

Makalenin geliş tarihi: 26.07.2020

1. Hakem rapor tarihi: 15.08.2020

2. Hakem rapor tarihi: 22.08.2020

Kabul tarihi: 15.09.2020

XGBOOST VE MARS YÖNTEMLERİYLE ALTIN FİYATLARININ KESTİRİMİ (Araştırma Makalesi)

Hayri ABAR (*)

Öz

Altın önemli bir ödeme, yatırım ve birikim aracı olduğundan fiyatının belirlenmesi ülkeler ve yatırımcılar için önemlidir. Bu nedenle bu çalışmada altın fiyatının kestirimi amaçlanmıştır. Bu amaçla altın fiyatı üzerinde etkili olduğu düşünülen gümüş fiyatı, ham petrol WTI vadeli işlemleri fiyatı, ABD Doları endeksi, S&P500 endeksi, ABD federal fonlar bileşik faiz oranı, ABD TÜFE değişkenleri oluşturulan modellerde girdi olarak kullanılmıştır. Kullanılan veriler Ocak 2015 – Haziran 2020 dönemine aittir. Altın fiyatı doğrusal olmayan bir seridir, bunun yanında durağandır. Altın fiyatının bu özellikleri fiyat kestirimlerinin elde edilmesini zorlaştırmaktadır. Bu nedenle klasik yöntemlerin yanında makine öğrenmesi yöntemlerinin ve parametrik olmayan yöntemlerin altın fiyatının kestiriminde kullanılması uygun olmaktadır. Bu çalışmada, kestirimlerin elde edilmesinde XGBoost, MARS ve lineer regresyon modelleri kullanılmıştır. Elde edilen sonuçlar modellere ait performans değerlendirme kriterleri kullanılarak karşılaştırılmış, XGBoost ve MARS modelleri için girdi değişkenlerin altın fiyatı üzerindeki etkileri belirlenmiştir. Kullanılan modeller arasında XGBoost modeli %99,6 başarılı kestirim oranı ile en başarılı sonuçların elde edilmesini sağlamıştır. MARS modeli için ise bu oran %97,8'dir. Bu oranlar kullanılan değişkenlerin altın fiyatı üzerinde önemli etkiye sahip olduğunu göstermektedir. Kullanılan değişkenler arasında altın fiyatı üzerinde en önemli etkiye sahip değişken ABD TÜFE değişkenidir. Ayrıca elde edilen bulgular XGBoost ve MARS yöntemlerinin altın fiyatı ve benzer seriler için kestirimlerin elde edilmesinde tercih edilebilecek yöntemler olduğunu göstermektedir.

Anahtar Kelimeler: Altın Fiyatı, Kestirim, Makine Öğrenmesi, Parametrik Olmayan Regresyon, XGBoost, MARS

Jel Kodları: C53, E44, G19

*) Dr. Öğr. Üyesi, Gaziantep Üniversitesi İİBF, İktisat Bölümü
(e-posta: hayriabar@gmail.com). ORCID ID: <https://orcid.org/0000-0002-2043-6747>

Prediction of Gold Prices by Xgboost and Mars Methods

Abstract

As gold is an important means of payment, investment and savings, determination of prices is important for countries and investors. Therefore, the prediction of the gold price is aimed in this study. For this purpose, the variables such as silver price, crude oil WTI futures price, US Dollar index, S&P500 index, US federal funds compound interest rate, and US CPI which are thought to have effect on the gold price, were used as inputs in the models. The data used belongs to the period January 2015 - June 2020. Gold price is a non-linear series, besides it is non-stationary. These features of the gold price make it difficult to obtain price predictions. For this reason, it is appropriate to use machine learning methods and non-parametric methods in prediction of the gold price in addition to classical methods. In this study, XGBoost, MARS and linear regression models were used to obtain the predictions. The results obtained were compared using the performance evaluation criteria of the models, and the effects of the input variables on the gold price for the XGBoost and MARS models were determined. Among the models used, the XGBoost model provided the most successful results with a 99.6% successful prediction rate. For the MARS model, this rate is 97.8%. These ratios show that the variables used have a significant effect on gold prices. Among the variables used, the variable that has the most important effect on gold prices is the US CPI. In addition, the findings show that the XGBoost and MARS methods are preferable methods to obtain estimates for gold price and similar series.

Keywords: Gold Price, Prediction, Machine Learning, Nonparametric Regression, XGBoost, MARS.

Jel Codes: C53, E44, G19

1. Giriş

Altın bireyler, firmalar ve devletler için eski çağlarda bir birikim ve ödeme aracı olarak günümüzde bu özelliklerinin yanında güvenli bir yatırım aracıdır. Ampirik çalışmalarda da pek çok ülke için altının bu özelliğini destekleyici sonuçlar elde edilmektedir (Baur ve McDermott, 2010). Birimler enflasyonun etkilerini azaltma, ekonomik krizden korunma, sağlam yatırımlar gibi birçok önemli nedenden dolayı altına yatırım yapmaktadır (Verma, Thampi ve Rao, 2020). Altın ülkelerin finansal gücünün bir göstergesi olarak kullanıldığından ülke merkez bankaları da altın rezervlerini artırmak amacıyla altın satın almaktadır (Sami ve Nazir, 2018). Altın ekonomik ve finansal krizlerde dahi değerini koruyabilen tek emtiadır (Sivalingam, Mahendran ve Natarajan, 2016). Kolayca nakde dönüştürülebilmesi de altının önemli bir özelliğidir. Yatırım aracı olmasının yanında takı olarak da kullanılabilir (Yüksel ve Akkoç, 2016). Altın fiyat dalgalanmalarından yararlanarak kar elde edilmesi amacıyla da kullanılabilir (Keskin Benli ve Yıldız, 2014). Özellikle yatırım çeşitlendirmesi için önemli bir araçtır. Altın fiyatı diğer yatırım

araçlarıyla da ilişkilidir. İlişki genellikle ters yönlü olduğundan diğer varlıkların fiyatı düşüğünde altın fiyatı artacak, bu şekilde yatırım çeşitlendirmesinden yararlanılabilecektir (Gangopadhyay, Jangir ve Sensarma, 2016). Bu nedenler altın fiyatının seyrinin para piyasaları ve ekonomi için oldukça önemli olmasına yol açmaktadır. Küresel ekonomi ile alakalı korku ve belirsizlik altın fiyatının yükselmesine yol açtığından altın ekonomik, finansal ve politik kriz dönemlerinde en çekici yatırım aracına dönüşmüş ve altın fiyatı dünyadaki ekonomik koşulların bir göstergesi haline gelmiştir (Ghalayini ve Farhat, 2020). Altın fiyatı kontrol edilemez; ancak gelecekteki kararlar için iyi bir şekilde değerlendirilebilir ve gelecekte alacağı değerler için öngörüler elde edilebilir (Khan, 2013). Öngörü elde edilmesi finans ve ekonomi alanında önemli bir çalışma başlığıdır (Carvalho ve Ribeiro, 2008). Son yıllarda ekonomik krizin üstesinden gelmek amacıyla pek çok ülke ve araştırmacı altın fiyatının öngörüsü için yoğun çaba harcadığından altın akademik alanda daha fazla dikkat çekmektedir (Weng ve diğerleri, 2020). Altın fiyatının doğru öngörüsü finansal yatırımcıların ve merkez bankalarının yatırım politikalarının belirlenmesinde ve riskten kaçınmada doğru kararlar almalarını sağlamaktadır (Livieris, Pintelas ve Pintelas, 2020). Altın farklı ulusal ekonomik aktivitelerde ve sosyal güvenlikte kullanılan stratejik kaynaktır ve fiyattaki dalgalanmalar fiyattaki oynaklığın artışına yol açtığından fiyat istikrarını etkilemekte ve fiyat tahmini yapmaya çalışan piyasa oyuncularının işlerini zorlaştırmaktadır (Bouri, Jain, Biswal ve Roubaud, 2017).

Altın fiyatı durağandışı olduğundan ve doğrusal olmadığından öngörülmesi ve kestirimi zor bir seridir. Bu zorluğun üstesinden gelmek amacıyla altın fiyatının öngörü ve kestiriminde farklı yöntemler tercih edilebilmektedir. Bu yöntemler klasik yöntemler ve alternatif yöntemler olarak iki başlıkta değerlendirilebilir. Klasik yöntemler içerisinde LR (Lineer Regresyon – Lineer Regresyon), VAR (Vector Autoregression – Vektör Otoregresyon), VECM (Vector Error Correction Model – Vektör Hata Düzeltme Modeli), ARIMA (Autoregressive Integrated Moving Average – Otoregresif Bütünleşik Hareketli Ortalamalar) ve GARCH (Autoregressive Conditional Heteroskedasticity – Otoregresif Koşullu Değişen Varyans) modelleri sayılabilir. Alternatif yöntemler olarak ise derin öğrenme, makine öğrenmesi ve parametrik olmayan regresyon yöntemleri sıralanabilir. Özellikle günümüzde bilgisayar teknolojisindeki gelişmeler alternatif yöntemlerin sıklıkla tercih edilmesine neden olmuştur. Makine öğrenmesi yöntemleri pek çok zaman serisi verisi için kestirim amaçlı kullanılmış ve başarılı sonuçlar elde edilmesini sağlamıştır. Makine öğrenmesi yöntemleri gürlü içerden kompleks zaman serilerine çok iyi uyum sağlayabilmektedir. Son on yılda altın fiyat ve hareketlerinin öngörüsü için makine öğrenmesi ve derin öğrenme yöntemlerinin kullanımı endüstriyel ve bilimsel alanda yaygınlık kazanmıştır (Livieris ve diğerleri, 2020). Çünkü fiyat kestirimi karmaşık bir alan olmasına rağmen makine öğrenme algoritmaları kestirimlerin kolay ve etkin bir şekilde elde edilmesini sağlamaktadır (Pandey, Misra ve Saxena, 2019). Alternatif yöntemler arasında ANN (Artificial Neural Networks – Yapay Sinir Ağları), RNN (Recurrent Neural Network – Özyinelemeli Sinir Ağları), LSTM (Long Short-Term Memory – Uzun Kısa Dönem Hafıza) ve SVM (Support Vector Machine – Destek Vektör Makinesi) gibi yön-

temler sayılabilir.

Bu yöntemlere ilaveten günümüzde boosting (artırma) algoritmaları da popülerlik kazanmaktadır. Bu algoritmaların temel yaklaşımı birden çok zayıf öğreniciyi bir araya getirerek güçlü bir öğrenici ortaya çıkarmaya dayanmaktadır. Bu şekilde geliştirilmiş pek çok algoritma mevcuttur. Bunlardan bir tanesi de XGBoost (Extreme Gradient Boosting – Ekstrem Gradyan Artırma) yöntemidir. XGBoost karar ağacı temelli bir makine öğrenme yöntemi olan ve sınıflama ve kestirim amaçlı kullanılabilen GB (Gradient Boosting – Gradyan Artırma) yöntemine benzer şekilde çalışmaktadır. Bu yöntem eğitim için klasik yöntemlere göre aşırı zaman harcamasına rağmen başarılı kestirimler elde edilmesini sağlamasıyla ön plana çıkmaktadır. Altın fiyatı serisi doğrusal olmayan bir seridir. Bağımsız değişkenin bağımlı değişken üzerindeki etkisi doğrusal olmadığında basit doğrusal regresyon modelinin doğrusallık varsayımı ihlal edilmiş olur (Alkan, Oktay, Genç ve Çelik, 2017). Kestirimlerde doğrusal modellerin kullanımı kestirim hatasının büyük olmasına yol açmaktadır. Modelin tek bir eğri (veya doğru) ile oluşturulması basit olmasına rağmen tahmin edilen model gözlemleri başarılı bir şekilde temsil etmeyebilir (Oktay, Genç ve Alkan, 2012). Bu nedenle bu serinin kestiriminde doğrusal olmayan modellerin tercih edilmesi uygun olmaktadır. Ayrıca doğrusal olmayan yapının fonksiyonel biçimi de bilinmemektedir. Bu özellikleri nedeniyle altın fiyat kestiriminde MARS (Multivariate Adaptive Regression Splines - Çok Değişkenli Uyarlanabilir Regresyon Eğrileri) modeli kullanılabilir modeller arasında yer almaktadır. Spline, çeşitli noktalara veya düğümlere sabitlenen ve bağımlı ve bağımsız değişken arasındaki ilişkiyi gösteren bir eğridir (Akin, Eyduvan, Eyduvan ve Reed, 2020). Spline fonksiyonları farklı parçalara sahiptir fakat her bir parçayı temsil eden eğri sürekli bir fonksiyondur (Oktay, Talas, Alkan ve Genç, 2012). Bu modellerde doğrusal dışı kalıbın önceden bilinmesi gerekmez. Yöntem, modelin eğitim aşamasında kestirimleri en iyi hale getirecek şekilde doğrusal dışı kalıbı yakalamaya çalışır.

Bu çalışmada amaç altın fiyatının (ons/\$) kestirimidir. Bu amaçla altın fiyatı üzerinde etkili olduğu düşünülen gümüş fiyatı (ons/\$), ham petrol WTI vadeli işlemleri fiyatı (varil/\$), ABD Doları endeksi, S&P500 endeksi, ABD federal fonlar bileşik faiz oranı, ABD TÜFE (tüm kalemler – tüm şehirli tüketiciler) değişkenleri girdi olarak kullanılmıştır. Kestirimlerin elde edilmesinde XGBoost, MARS ve lineer regresyon modelleri kullanılmıştır. Elde edilen sonuçlar modellere ait performans değerlendirme kriterleri kullanılarak karşılaştırılmış, XGBoost ve MARS modelleri için girdi değişkenlerin, altın fiyatı üzerindeki etkileri belirlenmiştir.

2. Literatür Özeti

Weng ve diğerleri (2020) çalışmalarında gümüş fiyatı, S&P 500 endeksi, petrol fiyatı, beş gecikmeli altın fiyatını kullanarak altın fiyatı öngörüsü yapmışlardır. Çalışmada 6 Temmuz 2010 – 4 Temmuz 2017 dönemi günlük veriler kullanılmıştır. Çalışmada GA-ROSELM yöntemi kullanılmıştır. Bu yöntem yazarlar tarafından GA (Genetic

– Algorithm - Genetik Algoritma) ve OS – ELM (On-Line Sequential Extreme Learning Machine – Online Ardışık Ekstrem Öğrenme Makinası) birlikte kullanılarak türetilmiştir. Türetilen yöntemde OS – ELM yöntemine ait bazı parametreler GA yardımıyla elde edilmektedir. Çalışmada GA ile parametrelerin belirlenmesinin sonuçları değerlendirilmiş ve türetilen yöntem diğer bazı yöntemlerle (ARIMA, SVM, BP, ELM, OS – ELM) karşılaştırılmıştır. GA kullanımı kestirimlerdeki rastgeleliği, modelin gizli nöron sayısına bağımlılığını azaltmakta ve kestirim modelinin genelleştirme yeteneğini artırmaktadır. Ayrıca türetilen model karşılaştırılan diğer modellerden daha başarılı sonuçlar elde edilmesini sağlamaktadır.

Verma ve diğerleri (2020) çalışmalarında Hindistan için altın fiyatını öngörmeyi amaçlamıştır. Kullanılan veriler Ocak 2015 – Aralık 2018 dönemi haftalık verilerdir. Veriler 0,1 – 0,9 aralığında normalize edilerek kullanılmıştır. Modellerde girdi olarak 5 haftalık gecikmeli altın fiyatı kullanılmıştır. Çalışmada GDM’de (Gradient Descent Method – Gradyan İniş Yöntemi) standart kuadratik hata fonksiyonunu kullanmak yerine beş farklı optimizasyon fonksiyonu kullanılmıştır. Çalışmada yazarlar tarafından geliştirilen yöntem sonuçları farklı yöntem sonuçlarıyla karşılaştırılmıştır. Elde edilen bulgulara göre Cauchy hata fonksiyonu temelli algoritma, yazarlar tarafından geliştirilen diğer yöntemlerden ve klasik yöntemlerden daha başarılı sonuçlar elde edilmesini sağlamıştır.

Livieris ve diğerleri (2020) çalışmalarında altın fiyatının öngörüsünde CNN (Convolutional Neural Network – Geleneksel Sinir Ağları) ve LSTM yöntemlerinin birleşimini kullanmışlardır. Kullanılan veriler Ocak 2014 – Nisan 2018 dönemi günlük verilerdir. Model iki bileşenden oluşmuştur. İlk bileşen geleneksel ve havuzlanmış katmanlardan oluşmuş ve girdi verisinin özelliklerini geliştirmek için karmaşık matematiksel işlemler bu kısımda gerçekleştirilmiştir. İkinci bileşen ise dense katmanları ve LSTM tarafından türetilen özellikleri kullanmaktadır. Çalışmada kullanılan iki CNN – LSTM modelinden biri en iyi performansın elde edilmesini sağlamıştır. Çalışmada elde edilen bulgulara göre yazarlar LSTM’nin altın fiyatının öngörüsü için iyi bir yöntem olmasına rağmen CNN ile birlikte kullanılmasının performansı artırdığını belirtmişlerdir.

Fong-Ching, Chao-Hui ve Chaochang (2020) çalışmalarında altın fiyatının öngörüsünde çoğunlukla kullanılan faktörlerin yanında metin madenciliği ile elde ettikleri karar skorunu da kullanmışlardır. Kullanılan diğer değişkenler altın fiyatı, gümüş fiyatı, platinyum ve paladyum fiyatı, ABD federal fon oranı, ABD Dolar endeksi, petrol fiyatı ve S&P 500 endeksidir. Çalışmada kullanılan veriler 1 Ocak 2016 – 31 Aralık 2017 dönemi günlük verilerdir. Çalışmada LSSVR ile GA birlikte kullanılmıştır. LSSVR için gerekli olan parametreler GA ile elde edilmiştir. Bu şekilde LSSVR yönteminin performansı artırılmıştır. Yazarlar, elde edilen bulgulara göre GA – LSSVR’nin üstün kestirim performansı sağladığını belirtmişlerdir.

Ghalayini ve Farhat (2020) çalışmalarında altın fiyatında değişikliğe neden olan faktörleri belirlemeyi ve altın fiyatı için öngörüü amaçlamışlardır. Kullanılan değişkenler altın talebi, altın arzı, açık faiz sözleşmeleri, enflasyon, Çin Yuani/ABD Doları, Japon

Yeni/ABD Doları, ABD Doları/Euro, faiz oranı, petrol fiyatı ve New York Menkul Kıymetler Borsası Endeksi'dir. Çalışmada Granger nedensellik testi yapılmış, değişkenler arasındaki ilişki için DOLS ve oynaklık için GARCH(1,1) modeli tahmin edilmiştir. Elde edilen bulgulara göre kısa dönemde Japon Yeni/ABD Doları altın fiyatının, altın fiyatı açık faizin Granger nedenidir. Ayrıca sonuçlar, altın fiyatı ile altın talebi, altın arzı, ABD Doları/SDR döviz kuru, enflasyon, spekülasyon, faiz oranı ve petrol fiyatının uzun vadeli ilişkili olduğunu göstermektedir.

Risse (2019) çalışmasında altın fiyatı için öngörülerini DWT (Discrete Wavelet Decomposition – Kesikli Dalgacık Ayırıştırması) ile SVR yöntemlerini birleşimi olan DWT – SVR ile elde etmiştir. Kullanılan veriler Ocak 1992 – Aralık 2016 dönemi aylık verilerdir. Çalışmada girdi olarak 9 farklı ekonomik ve finansal gösterge kullanılmıştır. Analizlerde kullanılan tüm değişkenler DWT ile farklı periyotlar için bileşenlerine ayrılmış ve elde edilen bileşenler SVR için girdi olarak kullanılmıştır. Elde edilen bulgulara göre tüm değişkenlerin birlikte kullanılmasının ekonomik ve istatistiksel performansı azaltmıştır. Girdi olarak özellikle üzerinde durulması gereken değişkenler enflasyon, ticaret ağırlıklı döviz kuru, GSCI ve CRB emtia endeksleridir. Değişkenlerin zaman ve frekansa göre ayrıştırılması öngörü performansını attığı tespit edilmiştir.

Öndes ve Oğuzlar (2019) çalışmalarında brent petrol fiyatı, gümüş fiyatı, ABD Doları/Euro paritesi, Euro Next 100 Endeksi ve Dow Jones Endeksi verilerini ve yapay sinir ağlarını kullanarak altın fiyatı için öngörü elde etmişlerdir. Kullanılan veriler 3 Ocak 2005 – 31 Ekim 2017 dönemi günlük verilerdir. ANN yöntemi ile elde edilen sonuçlar için başarı oranı %81,43 olarak elde edilmiştir. Çalışmada yapılan geçerlilik çözümlemesine göre altın fiyatı üzerinde en fazla gümüş fiyatı sonrasında sırasıyla ABD Doları/Euro paritesi, brent petrol fiyatı, Dow Jones Endeksi ve Euro Next 100 endeksinin etkili olduğu tespit edilmiştir.

Alameer, Elaziz, Ewees, Ye ve Jianhua (2019) çalışmalarında WOA (Balina Optimizasyon Algoritması - Whale Optimization Algorithm) ile MLP – NN (Çok Katmanlı Algılayıcı Sinir Ağları – Multilayer Perceptron Neural Network) yöntemlerinin hibriti yöntem (WOA – NN) kullanarak altın fiyatını öngörmeyi amaçlamışlardır. Yazarlar bu yöntemin sonuçlarını ANN ve ARIMA gibi yaygın kullanılan yöntemlerin yanında PSO – NN (Sinir Ağları İçin Parçacık Sürü Optimizasyonu – Particle Swarm Optimization For Neural Network), GA – NN (Sinir Ağları İçin Genetik Algoritma – Genetic Algorithm For Neural Network) ve GWO – NN (Sinir Ağları İçin Gri Kurt Optimizasyonu – Grey Wolf Optimization For Neural Network) gibi yöntemlerin sonuçlarıyla karşılaştırmışlardır. Kullanılan veriler Ekim 1987 – Ağustos 2017 dönemi aylık verilerdir. Çalışmada öngörülerde altın fiyatı, gümüş fiyatı, bakır ve demir fiyatı, petrol fiyatı, döviz kurları, Çin enflasyon oranı ve ABD enflasyon oranı verileri kullanılmıştır. Yazarlar kullanılan değişkenlerin altın fiyatı öngörüsünde başarılı sonuçlar verdiğini belirtmişlerdir. Ayrıca kullanılan yöntemler arasında en başarılı sonuçların elde edilmesini WOA – NN yöntemi sağlamıştır.

Sami ve Nazir (2018) çalışmalarında ANN ve doğrusal regresyon ile altın fiyat kestirimi yapmışlardır. Kullanılan veriler Ocak 2005 – Eylül 2016 dönemi günlük verilerdir. Çalışmada oluşturulan modellerde altın talep ve arzı açısından önemli ülkelerin borsa endeksleri ve faiz oranları, altına büyük yatırımlar yapan beş şirket verileri, bazı değerli metal fiyatları ve petrol fiyatı gibi pek çok girdi kullanılmıştır. Elde edilen sonuçlara göre ANN, doğrusal regresyondan daha başarılıdır fakat doğrusal regresyonun eğitim süresi ANN'nin eğitim süresine göre daha kısadır. Büyük bir şirketin hisse senedi fiyatının altın fiyatı üzerinde ABD ekonomisinden daha fazla etkili olduğu çalışmanın ilginç bulgularındandır.

Değirmenci ve Akay (2017) çalışmalarında ARIMA ve ARCH modellerini kullanarak BIST100 endeksi, altın fiyatı, petrol fiyatı ve döviz kuru için öngörüler elde etmişlerdir. Kullanılan veriler 2 Ocak 2009 – 25 Kasım 2016 dönemi haftalık verilerdir. Elde edilen bulgulara göre ele alınan dönemde en fazla getirinin borsadan, en düşük getirinin ise petrolden elde edildiği tespit edilmiştir. Tüm değişken için kurulan ARIMA modellerinin ARCH etkisi içerdiği tespit edilmiştir. Ayrıca altın hariç tüm seriler asimetrik etki içermektedir. Elde edilen bulgulara göre kullanılan modellerin analiz edilen seriler için öngörü elde edilmesinde başarılı bir şekilde kullanılabildiği ifade edilmiştir.

H.-H. Chen, Chen ve Chiu (2016) çalışmalarında ANN ile metin madenciliğini bir arada kullanarak altın fiyatı için öngörü elde etmişlerdir. Bu modelden elde edilen sonuçlar ARIMA modeli sonuçlarıyla karşılaştırılmıştır. Kullanılan veriler Ocak 1999 – Aralık 2006 dönemine ait olup aylık periyotlarda derlenmiştir. Oluşturulan modele altın fiyatını etkilediği düşünülen 8 değişken dâhil edilmiştir. Çalışmada altın fiyatının sadece nicel özelliklerden değil savaş, uluslararası ilişkiler, terör saldırıları gibi nitel özelliklerden de etkilenmesi nedeniyle metin madenciliğinin kullanıldığı ifade edilmiştir. Elde edilen bulgulara göre ANN'nin ARIMA'dan daha iyi sonuçlar elde edilmesini sağladığı ve metin madenciliğinin altın fiyat trendinin belirlenmesinde makul sonuçlar elde edilmesini sağladığı ifade edilmiştir.

Yüksel ve Akkoç (2016) çalışmalarında ANN ile altın fiyatı öngörüsü yapmışlardır. Kullanılan veriler 3 Ocak 2002 – 31 Ekim 2013 dönemi günlük verilerdir. Kurulan modelde gümüş fiyatı (G), brent petrol (BrP) fiyatı, ABD Doları/Euro paritesi (EUR/USD), Euro Next 100 endeksi (N100), ABD Dow Jones endeksi (DJI), 13 hafta vadeli ABD hazine bonosu faiz oranı (AHM) ve ABD tüfe (ABD TÜFE) kullanılmıştır. Kullanılan değişkenlerin altın fiyatı üzerindeki etkisinin belirlenmesi için duyarlılık analizi yapılmış ve etki büyüklüğüne göre büyükten küçüğe G, BrP, DJI, EUR/USD, N100, AHB ve ABD TÜFE sıralaması elde edilmiştir. AHB ve ABD TÜFE değişkenlerinin etkisi diğer değişkenlere göre oldukça küçüktür. MAPE oranına göre yaklaşık %98,46 oranında başarılı sonuçlar elde edilmiştir.

Sivalingam ve diğerleri (2016) çalışmalarında ELM ile altın fiyatı öngörüsü yapmışlardır. Kullanılan veriler 1 Ocak 2000 – 31 Nisan 2014 dönemi aylık verilerdir. Altın fiyatının öngörüsünde altının geçmiş fiyatı, gümüş fiyatı, petrol fiyatı, S&P 500 endeksi ve dö-

viz kuru kullanılmıştır. ELM'den elde edilen sonuçlar FNNWF (Feed Forward Networks Without Feedback – Geri Beslemesiz İleri Beslemeli Ağlar) ve FFBPNN (Feed Forward Back Propagation Networks – İleri Beslemeli Geri Yayılım Ağları) yöntemlerinden elde edilen sonuçlar ile karşılaştırılmıştır. Öngörü performansı açısından diğer yöntemlerle karşılaştırıldığında ELM yaklaşık %3 daha iyi sonuçlar elde edilmesini sağlamaktadır.

3. Yöntem

Bu çalışmada altın fiyatının kestirimi amaçlanmıştır. Bu amaçla kullanılabilir pek çok makine öğrenmesi yöntemi ve istatistiki yöntem bulunmaktadır. Bu yöntemler arasında XGBoost ve MARS yöntemleri etkin kestirim başarılarıyla ön plana çıkmaktadır. Bu nedenle altın fiyatının kestiriminde bu iki yöntem kullanılmıştır. Ayrıca karşılaştırma amacıyla LR modeli tahmini yapılmış ve bu yöneme ait kestirimler de elde edilmiştir.

3.1. XGBoost

XGBoost, Friedman (2001) tarafından geliştirilen GB yöntemi temelli bir algoritmadır. Hem XGBoost hem de GB, gradyan boosting yöntemini kullanmasına rağmen, XGBoost, aşırı öğrenmeyi kontrol etmek için daha düzenli bir model biçimlendirmesi kullanır ve daha iyi performans sağlar (Carmona, Climent ve Momparler, 2019). XGBoost son yıllarda makine öğrenmesi için geliştirilen en başarılı yöntemlerden biridir (Ma ve diğerleri, 2018). Günümüzde gereksiz maillerin tespiti, reklam eşleştirme sistemleri, dolandırıcılık tespit sistemleri, fizikte anomali olay tespit sistemleri gibi alanlarda bu yöntemler başarılı bir şekilde kullanılmaktadır (T. Chen ve Guestrin, 2016). Bu yöntemlerin önemli avantajı tahminin veriye dayalı olarak elde edilmesidir. Regresyon ve sınıflandırma ağaçları için boosting algoritmalarının kullanımı etkin sonuçlar elde edilmesini sağlamaktadır. XGBoost başarılı kestirimlerin elde edilmesini sağlamanın yanında diğer pek çok makine öğrenmesi yöntemine göre hesaplama kolaylığına sahiptir. Bu özelliği sayesinde çok boyutlu büyük veri analizinde etkin bir şekilde kullanılabilir. Pek çok Kaggle yarışmasını XGBoost ve bu tarz algoritmalar kazanmıştır. XGBoost yöntemi kullanılarak kazanılan yarışmalardaki problemlere örnek olarak mağaza satış tahmini, yüksek enerji fiziği olay sınıflandırması, web metin sınıflandırması, müşteri davranış tahmini, hareket algılama, reklam tıklama oranı tahmini, kötü amaçlı yazılım sınıflandırması, ürün kategorizasyonu, tehlike riski tahmini ve büyük çevrimiçi kurs bırakma oranı tahmini verilebilir (T. Chen ve Guestrin, 2016). Bu başarısı bu yöntemlerin bilimsel alanda da dikkat çekmesine yol açmıştır. Bu yöntemde değişkenlere ait bilgi kazançları kullanılarak, girdi değişkenlerin çıktı değişkeni üzerindeki etkisi belirlenebilir.

Örnek sayısı n ve girdi değişken sayısı m olan veri seti için $D = \{(x_i, y_i)\}$ ($|D| = n, x_i \in \mathcal{R}^m, y_i \in \mathcal{R}$), bir topluluk ağaç modeli çıktıyı kestirmek için K toplanır fonksiyon kullanır (T. Chen ve Guestrin, 2016).

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(y_i), f_k \in \mathcal{F}$$

Burada $\mathcal{F} = \{f(x) = w_{q(x)}\} (q: \mathcal{R}^m \rightarrow T, w \in \mathcal{R}^T)$ regresyon ağaçları (CART) uzayıdır. f_k , zayıf öğrenciler ve K , zayıf öğrenci sayısıdır (Zhou, Li, Shi ve Qian, 2019). Modelde kullanılan fonksiyon setini öğrenmek için aşağıdaki amaç fonksiyonu minimize edilir:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

Burada l , türevlenebilir kayıp fonksiyonu ve Ω , modelin aşırı öğrenmesini engelleyen düzeltme terimidir (Wang, Shi, Lyu ve Deng, 2017). Bu terimin modele dâhil edilmesi aşırı öğrenmeyi etkin bir şekilde engellemektedir (Ma ve diğerleri, 2018). Denklem 2'deki topluluk ağaç modeli Öklid uzayında geleneksel metotlarla optimize edilemez. Model toplamsal olarak eğitilir. $\hat{y}_i^{(t)}$ i . örneğin t . iterasyonda kestirimi olmak üzere aşağıdaki amaç fonksiyonunu minimize etmek üzere f_t 'nin eklenmesi gerekir:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

Genel ayarda amaç fonksiyonunu hızlı bir şekilde optimize etmek için ikinci dereceden yaklaşım kullanılabilir:

$$\mathcal{L}^{(t)} \cong \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

Burada $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ ve $h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$ olmak üzere kayıp fonksiyonundaki birinci ve ikinci derece gradyan istatistikleridir. t . adımda sabit terim çıkarılarak aşağıdaki daha basit amaç fonksiyonunu elde edilir:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

$I_j = \{i | q(x_i) = j\}$ yaprak j 'nin örnek kümesi olarak tanımlanır. Ω açılarak denklem 5 aşağıdaki gibi yazılabilir:

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$

$$= \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T$$

Sabit yapı $q(x)$ için j . yaprağın optimal ağırlığı w_j^* aşağıdaki gibi hesaplanır:

$$w_j^* = \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

ve uyan optimum değer:

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T$$

Her karar ağacı ayrıldığında, bir dal iki dala ayrılır. Bölme sonucu elde edilen sağ ve sol düğümün I_R ve I_L olduğunu varsayalım. Bölmenin bilgi kazancı (\mathcal{L}_{split}), bölünmeden önce ve sonra amaç fonksiyonları karşılaştırılarak aşağıdaki gibi hesaplanabilir:

$$\mathcal{L}_{split} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma$$

3.2. MARS

Birden çok veri noktası kullanılarak model tahmini genellikle başarılı kestirimlerin elde etmesini sağlasa da hatalı sonuçların elde edilmesine de yol açabilmektedir (Alkan, Genç, Oktay ve Çelik, 2013). Özellikle doğrusal olmayan ilişkilerin modellenmesinde klasik doğrusal modellerin kullanılması hata oranını artırmaktadır. Bu durumda doğrusal olmayan modellerin tercih edilmesi daha uygundur. MARS modeli, doğrusal olmayan modellerin oluşturulmasında kullanılacak yöntemler arasında yer almaktadır. Bu model, parametrik olmayan bir regresyon modelidir (Eyduvan, Akkus, Kazim, Tırınk ve Tariq, 2017). Bağımlı değişken ve bağımsız değişken arasındaki ilişki için belirli bir fonksiyonel kalıba gereksinim duyulmamaktadır (Y. Chen, Lin, Chen ve Wu, 2019). Yöntem, Friedman (1991) tarafından geliştirilmiştir. Bu yöntem, lineer regresyon modeli gibi kısıtlayıcı varsayımların sağlanmasını gerektirmez (Akin ve diğerleri, 2020). Bu modelde doğrusal olmama parçalı doğrusal regresyon doğrularının tahminiyle yani bağımsız değişken değerinin daha küçük alt değerlere bölünmesi ve farklı alt değerler için farklı regresyon katsayılarının tahminiyle sağlanır (Miguéis, Camanho ve Falcão e Cunha, 2013). Elde edilen farklı doğrular düğüm noktalarıyla birleştirilir. MARS yönteminin kestirim modellerinin oluşturulmasındaki gücü, izinsiz ağa giriş tespiti, elektrik fiyat öngörüsü, kanser teşhisi, yazılım mühendisliği ve kredi puanı hesaplama gibi birçok uygulamada gösterilmiştir (Lu, Lee ve Lian, 2012). Bu yöntemde modele değişkenler tek tek dâhil edilebildiği gibi birden çok değişken çarpılarak da dâhil edilebilmektedir.

Genel bir MARS modeli aşağıdaki gibi yazılabilir (Friedman, 1991):

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m B_m(x)$$

$$B_m(x) = \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]$$

Bu gösterimde a_0 , sabit terim; a_m 'ler regresyon parametreleri; $B_m(x)$ 'ler temel fonksiyonlar; x , bağımsız değişkenler ve t_{km} 'ler, düğüm noktalarıdır. M , temel fonksiyon sayısı, K_m , bölünme sayısı; s_{km} , ± 1 değerlerini alan ve dalların yönünü gösteren terimdir.

Model seçiminde aşağıda fonksiyonel biçimi sunulan, Craven ve Wahba (1978) tarafından önerilen LOF (Lack of fit - Uyumsuzluk) veya diğer bir ifadeyle GCV (Generalized Cross - Validation - Genelleştirilmiş Çapraz Doğrulama) kriteri kullanılmaktadır:

$$LOF(\hat{f}_M) = GCV(M) = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_M(x_i)]^2 / [1 - \frac{C(M)}{n}]^2$$

Bu gösterimde y_i , gözlenen değerler; $\hat{f}_M(x_i)$, kestirim değerleri ve n , gözlem sayısıdır. $C(M)$, modelin büyüklüğünün maliyet fonksiyonudur ve modelin aşırı öğrenmesini engeller ve B , $M \times n$ boyutlu veri matrisi olmak üzere $C(M) = \text{iz}(B(B^T B)^{-1} B^T) + 1$ eşitliğiyle hesaplanır.

MARS yönteminde modelin oluşturulmasında genelden özele doğru bir yaklaşım benimsenir. Öncelikle tüm alt bölünmeler iki kardeş alt bölüme ayrılır ve bölünme uyum iyiliği kriterleri kullanılarak optimize edilir. Bu süreç her bir alt bölgenin temel fonksiyonu ile çok sayıda ayrık alt bölgeler elde edilinceye kadar tekrarlı olarak devam eder. Bu şekilde aşırı öğrenmiş bir model oluşturulur. Daha sonra geriye doğru modelin uyumuna belirgin katkısı olmayan temel fonksiyonlar budanır. Budama sırasında hem uygunluk hem de modelin büyüklüğü kontrol altında tutulur. Bu şekilde modelin karmaşıklığı azaltılır (Katrıs, 2020).

MARS modeli oluşturulduktan sonra modelde kullanılan bağımsız değişkenlerin görece önemi belirlenebilmektedir. Bunu sağlamak amacıyla görece önemi belirlenmek istenen değişkeni içeren tüm terimler modelden çıkarılır ve değişkenin modelden çıkarılmasından kaynaklanan uyum iyiliği kaybı hesaplanır. Uyum iyiliği kaybı tüm değişkenler için bu şekilde hesaplanır ve her bir değişkene 0 ile 100 arasında skorlar verilir. Değişkenlerin aldığı skorun büyüklüğü değişkenin önemini gösterir (Lu ve diğerleri, 2012).

4. Değişkenler ve Veri Seti

4.1. Araştırma Etiği

Çalışmada bilimsel etik kurallarına uyulmuş; analizlerde kullanılan verilere ait kaynaklar belirtilmiş, veriler üzerinde herhangi bir tahrifat yapılmamıştır. Kaynak gösterimde kaynak gösterim kurallarına uyulmuştur.

4.2. Kullanılan Değişkenler

Altın pek çok amaçla kullanılan bir emtia olduğundan altın fiyatı finansal ve makroekonomik pek çok faktörle ilişkilidir. Bu nedenle literatürde altın fiyatının kestiriminde pek çok değişken kullanılmaktadır. Bu değişkenler arasında gümüş fiyatı (Alameer ve diğerleri, 2019; H.-H. Chen ve diğerleri, 2016; Fong-Ching ve diğerleri, 2020; Öndes ve Oğuzlar, 2019; Sami ve Nazir, 2018; Weng ve diğerleri, 2020), borsa endeksleri ve özellikle S&P Endeksi (Kocatepe ve Yıldız, 2016; Kristjanpoller ve Minutolo, 2015; Pierdzioch, Risse ve Rohloff, 2016; Risse, 2019; Sivalingam ve diğerleri, 2016), petrol fiyatı (Gangopadhyay ve diğerleri, 2016; Ghalayini ve Farhat, 2020; Mombeini ve Yazdani Chamzini, 2015; Yüksel ve Akkoç, 2016) ABD Doları fiyatı ile ilgili oran ve endeksler (Fong-Ching ve diğerleri, 2020; Kocatepe ve Yıldız, 2016; Mombeini ve Yazdani Chamzini, 2015; Sivalingam ve diğerleri, 2016), faiz oranı (Ghalayini ve Farhat, 2020; Mombeini ve Yazdani Chamzini, 2015; Sami ve Nazir, 2018) ve TÜFE veya enflasyon (Alameer ve diğerleri, 2019; Gangopadhyay ve diğerleri, 2016; Risse, 2019; Yüksel ve Akkoç, 2016) öne çıkan değişkenlerdir. Bu nedenle bu çalışmada bu değişkenler kullanılarak altın fiyatının kestirimi amaçlanmıştır. Çalışmada kullanılan değişkenler ve veri kaynakları aşağıdaki tabloda sunulmuştur:

Tablo 1. Kullanılan Değişkenler ve Veri Kaynakları

Kısaltma	Değişken	Kaynak
Altın	Altın Spot Fiyatı (ons/\$)	www.investing.com
Gümüş	Gümüş Fiyatı (ons/\$)	www.investing.com
Petrol	Ham Petrol WTI Vadeli İşlemleri Fiyatı (Varil/\$)	www.investing.com
Dolar	ABD Doları Endeksi	www.investing.com
Borsa	S&P500 Endeksi	www.investing.com
Faiz	ABD Federal Fonlar Bileşik Faiz Oranı	www.investing.com
TÜFE	ABD TÜFE (Tüm Kalemler – Tüm Şehirli Tüketiciler)	www.bls.gov

Kullanılan veriler TÜFE hariç 01 Ocak 2015 – 30 Haziran 2020 dönemi 1434 günlük verilerdir. TÜFE ise Ocak 2015 – Haziran 2020 dönemi aylık verilerdir. Oluşturulan model matematiksel olarak aşağıdaki gibi yazılabilir:

$$Altın = f(Gümüş, Petrol, Dolar, Borsa, Faiz, TÜFE)$$

5. Bulgular

Çalışmada her üç modelin tahmininde de *R* yazılımı kullanılmıştır. XGBoost modelinin tahmininde Tianqi Chen (2020) tarafından geliştirilen *xgboost* paketi kullanılmıştır. MARS modelinin tahmininde ise Milborrow (2019) tarafından geliştirilen *earth* paketi kullanılmıştır. Her iki yöntem de kendine özgü parametreler içermektedir. Parametrelerin doğru belirlenmesi modellerin kestirim başarısını doğrudan etkilemektedir. Her iki modele ait parametreler de Kuhn (2020) tarafından geliştirilen *caret* paketi yardımıyla belirlenmiştir. MARS modelinin tahmininde Eyduran ve Duman (2020) ve Boehmke ve Greenwell (2020) tarafından oluşturulan kaynaklardan yararlanılmıştır. XGBoost modelinin tahmininde ise Kaggle sayfasında yer alan "Visual XGBoost Tuning with caret" başlıklı çalışmadan yararlanılmıştır. Modellerin kestirim performanslarının karşılaştırılmasında Gulbe ve Eyduran (2019) tarafından geliştirilen *ahaGoF* paketinden yararlanılmıştır. Çalışmada kullanılan veriler rassal olarak %80 eğitim, %20 test veri olarak ikiye ayrılmıştır. Model tahminleri sadece eğitim verisi kullanılarak yapılmış, modellerin kestirim performansları hem eğitim hem de test modelleri için hesaplanmıştır. Modellerin karşılaştırılmasında ise test modellerine ait kriterler kullanılmıştır.

MARS modelinde belirlenmesi gereken 2 parametre bulunmaktadır. Bunlar:

nprune: Budanmış modelde yer alacak en fazla terim sayısı

degree: Değişkenlerin etkileşim derecesidir.

nprune parametresi için 2 ile 100 aralığından 23 olarak, "degree" ise 1 ile 3 aralığından 2 olarak belirlenmiştir.

XGBoost modelindeki parametreler ise aşağıdaki gibidir:

nrounds: En yüksek tekrar sayısı

max_depth: Her bir ağacın maksimum derinliği

eta: Öğrenme oranı

gamma: Ağacın bir yaprak düğümü üzerinde bir başka bölüm oluşturmak için gereken kayıp fonksiyonunda meydana gelecek en küçük azalma miktarı

colsample_bytree: Her bir ağacı oluştururken sütunların alt örnek oranı

min_child_weight: En küçük yaprak ağırlığı

subsample: Eğitim örneğinin alt örnek oranı

XGBoost modelinde değeri belirlenmesi gereken parametre sayısının fazla olması nedeniyle bu modeldeki tüm parametrelerin tek seferde sıradan bir masaüstü bilgisayar kullanılarak belirlenmesi aşırı zaman gerektirmektedir. Bu nedenle parametreler aşamalı olarak belirlenmiştir. Parametreler için denenen değerler ve seçilen en uygun değerler aşağıdaki tabloda sunulmuştur:

Tablo 2. XGBoost Modeli İçin Parametreler

Parametre	Alternatif Değerler	En Uygun Değer
nrounds	[100, 10000] aralığı	9700
max_depth	2, 3, 4, 5, 6	6
eta	0,01, 0,015, 0,025, 0,05, 0,1, 0,3	0,025
gamma	0, 0,05, 0,1, 0,5, 0,7, 0,9, 1	0,7
colsample_bytree	0,4, 0,6, 0,8, 1	1
min_child_weight	1, 2, 3	3
subsample	0,5, 0,75, 1	0,75

Tahmin edilen tüm modellere ait performans değerlendirme kriterleri aşağıdaki tabloda sunulmuştur:

Tablo 3. Modellere Ait Performans Değerlendirme Kriterleri Değerleri

Kriter	Eğitim			Test		
	XGBoost Değer	MARS Değer	LR Değer	XGBoost Değer	MARS Değer	LR Değer
RMSE	0,114	381,590	1376,369	89,704	447,641	1317,563
RRMSE	0,026	1,502	2,853	0,729	1,628	2,793
SDR	0,002	0,135	0,256	0,063	0,140	0,238
CV	0,030	1,500	2,850	0,730	1,630	2,780
PC	1,000	0,991	0,967	0,998	0,990	0,971
PI	0,013	0,754	1,451	0,365	0,818	1,417
MAPE	0,021	1,130	2,272	0,512	1,245	2,223
MAD	0,272	14,612	29,190	6,667	16,166	28,651
Rsqr	1,000	0,982	0,935	0,996	0,980	0,943
ARsqr	1,000	0,982	0,935	0,996	0,978	0,943
AIC	-2483,92	6858,17	8293,60	1294,50	1791,85	2065,68
CAIC	-2483,91	6858,91	8293,61	1294,54	1795,00	2065,72

Yukarıdaki tabloda görüldüğü gibi XGBoost ve MARS modelleri LR modeline göre çok daha başarılı kestirimlerin elde edilmesini sağlamıştır. Düzeltilmiş belirlenme katsayısına göre yorumlanacak olursa tahmin edilen XGBoost modeli altın fiyatında meydana gelen toplam değişimin %99,6'sını, MARS modeli ise %97,8'ni açıklamaktadır. LR için ise açıklanma oranı %94,3'tür. Bu sonuçlar tercih edilen yöntem ve değişkenlerin altın fiyatının belirlenmesi için oldukça uygun olduğunu göstermektedir. XGBoost ve MARS modellerinin her ikisi de çok başarılı olsa da XGBoost yöntemi daha başarılı sonuçların

elde edilmesini sağlamıştır. Her ne kadar XGBoost, bu çalışmada kullanılan diğer yöntemlere göre parametrelerin belirlenmesinin ve modelin eğitiminin aşırı zaman alması gibi bir dezavantaja sahip olsa da oldukça başarılı kestirimler elde edilmesini sağlamaktadır. Elde edilen bulgulara göre XGBoost yönteminin, altın fiyatının kestirimi için uygun bir yöntem olduğu söylenebilir.

XGBoost ve MARS yöntemleri kullanılan değişkenlerin bağımlı değişken üzerindeki etkisinin belirlenmesine ve değişkenlerin kendi aralarında sıralanmasına olanak sağlamaktadır. Değişkenlerin önem seviyesinin belirlenmesinde MARS modelinde değişkenlerin geliştirilmiş çapraz doğrulama skorunu katkıları; XGBoost modelinde ise değişkenlere ait bilgi kazançları kullanılmaktadır. Her iki model için değişkenlere ait skorlar ve sıralamalar aşağıdaki tabloda sunulmuştur:

Tablo 4. Değişkenlerin Görelî Önemleri

	XGBoost		MARS	
	Bilgi Kazancı	Sıra	GCV Katkısı	Sıralama
TÜFE	0,886	1	100	1
Gümüş	0,049	2	37,7	4
Faiz	0,024	3	56,2	3
Dolar	0,018	4	30,6	5
Petrol	0,017	5	17,8	6
Borsa	0,006	6	56,2	2

MARS yönteminde elde edilen skorlar, tüm değişkenler için birbirlerine nispeten yakın olmasına rağmen XGBoost yöntemi için diğer değişkenlere ait skorlar TÜFE için elde edilen skora göre oldukça küçüktür. Her iki yöntemde göre de altın fiyatının belirlenmesinde en önemli değişken TÜFE'dir. Diğer değişkenler ait sıralamalar ise iki yöntemde göre farklılaşmaktadır. XGBoost için sıralama TÜFE, Gümüş, Faiz, Dolar, Petrol ve Borsa şeklindedir. MARS için ise TÜFE, Borsa, Faiz, Gümüş, Dolar, Petrol şeklindedir.

6. Sonuç

Altın önemli bir ödeme, yatırım ve birikim aracı olduğundan fiyatının belirlenmesi ülkeler ve yatırımcılar için önemlidir. Fakat altın fiyatı durağandır ve doğrusal olmayan bir seyir izlemektedir. Ayrıca altın fiyatının fonksiyonel kalıbı önsel olarak bilinmemektedir. Bu nedenle altın fiyatının kestiriminde bu özelliklere uygun yöntemler kullanılmalıdır. Çalışmada tercih edilen XGBoost ve MARS yöntemleri bu gibi özelliklere sahip verilerin modellenmesinde başarılı bir şekilde kullanılabilir. Altın bir emtia olarak önemli bir yatırım aracı olduğundan piyasadaki diğer yatırım araçlarının ve emtiaların fiyatından etkilenebilmektedir. Bu amaçla altın fiyatı üzerinde etkili olduğu düşünülen gü-

müş fiyatı (ons/\$), ham petrol WTI vadeli işlemleri fiyatı (varil/\$), ABD Doları endeksi, S&P500 endeksi, ABD federal fonlar bileşik faiz oranı, ABD TÜFE (tüm kalemler – tüm şehirli tüketiciler) değişkenleri girdi olarak kullanılmıştır. Çalışmada lineer regresyon, XGBoost ve MARS modelleri tahmin edilmiş – eğitilmiştir. Bu modeller ile altın fiyatına ait kestirimler oluşturulmuş, elde edilen kestirimler kullanılarak modellere ait performans değerlendirme kriterleri hesaplanmıştır.

Lineer regresyon modeli başarılı kestirimlerin elde edilmesini sağlasa da MARS ve XGBoost modelleri nispeten daha başarılı sonuçların elde edilmesini sağlamıştır. Yöntemler arasında en iyi sonuçlar XGBoost modeli aracılığıyla elde edilmiştir. Bu sonuç, XGBoost modelinin kullanıldığı benzer çalışmalarda elde edilen sonuçları destekleyici niteliktedir. Kullanılan değişkenler ve XGBoost modeli %99,6; MARS modeli ise %97,8 başarı oranıyla altın fiyatının kestirilmesini sağlamıştır. Bu oranlar kullanılan değişkenler ve modellerin uygunluğunu göstermektedir.

MARS ve XGBoost modelleri kullanılan girdi değişkenlerinin çıktı değişkeni üzerindeki etkisinin önemini belirlenmesine olanak sağlamaktadır. Her iki yöntem için de girdi değişkenlerinin altın fiyatı üzerindeki etkisi belirlenmiştir. Her iki yöntem için elde edilen sonuçlar altın fiyatı üzerinde en önemli etkiye sahip faktörün ABD TÜFE olduğunu göstermektedir. Diğer değişkenler için yöntemler farklı sonuçlar elde edilmesini sağlamıştır. XGBoost için sıralama ABD TÜFE, gümüş fiyatı, ABD federal fonlar bileşik faiz oranı, ABD Doları endeksi, ham petrol WTI vadeli işlemleri fiyatı ve S&P500 endeksi şeklindedir. MARS için ise ABD TÜFE, S&P500 endeksi, ABD federal fonlar bileşik faiz oranı, gümüş fiyatı, ABD Doları endeksi, ham petrol WTI vadeli işlemleri fiyatı şeklindedir.

Bu çalışmada altın fiyatının belirlenmesinde kullanılan pek çok girdi ABD'ye aittir. Şüphesiz ki altın fiyatı ABD yanında Çin, Rusya ve Hindistan gibi yüksek miktarda altın talep eden ülkelere ait faktörlerden de etkilenmektedir. Bu açıdan altın fiyatının kestiriminde ABD ve diğer ülkelere ait faktörlerin ağırlıklandırılarak girdi olarak kullanılması altın fiyatının kestirimini amaçlayan araştırmacılara öneri olarak sunulabilir. Bunun yanında bu çalışmada kullanılmayan farklı yatırım araçlarına ait veriler de kullanılan modellerde girdi olarak kullanılabilir.

Kaynakça

- Akin, M., Eyduran, S. P., Eyduran, E. ve Reed, B. M. (2020). Analysis of macro nutrient related growth responses using multivariate adaptive regression splines. *Plant Cell, Tissue and Organ Culture (PCTOC)*, 140(3), 661-670.
- Alameer, Z., Elaziz, M. A., Ewees, A. A., Ye, H. ve Jianhua, Z. (2019). Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. *Resources Policy*, 61, 250-260.
- Alkan, Ö., Genç, A., Oktay, E. ve Çelik, A. K. (2013). Electricity consumption analysis using spline regression models: the case of a Turkish province. *Asian Social Science*, 9, 231-240.

- Alkan, Ö., Oktay, E., Genç, A. ve Çelik, A. K. (2017). An analysis of export-import coverage ratios in Turkey using spline regression models. *Ekonomika Istraživanja / Economic Research*, 30, 223-237.
- Baur, D. G. ve McDermott, T. K. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance*, 34(8), 1886-1898.
- Boehmke, B. ve Greenwell, B. (2020). *Hands-on machine learning with R* (1 ed.): Chapman and Hall/CRC
- Bouri, E., Jain, A., Biswal, P. C. ve Roubaud, D. (2017). Cointegration and nonlinear causality amongst gold, oil, and the Indian stock market: Evidence from implied volatility indices. *Resources Policy*, 52, 201-206.
- Carmona, P., Climent, F. ve Momparler, A. (2019). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 61, 304-323.
- Carvalho, A. ve Ribeiro, T. (2008). Do artificial neural networks provide better forecasts? Evidence from Latin American stock indexes. *Latin American Business Review*, 8(3), 92-110.
- Chen, H.-H., Chen, M. ve Chiu, C.-C. (2016). The integration of artificial neural networks and text mining to forecast gold futures prices. *Communications in Statistics - Simulation and Computation*, 45(4), 1213-1225.
- Chen, T. ve Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Retrieved from <https://doi.org/10.1145/2939672.2939785>
- Chen, Y., Lin, J., Chen, Y. ve Wu, J. (2019). Financial forecasting with multivariate adaptive regression splines and queen genetic algorithm-support vector regression. *IEEE Access*, 7, 112931-112938.
- Craven, P. ve Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4), 377-403.
- Değirmenci, N. ve Akay, A. (2017). Finansal verilerin ARIMA ve ARCH modelleriyle öngörüsü: Türkiye örneği. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 12.
- Eyduran, E., Akkus, O., Kazim, M., Tırınk, C. ve Tariq, M. (2017). *Use of multivariate adaptive regression splines (Mars) in predicting body weight from body measurements in Mengali rams*. Paper presented at the International Conference on Agriculture, Forest, Food Sciences and Technologies.
- Eyduran, E. ve Duman, H. (2020). R yazılımı ile multivariate adaptive regression splines (MARS) uygulaması ders notları.

- Fong-Ching, Y., Chao-Hui, L. ve Chaochang, C. (2020). Using market sentiment analysis and genetic algorithm-based least squares support vector regression to predict gold prices. *International Journal of Computational Intelligence Systems*, 13(1), 234-246.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist.*, 19(1), 1-67.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.*, 29(5), 1189-1232.
- Gangopadhyay, K., Jangir, A. ve Sensarma, R. (2016). Forecasting the price of gold: An error correction approach. *IIMB Management Review*, 28(1), 6-12.
- Ghalayini, L. ve Farhat, S. (2020). Modeling and forecasting gold prices: Research Square. <https://doi.org/10.21203/rs.3.rs-23825/v1> adresinden erişildi doi:10.21203/rs.3.rs-23825/v1
- Gulbe, A. ve Eyduran, E. (2019). ehaGoF: Calculates goodness of fit statistics (Version 0.1.0) [R Package].
- Katris, C. (2020). Prediction of unemployment rates with time series and machine learning techniques. *Computational Economics*, 55(2), 673-706.
- Keskin Benli, Y. ve Yıldız, A. (2014). Altın fiyatının zaman serisi yöntemleri ve yapay sinir ağları ile öngörüsü. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi / Dumlupınar University Journal of Social Sciences*, 213-224.
- Khan, M. M. A. (2013). Forecasting of gold prices (Box Jenkins approach). *International Journal of Emerging Technology and Advanced Engineering*, 3(3), 662-670.
- Kocatepe, C. İ. ve Yıldız, O. (2016). Ekonomik endeksler kullanılarak Türkiye'deki altın fiyatındaki değişim yönünün yapay sinir ağları ile tahmini. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 4, 926-934.
- Kristjanpoller, W. ve Minutolo, M. C. (2015). Gold price volatility: A forecasting approach using the artificial neural network–GARCH model. *Expert Systems with Applications*, 42(20), 7245-7251.
- Kuhn, M. (2020). caret: classification and regression training (Version 6.0-86) [R Package].
- Livieris, I. E., Pintelas, E. ve Pintelas, P. (2020). A CNN–LSTM model for gold price time-series forecasting. *Neural Computing and Applications*.
- Lu, C.-J., Lee, T.-S. ve Lian, C.-M. (2012). Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks. *Decision Support Systems*, 54(1), 584-596.
- Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q. ve Niu, X. (2018). Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost

- algorithms according to different high dimensional data cleaning. *Electronic Commerce Research and Applications*, 31, 24-39.
- Miguéis, V. L., Camanho, A. ve Falcão e Cunha, J. (2013). Customer attrition in retailing: An application of multivariate adaptive regression splines. *Expert Systems with Applications*, 40(16), 6225-6232.
- Milborrow, S. (2019). earth: multivariate adaptive regression splines (Version 5.1.2) [R package].
- Mombeini, H. ve Yazdani Chamzini, A. (2015). Modeling gold price via artificial neural network. *Journal of Economics, Business and Management*, 3, 699-703.
- Oktay, E., Genç, A. ve Alkan, Ö. (2012). İhracatın İthalatı Karşılama Oranlarının Parçalı Regresyonlarla Modellenmesi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 16, 1-15.
- Oktay, E., Talas, E., Alkan, Ö. ve Genç, A. (2012). Modeling with Linear Spline Regression of Turkish Tourism Demand. *Journal of Selçuk University Natural and Applied Science*, 1, 10-22.
- Öndes, H. ve Oğuzlar, A. (2019). Yapay sinir ağlarıyla altın (TL/kg) fiyatı tahmini. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler E-Dergisi*(72), 249-262.
- Pandey, A. C., Misra, S. ve Saxena, M. (2019, 8-10 Aug. 2019). *Gold and diamond price prediction using enhanced ensemble learning*. Paper presented at the 2019 Twelfth International Conference on Contemporary Computing (IC3).
- Pierdzioch, C., Risse, M. ve Rohloff, S. (2016). A boosting approach to forecasting gold and silver returns: Economic and statistical forecast evaluation. *Applied Economics Letters*, 23(5), 347-352.
- Risse, M. (2019). Combining wavelet decomposition with machine learning to forecast gold returns. *International Journal of Forecasting*, 35(2), 601-615.
- Sami, I. ve Nazir, K. (2018). Predicting future gold rates using machine learning approach. *International Journal of Advanced Computer Science and Applications*, 8, 92-99.
- Sivalingam, K. C., Mahendran, S. ve Natarajan, S. (2016). Forecasting gold prices based on extreme learning machine. *International Journal of Computers Communications & Control*, 11(3).
- Tianqi Chen, T. H. (2020). XGboost: extreme gradient boosting (Version 1.1.1.1) [R Package].
- Verma, S., Thampi, G. T. ve Rao, M. (2020). ANN based method for improving gold price forecasting accuracy through modified gradient descent methods. *IAES International Journal of Artificial Intelligence*, 9(1), 46-57.

- Visual XGBoost tuning with caret. (2020). Retrieved 08.07.2020, 2020, from <https://www.kaggle.com/pelkoja/visual-xgboost-tuning-with-caret>
- Wang, W., Shi, Y., Lyu, G. ve Deng, W. (2017). *Electricity consumption prediction using XGBoost based on discrete wavelet transform*. Paper presented at the 2nd International Conference on Artificial Intelligence and Engineering Applications (AIEA 2017).
- Weng, F., Chen, Y., Wang, Z., Hou, M., Luo, J. ve Tian, Z. (2020). Gold price forecasting research based on an improved online extreme learning machine algorithm. *Journal of Ambient Intelligence and Humanized Computing*.
- Yüksel, R. ve Akkoç, S. (2016). Altın fiyatlarının yapay sinir ağları ile tahmini ve bir uygulama. *Doğuş Üniversitesi Dergisi*, 17(1), 39-50.
- Zhou, Y., Li, T., Shi, J. ve Qian, Z. (2019). A CEEMDAN and XGBOOST-based approach to forecast crude oil prices. *Complexity*, 2019, 4392785.