

# Covid-19 Öncesi ve Sonrasındaki Bitcoin Fiyat Değişimlerinin Makine Öğrenmesi, Zaman Serileri Analizi ve Derin Öğrenme Yöntemleriyle Değerlendirilmesi

*Araştırma Makalesi/Research Article*

 Uğur KAYA<sup>1</sup>,  Fırat AKBA<sup>2</sup>,  İhsan Tolga MEDENİ<sup>1</sup>,  Tunç Durmuş MEDENİ<sup>1</sup>

<sup>1</sup>Yönetim Bilişim Sistemleri Bölümü, Ankara Yıldırım Beyazıt Üniversitesi, Ankara, Türkiye

<sup>2</sup>Bilgisayar Mühendisliği Bölümü, Ankara Üniversitesi, Ankara, Türkiye  
[u.06.kaya@gmail.com](mailto:u.06.kaya@gmail.com), [firatakba@gmail.com](mailto:firatakba@gmail.com), [tolgamedeni@gmail.com](mailto:tolgamedeni@gmail.com), [tuncmedeni@gmail.com](mailto:tuncmedeni@gmail.com)

(Geliş/Received:19.11.2019; Kabul/Accepted:19.07.2020)

DOI: 10.17671/gazibtd.648424

**Özet**— Son zamanlarda kullanımı oldukça yaygınlaşan blokzinciri teknolojisinin, İnternet teknolojisi ile beraber adı sıkça anılır olmaya başlamıştır. Blokzinciri teknolojisiyle geliştirilen Bitcoin, sanal para birimleri arasında en çok piyasa hacmini elinde bulunduran sanal para birimidir. Sanal para piyasalarının kontrolünü elinde bulunduran bir merkezi otoritenin olmaması sebebiyle fiyat manipülasyonlarına ve dışarıdan müdahalelere açık olan bu pazarda, en uçtaki yatırımcının yatırım yapabilmesi açısından yol gösterimine ihtiyaç duyulmaktadır. Son zamanlarda bu ihtiyacı karşılamak amacıyla birtakım yöntemler kullanılmaya başlanmıştır. Bu çalışmada makine öğrenmesi, zaman serileri analizi ve derin öğrenme yöntemleri kullanılarak Bitcoin fiyatlarındaki dalgalanma hakkında çeşitli tahminleme ve sınıflama yöntemleri beraber olarak değerlendirilmiştir. Bu bağlamda, koronavirüs pandemisi öncesi ve sonrasındaki Bitcoin kapanış fiyatları ve düşüş-yükseliş eğilimleri baz alınarak iki ayrı veri kümesi oluşturulmuştur. Bu veri kümeleri üzerinde tahmin ve sınıflama yöntemleri değerlendirilerek, başarıları karşılaştırılmıştır. Karşılaştırmalar sonucunda, pandemi öncesi verilerle yapılan çalışmada Destek Vektör Makineleri, pandemi sonrası verilerle yapılan çalışmada ise ARIMA en başarılı sonuçları vermiştir.

**Anahtar Kelimeler**— tahminleme, uzun kısa süreli bellek ağları algoritması, bütünleşik otoregresif hareketli ortalama yöntemi, destek vektör makineleri algoritması, bitcoin, blokzinciri teknolojisi, kriptopara, sanal para, koronavirüs

## Evaluation of Bitcoin Price Changes Before and After Covid-19 by Machine Learning, Time Series Analysis and Deep Learning Algorithms

**Abstract**— Blockchain technology, which has been become quite widespread in use recently, has become very popular with the Internet technology. Bitcoin, which has been developed with blockchain technology, is the virtual currency that holds the most market volume among virtual currencies. Due to the lack of a central authority that controls the virtual currency markets, this market is open to price manipulations and external interventions, so that guidance is needed for the end-investor to invest. Recently, a number of methods have started to be used to meet this need.

In this study, various forecasting and classification methods about fluctuation in Bitcoin prices were evaluated together using machine learning, time-series analysis and deep learning methods. In this context, two separate datasets have been created based on the Bitcoin closing prices and up-to-down trends before and after the coronavirus pandemic. The success of forecasting and classification methods on these two datasets were evaluated and compared. As a result of the comparisons, Support Vector Machines method for the study conducted with the data before the pandemic, and ARIMA method for the study conducted with the data after the pandemic, had the most successful results.

**Keywords**— forecasting, long-short term memory algorithm, auto regressive integrated moving average method, support vector machines algorithm, bitcoin, blockchain technology, cryptocurrency, virtual currency, coronavirus

## 1. GİRİŞ (INTRODUCTION)

Dijital yollarla gerçekleştirilen işlemlerin artmasıyla birlikte kriptopara birimleri de üretilmeye ve kullanılmaya başlanmıştır. Finansal bazda en önemli içerik olan parasal karşılık, sanal dünyada kendine kriptoparalarla birlikte yer bulabilmiştir. Kriptopara veya diğer adıyla sanal para, para gibi işlev gören ama paranın aksine ulusal sınırlar ve merkez bankalarından bağımsız bir değişim aracıdır [1]. Kriptopara üretiminin altında blokzincir adı verilen bir teknoloji yatmaktadır. Blokzinciri teknolojisi, yapılan tüm işlemlerin şifrelenerek kaydının tutulmasını ve bu kayıtların bloklar halinde saklanmasını sağlamaktadır [2]. Özellikle kriptopara uygulamalarıyla tanınan blokzincir teknolojisinin kullanımı son zamanlarda artmaya başlamıştır. Blokzinciri teknolojisinde kullanılan bloklar, özellikle işlemlerin güvenliği açısından büyük önem arz etmektedir. Bir blokzincir, eklenen yeni bloklarla birlikte genişlemektedir. Blokzincirindeki her bir blok, kendisinden önce gelen blokla ilişkilendirilmektedir. Blokların her biri, zaman damgasını(timestamp), kendinden önceki bloğun özet değerini(hash value) ve bu özet değeri doğrulamak için kullanılan değeri(nonce) barındırmaktadır [3]. Blokzincir teknolojisinde, her kullanıcı bir genel(public) ve bir özel(private) anahtara sahiptir. Bu anahtarlar kriptolojik olarak blokzincirinde depolanmaktadır. Ayrıca, blokzincir işlemlerinde akıllı sözleşmelerden(smart contracts), dijital imzalardan ve uzlaşma protokollerinden(consensus protocols) de yararlanılmaktadır [4]. Akıllı sözleşmeler, birtakım kurallara bağlı işlemler için kullanılmaktadır, fakat diğer işlemler için kullanımı bir zorunluluk değildir. Kullandığı bu tür yapılarla birlikte, blokzincirinin sahip olduğu merkezi olmayan dağıtık yapı, şeffaflık ve kriptografik algoritmalar zamanla kriptoparalara olan yönelimi arttırmıştır [5]. Blokzinciri teknolojisiyle geliştirilmiş ilk kriptopara birimi Bitcoin(BTC)'dir. BTC, kullanıcılara ait tüm tasarrufların korunduğu, yapılan her işlemin kayıt altına alındığı, halka açık ve uçtan uca(peer to peer) çalışan bir elektronik ödeme sistemidir [6]. BTC'den sonra birçok kriptopara birimi piyasaya sürülmüştür. BTC'nin ardından zayıf piyasa değerlerine sahip bu kriptoparaların piyasaya sürülmesi, kriptoparaların ticaret hacminde ani sivrilmelere ve fiyat manipülasyonlarına yol açmıştır [7]. Bu ani sivrilmeler ve manipülasyonlar, bazı etkenlerle birlikte BTC'nin fiyat akışında da görülmüştür. Örneğin, Zimbabwe'de yaşanan hiperenflasyon, BTC fiyatını önemli ölçüde etkilemiştir. 9 Kasım 2017 itibarıyla Harare merkezli BTC borsası Golix'te oluşan yüksek ticaret hacmi, BTC fiyatını dünya genelinde aynı dönemde geçerli olan ortalama BTC fiyatının 6.500 dolar üzerine taşıyarak 13.900 dolara yükseltmiştir [8]. BTC'nin fiyat akışını değiştirebilen bazı politik ve sosyal etkenler de vardır. Politik manada, bazı hükümet yetkilileri BTC'nin yasadışı geçişleri kolaylaştırabileceğini ve devletin para politikası uygulamak gibi faaliyetlerini bozabileceğini söylemiştir. Bu söylemler, BTC'ye bazı kısıtlama ve yaptırımlar yapılmasıyla sonuçlanmıştır [9]. Bu durum, kullanıcıların BTC'ye yönelimlerini olumsuz yönde etkilemiştir. Sosyal manada bakıldığında ise, liberaller ve merkez sol kesim tarafından BTC'ye alternatif bir para birimi gözüyle

yaklaşılması, BTC için olumlu sonuçlar doğurmuştur [10]. Fiyat manipülasyonlarına açık olan bu sektörde akışlara müdahale edebilecek güçte olan şüpheli işlemler de, BTC fiyatının aşırı düşmesine veya aşırı yükselmesine sebep olmaktadır. Buna ek olarak, alternatif ödeme teknolojilerinin potansiyel olarak sanal para hırsızlığına ve karapara aklama riskine açık olması da BTC fiyat akışını etkileyebilmektedir [11].

Dünya üzerindeki tüm ülkeler için büyük bir problem haline gelen koronavirüs(Covid-19) pandemisi, ülkelerin ekonomik politikalarının değişmesi gerekliliğini de beraberinde getirmiştir [12]. Koronavirüs pandemisinin geleneksel piyasalarda yaratmış olduğu çöküşün dolaylı olarak BTC'nin fiyatlarındaki dalgalanmaları olumlu yönde etkileyebileceği de iddia edilmiştir [13].

BTC fiyat akışını değiştirebilecek tüm bu etkenler göz önüne alındığında, fiyat akış yönünün belirlenmesi büyük bir ehemmiyet kazanmıştır. En uçtaki yatırımcının BTC'ye yatırım yapabilmesi için bir yol gösterimine ihtiyaç duyulmaktadır. Bu yol gösteriminin sağlanması için, gerekli analizlerin yapılması gerekmektedir. Bilgisayar tabanlı olarak analizlerin yapılabildiği bir karara varılmasını sağlayan yöntemler bulunmaktadır [14]. Sınıflandırma ve tahminleme algoritmaları, bu yöntemler kapsamındadır. Sınıflandırma(classification) algoritmaları, veri çıktılarının nitel olması durumunda verilerin hangi sınıfa ait olduğunu belirlemeye yarar [15]. Bu algoritmalar, sağlanan eğitim verilerinden sınıf dağılımını öğrenir. Ardından, test verilerini doğru bir biçimde sınıflandırmaya çalışır. Tahminleme(forecasting) algoritmaları, geçmiş verileri kullanarak gelecekteki eğilimlere dair çıkarımlar yapılmasını sağlar [16].

Çalışmada, makine öğrenmesi, zaman serileri analizi ve derin öğrenme yöntemleri kullanılarak koronavirüs pandemisi öncesi ve sonrası Bitcoin fiyatlarındaki dalgalanma hakkında çeşitli tahminleme ve sınıflama yöntemleri beraber olarak değerlendirilmiştir. Sınıflandırma yöntemi olarak SVM, tahminleme yöntemi olarak LSTM ve ARIMA algoritmaları kullanılmıştır. Kısa dönem zaman serisi tahminlemelerindeki uygunluğu nedeniyle ARIMA, güçlü katman yapıları nedeniyle LSTM ve sınıflandırıcılarının tahminleyici işlevi nedeniyle SVM algoritmaları seçilmiştir.

Bu çalışmanın amacı, koronavirüs pandemisi öncesi ve sonrasındaki haftalık etkiler üzerinden BTC fiyatlarının dönemsel düşüş, yükseliş eğilimlerini SVM, ARIMA ve LSTM yöntemleriyle tahmin edilerek, en başarılı yöntem veya yöntemlerin tespit edilmesi ve virtüs sürecindeki fiyat değişimlerinin değerlendirilmesidir. Takip eden bölümlerde, sırasıyla bu konudaki literatür taranıp daha önceki çalışmalara yer verilecektir. Ardından, kullanılan tüm yöntemler tanıtılacak ve uygulama aşamasında bu yöntemlerin nasıl uygulandığı hakkında bilgi verilecektir. Bu algoritmaların BTC fiyat eğilimlerini belirlemedeki model performansları birbiriyle karşılaştırılacaktır. Literatürde olan yayınlardan farklı olarak, ölçümlerde daha fazla performans metriği dikkate alınacaktır.

## 2. LİTERATÜR ÇALIŞMASI (LITERATURE REVIEW)

Kriptoparaların fiyat akışının belirlenmesi, kriptoparalara yatırım yapan yatırımcıların ilgisini büyük ölçüde çekmeyi başarmıştır. Bu mevzu, her geçen gün daha da önemli bir hale gelmeye başlamıştır. Birçok araştırmacı da bu konuda çalışmalarda bulunmuşlardır. Bu bölümde, özellikle son beş yılda kriptoparalarla ilgili yapılan çalışmaları detaylandırmayı amaçlamaktayız.

Jang ve Lee(2018) ileri sürdükleri çalışmada, makroekonomik değişkenleri Blockchain bilgilerine ekleyip BTC fiyatlarında zaman serisi analizi yapmışlardır. Çalışmada, Bayeşçi Sinir Ağları (Bayesian Neural Network-BNN) algoritması tahminleme için kullanılmıştır. Modelin, günlük fiyat ve günlük değişimleri baz alarak BTC fiyatlarındaki dalgalanmaları etkili bir şekilde gösterdiği görülmüştür. BNN algoritmasının, Doğrusal Regresyon(Linear Regression-LR) ve Destek Vektör Bağlayıcısı(Support Vector Regressor-SVR) yöntemlerinden tahminleme konusunda daha iyi olduğu iddia edilmiştir [17].

Katsiampa(2017) BTC fiyatlarındaki değişkenliği ortaya koymak amacıyla bazı Genelleştirilmiş Otoregresif Koşullu Değişen Varyans(GARCH) modellerini oluşturarak birbirleriyle kıyaslamıştır. Bu modellerin kıyaslanması sonucunda, fiyat değişkenliğinin uyum iyiliği anlamında optimal olarak açıklamasını yapabilecek bir Asimetrik Bileşenli Genelleştirilmiş Otoregresif Koşullu Değişen Varyans(AR-CGARCH) modeli geliştirmişlerdir. Modelin, koşullu varyansın kısa vadeli ve uzun vadeli bileşenlerine sahip olmasının önemi vurgulanmıştır. Ayrıca, modelin risk yönetiminde, portfolyo ve müşteri hassasiyeti analizinde yatırımcılar tarafından bir karar verme mekanizması olarak kullanılabilceği ifade edilmiştir [18].

Sutiksno, Ahmar, Kurniasih, Susanto ve Leiwakabessy(2018)  $\alpha$ -Sutte Indicator adını verdikleri yeni bir tahminleme yöntemi ile birlikte Bütünleşik Otoregresif Hareketli Ortalama(Auto Regressive Integrated Moving Average-ARIMA) ve Sinir Ağları Zaman Serisi Tahminlemesi(Neural Network Time Series Forecast-NNETAR) araçlarını kullanarak geçmiş BTC verileri üzerinde bir fiyat tahminleme çalışması yapmayı amaçlamışlardır. ARIMA(1,1,1) ve NNETAR(21,11) parametrelerini kullanan araştırmacılar, ölçümler için ortalama karesel hata(MSE) ve ortalama mutlak hata(MAE) değerlerini kullanmışlardır. Çalışma sonucunda,  $\alpha$ -Sutte Indicator yönteminin verdiği hata oranının ARIMA ve NNETAR algoritmalarının hata oranından daha düşük olduğunu gözlemlemişlerdir. BTC geçmiş verileri üzerinde tahminleme yapmak için oluşturmuş oldukları bu yöntemin oldukça uygun olduğunu tespit etmişlerdir [19].

Abu Bakar ve Rosbi(2017) BTC'nin 2013 Ocak ayından 2017 Ekim ayına kadar gelişen döviz kurunu incelemişlerdir. BTC/USD kuru üzerinden topladıkları verilerle ARIMA yöntemini kullanarak bir fiyat tahmin

çalışması yapmışlardır. BTC/USD kurunun durağanlık durumunu analiz etmek için otokorelasyon ve kısmi otokorelasyon fonksiyonları kullanılmıştır. Yıpranmanın yavaş olduğu ve kurun durağan olmadığı tespit edilmiştir. ARIMA (2,1,2) modelini kullanan araştırmacılar, bu yöntemin sözkonusu veriler için güvenilir sonuçlar verdiğini ifade etmişlerdir. Ayrıca, yüksek değişkenliğe sahip olan kurlarda ARIMA yönteminin kullanılması halinde hata tanınmalarında özel değerlendirmeler yapılmasına ihtiyaç duyulduğu belirtilmiştir [20].

Karakoyun ve Çıbıkdiken(2018), geçmiş BTC fiyatlarının verileri üzerinden bir sonraki otuz günün değerlerini ortaya çıkarmak için bir fiyat tahminlemesi uygulamışlardır. Çalışmada, ARIMA ve Uzun Kısa Süreli Bellek (Long-Short Term Memory-LSTM) algoritmaları kullanılmıştır. ARIMA(4,2,1) ve LSTM algoritmalarından çıkan sonuçların MAPE değerlerini, hata ölçüm oranı olarak kullanmışlardır. Hata oranları, ARIMA modeli için % 11,86, LSTM algoritması için % 1,4 olarak bulunmuştur. Derin öğrenme algoritmalarının fiyat tahminlemelerinde kullanılmak için uygun olduğu sonucuna varılmıştır [21].

McNally(2018), derin öğrenme ve zaman serisi tahminleme algoritmalarını BTC fiyatları üzerinde kullanmıştır. Tekrarlayan Sinir Ağları(Recurrent Neural Network-RNN), LSTM ve ARIMA algoritmaları kullanıldığı çalışmada, algoritmalara ait tahminleme başarıları kıyaslanmıştır. ARIMA modelinin, LSTM ve RNN algoritmalarının tahminleme performansları karşısında oldukça güçsüz bir tahminleme performansı gösterdiği belirtilmiştir. LSTM algoritması, RNN algoritmasından biraz daha yüksek bir performans göstermiştir. Ayrıca, BTC verilerindeki yüksek varyansın, tahminleme sonuçlarını büyük ölçüde zorlaştırdığının da altı çizilmiştir [22].

Sakız ve Kutlugün(2018), yapay sinir ağları yöntemlerini kullanarak 2015 Ocak ve 2018 Nisan ayları arasındaki aylık BTC fiyatlarının ortalamaları hesap edilerek oluşturulmuş bir veri üzerinde çalışmışlardır. Bu verileri kullanarak 2018 Mayıs ayına dair bir ortalama fiyat tahminlemesi yapmaya çalışmışlardır. Çıkan sonuçlara göre, BTC fiyatlarına dair tahminlenmiş değer ve gerçekleşen değer arasında büyük farklılık gözlenmiştir. Eğitilen verilerin fazla olmamasının ve yöntemlerde kullanılan parametre seçiminin bu büyük farklılığa etki ettiği sonucuna varılmıştır. Ayrıca, BTC piyasasında illegal faaliyetlerin bulunması ve mevcut talebin öngörülen talebe göre düşük olması gibi çıkarımlarda da bulunulmuştur [23].

Azari(2019), 2015 Eylül ayından itibaren üç yıllık bir periyottaki geçmiş kapanış fiyatlarını baz alarak BTC kuru için bir tahminleme çalışması yapmıştır. Çalışmada, BTC fiyatlarını tahminlemek amacıyla ARIMA modeli kullanılmıştır. Araştırmacılar, modelin hata oranı ölçümü için MSE değerini kullanmışlardır. MSE değerini minimize eden değeri saptamak için ARIMA modelinin değişik parametreleri denenmiştir. MSE değerlerine göre, BTC fiyatlarının yüksek sıçrama ve çöküşlerindeki

kırılganlığın uzun dönemlerde düşük tahminleme performansına sebep olduğu görülmüştür. Kısa dönemlerde ise ARIMA modelinin kullanılmaya uygun olduğu belirtilmiştir [24].

Olvera-Juarez ve Huerta-Manzanilla(2019), BTC fiyat tahminlemede kullanılan bazı karma yöntemleri incelemişlerdir. Araştırmacılar, ARIMA, LSTM ve RNN algoritmalarının BTC fiyatlarını tahminlemeye yönelik performanslarını incelemişlerdir. ARIMA algoritması durağan olmayan dönemlerde kullanıldığında, tatmin edici tahminlemelerin görülmediği belirtilmiştir. Ayrıca, LSTM ve RNN algoritmaları ile BTC fiyatlarını tahminlemek için uygun yapılar kurulabileceğini vurgulamışlardır [25].

Greaves ve Au(2015), BTC işlem grafiklerindeki düşüş ve yükseliş hareketlerini baz alarak bir fiyat tahminleme çalışması uygulamışlardır. Araştırmacılar, farklı yöntemler kullanılarak yapılan tahminlerdeki hata payını ölçmek üzere MSE metriğini kullanmışlardır. SVM ve Doğrusal Regresyon ile yaptıkları regresyon analizinde tahminlemeye yönelik MSE sonuçları düşük çıkmıştır. Lojistik Regresyon(LR), SVM ve Neural Network ile yaptıkları sınıflandırma modelinde ulaştıkları doğruluk seviyesi %53 ve %55 arasında seyretmiştir [26].

Almeida, Tata, Moser ve Smit(2015), önceki günlerin kapanış fiyatlarını ve hacimlerini baz alarak sonraki gün için BTC fiyatının eğilimini tahminlemişlerdir. Çalışmada, tahminleme yapmak üzere yapay sinir ağları(ANN) algoritması kullanılmıştır. Hacimlerin hesaba katıldığı ölçümlerin sonucu, sadece kapanış fiyatlarının yer aldığı ölçümlerin sonucundan daha düşük çıkmıştır. Buna ek olarak, basit yapay sinir ağ yapısı ve karmaşık yapay sinir ağ yapısı karşılaştırılmıştır. Basit ağ yapılarının tahminlemede daha yüksek performansa sahip olduğu görülmüştür. Farklı nöron sayıları ve geciktirme metrikleri kullanan araştırmacılar, nöron sayısı düşük olsa bile basit ağ yapılarındaki performansın pozitif bir ivme kaydettiğini vurgulamışlardır [27].

Hitam ve Ismail(2018), makine öğrenmesi algoritmalarının fiyat tahminleme başarılarını karşılaştırmak üzere bir çalışmada bulunmuşlardır. Fiyat tahminlemeleri, sözkonusu dönemde revaçta bulunan altı kriptopara birimi üzerinden yapılmıştır. Bu kriptopara birimleri, BTC, Ethereum, Litecoin, Nem, Ripple ve Stellar olarak belirlenip tahminlemeye dahil edilmiştir. Yapılan tahminlerde, SVM'in diğer sınıflayıcılara göre daha yüksek doğruluk oranına sahip olduğu gözlemlenmiştir. Ayrıca, SVM yönteminin kullanılan diğer sınıflayıcılara kıyasla daha düşük bir MAPE oranına sahip olduğu sonucuna ulaşılmıştır. Veri kümesindeki toplam veri sayısının ve eğitilen verinin kalitesinin, başarılı bir tahminleme için kilit noktalar olduğu belirtilmiştir [28].

Literatürde incelenen makalelerde yer alan tüm yöntemler, değerlendirme metrikleri ve bu yöntemlerden elde edilen sonuçlar derlenerek tabloda özetlenmiştir (Tablo 1).

Tablo 1. Literatüre ait özet tablo  
(Summary table of the literature)

Makale	Yöntem	Değerlendirme Metriği	Sonuç
[17]	BNN	MAPE	% 10,08
		RMSE	0,5750
[18]	AR-CGARCH	RMSE	0,8910
[19]	$\alpha$ -Sutte Indicator	MSE	121362,344
		MAE	299,766
[20]	ARIMA	MAPE	% 46,40
[21]	ARIMA	MAPE	% 11,86
	LSTM	MAPE	% 1,4
[22]	RNN	RMSE	0,9455
	LSTM	RMSE	0,9193
	ARIMA	RMSE	0,4626
[23]	ANN	MSE	0,00021042
[24]	ARIMA	MSE	118000,0
[25]	ARIMA+LSTM	Accuracy	%90,10
	ARIMA+RNN	Accuracy	%87,88
	ARIMA	Accuracy	%90,31
[26]	LR	Accuracy	%54,3
	SVM	Accuracy	%53,7
	NN	Accuracy	%55,1
[27]	ANN	MSE	0,03
[28]	SVM	MAPE	% 31

Literatür taraması sonucunda, SVM, ARIMA ve LSTM yöntemlerinin aynı veri kümesi üzerinde uygulandığı bir çalışma görülemedi. Bu sebepten, çalışmada kullanılmak üzere makine öğrenmesi tarafında SVM, derin öğrenme tarafında LSTM, zaman serisi tahminlemeleri için ise ARIMA yöntemleri seçilmiştir. Sözkonusu yöntemler, iki ayrı zaman dilimine ait (koronavirüs pandemisi öncesi ve koronavirüs pandemisi sonrası) veriler üzerinde uygulanmış olup koronavirüs pandemisinin BTC fiyatlarına olan etkisi de değerlendirilmeye çalışılmıştır.

### 3. YÖNTEMLER (METHODS)

#### 3.1. Veri Kümesi (Data Set)

Çalışmada, koronavirüs pandemisi öncesi ve koronavirüs pandemisi sonrası olmak üzere iki ayrı veri kümesi kullanılarak BTC fiyatlarının yükseliş ve düşüş eğilimleri literatürde kullanılmış en başarılı olarak gösterilen çeşitli yöntemler ile tahmin edilerek değerlendirilmiştir. Koronavirüs pandemisinin öncesini kapsayan veri kümesinde 5 Şubat 2018 ile 27 Ekim 2019 tarihleri arasında BTC/USD kurundaki gün sonu kapanış fiyatları esas alınmıştır. Koronavirüs pandemisinin sonrasını kapsayan çalışma için ise 5 Şubat 2018 ile 28 Haziran 2020 tarihleri arasında BTC/USD kurundaki gün sonu kapanış fiyatlarından yararlanılmıştır. Bahsedilen periyottaki günler, haftalık dilimlere çevrilmiştir. BTC, işlem hacminin büyüklüğü ve fiyat oynaklığının fazla olması nedeniyle gelişmekte olan ekonomileri etkileyebilecek güçtedir [29]. Günlük fiyat aralıklarında yüksek farklar olduğu gözlenen BTC'nin bir anda yükselip düşmeye başlaması sıkışma teorisiyle açıklanmaktadır [30]. Diğer yandan, aylık vadeli işlemlerin sonlarına denk gelen tarihlerde, piyasalardaki kriptopara hacmini düşürmeye

yönelik manipülasyonlar gözlemlenmiştir [31]. BTC, günlük ve aylık olarak manipülasyonlardan etkilenmektedir. Bu sebeplerden ötürü, günlük veya aylık olarak fiyat tahminleri yapmanın sistemin başarısını olumsuz etkileyeceği düşünülmüştür. Bu yüzden, çalışmada haftalık bir tahminleme sözkonusudur. Pandemi öncesi veriler 90 haftayı, pandemi sonrası çalışmayı içeren veriler ise 125 haftayı kapsamaktadır. BTC, haftanın tüm günlerinde işlem görebilmektedir. Her hafta için, haftanın ilk gününe(pazartesi) ait gün sonu kapanış fiyatı ve haftanın son gününe(pazar) ait gün sonu kapanış fiyatı göz önüne alınmıştır. Her haftaya ait eğilimler, düşüş ve yükseliş etiketleriyle derlenmiştir. Yapılan tahminleme, her hafta için o haftanın Pazartesi günü kapanış değeri ile Pazar günü kapanış değeri arasındaki eğilimi belirlemeye yöneliktir. Derleme sonucunda oluşan verideki öznitelikler “Hafta”, “Kapanış”, “Asıl Trend” olarak belirlenmiştir. Pandemi öncesi çalışma için kullanılan 90 haftalık verinin test kısmı için 30 haftanın ayrılmasının nedeni, üssel hareketli ortalamanın(Exponential Moving Average-EMA) 200 günlük periyodunun fiyat eğilimlerinde daha keskin ve daha doğru sonuçlar vermesinden kaynaklıdır. EMA, bir ortalama göstergesidir. Bu 200 günlük ortalama olarak 30 haftaya tekabül eder. EMA, borsadaki teknik analizlerde önemli bir yere sahiptir. BTC fiyatlarının haftalık trendlerini belirlemek ve onaylamak amacıyla borsada büyük bir öneme sahip olan trend çizgileri kullanılmıştır. Ayrıca, BTC fiyatlarının yükseliş trendinde olduğuna işaret eden boğa piyasasına(bullish market) girdiği dönemler ve aşağı yönlü düşüşe işaret eden ayı piyasasına(bear market) girdiği dönemler belirlenmiştir. Yayının geri kalanında verilecek grafik ve görsellerde kullanılacak ‘Hafta’ ifadesi hafta sayısını, ‘Kapanış’ ifadesi sözkonusu haftanın Pazar gününün kapanış fiyatını ve ‘Asıl Trend’ ifadesi ise mevcut haftanın düşüş ve yükseliş durumunu göstermektedir. Tahmin edilmiş değerler, yayının geri kalanında ‘Tahmin’ ifadesiyle belirtilmiştir. Buna ek olarak, etiketlenen düşüş ve yükseliş durumları, tahmin edilmiş değerlerle karşılaştırılmak üzere kullanılmıştır.

### 3.2. Veri Kaynağı (Source of Data)

Pandemi öncesi çalışma için belirli tarih aralıklarıyla filtrelenen 630 adet veri ve pandemi sonrası çalışma için 875 adet veri, www.coinmarketcap.com [32] adresinden toplanmıştır. Bu veriler haftalık olarak pandemi öncesi için 90 kayıt ve pandemi sonrası için 125 kayıt olarak tarafımızdan derlenmiştir. Çalışmada bu veriler kullanılmıştır. Sözkonusu adres birçok yatırımcı tarafından kriptopara birimlerinde güvenilir bir kaynak olarak görülmektedir.

### 3.3. Deney Aşamalarında Kullanılan Yöntemler (Methods Used in Experimental Stages)

Deney aşamalarında kullanmak için seçtiğimiz ARIMA, LSTM ve SVM algoritmaları bu kısımda detaylandırılmaktadır.

#### 3.3.1. Bütünleşik Otoregresif Hareketli Ortalama (Auto Regressive Integrated Moving Average-ARIMA)

ARIMA modeli, doğrusal zaman serisi modellerinde otoregresif(AR) ve hareketli ortalama(MA) olarak geçen modellerden yararlanır. Tek değişkenli zaman serisi tahminlerinde kullanılmaya uygun olan bu modelde üç parametre vardır. Bir ARIMA(p,d,q) modeli AR(p), I(d), MA(q) modellerinin birleşimiyle oluşur [33].

AR(p) notasyonu otoregresif modeli belirtir. Otoregresif model, bir değişkenin kendi gecikmeli(önceki) değerleri üzerinde bağımlı olarak gerilediğini göstermektedir. Bu model matematiksel olarak şu şekilde tanımlanabilir:

$$X_t = c + \sum_{i=1}^p \theta_i X_{t-i} + \epsilon_t \quad (1)$$

Formüldeki c ifadesi sabit değeri gösterir.  $X_t$  ifadesi bir zaman serisini belirtir ve t ifadesi zaman serisinin tamsayı indeksini gösterir.  $\theta_i$  ifadesi otoregresif modelin parametre değerlerini ifade eder, integralin üst kısmındaki p ifadesi ise modeldeki gecikme gözlemlerinin sayısını gösterir.  $\epsilon_t$  ifadesi, sıfır ortalamaya ve sabit varyansa sahip beyaz gürültülü hata terimini ifade eder.

MA(q) notasyonu hareketli ortalama modelini belirtir. AR(p) notasyonunun formülünden türetilen hareketli ortalama modeli, tahmin edilemeyen şimdiki ve geçmiş verilerin doğrusal olarak çıkış değişkenini ifade eder. Bu model şu şekilde gösterilebilir:

$$X_t = \mu + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (2)$$

Formüldeki  $\mu$  ifadesi modeldeki serinin ortalamasını, c ifadesi ise sabit değeri gösterir.  $\theta_i$  ifadesi hareketli ortalama modelinin parametre değerlerini ifade eder, integralin üst kısmındaki q ifadesi ise hareketli ortalamanın sırasını belirtir. Eşitlikteki  $X_t$  ifadesi zaman serisini ve t ifadesi ise zaman serisinin indeksini ifade eder.  $\epsilon_t$  ve  $\epsilon_{t-i}$  ifadeleri, sıfır ortalamaya ve sabit varyansa sahip beyaz gürültülü hata terimlerini ifade eder.

I(d) notasyonu ise gözlemlere ait fark alma mertebesini belirtir. ARIMA(p,d,q) modeli matematiksel olarak genelleştirilmiş formülü ile şu şekilde izah edilir:

$$X_t = \frac{(1 + \sum_{i=1}^q \theta_i L^i) \epsilon_t}{(1 - \sum_{i=1}^p \theta_i L^i) (1-L)^d} \quad (3)$$

Formüldeki L ifadesi gecikme operatörünü,  $\theta_i$  ifadesi otoregresif modelin parametre değerlerini,  $\theta_i$  ifadesi ise hareketli ortalama modelinin parametre değerlerini gösterir.  $X_t$  ifadesi zaman serisini, t ifadesi zaman serisinin indeksini ifade eder. Denklemdeki p ifadesi modeldeki gecikme gözlemlerinin sayısını, q ifadesi ise hareketli ortalamasının sırasını, d ifadesi ise modeldeki ham gözlemlerin kaç kez değiştirildiğini gösterir.  $\epsilon_t$  ifadesi, sıfır

ortalamaya ve sabit varyansa sahip beyaz gürültülü hata terimini ifade eder.

### 3.3.2. Uzun-Kısa Süreli Bellek (Long-Short Term Memory-LSTM)

LSTM modeli, 1997 yılında Hochreiter ve Schmidhuber tarafından tasarlanan, birleşme noktalarındaki eğim derecesi nedeniyle probleme yol açan yapıları ve uzun vadedeki bağımlılıkları hedef alan, doğrusal olmayan modellemelerde ve tahminlemelerde kullanılan komplike bir gizli birimler bütünüdür [34]. LSTM modelinde bir hücre durumu(cell state) yapısı vardır. Bu yapı, katman(gate) yapılarıyla bilgiler üzerinde kontrol sağlama yetkisine sahiptir. LSTM'nin temel katmanları, giriş katmanı(input gate), unutma katmanı(forget gate) ve çıkış katmanı(output gate) olarak üçe ayrılır.

Unutma katmanı, bilginin hücre durumunda ne kadar süre tutulacağını hesaplar. Bu katman, hücre durumundan hangi bilgilerin çıkarılacağına karar verilen katmandır. Bu katmanda 0 ve 1 arasında bir sonuç alınabilmesi için bir sigmoid fonksiyon kullanılır.

Giriş katmanı, bilginin hücre durumunda ne kadar süre girdi olarak işlem göreceğini hesaplar. Bu katman, hücre durumuna hangi yeni bilgilerin ekleneceğine karar verilen katmandır. Bu katmanda da yine aynı şekilde bir sigmoid fonksiyon kullanılır.

Giriş katmanının ardından, yeni duruma eklenebilecek aday verilerin tutulmasını sağlayan bir ara katman olarak tanh katmanı kullanılır. Unutma katmanı ve giriş katmanı yapılandırıldıktan sonra bu iki ana katman ve tanh ara katmanı birleştirilir ve güncellenmiş bir hücre durumu ortaya çıkar.

Çıkış katmanı, önemli bilgileri filtreleyip tahminlemede kullanılacak olan çıktıyı üreten katmandır. Hücre durumunun hangi kısımlarındaki bilgilerin çıkış katmanına aktarılacağına karar vermek için sigmoid fonksiyonu kullanılır. Hücre durumu yeniden tanh katmanından geçirilir ve çıkış katmanının çıktısıyla çarpılır. Böylece hücre çıktısı kazanılmış olur.

### 3.3.3. Destek Vektör Makineleri (Support Vector Machines-SVM)

SVM algoritması, 1992 yılında Boser, Guyon ve Vapnik tarafından geliştirilen, sınıflandırma ve regresyon problemlerinin çözümünde uygulanan ve aşırı öğrenmeyi azaltan yapıyla daha iyi sonuçlar veren güdümlü bir öğrenme algoritmasıdır [35]. SVM algoritması, doğrusal olan ve doğrusal olmayan türlere sahiptir. Hiperdüzlemlerdeki en optimal noktayı bulabilmek için kernel adı verilen çekirdek yapıları kullanılır [36].

SVM algoritmasında, sınıflandırma problemleri için Destek Vektör Sınıflandırması(Support Vector Classification-SVC) yapısı kullanılır. SVC yapısı, kullanıcı tarafından belirlenen etiketler(label) yardımıyla sistemin doğru şekilde bir öğrenme performansı

göstermesini ve sistemden çıkan sonuçların doğruluğunu onaylamayı sağlayan güdümlü bir öğrenme mekanizmasıdır [37]. Bu çalışmada, SVC'nin polinomik, doğrusal ve sigmoid kernel parametreleri kullanılmıştır.

SVM algoritmasında, regresyon için Destek Vektör Bağlayıcısı(Support Vector Regressor-SVR) adı verilen bir yapı bulunmaktadır. SVR, regresyon hatalarını en aza indirmek için deneysel riskleri ölçer ve bunun için de bir maliyet fonksiyonu kullanır [38].

SVM algoritmasındaki NuSVR modeli, maliyet(cost), kernel katsayısı(gamma), önbellek boyutu(cache size) ve nu(v) parametrelerini kullanarak tahmin edilmesi istenen çıktıların bağımsız bir örnek çiftler kümesi üzerinde öngörülmesini sağlamaktadır [39]. SVM algoritmasının modellerinden bir diğeri olan EpsilonSVR modeli ise maliyet, kernel katsayısı, önbellek boyutu ve epsilon(eğitim kaybı işlevinde gerçek değerle öngörülen noktalar arasında bir epsilon mesafesi içinde hiçbir cezanın ilişkilendirilmediği epsilon tüzünü belirtmektedir.) parametrelerini kullanarak işlevsel uygunluk problemlerine çözüm getirmektedir [40].

### 3.4. Performans Metrikleri (Performance Indicators)

Tahminleme bitiminde, modellerimizin tahminleme başarısını ölçmek amacıyla kullanılan bazı değerlendirme kriterleri bulunmaktadır. Karmaşıklık matrisi (confusion matrix), kesinlik (precision), hassasiyet (recall), F1 skor (F1 score), kök ortalama karesel hata (RMSE), ortalama mutlak hata yüzdesi (MAPE), Matthews Korelasyon Katsayısı (MCC) ve doğruluk(accuracy) kavramları bu kısımda detaylandırılmaktadır.

Karmaşıklık matrisi, mevcut veriler ile tahminlenmiş verilerin örneklenmesinin kıyaslamasıyla üretilen ve modelin doğru sonuçlar ortaya çıkarıp çıkarmadığını dört metrik kullanarak ifade eden bir tablodur [41]. Bu dört metrik, doğru pozitif(True Positive-TP), doğru negatif(True Negative-TN), yalancı pozitif(False Positive-FP) ve yalancı negatif(False Negative-FN) biçiminde karmaşıklık matrisinde yer alır. Ayrıca, karmaşıklık matrisinin bu metrikleri, diğer performans ölçümlerine de yardımcı olur. Karmaşıklık matrisi, sahip olduğu dört metrikle birlikte gösterilmiştir (Tablo 2).

Tablo 2. Karmaşıklık matrisi(Confusion matrix)

Karmaşıklık Matrisi	Değerler		
	Tahminlenmiş Değerler	Asıl Değerler	
	Pozitif	TP	FP
	Negatif	TN	FN

Kesinlik ölçümü, tahminlenmiş verideki pozitif noktaları mevcut veri ile kıyaslayıp bunlardan kaçının gerçekten doğru tahmin edildiğini bulmayı sağlar [42]. Kesinlik ölçümünün formülü şu şekildedir:

$$Kesinlik \text{ Ölçümü} = \frac{TP}{TP+FP} \quad (4)$$

Hassasiyet ölçümü, mevcut verideki pozitif durumların başarı tahminini göstermek için kullanılır [43]. Hassasiyet ölçümü şu formülle gösterilir:

$$Hassasiyet \text{ Ölçümü} = \frac{TP}{TP+FN} \quad (5)$$

Kesinlik ve hassasiyet ölçümünün tek bir performans ölçümünde gösterilmesini sağlamanın en standart yolu olan F1 skor metriği, bu ölçümlerin aritmetik ortalamasının yerine harmonik ortalamasının alınması ile elde edilir [44]. F1 skor şu şekilde formüle edilir:

$$F1 \text{ Skor} = \frac{2 * Kesinlik * Hassasiyet}{Kesinlik + Hassasiyet} \quad (6)$$

MAPE, zaman serileri ve regresyon modelleri için yapılan tahminlerin doğruluğunu ölçmeye yarar [45]. Bu ölçüm şu şekilde formüle edilir:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (7)$$

$A_t$  ifadesi gerçek değeri,  $F_t$  ifadesi ise tahmini değeri belirtir. Formüldeki  $n$  ifadesi veri sayısını belirtirken,  $t$  ifadesi ise verinin indeksini gösterir.

RMSE, bir modelin tahmin ettiği değerler ile gerçek değerler arasındaki farkın, yani tahmindeki hataların standart sapmasını ölçer [46]. Bu ölçümün formülü şu şekildedir:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (8)$$

Yukarıdaki formülde,  $n$  ifadesi veri sayısını,  $t$  ifadesi veri indeksini belirtir. Formüldeki  $e$  ifadesi ise gerçek değer ile tahminlenmiş değer arasındaki farkı belirtir.

MCC, makine öğreniminde ikili sınıflandırmaların kalitesini ölçmek üzere kullanılır [47]. MCC, -1 ile +1 arasında bir değer döndürür. Değerin +1 olması mükemmel bir tahmin sonucunu simgeler. Değerin -1 olması tahmini değer ve gerçek değer arasındaki büyük farkı simgeler. Değerin 0 olması durumunda, tahminlenen değer rastgele bir tahminden daha iyi değildir. MCC, karmaşıklık matrisinin dört metriği kullanılarak şu şekilde hesaplanabilir:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \quad (9)$$

Doğruluk ölçümü, gözlemlenmiş verilere ait doğru tahmin oranını gösterir [48]. Doğru tahmin oranı bulunurken karmaşıklık matrisinin metriklerinden yararlanır. Doğruluk ölçümü şu formül ile gösterilir:

$$Doğruluk \text{ Ölçümü} = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

#### 4. UYGULAMA (IMPLEMENTATION)

Yukarıda da belirtildiği üzere, pandemi öncesi çalışma kapsamında BTC fiyatlarının düşüş ve yükseliş durumlarını tahminlemek üzere BTC/USD kurunun 5 Şubat 2018 ile 27 Ekim 2019 arasındaki haftalık kapanış fiyatları baz alınarak bir veri kümesi oluşturulmuştur. Pandemi sonrası çalışma kapsamında ise BTC/USD kurunun 5 Şubat 2018 ile 28 Haziran 2020 arasındaki haftalık kapanış fiyatları esas alınarak ikinci bir veri kümesi oluşturulmuştur. Bu iki veri kümesi üzerinde ARIMA, LSTM ve SVM modelleri uygulanmıştır. Modeller, 64 bit Windows 10 işletim sistemi ve Intel Core i5 4210U merkezi işlemci birimine sahip bir bilgisayarda Merkezi İşlemci Birimi (Central Process Unit-CPU) baz alınarak test edilmiştir. Çalışmada, Grafik İşleme Birimi (Graphics Processor Unit-GPU) kullanılmaması teknik yetersizlikten kaynaklanmaktadır. Kodlama dili olarak Python, dağıtım platformu olarak Anaconda3, çalışma alanı olarak Spyder kullanılmıştır. Makine öğrenmesi kısmında Sklearn kütüphanesi kullanılmıştır [49]. Bilimsel hesaplamalarda Numpy kütüphanesinden yararlanılmıştır [50]. Verilerin analizinin yapılabilmesi için Pandas kütüphanesi kullanılmıştır [51]. Derin öğrenme ve makine öğrenmesi işlemleri için Tensorflow kütüphanesine de başvurulmuştur [52]. İki veri kümesi de, eğitim ve test olarak bölünmüştür (split). Pandemi öncesi çalışma için kullanılacak veri kümesinin test kısmı 30 hafta olarak belirlenmiştir. Bu kapsamda, pandemi öncesi verilerin %66.6'lık kısmı eğitim için, geri kalan kısmı test için kullanılmıştır. Eğitilen 60 haftalık veriyle, kalan 30 hafta için fiyat eğilim tahminlemesi yapılmıştır. Eğitilen verinin büyük çoğunluğunun ayı piyasasına ait olduğu gözlemlenmiştir. Pandemi sonrası çalışma için kullanılacak verinin test kısmı 15 hafta olarak belirlenmiş olup oransal olarak %88'lik kısmı eğitim için kullanılmıştır. Toplamda 110 haftanın eğitilmesiyle birlikte kalan 15 hafta üzerinde fiyat eğilim tahminlemesi uygulanmıştır.

ARIMA modelinde, birçok farklı AR(p), I(d) ve MA(q) değerleri denenip tahminleme için en uygun model uygulanmaya çalışılmıştır. Veri kümesi üzerinde Genişletilmiş Dickey-Fuller Birim Kök Testi(ADF), Kwiatkowski-Phillips-Schmidt-Shin Birim Kök Testi(KPSS) ve Philips-Perron Birim Kök Testi(PP) uygulanarak önerilen parametreler dikkate alınmıştır. Ayrıca, otokorelasyon ve kısmi otokorelasyon grafikleri çıkarılıp analiz edilmiştir. Modelin güven düzeyi %95 olarak ayarlanmıştır. Modelin ürettiği sonuçlar dikkate alınarak, modelin başarısını ölçmek üzere kesinlik, hassasiyet ve F1 skor performans metrikleri kullanılmıştır. Buna ek olarak, modelin RMSE, MAPE ve MCC değerleri de ölçülmüştür.

LSTM modelini uygulamak amacıyla ardışık(sequential) algoritma kullanılmıştır. Farklı döngü sayıları(epoch), parti boyutları(batch size), katman sayıları ve seyreltme değerleri(dropout) gibi parametreler denenmiş ve en uygun

model bulunmaya çalışılmıştır. Veriler üzerinde Kurtosis(basıklık) ve Skewness(çarpıklık) testleri yapılarak Gaussian model dağılımı ortaya çıkartılmıştır. Dickey-Fuller testi yapılarak belirli dilimlerdeki kritik değerler ve p değeri tespit edilmiştir. Ayrıca, modelin eğitim ve test kayıpları incelenmek üzere grafikleştirilmiştir. Modelin normalize edilmesi için "Adam Optimizer" adı verilen bir normalizasyon türü kullanılmıştır. Tahminlenen sonuçlar, kesinlik, hassasiyet ve F1 skor ölçümlerine tabii tutulmuştur. En başarılı modeli oluşturan parametreler kaydedilmiştir. Bununla beraber, MCC ve doğruluk değerleri de ölçülmüştür.

SVM modelinde kullanılmak üzere SVC ve SVR algoritmaları, bu algoritmaların alt algoritmaları ile birlikte kullanılmıştır. Uygulanacak SVC modelleri kernel parametreleri bakımından doğrusal, sigmoid ve polinomik olarak ayrılmış ve bu üç alt modelin maliyet ve önbellek boyutu parametreleri değiştirilip denenerek alt algoritmalar için en uygun parametreler belirlenmiştir. SVR için belirlenmiş alt algoritmalar olan NuSVR uygulanırken maliyet, nu(eğitim hatalarının oranının üst sınırı ve destek vektörlerinin alt sınırı) değeri, önbellek boyutu ve derece parametreleri en iyi modeli bulmak için sürekli bir şekilde denenmiştir. NuSVR'i uygulamak için radyal bazlı fonksiyon(rbf) kerneli kullanılmıştır. EpsilonSVR algoritmasında, doğrusal bir kernel modeli uygulanmıştır. Söz konusu algoritmaya ait epsilon, maliyet, önbellek boyutu ve derece parametreleri denenip en uygun model ortaya çıkartılmıştır. Verilerin ölçeklendirmesi gerçekleştirilmek üzere standart bir ölçeklendirme kullanılmıştır. Modellerin başarısı kesinlik, hassasiyet, F1 skor, doğruluk ve MCC değerleriyle ölçülmüştür.

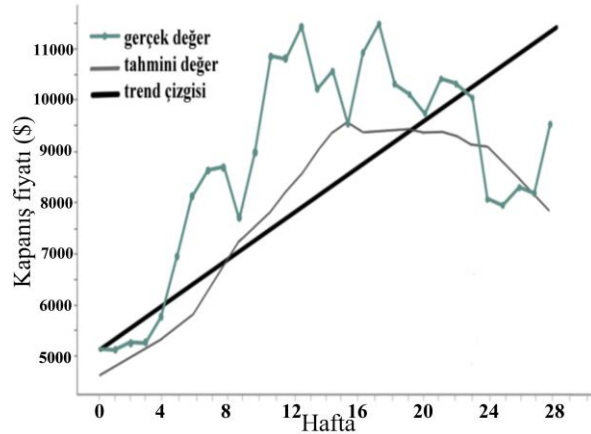
## 5. BULGULAR (FINDINGS)

Uygulanan modellerin performansını ölçmek için karmaşıklık matrisleri oluşturulmuştur. Algoritmalara ait kesinlik, hassasiyet, F1 skor, doğruluk ve MCC değerleri ölçülmüştür. Diğer algoritmalarından ayrı olarak ARIMA algoritması için RMSE ve MAPE değerleri de ölçülmüştür.

### 5.1. Koronavirüs Pandemisi Öncesi Çalışma İçin Bulgular (Findings for the study before coronavirus pandemic)

LSTM modeli üzerinde en başarılı sonucu almak için birçok parametre denenerek oluşturulan modeller test edilmiştir. Katman sayısındaki değişimin tahminleme sonucunu büyük oranda etkilediği görülmüştür. Döngü sayısı arttırıldığında ise belirli bir seviyeden sonra tahminleme performansının düştüğü gözlenmiştir. Bu gözlem, optimal bir tahminleme sonucu için döngünün düşük seviyelerde tutulması gerektiğini göstermiştir. Parti boyutunun arttırılmasının veya azaltılmasının tahminleme sonucunda kayda değer seviyede değişimlere yol açmadığı görülmüştür. Tüm gözlemler sonucunda, modelin başarı performansını en yüksek seviyeye çıkaran parametreler katman sayısında 5, döngü sayısında 20 ve parti boyutunda 64 olarak tespit edilmiştir. Yapılan ölçümler sonrası, BTC fiyatındaki düşüşlerde %67, yükselişlerde ise %73 F1 skor değerine ulaşılmıştır. Doğruluk değeri %70 olarak

gözlenmiştir. MCC değeri 0,4344 bulunmuştur. Tahminleme sonucunda ortaya çıkan değerlerle mevcut değerlerin grafikleştirilmiş hali şekilde (Şekil 1) gösterilmiştir.



Şekil 1. LSTM modelinin tahmin sonuçlarının karşılaştırılması (Comparing of forecasting results of LSTM model)

LSTM için uygulanan modellere ait hiperparametreler sonuçlarıyla birlikte aşağıdaki tabloda gösterilmiştir (Tablo 3). Yükseliş değeri için Y, düşüş değeri için D ifadesi kullanılmıştır.

Tablo 3. Uygulanmış hiperparametreler ve sonuçları (Applied hyperparameters and their results)

Parti boyutu	Katman sayısı	Döngü sayısı	Seyreltme Değeri	Doğruluk (%)	Trend	Kesinlik (%)	Hassasiyet (%)	F1 Skor (%)	MCC (0-1)
64	5	20	0,1	70	Y	63	86	73	0,4344
					D	82	56	67	
128	2	40	0,1	60	Y	58	50	54	0,19094
					D	61	69	65	
128	4	150	0,1	60	Y	56	71	63	0,2182
					D	67	50	57	
128	100	150	0,5	60	Y	55	79	65	0,23623
					D	70	44	54	
256	10	75	0,1	67	Y	61	79	69	0,3546
					D	75	56	64	
512	15	10	0,1	67	Y	61	79	69	0,3546
					D	75	56	64	
512	2	16	0,1	67	Y	62	71	67	0,3393
					D	71	62	67	
512	2	20	0,1	67	Y	62	71	67	0,3393
					D	71	62	67	
512	4	23	0,1	63	Y	60	64	62	0,26726
					D	67	62	65	

ARIMA modeli için önerilen AR(p), I(d) ve MA(q) parametrelerini görmek ve durağanlığı test etmek için ADF, KPSS ve PP testlerinden yararlanılmıştır. Önerilen parametrelerden farklı olarak birkaç parametre denemesi daha yapılmıştır. Korelasyon ve kısmi korelasyon analizlerinde değişkenler arasındaki ilişkinin çok zayıf olduğu görülmüştür. Parametre denemelerinin sonucunda bu üç parametrenin tahminlemeye etkisi üzerinde bazı sonuçlara ulaşılmıştır. Verinin doğrusal bir yayılıma sahip

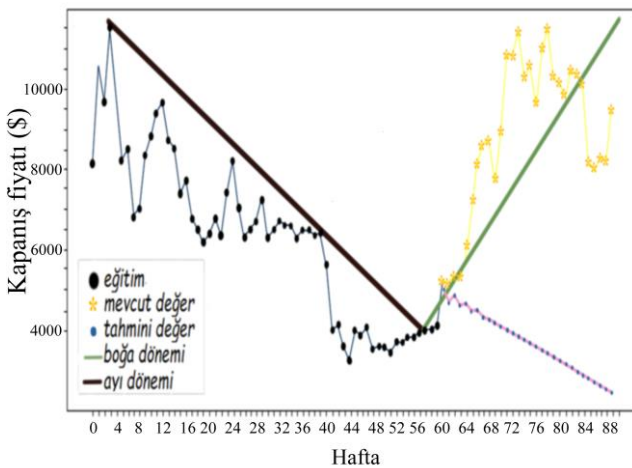


olmaması nedeniyle, I(d) parametresinin değeri değişikçe tahminlemede büyük değişimlere yol açtığı görülmüştür. Aynı sebepten dolayı, AR(p) ve MA(q) parametrelerinin de değişikçe tahminlemenin doğruluk oranına aşırı bir etkide bulunduğu tespit edilmiştir. Tüm bu gözlemler sonucunda oluşturulan en başarılı model, AR(p) parametresi için 2, I(d) parametresi için 1 ve MA(q) parametresi için 1 olmak üzere ARIMA(2,1,1) modelidir. Modelin F1 skoru, BTC fiyatlarındaki düşüşlerde %74, yükselişlerde %35 değerinde görülmektedir. MAPE değeri %78,57 ve RMSE değeri 0,6055 bulunmuştur. Doğruluk oranı %63 ve MCC değeri 0,35635 bulunmuştur. ARIMA modelinde %95 olarak belirlediğimiz güven düzeyine göre, sabit değer hariç AR ve MA parametrelerinin kabul edilebilir P değerlerine sahip olduğu tespit edilmiştir. ARIMA(2,1,1) modelindeki parametrelerin katsayıları, standart hataları, Z değeri ve P değeri aşağıdaki tabloda (Tablo 4) bulunmaktadır.

Tablo 4. ARIMA (2,1,1) model sonuçları  
(Results of ARIMA(2,1,1) model)

Parametre	Değerler			
	Katsayı	Standart Hata	Z	P> z
const	-8671,3616	8370,659	-1,036	0,305
ar.L1	-0,9995	0,143	-6,997	0,000
ar.L2	-0,2445	0,142	-1,723	0,090
ma.L1	1,0000	0,047	21,475	0,000

ARIMA modeliyle tahminlenmiş değerler ve mevcut değerler karşılaştırılmıştır. Düşüş tahminlerinde biraz da olsa başarılı olduğu gözlenmiştir. Fakat yükseliş tahminlerinde, modelin çok fazla bir başarı kaydedemediği görülmüştür. Yükseliş tahminlerinde başarı kaydedilememesinin nedeni, algoritmanın BTC fiyatlarının boğa piyasasına girdiği dönemdeki trend çizgilerini takip edemeyişinden kaynaklanmaktadır. Bu karşılaştırma aşağıdaki şekilde (Şekil 2) gösterilmiştir.



Şekil 2. ARIMA modelinin tahmin sonuçlarının karşılaştırılması (Comparing of forecasting results of ARIMA model)

Uygulanan tüm ARIMA modelleri ve sonuçları aşağıdaki tabloda verilmiştir (Tablo 5). Yükseliş değeri için Y, düşüş değeri için D ifadesi kullanılmıştır.

Tablo 5. Uygulanmış ARIMA modelleri ve sonuçları  
(Applied ARIMA models and their results)

Model	Trend	Kesimlik (%)	Hassasiyet (%)	F1 Skor (%)	MCC (0-1)	RMSE (0-∞)	MAPE (%)
ARIMA(2,1,1)	Y	100	21	35	0,35635	0,6055	78,57
	D	59	100	74			
ARIMA(2,1,2)	Y	100	7	13	0,1985	0,6582	92,85
	D	55	100	71			
ARIMA(4,1,1)	Y	50	14	22	0,026207	0,6831	94,32
	D	54	88	67			
ARIMA(8,1,1)	Y	50	21	30	0,03341	0,6831	94,32
	D	54	81	65			
ARIMA(4,1,2)	Y	75	21	33	0,22276	0,6324	87,53
	D	58	94	71			
ARIMA(2,1,4)	Y	100	7	13	0,1985	0,6582	92,85
	D	55	100	71			
ARIMA(3,1,1)	Y	80	29	42	0,2988	0,6055	78,57
	D	60	94	73			
ARIMA(4,2,3)	Y	71	36	48	0,2738	0,6055	78,57
	D	61	88	72			

SVM modelinde tahminleme uygulamak üzere SVC algoritmalarının kernel parametreleri doğrusal, çok terimli ve sigmoid olarak ayarlanmıştır. Bu üç alt algoritmanın her biri için farklı maliyet parametreleri denenmiştir. Böylece, tahminleme için en uygun model oluşturulmaya çalışılmıştır. Denemeler esnasında, maliyet parametresinin belirli aralıklarda tahminlemenin seyrini değiştirmedeği görülmüştür. Fakat bu aralıkların dışında verilen değerlerin çok büyük değişimlere yol açtığı gözlenmiştir. SVR algoritmalarında ise bir alt algoritma olarak bilinen NuSVR için maliyet, gamma ve nu parametreleri farklı değerler verilerek denemeye tabii tutulmuştur. Gamma değerlerindeki değişimin aşırı olmadığı tahminlemeye pek etkisi olmadığı gözlenmiştir. Buna rağmen maliyet parametreleri ve diğer parametreler arasındaki korelasyon sebebiyle, parametre değerleri değişikçe gerek pozitif yönde gerek negatif yönde tahminlemeye etki ettiği tespit edilmiştir. Bir diğer SVR alt algoritması olan EpsilonSVR için maliyet, gamma ve epsilon parametreleri birçok kez değiştirilerek uygun modele ulaşılmaya çalışılmıştır. Bu değişimler esnasında, epsilon değerinin artırılması durumunda, tahminlemenin başarılı olması için daha fazla vektöre ihtiyaç duyulduğu gözlenmiştir. Maliyet parametresinin epsilon parametresinin düşüklüğüne bağlı olarak tahminlemenin yönüne pozitif etki ettiği tespit edilmiştir. Buradan yola çıkarak bu iki parametre arasında negatif bir korelasyon bulunduğunu söyleyebiliriz.

LinearSVC modeli için uygun bulunan modelde maliyet parametresi 20, önbellek boyutu 100 olarak bulunmuştur. Düşüş tahminlerinde %86, yükseliş tahminlerinde ise %78

F1 skora ulaşılmıştır. MCC değeri 0,6999 ve doğruluk değeri %83'tür. SigmoidSVC modelinde ise en başarılı model, maliyet parametresi olarak 5 değerine, önbellek boyutu olarak 100 değerine sahip model olmuştur. Düşüş tahminleri %83, yükseliş tahminleri %75 F1 skor değerinde gözlenmektedir. MCC değeri 0,6142 ve doğruluk değeri %80 olarak bulunmuştur. PolynomialSVC modelinde maliyet parametresi 100, üçüncü dereceden, 100 önbellek boyutuna sahip model test edilmiştir. Düşüş tahmini %86, yükseliş tahmini ise %78 F1 skora sahiptir. MCC değeri 0,6999 ve doğruluk değeri %83'tür. SVR alt modellerine baktığımızda, NuSVR algoritması için maliyet parametresi 400, önbellek boyutu 10, nu değeri 0,1 olan, aralıklı gammaya sahip model kullanılmıştır. Bu modelin düşüş tahminleri %67, yükseliş tahminleri %78 F1 skor değerinde gözlemlenmiştir. Modelin MCC değeri 0,5641 ve doğruluk değeri %73'tür. Bir diğer SVR alt modeli olan EpsilonSVR algoritmasında ise maliyet değeri 5, önbellek boyutu 100, üçüncü dereceden, epsilon değeri 0,2 olan model kullanılmıştır. Düşüş tahminleri %86, yükseliş tahminleri %78 F1 skora sahiptir. MCC değeri 0,6999 ve doğruluk değeri %83'tür. SVM algoritmasının en başarılı modelinden çıkan sonuca göre oluşan düşüş ve yükseliş tahminleri, mevcut değerlerle karşılaştırılmış haliyle aşağıdaki tabloda (Tablo 6) verilmiştir.

Tablo 6. SVM modelinin sonuçlarının karşılaştırılması  
(Comparing of results of SVM model)

Hafta	Asıl Trend	Tahmin	Kapanış
1	Yükseliş	Yükseliş	5198,90
2	Düşüş	Düşüş	5167,72
3	Yükseliş	Yükseliş	5314,53
4	Düşüş	Düşüş	5285,14
5	Yükseliş	Yükseliş	5795,71
6	Yükseliş	Düşüş	6972,37
7	Yükseliş	Düşüş	8197,69
8	Yükseliş	Düşüş	8673,22
9	Düşüş	Düşüş	8742,96
10	Düşüş	Düşüş	7688,08
11	Yükseliş	Yükseliş	8994,49
12	Yükseliş	Düşüş	10855,37
13	Düşüş	Düşüş	10817,16
14	Yükseliş	Yükseliş	11450,85
15	Düşüş	Düşüş	10256,06
16	Düşüş	Düşüş	10599,11
17	Düşüş	Düşüş	9552,86
18	Yükseliş	Yükseliş	10970,18
19	Düşüş	Düşüş	11523,58
20	Düşüş	Düşüş	10345,81
21	Düşüş	Düşüş	10138,52
22	Düşüş	Düşüş	9757,97
23	Yükseliş	Yükseliş	10441,28
24	Yükseliş	Düşüş	10347,71
25	Düşüş	Düşüş	10070,39
26	Düşüş	Düşüş	8104,19
27	Düşüş	Düşüş	7988,16
28	Yükseliş	Yükseliş	8321,01
29	Düşüş	Düşüş	8222,08
30	Yükseliş	Yükseliş	9551,71

Tahminlemede kullanılan modellerin başarısını ölçmek için kullanılan performans metrikleri olan ham F1 skor, kesinlik, hassasiyet, MCC, doğruluk ve kümülatif hale getirilen F1 skor değerleri tabloda (Tablo 7) verilmiştir. Yükseliş değeri için Y, düşüş değeri için D ifadesi kullanılmıştır.

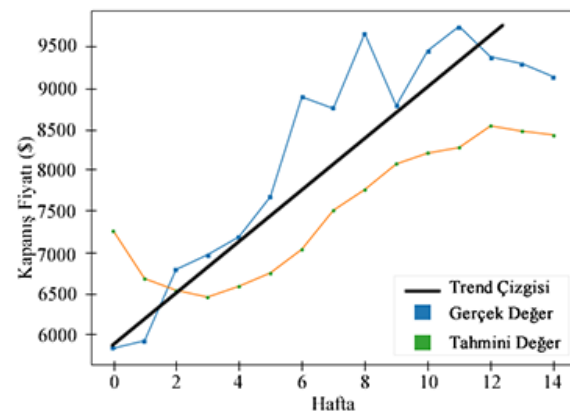
Tablo 7. Tüm modellere ait performans ölçüm sonuçları  
(Results of performance measurement concerning to the all models)

Model	Trend	F1 Skor (%)	Kesinlik (%)	Hassasiyet (%)	Küm. F1 Skor (%)	MCC (0-1)	Doğruluk (%)
LSTM (5,20,64)	Y	73	63	86	70	0,4344	70
	D	67	82	56			
ARIMA (2,1,1)	Y	35	100	21	55	0,35635	63
	D	74	59	100			
SVM Linear SVC	Y	78	100	64	82	0,6999	83
	D	86	76	100			
SVM Sigmoid SVC	Y	75	90	64	79	0,6142	80
	D	83	75	94			
SVM Poly-SVC	Y	78	100	64	82	0,6999	83
	D	86	76	100			
SVM NuSVR	Y	78	64	100	73	0,5641	73
	D	67	100	50			
SVM EpsilonSVR	Y	78	100	64	82	0,6999	83
	D	86	76	100			

SVM algoritmaları ortalama olarak 0,83 saniyede, LSTM algoritması 11,36 saniyede, ARIMA yöntemi ise 5,30 saniyede tahminleme işlemlerini gerçekleştirmişlerdir.

## 5.2. Koronavirüs Pandemisi Sonrası Çalışma İçin Bulgular (Findings for the study after coronavirus pandemic)

Koronavirüs pandemisi öncesini kapsayan çalışmaya ek olarak koronavirüs pandemisi sonrası için ayrı bir veri kümesi üzerinde aynı yöntemler uygulanmıştır. LSTM için en iyi performansa sahip model, katman sayısı 10, döngü sayısı 75 ve parti boyutu 256 olarak belirlenmiş olup modelin grafiği şekilde (Şekil 3) gösterilmiştir.



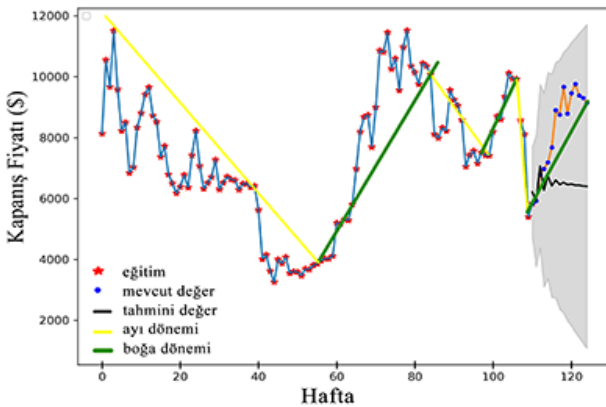
Şekil 3. LSTM modelinin tahmin sonuçlarının karşılaştırılması (Comparing of forecasting results of LSTM model)

Modele ait F1 skor değerleri, düşüş tahminleri için %71, yükseliş tahminleri için %75 olarak bulunmuştur. Modelin MCC değeri 0,4910 olarak hesaplanmıştır. Buna ek olarak, LSTM için uygulanan diğer modellere ait hiperparametreler sonuçlarıyla birlikte aşağıdaki tabloda gösterilmiştir (Tablo 8). Yükseliş değeri için Y, düşüş değeri için D ifadesi kullanılmıştır.

Tablo 8. Uygulanmış hiperparametreler ve sonuçları (Applied hyperparameters and their results)

Parti boyutu	Katman sayısı	Döngü sayısı	Seyretme Değeri	Doğruluk (%)	Trend	Keskinlik (%)	Hassasiyet (%)	F1 Skor (%)	MCC (0-1)
64	5	20	0,1	53	Y	50	57	53	0,0714
					D	57	50	53	
128	2	40	0,1	66	Y	67	57	62	0,3273
					D	67	75	71	
128	4	150	0,1	60	Y	50	71	59	0,0945
					D	60	38	46	
128	100	150	0,5	60	Y	50	71	59	0,0945
					D	60	38	46	
256	10	75	0,1	67	Y	67	86	75	0,4910
					D	83	62	71	
512	15	10	0,1	67	Y	55	86	67	0,2619
					D	75	38	50	
512	2	16	0,1	67	Y	62	71	67	0,3393
					D	71	62	67	
512	2	20	0,1	67	Y	43	43	43	0,0714
					D	50	50	50	
512	4	23	0,1	63	Y	56	71	63	0,2182
					D	67	50	57	

ARIMA modeli için denenen parametreler sonucunda oluşturulan en başarılı model, ARIMA(2,1,4) olmuştur. ARIMA(2,1,4) modeline ait tahminlenmiş ve mevcut değerlerin karşılaştırılması aşağıdaki şekilde (Şekil 4) gösterilmiştir.



Şekil 4. ARIMA modelinin tahmin sonuçlarının karşılaştırılması (Comparing of forecasting results of ARIMA model)

ARIMA(2,1,4) modelinin F1 skoru, BTC fiyatlarındaki düşüşlerde %94, yükselişlerde %92 değerinde seyretmiştir. Modelin MAPE değeri %14,28 ve RMSE değeri 0,2582

olarak bulunmuştur. Modelin doğruluk oranı %93 ve MCC değeri 0,8729 olarak bulunmuştur. ARIMA(2,1,4) modelinin, koronavirüs pandemisi sonrası çalışmayı kapsayan veriler üzerinde başarılı bir performans gösterdiği görülmüştür.

ARIMA modelleri uygulanan tüm parametreleri ve sonuçları ile aşağıdaki tabloda verilmiştir (Tablo 9). Yükseliş değeri için Y, düşüş değeri için D ifadesi kullanılmıştır.

Tablo 9. Uygulanmış ARIMA modelleri ve sonuçları (Applied ARIMA models and their results)

Model	Trend	Keskinlik (%)	Hassasiyet (%)	F1 Skor (%)	MCC (0-1)	RMSE (0-∞)	MAPE (%)
ARIMA(2,1,1)	Y	75	86	80	0,6071	0,4472	29,32
	D	86	75	80			
ARIMA(2,1,2)	Y	75	43	55	0,3425	0,5773	62,93
	D	64	88	74			
ARIMA(4,1,1)	Y	38	43	40	0,1964	0,7746	95,81
	D	43	38	40			
ARIMA(8,1,1)	Y	71	71	71	0,4643	0,5164	49,47
	D	75	75	75			
ARIMA(4,1,2)	Y	57	57	57	0,1964	0,6325	87,59
	D	62	62	62			
ARIMA(2,1,4)	Y	100	86	92	0,8729	0,2582	14,28
	D	89	100	94			
ARIMA(3,1,1)	Y	71	71	71	0,4643	0,5164	49,47
	D	75	75	75			
ARIMA(4,2,3)	Y	62	71	67	0,3393	0,5773	62,93
	D	71	62	67			

SVM algoritmasını alt modelleriyle birlikte koronavirüs pandemisi sonrası çalışma verilerine uygularken, koronavirüs pandemisi öncesi çalışma verilerine uygulanan modeller kullanılmıştır.

LinearSVC, SigmoidSVC ve PolynomialSVC modelleri için düşüş tahminlerinde %89, yükseliş tahminlerinde ise %83 F1 skora ulaşılmıştır. Bu üç modelin MCC değerleri 0,7559 ve doğruluk değerleri %87 olarak bulunmuştur.

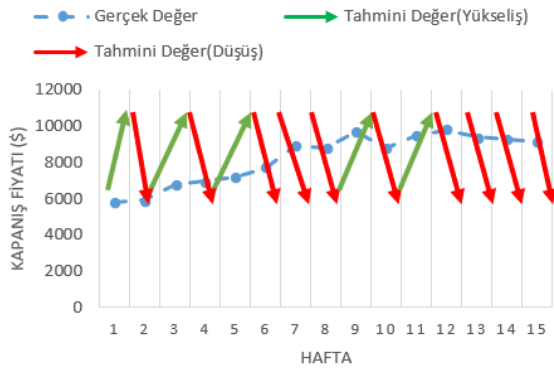
SVR alt modellerinden biri olan NuSVR algoritması, düşüş ve yükseliş tahminlerinde %67 F1 skor değerinde gözlemlenmiştir. NuSVR algoritmasına ait MCC değeri 0,3393 ve doğruluk değeri %67 olarak bulunmuştur. Bir diğer SVR alt modeli olan EpsilonSVR algoritmasında ise düşüş tahminleri %77, yükseliş tahminleri %82 F1 skora sahip olmuştur. Modele ait MCC değeri 0,6614 ve doğruluk değeri %80 olarak bulunmuştur.

SVM algoritmasının en başarılı modeline ait olan düşüş ve yükseliş tahminleri, mevcut değerlerle karşılaştırılmış haliyle tabloda (Tablo 10) verilmiştir.

Tablo 10. SVM modelinin sonuçlarının karşılaştırılması  
(Comparing of results of SVM model)

Hafta	Asıl Trend	Tahmin	Kapanış
1	Yükseliş	Yükseliş	5830,25
2	Düşüş	Düşüş	5922,04
3	Yükseliş	Yükseliş	6791,13
4	Düşüş	Düşüş	6971,09
5	Yükseliş	Yükseliş	7189,42
6	Yükseliş	Düşüş	7679,87
7	Yükseliş	Düşüş	8897,47
8	Düşüş	Düşüş	8756,43
9	Yükseliş	Yükseliş	9670,74
10	Düşüş	Düşüş	8790,37
11	Yükseliş	Yükseliş	9461,06
12	Düşüş	Düşüş	9758,85
13	Düşüş	Düşüş	9386,79
14	Düşüş	Düşüş	9303,63
15	Düşüş	Düşüş	9143,58

En başarılı SVM modeline ait tahminlenmiş ve mevcut değerlerin karşılaştırılması aşağıdaki şekilde (Şekil 5) gösterilmiştir.



Şekil 5. SVM modelinin tahmin sonuçlarının karşılaştırılması (Comparing of forecasting results of SVM model)

Koronavirüs pandemisi sonrası verileri kapsayan çalışmada, tahminleme için kullanılan ana modellerin ve bu ana modellerin alt modellerinin başarısını ölçmek için bazı performans metrikleri kullanılmıştır. Tüm modeller arasında F1 skoru ve doğruluk oranı en yüksek olan model %93 ile ARIMA(2,1,4) modeli olmuştur.

Çalışmada kullanılan bütün algoritmaların çalışma süreleri belirlenmiştir. Belirlenen çalışma sürelerine göre; SVM algoritmaları ortalama olarak 1,72 saniyede, LSTM algoritması 14,86 saniyede, ARIMA yöntemi ise 7,63 saniyede tahminleme işlemlerini gerçekleştirmiştir.

Tahminlemede kullanılan modellere ait ham F1 skorları, kesinlik değerleri, hassasiyet değerleri, MCC değeri, doğruluk değerleri ve kümülatif hale getirilen F1 skorları derlenerek tabloda (Tablo 11) verilmiştir. Tabloda, yükseliş değerlerini ifade etmek amacıyla Y harfi, düşüş değerlerini ifade etmek amacıyla D harfi kullanılmıştır.

Tablo 11. Tüm modellere ait performans ölçüm sonuçları  
(Results of performance measurement concerning to the all models)

Model	Trend	F1 Skor (%)	Kesinlik (%)	Hassasiyet (%)	Küm. F1 Skor (%)	MCC (0-1)	Doğruluk (%)
LSTM (10,75,256)	Y	75	67	86	73	0,4910	73
	D	71	83	62			
ARIMA (2,1,4)	Y	92	100	86	93	0,8729	93
	D	94	89	100			
SVM Linear SVC	Y	83	100	71	86	0,7559	87
	D	89	80	100			
SVM Sigmoid SVC	Y	83	100	71	86	0,7559	87
	D	89	80	100			
SVM Poly-SVC	Y	83	100	71	86	0,7559	87
	D	89	80	100			
SVM NuSVR	Y	67	62	71	67	0,3393	67
	D	67	71	62			
SVM EpsilonSVR	Y	82	70	100	80	0,6614	80
	D	77	100	62			

## 6. SONUÇ (RESULT)

BTC fiyatlarındaki düşüş ve yükseliş eğilimlerini belirleme amacıyla haftalık BTC kapanış fiyatları ve haftalık düşüş-yükseliş etiketleriyle oluşturduğumuz pandemi öncesini kapsayan ve pandemi sonrası kapsayan iki ayrı veri kümesi üzerinde ARIMA, LSTM ve SVM modelleri test edilip değerlendirilmiştir.

Sınıflandırma ve tahminleme modellerinin bu veriler üzerine uygulanmasının ardından bu modellerin başarıları ölçülmüştür. Tüm modellere ait kesinlik, hassasiyet, F1 skor, doğruluk ve MCC değerleri ölçülmüştür. Ek olarak, ARIMA modeli için RMSE ve MAPE ölçümleri de yapılmıştır. Böylece, en başarılı modeller göz önünde bulundurulmuştur. Bu performans metriklerini tercih etmemizin en önemli nedeni, diğer performans metrikleri ile karşılaştırıldıklarında tahminlemeye dair daha kesin ve net sonuçlar ortaya çıkarmasıdır.

Pandemi öncesini kapsayan veriler için en yüksek başarıya sahip modeller, SVM algoritmasına dayanan LinearSVC, PolynomialSVC ve EpsilonSVR modelleri olmuştur. Bu modeller, %82 kümülatif F1 skoru, %83 doğruluk ve 0,6999 MCC değeri ile tüm metrikler için en yüksek sonuca ulaşmışlardır. SVM algoritmalarının bu başarısı, sahip olduğu kernel parametrelerinin verilerdeki hata paylarını en aza indirmek üzere tasarlanmasıyla açıklanabilir. Ayrıca, SVM algoritmalarının Hitam ve Ismail'in(2018) yaptığı çalışmada da belirtildiği gibi ikili sınıflandırma problemleri için geliştirilmiş olması, sınıflandırıcılarının tahminleyici niteliğinin yüksek olması ve değiştirilebilir parametrelerin diğer algoritmaların parametrelerinden daha etkin rol oynaması da bu başarının nedeni olarak gösterilebilir. Buna ek olarak, eğitimde kullanılan verinin sayı ve kalite bakımından tahminlemeye uygun olması bir sebep olarak gösterilebilir. SVM algoritması, BTC fiyatlarının düşüş ve yükseliş

eğilimlerini belirleme açısından başarılı bir yöntem olarak gösterilebilir. Sonuç olarak, genelde literatürde sınıflandırma açısından başarılı bulunan SVM algoritmasının BTC fiyatlarının düşüş ve yükselişlerini tahmin etmede de oldukça başarılı olabileceği, bu çalışmadaki uygulama ile ortaya konulmaktadır.

Pandemi öncesini kapsayan verikümesi üzerinde uygulanan ARIMA(2,1,1) modeli, tahminleme açısından diğer modellerin gerisinde kalmıştır. MAPE değeri %78,57, RMSE değeri 0,6055 ve MCC değeri ise 0,35635 olarak bulunmuştur. Hata oranlarına ek olarak, %55 kümülatif F1 skoru ve %63 doğruluk oranı ile tahminlemedeki başarısızlığı gözlemlenmiştir. ARIMA modelindeki düşüş ve yükseliş tahminlerinin doğruluk bakımından birbirine yakın olmaması, parametrelerin sözkonusu veriler üzerinde Olvera-Juarez ve Huerta-Manzanilla(2019) tarafından vurgulanan durağanlık koşulunu sağlama yönünden yetersiz kalmasıyla açıklanabilir. BTC verilerinin yüksek değişkenliğe sahip olması, ARIMA ile yapılan zaman serisi tahminlemelerini zorlaştırmaktadır. Uyguladığımız ARIMA(2,1,1) modeli, belirli bir aralıktaki veriler için doğru tahminleme sonuçları vermesine rağmen, bazı aralıklar için kayda değer bir başarıya ulaşamamıştır. Buna rağmen, Hitam ve Ismail'in(2018) yaptığı çalışmayı destekler nitelikte, veri sayısının artırılması ve durağanlığın sağlanması durumunda ARIMA modelinin BTC fiyatlarındaki düşüş ve yükselişleri tahmin etme konusundaki başarı performansının biraz artabileceği düşünülmektedir. Abu Bakar ve Rosbi(2017), ARIMA yönteminin güvenilir sonuçlar verdiğini öne sürmesine rağmen, bu çalışma sonucunda BTC gibi yüksek değişkenliğe sahip bir kriptoparanın fiyatlarını tahmin etme konusunda uygun ve güvenilir bir yöntem olmadığı görülmüştür.

LSTM algoritmasıyla pandemi öncesini kapsayan veri kümesi üzerinde yapılan tahminlemede BTC fiyatlarındaki düşüş ve yükselişe dair model başarısı %70 kümülatif F1 skor değerine ulaşmıştır. MCC değeri 0,4344 ve doğruluk oranı %70'tir. Başarım olarak SVM algoritmalarından düşük, ARIMA algoritmasından yüksek sonuçlara sahip olmuştur. Bu model başarısının sebebi olarak, LSTM modelindeki katman ve döngü parametreleri gösterilebilir. Verilerin katmanlar kullanılarak tahminleme için en uygun hale getirilmesi ve sözkonusu verilerin döngü parametresi kullanılarak en iyi şekilde eğitilebileceği seviyenin belirlenmesi, tahminlemenin performansını olumlu yönde etkilemiştir. Hız bakımından ise yukarıda belirtildiği üzere en uzun eğitim süresine sahip model olmuştur. Model, özellikle fiyat yükselişlerini tahminlemede iyi sonuçlar vermesine rağmen trend değişimleri karşısında yetersiz kalmıştır. Bu yetersizlik, McNally'nin(2018) belirttiği BTC fiyatlarındaki yüksek varyansın tahminleme çalışmalarını zorlaştırdığı ifadesini de destekler niteliktedir. Uygulanan model sonucunda, LSTM algoritmasının BTC fiyatlarındaki düşüş ve yükseliş eğilimlerini tahminleme konusundaki performansı, SVM modellerinin performansının altında kalmıştır.

Pandemi öncesini kapsayan veriler için düşüş ve yükseliş dönemlerine ait keskin yükseliş değerlerinin yeterince başarılı bir şekilde tahmin edilememesinin nedeni, fiyat manipülasyonları olarak gösterilebilir. Ayrıca, algoritmaların bu manipülasyonlar dolayısıyla trend çizgilerini tahmin edemeyişi de bir neden olarak gösterilebilir. Kullanılan yöntemler arasında en yüksek başarı performansına sahip modeller SVM algoritmasına aittir. Çalışma sonucunda SVM'nin BTC fiyatlarının eğilimlerini belirleme konusunda oldukça kullanışlı modellere sahip olduğu tespit edilmiştir.

Koronavirüs pandemisi sonrasında kapsayan veri kümesinin üzerinde uygulanan modellerin sonuçlarına göre ise en başarılı yöntem, %93 kümülatif F1 skora sahip olan ARIMA (2,1,4) olarak göze çarpmaktadır. Daha önce, Hitam ve Ismail(2018) tarafından belirtilen ifadelerle göre veri sayısının artırılması durumunda ARIMA'nın başarı performansının artacağı düşüncesi ifade edilmişti. Pandemi öncesi verilerde kullanılmış toplam 90 haftalık veriye kıyasen pandemi sonrası veriler için 125 haftalık kullanılması, ARIMA'nın BTC fiyatları üzerindeki başarı performansına büyük ölçüde olumlu olarak etki etmiştir. Pandemi öncesini kapsayan veriler üzerinde %82 kümülatif F1 skora sahip olan LinearSVC, PolynomialSVC ve SigmoidSVC modellerinin kümülatif F1 skorları %86'ya yükselmiştir. Bu performans artışında, modelin eğitilmesi için kullanılan verinin daha fazla olması, SVM algoritmalarının sınıflandırıcılarının tahminleyici niteliğinin yüksek olması ve modelin sahip olduğu değiştirilebilir parametrelerinin de etkisi olduğu düşünülmektedir. SVM algoritmaları, iki farklı veri kümesinde de yüksek başarı oranlarına sahip olarak BTC değerini tahminlemede kullanılabileceğini tekrar göstermiştir. Pandemi öncesini kapsayan veri kümesi üzerinde uygulanan tahminlemede kümülatif F1 skoru %70 olan LSTM algoritması, pandemi sonrasında kapsayan veri kümesi üzerinde %73 kümülatif F1 skor değerine ulaşmıştır. LSTM modeliyle tahminleme yapılırken iki ayrı veri kümesinde de aynı parametreler denendiği için, bu model başarımının artışında veri sayısının artırılmasının etkili olduğu düşünülmektedir.

Pandemi öncesini kapsayan veriler üzerinde uygulanan çalışmanın sonuçlarına göre, konu ile ilgili uzman ve yatırımcılara SVM algoritmasını BTC değerini tahminlemede kullanmaları tavsiye edilebilir. Pandemi sonrasında kapsayan veriler üzerinde uygulanan çalışmanın sonuçlarına bakıldığında, SVM algoritmasının BTC değerini tahminlemede kullanılabileceği bir kez daha söylenebilir. Ayrıca, ARIMA (2,1,4) modelinin pandemi sonrasında kapsayan veriler üzerindeki başarı oranı, veri sayısının artırılması durumunda BTC değerini tahminlemede ARIMA modeline de başvurulabileceğini göstermiştir.

Bu çalışmayı literatürde sunulan yayınlardan ayıran en önemli nokta, modellerin performansını ölçmek için birçok değerlendirme kriterinin baz alınmasıdır. Ayrıca, pandemi öncesi ve sonrasında kapsayan veriler üzerindeki çalışmalarda sunulan SVM modellerinin performansı,

literatürde sunulan bütün SVM modellerinden yüksek olmuştur. Buna ek olarak, diğer çalışmalarda SVM algoritmasının alt algoritmalarına başvurulmamışken, bu çalışmada alt algoritmalar da denenmiştir. Pandemi öncesini kapsayan çalışmada SVM modelinin doğruluk oranı %83 olurken, Graves ve Au(2015) BTC fiyatlarının eğilimi konusunda yaptıkları çalışmada SVM yöntemiyle %53,7 doğruluk oranında kalmışlardır. Hitam ve Ismail(2018), SVM algoritmasını kullandıkları çalışmada sadece MAPE değerini gözlemlemişlerdir. Yapmış olduğumuz çalışmalarda ise model performansları birçok metrik kullanılarak ölçülmüştür. Ayrıca, pandemi sonrası veriler üzerinde gerçekleştirilen çalışmada ARIMA modelinin doğruluk oranı %93 olurken, Olvera-Juarez ve Huerta-Manzanilla(2019) BTC fiyatlarını tahminlemeye yönelik yaptıkları çalışmada ARIMA yöntemiyle %90,31 doğruluk oranında kalmışlardır. Buna ek olarak, Abu Bakar ve Rosbi(2017) BTC fiyat tahmini çalışmasında ARIMA yöntemini kullanarak %46,40 MAPE oranına ulaşırken, pandemi sonrasını kapsayan veriler kullanarak yapmış olduğumuz çalışmada ARIMA(2,1,4) modeli %14,28 MAPE oranıyla daha iyi bir performans sergilemiştir.

Çalışma, koronavirüs pandemisi öncesi ve sonrası BTC fiyatlamalarının yer aldığı iki farklı veri üzerinde yapılmıştır. Literatürde SVM, ARIMA ve LSTM algoritmalarının aynı anda bir veri üzerinde uygulanıp sonuçlarının karşılaştırıldığı bir yayın bulunmamıştır. Bu çalışmayla birlikte bu üç algoritma, pandemi öncesi ve pandemi sonrasını kapsayan veriler üzerinde uygulanmış ve model performansları sunulmuştur. Kullanılan bütün yöntemlerin parametreleri ilgili yerlerde verilmiştir. Kullanılan verilerin, kullanılan yöntemler aynı olsa bile başarı oranını etkilediği gözlemlenmiştir. Pandemi öncesi ve pandemi sonrası olmak üzere iki ayrı zaman dilimindeki fiyatlamaları içeren bu çalışmada, BTC fiyatlarının pandemi esnasında yükseliş beklentilerine rağmen bir yükselişte bulunmadığı da uyguladığımız yöntemlerle tespit edilebilmiştir.

Bu çalışmayla birlikte, daha uzun dönem tarihleri ele alınıp çalışmanın genişletilmesi planlanmaktadır. Ayrıca, diğer sanal para birimlerinin de dikkate alınarak bu birimlerin tahminlenmesi ve bu para birimlerinin birbirlerinin fiyatlarına etkisinin analizi de yine çalışılacak alanlar olarak araştırmacıların önüne çıkmaktadır.

#### KAYNAKLAR (REFERENCES)

- [1] V. A. Maese, A. W. Avery, B. A.Naftalis, S. P. Wink, Y. D. Valdez, "Cryptocurrency: A Primer", *Banking LJ*, 133, 468, 2016.
- [2] Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends", **In 2017 IEEE international congress on big data (BigData congress)**, 557-564, 2017.
- [3] M. Nofer, P. Gomber, O. Hinz, D. Schiereck, "Blockchain", *Business & Information Systems Engineering*, 59(3), 183-187, 2017.
- [4] M. Pilkington, "Blockchain technology: principles and applications", *In Research handbook on digital transformations*, 2016.
- [5] M. Tanrıverdi, M. Uysal, M. T. Üstündağ, "Blokzinciri Teknolojisi Nedir? Ne Değildir?: Alanyazın İncelemesi", *Bilişim Teknolojileri Dergisi*, 12(3), 203-217, 2019.
- [6] Internet: S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System", <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.221.9986>, 12.11.2019.
- [7] N. Gandal, J.T Hamrick, T. Moore, T. Oberman, "Price manipulation in the Bitcoin ecosystem", *Journal of Monetary Economics*, 95, 86-96, 2018.
- [8] Internet: M. Dzirutwe, "Think bitcoin's getting expensive? Try Zimbabwe", <https://www.reuters.com/article/us-zimbabwe-bitcoin/think-bitcoins-getting-expensive-try-zimbabwe-idUSKBN1DD0NF,27.3.2020>.
- [9] J. R. Henrickson, T. L Hogan, W. J. Luther, "The Political Economy of Bitcoin", *Economic Inquiry*, 54(2), 925-939, 2016.
- [10] J. Bohr, M. Bashir, "Who uses bitcoin? An Exploration of the Bitcoin Community", **Twelfth Annual International Conference on Privacy, Security and Trust (PST)**, Toronto, ON, Canada, 94-101, 2014.
- [11] R. Stokes, "Virtual Money Laundering: The case of Bitcoin and the Linden Dollar", *Information & Communications Technology Law*, 21(3), 221-236, 2012.
- [12] World Health Organization, Coronavirus disease 2019(COVID-19), **Situation Report**, 72, 2020.
- [13] Internet: <https://tr.cointelegraph.com/news/covid-19-has-reduced-the-risk-of-a-post-halving-price-dump>, 02.06.2020.
- [14] S. Savaş, N. Topaloğlu, M. Yılmaz, "Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri", *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23, 2012.
- [15] M. Gök, "Makine Öğrenmesi Yöntemleri ile Akademik Başarının Tahmin Edilmesi", *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, 5(3), 139-148, 2017.
- [16] Y. Peng, P. H. M. Albuquerque, J. M. C. de Sá, A. J. A. Padula, M. R. Montenegro, "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, 2018.
- [17] H. Jang, J. Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain I", *IEEE Access*, 6, 5427-5437, 2018.
- [18] P. Katsiampa(2017), "Volatility Estimation for Bitcoin: A Comparison of GARCH Models", *Economics Letters*, 158, 3-6,2017.
- [19] D. U. Sutiksno, A. S. Ahmar, N. Kurniasih, E. Susanto, A. Leiwakabessy, "Forecasting Historical Data of Bitcoin using ARIMA and  $\alpha$ -Sutte Indicator", *Journal of Physics: Conference Series*, 1028(1), 012194, 2018.
- [20] N. A. Bakar, S. Rosbi, "Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction", *International Journal of Advanced Engineering Research and Science*, 4(11), 130-137, 2017.

- [21] E. Ş. Karakoyun, A.O. Çibikdiken, "Comparison of ARIMA Time Series Model and LSTM Deep Learning Algorithm for Bitcoin Price Forecasting", **The 13th Multidisciplinary Academic Conference (The 13th MAC 2018)**, Prag, Czech Republic, 171-180, 2018.
- [22] S. McNally, J. Roche, S. Caton, "Predicting the Price of Bitcoin Using Machine Learning", **26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)**, 339-343, 2018.
- [23] B. Sakız, E. Kutlugün, "Bitcoin Price Forecast via Blockchain Technology and Artificial Intelligence Algorithms", **26th Signal Processing and Communications Applications Conference (SIU)**, 1-4, 2018.
- [24] A. Azari, "Bitcoin Price Prediction: An ARIMA Approach", 2019.
- [25] D. Olvario-Juarez, E. Huerta-Manzanilla, "Forecasting bitcoin pricing with hybrid models: A review of the literature", *International Journal of Advanced Engineering Research and Science*, 6(9), 161-164, 2019.
- [26] A. Greaves, B. Au, "Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin", 2015.
- [27] J. Almeida, S. Tata, A. Moser, V. Smit, "Bitcoin Prediction using ANN", *Neural Networks*, 1-12, 2015.
- [28] N. A. Hitam, A. R. Ismail, "Comparative Performance of Machine Learning Algorithms for Cryptocurrency Forecasting", *Indonesian Journal of Electrical Engineering and Computer Science*, 11(3), 1121-1128, 2018.
- [29] İ. Çütücü, Y. Kılıç, "Bitcoin Fiyatları ile Dolar Kuru Arasındaki İlişki: Yapısal Kırılmalı Zaman Serisi Analizi", *Yönetim ve Ekonomi Araştırmaları Dergisi*, 16(4), 349-366, 2018.
- [30] A. Sönmez, "Sanal Para Bitcoin", *The Turkish Online Journal of Design, Art and Communication (TOJDAC)*, 4(3), 1-14, 2014.
- [31] J. M. Griffin, A. Shams, "Is Bitcoin Really Un-Tethered?", Available at SSRN, 3195066, 2018.
- [32] Internet: <http://www.coinmarketcap.com>, 29.10.2019.
- [33] Z. A. Farhath, B. Arputhamary, L. Arockiam, "A Survey on ARIMA Forecasting Using Time Series Model", *Int. J. Comput. Sci. Mobile Comput.*, 5, 104-109, 2016.
- [34] K. Pichotta, R. J. Mooney, "Learning Statistical Scripts with LSTM Recurrent Neural Networks", **30th AAAI Conference on Artificial Intelligence (AAAI-16)**, 2016.
- [35] S. S. Panigrahi, J. K. Mantri, "Epsilon-SVR and Decision Tree for Stock Market Forecasting", **International Conference on Green Computing & Internet of Things**, Greater Noida, Delhi, 761-766, 2015.
- [36] C.F. Lin, S. D. Wang, "Fuzzy Support Vector Machines", *IEEE Transactions on Neural Networks*, 13(2), 464-471, 2002.
- [37] V. Jakkula, "Tutorial on Support Vector Machine (SVM)", *School of EECS*, Washington State University, 2006.
- [38] P. S. Yu, S. T. Chen, I. F. Chang, "Support vector regression for real-time flood stage forecasting", *Journal of Hydrology*, 328(3-4), 704-716, 2016.
- [39] D. Bhatt, P. Aggarwal, P. Bhattacharya, V. Devabhaktuni, "An Enhanced MEMS Error Modeling Approach Based on Nu-Support Vector Regression", *Sensors*, 12, 9448-9466, 2012.
- [40] F. Zhang, C. Deb, S. E. Lee, J. Yang, K. W. Shah, "Time Series Forecasting for Building Energy Consumption using Weighted Support Vector Regression with Differential Evolution Optimization Technique", *Energy and Buildings*, 126, 94-103, 2016.
- [41] M. Story, R. G. Congalton, "Accuracy Assessment: A User's Perspective", *Photogrammetric Engineering and Remote Sensing*, 52(3), 397-399, 1986.
- [42] S. J. Darnell, D. Page, J. C. Mitchell, "An automated decision-tree approach to predicting protein interaction hot spots", *Proteins*, 68, 813-823, 2007.
- [43] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation", *Journal of Machine Learning Technologies*, 2(1), 37-63, 2011.
- [44] P. A. Flach, M. Kull, "Precision-Recall-Gain Curves: PR Analysis Done Right", *Advances in Neural Information Processing Systems*, 28, 838-846, 2015.
- [45] U. Khair, H. Fahmi, S. Al Hakim, R. Rahim, "Forecasting error calculation with mean absolute deviation and mean absolute percentage error", **In Journal of Physics: Conference Series**, 1(930), 012002, 2017.
- [46] G. Brassington, "Mean absolute error and root mean square error: which is the better metric for assessing model performance?", **In EGU General Assembly Conference Abstracts**, 19, 3574, 2017.
- [47] S. Boughorbel, F. Jarray, M. El-Anbari, "Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric", *PLoS one*, 12(6), 2017.
- [48] M. Abdar, "Using Decision Trees in Data Mining for Predicting Factors Influencing of Heart Disease", *Carpathian Journal of Electronic & Computer Engineering*, 8(2), 2015.
- [49] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research (JMLR)*, 12, 2825-2830, 2011.
- [50] T. E. Oliphant, "A guide to NumPy", USA: Trelgol Publishing, 1, 85, 2006.
- [51] W. McKinney, "Pandas: a foundational Python library for data analysis and statistics", *Python for High Performance and Scientific Computing*, 14(9), 2011.
- [52] M. Abadi., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, "Tensorflow: A System for Large-Scale Machine Learning", **In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)**, 265-283, 2016.