım fonksiyonu veri setine ayarlanır ve  $P(x_k|C_i)$  Eşitlik 8 ile hesaplanır.

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (8)$$

$$P(x_k|C_i) = f(x_k, \mu_{C_i}, \sigma_{C_i}) \quad (9)$$

Burada,  $\mu_{C_i}, \sigma_{C_i}$  sırasıyla ortalama ve standart sapmayı ifade etmektedir.  $X$  Gözleminin sınıfı  $C_i$  olarak tahmin edilir.

#### 4. Ampirik Bulgular

Çalışmanın bu bölümünde YSA, DVM ve NB modellerine ilişkin yapı, parametre ve ulaşılan tahmin sonuçları sunulacaktır. Tablo 4'de modeller için girdi setlerini oluşturan teknik göstergelere ait bazı tanımlayıcı istatistikler sunulmaktadır.

Tablo 3: Göstergelere ilişkin özet istatistikler

Gösterge	Minimum	Maksimum	Ortalama	Standart sapma
MA	34706.677	118348.507	74514.625	17339.498
WMA	34482.403	119403.421	74601.188	17314.628
Mom	-16792.68	10920.66	318.826	3671.642
K%	14.08	80.193	51.864	14.501
D%	4.023	74.37	51.742	10.755
RSI	18.016	88.366	53.443	12.617
MACD	-3586.069	2491.367	184.149	1137.383
LW	-85.92	-19.807	-48.136	14.501
A/D	-0.67	1.166	0.503	0.144
CCI	-3356314565	16672664177	-7233062.423	297394750.298

Modellerin başarılarını test etmek ve birbirleriyle karşılaştırabilmek için sınıflandırma başarıları yanında, F istatistiği ve ROC istatistiği hesaplanmıştır. F istatistiği, doğru pozitif (DP), yanlış pozitif (YP), doğru negatif (DN) ve yanlış negatif (YN) istatistikleri kullanılarak Eşitlik (10)-(13) ile hesaplanmaktadır (Patel et al., 2015a):

$$Kesinlik_{pozitif} = \frac{DP}{DP+YP} \quad (10)$$

$$Kesinlik_{negatif} = \frac{DN}{DN+YN} \quad (11)$$

$$Hassasiyet_{pozitif} = \frac{DP}{DP+YP} \quad (12)$$

$$Hassasiyet_{negatif} = \frac{DN}{DN+YN} \quad (13)$$

Kesinlik, negatif ve pozitif kesinliğin ağırlıklı ortalaması iken hassasiyet pozitif ve negatif hassasiyetin ağırlıklı ortalamasıdır. F istatistiği Eşitlik (14) ile hesaplanmıştır.

$$F = \frac{2 * Kesinlik * Hassasiyet}{Kesinlik + Hassasiyet} \quad (14)$$

Tahmin edilen modellerin parametreleri deneyler sonucunda belirlenmiştir. Her üç model için veri seti normalize edilmiştir. En iyi sınıflandırma performansına sahip olan üç YSA modeli parametreleri Tablo 5'te sunulmaktadır.

Tablo 4: En iyi üç YSA modeline ilişkin parametreler

Model	lr (öğr. oranı)	ep (iterasyon)	mc (momentum sbt)	n(gizli katman sinir hücre sayısı)	Sınıflandırma performansı.
1	0.1	500	0.2	6	0.9914
2	0.1	500	0.6	10	0.9879
3	0.3	500	0.2	6	0.9841

YSA modeli için Tablo 2'de belirtildiği gibi çok sayıda parametre için denemeler yapılmış ve en düşük hata oranı ile tahminler üreten modeller belirlenmiştir. En iyi YSA modeli olarak öğrenme oranı 0,1; momentum sabiti 0,2 ve tek gizli katmanda altı sinir hücresi olan ağ belirlenmiştir. Ayrıca ağ 500 iterasyonda eğitilmiş ve test verisi üzerinde sınıflandırma performansı 0,9914 olarak hesaplanmıştır. Tüm tahmin modellerinin eğitim seti için hesaplanan ortalama mutlak hata (MAE) 0,01 düzeyinin altında gerçekleşmiştir. Bu sebeple aşağıda sunulan sınıflandırma performansları ve hata istatistikleri test verisine ait istatistiklerdir. Modellerin seçimi yapılırken de test verisi üzerindeki başarıları karar kriteri olmuştur. İlk üç model birbirine yakın sınıflandırma performansları göstermiştir.

YSA modeline benzer şekilde, veri seti DVM modeli için normalize edilmiş ve Tablo 3'te belirtilen parametreler ve düzeyleri belirlenmiştir. Kullanılan çekirdek fonksiyonuna göre polinomiyal ve radyal tabanlı olmak üzere üçer model tespit edilmiştir. Polinomiyal çekirdek fonksiyonlu DVM modeli radyal tabanlı çekirdek fonksiyonlu modele göre daha iyi bir sınıflandırma performansı göstermiştir. Tablo 6'da sunulan altı model arasında polinomiyal çekirdek fonksiyonlu modeller ilk iki sırayı, radyal tabanlı çekirdek fonksiyonlu model ise üçüncü sırayı almıştır. DVM modellerinde düzenleme parametresi (c) için 100 değeri en uygun değer olarak belirlenmiştir.

Tablo 5: En iyi üç DVM modeline ilişkin parametreler

	Çekirdek fonk.	d	$\gamma$	c	Perf.
1	Polinomiyal	2	-	100	0.9905
2	Polinomiyal	3	-	100	0.9896
3	Polinomiyal	1	-	100	0.9839
4	Radyal tabanlı	-	0.1	100	0.9871
5	Radyal tabanlı	-	2.0	100	0.9798
6	Radyal tabanlı	-	3.0	100	0.9751

NB algoritması ile yapılan sınıflandırmada sürekli verilerin kullanılmış olmasından dolayı Gauss dağılım fonksiyonu kullanılmıştır. NB algoritmasının sınıflandırma başarısı 0,9042 olarak hesaplanmıştır. Diğer modellerde olduğu gibi 10-kat çapraz doğrulama ile eğitim gerçekleştirilmiştir. NB algoritması için kesinlik ve hassasiyet istatistikleri kullanılarak F istatistiği hesaplanmıştır ve tablo 7'de görüldüğü gibi modeller karşılaştırılmıştır.

Tablo 7: Modellerin karşılaştırılması

	DP	YP	ROC	F-İst..	Perf.
YSA	0,991	0,009	0,998	0,991	0,991
DVM	0,991	0,009	0,991	0,991	0,990
NB	0,904	0,097	0,968	0,904	0,904

Tablo 7'de en iyi YSA, DVM ve NB modelleri karşılaştırılmıştır. F istatistiği iki model için de aynı değeri almıştır. Ancak ROC istatistiği YSA için 0,998 DVM için 0,991 olarak hesaplanmış ve NB için 0,904 de kalmıştır. ROC istatistiği genellikle sınıflandırma problemlerinde kullanılan bir istatistiktir. ROC eğrisi altında kalan alanın 1 değerine yakın olması tahminin başarılı olduğunun bir



göstergesidir. Sınıflandırma performansı açısından da YSA modeli 0,9914 ile en iyi sınıflandırma algoritması olarak tespit edilmiştir.

Doğru sınıflandırılan örneklere ilişkin hesaplanan değerler ise Tablo 8’de sunulmaktadır. YSA modeli toplamda 20 gözlemi yanlış sınıflandırırken, DVM 22, NB ise 223 gözlemi yanlış sınıflandırmıştır.

Tablo 6: Gözlemlerin sınıflandırılması

Sınıflar	YSA		DVM		NB	
	0	1	0	1	0	1
0	1070	16	1076	10	971	115
1	4	1239	12	1231	108	1135

## 5. Sonuç

Borsa hareketlerinin tahmini, etkin stratejiler geliştirmek için oldukça önemli olması yanında bir o kadar karmaşık ve tahmin edilmesi güç bir yapıya sahiptir. Borsa endeksi hareketlerinin tahmini için kullanılan modellerin tahmin başarısı kullanılan değişkenlerin temsil kabiliyetine ve modellerin doğru ve iyi tanımlanmış parametre kümelerine sahip olması ile yakından ilişkilidir.

Yapılan analizler sonucunda üç modelin finansal zaman serisi tahminleri için kullanılabileceği ancak tahmin hata oranlarının farklı olduğu tespit edilmiştir. Hesaplanan istatistikler YSA’nın en iyi tahmin modeli olduğunu göstermektedir. YSA modelinin parametre seçiminin zorluğu kullanımını zorlaştırırsa da doğru seçimlerin performansı ciddi şekilde arttırdığı, literatürdeki birçok çalışma ile doğrulanmıştır. DVM için ise deneysel hataların değil yapısal hataların minimize edilmesi, yerel minimumlara takılmadan küresel minimumların yani en uygun çözümlerin elde edilebilmesi ve parametre seçimin daha kolay yapılması modelin başarısının nedenlerinden olarak gösterilebilir. NB algoritmasının kullanım kolaylığı daha hızlı sonuç üretmesi gibi önemli özellikleri algoritmanın tercih edilmesinin nedenlerindedir. YSA ve DVM performansına erişemese de NB algoritmasının da iyi sonuçlar elde edebileceği açıktır.

Bu çalışma ile YSA, DVM ve NB algoritmaları kullanılarak borsa endeksi hareketlerinin tahmini ve yatırımcılara al-sat pozisyonu almada yardımcı bir tahmin modeli ortaya konmuştur. Bu çalışma için kaynak teşkil eden çalışmalarda ulaşılan sonuçlara göre çok daha iyi bir başarı performansı yakalanmıştır. Örneğin Kara vd. (2011)’de en yüksek test tahmin performansı 0,62 ile 0,83 değerleri arasında gerçekleşmiştir.

Son olarak elde edilen sonuçlar makine öğrenme yöntemlerinin finansal zaman serisi tahminleri konusunda umut verici olduğunu göstermektedir. Modellerin tahmin performanslarının iyileştirilebilmesi için diğer makine öğrenme veya geleneksel yöntemlerle birlikte kullanımlarının da faydalı olacağı düşünülmektedir. Finansal göstergelerin yanında makroekonomik değişkenlerin de girdi olarak kullanılması tahmin performanslarını etkileyebilir. Ayrıca, veri setinin genişletilmesi ile her yıl için ayrı bir tahmin yapıp sonuçların karşılaştırılması, ilgili yıllardaki veri setinin davranışına göre (kriz olup olmaması gibi) modellerin başarılarının incelenmesi ve sürekli veri setinin belli bazı kurallara oluşturularak kesikli veri haline dönüştürülüp analizlerin gerçekleştirilmesi önemli sonuçları ortaya çıkaracaktır. Öngörülebilir risklerin maksimum şekilde sisteme dâhil edilmesi ile elde edilecek tahminlerin yatırımcılara yol göstermesi bakımından önemli olduğu düşünülmektedir. Öngörülemez risklerin de her zaman var olacağı düşünüldüğünde ise hatasız bir tahminin olmayacağı açıktır. Yapılan tahminlerin kâğıt üzerinde kalmadan yatırımcı veya konunun paydaşları ile paylaşılması, hem araştırmacı hem de yatırımcılar için faydalı olacaktır. Ayrıca çalışmanın çıktılarının programlama dillerinden bir tanesi ile uzman bir sistem olarak tasarlanıp, sürdürülebilir ve dinamik bir sistem tasarlanabilir. Bu sistem ile yatırımcılara karar alma noktasından ciddi danışmanlık hizmetleri sunulabilir.

**Kaynakça**

- Abu-Mostafa, Y. S., ve Atiya, A. F. (1996). Introduction to Financial Forecasting. *Applied Intelligence*, 6(3), 205–213.
- Atsalakis, G. S., ve Valavanis, K. P. (2009). Surveying Stock Market Forecasting Techniques – Part II: Soft Computing Methods. *Expert Systems with Applications*, 36(3, Part 2), 5932–5941.
- Cao, L., ve Tay, F. E. H. (2001). Application of Support Vector Machines in Financial Time Series Forecasting. *Omega*, 29(4), 309–317.
- Chen, A.-S., Leung, M. T., ve Daouk, H. (2003). Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading The Taiwan Stock Index. *Computers & Operations Research*, 30(6), 901–923.
- Chen, Z., Matousek, R., ve Wanke, P. (2018). Chinese Bank Efficiency During the Global Financial Crisis: A Combined Approach Using Satisficing DEA and Support Vector Machines. *North American Journal of Economics and Finance*, 43(September 2017), 71–86.
- Chun, S.-H., ve Kim, S. H. (2004). Data Mining for Financial Prediction and Trading: Application to Single and Multiple Markets. *Expert Systems with Applications*, 26(2), 131–139.
- Enke, D., ve Thawornwong, S. (2005). The Use of Data Mining and Neural Networks for Forecasting Stock Market Returns. *Expert Systems with Applications*, 29(4), 927–940.
- Hsu, S. H., Hsieh, J. P. A., Chih, T. C., ve Hsu, K. C. (2009). A Two-Stage Architecture for Stock Price Forecasting by Integrating Self-Organizing Map and Support Vector Regression. *Expert Systems with Applications*, 36(4), 7947–7951.
- Hua, S., ve Sun, Z. (2001). Support Vector Machine Approach for Protein Subcellular Localization Prediction. *Bioinformatics*, 17(8), 721–728.
- Huang, W., Nakamori, Y., ve Wang, S.-Y. (2005). Forecasting Stock Market Movement Direction with Support Vector Machine. *Computers & Operations Research*, 32(10), 2513–2522.
- Kara, Y., Acar Boyacioglu, M., 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.
- Khemchandani, R., Jayadeva, ve Chandra, S. (2009). Knowledge Based Proximal Support Vector Machines. *European Journal of Operational Research*, 195(3), 914–923.
- Kim, K. J. (2003). Financial Time Series Forecasting Using Support Vector Machines. *Neurocomputing*, 55(1–2), 307–319.
- Kimoto, T., Asakawa, K., Yoda, M., ve Takeoka, M. (1990). Stock Market Prediction System with Modular Neural Networks. In *1990 IJCNN International Joint Conference on Neural Networks* (pp. 1–6 vol.1).
- Kumar, M., ve Thenmozhi, M. (2005). Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Fores. In *Forest, Indian Institute of Capital Markets 9th Capital Markets Conference Paper*. (pp. 1–16).
- Leung, M. T., Daouk, H., ve Chen, A.-S. (2000). Forecasting Stock Indices: A Comparison of Classification and Level Estimation Models. *International Journal of Forecasting*, 16(2), 173–190.
- Malkiel, B. G., ve Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383–417.
- Olson, D., ve Mossman, C. (2003). Neural Network Forecasts of Canadian Stock Returns Using

- Accounting Ratios. *International Journal of Forecasting*, 19(3), 453–465.
- Patel, J., Shah, S., Thakkar, P., ve Kotecha, K. (2015a). Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques. *Expert Systems with Applications*, 42(1), 259–268.
- Patel, J., Shah, S., Thakkar, P., ve Kotecha, K. (2015b). Predicting Stock Market Index Using Fusion of Machine Learning Techniques. *Expert Systems with Applications*, 42(4), 2162–2172.
- Tan, T. Z., Quek, C., ve Ng, G. S. (2007). Biological Brain-Inspired Genetic Complementary Learning for Stock Market and Bank Failure Prediction. *Computational Intelligence*, 23(2), 236–261.
- Tay, F. E. H., ve Cao, L. J. (2001). Improved Financial Time Series Forecasting by Combining Support Vector Machines with Self-Organizing Feature Map. *Intelligent-Data-Analysis*, 5, 339–354.
- Thawornwong, S., ve Enke, D. (2004). The adaptive Selection of Financial And Economic Variables for Use With Artificial Neural Networks. *Neurocomputing*, 56, 205–232.
- Vanstone, B., ve Finnie, G. (2009). An Empirical Methodology for Developing Stockmarket Trading Systems Using Artificial Neural Networks. *Expert Systems with Applications*, 36(3, Part 2), 6668–6680.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York, NY: Springer.
- Zhong, X., ve Enke, D. (2017). Forecasting Daily Stock Market Return Using Dimensionality Reduction. *Expert Systems with Applications*, 67, 126–139.
- Xu, X., Zhou, C., ve Wang, Z. (2009). Credit Scoring Algorithm Based on Link Analysis Ranking with Support Vector Machine. *Expert Systems with Applications*, 36(2, Part 2), 2625–2632.

---

## PREDICTING STOCK MARKET MOVEMENT BY USING MACHINE LEARNING

---

### *Extended Abstract*

**Aim:** The most challenging application of time series analysis is the estimation of the stock index and index movements. Although there are a lot of studies in the literature for stock market estimation, a significant portion of these studies are made for developed countries. Especially for Turkey, it is observed that a limited number of studies for emerging markets. Producing valid and reliable estimates for stock market movements is of great importance for developing effective market strategies. Thus, investors have the opportunity to protect against potential market risks and speculators.

The dynamic, non-linear, complex and chaotic structure of financial time series creates barriers to accurate estimates. In addition, stock market indices are affected by many macroeconomic factors such as political changes, the general outlook of the economy, investors' expectations and investment preferences and other indices. As a result of the literature research, very few studies have been found by using actual data and machine learning algorithms for Turkey. Thus, it is the main objective of this study to estimate the direction of the movements of the BIST 100 index and to compare the performance of the machine learning algorithms to be used. The models to be used for estimating and comparing their performances are machine-learning algorithms: i. Artificial neural networks (ANN) ii. Support vector machines (SVM) and iii. Naive Bayes (NB).

**Methods:** Three different machine learning algorithms were used to estimate the negative and positive movements of the BIST 100 index. As an input to these algorithms, ten technical indicators were used as input as declared in Kara et. Al. (2011). The stock market data consist of the closing prices covering the period 2009-2018. The data set was obtained from the Republic of Turkey Central Bank. In order to train the algorithms, ten-fold cross validation method used. Numerous statistics on classifications were used to evaluate model performances. In order to produce better estimates of the methods, the best estimation model was investigated by changing the specific parameters of each method in the training process.

**Findings:** As a result of the analysis, it was determined that all three models could be used for financial time series estimations but the estimation error rates were different. Calculated error statistics show that ANN is the best estimation model. Many studies show that the performance of ANN and SVM is higher than other methods. Although the difficulty of selecting parameters for the ANN model makes it difficult to use, it is seen that the correct choices significantly improve estimation performance. Minimization of structural errors with SVM, global minimums without having to be placed on local minimums and making parameter selection easier can be shown as reasons of the success of the model. Although Naïve Bayes classifier has a lower performance it has other advantages such as ease of use and ease of selecting parameter.

**Conclusion:** Finally, the results show that machine learning techniques are promising in terms of financial time series estimates. In order to improve the predictive performance of the models, other machine learning or their use with traditional methods are also considered to be useful. It will be beneficial for both the researcher and the investor to share the estimates with the stakeholders of the investor or the subject without being on paper. It would be present valuable results that using another macroeconomic variables. With this study, the estimation of stock market index movements by using ANN, SVM and NB algorithms and a helpful prediction model to take the buy-sell position to the investors are presented. According to the results obtained in other studies related to the subject, a better performance was achieved. For example, Kara et al. (2011), the highest test estimation performance was between 0,62 and 0,83.