

ENDEKS GETİRİSİ YÖNÜNÜN İKİLİ SINIFLANDIRMA YÖNTEMİYLE TAHMİN EDİLMESİ: İMKB-100 ENDEKSİ ÖRNEĞİ

A.Kerem ÖZDEMİR*
Seda TOLUN**
Ebru DEMİRÇİ***

ÖZET

Ampirik finansın temel araştırma konularından hisse senedi getirilerinin tahmininde istatistikî yöntemlere alternatif olarak daha az varsayım gerektirerek esneklik sağlayan ve gerçek riske dayanan genelleme kabiliyeti daha üstün olan makine öğrenmesi tekniklerinin giderek artan bir yoğunlukta kullanıldığı görülmektedir. Çalışmada ikili sınıflandırmaya imkan tanıyan lojistik regresyon ve destek vektör makineleri yöntemleri ile İMKB-100 endeksinin getirisinin yönü tahmin edilmiştir. Her iki yöntem için modelleme kümesi ve tahmin kümesi için doğru sınıflandırma oranları sırasıyla yaklaşık %75 ve %86 olarak elde edilmiştir. Destek vektör makineleri yönteminin yatırımcılar ve araştırmacılar tarafından hisse senedi getirilerinin tahmininde alternatif bir yöntem olarak etkin bir şekilde kullanılabileceği çalışmanın sonuçlarından görülmektedir.

Anahtar Kelimeler: Endeks Getiri Tahmini, Destek Vektör Makineleri, Lojistik Regresyon.

JEL Sınıflaması: G17, G12, C63

PREDICTING THE DIRECTION OF STOCK INDEX RETURN MOVEMENT USING BINARY CLASSIFICATION METHOD: AN APPLICATION ON THE ISE-100 INDEX

ABSTRACT

Machine learning techniques are seen to be used increasingly in estimating stock returns as an alternative to statistical methods with their ability to provide flexibility requiring relatively less assumptions and providing better generalization ability based on real risk. In this study, the direction of movement of the ISE-100 Index return is estimated using both logistic regression and support vector machines, making use of their ability to facilitate a binary classification. The correct classification ratio for modeling and holdout sets of logistic regression and support vector machines methods are obtained as 75% and 86% respectively. It can be seen from the results of the study that the support vector machines can be used effectively as an alternative method by investors and researchers in predicting the stock returns.

Keywords: Stock Index Return Prediction, Support Vector Machines, Logistic Regression.

JEL Classification: G17, G12, C63

* İstanbul Üniversitesi, İşletme Fakültesi Finans Anabilim Dalı, keremozd@istanbul.edu.tr

** İstanbul Üniversitesi, İşletme Fakültesi Sayısal Yöntemler Anabilim Dalı, stolun@istanbul.edu.tr

*** İstanbul Üniversitesi, Ulaştırma ve Lojistik Yüksekokulu, Ulaştırma ve Lojistik Yönetimi Anabilim Dalı, edemirci@istanbul.edu.tr

GİRİŞ

Hisse senedi getirilerinin tahmin edilebilirliği ampirik varlık fiyatlama alanında en çok ve en hararetli şekilde tartışılan konuların başında gelmektedir. Çeşitli istatistik/ekonometrik ve son dönemlerde daha yoğun olmak üzere makine öğrenmesi yöntemleri ile hisse senedi fiyatlarının veya getirilerinin tahmin edilebilirliğine yönelik geniş ve hacimli bir literatür oluşmuştur. Özellikle son 30 yıllık dönemde finans alanında yapılan ampirik çalışmalar çeşitli finansal ve makroekonomik değişkenlerin hisse senedi getirilerini açıklayabilme gücünü ortaya koymayı amaçlamıştır (Hashem ve Timmermann, 1995).

Makroekonomik değişkenleri kullanarak hisse senedi getirilerinin tahmin edilebilirliğini araştıran çalışmalarda sıklıkla faiz oranları (Campbell, 1990), imalat sanayi üretimi (Balvers vd., 1990), parasal büyüklükler (Geske ve Roll, 1983), döviz kurları (Griffin ve Stulz, 2001; Bailey ve Chung, 1995), enflasyon oranları (Fama, 1981; Fama ve Schwert, 1977; Siklos ve Kwok, 1999) ve tüketim harcamaları (Lettau ve Ludvigson, 2001) gibi değişkenler kullanılmıştır. Bu çalışmaların büyük bir kısmı hisse senedi getirilerinin makro değişkenlerle tahmin edilebileceğini iddia etmekle birlikte ampirik bulgular çelişkilidir. Araştırmacıların bir kısmı makro değişkenlerin hisse senedi getirilerini tahmin etme kabiliyetinin zaman boyutunda farklılıklar gösterdiğini (Durham, 2001), diğer bir kısmı ise belirli bir makro değişken kümesi ile hisse senedi getirileri arasında anlamlı ve güçlü ilişkiler olmadığını (Chan vd., 1998) göstermiştir. Netice itibarıyla, bu çalışmalar arasında söz konusu makro değişkenlerin neler olabileceğine dair bir mutabakatın oluştuğunu söylemek pek mümkün görünmemektedir.

Diğer taraftan hisse senedi fiyatlarının/getirilerinin seyrinin şirket karakteristikleri ve farklı değerlendirme parametreleri ile tahmin edilebileceğini ortaya koyan çok sayıda çalışma da yapılmıştır. İlgili literatürde temettü verimi (Fama ve French, 1988; Kothari ve Shanken, 1997), kârlılık oranları, nakit akışı, pazar değeri/defter değeri ve fiyat/kazanç oranı (Jaffe ve Keim, 1989) parametrelerinin hisse senedi fiyatlarını açıklama gücü olduğu vurgulanmıştır (Campbell, 1987; Fama ve French, 1988, 1993; Fama ve Schwert, 1977; Keim ve Stambaugh, 1986).

Zaman serisi analiz yöntemlerinde, özellikle serilerin yapısına dair çok güçlü varsayımlarda bulunulması ve hisse senedi fiyatlarıyla açıklayıcı değişkenler arasındaki ilişkilerin doğrusal olduğunun varsayılması (Enke ve Thawornwong, 2005:928) gibi ciddi metodolojik sorunlar bulunmaktadır. Bu sebepler, son 15-20 yıllık dönemde araştırmacıları, hisse senedi getirilerinin tahmininde, başta yapay sinir ağları olmak üzere, ekonometrik yöntemlere kıyasla daha az kısıtlayıcı varsayım içeren ve daha esnek yapılara sahip olan çeşitli makine öğrenmesi yöntemleri geliştirmeye sevk etmiş ve söz konusu yöntemlere dayalı olan literatür ciddi bir hacme ulaşmıştır (Na ve Sohn, 2011:9046).

Kim (2003) yaptığı çalışmada 1989–1998 döneminde KOSPI endeksinin yönünü günlük verilerle 12 teknik gösterge ile Destek Vektör Makineleri (DVM) yöntemini kullanarak tahmin etmiştir. Destek vektör makineleri yönteminin sonuçlarını CBR (Case Based Reasoning) ve BP Network yöntemleri ile karşılaştırarak DVM yönteminin finansal zaman serileri tahmininde iyi bir yöntem olduğunu belirtmiştir. KOSPI endeksinin günlük veriler kullanarak yönünün tahmin edildiği bir diğer çalışmada, 12 farklı teknik analiz göstergesinin girdi olarak

kullanıldığı çeşitli sinir ağıları modelleri %52 ile %62 oranları arasında değişen başarılı tahminler elde edilmiştir (Kim ve Han, 2000).

Diğer taraftan önde gelen uluslararası hisse senedi endekslerinin getirileri ile çeşitli yükselen ekonomi hisse senedi piyasası endeksleri arasındaki karşılıklı bağımlılık ve ilişki derecesini araştıran çalışmalarda, gelişmiş ekonomi hisse senedi piyasaları ile özellikle yükselen Asya piyasalarının kendi içlerinde entegre oldukları ve yükselen piyasa ekonomilerinin başta A.B.D. olmak üzere Japonya ve İngiltere hisse senedi piyasalarından etkilendiğini ve de özellikle ABD piyasalarının diğer gelişmiş ve gelişmekte olan piyasalar üzerinde baskın bir etkisi olduğu bulgulanmıştır (Masih ve Masih, 2001). Çeşitli gelişmiş ve yükselen piyasa ekonomileri arasında volatilité yayılma etkisini inceleyen çalışmalar ise çelişkili sonuçlar üretmektedir. Liu ve Pan (1997) – özellikle de Ekim 1987 A.B.D. hisse senedi piyasasındaki çöküş ertesinde daha da belirginleşen – ve Worthington ve Higgs (2004) yükselen Asya ekonomileri ile A.B.D., Japonya ve diğer gelişmiş ekonomilerin hisse senedi endeksleri arasında istatistikî açıdan anlamlı yayılma etkilerinin olduğu yönünde bulgular elde etmiştir. Beirne vd. (2009) çalışmalarında gelişmiş ülke hisse senedi piyasalarından 41 yükselen ekonomi piyasalarına yönelik anlamlı yayılma etkileri tespit etmiştir. Buna mukabil, Ng (2002) ve Baharumshah vd. (2003) ve Phylaktis ve Ravazzolo (2005) aksi yönde sonuçlara ulaşmışlardır.

Yükselen bir ekonomiye sahip Türkiye'ye olan ilgiyle birlikte endeksin getirisinin tahminine yönelik çalışmalar da artmıştır. Durukan (1999) tüketici fiyat endeksi, faiz oranı, sanayi üretim endeksi, döviz kuru ve para arzı değişkenlerinin Ocak 1986–Aralık 1996 döneminde İMKB-100 endeksi üzerine olan etkisini en küçük kareler yöntemini kullanarak incelemiştir. Çalışmada faiz oranının hisse senedi endeksi ile negatif yönlü ve güçlü bir ilişkisi olduğu, farklı model spesifikasyonlarına göre sanayi üretim endeksinin de anlamlı olduğu, ancak döviz kuru, para arzı ve enflasyon değişkenlerinin ise anlamlı olmadığı sonucuna ulaşılmıştır.

Diler (2003) İMKB 100 endeksinin günlük yönünü tahmin etmek için 1990–2003 dönemi için yedi teknik analiz göstergesi kullanarak yapay sinir ağıları yöntemi ile tahmin etmiş ve modelin tahmin başarısı %60,8 oranında gerçekleşmiştir.

Kara vd. (2011) yapay sinir ağıları ve destek vektör makineleri yöntemlerini kullanarak on farklı teknik analiz göstergesi ile İMKB 100 endeksinin yönünü tahmin etmeye çalışmışlardır. Yapay sinir ağlarının ortalama tahmin başarısının (%75,7), destek vektör makinelerinkinden (%71,2) anlamlı şekilde daha iyi olduğu sonucuna ulaşılmıştır.

Altay ve Satman (2005) yapay sinir ağıları ve regresyon yöntemleri ile günlük, haftalık ve aylık veriler kullanılarak İMKB 30 ve İMKB Tüm endekslerinin yönünü tahmin etmeye çalışmıştır. Ocak 1997–Şubat 2005 dönemlerinin ele alındığı çalışmada yapay sinir ağlarının endeksin yönünü daha iyi tahmin ettiği (tahmin başarı oranları günlük, haftalık ve aylık veriler için sırasıyla %57,8, %67,1 ve %78,3) ve alım-satım stratejilerinde de daha yüksek getiriler sağladığı gözlemlenmiştir.

Boyacıoğlu ve Avcı (2011) Ocak 1990–Aralık 2008 dönemi için yaptıkları çalışmada 6 makroekonomik değişkene ilâveten DJI (Dow Jones Industrial Average), DAX (Deutsche Aktien Index) ve BOVESPA (Bolsa de Volares de Sao Paulo) endekslerinin kapanış fiyatlarını

kullanarak İMKB-100 endeksinin ANFIS (Adaptive Neural Fuzzy Inference System) yöntemi ile % 98,3 oranında doğru tahmin etmiştir.

Avcı (2007) çeşitli yapay sinir ağları modelleri ile Ocak 1996–Haziran 2005 dönemi için İMKB-100 endeksinin günlük ve seanslık getirilerinin tahmin edilmesindeki etkinliğini incelemek amacıyla endeks değerleri ve getirileri ile işlem hacminin gecikmeli ve hareketli ortalama değerlerini girdi olarak kullandığı çalışmada modellerin performansının zamana bağlı olduğunu ve farklı dönemler için hiçbir modelin bir diğerine performans açısından üstünlük göstermediği sonucuna varmıştır. Elde edilen bir diğer bulgu da seans verilerinin günlük verilere kıyasla daha başarılı bir tahmin performansı sağladığı ve en az önemli olan değişkenlerin analizden çıkarılmasının modellerin tahmin performansını arttırdığı yönündedir.

Bu çalışmanın birinci bölümünde endeks getirisinin yönünü tahmin etmek amacıyla tercih edilen ikili sınıflandırma analizlerinden Destek Vektör Makineleri (DVM) ve Lojistik Regresyon (LR) yöntemlerinin teorik altyapısı üzerinde durulmuştur. İkinci bölümde ise veri seti ve analizlerde yer alan değişkenler açıklanarak, kullanılan iki ayrı yöntemin analiz sonuçları ve elde edilen bu sonuçların yorumlarına yer verilmiştir. Sonuç bölümünde, analiz bulgularının karşılaştırılması ve endeks getirisinin yönünün tahmin edilmesi üzerine yapılan bu çalışmanın literatüre katkısı üzerinde durulmuştur.

METODOLOJİ

Bu çalışmada endeks getirisinin yönünün tahmin edilmesi amaçlandığından ikili sınıflandırmaya imkan tanıyan yöntemlerden lojistik regresyon ve destek vektör makineleri kullanılmıştır.

1.1 LOJİSTİK REGRESYON

Lojistik regresyon ve doğrusal regresyon yöntemleri benzer özelliklere göstermesine rağmen en önemli farklılıkları lojistik regresyonun bağımlı değişkenin ikili şıklı değişken olması durumunda kullanılan bir yöntem olmasıdır. Bir diğer farklılık ise lojistik regresyonda varsayımlar açısından en küçük kareler tekniğine göre daha az kısıt olmasıdır.

İki şıklı durumda kukla değişken kullanılarak ilgilenilen şıkka “1” değerine “0” değeri verilir. Böylece dağılımın ortalaması $nP=1$ 'lerin toplamı olmaktadır. P ilgilenilen şıkkin gerçekleşme oranını, 1-P ise diğerinin gerçekleşme oranını göstermektedir. Çok şıklı durumda ise şıklara 1,2,3 vs. gibi değerler verilir (Orhunbilge, 2010:186).

Lojistik regresyon modelinin parametreleri, analitik olarak elde edilemediğinden, iteratif bir yöntem olan maksimum olabilirlik tekniğiyle tahmin edilmektedir. Lojistik regresyonun yönteminin kullanılabilmesi için gereken varsayımlar:

- Gözlemlerin bağımsız olması,
- Bağımsız değişkenlerle logit bağımlı değişken arasında doğrusal bir ilişki olması ve
- Bağımsız değişkenler arasında çoklu doğrusal bağlantı olmamasıdır.

Lojistik regresyon yönteminde bağımsız değişkenlerin kategorik değişkenlerle açıklanırken “0 yada 1 olma olasılığı” hesaplanmaktadır. Bağımlı değişken öncelikle logit değişkene (logaritmalar alınarak) dönüştürülür. Böylece şıklardan herhangi birinin gerçekleşme oranının tahmini yapılmaktadır.

Üstünlük Oranı değerlerinin doğal logaritması alındığında aşağıdaki denkleme ulaşılmaktadır (Sharma, 1996:319):

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1)$$

Denklem (1), olasılık değeri olarak (2)'deki gibi ifade edilebilir.

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad (2)$$

Lojistik regresyon parametreleri, analitik olarak elde edilemediğinden, iteratif bir yöntem olan maksimum olabilirlik (Maximum Likelihood) tekniği ile tahmin edilmektedir (Albayrak, 2006:439). Tahmin edilen logit bağımlı değişkenin değerleri ile gözlenen değerler arasındaki olabilirliğin maksimum yapılması amaçlanmaktadır.

1.2 DESTEK VEKTÖR MAKİNELERİ

Parametrik olmayan denetimli sınıflandırıcılardan biri olan Destek Vektör Makineleri (DVM), Vladimir Vapnik (1992) tarafından istatistikî öğrenme teorisi ve yapısal risk minimizasyonu bağlamında tanıtılmıştır (Li vd., 2006). DVM'ler, eğitim hataları toplamıyla sınırlandırılmış genelleme hatalarını gösteren yapısal risk minimizasyonunu gerçekleştirerek minimize edilmiş bir VC (Vapnik-Chervonenkis) boyutlu bir sınıflandırıcı yaratırlar. Bu açıdan, DVM ile bulunan fonksiyon, veriye yakınlık ve çözümün karmaşıklığı arasındaki geçiştir (Ali ve Abraham, 2002).

DVM'ler sınıflandırma görevini gerçekleştirmek için sınıfların istatistikî dağılım tahminlerine ihtiyaç duymayıp, sınıflandırma modelini marj maksimizasyonu kavramını kullanarak, yani eğitim kümesi ile ayırt etme yüzeyine en yakın noktaların arasındaki marjı maksimize etmek olarak tanımlanmaktadır. Öyle ki öğrenen makine sadece küçük ampirik risklere değil, aynı zamanda da iyi genelleme performansına da sahip olmaktadır (Zhang, 1999).

İkili sınıflandırma için genelleme sınırı, ampirik riski minimize eden fonksiyonu içeren tüm yaklaşım fonksiyonları için en az $1-\eta$ olasılıkla sağlanır (Cherkassky ve Shao, 2001). $0 \leq \eta \leq 1$ olacak şekilde beklenen risk üzerindeki sınır;

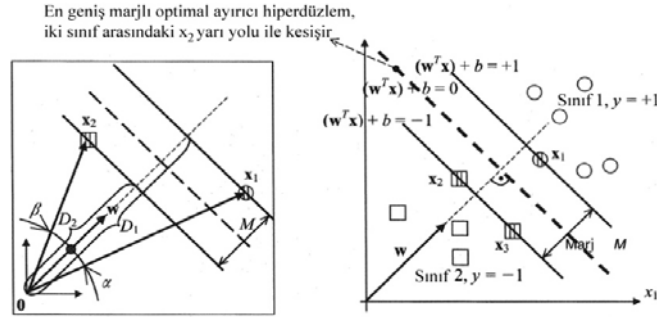
$$R(\alpha_n) \leq R_{amp}(\alpha_n) + \sqrt{\frac{h[(\ln(2l/h)+1) - \ln(\eta/4)]}{l}} \quad (3)$$

şeklinde ifade edilmektedir (Zheng ve Yang, 2004). $R(\alpha)$ sınırı, ilki $R_{amp}(\alpha)$ ampirik risk, ikincisi ise VC boyutu h (model kapasitesi ya da öğrenen makinenin karmaşıklığı olarak yorumlanabilir) ve eğitim örneği kütle sayısı l ile ilişkisi olan güven aralığı olmak üzere, iki bölümden oluşmaktadır (Chen vd., 2005). Eğitim veri sayısı arttıkça, yani $l \rightarrow \infty$ için (diğer parametreler sabit), gerçek risk $R(\alpha_n)$ ampirik risk $R_{amp}(\alpha_n)$ 'e çok yakındır. Diğer taraftan, olasılık (güven seviyesi/ $1-\eta$) 1'e yaklaştığında, genelleme sınırı fazla büyür.

$\min_{x_i \in \mathcal{X}} |w^T x_i + b| = 1$ eşitliğini sağlayan minimum marj normalize edilir. Böylece geometrik marj $1/\|w\|_2$ olmaktadır (Christianni ve Shawe-Taylor, 2000). Vektörler marjın ya üzerinde ya da dışında uzanmalıdır. Tam olarak marj üzerinde (destek hiperdüzlemler üzerinde) uzanan vektörlere destek vektörleri denir ve optimal ayırıcı hiperdüzlemin kurulumunda merkezi teşkil etmektedir.

DVM'lerin öğrenmesinde kernel yöntemleri, verileri girdi uzaylarına açık bir şekilde haritalamadan, bir nokta çarpımın daha yüksek boyutlu uzayda hesaplanmasına ve optimal bir ayırıcı hiperdüzlem oluşturulmasına izin verir. Kernel fonksiyonlarını kullanmak boyutsallık sorunundan kaçınılmasını sağlar. Bu durumda girdi verisi üzerindeki optimizasyon, bir kernel tabanlı dönüşümün kullanımını içermektedir (Ganapathiraju vd., 2004).

$$K(x_i, x_j) = \Phi(x_i) \bullet \Phi(x_j) \quad (4)$$



Şekil 1: En geniş marjlı optimal hiperdüzlem (Kecman, 2001:153)

Maksimum marjlı hiperdüzlemi bulmak bir optimizasyon probleminin çözümüne denktir. Ancak dual formülasyon orijinal formülasyona tercih edilir çünkü problemin iç çarpımlar cinsinden ifade edilmesine olanak sağlamaktadır (Martin, 2001).

$$\text{Min}L_d = -\sum_{i=1}^l \alpha_i + \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad (5)$$

Öyle ki

$$\sum_{i=1}^n y_i \alpha_i = 0 \text{ ve } 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l$$

$\alpha_i > 0$ olan tüm eğitim örnekleri destek vektörleri olarak adlandırılır. Destek vektörleri ya marj üzerinde [$0 < \alpha_i < C$ 'ye sahip ($y_i f(x_i) = 1$)] (sınırlandırılmamış destek vektörleri) ya da marj alanının içerisinde uzanırlar [$\alpha_i = C$ 'ye sahip ($y_i f(x_i) < 1$)] (sınırlandırılmış destek vektörleri). C , “uç değer” lere dayatılan cezayı belirten bağımsız bir parametre olarak maliyet fonksiyonunu $\|w\|^2 / 2$ kadar arttırmaktır. C parametresi kullanıcı tarafından sağlanmaktadır ve eğitim kümesi hatasını minimize etme ile marjı maksimize etme (model karmaşıklığı) arasındaki eşiği kontrol etmektedir.

DVM yaklaşımının temel görevinin kernellerin seçiminde yattığı bilinmektedir. Farklı kernel fonksiyonları seçmek farklı DVM'ler üretmektir. DVM literatüründe polinomial DVM'ler, RBF DVM'ler, iki katman sinir ağları DVM'ler ve daha pek çokları bulunmaktadır (Tan ve Wang, 2004).

1 VERİ SETİ VE ANALİZ

Bu çalışmada İMKB ulusal-100 endeksinin getirisinin yönünü tahmin etmek amacıyla ikili sınıflandırma yöntemi kullanılarak hem lojistik regresyon ve hem de destek vektör sınıflandırması yöntemleriyle modellenmiştir.

1.1 VERİ SETİ

Söz konusu modellerde İMKB ulusal-100 endeksinin getiri yönünü tahmin etmek amacıyla yukarıda atıfta bulunulan çalışmaların bulguları çerçevesinde hem önde gelen uluslararası endeksler hem de birtakım makroekonomik ve finansal göstergeler bağımsız değişken olarak kullanılmıştır. Çalışma Şubat 1997'den Aralık 2010'a kadar olan dönemi kapsayan aylık verileri içermektedir. Analizde, sekiz tane makroekonomik gösterge ile MSCI European Index (16 Avrupa ülkesinin hisse senedi piyasasının performansını gösteren piyasa değeri ağırlıklı bir endekstir.), MSCI Emerging Markets Index (21 yükselen piyasa ekonomisinin hisse senedi piyasa performansını gösteren piyasa değeri ağırlıklı bir endekstir.) ve S&P500 endeksleri olmak üzere toplamda 11 tane değişken kullanılmıştır.

Tüketici fiyatları endeksine dayalı enflasyon oranları, TCMB tarafından gösterge niteliğinde açıklanan TL/Dolar döviz satış kuru, reel efektif döviz kuru endeksi, TCMB döviz rezervleri, bankalararası para piyasası bir gecelik faiz oranları ile ABD Federal Funds Rate farkı, imalat sanayi kapasite kullanım oranı, İstanbul altın borsası endeksi ve son olarak da West Texas Intermediate spot petrol fiyatı kullanılmıştır. Faiz farkları haricindeki diğer tüm değişkenler değişim oranı olarak analize sokulmuştur. MSCI endeksleri Barra veri tabanından diğer veriler de Türkiye İstatistik Kurumu, Türkiye Cumhuriyet Merkez Bankası ve Federal Reserve System veri tabanlarından elde edilmiştir.

Çalışmada İMKB ulusal-100 endeksinin getiri yönünü tahmin etmek amacıyla ikili sınıflandırma yapılmıştır. Endeks getirisi bir önceki ay sonundaki değerine göre pozitif yönlü olduğunda “1”, negatif yönlü olduğunda ise “0” olarak kodlanmıştır. Toplam 167 aylık veri seti, modellerin kurulduğu 138 veriden oluşan eğitim kümesine ve modellerin geçerliliğinin test edildiği 29 veriden oluşan tahmin kümesine ayrılmıştır. Her iki küme de bağımlı değişkenin aldığı değerler açısından benzer dağılım göstermektedir. Analiz sonuçları her iki veri seti için de sunulmuştur.

1.2 LOJİSTİK REGRESYON ANALİZ SONUÇLARI

Lojistik regresyon sonuçlarına göre İMKB-100 endeksinin yönünün belirlenmesinde etkili olduğu düşünülen değişkenlerden MSCI Emerging Market Endeksi, tüketici fiyatları endeksine dayalı enflasyon oranları, TL/Dolar döviz satış kuru ve reel efektif döviz kuru endeksinin %5, MSCI Europe Endeksi ve S&P 500’ün %10 ve TCMB döviz rezervinin %25 anlamlılık seviyesinde modele alınmasının anlamlı olduğu ortaya konulmuştur. Analiz Sonuçları Tablo 1’de gösterilmektedir.

Tablo 1: Lojistik Regresyon Analizi Sonuçları

	β_i	S.H.	Wald	df	Anlamlılık
MSCI Europe Endeksi	14,806	8,310	3,175	1	,075
MSCI Emerging Market Endeksi	15,055	5,752	6,851	1	,009
S&P 500	-14,714	9,453	3,423	1	,120
TÜFE	27,378	14,011	3,818	1	,051
TL/DOLAR Döviz Kuru	-41,077	10,755	14,587	1	,000
Reel Efektif Döviz Kuru Endeksi	-18,029	8,089	4,967	1	,026
TCMB Döviz Rezervleri	4,940	4,069	1,474	1	,225
Sabit Terim	,168	,354	,226	1	,635

Modelin anlamlılığı Hosmer-Lemeshow testi ile sınanmıştır. Modelin anlamlılık değeri ($p=0,591$) $> 0,05$ olduğu için model anlamlı bulunmuştur.

H_0 : Tahmin denklemi anlamlıdır.

H_1 : Tahmin denklemi anlamlı değildir.

Tablo 2: Lojistik Regresyon Modelinin Sınıflandırma Başarısı

		Modelleme Kümesi			Tahmin Kümesi		
		İMKB-100		Doğru Sınıflandırma Oranı (%)	İMKB-100		Doğru Sınıflandırma Oranı (%)
		1	0		1	0	
İMKB-100	1	61	16	79,2	15	1	93,8
	0	18	43	70,5	3	10	76,9
Toplam		138		75,4	29		86,2

Lojistik Regresyon modelinin Tablo 2’de sunulan sınıflandırma başarısı incelendiğinde, İMKB-100 endeksi getirisinin yönünün doğru sınıflandırma oranının modelleme kümesinde %75,4 iken tahmin kümesinde %86,2’ye yükseldiği gözlenmiştir. İMKB-100 endeksinin negatif getiri yönünün ise modelleme kümesinde %70,5 oranında doğru sınıflandırılırken pozitif yönü %79,2 oranında doğru sınıflandırılmıştır. Bu oranlar tahmin kümesi için sırasıyla %76,9 ve % 93,8 olarak gerçekleşmiştir.

1.3 DESTEK VEKTÖR MAKİNELERİ ANALİZ SONUÇLARI

Destek Vektör Makineleri yöntemiyle yapılan uygulama analiz öncesi sürece göre iki kısımda ele alınmıştır. Birinci kısımda, elde edilen tüm değişkenler analize sokulmuştur. İkinci kısımda ise değişken seçimi yapılmış ve bunun sonucunda sadece modelde kalmasına karar verilen değişkenler analizde yer almıştır. Yapılan analizlerde, 2001 yılında Chih-Chung Chang ve Chih-Jen Lin tarafından geliştirilmiş olan ve literatürde en çok kullanılan destek vektör makinesi kütüphanesi olan LIBSVM kullanılmıştır.

Analiz öncesinde kullanılan tüm değişkenler, modelin veri kümesinde yer alan daha büyük sayısal aralıktaki değişkenlerin daha küçük sayısal aralıktaki değişkenlerin üzerinde baskınlık kurmasından kaçınmak amacıyla, [-1,1] aralığında ölçeklendirilmiştir.

Kernel tipi olarak Radyal Tabanlı Kernel (RBF) tercih edilmiştir. RBF parametreleri (C ve γ) optimize edilirken çapraz geçerlilik (cross-validation) yöntemi uygulanmıştır. Bu aynı

zamanda veriye aşırı uyumun olmamasını sağlamaktadır. Böylece eğitim kümesi (modelleme kümesi), k-1 adet alt eğitim kümesi, kalan kısmı ise test kümesi olacak şekilde belirlenen k adet alt kümeye ayrılmaktadır. Çapraz geçerlilik uygulanmadan önce veri her bir alt kümenin dağılımları benzer olacak şekilde rassal olarak parçalanmaktadır. Uygulamada incelenen modellerde k=10 olarak belirlenmiştir. Test edilen parametre kombinasyonları $C=2^{-3}, 2^{-2}, \dots, 2^6$ ve $\gamma=2^{-4}, 2^{-3}, \dots, 2^3$ kümelerinden oluşturulmuştur. Her iki analiz için de en iyi sonuç veren parametre kombinasyonundan edilen sonuçlar gösterilmiştir.

Tablo 3: Tüm Değişkenlerle Kurulan DVM Modelinin Sınıflandırma Başarısı

		Eğitim Kümesi			Tahmin Kümesi		
		İMKB-100		Doğru Sınıflandırma Oranı (%)	İMKB-100		Doğru Sınıflandırma Oranı (%)
İMKB-100		1	0		1	0	
	1	64	13	83,1	15	1	93,8
	0	23	38	62,9	5	8	61,5
Toplam		138		73,9	29		79,3

Tablo 4: Değişken Seçimi Sonucunda Kurulan DVM Modelinin Sınıflandırma Başarısı

		Eğitim Kümesi			Tahmin Kümesi		
		İMKB-100		Doğru Sınıflandırma Oranı (%)	İMKB-100		Doğru Sınıflandırma Oranı (%)
İMKB-100		1	0		1	0	
	1	64	13	83,1	16	0	100
	0	20	41	67,2	4	9	69,2
Toplam		138		76,1	29		86,2

Değişken seçimi yapılmadan uygulanan analizde veri kümesinde bahsedilen tüm değişkenler yer almaktadır. Analiz sonuçları Tablo 3'te verilmiştir. Değişken seçimi uygulanarak yapılan analiz için ise denetimli bir filtreleme yöntemi olan CfsSubsetEvaluation ve arama yöntemi olarak da genetik algoritma kullanılmıştır. CfsSubsetEvaluation her bir özelliğin birbirleriyle olan ilişki derecesine bağlı olarak bireysel tahmin kabiliyetlerini göz

önünde bulundurarak, bir değişken kümesi değerlemektedir. Genetik algoritma ile yapılan arama sonucunda, MSCI Emerging Markets Endeksi, MSCI European Endeksi, S&P 500 ve TL/Dolar döviz kuru değişkenleri modele alınmıştır. Analiz sonuçları Tablo 4’te verilmiştir.

Hem tüm değişkenlerle kurulan DVM modeli hem de değişken seçimi sonucunda analize alınan dört tane değişkenle kurulan DVM modeli tahmin kümesinde % 86.2 oranında verileri doğru sınıflandırmaktadır. Her iki analiz sonucunda da “1” ile kodlanan pozitif yöndeki endeks getirilerinin “0” ile kodlanan negatif yöndeki endeks getirilerine kıyasla çok daha iyi tahmin edilebildiği belirgin olarak göze çarpmaktadır. Yanı sıra, eğitim kümesine oranla tahmin kümesinin sonuçlarının da artış göstermesi model kurma aşamasında veriler eğitilirken (138 veri) uygulanacak kernel fonksiyonunun parametrelerinin belirlenmesinde kullanılan çapraz geçerlilik işleminin başarıyla uygulandığını ve eğitim verisine aşırı uyum (overfitting) sorununu ortadan kaldırdığını desteklemektedir.

SONUÇ

İMKB-100 endeksi getirisini tahmin etmek amacıyla çalışmada kullanılan yöntemlerin sonuçlarının birbirine çok yakın olduğu görülmektedir. Lojistik regresyon sonucunda modelleme kümesi ve tahmin kümesi için doğru sınıflandırma oranları sırasıyla %75,4 ve %86,2 olarak gerçekleşirken, DVM yönteminin sonuçları ise sırasıyla %76,1 ve %86,2’dir. Yöntemlerin tahmin güçleri incelendiğinde lojistik regresyonun endeksin negatif getiri yönünü daha iyi tahminlerken, DVM yöntemiyle kurulan her iki modelde de endeksin pozitif getiri yönünü daha iyi tahmin ettiği görülmektedir.

Makine öğrenmesi tekniklerinin istatistikî yöntemlerle karşılaştırıldığında daha az varsayım gerektirerek araştırmacıya esneklik sağlaması ve gerçek riske dayanan genelleme kabiliyetinin daha üstün olması, son dönemlerde makine öğrenmesi tekniklerinin tahmin modellerinde kullanılmasını yaygınlaştırmıştır. Literatürde hisse senedi endeksinin fiyat ya da getirisinin tahmin edilmesinde yapay zeka yöntemlerinin daha sık kullanıldığı görülmektedir. Bu çalışma ile DVM yönteminin yatırımcılar ve araştırmacılar tarafından bu konuda yapılan çalışmalarda alternatif bir yöntem olarak etkin bir şekilde kullanılabileceği çalışmanın sonuçlarıyla da desteklenmiştir.

KAYNAKÇA

- ALBAYRAK, Ali Sait (2006), *Uygulamalı Çok Değişkenli İstatistik Teknikler*, Asil Yayın Dağıtım Ltd.Şti., Ankara.
- ALİ, Shawkat ve Ajith ABRAHAM (2002), “An Empirical Comparison of Kernel Selection for Support Vector Machines”, *Proceedings of the 2nd International Conference on Hybrid Intelligent Systems, Soft Computing Systems: Design, Management and Application*, IOS Press, Amsterdam, The Netherlands.

- ALTAY, Erdiñç ve M. Hakan SATMAN (2005), “Stock Market Forecasting: Artificial Neural Network and Linear Regression Comparison in an Emerging Market”, *Journal of Financial Management and Analysis*, 18(2):18-33.
- ANG, A. ve G. BEKAERT (2001), “Stock Return Predictability: Is it There?”, *National Bureau of Economic Research Working Paper*, No. 8207.
- AVCI, Emin (2007), “Forecasting Daily and Sessional Returns of the ISE-100 Index with Neural Network Models”, *Doğuş Üniversitesi Dergisi*, 8 (2):128-142.
- BAHARUMSHAH, A.Z.; T. SARMIDI ve H.B. TAN (2003), “Dynamic Linkages of Asian Stock Markets: An Analysis of Pre-Liberalization and Post-Liberalization Eras”, *Journal of the Asia Pacific Economy*, 8:180-209.
- BAILEY, Warren ve Y. Peter CHUNG (1995), “Exchange Rate Fluctuations, Political Risk, and Stock Returns: Some Evidence from an Emerging Market”, *Journal of Financial and Quantitative Analysis*, 30(4):541-561.
- BALVERS, Ronald J.; Thomas F. CASIMANO ve Bill MACDONALD (1990), “Predicting Stock Returns in an Efficient Market”, *Journal of Finance*, 45(4):1109-1128.
- BEIRNE, John; Guglielmo Maria CAPORALE; Marianne SCHULZE-GHATTAS ve Nicola SPAGNOLO (2009), “Volatility Spillovers and Contagion from Mature to Emerging Stock Markets”, *Brunel University Department of Economics and Finance Working Paper Series*, no. 09-05.
- BOYACIOĞLU, Melek Acar ve Derya AVCI (2010), “An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the Prediction of Stock Market Return: The Case of the Istanbul Stock Exchange”, *Expert System with Applications*, (37):7908-7912.
- CAMPBELL, John Y. (1990), “Measuring the Persistence of Expected Returns”, *American Economic Review Papers and Proceedings*, ss.43-47.
- CAMPBELL, John Y. ve Robert J. SHILLER (1988), “Stock Prices, Earnings, and Expected Dividends”, *Journal of Finance*, 43(3):661-676.
- CAMPBELL, John Y. ve Robert J. SHILLER (1998), “Valuation Ratios and Long-Run Stock Market Outlook”, *Journal of Portfolio Management*, 24:11-26.
- CHAN, K.C.; J. KARCESKI ve J. LAKONISHOK (1998), “The Risk and Return from Factors”, *Journal of Financial and Quantitative Analysis*, 33:159-188.
- CHEN, Pai-Hsuen; Chih-Jen LIN ve Bernhard SCHÖLKOPF (2005), “A Tutorial on v-Support Vector Machines”, *Applied Stochastic Models in Business and Industry*, John Wiley&Sons, USA, 21(2):111-136.
- CHERKASSKY, Vladimir ve X. SHAO (2001), “Signal Estimation and Denoising Using VC-theory”, *Neural Networks*, Elsevier, 14(1):37-52.
- CHU, Hsing-Hui; Tai-Liang CHEN; Ching-Hsue CHENG ve Chen-Chi Huang (2009), “Fuzzy Dual-factor Time-series for Stock Index Forecasting”, *Expert Systems with Applications*, (36):165-171.

- CRISTIANNI, Nello ve John SHAWE-TAYLOR (2000) *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, UK.
- DİLER, Ali İhsan (2003), “İMKB Ulusal 100 Endeksinin Yönünün Yapay Sinir Ağları Hata Geriye Yayma Yöntemi ile Tahmin Edilmesi”, *İMKB Dergisi*, 7(25-26):65-81.
- DURHAM, J.B. (2001), “The Effects of Monetary Policy on Monthly and Quarterly Stock Market Returns: Cross-Country Evidence and Sensitivity Analyses”, *The Federal Reserve Board Finance and Economics Discussion Series Paper*, No.2001-42.
- DURUKAN, M. Banu (2009), “İstanbul Menkul Kıymetler Borsasında Makro Ekonomik Değişkenlerin Hisse Senedi Piyasasına Etkisi”, *İMKB Dergisi*, 3(11):18-47.
- EGELİ, Birgül; Meltem ÖZTURAN ve Bertan BADUR (2003), “Stock Market Prediction Using Artificial Neural Networks”, *Proceedings of the 3rd Hawaii International Conference on Business*, Hawaii, USA.
- ENKE, David ve Suraphan THAWORNWONG (2005), “The Use of Data Mining and Neural Networks for Forecasting Stock Market Returns”, *Expert Systems with Applications*, 29:927-940.
- FAMA, Eugene F. (1981), “Stock Returns, Real Activity, Inflation and Money”, *American Economic Review*, 71:545-565.
- FAMA, Eugene F. ve G.W. SCHWERT (1977), “Asset Returns and Inflation”, *Journal of Financial Economics*, 5:115-146.
- FAMA, Eugene F. ve Kenneth FRENCH (1988), “Dividend Yields and Expected Stock Returns”, *Journal of Financial Economics*, 22(1):3-25.
- FAMA, Eugene F. ve Kenneth FRENCH (1993), “Common Risk Factors in the Returns on Stocks and Bonds”, *Journal of Finance*, 33(1):3-56.
- GANAPATHIRAJU, Aravind; Jonathan E. HAMAKER ve Joseph PICONE (2004), “Applications of Support Vector Machines to Speech Recognition”, *IEEE Transactions on Signal Processing*, 52(8):2348-2355.
- GESKE, R. ve R. ROLL (1983), “The Fiscal and Monetary Linkages between Stock Returns and Inflation”, *Journal of Finance*, 38:1-34.
- GRIFFIN, John M. ve René M. STULZ (2001), “International Competition and Exchange Rate Shocks: A Cross-country industry analysis of stock returns”, *The Review of Financial Studies*, 14(1):215-241.
- JAFFE, Jeffrey; Donald B. KEIM ve Randolph WESTERFIELD (1989), “Earnings Yield, Market Values, and Stock Returns”, *Journal of Finance*, 44(1):135-148.
- JEON, B.N. ve B.S. JANG (2004), “The Linkage Between the US and Korean Stock Markets: The Case of NASDAQ, KOSDAQ, and the Semiconductor Stocks”, *Research Journal of International Business and Finance*, 16:319-340.

- KARA, Yakup; Melek Acar BOYACIOĞLU ve Ömer Kaan BAYKAN (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*, 38:5311-5319.
- KECMAN, Vojislav (2001), *Learning and Soft Computing: Support Vector Machines, Neural Networks and Fuzzy Logic Models*, MIT Press, Cambridge, Massachusetts, England.
- KEIM, Donald B. ve Robert F. Stambaugh (1986), “Predicting Returns in the Stock and Bond Markets”, *Journal of Financial Economics*, 17(2):357-390.
- KIM, Kyoung-jae (2003), “Financial Time Series Forecasting Using Support Vector Machines”, *Neurocomputing*, 55:307-319.
- KIM, Kyong-jae ve Ingoo HAN (2000), “Genetic Algorithms Approach to Feature Discretization in Artificial Neural Networks for the Prediction of Stock Price Index”, *Expert Systems with Applications*, 19:125-132.
- KOTHARI, S.P. ve Jay SHANKEN (1997), “Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis”, *Journal of Financial Economics*, 44(2):169-203.
- LETTAU, Martin ve Sydney LUDVIGSON (2001), “Consumption, Aggregate Wealth, and Expected Stock Returns”, *Journal of Finance*, 56(3):815-849.
- LI, Yuanguai; Weidong Zhang ve Chen Lin (2006), “Simplify Support Vector Machines by Iterative Learning”, *Neural Information Processing-Letters and Reviews*, 10(1):11-17.
- LIU, Y. Angela ve Ming-Shiun PAN (1997), “Mean and Volatility Spillover Effects in the U.S. and Pacific-Basin Stock Markets”, *Multinational Finance Journal*, 1(1):47-62.
- LU, Chi-Jie (2010), “Integrating Independent Component Analysis-based Denoising Scheme with Neural Network for Stock Price Prediction”, *Expert Systems with Applications*, 37:7056-7064.
- MARTIN, Shawn (2001), *Techniques in Support Vector Classification*, Dissertation, Colorado State University, Fort Collins, Colorado.
- NA, Sung Hoon ve So Young SOHN (2011), “Forecasting Changes in Korea Composite Stock Price Index (KOSPI) Using Association Rules”, *Expert Systems with Applications*, 38:9046-9049.
- NG, T.H. (2002), “Stock Market Linkages in Southeast Asia”, *Asian Economic Journal*, 16(4):353-377.
- MASIH, Abul M.M. ve Rumi MASIH (1999), “Are Asian Stock Market Fluctuations due mainly to Intra-regional Contagion Effects? Evidence Based on Asian Emerging Markets”, *Pacific-Basin Finance Journal*, 3-4:251-282.
- MASIH, Abul M.M. ve Rumi MASIH (2001), “Long and Short Term Dynamic Causal Transmission amongst International Stock Markets”, *Journal of International Money and Finance*, 20:563-587.

- ORHUNBİLGE, Neyran (2010), *Çok Değişkenli İstatistik Yöntemler*, İstanbul Üniversitesi, İşletme Fakültesi, İstanbul.
- PARK, Kwangwoo ve Ronald A. RATTI (2005), “Real Activity, Inflation, Stock Returns and Monetary Policy”, *The Financial Review*, 35(2):59-78.
- PEARCE, D.K. ve V.V. ROLEY (1983), “The Reaction of Stock Prices to Unanticipated Changes in Money: A Note”, *Journal of Finance*, 38:1323-1333.
- PESARAN, M. Hashem ve Allan TIMMERMANN (1995), “Predictability of Stock Returns: Robustness and Economic Significance”, *Journal of Finance*, 50(4):1201-1228.
- PHYLAKTIS, K. ve F. RAVAZZOLO (2005), “Stock Market Linkages in Emerging Markets: Implications for International Portfolio Diversification”, *Journal of International Financial Markets, Institutions and Money*, 15(2):91-106.
- SHARMA, Subhash (1996), *Applied Multivariate Techniques*, John Wiley & Sons, Inc., Canada.
- SIKLOS, P. ve B. KWOK, (1999), “Stock Returns and Inflation: A New Test of Competing Hypotheses”, *Applied Financial Economics*, 9:567-581.
- TAN, Ying ve Jun WANG (2004), “A Support Vector Machine with a Hybrid Kernel and Minimal Vapnik-Chervonenkis Dimension”, *IEEE Transactions on Knowledge and Data Engineering*, IEEE Educational Activities Department, 16(4):385-395.
- WANG, Yi-Fan Wang, Shihmin CHENG ve Mei-Hua HSU (2009), “Incorporating the Markov chain concept into fuzzy stochastic prediction of stock indexes”, *Applied Soft Computing*, (10):613-617.
- WORTHINGTON, Andrew ve Helen HIGGS (2004), “Transmission of Equity Returns and Volatility in Asian Developed and Emerging Markets: a Multivariate GARCH Analysis”, *International Journal of Finance & Economics*, 9(1):71-80.
- WU, C. ve Y. SU (1998), “Dynamic Relations among International Stock Markets”, *International Review of Economics and Finance*, 7:63-84.
- ZHANG, Xuegong (1999), “Using Class-Center Vectors to Build Support Vector Machines”, *Proceedings of the 1999 IEEE Signal Processing Society Workshop. Neural Networks for Signal Processing IX*, IEEE, August:3-11.
- ZHENG, En-Hui; Min YANG (2004), “Statistical Learning Theory”, *Proceedings of the Third International Conference on Machine Learning and Cybernetics*, Shanghai, IEEE, Vol.2, 26-29 August 2004, ss.3095-3099.