
Araştırma Makalesi / Research Article

Büyük Dünya Endeksleri Kullanılarak BIST-100 Endeksi Değişim Yönünün Makine Öğrenmesi Algoritmaları ile Sınıflandırılması

Enes FİLİZ^{1*}, Serkan AKOĞUL², Hasan Aykut KARABOĞA³

¹Fırat Üniversitesi Fen Fakültesi İstatistik Bölümü, Elazığ

²Pamukkale Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü, Denizli

³Amasya Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü, Amasya

(ORCID: 0000-0002-8006-9467) (ORCID: 0000-0002-0346-4308) (ORCID: 0000-0007-8877-3267)

Öz

Borsa İstanbul 100 (BIST-100) endeksi, diğer büyük dünya endeksleri ile birlikte finans piyasalarının küreselleşme değişiminin bir parçası olmuştur. Endeksler arasındaki ilişkinin analizi yatırımcılara büyük avantajlar sağlayacaktır. Bu durumdan yola çıkarak çeşitli makine öğrenmesi algoritmaları ile büyük dünya endeksleri ve bazı makroekonomik göstergeler kullanılarak BIST-100 endeksinin değişim yönünün (artış-azalış) sınıflandırılması amaçlanmıştır. Bu amaç doğrultusunda BIST-100 endeksinin değişim yönünün sınıflandırmasında etkin rol oynayan değişkenler belirlenmiş ve belirlenen bu değişkenler yardımıyla sınıflandırma başarılarında değişim olup olmadığı incelenmiştir. Tüm değişkenler ile yapılan sınıflandırmada lojistik regresyonun %70,6; öznelik seçimi ile yapılan sınıflandırmada da Destek Vektör Makinesi PUK çekirdeği algoritmasının %71,9 daha doğru sınıflandırma başarısı gösterdiği belirlenmiştir. Böylelikle daha az sayıda değişken ile daha yüksek sınıflandırma başarısı elde edilmiştir.

Anahtar kelimeler: BIST-100 endeksi, Makine öğrenmesi, Öznelik seçimi, Sınıflandırma algoritmaları.

Classification of BIST-100 Index Change Direction with Machine Learning Algorithms Using Major World Indexes

Abstract

Borsa İstanbul 100 (BIST-100) index has become a part of the globalization of the financial markets along with other major world indexes. The analysis of this relationship will bring great advantages to investors. For this purpose, it is aimed to classify the change (increase-decrease) of the BIST-100 index with machine learning algorithms using major world indexes and some macroeconomic indicators. In first step, effective variables in the classification of the BIST-100 index were determined. In the next step, with the help of these variables, it was examined whether there was a change in classification success. The classification with all variables, logistic regression's classification success was 70.6%. Support Vector Machine PUK core algorithm showed 71.9% correct classification success with selected features. Thus, higher classification success was achieved with less number of variables.

Keywords: BIST-100 index, Machine learning, Feature selection, Classification algorithms.

1. Giriş

Hisse senedi piyasaları bir ülkenin ekonomik gelişmesi ve büyümesinin en önemli göstergelerindendir [1]. Hisse senedi piyasasındaki gelişmeler ülke reel piyasasının uzun dönemli büyümesinin de önemli rol oynamaktadır [2, 3]. Uzun yıllardır yapılan araştırmalar hisse senetleri piyasası ile makroekonomik değişkenler arasındaki ilişkiyi ortaya çıkarmıştır. Bu durum hisse senetlerinin, dolayısıyla da borsa endekslerinin mikro ve makroekonomik değişkenlerle güçlü ilişkisini vurgulamaktadır [1, 4, 5].

*Sorumlu yazar: efiliz@firat.edu.tr

Geliş Tarihi: 01.03.2021, Kabul Tarihi: 03.05.2021

Genel ekonomik koşullar, politik olaylar, yatırımcı psikolojisi ve büyük yatırımcıların farklı ülke borsalarına ilgisi borsa endekslerini etkilemektedir [6, 7]. Döviz kurlarının, altın fiyatlarının ve yabancı borsa endekslerinin de ülkenin borsa endeksleri üzerinde etkisi vardır [8, 9]. Teknik indikatörlerin yanında politika faizi, döviz kurları, enflasyon ve hükümet politikaları borsa endeksini etkileyen değişkenlerdir [4, 10]. Yani tüm bu belirsizlikler endeks yönü ya da hisse fiyatı tahminlerini zorlaştırmaktadır [4].

Alış ve satış emirlerinin saliselerle ölçüldüğü günümüzde hisse senedi piyasaları oldukça oynak hale gelmiştir. Birçok parametreden etkilenen hisse senedi yatırımları bu nedenle yüksek risk içermektedir [7, 11]. Yüksek risk genel olarak yüksek getiri potansiyelini de beraberinde getirmektedir. Yüksek getiri potansiyeli nedeniyle yabancı yatırımcılar BIST' e yoğun ilgi göstermektedir [12].

Makine öğrenmesine dayanan yöntemler ile modelleme yaparak yatırımcılar olağanüstü durumlara karşı kendini korumaya almakta hatta kazanç sağlayabilmektedir [6, 13-15]. Endeks yönünün bilinmesi ile alıcı ve satıcılar alım satım stratejileri belirleyebilmektedir [16]. Yatırımcıların birikimlerini korumak ve yatırımlarını daha kârlı hale getirmek için piyasanın yönünü başarılı şekilde tahmin edebilmesi gerekmektedir.

Makine öğrenmesi yöntemlerinin klasik yöntemlere göre daha esnek yapıda olması ve daha az varsayıma ihtiyaç duyması finansal araştırmalarda yoğun olarak kullanılmasını sağlamaktadır [17, 18]. Makine öğrenmesi yöntemlerinin bir diğer üstün özelliği ise doğrusal olmayan ilişkileri modelleme başarısıdır [19]. Yöntemlerin kullanım kolaylığı ile tahmin ve genelleme yapabilme yetenekleriyle sağladığı üstünlükler de yaygınlaşmasını kolaylaştırmıştır [20].

Finansal veri madenciliği, yapay zekâ algoritmalarının finans piyasalarında kullanımını sağlamaktadır. Finansal veri madenciliğine yapılan yatırımlar, dünya borsalarının, petrol ve emtiaların ve yabancı para cinsinden yatırım araçlarının birbirine daha bağlı hale gelmesini sağlamıştır. Endeksler ve diğer yatırım araçları arasındaki ilişkinin tespiti ileriye dönük tahmin yapabilmeyi de kolaylaştıracaktır. Literatürde BIST-100 endeksinin sınıflandırılmasında ve tahmininde birçok algoritma kullanılmıştır. Bu çalışmalarda çoğunlukla yapay sinir ağları (YSA), destek vektör makinesi (DVM), lojistik regresyon (LR), k-en yakın komşu algoritması (k-NN), C4.5 karar ağacı algoritması (KA-C4.5), lineer diskriminant algoritmaları (LDA) ve Naif Bayes algoritması (NB) kullanılmıştır. Örneğin Diler (2003), BIST-100 endeksinin bir sonraki günlük yönünü YSA kullanarak %60,81 doğruluk oranı ile tahmin etmiştir [21]. Avcı (2007) da çalışmasında BIST-100 endeksinin seanslık ve günlük getirilerini YSA ile tahmin etmiştir. Çalışma sonucunda seanslık tahminin daha başarılı sonuç ürettiğini belirtmiştir [22]. Benzer şekilde Kutlu ve Badur (2009), döviz ve diğer borsaların verilerini kullanarak endeks değerinin ileri beslemeli yapay sinir ağları ile modellenebileceğini göstermişlerdir [11]. Sakarya vd. (2015), BIST-100 Endeksi'nin 2008 küresel krizi döneminde günlük ve haftalık getirilerinin YSA kullanılarak öngörülebilirliği göstermişlerdir. Girdi değişkeni olarak yedi farklı ekonomik parametre ile YSA'nın bu alanda oldukça başarılı bir şekilde kullanılabileceğini ve bilinmeyen örnekler için bile %5'in altında bir doğruluk marj hatası ile sonraki gün ve gelecek hafta için değeri doğru bir şekilde öngördüğünü göstermişlerdir [23].

BIST-100 endeksinin günlük hareket yönü tahmini için on teknik göstereyi kullanan Kara vd. (2011), YSA ile %75,74 ve DVM ile %71,52 sınıflandırma başarısı elde etmiştir [16]. Yakut vd. (2014) ise BIST-100 endeksi hareket yönünü YSA ve DVM ile modellemiştir. Bağımsız değişken olarak gecelik faiz oranı, dolar kuru, NIKKEI, BOVESPA, FTSE, CAC ve DAX kullanılmış ve yöntemlerin tahmin başarısında anlamlı bir farklılık bulunmamıştır [24]. Özdemir vd. (2011) ise BIST-100 endeksi getiri yönünü LR ve DVM yöntemleri ile %86 doğrulukla sınıflandırmıştır [18]. Tayyar ve Tekin (2013) de BIST-100 endeksi hareket yönünü DVM ve LR sınıflandırmışlardır. DVM'nin %70 sınıflama başarısı ile daha iyi performans gösterdiğini ortaya koymuşlardır [25]. Pabuçcu (2019), BIST-100 endeksinin geleceğe dönük hareketlerini tahmin edebilmek için 2009-2018 tarihsel verisini kullanarak YSA, DVM ve NB sınıflandırma algoritmalarını kullanmış ve algoritmanın başarılarını karşılaştırmıştır. Çalışma sonucunda üç yöntemde başarılı ve birbirine yakın sonuçlar ürettiğini vurgulamıştır [26].

Diğer çalışmalardan farklı olarak Filiz ve Öz (2017) BIST-100 endeksindeki değişimi makine öğrenme yöntemlerinden k-NN, NB, LR ve KA-C4.5 ile sınıflandırmışlardır. Girdi değişkeni olarak BIST-100 endeks değerlerinin değişimini etkileyen faktörler alındığında en başarılı yöntemi KA-C4.5 algoritması (%66,2) olarak bulmuşlardır [27]. k-NN algoritmasının kullanıldığı bir başka çalışmada ise Kemalbay ve Alkış (2020) BIST-100 endeksinin hareket yönünü tahmin etmiştir. Yöntem olarak çoklu LR ve k-NN algoritmasını kullanmışlardır. Veri seti 2010-2016 tarihleri arasında 1700 işlem günü olup

%95 eğitim ve %5 test verisi olmak üzere model kurmuşlardır. LR modelinin %81 doğruluk oranı ile iyi tahmin performansı gösterdiğini ortaya koymuşlardır [14]. Çalışmaların büyük bir çoğunluğunda YSA daha başarılı sonuçlar ortaya koymuştur [28]. Sınıflandırma algoritması olarak LR' de oldukça başarılı sonuçlar üretmektedir [14]. Bazı çalışmalarda ise DVM yöntemi LR yönteminden daha başarılı sınıflandırma yapmıştır [25]. Parametrelerin belirlenmesi ve modelin kullanım zorluğu bakımından YSA oldukça dezavantajlı konumdadır. Kullanım kolaylığı bakımından NB diğer makine öğrenmesi yöntemlerine göre daha kolaydır. Buna karşın DVM model hatalarının en küçük hale getirilmesinde diğer yöntemlere göre üstündür [26]. Ancak DVM için kullanılan çekirdek fonksiyonu modelleme başarısını etkilemektedir [25]. Özetle makine öğrenmesi yöntemlerinin birbiri üzerinde farklı üstünlükleri vardır. Bu üstünlükler modelleme başarılarını etkilemektedir. Bunun için modellerin birçoğunun birlikte kullanımı modelleme başarısını arttıracaktır [29].

Bu çalışmada, literatürde verilen çalışmalardan farklı olarak büyük dünya endeksleri, altın (ons), dolar ve Euro kuru verilerinden yararlanarak makine öğrenmesi algoritmaları ile BIST-100 endeksinin değişim yönünün sınıflandırılması ve sınıflandırmada etkili özniteliklerin belirlenmesi amaçlanmıştır. Çalışmada kullanılacak olan sınıflandırma algoritmaları: NB, LR, YSA, DVM-PUK ve rastgele orman karar ağacı (KA-RF) algoritmalarıdır. Ayrıca sınıflandırmada etkili özniteliklerin seçimi içinde CfsSubset öznitelik seçim algoritması kullanılmıştır.

2. Materyal ve Metot

2.1. Veri Seti

İstanbul Borsası, 2013 yılında yapılan bir değişiklikle İstanbul Menkul Kıymetler Borsası, Vadeli İşlemler Borsası, Opsiyon Borsası ve Altın Borsası tek çatı altında toplanarak Borsa İstanbul (BIST) adını almıştır. BIST-100 endeksi, Borsa İstanbul Pay Piyasası için temel endeks olarak kullanılmaktadır. Borsa İstanbul pazarlarında işlem gören, piyasa değeri ve işlem hacmi en yüksek 100 payın ortak performansının ölçülmesi amacıyla oluşturulmuştur.

BIST-100 endeksinin değişim yönünün sınıflandırılmasında büyük dünya endeksleri ile küresel olarak piyasalarda kullanılan altın (ons), dolar ve Euro'nun TL cinsinden kur bilgilerinden yararlanılmıştır. Analizler, 01.01.2006 – 01.12.2020 tarihleri arasında işlem görmüş veriler kullanılarak yapılmıştır. Ele alınan veriler zaman serisi olarak incelendiğinde durağanlaştırma zorluğu ortaya çıkmaktadır. Verinin durağan olmayan bir yapıdan kesikli bir yapıya dönüştürülmesi modelleme ve tahmin yapılmasını kolaylaştırmaktadır [4]. Bu nedenle değişkenlerin bir önceki güne göre artış ya da azalışları 1 ve 0 olarak kodlanmıştır. Çalışmada kullanılan bağımlı değişken ve bağımsız değişkenler Tablo 1'de verilmiştir.

Tablo 1. Çalışmada kullanılan değişkenler*

Bağımlı değişken	Bağımsız değişkenler	
<i>BIST-100 (Türkiye)</i>	<i>S&P-500 (Amerika)</i>	<i>Moex (Rusya)</i>
	<i>Nasdaq (Amerika)</i>	<i>Ta-35 (İsrail)</i>
	<i>Russell-2000 (Amerika)</i>	<i>Tadawul (Saudi Arabia)</i>
	<i>S&P/Tsx (Kanada)</i>	<i>Nikkei-225 (Japonya)</i>
	<i>Bovespa (Brezilya)</i>	<i>S&P/Asx-200 (Avustralya)</i>
	<i>Ipc (Meksika)</i>	<i>Shangai (Çin)</i>
	<i>Dax (Almanya)</i>	<i>Hang-Seng (Hong-Kong)</i>
	<i>Cac-40 (Fransa)</i>	<i>Kospi (Güney Kore)</i>
	<i>Aex (Hollanda)</i>	<i>Idx-Composite (Endonezya)</i>
	<i>Ibex-35 (İspanya)</i>	<i>Nifty-50 (Hindistan)</i>
	<i>Ftse-Mib (İtalya)</i>	<i>Cse (Sri Lanka)</i>
	<i>Smi (İsviçre)</i>	<i>Dolar kuru</i>
	<i>Bel-20 (Belçika)</i>	<i>Euro kuru</i>
	<i>Omx-30 (İsveç)</i>	<i>Altın (ons)</i>

*Veriler <https://tr.investing.com/> sitesinden alınmıştır [30].

2.2. Sınıflandırma Algoritmaları

Naif Bayes algoritması (NB), makine öğrenmesinin danışmanlı öğrenme alt sınıfında yer almaktadır. Bir tür Bayes ağı olan bu algoritmanın en başarılı şekilde çalışması için iki varsayımın gerçekleşmesi gerekmektedir. İlk olarak, sınıfların belirli şartlar altında birbirlerinden bağımsız olmasıdır. İkincisi ise çalışmanın sonuçlarını etkileyebilecek olan değişkenlerin belirli olmasıdır [31].

Lojistik regresyon algoritması (LR), temel amacı bağımlı ve bağımsız değişkenler arasındaki ilişki veya ilişkileri ortaya çıkarmak olan klasik istatistiksel bir yöntemdir. Bu yöntemde doğrusal regresyonda geçerli olan varsayımlar aranmadığından doğrusal regresyona göre daha esnek ve daha fazla tercih edilmektedir. Ayrıca bağımlı değişkenin kategorik olması lojistik regresyon analizini doğrusal regresyondan ayırır [32].

Yapay Sinir Ağları (YSA), insan beyninin işlem süreçlerini taklit ederek çalışan ve geçmişteki deneyimlerin sonuçlarına göre hareket eden en önemli makine öğrenmesi algoritmalarındandır. Bu nedenle, karmaşık doğrusal olmayan durumlarla çalışırken bile, YSA modelleri standart istatistik yöntemlerinde olduğu gibi varsayımlara ihtiyaç duymaz [33].

Destek Vektör Makinesi Puk Çekirdeği (DVM-PUK), verileri iki kategoriye ayırmak için n-boyutlu bir hiper düzlem oluşturan güçlü bir yöntemdir [34]. DVM kendi içinde çekirdek fonksiyonlarına sahiptir. Çekirdek fonksiyonu, etkin bir benzerlik ölçüsüdür ve lineer bir modeli doğrusal olmayan bir modele dönüştürebilir. Çalışmada destek vektör makinesi PUK çekirdeği kullanılmıştır.

Rastgele Orman Karar Ağacı algoritması (KA-RF), karar ağaçlarını geliştirmek için eğitim verilerinin değişimlerinden yararlanarak işlem yapan bir yöntemdir. Eğitim verilerinin değişimlerini belirlemek için orijinal eğitim verilerinin parçaları rassal olarak değiştirilir. Her ağaç budama işlemi yapılamadan önce olabildiğince ilerletilmeye odaklanır. Daha sonra her ağacın kendi sınıflandırmasını yapmasına izin verilir ve böylece karar süreci uygulama işlemi yapılır [35].

k-katlı çapraz doğrulama, makine öğrenmesi algoritmalarının çalışma prensibinde önemli noktalardan birisi de eğitim ve test veri setinin ayrımıdır. k-katlı çapraz doğrulama da veri seti k eşit parçaya bölünür. Burada k-1 tanesi eğitim veri seti geri kalan kısmı ise test veri seti olarak belirlenir. Bu işlem k kez tekrarlanır. Her bir parça test veri seti olarak alınır. Sonuç olarak bu k kez tekrar edilen işlemler sonrası elde edilen başarı değerlerinin ortalaması alınarak sınıflandırma algoritmasının başarısı elde edilir. Bu çalışmada da 10 katlı çapraz doğrulama işlemi yapılmıştır [36].

2.3. Öznitelik Seçimi

Öznitelik seçimi, makine öğrenmesi algoritmaları bünyesinde kullanılan değişken indirgeme durumudur. Öznitelik seçimi ile veri setindeki ilişkisiz ya da sonuca az etkisi olan değişkenler çalışmadan çıkarılmaktadır. Önemli nokta sınıflandırma yaparken daha az değişkenle daha yüksek ya da fazla düşüş olmadan benzer sonuçlar elde etmektir. Çalışmada CfsSubset öznitelik seçim algoritmasından yararlanılmıştır.

CfsSubset öznitelik seçimi algoritması, en iyi öznitelikleri belirlemek için veri setindeki değişkenleri korelasyon yardımı ile değerlendirilmektedir. Bu algoritma aralarında düşük korelasyonlu, sınıf etiketleri arasında ise yüksek korelasyonlu öznitelikleri seçmektedir. CfsSubset algoritması veri setine en uygun sayıda öznitelik belirlemektedir. Seçilen özniteliklerin önem sıralaması için Ki-Kare testinden yararlanılabilir. Ki-Kare testi, gözlenen ve beklenen frekanslar arasındaki farkın anlamlı olup olmadığını test etmektedir. Hesaplanan χ^2 değerine göre büyükten küçüğe doğru sıralanmasının ardından öznitelikler önem sırasına göre de sıralanmış olacaktır [37, 38].

2.4. Sınıflandırma Kriterleri

Çalışmada BIST-100 endeksinin değişim yönünün sınıflandırma başarısını elde edebilmek için farklı makine öğrenmesi algoritmaları kullanılacaktır. Bu sınıflandırma algoritmalarının başarılarının belirlenebilmesi için sınıflandırma ölçütlerine ihtiyaç duyulmaktadır. Bu çalışmada Doğruluk, Kappa istatistiği, Ortalama Hataların Karekökü, ROC Eğrisi, Matthew Korelasyon Katsayısı sınıflandırma ölçütlerinden yararlanılmıştır.

Doğruluk (Acc), modeldeki doğru sınıflandırılmış olan tüm örneklerin, toplam örnek sayısına oranı ile tespit edilmektedir [39].

Kappa istatistiği (κ), kategorik değişkenler için yapılan analizlerin anlaşılmasında kullanılan ve ki-kare tablosuna dayanan uygun bir istatistiki veridir [40]. κ istatistiği 1'e yaklaştıkça, sınıflandırma algoritmasının başarısının arttığı söylenebilir.

Ortalama Hataların Karekökü (Rmse), bir modelin tahmin edilen ve gözlenen değerleri arasındaki farkları göstermeye yardımcı olmaktadır. Kök ortalama kare hatası, tahmin edilen ve gözlemlenen değerler arasındaki kare farklarının ortalama kareköküne eşittir. Bu değer az olması istenilen bir durumdur.

ROC Eğrisi (ROC), sınıflandırma algoritmalarının performansını ölçmek için kullanılan ROC eğrisinin altındaki alan sınıflandırıcının başarısı ile ilgili bilgi vermektedir [41]. ROC eğrisinin altında kalan alanın değeri ne kadar yüksek olursa, algoritmanın yaptığı sınıflandırmanın da o kadar başarılı olduğu söylenebilir. Ayrıca algoritmanın başarısını görsel olarak göstermektedir.

Matthew Korelasyon Katsayısı (Mcc), -1 ile 1 arasında değerler alan Mcc karşılaştırma matrisindeki bileşenler kullanılarak elde edilmektedir. Pozitif değer elde eden Mcc'lerin doğru tahminler ürettiği sonucuna varılabilmektedir [42].

3. Uygulama

BIST-100 endeksinin değişim yönünün sınıflandırılmasında eldeki bağımsız değişkenler yardımıyla çalışma 3 aşamadan oluşmaktadır. Analizlerde Waikato Üniversitesi tarafından geliştirilen Weka programından yararlanılmıştır. Bu program, makine öğrenmesi algoritmalarını ve öznitelik seçim algoritmalarını bünyesinde barındırmaktadır. Sınıflandırma ve kümeleme analizlerinde kullanımı uygundur [29]. Çalışmanın aşamaları şu şekilde verilmiştir:

Aşama1. BIST-100 endeksinin değişim yönünün tüm değişiklikler kullanılarak makine öğrenmesi algoritmaları ile sınıflandırılması,

Aşama2. Sınıflandırmada etkin rol oynayan özniteliklerin belirlenmesi ve istatistiksel olarak test edilmesi,

Aşama3. Etkin öznitelikler kullanarak makine öğrenmesi algoritmalarının sınıflandırma başarılarında düşüş ya da yükselişin incelenmesi.

Çalışmanın ilk aşamasında, kullanılan 28 bağımsız değişken göz önünde bulundurularak sınıflandırma algoritmaları ile veriler analiz edilmiştir. Aşama1'e göre kullanılan beş sınıflandırma algoritması ile sınıflandırma başarıları Tablo 2' de verilmiştir. Daha önce belirtilmiş olan sınıflandırma kriterlerine göre (Acc, κ , Rmse, ROC ve Mcc) algoritmaların sınıflandırma performansları incelenmiştir.

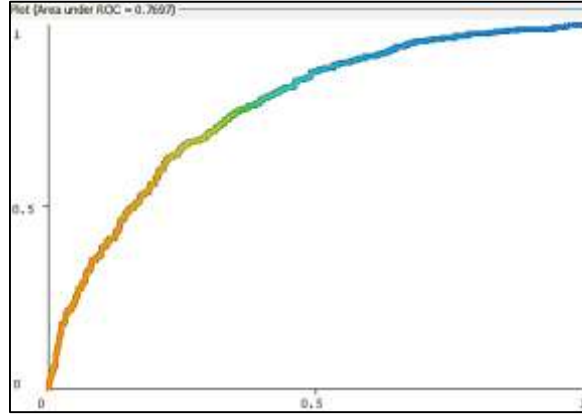
Tablo 2. Aşama1 için elde edilen sınıflandırma sonuçları

	NB	LR	YSA	DVM-PUK	KA-RF
Acc(%)	67,70	70,60	62,70	69,60	67,80
κ	0,350	0,406	0,249	0,383	0,350
Rmse	0,524	0,441	0,583	0,551	0,455
ROC	0,749	0,770	0,669	0,690	0,740
Mcc	0,350	0,406	0,249	0,387	0,350

Tablo 2'de görüldüğü gibi, Acc'ye göre LR (%70,6) en iyi algoritma olarak belirlenmiştir. LR için κ (0,4057), Rmse (0,4410), ROC (0,7700) ve Mcc (0,4060) değerleri bu sonucu desteklemektedir.

Şekil 1'de, BIST-100 endeksinin sınıflandırılmasında Aşama1 için sınıflandırmada başarılı algoritmanın ROC alanı gösterilmiştir. ROC alanı değerleri tüm algoritmalar için incelendiğinde 0,657'nin üzerinde olduğunu söylemek mümkündür. LR ise en başarılı ROC alanı değerini vermiştir (0,770).

Aşama2 için CfsSubset öznitelik seçim algoritması yardımıyla yapılan analizler sonucunda Tablo 3'te belirtilen değişkenler BIST-100 endeksinin değişim yönünün sınıflandırılmasında etkin olan öznitelik olarak belirlenmiştir.



Şekil 1. Aşama1 için LR algoritmasının ROC eğrisi

Tablo 3. Aşama2 için elde edilen sınıflandırmada etkin öznitelikler

<i>Bovespa</i>	<i>Moex</i>
<i>Ipc</i>	<i>Nifty-50</i>
<i>Cac-40</i>	<i>Dolar kuru</i>
<i>Ibex-35</i>	<i>Euro kuru</i>
<i>Omx-30</i>	

Aşama2'ye göre, BIST-100 endeksinin değişim yönünün sınıflandırılmasında etkin olan öznitelikler Ki-kare analizi ile test edilmiş ve sonuçlar Tablo 4'te verilmiştir.

Tablo 4. Aşama2 için etkin özniteliklerin Ki-kare analizi

Öznitelik	χ^2 Değeri	p-değeri
Dolar kuru	202,982	
CAC-40	127,064	
Omx-30	125,992	
Ipc	121,325	
Ibex-35	112,637	0,000
Euro kuru	106,116	
Bovespa	100,034	
Moex	99,067	
Nifty-50	98,859	

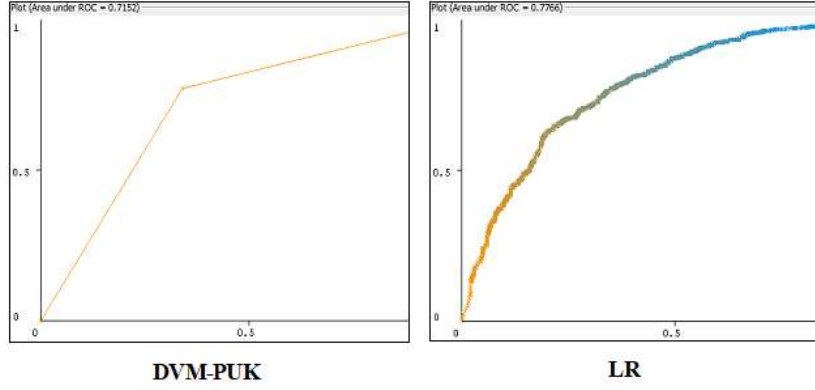
Tablo 4'e göre BIST-100 endeksinin değişim yönü ile etkin öznitelikler arasında istatistiksel olarak anlamlı ilişki tespit edilmiştir ($p=0,000$). Bu öznitelikler χ^2 değerlerine göre sıralanmıştır. En etkili öznitelik dolar kuru ($\chi^2 = 202,982$) olarak belirlenmiştir.

Aşama3'e göre etkin öznitelikler kullanarak makine öğrenmesi algoritmalarının sınıflandırma başarıları Tablo 5'te verilmiştir.

Tablo 5. Aşama3 için elde edilen sınıflandırma sonuçları

	NB	LR	YSA	DVM-PUK	KA-RF
Acc(%)	70,4	71,2	70,2	71,9	70,3
κ	0,404	0,418	0,396	0,433	0,401
Rmse	0,458	0,437	0,456	0,530	0,466
ROC	0,772	0,777	0,737	0,715	0,732
Mcc	0,404	0,419	0,399	0,434	0,401

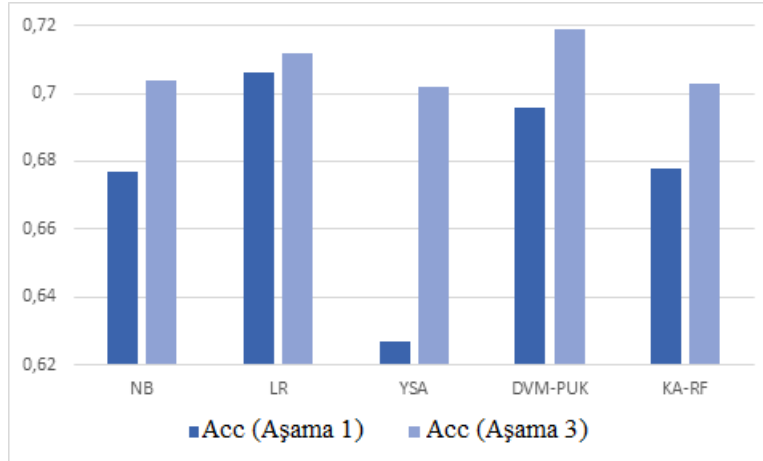
Tablo 5'te görüldüğü gibi, Acc'ye göre DVM-PUK (%71,9) en iyi algoritma olarak belirlenmiştir. DVM-PUK için κ (0,4328) ve Mcc (0,4340) değerleri bu sonucu desteklemektedir. Ayrıca, LR (%71,2) DVM-PUK ile karşılaştırıldığında benzer performansa sahip olduğu söylenebilir. LR sonucunu Rmse (0,4369) ve ROC alanı (0,7770) ölçütlerinin desteklediği söylenebilir.



Şekil 2. Aşama3 için DVM-PUK ve LR algoritmalarının ROC eğrileri

Şekil 2’de, BIST-100 endeksinin değişim yönünün sınıflandırılmasında Aşama3 için sınıflandırmada başarılı algoritmanın ROC alanı gösterilmiştir. ROC alanı değerlerinin tüm algoritmalar için 0,7150’ nin üzerinde olduğunu söylemek mümkündür. LR (0,777) ise en başarılı ROC alanı değerini vermiştir.

BIST-100 endeksinin değişim yönünün sınıflandırılmasına yönelik olarak Aşama1 ve Aşama3 de gerçekleştirilen sınıflandırmalar için sınıflandırma ölçütlerinden Acc temel alınarak algoritmaların başarısı Şekil 3’te verilmiştir. Şekil 3’e göre öznitelik seçimi ile tüm algoritmaların sınıflandırma başarısının arttığı görülmektedir.



Şekil 3. BIST-100 endeksinin değişim yönü için sınıflandırma sonuçları

4. Bulgular ve Tartışma

Borsa endeksleri teknolojik gelişmeler ile her geçen gün daha yoğun olarak iç içe geçmektedir. Haberlerin anlık yayılımı ve piyasalarda etkilerinin anında görülmesi alım satım kararlarının anında yön değiştirmesine neden olmaktadır. Yatırımcıların kendini koruması için kararlarını belirli bir modele ve değişken kümesine dayandırarak alması gerekmektedir. Bu amaçla çalışmamızda dünyanın gelişmiş ve gelişmekte olan borsalarının endeksleri ile borsayı etkilediği bilinen altın, dolar ve Euro kurları ile BIST-100 ilişkisi incelenmiştir. Sınıflandırma için etkili özniteliklerin de belirlendiği bu çalışma, öznitelik seçimi kullanılması ile literatürden farklılık göstermektedir. Ayrıca etkili özniteliklerin istatistiksel olarak da değerlendirilmesi için Ki-kare testinden yararlanılmıştır.

BIST-100 endeksinin değişim yönünün sınıflandırılmasında Aşama1 de LR en yüksek doğruluğa sahip algoritma olarak bulunmuştur. LR’nin Acc değeri %70,6 olduğu ve bu değer 28 değişken kullanarak elde edildiği görülmektedir. Yani bağımsız değişkenlerin değerlerinin bilinmesi durumunda BIST-100 endeksinin değişim yönünün belirlenmesi %70,6 oranında doğru bir şekilde yapılabilecektir. Literatürde yapılan incelemelere göre mevcut pek çok çalışmada, LR’nin hisse senedi piyasalarında yapılan sınıflandırılma çalışmalarında kullanılan en uygun algoritmalarından biri olduğu

belirlenmiştir [18, 19, 28]. LR farklı alanlarda yapılan çalışmalarda da diğer makine öğrenmesi yöntemlerine göre daha iyi performans gösterebilmektedir [36, 43].

BIST-100 endeksinin değişim yönünün sınıflandırma başarısını etkileyen öznitelikler bilindiğinde önlemler alınarak veya mevcut koşulları değiştirerek sınıflandırma başarısında bir düşüş olmayacağı ya da arttırılacağı varsayılabilir. Önem düzeyi düşük özniteliklerin modelden çıkarılması modellerin başarısını arttırmaktadır [22]. Bu amaçla Aşama2’de BIST-100 endeksinin değişim yönünün sınıflandırma başarısını etkileyen en önemli öznitelikleri belirlemek için öznitelik seçim algoritması kullanılmıştır. Öznitelik seçim algoritmasını kullanmanın amacı, sınıflandırma başarısından ödün vermeden daha az öznitelik kullanmaktır. BIST-100 veri seti için CfsSubset öznitelik seçim algoritması kullanılarak elde edilen sonuçlara göre ‘Bovespa’, ‘Ipc’, ‘Cac-40’, ‘Ibex-35’, ‘Omxs-30’, ‘Moex’, ‘Nifty-50’, ‘dolar’, ‘Euro’ etkili öznitelikler olarak belirlenmiştir. Ki-kare analizine göre en etkili özneliğin dolar kuru olduğu görülmüştür. Etkili öznitelikler literatürdeki pek çok çalışmada da önemli öznitelikler olarak bulunmuştur [8, 9, 13, 18, 24].

Son olarak Aşama3 de BIST-100 endeksinin değişim yönünün 9 öznitelik ile Acc oranı %71,9 olan DVM-PUK algoritması en iyi sınıflandırma algoritması olarak belirlenmiştir. Yani BIST-100 endeksinin değişim yönünün sınıflandırılmasını etkileyen bağımsız değişkenlerin değerlerinin bilinmesi durumunda BIST-100 endeksinin değişim yönünün belirlenmesi %71,9 oranında doğru bir şekilde yapılabilecektir. Daha az sayıda öznitelik kullanılarak sınıflandırma başarısında tatminkâr sonuçlar elde edilmiştir.

BIST-100 veri setinde BIST-100 endeksinin değişim yönünün sınıflandırılmasında çok sayıda makine öğrenmesi algoritması ve öznitelik seçim algoritması kullanılması bu çalışma açısından oldukça önemlidir. Ayrıca, BIST-100 literatürü ile ilgili daha önceki çalışmalar incelendiğinde, kümeleme, tahmin ve regresyon amacıyla sıklıkla tercih edilen makine öğrenmesi algoritmalarının birçoğu kullanılmıştır. Diğer önemli bir nokta ise BIST-100 sınıflandırılmasında sadece en etkin değişkenlerin dikkate alınması gerektiğini ve tüm değişkenlerin kullanılmasına gerek olmadığını gösterilmektedir. İlerleyen çalışmalarda ilk aşamada kullanılan değişkenlerin sayısının arttırılması ve en etkili makro ve mikro değişkenlerin belirlenmesi yatırımcılar, ekonomistler ve ekonomi politikası belirleyicileri için yararlı olacaktır.

Çalışmada BIST-100 endeksinin değişim yönünün sınıflandırılmasında yalnızca büyük dünya endeksleri, altın, dolar ve Euro kuru kullanılmıştır. Bu ise çalışmanın kısıtlarındandır. Zira birçok mikro ve makro değişkenin endeks üzerinde etkili olabileceği düşünülmektedir. Diğer bir sınırlama da BIST-100 endeksinin değişim yönünün sınıflandırılmasını etkileyen özniteliklerin, sadece BIST-100 için belirlenmiş olmasıdır.

Yazarların Katkısı

Yazarlar makaleye eşit oranda katkı sağlamışlardır.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

Kaynaklar

- [1] Oskooe P. 2010. Emerging Stock Market Performance and Economic Growth. American Journal of Applied Sciences, 7 (2): 265-269.
- [2] Levine R., Zervos S. 1996. Stock Market Development and Long-Run Growth. The World Bank Economic Review, 10 (2): 323-339.
- [3] Masoud N.M.H. 2013. The Impact of Stock Market Performance upon Economic Growth. International Journal of Economics and Financial Issues, 3 (4): 788-798.

- [4] 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.
- [5] Pesaran M.H., Timmermann A. 1995. Predictability of Stock Returns: Robustness and Economic Significance. *The Journal of Finance*, 50 (4): 1201-1228.
- [6] Tan T.Z., Quek C., Ng G.S. 2007. Biological Brain-Inspired Genetic Complementary Learning for Stock Market and Bank Failure Prediction1. *Computational Intelligence*, 23 (2): 236-261.
- [7] Wang J.-Z., Wang J.-J., Zhang Z.-G., Guo S.-P. 2011. Forecasting stock indices with back propagation neural network. *Expert Systems with Applications*, S0957417411007494.
- [8] Savaş İ., Can İ. 2011. Euro-Dolar Paritesi ve Reel Döviz Kuru'nun İMKB 100 Endeksi'ne Etkisi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 6 (1): 323-339.
- [9] Bali S., Cinel M. 2011. Altın fiyatlarının İMKB 100 endeksi'ne etkisi ve bu etkinin ölçümlenmesi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 25 (3-4): 45-63.
- [10] Lehkonen H., Heimonen K. 2015. Democracy, political risks and stock market performance. *Journal of International Money and Finance*, 59: 77-99.
- [11] Kutlu B., Badur B. 2009. Yapay sinir ağları ile borsa endeksi tahmini. *Yönetim Dergisi: İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadi Enstitüsü*, 20 (63): 25-40.
- [12] Gunduz H., Cataltepe Z. 2015. Borsa İstanbul (BIST) daily prediction using financial news and balanced feature selection. *Expert Systems with Applications*, 42 (22): 9001-9011.
- [13] Aydın A.D., Cavdar S.C. 2015. Comparison of Prediction Performances of Artificial Neural Network (ANN) and Vector Autoregressive (VAR) Models by Using the Macroeconomic Variables of Gold Prices, Borsa İstanbul (BIST) 100 Index and US Dollar-Turkish Lira (USD/TRY) Exchange Rates. *Procedia Economics and Finance*, 30: 3-14.
- [14] Kemalbay G., Alkiş B.N. 2020. Borsa endeks hareket yönünün çoklu lojistik regresyon ve k-en yakın komşu algoritması ile tahmini. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 26 (8).
- [15] Kumar M., Thenmozhi M. 2006. Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest (SSRN Scholarly Paper ID 876544). *Social Science Research Network*. <https://doi.org/10.2139/ssrn.876544>
- [16] Kara Y., Acar Boyacioglu M., 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.
- [17] Na S.H., Sohn S.Y. 2011. Forecasting changes in Korea Composite Stock Price Index (KOSPI) using association rules. *Expert Systems with Applications*, 38 (7): 9046-9049.
- [18] Özdemir A.K., Tolun S., Demirci E. 2011. Endeks Getirisi Yönünün İkili Sınıflandırma Yöntemiyle Tahmin Edilmesi: İMKB 100 Endeksi Örneği. *Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4 (2): 45-59.
- [19] Kara İ., Ecer F. 2018. BIST endeks hareket yönünün tahmininde sınıflandırma yöntemlerinin performanslarının karşılaştırılması. *The Journal of Academic Social Sciences*, 83 (83): 514-524.
- [20] Filiz E., Karaboğa H.A., Akogul S. 2017. BIST-50 endeksi değişim değerlerinin sınıflandırılmasında makine öğrenmesi yöntemleri ve yapay sinir ağları kullanımı. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 26 (1): 231-241.
- [21] Diler A.İ. 2003. İmkb ulusal-100 endeksinin yönünün yapay sinir ağları hata geriye yayma yöntemi ile tahmin edilmesi. *İMKB Dergisi*, 7 (25-26): 66-81.
- [22] Avcı E. 2007. Forecasting daily and sessional returns of the ISE - 100 index with neural network models. *Doğuş Üniversitesi Dergisi*, 8 (2): 128-142.
- [23] Sakarya S., Yavuz M., Karaoglan A.D., Özdemir N. 2015. Stock Market Index Prediction with Neural Network during Financial Crises: A Review on Bist-100. *Financial Risk and Management Reviews*, 1 (2): 53-67.
- [24] Yakut E., Elmas B., Yavuz S. 2014. Yapay Sinir Ağları ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19 (1): 139-157.
- [25] Tayyar N., Tekin S. 2013. İMKB-100 Endeksinin Destek Vektör Makineleri ile Günlük, Haftalık ve Aylık Veriler Kullanarak Tahmin Edilmesi. *Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 13 (1): 189-217.

- [26] Pabuçcu H. 2019. Borsa Endeksi Hareketlerinin Tahmini: Trend Belirleyici Veri. Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi, 22 (1): 246-256.
- [27] Filiz E., Öz E. 2017. Classification Of BIST -100 Index' Changes Via Machine Learning Methods. Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi, 39 (1): 117-129.
- [28] Yakut E., Gemici E. 2017. Predicting Stock Return Classification through LR, C5.0, CART and SVM methods, and Comparing the Methods Used: An Application at BIST in Turkey. Ege Academic Review, 17 (4): 461-479.
- [29] Witten I.H., Frank E., Hall M.A., Pal C.J. 2017. Data mining: Practical machine learning tools and techniques with Java implementations (Fourth Edition). Morgan Kaufmann.
- [30] <https://tr.investing.com/> (Erişim Tarihi: 01.12.2020).
- [31] John G., Langley P. 1995. Estimating Continuous Distributions in Bayesian Classifiers. In proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers. San Matco.
- [32] Hosmer Jr D.W., Lemeshow S., Sturdivant R.X. 2013. Applied logistic regression (C. 398). John Wiley & Sons.
- [33] Şen Z. 2004. Yapay sinir ağları ilkeleri (1.). Su Vakfi, İstanbul.
- [34] Haykin S. 1994. Neural networks: A comprehensive foundation. Mc Millan.
- [35] Chen X.-W., Liu M. 2005. Prediction of protein-protein interactions using random decision forest framework. Bioinformatics, 21 (24): 4394-4400.
- [36] Filiz E., Öz E. 2019. Finding the best algorithms and effective factors in classification of Turkish science student success. Journal of Baltic Science Education, 18 (2): 239-253.
- [37] Gennari J.H., Langley P., Fisher D. 1989. Models of incremental concept formation. Artificial Intelligence, 40 (1): 11-61.
- [38] Gümüşçü A., Aydılek İ.B., Taştaltın R. 2016. Mikro-dizilim Veri Sınıflandırmasında Öznitelik Seçme Algoritmalarının Karşılaştırılması. Harran Üniversitesi Mühendislik Dergisi, 1 (1): 1-7.
- [39] Balaban M.E., Kartal E. 2015. Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları (1.). Çağlayan Kitabevi, İstanbul.
- [40] Donner A., Klar N. 1996. The statistical analysis of kappa statistics in multiple samples. Journal of Clinical Epidemiology, 49 (9): 1053-1058.
- [41] Bradley A.P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30 (7): 1145-1159.
- [42] Depren S.K., Aşkın Ö.E., Öz E. 2017. Identifying the Classification Performances of Educational Data Mining Methods: A Case Study for TIMSS. Educational Sciences: Theory & Practice, 17 (5): Article 5.
- [43] Christodoulou E., Ma J., Collins G.S., Steyerberg E.W., Verbakel J.Y., Van Calster B. 2019. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of Clinical Epidemiology, 110: 12-22.