

Hisse Senedi Fiyat Tahmininde Genetik Algoritma ile Değişken Seçimi Feature Selection for Stock Price Predicting with Genetic Algorithms

Yrd. Doç. Dr. Mehmet Özçalıcı

Öz

Hisse senedi fiyatı yön tahmini, piyasaların sahip olduğu karmaşık yapı itibarıyla zor bir problemdir ve bu nedenle de araştırmacıların ilgisini çeken bir konudur. Hisse senedi fiyat tahmininde kullanılan yöntemlerden bir tanesi destek vektör makineleridir (DVM) ve literatürde başarılı bir şekilde kullanıldığı görülmektedir. DVM ile başarılı bir tahmin çalışması gerçekleştirmek için kullanılacak girdi değişkenlerinin dikkatli bir şekilde seçilmesi gerekmektedir. Genetik algoritma, DVM için değişken seçiminde kullanılan yöntemlerden bir tanesidir. Yapılan çalışmalar araştırıldığında genetik algoritma (GA) tabanlı DVM yönteminde farklı kernel fonksiyonlarının performansı ve farklı sayılardaki değişkenlerin tahmin performansının detaylı bir şekilde incelenmediği ortaya çıkmıştır. Bu açığı kapatmak için emtia tabanlı bir fonun tarihi fiyat bilgileri kullanılmış ve GA-DVM yöntemi ile değişken seçimi gerçekleştirilmiştir. En yüksek doğru tahmin oranı, 10 adet değişkenin seçildiği doğrusal kernel fonksiyonuna sahip DVM modelinde gerçekleştiği ortaya çıkmıştır. Aynı zamanda modellerin yarısından fazlasında al ve tut stratejisinden daha yüksek oranda bir getiri sağlamak mümkün olmuştur. Çalışmada elde edilen başarılı sonuçlar, önerilen yöntemin hisse senedi fiyat tahmininde karar destek sistemi olarak kullanılabileceğini göstermektedir.

Anahtar Kelimeler: Hisse Senedi Fiyat Tahmini, Teknik Analiz Göstergeleri, Genetik Algoritmalar, Destek Vektör Makineleri, Değişken Seçimi

Abstract

Stock price predicting attracts interests of researches due to the complex structure of stock price markets. One of the techniques that is used for stock price forecasting is Support Vector Machines (SVM) and it is successfully used in stock price predicting. In order to obtain successful predicting results, input selection must be handled carefully. Genetic Algorithms (GA) is one of the methods that is used for feature selection. Literature review revealed that the performance of different kernel function as well as the performance of number of features did not deeply investigated in GA based SVM models. In order to fill this gap, historical price information of an emtia-based fund is used and features are selected with genetic algorithms. The highest hit rate is obtained in linear kernel function model in which 10 features are extracted. Moreover higher financial returns are obtained in more than half of the prediction models. The successful results obtained in this study shows that proposed model can be used as a decision support system in stock price prediction.

Keywords: Stock Price Predicting, Technical Analysis Indices, Genetic Algorithms, Support Vector Machines, Feature Selection

Giriş

Hisse senedi fiyatlarının önceden tahmin edilmesi zordur. Bunun nedenlerinden biri hisse senedi piyasalarının karmaşık yapısıdır. Buna rağmen hisse senedi fiyatını önceden daha doğru tahmin etmeye yönelik çalışmalar devam etmektedir. Bu problem üzerindeki çalışmaların devam etmesinin nedenlerinden biri de problemin çözümünün kâr sağlamasıdır. Hisse senedi fiyatlarındaki eğilimin önceden doğru bir şekilde tahmin edilmesi, alım satım stratejilerini önemli düzeyde getiri sağlayabilecek şekilde şekillendirebilir, bu durum da daha yüksek kazançta neden olabilir.

Hisse senedi fiyatını tahmin etmek için istatistiksel yöntemler kullanılmıştır. Bunlar regresyon, lojistik regresyon (Özdemir vd. 2011), diskriminant analizi (Öz vd, 2011) ve ARCH, GARCH (Akar, 2007; Karabacak vd 2014) gibi istatistiksel modellerdir. Hisse senedi fiyat tahmininde geleneksel yöntemleri kullanan çalışmalar (Atsalakis ve Valavanis 2010) detaylı bir şekilde gözden geçirilmiştir. Bu modellerin varsayımlara dayanması, karmaşık olan hisse senedi piyasalarını modellemede yetersiz kalmalarına neden olmuştur.

Hisse senedi fiyatını tahmin etmek için aynı zamanda makine öğrenmesi (machine learning) veya esnek hesaplama (soft computing) teknikleri de kullanılmıştır. Bu yöntemler, ANN (Toraman, 2008; Kara vd, 2011), SVM (Abolhassani ve Yaghoobi, 2010; Ince ve Trafalis, 2007; Ni vd. 2011; Pai ve Lin, 2005; Yu vd. 2005; Kara vd. 2011; Tayyar ve Tekin, 2013; Kim, 2003) ve ANFIS (Atsalakis ve Valavanis, 2009; Atsalakis vd, 2011) gibi yöntemlerdir. Bu gibi yöntemler ile elde edilen tahmin sonuçlarının, istatistiksel yöntemlerle elde edilen tahmin sonuçları ile karşılaştırıldığında daha başarılı oldukları görülmektedir. Ayrıca esnek hesaplama yöntemlerini kullanmak suretiyle hisse senedi fiyat tahmini gerçekleştiren çalışmalar (Atsalakis ve Valavanis, 2009) detaylı olarak incelenmiştir.

1965 yılında, ünlü finans yazarı Fama French tarafından ismi belirlenen Etkin Piyasalar Hipotezi (EPH) 'ne göre hisse senedi fiyatları, bir önceki fiyat hareketlerinden bağımsız ve rassal bir şekilde meydana gelmektedir (Fama, 1965a). Bu hipoteze göre hisse senedi fiyatlarını öngörmek mümkün değildir. Hisse senedi piyasasında yeni bir bilgi ortaya çıktığında bu bilgi piyasadaki oyuncular tarafından anında asimile edilecektir (Fama, 1965b). Bu nedenle söz konusu

bilgiyi kullanmak suretiyle piyasada olağandışı getiri elde etmek mümkün olmayacaktır. Etkin piyasalar hipotezi ile ilgili yapılan çalışmaların tarihini 1900'lü yıllara kadar dayandırmak mümkündür (EPH tarihi hakkında detaylı bilgi (Dimson ve Mussavian, 1998) çalışmasında yer almaktadır). Yukarıda değinilen esnek hesaplama ve makine öğrenmesi yöntemlerini kullanan yazarlar piyasaların etkin olamayacağını ve hisse senedi fiyatlarının belirli bir düzeyde önceden tahmin etmenin mümkün olduğunu düşündürecek düzeyde başarılı sonuçlar elde etmişlerdir.

Hisse senedi fiyat tahmini ile ilgili yapılan çalışmalar incelendiğinde, hem hisse senedi fiyatlarının tahmin edildiği, hem de eğilimin tahmin edildiği görülmektedir. Genel olarak hisse senedindeki fiyat hareketlerinin, gerçek fiyat bilgilerinden daha kolay tahmin edileceği öngörülmektedir. Başka bir ifade ile hisse senedinin fiyatının yarın bugünkünden daha yüksek (veya düşük) olacağını öngörmek, hisse senedi fiyatının yarın örneğin 3,82 TL olacağını tahmin etmekten daha kolaydır. Her iki durumda kullanılması gereken tahmin yöntemleri birbirlerinden farklı özelliklere sahiptir.

Diğer taraftan kullanılan değişkenler açısından, hisse senetleri fiyat tahmin çalışmalarını iki grupta toplamak mümkündür. İlk grupta makroekonomik göstergeler tahmin çalışmasında girdi olarak kullanılmaktadır. Bu göstergeler başka ülkelerin borsaları, döviz kurları, işsizlik oranları, dış ticaret verileri, emtia fiyatları ve işletmelerin finansal tablolarından elde edilen bilgilerdir (Fama ve French, 1988; Tsai ve Hsiao, 2010; Özdemir vd, 2011; Öz vd. 2011; Lahmiri, 2011).

Hisse senedi fiyat tahmininde kullanılan diğer grup değişkenler ise teknik analizde kullanılan değişkenlerdir. Teknik analizi ise hisse senedinin geçmiş fiyatlarının (genellikle grafik şeklinde) kaydedilmesi ve bu kayda alınan geçmiş bilgilere bakmak suretiyle, gelecekte olması muhtemel kârlı eğilimlerin tahmin edilmesi ile uğraşan bir bilim olarak tanımlamak mümkündür (Edwards vd., 2007: s.4). Çok fazla sayıda teknik gösterge mevcuttur (Ni vd. 2011; Yu vd. 2005; Ince ve Trafalis, 2007; Kara vd. 2011; Abolhassani ve Yaghoobi, 2010). Teknik göstergelerin sayısının fazla olması hangi hisse senedi için hangi değişkenlerin fiyat tahmin çalışmasında kullanılması gerektiği problemini ortaya çıkarmaktadır.

Her hisse senedinin kendine ait özellikleri nedeniyle tahmin modelinde başarılı olacak değişkenlerin birbirlerinden farklı olması beklenmektedir. Bir hisse senedi fiyatının tahmininde başarılı bir şekilde kullanılabilen bir değişken, bir başka hisse senedi için kullanıldığında aynı başarıyı sağlamayabilir. Bu nedenle hisse senedi fiyat tahmininde evrensel bir değişken kümesinden bahsetmek doğru olmayacaktır. Üstelik A hisse senedinin fiyatının tahmin edilmesinde t zamanında başarılı olmuş bir değişken, bir başka zaman diliminde başarılı olmayabilir. Başka bir ifade ile hisse senetleri için en iyi tahmini gerçekleştirecek göstergelerin bir şekilde seçilmesi (Tsai ve Hsiao, 2010) gerekmektedir.

Çok fazla teknik göstergenin varlığı olası bütün değişken alt kümelerinin boyutunu da artırmaktadır. Belirli bir hacimdeki değişken kümesinden, belirli sayıda kaç adet değişken alt kümesinin seçileceği, kombinasyon formülü yardımıyla aşağıdaki şekilde hesaplanmaktadır;

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Formülde n ifadesi değişkenlerin seçileceği değişken kümesi, r ifadesi ise kaç adet değişkenin seçileceğini temsil etmektedir. Örneğin $n = 169$ ve $r = 5$ olduğunda

$$\begin{aligned} \binom{169}{5} &= \frac{169!}{5! \cdot (169-5)!} = \frac{169 \cdot 168 \cdot 167 \cdot 166 \cdot 165}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \\ &= 1.0822e + 09 \end{aligned}$$

Yukarıdaki hesaplama göre 169 adet değişken kümesinden 5 adet değişken seçmek istediğimizde 1 milyardan daha fazla olası bileşen karşımıza çıkmaktadır. Bu değişken alt kümelerinden hangisinin hisse senedi fiyat tahmininde başarılı olacağını belirlemek bir problem olarak karşımıza çıkmaktadır. Her bir olası bileşeni tek tek denemek ve optimal olanı bu sonuçlar içerisinde seçmek, günümüz hesaplama teknolojisini kullanıldığında, rasyonel bir zaman diliminde gerçekleşmeyecektir. Bu nedenle daha etkin bir optimizasyon yöntemine ihtiyaç duyulmaktadır.

Değişken seçimi (veya özellik seçimi) için herkes tarafından kabul edilen bir yöntem söz konusu değildir. Meta-sezgisel yöntemler hisse senedi fiyat tahmininde değişken seçimi için kullanılabilir. Genetik algoritmalar en eski ve en popüler meta-sezgisel optimi-

zasyon yöntemidir ve popülasyon tabanlı bir arama stratejisi uygulanmaktadır.

Meta-sezgisel yöntemlerin hisse senedi fiyat tahmin problemlerinde bütünleşik olarak kullanılması yeni bir konu değildir. Aboulhassani ve Yaghoobi (2010) particle swarm optimizasyon yöntemi ile destek vektör makineleri (DVM) için değişken seçimini gerçekleştirmiştir. Önerdikleri melez yöntemin DVM'nin başarısını artırdığını raporlamışlardır. Ni vd (2011) ise hisse senedi eğilim tahmininde, fractal özellik seçimi ile destek vektör makinelerini bir arada kullanmışlardır. Yu vd 2005 ise DVM yöntemi ile fiyat tahminin gerçekleştirmek için genetik algoritmaları kullanmışlardır. Çalışmalarında sadece 11 adet değişken seçimi üzerinde durmuşlardır. Zhao vd (2011) ise değişken seçimi ve parametre optimizasyonunu aynı anda gerçekleştirecek bir kromozom tasarlamışlardır. Ancak önerdikleri yöntemi hisse senedi fiyat tahmini problemi için değil, literatürde yer alan veri madenciliği veri setleri için uygulamışlardır ve genetik algoritmanın DVM'nin performansını artırdığını raporlamışlardır. Tsai ve Hsiao (2010) ise birden fazla değişken seçim yönteminin hisse senedi fiyat tahminindeki performansını incelemişlerdir. Tahmin yöntemi için ise yapay sinir ağlarını kullanmışlardır.

Genetik algoritmanın DVM yöntemi için değişken seçiminde kullanıldığı ve yukarıda değinilen çalışmalarda eksik kalan bazı noktalar olduğu görülmektedir. Örneğin farklı kernel fonksiyonlarındaki GA-DVM modelinin performansı incelenmemiştir. Ayrıca farklı sayılarda değişken seçildiğinde yine söz konusu modelin performansının ne olacağı incelenmemiştir. Üstelik söz konusu çalışmalarda GA-DVM nin finansal performansından bahsedilmemektedir. GA-DVM yöntemi sadece istatistiksel başarı oranı ile incelenmektedir. Modelin öngördüğü şekilde hareket eden bir yatırımcının getiri durumu, başka bir ifade ile yatırımın finansal performansı incelenmemiştir. Bu çalışmada, genetik algoritma tabanlı destek vektör makineleri melez yönteminin, farklı kernel fonksiyonları ve farklı değişken sayılarındaki finansal ve istatistiksel performansını bir arada incelemek suretiyle literatürdeki bu açığın kapanması hedeflenmiştir.

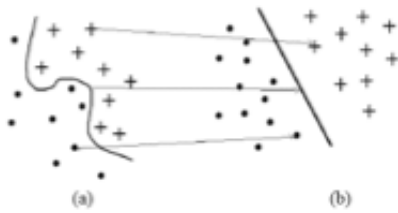
Bu çalışmada geçmiş fiyat bilgilerinden yola çıkmak suretiyle hisse senedi fiyatında bir gün sonra meydana gelecek eğilimi (trendi) mümkün olduğunca gerçeğe yakın bir şekilde tahmin etmek hedeflenmek-

tedir. Bu amaçla bilgi işlemsel zeka yöntemlerinden iki tanesi bir arada kullanılacaktır. Önerilen yöntem genetik algoritma tabanlı melez bir yöntemdir. Belirli bir dönemde destek vektör makineleri yöntemi ile başarılı tahmini gerçekleştirilmesi beklenen en iyi değişken altkümüsi genetik algoritmalar yardımıyla belirlenmiştir.

Destek Vektör Makineleri

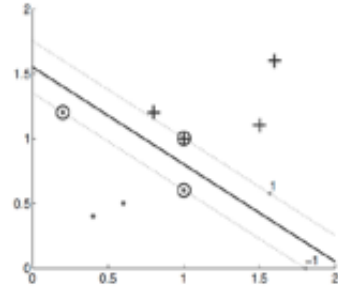
Destek vektör makineleri makine öğrenmesi yöntemlerinden bir tanesidir ve yapısal risk minimizasyonu prensibine dayanmaktadır. Destek Vektör Makineleri (DVM), farklı sınıf etiketlerine sahip gözlemleri çok boyutlu bir uzayda sınıflandırmaya yarayan bir yöntemdir (Nisbet vd. 2009). İlk kez Vapnik (1995) tarafından literatüre tanıtılan DVM; sınıflandırmanın yanı sıra regresyon analizi ve uçdeğerlerin belirlenmesinde de kullanılabilir.

DVM ilk önce eğitim setini kullanmak suretiyle bir model oluşturmaktadır. Bu model sınıflandırılması en zor örnekler üzerine inşa edilmektedir. Bu zor örnekler destek vektörleri olarak adlandırılmaktadır. DVM yöntemi doğrusal bir şekilde ayrılabilen ve ayrılabilen veri setlerinde kullanılabilir. Doğrusal bir şekilde ayrılabilen veri setinin sınıflandırılmasında, girdi seti çok boyutlu başka bir uzaya dönüştürülmektedir. Bu çok boyutlu özellik uzayında doğrusal sınıflandırma gerçekleştirilebilir. Az sayıda gözlemlerle eğitim yapıldığında bile DVM yeterli doğrulukta sonuçlar ortaya çıkarabilmektedir (Venogopal vd, 2009).



Şekil 1. Girdi Düzlemi ve Hiper Düzlem

DVM' in temel fikri girdi setini çok boyutlu bir uzayda yeniden inşa etmektir. Şekil 1'in (a) kısmında iki boyutlu uzayda girdi seti grafik halinde gösterilmektedir. Söz konusu uzayda verilerin doğrusal bir şekilde sınıflandırılması mümkün değildir. Şeklin (b) kısmında ise, veri setinin doğrusal bir şekilde sınıflandırılmasına olanak sağlayan çok boyutlu başka bir uzay (hiper düzlem) görülmektedir. Bu yeni çok boyutlu uzayda veri setinin doğrusal sınıflandırılması mümkün hale gelmektedir.



Şekil 2. Optimal Ayrıştırıcı Düzlem, Marj, Destek Vektörler

Şekil 2'de gösterildiği gibi, destek vektör makineleri grupları doğrusal bir şekilde ayırmaya olanak sağlayan marjini maksimum kılmaya çalışmaktadır. Yukarıdaki şekilde nokta ve artı ile işaretlenen iki grup gözlem söz konusudur. Şekildeki kalın çizgi optimal ayrıştırıcı düzlemi (optimal separating hyperplane (OHS)) temsil etmektedir. Kesik çizgiler ise her iki taraftaki marjı temsil etmektedir. DVM'de amaç bu marjı maksimum hale getirmektir. Bu çizgilerin üzerindeki gözlemler ise destek vektörleri olarak adlandırılmaktadır. Kesik çizgiler arasındaki uzaklık ise OSH marjini olarak adlandırılmaktadır. Bu marj (uzaklık) ne kadar yüksekse sınıflandırma o kadar iyi yapılmış demektir.

Aşağıdaki gibi ifade edilen bir girdi vektörü söz konusu ise;

$$x_i \in R^d (i = 1, 2, \dots, N)$$

Formüldeki N örnek sayısını temsil etmektedir. Her bir gözlemin hangi sınıfa ait olduğu bilindiği varsayalım. Bu durumda DVM nin esas hedefi varolan örnekleri kullanmak suretiyle ikili bir sınıflandırma kurmaktır. DVM ile birlikte $x_i \in R^d$ ile ifade edilen gridi vektörü $\varphi(x_i) \in H$ ile ifade edilen yüksek boyutlu özellik uzayına dönüştürülmektedir. Söz konusu dönüştürme işlemi kernel (çekirdek) fonksiyonu adı verilen fonksiyonlar yardımıyla yapılmaktadır. Girdi vektörlerinin sınıflandıran fonksiyon aşağıdaki gibidir.

$$f(x) = \sum_i \alpha_i k(s_i, x) + b$$

Formüldeki s_i destek vektörleridir, α_i ağırlıklardır, b eğim ve k kernel fonksiyonunu ifade etmektedir.

Farklı kernel fonksiyonları mevcuttur (Tablo 1). Bunlar; doğrusal, karesel, polinomal, radyal bazlı (RBF) ve multilayer perceptron (MLP) kernel fonksiyonu. Hangi fonksiyonun ne zaman kullanılması gerektiğine ilişkin literatürde bir bilgi mevcut değildir.

Tablo 1. Kernel Fonksiyonları, Denklemleri ve Parametreleri

	Denklem	Parametre
Doğrusal	$G(x, y) = x \cdot y$	-
Polinomal	$G(x, y) = (1 + x \cdot y)^p$	p
RBF	$G(x, y) = \exp\left(-\frac{\ x - y\ ^2}{2\sigma^2}\right)$	γ
MLP	$G(x, y) = \tanh(w \cdot x \cdot y + w_0)$	Ölçek [1 -1]

Genetik Algoritmalar (GA)

Doğal evrimden esinlenmek suretiyle geliştirilmiş sezgisel optimizasyon yöntemi olan popülasyon tabanlı genetik algoritmalar, John Holland'ın 1975 yılındaki "Adaptation in Natural and Artificial Systems" adlı kitabında ilk kez tanıtılmıştır (Holland, 1992).

Genetik algoritma bir grup bireyden oluşan popülasyon tabanlı optimizasyon yöntemidir. Bu popülasyon, nesli için ile temsil edilmektedir. GA terminolojisinde birey aday çözümü temsil etmektedir. Aday çözümlerin hangisinin optimal sonuç verdiğini değerlendirmek için uygunluk fonksiyonu adı verilen fonksiyon kullanılmaktadır. Söz konusu fonksiyon problemin özünü oluşturmaktadır. Uygunluk fonksi-

yonu aday çözümlerin değerini çıktı olarak sunmaktadır. Her birey (aday çözüm) uygunluk fonksiyonundan aldığı değerle değerlendirir. Minimum problemlerinde en düşük uygunluk fonksiyonunu döndüren aday çözüm, optimal olarak düşünülebilir. Optimal çözümün bulunabilmesi için ana kütledeki bireyler üzerinde değişiklikler yapılmalıdır. Bu değişiklikler genetik algoritmada tek birey üzerinde değişiklik yapan mutasyon ve iki farklı bireyin farklı parçalarını bir araya getiren çaprazlama ile isimlendirilen işlemler ile gerçekleştirilmektedir. Birkaç nesil (yenileme) sonucunda, algoritmanın optimal veya optimale yakın en iyi bireyi oluşturması beklenmektedir. Uygulamalarda farklılıklar olsa da genetik algoritmaların temel yapısı Şekil 3'deki gibidir.

Prosedür: Genetik Algoritmalar

Başla

$t = 0$;

Başlangıç $A(t)$ anakütlesinin oluşturulması;

$A(t)$ deki her bireyin uygunluk değerinin hesaplanması;

while bitirme şartı sağlanmadığı müddetçe **do**

başla

$Y(t)$ 'yi elde etmek için $A(t)$ 'nin çaprazlanması

$Y(t)$ 'deki bireylerin uygunluk değerlerinin hesaplanması;

$A(t)$ ve $Y(t)$ anakütlelerinden $A(t + 1)$ 'in oluşturulması

$t = t + 1$;

bitir

bitir

Şekil 3. Genetik Algoritmaların Temel Prosedürü

Genetik Algoritmalarla işlemler genotip üzerinde yapılmaktadır. Genotipler sıfır ve bir değerlerinden oluşmaktadır. Fenotip ise özelliğin gerçek değerini temsil etmektedir. Her bir genin yapısı fenotip parametrelerinde kodlanmıştır. Fenotip parametreleri aynı zamanda genotip ve fenotip arasındaki değişimi sağlayacak bilgileri içermektedir. Genotip ve fenotip dönüşümü, modelin parametrelerinin genetik algoritmanın işlem yapabildiğini sağlayacak şekilde dönüştürülmesi ve oluşturulan yeni bireylerin uygunluk değerlerinin belirlenmesi için gereklidir (Svanandam ve Deepa, 2008: 39).

Değişken Seçimi

Değişken seçimi (Feature Selection) etkin bir tahmin performansı için veri setinde bulunan faydasız değişkenlerin elenmesini sağlamaktadır. Eğer başlangıç veri setinde yer alan ilgisiz ve fazla değişkenler listeden çıkartılırsa, öğrenme algoritmasının çalışma süresi de azalacaktır (Tsai ve Hsiao, 2010).

Değişken seçiminin belli başlı üç adet amacı vardır: değişkenlerin tahmin performansını güçlendirmek, daha hızlı ve etkili tahminleyiciler bulmak ve veri setini etkileyen esas süreçleri anlamayı sağlamak (Guyon ve Elisseeff, 2003: 1157).

Hisse senetleri geçmiş fiyat bilgilerini kullanmak suretiyle çok fazla sayıda teknik göstergenin hesaplanması mümkündür. Çok fazla sayıda değişken aynı zamanda yönetilmesi zor bir veri seti anlamına gelmektedir. Bu nedenle, mümkün olan en az kayıpla veri setinin daha yönetilebilir boyuta indirgenmesi gerekmektedir. Başlangıç havuzunda bir araya getirilen veriler gereksiz olabilir veya yanıltıcı unsurlar barındırabilir. Bu nedenle analizlerin ileri safhalarına geçmeden önce bu tür unsurların, veri setinden çıkarılması gerekmektedir. Bu tür temizleme işlemi sadece analizin daha hızlı gerçekleşmesini sağlamaz aynı zamanda elde edilen sonuçların daha anlaşılır olmasını sağlar ve sonuçların kalitesini arttırabilir (Jensen ve Shen, 2008: 62).

Değişken seçim algoritmaları, değerlendirme prosedürüne göre filtre yaklaşımı veya wrapper yaklaşımı olmak üzere iki kategoriye ayrılabilir. Eğer seçim süreci bir öğrenme algoritmasından tamamen bağımsızsa filtre yaklaşımı söz konusudur. Bu sınıfta ileri ve geri seçim gibi yöntemler bulunmaktadır. Diğer tarafta değerlendirme süreci bir hedefe bağlıysa,

bu durumda wrapper yaklaşımdan bahsedilmektedir. Bu yöntemde alt kümenin uygunluğu bir öğrenme algoritmasına bağlıdır (Jensen ve Shen, 2008: 69). Bu yöntemlere örnek olarak meta-sezgisel algoritmalar örnek olarak verilmektedir. Bu çalışmada değişken seçimi için meta-sezgisel algoritmaların en çok bilineni olan genetik algoritmalar kullanılacaktır.

Yöntem

Veri Seti

Çalışmada yatırımcıların altın piyasasında yatırım yapmasına olanak sağlayan altın bazlı alım satım gerçekleştiren ve hisseleri Amerika, Singapur, Japonya ve Hong Kong borsalarında da işlem gören SPDR Gold Shares isimli fonun New York Stock Exchange borsasında işlem gördüğü fiyatlar veri seti olarak kullanılmıştır. Söz konusu fonun seçilmesinin esas nedeni birden fazla borsada işlem görmesi nedeniyle işlem hacminin oldukça fazla olması, aynı zamanda altın bazlı işlemler yapıyor olmasıdır. Dünya çapında işlem gören bir hisse senedinin fiyatında kitlelerin etkisi daha fazla olacaktır. Söz konusu fon altın tabanlı olduğu için fonun değeri büyük oranda bir emtia olan altın tarafından belirlenmektedir.

Türkiye'de faaliyet gösteren Borsa İstanbul'da işlem gören hisse senetleri ile ilgili detaylı bir çalışma Özçalıcı (2015) e aittir. Yazar çalışmasında BIST30'da işlem gören 30 adet hisse senedinin fiyatını yapay sinir ağı tabanlı melez bir yöntemle tahmin etmeye çalışmış ve test veri setinde önceden doğru tahmin oranı en fazla %60 olmuştur. Daha geniş bir piyasada işlem gören bir hisse senedi fiyat bilgisini kullanmak tahminin başarı oranında artış sağlayabilecektir.

Çalışmada kullanılan eğitim veri seti 04 Ekim 2014 ile 17 Nisan 2015 tarihleri arasında kapsamaktadır. Söz konusu dönemde 2400 işlem günü bulunmaktadır. Veri seti, Yahoo.com sitesinden MATLAB yazılımının datafeed araç kutusu kullanmak suretiyle elde edilmiştir. Veri setinde her gün için söz konusu hisse senedinin açılış, kapanış, en yüksek, en düşük, kapanış fiyatı ve işlem hacmi bilgileri bulunmaktadır.

Test veri seti ise 20 Nisan 2015 tarihi ile 14 Temmuz 2015 tarihleri arasındaki 60 işlem gününü içermektedir.

Değişkenler

Çalışmada mümkün olduğu kadar fazla sayıda değişken ve hisse senedi fiyat tahmininde literatürde önceden de kullanılan teknik göstergeler bir araya getirilmeye çalışılmıştır. Değişken setinde yer alan ilk 134 adet değişken geçmiş fiyat bilgileri ve hareketli ortalama çentiklerinden oluşmaktadır. Geri kalan 35 adet değişken ise teknik göstergelerden oluşmaktadır. Toplamda 169 adet değişken bir araya getirilmiştir. Değişkenlerin isimleri ve sıralamasına Tablo 2 'de yer verilmiştir.

Her bir teknik göstergeye ilişkin formüller ve detaylı açıklamalar (Özçalıcı, 2015) kaynağında mevcuttur. Burada çalışmanın hacmini artırmaktan kaçınmak için formüllere yeniden yer verilmeyecektir.

Veri seti daha iyi performans elde edebilmek adına 0 ile 1 arasında aşağıdaki formülü kullanmak suretiyle standartlaştırılmıştır:

$$X_{yeni} = \frac{X_{eski} - X_{min}}{X_{max} - X_{min}}$$

Formülde, X_{yeni} standartlaştırıldıktan sonraki değeri, X_{eski} standartlaştırma işleminden önceki değeri, X_{min} ve X_{max} değişkenin sırasıyla minimum ve maksimum değerlerini ifade etmektedir.

Tablo 2. Kullanılan Değişkenlerin Sıralaması

Sıra No.	Değişken İsmi		
1	(t - 1) gününe ilişkin kapanış fiyatı	86	10 günlük üssel hareketli ortalama
2	(t - 2) gününe ilişkin kapanış fiyatı	87	11 günlük üssel hareketli ortalama
3	(t - 3) gününe ilişkin kapanış fiyatı	88	12 günlük üssel hareketli ortalama
4	(t - 4) gününe ilişkin kapanış fiyatı	89	13 günlük üssel hareketli ortalama
5	(t - 5) gününe ilişkin kapanış fiyatı	90	14 günlük üssel hareketli ortalama
6	(t - 6) gününe ilişkin kapanış fiyatı	91	15 günlük üssel hareketli ortalama
7	(t - 7) gününe ilişkin kapanış fiyatı	92	16 günlük üssel hareketli ortalama
8	(t - 8) gününe ilişkin kapanış fiyatı	93	17 günlük üssel hareketli ortalama
9	(t - 9) gününe ilişkin kapanış fiyatı	94	18 günlük üssel hareketli ortalama
10	(t - 10) gününe ilişkin kapanış fiyatı	95	19 günlük üssel hareketli ortalama
11	(t - 1) gününe ilişkin en yüksek fiyat	96	20 günlük üssel hareketli ortalama
12	(t - 2) gününe ilişkin en yüksek fiyat	97	21 günlük üssel hareketli ortalama
13	(t - 3) gününe ilişkin en yüksek fiyat	98	22 günlük üssel hareketli ortalama
14	(t - 4) gününe ilişkin en yüksek fiyat	99	23 günlük üssel hareketli ortalama
15	(t - 5) gününe ilişkin en yüksek fiyat	100	24 günlük üssel hareketli ortalama
16	(t - 6) gününe ilişkin en yüksek fiyat	101	25 günlük üssel hareketli ortalama
17	(t - 7) gününe ilişkin en yüksek fiyat	102	26 günlük üssel hareketli ortalama
18	(t - 8) gününe ilişkin en yüksek fiyat	103	27 günlük üssel hareketli ortalama
19	(t - 9) gününe ilişkin en yüksek fiyat	104	28 günlük üssel hareketli ortalama
20	(t - 10) gününe ilişkin en yüksek fiyat	105	29 günlük üssel hareketli ortalama
21	(t - 1) gününe ilişkin en düşük fiyat	106	30 günlük üssel hareketli ortalama
22	(t - 2) gününe ilişkin en düşük fiyat	107	3 günlük üçgensel ortalama
23	(t - 3) gününe ilişkin en düşük fiyat	108	4 günlük üçgensel ortalama
24	(t - 4) gününe ilişkin en düşük fiyat	109	5 günlük üçgensel ortalama
25	(t - 5) gününe ilişkin en düşük fiyat	110	6 günlük üçgensel ortalama
26	(t - 6) gününe ilişkin en düşük fiyat	111	7 günlük üçgensel ortalama
27	(t - 7) gününe ilişkin en düşük fiyat	112	8 günlük üçgensel ortalama
28	(t - 8) gününe ilişkin en düşük fiyat	113	9 günlük üçgensel ortalama
29	(t - 9) gününe ilişkin en düşük fiyat	114	10 günlük üçgensel ortalama
30	(t - 10) gününe ilişkin en düşük fiyat	115	11 günlük üçgensel ortalama
31	(t - 1) gününe ilişkin açılış fiyatı	116	12 günlük üçgensel ortalama
32	(t - 2) gününe ilişkin açılış fiyatı	117	13 günlük üçgensel ortalama
33	(t - 3) gününe ilişkin açılış fiyatı	118	14 günlük üçgensel ortalama
34	(t - 4) gününe ilişkin açılış fiyatı	119	15 günlük üçgensel ortalama
35	(t - 5) gününe ilişkin açılış fiyatı	120	16 günlük üçgensel ortalama
36	(t - 6) gününe ilişkin açılış fiyatı	121	17 günlük üçgensel ortalama
37	(t - 7) gününe ilişkin açılış fiyatı	122	18 günlük üçgensel ortalama

Tablo 2. Kullanılan Değişkenlerin Sıralaması (Devamı)

38	(t - 8) gününe ilişkin açılış fiyatı	123	19 günlük üçgenel ortalama
39	(t - 9) gününe ilişkin açılış fiyatı	124	20 günlük üçgenel ortalama
40	(t - 10) gününe ilişkin açılış fiyatı	125	21 günlük üçgenel ortalama
41	(t - 1) gününe ilişkin işlem hacmi	126	22 günlük üçgenel ortalama
42	(t - 2) gününe ilişkin işlem hacmi	127	23 günlük üçgenel ortalama
43	(t - 3) gününe ilişkin işlem hacmi	128	24 günlük üçgenel ortalama
44	(t - 4) gününe ilişkin işlem hacmi	129	25 günlük üçgenel ortalama
45	(t - 5) gününe ilişkin işlem hacmi	130	26 günlük üçgenel ortalama
46	(t - 6) gününe ilişkin işlem hacmi	131	27 günlük üçgenel ortalama
47	(t - 7) gününe ilişkin işlem hacmi	132	28 günlük üçgenel ortalama
48	(t - 8) gününe ilişkin işlem hacmi	133	29 günlük üçgenel ortalama
49	(t - 9) gününe ilişkin işlem hacmi	134	30 günlük üçgenel ortalama
50	(t - 10) gününe ilişkin işlem hacmi	135	Toplama / Dağıtım Salınımı
51	3 günlük basit hareketli ortalama	136	Hareketli Ortalamaların Birleşmesi Ayrılması
52	4 günlük basit hareketli ortalama	137	136'nın 9 günlük hareketli ortalaması
53	5 günlük basit hareketli ortalama	138	Açılış Fiyatı İvme
54	6 günlük basit hareketli ortalama	139	En Yüksek Fiyat İvme
55	7 günlük basit hareketli ortalama	140	En Düşük Fiyat İvme
56	8 günlük basit hareketli ortalama	141	Kapanış Fiyatı İvme
57	9 günlük basit hareketli ortalama	142	Açılış Fiyatı Momentum
58	10 günlük basit hareketli ortalama	143	En Yüksek Fiyat Momentum
59	11 günlük basit hareketli ortalama	144	En Düşük Fiyat Momentum
60	12 günlük basit hareketli ortalama	145	Kapanış Fiyatı Momentum
61	13 günlük basit hareketli ortalama	146	Chaikin Değişkenlik
62	14 günlük basit hareketli ortalama	147	Stokastik %K Değeri
63	15 günlük basit hareketli ortalama	148	Stokastik %D Değeri
64	16 günlük basit hareketli ortalama	149	Yavaşlatılmış Stokastik %K Değeri
65	17 günlük basit hareketli ortalama	150	Yavaşlatılmış Stokastik %D Değeri
66	18 günlük basit hareketli ortalama	151	William %R Değeri
67	19 günlük basit hareketli ortalama	152	Görelî Güç İndeksi
68	20 günlük basit hareketli ortalama	153	Bollinger Bandı - Orta Bant
69	21 günlük basit hareketli ortalama	154	Bollinger Bandı - Üst Bant
70	22 günlük basit hareketli ortalama	155	Bollinger Bandı - Alt Bant
71	23 günlük basit hareketli ortalama	156	En Yüksek Yüksek Değer
72	24 günlük basit hareketli ortalama	157	En Düşük Düşük Değer
73	25 günlük basit hareketli ortalama	158	Medyan Fiyat
74	26 günlük basit hareketli ortalama	159	Fiyat Değişim Oranı
75	27 günlük basit hareketli ortalama	160	Tipik Fiyat
76	28 günlük basit hareketli ortalama	161	Ağırlıklı Kapanış Fiyatı
77	29 günlük basit hareketli ortalama	162	William Toplama / Dağıtım Göstergesi
78	30 günlük basit hareketli ortalama	163	Chaikin Salınım
79	3 günlük üssel hareketli ortalama	164	Negatif Hacim İndeksi
80	4 günlük üssel hareketli ortalama	165	Pozitif Hacim İndeksi
81	5 günlük üssel hareketli ortalama	166	Toplama / Dağıtım Çizgisi
82	6 günlük üssel hareketli ortalama	167	Denge İşlem Hacmi
83	7 günlük üssel hareketli ortalama	168	Fiyat-Hacim Trendi
84	8 günlük üssel hareketli ortalama	169	Hacim Değişim Oranı
85	9 günlük üssel hareketli ortalama		

Çıktı Değerleri

Çalışmada çıktı değişkeni (H), fiyat hareketleri olarak adlandırılmaktadır ve 1 ve 0 değerlerinden oluşmaktadır;

$$H = \begin{cases} 1, & \text{fiyat yükselişi, } F_{t+1} > F_t \\ 0, & \text{fiyat düşüşü, } F_{t+1} \leq F_t \end{cases}$$

ifadede F_{t+1} , $t+1$ gününe ilişkin kapanış fiyatını, F_t ise t gününe ilişkin kapanış fiyatını temsil etmektedir. H değerinin 1 olması fiyatın bir sonraki gün yükseleceğinin tahmin edildiğini (veya yükseldiğini) ve sıfır çıkması fiyatın bir sonraki gün bugünkünden daha düşük ortaya çıkacağını tahmin edildiğini (veya ortaya çıktığını) göstermektedir.

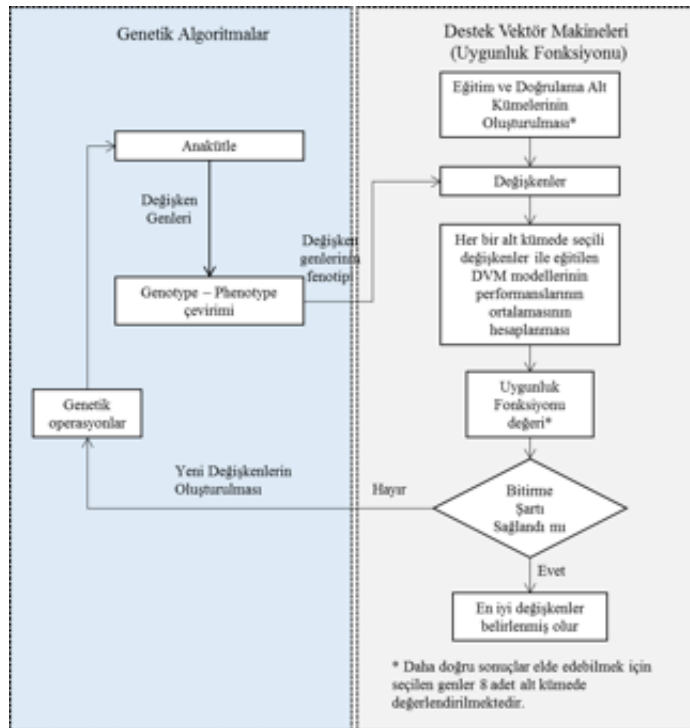
Çalışmanın Modeli

Çalışmada ana tahmin yöntemi destek vektör makineleridir. Bu modeli kullanmak suretiyle en iyi tahmini gerçekleştirecek değişken bileşenleri genetik algoritma ile seçilmektedir. Genetik algoritma ile optimal değişkenlerin seçim süreci Şekil 4'de görselleştirilmiştir.

Yapay sinir ağları gibi makine öğrenmesi modellerinde aşırı uyum (overfitting) durumu söz konusudur. Aşırı uyum, özellikle dikkate alınması gereken bir problem olarak karşımıza çıkmaktadır. Aşırı uyum, bir makine öğrenmesi modelinin, genelleştirme yeteneğini kaybetmesi ve kötü sonuçlar oluşturmasıdır

(Han ve Jiang, 2014). Tipik bir aşırı uyum problemi, tahmin modelinin parametrelerini eğitim setine aşırı uyum sağlamaktadır. Fakat model daha önce karşılaşmadığı bir veri setinde uygulandığında çok kötü performans sergilemektedir.

Genetik algoritmada adayların uygunluk değerini belirleyen uygunluk fonksiyonunda ilk olarak veri seti alt kümelere ayrılmaktadır. Tek bir alt kümede değerlendirme yapmak hatalı veya tesadüfi sonuçlar elde edilmesine neden olabilir. Aşırı uyum problemi ortadan kaldırmak için, veri seti 8 adet sıralı alt kümeye ayrılmaktadır. Seçilen değişkenler sıralı alt kümelere değerlendirilme tabi tutulmaktadır.



Şekil 4. Çalışmanın Modeli

Çalışmada genetik algoritmalar destek vektör makineleri ile en iyi tahmin sonucunu verecek değişkenleri seçmek için kullanılmaktadır. Önerilen melez yöntemin şematik gösterimi Şekil 4'deki gibidir. Çalışmada, farklı sayılardaki değişkenlerin performansını incelemek amacıyla, sırasıyla 2, 3, 5 ve 10 adet

değişken seçilmesi durumunda tahmin modelinin test veri seti üzerindeki performansı hesaplanmıştır. Değişken sayıları rastgele seçilmiştir. En fazla 10 adet değişken seçilmesinin nedeni ise mümkün olduğunca az değişken seçilmesi ve böylelikle veri setindeki karmaşıklığın minimum düzeyde tutulmasıdır.

Genetik algoritmaların uygunluk fonksiyonu ise eğitim setini ilk önce 8 adet sıralı alt kümeye ayırmaktır. Bu alt kümeler de kendi içinde eğitim ve test veri setinden oluşmaktadır. Son 60 adet fiyat test veri seti olarak adlandırırken, ilk 240 adet fiyat eğitim veri seti olarak kullanılmaktadır. Bir değişken alt kümesi belirlendiğinde, sırasıyla her bir alt kümenin eğitim veri seti kullanmak ve ilgili kernel fonksiyonu ve parametreler kullanmak suretiyle model eğitilmektedir. Eğitilen model ilgili alt kümenin test setinde değerlendirmeye tabi tutulmaktadır. Her bir alt kümenin test veri setinde gösterdiği performans, başka bir ifade ile hatalı tahmin oranı yüzdelere oranı alınmaktadır. Bu ortalama değeri uygunluk fonksiyonunun çıktısını oluşturmaktadır.

Parametreler

Daha önce de belirtildiği gibi destek vektör makinelelerinde farklı türdeki kernel fonksiyonun kullanılması mümkündür. Kernel fonksiyonların aynı zamanda kendi içlerinde de parametreleri mevcuttur. Örneğin polinomal kernel fonksiyonunda polinom derecesi veya radyal bazlı fonksiyonda sigma değeri gibi. Çalışmada bir araya getirilebilen kernel fonksiyonları

tek tek tahmin amacıyla kullanılmıştır. Bunun yanında çalışmanın kapsamını makul düzeyde tutabilmek açısından kernel fonksiyonların parametreleri MATLAB yazılımındaki varsayılan parametreler olarak kullanılmıştır. Genetik algoritma ve destek vektör makineleri için parametreler Tablo 3'de yer almaktadır.

Çalışmada kullanılan genetik algoritma için de çeşitli parametreler söz konusudur. Nesil sayısı kaç adet yeni ana kütle yaratıldıktan sonra genetik algoritmanın çalışmayı durduracağını belirlemektedir. Çaprazlama fonksiyonu *crossoverintermediate* olarak belirlenmiştir. Bu çaprazlama yönteminde yeni bireylere, ebeveynlerin sahip olduğu değerlerin ağırlıklı ortalaması atanmaktadır. Çalışmada tamsayı kısıtlar mevcut olduğu için, ağırlıklı değeri 1 olarak belirlenmiştir. Mutasyon fonksiyonu ise *mutationadaptfeasible* olarak belirlenmiştir. Bu mutasyon fonksiyonunda uygun sınırlar dahilinde mutasyon işlemi gerçekleştirilmektedir. Elit sayısı ise mutasyon veya çaprazlamaya maruz kalmadan bir önceki popülasyondan kaç tane en başarılı bireyin bir sonraki nesle aktarılacağını belirlemektedir.

Tablo 3. GA ve DVM Parametreleri

Genetik Algoritma		Destek Vektör Makineleri	
Nesil sayısı	100	Hiper düzlemin ayrıştırılmasında kullanılan yöntem	En küçük kareler yöntemi
Çaprazlama Fonksiyonu	crossoverintermediate	Maximum yenileme sayısı	15000
Çaprazlama ile belirlenecek birey oranı	0,8	MLP ölçek parametresi	[1 -1]
Mutasyon Fonksiyonu	mutationadaptfeasible	Polinom derecesi	3 ve 5
Popülasyon	*	RBF sigma	1
Elit Sayısı	$0.05 \times \min(\max(10 \times nvars, 40), 100)$ **		
* 2 ve 3 adet değişken seçiminde popülasyon 40, 5 adet değişken seçiminde 50 ve 10 adet değişken seçiminde 100 adet bireyden oluşmaktadır.			
** nvars değişken sayısı (2,3,5,10)			

Analiz Sonuçları

Tahmin edilen fiyatlar ile gerçek fiyat hareketlerinin ne derecede doğru olduğunu ortaya çıkarmak için hata matrisinden yararlanılmaktadır (Tablo 4). Fiyat bir sonraki gün bugünden yüksek veya düşük olabilir

ve bir sonraki güne ilişkin fiyat yükselişi veya azalışı tahmin edilebilir. Örneğin bir sonraki gün fiyat yükselişi tahmin edilirken, gerçekte de fiyat yükselişinin olduğu durum TP (True positive) olarak adlandırılmaktadır.

Tablo 4. Hisse Senedi Fiyat Eğilimi Tahmini için Hata Matrisi

Tahmin	Gerçek	
	Artış	Azalış
Artış	Fiyatın Yükselmesi durumu ve fiyat yükselmesinin tahmin edilmesi (TP – True positive)	Fiyat düşerken fiyat artışının tahmin edilmesi (FP – False positive)
Azalış	Fiyat düşerken fiyat artışının tahmin edilmesi (FN – False negative)	Fiyat düşerken fiyat azalışının tahmin edilmesi (TN – True Negative)

Doğru tahmin oranı (DTO) aşağıdaki formül yardımıyla hesaplanmıştır.

$$DTO = \frac{TP + TN}{TP + TN + FP + FN}$$

DTO değeri doğru tahmin oranı ifadesini temsil etmektedir. Bu ifadenin bir (1) çıkması bütün fiyat hareketlerinin önceden doğru bir şekilde tahmin edildiğini gösterirken, bu değerın sıfır olması hiçbir fiyat hareketinin önceden doğru bir şekilde tahmin edilemediğini göstermektedir.

Çalışmada elde edilen bütün sonuçlar Tablo 5’de bir arada yer almaktadır. Genetik algoritma ile eğitim süreci boyunca gerçekleşen hata oranı “GA süreci Hata oranı” isimli sütunda yer almaktadır. Söz konusu değerler, pratik tahmin çalışması için hiçbir önem arz etmemektedir. Çünkü söz konusu veri seti üzerinde destek vektör makinelerinin eğitimi gerçekleştirilmektedir. Bu sütunda erişilen en yüksek performans değeri 59,58% (Polinom(3)-10) olmuştur. Modelin daha önce karşılaşmadığı bir veri seti üzerindeki performansı ise modelin gerçek performansını oluşturmaktadır. “Test süreci Doğru Tahmin Oranı” adı verilen sütunda eğitim sürecinde daha önce hiç karşılaşılmamış bir veri seti üzerinde tahmin gerçekleştirilmektedir. Bu değerlerin pratik önemi söz konusudur çünkü bu değerler örneklem dışı performansı temsil etmektedir. Örneklem dışı erişilebilen en yüksek doğ-

ru tahmin oranı ise 61,67% (Doğrusal-10) olmuştur.

Test süreci getiri oranı isimli sütunda ise modellerin finansal performansı yer almaktadır. Finansal performans hesaplanırken, ilgili parametre ve değişkenlerle oluşturulan modelin test süreci için ürettiği yön tahminleri kullanılmıştır. Modelin fiyat artışı öngördüğü bir gün (t) alım gerçekleştirilmiş ve bir sonraki gün (t+1) satış gerçekleştirilmiştir. Aşağıdaki formül yardımıyla getiri oranları hesaplanmıştır.

$$g_t = \frac{f_{t+1} - f_t}{f_t}$$

formülde, g getiri oranını, f_t t gününe ilişkin standartlaştırılmamış (gerçek) kapanış fiyatını temsil etmektedir. 60 günlük test süreci boyunca bütün alım satımlar neticesi gerçekleşen kâr yüzdeleri toplanmıştır. Çalışmada kolaylık sağlaması açısından alım satımlarda, vergi ve komisyon giderleri göz ardı edilmiştir.

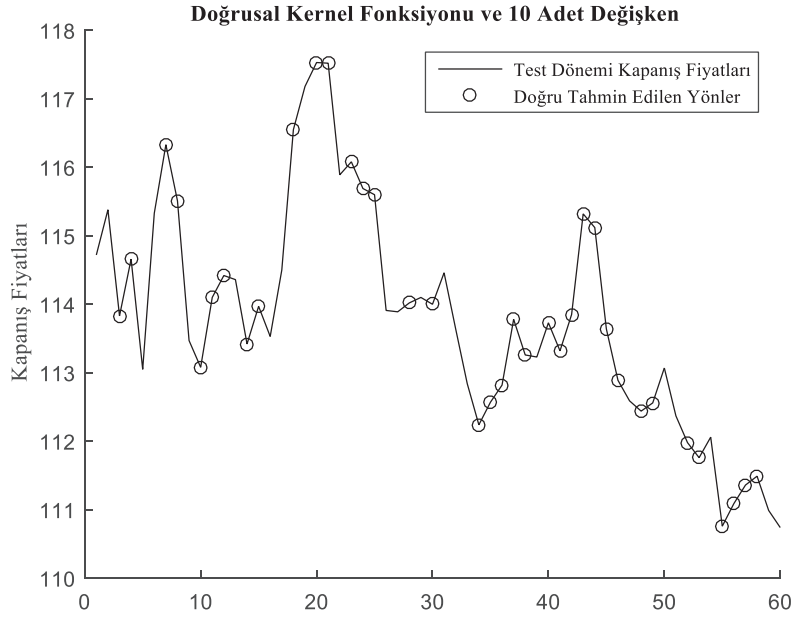
Çalışmada ayrıca al ve tut stratejisinin getiri oranı da hesaplanmıştır. Pasif bir strateji olan al ve tut stratejisinde yatırımcı hisse senedini satın almakta, belirli bir süre elde tuttuktan sonra satmaktadır. Bu alım satımdan elde edilen getiri oranı yine yukarıdaki formül yardımıyla hesaplanmaktadır. Çalışmada altmış günlük test sürecinin ilk fiyatından hisse senedi satın alındığında ve altmışıncı günkü kapanış fiyatı ile satış yapıldığında elde edilen getiri oranı (-0,0359) olmaktadır. Başka bir ifade ile pasif stratejiyi uygulayan bir yatırımcı %3,59 zarar etmektedir.

Tablo 5. İstatistiksel ve Finansal Tahmin Sonuçları

	Değişken Sayısı	GA Süreci Hata Oranı	Test Süreci Doğru Tahmin Oranı	Test Süreci Getiri oranı	Değişken İndeksleri
Doğrusal	2	0,5479	0,5833	0,0136	[150 151]
	3	0,5417	0,5333	-0,0029	[37 42 124]
	5	0,5625	0,5000	0,0022	[10 11 37 51 126]
	10	0,5687	0,6167	0,0097	[27 38 43 55 79 103 130 137 143 155]
	Bütün Değişkenler	-	0,5667	-0,0246	Hepsi
Karesel	2	0,5437	0,5500	0	[77 83]
	3	0,5542	0,5500	0	[64 72 85]
	5	0,5479	0,6000	0,0361	[17 80 101 134 147]
	10	0,5604	0,5667	-0,0129	[19 22 25 39 145 149 152 156 159 167]
	Bütün Değişkenler	-	0,5333	0,0088	Hepsi
Polinom (3)	2	0,5604	0,5500	0	[40 167]
	3	0,5521	0,5500	0	[62 106 167]
	5	0,5812	0,5333	0	[39 67 98 152 156]
	10	0,5958	0,4500	-0,045	[12 39 60 76 110 125 127 136 141 156]
	Bütün Değişkenler	-	0,5833	-0,0191	Hepsi
Polinom (5)	2	0,5687	0,5500	-0,0062	[8 146]
	3	0,5708	0,5333	-0,0039	[34 50 69]
	5	0,5771	0,5667	0,0085	[17 41 116 155 168]
	10	0,5896	0,4833	-0,0625	[39 47 54 60 66 81 97 116 145 155]
	Bütün Değişkenler	-	0,4833	-0,0114	Hepsi
MLP	2	0,5583	0,5000	0,0111	[50 167]
	3	0,5833	0,4167	-0,0673	[90 108 145]
	5	0,5729	0,5167	0,0199	[21 33 34 93 168]
	10	0,5854	0,4667	-0,0056	[28 38 51 54 56 87 100 110 122 152]
	Bütün Değişkenler	-	0,4667	-0,0253	Hepsi
Radyal Bazlı	2	0,5646	0,5667	0,0088	[47 165]
	3	0,575	0,5500	0	[23 134 167]
	5	0,5917	0,5500	0	[30 104 107 125 167]
	10	0,5917	0,5500	0	[39 81 86 91 98 104 105 108 125 167]
	Bütün Değişkenler	-	0,4833	0,0128	Hepsi

Analiz sonuçlarını görsel bir şekilde inceleyebilmek için Şekil 5'de en doğru tahmini gerçekleştiren modelin tahmin sonuçlarına yer verilmiştir. Grafikteki çizgi tarihi kapanış fiyatlarını temsil etmektedir. İçi boş daireler ise modelin doğru bir şekilde öngörebildiği eğilimleri temsil etmektedir. Dairenin yer aldığı

gündeki eğilim (bir sonraki gündeki fiyat artışı veya azalışı) tahmin modeli tarafından doğru bir şekilde öngörülebilmektedir. Şekil, en yüksek doğru tahminin gerçekleştiği, doğrusal kernel fonksiyonu ve 10 adet değişkeni içeren model için çizilmiştir.



Şekil 5. Test Süreci Fiyat Grafiği ve Doğrusal Kernel Fonksiyonu - 10 Adet Değişken Modelinin Doğru Tahmin Ettiği Eğilim



Şekil 6. En Sık Seçilen Değişkenler

Şekil 6'de ise, çalışmada önerilen model tarafından en sık seçilen değişkenler yer almaktadır. Şekle göre en sık seçilen değişken 7 sefer ile denge işlem hacmi olmuştur. Aynı şekilde t-9 gününe ilişkin açılış fiyatı 5 model tarafından optimal değişken olarak seçilmiştir.

Sonuç ve Tartışma

Çalışmada destek vektör makineleri için genetik algoritmalar yardımıyla optimal değişken alt küme seçim işlemi gerçekleştirilmiştir. DVM modeli için GA ile değişken seçimi literatürde daha önce çalışılan bir konudur. Bunun yanında DVM ile tahmin çalışmasında bazı parametrelere karar verilmesi gerekmektedir. Literatürde farklı kernel fonksiyonlarının ve farklı sayıda değişken seçiminin gerçekleştirildiği modellerin performansının incelenmediği belirlenmiştir. Literatürdeki bu açığı kapatmak için farklı kernel fonksiyonları ve farklı sayılarda değişkenleri kullanmak suretiyle 30 adet tahmin model oluşturulmuştur. Çalışma sonunda doğrusal kernel fonksiyonuna sahip ve 10 adet değişkenin seçildiği modelin örneklem dışı performansı %61,67 olarak hesaplanmıştır.

Çalışmada önerilen modellerin sonuçları, herhangi bir değişken seçim yönteminin uygulanmadığı ve 169 adet değişkenin hepsinin kullanıldığı alternatif model ile karşılaştırılmıştır. Bu karşılaştırmanın sonucunda, değişkenlerin seçildiği modellerin, fiyat tahmininde daha başarılı oldukları sonucu ortaya çıkmaktadır. Başka bir ifade ile bütün değişkenlerin bir arada kullanılması, başarı oranını düşürmektedir. Bunun olası nedeni birbirini tekrar eden ve aynı bilgiyi taşıyan (yüksek korelasyon değerine sahip) değişkenlerin veri setinde var olması olabilir. Buradan hareketle genetik algoritmaların, gereksiz olarak adlandırılacak değişkenleri elemekte başarılı olduğunu söylemek mümkündür.

Çalışmanın en önemli sonuçlarından bir tanesi veri sayısının azaltılmasıdır. 169 adet veri seti içinden başarılı tahmin sonuçlarını gerçekleştiren daha az sayıda değişkenin (en fazla 10) seçimi gerçekleştirilmiştir. Başka bir ifade ile daha az sayıda değişken ile aynı performans (veya daha iyi performans) sağlanmıştır.

Çalışmada ortaya çıkan en önemli sonuçlardan bir diğeri de önerilen modeli kullanmak suretiyle, gerileyen bir piyasada dahi pozitif getiri elde edilebilme-

sidir. Pasif al ve tut stratejisini uygulayan bir yatırımcı %3,59 zarar etmiştir. Karesel kernel fonksiyonu olan ve 5 adet değişkenin seçildiği modelde ise aynı dönemde %3,61 getiri elde edilebilmiştir. Fiyatların düşüyor olmasına rağmen, toplamda çalıştırılan 30 adet modelden 18 tanesinde pozitif bir getiri oranı sağlanmış veya olası bir zarar engellenmiştir. Bu durumda önerilen modelin, finansal açıdan başarılı olduğunu söylemek mümkündür.

60 günlük süreçte alım satımlar sonucu oluşan getiri oranlarına bakıldığında, en yüksek getiri oranının %3,49 ile doğrusal kernel fonksiyona sahip iki adet değişkenin seçildiği modelde gerçekleştirildiği görülmektedir. Çalışmada basit bir getiri oranı kullanılmıştır. Daha karmaşık ve riske karşı yatırımın değerini koruyabilecek alım satım algoritmalarının kullanılması getiri oranını artırabilir. Aynı şekilde daha sonraki çalışmalarda kernel fonksiyonun parametrelerinin tahmin performansındaki etkisi araştırılabilir.

Sonuç olarak genetik algoritma ile destek vektör makineleri için değişken seçiminin başarılı bir şekilde gerçekleştiğini söylemek mümkündür. Söz konusu modelin en önemli avantajı, değişken seçiminin bilgisayar tabanlı yapılması, başka bir ifade ile değişkenlerin tarafsız bir şekilde seçilmesinin gerçekleşmesidir. Değişken seçiminin uzmanlar tarafından yapılması da söz konusudur, fakat uzmanların kararları kişisel deneyimlerinden etkileneceğinden seçimin yanlı olma durumu söz konusu olacaktır. Aynı zamanda uzman bilgisine erişmek bazen pahalı bazen de hiçbir şekilde mümkün olmayabilir. Çalışmadaki gibi önerilen bir değişken seçim yönteminin kullanılması uzman bilgisine olan ihtiyacı ortadan kaldırmaktadır. Bu açıdan önerilen yöntemin, hisse senedi piyasalarında işlem yapan kişi ve kurumlar için etkin bir karar destek sistemi olacağını söylemek mümkündür.

Kaynakça

- Abolhassani, A.M.T. ve Yaghoobi, M. (2010). Stock price forecasting using PSOSVM. *3rd International Conference on Advanced Computer Theory and Engineering* 3:V3-352-V3-356.
- Akar, C. (2007). Volatilite modellerinin öngörü performansları: ARCH, GARCH ve SWARCH karşılaştırılması, *İşletme Fakültesi Dergisi*, 8(2):201-217.

- Atsalakis, G.S. ve Valavanis, K.P. (2010). Surveying Stock Market Forecasting Techniques – Part I: Conventional Methods, *Journal of Computation Optimization in Economics and Finance*, 2 (2).
- Atsalakis, G.S., Dimitrakakis, E.M. ve Zopounidis, C.D. (2011). Elliott Wave Theory and neuro-fuzzy systems in stock market prediction: The WASP System, *Expert Systems with Applications*, 38(8):9196-9206.
- Atsalakis, G.S. ve Valavanis, K.P. (2009). Surveying stock market forecasting techniques – Part II : Soft computing methods, *Expert Systems with Applications*, (36):5932-5941.
- Dimson, E. ve Mussavian, M. (1998). A brief history of market efficiency, *European Financial Management*, 4(1):91-193.
- Edwards, R.D., Magee, J. ve Bassetti, W.H.C. (2007). *Technical Analysis of Stock Trends*, 9. Bası, Taylor & Francis, New York.
- Fama, F. (1965a). Random walks in stock market prices, *Financial Analyst Journal*, 51(1):55-59.
- Fama, F. (1965b). The Behaviour of stock-market prices, *The Journal of Business*, 38(1):34-105.
- Fama, F. ve French, K. (1988). Dividend yields and expected stock returns, *Journal of Financial Economics*, (22):3-25.
- Guyon, I. ve Elisseeff, A. (2003). An introduction to variable and feature selection, *Journal of Machine Learning Research*, (3):1157-1182.
- Han, H. ve Jiang, X. (2014). Overcome support vector machine diagnosis overfitting, *Cancer Informatics*, 13(S1):145-158.
- Holland, J.H., (1992). *Adaptation in Natural and Artificial Systems*, THE MIT Press, London.
- Jensen, R. ve Shen, Q., (2008). *Computational Intelligence and Feature Selection Rough and Fuzzy Approaches*, IEEE Press, New Jersey.
- Ince, H. ve Trafalis, T.B. (2007). Kernel principal component analysis and support vector machines for stock price prediction, *IIE Transactions*, 39:629-637.
- 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), ss. 5311-5319.
- Karabacak, M., Meçik, O. ve Genç, E. (2014). Koşullu değişen varyans modelleri ile BİST 100 endeks getirisi ve altın getirisi volatilitésinin tahmini. *Uluslararası Alanya İşletme Fakültesi Dergisi*. 6(1):79-90.
- Kim, K. (2003). Financial time series forecasting using support vector machines, *Neurocomputing*. 55:307-319.
- Lahmire, S. (2011). A comparison of PNN and SVM for Stock Market Trend Prediction using economic and technical information. *International Journal of Computer Applications*. 29(3):24-30.
- Ni, L.P, Ni, Z.W ve Gao, Y.Z. (2011). Stock trend prediction based on fractal feature selection and support vector machine. *Expert Systems with Applications*. 38(5):5569-5576.
- Nisbet, R., Elder, J., Miner, G., (2009). *Handbook of Statistical Analysis and Data Mining Applications*. Elsevier Inc, Academic Press.
- Öz. B., Ayriçay, Y. ve Kalkan, G. (2011). Finansal oranlarla hisse senedi getirilerinin tahmini: IMKB 30 endeksi hisse senetleri üzerine diskriminant analizi ile bir uygulama. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*. 11(3):51-64.
- Özçalıcı, M. (2015). *Hisse Senedi Fiyat Tahminlerinde Bilgi İşlemsel Zeka Yöntemleri: Uzman Bir Sistem Aracılığıyla BİST Uygulaması*, Kahramanmaraş Sütçü İmam Üniversitesi/Sosyal Bilimler Enstitüsü, (Yayınlanmamış Doktora Tezi). Kahramanmaraş.

- 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), ss. 45-59.
- Pai, P.F. ve Lin, C.S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, 33:497-505.
- Svanandam, S.N. ve Deepa, S.N., (2008). *Introduction to Genetic Algorithms*, Springer, New York.
- Tayyar, N. ve Tekin, S. (2013). İMKB-100 endeksinin destek vektör makineleri ile günlük, haftalık ve aylık veriler kullanarak tahmin edilmesi. *AİBÜ Sosyal Bilimler Enstitüsü Dergisi*, 13(13). 189-217.
- Toraman, C., (2008). Demir-Çelik Sektöründe Yapay Sinir ağları ile Hisse Senedi Fiyat Tahmini: Erdemir A.Ş. ve Kardemir A.Ş. Üzerine Bir Tahmin, *Muhasebe ve Finansman Dergisi*, 39, ss. 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*, 50(1):258-269.
- Vapnik, V. N., (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA.
- Venogopal, K. R., Srinivasa, K. G., Patnaik, L. M., (2009). *Soft Computing for Data Mining Applications*. Springer-Verlag.
- Yu, L., Wang, S. ve Lai, K.K. (2005). Mining stock market tendency using GA-based support vector machines. *Internet and Network Economics Lecture Notes in Computer Sciences*, (3828):336-345.
- Zhao, M., Fu, C., Ji, L., Tang, K. ve Zhou. M. (2011). Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert Systems with Applications*. 38(5):5197-5204.