

HİSSE SENETLERİNİN DOĞRU TAHMİN ORANLARI İLE KÜMELENDİRİLMESİ

Mehmet ÖZÇALICI *

Özet

Hisse senedi fiyat tahmini farklı alanlardaki araştırmacıların dikkatini çeken, birçok avantajı bulunan popüler bir konudur. Literatürde bir gün sonraki hisse senedi fiyatını veya bir gün sonraki fiyat hareketini tahmin etmeye yönelik çalışmalar gerçekleştirilmektedir. Bununla birlikte literatürde hangi özelliklere sahip hisse senetlerinin daha yüksek doğru performans ile tahmin edileceği yönünde bir çalışma mevcut değildir. Bu çalışmada yüksek doğrulukla tahmin edilen senetlerin özelliklerini belirlemek amacıyla kümeleme analizi kullanılmıştır. Çalışmada BIST50 Endeksinde listelenen 50 adet hisse senedinin Ocak 2010 ile Kasım 2015 tarihleri arasındaki fiyat ve hacim bilgilerini kullanmak suretiyle 196 adet teknik gösterge hesaplanmıştır. Çalışmada tahmin yöntemi olarak hisse sendi fiyat tahmininde yeni yeni kullanılmaya başlanan Aşırı Öğrenme Makineleri (AÖM) kullanılmıştır. Söz konusu yöntem Yapay Sinir Ağlarına göre oldukça hızlı sonuç vermektedir. AÖM yöntemi için parametre optimizasyonu ve değişken seçimi genetik algoritma ile gerçekleştirilmiştir. Her bir hisse senedi için fiyat tahmini gerçekleştirildikten sonra, elde edilen tahmin oranları, hisse senetlerinin risk ve getiri değerleri ile birlikte k-ortalamlar yöntemi ile kümelere ayrılmıştır. Sonuçta daha yüksek risk ve daha düşük getiriye sahip senetlerin diğer gruptaki senetlerden ortalama olarak daha yüksek doğru oranı ile tahmin edildiği belirlenmiştir.

Anahtar kelimeler: Genetik Algoritmalar, Aşırı Öğrenme Makineleri, k-ortalamlar yöntemi, hisse senedi fiyat tahmini

JEL Sınıflaması: C45, N20, C53

CLUSTERING STOCKS WITH HIT RATE FORECASTING PERFORMANCE

Abstract

Forecasting stock prices is a popular subject that receive attention of researchers from different fields and has many advantages. In literature, forecasting the price itself or the price movement form one day ahead is researched. However, there is no research available that tries to determine the characteristics of

* Yrd. Doç. Dr., Kilis 7 Aralık Üniversitesi İİBF, Uluslararası Ticaret ve Lojistik Bölümü, mozcatici@kilis.edu.tr

stocks which are estimated with a high hit rate performance measurement. This study utilizes clustering analysis to determine the characteristics of stocks which are suitable for high performance forecasting. In this study 196 technical indicators are calculated using the price and volume data of stocks listed in BIST50 Index from January 2010 to November 2015. The forecasting technique used in this study is a novel technique called Extreme Learning Machine (ELM) which is much faster than its counterpart Artificial Neural Networks. The parameter optimization for ELM and feature selection is optimized using Genetic Algorithms. The forecasting performance of each stock is used in k-means technique along with risk and return data. Results indicate that stocks with higher risk and lower return has higher forecasting performance when compared with other cluster which has lower forecasting performance.

Keywords: Genetic Algorithm, Extreme Learning Machine, k-means clustering, stock price forecasting

JEL Classification: C45, N20, C53

I. Giriş

Hisse senedi fiyat tahmini literatürde sıklıkla çalışılan konulardan biridir. Özellikle son zamanlarda geliştirilen analizler ile birlikte bu alanın popülaritesi daha da arttırmıştır. Literatürde yapılan çalışmalar incelendiğinde çalışmaların genellikle doğru tahmin oranını yükseltmeye odaklandığı görülmektedir. Henüz hangi özelliklere sahip hisse senetlerinin daha yüksek doğru oranı ile tahmin edileceği üzerine bir çalışma gerçekleştirilmemiştir.

Söz konusu problem önemlidir çünkü daha yüksek doğru tahmini ile tahmin edilecek senetlerin bilinmesi yatırımcıların doğru yatırım kararları almasına yardım edecek ve bu durum da ülkenin ekonomik kaynaklarının boşa sarf edilmesini engelleyecektir. Bunun yanı sıra ortaya çıkacak bu bilgi sonraki akademik çalışmalara yol gösterici nitelikte olacaktır.

Bu çalışmada yüksek doğrulukla tahmin edilecek senetlerin özelliklerini belirlemek için ilk önce kullanıcının uzman bir sistem aracılığıyla hisse senedi fiyatları tahmin edilecek ve daha sonra ortaya çıkan doğru tahmin oranları, risk ve getiri değerleri ile birlikte kümeleme analizinde girdi seti olarak kullanılacaktır.

Literatürde hisse senedi fiyatlarını veya bir gün sonraki eğilimi tahmin etmek için yapay sinir ağlarının başarı ile kullanıldığı görülmektedir¹. Buna karşın yapay sinir ağları ile analiz

1 C.M. Anish- Babita Majhi, "Hybrid Nonlinear Adaptive Scheme for Stock Market Prediction Using Feedback FLANN and Factor Analysis". *Journal of the Korean Statistical Society*, 45(1), 2016, s.64-76; Michel Ballings ve diğerleri, "Evaluating Multiple Classifiers for Stock Price Direction Prediction", *Expert Systems with Applications*, 42(20), 2015, s.7046-7056; Reza Hafezi ve diğerleri, "A Bat-Neural Network Multi-Agent System (BNNMAS) for Stock Price Prediction: Case Study of Dax Stock Price", *Applied Soft Computing*, 29, 2015, s.196-210; Mehmet Özçalıcı, "Hisse Senedi Fiyat Tahminlerinde Bilgi İşlemsel Zeka Yöntemleri: Uzman Bir Sistem Aracılığıyla BİST Uygulaması", Kahramanmaraş, Sütçü İmam Üniversitesi Sosyal Bilimler Enstitüsü, 2015, s.154, (Yayınlanmamış Doktora Tezi); Javad Zahedi- Mohammad Mahdi Rounaghi, "Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran Stock Exchange", *Physica A: Statistical Mechanics and its Applications*, 2015, 438, 178-187; Emin Avcı, "Stock Return Forecasts with Artificial Neural Network Models", *Marmara Üniversitesi İİBF Dergisi*, 26(1), 2009, s.443:461.

gerçekleştirilirken, kullanıcının eğitim parametrelerine karar vermesi gerekliliği ve öğrenme sürecinin yavaş olması yeni yöntemlerin geliştirilmesini zorunlu kılmıştır.

Aşırı Öğrenme Makineleri (AÖM) yapay sinir ağları tabanlı bir analiz yöntemidir ve yapay sinir ağlarının yukarıda değinilen olumsuz taraflarını ortadan kaldırmak üzere Huang tarafından literatüre tanıtılmıştır² ve geleneksel Yapay Sinir ağları veya Destek Vektör makineleri gibi yöntemlerinden daha iyi genelleştirebilme özelliğine sahiptir³. AÖM hisse senedi fiyat tahmini için yeni yeni kullanılmaya başlanmıştır⁴. AÖM eğitim için parametre ayarlaması gerektirmese de, ara katmanda yer alması gereken nöron sayısı ve aktivasyon fonksiyonu çeşidine kullanıcının karar vermesi gerekir ve bu parametreler tahmin performansını etkileyebilir.

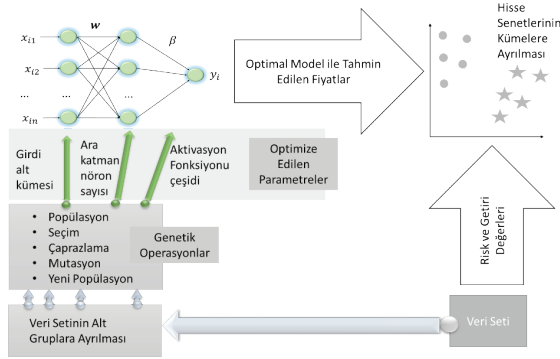
Bunun yanı sıra kısa süreli hisse senedi fiyat tahminlerinde teknik göstergelerin ve uzun süreli fiyat tahminlerinde ise temel göstergelerin kullanılması gerektiği literatürde kabul görmektedir⁵. Bu çalışmada kısa süreli tahmin gerçekleştirileceği için teknik göstergelerin kullanılması uygundur. Çok fazla sayıda teknik gösterge hesaplamak mümkündür. Örneğin bu çalışma için 196 adet teknik gösterge hesaplanmıştır. Bu değişken kümesinin içinden 10 tanesi $\left(\frac{196!}{10!(186!)}\right)$ farklı şekilde seçilebilir. Ancak bütün olasılıkları tek tek deneyip, kaydedip aralarından en iyisini seçmek rasyonel bir zamanda veya maliyette gerçekleşmez. Bu problemi çözmek için meta-sezgisel bir algoritma olan genetik algoritmalar kullanılabilir. Literatürde tanıtıldığından beri genetik algoritma birçok optimizasyon probleminin çözümü için başarıyla kullanılmaktadır⁶. Genetik algoritmalar hisse senedi tahmininde değişken seçimi için de başarıyla kullanılmaktadır⁷.

Çalışma dört bölümden oluşmaktadır. Bu giriş bölümünden sonra ikinci bölümde çalışmada kullanılan yöntemler tanıtılmıştır. Üçüncü bölümde ise çalışmada kullanılan veri seti tanıtılmış ve analiz sonuçlarına yer verilmiştir. Dördüncü bölümde ise sonuç ve tartışmanın yanı sıra sonraki çalışmalar için öneriler yer almaktadır.

-
- 2 Guang-Bin Huang ve diğerleri, "Extreme Learning Machine: Theory and Applications", *Neurocomputing*, 70(1-3), 2006, s.489-501.
 - 3 Guang-Bin Huang ve diğerleri, "Extreme Learning Machines: A Survey", *International Journal of Machine Learning and Cybernetics*, 2(2), 2011, s.107-122; Gao Huang ve diğerleri, "Trends in Extreme Learning Machines: A Review", *Neural Networks*, 61, 2015, s.32-48.
 - 4 Rajashree Dash ve diğerleri, "A Self Adaptive Differential Harmony Search Based Optimized Extreme Learning Machine for Financial Time Series Prediction", *Swarm and Evolutionary Computation*, 19, 2014, s.25-42; Xiaodong Li ve diğerleri, "Empirical Analysis: Stock Market Prediction via Extreme Learning Machine", *Neural Computing and Applications*, 27(1), 2014, s.67-78.
 - 5 George S. Atsalakis- Kimon P. Valavanis, "Forecasting Stock Market Short-Term Trends Using a Neuro-Fuzzy Based Methodology", *Expert Systems with Applications*, 36(7), 2009, s.10696-10707.
 - 6 Xin-She Yang, *Nature-Inspired Metaheuristic Algorithms*, Frome, Luniver Press, 2010, s.6.
 - 7 Kyoung jae Kim- Ingoo Han, "Genetic Algorithms Approach to Feature Discretization in Artificial Neural Networks for the Prediction of Stock Price Index", *Expert Systems with Applications*, 19(2), 2000, s.125-132; Mehmet Özçalıcı a.g.e., 2015.

2. Metodoloji

Çalışmanın modeli Şekil 1'de görülmektedir. İlk olarak ham veri seti bir araya getirilmektedir. Daha sonra girdi ve çıktı değişkenleri hesaplanmaktadır. Veri seti ayrıca eğitim ve test olmak üzere iki gruba ayrılmaktadır. Çalışmada hisse senedi fiyat tahmini gerçekleştirecek modelin parametreleri ve değişken seçimi genetik algoritma ile gerçekleştirilecektir. Başka bir ifade ile Aşırı Öğrenme Makinelerinde ara katmandaki nöron sayısı ve aktivasyon fonksiyonu ile optimal tahmin değerini gerçekleştireceği düşünülen değişken alt kümesi genetik algoritma ile optimize edilmektedir. Optimize edilen model ile fiyat tahminleri gerçekleştirildikten sonra hisse senetlerinin risk ve getiri değerleri ile birlikte kümeleme analizinde kullanılmaktadır. Bu işlem ile birlikte, yüksek doğrulukla tahmine elverişli senetlerin özellikleri belirlenmeye çalışılmaktadır.



Şekil 1. Çalışmanın Modeli

2.1. Aşırı Öğrenme Makineleri

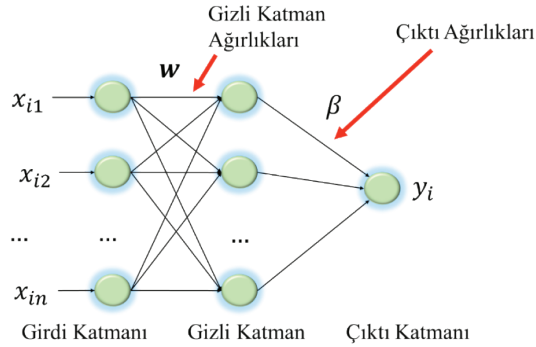
Aşırı öğrenme makineleri temelde tek ara katmanı olan, çok katmanlı yapay sinir ağları modelidir (Şekil 2). Geleneksel tek katmanlı ileri beslemeli yapay sinir ağları modelinde adet gözlem gizli katmanında adet nöron bulunan standart ileri beslemeli yapay sinir ağları şu şekilde modellenebilir:

$$\sum_{i=1}^k \beta_i g(\mathbf{w}_i \mathbf{x}_i + b_i) = 0_j, j = 1, \dots, n \quad (1)$$

Bu modelde çıktı ağırlığı, girdi ağırlık matrisini ve ifadesi nöronundaki eşik (bias) değerlerini temsil etmektedir⁸. Geleneksel yapay sinir ağları yönteminde bu modeldeki ağırlıkları bir döngü içinde hesaplanmaktadır. Buna karşın, modeldeki β ağırlıklarını analitik bir şekilde tek seferde hesaplayacak yeni bir yöntem önerilmiştir⁹. Bu yöntemin adı AÖM olarak adlandırılmaktadır. Modeldeki ağırlıklar döngü içerisinde değil de tek seferde belirlendiğinden ötürü yapay sinir ağlarına göre oldukça hızlı bir şekilde ağırlıklar belirlenmektedir.

8 Xiaodong Li ve diğerleri, **a.g.m.**, 2014.

9 Guang-Bin Huang ve diğerleri, **a.g.m.**, 2006.



Şekil 2. Aşırı Öğrenme Makineleri

2.2. Genetik Algoritmalar

1960'lı yıllardan beri yaşayan organizmaları taklit etmek suretiyle, zor olan optimizasyon problemlerini çözebilecek güçlü algoritmalar geliştirilmeye çalışılmaktadır. Bu hesaplama tekniklerine evrimsel hesaplama da denilmektedir. Genetik algoritmalar, evrimsel hesaplamalar içerisinde en bilinen yöntemdir¹⁰.

Çalışmada aşırı uyum (overfitting) problemini engellemek adına farklı bir uygunluk fonksiyonu tasarlanmıştır. Çalışmada kullanılan uygunluk fonksiyonunda eğitim veri setini sıralı 9 parçaya ayrılmaktadır ve her bir parça kendi içinde eğitim (240) ve test (60) olarak iki parçaya ayrılmaktadır. Sıralı alt parçaların her birinde eğitim setinde AÖM yöntemi belirlenen parametrelerle eğitilmektedir ve test setinde teste tabi tutulmak suretiyle o alt parçadaki hata hesaplanmaktadır. Uygunluk fonksiyonun çıktısı ise bu sıralı setlerdeki hataların aritmetik ortalamasıdır. Uygunluk fonksiyonunu sahte kodu Şekil 3'deki gibidir:

```

Function çıktı = uygunluk_fonksiyonu(çözüm, eğitim seti)
Aktivasyon Fonksiyonu = çözüm(1);
Gizli Katman Nöron Sayısı = çözüm(2);
Değişken indeksleri = çözüm(3 ile 12 arası)
Girdi seti = eğitim seti ( Değişken indeksleri );
Eğitim setini sıralı 9 adet alt parçaya ayır.
Her bir alt parça için
    Eğitim setini belirle;
    Test setini belirle;
    Aktivasyon fonksiyonu ve gizli katman nöron sayısı ile AÖM modeli çalıştır
    Test setinde modeli test et ve hatasını hesapla
Bitir
Çıktı = Hataların aritmetik ortalaması
  
```

Şekil 3. Çalışmada Kullanılan Uygunluk Fonksiyonunun Sahte Kodu

10 Mitsuo Gen- Runwei Cheng, **Genetic Algorithms and Engineering Optimization**, John Wiley & Sons, 2000.

Çalışmada kullanılan diğer genetik algoritma parametreleri şu şekildedir: Anakitle sayısı = 200; Çaprazlama oranı = 0.8; mutasyon oranı = 0.2 ve nesil sayısı = 100. Ayrıca kromozom gerçek değerlerle kodlanmıştır. Başka bir ifade ile 0 ve 1 değerlerinden oluşan ikili kodlama kullanılmamıştır.

2.3. K – Ortalamalar Kümeleme Yöntemi

Kümeleme analizi, birbirlerine benzeyen veri parçalarını kümelere (gruplara, segmentlere) ayırma işlemidir¹¹ ve veri madenciliğinin önemli bir konusudur. Literatürde k-ortalama yönteminin birçok alanda başarıyla kullanılmakta olduğu görülmektedir. Tüketici profileme¹² ve perakende mallarının gruplandırılması¹³ örneklerinde olduğu gibi.

K-ortalama yöntemi hiyerarşik olmayan kümeleme yöntemleri arasında önem taşır ve yaygın bir biçimde kullanılır¹⁴. K-ortalama yöntemi sürekli olarak kümelerin yenilediği ve en uygun çözüme ulaşana kadar devam eden döngüsel bir algoritmadır ve veri setini k adet kümede ve kümelerin ortalamalarına göre kümelere ayırır¹⁵.

Kümeleme yöntemlerinin önemli bir kısmı, gözlem değerleri arasındaki uzaklıkların hesaplanması esasına dayanmaktadır bu nedenle iki nokta arasındaki uzaklığı hesaplayan fonksiyonlara ihtiyaç duyulmaktadır¹⁶. Çok farklı özelliklere sahip uzaklık ölçüleri geliştirilmiştir. Bu çalışmada literatürde sıklıkla kullanılan Öklid uzaklık ölçüsü kullanılmıştır. Bu uzaklık ölçüsüne göre iki gözlem arasındaki uzaklık, değişken sayısını temsil etmek üzere, şu şekilde hesaplanmaktadır:

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2)$$

3. Bulgular

3.1. Veri Seti

Çalışmada veri seti olarak 2015 yılının son çeyreğinde BIST50 Endeksinde listelenen 50 adet hisse senedinin Ocak 2010 ile Kasım 2015 tarihleri arasındaki fiyat ve işlem hacmi bilgileri kullanılmıştır. Söz konusu bu işlem hacmi ve fiyat bilgilerini kullanmak suretiyle teknik göstergeler

11 Yalçın Özkan, **Veri Madenciliği Yöntemleri**, Papatya Yayıncılık Eğitim, İstanbul, 2013, s.131.

12 Hüseyin İnce ve diğerleri, "Öz-Düzenlemeli Harita Ağları ile K-Ortalama Kümeleme Analizinin Karşılaştırılması: Tüketici Profileme Örneği". **Gazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi**, 28(4), 2013, s.723-731.

13 Kusrini Kusrini, "Grouping of Retail Items by Using K-Means Clustering", **Procedia Computer Science**, 72, 2015, s.495-502.

14 Özkan. **a.g.k.** s.149.

15 Gökhan Silahtaroglu, **Veri Madenciliği (Kavram ve Algoritmaları)**, Papatya Yayıncılık Eğitim, İstanbul, 2013.

16 Özkan. **a.g.k.** s.132.

hesaplanmıştır. Çalışmada melez modelde değişken seçimi de gerçekleştirileceğinden, ulaşılabilen bütün teknik göstergeler bir araya getirilmiştir. Teknik göstergelerin sıralaması ve isimleri Tablo 1'de yer almaktadır. Her bir teknik göstergenin formülüne tekrardan kaçınmak ve çalışmanın hacmini arttırmamak için burada yer verilmeyecektir. Çalışmada yer alan göstergelerin formülleri ve detaylı hesaplama adımları için (Achelis, 2001) kaynağına başvurulabilir¹⁷.

Tablo 1. Çalışmada Kullanılan Teknik Göstergelerin Sıralaması ve İsimleri

Sıra	Değişken İsmi	Sıra	Değişken İsmi
1	Toplama Salınım Salınımı	23	Bollinger üst band
2	Chaikin Salınımı	24	Bollinger alt band
3	Hareketli Ortalamalar Birleştirilmesi Ayrıştırılması	25	En yüksek yüksek fiyat
4	3'ün 9 günlük ortalaması	26	En düşük düşük fiyat
5	Açılış fiyatı ivmesi	27	Medyan Fiyat
6	En yüksek fiyat ivmesi	28	Denge işlem hacmi
7	En düşük fiyat ivmesi	29	Tipik fiyat
8	Kapanış fiyatı ivmesi	30	Ağırlıklandırılmış kapanış fiyatı
9	Açılış fiyatı momentum	31	William Birleştirme Ayrıştırma
10	En yüksek fiyat momentum	32 -	Son 20 günün kapanış fiyatı
11	En düşük fiyat momentum	51	
12	Kapanış fiyatı momentumu	52 -	Son 20 günün En yüksek fiyatı
13	Chaikin Volatility	71	
14	Hızlandırılmış %K stokastik	72 -	Son 20 günün En düşük fiyatı
15	Hızlandırılmış %D stokastik	91	
16	Yavaşlatılmış %K stokastik	92 -	Son 20 günün Açılış fiyatı
17	Yavaşlatılmış %D stokastik	111	
18	William's %R	112 -	Son 20 günün İşlem Hacmi
19	Negatif hacim endeksi	131	
20	Görelî güç endeksi	132 -	1-20 günlük basit hareketli ortalama
21	Birleştirme ayrıştırma çizgisi	151	
22	Bollinger orta band	152 -	1-20 günlük üssel hareketli ortalama
		171	
		172 -	1-20 günlük üçgensel hareketli ortalama
		191	
		192	Hareket kolaylığı
		193	Detrend fiyat salınımı
		194	Kitle endeksi
		195	Para Akış endeksi
		196	Emtia Kanal endeksi

17 Steven Achelis (2001). *Technical Analysis from A to Z*, McGraw Hill.

3.2. Optimal Parametreler ve Değişken İndeksleri

AÖM yöntemi kullanılırken kullanıcının karar vermesi gereken parametrelerden biri aktivasyon fonksiyonudur. Bu çalışmada aktivasyon fonksiyonu da genetik algoritma tarafından optimize edilmiştir. Çalışmada test edilen aktivasyon fonksiyonları sırasıyla Tablo 2'deki gibidir.

Tablo 2. Çalışmada Test Edilen Aktivasyon Fonksiyonları

Sıra	Aktivasyon Fonksiyonu	Açıklama
1	sig	Sigmoidal aktivasyon fonksiyonu
2	sin	Sinus aktivasyon fonksiyonu
3	hardlim	Hardlimit aktivasyon fonksiyonu
4	tribas	Üçgensel aktivasyon fonksiyonu
5	radbas	Radyal-bazlı aktivasyon fonksiyonu

Çalışmada optimize edilen bir diğer parametre de ara katmanda yer alması gereken nöron sayısıdır. Ara katmanda yer alması gereken nöron sayısını önceden belirleyecek bir yöntem söz konusu değildir ve problemin karmaşıklık derecesine göre ara katmanda kullanılması gereken nöron sayısı değişmektedir. Çalışmada, ara katmanda yer alması gereken nöron sayısının alt sınır 1 ve üst sınırı 30 olarak belirlenmiştir.

Tablo 3'de her bir hisse senedi için genetik algoritma tarafından optimize edilmiş, en uygun aktivasyon fonksiyonu, optimal nöron sayısı ve değişken indeksleri yer almaktadır. Tablo 3'de dikkat çeken sonuçlardan biri, bütün hisse senetleri için sigmoidal aktivasyon fonksiyonun optimal olarak belirlenmesidir. Bunun yanında ara katmanda yer alması gereken optimal nöron sayıları her bir hisse senedi için değişiklik arz etmektedir.

Tablo 3. Her Bir Hisse Senedi İçin Optimal Aktivasyon Fonksiyonu, Ara Katmandaki Nöron Sayısı Ve Teknik Gösterge İndeksleri

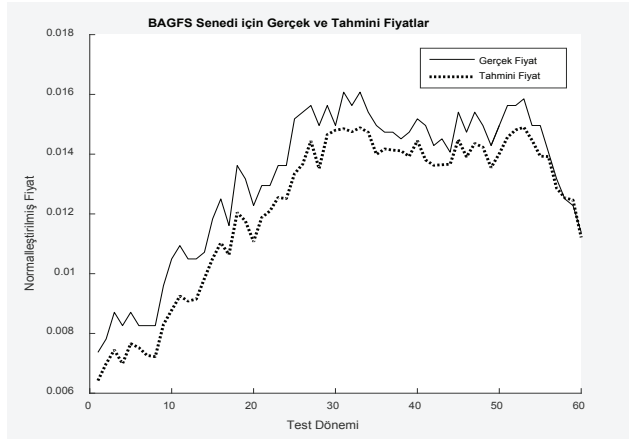
	*	**	Teknik Gösterge Sıra Numarası									
AEFES	1	14	27	38	56	71	73	83	100	135	152	166
AFYON	1	10	30	44	48	62	78	106	138	152	163	166
AKBNK	1	15	8	10	79	101	132	134	138	150	175	193
AKSEN	1	13	23	60	69	82	85	92	98	132	144	166
ARCLK	1	14	48	55	61	64	71	132	144	148	157	196
ASELS	1	15	16	44	54	77	102	112	122	144	152	153
BAGFS	1	11	33	61	83	99	103	141	149	151	152	162

BIMAS	1	13	36	44	49	60	104	146	152	169	175	176
BJKAS	1	22	12	16	30	42	54	73	84	95	105	142
CCOLA	1	8	26	29	33	37	73	89	149	162	168	181
DOAS	1	13	30	67	75	92	117	121	135	140	176	185
DOHOL	1	9	54	62	109	110	125	132	138	141	153	180
ECILC	1	13	63	67	80	82	102	104	151	152	156	162
EGEEN	1	11	58	72	75	97	98	134	143	165	172	196
EKGYO	1	8	52	64	94	95	101	132	152	159	168	176
ENKAI	1	13	23	28	84	87	89	117	127	149	152	188
EREGL	1	13	35	54	81	142	146	148	152	158	179	187
FROTO	1	15	1	29	40	74	108	111	142	156	163	194
GARAN	1	21	6	15	59	86	92	107	118	132	134	143
GOLTS	1	12	43	45	70	86	89	97	105	142	152	191
GUBRF	1	12	42	76	78	99	102	116	151	152	160	193
HALKB	1	13	56	100	133	136	154	158	159	161	164	172
ISCTR	1	12	30	82	102	120	132	140	155	164	173	184
ISGYO	1	10	24	69	72	95	107	133	143	152	159	171
KCHOL	1	14	49	85	87	144	147	155	161	172	174	187
KONYA	1	12	30	49	54	55	57	59	112	142	172	173
KOZAL	1	15	24	47	67	75	91	92	95	133	158	172
KRDMD	1	8	29	84	98	123	132	135	152	165	174	188
MGROS	1	11	4	25	34	54	81	132	142	161	167	172
NETAS	1	13	31	50	59	61	75	94	104	119	163	168
OTKAR	1	14	17	31	50	54	82	95	139	152	158	168
PETKM	1	10	67	76	91	97	98	132	141	152	155	161
PGSUS	1	11	100	101	102	105	119	140	144	152	163	184
SAHOL	1	11	4	134	137	147	149	152	156	178	180	183
SISE	1	10	22	32	37	63	132	142	146	149	155	194
TATGD	1	10	49	84	95	118	135	149	154	172	182	186
TAVHL	1	10	71	95	132	146	154	157	167	169	171	181
TCELL	1	11	31	35	60	81	107	142	158	172	184	188
THYAO	1	12	12	41	97	101	111	118	132	159	161	167
TKFEN	1	12	52	57	65	92	128	132	133	136	138	140
TOASO	1	12	24	34	47	69	73	89	112	162	172	173
TRKCM	1	18	51	53	58	74	96	106	126	152	173	180
TSKB	1	10	41	60	62	132	141	152	155	161	172	175
TTKOM	1	12	59	63	66	90	132	135	144	149	160	173
TTRAK	1	12	39	47	61	69	87	106	107	132	136	175
TUPRS	1	14	82	102	105	122	159	161	165	168	172	184
ULKER	1	16	14	25	55	77	100	133	137	170	185	196
VAKBN	1	14	55	73	109	134	151	161	165	172	174	189
VESTL	1	10	33	38	40	46	75	109	140	152	161	181
YKBNK	1	8	11	30	60	76	95	154	160	164	166	182

*: Aktivasyon fonksiyonu çeşidi.

** : Ara katmandaki nöron sayısı.

Çalışmada elde edilen tahmin sonuçlarını görselleştirmek adına en iyi tahmin performansını sergileyen BAGFS hisse senedinin tahmini fiyatları ile gerçek fiyatları grafik üzerinde Şekil 4'deki gibi görselleştirilmiştir. Şekil 3'den de görülebileceği gibi normalleştirilmiş tahmini fiyatlar ile normalleştirilmiş gerçek fiyatlar birbirlerine yakındır.



Şekil 4. BAGFS Hisse Senedi İçin Gerçek Ve Tahmini Fiyatlar

3.3. Doğru Tahmin Oranları ve Kümeleme Analizinde Kullanılacak Değişkenler

Her bir hisse senedi için eğitim setinde optimal yapı belirlendikten sonra, bu yapı her bir hisse senedinin eğitim setinde son bir kez eğitilmiş ve test setindeki girdi değişkenleri ile teste tabi tutulmuştur. Bu test setindeki performans gerçek tahmin performansını oluşturmaktadır, çünkü model, daha önce karşılaşmadığı bir veri seti üzerinde tahmin gerçekleştirmektedir. Her bir hisse senedinin test setindeki tahmin performansı Tablo 4'ün son sütununda yer almaktadır.

Hisse senedi fiyatları 50 adet hisse senedi için tahmin edildikten sonra, kapanış fiyatları kümeleme analizinde kullanılacak ve detaylı bilgilere ulaşılmaya çalışılacaktır. Hisse senedi fiyatları ile ilgili iki önemli kavram mevcuttur. Bunlar hisse senetlerinin ortalama getirisi ve risk değerleridir. Ortalama getiri Formül 3 yardımıyla hesaplanmış, risk ise hisse senetlerinin getirilerinin standart sapması olarak hesaplanmıştır.

$$\bar{r} = \frac{\sum_{i=1}^n r_i}{n} \quad (3)$$

Formülde i gününe ilişkin hisse senedi getirisini, n ise toplam seans sayısını temsil etmektedir. Hisse senedinin getirisi (\bar{r}) ise şu şekilde hesaplanmıştır

$$r_i = \log \left(\frac{p_{t+1}}{p_t} \right) \quad (4)$$

Formülde t gününe ilişkin kapanış fiyatını temsil etmektedir.

Optimal modellerin test setindeki performansı ile birlikte hisse senetlerinin ortalama getiri ve risk değerleri her bir hisse senedi için Tablo 4'deki gibidir. Başka bir ifade ile Tablo 4'de yer alan değerler kümeleme analizinin girdi setini oluşturmaktadır.

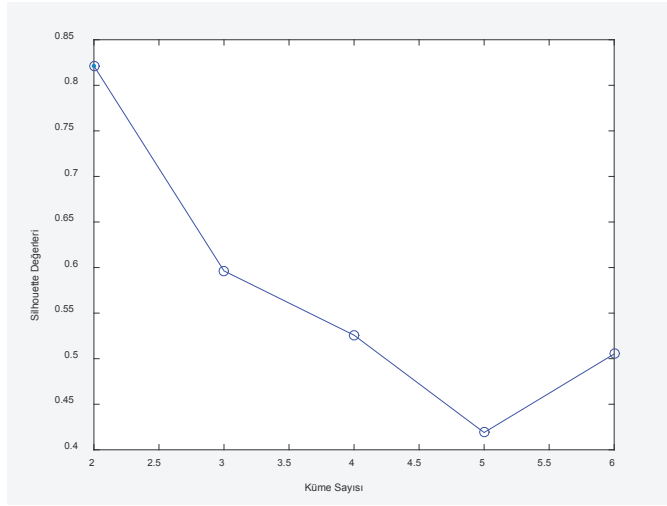
Tablo 4. Kümelemeye Esas Teşkil Eden Veri Seti

	Senet İsmi	Ortalama Getiri ($\times 10^{-3}$)	Risk	Doğru tahmin oranı	Senet İsmi	Ortalama Getiri ($\times 10^{-3}$)	Risk	Doğru tahmin oranı
1	AEFES	0,029	0,014	0,458 26	KONYA	0,308	0,020	0,576
2	<u>AFYON</u>	<u>-1,702</u>	<u>0,068</u>	<u>0,542</u> 27	KOZAL	-0,214	0,027	0,610
3	AKBNK	-0,049	0,015	0,339 28	KRDMD	0,194	0,018	0,525
4	AKSEN	-0,257	0,016	0,508 29	MGROS	-0,221	0,015	0,644
5	ARCLK	0,291	0,015	0,508 30	<u>NETAS</u>	<u>-0,623</u>	<u>0,059</u>	<u>0,525</u>
6	ASELS	-0,003	0,024	0,712 31	OTKAR	0,520	0,017	0,576
7	<u>BAGFS</u>	<u>-0,855</u>	<u>0,038</u>	<u>0,847</u> 32	PETKM	-0,349	0,033	0,576
8	BIMAS	-0,193	0,023	0,525 33	PGSUS	-0,969	0,015	0,441
9	BJKAS	-0,275	0,039	0,441 34	SAHOL	0,080	0,014	0,610
10	CCOLA	0,319	0,016	0,661 35	SISE	0,120	0,016	0,356
11	DOAS	0,224	0,022	0,610 36	TATGD	1,615	0,015	0,610
12	DOHOL	-0,306	0,019	0,390 37	TAVHL	0,446	0,015	0,610
13	ECILC	-0,055	0,013	0,559 38	TCELL	0,046	0,013	0,525
14	EGEEN	0,982	0,024	0,390 39	THYAO	0,172	0,016	0,644
15	EKGYO	-0,058	0,015	0,508 40	TKFEN	-0,125	0,014	0,559
16	ENKAI	-0,150	0,015	0,373 41	TOASO	0,379	0,017	0,492
17	EREGL	-0,085	0,016	0,695 42	TRKCM	-0,128	0,015	0,593
18	FROTO	0,357	0,015	0,576 43	TSKB	-0,177	0,017	0,542
19	GARAN	-0,025	0,015	0,542 44	TTKOM	0,026	0,013	0,508
20	GOLTS	-0,061	0,019	0,373 45	TTRAK	0,665	0,016	0,559
21	GUBRF	-0,219	0,032	0,390 46	TUPRS	0,276	0,014	0,542
22	HALKB	-0,069	0,016	0,322 47	ULKER	0,559	0,016	0,576
23	ISCTR	-0,063	0,015	0,475 48	VAKBN	-0,036	0,016	0,508
24	ISGYO	-0,092	0,015	0,712 49	VESTL	0,159	0,019	0,407
25	KCHOL	0,240	0,014	0,559 50	YKBNK	-0,097	0,015	0,542

3.4. Optimal Küme Sayısı

Optimal küme sayısının belirlenmesinde genellikle kümeleme analizinin kalitesini ölçen ölçeklerden yararlanılmaktadır. Bu ölçekler, kümelerin merkezlerinin birbirlerinden ne kadar uzak olduğunu ve kümelerin ne kadar homojen olduğunu birlikte ölçmeye çalışmaktadır. Bu yöntemlerden bir tanesi de Silhouette değeridir. Çalışmada Tablo 4'deki veri seti ayrı ayrı 2 ile 6 adet kümeye ayrıldığı analizlere tabi tutulmuş ve Silhouette katsayıları hesaplanmıştır. Söz konusu değerler Şekil 5'de gösterilmektedir. Silhouette değerleri ne kadar yüksek çıkarsa, kümeleme o

kadar iyi yapılmış demektir¹⁸. Söz konusu altı model arasında en yüksek Silhouette değeri, veri setininin 2 adet kümeye ayrıldığı modelde gerçekleştirilmiştir. Dolayısıyla optimal küme sayısı 2 olarak belirlenmiştir.



Şekil 5. Farklı Kümeleme Analizleri İçin Silhouette Değerleri

3.5. Kümeleme Analizi

Optimal küme sayısı 2 olarak belirlendikten sonra, Tablo 4'deki veri seti k-ortalamlar yöntemi ile 2 adet kümeye ayrılmıştır. Tablo 4'deki altı çizili, koyu ve italik değerler ikinci kümede yer alırken, diğer hisse senetleri ilk kümede yer almaktadır. Kümelerin ortalama değerleri Tablo 5'deki gibidir.

Tablo 5. Kümelerin Tanımlayıcı İstatistikleri

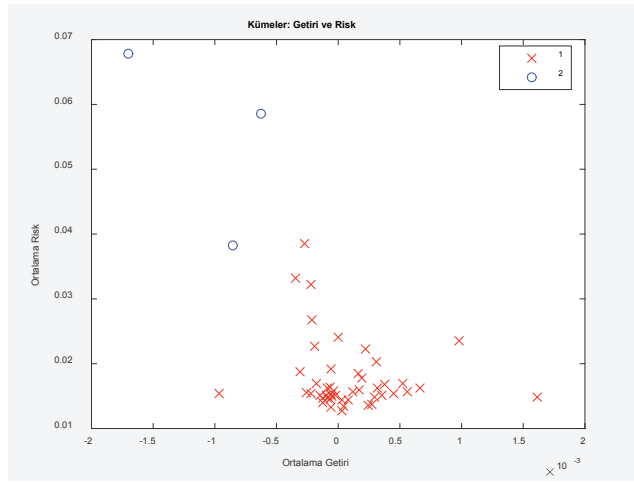
	Ortalama Getiri ($\times 10^{-3}$)	Ortalama Risk	Ortalama Doğru Tahmin Oranı
Birinci Küme	0.0794	0.0177	0.5269
İkinci Küme	-1.0599	0.0549	0.6384

Tablo 5'e göre birinci kümedeki hisse senetlerinin fiyatları ortalama olarak 52.69% oranında önceden doğru bir şekilde tahmin edilebilmiş ve ortalama risk değeri 0.0177 olarak hesaplanmıştır. Ortalama getiri değeri ise yaklaşık 0.0001 olarak hesaplanmıştır.

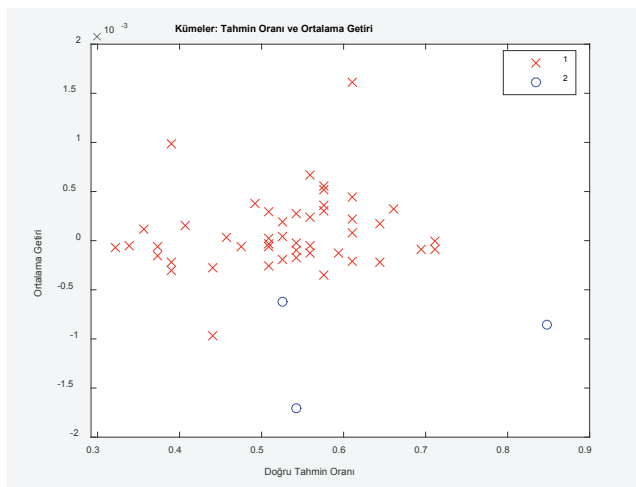
18 Peter J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis", **Journal of Computational and Applied Mathematics**, 1987, 20, s53-65.

Tablo 5'e göre ikinci kümedeki hisse senetlerinin kapanış fiyatları genel olarak daha yüksek doğrulukla önceden tahmin edilmiştir. Ayrıca bu gruptaki senetlerin ortalama getirisi negatiftir ve birinci küme ile karşılaştırıldığında daha yüksek riske sahip hisse senetleridir.

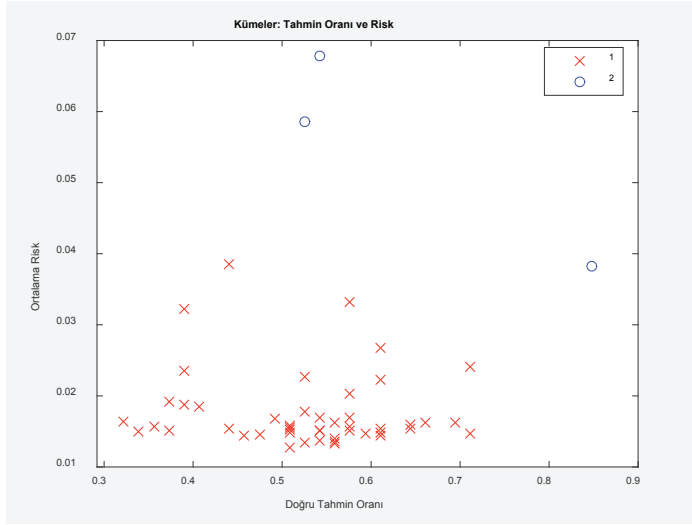
Şekil 6, 7 ve 8'de kümeler dağılım (serpilme) grafiklerinde görselleştirilmiştir. Şekillerde eksenlerde sırasıyla getiri ve risk, doğru tahmin oranı ve ortalama getiri ile doğru tahmin oranı ve ortalama risk değerleri yer almaktadır. Şekillerde farklı kümelere ayrıştırılmış hisse senetleri farklı işaretlerle gösterilmiştir. Her şekilde kümelerin homojen oldukları ve birbirlerinden farklı oldukları görülmektedir. Başka bir ifade ile kümelerin homojen olduğu anlaşılmaktadır.



Şekil 6. Getiri Ve Risk Değerlerine Göre Kümeler



Şekil 7. Doğru Tahmin Oranı Ve Ortalama Getiri Değişkenlerine Göre Kümeler



Şekil 8. Doğru Tahmin Oranı Ve Ortalama Risk Değerlerine Göre Kümeler

4. Sonuç

Çalışmada fiyat tahmin yöntemi olarak aşırı öğrenme yöntemi seçilmiştir. Bunun nedeni aşırı öğrenme yönteminin, son derece hızlı sonuç vermesidir. 50 adet hisse senedi için 100 adet döngü ve her döngüde uygunluk fonksiyonunun 200 kere çalıştırılacağı düşünüldüğünde (uygunluk fonksiyonu da kendi içinde 9 kere tahmin gerçekleştirmektedir) AÖM yönteminin sağlayacağı zaman tasarrufu ortaya çıkacaktır. AÖM yöntemini kullanmak için aktivasyon fonksiyonu ve ara katmandaki nöron sayısına karar verilmesi gerekir. Aynı zamanda hangi girdi değişkenlerinin de tahminde kullanılması gerektiğinin belirlenmesi gerekmektedir. Bu kararların alınmasında Genetik Algoritma kullanılmıştır. Sonuçta bilgisayar tabanlı uzman bir sistem ile her bir hisse senedi için optimal yapı belirlenmiş ve tahmin gerçekleştirilmiştir. Elde edilen tahmin sonuçları kümeleme analizine tabi tutulmuştur. Sonuçta 50 adet hisse senedi iki adet kümeye ayrılmıştır. Kümelerin ortalama değerleri birbirleri ile karşılaştırıldığında, birinci kümede yer alan senetlerin ortalama getirisinin daha yüksek olduğu, daha düşük risk değerine sahip oldukları ve daha düşük doğrulukla tahmin edilebildikleri ortaya çıkmıştır. İkinci kümedeki senetlerin getirisi ise daha düşüktür, buna karşın daha yüksek riske sahip ve yüksek doğrulukla tahmin edilebilen senetler olduğu belirlenmiştir.

Literatürde önceki yapılan çalışmalar incelendiğinde, hisse senedi fiyatlarının genellikle teknik göstergeler veya temel göstergeler ile tahmin edilmeye çalışıldığı görülmektedir. Bu çalışmada bir adım daha ileriye gidilmiş ve uzman bir sistem ile fiyat tahmini gerçekleştirildikten sonra elde edilen tahmin oranları, hisse senetlerinin risk ve getiri değerleri ile birlikte kümeleme analizinde girdi olarak kullanılmıştır. İlerleyen çalışmalarda karar ağaçları yardımıyla kurallar oluşturmak suretiyle, daha yüksek doğrulukla tahmin edilecek senetlerin özellikleri daha detaylı bir şekilde analiz edilebilir.

Kaynakça

- ACHELIS, Steven, **Technical Analysis from A to Z**, McGraw Hill, 2001.
- ANISH, C.M. – Majhi, Babita, “Hybrid Nonlinear Adaptive Scheme for Stock Market Prediction Using Feedback FLANN and Factor Analysis”. **Journal of the Korean Statistical Society**, 45(1), 2016, s.64-76.
- ATSALAKIS, George S. – Valavanis, Kimon P., “Forecasting Stock Market Short-Term Trends Using A Neuro-Fuzzy Based Methodology”. **Expert Systems with Applications**, 36(7), 2009, s.10696-10707.
- AVCI, Emin, “Stock Return Forecasts with Artificial Neural Network Models”, **Marmara Üniversitesi İİBF Dergisi**, 26(1), 2009, s.443-461.
- BALLINGS, Michel, Van Den Poel, Dirk, Hespels, Nathalie ve Gryp, Ruben, “Evaluating Multiple Classifiers for Stock Price Direction Prediction”, **Expert Systems with Applications**, 42(20), 2015, s.7046-7056.
- DASH, Rajashree, Dash, P.K. ve Bisoi Ranjeeta, “A Self Adaptive Differential Harmony Search Based Optimized Extreme Learning Machine for Financial Time Series Prediction”, **Swarm and Evolutionary Computation**, 19, 2014, s.25-42.
- GEN, Mitsuo – Cheng, Runwei, **Genetic Algorithms and Engineering Optimization**, John Wiley & Sons, 2000.
- HAFEZI, Reza, Shahrabi, Jamal ve Hadavandi, Esmail, “A Bat-Neural Network Multi-Agent System (Bnnmas) for Stock Price Prediction: Case Study of Dax Stock Price”, **Applied Soft Computing**, 29, 2015, s.196-210.
- HUANG, Guang-Bin, Wang, Dian Hui ve Lan, Yuan, “Extreme Learning Machines: A Survey”, **International Journal of Machine Learning and Cybernetics**, 2(2), 2011, s.107-122.
- HUANG, Guang-Bin, Zhu, Qin-Yu ve Siew, Chee-Kheong, “Extreme Learning Machine: Theory and Applications”, **Neurocomputing**, 70(1-3), 2006, s.489-501.
- HUANG, Gao, Huang, Guang-Bin, Song, Shiji ve You, Keyou, “Trends in Extreme Learning Machines: A Review”, **Neural Networks**, 61, 2015, s.32-48.
- İNCE, Hüseyin, Imamoğlu, Salih Zeki ve Keskin, Halit, “Öz-Düzenlemeli Harita Ağları ile K-Ortalama Kümeleme Analizinin Karşılaştırılması: Tüketici Profilleme Örneği”. **Gazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi**, 28(4), 2013, s.723-731.
- KIM, Kyoung jae – Han, Ingoo, “Genetic Algorithms Approach to Feature Discretization in Artificial Neural Networks for the Prediction of Stock Price Index”, **Expert Systems with Applications**, 19(2), 2000, s.125-132.
- KUSRINI, Kusrini, “Grouping of Retail Items by Using K-Means Clustering”, **Procedia Computer Science**, 72, 2015, s.495-502.
- LI, Xiaodong, Xie, Haoran, Wang, Ran, Cai, Yi, Cao, Jingjing, Wang, Feng, Min, Huaqing ve Deng, Xiaotie. “Empirical Analysis: Stock Market Prediction via Extreme Learning Machine”, **Neural Computing and Applications**, 27(1), 2014, s.67-78.
- ÖZÇALICI, Mehmet, “Hisse Senedi Fiyat Tahminlerinde Bilgi İşlemsel Zeka Yöntemleri: Uzman Bir Sistem Aracılığıyla BİST Uygulaması”, Kahramanmaraş, Sütçü İmam Üniversitesi Sosyal Bilimler Enstitüsü, 2015 (Yayınlanmamış Doktora Tezi).
- ÖZKAN, Yalçın, **Veri Madenciliği Yöntemleri**, Papatya Yayıncılık Eğitim, İstanbul, 2013.
- ROUSSEEUW, Peter J., “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis”, **Journal of Computational and Applied Mathematics**, 1987, 20, s53-65.
- SILAHTAROĞLU, Gökhan, **Veri Madenciliği Kavram ve Algoritmaları**, Papatya Yayıncılık Eğitim, İstanbul, 2013.
- YANG, Xin-She, **Nautre-Inspired Metaheuristic Algorithms**, Frome, Luniver Press, 2010.
- ZAHEDI, Javad – Rounaghi, Mohammad Mahdi, “Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran Stock Exchange”, **Physica A: Statistical Mechanics and its Applications**, 438, 2015, 178-187.

