

# LR, C5.0, CART, DVM Yöntemlerini Kullanarak Hisse Senedi Getiri Sınıflandırma Tahmini Yapılması ve Kullanılan Yöntemlerin Karşılaştırılması: Türkiye’de BIST’de Bir Uygulama

*Predicting Stock Return Classification through LR, C5.0, CART and SVM methods, and Comparing the Methods Used: An Application at BIST in Turkey*

Emre YAKUT<sup>1</sup>, Eray GEMİCİ<sup>2</sup>

## ÖZET

Hisse senedi getiri sınıflandırma tahmini her zaman için yatırımcıların ve analizcilerin ilgisini çeken bir araştırma alanı olmuştur. Bu çalışmada BIST 100 endeksinde işlem gören kimya, kauçuk ve plastik ürünleri imalatı sanayinde yer alan, faaliyetleri 2009-2014 yılları arasında süreklilik gösteren 18 şirketin hisse senedi getirilerinde etkili olan faktörler belirlenerek, hisse senedi getirileri tahmin edilmeye çalışılmıştır. Söz konusu verilerin veri madenciliği yöntemlerinden olan LR analiz, C5.0 algoritması, CART algoritması ve DVM yöntemleri kullanılarak analiz işlemleri gerçekleştirilmiş, hisse senedi getiri sınıflandırma tahmininde anlamlı ve faydalı bilgileri ortaya çıkarmak için karar ağacına ait kurallar elde edilmiştir. Yapılan analizler sonucunda LR analizi %75, C5.0 algoritması %88, CART algoritması %89,8 ve DVM analizi %75,9'luk doğru sınıflandırma başarısı gerçekleştirmiştir. Pozitif ve negatif hisse senedi getiri sınıflandırma tahminine etki eden en önemli değişkenlerin "piyasa/defter değeri değişkeni", "TÜFE değişkeni" ve "brüt kar marjı değişkeni" olduğu saptanmıştır. Yatırımcılar ve analizciler için önerdiğimiz modelin değişkenleri ile birlikte hisse senedi getiri tahmininde kullanılmasının uygun olabileceği gözlenmiştir.

**Anahtar Kelimeler:** Hisse Senedi Getiri Sınıflandırma Tahmini, LR Analizi, C5.0, CART Algoritması, DVM.

## ABSTRACT

Predicting stock return classification is a research field that has always attracted the attention of investors and analysts. In this study, the factors that affected the stock returns were determined for 18 companies which were dealt at BIST 100 index, and operating to manufacture chemistry, rubber and plastic products, and were active between 2009 and 2014, after which their stock returns were predicted. Of all the data mining methods, relevant data were collected, and analyzed through LR analysis, C5.0 algorithm, CART algorithm and SVM methods. Accordingly, some rules were obtained out of the decision tree in order to reveal significant and beneficial information for predicting the stock return classification. As a result of the analyses, LR analysis showed 75% of success, C5.0 algorithm showed 88% of success, CART algorithm showed 89.8% of success, and SVM analysis showed 75.9% of success for accurate classification. It was determined that the most important variables influencing positive and negative predictions of the stock return classification were 'market/book value variable', 'CPI variable' and 'gross profit margin variable'. Using the studied model and its variables for predicting stock return classification was observed to be convenient and was suggested for investors and analysts.

**Keywords:** Predicting Stock Return Classification, LR Analysis, C5.0, CART Algorithm, SVM.

<sup>1</sup>Yrd. Doç. Dr., Osmaniye Korkut Ata Üniversitesi, emreyakut@osmaniye.edu.tr

<sup>2</sup>Yrd. Doç. Dr., Gaziantep Üniversitesi, gemici@gantep.edu.tr

## 1. GİRİŞ

Fama (1970), etkin piyasayı; fiyatların elde edilebilen bilgileri tam olarak yansıttığı bir piyasa olarak tanımlamıştır. Sermaye piyasaları üzerinde normal üstü getiri sağlama anlamında kullanılan anomali çerçevesinde yapılan ampirik çalışmalar, yatırımcıların bazı işlem stratejilerini izleyerek ortalamanın üzerinde getiri elde ettiğini ortaya koymuştur. Bu konuda Türkiye’de yapılan çalışmalar, genellikle sermaye piyasalarının zayıf formda etkin olup olmadığını belirlemek üzere yapılmıştır. Elde edilen sonuçların bazıları (Bekçioğlu ve Ada, 1985; Kıyılar, 1997; Sevim vd. 1997; Kahraman ve Erkan, 2005; Tunçel, 2007) gelişmekte olan borsamızın zayıf formda dahi etkin olmadığına yani geçmiş fiyat hareketleri yardımıyla yapılan teknik analizin yatırımcılara normalüstü getiri sağlanabileceğini göstermiştir. Borsamızın zayıf formda etkin olduğunu destekleyen çalışmalara da rastlamak mümkündür; (Buguk ve Brorsen, 2003; Bekçioğlu vd. 2004; Zengin ve Kurt, 2004; Eken ve Adalı, 2008; Ergül, 2009; Atan vd. 2009; Zeren vd. 2013).

Yatırımcıların beklenen getirilerini belirleyen faktörlerin bir kısmının işletmeden kaynaklanan ve sistematik olmayan risk adıyla bilinen faktörler olduğu, diğer kısmının ise ekonominin genel seyrinden kaynaklanan makro olaylara bağlı faktörlerden kaynaklandığı finans alanında kabul gören bir yaklaşımdır. Bütün menkul kıymetlerle piyasa arasında var olan doğrusal ilişkinin doğrusal regresyon modeliyle ifade edilebileceği, burada menkul kıymet getirisinin hesaplanmasında, menkul kıymet getirisinin bağımlı değişken, piyasa endeksinin veya daha fazla sayıda makroekonomik değişkenin ise bağımsız değişkenler olarak kullanıldığı çoklu regresyon modeliyle hesaplanabildiği yine finans alanında kabul gören bir yaklaşımdır. Dolayısıyla, finansal varlık getirilerinin piyasa endeksi dışında çok sayıda değişkenin etkisi altında olduğu kabul edilmektedir. Bu değişkenler, sanayi sektörüne ait endeksler olabileceği gibi, faizler, enflasyon gibi diğer makro değişkenler olabilmektedir (Elton ve Gruber, 1981).

Çalışmada hisse senedi getirilerini tahmin ederken şirketin kendisi tarafından oluşturulan mikro ekonomik değişkenlerle, TÜFE gibi sistematik risk kaynakları bir arada değerlendirilmiş böylelikle hisse senedi getirilerinde etkili olan faktörler belirlenerek, şirketlerin hisse senedi getirileri tahmin edilmeye çalışılmıştır.

## 2. LİTERATÜR ÖZETİ

Hisse senedi getiri tahminine yönelik literatürde farklı analiz yöntemleri denenmiştir. Bu yöntemlerden yapay sinir ağları (YSA) yöntemine dayalı olarak yapılan çalışmalar (Tektaş ve Karataş, 2004; Altay ve Satman, 2005; Panda ve Narasimhan, 2006; Avcı ve Çinko, 2008; Yıldız vd. 2008; Toraman, 2008; Avcı, 2009; Ege ve Bayrakdaroğlu, 2009; Akcan ve Kartal; Cao vd. 2011; Dayı ve Ata, 2012; De Olivera vd. 2013) ve diğer istatistiksel tekniklere dayalı olarak yapılan çalışmalar (Öz vd. 2011; Özdemir vd. 2011; Dutta vd. 2012; Tayyar ve Tekin, 2013) hisse senedi getirisini tahmin etmek için gerçekleştirilmiştir. Çalışmaların çoğunda YSA yöntemi diğer analiz tekniklerine göre daha iyi sonuçlar ürettiği gözlenmiştir.

Veri madenciliği ile ilgili hisse senedi getiri tahminine ilişkin çalışmalar da mevcuttur. Enke ve Thawornwong (2005), Amerika’nın S&P 500 borsasında pozitif ve negatif getirisi olan hisse senetlerini sınıflandırmak için veri madenciliği yöntemlerinden gain algoritmasını kullanmışlardır. Huang vd. (2008), Kore ve Tayvan borsasında 1990-1991 yılları arasındaki günlük hisse senedi fiyat endeksinden hareketle, hisse senedi fiyat endeksini sınıflandırmak için 23 teknik endeksi kullanarak DVM, YSA, K en yakın komşu algoritması, C4.5 algoritması ve LR tekniklerinin sınıflandırma sonuçlarını karşılaştırmışlardır. İnce ve Trafalis (2008), Newyork’ta NASDAQ borsasındaki 10 farklı firmanın 2000-2006 yılları arasındaki dönemde hisse senedi fiyat hareketlerini belirlemek için SVM ve YSA analizlerini uygulamışlardır. Risk kavramının karşılaştırma kriteri baz alındığında SVM tekniğinin YSA metodu kadar iyi sonuç verdiğini ifade etmişlerdir. Lai vd. (2009), Tayvan borsasında hisse senedi fiyat hareketini belirlemek için bulanık karar ağacı, genetik algoritma ve veri kümeleme tekniğini entegre ederek önerdikleri modelin ortalama %82’lik bir performans gösterdiğini bahsetmişlerdir. Ou ve Wang (2009), Hong Kong borsasında Hang Seng endeksinin hareket yönünü tahmin etmek için veri madenciliği tekniklerinin sınıflandırma sonuçlarını karşılaştırmışlardır. Yapılan analizler sonucunda DVM yönteminin diğer yöntemlere göre daha iyi sonuç verdiklerini ileri sürmüşlerdir. Cheng vd. (2010), Tayvan borsasında hisse senedi getirisindeki değişime ilişkin araştırmaya yönelik C4.5 algoritması ile olasılıklı yapay sinir ağları yöntemlerinin bir birleşimi olan hibrid sınıflandırma tekniğini kullanarak %76’lık doğru sınıflandırma başarısı elde etmişlerdir. Tsai ve Hsiao (2010), Tayvan borsasında işlem gören elektronik endüstrisine ait hisse senetlerinin fiyatını tahmin

etmek ve tahminde etkili olan faktörleri belirlemek için temel bileşenler analizi, genetik algoritma ve CART algoritmasının birleşiminden oluşan bir filtreleme tekniğini kullanarak %79’luk başarı gerçekleştirmişlerdir. Chang (2011), Tayvan borsasında işlem gören 10 farklı hisse senedinin 2008-2009 yılları arasındaki fiyatını belirlemek için ANN, CART algoritması, ANN-CART birleşimli hibrit bir metot kullanarak tahmin etmeye çalışmışlardır ve sonuç olarak ANN yönteminin diğer iki yöntemle göre daha iyi sonuç verdiğini ifade etmişlerdir. Kara, vd. (2011), Türkiye’de BIST 100 endeksinde 1997-2007 yılları arasındaki hisse senetlerine ait fiyat endeksinin yönünü tahmin etmek için YSA ve DVM modellerini kullanmışlardır. Yaptıkları analizler sonucunda YSA modelinin DVM modeline göre daha başarılı olduğu sonucuna ulaşmışlardır. Tsai vd. (2011), Tayvan borsasında 2002-2006 yılları arasındaki 5 yıllık dönem için 2511 şirketin (1052 pozitif ve 1459 negatif getiri) hisse senedi getirilerinin tahmin edilmesinde 19 adet finansal oran ve 11 adet ekonomik göstergeler olmak üzere toplam 30 adet değişkeni kullanmışlardır. Hisse senedi getirilerini pozitif ve negatif olarak sınıflandırmaya ilişkin YSA, CART ve LR tekniklerini karşılaştırmışlardır. Yapılan analizler neticesinde YSA tekniğinin, CART ve LR tekniklerinden daha yüksek doğrulukla sınıflandırma gerçekleştirdiğini tespit etmişlerdir. Hargreaves ve Hao (2013), Avustralya borsasında son üç aya ait 200 şirketin hisse senedi performanslarının tahmini için 5 adet değişken ile C5.0 algoritması, CHAID algoritması, YSA ve Lojistik Regresyon analizlerini karşılaştırarak, tüm modeller içerisinde C5.0 algoritmasının daha yüksek doğrulukla en iyi performansa sahip olduğunu ileri sürmüşlerdir. Zhang, vd. (2014), Şangay borsasında hisse senedi tahmin modelinin performansını artırmak için özellik seçimine dayalı nedensel özellik seçim (causal feature selection-CFS) tekniği ile birlikte özellik seçim algoritmalarından temel bileşenler analizi, CART algoritması ve LASSO tekniklerini kullanarak hisse senedi tahmin modelleri geliştirmişlerdir. Hisse senedi tahmin modelleri sonucunda CFS tekniğinin diğer seçim algoritmalarına göre daha uygun ve istikrarlı olduğunu ifade etmişlerdir. Barak ve Modarres (2015), Tahran borsasında 2002-2010 yılları arasındaki dönemde 400 şirkete ait hisse senetlerin getiri ve risk değerlerini tahmin etmek için 3 aşamalı bir model geliştirmişlerdir. Veri madenciliği tekniklerinden karar ağaçları ve YSA yöntemleriyle risk ve getiri tahmininde etkili olan göstergeleri belirlemişlerdir. Patel vd. (2015), Hindistan borsasında 2003-2012 yılları arasında işlem gören iki farklı hisse senedi ve

iki farklı borsa fiyat endeksinde ait veriler ile birlikte 10 adet teknik parametre kullanarak hisse senedi fiyat hareketini yukarı ve aşağı yönlü olacak şekilde YSA, SVM, Random forest ve Naive-Bayes yöntemleriyle tahmin etmeye çalışmışlardır. Analizler sonucunda Naive-Bayes yönteminin diğer üç yöntemle göre daha başarılı olduğunu açıklamışlardır.

### 3. YÖNTEM VE ANALİZ

Bu çalışmada Türkiye’de 2009-2014 yılları arasındaki BIST 100 endeksinde işlem gören kimya, kauçuk ve plastik ürünleri imalatı sanayinde yer alan 18 şirketin hisse senedi getirilerinde etkili olan faktörlerin belirlenerek, şirketlerin pozitif ve negatif hisse senedi getirilerinin tahmin edilmesi amaçlanmıştır. Endeks bazında bakıldığında kimya, kauçuk ve plastik ürünleri imalatı sanayinde yer alan şirketlerin bazıları bahsi geçen dönem içerisinde BIST-100 endeksi içinde yer alırken bazı tarihlerde BIST-100’e giremediği için çalışmada 18 şirketin verileri dikkate alınmıştır. Şirketlere ait finansal oranlar ve hisse senedi fiyatları, FINNET tarafından geliştirilen Analiz Expert programı veri tabanından elde edilmiştir. Araştırmanın yöntemi olarak lojistik regresyon analizi, C5.0 algoritması, CART algoritması ve DVM yöntemleri seçilmiştir.

#### 3.1. Lojistik Regresyon Analizi

Verilerin kategorik (nominal/ordinal) olması durumunda parametrik olmayan istatistiksel yöntemler kullanılmaktadır. Lojistik regresyon analizinde amaç, kategorik bağımlı değişkenin değerini tahmin etmek olduğundan, iki ya da daha fazla gruba ilişkin “üyelik” tahmini yapılmaktadır. Buna göre analizin amaçlarından birinin sınıflandırmak, diğerinin ise bağımlı ve bağımsız değişkenler arasındaki ilişkileri araştırmak olduğu söylenebilmektedir (Mertler ve Vannatta, 2005; Çokluk vd. 2012). Lojistik regresyon, istatistikte kullanılan bir model oluşturma tekniği olup, bağımlı değişkenin kategorik olarak; ikili (binary, dichotomous), üçlü ve çoklu kategorili olması durumunda bağımsız değişkenlerle sebep-sonuç ilişkisini belirlemede yararlanılan bir yöntemdir (Özdamar, 2004). Lojistik regresyon analizinde bağımsız değişkenlerin normal dağılıma uygun olması ve bütün gruplar için kovaryans matrislerinin eşit olması varsayımı aranmamaktadır (Kalaycı, 2008). Dolayısıyla, Lojistik regresyonda bağımlı değişkenin değerinin tahmin edilmesi yerine bağımlı değişkenin 1 değerini alma olasılığı tahmin edilmeye çalışılmaktadır. Elde edilen sonuç bir olasılık değeri olduğu için sadece 0 ile 1 arasında bir değer alabilmektedir (Alpar, 2013).

$X$  : bağımsız değişkene ilişkin veri matrisi için  $X=$   
 $x$  olduğunda  $Y=1$  olma olasılığı  $\pi$ 'dir.

Lojistik regresyon modelinin spesifik biçimi aşağıdaki gibidir (Hosmer ve Lemeshow, 2000):

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (3.1)$$

Eşitlik (3.1)'de verilen doğrusal olmayan lojistik regresyon fonksiyonuna logit dönüşüm uygulanır. Logit işlemi gerçekleştirilirken bir olayın odds'unun doğal logaritması alınır. Odds bir olayın gerçekleşme olasılığının gerçekleşmeme olasılığına olan oranıdır. Başka bir ifadeyle odds  $Y=1$  için  $P(Y=1)/(1 - P(Y=1))$ 'dir (Hosmer ve Lemeshow, 2000; Menard, 2002).

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x \quad (3.2)$$

Logit dönüşüm olan  $g(x)$  fonksiyonu, bir lineer regresyon modelinin arzu edilen özelliklerinin çoğunu kapsar. Logit ya da  $g(x)$ ,  $x$  bağımsız değişkenine bağlı olarak  $-\infty$  ile  $+\infty$  değer alabilen lineer bir denklemdir. Lojistik regresyon modeli, bağımsız değişkenin odds'u cinsinden aşağıdaki yazılabilir.

$$\frac{\pi(x)}{1 - \pi(x)} = e^{\beta_0 + \beta_1 x} \quad (3.3)$$

Odds'un doğal logaritması alındığında model doğrusal hale dönüşmüş olur.

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \ln e^{\beta_0 + \beta_1 x} = \beta_0 + \beta_1 x \quad (3.4)$$

### 3.1.1. Lojistik Regresyon Katsayılarının Tahmini

Lojistik regresyon analizinde katsayıların tahmininde maksimum olabilirlik yöntemi kullanılmaktadır. Genel anlamda maksimum olabilirlik yöntemi gözlenen veri setini elde etme olasılığını maksimum yapan bilinmeyen parametreler için değerler vermektedir. Bu yöntemi uygulamak için başta olabilirlik fonksiyonu (likelihood function) olarak adlandırılan bir fonksiyon yazılmalıdır. Olabilirlik fonksiyonu, bilinmeyen parametrelerin bir fonksiyonu olarak gözlenen verinin olasılığını açıklamaktadır. Maksimum olabilirlik yöntemi, bu fonksiyonu maksimize eden değerlerin seçimidir. Başka bir ifadeyle bir olayın olma olasılığını maksimum yapılmaya çalışılır (Hosmer ve Lemeshow, 2000; Alpar, 2013). Olabilirlik fonksiyonu (likelihood function) verilen terimlerle birlikte aşağı-

daki gibi yazılabilmektedir (Hosmer ve Lemeshow, 2000):

$$l(\beta) = \prod_{i=1}^{\pi} \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1 - y_i} \quad (3.5)$$

Maksimum olabilirlik ilkesi, (3.5) denklemindeki ifadeyi maksimum yapan değerini tahmin etmek için kullanılmaktadır. (3.5) denkleminin logaritması alınarak tekrar düzenlenirse eşitlik (3.6)'daki gibi yazılabilir. Aşağıdaki eşitlik log likelihood olarak tanımlanmaktadır. (Hosmer ve Lemeshow, 2000). Log likelihood (log olabilirlik) istatistiğinin büyük değerleri daha zayıf istatistiksel modelleri göstermektedir. Log likelihood değeri arttıkça daha çok açıklanamayan gözlemler vardır (Field, 2005). Log likelihood (log olabilirlik) -2 ile çarpılarak -2LL olarak kısaltılmaktadır. Log likelihood değeri negatif olmasına karşın, -2LL pozitifdir (Menard, 2002).

$$l(\beta) = \ln[l(\beta)] = \sum_{i=1}^{\pi} \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (3.6)$$

### 3.1.2. Lojistik Regresyon Katsayılarının Anlamlılığının Testi

Lojistik regresyon katsayıları belirlendikten sonra modeldeki değişkenlerin istatistiksel açıdan anlamlılığı test edilmektedir. Modeldeki bağımsız değişkenlerin bağımlı değişken ile anlamlı olup olmadığını belirlemek için istatistiksel bir hipotez testi ve formülü kurulmaktadır. Lojistik regresyondaki temel ilke modelde yer alacak değişkenin olup olmaması durumuna göre modelden elde edilen tahmini değerleri bağımsız değişkenin gözlenen değerleri ile karşılaştırmaktır. Modeldeki değişkenlerin anlamlılığı Olabilirlik Oranı, Wald ya da Skor testlerinden biri ile test edilebilir (Hosmer ve Lemeshow, 2000; Alpar, 2013).

Olabilirlik fonksiyonunu kullanarak tahmin edilen ve gözlenen modeli karşılaştırmada aşağıdaki formül kullanılmaktadır (Hosmer ve Lemeshow, 2000):

$$D = -2 \ln \left[ \frac{\text{Tahminlenen modelin olabilirliği}}{\text{Tam modelin olabilirliği}} \right] \quad (3.7)$$

Yukarıdaki parantez içerisinde yer alan miktar olabilirlik oranı (likelihood ratio) olarak adlandırılmaktadır. Parantez içindeki ifade yerine (3.6) denklemini yazılıp tekrar düzenlenirse aşağıdaki eşitlik elde edilebilmektedir.

$$D = -2 \sum_{i=1}^{\pi} [y_i \ln \left( \frac{\hat{\pi}}{y_i} \right) + (1 - y_i) \ln \left( \frac{1 - \hat{\pi}}{1 - y_i} \right)] \quad (3.8)$$

Yukarıdaki denklemde  $D$  istatistiği (eşitlik (3.8)) sapma ölçütü olarak adlandırılmaktadır. Lojistik model için sapma ölçütü lineer regresyon analizindeki hata kareler toplamına (SSE) benzetilmektedir. Modelde yer alan bağımsız değişkenlerin anlamlılığı bağımsız değişkenin dahil olduğu modelin saptması ile bağımsız değişkenin yer almadığı modelin saptması arasındaki fark alınarak  $G$  istatistiği (eşitlik (3.9)) ile hesaplanmaktadır.  $\beta_1$  katsayısının sıfıra eşit olduğu hipotezi altında,  $G$  istatistiği bir serbestlik dereceli ki-kare dağılımı göstermektedir. Bu test olabirlik oran (likelihood ratio) testi ya da sapma (deviance) testi olarak da adlandırılmaktadır. (Hosmer ve Lemeshow, 2000; Field, 2005). Ki-kare dağılımı gösteren  $G$  istatistiği istatistiksel olarak anlamlı çıkarsa ( $p0,05$ ),

$H_0 = \beta_0 = \beta_1 = \beta_2 = \dots \beta_k = 0$  hipotezi reddedilerek bağımsız değişkenler ile ilgili bilginin  $P(Y = h)$  olasılığı için daha iyi tahminler yapmasına imkan sağlamaktadır.  $G$  istatistiği, lineer regresyon analizindeki regresyon kareler toplamının yanı sıra (SSR) çok değişkenli  $F$  testine benzemektedir (Menard, 2002).

$$G = LR = D(\text{değişkensiz model}) - D(\text{değişkenli model})$$

$$G = LR - 2 \ln \left[ \frac{\text{değişkensiz olabirlik}}{\text{değişkenli olabirlik}} \right] \quad (3.9)$$

Modeldeki bağımsız değişkenlerin anlamlılığını test eden tekniklerden birisi olan Wald istatistiği en çok olabirlik yönteminden yararlanmaktadır (Hosmer ve Lemeshow, 2000). Wald istatistiği  $\beta$ 'nin anlamlılığına ilişkin bir ölçüdür ve her bir değişkenin modele katkısını ifade etmektedir. (Çokluk vd. 2012). Lineer regresyondaki  $t$  testinde olduğu gibi, Wald istatistiği bağımsız değişkene ilişkin  $\beta$  katsayısının sıfırdan farklı anlamlılık gösterip göstermediğini tespit edilmesinde kullanılmaktadır. Wald istatistiği (eşitlik (3.10)),  $\beta$  katsayısının kendi standart hatasına oranlanmasıyla hesaplanmaktadır.  $\beta$  (Field, 2005). Wald test istatistiği standart normal dağılıma uymaktadır. 'nin en çok olabirlik tahminin karesi standart hatanın karesine bölüldüğünde (eşitlik (3.11)) Wald istatistiği bir serbestlik dereceli ki-kare dağılımına uymaktadır (Alpar, 2013).

$$W = \frac{\hat{\beta}}{S(\hat{\beta}_j)} \quad (3.10)$$

$$W = \frac{\hat{\beta}_j^2}{S(\hat{\beta}_j)^2} \sim \chi^2 \quad (3.11)$$

### 3.2. C5.0 Algoritması

Karar ağacı bir dizi basit karar kuralları uygulayarak başarılı bir şekilde toplu haldeki büyük veri kayıtlarını gruplamak için kullanılabilir olan bir yöntemdir (Berry ve Linoff, 2004). Karar ağaçları sınıflandırma amaçlı veri madenciliği tekniklerindedir. Heterojen yapıdaki niteliklere (attributes) ait bir veri seti verildiğinde, karar ağacı veri setini daha küçük ve daha homojen alt kümelere ayırmak için bu niteliklerin değerlerini kullanmaktadır. Karar ağaçları uygulaması öncelikle ağaçların modellenmesi veya dallanması için hangi tekniklerin nasıl uygulanacağı konusunda farklılıklar göstermektedir. Bu modelleme teknikleri yaygın olarak karar ağacı tabanlı algoritmalar veya karar ağacı algoritmaları olarak adlandırılmaktadır (Shahnaz, 2006). Oluşan karar ağaçları model için tersine çevrilmiş ağaç şekline benzetilmektedir. Karar ağacı karar verme kuralları için düğümler ve bu düğümleri bir araya getiren dallardan oluşmaktadır. Ağacın tepesinde kök, altında ise dal ve yapraklardan meydana gelmektedir (Seyrek ve Ata, 2010). En yaygın bilinen karar ağacı algoritmaları ID3, C4.5, C5.0 ve CART olarak bilinen algoritmalar (Shahnaz, 2006). Bu çalışmada karar ağacının elde edilmesinde C5.0 sınıflandırma algoritması ele alınmıştır. C5.0 algoritması bellek kullanımı ve çalışma zamanı açısından C4.5 algoritmasından daha iyi olup, aynı zamanda karar ağaçlarını en aza indirmek ve karar kurallarını üretme açısından daha iyi sonuçlar üretmektedir (Shahnaz, 2006).

C4.5 ve C5.0 algoritması optimal olarak düğümleri ayırmak için bilgi kazancı ve entropi azaltma kavramını kullanmaktadır.  $X$  değişkeni (attribute) için  $k$  adet olasılıklar sırasıyla  $p_1, p_2, p_3, \dots, p_k$  olarak adlandırılmaktadır.  $X$  değişkeni ya da niteliği için entropi aşağıdaki eşitlikte verilmiştir (Quinlan, 1993a; Quinlan, 1996b; Larose ve Larose, 2014).

$$\text{Entropi} = H(X) = - \sum_{j=1}^k p_j \log_2(p_j) \quad (3.12)$$

Eğitim seti içerisinde yer alan  $X$  niteliğine bağlı olarak hedef niteliği  $T$  alt kümelerinin  $T_1, T_2, \dots, T_k$  olarak alt bölmelere ayrıldığı varsayılmaktadır. Her bir  $T$ 'nin sınıfını belirlemek için gerekli olan bilgilerin ağırlıklı ortalaması entropilerin ağırlıklandırılmış toplamları olarak hesaplanabilir. Bilgilerin ağırlıklı ortalaması aşağıdaki eşitlikte verilmiştir.

$$H_S(T) = \sum_{i=1}^k p_i H_S(T_i) \quad (3.13)$$

Ardından ayırma işlemini gerçekleştirmek için bilgi kazancı (information gain) hesaplanmaktadır. Böylelikle C4.5 ve C5.0 algoritması her bir karar verme düğümünde en büyük bilgi kazancına sahip olan ayırma ölçütünü belirleyerek optimal ayırma işlemini gerçekleştirmektedir. Bilgi kazancı aşağıdaki eşitlikte verilmiştir (Quinlan, 1993a; Quinlan, 1996b; Larose ve Larose, 2014):

$$\text{Information Gain } (S) = H(T) - H_S(T) \quad (3.14)$$

### 3.3. CART Algoritması

Sınıflandırma ve regresyon ağaçları (CART) metodu Breiman vd. (1984) tarafından önerilmiştir (Larose ve Larose, 2014; Han ve Kamber, 2006). CART algoritması ikili bölünme esasına dayanarak yeni bölünmeler olduğu sürece devam etmektedir. CART algoritması aday bölünmeler gibi ağaçlara benzetilmektedir. (Berry ve Linoff, 2004). Bu algoritma çeşidi non-parametrik olup sürekli ya da kesikli olarak tanımlanan verilerden ikili karar ağacı oluşturmaktadır. Her bir değişken için, CART algoritması iki ayrı ve birbirini tamamlayan kümeler içerisindeki tüm olası aday bölünmelerle ilgilenmektedir (Grabczewski, 2014). CART algoritması tüm mevcut değişkenlerin ve tüm olası bölünmelerin ayrıntılı bir şekilde yer aldığı karar düğümlerini meydana getirecek ağacı oluşturmaktadır. CART algoritması aşağıdaki kriterlerin değerlerine göre optimum bölünme işlemi gerçekleştirmektedir (Larose ve Larose, 2014):

$t$ . düğümdeki  $s$ . aday bölünmenin uygunluk ölçümü olan  $\Phi(s|t)$  hesaplanır.  $t_L$ :  $t$  düğümünün sol taraftaki bölünmesi,  $t_R$ :  $t$  düğümünün sağ taraftaki bölünmesi olmak üzere,

$$\Phi(s|t) = 2P_L P_R \sum_{(j=1)}^{\pi} |P(j|t_L) - P(j|t_R)| \quad (3.15)$$

$$P_L = \frac{t_L \text{deki kayıtların sayısı}}{\text{Eğitim setindeki kayıtların sayısı}} \quad (3.16)$$

$$P_R = \frac{t_R \text{deki kayıtların sayısı}}{\text{Eğitim setindeki kayıtların sayısı}} \quad (3.17)$$

$$P(j|t_L) = \frac{t_L \text{deki } j \text{ sınıflarının sayısı}}{t \text{deki kayıtların sayısı}} \quad (3.18)$$

$$P(j|t_R) = \frac{t_R \text{deki } j \text{ sınıfların sayısı}}{t \text{deki kayıtların sayısı}} \quad (3.19)$$

Optimal bölünme  $t$  düğümündeki olası tüm aday bölünmeler için  $\Phi(s|t)$  uygunluk ölçüsü maksimize edilerek gerçekleştirilmektedir.

### 3.4. Destek Vektör Makineleri Analizi

Destek Vektör Makineleri (DVM), Vapnik ve Chervonenkis (1974) tarafından optimal ayırma bir hiper düzlemi kullanarak geliştirdikleri makine öğrenmesinin yeni bir türü olarak ortaya çıkmıştır (Vapnik, 1995). DVM sınıf sınırları arasındaki mesafeyi maksimize ederek yanlış sınıflandırma olasılığını azaltmaya çalışan, yeni bir yaklaşım modeli olan maksimum marjlı sınıflandırıcı olarak ifade edilmektedir. Maksimum marjlı sınıflandırıcılar kuadratik programlama teknikleri üzerinden çözülebilecek olan bir dışbükey optimizasyon problemidir (Hamel, 2009). Genel olarak sınıflandırıcıları geliştirmek için iki yaklaşım vardır: birincisi veri dağılımlarının önsel bilgi varsayıldığı parametrik yaklaşım, ikincisi ise veri dağılımlarının önsel bilgi varsayılmadığı parametrik olmayan yaklaşımdır. Çekirdek tabanlı yöntemler olarak bilinen destek vektör makineleri sınıflandırma ve fonksiyon yaklaşım problemleri için uygulanmaktadır (Abe, 2005). DVM, sigmoid kernel fonksiyonu kullanan iki katmanlı ileri beslemeli bir yapay sinir ağı dizaynına sahip bir yaklaşımdır (Haykin, 1999). Bir DVM eğitiminde, eğitim verisi için sınıflandırma performansı maksimize edilmeye çalışılmaktadır. Verileri birbirinden ayıran düzlemlere optimal ayırma hiper düzlemi denilmektedir (Hamel, 2009). Hiper düzlemler arasındaki mesafe marjindir ve optimal karar düzlemi marjinin merkezinde yer almaktadır. Destek vektörler ait olduğu sınıfın hiper düzleminde bulunur. Optimal karar düzlemi destek hiper düzlemlerine eşit uzaklıktadır. Marjinin büyüklüğü her bir sınıftaki destek vektörlerle çevrelenerek kısıtlanmıştır. DVM için maksimum marjlı sınıflandırma yaklaşımındaki amaç, marjini maksimize eden karar düzleminin pozisyonunu bulmaktır (Hamel, 2009).

DVM eğitimi için,  $(x_1, y_1), \dots, (x_n, y_n), x \in R^n, y \in \{+1, -1\}$  olmak üzere,  $(x_i, y_i)$  çiftlerinden oluşan bir eğitim kümesi verildiğinde farklı sınıflar bir hiper düzlem tarafından ayrılabilirler (Vapnik, 1995).

$$(x \cdot w) - b = 0 \quad (3.20)$$

Burada bahsedilen  $w$  hiper düzlemin ağırlık vektörünü ve  $b$  sabit terimini göstermektedir. İkili sınıflandırıcı problemlerinde veri kümelerini birbirinden ayırmak için iki destekli hiper düzleme ihtiyaç

vardır: ilki sınıfı +1 ile belirtilen sınıf yönündeki hiper düzlem ve ikincisi -1 ile belirtilen sınıf yönündeki hiper düzlemdir. Bu iki hiper düzlemin ortasında verileri birbirinden ayıran optimal karar düzlemi yer alır. İkili sınıflandırıcı problemlerinde bu iki düzlem arasındaki mesafe marjin olarak adlandırılmaktadır (Hamel, 2009). Optimal karar düzlemi bir taraftan  $y_i = +1(w \cdot x + b) > 0$ , diğer taraftan  $y_i = -1(w \cdot x + b) < 0$  olarak etkilenmiş noktalar ile iki sınıflı verileri birbirinden ayırmaktadır. (Campbell ve Ying, 2011). Bu eşitsizler bir arada ifade edilebilmektedir.

$$y_i(w \cdot x + b) - 1 \geq 0, \forall i \quad (3.21)$$

Eşitlik (3.21)'den hareketle  $H_1 : x_i \cdot w + b = 1$  olan  $w$  ağırlığı ve orjinden dik uzaklığı  $|1 - b| / \|w\|$  olan bir hiper düzlemi ifade etmektedir. Benzer şekilde  $H_2 : x_i \cdot w + b = -1$  olan  $w$  ağırlığı ve orjinden dik uzaklığı  $|-1 - b| / \|w\|$  olan diğer bir hiper düzlemdir. Böylece bir verinin optimal karar düzlemine uzaklığı  $d_+ = d_- = |x_i \cdot w + b| / \|w\| = 1 / \|w\|$  ve marjin sadece  $2 / \|w\|$ 'dir.  $H_1$  ve  $H_2$  düzlemleri birbirine paralel olup iki arasında eğitim verileri yer almamaktadır. Eşitlik (3.21)'de ki kısıtlayıcılar ile birlikte  $\|w\|^2$  minimize edilerek, maksimum marjı veren hiper düzlem çifti bulunabilmektedir (Vapnik, 1995; Burges, 1998; Scholkopf ve Smola, 2002; Abe, 2005; Campbell ve Ying, 2011; Alpaydın, 2011):

$$\min \frac{1}{2} \|w\|^2 \quad \text{kısıt} : y_i(w \cdot x_i + b) \geq +1, \forall i \quad (3.22)$$

Veriler doğrusal olarak ayrılmadığı durumlarda yukarıda bahsedilen algoritmalar uygun çözüm vermeyecektir. Bu durumda pozitif aylak değişken  $\xi_i, i = 1, \dots, l$  eklenerek çözüm gerçekleştirilebilir.  $0 < \xi_i < 1$  için  $x_i$  verileri doğru olarak sınıflandırılıyor-dur buna karşın

$\xi_i \geq 1$  için  $x_i$  verileri optimal karar düzleminin diğer tarafında yer aldığından yanlış sınıflandırılmaktadır.  $\sum_{i=1}^n \xi_i$  yanlış sınıflandırılan örnek sayısını göstermektedir. Yanlış sınıflandırma miktarını azaltmak için bir ceza terimi olan  $C$  ile amaç fonksiyonu ve kısıtlayıcılar tekrar düzenlenirse elde edilen kısıtlayıcı optimizasyon problemi aşağıdaki gibi olacaktır (Burges, 1999; Alpaydın, 2011; Campbell ve Ying, 2011).

$$\min \left[ \frac{1}{2} w \cdot w + C \sum_{i=1}^m \xi_i^2 \right] \quad (3.23)$$

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (3.24)$$

Veriler doğrusal olarak ayrılmazlarsa elde edilen sınıflandırıcılar yüksek genelleme yeteneğine sahip olmayabilirler. Bu durumda doğrusal olarak genelleştirme yeteneğini artırmak için orijinal girdi uzayı özellik uzayı olarak adlandırılan daha yüksek boyutlu bir uzaya dönüştürülmektedir (Vapnik, 1995; Abe, 2005). Genişletme ve yüksek boyuta dönüştürme  $\phi$  ile gerçekleşmektedir. Kernel fonksiyonu denklemi eşitlik (3.25)'de verilmiştir (Burges, 1999).

$$K(x_i, x_j) = \sum_i^m \phi(x_i) \cdot \phi(x_j) \quad (3.25)$$

DVM bazı kernel fonksiyonları kullanarak çeşitli öğrenme makineleri oluşturabilirler. Kullanılan kernel fonksiyonların bazıları aşağıdaki eşitlikte verilmiştir (Campbell ve Ying, 2011).

$$K(x_i, x_j) = (x_i, x_j)^d, \text{ Doğrusal Kernel Fonksiyon} \quad (3.26)$$

$$K(x, x) = (x, x + c)^d, \text{ Polinom Kernel Fonksiyon} \quad (3.27)$$

$$K(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)}{2\sigma^2}\right), \text{ Gauss Kernel Fonksiyon} \quad (3.28)$$

$$K(x_i, x_j) = \tanh(k(x_i, x_j) + c), \text{ Sigmoid Kernel Fonksiyon} \quad (3.29)$$

#### 4. VERİ SETİ VE DEĞİŞKENLER

Araştırmada Türkiye'de 2009-2014 yılları arasında BIST 100 endeksinde işlem gören kimya, kauçuk ve plastik ürünleri imalatı sanayinde sürekli yer alan 18 şirketin verileri kullanılmıştır. Ele alınan şirketlerin yıllık bilanço ve gelir tabloları incelenerek toplam 108 gözlem için %80'i ile eğitim seti ve %20'si ile test seti oluşturulmuştur. Şirketlerin pozitif ve negatif hisse senedi getiri sınıflandırma tahminine yönelik literatürde yer alan makaleler incelenmiş, sınırlı mevcut veriler dikkate alınarak değişkenler oluşturulmuştur (Barak ve Modarres, 2015; Zhang, vd. 2014; Ege ve Bayrakdaroğlu, 2009; Tsai vd. 2011; Tsai ve Hsiao, 2010; Lam, 2004). Analiz aşamasında 20 finansal oranın yanı sıra makroekonomik değişkenlerden olan Tüketici Fiyat Endeksi "TÜFE" ile birlikte toplam 21 bağımsız değişken kullanılmıştır. Analiz aşamasında bağımsız değişkenler likidite oranları, faaliyet oranları, karlılık oranları, finansal yapı oranları ve borsa performans oranları olmak üzere 5 grupta toplanmıştır. Analizde kullanılan 20 finansal oran, TÜFE ve analiz kapsamına alınan şirketler aşağıdaki Tablo 1'de gösterilmiştir.

**Tablo 1:** Analizde Kullanılan Finansal Oranlar ve Araştırma Kapsamındaki Şirketler

Finansal Oranlar		Kısaltma	Analizde Kullanılan Şirketler İsimleri
Likidite Oranları		AKSA	Aksa
x1	cari oran	ALKIM	Alkim Kimya
x2	likidite oranı	AYGAZ	Aygaz
x3	nakit oranı	BAGFS	Bagfaş
<b>Faaliyet Oranları</b>		<b>BRISA</b>	<b>Brisa</b>
x4	duran varlık/toplam aktif	DEVA	Deva Holding
x5	aktif devir hızı	DYOB	Dyo Boya
x6	alacak devir hızı	EGGUB	Ege Gübre
x7	dönen varlık devir hızı	GOODY	Good-Year
x8	stok devir hızı	GUBRF	Gübre Fabrik.
x9	özsermaye devir hızı	HEKTS	Hektaş
<b>Karlılık Oranları</b>		<b>MRSHL</b>	<b>Marshall</b>
x10	aktif karlılık	PETKM	Petkim
x11	brüt kar marjı	PIMAS	Pimaş
x12	esas faaliyet kar marjı	SASA	Sasa Polyester
x13	net kar marjı	SODA	Soda Sanayii
x14	özsermaye karlılık	TRCAS	Turcas Petrol
<b>Finansal Yapı Oranları</b>		<b>TUPRS</b>	<b>Tüpraş</b>
x15	borç/özsermaye		
x16	kısa vad.borç/toplam borç		
x17	özsermaye /aktif		
x18	kaldıraç oranı		
<b>Borsa Performans Oranları</b>			
x19	fiyat/kazanç		
x20	piyasa/defter değeri		
x21	TÜFE		

Hisse senedi fiyat eğilimi bağımlı değişken olduğu için literatürde üç farklı yaklaşımla kategorileştirilmiştir (Zhang, vd., 2014:52):

1) Hisse senedi kapanış fiyatının bir önceki fiyata eşit ya da büyük olup olmaması durumuna göre 0 ya da 1 (Tsai ve Hsiao, 2010),

2) Hisse senedi fiyatı için bir eşik değer belirlenmekte olup, hisse senedi fiyatı eşik değer üzerinde ise yukarı eğilimli "1", altında ise düşüş eğilimli "-1" ya da değişim sabit bir durumda ise "0", (Ren, Zangham ve Rahimi, 2006; Chang vd. 2011),

3) Hisse senedi fiyatı karşılaştırma ölçütü olarak piyasanın fiyatını (market benchmark) aşım aşmama durumuna göre 1 ya da 0 ikili değerlerini almaktadır (Thawornwong ve Enke, 2004).

Zhang vd. (2014), yaptıkları çalışmada üçüncü yaklaşımdan hareketle hisse senedi fiyat eğilimini belirlemek için Şangay Kompozit Endeksi ile karşılaştırmışlardır. Zhang vd. (2014), bir hisse senedinin fiyatını Şangay Kompozit Endeksi ile

karşılaştırdıklarında yüksek ise aşırı getiri aksi takdirde negatif getiri olarak belirlemişlerdir. Bu çalışmanın ışığında analizimizde bir şirketin hisse senedi getirisi sektörün hisse senedi getirisine göre kıyaslandığında yüksek ise o şirketin hisse senedi getirisi pozitif "1"; buna karşın bir şirketin hisse senedi getirisi sektöre göre kıyaslandığında düşük ise o şirketin hisse senedi getirisi negatif "0" olarak kodlanarak bağımlı değişken oluşturulmuştur. Hisse senedi getirisi yıllık olacak şekilde hesaplanarak belirlenmiştir. Ardından sektörün hisse senedi getirisi ile karşılaştırılarak kategorileştirilmiş ve bağımlı değişken oluşturulmuştur. Hisse senedi getirilerinin hesaplanmasında, Zivot (2007)'de yer verilen yöntem benimsenmiş,  $R_t = (P_t - P_{t-1})/P_{t-1}$  formülü tercih edilmiştir. Burada,  $R_t$  basit net getiriyi,  $P_t$  varlığın 12. aylardaki kar payı ödemesiz kapanış fiyatlarını,  $P_{t-1}$  ise bir önceki yılın son işlem gününün kapanış fiyatını açıklamaktadır (Öz vd. 2011). Aynı döneme ait bağımlı ve bağımsız değişkenler belirlendikten sonra 2009-2014 yılları arasında toplam 108 gözlem analizler için hazır hale getirilmiştir. Şirketlerin 53'ü negatif hisse getirisi ve 55'i ise pozitif hisse senedi getirisine sahip şirketlerdir. Bağımlı değişken kategorik değişken olduğu için açıklayıcı değişkenlerde gecikme kullanılmamıştır.

## 5. ARAŞTIRMANIN MODELİ

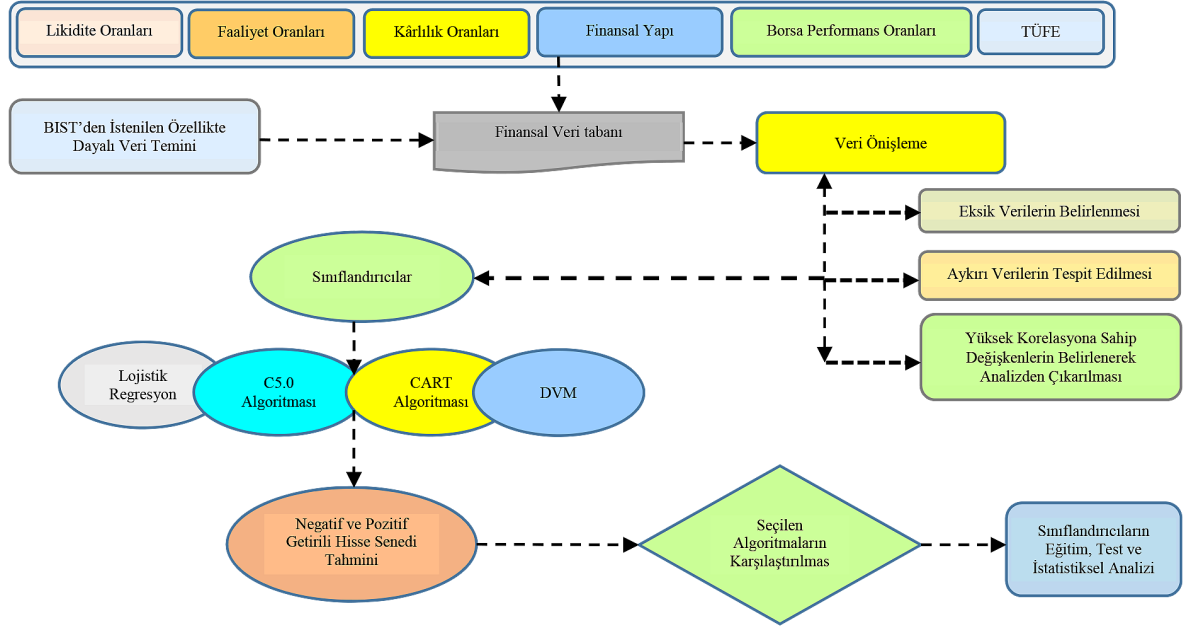
Araştırmanın bu aşamasında Şekil 1'de görüldüğü gibi analizlerde kullanılmak üzere gerekli düzenlemelerin yapıldığı modelleme kısmından bahsedilmiştir. Şekil 1'de analizlere ait modelin kavramsal yapısı verilmiştir.

Şekil 1'de görüldüğü üzere modelin kavramsal yapısında veri ön işleme gerçekleştirilirken 24 firmanın 6'sı belirlenen dönem içerisinde yer almadığı için analiz kapsamına dahil edilmemiştir. Bağımsız değişkenler arasında çoklu doğrusal bağlantı varlığını ortaya çıkarmak için Pearson testine dayalı korelasyon analizi gerçekleştirilmiştir. Pearson testine dayalı %80'in üzerinde korelasyona sahip olan değişkenler analizden çıkarılmıştır (Alpar, 2013). Bu değişkenler  $X_1$  "cari oran",  $X_2$  "likidite oranı",  $X_{15}$  "borç/özsermaye" ve  $X_{17}$  "özsermaye/aktif" değişkenleridir.

Veri ön işleminin ardından 18 firmanın toplam 108 adet veri kaydı SPSS 18 ve Clementine 12.0 yazılımına girilmiştir.

Lojistik regresyon analizi ile pozitif ve negatif hisse senedi getiri sınıflandırma tahminine etki eden faktörler belirlenmeye çalışılmış, hisse senedi getirileri tahmin edilmiştir.





Şekil 1: Analizler için Kullanılan Modelin Kavramsal Yapısı

Denetimli bir sınıflandırma algoritması olarak bilinen C5.0 algoritması ile CART algoritması tarafından, hedef (target) değişkenin yer aldığı veriler ile model eğitilmiş ve karar ağaçları ile kurallar elde edilmiştir. DVM analizi uygulamada kural üretmeyen buna karşın sınıflandırma tahmininde kullanılan bir analiz olarak seçilmiştir. Lojistik regresyon analizi, C5.0, CART algoritması ve DVM için verilerin % 80 ile model eğitilmiş geri kalan % 20'si ile test işlemi gerçekleştirilerek analiz işlemleri gerçekleştirilmiştir. Bir sonraki aşamada sınıflandırıcıların eğitim, test ve istatistiğe ilişkin belirleyicilik (specificity) ve duyarlılık (sensitivity) performansları karşılaştırılmıştır. Böylelikle çalışmada yatırımcılar ve analizciler için hisse senedi getiri başarısına etki edecek faktörlerin belirlenmesine yönelik yöntemler incelenmiştir.

### 5.1. Lojistik Regresyon Analizi

Lojistik regresyon analizi için verilerin değerlendirilmesi SPSS 18.0 istatistiksel analiz programıyla yapılmıştır. Lojistik regresyon analizinde işletmelerin pozitif ve negatif hisse senedi getiri sınıflandırma tahminine etki eden faktörlerin belirlenmesi için Enter metodu seçilmiştir. -2LL ve  $R^2$  değerleri hesaplanmıştır. %95 anlamlılık seviyesinde -2LL değeri 114,623'tür. Cox-Snell  $R^2$  değeri incelendiğinde bağımsız değişkenler analize girildiğinde, lojistik modelin bağımlı değişkendeki varyansının %27,7'sini açıklamaktadır. Nagelkerke  $R^2$  değeri biraz daha

yüksek olup %37'dir. Modelin uyum iyiliği testi için Hosmer ve Lemeshow uyum iyiliği testi gerçekleştirilmiştir. Lojistik modelin ki-kare 8 serbestlik derecesiyle 9,14 değeri, =15,507 değerinden küçük olduğundan dolayı, modelin veri uyumunun yeterli düzeyde olduğunu tespit edilmiştir. Lojistik regresyon katsayılarının değeri olan odds oranı hesaplanmıştır. Odds oranı bağımsız değişkenlerin logit üzerindeki etkisini göstermek için doğrudan yorumlanabilir. Odds'daki yüzdelik değişim ( $100 \times [OR-1]$ ) formülü kullanılarak hesaplanabilir. Odds oranı (OR) 1'den büyük ise bağımsız değişkenin logit'te artırıcı bir etkisi olduğunu, OR değerinin 1'den küçük olması bağımsız değişkenin logit'te azaltıcı bir etkisi olduğunu ve OR değerinin 1'e eşit olması bağımsız değişkenin logit'te değişime neden olmadığı şeklinde değerlendirilebilir (O'Connell, 2006; Alpar, 2013). Tablo 2'de bağımsız değişkenlerin katsayılarına ait standart hatalar, Wald istatistikleri, anlamlılık düzeyleri (Sig) ve Exp () istatistikleri verilmektedir.

Tablo 2'nin ilk sütununda lojistik regresyon analizinin  $\beta$  katsayıları gösterilmektedir. İşletmelerin pozitif ve negatif hisse senedi getiri sınıflandırma tahmini için Wald testi kullanılarak bağımsız değişkenlerin katsayılarının anlamlılığı test edilmiştir.

Wald testi sonucuna göre  $X_{11}$  değişkeni "brüt kar marjı" ve  $X_{20}$  değişkeni "piyasa/defter değeri" 0,05 önem seviyesinde;  $X_{21}$  değişkeni "TÜFE" 0,01 önem seviyesinde anlamlı olduğu tespit edilmiştir.

**Tablo 2:** Lojistik Regresyon Analizi Sonucu Belirlenen Değişkenler

Lojistik Model	B Katsayısı	Std.hata	Wald	Sd.	Anl.	Exp(B) Odds Oranı
X <sub>3</sub>	,128	,769	,028	1	,868	1,137
X <sub>4</sub>	-3,172	4,673	,461	1	,497	,042
X <sub>5</sub>	-2,100	1,920	1,196	1	,274	,122
X <sub>6</sub>	,017	,053	,098	1	,755	1,017
X <sub>7</sub>	-,278	,710	,154	1	,695	,757
X <sub>8</sub>	,076	,089	,722	1	,396	1,079
X <sub>9</sub>	,002	,003	,250	1	,617	1,002
X <sub>10</sub>	,034	,103	,106	1	,745	1,034
X <sub>11</sub>	-11,371	4,786	5,645	1	,018	,000
X <sub>12</sub>	2,519	6,004	0,176	1	0,675	12,414
X <sub>13</sub>	-,072	,741	,009	1	,923	,931
X <sub>14</sub>	1,014	2,647	,147	1	,702	2,757
X <sub>16</sub>	,287	2,299	,016	1	,901	1,332
X <sub>18</sub>	-2,435	3,702	,433	1	,511	,088
X <sub>19</sub>	,000	,002	,054	1	,816	1,000
X <sub>20</sub>	1,316	0,531	6,147	1	0,013	3,729
X <sub>21</sub>	-0,034	0,012	7,917	1	0,005	0,967
Sabit	11,506	5,336	4,649	1	,031	99351,494

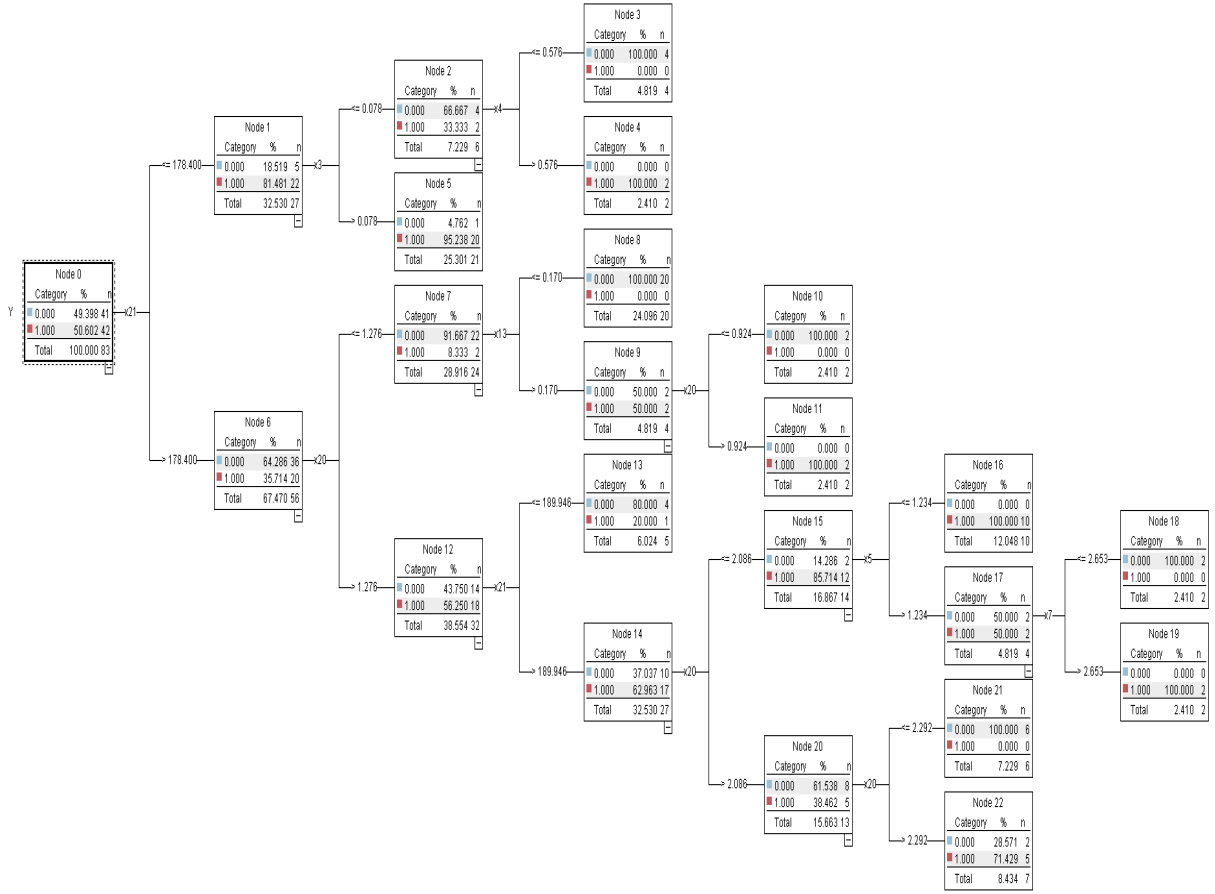
İstatistiksel açıdan anlamlı değişkenler incelendiğinde hisse senedi getiri başarılarına X<sub>20</sub> değişkenin “brüt kar marjı” pozitif etki yaptığı; X<sub>11</sub> değişkenin “brüt kar marjı” ve X<sub>21</sub> değişkenin “TÜFE” negatif etki yaptığı tespit edilmiştir. İstatistiksel açıdan anlamlı değişkenler dikkate alındığında hisse senedi getiri başarılarının belirlenmesinde X<sub>20</sub> değişkenin “piyasa/defter değeri” daha fazla etkiye sahip olduğu belirlenmiştir. Diğer bir ifade ile X<sub>20</sub> değişkenindeki “piyasa/defter değeri” bir birimlik artışın Exp() diğer tüm bağımsız değişkenler sabit kalmak şartıyla hisse senedi getirisi negatif olan işletmelere karşın, hisse senedi getirisi pozitif olan işletmelerin odds’unda pozitif etkiye sahip olduğu tespit edilmiştir. X<sub>20</sub> değişkenin “piyasa/defter değeri” üstel lojistik değeri olarak gerçekleşmiştir. Bu değişkendeki bir birimlik artışın hisse senedi getirisi pozitif olanların negatif olanlara göre 3,729 kat daha fazla olduğunu göstermektedir. Benzer bir ifade ile X<sub>11</sub> “brüt kar marjı” ve X<sub>21</sub> “TÜFE” değişkenlerinin üstel lojistik değerleri ve olarak bulunmuştur. Brüt kar marjı ve TÜFE değişkenlerindeki bir birimlik artışın hisse senedi getirisi pozitif

olan işletmelerin odds’u için brüt kar marjı değişkeni %100 (1-0,000=1\*100) ve TÜFE değişkeni ise %3,3 (1-0,967=0,03\*100) oranında azalışa sebep olmuştur.

## 5.2. C5.0 Algoritması ile Karar Ağacı ve Analizi

Karar ağacının oluşturulmasında veri madenciliği yöntemlerinden olan entropiye dayalı algoritmalarından C5.0 algoritması seçilmiştir. Şekil 2’de C5.0 algoritması kullanılarak oluşturulan karar ağacı gösterilmiştir. Şekil 2’de C5.0 algoritması kullanılarak oluşturulan karar ağacı aşağıda verilmiştir.

Şekil 2’de görüldüğü üzere C5.0 algoritması tarafından oluşturulan modelin eğitim setine ait 83 şirketin 41’i negatif hisse senedi getirisi ve 42’si ise pozitif hisse senedi getirine sahip firmalardır. Karar ağacının oluşturulmasında hisse senedi getirilerinde en kuvvetli etkiye sahip olan değişken X<sub>20</sub> değişkeni “piyasa/defter değeri” olarak tespit edilmiştir. C5.0 algoritması ile elde edilen karar ağacından 12 kural üretilmiştir. Elde edilen bu kurallar Tablo 3’te gösterilmiştir.



Şekil 2: C5.0 Algoritması ile Hisse Senedi Getiri Sınıflandırma Tahmine İlişkin Karar Ağacı

Tablo 3: C5.0 Algoritması tarafından Elde Edilen Kurallar

Hisse senedi getirisi pozitif ve negatif kategorisi için elde edilen kurallar				
Kural Sayısı	Eğer	% oran	İşletme Sayısı	
1	$X_{21} \leq 178.4 \& X_3 > 0.078$	95,24%	20	Pozitif Hisse Senedi Getirisi
2	$X_{21} > 189.946 \& X_{20} > 1.276 \& X_{20} \leq 2.086 \& X_5 \leq 1.234$	100,00%	10	
3	$X_{21} > 189.946 \& X_{20} > 2.292$	71,43%	5	
4	$X_{21} \leq 178.4 \& X_3 \leq 0.078 \& X_4 > 0.576$	100,00%	2	
5	$X_{21} > 178.4 \& X_{20} \leq 1.276 \& X_{13} > 0.170 \& X_{20} > 0.924$	100,00%	2	
6	$X_{21} > 189.946 \& X_{20} > 1.276 \& X_{20} \leq 2.086 \& X_5 > 1.234 \& X_7 > 2.653$	100,00%	2	
7	$X_{21} > 178.4 \& X_{20} \leq 1.276 \& X_{13} \leq 0.17$	100,00%	20	Negatif Hisse Senedi Getirisi
8	$X_{21} > 189.946 \& X_{20} > 2.086 \& X_{20} \leq 2.292$	100,00%	5	
9	$X_{21} \leq 178.4 \& X_3 \leq 0.078 \& X_4 \leq 0.576$	100,00%	4	
10	$X_{21} > 178.4 \& X_{21} \leq 189.946 \& X_{20} > 1.276$	80,00%	4	
11	$X_{21} > 178.4 \& X_{13} > 0.17 \& X_{20} \leq 0.924$	100,00%	2	
12	$X_{21} > 189.946 \& X_{20} > 1.276 \& X_{20} \leq 2.086 \& X_5 > 1.234 \& X_7 \leq 2.653$	100,00%	2	

Tablo 3'te görüldüğü üzere C5.0 algoritması ile bir şirketin hisse senedi getiri sınıflandırma tahminine ilişkin en önemli iki kural aşağıda verilmiştir.

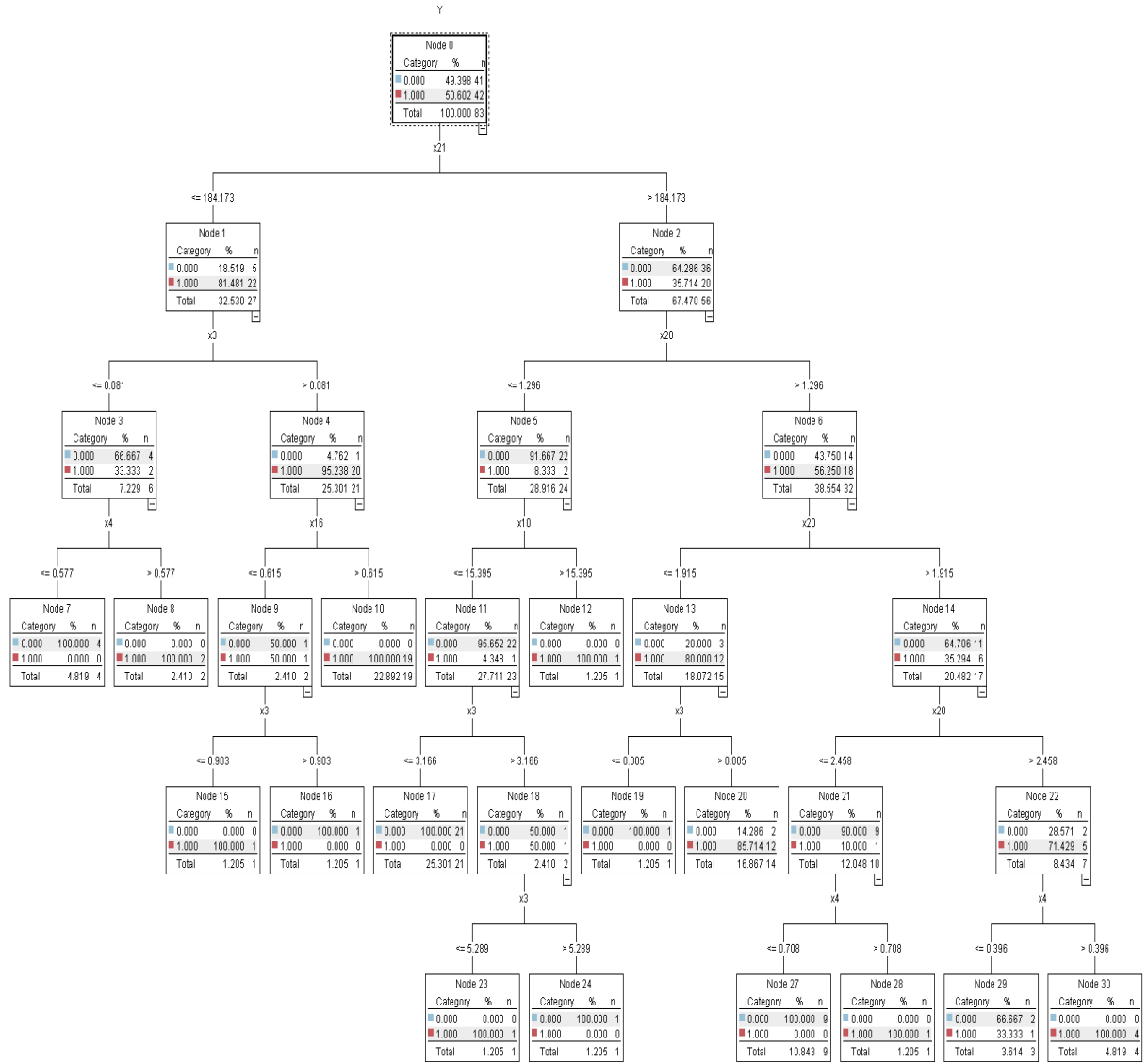
**Kural 1:** Eğer bir şirketin  $X_{21}$  değişkeni "TÜFE" 178,4'e eşit ya da küçük,  $X_3$  değişkeni "nakit oranı" 0,078'den büyük ise o şirket %95,24 (20 şirket) oran ile hisse senedi pozitif bir getiriye sahip olan bir işletmedir.

**Kural 7:** Eğer bir şirketin  $X_{21}$  değişkeni "TÜFE" 178,4'den büyük,  $X_{20}$  değişkeni "piyasa/defter değeri" 1.276'ya eşit ya da küçük ve  $X_{13}$  değişkeni "net kar marjı" 0.17 eşit ya da küçük ise o şirket %100 (20 şirket) oran ile hisse senedi negatif bir getiriye sahip olan bir işletmedir.

### 5.3. CART Algoritması ile Karar Ağacı ve Analizi

CART analizi kullanılırken karar ağacının oluşturulmasında Twoing algoritması kullanılmıştır. Şekil 3'te CART algoritması kullanılarak oluşturulan karar ağacı gösterilmiştir.

Şekil 3'te karar ağacında görüldüğü üzere C5.0 algoritmasında olduğu gibi CART algoritmasında da hisse senedi sınıflandırma tahmininde en kuvvetli etkiye sahip olan değişkenin  $X_{20}$  değişkeni "piyasa/defter değeri" olduğu tespit edilmiştir. CART algoritması ile 15 kural türetilmiştir. CART algoritması ile oluşan bu kurallar Tablo 4'te verilmiştir.



**Şekil 3:** CART Algoritması ile Hisse Senedi Getiri Sınıflandırma Tahmine İlişkin Karar Ağacı

**Tablo 4:** CART Algoritması tarafından Elde Edilen Kurallar

Hisse senedi getirisi pozitif ve negatif kategorisi için elde edilen kurallar				
Kural Sayısı	Eğer	% oran	İşletme Sayısı	
1	$X_{21} \leq 184.173 \ \& \ X_3 > 0.081 \ \& \ X_{16} > 0.615$	100,00%	19	Pozitif Hisse Senedi Getirisi
2	$X_{21} > 184.173 \ \& \ X_{20} > 1.296 \ \& \ X_{20} \leq 1.915 \ \& \ X_3 > 0.005$	85,71%	12	
3	$X_{21} > 184.173 \ \& \ X_{20} > 2.458 \ \& \ X_4 > 0.396$	100,00%	4	
4	$X_{21} \leq 184.173 \ \& \ X_3 \leq 0.081 \ \& \ X_4 > 0.577$	100,00%	2	
5	$X_{21} > 184.173 \ \& \ X_{20} \leq 1.296 \ \& \ X_{10} > 15.395$	100,00%	1	
6	$X_{21} \leq 184.173 \ \& \ X_3 > 0.081 \ \& \ X_3 \leq 0.903 \ \& \ X_{16} \leq 0.615$	100,00%	1	
7	$X_{21} > 184.173 \ \& \ X_{20} \leq 1.296 \ \& \ X_{10} \leq 15.395 \ \& \ X_3 > 3.166 \ \& \ X_3 \leq 5.289$	100,00%	1	
8	$X_{21} > 184.173 \ \& \ X_{20} > 1.915 \ \& \ X_{20} \leq 2.458 \ \& \ X_4 > 0.708$	100,00%	1	
9	$X_{21} > 184.173 \ \& \ X_{20} \leq 1.296 \ \& \ X_{10} \leq 15.395 \ \& \ X_3 \leq 3.166$	100,00%	21	Negatif Hisse Senedi Getirisi
10	$X_{21} > 184.173 \ \& \ X_{20} > 1.915 \ \& \ X_{20} \leq 2.458 \ \& \ X_4 \leq 0.708$	100,00%	9	
11	$X_{21} \leq 184.173 \ \& \ X_3 \leq 0.081 \ \& \ X_4 \leq 0.577$	100,00%	4	
12	$X_{21} > 184.173 \ \& \ X_{20} > 2.458 \ \& \ X_4 \leq 0.396$	66,67%	2	
13	$X_{21} \leq 184.173 \ \& \ X_{16} \leq 0.615 \ \& \ X_3 > 0.903$	100,00%	1	
14	$X_{21} > 184.173 \ \& \ X_{20} > 1.296 \ \& \ X_{20} \leq 1.915 \ \& \ X_3 \leq 0.005$	100,00%	1	
15	$X_{21} > 184.173 \ \& \ X_{20} \leq 1.296 \ \& \ X_{10} \leq 15.395 \ \& \ X_3 > 3.166 \ \& \ X_3 > 5.289$	100,00%	1	

Tablo 4’te görüldüğü üzere CART algoritması ile bir şirketin hisse senedi getirisi sınıflandırma tahminine ilişkin en önemli iki kural aşağıda verilmiştir.

**Kural 1:** Eğer bir şirketin  $X_{21}$  değişkeni “TÜFE” 184.173’e eşit ya da küçük,  $X_3$  değişkeni “nakit oranı” 0,081’den büyük ve  $X_{16}$  değişkeni “kısa vadeli borç/ toplam borç” 0.615’ten büyük ise o şirket %100 (19 şirket) oran ile hisse senedi pozitif bir getiriye sahip olan bir işletmedir.

**Kural 9:** Eğer bir şirketin  $X_{21}$  değişkeni “TÜFE” 184.173’den büyük,  $X_{20}$  değişkeni “piyasa değeri/defter değeri” 1.296’ya eşit ya da küçük,  $x_{10}$  değişkeni “aktif karlılık” 15.395’e eşit ya da küçük ve  $X_3$  değişkeni “nakit oranı” 3.166’ya eşit ya da küçük ise o şirket %100 (21 şirket) oran ile hisse senedi negatif bir getiriye sahip olan bir işletmedir.

#### 5.4. DVM Analizi

Analizin bu kısmında bir işletmenin pozitif ve negatif hisse senedi getirisi sınıflandırma tahmini için DVM yöntemi uygulanmıştır. DVM analizi için kernel fonksiyonlarından doğrusal, polinom, radyal (gauss), ve sigmoid fonksiyonları denenerek sınıflandırma

sonuçları elde edilmiştir. Tablo 5’te kernel fonksiyonları ile ilişkili eğitim ve test skorları verilmiştir.

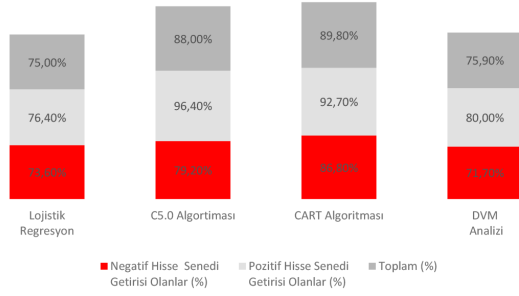
**Tablo 5:** Kernel Fonksiyonları ile İlişkili DVM Sınıflandırıcılarının Performansları

Kernel Tipi	Eğitim Performansı (%)	Test Performansı (%)
Doğrusal	74,7	80
Polinom	96,39	68
Radyal (Gauss) Tabanlı	72,29	80
Sigmoid	51,81	64

Tablo 5’te görüldüğü üzere en yüksek test performansını gösteren DVM sınıflandırıcısı doğrusal ve radyal tabanlı kernel fonksiyonları olmuştur. En iyi test performansını gösteren ve beraberinde eğitim performansı daha yüksek olan doğrusal kernel fonksiyonu seçilerek DVM analizi gerçekleştirilmiştir.

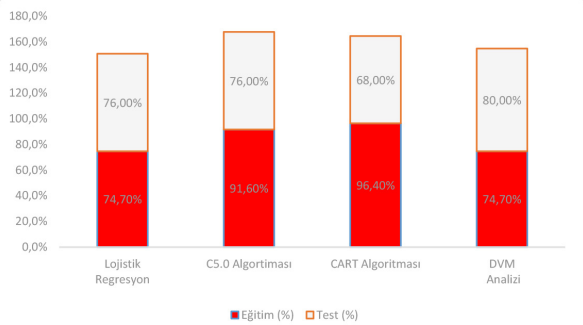
#### 5.5 Analiz Yöntemlerinin Karşılaştırılması

Grafik 1’de analiz yöntemlerine göre işletmelerin hisse senedi getirisi sınıflandırma performans sonuçları gösterilmiştir.



**Grafik 1:** Analiz Yöntemlerine Göre Hisse Senedi Getiri Sınıflandırma Performans Sonuçları

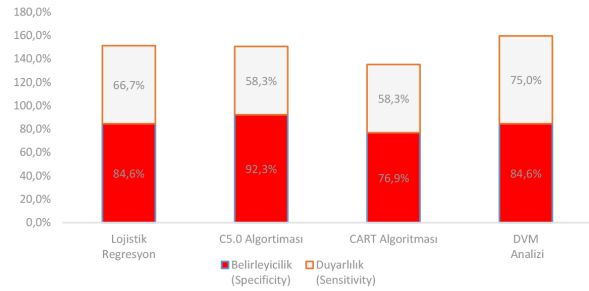
Grafik 1’de görüldüğü üzere negatif hisse senedi getirisine sahip işletmeleri %86,8’lik en yüksek başarı düzeyi ile CART algoritması; pozitif hisse senedi getirisine sahip işletmeleri %96,4’lük en yüksek başarı düzeyi ile C5.0 algoritması gerçekleştirmiştir. CART algoritması için toplam doğru sınıflandırma başarı düzeyi %89,8’dir. Analiz yöntemlerinin hisse senedi getiri sınıflandırma performans sonuçlarına bakılarak karşılaştırılması hatalı olabilir. Bundan dolayı analiz yöntemlerine ait eğitim ve test performansları ile belirleyicilik (specificity) ve duyarlılık (sensitivity) istatistiklerinin yorumlanması gerekir.



**Grafik 2:** Analiz Yöntemlerine Göre Eğitim ve Test Performans Sonuçları

Grafik 2’de analiz yöntemlerine ait eğitim ve test setinin performans sonuçları verilmiştir. Test verisi için en yüksek sınıflandırma düzeyini % 80 ile DVM analizi gerçekleştirirken, en kötü sınıflandırma düzeyini %68 ile CART algoritması gerçekleştirmiştir. Buna karşın CART algoritması eğitim seti için %96 ile en yüksek sınıflandırma performansını göstermiştir. Bu durum toplam sınıflandırma performansına dayalı analizlerin test performanslarının karşılaştırılması açısından

yanıltıcı olabilir. Bu nedenden dolayı istatistiğe ilişkin belirleyiciliğe (specificity) ve duyarlılığa (sensitivity) dayalı sınıflandırıcıların test performansları karşılaştırılır. Belirleyicilik istatistiği (specificity): hisse senedi getirisi pozitif olan toplam işletme içerisinde doğru sınıflandırılarak tahmin edilmiş hisse senedi getirisi pozitif olan işletme sayısıdır. Duyarlılık istatistiği (sensitivity): hisse senedi getirisi negatif olan toplam işletme içerisinde doğru sınıflandırılarak tahmin edilmiş hisse senedi getirisi negatif olan işletme sayısıdır (Han ve Kamber; 2006; Öğüt, vd. 2009). Bu açıdan ele alındığında yatırımcı gerçekte negatif hisse senedi getirisini pozitif hisse senedi getirisi olarak tahmin edildiğinde yapmış olduğu yatırımın değeri azalacaktır. Yatırımcı eğer pozitif hisse senedi getirisini negatif hisse senedi getirisi olarak tahmin edildiğinde kârlı yatırım fırsatını kaçıracaktır. Böylece, daha yüksek duyarlılık (sensitivity) istatistiği belirleyicilik istatistiğine (specificity) göre daha önemli olabilmektedir. Grafik 3’de analiz yöntemlerine göre belirleyicilik (specificity) ve duyarlılık (sensitivity) istatistikleri verilmiştir.



**Grafik 3:** Analiz Yöntemlerine Göre Belirleyicilik ve Duyarlılık İstatistiklerinin Karşılaştırılması

Grafik 3’te görüldüğü üzere analiz yöntemleri açısından duyarlılık (sensitivity) istatistiğinde en iyi analiz yöntemini %75 ile DVM analizi; belirleyicilik (specificity) istatistiğinde en iyi analiz yöntemini %92,3 ile C5.0 algoritması gerçekleştirmiştir. Böylece işletmelerin pozitif ve negatif hisse senedi getiri sınıflandırma tahmininde duyarlılık (sensitivity) istatistiği için DVM analizinin, belirleyicilik (specificity) istatistiği için C5.0 algoritmasının daha yüksek doğrulukla sınıflandırma gerçekleştirdikleri tespit edilmiştir. Tablo 6’da hisse senedi getiri sınıflandırmasına yönelik benzer çalışmalara yer verilmiştir.

**Tablo 6:** Benzer çalışmalarla sonuçların karşılaştırması

Yazar/yıl	Veri Seti	Girdi Verileri	Analiz Yöntemleri	En iyi Sınıflandırma Doğruluğu
Barak ve Modarres, (2015)	İran Borsası	44 finansal oran ve borsa endeksi	LAD, CART, J48, DTNM karar ağaçları.....	%80,24
Patel, vd., (2015)	Hindistan Borsası	10 teknik parametre	ANN, SVM, Random forest ve Naive-Bayes	%90,19
Zhang, vd., (2014)	Şangay Borsası	50 finansal oran ve borsa endeksi	LR, NB, BN,NN, SVM, J48, RF	%54,70
Hargreaves ve Hao (2013)	Avustralya Borsası	5 finansal oran ve borsa endeksi	C5.0, CHAID, YSA, LR	%94,20
Dutta, vd., (2012)	Hindistan Borsası	8 finansal oran ve borsa endeksi	LR	%74,60
Tsai, Lu ve Yen (2012)	Tayvan Borsası	61 duran varlıklar değişkeni	PCA, LR, DT, AR, GA, ANN	%75,06
Tsai, vd., (2011)	Tayvan Borsası	19 finansal oran ve 11 makroekonomik değişken	ANN, CART, LR	%66,67
Chang (2011)	Tayvan Borsası	10 adet hisse senedi fiyat	ANN, CART, Hybrid Model	%97,00
Kara, Boyacıoğlu, Baykan, (2011)	İstanbul Borsası	10 adet teknik gösterge	ANN, SVM	%75,74
Tsai ve Hsiao (2010)	Tayvan Borsası	14+17 olmak üzere finansal ve makroekonomik değişken	PCA, GA, CART	%79
Lai, vd., (2009)	Tayvan Borsası	6 temel ve 7 teknik endeks değişkeni	FDT, GAFDT	%82
Lu ve Chen,, (2009)	Tayvan Borsası	17 karakteristik değişken	C5.0	%91
<b>Önerdiğimiz model</b>	<b>Borsa İstanbul</b>	<b>20 finansal oran ve TÜFE</b>	<b>LR, C5.0, CART, DVM</b>	<b>%89,8</b>

Tablo 6'da hisse senedi getiri sınıflandırmasını amaçlayan literatürdeki benzer çalışmalar özetlenmiş ve önerdiğimiz modelin sonuçları karşılaştırılmıştır. Barak ve Modarres (2015)'in önerdiği modelin doğru sınıflandırma performansı %80,24 iken, Patel vd.(2015)'in önerdiği model %90,19'dur. Buna karşın bizim önerdiğimiz modelin en yüksek doğru sınıflandırma sonucu %89,8'dir. Analiz yöntemleri açısından kıyaslama yapılırken yatırımcılar açısından duyarlılık istatistiğinin karşılaştırılması daha doğrudur. Barak ve Modarres (2015)'in önerdiği J48 karar ağacının duyarlılık istatistiği %70,9 iken önerdiğimiz DVM analizinin duyarlılık istatistiği %75'dir. Ayrıca çalışmaya ait verilerimizin büyüklüğü diğer yerli çalışmalar ile karşılaştırıldığında; Ege ve Bayrakdaroğlu (2009), 18 şirketin 2004 yılı verilerini kullanarak hisse senedi getiri başarıları tahmininde lojistik regresyon analizi ile %72,2'lik sınıflandırma başarıları gerçekleştirmişlerdir. Özdemir vd. (2011), İMKB-100 endeksinde 1997-2010 yılları arasında işlem gören şirketler için endeks getiri tahmini hesaplamışlardır. 167 aylık veri setine yönelik DVM için %75'lik ve lojistik regresyon analiz için %86'lık doğru sınıflandırma başarıları tespit etmişlerdir. Öz vd. (2011), İMKB 30 endeksinde 2005-2007 yılları

arasında işlem gören 24 şirketin hisse senedi getiri tahmin için diskriminant analizi kullanmışlardır. Bir yıl öncesinden %75 ve iki yıl öncesinden %91,7'lik doğru sınıflandırma başarıları gerçekleştirmişlerdir. Çalışmamızda ise 2009-2014 yılları arasında 18 şirketin 108 kaydı için hisse senedi getiri sınıflandırmaları tahmin edilmiş ve sırasıyla LR ile %75, C5.0 ile %88, CART ile %89,8 ve DVM ile %75,9'luk doğru sınıflandırma başarıları elde edilmiştir. Önerdiğimiz modelin değişkenleri ile birlikte pozitif ve negatif hisse senedi getiri sınıflandırma tahmininde kullanılmasının uygun olabileceği düşünülmektedir. Oluşturulacak olan modelin değişkenlerine ait ayrıntılı analizi ile birlikte alternatif bir analiz tekniği kullanarak daha yüksek sınıflandırma sonucu elde edilebilir.

İleriye yönelik tahminlerde bulunmak için karar ağacı kuralları bir şirketin pozitif ya da negatif getiriye sahip olup olmayacağı hakkında bilgi üretebilmektedir. Lojistik regresyon analizinde ise bağımsız değişkenlerin değerleri modelde yerine konmak suretiyle elde edilecek bağımlı değişkenin olasılık değeri 0,50'nin üzerinde ise pozitif hisse senedi getirisi altında ise negatif hisse senedi getirisi şeklinde yorumlanabilecektir.

## SONUÇ

Bu çalışmada 2009-2014 yılları arasındaki Türkiye'de BIST-100 endeksi içerisinde işlem gören kimya, kauçuk ve plastik ürünleri imalatı sanayinde yer alan şirketlerin pozitif ve negatif hisse senedi getiri sınıflandırma tahminine etki eden faktörlerin belirlenerek, hisse senedi getirilerinin tahmin edilmesi amaçlanmıştır. Lojistik regresyon analizi, C5.0 algoritması, CART algoritması ve DVM yöntemleri kullanılarak şirketlere ait 20 finansal oran ve TÜFE değişkenleri ile analiz işlemleri gerçekleştirilmiştir. Lojistik regresyon analizi, C5.0 algoritması ve CART algoritmasının analiz sonuçları dikkate alındığında pozitif ve negatif hisse senedi getiri sınıflandırma tahminine etki eden en önemli faktörün "piyasa/defter değeri değişkeni" ardından "TÜFE değişkeni" olduğu belirlenmiş ayrıca lojistik regresyon analizinde "brüt kar marjı değişkeni" de istatistiksel açıdan anlamlı olduğu tespit edilmiştir. C5.0 algoritması ile 12 ve CART algoritması ile 15

kural üretilmiştir. Analiz yöntemlerinin eğitim performansı açısından incelendiğinde veri madenciliği tekniklerinin lojistik regresyon analizine göre daha iyi performans gösterdiği tespit edilmiştir. Yatırımcı açısından bakıldığında DVM analizi duyarlılık (sensitivity) istatistikinde en iyi sınıflandırma sonucunu göstermiştir. Ayrıca önerdiğimiz modelin sonuçları literatürdeki benzer çalışmalarla karşılaştırılmıştır. Analizlerden elde edilen sonuçlara göre hisse senedi getiri sınıflandırma tahmininde LR, C5.0, CART algoritması ve DVM analizlerinin literatürdeki diğer analiz yöntemleri kadar önemli performans göstermiştir. Sonuç itibarı ile bu çalışma hisse senedi yatırım kararlarında yatırımcılara ve analizcilere faydası olabileceği düşünülmektedir. Bundan sonraki çalışmalarda BIST 100 endeksine dahil işletmeleri kapsayacak şekilde çalışmanın örneklemini geniş tutularak parçacık sürü optimizasyonu ve genetik algoritma yöntemleri uygulanarak hisse senedi getirileri optimize edilmeye çalışılacaktır.

## KAYNAKÇA

- Abe, S. (2005) Support Vector Machines For Pattern Classification. London, Springer.
- Akcan, A. ve Kartal, C. (2011) "İMKB Sigorta Endeksini Oluşturan Şirketlerin Hisse Senedi Fiyatlarının Yapay Sinir Ağları İle Tahmini" *Muhasebe ve Finansman Dergisi*, 27-40.
- Akpınar, H. (2014) Data Veri Madenciliği Veri Analizi, Papatya Yayıncılık, İstanbul.
- Alpar, R. (2013) Çok Değişkenli İstatistiksel Yöntemler, 4. Baskı. Ankara, Detay Yayıncılık.
- Alpaydın, E. (2011) Yapay Öğrenme, İstanbul, Boğaziçi Üniversitesi Yayınevi.
- Altay, E. ve Satman, M. H. (2005) "Stock Market Forecasting: Artificial Neural Network And Linear Regression Comparison In An Emerging Market" *Journal of Financial Management & Analysis*, 18(2).
- Atan, M., Atan, S. ve Özdemir, Z. A. (2009) "Hisse Senedi Piyasasında Zayıf Formda Etkinlik: İMKB Üzerine Ampirik Bir Çalışma" *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 24(2), 33-48.
- Avcı, E. (2009) "Stock Return Forecasts With Artificial Neural Network Models" *Marmara Üniversitesi İ.İ.B.F Dergisi*, 26(1), 443-461.
- Avcı, E. ve Çinko, M. (2008) "Endeks Getirilerinin Yapay Sinir Ağları Modelleri İle Tahmin Edilmesi: Gelişmekte Olan Avrupa Borsaları Uygulaması" *İktisat İşletme ve Finans*, 137-114 ,(266)23.
- Ayhan, S. ve Erdoğan, Ş. (2014) "Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi" *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, 9(1), 175-198.
- Barak, S. ve Modarres, M. (2015) "Developing an approach to evaluate stocks by forecasting effective features with data mining methods" *Expert Systems with Applications*, 42(3), 1325-1339.
- Bekçioğlu, S. ve Ada, E. (1985) "Menkul Kıymetler Piyasası Etkin mi?" *Muhasebe Enstitüsü Dergisi*, Sayı 41.
- Bekçioğlu, S., Öztürk, M. ve Doğanlı, B. (2004) "Türk Hisse Senedi Piyasasının Zayıf Şekilde Etkinliğinin Test Edilmesi" *Muhasebe ve Finansman Dergisi*, 22: s.39-48.
- Berry, M.J. ve Linoff, G.S. (2004) Data mining techniques: for marketing, sales and customer relationship management, John Wiley & Sons.
- Buguk, C. Ve Brorsen, B. (2003) "Testing weak-form market efficiency: Evidence from the Istanbul Stock Exchange" *International Review of Financial Analysis*, 12(5), 579-590.
- Burges, C. J. C. (1998) A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery: Boston, USA: Kluwer Academic Publishers.
- Campbell, C. ve Ying, Y. (2011) Learning with support vector machines. Synthesis Lectures on Artificial Intelligence and Machine Learning. San Rafael: Morgan & Claypool Publishers.



- Cao, Q., Parry, M. E., ve Leggio, K. B. (2011) “The three-factor model and artificial neural networks: predicting stock price movement in China” *Annals of Operations Research*, 185(1), 25-44.
- Chang, P. C., Fan, C. Y., ve Lin, J. L. (2011) “Trend discovery in financial time series data using a case based fuzzy decision tree” *Expert Systems with Applications*, 38(5), 6070-6080.
- Chang, T. S. (2011). “A comparative study of artificial neural networks, and decision trees for digital game content stocks price prediction” *Expert Systems with Applications*, 38(12), 14846-14851.
- Cheng, J. H., Chen, H. P., ve Lin, Y. M. (2010) “A hybrid forecast marketing timing model based on probabilistic neural network, rough set and C4.5” *Expert systems with Applications*, 37(3), 1814-1820.
- Çokluk, Ö., Şekercioğlu, G., ve Büyüköztürk, Ş. (2012) Sosyal Bilimler İçin Çok Değişkenli İstatistik SPSS ve Lisrel Uygulamaları, Pegem Akademi, Ankara.
- Dayı, F. ve Ata, H.A. (2012) Yapay Sinir Ağı ile Hisse Senedi Getirisi Tahmini: Bir Firma Uygulaması, 16. *Finans Sempozyumu*, 10-13 Ekim 2012, Erzurum, 181-194.
- de Oliveira, F. A., Nobre, C. N., ve Zarate, L. E. (2013) “Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index–Case study of PETR4, Petrobras, Brazil” *Expert Systems with Applications*, 40(18), 7596-7606.
- Dutta, A., Bandopadhyay, G., ve Sengupta, S. (2012) “Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression” *International Journal of Business & Information*, 7(1).
- Ege, İ. ve Bayrakdaroğlu, A. (2009) “İMKB Şirketlerinin Hisse Senedi Getiri Başarılarının Lojistik Regresyon Tekniği İle Analizi” *ZKÜ Sosyal Bilimler Dergisi*, Cilt 5, Sayı 10, ss. 139–158.
- Eken, H. ve Adalı, S. (2008). Piyasa Etkinliği ve İMKB: Zayıf Formda Etkinliğe İlişkin Ekonometrik Bir Analiz. *Muhasebe ve Finansman Dergisi*, (37).
- Elton Edwin J. ve Martin J. Gruber (1981) *Modern Portfolio Theory*, Second Edition, John Wiley & Sons.
- 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.
- Ergül, N. (2009) “Ulusal Hisse Senedi Piyasasında Etkinlik” *Yönetim Bilimleri Dergisi* (7:1 *Journal of Administrative Sciences*, ss.101-117.
- Fama, E. F. (1970) “Efficient capital markets: A review of theory and empirical work” *The Journal of Finance*, 25(2), 383-417.
- Field, A. (2005) *Discovering statistics using SPSS*, 2nd Edition, London, Sage Publications.
- Grabczewski, K. (2014) *Meta-learning in decision tree induction*. Springer.
- Hair, J.F., Black, W.C., Babin, B., Anderson, R.E., ve Tatham, R.L. (2010) *Multivariate data analysis*, Seventh Edition, Pearson Prentice Hall Publishing, New Jersey.
- Hamel, L.H. (2009) *Knowledge Discovery with Support Vector Machines*. New Jersey: Wiley Publication.
- Han, J. ve Kamber, M. (2006) *Data Mining: Concepts and Techniques*, The Morgan Kaufmann.
- Hargreaves, C. ve Hao, Y. (2013) “Prediction of Stock Performance Using Analytical Techniques” *Journal of Emerging Technologies in web Intelligence*, Vol.5., No.,2, 136-142.
- Haykin S. (1999) *Neural Networks: A Comprehensive Foundation* [Elektronik Sürüm], Prentice Hall Inc, New Jersey.
- Hosmer, D.W. ve Lemeshow, S. (2000) *Applied Logistic Regression*, Second Edition, A Wiley-Interscience Publication, New York.
- Huang, C.J., Yang, D.X., ve Chuang, Y.T. (2008) Application of Wrapper Approach And Composite Classifier To Stock Trend Prediction, *Expert Systems with Applications*, 34(4), 2870-2878.
- Ince, H. Ve Trafalis, T. B. (2008) “Short term forecasting with support vector machines and application to stock price prediction” *International Journal of General Systems*, 37(6), 677-687.
- Kahraman, D. ve Erkan, M. (2005) “İstanbul Menkul Kıymetler Borsası’nda Tesadüfi Yürüyüş Testi” *Celal Bayar Üniversitesi, İİBF, Yönetim ve Ekonomi Dergisi*, 12(1), 11-24.
- Kalaycı, Ş. (2008) *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*, 3.Baskı, Ankara, Asil Yayın.
- Kara, Y., Boyacıoğlu, M. A., ve Baykan, Ö. K. (2011) “Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange” *Expert systems with Applications*, 38(5), 5311-5319.
- Kıyılar, M. (1997) Etkin pazar kuramı ve etkin pazar kuramının İMKB’de irdelenmesi: test edilmesi. Sermaye Piyasası Kurulu, Ankara.
- Lai, R. K., Fan, C. Y., Huang, W. H., ve Chang, P. C. (2009) “Evolving and clustering fuzzy decision tree for

- financial time series data forecasting" *Expert Systems with Applications*, 36(2), 3761-3773.
- Lam, M. (2004) "Neural network techniques for financial performance prediction: integrating fundamental and technical analysis" *Decision Support Systems*, 37(4), 567-581.
- Larose, D.T., ve Larose, C.D. (2014) *Discovering Knowledge In Data An Introduction To Data Mining*, New Jersey: John Wiley & Sons.
- Lu, C. L. ve Chen, T. C. (2009) "A study of applying data mining approach to the information disclosure for Taiwan's stock market investors" *Expert Systems with Applications*, 36(2), 3536-3542.
- Menard, S. (2002) *Applied Logistic Regression Analysis*, Second Edition, Sage Publication, London.
- Mertler, C.A. ve Vannatta, R.A. (2005) *Advanced and Multivariate Statistical Methods: Practical Application and Interpretation*, CA: Pyrczak Publishing.
- O'Connell, A.A. (2006) *Logistic Regression Models for Ordinal Response Variables*. Thousand Oaks, Sage Publications, London.
- Ou, P. ve Wang, H. (2009) "Prediction of stock market index movement by ten data mining techniques" *Modern Applied Science*, 3(12), 28-42.
- Öğüt, H., Doğanay, M. M., ve Aktaş, R. (2009) "Detecting stock-price manipulation in an emerging market: The case of Turkey" *Expert Systems with Applications*, 36(9), 11944-11949.
- Öz, B., Ayrıçay, Y., ve Kalkan, G. (2011) "Finansal Oranlarla Hisse Senedi Getirilerinin Tahmini: İMKB 30 Endeksi Hisse Senetleri Üzerine Diskriminant Analizi İle Bir Uygulama" *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 11(3), 51-64.
- Özdamar, K. (2004) *Paket Programlar İle İstatistiksel Veri Analizi*. Genişletilmiş 5.Baskı. Eskişehir, Kaan Kitabevi.
- 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.
- Özkan, Y. (2008) *Veri Madenciliği Yöntemleri*, Papatya Yayıncılık, İstanbul.
- Panda, C., ve Narasimhan, V. (2006) "Predicting Stock Returns An Experiment of the Artificial Neural Network in Indian Stock Market" *South Asia Economic Journal*, 7(2), 205-218.
- Patel, J., Shah, S., Thakkar, P., ve 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.
- Ren, N., Zargham, M., ve Rahimi, S. (2006) "A decision tree-based classification approach to rule extraction for security analysis" *International Journal of Information Technology & Decision Making*, 5(01), 227-240.
- Quinlan, J.R. (1993) *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.
- Quinlan, J.R. (1996) "Improved use of continuous attributes in C4.5" *Journal of Artificial Intelligence Research*, 4:77-90.
- Scholkopf, B. ve Smola A.J. (2002) *Learning with Kernels*, Cambridge: MIT Press.
- Sevim, Ş., Yıldız, B., ve Akkoç, S. (1997) "Aşırı Tepki Verme Hipotezi ve İMKB Üzerine Ampirik Bir Çalışma" *UMKB Dergisi*, Cilt 9 Sayı 35, ss.23-39.
- Seyrek, İ.H., ve Ata, H. A. (2010) "Veri Zarflama Analizi ve Veri Madenciliği ile Mevduat Bankalarında Etkinlik Ölçümü", *BDDK Bankacılık ve Finansal Piyasalar Dergisi*, Cilt 4, Sayı:2, ss 67-84.
- Shahnaz, F. (2006) *Decision Tree Based Algorithms*, Michael W. Berry (Ed.), *Lecture Notes in Data Mining*, USA: World Scientific Publisher.
- Tayyar, N. ve 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), 189-218.
- Tektaş, A., ve Karataş, A. (2004) "Yapay Sinir Ağları ve Finans Alanına Uygulanması: Hisse Senedi Fiyat Tahminlemesi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 337-349 ,(4-3) 18.
- Thawornwong S. ve Enke, D. (2004) "The adaptive selection of financial and economic variables for use with artificial neural networks" *Neurocomputing*, 56, 205-232.
- Toraman, C. (2008) "Demir-Çelik Sektöründe Yapay Sinir Ağları İle Hisse Senedi Fiyat Tahmini: Erdemir A.Ş. ve Kardemir A.Ş. Üzerine Bir Tahmin Uygulaması" *Muhasebe ve Finansman Dergisi*, Sayı:39, s.44-57.
- Tsai, C. F. ve Hsiao, Y. C. (2010) "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches" *Decision Support Systems*, 269-258 ,(1)50.
- Tsai, C. F., Lu, Y. H., ve Yen, D. C. (2012) "Determinants of intangible assets value: The data mining approach" *Knowledge-Based Systems*, 31, 67-77.
- Tsai, C.F., Lin, Y.C., Yen, D. C., ve Chen, Y.M. (2011) "Predicting stock returns by classifier ensembles" *Applied Soft Computing*, 11, 2452-2459.

Tunçel, A. K. (2007) "Rassal Yürüyüş (Random Walk) Hipotezinin İMKB'de Test Edilmesi: Koşu Testi Uygulaması" *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 9(2), 1-18.

Vapnik, V. N. (1995) *The Nature of Statistical Learning Theory*, NewYork: Springer-Verlag.

Vapnik, V.N. ve Chervonenkis A. (1974) *Theory of Pattern Recognition (in Russian)*, Nauka, Moscow. (German translation: W.N. Wapnik, A. Ja. Tschervonenkis (1979), *Theorie der Zeichenerkennung*, Akademie-Verlag, Berlin.

Yildiz, B., Coskun, M., ve Yalama, A. (2008) "Forecasting the Istanbul Stock Exchange National 100 Index Using an Artificial Neural Network" *Proceedings of World Academy of Science: Engineering & Technology*, 48.

Zengin, H. ve Kurt, S. (2004) "İMKB'nin Zayıf ve Yarı Güçlü Formda Etkinliğinin Ekonometrik Analizi" *Öneri Dergisi*, 21(6), 145-152.

Zeren, F., Kara, H., ve Arı, A. (2013) "Piyasa Etkinliği Hipotezi: İMKB İçin Ampirik Bir Analiz" *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (36), 141-148.

Zhang, X., Hu, Y., Xie, K., Wang, S., Ngai, E. W. T., ve Liu, M. (2014) "A causal feature selection algorithm for stock prediction modeling" *Neurocomputing*, 142, 48-59.

Zivot, E. (2007) *Introduction to Computational Finance and Financial Econometrics*, Chapter 1 Asset Return Calculations.