

BIST-50 ENDEKSİ DEĞİŞİM DEĞERLERİNİN SINIFLANDIRILMASINDA MAKİNE ÖĞRENMESİ YÖNTEMLERİ VE YAPAY SİNİR AĞLARI KULLANIMI

Enes FILİZ¹
Hasan Aykut KARABOGA²
Serkan AKOGUL³

ÖZET

Çalışmada, BIST-50 endeksini etkileyen faktörlerden yararlanarak bazı makine öğrenmesi yöntemleri ve yapay sinir ağları ile sınıflandırmalar yapılmıştır. Literatürde sıklıkla kullanılan BIST-50 endeksinin değişimine etki eden faktörlerden DAX, FTSE, S&P 500, BISTBANK, BISTMALI, BISTSINAI, GOLDINDEX, Euro/Dolar Paritesi, Ham Petrol Fiyatı ve Faiz oranları alınmıştır. Sınıflandırmada makine öğrenmesi yöntemlerinden, k en yakın komşu algoritması (k-NN), Naive (basit) Bayes sınıflandırıcısı, C4.5 sınıflandırma algoritması ve yapay sinir ağları (YSA) kullanılmıştır. Weka3.8 programı yardımıyla analizler yapılmıştır. Bu analizler sonucunda %92.71 oranı ile C4.5 sınıflandırma algoritması doğru sınıflandırmada en iyi performansı göstermiştir.

Anahtar kelimeler: BIST-50 endeksi, Makine Öğrenmesi, Yapay Sinir Ağları, Sınıflandırma.

BIST-50 INDEX CHANGE VALUES CLASSIFICATION USING MACHINE LEARNING METHODS AND ARTIFICIAL NEURAL NETWORKS

ABSTRACT

In this study, BIST-50 index was classified with some selected machine learning algorithms and artificial neural networks. For this purpose, DAX, FTSE, S & P 500, BISTBANK, BISTMALI, BISTSINAI, GOLDINDEX, Euro / Dollar Parity, Crude Oil Price and interest rates were taken which are commonly used in the literature as factors that affect the BIST-50 index. Furthermore, k nearest neighbor algorithm (k-NN), Naive Bayes classifier, C4.5 classification algorithm and artificial neural networks (ANN) were used as classifier machine learning methods. Analyzes were performed with Weka3.8 program. At the end of analyzes, C4.5 classification algorithm has revealed the best performance with 92.71% in correct classification.

Keywords: BIST-50 index, Machine Learning, Artificial Neural Networks, Classification.

¹ Arş. Gör., Yıldız Teknik Üniversitesi, F.E.F., İstatistik, enesf@yildiz.edu.tr

² Arş. Gör., Yıldız Teknik Üniversitesi, F.E.F., İstatistik, karaboga@yildiz.edu.tr

³ Arş. Gör., Yıldız Teknik Üniversitesi, F.E.F., İstatistik, sakogul@yildiz.edu.tr

1. Giriş

Teknolojik gelişmelerdeki ivme ile birlikte dünya piyasaları birbirlerine hızla entegre olmuştur. Gerçekleşen gelişmeler şüphesiz İstanbul Borsası'nı da etkilemiştir. Borsanın bu gelişim ve dönüşüme uyumu için 2013 yılında yapılan bir değişiklikle İstanbul Menkul Kıymetler Borsası, Vadeli İşlemler Borsası, Opsiyon Borsası ve Altın Borsası tek çatı altında toplanarak Borsa İstanbul (BİST) adını almıştır (www.borsaistanbul.com). Bu değişim finansal piyasalarımızın uyumunu ve işlem yapma hızını arttırmıştır. Ayrıca hesaplanan endekslerin çeşitlenmesi ve daha iyi hesaplanması piyasa endekslerine olan güveni arttırmıştır.

Borsanın en temel yatırım araçlarından biri hisse senetleridir. Yatırımcılara yol gösterici olması için çeşitli endeksler hesaplanmaktadır. Bu endeksler yatırımcıların öngörü yapmasını sağlamanın yanında, genel piyasa ve sektör durumlarını anlamayı da kolaylaştırmaktadır. Genellikle birçok yatırımcı diğer yatırım araçlarına göre daha riskli olan hisse senetlerini tercih etmektedir. Şüphesiz bundaki en önemli etken riskin yüksek olmasına karşın getirinin de o denli yüksek olmasıdır. Ancak riskleri azaltmak için kullanılan çeşitli modeller ve analiz metotları mevcuttur (Karaboğa, 2015).

Borsa verilerinin, özellikle de hisse senetlerinin lineer olarak modellenememesi yapay zekâ ve makine öğrenmesi tekniklerinin kullanımının popüler hale gelmesini sağlamıştır (Egeli vd., 2003; Bahrammirzaee, 2010; Wang vd., 2011; Atsalakis & Valavanis, 2009). Verilerin lineer olarak modellenememesi, aykırı gözlemlerin ve değişen varyans sorununun modelin tahmin başarısını azaltmasından, otokorelasyon sorunundan veya modelleme hatalarından kaynaklanabilmektedir (Çelik, 2013). Ancak yine de bu yöntemler birbirini tamamlayıcı ve geliştirici etki göstermektedir. Dikkat edilecek olursa finansal analizlerde yoğun olarak kullanılan makine öğrenmesi yöntemleri ile veri madenciliği teknikleri iç içedir (Mitchell, 1997). Makine öğrenmesinin kullanımının kolaylaşması ile veri madenciliğine olan ilgi gittikçe artmaktadır. Ayrıca, insan beyninin makineye uyarlanması fikri ortaya çıktıktan sonra yapay zekâ ile ilgili önemli çalışmalar yapılmıştır.

Finansal piyasalarla ilgili literatürde yapay sinir ağları; döviz kurları veya hisse senedi fiyat tahmininde (Akcan & Kartal, 2011; Toraman, 2008; Erdoğan & Özyürek, 2012; Tektaş & Karataş, 2004; Karaboğa, 2015), hisse senedi getirilerinin tahmininde (Ünlü vd., 2009), çeşitli endeksler ve endeks getirilerinin tahmininde (Karaatlı vd., 2005; Aygören vd., 2012; Diler, 2003; Avcı, 2007) yaygın olarak kullanılmıştır. Sonuç olarak makine öğrenmesi ve yapay zekâ tekniklerinin daha az varsayım ile çalışmasından ötürü esneklik, genelleme, risk değerlendirme ve tahmin bakımından sağladığı üstünlük, kullanımını yaygınlaştırmıştır. Bu nedenle hisse senetlerinin, endekslerin fiyat ya da getirisinin tahmin edilmesinde yapay zekâ yöntemlerinin daha sık kullanıldığı görülmektedir. Bahrammirzaee (2010), çalışmasında en önemli yapay zeka tekniklerinden olan yapay sinir ağları, uzman sistemler ve hibrit akıllı sistemlerin finansal piyasalarda kullanımı ile ilgili kapsamlı bir literatür çalışması gerçekleştirmiştir. Bu amaçla finansal piyasaları kredi değerlendirme, portföy yönetimi ile finansal tahmin ve planlama olarak üç kısma ayırmıştır. Sonuçlar kesin olmamakla birlikte, finansal problemlerde doğrusal olmayan modeller bakımından yapay zeka yöntemleri geleneksel istatistiksel yöntemlere göre daha üstün olduğu belirtilmiştir. Pang vd. (2002), danışmanlı öğrenme yöntemleri ile ilgili en önemli çalışmalardan birini gerçekleştirmiştir. Bu çalışmada, film değerlendirmelerinde destek vektör

makinelere, Naive Bayes ve maksimum entropi sınıflandırıcıları kullanılmıştır. Solmaz vd. (2013), sınıflandırma yöntemlerinden destek vektör makinelere, lineer diskriminant analizi ve ileri beslemeli yapay sinir ağları; kümeleme yöntemlerinden de bulanık c-ortalama ve k-ortalama algoritmaları ile tiroit teşhisinde kullanımlarını incelemiştir. Çalışmanın sonucunda sınıflandırma yöntemlerinin kümeleme yöntemlerine göre daha başarılı olduğu ortaya konmuştur. Bulut (2016), çalışmasında karar ağaçları, k-en yakın komşu sınıflandırıcısı, Naive Bayes, destek vektör makinelere ve lojistik regresyon modelini kullanarak, bu yöntemlerin dengesiz veri setlerindeki başarılarını karşılaştırmıştır.

Literatür çalışmalarında makine öğrenmesi yöntemlerinin çoğunlukla BIST-100 veya BIST-30 endeksi tahmininde kullanıldığı görülmüştür. Son zamanlarda yapılan birçok çalışmada farklı yöntemler kullanılarak BIST-50 endeksi üzerinden piyasa analizleri yapılmıştır. Bu bağlamda Kayalidere vd. (2009), çalışmalarında İMKB’de fiyat ve hacim arasındaki nedensellik ilişkisini incelemiştir. Bu çalışmada Ocak 2001-Eylül 2008 döneminde İMKB-30 ve İMKB-50 endeksinde yer alan hisse senetleri kullanılmıştır. Seriler arasındaki nedensellik ilişkisi Granger Testi ile araştırılmıştır. Arslantaş ve Fındıklı (2010), çalışmalarında İMKB-50’ de bulunan şirketlerin yönetim kurulu yapılanmalarını incelemiştir. Akçayır vd. (2014), BIST-50 endeksinde yer alan hisse senetlerinin 2013 yılında 2 aylık kapanış verilerini kullanarak portföy seçim yöntemi ve Markowitz’ in ortalama-varyans modelinin uygulanabilirliğini test etmişlerdir.

Ayrıca BIST verileri, makine öğrenmesi ve yapay sinir ağları yöntemlerinin kullanıldığı çalışmalara sıklıkla konu olmuştur. Bu çalışmalarda genellikle BIST verilerinin sınıflandırılması ya da tahmin edilmesi amaçlanmaktadır. Özdemir vd. (2011), çalışmalarında İMKB-100 endeksinin getiri yönünü lojistik regresyon ve destek vektör makinelere yöntemlerini kullanarak tahmin etmiştir. Yakut vd. (2014), çalışmalarında BIST endeksinin, ileri beslemeli yapay sinir ağları ve destek vektör makinelere yöntemleri ile tahmin edilmesinde modellenebileceğini ortaya koymuştur. İlarıslan (2016), k-NN (k-en yakın komşu) algoritması ile BIST’ de işlem gören Afyon Çimento hisse senedi fiyatlarının bir gün sonrası için tahmin edilebileceğini %97 doğruluk oranı ile göstermiştir.

BIST-50 endeksi, ekonominin durumunu anlamak için önemli bir yere sahiptir. Bu endeksin artış ve azalışlarının doğru tahmini ya da doğru sınıflandırması yatırımcılar için karar vermede yol göstericidir. Bu nedenle çalışmanın uygulama kısmında, BIST-50 endeksi üzerinde etkili olan değişkenlerin bir önceki güne göre artış ve azalışları kullanılarak sınıflandırma yapılmıştır. Ayrıca literatürde yaygın olarak kullanılan makine öğrenmesi yöntemlerinden k-en yakın komşu algoritması (k-NN), Naive Bayes algoritması, C4.5 sınıflandırma algoritması ve yapay sinir ağlarının (YSA) başarıları da karşılaştırılmıştır. BIST-50 endeksinin değişimini etkileyen değişkenlerin seçimi ile ilgili bilgi verilmiştir.

2. Yöntemler

2.1 k-En Yakın Komşu Algoritması (k-NN)

k-en yakın komşu algoritması (k-Nearest Neighbour, k-NN), ilk kez Fix ve Hodges tarafından ortaya konmuştur. k-NN, parametrik olmayan ve sınıflandırma yapan eğitilmiş bir öğrenme algoritmasıdır (Balaban & Kartal, 2015, s60-61). Bu yöntem, var olan veri

setine yeni bir veri eklendiğinde, onun en yakın k komşusuna bakarak bu verinin sınıfını belirler. Yeni eklenen verinin daha önceki verilerden k tanesine uzaklığı incelenir. Bu iki veri arasındaki mesafe için genellikle Manhattan, Minkowski, Mahalanobis, ve Öklid gibi uzaklık ölçüleri kullanılır. Veriye en yakın mesafe neresi ise yeni veri o sınıfa atanır (Daş & Türkoğlu, 2014). Literatürde k -NN algoritmasının k değeri için farklı formüller yer almaktadır. Genel formülü aşağıdaki gibidir.

$$f(x) = \operatorname{argmax}_{c \in C} = \sum_{i=1}^k w_i \delta(c, f(x_i))$$

$f: R \rightarrow C$ olan $f(x)$ fonksiyonu x test noktasının sınıfını bulur. w_i , her bir örneğin hesaplamadaki ağırlığını, C ise eğitim setindeki sınıf etiketlerinin kümesi gösterir. $f(x_i)$, verideki x_i noktasının etiketinin ne olacağını belirtir. Eğer c $vef(x_i)$ birbirine eşit olursa $\delta(c, f(x_i)) = 1$, aksi halde ise $\delta(c, f(x_i)) = 0$ olur (Bulut, 2016).

2.2 Naive (Basit) Bayes Sınıflandırıcısı

Naive Bayes sınıflandırıcısı, makine öğrenmesinin danışmanlı öğrenme alt sınıfındadır. Yani veri setindeki örnek verilerin hangi sınıfa ait olduğu bellidir. Bu istatistiksel yöntem, her niteliğin sonuca etkisinin koşullu olasılığının hesaplanmasına dayanır (Balaban & Kartal, 2015, s.68-72). Ayrıca bu algoritma genellikle eldeki verinin, birleşik olasılıklar ile sınıflandırılmasında kullanılmaktadır (Türkoğlu vd., 2006). Naive Bayes sınıflandırıcısı, makine öğrenmesi yöntemleri ve veri madenciliği için en verimli ve etkin tümevarımsal öğrenme algoritmalarından biri olarak ön plana çıkmaktadır (Zhang, 2004).

Genel formülü aşağıdaki gibidir;

$$P(X|C) = \prod_{i=1}^n P(X_i|C)$$

Burada $i = 1, 2, \dots, n$ olmak üzere $X = (X_1, \dots, X_n)$ örnek uzayı, C ise örnek uzaydaki sınıf değerini belirtir (Balaban & Kartal, 2015, s.70).

2.3 C4.5 Sınıflandırma Algoritması

Karar ağaçları yöntemleri sınıflandırma problemlerinde sıklıkla kullanılan metotlardandır. Diğer makine öğrenmesi yöntemlerine göre uygulanabilirliği ve anlaşılması daha basittir. Sınıflandırma yapılırken veri seti kullanılarak karar ağacı oluşturulur, veriler teker teker ağaca uygulanır ve tahmin için gerekli olan kurallar ortaya çıkarılır (Silahtaroglu, 2013, s.68-69).

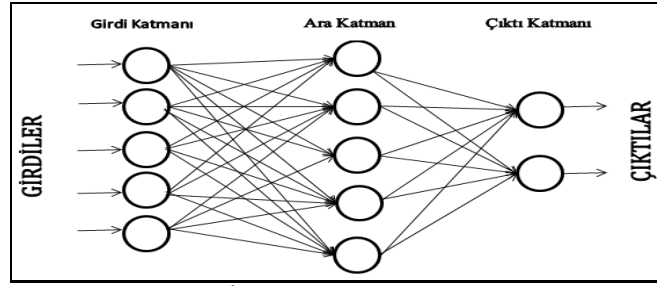
Karar ağaçlarında, ID3, C4.5, C 5, CART, SLIQ, CHAID, Random Forest gibi algoritmalar kullanılmaktadır. Bu algoritmaların temel amacı genelleştirme hatasını en aza indirerek eldeki veri setinden karar ağacı ortaya çıkarmaktır (Rokach & Maimon 2005).

C4.5 algoritması ilk olarak Quinlan tarafından 1993 yılında ortaya atılmıştır (Silahtaroglu, 2013, s.80-81). Bu algoritma, makine öğrenmesi yöntemlerinden bilgiyi temsil etmek amacıyla kullanılan klasik bir yoldur. Ayrıca veri yapılarını ifade etmek için güçlü bir çözüm yolu ortaya koyar (Nizam & Akın, 2014). Dahası C4.5 algoritması sayısal değerler alan nitelikler ile çalışma olanağı sağlar. Bu sayede nitelikleri iki ya da daha fazla biçimde ayırma fırsatı sunar (Balaban & Kartal, 2015, s.112).

2.4 Yapay Sinir Ağları (YSA)

Yapay Sinir Ağları (YSA), birçok sinir hücresinden oluşan ve bu hücrelerin eşanlı olarak hesaplama yapabilme yetenekleri bulunan bir bilgi işleme sistemidir. Bu sistem, seçilen algoritmaya göre girilen bilgileri çıktı ile eşleştiren, sınıflandıran veya tahmin amaçlı kullanan matematiksel modelleme tekniği olarak da adlandırılabilir. Sistemin sahip olduğu üstün özellikler işletmeden ekonomiye, inşaatın veri madenciliğine, bilgisayar mühendisliğinden harita mühendisliğine birçok alanda kullanılmasını sağlamaktadır.

Yapay sinir ağlarının insan beyninin taklit edilmesini sağlayan bilgisayar programları olmaları bu yöntemi popüler hale getirmiştir. Sistemin otonom çalışabilmesi de ayrıca gücünü arttırmaktadır. Bununla birlikte YSA' nın belirsiz, gürültülü ve eksik verileri de işleyebilmesi YSA' yı tahmin ve modelleme bakımından diğer tekniklere göre daha üstün kılmaktadır. Şekil 1'de YSA' nın en çok tercih edilen ağlarından biri olan ve çalışmamızda kullanılan Çok Katmanlı Algılayıcı (ÇKA) örneği verilmiştir. Bu model girdilerin alınarak belirli fonksiyonlar yardımıyla işlenip çıktı olarak sunulmasını sağlamaktadır (Karaboğa, 2015; Hamzaçebi, 2011; Elmas, 2010).



Şekil 1 İki katmanlı bir ÇKA örneği

3. Uygulama

Çalışmada makroekonomik göstergeler ve dünyada gerçekleşen siyasi, ekonomik ya da sosyal olaylar göz önünde bulundurularak değişkenler seçilmiştir. *Türkiye Cumhuriyet Merkez Bankası* tarafından açıklanan göstergeler, BIST içindeki şirketlerin yatırımlarında ve finansal planlamalarında büyük önem arz etmektedir. Bu göstergelerin başında faiz oranları, döviz kuru, BIST altın endeksi (GOLDINDEX) gibi önemli değişkenler gelmektedir. (Aydeniz, 2009; Özdemir vd., 2011) Ayrıca küreselleşme nedeniyle dünyada meydana gelen ekonomik ve siyasi olaylar dünya borsalarını etkilemektedir. Literatürde BIST endeksindeki değişimler ile gelişmiş ya da gelişmekte olan ülkelerin borsaları arasında uzun vadeli ilişkiler olabileceği düşünülmektedir (Vuran, 2010). Ekonomik bakımdan dünyanın en büyük şirketlerini barındıran Amerikan Borsa Endeksi (S&P 500), Almanya Borsa Endeksi (DAX) ve İngiltere Borsa Endeksi (FTSE) çalışmada işleme alınmıştır. “Siyah Altın” olarak kabul edilen ham petrol, sanayi, ulaşım, otomobil, kozmetik, enerji, kimya, ilaç sektörleri dâhil çok çeşitli alanlarda kullanılmakta olup Dünya ekonomisinde de çok önemli yere sahiptir. Ham petrol fiyatları, ekonomik, siyasi ve sosyal gelişmelerden etkilenmektedir (<https://www.gcmforex.com>). Borsa İstanbul Bankacılık endeksi (BISTBANK), Borsa

İstanbul Mali (finans) endeksi (BISTMALI), Borsa İstanbul Sanayi endeksi (BISTSINAI) BIST endeksinin öncü sektör endeksleri arasındadır. Herhangi bir sektör bazında ortaya çıkabilecek olumsuz durumlar sektörleri etkileyebilmektedir (Kendirli & Çankaya, 2016; Kasman, 2003). Analizler, 03.01.2006–01.12.2016 tarihleri arasında toplamda 2591 işlem gününün verileri kullanılarak yapılmıştır. Kullanılan değişkenler sırasıyla BIST-50, DAX, FTSE, S&P 500, BISTBANK, BISTMALI, BISTSINAI, GOLDINDEX, Euro/Dolar Paritesi, *Ham Petrol fiyatı ve **Faiz oranlarıdır. BIST-50 endeksi, DAX, FTSE, S&P 500, BISTBANK, BISTMALI, BISTSINAI ve petrol fiyatları verileri "<https://tr.investing.com/>" isimli piyasa hareketlerini inceleyen internet sitesinden, Faiz oranları, Euro/Dolar Paritesi ve GOLDINDEX verileri ise *Türkiye Cumhuriyet Merkez Bankası* internet sitesinden alınmıştır.

Değişkenlerin bir önceki güne göre artış ya da azalışları 1 ve 0 olarak kodlanmıştır. Fakat faiz oranları aylık olarak açıklandığı için ay içindeki tüm günlere eşit olarak o ayki faiz oranı değeri yazılmıştır. Ayrıca BIST-50 endeksi bağımlı değişken, diğer değişkenler ise bağımsız değişken olarak işleme alınmıştır. Sonuçlar Weka3.8 programı yardımıyla ortaya çıkarılarak bu çalışma için en başarılı sınıflandırma yöntemi belirlenmiştir. Karar ağaçlarından C4.5 algoritması, Bayes sınıflandırma yöntemlerinden Naive Bayes, en yakın komşu algoritmalarından k-NN ve Yapay zeka algoritmalarından da YSA çalışmada kullanılmıştır. Bu yöntemler farklı algoritmalara sahip olduğu için tercih edilmiştir.

Makine öğrenmesi yöntemlerinde veri seti eğitim ve test grubu olarak ikiye ayrılır. Literatürde eğitim seti ve test grubu sırasıyla %75 - %25, %90 - %10 vb. gibi değerler alınmıştır, fakat daha yaygın kullanılan k-katlı çapraz doğrulama yöntemidir. k-katlı çapraz doğrulama yönteminde veri seti k eşit parçaya ayrılır ve bu k adet parçanın k-1 tanesi ile eğitim, kalan bir tanesi ile test gerçekleştirilir. Bu işlem her bir parçanın test kümesi olarak alınması suretiyle k kez tekrarlanır. Elde edilen tüm sonuçların ortalaması alınarak sınıflandırma ölçülerinin değerleri belirlenmiş olur (Öz vd., 2016).

Bu çalışmada literatürde sıklıkla kullanılan, k değeri için 10, 5 ve 2 alınmıştır. k=10 için 2591 işlem gününün verileri ile 10-katlı çapraz doğrulama yapılmıştır. 10 parçanın 9 tanesi eğitim seti, 1 tanesi test grubu olarak seçilmiştir. 2332 işlem günü verisi eğitim setinde, 259 işlem günü verisi de test grubunda yer almaktadır. Bu işlem her defasında test grubunu değiştirmek suretiyle 10 kez tekrar edilir. Aynı işlemler k=5 ve k=2 içinde yapılmıştır (Erpolat & Öz, 2010).

Çalışmada kullanılacak doğru ya da yanlış sınıflandırılmış değerleri gösteren kontenjans tablosu Tablo 3.1' de verilmiştir.

Tablo 3.1 Kontenjans tablosu

		Tahmin	
		a	b
Gerçek Değer	a	DP (Doğru Pozitif)	YN (Yanlış Negatif)
	b	YP (Yanlış Pozitif)	DN (Doğru Negatif)

* West Texas Intermediate Ham Petrol Vadeli İşlemleri (varil fiyatı)

** 2006 yılı Ocak ayından itibaren bankalarca Türk lirası üzerinden açılan mevduat ağırlıklı ortalama faiz oranları (1 aya kadar vadeli)

Burada;

DP(Doğru Pozitif); gerçekte artış gösteren örneklerin model tarafından artış olarak tahmin ettiği değer sayısıdır.

YP(Yanlış Pozitif); gerçekte azalış gösteren örneklerin model tarafından artış olarak tahmin ettiği değer sayısıdır.

DN(Doğru Negatif); gerçekte azalış gösteren örneklerin model tarafından azalış olarak tahmin ettiği değer sayısıdır.

YN(Yanlış Negatif); gerçekte artış gösteren örneklerin model tarafından azalış olarak tahmin ettiği değer sayısıdır.

Tablo 3.1’de yer alan ifadeler kullanılarak *ACC* (*Doğruluk*) değeri hesaplanır. Bu değer, doğru sınıflandırma oranlarını karşılaştırmada etkin rol oynamaktadır. *ACC*; modeldeki doğru sınıflandırılmış olan tüm örneklerin, toplam örnek sayısına oranı ile tespit edilir. Literatürde yaygın olarak kullanılan formül aşağıda verilmiştir (Balaban & Kartal, 2015, s.49; Nizam & Akın, 2014).

$$ACC = \frac{DP + DN}{DP + DN + YP + YN}$$

Tablo 3.2’ de çalışmada kullanılan yöntemlerin doğru ya da yanlış sınıflandırılmış örnek sayıları verilmiştir. Bu sayılar kullanılarak yöntemlerin doğru sınıflandırma oranları Tablo 3.3’ te gösterilmiştir.

Tablo 3.2 Yöntemler için Kontenjans Tablosu (k=10, k=5, k=2)

	Sınıflar	k-NN		Naive Bayes		C4.5		YSA	
		a	b	a	B	a	b	a	b
k=10	a=1	1204	134	1215	123	1239	99	1231	108
	b=0	127	1126	78	1175	97	1156	99	1154
k=5	a=1	1206	132	1214	124	1239	99	1230	108
	b=0	123	1130	78	1175	99	1154	106	1147
k=2	a=1	1202	136	1214	124	1236	102	1230	108
	b=0	134	119	76	1177	87	1166	108	1145

Tablo 3.3 Yöntemlerin Sınıflandırma Başarıları (%) (k=10, k=5, k=2)

		k-NN	Naive Bayes	C4.5	YSA
		k=10	Doğruluk (ACC)	89,93	92,24
k=5	Doğruluk (ACC)	90,16	92,20	92,36	91,74
k=2	Doğruluk (ACC)	89,58	92,28	92,71	91,66

Not: En başarılı sınıflandırma oranı koyu renkle belirtilmiştir.

Tablo 3.3' e göre, k=10 için k-NN algoritmasının sınıflandırma başarısı %89.93, Naive Bayes algoritmasının başarılı sınıflandırma oranı %92.24, C4.5 algoritmasının başarılı sınıflandırma oranı %92.44, YSA ise başarılı sınıflandırma oranı %92.05 olduğu görülmektedir. k=5 için k-NN %90.16, Naive Bayes %92.20, C4.5 %92.36, YSA ise %91.74 olarak bulunmuştur. k=2 için ise k-NN %89.58, Naive Bayes %92.28, C4.5 %92.71, YSA da %91.66 sonuçları elde edilmiştir.

Ayrıca incelenen bir diğer durum bazı öznelik seçim algoritmaları ile BIST-50 endeksini etkileyen değişkenlerin önem sırasıyla bulunmasıdır. Kullanılan algoritmalar One R, Gain Information ve Korelasyon tabanlı öznelik seçim algoritmalarıdır. Kullanılan üç algortmada aynı sonuçları vermiştir. Bu algoritmalara göre kullanılan değişkenler bakımından, BIST-50 endeksinin değişim yönünü etkileyen faktörlerin önem sırası BISTMALI, BISTBANK, BISTSINAI, FTSE, DAX, S&P 500, Ham Petrol fiyatı, Euro/Dolar Paritesi, GOLDINDEX ve Faiz oranları olarak bulunmuştur.

4. Sonuç

Ülkenin lokomotifi konumunda bulunan en büyük 50 şirket İstanbul Borsası'nın BIST-50 endeksinde taranmaktadır. Borsa İstanbul A.Ş. tarafından belirlenen koşulları yerine getiren şirketler bu endekse dâhil edilirken, bu koşulları yerine getiremeyen şirketler endeksten çıkartılmaktadır. Ekonominin durumunu anlamak için de kullanılan bu endeks gösterge niteliğindedir. Çünkü BIST-50 endeksine hemen hemen tüm sektörlerden belirli ağırlıklarda katkı yapılmaktadır. Borsa yatırımcılarının endeksin artış azalışlarını doğru tahmin etmesi, doğru sınıflandırması borsada doğru pozisyon almalarını ve böylece kazançlarını arttırmalarını sağlayacaktır. Bu nedenle literatürde BIST-50 endeksi üzerinde etkili olan değişkenlerin günlük artış ve azalışları kullanılarak farklı sınıflandırma yöntemlerinin başarısı irdelenmiştir. Çalışmada, BIST-50 endeksini etkileyen faktörlerden yararlanarak bazı makine öğrenmesi yöntemleri ve yapay sinir ağları ile sınıflandırmalar yapılmıştır.

Tablo 3.3' e göre k=5 için k-NN algoritmasının (%90.16), k=2 için Naive Bayes (%92.28) ve C4.5 algoritmasının(%92.71), k=10 için YSA'nın (%92.05) en iyi sonucu verdiği söylenebilir. Sonuçlara göre tüm k değerleri için en başarılı sınıflandırmayı C4.5 algoritması ve en düşük başarılı sınıflandırmayı da k-NN algoritması gerçekleştirmiştir. Ayrıca k değerine göre bakıldığında; k-NN için en iyi değeri k=5, Naive Bayes ve C4.5 algoritması için k=2, YSA için de k=10 vermiştir. Benzer değişken yapısına sahip çalışmalarda C4.5 algoritması için k değerinin küçük seçilmesi başarılı sonuçlar verebilir. Bir diğer sonuç ise, bazı öznelik seçim algoritmaları ile BIST-50 endeksinin değişim yönünü etkileyen faktörlerin önem sırası BISTMALI, BISTBANK, BISTSINAI, FTSE, DAX, S&P 500, Ham Petrol fiyatı, Euro/Dolar Paritesi, GOLDINDEX ve Faiz oranları olarak bulunmuştur.

Sonuç olarak BIST-50 endeksini etkileyen değişkenlerin değerleri bilindiği durumda BIST-50 endeksinin değişiminin yönünü yaklaşık %89-%93 arası başarı ile makine öğrenmesi ve YSA yöntemi yardımıyla doğru sınıflandırıldığı söylenebilir. BIST-50 endeksinin yönünü bulabilmek yatırımcıya gün içinde daha başarılı kararlar vermesine ve doğru günlük planlama yapmasına yardımcı olacaktır. Sınıflandırma başarısı en yüksek algoritmanın %92.71 (Tablo 3.3'te koyu ile belirtilmiştir) ile C4.5 olması, bu algoritma gibi basit ve anlaşılabilir makine öğrenmesi yöntemlerinin yatırımcılar tarafından kolaylıkla kullanılabileceğini ortaya koymaktadır.

KAYNAKLAR

Akcan, A. ve Kartal, C., (2011), The Forecasting of Stock Prices in ISE Insurance Index with Artificial Neural Networks, *The Journal of Accounting and Finance*, 51:27-40.

Akçayır, Ö., Doğan, B. ve Demir, Y., (2014), Elton-Gruber Kısıtlı Markowitz Kuadratik Programlama Modeli İle Portföy Optimizasyonu: BIST-50 Üzerine Bir Uygulama, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19(3):333-352.

Arslandaş, C. C. ve Fındıklı, M. A., (2010), İMKB-50'de Yer Alan Şirketlerin Yönetim Kurulu Yapılanmaları, *Istanbul University Journal of the School of Business Administration*, 39(2): 258-275.

Atsalakis, G. S. ve Valavanis, K. P., (2009), Surveying stock market forecasting techniques–Part II: soft computing methods, *Expert System Application*, 36:5932–5941.

Avcı, E., (2007), Forecasting Daily And Sessional Returns Of The ISE-100 Index with Neural Network Models, *Doğuş Üniversitesi Dergisi*, 128-142.

Aydeniz, E. Ş., (2009), Makroekonomik Göstergelerin Firmaların Finansal Performans Ölçütleri Üzerindeki Etkisinin Ölçülmesine Yönelik Bir Araştırma: İmkb'ye Kote Gıda Ve İçecek İşletmeleri Üzerine Bir Uygulama, *İktisadi ve İdari Bilimler Dergisi*, 27(2):263-277.

Aygören, H., Sarıtaş, H. ve Moralı, T., (2012), Forecasting ISE 100 Indice Using Artificial Neural Networks and Newton Numerical Search Models, *International Journal of Alanya Faculty of Business*, 4(1):73-88.

Bahrammirzaee, A., (2010), A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems, *Neural Computing and Applications*, 19:1165–1195.

Balaban, M., E. ve Kartal, E., (2015), *Veri Madenciliği Ve Makine Öğrenmesi*, Çağlayan Kitabevi, İstanbul.

Borsa İstanbul, <http://www.borsaistanbul.com/>, Erişim Tarihi: 15.01.2017

Bulut, F., (2016), Performance evaluations of supervised learners on imbalanced datasets, *In Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), IEEE*, 1-4.

Çelik, R., (2013), Kelebek Şeklinde Dağılan Artıklar İçin Genel Varyans Dengeleme Yöntemlerinin Karşılaştırılması, *Trakya University Journal of Social Science*, 15(1):63-84.

Daş, B. ve Türkoğlu, İ., (2014), DNA Dizilimindeki Nükleotit Çiftlerinin Frekans Değerlerine Göre Farklı Sınıflandırma Yöntemleri ile Karşılaştırılması, *Tıp Teknolojileri Ulusal Kongresi (TIPTEKNO'2014)*, 191-194.

Diler, A. İ., (2003), İMKB ULUSAL-100 Endeksinin Yönünün Yapay Sinir Ağları Hata Geriye Yayma Yöntemi İle Tahmin Edilmesi, *İMKB Dergisi*, 7(25-26):65-84.

Egeli, B., Ozturan, M. ve Badur, B., (2003), Stock Market Prediction Using Artificial Neural Networks, *Proceedings of the 3rd Hawaii International Conference on Business*, Hawaii, USA.

Elmas, Ç., (2010), *Yapay Zekâ Uygulamaları*, 2. Baskı, Seçkin Yayıncılık, Ankara.

Erdoğan, E. ve Özyürek, H., (2012), Estimating Stock Prices By Neural Networks, *Sosyal ve Beşeri İlimler Dergisi*, 4:1309-8012.

Erpolat, S. ve Öz, E., (2010), Kanser Verilerinin Sınıflandırılmasında Yapay Sinir Ağları ile Destek Vektör Makinelerinin Karşılaştırılması, *İstanbul Aydın Üniversitesi Fen Bilimleri Dergisi*, 2: 71-83.

Fix, E. ve Hodges Jr, J. L., (1951). Discriminatory analysis-nonparametric discrimination: consistency properties. *California Univ Berkeley*.

GCMFOREX, <https://www.gcmforex.com>, Erişim Tarihi: 15.01.2017

Hamzaçebi, C., (2011), *Yapay Sinir Ağlarının Tahmin Amaçlı Kullanımı: Matlab ve Neurosolutions Uygulamalı*, Ekin Yayınevi, Bursa.

İlarslan, K., (2016), k-En Yakın Komşu (k-NN) Algoritması ile Hisse Senedi Fiyatlarının Tahmin Edilmesi: BİST' den Örnek Bir Uygulama, *Asos journal*, 4(30):375-392.

Investing, <http://tr.investing.com>, Erişim Tarihi: 15.01.2017

Karaath, M., Güngör, İ., Demir, Y. ve Kalaycı, S., (2005), Estimating stock market movements with neural network approach, *Journal of Balikesir University*, 2(1):22-48.

Karaboğa, H. A., (2015), *Dalgacık Dönüşümü Kullanılarak Hisse Senedi Fiyat Tahmini Üzerine Bir Uygulama*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Kasman, S., (2003), The relationship between exchange rates and stock prices: A causality analysis, *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 5(2): 70-79.

Kayalıdere, U. A. K., Kargın, S. ve Aktaş, R., (2009), İMKB'de Fiyat ve Hacim Arasındaki Nedensellik İlişkisi, *Celal Bayar Üniversitesi SBE Sosyal Bilimler Dergisi*, 7:115-124.

Kendirli, S. ve Çankaya, M., (2016), Döviz Kuru ve Enflasyonun BİST Banka Endeksi Üzerindeki Etkisi, *MANAS Sosyal Araştırmalar Dergisi/MANAS Journal of Social Studies*, 5(3): 215-227.

Mitchell, T. M., (1997), *Machine Learning*, Burr Ridge, IL: McGraw Hill, 45(37):870-877.

Nizam, H. ve Akın, S. S., (2014), Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması, *XIX. Türkiye'de İnternet Konferansı*

Öz, E., Kurt, S., Asyalı, M. H., Kaya, H. ve Yücel, Y., (2016), Feature based quality assessment of DNA sequencing chromatograms, *Applied Soft Computing*, 41: 420-427.

Özdemir, A. K., Tolun, S. ve Demirci, E., (2011), Endeks Getirisi Yönünün İkili Sınıflandırma Yöntemiyle Tahmin Edilmesi: İMKB-100 Endeksi Örneği, *Niğde Üniversitesi İİBF Dergisi*, 4(2):45-59.

Pang, B., Lee, L. ve Vaithyanathan, S., (2002), Thumbs up? Sentiment Classification using Machine Learning Techniques, *In Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, Association for Computational Linguistics, 10:79-86.

- Quinlan, J. R., (1996), Improved use of continuous attributes in C4. 5, *Journal of artificial intelligence research*, 4: 77-90.
- Rokach, L. ve Maimon, O., (2005), *Decision Trees*, Data Mining and Knowledge Discovery Handbook, Springer, 165-192.
- Silahtaroglu, G., (2013), *Veri Madenciliği Kavram ve Algoritmaları*, Papatya Yayıncılık, İstanbul.
- Solmaz, R., Günay, M. ve Alkan, A., (2013), Uzman Sistemlerin Tiroit Teşhisinde Kullanılması, *XV. Akademik Bilişim Konferansı Bildirileri*, 23-25 Ocak 2013, Antalya.
- Tektaş, A. ve Karataş, A., (2004), Yapay Sinir Ağları ve Finans Alanına Uygulanması: Hisse Senedi Fiyat Tahminlemesi, *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 3(4):337-349.
- Toraman, C., (2008), Stock Price Forecasting by Using Artificial Neural Networks in Iron-Steel Sector: A Forecasting Application on Erdemir Co. and Kardemir Co., *The Journal of Accounting and Finance*, 39:44-57.
- Türkiye Cumhuriyeti Merkez Bankası, <http://www.tcmb.gov.tr>, Erişim Tarihi: 15.01.2017.
- Türkoğlu, F., Diri, B. ve Amasyalı, M. F.,(2006), Farklı Özellik Vektörleri ile Türkçe Dokümanların Yazarlarının Belirlenmesi, *15. Turkish Symposium on Artificial Intelligence and Neural Network (TAINN)*, Muğla.
- Ünlü, U., Yıldız, B. ve Yalama, A., (2009), İlk Halka Arzlarda Uzun Dönem Getirilerinin Tahmini: Yapay Sinir Ağları İle IMKB İçin Ampirik Bir Çalışma, *İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi*, 10:29-47.
- Wang, J. Z., Wang, J. J., Zhang, Z. G. ve Guo, S. P., (2011), Forecasting Stock Indices with Back Propagation Neural Network, *Expert System Application*, 38:14346–14355.
- Vuran, B., (2010), IMKB 100 endeksinin uluslararası hisse senedi endeksleri ile ilişkisinin eşbütünleşim analizi ile belirlenmesi, *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 39(1): 154-168.
- Yakut, E., Elmas B. ve 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.
- Zhang, H., (2004), The Optimality of Naive Bayes, *A A*, 1(2):3.