



DERİN ÖĞRENME VE EKONOMETRİK MODEL İLE BİTCOİN FİYAT TAHMİNİ: LSTM ve ARIMA

Yasin BÜYÜKKÖR¹

Öz

Dünyada en çok rağbet gören kripto para birimi olması nedeniyle Bitcoin (BTC), yatırımcılar ve araştırmacılar için son yıllarda dikkat çekici hale gelmiştir. Merkezi bir para birimi olmaması ve spekülasyonlara açık olması BTC fiyatında yüksek oynaklığa sebep olmaktadır. BTC fiyatının oynaklığının dikkate alınarak tahminlenmesi özellikle yatırımcılar için büyük önem taşımaktadır. Son yıllarda Makine Öğrenmesi (ML) yöntemlerinin gelişmesiyle birlikte birçok finansal alanda olduğu gibi kripto paraların fiyat tahminlemede sıklıkla ML yöntemlerine başvurulmaktadır. ML yöntemleri geleneksel ekonometrik yöntemlerin aksine veri setinde meydana gelen dalgalanmaları herhangi bir varsayıma ihtiyaç duymadan dikkate almakta ve çoğu zaman daha iyi sonuçlar vermektedirler. Bu çalışmada, 01.01.2018 ile 21.12.2023 tarihleri arasında BTC fiyatı geleneksel ekonometrik yöntem olan ARIMA ile ML yöntemi olan LSTM kullanılarak tahminlenmeye çalışılmıştır. Yöntemler karşılaştırılırken performans kriterleri olarak RMSE, MAE ve MAPE kriterleri kullanılmıştır. Çalışmanın sonuçlarına göre LSTM yöntemi en düşük RMSE ve MAPE değerlerine sahip olmuştur.

Anahtar Kelimeler: Bitcoin, Derin Öğrenme, Makine Öğrenmesi, Uzun Kısa Hafızalı Bellek, Finansal Zaman Serileri

JEL Sınıflandırması: C01, C22, C13, C53, C58

BITCOIN PRICE PREDICTION USING DEEP LEARNING AND ECONOMETRIC MODEL: LSTM AND ARIMA

Abstract

As the most popular cryptocurrency in the world, Bitcoin (BTC) has attracted the attention of investors and researchers in recent years. Its decentralized nature and the exposure to speculation lead to high volatility in the BTC price. Predicting the BTC price taking into account its volatility is of great importance, especially for investors. In recent years, with the development of Machine Learning (ML) methods, ML methods are frequently used in price predicting of cryptocurrencies as in many other financial areas. In contrast to traditional econometric methods, ML methods take into account the fluctuations in the data set without the making any assumptions and often give better results. In this study, the price of BTC between 01.01.2018 and 21.12.2023 is predicted using the traditional econometric method ARIMA and ML method LSTM. While comparing the methods, RMSE, MAE and MAPE criteria were used as performance criteria. According to the results of the study, LSTM method has the lowest RMSE and MAPE values.

Keywords: Bitcoin, Deep Learning, Machine Learning, Long Short-Term Memory, Financial Time Series

JEL Classification: C01, C22, C13, C53, C58

¹ Dr. Öğretim Üyesi, Karamanoğlu Mehmetbey Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, e-posta: yasinbuyukkor@kmu.edu.tr, ORCID iD: <https://orcid.org/my-orcid?orcid=0000-0002-1006-0539>

1. GİRİŞ

Nakamoto (2008) tarafından geliştirilen kripto para birimi Bitcoin (BTC), eşler arasında para (peer to peer) aktarma sistemi olarak tanıtılmıştır. BTC, merkezi olmayan yapısı (decentralized) ve izlenebilir olması gibi yenilikçi yapısı göz önünde bulundurulduğunda hem yatırımcıların hem medyanın hem de araştırmacıların kısa sürede dikkatini çekmiştir (Liu vd., 2021; Urguhart, 2016). Yenilikçi özelliklere sahip olması BTC'nin en başarılı ve en popüler kripto para birimi olmasını beraberinde getirmiştir (Katsiampa, 2017). Kripto para birimlerini listelenmesinde önemli bir yere sahip olan coinmarketcap² (06.03.2024) web sitesinin verilerine göre yaklaşık 2.5 trilyon dolar market büyüklüğüne sahip kripto para piyasasının %52.3'ünü BTC oluşturmaktadır. BTC ağında eşler arasında para aktarımı yapılırken kriptografi (şifreleme) modelleri kullanılarak ağ içerisindeki para aktarımı ağda bulunan kullanıcılar tarafından doğrulanır. Böylece herhangi bir merkezi kuruluşa bağlı olmadan hem göndericinin hem de alıcının kimliği saklı tutulabilir (Cocco vd., 2017).

Geleneksel para birimlerine göre artı ve eksi yönleri olması nedeniyle çoğu zaman BTC tartışmalara konu olmuştur. Günümüzde halen devam eden geleneksel otoritenin sarsılması ve kara paranın el değiştirmesi gibi tartışmalar olsa bile Microsoft, Amazon, Dell, AT&T ve Reddit gibi teknoloji ve e-ticaret şirketleri müşterilerini memnun etmek amacıyla BTC'yi bir ödeme aracı olarak kabul etmektedirler (Oprea vd., 2024).

BTC bir para birimi olmasına rağmen makroekonomik değişkenlerden diğer para birimlerine göre daha az etkilenmektedir (Ciaian vd., 2016). BTC fiyatını doğru olarak tahmin edebilmek yatırımcıların karar verme süreçlerine fayda sağlarken ülke yönetimleri için düzenleyici politikaların oluşturulmasına imkân sağlamaktadır (Matkovskyy ve Jalan, 2019). Ancak BTC fiyatının tahmin edilmesinde en büyük zorluğu fiyatındaki aşırı dalgalanmalar ve spekülasyona açık fiyat oluşumları oluşturmaktadır (McIntyre ve Harjes, 2016). Aşırı dalgalanması ve dinamik yapısı nedeniyle BTC fiyatının tahmin edilmesinde son yıllarda makine öğrenmesi yöntemleri öne çıkmaktadır. Geleneksel ekonometrik yöntemlerin gerektirdiği varsayımların kripto paralar üzerinde rahatlıkla uygulanamaması nedeniyle daha az veya hiçbir varsayımın olmadığı derin öğrenme ve makine öğrenmesi yöntemleri gelişen teknolojiyle birlikte araştırmacı ve yatırımcılar tarafından sıklıkla tercih edilmektedir.

Bu çalışmada, BTC Kapanış fiyatı zaman serisi verileri 01.01.2018 ile 21.12.2023 tarihleri arasındaki fiyatları tahmin edilmeye çalışılmıştır. Tahmin yöntemi olarak uzun dönemli bağımlılıkları dikkate alan Uzun Kısa Hafızalı Bellek (LSTM) makine öğrenmesi yöntemi ile geleneksel Otoresgresif Entegre Hareketli Ortalama (ARIMA) yöntemlerinin öngörümlere performansları karşılaştırılmıştır. Karşılaştırma kriterleri olarak Ortalama Karesel Hatanın Karekökü (RMSE), Ortalama Mutlak Hata (MAE) ve Ortalama Mutlak Yüzde Hata (MAPE) kullanılmıştır. ARIMA yöntemi veri sayısının az olduğu uygulamalarda güçlü bir model oluştururken LSTM yöntemi daha büyük veri setlerinde katmanlı yapılar kullandığından kripto para verilerinin doğrusal olmayan ve oynak yapılarını daha kolay yakalayabilmektedir.

Çalışma toplam beş bölümden oluşmaktadır. Giriş bölümünün ardından ikinci bölümde kripto para fiyatının tahminlenmesi ile ilgili geçmiş çalışmalar, üçüncü bölümde ARIMA ve LSTM modellerinin teorik yapıları, dördüncü bölümde verilerin analizi ve son bölümde ise sonuç ve değerlendirme kısmına yer verilmiştir.

2. LİTERATÜR TARAMASI

Borsanın doğuşundan itibaren yatırımcılar kar elde edebilmek için varlıkların fiyatını tahmin

² www.coinmarketcap.com

edebilecek doğru modelleri bulmaya çalışmaktadırlar (Fang vd., 2021). Özellikle 1970'li yıllardan itibaren araştırmacılar bir varlığın değerini tahmin etmek amacıyla ARIMA veya Mevsimsel ARIMA (SARIMA) gibi geleneksel istatistiksel ve ekonometrik modelleri kullanmaya başlamışlardır (Dooley ve Lenihan, 2005; Ediger ve Akar, 2007). Ancak 1990'lı yılların ortalarında internetin ve bilgisayar teknolojilerinin gelişmesiyle birlikte geleneksel yöntemlerin varsayımlarından etkilenmeyen makine öğrenmesi yöntemleri zaman serisi verilerinin tahminlenmesinde sıklıkla kullanılmaya başlanmıştır (Demirci ve Karaatlı, 2023; Kaya vd., 2020; Kercheval ve Zhang, 2015).

BTC fiyatının tahminlenmesi söz konusu olduğunda geçmiş çalışmalar çoğunlukla BTC fiyatını belirleyen etkenler üzerinde yoğunlaşmışlardır. Bu nedenle BTC fiyat tahminlenmesinde makine öğrenmesi yöntemlerini kullanan çalışmalar kısıtlıdır. BTC fiyat tahminlemesi yapan çalışmalar; Saad ve Mohaisen (2017), BTC fiyatına yüksek etki yapan ağ hareketlerini incelemişlerdir. Belirledikleri ağ yapılarını bağımsız değişkenler olarak kullanmışlar ve BTC fiyatını Doğrusal Regresyon (LRA), Rassal Orman (RF) ve Gradient Descent (GD) yöntemleriyle analiz etmişlerdir. Analiz sonuçlarına göre LRA yöntemi en iyi RMSE değerine sahip olmuştur. Jang ve Lee (2017), BTC fiyatını çeşitli Blok Zincir değişkenleri, S&P 500 endeksi, Eurostoxx endeksi, DOW 30 endeksi, NASDAQ endeksi, Ham Petrol, Altın fiyatı, Nikkei 225 endeksi ve FTSE 100 endeksi gibi makroekonomik değişkenler ve İngiliz Poundu (GBP), Japon Yeni (JPY), İsviçre Frangı (CHF), Çin Yuan'ı (CNY) ve Euro (EUR) gibi para birimlerini kullanarak tahminlemişlerdir. Fiyat tahmini için LRA, Destek Vektör Makinası (SVM) ve Geri Beslemeli Sinir Ağları (BPNN) yöntemlerini kullanmışlardır. En iyi düşük RMSE ve MAPE değerlerini BPNN yöntemi ile elde etmişlerdir. Aggarwal vd. (2019) çalışmalarında BTC fiyatını etkileyen sosyal ve ekonomik faktörleri incelemişlerdir. Bu faktörlerin BTC fiyatına etkilerini Evrişimli Sinir Ağları (CNN), LSTM ve GRU derin öğrenme yöntemleriyle tahmin etmişlerdir. Analiz sonuçlarına göre LSTM modelinin en düşük RMSE değerine sahip olduğunu göstermişlerdir. Awoke vd. (2020) BTC fiyatının tahminlenmesinde LSTM ve Gated Recurrent Unit (GRU) gibi derin öğrenme yöntemlerini kullanmışlardır. Araştırmanın bulgularına göre LSTM, GRU'dan daha uzun sürede eğitim sürecini tamamlamaktadır. Ayrıca RMSE ve MAPE kriterlerinde GRU daha düşük seviyelere ulaşabilmektedir. Shin vd. (2021) BTC fiyatını tahmin etmek için çeşitli zaman aralıklarının modellenmesini LSTM yöntemi ile tahminlemişlerdir. Saatlik ve dakikalık verilerin değişkenliklerinin günlük verilere göre daha fazla olduğunu göstermişlerdir. Akyıldırım vd. (2021) BTC vadeli işlem fiyatlarının orta fiyat hareketlerinin tahmin edilmesinde çeşitli ML yöntemleri kullanmışlardır. 5 ve 60 dakikalık yüksek frekanslı veriler kullanılarak ML yöntemlerinin göre performansları karşılaştırılmıştır. ML yöntemlerinin ARIMA ve Rassal Yürüyüş (RW) yöntemlerinden daha iyi performansla sahip olduğu göstermişlerdir. Latif vd. (2023) BTC fiyatının tahmin edilmesinde ARIMA ve LSTM yöntemlerini karşılaştırmışlardır. Kullanmış oldukları iki farklı eğitim ve test veri setlerinin ilkinde ARIMA üstün gelirken ikinci veri setinde LSTM daha üstün gelmiştir. Chen (2023), BTC'nin bir sonraki gün fiyatını 2015-2018 yılları arasındaki günlük verileri kullanarak RF ve LSTM yöntemleriyle tahminlemiştir. RMSE ve MAPE kriterlerine göre en iyi yöntemin RF olduğu göstermiştir. Cheng vd. (2024) çalışmalarında LSTM, SARIMA ve Facebook Prophet (FBP) modelleri kullanılarak BTC fiyatı ve Garman-Klass (GK) oynaklığının tahminini gerçekleştirmek için finansal zaman serileri ve makine öğrenimi için ampirik bir analiz yapmışlardır. Çalışmanın bulgularına göre LSTM, MSE ve MAE kriterlerine göre hem SARIMA hem de FBP yöntemine göre kayda değer bir şekilde iyileşme sağlamıştır.

BTC fiyat ve yön tahmini ile ilgili geçmiş çalışmalardan bazıları konu, içerik, değişken ve analiz aşamasında kullanılan kripto para birimlerine göre Tablo 1'de sınıflandırılmıştır.

Tablo 1: BTC ve diğer kripto para birimleri kullanılarak yapılan geçmiş çalışmalar

Yazar Adı	Konu	Yöntem	Değişkenler	Kripto Para	Sonuç
Rathore vd., (2022)	Bitcoin'in Gelecek Fiyat Tahmini	FBP, ARIMA, LSTM ve Naive Model	BTC	BTC	Önerilen FBP Modeli Diğer yöntemlerden daha üstün performansa sahip olmuştur.
Chen vd., (2020)	Günlük ve yüksek frekanslı (5 dk.) BTC fiyat tahmini	LR, LDA, RF, XGB, QDA, SVM ve LSTM	Kripto Para Piyasası Göstergeleri, Google Trend Arama Verileri Hacim Endeksi, Baidu Media	BTC	Günlük Fiyat Tahmin için LR yöntemi en iyiyken yüksek frekanslı veriler için LSTM en iyi performansa sahiptir.
Peng vd., (2018)	Üç farklı kripto paranın günlük ve saatlik verileri ile üç farklı para biriminin	GARCH Ailesi ve SVR- GARCH	Günlük ve Saatlik Kripto paraların Dolar cinsinden değeri ve Euro, BP ve JY	BTC, ETH,DASH, Euro, BP ve JY	Yöntemler Diebol-Mariano ve Hansen's Model Confidence Set kriterleri ile karşılaştırılmıştır. Sonuç olarak SVR-GARCH modeli en iyi performansa sahiptir.
Liu vd. (2021)	BTC fiyatının SDAE, BPNN, PCA-SVR ve SVR ile tahmini	SDAE, BPNN ve SVR	20 farklı kripto para piyasası göstergesi ile 20 adet Makroekonomik Gösterge	BTC	Önerilen SDAE yöntemi diğer ML yöntemlerine göre bütün performans kriterlerinde daha iyi sonuçlara sahiptir.
Mallaqui ve Fernandes (2019)	BTC fiyatı ve yönünün özellik seçim yöntemi ile belirlenen değişkenler	ANN, SVM, RNN ve k-Means Clustering	86 gösterge arasından Özellik Seçimi ile 20 en iyi gösterge seçilmiştir.	BTC	Özellik seçimi ile belirlenen değişkenler ve ML yöntemleri kullanılarak yapılan tahminler geçmiş çalışmaların doğruluklarını %10 artırmıştır.
Jaquart vd. (2021)	BTC fiyat yönünün 1-60 dakikalık uzunluklar için belirlenmesi	GRU, LSTM, FNN, GBC ve RF	TG, Makroekonomik Göstergeler, Blok Zinciri Göstergeleri ve İlgili Temelli Göstergeler	BTC	GBC ve RNN tabanlı yöntemler BTC hareket yönünü yakalamada diğer yöntemlerden daha yeteneklidirler.
Fleischer vd. (2022)	Kripto paraların fiyat tahmini	LSTM ve ARIMA	EOS, BTC, ETH ve DOGE	EOS, BTC, ETH ve DOGE	LSTM, ARIMA yöntemine göre daha iyi sonuçları daha kısa sürede elde etmektedir.
Ji vd. (2019)	BTC yön ve fiyat tahmini	DNN, RNN, LSTM, CNN, DRN ve bu yöntemlerin hibrit hale	BTC	BTC	Fiyat tahmini için LSTM ve yön tahmini için DNN en iyi sonuçları vermektedir.
Lamothe-Fernández vd. (2020)	BTC Fiyat Tahmini	DRCNN, DNDT ve DSVR	Kripto Para Piyasası Göstergeleri, Makroekonomik ve Finansal olmak üzere Toplam 29 Değişken	BTC	DRCNN modeli BTC fiyatının gelecek tahmininde diğer yöntemlere göre daha yüksek doğruluğa sahip olmuştur.
Maleki vd. (2023)	BTC fiyatını ETH, ZEC ve Litecoin kullanarak tahmini	ARIMA, Lineer Regresyon, SVR, GBR ve RFR	ETH, ZEC ve Litecoin	BTC	BTC fiyat tahmininde ZEC en önemli değişken olarak belirlenmiştir.

LDA: Doğrusal Diskriminant Analizi, **XGB:** eXtreme Gradient Boosting, **QDA:** Kuadratik Diskriminant Analizi, **SVR:** Destek Vektör Regresyonu, **GARCH:** Genelleştirilmiş Otoregresif Koşullu Varyans Modeli, **ETH:** Ethereum, **SDAE:** Stacked Denoising Autoencoding, **PCA:** Temel Bileşenler Analizi, **ANN:** Artificial Neural Network, **RNN:** Recurrent Neural Network, **TG:** Teknik Göstergeler, **FNN:** FeedForward Neural Network, **GBC:** Gradient Based Classifiers, **EOS:** Electro-Optical System, **DOGE:** Dogecoin, **DNN:** Deep Neural Network, **CNN:** Convolutional Neural Network, **DRN:** Deep Residual Network, **DRCNN:** Deep Recurrent CNN, **DNDT:** Deep Neural Decision Tree, **DSVR:** Deep SVR, **GBR:** Gradient Boosting Regression, **RFR:** RF Regression, **SANN:** Stacked ANN, **ZEC:** ZCASH Coin

3. VERİ VE YÖNTEMLER

3.1. Otoregresif Entegre Hareketli Ortalama (ARIMA)

Box ve Jenkins (1970) tarafından geliştirilen Otoregresif Entegre Hareketli Ortalama (ARIMA), zaman serisi verilerinde sıklıkla kullanılan geleneksel bir ekonometrik yöntemdir. Otoregresif (AR) ve Hareketli Ortalama (MA) süreçlerinin bir kombinasyonu olarak bilinir. Buna ek olarak Integrated (*I*) ise seriyi durağanlaştırmak için kullanılan fark alma operatörünü göstermektedir. Bir $ARIMA(p,d,q)$ modelinde p ; AR modelinin derecesi, d ; zaman serisinde alınan farkın derecesini ve q ; MA modelinin derecesini göstermektedir. $AR(p)$ modeli:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (1)$$

ile gösterilir. Eşitlik 1'de, Y_t ; zaman serisi verilerini, Y_{t-i} ; zaman serisinin gecikmelerini, ϕ_i gecikmelerin katsayılarını ve ε_t hata terimini göstermektedir. $MA(q)$ süreci ise:

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2)$$

olarak yazılabilir. Eşitlik 2'de, θ_i MA sürecinin katsayılarını göstermektedir. $AR(p)$ ve $MA(q)$ süreçleri bir araya getirilirse $ARIMA(p,d,q)$:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3)$$

yazılır. $ARIMA(p,d,q)$ modeli kurulurken sırasıyla:

- i) Öncelikle seri durağan değilse durağan hale getirmek için fark alınır. Daha sonra Otokorelasyon (ACF) ve Kısmi Otokorelasyon (PACF) grafikleri çizilerek AR modelinin derecesi p ve MA modelinin derecesi q belirlenir.
- ii) Uygun derecelere sahip $ARIMA(p,d,q)$ modeli kurularak model parametrelerinin tahminleri yapılır ve katsayıların istatistiksel olarak anlamlı olup olmadıkları kontrol edilir. Eğer kurulan $ARIMA(p,d,q)$ modelinde anlamsız katsayılar varsa ilk aşamaya geri dönlür.
- iii) $ARIMA(p,d,q)$ modelinin artıkları incelenir. Artıkların beyaz gürültü olması beklenir.
- iv) Son aşamada elde edilen model kullanılarak geleceğe yönelik öngörüleme (forecasting) yapılır.

3.2. Uzun Kısa Hafızalı Bellek (LSTM)

Uzun Kısa Hafızalı Bellek (LSTM), Tekrarlayan Sinir Ağlarında (RNN) meydana gelen gradient kaybolması probleminin çözülmesi amacıyla Hochreiter ve Schmidhuber (1997) tarafından geliştirilmiştir. RNN, ardışık veriler arasındaki uzun dönemli (long term) bağımlılıkları yakalamada sorun yaşarken, LSTM bu problemin üstesinden hafıza hücreleri (memory cell) ve kapı yapıları (gate structure) kullanarak gelmektedir (Patel vd., 2020; Patel vd., 2020; Hamayel ve Owda, 2021). LSTM'de kapı yapıları ile elde edilen bir bilgi hücreler arasında sürekli olarak aktarılır. Şekil 1'den görülebileceği gibi tek bir sinir ağı katmanı yerine LSTM birbirleriyle etkileşim halinde olan dört sinir ağı katmanından oluşur.

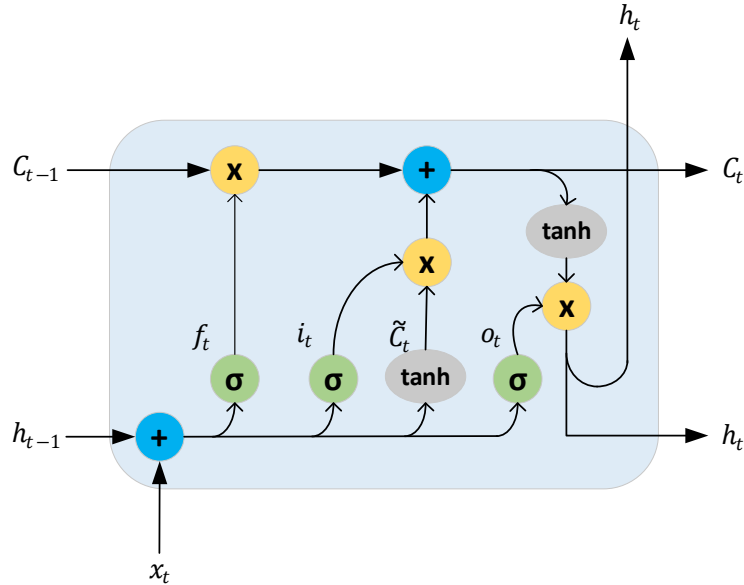
Genel olarak bir LSTM birimi Giriş Kapısı (input gate), Unutma Kapısı (forget gate) ve Çıktı Kapısı (output gate) aracılığıyla sürekli olarak güncellenen bilgileri depolayan bir bellek hücresinden oluşur. Böylece geçmiş bilgilerin ve gelecekteki bilgilerin temsil edilmesi kolaylaşırken ardışık verilerde uzun dönemli bağımlılıkların ortaya çıkarılmasında RNN'ye göre daha iyi bir performans gösterir (Chen vd., 2020).

LSTM'nin matematiksel formülasyonu:

$$X = \begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix}$$

$$\begin{aligned} f_t &= \delta(W_f \cdot X + b_f) \\ i_t &= \delta(W_i \cdot X + b_i) \\ o_t &= \delta(W_o \cdot X + b_o) \\ \tilde{C}_t &= \tanh(W_c \cdot [h_t, x_t] + b_c) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ h_t &= o_t \odot \tanh(C_t) \end{aligned} \quad (4)$$

Eşitlik 4'te x_t , t zamanındaki girdiyi; h_t , t zamanındaki gizil yapıyı; W_f , W_i , W_o ve W_c LSTM'nin ağırlık matrislerini; b_f , b_i , b_o ve b_c sapma vektörünü; δ aktivasyon fonksiyonunu ve \odot , nokta çarpım operatörüdür. Şekil 1'de LSTM modelinin genel yapısı görselleştirilmiştir.



Şekil 1. LSTM Modeli Yapısı

LSTM çok geniş bir alanda farklı görevleri yerine getirmek amacıyla kullanılmaktadır. Kullanım alanları arasında doğal dil işleme, el yazısı tanıma, müzik geliştirme ve bu makalede olduğu gibi finansal zaman serisi verilerinin analizi gösterilebilir (Jaquart vd., 2021:52).

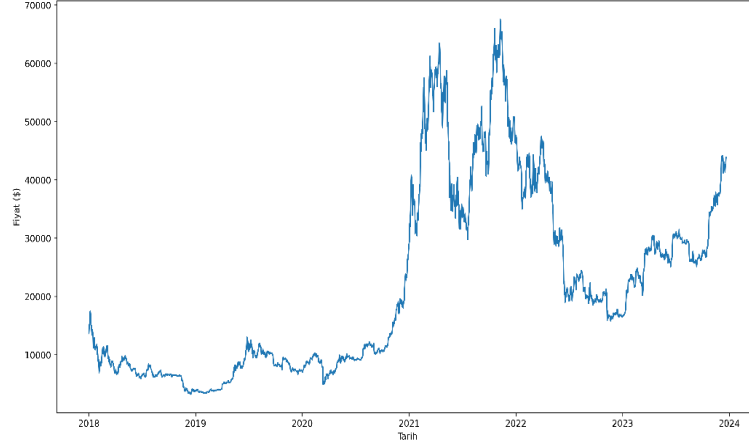
4. UYGULAMA

4.1. Veri Seti

BTC fiyatını öngörümlemek ve ARIMA ile LSTM modellerinin öngörümleme performanslarını karşılaştırmak amacıyla 01.01.2018 ile 21.12.2023 tarihleri arasında BTC (USD) veri seti Yahoo Finance'den (<http://finance.yahoo.com>³) elde edilmiştir. Elde edilen veriler BTC fiyatının günlük olarak Açılış, Kapanış, En Yüksek, En Düşük, Hacim ve Tarih verilerini içermektedir. Bahsedilen tarihler arasında BTC Kapanış fiyatının zaman serisi grafiği Şekil 2'de verilmiştir. Ayrıca zaman serisi verilerinin tanımlayıcı istatistikleri Tablo 2'de yer almaktadır. Geleneksel borsa veya hisse senedi

³ <https://www.finance.yahoo.com/>

verilerinden farklı olarak BTC fiyatı haftanın her günü işlem gördüğünden piyasanın kapalı olduğu günlerde de fiyatlanmaya devam etmiştir. Böylece toplam veri sayısı 2181 olarak belirlenmiştir.



Şekil 2: BTC Kapanış fiyatı

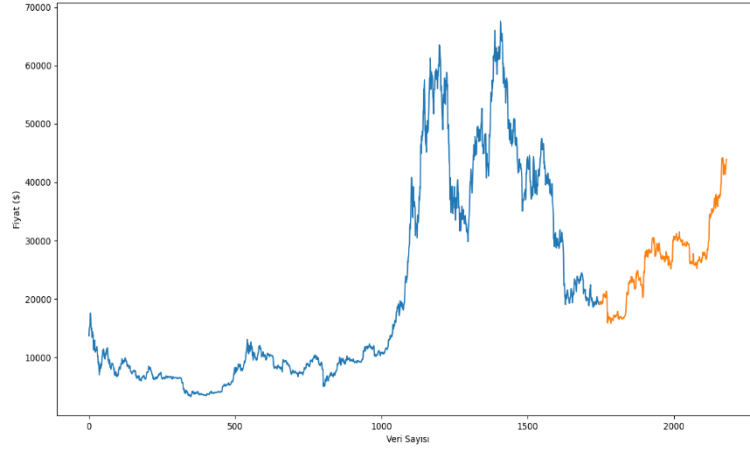
Tablo 2: BTC veri setine ait özet istatistikler

Gözlem Sayısı	Ortalama	Standart Sapma	Minimum	Maksimum	25%	50%	75%
2181	21660.99	15981.62	3236.76	67566.82	8309.28	16919.8	31726.39

BTC veri setine ait 2181 günlük Kapanış fiyatı, makine öğrenmesi yöntemlerine uygun olarak eğitim ve test veri setleri olarak 2 bölüme ayrılmıştır. Genel olarak makine öğrenmesi yöntemlerinde eğitim ve test veri setleri rassal olarak bölümlenirken üzerinde çalışılan veri seti bir zaman serisi olduğundan ve geçmiş veriler ile ilişkili olduğundan rassal olarak bölümlenmemiştir. Zaman serisi uygulamalarına uygun olarak veri setinin ilk %80'lik kısmı 01.01.2018 ile 10.10.2022 tarihleri arasında eğitim verisi olarak ayrılmış ve %20'lik kısmı ise 11.10.2022 ile 21.12.2023 tarihleri arasında test veri seti olarak belirlenmiştir. Böylece eğitim veri seti 1744 adet gözlemden oluşurken test veri seti 437 gözlemden oluşmaktadır. Eğitim ve test veri setlerine ait tarih ve gözlem sayıları Tablo 2'de ve Şekil 3'te verilmiştir.

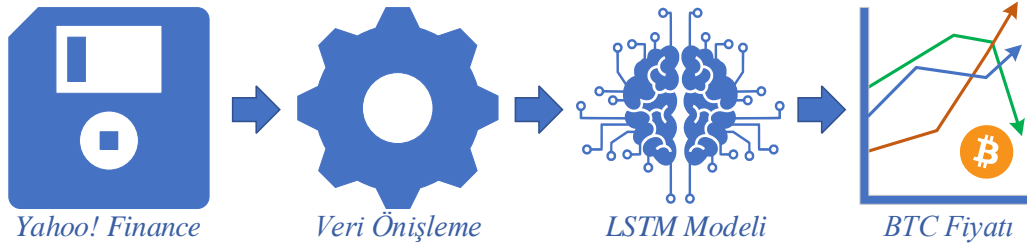
Tablo 2: Eğitim ve Test veri setine ait tarih ve gözlem sayıları

Veri seti	Eğitim Veri Seti	Test Veri Seti
BTC	01.01.2018 ile 10.10.2022 1744 Gözlem (%80)	11.10.2022 – 21.12.2023 437 Gözlem (%20)



Şekil 3: BTC eğitim ve Test veri seti

Şekil 4’te BTC fiyatı zaman serisi verilerinin tahminlenmesi için kullanılan bir süreç şeması verilmiştir. Bu şemaya göre öncelikle Yahoo Finance’den elde edilen veriler bir veri ön işleme sürecine tabii tutularak zaman serisi verisine uygun olarak Tarih ve Kapanış fiyatı ayırt edilmiştir. Bu aşamadan sonra uygun olarak eğitim ve test veri setleri ayrılarak LSTM modeli kurulmuş ve modelin tahmin grafiği elde edilmiştir.



Şekil 4: BTC fiyat tahmini için infografik

BTC fiyatının tahminlenmesi sürecinde Phyton 3.9.16 programı ve E-views paket programından yararlanılmıştır. Phyton’da ARIMA için yardımcı kütüphaneler *auto.arima*, LSTM için *pandas*, *keras*, *tensorflow*, *Sequential* ve *LSTM* kütüphaneleri kullanılmıştır.

4.2. Veri Ön İşleme

Çalışmada kullanılan yöntemlerin tahmin performanslarını en iyilemek amacıyla yöntemlere ait optimal parametreler (hyperparameter tuning) Grid Arama (Grid Search) yöntemi kullanılarak araştırılmıştır. Kullanılan veri seti bir zaman serisi verisi olduğu için tahminleme yapılırken Kayan Pencereler (Sliding Window) yöntemi kullanılmıştır. Bu yöntemde veri setinde önceki n adet gözlem kullanılarak $(n+1)$. gözlem tahminlenir. Bir sonraki aşamada $2, \dots, (n+1)$ gözlem kullanılarak $(n+2)$. gözlem tahminlenir. Böylece son veri tahmin edilene kadar gözlemler kaydırılır (Hu vd., 1999). Bu yöntemin avantajı veri setinin geçmiş değerleri ile arasındaki ilişkinin korunmasını sağlamaktır.

4.2.1. LSTM veri ön işleme: Çalışmada kullanılan veri setinin dalgalı yapısı nedeniyle LSTM modeli uygulanmadan önce veriler ölçeklenmiştir. Makine öğrenmesi yöntemlerinde genellikle kullanılan Min- Max Ölçeklendirme yönteminde veriler 0-1 aralığına ölçeklendirilmektedir. Ancak ölçeklendirilme yapılsa bile sapan gözlemler ağırlıklarını ölçeklendirilmiş veri seti içerisinde korumaktadır. Verilere uygulanacak ölçeklendirme yöntemi seçilirken veride var olan sapan gözlemler

nedeniyle daha uygun olacağı düşünülen Robust Ölçeklendirme (Robust Scaler) kullanılmıştır (de Amorim vd., 2023; Mudassir vd., 2020). Robust Ölçeklendirme yöntemi Eşitlik 5'te verilmiştir.

$$x_{scaled} = \frac{x_i - Q_2(x_i)}{Q_3(x_i) - Q_1(x_i)} \quad (5)$$

Eşitlik 5'te Q_2 verilerin medyanını, Q_1 ve Q_3 sırasıyla verilerin birinci ve üçüncü kartillerini göstermektedir.

BTC fiyatını tahmin etmek için kullanılan LSTM yöntemine ait hiper parametreler ve bu parametrelere ait belirlenen aralıklar Tablo 3'te yer almaktadır. Parametre aralıkları belirlenirken geçmiş çalışmalar dikkate alınmış ve Grid Arama yöntemi kullanılarak bütün parametreler için sonuçlar elde edilmiştir. Literatürde yer alan çalışmalar dikkate alındığında zaman serisi verilerinde doğrusal olmayan ilişkileri yakalamak için iki katmanlı LSTM modelinin kurulmasının yeterli olduğu görülmüştür (McNally vd., 2018). Bu nedenle bu çalışmada iki katmanlı LSTM modeli kurulmuştur.

Tablo 3: LSTM modeli parametreleri

Katman	Optimal Parametreler	Parametre Aralığı	Optimal Parametre
Katman 1	LSTM neuron 1	32, 64, 96, 128, 256, 512	64
Katman 2	LSTM neuron 2	32, 64, 96, 128, 256, 512	64
	Dropout rate	0, 0.1, 0.2, 0.3, 0.4, 0.5	0
Katman 1	Dense neuron 1	4, 8, 16, 32, 64, 128, 256	4
Katman 2	Dense neuron 2 (Çıkış Katmanı)	1	1
	Batch size	16, 32, 64, 128, 256	32
	Epochs	10, 30, 50, 100, 200, 500	200
	Learning rate	0.0001, 0.001, 0.01, 0.1	0.001
	Optimizer	Adam, SGD, Adamax, Nadam, RMSprop	Adam

4.2.2. ARIMA veri ön işleme: Geleneksel zaman serisi modellerinden olan ARIMA yöntemi kullanılarak BTC fiyatının tahminlenebilmesi için veri setinin durağan olması gerekmektedir. Bir zaman serisinin durağan olması, veri setindeki gözlemlerin zamana göre sabit kalması anlamına gelmektedir. ARIMA yöntemi uygulanmadan önce veri setinin durağanlığı Dickey- Fuller (ADF) testi ile incelenmiştir. Veri setine uygulanan ADF test sonucuna göre test istatistiği -1.346 ve p değeri 0.607 elde edilmiştir. Elde edilen sonuca göre veri setinin durağan olmadığını söyleyen sıfır hipotezi (H_0) reddedilememiştir ve veri setinin durağan olmadığına karar verilmiştir. ARIMA modelinde otoregresif parametre (AR) p ve hareketli ortalama parametresi (MA) q belirlenmeden önce veri setinin durağan hale getirilmesi önemlidir. Veri setini durağan hale getirmek için veri seti setinde gerekli olduğu kadar (durağan hale gelene kadar) fark (difference) alınır. BTC veri setinin birinci farkı alındıktan sonra veri setine tekrar ADF testi uygulanmış ve test istatistiği -8.144, p değeri 0.000 elde edilmiştir. Böylece veri seti durağan hale getirilmiştir.

BTC verisinin durağan hale getirilmesinden sonra AR(p) ve MA(q) parametrelerini belirlemek amacıyla önceden seçilen aralıklar kullanılarak en düşük Akaike Information Criteria (AIC) değerine sahip parametreler belirlenmiştir. ARIMA yöntemi için optimal parametre aralıkları ve en iyi parametre değerleri Tablo 4'te verilmiştir.

Tablo 4: ARIMA model parametreleri

Model	Optimal parametreler	Parametre Aralığı	Optimal Parametre
ARIMA	p	0, ...,10	2 ve 2
	d	Önceden Belirlenmiş	1
	q	0, ...,10	2 ve 3

Tablo 4 incelendiğinde en iyi modelin birden fazla olduğu görülmektedir. Bu nedenle uygulamada ARIMA(2,1,2) ve ARIMA(2,1,3) modelleri kullanılmıştır.

4.3. Karşılaştırma Kriterleri

Regresyon ve zaman serisi tahminlemede, elde edilen modellerin performanslarını karşılaştırmak amacıyla genellikle Ortalama Karesel Hatanın Karekökü (RMSE), Ortalama Mutlak Hata (MAE) ve Ortalama Mutlak Yüzde Hata (MAPE) gibi hata metrikleri kullanılmaktadır (Çılgın ve Özdemir, (2023)). Tahmin sonuçları dikkate alındığında her üç hata metriğinin küçük olanı tercih edilmektedir. Ancak hata metriklerinden herhangi birinin küçük olması diğer hata metriklerinin de küçük olması anlamına gelmeyebilir. Bu nedenle modeller arasında tercih yapılırken her üç hata metriğinin birlikte değerlendirilmesi daha doğru bir sonuç elde etmek için önemlidir. Söz konusu hata metrikleri Eşitlik 5'te verilmiştir.

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ MAE &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ MAPE &= \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \end{aligned} \quad (5)$$

Eşitlik 5'te n gözlem sayısını, y_i gerçek değeri ve \hat{y}_i tahmin değerini göstermektedir.

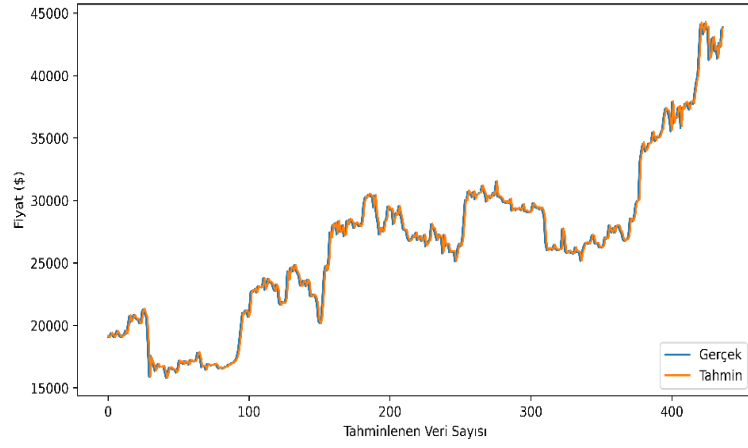
4.4. Uygulama Sonuçları

BTC fiyatının tahminlenmesi için kullanılan LSTM ve ARIMA modellerinin tahmin performansları Tablo 5'te verilmiştir. Tablo 5 incelendiğinde geleneksel ekonometrik model ARIMA ile makine öğrenmesi modeli olan LSTM'nin her üç performans kriterinde birbirlerine yakın sonuçlara sahip olduğu görülmektedir. Ancak LSTM modelinin performansının az da olsa ARIMA modellerine göre daha iyi performansa sahip olduğu görülmektedir. RMSE ve MAPE kriterleri dikkate alındığında LSTM modeli daha iyiyken MAE kriteri dikkate alındığında ARIMA(2,1,2) modelinin daha iyi olduğu söylenebilir.

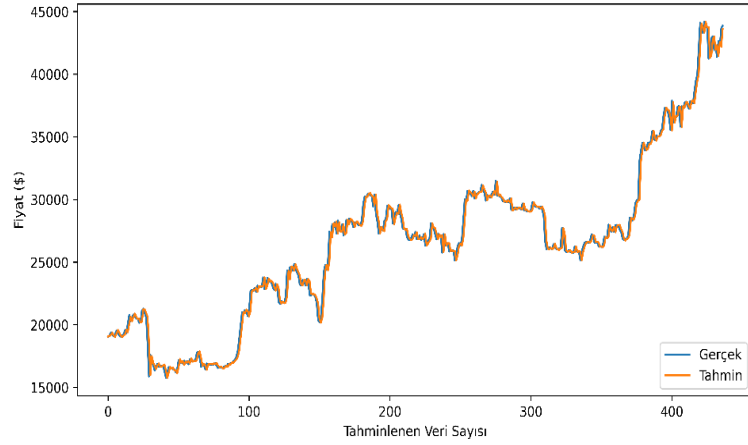
Tablo 5: LSTM ve ARIMA modeli uygulama sonuçları

Model	RMSE	MAE	MAPE
LSTM	628.27	412.55	0.015
ARIMA (2,1,2)	632.59	410.34	0.016
ARIMA (2,1,3)	635.71	412.36	0.017

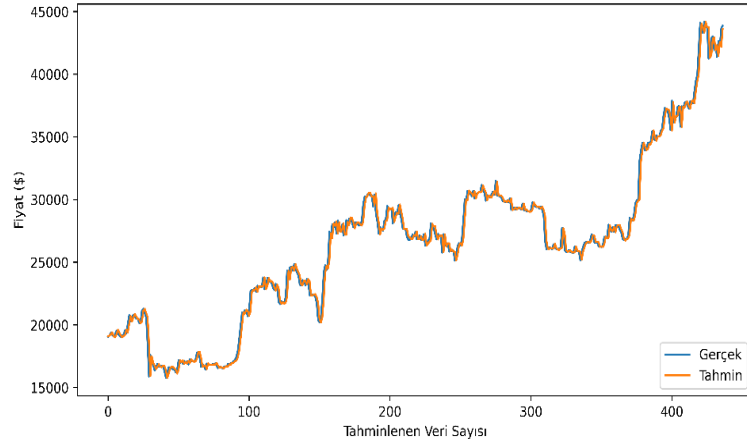
Uygulamada kullanılan modellerin 11.10.2022 ile 21.12.2023 tarihleri arasında kapsayan test veri seti üzerindeki tahmin performanslarının görselleri sırasıyla Şekil 5, Şekil 6 ve Şekil 7’de verilmiştir. Görseller dikkate alındığında en iyi tahmin performansına LSTM modelinin ulaştığı görülebilir.



Şekil 5: LSTM modeline ait gerçek ve tahmin BTC fiyatı



Şekil 6: ARIMA(2,1,2) modeline ait gerçek ve tahmin BTC fiyatı



Şekil 7: ARIMA(2,1,3) modeline ait gerçek ve tahmin BTC fiyatı

SONUÇ

Bu çalışmada son yıllarda adından sıkça söz ettiren ve en yüksek piyasa hakimiyetine sahip olan kripto para birimi BTC fiyatı 01.01.2018 ile 21.12.2023 tarihleri arasında tahmin edilmiştir. Çalışmada uzun dönemli bağımlılıkları dikkate alan makine öğrenmesi yöntemi LSTM ve geleneksel ekonometrik yöntem olan ARIMA karşılaştırılmıştır. Her iki yöntemin tahmin performanslarını artırmak amacıyla önceki çalışmalar dikkate alınarak yöntemlerin parametreleri optimize edilmiştir. Söz konusu modeller RMSE, MAE ve MAPE kriterleri göz önünde bulundurularak karşılaştırılmıştır. Genel olarak LSTM, ARIMA modellerine göre daha iyi RMSE ve MAPE değerleri elde ederken MAE kriteri dikkate alındığında ARIMA(2,1,2) modeli daha iyi sonuçlar elde etmiştir. Latif vd. (2023)'e göre eğer veri setini trendi yukarı yönlü ise ARIMA yöntemi en az makine öğrenmesi kadar iyi sonuçlar üretebilir. Bu çalışmada BTC veri setinin yukarı yönlü trende sahip olması ve veri sayısının fazla olması nedeniyle ARIMA ve LSTM model sonuçları birbirlerine yakındır.

Kripto para fiyatları gibi zaman serisi verilerinin doğrusal olmayan yapısı ve oynaklıkları nedeniyle tahmin edilmesinde makine öğrenmesi yöntemlerinin kullanılmasına literatürde sıklıkla rastlanılmaktadır. Her ne kadar makine öğrenmesi yöntemlerinin daha iyi tahmin sonuçlarına sahip olacağı beklense de veri sayısı fazla olduğunda LSTM gibi eğitim süreci uzun ve fazla sayıda parametreye sahip yöntemlerin pratik olarak kullanılması her zaman mümkün olmamaktadır. Ancak LSTM modeli birçok makine öğrenmesi yönteminin aksine uzun dönemli hafızayı dikkate aldığından veride var olan doğrusal olmayan yapıyı başarılı bir şekilde yakalayabilmektedir. Bu nedenle birçok araştırmacı ve yatırımcı için dikkate alınması gereken önemli modellerden biridir. Buna ek olarak geleneksel ekonometrik yöntemlerin göz ardı edilmemesi ve mutlaka sonuçlarının elde edilmesi karşılaştırma açısından faydalı olabilmektedir.

Etik Beyan

“Derin Öğrenme ve Ekonometrik Model ile Bitcoin Fiyat Tahmini: LSTM ve ARIMA” başlıklı çalışmanın yazılması ve yayınlanması süreçlerinde Araştırma ve Yayın Etiği kurallarına riayet edilmiş ve çalışma için elde edilen verilerde herhangi bir tahrifat yapılmamıştır. Çalışma için etik kurul izni gerekmemektedir.

Büyükkör, Y. (2024). Derin Öğrenme ve Ekonometrik Model ile Bitcoin Fiyat Tahmini: LSTM ve ARIMA. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 26(47), 978-993.

Katkı Oranı Beyanı

Çalışmadaki yazar çalışmanın yazılmasından taslağın oluşturulmasına kadar tüm süreçlere katkı yapmış ve nihai halini okuyarak onaylamıştır.

Çatışma Beyanı

Yapılan bu çalışma gerek bireysel gerekse kurumsal/örgütsel herhangi bir çıkar çatışmasına yol açmamıştır.

KAYNAKÇA

- Aggarwal, A., Gupta, I., Garg, N. ve Goel, A. (2019, August). Deep Learning Approach to Determine The Impact Of Socio Economic Factors On Bitcoin Price Prediction. In *2019 Twelfth International Conference on Contemporary Computing (IC3)* (pp. 1-5). IEEE.
- Akyildirim, E., Cepni, O., Corbet, S. ve Uddin, G. S. (2023). Forecasting Mid-Price Movement of Bitcoin Futures Using Machine Learning. *Annals of Operations Research*, 330(1), 553-584.
- Awoke, T., Rout, M., Mohanty, L. ve Satapathy, S. C. (2020). Bitcoin Price Prediction and Analysis Using Deep Learning Models. In *Communication Software and Networks: Proceedings of INDIA 2019* (pp. 631-640). Singapore: Springer Singapore.
- Box, George; Jenkins, Gwilym (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Ciaian, P., Rajcaniova, M. ve Kancs, D. A. (2016). The Economics of Bitcoin Price Formation. *Applied economics*, 48(19), 1799-1815.
- Chen, J. (2023). Analysis Of Bitcoin Price Prediction Using Machine Learning. *Journal of Risk and Financial Management*, 16(1), 51.
- Chen, Z., Li, C. ve Sun, W. (2020). Bitcoin Price Prediction Using Machine Learning: An Approach to Sample Dimension Engineering. *Journal of Computational and Applied Mathematics*, 365, 112395.
- Cheng, J., Tiwari, S., Khaled, D., Mahendru, M. ve Shahzad, U. (2024). Forecasting Bitcoin Prices Using Artificial Intelligence: Combination Of ML, SARIMA, And Facebook Prophet Models. *Technological Forecasting and Social Change*, 198, 122938.
- Cocco, L., Concas, G., ve Marchesi, M. (2017). Using An Artificial Financial Market for Studying a Cryptocurrency Market. *Journal of Economic Interaction and Coordination*, 12, 345-365.
- Çılğın, C. ve Özdemir, M. O. (2023). Time Series Forecasting of Covid-19 Confirmed Cases in Turkey with Stacking Ensemble Models. *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (26), 504-520.
- de Amorim, L. B., Cavalcanti, G. D. ve Cruz, R. M. (2023). The Choice of Scaling Technique Matters For Classification Performance. *Applied Soft Computing*, 133, 109924.
- Demirci, E. ve Karaatlı, M. (2023). Kripto Para Fiyatlarının Lstm ve Gru Modelleri İle Tahmini. *Journal of Mehmet Akif Ersoy University Economics and Administrative Sciences Faculty*, 10(1), 134-157.

- Büyükkör, Y. (2024). Derin Öğrenme ve Ekonometrik Model ile Bitcoin Fiyat Tahmini: LSTM ve ARIMA. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 26(47), 978-993.
- Dooley, G. ve Lenihan, H. (2005). An Assessment of Time Series Methods In Metal Price Forecasting. *Resources Policy*, 30(3), 208-217.
- Ediger, V. Ş. ve Akar, S. (2007). ARIMA Forecasting of Primary Energy Demand By Fuel In Turkey. *Energy policy*, 35(3), 1701-1708.
- Fang, F., Chung, W., Ventre, C., Basios, M., Kanthan, L., Li, L. ve Wu, F. (2024). Ascertainin Price Formation in Cryptocurrency Markets With Machine Learning. *The European Journal of Finance*, 30(1), 78-100.
- Fleischer, J. P., von Laszewski, G., Theran, C. ve Parra Bautista, Y. J. (2022). Time Series Analysis of Cryptocurrency Prices Using Long Short-Term Memory. *Algorithms*, 15(7), 230.
- Hamayel, M. J. ve Owda, A. Y. (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM And Bi-LSTM Machine Learning Algorithms. *AI*, 2(4), 477-496.
- Hochreiter, S. (1997). Long Short-Term Memory. *Neural Computation MIT-Press*.
- Hu, M. Y., Zhang, G., Jiang, C. X. ve Patuwo, B. E. (1999). A Cross-Validation Analysis Of Neural Network Out-Of-Sample Performance In Exchange Rate Forecasting. *Decision Sciences*, 30(1), 197-216.
- Jang, H. ve Lee, J. (2017). An Empirical Study on Modelling And Prediction Of Bitcoin Prices With Bayesian Neural Networks Based On Blockchain Information. *IEEE Access*, 6, 5427-5437.
- Jaquart, P., Dann, D. ve Weinhardt, C. (2021). Short-Term Bitcoin Market Prediction Via Machine Learning. *The journal of finance and data science*, 7, 45-66.
- Ji, S., Kim, J. ve Im, H. (2019). A Comparative Study of Bitcoin Price Prediction Using Deep Learning. *Mathematics*, 7(10), 898.
- Katsiampa, P. (2017). Volatility Estimation for Bitcoin: A Comparison of GARCH Models. *Economics letters*, 158, 3-6.
- Kercheval, A. N. ve Zhang, Y. (2015). Modelling High-Frequency Limit Order Book Dynamics with Support Vector Machines. *Quantitative Finance*, 15(8), 1315-1329.
- Lamothe-Fernández, P., Alaminos, D., Lamothe-López, P. ve Fernández-Gámez, M. A. (2020). Deep Learning Methods For Modeling Bitcoin Price. *Mathematics*, 8(8), 1245.
- Latif, N., Selvam, J. D., Kapse, M., Sharma, V. ve Mahajan, V. (2023). Comparative Performance of LSTM And ARIMA For the Short-Term Prediction Of Bitcoin Prices. *Australasian Accounting, Business and Finance Journal*, 17(1), 256-276.
- Liu, M., Li, G., Li, J., Zhu, X. ve Yao, Y. (2021). Forecasting The Price of Bitcoin Using Deep Learning. *Finance research letters*, 40, 101755.
- Maleki, N., Nikoubin, A., Rabbani, M. ve Zeinali, Y. (2023). Bitcoin Price Prediction Based On Other Cryptocurrencies Using Machine Learning And Time Series Analysis. *Scientia Iranica*, 30(1), 285-301.
- Mallqui, D. C. ve Fernandes, R. A. (2019). Predicting The Direction, Maximum, Minimum and Closing Prices of Daily Bitcoin Exchange Rate Using Machine Learning Techniques. *Applied Soft Computing*, 75, 596-606.
- Matkovskyy, R. ve Jalan, A. (2019). From Financial Markets to Bitcoin Markets: A Fresh Look At The Contagion Effect. *Finance research letters*, 31, 93-97.

- Büyükkör, Y. (2024). Derin Öğrenme ve Ekonometrik Model ile Bitcoin Fiyat Tahmini: LSTM ve ARIMA. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 26(47), 978-993.
- McIntyre, K. H. ve Harjes, K. (2016). Order Flow and The Bitcoin Spot Rate. *Applied Economics and Finance*, 3(3), 136-147.
- McNally, S., Roche, J. ve Caton, S. (2018, March). Predicting The Price of Bitcoin Using Machine Learning. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)* (pp. 339-343). IEEE.
- Mudassir, M., Bennbaia, S., Unal, D. ve Hammoudeh, M. (2020). Time-Series Forecasting of Bitcoin Prices Using High-Dimensional Features: A Machine Learning Approach. *Neural computing and applications*, 1-15.
- Nakamoto, S. (2008). Bitcoin: A Peer-To-Peer Electronic Cash System. *Satoshi Nakamoto*.
- Oprea, S. V., Georgescu, I. A. ve Bâra, A. (2024). Is Bitcoin Ready to Be A Widespread Payment Method? Using Price Volatility and Setting Strategies for Merchants. *Electronic Commerce Research*, 1-39.
- Patel, K., Mehta, D., Mistry, C., Gupta, R., Tanwar, S., Kumar, N. ve Alazab, M. (2020). Facial Sentiment Analysis Using AI Techniques: State-Of-The-Art, Taxonomies, And Challenges. *IEEE access*, 8, 90495-90519.
- Patel, M. M., Tanwar, S., Gupta, R. ve Kumar, N. (2020). A Deep Learning-Based Cryptocurrency Price Prediction Scheme for Financial Institutions. *Journal of information security and applications*, 55, 102583.
- Peng, Y., Albuquerque, P. H. M., de Sá, J. M. C., Padula, A. J. A. ve Montenegro, M. R. (2018). The Best of Two Worlds: Forecasting High Frequency Volatility for Cryptocurrencies And Traditional Currencies With Support Vector Regression. *Expert Systems with Applications*, 97, 177-192.
- Rathore, R. K., Mishra, D., Mehra, P. S., Pal, O., Hashim, A. S., Shapi'i, A., ... ve Shutaywi, M. (2022). Real-World Model for Bitcoin Price Prediction. *Information Processing & Management*, 59(4), 102968.
- Saad, M., Choi, J., Nyang, D., Kim, J. ve Mohaisen, A. (2019). Toward Characterizing Blockchain-Based Cryptocurrencies for Highly Accurate Predictions. *IEEE Systems Journal*, 14(1), 321-332.
- Shin, M., Mohaisen, D. ve Kim, J. (2021, January). Bitcoin Price Forecasting Via Ensemble-Based LSTM Deep Learning Networks. In *2021 International Conference on Information Networking (ICOIN)* (pp. 603-608). IEEE.
- Kaya, U., Akba, F., Medeni, İ. ve Medeni, T. (2020). Covid-19 Öncesi Ve Sonrasındaki Bitcoin Fiyat Değişimlerinin Makine Öğrenmesi, Zaman Serileri Analizi Ve Derin Öğrenme Yöntemleriyle Değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 13(3), 341-355.
- Urquhart, A. (2016). The Inefficiency of Bitcoin. *Economics Letters*, 148, 80-82.

Extended Abstract

Bitcoin Price Prediction Using Deep Learning and Econometric Model: LSTM and ARIMA

Bitcoin (BTC), the world's most popular cryptocurrency, has become interesting for both investors and researchers in recent years. Developed by Satoshi Nakamoto (2008), BTC was introduced as a peer-to-peer (P2P) money transfer system and distinguished from traditional currencies due to its decentralized structure. While transferring money between peers, BTC uses cryptology models to ensure that the money transfer process is verified by users within the network. This method ensures the anonymity of both the sender and the receiver on the network.

As it differs from traditional currencies, BTC tends to be seen as a risk, particularly by authorities, and causes controversy discussion. These arguments mainly centers on the weakening of the authority and the flow of black money. However, technology and e-commerce companies such as Microsoft, Amazon, and Dell accept Bitcoin as a method of payment in order to satisfy their customers.

Accurate prediction of Bitcoin (BTC) price improve investors' decision-making processes and facilitate the establishment of regulatory rules by governments. However, the extreme volatility and speculative nature of Bitcoin's price make accurately prediction difficult. In recent years, machine learning (ML) approaches have emerged as an effective approach to dealing with this difficulty, due to their capacity to analyze complicated non-linear relationships and adapt to the BTC market's changing dynamics. The difficulty in applying traditional econometric methods to cryptocurrencies due to their fundamental characteristics has led researchers and investors to increasingly favor machine learning and deep learning techniques, which require fewer or no assumptions.

This study compares the predictive performance of two different approaches to Bitcoin (BTC) closing price prediction: the traditional econometric Autoregressive Integrated Moving Average (ARIMA) model and the deep learning Long Short-Term Memory (LSTM) model. A complete comparison of these methods is performed using daily BTC closing price data from January 1, 2018 to December 21, 2023, for a total of 2181 observations. The prediction performance of both ARIMA and LSTM models is evaluated using performance metrics that include root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The data for this study was obtained from the Yahoo Finance website (<http://finance.yahoo.com>). It includes daily Open, Close, High, Low, Volume, and Date information for the BTC price. Other variables were removed from the analysis as only the BTC closing price was considered. While traditional machine learning algorithms typically split data randomly into training and testing sets, the time series nature of BTC price data and the significance of past information for analysis required a different approach. Consequently, the dataset was divided based on the time series data. As a result, the first 80% of the data (from January 1, 2018, to October 10, 2022) was split as the training set, and the remaining 20% (from October 11, 2022, to December 21, 2023) as the testing set. Python 3.9.16 and EViews software were utilized for BTC price prediction. The *auto.arima* library was used for ARIMA modeling in Python, while the *pandas*, *keras*, *tensorflow*, *Sequential*, and *LSTM* libraries were employed for LSTM model.

To improve the prediction performance of the used algorithms, the hyperparameters were investigated using the Grid Search approach. The autoregressive parameter p and moving average parameter q were tested with integers ranging from 1 to 10 using the Grid Search method, while the lag parameter d was set to 1. Before constructing an LSTM model, the data was first scaled using the Robust Scale method. An LSTM model was then constructed with two LSTM layers, one Dropout layer, and two Dense layers. Also, as the dataset used was a time series, prediction was performed using the Sliding Windows approach to preserve the relationship between observations. The performance of the methods was evaluated using RMSE, MAE, and MAPE performance criteria.

According to the application results, the LSTM model, which takes into account the long-term dependencies in the data, gave better RMSE and MAPE values than both ARIMA models, while the ARIMA(2,1,2) model gave better MAE values. Considering the studies in the literature, if the trend of the data set is increasing, the ARIMA model produces results at least as good as machine learning methods. In this study, since the trend of the BTC data set is increasing, ARIMA and LSTM models gave similar results. ML methods are frequently used to predict cryptocurrency prices due to their non-linear structure and volatility. However, even though ML methods can be easily applied because they are not based on any assumptions, traditional econometric methods should not be ignored, and results must be obtained to compare with ML methods.
