

## YAPAY SİNİR AĞLARI TAHMİNLEME MODELİ İLE HİSSE SENEDİ FİYAT ENDEKSİ TAHMİNLEMESİ ÜZERİNE BİR ÇALIŞMA

M. Çiğdem AKBAŞ\*

**ÖZ:** Menkul kıymetler piyasasında hisse senedi fiyat endekslerinin gelecekteki değerlerinin tahmin edilmesi için yapay sinir ağlarını esas alan iki farklı modelleme yaklaşımı sunulmaktadır. Birinci tahminleme modelinde, yapay sinir ağları kullanılarak verileri mevcut olan karar değişkenlerinin tümünün modellemeye dahil edildiği bir zaman serisi tahminleme modeli uygulanmaktadır. İkinci modelde, regresyon ağaçları ve yapay sinir ağlarıyla bütünleşik bir tahminleme modeli tasarlanmaktadır. Regresyon ağaçları, yapay sinir ağlarında girdi olarak kullanılacak olan en önemli karar değişkenlerinin belirlenmesi için özellik seçme işlemi gerçekleştirmektedir. Önemli karar değişkenlerine bağlı olarak oluşturulan modelde menkul kıymetler borsası endekslerinin geleceğe yönelik tahminlemesi yapay sinir ağlarıyla gerçekleştirilmektedir. Yapay sinir ağlarıyla tahminleme modellerinde, zaman serileri olarak elde edilen veriler eğitim, doğrulama ve test verisi olarak gruplandırıldıktan sonra endekslerin gelecekteki değerleri iki aşamalı yapay sinir ağlarıyla tahmin edilmektedir. İleri sürülen tahminleme modelleri için performans değerlendirmesi işlem süresi ve istatistiksel göstergelere dayalı tahminleme doğruluğu kriterlerine göre yapılmaktadır.

**Anahtar Kelimeler:** *Yapay Sinir Ağları, Regresyon Ağaçları, Önem Analizi, Tahminleme Modeli, Borsa Endeksi*

**Makalenin türü:** Araştırma

**Jel Sınıflandırması:** E47

**DOI:** 10.54969/abuijss.1208857

**Geliş tarihi:** 23.11.2022 / **Kabul Tarihi:** 17.12.2022 / **Yayın Tarihi:** 31.12.2022

## A STUDY ON STOCK PRICE INDEX PREDICTION WITH ARTIFICIAL NEURAL NETWORKS PREDICTION MODEL

**ABSTRACT:** Two different modeling approaches based on artificial neural networks are presented to predict the future values of stock price indices in the securities market. In the first forecasting model, a time series forecasting model is applied using artificial neural networks, in which all the decision variables with available data are included in the modeling. In the second model, an integrated prediction model is designed with regression trees and artificial neural networks. Regression trees perform feature selection to determine the most important decision variables to be used as inputs in artificial neural networks. In the model created depending on the important decision variables, the future estimation of the stock market indices is carried out with artificial neural networks. In the prediction models with artificial neural networks, after the data obtained as time series are grouped as training, validation and test data, the future values of the indices are estimated with two-stage neural networks. Performance evaluation for the proposed estimation models is made according to the processing time and estimation accuracy criteria based on statistical indicators.

**Key Words:** *Artificial Neural Networks, Regression Trees, Importance Analysis, Forecasting Model, Stock Market Index*

**Type of article:** Research

**Jel Classification:** E47

**DOI:** 10.54969/abuijss.1208857

**Received:** 23.11.2022 / **Accepted:** 17.12.2022 / **Published:** 31.12.2022

\* Öğr. Gör. Dr., Antalya Bilim Üniversitesi, Antalya, [cigdem.akbas@antalya.edu.tr](mailto:cigdem.akbas@antalya.edu.tr), ORCID:0000-0002-1425-9828

**Kaynak gösterimi için:**

AKBAŞ M.Ç. (2022). Yapay Sinir Ağları Tahminleme Modeli ile Hisse Senedi Fiyat Endeksi Tahminlemesi Üzerine Bir Çalışma. Antalya Bilim Üniversitesi Uluslararası Sosyal Bilimler Dergisi, 3 (2), ss.97-110. DOI: 10.54969/abuijss.1026807.

## 1. GİRİŞ

Geleceğe dönük yönetsel kararların verilmesinde karşılaşılan belirsizlik ve risk unsurları tahminleme modellerine ihtiyaç duyulmasına yol açmaktadır. Değişken, karmaşık ve dinamik bir yapıya sahip olan menkul kıymetler piyasası yatırımlarında risklerin kazançla dönüştürülmesi amacıyla tahminleme modellerinin kullanılması, hisse senedi, tahvil ve bonolardan oluşan menkul kıymetlerin fiyat ve getirilerini tahmin edilebilir kılmaktadır. Menkul kıymetler piyasasında yatırımcılar için daha yüksek getiri elde etme veya gerçekleşmesi mümkün olan zararları önleme imkânı sağlayan tahminleme modellerinin tasarlanması bilimsel yazında karşılaşılan önemli bir araştırma konusudur.

Bilimsel yazında hisse senedi fiyat tahminlemede kullanılan geleneksel yöntemler; teknik analiz ve temel analiz yöntemleri olmak üzere iki ana gruptan oluşmaktadır. Hisse senedi piyasalarındaki değişkenlikler ve dalgalanmalar fiyat tahminlemesini güçleştirmekte, menkul kıymetler borsası, hisse senetleri ve makro ekonomi göstergeleriyle ilgili verilerin kullanımıyla tasarlanan yüksek tahminleme doğruluğu sağlayan gelişmiş tahminleme modellerine ihtiyaç duyulmasına sebep olmaktadır (Başoğlu vd., 2009; Hsu, 2011). Teknik analiz modeli, sadece sıralanmış finansal verilerden yararlanarak geleceğe dönük tahminlemelerin gerçekleştirilmesi yaklaşımını benimsemektedir (Kumar ve Ningombam, 2018). Sadece finansal verileri tahminleme modelinde girdi olarak kullanan teknik analiz modelinde model tasarımı kolay yapılabilmektedir, ancak model girdilerindeki çeşitliliğin az olması teknik analiz modelinin piyasadaki anlık değişikliklere yüksek doğrulukta yanıt vermesini güçleştirmektedir. Teknik analizden farklı olarak, temel analiz modelinde daha fazla piyasa verisinden girdi olarak yararlanılmaktadır (Borovkova ve Tsiamas, 2019). Modellemede kullanılacak girdi çeşitliliğinin fazla olması temel analiz modelindeki en uygun girdilerin belirlenmesini zorlaştırmaktadır.

Zaman serileri modellerinde kullanılmakta olan geleneksel istatistiksel modelleme yaklaşımında, hareketli ortalama, basit üstel yumuşatma, bütünleşik otoregresif hareketli ortalama ve doğrusal regresyon tekniklerinden oluşan doğrusal tahminleme modelleri uygulanmaktadır (Bollerslev, 1986; Hsieh, 1991; Rao ve Gabr, 1984; Jarrett ve Schilling, 2008; Olaniyi vd., 2011)). Teknik analiz ve temel analiz modellerinde finansal verilerin tahmin edilmesi konusunda son yıllardaki gelişmelerle yapay sinir ağları, destek vektör makineleri, ağaç tabanlı tahmin modeli ve k-en yakın komşu yöntemlerinin kapsayan yapay zekâ teknikleriyle oluşturulan doğrusal olmayan zaman serisi regresyon ve sınıflandırma yöntemleri kullanılmaktadır. Leung vd. (2000), sınıflandırma modellerinden olan ayırtaç analizi, logit modeli, probit modeli ve olasılıksal sinir ağı ile seviye tahminleme modellerinden uyarlanabilir üstel yumuşatma, Kalman filtresiyle vektör otoregresyonu, çok değişkenli transfer fonksiyonu ve yapay sinir ağını kullanarak hisse senedi indeks hareketinin tahmin edilmesinde uygulanmak üzere modelleme yaklaşımları sunmuştur. Ou ve Wang (2009), yatırımcıların karlarını enbüyüklemede büyük öneme sahip olan hisse senedi fiyat hareketi tahminlemede yüksek tahmin doğruluğu elde etmek için doğrusal ayırtaç analizi, ikinci dereceden ayırtaç analizi, k-en yakın komşu sınıflandırması, çekirdek tahmini tabanlı Naif Bayes, logit modeli, ağaç tabanlı sınıflandırma, yapay sinir ağı, Gauss işlemi ile Bayesgil sınıflandırma, destek vektör makinası ve en küçük kareler destek vektör regresyonundan oluşan veri madenciliği tekniklerini kullanmışlar, elde ettikleri sonuçlara göre en yüksek tahminleme doğruluğunu destek vektör makinası ve en küçük kareler destek vektör regresyonu modelleriyle elde etmişlerdir.

Yapay sinir ağları, hisse senedi fiyatı ve borsa endeksi değerinin tahmin edilmesinde çok sayıda uygulamaları olan bir yapay zekâ yöntemi olarak dikkat çekmektedir. Tektaş ve

Karataş (2004), günlük ve haftalık verilerle karşılaştırmalı olarak yapay sinir ağları ve doğrusal regresyon modellerinin finans alanındaki uygulanabilirliklerini karşılaştırmak amacıyla Türkiye’de hisse senedi fiyat tahminlemede bir uygulama gerçekleştirmiş ve yapay sinir ağının daha yüksek tahmin netliği sağladığı sonucunu elde etmiştir. Kara vd. (2011), yapay sinir ağları ve destek vektör makinaları sınıflandırma yöntemlerini kullanarak İstanbul Menkul Kıymetler Borsası (İMKB) ulusal 100 endeksinin hareket yönünün günlük verilerle tahmin edilmesini sağlamış ve yapay sinir ağlarının tutarlı bir şekilde daha yüksek tahmin doğruluğu sağladığı sonucunu elde etmiştir. Erdoğan ve Özyürek (2012), İMKB 100 endeksine dahil olan beyaz eşya firmalarının günlük hisse senedi fiyat tahminlemede yapay sinir ağları yöntemini kullanmıştır. Ayodele vd. (2012), menkul kıymetler piyasasındaki hisse senedi fiyat endekslerinin tahmininde tahminleme doğruluğunun artırılması için yapay sinir ağlarının kullanıldığı bir model sunmuştur. Yürük (2021), Türk Hava Yolları hisse senedinin tahminini yapay sinir ağları metodunu kullanarak gerçekleştirmiştir. Çınaroğlu ve Avcı (2020), Borsa İstanbul’da işlem gören Türk Hava Yolları şirketine ait hisse senedi değerlerini yapay sinir ağları yaklaşımı ile tahmin eden bir model kullanmıştır. Bilimsel yazındaki araştırmalardan farklı olarak bu makalede borsa endeksi tahminlemesi için sadece yapay sinir ağlarının kullanıldığı ve regresyon ağaçları ve yapay sinir ağlarının birlikte kullanıldığı modeller olmak üzere iki farklı tahminleme modeli tasarlanmaktadır. Yapay sinir ağları tahminleme modelinde, borsa endeksi tahminlemesi için zaman serilerinden elde edilen verilere bağlı kalınarak tasarlanan bir modelleme yaklaşımı kullanılmakta ve çok sayıda piyasa girdisinden yararlanarak modellemedeki girdi çeşitliliğinin fazla tutulması amaçlanmaktadır. Diğer taraftan, ikinci modelleme yaklaşımında, karar değişkenlerinden oluşturulan model girdilerinin önemli olanlarının belirlenmesinde regresyon ağaçları yöntemi kullanılmakta, önem değeri belirli bir eşik değeri geçen girdiler tahminleme modelinde kullanılacak olan karar değişkeni olarak seçilmekte ve modellemedeki işlem karmaşıklığını azaltmak ve işlem süresini kısaltmak maksadıyla sadece önemli karar değişkenlerinin yapay sinir ağı tahmin modeline dahil edilmesiyle borsa endeksinin gelecekteki değerleri tahmin edilmektedir. Her iki modelleme yaklaşımının karşılaştırması işlem süreleri ve istatistiksel göstergelerle hesaplanarak belirlenen tahminleme doğruluklarına göre yapılmaktadır.

Makalenin ilerleyen kısımlarındaki konu içerikleri bu paragrafta verilmektedir. Bir sonraki bölümde, modelleme yöntemleri olan regresyon ağaçları ve yapay sinir ağları anlatılmakta ve modelleme verileri ile modellemede kullanılan değişkenler tanımlanmaktadır. Üçüncü bölümde, Borsa İstanbul’da işlem görmekte olan bir endeksin değerlerinin tahmin edilmesi konusunda modelleme sonuçları açıklanmakta ve tartışılmaktadır. Son olarak, dördüncü bölümde makalenin sonuçları ve geleceğe yönelik araştırma konuları açıklanmaktadır.

## 2. Tahminleme Modelinde Kullanılan Yöntemler, Veriler ve Değişkenler

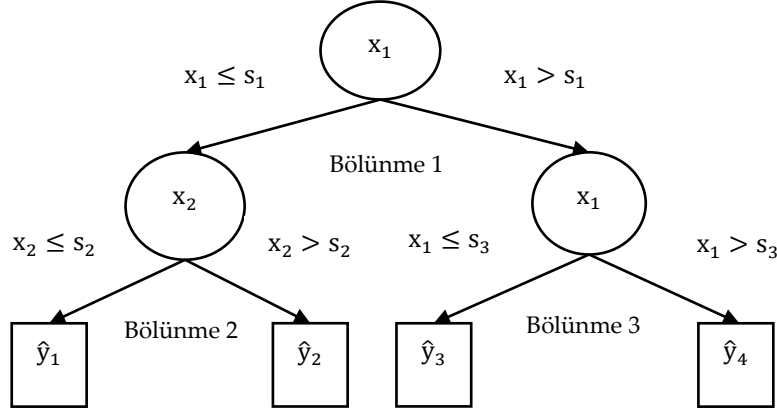
Tahminleme modelinin tasarlanmasında kullanılan yöntemler regresyon ağaçları ve yapay sinir ağları olup, regresyon ağaçları özellik seçme ve yapay sinir ağları ise gelecek değerlerini tahmin etme yöntemi olarak uygulanmaktadır.

### 2.1. Regresyon ağaçları

Regresyon ağaçları, kök düğüm ile yaprak düğümler arasında alınmakta olan kararları takip etmek suretiyle yanıt değişkeninin alacağı değerlerin tahmin edilmesinde kullanılmaktadır. Kök düğüm ile yaprak düğümler arasındaki karar alma işlemleri Şekil 1’de sunulan regresyon ağacı üzerinde gösterilmektedir. Kök düğüm regresyon ağacının en üst noktasındaki düğümü temsil etmekte olup, eğitim verisini kapsamaktadır.  $\hat{y}_i$  yanıt

değişkeninin tahmini değerini gösterirken  $i \in R_m$  varış düğümlerini temsil etmekte olan her bir yaprak düğüme atanmaktadır.

**Şekil 1. Regresyon ağacı**



Regresyon ağaçlarındaki temel düşünce; tahminleyici uzayının  $R_1, R_2, \dots, R_m$  olarak adlandırılan  $M$  tane bölgeye ayrılması ve eğitim verisini kullanarak yanıt değişkeninin tahmini değerini olan  $\hat{y}_i$ 'nin her bir bölgeye düşen  $y_i$  yanıt değişkeni değerlerinin ortalaması olarak hesaplanmasıdır. Regresyon ağaçları modelinin tahminleyicisi olan  $T(x; \theta)$ 'nin hesaplanması için Eşitlik-1 kullanılmaktadır.  $\theta$ , regresyon ağaçları için bölünen değişkenler, her bir düğümdeki kesme noktaları ve varış düğüm değerlerinin parametrelerini temsil etmektedir (Breiman vd., 1984).

$$T(x; \theta) = \hat{y}_i = \sum_{m=1}^M \left( \frac{1}{N_m} \sum_{i \in R_m} y_i \right) I(x \in R_m), \quad (1)$$

$$i = 1, 2, \dots, N \text{ and } m = 1, 2, \dots, M.$$

Ağaç yetiştirme safhasında bölünen düğüm noktaları, yanıt değişkeninin eğitim verisi ve her bir düğüm noktasındaki örnek ortalamaları için aldığı değerlerin arasındaki ortalama hata kare değerindeki en yüksek azalma değerini göz önüne alarak seçilmektedir. Eğitim hatasının değeri Eşitlik-2'de verilen ve ortalama hata karenin enküçülenmesi olarak tanımlanan formül ile hesaplanmaktadır.  $N_m$ ,  $R_m$  bölgesindeki gözlem sayısını göstermektedir.  $M$ , bölge sayısını temsil etmektedir.

$$\underbrace{Enk.}_{\{R_m\}_{m=1}^M} \frac{1}{N_m} \sum_{i \in R_m} (y_i - \bar{y}_m)^2, \quad (2)$$

$$i = 1, 2, \dots, N \text{ and } m = 1, 2, \dots, M.$$

## 2.2. Yapay sinir ağları

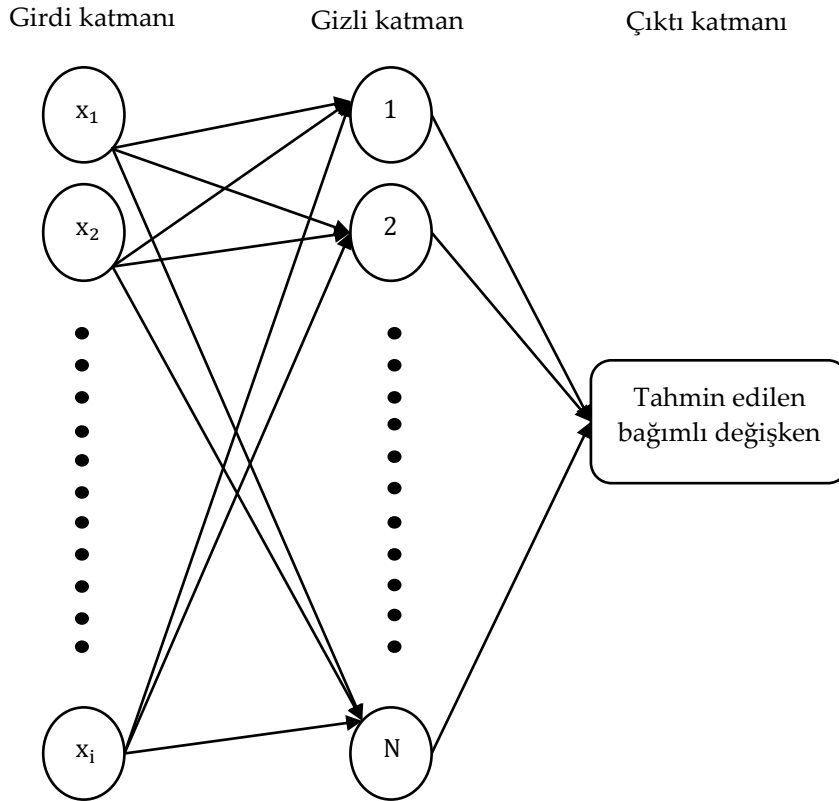
Yapay sinir ağları, biyolojik sinir sisteminin karmaşık hesaplamalı işlemleri gerçekleştirebilen matematiksel bir ifadesidir. Bu makalede, tahminleme modelinde çok katmanlı algılayıcı sinir ağları kullanılmıştır. Çok katmanlı algılayıcı sinir ağları çoklu ve

birbiriyle tam olarak bağlantılı olan girdi, gizli ve çıktı katmanlarından oluşmaktadır. Gizli ve çıktı katmanlarının içerisindeki her bir düğüm noktası nöron olarak adlandırılmakta ve nöronlar içerisinde doğrusal veya sigmoid aktivasyon fonksiyonları tanımlanmaktadır. Tek gizli katmana ve tek çıktıya sahip olan çok katmanlı algılayıcı sinir ağının temsili Şekil 2’de gösterilmektedir.

Çok katmanlı algılayıcı sinir ağlarının öğrenme işlemi,  $(t_i, p_i)$  ile tanımlanan hedeflenen ve tahmin edilen bağımlı değişken değerlerinin karşılaştırılmasını temel olarak  $w$  ağırlık değerinin ayarlanmasıyla etki altına alınmaktadır.  $x_i$  girdiler için gözlemlenen değerleri ve  $t_i$  hedeflenen çıktı değerlerini gösterirken  $T = \{(x_i, t_i): i = 1, \dots, n\}$  ile eğitim veri kümesinde  $n$  tane veri çiftini temsil etmektedir. Hedeflenen ve tahmin edilen çıktı değişkenlerinin birbirine yakın olması amaçlanmakta ve Eşitlik-3’te gösterilen doğrusal olmayan hata fonksiyonu ile kontrol edilmektedir.  $N$  gizli katman sayısını göstermekte ve  $p_i = f(x_i, w)$  tahmin edilen çıktı değişkenlerinin değeri olarak tanımlanmaktadır.

$$E(w) = \frac{1}{2} \sum_{i=1}^n [t_i - p_i]^2, \quad (3)$$

Şekil 2. Tek gizli katmana sahip olan çok katmanlı algılayıcı sinir ağı



$x_i$  bağımsız değişkenleriyle tanımlanmakta olan girdi katmanındaki her bir girdi karşılık gelen  $w_{ij}$  ağırlık değeriyle çarpımı gerçekleştirilerek her bir nöron için  $y_j$  ön aktivasyon fonksiyonları oluşturulmaktadır. Her bir nöron için belirlenen ön aktivasyon değerleri  $f_j$  doğrusal olmayan aktivasyon fonksiyonlarına atanmakta, her bir nöronun çıktısı karşılığı olan  $w_{jk}$  ağırlık değeriyle çarpılmakta ve çıktı katmanındaki nöronun  $y_k$  ön aktivasyon fonksiyonu tüm ağırlıklandırılmış değerlerin toplamı olarak elde edilmektedir.

Nihai olarak  $p_k$  hedeflenen çıktı değerleri Eşitlik-4'e göre  $f_k$  doğrusal aktivasyon fonksiyonunda ön aktivasyon değerlerinin kullanımıyla hesap edilmektedir.  $b_j$  ve  $b_k$  gizli ve çıktı katmanlarındaki her bir nöronu için sapma değerlerini göstermektedir (Haykin, 2009).

$$p_k = f_k(b_k + \sum_j f_j (b_j + \sum_i x_i w_{ij}) w_{jk}), \quad (4)$$

Dereceli azalma algoritmaları (Haykin, 2009) kullanılarak Eşitlik-3'te sunulmakta olan  $E(w)$  hata fonksiyonunu enküçükleyen ağırlık değerleri kümesinin hesaplanması mümkündür. Dereceli azalma algoritmalarında, kullanılmakta olan her bir parametreye göre türev alınarak hata fonksiyonunun enküçük değeri araştırılmaktadır. Çok katmanlı algılayıcı sinir ağlarının geri yayılım özelliğini sağlamasında kullanılan geri yayılım algoritmasıyla çıktı katmanında tanımlı olan bağımlı değişkenlerin değerlerinin tahmin edilmesi ve hata değeri minimizasyonunda saptanan ağırlık değerleri ayarlanması sağlanmaktadır. Bu makalede sayısal enküçükleme algoritmaları olan dereceli azalma yöntemi ve Gauss-Newton metodunu kombine eden Levenberg-Marquardt algoritması yapay sinir ağının eğitilmesinde kullanılmakta olup, Levenberg-Marquardt algoritmasıyla ilgili denklemler Hagan ve Menhaj (1994)'te detaylı olarak anlatılmaktadır. Levenberg-Marquardt algoritmasıyla çok katmanlı algılayıcı sinir ağındaki ağırlık değerleri Eşitlik-5 ile bulunmaktadır (Hagan ve Menhaj, 1994).  $J$  Jakobyen matrisi göstermekte iken  $\mu$  sabit bir değer olup, ağırlık değerlerinin güncellenmesi için Hessian matrise yakınsama yapılması amaçlanmaktadır.

$$w_{k+1} = w_k - (J_k^T J_k + \mu I)^{-1} J_k e_k, \quad (5)$$

### 2.3. Modelleme Doğruluğu

Yüzdesele hata (YH), karekök ortalama hata kare (KKOHK) ve belirlilik katsayısı ( $R^2$ ) tahminleme modellerinin modelleme doğruluğunu belirlemek için kullanılmaktadır. İlgili formüller Eşitlik-?, ? ve ?'de gösterilmektedir.  $\hat{y}_i$  karar değişkeninin tahmin edilen değerlerini,  $y_i$  karar değişkeninin gözlemlenen değerlerini ve  $y_m$  karar değişkeninin gözlemlenen değerlerinin ortalamasını temsil etmektedir.  $N$  test verisindeki gözlem sayısıdır.  $R^2$  regresyon modeliyle açıklanan değişkenliğin  $y_i$  değerleri bünyesindeki toplam değişkenliğe oranını göstermektedir.  $R^2$  değeri 1'e yaklaştıkça, tahminleme modelinin tahmin doğruluğu da yükselmektedir.

$$YH = \frac{1}{N} \sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{y_i}, \quad (6)$$

$$KKOHK = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - y_m)^2}, \quad (8)$$

### 2.4. Veri

Bu makalede, Borsa İstanbul Menkul Kıymetler Piyasasında işlem görmekte olan XBank bankacılık endeksinin değerlerini 02/01/2019 ile 10/11/2022 tarihleri arasındaki tarihsel sıradaki günlük değerleri tahminleme modellerinde veri olarak kullanılmaktadır. Veri setinde 965 adet veri bulunmakta olup, zaman serileri veri seti ön işlemlerden geçirilerek

hatalardan ve aykırı değerlerden arındırılmaktadır. Zaman serisi verilerindeki kronolojik sıralamayı bozmadan eğitim verilerinin oranları 70% ile 90% aralığında ve test verilerinin oranları 10% ile 30% aralığındaki oranlarda en yüksek tahminleme doğruluğu elde edilene kadar rastgele ve indekse göre denemeler yapılarak seçilmektedir. XBank bankacılık endeksinin günlük endeks verileri Borsa İstanbul'a ait olan <https://www.borsaistanbul.com/tr/sayfa/49/veriler> internet sitesinden alınmaktadır.

## 2.5. Değişkenler

Tahminleme modelinde 11 tane bağımsız değişken kullanılmaktadır. Bağımsız değişkenler; Türkiye Cumhuriyet Merkez Bankası'ndan Elektronik Fon Transferi-Elektronik Menkul Kıymet Transferi Çıkış Tutarı (TP.EFTEMKT2.TUTAR.TEC), Türkiye Cumhuriyet Merkez Bankası'ndan Perakende Ödeme İşlemleri Çıkış Tutarı (TP.OSGMPOS.TUTAR.TPC), Reeskont İşlemlerinde Uygulanan İskonto Oranı (TP.REESAVANS.RIO), Avans İşlemlerinde Uygulanan Faiz Oranı (TP.REESAVANS.AFO), Türk Lirası Gecelik Referans Faiz Oranı (TP.BISTTLREF.ORAN), Bankalar Serbest Mevduatı (TP.PPIBSM), Gün Başı Toplam Likidite (TP.PPIGBTL), Müşterilerle Yapılan Toplam Repo ve Ters Repo İşlemleri (TP.RP01.MYR + TP.RP03.MKR), Mali Kuruluşlar arası Toplam Repo ve Ters Repo İşlemleri (TP.RP02.MYT + TP.RP04.MKT), Amerikan Doları Döviz Satışı (TP.DK.USD.S YTL) ve Türkiye Cumhuriyet Merkez Bankası Net Fonlaması (TP.APIFON3) olarak adlandırılmaktadır. Bunun yanında, XBank bankacılık endeksi olarak adlandırılan bir tane bağımlı değişken de modele dahil edilmektedir. Bağımsız ve bağımlı değişkenler en küçük ve en büyük değerleriyle birlikte sırasıyla Tablo 1 ve Tablo 2'de sunulmaktadır.

**Tablo 1. Bağımsız değişkenler**

Bağımsız değişkenler (Birim) ( $x_i$ )	Enküçük	Enbüyük
TP.EFTEMKT2.TUTAR.TEC (YTL) ( $x_1$ )	23721319895.30	430585916958.66
TP.OSGMPOS.TUTAR.TPC (YTL) ( $x_2$ )	5803400381.46	124969895515.85
TP.REESAVANS.RIO (%) ( $x_3$ )	9.00	18.50
TP.REESAVANS.AFO (%) ( $x_4$ )	10.00	19.50
TP.BISTTLREF.ORAN (%) ( $x_5$ )	7.50	25.49
TP.PPIBSM (YTL) ( $x_6$ )	9715.80	216690.50
TP.PPIGBTL (YTL) ( $x_7$ )	-136527.70	180110.90
TP.DK.USD.S YTL (%) ( $x_8$ )	5.20	18.63
TP.RP01.MYR + TP.RP03.MKR (YTL) ( $x_9$ )	17530989.00	168294738.00
TP.RP02.MYT + TP.RP04.MKT (TL) ( $x_{10}$ )	15859027.00	165461359.71
TP.APIFON3 (YTL) ( $x_{11}$ )	-14992.2	625094.00

**Tablo 2. Bağımlı değişkenler**

Bağımlı değişkenler ( $y_i$ )	Enküçük	Enbüyük
XBank bankacılık endeksi ( $y_1$ )	1063.49	4844.36

### 3. Bulgular ve Tartışma

Borsa İstanbul'da işlem görmekte olan Xbank bankacılık endeksinin tahmininde iki farklı model makale kapsamında kullanılmaktadır. Önerilen modellerden birincisi yapay sinir ağları yöntemiyle oluşturulan iki aşamalı tahminleme modelidir. İkinci modelde ise regresyon ağaçları ve iki aşamalı yapay sinir ağlarını bütünleşik olarak uygulayarak oluşturulan bir tahminleme modeli geliştirilmektedir. Regresyon ağaçları algoritmasıyla özellik seçme işlemi gerçekleştirilerek tahminleme modelinde kullanılacak olan karar değişkenlerinin sadece önemli olanlarının kullanılması sağlanırken, iki aşamalı yapay sinir ağları yöntemiyle Xbank bankacılık endeksi değerinin  $n$  gün ilerisi için gelecekte alacağı tahmini değerler belirlenmektedir.

Regresyon ağaçları, en büyük bölünme sayısı, en küçük yaprak sayısı ve en küçük ebeveyn sayısı parametreleri için atanan değerlere bağlı olarak farklı derinliklerde tasarlanabilmektedir. Regresyon ağaçlarında, eğitim verisi sayısı  $n$  ile gösterildiğinde en büyük bölünme sayısı  $n-1$  değerini alırken, en küçük yaprak sayısı 1 ile 5 ve en küçük ebeveyn sayısı 5 ile 50 aralığında değerler almaktadır. Tahminleme modelinde kullanılacak olan karar değişkenlerinin sahip olduğu yüksek boyutun en önemli karar değişkenlerini belirleyerek küçültülmesinde regresyon ağaçları algoritması kullanılmaktadır.

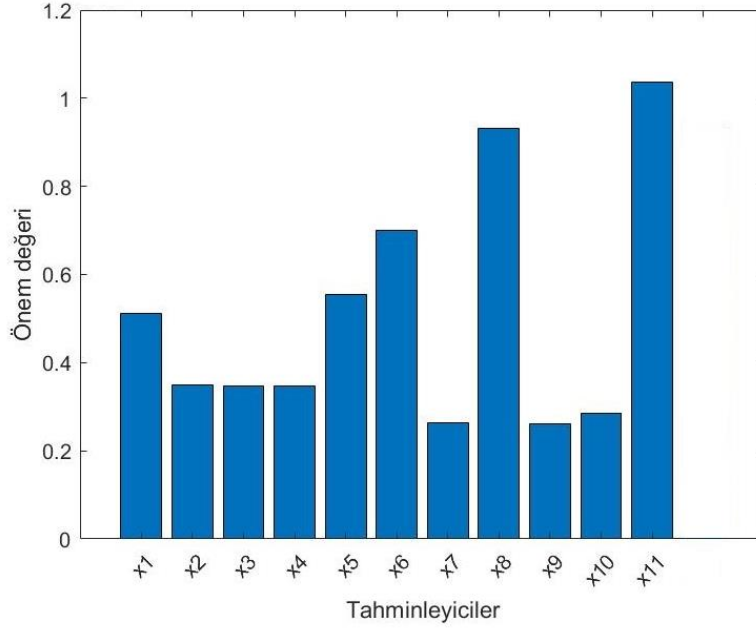
Tablo 3'te en önemli karar değişkenlerinin seçilmesi için önem değerlerinin belirlenmesinde kullanılan regresyon ağacı için kullanılan parametrelerin aldığı değerler gösterilmektedir.  $X_1$ ,  $X_5$ ,  $X_6$ ,  $X_8$  ve  $X_{11}$  karar değişkenleri en önemli tahminleyiciler olarak belirlenmektedir. Her bir tahminleyici için belirlenen önem değeri Şekil 3'te sunulmaktadır. Önem değerleri için eşik değer 0,5 olarak alınmıştır. 0,5'ten daha büyük önem değerine sahip olan karar değişkenleri en önemli tahminleyiciler olarak seçilerek tahminleme modelinde kullanılmaktadır.

**Tablo 3. Önem analizi için kullanılan regresyon ağacında uygulanan parametre değerleri**

Model adı	En büyük bölünme sayısı	En küçük yaprak sayısı	En küçük ebeveyn sayısı	Karar değişkeni sayısı	Hata fonksiyonu
Regresyon ağacı	771	1	15	12	OHK



**Şekil 3. XBank bankacılık endeksinin tahmininde kullanılacak olan tahminleyicilerin regresyon ağaçlarıyla bulunan önem değerleri**



Yapay sinir ağı modeli olarak geri bildirimli ve çok katmanlı algılayıcı sinir ağı uygulanmaktadır. Eğitim, doğrulama ve test verileriyle rastgele ve saptanmış sıralamaya göre denemeler gerçekleştirilmektedir. Yapay sinir ağının eğitim algoritması Levenberg-Marquardt algoritması olarak kullanılmaktadır. Hata fonksiyonu OHK fonksiyonudur.

Makale kapsamında iki farklı tahminleme modeli tasarlanmaktadır. Birinci modelde, yapay sinir ağının tasarımında 11 tane karar değişkeni kullanılmakta ve modeldeki tek amaç fonksiyonu olan XBank bankacılık endeksinin gelecekte alacağı değerler tahmin edilmektedir. İkinci model, regresyon ağaçları ve yapay sinir ağlarının birlikte uygulandığı melez bir tahminleme modelidir. İkinci modelleme yaklaşımındaki regresyon ağaçları algoritması önemli karar değişkenlerinin seçilmesinde kullanılmakta ve karar değişkeni sayısı azaltılarak tahminleme modelinin karmaşıklık boyutunun azaltılması amaçlanmaktadır. Birinci ve ikinci modellerdeki yapay sinir ağı tahminleme modeli birbiriyle özdeştir. Her iki tahminleme modelinde iki aşamalı yapay sinir ağı tasarımı kullanılmaktadır. Birinci aşamada veriler yapay sinir ağından geçirilerek XBank bankacılık endeksinin değerleri tahmin edilmekte ve yapay sinir ağının ikinci aşamasında birinci aşamada bulunan değerler modelde kullanılarak  $n$  gün ilerisi için XBank bankacılık endeksinin günlük kapanış değerleri tahmin edilmektedir. Önerilmekte olan her iki tahminleme modelinde yapay sinir ağları için eğitim, doğrulama ve test verileri hem belirlenmiş hem de rastgele indislerle oluşturularak denemeler yapılmaktadır. Gizli katman sayısı 1 ile 50 arasındaki değerlerde denenmektedir. Gizli katmanlarda kullanılan nöron sayısı 5 ile 300 arasındaki değerlerde atanmaktadır. Gizli katmanlardaki nöronlarda transfer fonksiyonu olarak tanjant-sigmoid ve log-sigmoid fonksiyonları uygulanırken, çıktı katmanında doğrusal transfer fonksiyonu bulunmaktadır. Eniyi tahminleme performansının bulunması için her iki yapay sinir ağında 50 kez denemeler yapılmaktadır.

XBank bankacılık endeksinin günlük kapanış değerleri tahmininde kullanılan modeller için modelleme performansının test edilmesinde elde edilen sonuçlar Tablo 4'te sunulmaktadır. Tahminleme sonuçları ve gözlemlenen değerlerin test dönemindeki değerleri

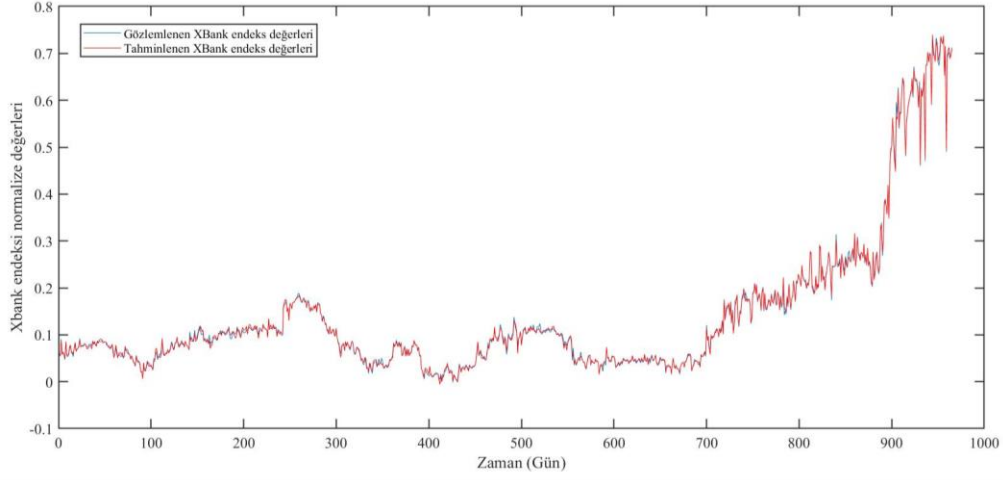
kullanılarak YH, KKOHK ve  $R^2$  değerleri hesaplanmaktadır. Tablo 4'teki sonuçlar göz önüne alındığında iki aşamalı çok katmanlı algılayıcı sinir ağı modelinin tahmin doğruluğunun daha yüksek olduğu anlaşılmaktadır. Diğer taraftan, iki aşamalı çok katmanlı algılayıcı sinir ağı modeliyle tahminleme işlemi 2.3 saniye sürerken, iki aşamalı melez regresyon ağacı-çok katmanlı algılayıcı sinir ağı tahminleme modelinde tahminleme süresi 2.5 saniyede tamamlanmaktadır. Bu durumda, her iki modelde tatmin edici ve yüksek tahmin doğruluğunda sonuçlar elde edilmesine karşın hem tahmin doğruluğu hem de işlem süresi kriterlerine göre iki aşamalı çok katmanlı algılayıcı sinir ağı modeli daha iyi tahmin sonuçları vermektedir. İki aşamalı çok katmanlı algılayıcı sinir ağı modeli verileri rassal olarak dağıtılmış sırasıyla %60, % 20 ve % 20 oranlarında kullanılmış olan eğitim, doğrulama ve test verileriyle tasarlanmakta, gizli katman sayısı 1, nöron sayısı 300, nöronlar için transfer fonksiyonu olarak tanjant-sigmoid ve çıktı katmanında doğrusal transfer fonksiyonu olarak kullanılarak eniyi sonuçları veren model elde edilmektedir.

**Tablo 4. Tahminleme modellerinin tahminleme doğruluğu**

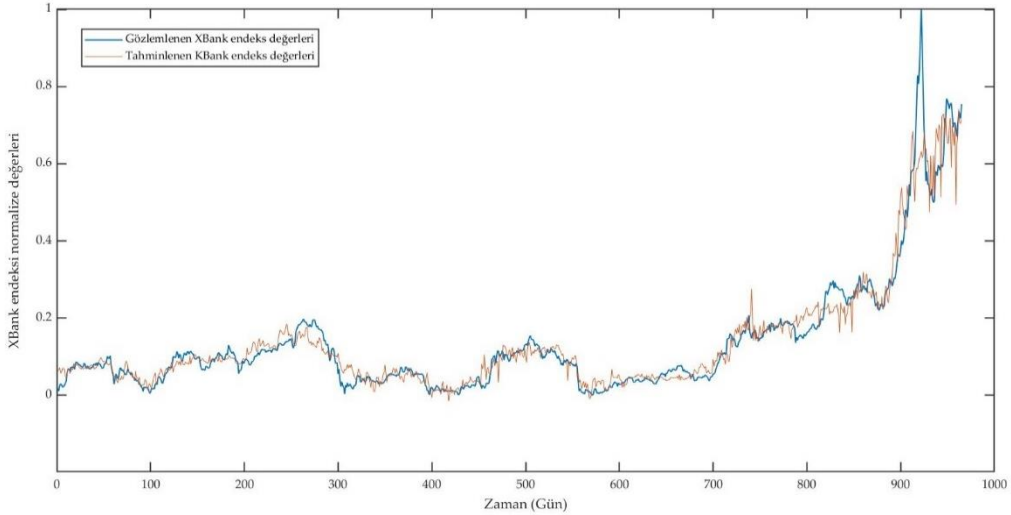
Yanıt değişkeni	İstatistiksel göstergeler	İki aşamalı çok katmanlı algılayıcı sinir ağı modeli	İki aşamalı melez regresyon ağacı-çok katmanlı algılayıcı sinir ağı tahminleme modeli
XBank bankacılık endeksinin günlük kapanış değeri	YH	0,0039	0,0266
	KKOHK	0,0042	0,0299
	$R^2$	1,00	0,99

Şekil 4 ve Şekil 5'te sırasıyla iki aşamalı çok katmanlı algılayıcı sinir ağı modeli ve iki aşamalı melez regresyon ağacı-çok katmanlı algılayıcı sinir ağı tahminleme modelleri için 965 tane gözlem gününü kapsayacak şekilde eğitim, doğrulama ve test verilerinin gözlemlenen ve tahmin edilen değerleri sunulmaktadır. Şekil 4'te verilen sonuçlara göre iki aşamalı çok katmanlı algılayıcı sinir ağı modelinin örüntü, tepe ve dip noktaları tam olarak tahminleme sonucunda bulunduğu görülmektedir. Buna karşın, Şekil 5'te gösterildiği gibi iki aşamalı melez regresyon ağacı-çok katmanlı algılayıcı sinir ağı tahminleme modelinin genel olarak örüntü ve değer tahmininde yüksek tahmin doğruluğu sağladığı, fakat bazı tepe ve dip noktalarını iki aşamalı çok katmanlı algılayıcı sinir ağı modeli gibi tam doğrulukta tahmin edemediği kanaatine varılmaktadır.

**Şekil 4. XBank bankacılık endeksinin günlük kapanış değerlerinin iki aşamalı çok katmanlı algılayıcı sinir ağı modeliyle tahmin edilen değerleri**

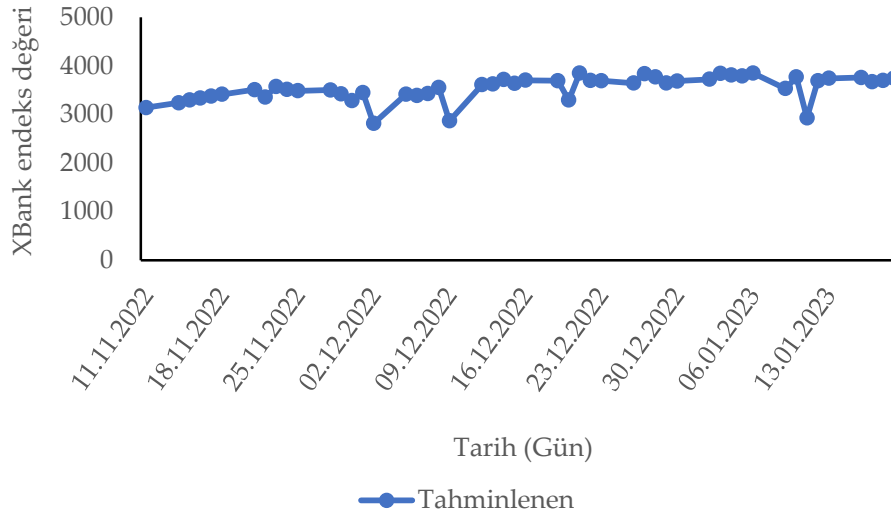


**Şekil 5. XBank bankacılık endeksinin günlük kapanış değerlerinin iki aşamalı melez regresyon ağacı-çok katmanlı algılayıcı sinir ağı tahminleme modeliyle tahmin edilen değerleri**



XBank bankacılık endeksinin günlük kapanış değerlerinin 50 gün ilerisi için tahmini değerleri, yeni modelleme yöntemi olarak tespit edilen iki aşamalı çok katmanlı algılayıcı sinir ağı modeli için Şekil-6'da gösterilmektedir. Şekil-6'daki değerler, veri setinde bulunan son gözlem tarihinden itibaren henüz gözlemlenmemiş ve gelecekte gözlemlenecek olan 50 gün için XBank bankacılık endeksinin tahmini günlük kapanış değerleridir.

**Şekil 6. XBank bankacılık endeksinin günlük kapanış değerlerinin iki aşamalı çok katmanlı algılayıcı sinir ağı modeliyle en son gözlem gününden itibaren 50 gün ilerisi için tahmin edilen değerleri**



#### 4. Sonuç

Menkul kıymetler borasında işlem gören endekslerin geleceği dönük kapanış değerlerinin tahmin edilmesinde kullanılmak üzere iki aşamalı çok katmanlı algılayıcı sinir ağı modeli ve iki aşamalı melez regresyon ağacı-çok katmanlı algılayıcı sinir ağı olmak üzere iki farklı yapay zekâ tahminleme modeli bu makalede sunulmaktadır. Önerilen tahminleme modelleri Borsa İstanbul'da işlem görmekte olan XBank bankacılık endeksi için son gözlem tarihinden itibaren 50 gün boyunca gerçekleşecek olan kapanış değerlerinin tahmin edilmesi için uygulanmaktadır. Tahminleme modellerinin tahmin doğruluğu yüzdesel hata, karekök ortalama hata kare ve belirlilik katsayısı olmak üzere 3 farklı istatistiksel gösterge ile test edilmektedir. Önemli karar değişkenlerinin belirlenmesinde seçme algoritması olarak regresyon ağaçları kullanılmaktadır. İstatistiksel göstergelerin aldığı değerlere göre, XBank bankacılık endeksinin tahmininde en doğru tahminleme sonuçları iki aşamalı çok katmanlı algılayıcı sinir ağı modeliyle elde edilmekte olup, sırasıyla iki aşamalı çok katmanlı algılayıcı sinir ağı modeli ve iki aşamalı melez regresyon ağacı-çok katmanlı algılayıcı sinir ağı yöntemleri için işlem süreleri 2.3 saniye ve 2.5 saniye olarak birbirine yakın değerlerde bulunmaktadır.

Karmaşık bir yapıya sahip olan menkul kıymetler piyasasında endeks değerlerinin tahmin edilmesinde yapay zekâ yöntemlerinden olan yapay sinir ağlarının yüksek tahmin doğruluğu, uygulama kolaylığı ve kısa işlem süreleriyle uygulanabildiği bu makalede gösterilmektedir. Gelecekte yapılacak olan çalışmalarda, yüksek tahminleme doğruluğu ve kısa işlem süreleri sağlayan melez yapay zekâ modelleme yaklaşımları tasarlanması önerilmektedir.

**Kaynakça**

- Ayodele, A., Ayo, C. K., Adebisi, M. O. ve Otokiti, S. O. 2012. "Stock Price Prediction Using Neural Network with Hybridized Market Indicators." *Journal of Emerging Trends in Computing and Information Sciences*, 3 (1): 1-9.
- Baçoğlu, U., Ceylan, A. ve Parasız, İ. 2009. *Finans: Teori, Kurum, Uygulama*. Bursa: Ekin Basım Yayın Dağıtım.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327 [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Pacific Grove: Wadsworth, OH, USA, 1984. <https://doi.org/10.1201/9781315139470>
- Ergeç, Funda. 1996. "Markov Analizi ile Hisse Senedi Fiyatının Tahmin Edilmesi." *İstanbul Üniversitesi İşletme Fakültesi Dergisi*. 25 (2): 123-151. <https://doi.org/10.19168/jyu.32384>
- Hagan, M.T.; Menhaj, M.B. Training feedforward networks with the marquardt algorithm. *IEEE Trans. Neural Netw.* 1994, 5, 989–993. DOI: 10.1109/72.329697
- Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; Pearson Prentice Hall: Hoboken, NJ, USA, 2009.
- Hsieh, D. A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *The Journal of Finance*, 46, 1839–1877. <https://doi.org/10.1111/j.1540-6261.1991.tb04646.x>
- Hsu, Chih Ming. 2011. "A Hybrid Procedure with Feature Selection for Resolving Stock/Futures Price Forecasting Problems." *Neural Computing and Applications*, 1-21.
- Jarrett, J. E. ve Schilling, J. 2008. "Daily Variation and Predicting Stock Market Returns for the Frakfurter Börse (Stock Market)." *Journal of Business Economics and Management*, 9 (3): 189-198.
- 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: 5311-5319. <https://doi.org/10.1016/j.eswa.2010.10.027>
- Kim, M.H.; Song, H.B. Analysis of the global warming potential for wood waste recycling system. *J. Clean. Prod.* 2014, 69, 199-207. <https://doi.org/10.1016/j.jclepro.2014.01.039>
- Leung, M. T., Daouk, H. ve Chen, A. S. 2000. "Forecasting Stock Indices: A Comparison of Classification and Level Estimation Models." *International Journal of Forecasting*, 16: 173-190. <https://EconPapers.repec.org/RePEc:eee:intfor:v:16:y:2000:i:2:p:173-190>
- Olaniyi, S. A. S., Adewole, K. S. ve Jimoh, R. G. 2011. "Stock Trend Prediction Using Regression Analysis- A Data Mining Approach." *Journal of Systems and Software*, 1 (4): 154-157.
- Ou, P. ve Wang, H. 2009. "Prediction of Stock Market Index Movement by Ten Data Mining Techniques." *Modern Applied Science*, 12 (3): 28-42. DOI:10.5539/mas.v3n12p28
- Rao, T., & Gabr, M. (1984). *Introduction to bispectral analysis and bilinear time series models*. Lecture notes in statistics (Vol. 24).
- S. Borovkova, I. Tsiamas, An ensemble of LSTM neural networks for highfrequency stock market classification, *J. Forecast.* 38 (2019) 600–619, <https://doi.org/10.1002/for.2585>.
- S. Kumar, D. Ningombam, Short-term forecasting of stock prices using long short term memory, in: 2018 International Conference on Information Technology (ICIT), 2018, pp. 182–186, <https://doi.org/10.1109/ICIT.2018.00046>.

- Tektaş, A. ve Karataş, A. 2004. "Yapay Sinir Ağları ve Finans Alanına Uygulanması: Hisse Senedi Fiyat Tahminlemesi." Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 3 (4): 337-349. Retrieved from <https://dergipark.org.tr/tr/pub/atauniiibd/issue/44281/454444>
- Vasanthi, S., Subha, V. ve Nambi, S. T. 2011. "An Empirical Study On Stock Index Trend Prediction Using Markov Chain Analysis." Journal on Banking Financial Services and Insurance Research, 1 (1): 72-91.
- Yin, C.; Rosendahl, L.A.; Kaer, S.K. Grate-firing of biomass for heat and power production, Prog. Energy Combust. Sci. 2008a, 34, 725–754.  
<https://doi.org/10.1016/j.pecs.2008.05.002>.