

# Zaman Serisi Verilerini Kullanarak Makine Öğrenmesi Yöntemleri ile Bitcoin Fiyat Tahmini

## Prediction of Bitcoin Prices with Machine Learning Methods using Time Series Data

Seçkin KARASU, Aytaç ALTAN, Zehra SARAÇ, Rıfat HACIOĞLU

Elektrik Elektronik Mühendisliği Bölümü

Bülent Ecevit Üniversitesi

Zonguldak, Türkiye

seckin.karasu@beun.edu.tr, aytacaltan@beun.edu.tr, saraczehra@yahoo.com, hacirif@beun.edu.tr

**Özetçe—** Bu çalışmada, makine öğrenmesi yöntemlerinden Doğrusal Regresyon (DR) ve Destek Vektör Makinesi (DVM) ile 2012-2018 yılları arasındaki günlük Bitcoin kapanış fiyatlarından oluşan zaman serisi kullanılarak Bitcoin tahmini yapılmaktadır. En az hata içeren tahmin modeli doğrusal ve polinom çekirdek fonksiyonlar içeren DVM gibi farklı parametre kombinasyonları ile test edilerek elde edilmektedir. Farklı pencere uzunlukları için Bitcoin fiyat tahmini, farklı ağırlık katsayılarına sahip süzgeçler kullanılarak yapılmaktadır. Veri setinden bağımsız başarıyı yüksek bir model oluşturulabilmesi için eğitim aşamasında 10 kat çapraz doğrulama yöntemi kullanılmaktadır. Elde edilen modelin başarımı, istatistiksel göstergelerden Ortalama Mutlak Hata, Ortalama Karesel Hata, Ortalama Hata Kareleri Karekökü, Pearson korelasyon yardımıyla ölçülmektedir. Önerilen DVM modelinin DR modeline kıyasla Bitcoin veri seti için tahmin başarımının yüksek olduğu görülmektedir.

**Anahtar Kelimeler—**Bitcoin; Kriptopara Birimi; Fiyat Tahmini; Destek Vektör Makinesi; Makine Öğrenmesi.

**Abstract—** In this study, Bitcoin prediction is performed with Linear Regression (LR) and Support Vector Machine (SVM) from machine learning methods by using time series consisting of daily Bitcoin closing prices between 2012-2018. The prediction model with include the least error is obtained by testing with different parameter combinations such as SVM with including linear and polynomial kernel functions. Filters with different weight coefficients are used for different window lengths. For different window lengths, Bitcoin price prediction is made using filters with different weight coefficients. 10-fold cross-validation method in training phase is used in order to construct a model with high performance independent of the data set. The performance of the obtained model is measured by means of statistical indicators such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Pearson Correlation. It is seen that the price prediction performance of the proposed SVM model for Bitcoin data set is higher than that of the LR model.

**Keywords—** Bitcoin; Cryptocurrency; Price Prediction; Support Vector Machine (SVM), Machine Learning.

### I. GİRİŞ

Mal ve hizmet alışverişinde kullanılan paranın evrimi değiş-tokuş yöntemlerinden sırasıyla emtia para, altın ve gümüş, değerli kağıt, güvene dayalı itibari para ve dijital/sanal paralara doğru yol almaktadır [1]. Nakamoto'nun sistematik yapısal özelliklerini temel alarak eşsiz protokolü ile finansal piyasaya başarılı şekilde sunulan kriptopara birimi olan Bitcoin, tamamen merkezi olmayan ödeme sistemi olarak tanımlanmaktadır [2, 3]. Bu çalışmada, 2012-2018 yılları arasındaki günlük Bitcoin kapanış fiyatlarından oluşan zaman serisi kullanılarak makine öğrenmesi yöntemlerinden Doğrusal Regresyon (Linear Regression-LR) ve Destek Vektör Makinesi (Support Vector Machine-SVM) ile Bitcoin fiyat tahmini yapılmaktadır.

Kim tarafından yapılan finansal zaman serisi tahmini çalışmasında, SVM yöntemi ile elde edilen model durum temelli muhakeme (Case-Based Reasoning-CBR) ve geri beslemeli sinir ağı modelleri kıyaslanmıştır. Çalışmada, farklı çekirdek parametresi ve üst limit değerleri için farklı SVM modelleri oluşturulmuş ve oluşturulan modelin finansal zaman serisi tahmininde diğer modellere kıyasla tatminkâr başarımlar sergilediği gösterilmiştir [4]. Patel ve arkadaşları tarafından makine öğrenmesi yöntemlerine dayalı iki aşamalı yaklaşımdan oluşan melez bir model önerilmiştir. Önerilen modelin ilk aşamasında, Destek Vektör Regresyon (Support Vector Regression-SVR) yöntemi ile gelecekteki istatistiksel parametre değerlerinin tahmini yapılmıştır. Modelin ikinci aşamasında, tahmin edilen parametre değerlerini giriş olarak kullanan Yapay Sinir Ağı (Artificial Neural Network-ANN), Rastgele Orman (Random Forest-RF) ve SVR yöntemleri ile iki ayrı borsanın farklı gelecek adım değerleri için endeks tahmini yapılmıştır. Önerilen melez modelin girişinde 10 adet teknik gösterge değerleri kullanılmış, çıkışında ise borsa endeksinin değerinin tahmini yapılmıştır [5]. Wei tarafından yapılan çalışmada, Görgül Kip Ayrışımı (Empirical Mode Decomposition-EMD) ve Uyarlamalı Ağ Tabanlı Bulanık

Çıkarım Sistem (Adaptive Network-based Fuzzy Inference System-ANFIS) birleşimine dayalı bir borsa tahmin modeli geliştirilmektedir. Geliştirilen modelin ilk aşamasında borsa verilerine uygulanan EMD yöntemi ile içsel mod fonksiyon (Intrinsic Mode Function-IMF) bileşenlerine ayrılmaktadır. Ayrılan bileşenler oluşturulan ANFIS model girişine uygulanarak bir sonraki adım için borsa endeks değerinin tahmini yapılmaktadır. Elde edilen melez modelin başarımının Chen'in modeli [6], Yu'nun modeli [7], Otoregresif (AutoRegressive-AR), SVR ve ANFIS modellerine kıyasla yüksek olduğu belirtilmektedir [8].

Son zamanlarda, kriptopara piyasasında 70 milyar Amerikan Dolarına (United States Dollar-USD) yaklaşan sermayesiyle en yüksek orana sahip olan Bitcoin'in fiyat tahmini üzerine çalışmalar hız kazanmıştır [9]. İndera ve arkadaşları tarafından Bitcoin fiyat tahmini için yapılan çalışmada, Doğrusal Olmayan Otoregresif Dış Kaynaklı (Nonlinear AutoRegressive eXogenous-NARX) yöntemine dayalı Çok Katmanlı Algılayıcı (Multi Layer Perceptron-MLP) modeli kullanılmıştır. Çalışmada, teknik göstergelerden hareketli ortalama (Moving Average-MA) ile açılış, kapanış, en düşük ve en yüksek fiyat verilerini model girişi olarak kullanmışlardır. Parçacık Sürü Optimizasyon (Particle Swarm Optimization-PSO) yöntemi ile model parametrelerini optimize ederek Bitcoin fiyatını tahmin edebilen model önermişlerdir [10]. Almeida ve arkadaşları bir sonraki gün için Bitcoin fiyat yönünü tahmin etmede önceki günlerdeki fiyat ve işlem hacmi verilerini kullanarak ANN'ye dayalı tahmin modeli kullanmışlardır. Çalışmada, oluşturdukları tek katmanlı ileri beslemeli ağ model başarımının çevrimiçi karmaşık modellere göre yüksek olduğunu ifade etmişlerdir [11]. Pichl ve Kaizoji tarafından yapılan çalışmada, on günlük hareketli ortalama kullanılarak iki ara katmanlı ileri beslemeli sinir ağı ile bir sonraki gün için logaritmik kazanç tahmin edilmiştir. Oluşturulan modelin gerçek logaritmik kazanç dağılımını yakalama yeteneğine sahip olduğu belirtilmiştir [9]. Madan ve arkadaşları makine öğrenmesi yöntemleri ile Bitcoin fiyatı ile ilişkili 25 farklı özelliği kullanarak günlük verilerden oluşan beş yıllık Bitcoin fiyat verisi üzerinde fiyat değişimi tahmini yapmışlardır. Elde edilen algoritmanın başarımının %98.7 olduğunu ifade etmişlerdir. Bitcoin fiyat tahminlerini çeşitli hassasiyet düzeylerinde ve gürültüye göre değerlendirmek için 10 dakikalık ve 10 saniyelik zaman noktalarında kaldırma verilerini kullanmışlardır. 10 dakikalık zaman aralıklarında gelecekteki fiyat değişimi tahmini için elde edilen modelin başarımının %50-55 arasında olduğu belirtilmiştir [2]. Jang ve Lee tarafından Bitcoin arz ve talebi ile ilişkili Blockchain bilgisine dayalı Bayesci Sinir Ağı (Bayesian Neural Network-BNN) kullanılarak Bitcoin fiyatlarının tahmini ve modellenmesi üzerine deneysel çalışma yapılmıştır. Oluşturulan modeli diğer doğrusal ve doğrusal olmayan modellerle kıyaslayarak model başarımı ölçmüşlerdir [3]. Greave ve Au tarafından yapılan çalışmada, Blockchain ağına dayalı öznitelikler kullanılarak gelecek Bitcoin yükseliş ve düşüş fiyat hareketleri makine öğrenmesi yöntemi ile sınıflandırılmıştır. Oluşturulan modelin başarımının %55 olduğu belirtilmiştir [12].

Bu çalışmada, LR, doğrusal SVM (L-SVM) ve polinom SVM (P-SVM) modelleri ile bir sonraki Bitcoin kapanış fiyatı tahmini çalışması yapılmaktadır. Modeller iki giriş ve bir çıkıştan oluşmaktadır. Model girişinde MA veya WMA filtrelerinden biri ile birlikte A/D osilatörü uygulanmaktadır. Bu göstergelerin hesaplanması için kapanış, en yüksek veya en düşük fiyat zaman serileri kullanılmaktadır. Çalışmada MA ve WMA filtrelerinde gecikme adım uzunlukları 2, 3, 4, 5 ve 6 olarak alınarak tahmin sürecindeki etkisi de incelenmektedir. Model sonuçları için MSE değeri ile oluşan hata ölçülmektedir. Çalışmada, gerçekleştirilen işlem adımları model oluşturma ve tahmin yöntemi başlığı altında II. bölümde açıklanmakta, farklı modeller ve filtreleme uzunluklarındaki filtreleme yöntemleri için Bitcoin fiyat tahminine ilişkin benzetim sonuçları III. bölümde verilmekte ve elde edilen sonuçlar IV. bölümde tartışılmaktadır.

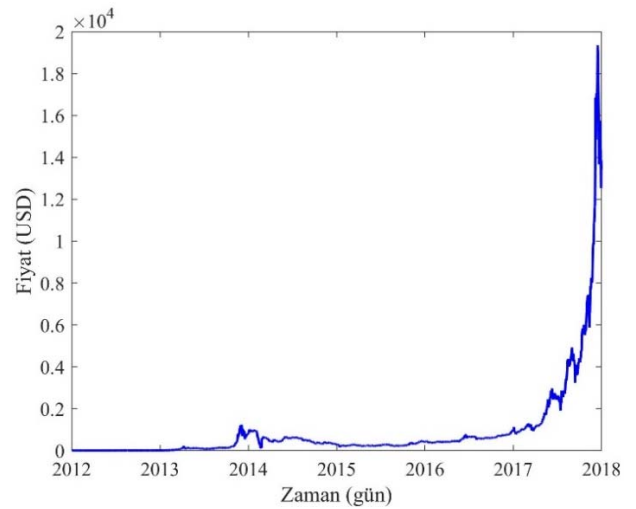
## II. BITCOIN FİYAT TAHMİN MODELİNİN OLUŞTURULMASI

Bu çalışmada, bir sonraki gün için Bitcoin kapanış fiyatının tahmininde 2012-2018 yılları arasındaki zaman serisi verisi kullanılmaktadır. Kullanılan zaman serisinin yıllara göre istatistiksel değerleri Tablo 1'deki gibi verilmektedir.

TABLO I. 2012-2018 YILLARI ARASINDAKİ BITCOIN FİYATININ İSTATİSTİKSEL DEĞERLERİ

Yıllar	Minimum	Maksimum	Ortalama	Standart Sapma
2012	4.22	13.70	930.36	3.21
2013	13.28	1237.55	200.15	260.42
2014	111.56	1014.74	522.89	176.00
2015	164.92	463.18	272.02	58.84
2016	368.02	972.17	567.00	138.16
2017	785.43	19345.49	3981.07	3981.71

Şekil 1'de sırası ile 2012-2018 yılları arasındaki Bitcoin günlük kapanış fiyatı ve hacim grafiği verilmektedir. Tahmin modelinin girişine MA ve ağırlıklı hareketli ortalama (Weighted Moving Average-WMA) filtre, filtre uzunluğu 2 ile 6 arasında değiştirilerek uygulanmaktadır. Ayrıca, A/D osilatörü (Accumulation / Distribution) tahmin modeline ikinci giriş olarak verilmektedir. Tahmin modelinin çıkışı bir sonraki günün tahmini Bitcoin kapanış fiyatı olarak belirlenmektedir.



Şekil. 1. 2012-2018 yılları arasındaki Bitcoin kapanış fiyatı değişimi.

Tahmin modeli, sırasıyla LR, L-SVM ve P-SVM ile oluşturulmaktadır. Tahmin yöntemleri için hata değerleri ölçülmekte ve oluşturulan modeller için hata değişimleri belirlenmektedir.

#### A. Filtreleme İşlemleri

Uzunluğu  $N$  ve filtre katsayısı  $w$  ile gösterilen bir filtre için,  $y$  bağımlı değişkeninden  $y_n^+$  filtrelenmiş değeri

$$y_n^+ = \frac{\sum_{i=1}^N w_i y_i}{\sum_{i=1}^N w_i} \quad (1)$$

ile filtreleme işlemi gerçekleştirilmektedir. Denklem (1)'de MA ve WMA filtre için Tablo 2'deki  $w$  katsayıları kullanılmaktadır.

TABLO II. FİLTRE ÇEŞİTLERİ VE KATSAYILARI

Filtre Çeşidi	Filtre Katsayıları
MA Filtre	$w_i = \{1 \quad i \leq N\}$
WMA Filtre	$w_i = \{i \quad i \leq N\}$

MA filtre için sinyal üzerindeki değerlerin ortalaması, sinyal üzerine uygulanacak filtre uzunluğu kadar geri gidilerek hesaplanmaktadır. WMA filtrede sinyal üzerindeki seçilen uzunluk için filtrelenecek sinyalin en yakın değerinden önceki değere doğru doğrusal olarak azalan ağırlık değerleri kullanılmaktadır [13].

#### B. A/D Osilatörü

Zaman serilerinde döngüsel etkileri gözlemleyebilmek için osilatör tabanlı göstergeler kullanılmaktadır [4]. Bu çalışmada döngüsel etkiler A/D osilatörü ile belirlenerek tahmin modeline giriş olarak uygulanmaktadır. A/D osilatör değeri

$$y_n^+ = \frac{h_i - y_{i-1}}{h_i - l_i} \quad (2)$$

ile belirlenmektedir. Denklem (2)'de  $h$  ile en yüksek fiyat değişkeni,  $l$  ile en düşük fiyat değişkeni ifade edilmektedir.  $y_n^+$  A/D osilatörünün yeni değeri olmak üzere, en yüksek fiyat değeri ile önceki kapanış fiyatı farkının, en yüksek fiyat değeri ile en düşük fiyat değeri farkına oranı olarak hesaplanmaktadır.

#### C. Doğrusal Regresyon Modeli

Doğrusal regresyon yöntemi bağımlı değişkeni tahmin etmek için bağımsız değişken ile bağımlı değişken arasında doğrusal bir ilişki kuran istatistiksel bir yöntemdir [14]. Doğrusal regresyon eşitliği

$$y = a_1 x + b_1 \quad (3)$$

ile verilmektedir. Denklem (3)'te  $x$  bir sonraki gün için fiyat tahmincisini,  $y$  cevap değişkenini,  $a_1$  regresyon çizgisinin eğimini ve  $b_1$  regresyon çizgisinin ordinat eksenini kestiği noktayı ifade etmektedir.

#### D. Destek Vektör Makineleri

Destek vektör makineleri, Vladimir Vapnik ve Alexey Chervonenkis tarafından geliştirilen, karar vermede bağımlı destek noktaları arasındaki genişliği en yükseğe çıkarmak için kullanılan ve örüntü tanıma, zaman serileri analizi,

sınıflandırma gibi uygulamalarda başarıyla kullanılan gözetimli öğrenme algoritmalarıdır [15]. SVM hem doğrusal hem de doğrusal olmayan verilere uygulanabilmektedir. Doğrusal olmayan  $x$  veri dizisini daha yüksek boyuta taşıyacak olan fonksiyon haritalama fonksiyonu olarak tanımlanmakta ve  $\Phi(x)$  ile gösterilmektedir. Bu durumda, regresyon süreci, çekirdek fonksiyonu  $K(x_i, x_j)$  ile yüksek boyutlu alana taşınmaktadır [15]. Bu çalışmada, Tablo 3'te verilen doğrusal ve polinom çekirdek fonksiyonları SVM modelinde kullanılmaktadır. Tablo 3'te  $p$  polinom derecesini ifade etmektedir. Burada  $p=2$  olarak alınmakta ve  $C$  katsayısı 0,0207 olarak seçilmektedir.

TABLO III. KULLANILAN ÇEKİRDEK FONKSİYONLARI

Çekirdek Tipi	Eşitlik
Doğrusal	$K(x_i, x_j) = \Phi(x_i, x_j)^T \Phi(x_i, x_j)$
Polinom	$K(x_i, x_j) = (\Phi(x_i, x_j)^T \Phi(x_i, x_j) + 1)^p$

#### E. İstatistiksel Hata Ölçütleri

Bitcoin kapanış fiyatı tahmini sonucunda oluşan hatayı hesaplamak için Tablo 4'te verilen istatistiksel hata ölçütleri kullanılmaktadır. Belirli  $t$  zamanı için  $r_t$  gözlemlenen ve  $p_t$  tahmin edilen zaman serisi olmak üzere hata  $e_t$  (4) eşitliği ile ifade edilmektedir [16].

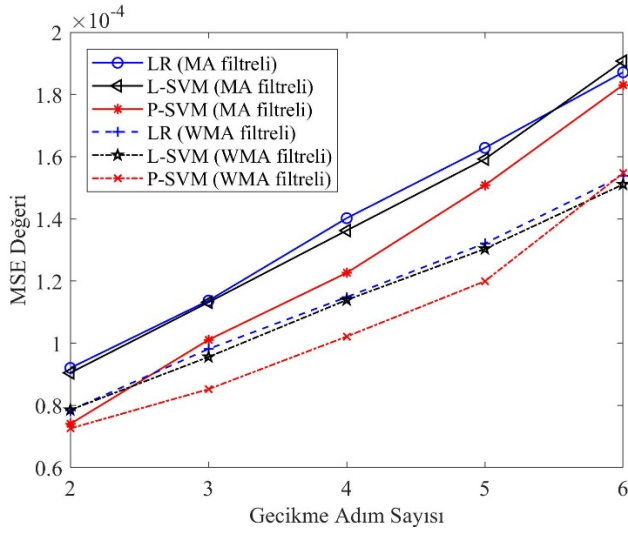
$$e_t = r_t - p_t \quad (4)$$

TABLO IV. İSTATİSTİKSEL HATA ÖLÇÜTLERİ

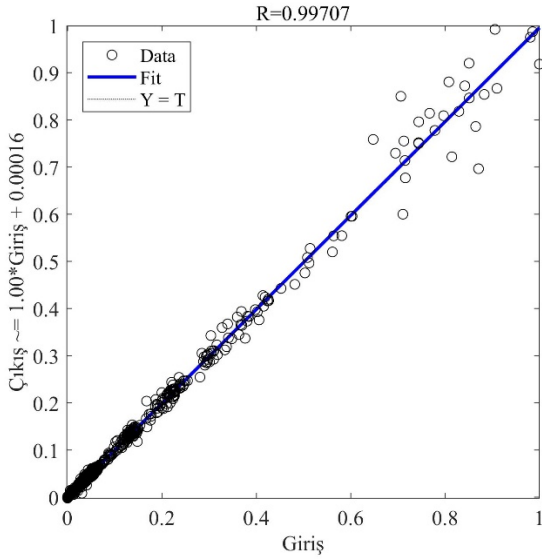
Açıklaması	İstatistiksel Hata Ölçüt Formülü
Ortalama Mutlak Hata (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n  e_t $
Ortalama Karesel Hata (MSE)	$MSE = \frac{1}{n} \sum_{i=1}^n e_t^2$
Ortalama Hata Kareleri Karekökü (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_t^2}$

### III. BITCOIN FİYAT TAHMİNİ BENZETİM SONUÇLARI

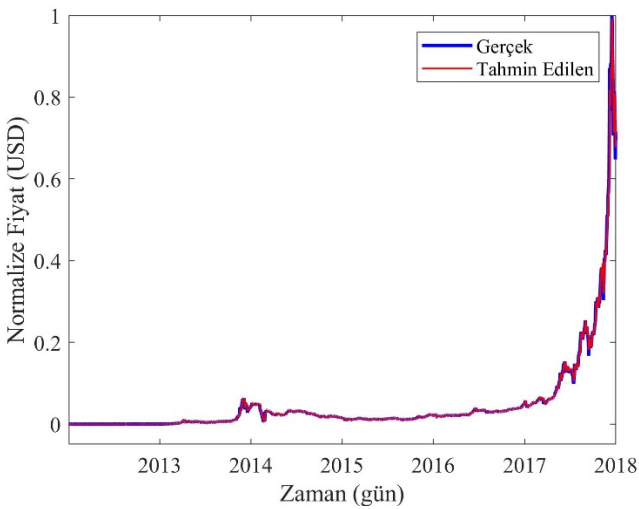
Benzetim çalışmasında makine öğrenmesi yöntemlerinden LR, L-SVM ve P-SVM ile bir sonraki ağıdındaki Bitcoin kapanış fiyatının tahmini yapılmaktadır. 2012-2018 yılları arasındaki 2192 adet günlük veriden oluşan veri seti kullanılmaktadır [17]. Veri seti günlük kapanış, en yüksek fiyat, en düşük fiyat ve işlem hacmi değişkenlerinden oluşmaktadır. Model eğitimi aşamasında 10-kat çapraz doğrulama yöntemi kullanılmaktadır. MA ve WMA filtreleri için gecikme adım sayısı 2 ile 6 arasında değiştirilmekte ve en düşük hatayı veren model elde edilmektedir. Ayrıca, model girişinde A/D osilatör değeri kullanılmaktadır. Farklı model parametreleri için MSE değişim grafiği Şekil 2'de verilmektedir. Gecikme adım sayısının azaltılması ile sistemin tahmin başarımının arttığı ve hata değerinin azaldığı görülmektedir. Gecikme sayısı 2 olan WMA filtreli durumda P-SVM yönteminin 0.00075 MSE değeri ile diğer modellere göre yüksek başarıma sahip olduğu görülmektedir. P-SVM tahmin modeli için hatanın hangi değerler etrafında gerçekleştiği dağılım grafiği ile araştırılmakta ve giriş ve çıkış arasındaki ilişki Pearson korelasyon katsayısı ile Şekil 3'te verilmektedir.



Şekil 2. LR, L-SVM ve P-SVM modellerinin MA ve WMA filtre durumları için gecikme adım sayısına göre MSE değişimleri.



Şekil 3. Gecikme adım sayısı 2 olan WMA filtre ve P-SVM modeli için regresyon eğrisi.



Şekil 4. Gecikme adım sayısı 2 olan WMA filtre ve P-SVM modeli için tahmin değerleri.

#### IV. SONUÇ

Bu çalışmada, LR, L-SVM ve P-SVM modelleri ile bir günlük kapanış fiyatı, en yüksek fiyat ve en düşük fiyat zaman serileri kullanılarak, farklı gecikme adımı içeren MA ve WMA filtreleri için Bitcoin kapanış fiyatı tahmini yapılmıştır. Çalışmada, ayrıca A/D osilatörü model girişi olarak kullanılmıştır. 2012-2018 yılları arasındaki 2192 adet veri noktası için modellerin eğitim ve test süreci 10-kat çapraz doğrulama yöntemi ile yapılmıştır. Elde edilen modellere ait tahmin değerleri birbirleri ile kıyaslanmış, en iyi sonucu veren tahmin modelinin gecikme sayısı 2 olan WMA filtreli P-SVM olduğu belirlenmiştir. Sonuç olarak kullanılan yöntemin yüksek başarımla gösterdiği ve ileriki çalışmalara ışık tutacağı düşünülmektedir.

#### KAYNAKLAR

- [1] Çarkacıoğlu, A., "Kripto-Para-Bitcoin", *Sermaye Piyasası Kurulu: Araştırma Raporu*, 2016, p 1-73.
- [2] Madan, I., Saluja, S. and Zhao, A., "Automated Bitcoin Trading via Machine Learning Algorithms", *Tech. Notes in Stanford Univ.*, 2015.
- [3] Jang, H. and Lee, J., "An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks based on Blockchain Information", *IEEE Access*, Vol. 98, 2017, p 1-11.
- [4] Kim, K. J., "Financial Time Series Forecasting using Support Vector Machines", *Neurocomputing*, 55(1-2):307-319, 2003.
- [5] Patel, J., Shah, S., Thakkar, P. and Kotecha, K., "Predicting Stock Market Index using Fusion of Machine Learning Techniques", *Expert Systems with Applications*, 42(4):2162-2172, 2015.
- [6] Chen, S. M. and Chung, N. Y., "Forecasting Enrollments using High-Order Fuzzy Time Series and Genetic Algorithms", *Int. Intell. Syst.*, Vol. 21, 2006, p 485-501.
- [7] Yu, H. K., "Weighted Fuzzy Time-Series Models for TAIEX Forecasting" *Phys. A Vol. 349*, 2005, p 609-624.
- [8] Wei, L. Y., "A Hybrid ANFIS Model based on Empirical Mode Decomposition for Stock Time Series Forecasting", *Applied Soft Computing*, Vol. 42, 2016, p 368-376.
- [9] Pichl, L. and Kaizoji, T., "Volatility Analysis of Bitcoin", *Quantitative Finance and Economics*, Vol. 1, 2017, p 474-485.
- [10] Indera, N. I., Yassin, I. M., Zabidi, A. and Rizman, Z. I., "Non-Linear Autoregressive with Exogeneous Input (NARX) Bitcoin Price Prediction Model using PSO-Optimized Parameters and Moving Average Technical Indicators", *J. of Fund. and App. Sc.*, 9(3S):791-808, 2017.
- [11] Almeida, J., Tata, S., Moser, A. and Smit, V., "Bitcoin Prediction using ANN", *Neural Networks*, 2015, p 1-12.
- [12] Greaves, A. and Au, B., "Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin", *Tech. Notes in Stanford Univ.*, 2015.
- [13] Kaur, T., Kumar, S. and Segal, R., "Application of Artificial Neural Network for Short Term Wind Speed Forecasting", *Biennial Int. Conf. on Pow. and Energy Syst.: Towards Sust. Energy*, 1-5, 2016.
- [14] Tan, P. N., Kumar, V. and Steinbach, M., "Introduction to Data Mining", Pearson, 2005.
- [15] Ahmed, S., Khalid, M. and Akram, U., "A Method for Short-Term Wind Speed Time Series Forecasting using Support Vector Machine Regression Model", *IEEE Int. Conf. on Clean Elect. Pow.*, 190-195, 2017.
- [16] Sun, S., Qiao, H., Wei, Y. and Wang, S., "A New Dynamic Integrated Approach for Wind Speed Forecasting", *Applied Energy*, Vol. 197, 2017, p 151-162.
- [17] URL: <https://query1.finance.yahoo.com/v7/finance/download/BTC-USD>