

YAPAY ZEKA YÖNTEMLERİ İLE HİSSE SENEDİ FİYAT ÖNGÖRÜLERİ

Stock Price Predictions Using Artificial Intelligence Methods

Efe ARDA* & Güray KÜÇÜKKOCAOĞLU**

Öz

Finansal varlık fiyatlarının geleceğinin tahmin edilmesi literatür ve uygulamada uzun zamandır ilgi çeken bir konudur. Son yıllarda, borsaya kote şirketlerin hisse senetlerinin fiyat hareketleri öngörme ve geleceğe dönük değerlerini tahmin etme hedefi için yapay zeka algoritmalarının başarılı yöntemler sundukları farklı akademik çalışmalarca ortaya konulmuştur. Belirtilen akademik çalışmaların büyük çoğunluğu yurt dışında bulunan piyasalarda yapılmıştır. Bu durumun geçerliliğini BIST 30 endeksi hisselerinde test etmek için bu çalışmada yedi farklı yapay zeka algoritması programlanmış, 30 hissenin 2014-2016 yılları günlük kapanış fiyatı verileri ile algoritmalar eğitilmiş ve bir firma için üretilen kapanış değerleri tahminleri gerçekleşen değerlerle kıyaslanmıştır. Veri seti için 02/01/2014 ve 30/12/2016 tarihleri arasında işlem yapılan 755 iş günü kullanılmıştır. Kullanılan öğrenme sürelerinin performans üzerindeki etkilerini görmek için öğrenme/tahmin oranları %80/20, %90/10, %99/1 olarak belirlenen üç farklı deney yapılmıştır. Çalışmanın sonucunda doğrusal regresyon temelli algoritmaların BIST30 hisse senedi fiyat hareket yönünü tahmin etmede, nöral ağ ve Poisson regresyonu yöntemlerinin ise kapanış fiyatı değerini tahmin etmede etkili oldukları görülmüştür.

Anahtar Kelimeler:

Finansal Zaman Serileri, Yapay Zeka, Makine Öğrenmesi, Hisse Senedi Fiyat Tahmini, Regresyon.

JEL Kodları:

C45, C53, D53, E44

Abstract

Forecasting the future of stock market prices has been an interesting topic for researchers and professionals for a long time. Lately, numerous academic papers have shown that artificial intelligence algorithms generate some successful forecasts for stock prices. Most of these referenced research papers are conducted in markets outside Turkey. To test this hypothesis in BIST 30 index companies, seven different artificial intelligence algorithms have been programmed and trained with a dataset of daily closing prices between 2014 and 2016. The dataset consists of 755 market days starting 02/01/2014 and ending 30/12/2016. Then, forecasted numbers have been compared to actual prices for one particular stock. To see the effects of amount of learning days used to performance, 3 experiments have been conducted with learning to prediction ratios of %80/20, %90/10 and %99/1. In conclusion, it is seen that linear regression-based algorithms perform well to predict the price movements while neural network and Poisson regression algorithms perform well to predict closing price values for BIST 30 stocks.

Keywords:

Financial Time Series, Artificial Intelligence, Machine Learning, Stock Price Prediction, Regression.

JEL Codes:

C45, C53, D53, E44

* Dr., Ericsson, efe.arda@ericsson.com, ORCID: 0000-0002-6117-1909

** Prof. Dr., Başkent Üniversitesi, İİBF, İşletme Bölümü, gurayk@baskent.edu.tr, ORCID: 0000-0001-6170-3269

1. Giriş

Finans alanında arařtırmacıların ve uygulayıcıların karşılařtığı en zorlu problem belirsizliktir. Belirsizlik, finansal teorinin ayrılmaz bir parçası olan risk faktörünü de beraberinde getirmektedir. Finans piyasaları; yerel ve global ekonomik, siyasal, politik ve bunun gibi birçok faktörden etkilenmektedir. Tüm bu faktörler piyasanın geleceğini tahmin etmeyi zorlařtırmaktadır. Finans alanında çalışanların, arařtırmacıların ve yatırımcıların ana hedefi piyasadaki durumun gelecekte nasıl olacağı hakkında bilgi sahibi olmaktır. Francis Edgeworth ve Vilfredo Pareto tarafından geliştirilen etkinlik kavramı, hisse senedi fiyatlarının piyasadaki mevcut bilgileri yansıması olarak tanımlanır. Etkin piyasalar hipotezine göre bilgi etkinliğinin sağlandığı durumlarda yatırımcılar piyasa ortalamasının üzerinde kazanç elde edemezler (Fama, 1970). Menkul kıymet fiyat deęişimleri, Louis Bachelier’in rassal yürüyüş hipotezinde açıklandığı üzere rassal olarak gerçekleşir ve tahmin edilemez (Bachelier, 1900). Bu çalışma belirtilen hipotezlere belirli eleştirilerde bulunmaktadır.

Rassal yürüyüş hipotezinin üzerine kurulu olduğu “rasgelelik” kavramı yüzyıllardır tartışılan, epistemik erişim yetersizliği ve ontolojik bir özellik olarak ele alınan felsefi bir sorudur (Hromkovic, 2005; Rosenberg ve McIntyre, 2011). 21. Yüzyılda teknolojik ilerlemelerle birlikte finansal veriler bilgisayarlar tarafından yaratılmaktadır. Bilgisayar bilimleri için ise rasgele bir deęişken yaratmak tam olarak mümkün deęildir ve genelde istatistiksel olarak rassallığı taklit eden “rassal varsayılan” (psuedo random) algoritmalar kullanılmaktadır. Bu algoritmalar, insanın bilgiyi işleme ve saklama kapasitesinin sınırlı olması sebebi ile insan algısı için rassal sonuçlar üretmektedir. Aynı sonuçlar işlem ve hafıza kapasitesi insandan çok daha üstün olan bilgisayar üzerinde çalışan yapay zeka algoritmalarına verildiğinde rasgele görünümünün altında yatan kalıplar ortaya çıkmakta ve tahmin edilebilirlikleri artmaktadır (Kolmogorov ve Uspensky, 1987). Bu durum rassallığın bir epistemik erişim yetersizliği olduğu görüşünü savunmaktadır. Rastlantısal olarak tanımlanan birçok durum, aslında eldeki bilgi ve bilgi işleme hızının yetersiz kaldığı durumlardır. Bu sebeple finansal verilerde “Rassal gibi davranan fakat detaylı incelendiğinde rassal olmayan” durumlar söz konusu olabilmektedir.

Menkul kıymet fiyatlarının tahmini ile ilgili akademide kabul görmüş bir dięer hipotez de Etkin Piyasalar Hipotezi’dir. Eugene Fama’nın öne sürdüğü Etkin Piyasalar Hipotezince piyasaya açık ve hatta açık olmayan (insider trading) bilgilerin arbitraj yaratma amacı ile kullanılması mümkün deęildir, belirtilen bilgiler halihazırda fiyat bilgisine dahildir (Fama, 1970). Etkin piyasalar hipotezi ekonomi ve finans alanlarında ampirik veriler ile en fazla desteklenen hipotezlerdendir. Etkin piyasalar hipotezi New York Borsası, Avustralya Borsası ve Londra Borsası gibi dünyanın en büyük borsalarında yapılan akademik çalışmalar ile test edilmiş ve güçlü etkinlik durumu büyük çoğunlukla onaylanmıştır (Baciu, 2012; Guerrien ve Gun, 2011; Millionis ve Moschos, 2000). Fakat belirtilen hipotez piyasa etkinliğinin maksimum olduğu seviyelerde geçerlidir. Günümüz menkul kıymet piyasalarında etkinlik seviyesi zayıf / orta güçlü / güçlü gibi farklı deęerler alabilmektedir. İstanbul Menkul Kıymetler Borsası üzerinde yapılan çalışmalarda zayıf form piyasa etkinliği görülmektedir (Çelik ve Taş, 2007; Çelik, Kurtaran ve Kurtaran, 2018; Özdemir, Atan ve Atan, 2016). Bu durumda ilginin fiyata yansımaları, spekülasyon ve benzeri sebeplerden ötürü gecikebilir ve kısa süreli de olsa bir arbitraj yaratılmış olur (Fama, 1970). Etkin piyasalar hipotezi 1970’te öne sürülmüştür ve geçen yarım asırlık süre içinde hem finansal enstrümanlar hem de telekomünikasyon ve bilgisayar alanındaki gelişmeler ile yatırım yapma işlemi birçok deęişiklikten geçmiş, bu süre zarfında

hipotezin öne sürdüğü kořulların sađlanmadığı durumlar görölmüřtür. Hipotezin öne sürdüğü gibi tüm firmalar hakkındaki bilgilerin řeffaf ve verimli bir řekilde halka açık olması durumunda Enron, MCI WorldCOM ve Lehman Brothers gibi bir gecede 60 milyar dolar üzerinde deđer kayıplarının yařanması mümkün olmamalıdır (Benston ve Hartgraves, 2002; Leetaru, 2008; Swednerg, 2010). Ayrıca hipotez yatırımcıların rasyonel olduklarını ve riske uyarlanmış getiri beklediğini varsayar fakat firma ismi ve büyüklüğü gibi sebeplerle hareket eden, rasyonel olmayan yatırımcılar da piyasada aktif olarak görölmektedir (Bayraktar, 2012).

Regresyon veya sınıflandırma hedefinde olan işlemlerde 21. yüzyılın en kuvvetli araçlarından biri yapay zeka algoritmalarıdır. Finansal zaman serileri öngöröleri yapay zeka için birer regresyon problemidir. Mevcut piyasa durumunu teknik analiz yöntemleri ile inceleyip mevcut durumun “al” ya da “sat” durumu olarak kategorize edilmesi ile sınıflandırma problemi olarak da tanımlanabilir fakat belirtilen durum bu çalışmanın kapsamında değildir. Yapay zeka alanı ve sunduğu yöntemler finans alanına kıyasla yeni yöntemler oldukları için uygulamaları henüz birçok disiplinde kullanılmaya başlanmamış veya yüksek performans evresine geçmemiştir. Finans alanında yapay zeka öngöröleri üzerine yapılan çalışmalar incelendiğinde öne çıkan çalışmaların son 20 yılda yapılmış oldukları görölmektedir (Enke ve Thawornwong 2005; Lee ve Ko 2009; Marcek, Marcek, ve Babel 2009; Theofilatos, Karathanasopoulos, Sermpinis and Amorgianiotis, 2012; Yan, Wang, Yu ve Li 2005). Çalışmalar finansal piyasalarda farklı kořullarda etkin performans gösteren farklı algoritmalar öne sürmektedir. Belirtilen çalışmalar BIST endeksleri üzerinde test edilmemiş olup, bu çalışma bu eksikliği doldurmaya yardımcı olma hedefindedir.

Bu motivasyon ile bu çalışmanın hipotezi halka açık geçmiş piyasa verilerini kullanarak hisse senedi fiyatlarını etkin olarak tahmin edebilecek yapay zeka algoritmalarının tasarlanabileceğini göstermektir. H_0 hipotezi yapay zeka tahminlerinin etkili olmayacağı olarak tanımlanmaktadır. Tahminlerin geçerliliğini ortaya koymak için tahmin edilen deđerler gerçekleşen deđerlerle kıyaslanmıştır, ortalama hata deđerleri, ortalama hata yüzdeleri ve R^2 deđerleri incelenmiştir.

2. Literatür

Yapay zeka yöntemlerini kullanarak hisse senedi tahminleri yapma konusunda çok sayıda ampirik çalışma bulunmaktadır. Tsai ve Wang'ın (2009) nöral ađ ve karar ađacı modellerini kullanan algoritmaları Tayvan piyasasında bir hisse senedi için %77 oranında dođru öngörü yapılabildiğini göstermiştir. Shen, Jiang ve Zhang (2012), çalışmalarında bir sonraki günkü hisse durumunu NASDAQ için %74.4, S&P için %77.6 ve DJIA için %77.6 oranında dođru tahmin edebildiklerini göstermişlerdir. Krollner, Vanstone ve Finnie (2010) ise finansal zaman serileri öngöröleri için farklı makine öğrenmesi algoritmalarını kıyasladıkları çalışmalarında nöral ađ yönteminin en etkin olduğunu öne sürmüşlerdir. Arařtırma sürecinde BIST hisse senedi verileri ile yapılan çalışmaların noksanlığı öne çıkmıştır.

Bu çalışmada yedi farklı algoritma programlanmış, her algoritma için ampirik literatür Regresyon Modelleri bölümünde sunulmuřtur. Regresyon modelleri, makine öğrenmesi yönteminin bir adımı olduđu için öncesinde makine öğrenmesi yöntemi ve konunun ana başlığı olan yapay zeka alanı açıklanmıştır.

2.1. Yapay Zeka

“Yapay Zeka” terimi ilk kez 1956 yılında John McCarthy tarafından konu hakkındaki bir akademik konferansta kullanılmıştır. Makinelerin düşünmesi konsepti ise bundan da önceki çalışmalara dayanmaktadır. Vannevar Bush’un 1945’te yayımladığı “As We May Think” isimli çalışması insan zeka ve algısını kuvvetlendiren yardımcı sistemleri barındırır. 5 yıl sonrasında yapay zeka tarihinin öne çıkan isimlerinden Alan Turing, insanları taklit eden ve satranç gibi insan zekası gerektiren oyunlarda karar alabilen makinelerden bahsetmiştir. (Smith, McGuire, Huang ve Yang, 2006) Bilgisayar bilimlerinin bir alt dalı olan yapay zeka başlığı altında son 60 yıl içinde arama algoritmaları, makine öğrenmesi yöntemleri ve dil analizi gibi onlarca alt dal ortaya çıkmıştır. Yapay zeka bir çok insanın beklediğinin aksine konuşan makineler veya Mars’a uzay gemisi göndermekten çok; kişiye özel pazarlama, arama motorları gibi fark edilmeyen işlevlerde kullanılmaktadır. (Adams, 2004)

2.2. Makine Öğrenmesi

Makine öğrenmesi, bu iş için kullanılan donanım ürünlerini üreten NVIDIA firması tarafından “Verileri ayrıştırmak, onlardan öğrenmek ve ardından dünyadaki bir durum hakkında bir belirleme veya tahmin yapmak için algoritmalar kullanma işlemi” olarak, Stanford Üniversitesi tarafından “Bilgisayarlara sabit kurallar ile programlanmadan karar alma yetkisi kazandırma bilimi”, Amerikan danışmanlık firması Mckinsey & Co tarafından “Kural tabanlı programlama olmadan veriden öğrenme sağlayan algoritmalar.”, Carnegie Mellon Üniversitesi tarafından ise “Tüm öğrenme işlemine etki eden yasaları araştıran ve kazandığı tecrübe ile otomatik olarak gelişen algoritmaları inceleyen bilgisayar bilimi” olarak tanımlanmaktadır. Bu tanımlardaki önemli noktaları birleştirerek “Makine Öğrenmesi, bilgisayarların insan gibi öğrenme ve davranmaları için, gözlem ve gerçek dünya etkileşimi verileri ile besleyerek, özerk bir şekilde performanslarını zaman içinde geliştirmelerini sağlama bilimidir” denilebilir. (Arda, 2020)

Algoritmalarda kullanılan makine öğrenmesi yöntemi yedi adımdan oluşan bir işlem olarak açıklanabilir. İşlem süresini makul tutmak adına öğrenme işlemi 30 firma hisse senedi kapanış fiyatlarını kullanarak gerçekleştirilmiş, sonrasında rasgele seçilen bir firmanın hisse senedi kapanış fiyatları tahmin edilmiştir. Kullanılan yedi adım, bu çalışmada uygulandığı şekilde aşağıda açıklanmıştır:

1. Veri Toplama: Bilgisayarın öğrenebileceği verileri toplama işlemidir. Çalışma için Borsa İstanbul Datastore tarafından sağlanan, 02.01.2006 – 30.12.2016 arası gerçekleşen BIST30 kapanış değerleri kullanılmıştır.

2. Veri Hazırlama: Öğrenme süresini makul tutmak ve öğrenmeyi güncel veriler üzerinde gerçekleştirmek adına verilerin tamamı değil 02.01.2014 - 30.12.2016 aralığı kullanılmıştır. Verilerde eksik / hatalı değer var ise bu bölümde düzeltilir.

Sonrasında veri öğrenme ve değerlendirme verisi olarak ikiye ayrılır:

Öğrenme Verisi: Öğrenme işlemi için bilgisayara verilen veri kümesi. Çalışma kapsamında öğrenme verisi yüzdeleri %80, %90, ve %99 olarak kullanılmıştır. “Uzun Vade Tahminleri” deneyinde %80, “Orta Vade Tahminleri” deneyinde %90, “Kısa Vade Tahminleri” deneyinde %99 değeri kullanılmıştır.

• Deęerlendirme Verisi: Öğrenme performansını ölçmek için bilgisayara daha önce verilmemiş deęerlerden oluşan veri kümesidir. Çalışma kapsamında deęerlendirme verisi yüzdeleri %20, %10 ve %1 olarak kullanılmıştır. “Uzun Vade Tahminleri” deneyinde %20, “Orta Vade Tahminleri” deneyinde %10, “Kısa Vade Tahminleri” deneyinde %1 deęeri kullanılmıştır.

3. Model Seçme: Veri setinden hangi verilerin alıp ne şekilde kullanılacağına karar verilen bölümdür. Finansal zaman serilerinde kullanılan modeller regresyon algoritmalarıdır. Bu çalışmada kullanılan modeller doğrusal regresyon, Bayes doğrusal regresyon, karar ağacı regresyonu, destekli karar ormanı regresyonu, hızlı orman yüzdelik regresyonu, nöral ağ regresyonu ve Poisson regresyonudur.

4. Öğretme: Bilgisayarın önündeki örneğin ne olduğunu öğrendiđi adımdır. Bilgisayar öğrenebilmek için elindeki verileri kullanarak bir karar doğrusu / eğrisi (modele göre) yaratır. Her örnek için bu doğru / eğriyi günceller. Sonrasında bu test verileri matris şeklinde kullanılarak ağırlık (weight) ve kayma (bias) deęerleri üretilir. Belirtilen ağırlık ve kayma deęerleri başlangıçta rasgele seçilir. Her yeni örnek ile ağırlık ve kayma deęerleri güncellenir. Bu işleme “öğrenme döngüsü” denir. Bu döngü ne kadar fazla yapılırsa o kadar fazla veride öğrenme işlemi gerçekleştirilir ve performans o kadar yükselir. Bu durum bir doyum noktasına ulařana kadar sürer. Doyum noktasına ulařtıktan sonra ise aşırı öğrenme (over-fitting) denilen durum ortaya çıkar ve algoritma olmayan kalıp ve özellikler bulmaya zorlanacağı için performans düşer.

5. Deęerlendirme: Öğretme işlemi bittikten sonra algoritmanın performansı deęerlendirme verileri ile test edilir. Deęerlendirme bölümünde kullanılan veriler öğretme bölümünde kullanılmamış olmalıdır. İkinci adımda ayrılan veriler bu bölümde kullanılır.

6. Parametre Ayarlaması: Deęerlendirme bölümünde elde edilen sonuç, tamamen model seçmede kullanılan parametrelere baęlıdır. Deęerlendirme evresinden sonra istenilen performans seviyelerine ulařmak için parametrelerde ayarlama yapmak gerekebilir.

7. Tahmin: Tüm eğitim adımları tamamlandıktan sonra bilgisayar önüne gelen gerçek dünya örnekleri üzerinde tahmin yapmaya hazırdır.

Bilgisayarlarda öğrenme, gözetimli öğrenme (supervised learning) ve gözetimsiz öğrenme (unsupervised learning) olarak ikiye ayrılır. Finans alanının sunduđu veri seti ve hedeflenen yatırım kriterleri gözetimsiz öğrenme yöntemleri için daha uygundur (Alpaydın, 2010).

2.3. Regresyon Modelleri

Bu bölümde “model seçme” adımıında açıklanan regresyon modelleri, matematiksel alt yapıları ile sunulmuştur. İlk olarak doğrusal regresyon modeli, sonrasında aynı modeli bir dağılım grafiđi olarak inceleyen Bayes doğrusal regresyon modeli kullanılmıştır. Sonrasında satranç algoritmalarında kullanılan ağaç tipi algoritmalar incelenmiş, karar ağacı (decision tree), destekli karar ormanı (boosted decision forest) ve hızlı orman yüzdelik dağılım (fast forest quantile) regresyon algoritmaları denenmiştir. Altıncı olarak güncel çalışmalarda etkinliđi gösterilen nöral ağ regresyonu programlanmıştır. Son olarak, finans alanında kullanılan veriler normal dağılıma uymadığı için Poisson regresyonu kullanılmıştır.

2.3.1. Doğrusal Regresyon

Doğrusal regresyon modeli, bağımlı değişkeninin (y), bir dizi açıklayıcı değişken (x) ile çarpılan ağırlıkların (β) doğrusal bir birleşimi olduğunu varsayar. Formül aynı zamanda tesadüfi örnekleme gürültüsünü hesaba katan bir hata terimi (ε) içerir. Örnek olarak iki bağımlı değişken içeren bir sistem için model aşağıdaki gibidir:

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \varepsilon \quad (1)$$

Doğrusal regresyon modeli matris denklemlerini kullanarak herhangi bir sayıda bağımsız değişken için ifade edilebilir (Draper ve Smith, 1998).

$$y = \beta^T X + \varepsilon \quad (2)$$

Doğrusal regresyon, anlaşılabilirliği ve finans alanında sıkça kullanılması sebebi ile kullanılacak ilk model olarak seçilmiştir. Normal dağılımlı veriler için uygun olan doğrusal regresyonun ortalama değerler üzerine kurulu olduğu için yüksek ve alçak değerlerin birbirini dengelediği uzun dönem tahminlerinde etkinliği referans gösterilen çalışmalarda gözlemlenmiştir (Altay ve Satman, 2005; Goia, May, ve Fusai, 2009; Heshmaty ve Kandel, 1985).

2.3.2. Bayes Doğrusal Regresyon

Bayes doğrusal regresyon koşullu olasılık teoreminin doğrusal regresyon modeline uygulanması ile elde edilen bir regresyon modelidir. Bayes bakış açısında bir önceki modelde ($y=$) şeklinde tanımlanan nokta tahminlerinden ziyade, olasılık dağılımlarını kullanarak doğrusal regresyon formüle edilir. Yani y 'nin, tek bir değer olarak tahmin edilmediği, ancak olasılık dağılımından alındığı varsayılır.

$$y \sim N(\beta^T X, \sigma^2 I) \quad (3)$$

Bayes Doğrusal Regresyonun amacı, bağımlı değişken için tek bir “en iyi” değer bulmak değil, bağımlı değişkeni değerinin olasılık dağılımını belirlemektir. Sonuç bir olasılık dağılımı belirtmekle birlikte model parametreleri de bir dağılım olarak girdi alınır. Bu dağılımlı model aşağıdaki şekilde gösterilebilir (Yan ve Su, 2009):

$$P(\beta|y, X) = \frac{P(y|\beta, X) * P(\beta|X)}{P(y|X)} \quad (4)$$

Kullanılan model bir doğru değil dağılım olduğu için, çıktısı da bir dağılım grafiği olur.

Model için yapılan araştırmalarda yurt dışı piyasalarında yapılan çalışmalar kullanılmıştır. (Koop, 2003; Koop ve Korobilis, 2009; Pole, West ve Harrison, 1994). Konu hakkında Türkiye piyasalarında yapılan çalışmaların noksanlığı göze çarpmaktadır. Şıklar’ın çalışmasında Bayes yaklaşımının regresyon modeli için kullanımını anlatsa da uygulamalı örnek bulunamamıştır ve bu çalışmanın da tüm referans kaynakları yurt dışında yapılan çalışmaları işaret etmektedir. (Şıklar, 1999)

2.3.3. Karar Ađacı Regresyonu

Yapay zekanın en kuvvetli olduđu oyunlardan biri olan satranç oyunu algoritmalarının başarısı bu yöntemi kullanmaya ilham sağlamıřtır. Satranç algoritmaları, satrançtaki her hamle sonrası oluřan tahtayı bir bođum olarak ifade ederek strateji üretme amaçlı karar ađaçları üretmektedir. Oyunda gerçekteşebilecek olası hamle sayısı Shannon serisince üstsel olarak arttıđı için tüm olası hamleleri birer bođum olarak ifade etmek mümkün deđildir (Romashchenko, Shen ve Vereshchagin, 2000). Stratejik olarak mantıklı bođumlar seçilir ve bu bođumların alt dalları olan bođumlar mantıklı hamleler olacađı için işleme gücü ve hafıza gibi kaynaklar yüksek çođunlukla bu alt bođumlar üzerinde kullanılır. Aynı mantık finans alanında regresyona etki eden faktörlerin önemli bođumlar olarak seçildiđi bir sistemle uygulanabilir düşüncesi ile karar ađacı regresyon algoritmaları bu çalışmada test edilmiştir.

Karar ađacı, tahmin yapmak için belirli özelliklerden karar verdirici çıkarımları yapan gözetimsiz bir makine öğrenmesi yöntemidir. Modelin adından da anlaşıldıđı gibi bu model veriyi parçalayıp bu parçaların özelliklerini kullanarak öngörü yapar. Tümevarımsal bir öğrenme yöntemidir.

Çalışmanın eğitim setindeki özelliklere dayanarak karar ađacı modeli, örneklerin sınıf etiketlerini anlamak için bir dizi soru öğrenir. Yorumlanabilirliđin önemli olduđu durumlarda karar ađacı kullanışlı bir modeldir. Kategorik hedefler yerine reel sayı ve aktivasyon fonksiyonları kullanılarak karar ađaçları regresyon için de kullanılabilir (Xu, Watanachaturaporn, Varshney ve Arora, 2005).

Karar ađacı konusunu içeren, referans gösterilen çalışmalar yazılan algoritma için kullanılmıştır (Jordan, Ghahramani ve Saul, 1997; Meek, Chickering ve Heckerman, 2002). BIST 30 üzerine yapılan çalışmalar incelendiđinde, 1990-2010 verileri üzerinde sınıflandırma ve regresyon ađaçları modeli uygulanan bir çalışma bulunmuştur fakat bu çalışmanın amacı fiyat öngörülerini yapmak deđil algoritmanın sınıflandırıcı özelliklerini kullanarak regresyon için önemli nitelikleri öne çıkarmaktır (Dondurmacı ve Çınar, 2014).

2.3.4. Destekli Karar Ormanı Regresyonu

Destekli karar ormanı regresyonu; Stokastik gradyan artırılmalı hesaplama yaklařımı, TreeNet ve MART (Multiple Additive Regression Trees) isimleriyle de bilinmektedir. Bu teknik 21.yüzyılın başlarında öngörülü veri madenciliđi için en güçlü yöntemlerden biri olarak ortaya çıkmıştır. Aslında sınıflandırma problemleri için yazılmış olan bu algoritma aynı zamanda regresyon için de kullanılmaktadır.

Destekli karar ormanı regresyonu baz alınan modeldeki alt sekanslarda bulunan kararları birleřtirerek tahmin yürüten ilaveli bir modeldir. Genel fikir, ardışık her bir ađacın, önceki ađacın tahmini kalıntılarından yaratıldıđı, basit ađaçların kombinasyonu olan bir dizi hesaplamaktır. Bu model matematiksel olarak bir “g” fonksiyonunun daha basit “f” fonksiyonları toplamı olarak yazılması şekliyle gösterilebilir:

$$g(x) = f_0(x) + f_1(x) + f_2(x) + \dots \quad (5)$$

Rastlantısal karar ađacı algoritmasından farklı olarak, her biri bir alt veri örneđi kullanarak destekli karar ormanı regresyonunda gradyan artırma adı verilen belirli bir model oluřturma tekniđi kullanır.

Örnek olarak $F(x)$ fonksiyonunu optimize etmek istediğimizi var sayalım. Bu durumda gradyan artırma algoritması yinelemeli olarak “ η ” adım için aşağıda belirtilen modeli hesaplayacaktır:

$$x_{t+1} = x_t - \eta \left. \frac{\partial f}{\partial x} \right|_{x = x_t} \quad (6)$$

Bu gradyant artırımını “ g ” fonksiyonuna uygulayıp her adımda yaşanan ampirik kaybı $L(y_i, g(x_i))$ olarak ifade edildiğinde modelin tamamı aşağıdaki şekilde gösterilebilir:

$$f_t = \arg \min_f \sum_{i=1}^N \left[\left. \frac{\partial L(y_i, g(x_i))}{\partial g(x_i)} \right|_{g = g_t} - f(x_i) \right]^2 \quad (7)$$

Ağaçların bu tür "ilave ağırlıklı genişlemeleri", açıklayıcı değişkenler ve ilgilenilen bağımlı değişken arasındaki ilişkilerin spesifik niteliği çok karmaşık olsa bile, nihayetinde öngörülen değerlerin gözlenen değerlere iyi bir şekilde uyduğu gösterilebilir. Bu nedenle, gradyant artırma yöntemi oldukça genel ve güçlü bir makine öğrenme algoritmasıdır (Poyarkov, Drutsa, Khalyavin, Gusev ve Serdyukov, 2016).

İncelenen akademik çalışmalarda destekli karar ormanı modellerinin özellikle doğrusal olmayan fonksiyonlarda başarılı olduğu gözlemlenmiştir. Algoritmanın bu çalışmada kullanılması, doğrusal ve doğrusal olmayan regresyon yöntemlerinin kıyaslanması amacı ile önemlidir (Johnson ve Zhang, 2014; Rokach, 2016; Siringnana ve Spiliopoulos, 2017). Türkiye’de bulunan piyasalar üzerine yapılan çalışmalar incelendiğinde Güdelek’in (2019) çalışmasında bu yöntemin “al-sat” kararlarını %60 oranında doğru tahmin edildiği görülmektedir.

2.3.5. Hızlı Orman Yüzdellik Regresyonu

Aynı Bayes doğrusal regresyon gibi yüzdellik regresyon (quantile regression) tek bir ortalama tahmin değeri yerine tahmin edilen değerlerin dağılımı hakkında bilgi edinmek için yararlı bir regresyon yöntemidir. Bu yöntem fiyat tahmini, öğrenci performans tahminleri gibi birçok alanda kullanılmaktadır. İncelenen çalışmalarda 2001 yılından beri kullanılan bu algoritmanın parametrik olmayan ve çok sayıda yordayıcı değişken içeren durumlarda etkin olduğu görülmektedir (Hao ve Naiman, 2007; Meinshausen, 2006; Yu ve Moyeed, 2001).

Doğrusal regresyon modelleri, tek bir tahmin kullanarak bir sayısal değişkenin değerini tahmin etmeye çalışırken, Bayes ve yüzdellik regresyon modelleri hedef değişkenin aralığını veya tüm dağılımın tahmin etmek için kullanılmaktadır. Ağaç temelli dağılım regresyon modelleri, parametrik olmayan dağılımları tahmin etmede kullanılabilecekler ek avantajlara sahiptir.

Kullanılan matematiksel model aşağıdaki şekildedir:

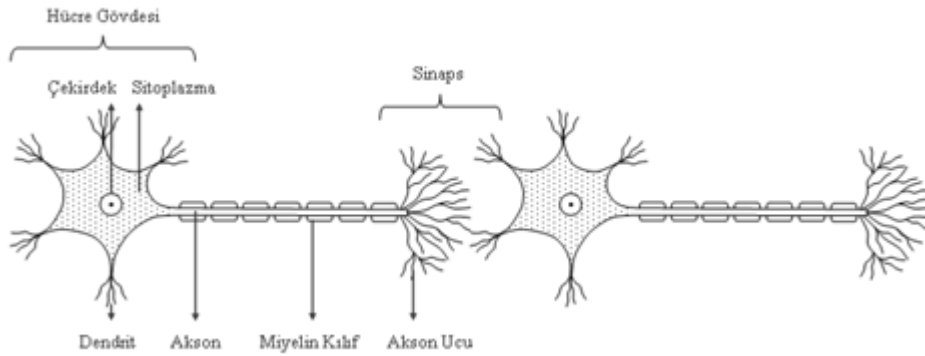
$$Y_m^p = \sum_{j=1}^n \omega_j(x_m) Y_j \quad (8)$$

Buna bağlı “ n ” derece dağılım formülü: Programlanan modeli için 75. Yüzdellik değer kullanılmıştır.

$$\hat{F}(y|X = x) = \sum_{j=1}^n \omega_j(x)I(Y_j < y) \quad (9)$$

2.3.6. Nöral Ağ Regresyonu

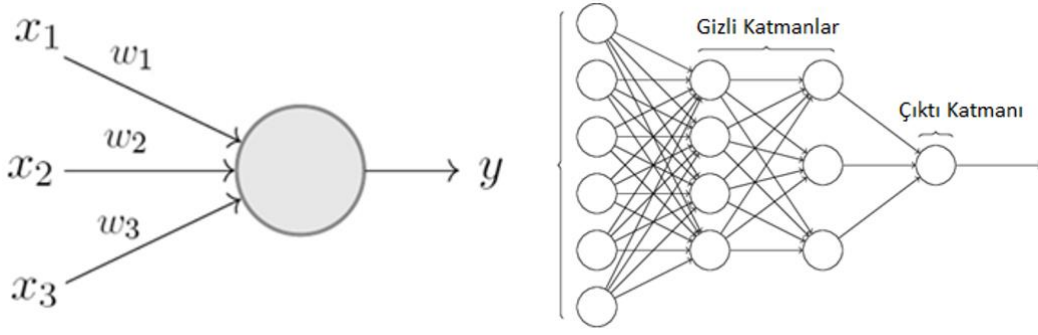
Nöral ağ regresyonu, sinir sisteminde bulunan nöronları model alan bir regresyon yöntemidir. Sinir sistemi “nöron” denilen hücrelerden oluşur. Bu nöronlar kendilerinden önceki ve sonraki nöronlara baęlı bir ağ oluşturur. Sinirbilimde, nöronlar üzerinde taşınan sinyallere aksiyon potansiyeli adı verilir. Bu aksiyon potansiyeli, akson boyunca akarak ilerler ve akson ucunda, telodendritler adı verilen dallı bölgeye ulaşır. Burada bulunan sinaps adı verilen boşluęa salınan nörotransmitterler aracılığıyla sinaps sonrası, yani postsinaptik nöron yapısına iletilir. Böylece tek bir aksiyon potansiyeli, nörondan nörona sadece elektro biyokimyasal süreçlerle iletilmiş olur (Minsky ve Papert, 2017). Şekil 1, bu süreci görsel olarak aktarmaktadır.



Şekil 1. Nöron Yapısı

Kaynak: Minsky ve Papert (1969). (Yazar tarafından uyarlanmıştır.)

Bilgisayar bilimleri bu sistemi basit bir matematiksel şekilde modeller. Nöronlarda, belirli bir eşik deęerin üzerinde bir aksiyon potansiyeli gelir ise, akson ucundaki nörotransmitter kesecikler patlar ve ortaya çıkan kimyasalların bir sonraki hücre sinapsı tarafından emilmesi ile elektrik sinyali iletilir. Eğer aksiyon potansiyeli belirli bir eşik deęerini geçemez ise bu nörotransmitter kesecikler patlamaz ve aksiyon potansiyeli iletilmez. Algoritmaların örnek aldığı model bu kısımdır (Minsky ve Papert, 2017).



Şekil 2. Tek ve Çok Katmanlı Yapay Nöron

Kaynak: Hastie, Tibshirani and Friedman (2013). (Yazar tarafından uyarlanmıştır).

Bilgisayar bilimlerinde nörondan ilham alınan bu yapay nöronlara “perceptron” adı verilmektedir. Aynı sinapsların birden fazla nörondan nörotransmitter madde alması gibi, bir veya daha fazla girdi bir noda alınır. Şekildeki x_1 , x_2 ve x_3 değerleri bu girdilere örnektir. Belirtilen girdilerin ağırlık vektörleri (w_1 , w_2 , w_3) ile skalar çarpımı, belirlenen eşik değerinden büyük ise çıktı 1, küçük ise çıktı 0 olur. Bu durum beyindeki aksiyon potansiyelinin iletilmesi veya iletilmemesini sembolize eder. Yukarıdaki örnekte görüldüğü gibi bu model tek bir nöronu sembolize eder ve bu sebeple “tek katmanlı perceptron (single layer perceptron)” olarak bilinir. Bu belirtilen perceptronların bir ağ oluşturduğu kümeye ise çok katmanlı perceptron denir (Hastie vd., 2013). Her iki yapı da Şekil 2’de gösterilmiştir.

Nöral ağ yöntemi özellikle patern yakalama becerisi ile öne çıkar. Bu özellik otomatik spam e-posta yakalama, bilgisayar ağ atakları güvenlik sistemleri, kişiye özel pazarlama, yüz tanıma, ses tanıma, fotoğraf içindeki objeleri ayırma gibi birçok uygulamada kullanılmaktadır. Finans alanında nöral ağ yöntemi döviz kuru tahminleri, tahvil değer öngörülleri, kredi onaylama, pay fiyatı öngörülleri gibi birçok alanda kullanılır. Piyasa öngörülleri alanında yapılan nöral ağ kullanan çalışmalar incelendiğinde yüksek tahmin hassasiyeti görülmektedir (Guresen, Kayakutlu ve Daim, 2011; Moghaddam, Moghaddam ve Esfandyari, 2016; Selvamuthu, Kumar ve Mishra, 2019).

2.3.7. Poisson Regresyonu

Poisson regresyonu, sayısal değerleri tahmin etmek için kullanılan regresyon modellerinde tipik olarak sayıya yöneliktir. Dağılım bir uçustaki hasta insan sayısı, belirli zaman aralığında beklenen sipariş sayısı, ihtimal tablosu üretimi gibi amaçlar için kullanılabilir. Burada dikkat edilmesi gereken nokta, dağılımın yoğunlaştığı ve seyrekleştiği noktalarıdır. Örnek olarak bir restorana gelen sipariş verileri incelendiğinde kahvaltı, öğlen yemeği ve akşam yemeği saatlerinde yoğunluk, ara saatlerde seyreklik olarak göze çarpacaktır. Benzer paternler borsa açılış kapanış saatleri arasında da görülmektedir. Finans verilerinin normal dağılıma uymadığı durumu ile ilgili bulgular birçok çalışmada gösterilmiştir (Mazzola ve Muliere, 2011; Nannavecchia, 2015; Schoutens, 2002).

Poisson regresyon modeli yalnızca; bağımlı değişken Poisson dağılımına sahipse, tüm sayımlar pozitif değerler ise ve değerler tam sayı ise kullanılabilir. Finansal verilerde

kullanılan deęerler kesirlidir. Tam sayı kullanılma zorunluluęu sebebi ile kesirli para birimi deęerleri (TL, dolar) yerine tam sayı olan deęerler (Kuruř, sent) kullanılmaktadır.

Poisson daęılımı ařaęıda belirtilen olasılık ktle fonksiyonunu kullanmaktadır:

$$P_x(k) = \frac{e^{-(\lambda t)} * (\lambda t)^k}{k!} = \text{Poisson } (\lambda t) \quad (10)$$

Gsterilen fonksiyonda:

$P_x(k)$ = “t” zamanda “k” olayının grlme olasılıęı

λt = birim zamanda gerekleřen olay

k = olay sayısı'nı gstermektedir.

İncelenen alıřmalarda Poisson ve Poisson bazlı algoritmaların finans alanındaki farklı problemlerde etkili olduęu grlmřtir. Faria ve Goncalves (2013) Portekiz bankaları iin yaptığı mřteri kredi notu alıřmalarında Poisson regresyonu ile etkili sonular gstermektedir. Heinen'in (2008) “Autoregressive Conditional Poisson” algoritması New York Stock Exchange piyasası IBM firması verileri tahminleri yksek R^2 deęerleri sunmaktadır.

3. Yntem

ncelikle tm algoritmaların rettięi ngrler iin maksimum hata, minimum hata, ortalama hata, maksimum hata yzdesi, minimum hata yzdesi ve ortalama hata yzdesi verileri incelenmiřtir. Bylelikle algoritmaların gerek deęerlere TL olarak ne kadar yaklařabildięi ve belirtilen bu farkların yzdeleri gsterilmiřtir.

• Ortalama hata: Algoritmanın bulduęu tm sonular ve gerek deęerler arasındaki kuruř cinsinden farkın aritmetik ortalamasıdır.

• Maksimum hata: Algoritmanın bulduęu tm sonular ve gerek deęerler arasındaki kuruř cinsinden en byk farktır.

• Minimum hata: Algoritmanın bulduęu tm sonular ve gerek deęerler arasındaki kuruř cinsinden en kk farktır.

• Ortalama hata yzdesi: Tm tahmin ve gerek deęer iin aradaki farkın gerek deęere blmnn aritmetik ortalamasıdır.

• Maksimum hata yzdesi: En byk hatanın gerek deęere blmdr.

• Minimum hata yzdesi: En kk hatanın gerek deęere blmdr.

Sonrasında regresyon algoritmasının deęiřkeni aıklamadaki bařarısını gstermek iin determinasyon katsayısı (R^2) deęerleri sunulmuřtur. Determinasyon katsayısının yksek olması regresyonun bařarılı olduęu anlamına gelmez, fakat fiyat iniř ıkıřlarını tespit etmek iin kullanılabilir ve bu sebeple kısa / uzun pozisyon alma kararları iin nemli bir veridir. Bu grafikleri anlařılabilir tutmak iin yalnızca az sayıda tahmin ieren kısa dnem tahminleri grafikleri sunulmuřtur. Etik kurul izni ve/veya yasal/zel izin alınmasına gerek olmayan bu alıřmada arařtırma ve yayın etięine uyulmuřtur.

4. Bulgular

4.1. Uzun Vade Öngörülleri

Birinci deneyde 755 günlük verinin 603 günü ile programlar eğitilmiş, kalan 152 gün tahmin edilmiştir. Çalışmadaki en uzun vadeli yatırım ufkuna sahip olan bu deneyde en kısa öğrenme süresi kullanılmıştır. Uzun vade öngörülleri ilişkin sonuçlar Tablo 1’de yer almaktadır.

Tablo 1. Uzun Vade Öngörülleri

	Maks. Hata Maks. Hata %	Min Hata Min Hata %	Ortalama Hata Ortalama Hata %	R ²
Doğrusal	72.508	0.3825	29.5765	0.7305
Regresyon	%8.6096	%0.0488	%3.7006	
Bayes Doğrusal	70.5940	0.4240	29.6440	0.7396
Regresyon	%8.3776	%0.0532	%3.7102	
Karar Ağacı	80.1210	2.3130	31.4670	0.6743
Regresyonu	%9.0530	%0.3075	%3.8993	
Destekli Karar Ormanı	63.093	4.1820	28.9330	0.7305
Regresyonu	%7.4843	%0.5499	%3.5879	
Hızlı Orman Yüzdellik	49.0000	0.0000	17.3960	0.8164
Regresyonu	%5.8125	%0.0000	%2.1414	
Nöral Ağ	79.6500	3.2620	35.4540	0.7240
Regresyonu	%0.4337	%4.3867	%72.4042	
Poisson	66.9990	7.3240	38.3140	0.8032
Regresyon	%7.6379	%0.9353	%4.7690	

Kullanılan veri döneminde tahmin edilen hisse senedinin fiyat aralığı 7.50 – 8.50 TL’dir. Bu süre zarfında yapılan tahminlerde ortalama hata kolonu karşılaştırıldığında, bu deney için en etkin tahminleri 17 kuruş ortalama hata ve %2,14 ortalama hata yüzdesi ile hızlı orman yüzdellik regresyon algoritmasının ürettiği görülmektedir. En iyi ikinci algoritmanın ise 29 kuruş ortalama hata ve %3,58 ortalama hata yüzdesi ile destekli karar ormanı regresyonu algoritması olduğu görülmektedir. Kalan algoritmaların sonuçları birbirine yakın olmakla birlikte maksimum hata yüzdelerindeki farklılıklar sebebi ile Bayes doğrusal regresyon üçüncü, doğrusal regresyon algoritması dördüncü, karar ağacı algoritması ise beşinci sıradadır. Belirtilen algoritmalar birbirinin geliştirilmiş versiyonları kabul edilebileceği için bu sonuçların beklenen sonuçlar olduğu söylenebilir.

Determinasyon katsayısı kolonu incelendiğinde yine hızlı orman yüzdellik dağılımı algoritmasının değişkenler arasındaki ilişkiyi en iyi açıklayan algoritma olduğu görülmektedir. Fiyat artış yönünü açıklamada olan başarısı, kısa ve uzun pozisyon alma kararları için kullanışlıdır.

Poisson ve nöral ağ algoritmalarının bu çalışmadaki en kötü sonuçları üretmesi şaşırtıcıdır. Bu iki algoritmanın referans gösterilen makalelerde gösterdikleri yüksek performans sebebi ile daha yüksek performans göstermeleri gerektiği beklentisi vardır.

4.2. Orta Vade Öngörülleri

İkinci deneyde 755 günlük verinin 680 günü ile programlar eğitilmiş, kalan 75 gün tahmin edilmiştir. Orta vade öngörülleri ilişkin sonuçlar Tablo 2’de yer almaktadır.

Tablo 2. Orta Vade Öngörüleri

	Maks. Hata	Min Hata	Ortalama Hata	R²
	Maks. Hata %	Min Hata %	Ortalama Hata %	
Doğrusal	26.0889	0.0611	10.4295	0.7588
Regresyon	%0.0341	%0.0000	%1.3274	
Bayes Doğrusal	25.5078	0.1228	9.6346	0.8021
Regresyon	%3.2004	%0.0000	%1.2251	
Karar Ağacı	44.3750	0.2500	19.2441	0.7531
Regresyonu	%5.6313	%0.0321	%2.4129	
Destekli Karar Ormanı	29.4278	0.0746	9.41795	0.7844
Regresyonu	%3.4949	%0	%1.1855	
Hızlı Orman Yüzdeleri	35.9999	0.0000	14.7466	0.6749
Regresyonu	%4.6632	%0.0000	%1.8709	
Nöral Ağ	32.5481	0.0574951	11.9374	0.7317
Regresyonu	%4.4162	%0.0000	%1.5178	
Poisson	44.7210	0.1001	22.5274	0.6385
Regresyon	%5.3430	%0.0128	%2.7990	

Kullanılan veri döneminde tahmin edilen hisse senedinin fiyat aralığı 7.50 – 8.50 TL'dir. Bu süre zarfında yapılan tahminlerde ortalama hata kolunu karşılaştırıldığında bu deney için en etkin tahminleri 9 kuruş ortalama hata ve %1,18 ortalama hata yüzdesi ile destekli karar ormanı regresyonu algoritmasının ürettiği görülmektedir. En iyi tahmin üreten ikinci ve üçüncü algoritmalar ise %1,2251 ortalama hata ve %1,3274 ortalama hata yüzdesi ile 9.6 ve 10.4 kuruş hataya tekabül eden Bayes doğrusal ve doğrusal regresyon algoritmaları oldukları görülmektedir. Bu iki algoritmanın ortalama hata yüzdeleri incelendiğinde ise Bayes doğrusal regresyonun ufak bir farkla performansının daha yüksek olduğu görülmektedir. Bu sebeple Bayes doğrusal regresyon algoritması ikinci, doğrusal regresyon algoritması üçüncü sıradadır. Dördüncü sırada ise doğrusal algoritmalar çok yakın bir ortalama hata oranı olan %1,5178 yani 11.9 kuruş ortalama hata ile nöral ağ Regresyonu algoritması öne çıkmaktadır. %1,8709 hata ortalaması ve 14.7 kuruş ortalama hata ile hızlı orman yüzdeleri regresyonu algoritması beşinci, %2,4129 hata oranı ve 19.2 kuruş ortalama hata ile karar ağacı regresyonu algoritması altıncı, %2,7990 hata oranı ve 22.5 kuruş ortalama hata ile ise Poisson regresyonu algoritması son sıradadır. Deney sonuçları bir önceki deney ile karşılaştırıldığında tüm tahminlerin ilk deneye kıyasla daha etkin olduğu görülmektedir. Birinci deneydeki en düşük ortalama hata %2,14 ile hızlı orman yüzdeleri dağılım algoritması iken, ikinci deneyde %1,18 hata oranı ile destekli karar ormanı algoritmasıdır. Dikkat çeken bir diğer sonuçta iki deneyde öne çıkan algoritmaların farklı algoritmalar olmasıdır. Minimum hata kolunu incelendiğinde 0 hata minimum hata yapan destekli karar ormanı, hızlı orman yüzdeleri dağılımı ve nöral ağ algoritmalarının değeri tam doğru tahmin ettiği durumlar olduğunu göstermektedir.

Determinasyon katsayısı kolunu incelendiğinde Bayes doğrusal regresyon algoritmasının değişkenler arasındaki ilişkiyi en iyi açıklayan algoritma olduğu görülmektedir. Kısa ve uzun pozisyon alma kararları için bu çalışmada en etkin sonuçları göstermiştir.

4.3. Kısa Vade Öngörüleri

Üçüncü ve son deneyde ise 755 günlük verinin 747 günü ile programlar eğitilmiş, kalan 8 gün tahmin edilmiştir. Çalışmadaki en kısa vadeli yatırım ufkuna sahip olan bu deneyde

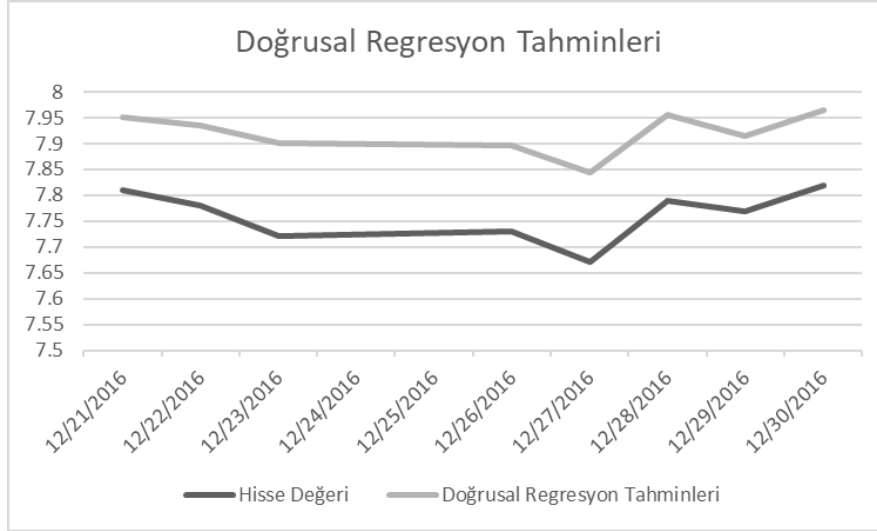
maksimum öğrenme süresi kullanılmıştır. Kısa vade öngörülleriine ilişkin sonuçlar Tablo 3’te yer almaktadır.

Tablo 3. Kısa Vade Öngörülleri

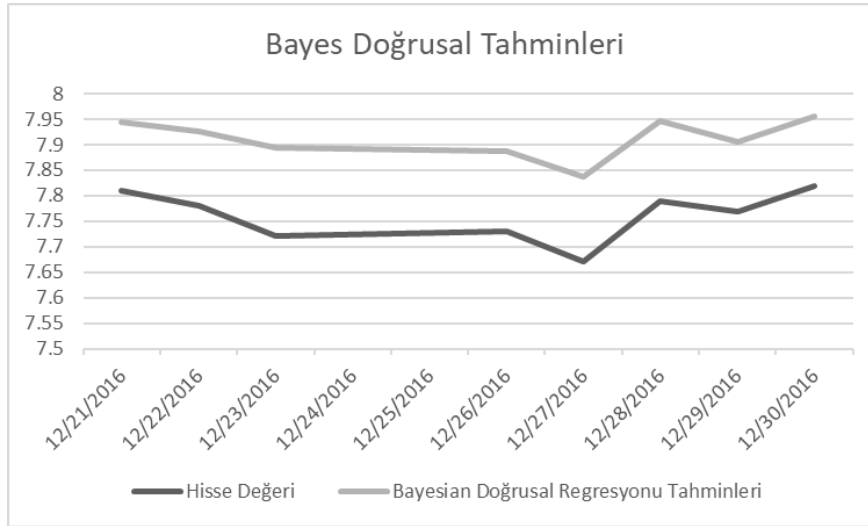
	Maks. Hata	Min. Hata	Ortalama Hata	R²
	Maks. Hata %	Min. Hata %	Ortalama Hata %	
Doğrusal Regresyon	18.1630	14.3757	16.1376	0.9479
Bayes Doğrusal Regresyon	17.3958	13.4689	15.3326	0.9459
Karar Ağacı Regresyonu	12.0063	5.1938	9.7771	0.9069
Destekli Karar Ormanı Regresyonu	16.2827	1.8347	7.5513	0.7100
Hızlı Orman Yüzdellik Regresyonu	20.0000	13.0000	17.5714	0.7717
Nöral Ağ Regresyonu	6.3093	1.8530	3.7479	0.8845
Poisson Regresyonu	4.0817	0.0000	2.0550	0.7183
	%0.5253	%0.0000	%0.2654	

Kullanılan veri döneminde tahmin edilen hisse senedinin fiyat aralığı 7.60 – 7.90 TL’dir. Bu süre zarfında yapılan tahminlerde ortalama hata kolunu karşılaştırıldığında bu çalışma için en etkin tahminleri 2 kuruş ortalama hata ve %0,2654 ortalama hata yüzdesi ile Poisson regresyonu algoritmasının ürettiği görülmektedir. İkinci sırada ise 3 kuruş ortalama hata ve %0,4842 ortalama hata yüzdesi ile nöral ağ regresyonu algoritması bulunmaktadır. Minimum hata kolunu incelendiğinde fiyat 0 minimum hata yapan bu iki algoritmanın fiyatı tamamen doğru tahmin ettiği durumlar olduğu görülmektedir. Bu iki algoritmanın bir önceki çalışmalardaki performansları incelendiğinde son sırada oldukları görülmektedir. İki algoritma da artan öğrenme süresi ile performanslarını yükseltmiştir. Üçüncü ve dördüncü sırada 7 kuruş ortalama hata ve %0,971 ortalama hata yüzdesi ile destekli karar ağacı regresyonu algoritması ve 9 kuruş ortalama hata ve %1,26 ortalama hata yüzdesi ile karar ağacı regresyonu algoritmaları bulunmaktadır. Bu algoritmalar da artan öğrenme süresi ile daha yüksek performans göstermişlerdir. Beşinci ve altıncı sırada 15 kuruş ortalama hata ve %1,9 ortalama hata yüzdesi ile Bayes doğrusal regresyon ve 16 kuruş ortalama hata ve %2,08 ortalama hata yüzdesi ile doğrusal regresyon algoritmaları bulunmaktadır. Kısa vadede son sırayı 17 kuruş ortalama hata ve %2,267 ortalama hata yüzdesi ile hızlı orman yüzdellik dağılım regresyonu almıştır.

Tüm deneyler kıyaslandığında Bayes doğrusal regresyon algoritmasının düzenli olarak doğrusal regresyon algoritmasından üstün performans gösterdiği görülmektedir. Dikkat çeken bir özellik, doğrusal algoritmaların yüksek determinasyon katsayısı değerleri ve neredeyse tüm hareket yönlerini doğru tahmin etmiş olmalarıdır. Bayes doğrusal ve doğrusal regresyon algoritmaları değeri doğru tahmin edememiş olsalar da 7 hareket yönünün 6’sını doğru tahmin etmişlerdir. Kısa ve uzun pozisyon alma konusunda başarıları böylelikle görülebilir. İki algoritmanın tahminleri ve gerçek değerlerle karşılaştırmaları Şekil 3 ve Şekil 4’te gösterilmiştir.



Şekil 3. Doğrusal Regresyon Tahminleri / Hisse Değeri



Şekil 4. Bayes Doğrusal Regresyon Tahminleri / Hisse Değeri

Hızlı Orman Yüzdelik Dağılım algoritması eğitim / tahmin sürelerinden bağımsız olarak hep benzer sonuçlar göstermiştir: Ortalama hata oranı birinci çalışmada %2,14, ikinci çalışmada %1,87, üçüncü çalışmada ise %2.26'dır. Birinci çalışmada ilk sırayı alan bu algoritma, son çalışmada sonuncu sıradadır. Bu algoritma için artan öğrenme süresi performans artışı anlamına gelmemektedir.

5. Sonuç

Bu çalışmada, BIST 30'da bulunan 30 firmanın hisse senetlerine ait geçmiş fiyat verileri kullanılarak, etkin gelecek tahminleri yapabilmek amacıyla 7 farklı yapay zeka algoritması programlanmış ve geçmiş piyasa kapanış verileri ilgili algoritmalara öğretilmiş ve seçilen bir firmanın hisse senedi fiyatları tahmin edilmiştir. Çalışmada programlanan algoritmalar sırasıyla doğrusal regresyon, Bayes doğrusal regresyon, karar ağacı regresyonu, hızlı karar ormanı

regresyonu, hızlı orman yüzdellik dağılım regresyonu, nöral ağ regresyonu ve Poisson regresyonudur. Veri seti bir zaman serisi olduğu için gözetimsiz makine öğrenmesi yöntemi kullanılmıştır. Kısa, orta ve uzun vade yatırımları simüle edecek şekilde üç farklı deney yapılmıştır. Birinci deneyde 755 günlük verinin 603 günü ile programlar eğitilmiş, kalan 152 gün tahmin edilmiştir. İkinci deneyde 755 günlük verinin 680 günü ile programlar eğitilmiş, kalan 75 gün tahmin edilmiştir. Son deneyde ise 755 günlük verinin 747 günü ile programlar eğitilmiş, kalan 8 gün tahmin edilmiştir. Ardından algoritmaların ürettiği tahmini gelecek değerler, gerçekleşen değerler ile kıyaslanmıştır. Deneylerin sonucunda aşağıdaki bulgular elde edilmiştir.

Her koşulda etkili tahmin yapan tek bir yapay zeka algoritması yoktur. Her algoritma farklı ortamlarda, farklı öğrenme süresi ve volatilité değerleri ile karşılaştığında farklı performans göstermektedir. Hızlı orman yüzdellik regresyonu algoritması sonuçları daha uzun öğrenme süresinin daha etkin performans anlamına gelmediğini göstermektedir. Bu durum yapay zekada aşırı öğrenme (overfitting) olarak geçmektedir. Öğrenme süresinde doyuma ulaşıldıktan sonra öğrenmeye devam edilmesi algoritmayı var olmayan paternler bulmaya zorlamaktadır. Doğrusal regresyon algoritmalarının yeterince öğrenme gerçekleştikten sonra artan determinasyon katsayısı ile fiyat hareket yönlerini etkili tahmin ettikleri gözlemlenmiştir. Nöral ağ ve Poisson regresyonları referans gösterilen çalışmalarda gözlemlendiği gibi yeterince öğrenme gerçekleştikten sonra etkin performans göstermektedir. Sonuç olarak fiyat hareket yönü ve hareketin şiddetini tahmin etmek için belirtilen yöntemler kullanılabilir. Hisse senedi fiyat tahminlerinde yapay zeka yöntemlerinin etkili olmayacağını ifade eden H_0 hipotezi reddedilmiştir.

Bu alanda gelecekte yapılacak çalışmalara şu önerilerde bulunulabilir. Bazı algoritmalar devamlı olarak gerçek değer üzerinde/altında tahminler yapmaktadır. Bu algoritmalar birlikte kullanılıp daha üstün performans göstermeleri sağlanabilir. Yeni bilgisayar donanımları çıkması ile bugün sonuç alınamayacak süre isteyen algoritmalar gelecekte kullanışlı olabilir. Yeni donanımlar çıktıkça bugüne kadar doğru veya yanlış kabul edilen birçok bilgi yeniden gözden geçirilmelidir. Son zamanlarda hız kazanan kuantum işlemci çalışmaları, işlem kapasitesi ve örnek uzay simülasyonları bağlamında mevcut bilgisayarlardan çok daha kuvvetli bir görüntü çizmektedir. Şüphesizdir ki, bu işlemcilerle çok daha fazla veri çok daha hızlı işlenecek ve daha etkili sonuçlar elde edilecektir, donanımsal özellikler sonuçlar için önemlidir. Bu çalışmada farklı öğrenme süreleri denenmiş, bazı algoritmalarda performans öğrenme süresi ile artmış, bazı algoritmalarda düşmüştür. Farklı öğrenme süreleri denenerek her algoritmanın maksimum performansı için doyuma ulaşma süresi araştırılabilir.

Araştırmacıların Katkı Oranı Beyanı

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

Çıkar Çatışması Beyanı

Bu çalışmada herhangi bir potansiyel çıkar çatışması bulunmamaktadır.

Kaynakça

- Adams, R. (2004). Intelligent advertising. *AI & Society*, 18, 59-81. <https://doi.org/10.1007/s00146-003-0259-9>
- Alpaydin, E. (2010). *Introduction to machine learning*. Cambridge, Massachusetts London: The MIT Press.
- Altay, E. and Satman, H. (2005). Market forecasting: Artificial neural network and linear regression comparison in an emerging market. *Journal of Financial Management and Analysis*, 18(2), 18-33. Retrieved from <https://papers.ssrn.com/>
- Arda, E. (2020). *Yapay zeka yöntemleri ile finansal zaman serisi öngörülerini* (Yayımlanmamış doktora tezi). Başkent Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Bachelier, L. (1900). *Theory of speculation* (Yayımlanmamış doktora tezi). Paris: University of Paris.
- Baciu, O. (2012). Ranking capital markets efficiency: The case of twenty European stock markets. *Journal of Applied Quantitative Methods*, 9(3), 24-33. Retrieved from <http://www.jaqm.ro/>
- Bayraktar, A. (2012). Etkin piyasalar hipotezi. *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(1), 37-47. Eriřim adresi: <http://aksarayıbd.aksaray.edu.tr/>
- Benston, G. and Hartgraves, A. (2002). Enron: What happened and what we can learn from it. *Journal of Accounting and Public Policy*, 21(2), 105-127. [https://doi.org/10.1016/S0278-4254\(02\)00042-X](https://doi.org/10.1016/S0278-4254(02)00042-X)
- Çelik, M., Kurtaran, A. ve Kurtaran, A. (2018). Zayıf formda piyasa etkinliğinin Türkiye hisse senedi piyasasında test edilmesi [Özel Sayı]. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 457-474. doi:10.18092/ulikidince.456639
- Çelik, T. T. ve Taş, O. (2007). Etkin piyasa hipotezi ve geliřmekte olan hisse senedi piyasaları. *İstanbul Teknik Üniversitesi Dergisi*, 4(2), 11-22. Eriřim adresi: <http://itudergi.itu.edu.tr/>
- Dondurmacı, G. ve Çınar, A. (2014). Finans sektöründe veri madencilięi uygulaması. *Akademik Sosyal Arařtırmalar Dergisi*, 2(1), 258-271. doi:10.16992/ASOS.138
- Draper, N. R. and Smith, H. (1998). *Applied regression analysis* (3. ed.). New Jersey: Wiley Publishing.
- Enke, D. and Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, 29(4), 927-940. <http://doi.org/https://doi.org/10.1016/j.eswa.2005.06.024>
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. doi:10.2307/2325486
- Faria, S. and Gonçaves, F. (2013). Financial data modeling by Poisson mixture regression. *Journal of Applied Statistics*, 40(10), 2150-2162, <https://doi.org/10.1080/02664763.2013.807332>
- Goia, A., May, C. and Fusai, G. (2009). Functional clustering and linear regression for peak load forecasting. *International Journal of Forecasting*, 26(4), 700-711. doi:10.1016/j.ijforecast.2009.05.015
- Guerrien, B. and Gun, O. (2011). Efficient market hypothesis: What are we talking about? *Real World Economics Review*, 56, 19-30. Retrieved from <http://rwer.wordpress.com/>
- Guresen, E., Kayakutlu, G. and Daim, T. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 38(8), 10389-10397. doi:10.1016/j.eswa.2011.02.068
- Güdelek, M. (2019). *Zaman serisi analizi ve tahmini: Derin öğrenme yaklaşımı* (Yayımlanmamış yüksek lisans tezi). TOBB Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Hao, L. and Naiman, D. (2007). *Quantile regression* (1. ed.). California: Sage Publications.
- Hastie, T., Tibshirani, R. and Friedman, J. (2013). *The elements of statistical learning: Data mining, inference and prediction* (2. ed.). New York: Springer Publishing.

- Heinen, A. (2008). *Modelling time series count data: An autoregressive conditional Poisson model* (SSRN Working Paper). Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1117187
- Heshmaty, B. and Kandel, A. (1985). Fuzzy linear regression and its applications to forecasting in uncertain environment. *Fuzzy Sets and Systems*, 15(2), 159-191. [https://doi.org/10.1016/0165-0114\(85\)90044-2](https://doi.org/10.1016/0165-0114(85)90044-2)
- Hromkovic, J. (2005). *Design and analysis of randomized algorithms* (1. ed.). New York: Springer Publishing.
- Johnson, R. and Zhang, T. (2014). Learning nonlinear functions using regularized greedy forest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5), 942-954. doi:10.1109/TPAMI.2013.159
- Jordan, M., Ghahramani, Z. and Saul, L. (1997, December). Hidden Markov decision trees. In M. I. Jordan and T. Petsche (Eds.), *NIPS'96* (pp. 501-507). Paper presented at the Proceedings of the 9th International Conference on Neural Information Processing Systems, Denver, Colorado. Retrieved from <https://dl.acm.org/doi/abs/10.5555/2998981.2999052>
- Kolmogorov, A. N. and Uspensky, V. A. (1987). Algorithms and randomness. *Theory of Probability and Its Applications*, 32(3), 389-412. Retrieved from www.siam.org
- Koop, G. (2003). *Bayesian econometrics* (1. ed.). London: John Wiley & Sons Inc.
- Koop, G. and Korobilis, D. (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4), 267-358. <http://dx.doi.org/10.1561/08000000013>
- Krollner, B., Vanstone, B. and Finnie, G. (2010) *Financial time series forecasting with machine learning techniques: A survey*. Paper presented at the Proceedings of the 18th European Symposium on Artificial Neural Networks: Computational Intelligence and Machine Learning. Retrieved from https://pure.bond.edu.au/ws/files/27498056/Financial_time_series_forecasting_with_machine_learning_techniques.pdf
- Lee, C.-M. and Ko, C.-N. (2009). Time series prediction using RBF neural networks with a nonlinear time-varying evolution PSO algorithm. *Neurocomputing*, 73(1), 449-460. <https://doi.org/10.1016/j.neucom.2009.07.005>
- Leetaru, K. (2008). An open source study of international media coverage of the WorldCom scandal. *The Journal of International Communication*, 14(2), 66-86. <https://doi.org/10.1080/13216597.2008.9674733>
- Marcek, D., Marcek, M. and Babel, J. (2009) Granular RBF NN approach and statistical methods applied to modelling and forecasting high frequency data. *International Journal of Computational Intelligence Systems*, 2(4), 353-364. doi:10.1080/18756891.2009.9727667
- Mazzola, E. and Muliere, P. (2011). Reviewing alternative characterizations of Meixner process. *Probability Surveys*, 8, 127-154. doi:10.1214/11-PS177
- Meek, C., Chickering, D. and Heckerman, D. (2002). Autoregressive tree models for time-series analysis. In R. Grossman, J. Han, V. Kumar, H. Mannila and R. Motwani (Eds.), *2002 SIAM International Conference on Data Mining* (pp. 229-244). Paper presented at the Proceedings of the 2002 Society for Industrial and Applied Mathematics International Conference on Data Mining, Virginia, USA.
- Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(35), 983-999. Retrieved from <https://www.jmlr.org/>
- Millionis, A. and Maschos, D. (2000). On the validity of the weak-form efficient markets hypothesis applied to the London stock exchange. *Applied Economics Letters*, 7(7), 419-421. doi:10.1080/135048500351087
- Minsky, M. and Papert, S. A. (1969). *Perceptrons* (1. ed.). Massachusetts: The MIT Press.
- Minsky, M. and Papert, S. A. (2017). *Perceptrons: An introduction to computational geometry*. Massachusetts: MIT press.

- Moghaddam, A., Moghaddam, M. and Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics Finance and Administrative Science*, 21(41), 89-93. doi:10.1016/j.jefas.2016.07.002
- Nannavecchia, A. (2015). The Meixner process for financial data. *Megatrend Review*, 12(2), 33-44. doi:10.5937/MegRev1502033N
- Özdemir, Z., Atan, S. ve Atan, D. (2016). Hisse senedi piyasasında zayıf formda etkinlik: İMKB üzerine ampirik bir çalışma. *Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, 24(2), 33-48. Erişim adresi: <https://dergipark.org.tr/pub/ije>
- Pole, A., West, M. and Harrison, J. (1994). *Applied Bayesian forecasting and time series analysis*. New York: Chapman & Hall.
- Poyarkov, A., Drutsa, A., Khalyavin, A., Gusev, G. and Serdyukov, P. (2016). Boosted decision tree regression adjustment for variance reduction in online controlled experiments. In B. Krishnapuram and M. Shah (Eds.), *22nd ACM SIGKDD International Conference (235-244)*. Paper presented at the Proceedings of the 22nd Knowledge Discovery and Data Mining International Conference.
- Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion*, 27, 111-125. doi:10.1016/j.inffus.2015.06.005
- Romashchenko, A., Shen, A. and Vershchagin, N. (2000). Combinatorial interpretation of Kolmogorov complexity. In *15th Annual IEEE Conference on Computational Complexity (131-137)*. Paper presented at the Proceedings of the Annual Institute of Electrical and Electronics Engineers Conference on Computational Complexity, Florence, Italy.
- Rosenberg, A. and McIntyre, L. (2011). *Philosophy of science*. Oxfordshire: Routledge.
- Schoutens, W. (2002). *The Meixner Process: Theory and applications in finance* (Semantic Scholar Working Paper). Retrieved from <https://www.eurandom.tue.nl/reports/2002/004-abstract.pdf>
- Selvamuthu, D., Kumar, V. and Mishra, A. (2019). Indian stock market prediction using artificial neural networks on tick data. *Financial Innovation*, 5(1), 1-12. Retrieved from <https://link.springer.com/>
- Shen, S., Jiang, H. and Zhang, T. (2012) *Stock market forecasting using machine learning algorithms* (Stanford University Working Paper). Retrieved from <http://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf>
- Siringnano, J. and Spiliopoulos, K. (2017). Stochastic gradient descent in continuous time. *Society for Industrial and Applied Mathematic Journal on Financial Mathematics*, 8(1), 933-961. doi:10.1137/17M1126825
- Smith, C., McGuire, B., Huang, T. and Yang, G. (2006). *The history of artificial intelligence* (Washington University Working Paper). Retrieved from <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>
- Swednerg, R. (2010). *The structure of confidence and the collapse of Lehman Brothers*. Bingley: Emerald Group Publishing Limited.
- Şıklar, E. (1999). Regresyon analizinde Bayeşçi yaklaşım. *Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 15(1), 113-122. Erişim adresi: <https://dergipark.org.tr/pub/anadoluibfd>
- Theofilatos, K., Karathanasopoulos, A., Sermpinis, G. and Amorgianiotis, T. (2012). Modelling and trading the DJIA financial index using neural networks optimized with adaptive evolutionary algorithms. *Communications in Computer and Information Science*, 311, 453-462. doi:10.1007/978-3-642-32909-8_46
- Tsai, C.-F. and Wang, S.-P. (2009). *Stock price forecasting by hybrid machine learning techniques*. Paper presented at the International MultiConference of Engineers and Computer Scientists. Retrieved from <http://www.iaeng.org/>
- Xu, M., Watanachaturaporn, P., Varshney, P. and Arora, M. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, 97(3), 322-336. doi:10.1016/j.rse.2005.05.008

- Yan, X. and Su, X. (2009). *Linear regression analysis: Theory and computing*. London: World Scientific Publishing Company
- Yan, X.-B., Wang, Z., Yu, S.-H. and Li, Y.-J. (2005) *Time series forecasting with RBF neural network*. Paper presented at the 2005 International Conference on Machine Learning and Cybernetics. Retrieved from <https://ieeexplore.ieee.org/document/1527764>
- Yu, K. and Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4), 437-447. doi:10.1016/S0167-7152(01)00124-9

STOCK PRICE PREDICTIONS USING ARTIFICIAL INTELLIGENCE METHODS

EXTENDED SUMMARY

Aim of the Study

In this article, seven different artificial intelligence algorithms have been programmed with the goal of making accurate stock market price predictions. Algorithms have been trained with past BIST30 daily closing prices, and their predictions are compared to real values. The algorithms tried in this article are Linear Regression, Bayes Linear Regression, Decision Tree Regression, Decision Forest Regression, Fast Forest Quantile Regression, Neural Network Regression and Poisson Regression.

Literature

According to Louis Bachelier's "Random Walk Theory" the prices of stocks follow a random path and cannot be predicted. The concept of randomness that this theory is built on is however a philosophical question, generally examined as a lack of epistemic access or an ontologic property. In the 21. Century most of the financial data is produced by computers, and for computers it is not quite possible to create "random" numbers. Generally, "pseudo-random" algorithms are used. By processing the financial series with machine learning algorithms, it may be possible to uncover patterns that are too complex for humans to see.

Another well-known hypothesis that claims that stock prices cannot be predicted is "Efficient Markets Hypothesis" by Eugene Fama. This hypothesis claims all existing information is reflected in the current price, therefor cannot be used for further predictions. This hypothesis assumes that market efficiency is as high as possible, whereas markets with low / medium efficiency are present. This article challenges these hypotheses due to explained reasons and tries to show artificial intelligence methods can be used for accurate financial timeseries predictions.

Methodology

Seven supervised machine learning algorithms have been coded and trained with the closing prices of BIST 30 companies throughout a time period of 3 years. Then the predictions of the algorithms are compared with real values to show prediction accuracy.

Three different experiments were made. In the first experiment %80 of the data was used for training and remaining %20 of the data was predicted. Secondly %90 of the data used for training and %10 of the data was predicted. Lastly %99 of the data was used for training and %1 of the data was predicted.

Findings

One artificial intelligence algorithm that always makes great predictions does not exist. Each algorithm shows different performance levels in different markets, learning/training ratios, volatility levels. For long-term predictions, algorithms that use median scores such as linear regression and fast forest quantile regression algorithms show the best results. These algorithms are effective in predicting up and down price movements. For short -term predictions, neural network and Poisson regression algorithms produce the best results. These algorithms are effective at forecasting the actual price.

Conclusion

Contrary to the popular random-walk and market efficiency theories, it is shown that artificial intelligence methods can indeed be used for financial time series forecasting. 7 different artificial intelligence algorithms have been tested in this paper. Other articles claim different algorithms such as Hidden Markov Model, Support Vector Machines, Long-Short Term Memory Neurons, Resilient Back Propagation models show accurate results in different markets. With newer hardware and software breakthroughs, the memory and processing power of computers increase exponentially. Therefore what seems to be too complex to calculate 20 years ago can easily be calculated in time. Different training / predicting ratios can be used with much more data in the future.