

**T.C.  
ERCIYES ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE ZAMAN  
SERİLERİNDE DEĞER TAHMİNLEME**

**Hazırlayan  
Mustafa ÜLKER**

**Danışman  
Dr. Öğr. Üyesi Fehim KÖYLÜ**

**Yüksek Lisans Tezi**

**Ekim 2023  
KAYSERİ**



**T.C.  
ERCIYES ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE ZAMAN  
SERİLERİNDE DEĞER TAHMİNLEME  
(Yüksek Lisans Tezi)**

**Hazırlayan  
Mustafa ÜLKER**

**Danışman  
Dr. Öğr. Üyesi Fehim KÖYLÜ**

**Ekim 2023  
KAYSERİ**

## BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi belirtirim.

Mustafa ÜLKER

İmza:

“Makine Öğrenmesi Yöntemleriyle Zaman Serilerinde Değer Tahminleme” adlı Yüksek Lisans tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi’ne uygun olarak hazırlanmıştır.

**Hazırlayan**  
Mustafa ÜLKER

**Danışman**  
Dr. Öğr. Üyesi Fehim KÖYLÜ

Bilgisayar Mühendisliği ABD Başkanı  
Prof. Dr. Veysel ASLANTAŞ

## TEŐEKKÜR

Sayın Dr. Fehim Kyl,

Yksek lisans tezim iin beni ynlendirdiđiniz, desteđiniz ve rehberliđiniz iin size teŐekkr etmek istiyorum. Tez alıŐmam boyunca, konu seimi, araŐtırma planlaması, veri toplama, analiz ve sonuların yorumlanması gibi birok konuda beni ynlendirdiniz.

Sizin sayesinde tez konumunun kapsamını belirleyebildim ve araŐtırma yntemleri konusunda daha fazla bilgi edinebildim. Sizin deneyimli bakıŐ aınız, beni tezimde dođru ynde hareket etmem iin cesaretlendirdi. Ayrıca tezimdaki zorlukları aŐmam iin srekli motivasyon sađladınız.

Sizin benim tez danıŐmanım olmanız, benim iin bir ayrıcalıktı. Beni her zaman ynlendirdiniz ve tez alıŐmamda dođru kararları almamda yardımcı oldunuz. Siz sayesinde bu tez, bir đrencinin hayal edebileceđi en iyi sonuları verdi.

Bu harika deneyim iin teŐekkr ederim. İlham veren bir danıŐman olduđunuz ve benim hayatımda unutulmaz bir iz bıraktıđınız iin minnettarım.

Saygılarımla,

# MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE ZAMAN SERİLERİNDE DEĞER TAHMİNLEME

**Mustafa ÜLKER**  
**Erciyes Üniversitesi, Fen Bilimleri Enstitüsü**  
**Yüksek Lisans Tezi, Ekim 2023**  
**Danışman : Dr. Öğr. Üyesi Fehim KÖYLÜ**

## ÖZET

Belirsiz olan geleceğin tahmin edilmesi gerek işletmeler gerekse şahıslar için hayati öneme sahiptir [1]. Borsalarda hisse senetlerinin veya kripto paraların bir sonraki zaman dilimindeki fiyatlarının yüksek oranda doğru tahmin edilebilmesi al-sat yapanlar için mükemmel bir kazanç fırsatıdır. Kripto para borsalarının 7/24 açık olması, veri çekme, analiz oluşturma ve gündelik haberlerden işletme borsalarına göre daha az etkilenmesi bu çalışmanın kripto para borsaları üzerinde yapılmasını teşvik etmiştir. Bitcoin üzerinden yapılan çalışmada makine öğrenmesi algoritmalarının kullanımının fiyat tahminlemesiyle yapılan yatırımın kar elde ettirebileceğinden bahsedilebilir. Bu çalışmada kripto para borsasından alınan Ağustos 2020-Mayıs 2023 tarihleri arasına ait Bitcoin/USDT çifti fiyat hareketi (OHLCV, açılış fiyatı, yüksek fiyat, düşük fiyat, kapanış fiyatı ve hacim) verileri kullanıldı. OHLCV veri kümesi finans alanında kullanılan indikatörler kullanılarak çok parametrelili zenginleştirilmiş veri kümesi haline getirildi. Ağustos 2020-Haziran 2022 tarihleri arasındaki veri kümesi üzerinde makine öğrenmesi algoritmalarından Rastgele Orman(Random Forest) ve Aşırı Gradyan Artırma (XGBoost) algoritmaları kullanılarak regresyon modeli eğitimi yapıldı. Test veri kümesi olarak Haziran 2022-Mayıs 2023 tarihleri arasındaki borsa fiyat değerleri kullanıldı. Eğitilen regresyon modelleri test veri kümesi üzerinde farklı sayıda geçmiş veri (lag) bir sonraki değeri tahmin etmesi sağlanarak doğruluk performansları karşılaştırıldı. Testler içerisinde en düşük hata oranına sahip 23 gecikme (lag) kullanılan rastgele orman algoritması ile borsa işlemleri yapıldığında elde edilen cüzdan bakiyesi analiz edildi.

**Anahtar Kelimeler:** bilgisayar öğrenimi, veri madenciliği, regresyon analizi, rastgele orman algoritması, aşırı gradyan artırma

## VALUE ESTIMATION IN TIME SERIES BY MACHINE LEARNING METHODS

Mustafa ÜLKER

Erciyes University, Graduate School of Natural and Applied Sciences

M.Sc. Thesis, October 2023

Supervisor: Asst. Prof. Dr. Fehim KÖYLÜ

### ABSTRACT

Predicting the uncertain future is of vital importance for both businesses and individuals [1]. Being able to predict the prices of stocks or cryptocurrencies in the next time period with high accuracy in the stock markets is an excellent earning opportunity for those who buy and sell. The fact that cryptocurrency exchanges are open 24/7 and are less affected by data extraction, analysis and daily news than business exchanges encouraged this study to be conducted on cryptocurrency exchanges. In the study conducted on Bitcoin, it can be mentioned that the use of machine learning algorithms can make a profit from the investment made by predicting the price. In this study, Bitcoin/USDT pair price movement (OHLCV, opening price, high price, low price, closing price and volume) data from the cryptocurrency exchange between August 2020 and May 2023 was used. The OHLCV dataset was turned into a multi-parameter enriched dataset using indicators used in the field of finance. Regression model training was performed on the dataset between August 2020 and June 2022 using Random Forest and Extreme Gradient Boosting (XGBoost) algorithms, which are machine learning algorithms. Stock market price values between June 2022 and May 2023 were used as the test data set. The accuracy performances of the trained regression models were compared by allowing them to predict the next value with different numbers of historical data (lag) on the test data set. The wallet balance obtained when stock market transactions were made was analyzed with the random forest algorithm using 23 lags, which had the lowest error rate among the tests.

**Keywords:** machine learning, data mining, regression analysis  
random forest, xgboost



## İÇİNDEKİLER

### MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE ZAMAN SERİLERİNDE DEĞER TAHMİNLEME

BİLİMSEL ETİĞE UYGUNLUK SAYFASI . . . . .	ii
YÖNERGEYE UYGUNLUK SAYFASI . . . . .	iii
KABUL VE ONAY . . . . .	iv
TEŞEKKÜR . . . . .	v
ÖZET . . . . .	vi
ABSTRACT . . . . .	vii
İÇİNDEKİLER . . . . .	viii
KISALTMALAR . . . . .	x
TABLolar LİSTESİ . . . . .	xi
ŞEKİLLER LİSTESİ . . . . .	xii
GİRİŞ . . . . .	1

#### 1. BÖLÜM

##### 1.KAVRAMLAR

1.1. Para . . . . .	6
1.1.1. Bitcoin Nedir . . . . .	7
1.1.1.1. Blok Zinciri . . . . .	7
1.1.1.2. Fikir Birliği Mekanizmaları (Konsensus) . . . . .	10
1.1.1.3. Bitcoinin Değeri . . . . .	12
1.2. Makine Öğrenmesi . . . . .	12
1.2.1. Makine Öğrenmesi Yöntemlerinin Kategorileri . . . . .	13
1.2.1.1. Denetimli Öğrenme . . . . .	14
1.2.1.2. Denetimsiz Makine Öğrenmesi . . . . .	15
1.2.1.3. Yarı Denetimli Makine Öğrenmesi . . . . .	15

#### 2. BÖLÜM

##### 2.METOTLAR

2.1. Veri Madenciliği . . . . .	17
---------------------------------	----

2.2. Makine Öğrenmesi Algoritmaları . . . . .	18
2.2.1. Rastgele Orman(RF) ve Karar Ağacı . . . . .	18
2.2.2. Aşırı Gradyan Artırma (XGBoost) Algoritması . . . . .	19
2.3. Regresyon Analizi . . . . .	20
2.3.1. Rastgele Orman Regresyonu (RFR) . . . . .	22
2.3.2. Otomatik Regresyon (AR) . . . . .	22
2.4. Performans Ölçümleri . . . . .	24
2.5. Materyaller . . . . .	24
2.6. Veri kümesi . . . . .	25
2.7. Veri Önışleme . . . . .	26
2.8. İndikatör hesaplamalarıyla veri kümesi zenginleştirilmesi . . . . .	26
<b>3. BÖLÜM</b>	
<b>3.DENEYLER</b>	
3.1. Algoritma parametre analizi ve performans ölçümleri . . . . .	35
<b>4. BÖLÜM</b>	
<b>SONUÇ</b>	
4.1. Bulgular ve Sonuçlar . . . . .	43
4.2. Çıkarımlar . . . . .	46
4.3. Tartışma ve Öneriler . . . . .	47
<b>KAYNAKLAR</b> . . . . .	49
<b>EKLER</b> . . . . .	60
<b>ÖZGEÇMİŞ</b> . . . . .	61

## KISALTMALAR

<i>SVM</i>	: Support Vector Machine(Destek vektör makinası)
<i>RF</i>	: Random Forest(Rastgele Orman)
<i>LR</i>	: Linear Regression(Doğrusal Regresyon)
<i>MLP</i>	: Multilayer Layer Perceptron (Çok Katmanlı YSA)
<i>AFER</i>	: Average Forecasting Error Rate(OTHO)
<i>OTHO</i>	: Ortalama Tahmin Hatası Oranı
<i>MSE</i>	: Ortalama Kare Hata
<i>RMSE</i>	: Kök Ortalama Kare Hata
<i>ARIMA</i>	: Autoregressive Integrated Moving Average(OEHO)
<i>OEHO</i>	: Otoresif Entegre Hareketli Ortalama
<i>AFER</i>	: Average Forecasting Error Rate(OTHO)
<i>OTHO</i>	: Ortalama Tahmin Hata Oranı
<i>VGG16</i>	: Sanal Geometri Grubu 16 katmanlı Evrişimsel sinir ağı
<i>NARX</i>	: Doğrusal Olmayan Otomatik Regresyon Dışsal denetimi
<i>OHLCV</i>	: Açılış(O),Yüksek(H),Düşük(L),Kapamış(C),Hacim(V)
<i>ARCH</i>	: Otoresif Koşullu Değişen Varyans
<i>ECB</i>	: European Central Bank (Avrupa Merkez Bankası)
<i>KA</i>	: Karar Ağaçları
<i>KNN</i>	: K-En Yakın Komşular
<i>OOB</i>	: Out Of Bag(Test veri kümesi)
<i>RBF</i>	: Radyal Tabanlı Fonksiyon
<i>YSA</i>	: Yapay Sinir Ağları
<i>LSTM</i>	: Uzun Kısa Vadeli Bellek
<i>LAG</i>	: Kaç önceki gözlemin kullanılması parametresi
<i>MSE</i>	: Mean Squared Error(En küçük kareler ortalaması)
<i>RMSE</i>	: Root Mean Squared Error(Kök ortalama kareler hatası)
<i>MA</i>	: Moving Average (Hakeretli Ortalama)
<i>EMA</i>	: Exponential Moving Average (Üssel MA)
<i>DEMA</i>	: Double Exponential Moving Average (Çift EMA)
<i>BTC</i>	: Bitcoin
<i>USD</i>	: Amerikan Doları
<i>USDT</i>	: Amerikan Dolarına endekslenen Tether Sabit Parası
<i>TWh</i>	: TeraWatt Saat
<i>NumberOfTrades</i>	: Yapılan ticaret sayısı [?]
<i>OpenTime</i>	: Açılış zamanı
<i>OpenPrice</i>	: Bitcoin açılış fiyatı
<i>HighPrice</i>	: Bitcoinin ulaştığı en yüksek fiyat
<i>LowPrice</i>	: Bitcoinin en düştüğü en düşük fiyat
<i>ClosePrice</i>	: Bitcoin kapanış fiyatı
<i>VolumeBTC</i>	: Bitcoin alım-satım hacmi
<i>CloseTime</i>	: Kapanış zamanı
<i>QuoteAssetVolume</i>	: Ticaretin USD hacmi
<i>TakerBuyVolume</i>	: Bitcoin satın alma hacmi
<i>TakerBuyQuoteAssetVolume</i>	: Satın alma USD hacmi
<i>Ignore</i>	: Geçerliliği

**TABLolar LİSTESİ**

Tablo 2.1.	XGBoost işlem adımları. . . . .	20
Tablo 2.2.	Çalışmada kullanılan ortalama tabanlı indikatörler. . . . .	27
Tablo 2.3.	Çalışmada kullanılan trend tabanlı indikatörler. . . . .	28
Tablo 2.4.	Çalışmada kullanılan oransal parametreler. . . . .	33
Tablo 2.5.	Çalışmada kullanılan mumlar. . . . .	33
Tablo 3.1.	RF Tahmin Modeli parametreleri . . . . .	37
Tablo 3.2.	XGBoost tahmin modeli parametreleri . . . . .	39
Tablo 3.3.	ForecasterAutoreg parametre değerleri. . . . .	41
Tablo 4.1.	Algoritma ve lag çiftinin RMSE etkisi. . . . .	44
Tablo 4.2.	RF lag23 başarımlı hesaplaması yapılmıştır. (Aylık ortalamalar şeklinde özet tablodur.) . . . . .	47

## ŞEKİLLER LİSTESİ

Şekil 1.1.	Limitin 21.000.000 adet olduğu Bitcoin sayısının yıllara göre artışı gösterilmektedir . . . . .	8
Şekil 1.2.	Örnek hash blok gösterilmiştir. . . . .	9
Şekil 1.3.	Bitcoin blok zinciri boyutunun yıllara göre büyümesi. . . . .	9
Şekil 1.4.	Blok zincirindeki blokların birbiri ile ilişkisi. . . . .	11
Şekil 1.5.	Bitcoin/\$ fiyatının yıllara göre değişimi [2]. . . . .	12
Şekil 1.6.	Makine öğrenmesi yöntemlerinin ortak işlem adımları. . . . .	13
Şekil 1.7.	Makine öğrenmesi yöntemlerinin çeşitleri. . . . .	14
Şekil 1.8.	Denetimli makine öğrenmesi çalışma şekli [3]. . . . .	14
Şekil 1.9.	Denetimsiz makine öğrenmesi çalışma şekli [3] . . . . .	15
Şekil 1.10.	Yarıdenetimli makine öğrenmesi çalışma şekli [3]. . . . .	15
Şekil 2.1.	24 saatlik hacimlerine göre ilk 6 kripto para borsası [4] . . . . .	25
Şekil 2.2.	Tarihe göre bitcoin fiyatında 26.08.2020-11.06.2022 arası eğitim verisi ve 11.06.2022-09.05.2023 arası test verileri gösterilmiştir. . . . .	26
Şekil 2.3.	Birikimli Hacim Farkı indikatörünü mum çubuklarıyla gösterilmiştir [5]. . . . .	29
Şekil 2.4.	Mumlar ve sonraki periyotlarda tahmin edilen trend yönleri gösterilmiştir. . . . .	34
Şekil 3.1.	Makine öğrenmesi eğitim ve test aşamasında yapılan işlemler gösterilmiştir. . . . .	36
Şekil 4.1.	RF (rmse:832) ve XGBOOST (rmse:1560) algoritmalarının lag 1'e göre tahminleri karşılaştırılmıştır. . . . .	45
Şekil 4.2.	RF (rmse:850) ve XGBOOST (rmse:3258) algoritmalarının lag 180'e göre tahminleri karşılaştırılmıştır. . . . .	45
Şekil 4.3.	RF ve XGBOOST lag performans analizi yapılmıştır. . . . .	46

Şekil 4.4. RF (rmse:730) ve XGBOOST (rmse:1513) algoritmalarının lag 23'e göre tahminleri karşılaştırılmıştır. . . . .	46
--	----



## GİRİŞ

Zaman serisi, zaman içinde meydana gelen gözlemlerin bir araya toplanmasıdır [6]. Dolayısıyla, zamanla etiketlenmiş ve değişen değerleri tutan bir veri kümesidir. Ekonomik alanda maden ve gıda gibi emtia fiyatları, döviz kurları, hisse senedi fiyatları, kripto para değerleri zaman serisi ile ifade edilmektedir. Bu açıdan karar vermek, karar süreci ve sonraki süreci değerlendirmek için zaman serisinin analizi yapılmaktadır. Günümüzde bilgisayarların hesaplama kabiliyetlerinin gelişmesi ile birlikte zaman serisinin analizi için istatistik, makinalı öğrenme, veri madenciliği, yapay sinir ağları gibi çeşitli yöntemler kullanılmaktadır. Zamana bağlı olarak elde edilen değerlerin birbiri ile ilişkili olacağı kabulü ile analiz yapılmaktadır. Finans başta olmak üzere, çeşitli alanlarda yaygın olarak kullanılan zaman serisi verileri için tahminlerin doğru olması çok önemlidir. Zaman serilerinin özellikleri için verilerdeki her bilgi kritiktir. Özellikle, trend bilgisine sahip veriler üzerinde tahmin yapmak karmaşık bir süreçtir. Belirsiz olan geleceğin tahmin edilmesi gerek işletmeler gerekse şahıslar için hayati önem taşımaktadır [1].

Makine öğrenmesi, bir bilgisayar programının belirli bir görevi yerine getirmek için verilerden otomatik olarak bir model öğrenmesi için kullanılan bir yapay zeka dalıdır. Bu teknoloji, algoritmalar kullanarak büyük veri kümelerini analiz eder ve bu verilerden desenleri ve ilişkileri tanımlar. Bu bilgi daha sonra, yeni verileri analiz etmek ve tahminler yapmak için kullanılabilir hale getirilir.

Makine öğrenmesi, sınıflandırma, kümeleme, regresyon, öngörü ve çevrimiçi öğrenme gibi birçok alanda kullanılabilir. Bu yöntemler, tıp, finans, imalat, perakende, enerji ve birçok diğer endüstride uygulanmaktadır. Makine öğrenmesi teknikleri, insanların yapamayacağı kadar büyük veri kümelerini analiz ederek yeni anlamlar keşfetmelerine yardımcı olabilir ve karar verme süreçlerini optimize edebilir.

Daha önce yapılan araştırmalarda, makine öğrenmesi teknikleri kullanılarak, zaman

serisi verileri işlenip zaman serisinin gelecekteki adımlarının tahmin edilebildiği görülmektedir [7].

Makine öğrenme algoritmaları, verileri kullanarak bir model ya da bir desen(pattern) oluştururlar, bunu gelecekte karşılaşılabileceği yeni durumları tahmin etmek için kullanır [8]. Örneğin, anlık olarak piyasayı izleyen bir makinenin, geçmiş verilerle oluşturduğu model ile gelecek piyasa analizleri yapması bir makine öğrenmesi olarak kabul edilebilir. Yani, makine öğrenmesi, verilerin analizi, model oluşturma ve bu modeli kullanarak tanıma adımlarının tekrarlı bir şekilde kullanılarak geliştirildiği bir öğrenme ve kullanma süreci olarak değerlendirilebilir [9].

Borsalarda hisse senetlerinin veya 2009 yılında “Bitcoin: A Peer-to-Peer Electronic Cash System” [10] isimli makale ile tanıştığımız ilk kripto para olan Bitcoin ve diğer alternatif kripto paraların bir sonraki zaman dilimindeki fiyatlarının yüksek oranda doğru tahmin edilebilmesi al-sat yapanlar için bir kazanç fırsatı olarak değerlendirilebilir. Bu alım satımlar için açılan kripto para borsalarının 7/24 açık olması, veri çekme, analiz oluşturma ve gündelik haberlerden işletme borsalarına göre daha az etkilenmesi bu çalışmanın kripto para borsaları üzerinde yapılmasını teşvik etmiştir.

**Literatür Özeti** Neden Bitcoin sorusuna yönelik Yermack (2013) tarafından Bitcoin’in çok değerli bir para birimi olabileceği belirtilmiştir [11]. Fiyat dalgalanmalarının yüksek olduğu finansal varlıklarda riskin yüksek olduğu kadar kazanç oranı da yüksek olabilir. Kristoufek (2015) Çin piyasasının hareketlerinin Bitcoin fiyatının üzerinde potansiyel etkilerini araştırmış ve benzersiz bir spekülasyon varlık olduğunu belirtmiştir [12]. Dolayısıyla Bitcoin’in çok tercih edilen bir finansal varlığa dönüştüğü ifade edilebilir. Zaten spekülasyon olan bu finansal varlığın düşüş veya yükseliş eğiliminde olduğunun tahminlenmesi hatta fiyatının tahmin edilmesi çok daha fazla ilgi uyandırmaktadır.

Bitcoin fiyatının değişiminin düşüş ve yükseliş olmak üzere yönünün tahminlenmesi sınıflandırma problemi olarak literatürde bir çok çalışmada ele alınmıştır. Destek vektör makineleri (SVM) [13], binom genelleştirilmiş doğrusal model (Bi-GLM) [14], rastgele orman (RF) [15], yapay sinir ağı (ANN) [16], tekrarlayan sinir ağı (RNN) [17], lojistik regresyon (LR) [18] vb. yöntemler fiyat tahmini için uygulanmıştır.



Madan vd. (2015) yaptıkları çalışmada makine öğrenme algoritmaları kullanarak günlük fiyat değişiminin işaretini Bi-GLM ile %98,7 doğrulukla, SVM ile %27 doğrulukla, RF ile %95 doğrulukla tahmin edebildiklerini, 10 dakikalık fiyat değişiminin işaretini ise %50-55 doğrulukla tahmin edebildiklerini belirtmiştir [?]. McNally (2016) tarafından yapılan çalışmada RNN ile %52 oranında başarı elde edilmiştir [19]. Indera vd. (2018) yaptıkları çalışmada karar ağacı, tahmin için toplam özelliklerin %30'unu seçmek için kullanmış, SVM ile eğitilen model trend yönü belirlemede %56 başarı sağlamıştır [20]. Alex Greaves ve Benjamin Au(2015) tarafından yapılan çalışmadaysa SVM ile %53,7, LR ile %54,3, ANN yöntemi ile %55,1 oranında başarı elde edilmiştir [21].

Mem vd. (2017) tarafından yapılan çalışmada fiyat tahminlemesinde ve aynı zamanda fiyat üzerindeki etkisi yüksek olan hiper parametreleri bulmak için aktivasyon fonksiyonu düzeltilmiş doğrusal ünite (ReLU) olan evrimsel sinir ağları (CNN) kullanmış ve LR %65, SVM %67, çok katmanlı perseptron (MLP) %68 oranlarında doğru sınıflandırma yaparken CNN ile %96 doğrulukla sınıflandırma yaptığı göstermiştir [22].

Zaman serisi analizi dışında farklı yöntemlerle de Bitcoin fiyatının yönü tahminlemesi yapılmıştır. Kim vd. (2016) tarafından yapılan çalışmada kullanıcıların yorumlarını çok olumlu, olumlu, nötr, olumsuz, çok olumsuz olacak şekilde sınıflandırmış ve sosyal medyada yazılan yazıların ve yorumların işlem sayısına etkisinin olduğunu göstermişlerdir [23]. İşlem sayısının, dolaylı olarak işlem hacmini ve dolayısıyla fiyatı etkileyen bir unsur olduğunu göstermektedir. Steqvist (2017) tarafından yapılan çalışma da Twitter tweetleri kullanılarak %83 oranında bir tahminsel başarıya ulaşılmıştır [24]. Temel analizin teknik analiz kadar önemli olduğunu vurgulanmaktadır.

2020 yılında yapılan çalışmada Bitcoin market üstünlüğü, altın fiyatları ve Bitcoin madencilik zorluğu gibi parametreler kullanarak hem istatistik hem de makineli öğrenme yöntemleri ile fiyat tahmini yapmıştır. LR, lineer ayrılık analizi, RF, aşırı gradyan artırma (XGBoost), kuadratik ayrılık analizi, SVM ve uzun kısa süreli hafıza yöntemleri ile tahmin yapmıştır. Ortaya koyduğu sonuçta yüksek frekanslı veri üzerinde makinalı öğrenme yöntemlerinin istatistik yöntemlere göre daha yüksek

doğruluk değeri elde ettiğini ifade etmektedir [25].

Makine öğrenmesinde model eğitilirken kullanılan parametreler, giriş değişkenlerinin hedef değişkene olan etkisini belirlemede önemli rol oynamaktadır. Yassin vd. (2017) yaptıkları çalışmada, Apple şirketinin haftalık hisse senedi fiyatlarının 1995-2013 zaman dilimi boyunca tahmin edilmesi için MLP tabanlı olan sistemin dinamik giriş/çıkış okumalarından matematiksel bir model oluşturan ve bu modele dayanarak sistemin gelecekteki davranışını tahmin eden bir sistem tanımlama modeli doğrusal olmayan otomatik regresyon (NARX) modeli kullanmıştır. Sonuç olarak, tahmin modeli ile gerçek veriler arasında iyi bir uyum olduğunu belirtmiş gecikme hiper parametresi (lag) sayısının optimizasyonunda 13-20 aralığında korelasyonun çok düşük olduğunu göstermiştir [26].

Literatürde istatistik ve formasyon tabanlı yapılan çalışmalar da bulunmaktadır. Tadvi (2018) tarafından yapılan çalışmada Elliot dalga teorisi kullanmış ve 5. dalgada farklı öngörücü tahminlerde bulunulmuştur [27]. Buna göre Aralık 2019'da Bitcoin fiyatının 28.000 dolar olacağını tahmin etmiştir fakat 2019 Aralık ayı en yüksek BTC fiyatı 7500 civarında olmuştur.

Sakız ve Kutlugün (2018) tarafından yapılan çalışmada Bitcoin'in boğa sezonu yaşadığı yıl olan 2017 yılı aylık Bitcoin fiyat ortalaması verisiyle ANN eğitimi yaparak Şubat 2018'de Bitcoin fiyatının 170 milyon 130 bin dolar olduğu tahminini yapmış ve çalışmalarında kullandıkları geçmiş Bitcoin fiyatları veri kümesinin Bitcoin fiyat tahmininde yararlanılamayacağı sonucuna ulaşmışlardır [28].

Shah, Devavrat ve Zhang, Kang (2014) Bayes doğrusal regresyon [29] yöntemi kullanarak yaptıkları çalışmalarında 3781 Yuan yatırımla, 50 günde, 2872 işlem yapmış ve 3362 Yuan kazanç sağlayarak %89 oranında bir kazanç sağlamışlardır [30].

Amjad vd. (2017), bütünleşik otoregresif hareketli ortalama (ARIMA) [31] klasik zaman serisi analizlerinin RF algoritmasına göre kötü tahminler yaptığını belirtmiştir [32]. Bu çalışma bize, dalgalı bir fiyatın doğrusal regresyonlarla tahmin edilmesinin doğru sonuçlar vermeyeceğini gösterebilir.

Amjad vd. (2018) tarafından yapılan çalışmada kuantum bazlı evrimsel algoritma (QEA) kullanılmış, 3 aylık süre boyunca günlük tahminlerde ortalama tahmin hata

oranı (OTHO) %0.96 olarak bulunmuştur [33]. Tahminlerinin gerçek değerlere ne kadar yakın olduğunu ölçen bir performans metriği olan OTHO eğer düşükse, tahmin modelinin daha doğru tahminler yaptığını gösterirken, daha yüksek değerleri tahminlerin daha az doğru olduğunu göstermektedir.

Şengül vd. (2022) yılında, k-en yakın komşular (KNN) [34] temelli çalışmada bir günlük tahmininde kök ortalama kare hatası (RMSE) 1025 ve RF ile eğittiği modelin bir günlük tahmininde ise RMSE değerini 905 olarak ölçmüş ve en az hata değerine sahip olan algoritmanın RF regresyonu olduğunu belirtmiştir [35].

### **Çalışmanın Amacı**

Bu kapsamda tezin amacı zaman serilerinde, insanların yatırımları için erken karar verebilmesini sağlayacak bir yöntemin araştırılması olacaktır. Rastgele orman ve aşırı gradyan artırma yöntemleriyle zaman serilerinde regresyon işlemi ile değer tahminlemesi yapılacaktır. Gecikme hiperparametresi değiştirilerek, algoritmaların yapılacak olan tahminlerinde karşılaştırmalı olarak performans ölçümleri yapılacaktır.

### **Tezin Organizasyonu**

Bu tez çalışması aşağıdaki biçimde düzenlenmiştir.

1. Bölümde, kavramların tanımlanması, teknik bilgilerin verilmesi yapılmıştır.
2. Bölümde, materyaller ve veri kümesinin tanımlanması, veri kümesinin eğitim ve test amaçlı olarak bölünmesi, metotların belirtilmesi ve deneylerin yapılması açıklanmıştır.
3. Bölümde, yapılan deneyler sonucunda elde edilen verilerin gerçek verilerle karşılaştırılması ve hata payları değerlendirilmiştir.
4. Bölümde ise sonuçlar, çıkarımlar, öneriler ve gelecekte yapılması planlanan çalışmalar verilmiştir.

## 1. BÖLÜM

### 1.KAVRAMLAR

#### 1.1. Para

Satın alınan mal ve hizmetlerin karşılığının ödenmesinde genel olarak kabul edilen bir değişim aracı ve bir hesap birimidir [36]. Para, günümüzde bireylerin temel ihtiyaçlarını dahi karşılamaları için mecburen sahip olmaları gereken sosyal araçtır. Bir toplum varsa para nihai gerekliliktir [37]. M.Ö. 7. yüzyılda deniz ticaretinde gelişmiş bir toplum olan Lidyalılar'ın altın ve gümüşten yapılmış sikkeyi ticarete kullanmaya başlamaları bu gerekliliği bize anlatmaktadır [38].

Zaman içerisinde, para için kullanılan değerli madenlerin az oluşu, üretiminin maliyetli olması ve aşınıp değer kaybetmesi gibi sebeplerle farklı para çeşitleri kullanılmıştır [39]. Değeri yüksek olan altın, gümüş gibi madenlerin taşınması, çalınma riskinin yüksek olması nedeni ile insanlar sarraflara paralarını emanet etmeye başlayarak karşılığında emanetin değeri yazılı kağıtlar almışlardır. Ticaretin ilerlemesi ile emanet karşılığında alınan kağıtların yerini makbuzlar almış oldu. Böylece zamanla altın ve gümüşlerin yerini makbuz olarak temsili paraya dönüşmüş oldu. Ancak bu süreçte kıymetli madenlerin emanet edildiği sarrafların altınları başkalarına vermesi veya makbuzların karşılığının çıkmamasından dolayı devlet bu duruma çözüm getirerek bu paraları altın karşılığı olarak kendisi basmış ve ticarete kullandırmıştır [40].

Konu insanlık için bir gereklilik olan para olduğunda daha fazla özellik taşıması beklenebilir. Paranın beş ana özellik taşıması gerekmektedir. Bunlar, taşınabilirlik, dayanıklılık, bölünebilirlik, homojenlik ve taklit edilmemesi özellikleridir [41]. Teknolojinin gelişmesiyle paranın bu özelliklerinin hepsini taşıyan bir sistem olarak

Bitcoin teknolojisi sunulmuştur.

### **1.1.1. Bitcoin Nedir**

2009 yılında “Bitcoin: A Peer-to-Peer Electronic Cash System” [10] isimli makale ile tanıştığımız ilk kripto para olan Bitcoin kullanımı aynı çerçevede yaygınlaşmıştır. Bitcoin merkeziyetsizdir, dağıtık mimariye sahiptir, gizlilik sağlar, güvenlidir, dayanıklıdır, doğrulama işlemleri yapılmaktadır, performansı yüksektir, güncelleme ve silme işlemlerine izin verilmez, sorgulamaya açıktır [42].

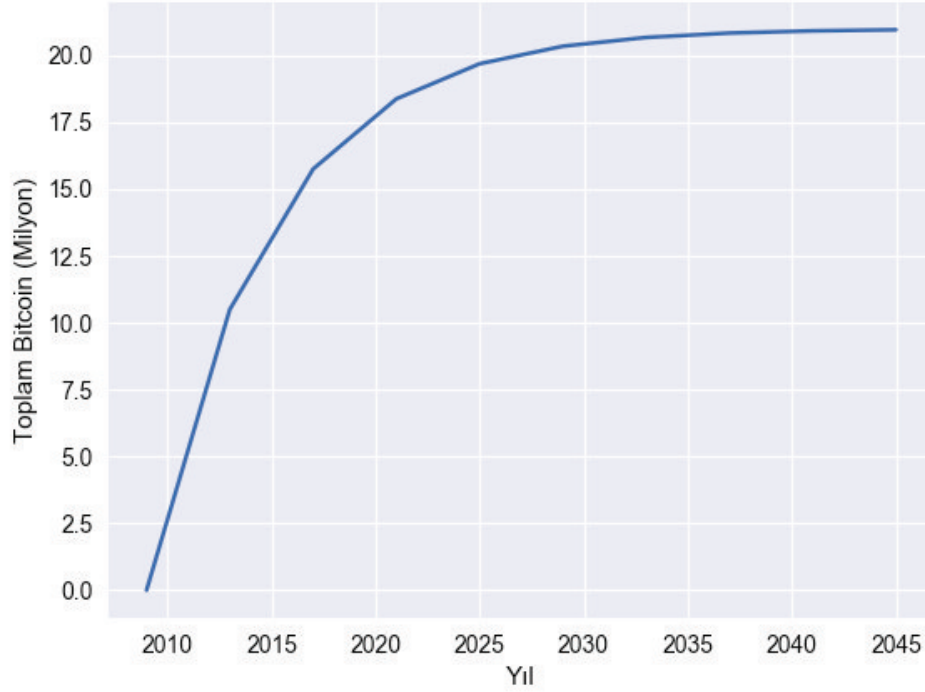
Merkeziyetsizlik, her hangi bir otorite tarafından yönetilmemesi anlamına gelmektedir. Kayıtların eklenmesi ve çıkarılması işlemlerin onaylanması bir merkez tarafından onay beklemediği için merkeziyetsizlik olarak adlandırılmıştır. Bütün transferlerin tutulduğu kayıt defteri isteyen herkes tarafından yedeklenebilmektedir. Bütün yedekler birbirleri ile haberleşerek güncel bloklar dağıtık mimarideki tüm bağımsız uçlar (node) tarafından kaydedilir. Kişilerin kimlik bilgileri veya Id numaraları kullanılmamaktadır, bunun yerine bir cüzdan numarası kullanan kişilerin kim oldukları gizlidir. Doğrulamaya ve sorgulamaya açık bir yapısı vardır. Hangi cüzdandan hangi cüzdana ne kadar BTC gönderildiği sorgulanabilmektedir. Tüm bu dayanaklarla ve özellikle her işlemin dünyadaki tüm kayıt tutanlarca kayıt altına alınarak bilgilerin güvenliğini sağlaması ile Bitcoin teknolojisinin güvenli olduğu sonucuna varılmaktadır.

Şifreleme uçtan uca güvenli gönderinin kimliği tespit

Özetleme Amaçlar 1. veri bütünlüğü doğrulamak 2. başka bir kümeye eşleştirme sha 256 64 karakter uzunluğunda bir bit bile değişse daha önce karşılaşılmamış tekil bir dizi (öz)

#### **1.1.1.1. Blok Zinciri**

Blok zinciri veya daha çok bilinen ismiyle blockchain, birçok uç birimde eş zamanlı tutulan defter teknolojisini kullanarak mevcut internet yapısının kripto destekli olarak yeniden tanımlanmasıdır [43]. Hash işlemi, kriptografi alanında güvenlik hedefi olarak kabul edilen bir sayıya ulaşmaktır. Hash fonksiyonlarının girişine ne



Şekil 1.1. Limitin 21.000.000 adet olduğu Bitcoin sayısının yıllara göre artışı gösterilmektedir

verilirse verilsin, çıktısındaki sonuç değeri hep aynı uzunlukta ve benzersiz bir değer olmaktadır [44].

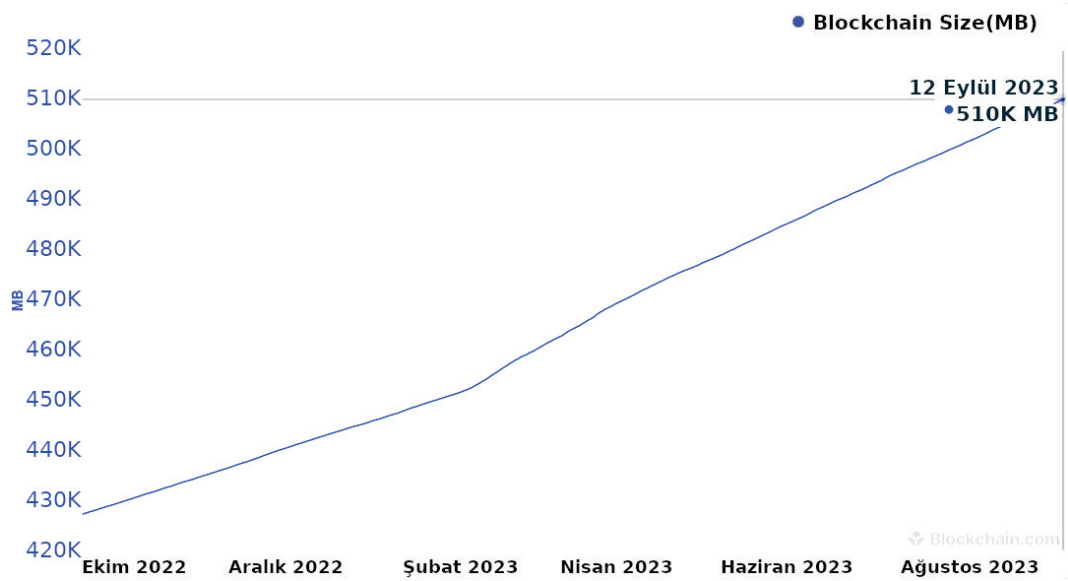
Bitcoin bloklarında kullanılan SHA-256 hash algoritması işlemi ile üretilen hash sayısı, bloktaki verinin, girdinin boyutundan bağımsız olarak 64 karaktere(256 bit) indirgenmesidir [45]. Her bloğa ait gerçekleşen transfer işlemlerinin hesaplanan hash sayısı blokların birbirine bağlanmasını sağlar. Önceki bloğun SHA-256 hash fonksiyonu çıktısı bir sonraki bloğun başına yazılır. Örnek blok Şekil 1.2'ta verilmiştir. Burada ilk satır önceki hash değeridir. Sonra işlemler listelenmiştir. Son satırda ise nonce (Number Only Used Once) 51 olarak seçilmiştir. Yeni blok içindeki veriler (bir önceki bloğun hash sayısı, nonce sayısı ve o bloktaki verinin tamamı) SHA-256 ile hash işlemine girdiğinde EC8D93C35617716F5B76408690D0667F2E059876BEE8A9CEF1651A26B01AC20B çıktısı üretecektir. Blok zincirinin devam etmesi için bu değer bir sonraki blok başına yazılarak devam edilmektedir.

EEE62B2F5FC81FFD94DC44EEFD045902FBE105AD265519A8DBAAE437DC4DFF72

**Mustafa'dan Ahmet'e 5 BTC gönderildi.**  
**Ahmet'ten Selim'e 3 BTC gönderildi.**  
**Ahmet'ten Osman'a 2 BTC gönderildi.**  
**Nonce=51**

Şekil 1.2. Örnek hash blok gösterilmiştir.

Bitcoin gibi kripto paralarda güvenliği artıracak şekilde fazladan prosedürler istenilmektedir. Bir önceki hash değeri ve bütün işlemlerin yanı sıra Bitcoin blok zincirinde, her yeni hash girişiminde kullanılan ve her seferinde değiştirilen rastgele bir sayı olan nonce, hash işlemi öncesi talep edilmektedir. Nonce değeri öyle bir değer olmalı ki blok hash çıktısının ilk 4 basamağı 0 olmalıdır. 22 Eylül 2023 tarihi itibarıyla Bitcoin işlemlerinde hesaplama zorluğunun artışı nedeniyle sıfır sayısı artmıştır ve blok, 0000000000000000000000000000000034dc521e2b9d0654ef4d6d74d5ad86265ae138dfd0a61 hash çıktısı üretmiştir [46]. 10 dakikada 1 blok bulunması hedeflenen Bitcoin blok zinciri teknolojisi bu zorluğu otomatik olarak ayarlamaktadır. Madencilik arttıkça ve azaldıkça istenilen hedef 0(sıfır) değerlerinin basamak sayısı otomatik olarak değiştirilebilir.



Şekil 1.3. Bitcoin blok zinciri boyutunun yıllara göre büyümesi.

Bitcoin madencilği olarak adlandırılan hash hesaplama sürecinde verilen ödüller her seviyesi için 210,000 blok üretilecek ve bu üretilen blokların daha sonra ödülü sürekli yarıya düşecektir. Alınan ödül Bitcoinler 4 yılda bir yarıya inmektedir [10, 47]. Son yarılama 11 Mayıs 2020 tarihinde 630.000'inci blokta gerçekleşmiş ve ödül 6,25 BTC'ye düşmüştür. Bitcoin sayısının yıllara göre artış hızının giderek azalması ve limitin 21.000.000 adet olduğu Nakamoto tarafından belirlenmiştir [10]. Bu ifadenin detayı  $10.500.000 + 5.250.000 + 2.625.000 + 1.312.500 + \dots$  şeklinde devam eden bir toplama işlemidir. 50 BTC üzerinden ödüllendirilen 210.000 blok + 25 BTC üzerinden ödüllendirilen 210.000 blok + ... şeklinde devam eden toplamlar şeklindedir. Şekil 1.1'de grafik olarak bu ifade gösterilmektedir.

Blokların birleşimiyle oluşan defter bir blok zinciridir. İşlem olup olmadığına bakılmaksızın yeni bloklar zincire eklendiği için blok zinciri her geçen gün giderek büyümektedir. Yakın zamanda petabayt boyutuna ulaşması tahmin edilmektedir. Mevcut blockchain boyutu 31.12.2022 itibariyle 435 GB'a ulaşmıştır. Bitcoin blok zincirinin büyümesi Şekil 1.3'de gösterilmektedir.

Zincirdeki ilk blok olan "Genesis Blok", zincirin başlangıç bloğu olduğu ve bu nedenle kendinden önce başka bir blok olmadığı için içeriğinde önceki hash çıktısı bulundurmaz [48]. Şekil 1.4'te blok zincirindeki blokların birbirine nasıl bağlı olarak yerleştirildiği gösterilmiştir.

### 1.1.1.2. Fikir Birliği Mekanizmaları (Konsensus)

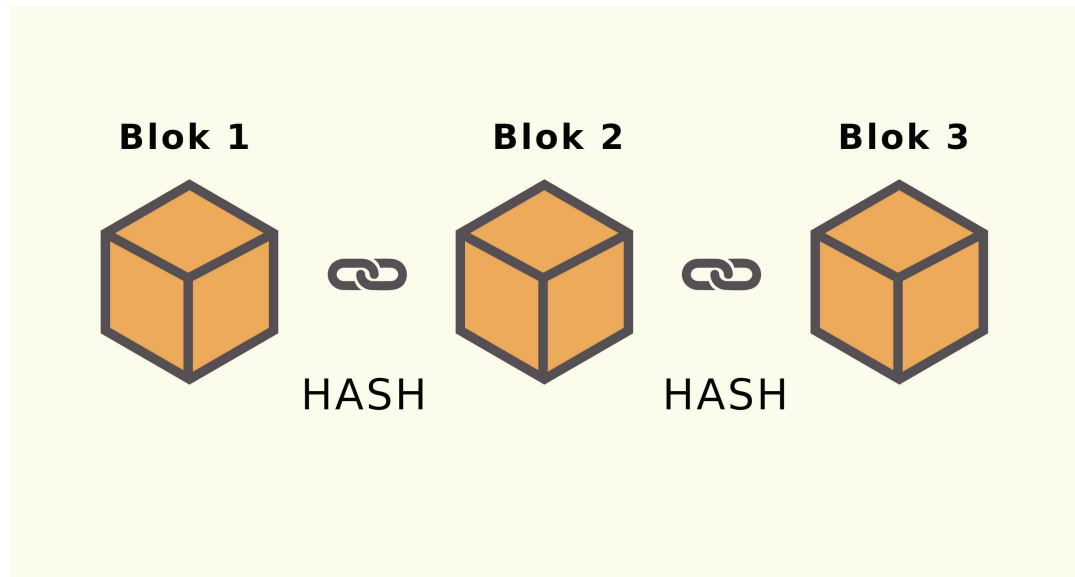
Blok zincirinde yapılan bir işlemi doğrulamak ve aynı işlemin tekrarının önüne geçmek için kontrolü sağlayan bir fikir birliği mekanizmasına ihtiyaç duyulmaktadır. Blok zinciri teknolojisinde birden fazla doğrulama mekanizması mevcuttur. Fikir birliği mekanizmaları literatürde mutabakat, doğrulama ve konsensüs terimleri ile ifade edilmektedir. Ağa veri ekleme işlemi doğrulayıcı düğüm olarak adlandırılan madenciler tarafından yapılmaktadır. Madenciler hash bulma problemini güçlü makineler ile çözmeye çalışır ve bu problemi ilk çözen madenci bloğu ağa ekler ve ödül olarak bitcoin alır. Bu yöntem iş ispatı (proof of work(PoW)) uzlaşma mekanizması olarak adlandırılmaktadır. Aynı anda farklı madenciler tarafından bloklar bulunabilir. Bu durumda yeni bloklar uzun olan blok yönüne devam



etmelidir. Burada, güçlü makina havuzuna sahip sistemlerin ağı domine etmesi ve diğer madencilerin ödül alamaması durumu ile karşılaşmaktadır. PoW uzlaşma mekanizmasında blok bulmak için yüksek hızda işlemler yapan madenciler enerji tüketimi gerçekleştirmektedirler. Bitcoin'in değeri ve üretim maliyeti olan elektrik tüketimine oranla karlılığı Sedlmeir(2021) tarafından incelenmiş ve enerji tüketiminin günümüzde çok fazla olduğu dile getirilmiştir [49]. Bitcoin fiyatının en yüksek değere sahip olduğunda bitcoin üretmek için kullanılan dünya genelindeki elektrik tüketimi, gerekli olan 50TWh'tan 4 kat daha fazla olduğu ve yaklaşık 200TWh olduğu tahmin edilmektedir [50].

2022 yılında Ethereum ağının geçtiği hisse kanıtı (proof of stake (PoS)) ise kullanıcılara, sahip oldukları kripto paraların miktarına göre yeni blokları oluşturma hakkı verir. Yani, PoS'ta ödüllendirme işlemleri madencilik gücüne değil, kripto paraların sahipliğine dayalıdır.

Son yapılan araştırmalarda uygunluk kanıtı (proof of optimization (PoO)) konsensüs modelinin blok oluşturma süresi farklı veri kümesi çözümlerine göre 2 s ile 60 s arasında değiştiği gösterilmiştir. PoO'nun analiz sonuçlarına göre, blok zinciri içinde madenciler arasındaki blok oluşturma oranının daha istikrarlı olduğu ve adalet indeksinin ortalama olarak 0.90'ın üzerinde olduğu görülmüştür. Yüksek donanım



Şekil 1.4. Blok zincirindeki blokların birbiri ile ilişkisi.



Şekil 1.5. Bitcoin/\$ fiyatının yıllara göre değişimi [2].

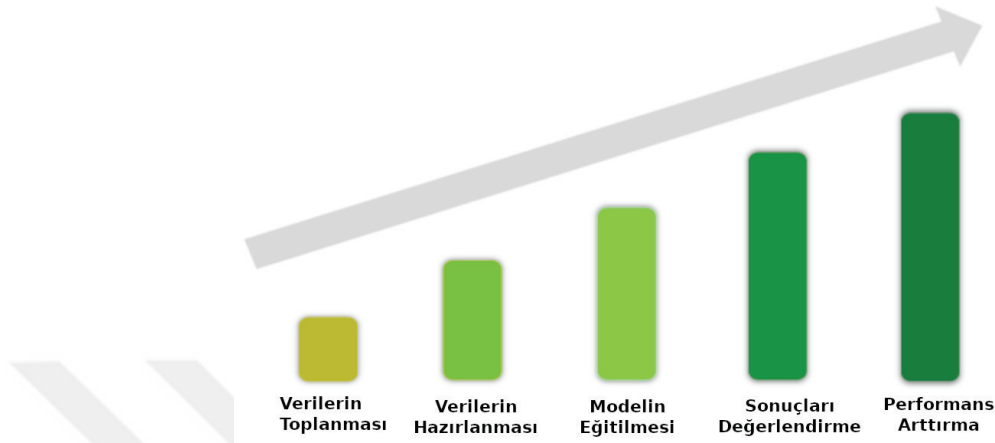
sahip uçların ağı domine etmesini önlemek ve düşük donanımlı uçların blok zinciri içinde bloklar oluşturma hakkını sağlamak amacıyla PoO önerilmektedir [51].

### 1.1.1.3. Bitcoinin Değeri

Geçen binlerce yılda paranın değışımi, temelde altının daha hızlı, daha güvenli, daha ekonomik kullanılabilir bir şeklinin olduğu öne sürülerek gerçekleşmektedir. Ancak, ikinci Dünya Savaşı sonrası ABD doları küresel rezerv para birimi statüsü edinmiş ve bu konum farklı dönemlerdeki farklı uyarlamalarla günümüze kadar gelmiştir [52]. Mevcut durumda, finans piyasalarında USD olarak değerlemeler yapılmaktadır. 1 Ocak 2009 yılında 0.001 USD değer ile başlamış, Şekil 1.5'te gösterildiği gibi 2021 yılında, bitcoin fiyatı yaklaşık 70.000 USD'ye ulaşmıştır. Bitcoin fiyatının enerji tüketimine etkisinin olduğu görülmektedir. Diğer taraftan enerji tüketiminin maliyetlerinin bitcoin fiyatı üzerinde etkisi araştırılmaktadır [53].

## 1.2. Makine Öğrenmesi

Ordinaryüs Profesör Cahit Arf'ın 1959 yılında makineler düşünebilir mi, nasıl düşünebilir? [54] konuşmasına bir çok makaleler yazılmıştır. Makine öğrenmesi konusunda yapılan bütün çalışmalarla makinelerin düşündüğü ve öğrendiği belirtilmektedir. Eğitim, amaçlara yönelik yetiştirme süreci olup öğrenme ise kalıcı değişiklik anlamına gelmektedir. Bu durumda makine öğrenmesi için bir amaç gereklidir. Amaç seçildiğinde uygun yöntemlerle önce eğitim başlar ve bu eğitim



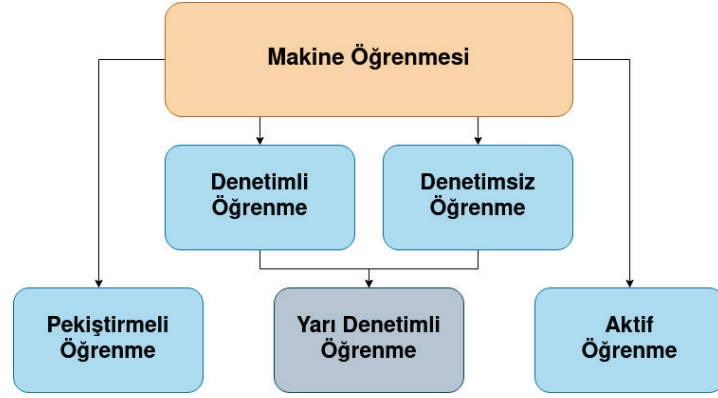
Şekil 1.6. Makine öğrenmesi yöntemlerinin ortak işlem adımları.

sırasında öğrenme gerçekleşir [55].

Makine öğrenmesi algoritmaları, ilgili algoritmaya bağımsız değişkenler üzerinden bağımlı değişkeni tahmin etme gibi bir yaklaşıma sahiptir [56]. Diğer bir deyişle geçmiş deneyimlerden faydalanarak model öğrenerek geleceğe yönelik tahmin yapmaktadır [35]. Bilgisayar mevcut geçmiş verilerden yola çıkarak bir model geliştirir ve bu modeli kullanarak bir tahmin veya karar verir [57]. Bu nedenle makine öğrenmesi, yapay zekanın bir alt uygulamasıdır. Makine öğrenmesi, veriler üzerinde istatistiksel işlemler yaparak bu verilerden öğrenme yapabilen, kararlar alabilen ve tahminlerde bulunabilen sistemler geliştirme alanıdır. Veri madenciliği, yapay zeka ve istatistik gibi alanlarla birlikte çalışır. Şekil 1.6'de makine öğrenmesi yöntemlerinin işlem adımları gösterilmiştir.

### 1.2.1. Makine Öğrenmesi Yöntemlerinin Kategorileri

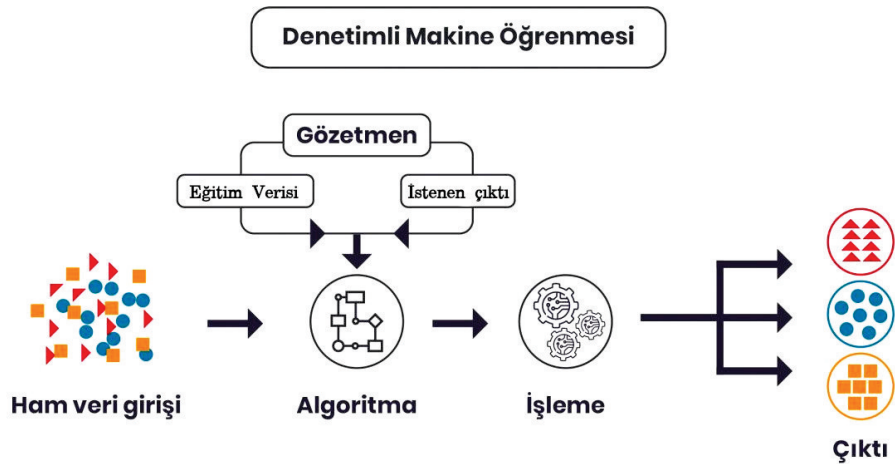
Makine öğrenmesi yöntemleri uygulandığı problemin amaçlarının ve elde edilen çıktılarının farklılığından dolayı alt sınıflar altında incelenmektedir. Şekil 1.7'deki öğrenme çeşitleri olan denetimli öğrenme (supervised learning), denetimsiz öğrenme (unsupervised learning) ve takviyeli öğrenme (reinforcement learning) şeklinde üç ana gruba ayrılan makineli öğrenme alanında daha sonra yarı denetimli öğrenme (semi-supervised learning), aktif öğrenme (active learning) kategorileri de eklenmiştir [58–62].



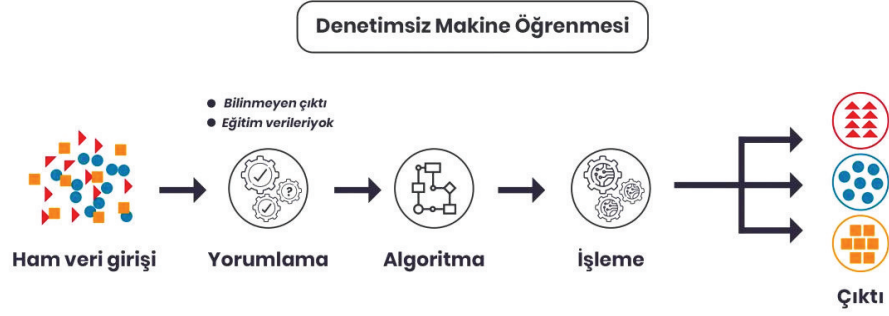
Şekil 1.7. Makine öğrenmesi yöntemlerinin çeşitleri.

### 1.2.1.1. Denetimli Öğrenme

Önceden sınıfı bilinen birimlerin açıklayıcı değişkenlerle verinin çalışma alanına göre sınıflandırılması veya regresyon amacıyla veri seti iki kısma ayrılmaktadır. Eğitim aşamasında modelin parametrelerinin tahmini yapılırken, test aşamasında yeni elde edilecek verilerin doğru tahmin edilip edilmediğine dair performans kriterleri hesaplanmaktadır. Bu nedenle, bu yöntemlere “denetimli öğrenme yöntemleri” adı verilmektedir. SVM, RF, ANN, vb. yöntemler en bilinen denetimli makine öğrenmesi teknikleridir [63]. Şekil 1.8’te denetimli makine öğrenmesi gösterilmiştir. Gözetmen denetiminde makine öğrenmesi gerçekleştirilmektedir.



Şekil 1.8. Denetimli makine öğrenmesi çalışma şekli [3].



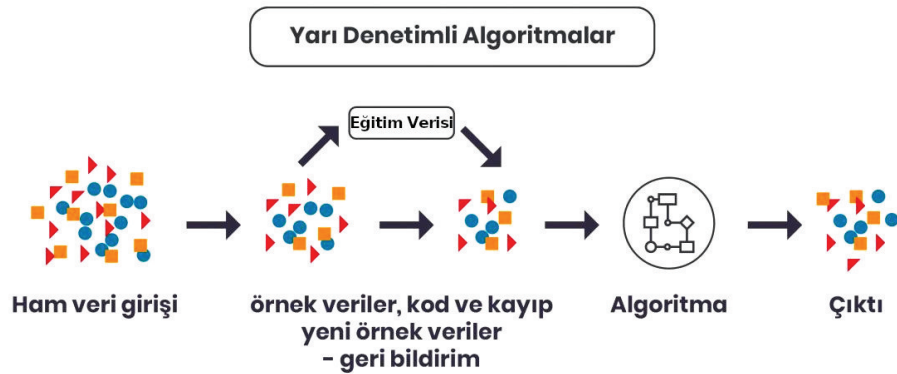
Şekil 1.9. Denetimsiz makine öğrenmesi çalışma şekli [3]

### 1.2.1.2. Denetimsiz Makine Öğrenmesi

Denetimsiz makine öğrenmesi, etiketsiz verilerin kullanılarak sistemin eğitilmesini sağlamaktadır. Denetimli öğrenmeye kıyasla verilerin sadece sınıflandırılması değil özellik setini de alır [64]. Şekil 1.9'te denetimsiz olarak üretilen veriler ışığında tekrarlanan döngülerle makine öğrenmesi gösterilmektedir.

### 1.2.1.3. Yarı Denetimli Makine Öğrenmesi

Yarı denetimli makine öğrenmesi, işaretlenmiş az sayıda veri ve işaretlenmemiş çok sayıda veri olduğu zaman kullanımı tercih edilmektedir. Bu yöntem, az sayıdaki işaretlenmiş verileri kullanarak çok sayıdaki işaretlenmemiş veriyi tahmin



Şekil 1.10. Yarıdenetimli makine öğrenmesi çalışma şekli [3].

ve sınıflandırma yapmada kullanılmaktadır. Yarı denetimli makine öğrenmesi ne kadar deneme yapabildiğini ve hatayı öğretmektedir. Birçok kez deneme yapmasının sebebi geçmiş deneyimlerden öğrenerek en iyi performansı elde etmesidir [65]. Şekil 1.10'te yönetici tarafından geliştirilen örnek veriler ve denetimsiz olarak üretilen veriler ışığında tekrarlanan döngülerle makine öğrenmesi gösterilmektedir. Şekil 1.10'te yarı denetimli makine öğrenmesi gösterilmektedir.



## 2. BÖLÜM

### 2.METOTLAR

#### 2.1. Veri Madenciliği

Problem dahilinde elde edilen ham veriyi azaltma, çoğaltma, temizleme, zenginleştirme ve sonrasında veriyi analiz ederek çıkarımda bulunma, sınıflandırma ve tahmin yapma kavramlarının tamamı veri madenciliği olarak adlandırılmaktadır. Veri Madenciliğinde, çalışma alanını ve çalışmanın amacını problemin türü belirler. Seçilen makine öğrenmesi yönteminin çalışma alanı problemin türüne göre değişir. Veri madenciliği, analiz edilerek bilgi çıkarılması zor olan büyük veri kümelerini analiz etmek ve anlamlı ve faydalı olabilecek bilgi çıkarmak; bir model oluşturarak yeni bir veri kümesi hakkında yorum yapmayı ve bu veri hakkında tahminde bulunmayı amaçlayan yöntemler bütünüdür [66]. Bunun yanı sıra kümeleme, genellikle denetimsiz öğrenme ile verilerin etiketlenmesini sağlamak amacıyla kullanılmaktadır. Zaman serilerinde analiz yöntemleri tanımlayıcı ve keşfedici analiz [67], eğri uydurma [68], fonksiyonel yaklaşım [69], tahmin ve öngörü [70], sınıflandırma [71], segmentasyon [72] ve işaret kestirimi [73] olmak üzere alt gruplara ayrılmaktadır. Makinalı öğrenme yöntemleri olarak ANN [74], SVM [35, 75], bulanık mantık [76, 77], genetik programlama [78] [79], metasezgisel algoritmalar [80, 81], saklı Markov modeli [82], karar ağaçları [83], RF [15] gibi yöntemler kullanılmaktadır. Genel olarak, veri madenciliği ile problemin türüne göre bazı durumlarda sınıflandırma yapılması, bazı durumlarda regresyon yapılması amaçlanır. Makine öğrenmesi ve veri madenciliği alanları iç içe geçmiş durumdadır. Veri madenciliği problem alanını tarif ederken makine öğrenmesi yöntemleri amaçlar doğrultusunda kullanılır. Bundan dolayı bu başlıkta incelenmiştir.

## 2.2. Makine Öğrenmesi Algoritmaları

Denetimli makine öğrenmesi için Bayes yöntemi, karar ağaçları, SVM, ANN, k-NN, RF ve derin öğrenme gibi yöntemler kullanılmaktadır. Bu çalışmada RF ve XGBoost regresyon analizi için kullanılmıştır.

### 2.2.1. Rastgele Orman(RF) ve Karar Ağacı

RF algoritması denetimli bir sınıflandırma veya regresyon algoritmasıdır. Algoritma, rastgele olarak bir orman oluşturur ve bu ormandaki ağaç sayısı ve elde edebileceği sonuç arasında doğrudan bir ilişki bulunmaktadır. Ağaç sayısı arttıkça doğruya yakın bir sonuç elde edilir [84]. RF algoritması, hızlı eğitim, mükemmel performans ve her türlü datayı elde edebilmek için büyük esnekliği olan popüler makine öğrenme algoritmalarından biridir [85,86].

Rastgele bir orman, ağaç yapılı bir koleksiyondan hesaplanan bir sınıflandırıcıdır. Sınıflandırıcılar Eşitlik (2.1)'deki gibi ifade edilir. Burada  $\Theta_k$  bağımsız ve özdeş dağıtılmış rastgele örneklerdir. Ormandaki her ağaç,  $x$  girdisindeki en popüler sınıf için bir birim oy kullanır [85,87].

$$h_k(x, \Theta_k), k = 1, \dots \quad (2.1)$$

RF algoritması, eğitim verisi, eğitim etiketleri, ağaç sayısı ve özellik sayısı gibi girdi parametrelerini alır. Algoritma 2.1 RF algoritmasının, Algoritma 2.2 karar ağacı algoritmasının temel mantığını göstermektedir. İlk olarak, boş bir "ağaçlar listesi" tanımlanır. Ardından, belirtilen ağaç sayısı kadar döngü yapılır. Her döngü adımında, rastgele seçilen özellikler ve örnekler kullanılarak yeni bir karar ağacı çıkartılır. Yeni ağaç, ağaçlar listesine eklenir. Döngü tamamlandığında, ağaçlar listesi rastgele orman algoritmasının çıktısı olarak döndürülür.

RF algoritması içerisinde tekrarlı biçimde çağrılan karar ağacı çıkartma işleminin algoritması ise şu şekildedir; Karar ağaçları madencilik sırasında pek çok test gerçekleştirerek, bağımlı değişken değerini (hedefi) tahmin etmek için en iyi sırayı keşfetmeye çalışır. Yapılan her bir test karar ağacındaki dalları oluşturur ve diğer testlerin daha sonra bu dallara eklenebilmesi için alt yapı elde edilmiş olur [88].



Karar ağaçları, örnek özniteliklerini girdi parametresi olarak alır. İlk olarak, boş bir ağaç tanımlanır. Durdurma koşulu kontrol edilerek sağlanmadığı durumda en iyi özelliği seçer ve eldeki eğitim örneklerini bu özelliğe göre farklı dallara böler. Eğer durdurma koşulu sağlanıyorsa, Ağaç'a yaprak düğümleri eklenerek ağaç geri döndürülür. Durdurma koşulu sağlanmıyorsa sağlanıncaya kadar ağaç üzerinde yeni bir iç düğümü eklenir, sol ve sağ alt ağaçlar için özyinelemeli olarak karar ağacı algoritması tekrar uygulanır. İşlem tamamlandığında ağaç geri döndürülür.

```

Veri: EğitimVerisi, EğitimEtiketleri, AğaçSayısı, ÖzellikSayısı
Ağaçlar  $\leftarrow []$ 
for  $i = 1$  to AğaçSayısı do
  ÖrnekSayısı  $\leftarrow$  EğitimVerisi.örnekSayısı
  Özellikler  $\leftarrow$  RastgeleSeçilenÖzellikler(ÖzellikSayısı)
  Örnekler  $\leftarrow$  RastgeleSeçilenÖrnekler(EğitimVerisi, ÖrnekSayısı)
  Ağaç  $\leftarrow$  KararAğacı(Örnekler, Özellikler)
  Ağaçlar.ekle(Ağaç)
end for
return Ağaçlar

```

**Algoritma 2.1:** RF Algoritması çalışmasının temel mantığı.

```

Girişler:  $D$ : Eğitim veri kümesi.
            $a$ : Öznitelik değerleri.
Çıkış :  $KararAğacı$ : Keşfedilen karar ağacı modeli.
Veri:  $m$ : Öznitelik sayısı.
        $KararAğacı = \{ \}$ 
       Temel durumları kontrol et
       for ( $i = 1 \dots m$ ) do
          $a_i$  için normalleştirilmiş bilgi kazancını hesapla.
       end for
       En yüksek bilgi kazancını veren  $a_{best}$  ile düğüm ekle.
       Yapılan dallanma üzerinde alt kümelere aynı biçimde yeni düğümler ekle.
       Karar ağacı üretildikten sonra budama işlemi gerçekleştir.
return  $KararAğacı$ 

```

**Algoritma 2.2:** Karar Ağacı.

### 2.2.2. Aşırı Gradyan Artırma (XGBoost) Algoritması

XGBoost algoritması, gradyan artırma algoritmasının (Gradient Boosting Machine-GBM) iyileştirilmesi ile elde edilmiştir. Aşırı öğrenmenin önüne geçebilmesi, boş verileri yönetebilmesi, hızlı olması ve en önemlisi yüksek tahmin gücüne sahip olma

Tablo 2.1. XGBoost işlem adımları.

İşlem Adımı	Açıklama
Adım 1	FOR döngüsü kullanılır ve iterasyon değişkeni 1'den başlayarak iterasyonSayisi'na kadar artırılır.
Adım 2	Her bir iterasyonda, negatifGradyanlar fonksiyonu kullanılarak negatif gradyanlar hesaplanır.
Adım 3	Ardından, KararAgaci fonksiyonu kullanılarak bir karar ağacı oluşturulur ve bu ağaç ağaçlar listesine eklenir.
Adım 4	Sonraki adımda, ağaçlarıTahminEt fonksiyonu ile oluşturulan tüm ağaçlar kullanılarak tahminler yapılır.
Adım 5	Elde edilen tahminler, önceki tahminlere öğrenmeOrani ile çarpılarak güncellenir.
Adım 6	FOR döngüsü tamamlandığında, ağaçlar listesi geri döndürülür.
Adım 7	Bir sonraki adımda, negatifGradyanlar fonksiyonu tanımlanır Bu fonksiyon, etiketler ve tahminler arasındaki gradyanları hesaplar.
Adım 8	KararAgaci fonksiyonu tanımlanır. Bu fonksiyon veri, artıklar ve maksimum derinlik parametrelerini alır.
Adım 9	Fonksiyon içinde, maksimum derinlik kontrol edilir eğer derinlik 0 ise veya tüm veri noktaları aynı sınıfa sahipse bir yaprak düğümü oluşturulur.
Adım 10	Eğer bu koşullar sağlanmazsa, en iyi bölme noktası ve kazancı bulunur veri bölünür ve sol ve sağ alt ağaçlar oluşturulur.
Adım 11	Fonksiyon, oluşturulan düğümü geri döndürür.
Adım 12	ağaçlarıTahminEt fonksiyonu tanımlanır Bu fonksiyon, oluşturulan tüm ağaçları ve veriyi alır.
Adım 13	Fonksiyon içinde, her bir ağaç için tahminler yapılır sonuçlar toplanır.
Adım 14	Tahminler sonuç olarak geri döndürülür.

nedenlerinden dolayı diğer denetimli öğrenme algoritmalarına göre avantaja sahip bir yöntemdir. XGBoost başarılı tahmin gücü yanında hesaplama kolaylığına da sahiptir. Bu hesaplama kolaylığı ile çok boyutlu veri analizlerinde kullanılmaktadır [89]. Algoritma 2.3'de XGBoost algoritmasının kaba kodu verilmiştir. Algoritmanın adımları Tablo 2.1'de belirtilmiştir.

### 2.3. Regresyon Analizi

Regresyon, metotlar arasında iki ya da daha çok değişken arasındaki ilişkiyi modelleyen en yaygın metottur. Bağımlı değişkenin tahmin edilmesi için kurulan model giriş olarak tek bağımsız değişken kullanılıyorsa tekli regresyon, birden fazla

**Fonksiyon XGBoost(egitimVerisi, egitimEtiketleri, iterasyonSayisi, ogrenmeOrani, maksDerinlik)**

**Başla**

egitimVerisiDMatrix = DMatrix(egitimVerisi, etiket=egitimEtiketleri)

agaclar = []

oncekiTahminler = sifirDizisi(egitimEtiketleri.uzunluk)

**for** iterasyon = 1 to iterasyonSayisi **do**

    artik = negatifGradyanlar(egitimEtiketleri, oncekiTahminler)

    agac = KararAgaci(egitimVerisiDMatrix, artik, maksDerinlik)

    agaclar.ekle(agac)

    yeniTahminler = agaclariTahminEt(agaclar, egitimVerisi)

    oncekiTahminler = oncekiTahminler + ogrenmeOrani \* yeniTahminler

**end for**

**return** agaclar

**Fonksiyon negatifGradyanlar(etiketler, tahminler)**

**Başla**

gradyanlar = []

**for** her bir etiket, tahmin **do**

    gradyan = -2 \* (etiket - tahmin)

    gradyanlar.ekle(gradyan)

**end for**

**return** gradyanlar

**Fonksiyon KararAgaci(veri, artik, maksDerinlik)**

**Başla**

**if** maksDerinlik = 0 veya dugumlerdeki tum veri noktaları aynı sınıfa sahipse

**then**

    yeniDugum = YaprakDugumu(artik)

**else**

    eniyeBolmeNoktasi, eniyiBolmeKazanci = eniyiBolmeNoktasiniBul(veri, artik)

    solVeri, sagVeri = veriyiBol(veri, eniyiBolmeNoktasi)

    solAltAgaci = KararAgaci(solVeri, solVeri.olcut, maksDerinlik - 1)

    sagAltAgaci = KararAgaci(sagVeri, sagVeri.olcut, maksDerinlik - 1)

    yeniDugum = IcDugumu(eniyiBolmeNoktasi, solAltAgaci, sagAltAgaci)

**end if**

**return** yeniDugum

**Fonksiyon agaclariTahminEt(agaclar, veri)**

**Başla**

tahminler = sifirDizisi(veri.uzunluk)

**for** her bir agac **do**

    agac = agaclar[i]

    tahminler = tahminler + ogrenmeOrani \* agaciTahminEt(agac, veri)

**end for**

**return** tahminler

**Algoritma 2.3:** XGBoost Algoritması kaba kodu.

ise çoklu regresyon analizi olarak adlandırılır. Bilinen normal dağılan sayısal bir değişken kullanılarak bilinmeyen ve aralarında ilişki olan başka bir normal dağılan sayısal değişkeni tahmin için uygulanırsa basit doğrusal regresyon (simple linear regression), birden fazla değişken kullanılarak bir değişkeni tahmin etmek amacıyla modelleme yapıldığında ise "çoklu doğrusal regresyon" (multiple linear regression) olarak tanımlanır [90–92]. Çoklu doğrusal regresyonda  $n$  adet bağımsız değişken ( $X_1, X_2, X_3 \dots X_n$ ) ile bağımlı değişken ( $Y$ ) arasındaki doğrusal ilişki eşitlik (2.2)'deki gibidir [93].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (2.2)$$

Burada  $Y$  bağımsız değişkeni,  $X_n$  bağımlı değişkenlerin doğrusal bir fonksiyonu olarak ele alınır. Eşitlikteki  $\beta_n$  değerleri modeldeki katsayılar olup,  $\beta_0$  doğrunun  $Y$  eksenini kestiği noktayı göstermektedir. Eşitlikteki  $\epsilon$  ise hata terimi olarak tanımlanmıştır. Buradaki katsayılar en küçük kareler yöntemi ile bulunmaktadır.

Rastgele orman regresyonu, polinom regresyon [94], lojistik regresyon [95], ridge ve lasso regresyonu [96], elasticnet regresyon [97], poisson regresyon [98], gamma regresyon [99], çoklu çıktı regresyonu [100] modelleri de literatürde mevcuttur.

### 2.3.1. Rastgele Orman Regresyonu (RFR)

RFR tahmin konusunda makine öğrenmesi yöntemleri arasında regresyon problemlerinde en iyi sonuçları veren topluluk algoritmaları arasındadır [101].

$K$  adet ağacın tahmin değerlerinin ortalamasıyla modelin nihai tahmin sonucunu vermektedir. Modelin tahmin sonucu şekil (2.3)'daki gibi ifade edilmektedir.

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^N \hat{y}_i \quad (2.3)$$

### 2.3.2. Otomatik Regresyon (AR)

Otomatik regresyon, seçilen regresyon modelinin makine öğrenmesi alanlarında kullanılmasıyla regresyonun otomatik olarak belirlendiği veya ayarlandığı bir süreç olarak ifade edilebilir.

Zaman serisinin kendi önceki değerlerine olan bağımlılığını belirleyen parametreleri vardır ve otomatik regresyon modellerinde, lag parametresi, kaç önceki gözlemin

modele dahil edileceğini belirler. Örneğin, Lag(1) modelinde, sadece bir önceki gözlem değeri, mevcut gözlemi tahmin etmek için kullanılır. Lag(2) modelinde, hem bir önceki hem de iki önceki gözlem değeri, tahmin için kullanılır. Genel olarak, AR(p) modelinde, p tane önceki gözlem değeri, tahmin için kullanılır. [102]

Lag parametresi, modelin doğruluğunu ve performansını etkileyen önemli bir parametredir ve uygun bir lag parametresi seçimi, modelin doğru tahminler yapabilmesi için önemlidir.

$$Y_i = \beta_0 + \beta_1 X_1 + \epsilon \quad (2.4)$$

Eşitlik (2.4)'de yer alan denklemde  $Y_i$  bağımlı değişkeni,  $X_1$  bağımsız değişkeni ifade etmektedir.  $\beta_0$  ve  $\beta_1$  regresyon katsayılarını göstermektedir.  $\beta_0$  sabit katsayı terimidir.  $\epsilon$  değeri ise hatayı ifade etmektedir.

Regresyon modelinin parametrelerini belirlemek için optimizasyon yöntemleri kullanılabilir. Burada hangi yöntemin kullanılacağı, regresyon analizindeki özel duruma, veri setine ve hedefe bağlıdır. Her bir optimizasyon yönteminin avantajları ve dezavantajları vardır, bu nedenle uygulanacak yöntemin problem ve veri setine uygunluğu dikkate alınmalıdır.

En küçük kareler yöntemi (Least Squares Method) [103]: En yaygın kullanılan yöntemlerden biridir. Hataların karelerinin toplamını minimize ederek parametreleri hesaplar.

Gradyan inişi (Gradient Descent) [104]: Hata fonksiyonunu minimize etmek için iteratif bir optimizasyon yöntemidir. Parametre güncellemelerini, hata fonksiyonunun negatif gradyan yönünde yaparak ilerler.

İteratif ağırlık ayarlama (Iteratively Reweighted Least Squares - IRLS) [105]: Lojistik regresyon ve poisson regresyon gibi bağımlı değişkenin dağılımının belirli olduğu durumlarda yaygın olarak kullanılan bir yöntemdir. Tekrarlanan en küçük kareler yöntemi gibi çalışır, ancak hataların ağırlıkları iteratif olarak yeniden hesaplanır.

Genetik Algoritma [106]: Evrimsel hesaplama temel alınarak oluşturulan bir optimizasyon yöntemidir. Parametreleri genetik operatörler kullanarak popülasyon

üzerinde optimize eder.

## 2.4. Performans Ölçümleri

Yapılan çalışmaların doğruluğunun test edilmesi gerekmektedir. çeşitli performans ölçümleri kullanılarak modelin çalıştırılması sonucu alınan test çıktıları, gerçek değerleri ile kontrol edilerek sayısal bir değer hesaplanır. En küçük kareler ortalaması(MSE) [107] ve hata kareler ortalamasının karekökü (RMSE) [108] bu amaçla sıklıkla kullanılan performans ölçütüdür. Makine öğrenme yöntemlerinin başarılarını değerlendirirken hangi performans ölçütleri kullanılmalıdır [109] başlıklı yayında bu ve diğer yöntemler açıklanmıştır. RMSE, MSE'nin kareköküdür.

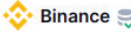


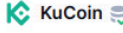


$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.5)$$

Eşitlik (2.5)'te RMSE formülü verilmiştir.  $x_i$ :  $i$ . gözlemlenen değeri,  $y_i$ : buna karşılık gelen tahmin edilen değeri,  $n$ : gözlemlerin sayısını ifade etmektedir.

## 2.5. Materyaller

Guido Van Rossum tarafından geliştirilen bir programlama dili olan Python, birçok uygulama için kullanılabilen genel amaçlı bir dildir [110,111]. Scikit-learn (sklearn), python programlama dilinde popüler ve güçlü bir makine öğrenimi kütüphanesidir. Scikit-learn kütüphanesi, çeşitli makine öğrenimi algoritmalarını, veri işleme ve model değerlendirme araçlarını içeren kapsamlı bir araç seti sunar. Çeşitli denetimli ve denetimsiz öğrenme algoritmaları olan sınıflandırma, regresyon, kümeleme, boyut indirgeme ve daha fazlası gibi çeşitli fonksiyonlar sunar. Örneğin, karar ağaçları, rastgele ormanlar, doğrusal ve lojistik regresyon gibi algoritmaları içerir [112]. Ta-lib, python için stok market verilerini analiz etmede kullanılan teknik analiz kütüphanesidir [113]. Matplotlib, python için yazılan grafik çizim kütüphanesidir [114].

Şekil 2.1.'de görüldüğü üzere tez hazırlanırken Binance borsasının dünyanın en büyük kripto para ticaret hacmine sahip olduğu görülmektedir. Volatilitenin temel sebebi Binance'dır [115]. Yüksek değişkenlik ve hacme sahip olması nedeniyle bu

# ▲	Borsa Adı	Skor	İşlem Hacmi(24s)	Ort.Likidite	Haftalık ziyaret
1	 Binance	9.9	\$8,592,939,532 ▼ 5.68%	860	13,540,566
2	 Coinbase Exchange	8.4	\$851,835,073 ▼ 32.49%	741	35,502
3	 Kraken	8.0	\$353,364,598 ▼ 35.27%	744	1,080,004
4	 KuCoin	7.6	\$596,108,168 ▼ 8.77%	568	2,190,836
5	 Bitfinex	7.2	\$124,458,209 ▲ 4.58%	607	525,612
6	 OKX	7.2	\$1,597,519,398 ▼ 11.35%	553	2,197,438

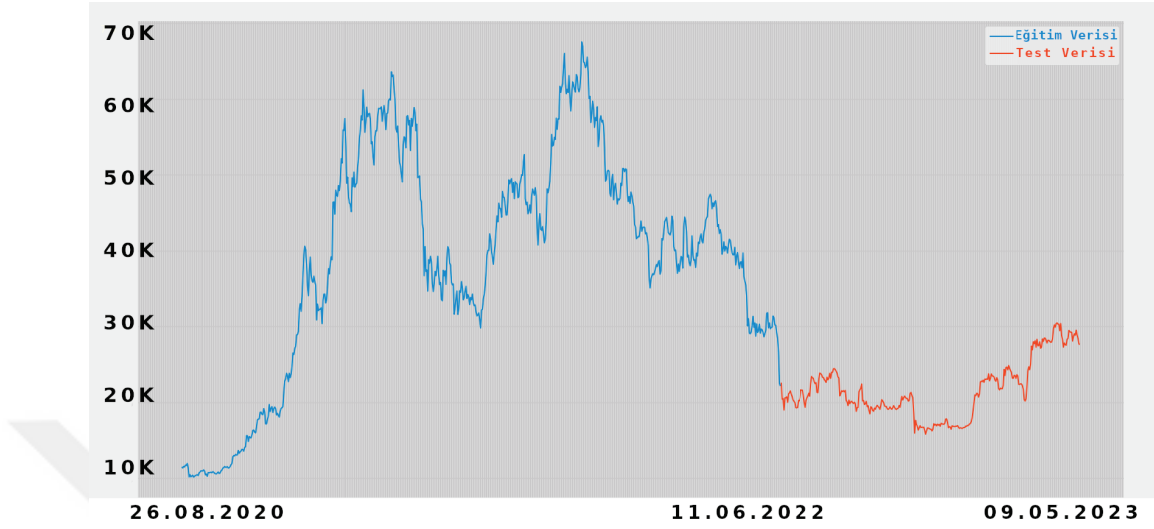
Şekil 2.1. 24 saatlik hacimlerine göre ilk 6 kripto para borsası [4] listelenmiştir (7.05.2023 itibariyle).

tezde veri kümesi Binance borsası üzerinden alınmıştır. Tezin amacı doğrultusunda fiyat tahminleme diğer çalışmalarda da genellikle kapanış fiyatları üzerinden yapıldığı için kapanış fiyatları üzerinden değerlendirilmiştir.

## 2.6. Veri kümesi

Veri kümesi olarak mum(candle) grafikler için sağlanan peryot içinde gerçekleşen sınır değerleri ifade eden ve açılış değeri (O), yüksek değer (H), düşük değer (L), kapanış değeri (C) ile işlem hacmi miktarı (V) değerlerinin bir arada tutulduğu OHLCV veri seti tercih edilmiştir. Mum grafikler, yatay ekseninde zamanı, dikey ekseninde ise fiyat verilerini gösterir ve bu parametrelerin herbirini gösterebildiği için daha fazla bilgi içerir. Varlığın belirli bir zaman diliminde ulaştığı en yüksek, en düşük fiyatı, açılış ve kapanış fiyatlarını toplu olarak gösterebilmektedir.

Binance borsasından BTC/USDT çiftinin 26.08.2020-9.05.2023 tarihleri aralığındaki 1 günlük peryotlu OHLCV verileri indirilmiştir. Şekil 2.2.'de Binance'dan indirilen geçmiş verilerin eğitim ve test için hangi aralıkların kullanıldığı gösterilmektedir. Veri kümesi [987,6] boyutlu dizinin, 26 Ağustos 2020 - 11 Haziran 2022 tarihleri arasındaki [657,6] boyutluk kısmı farklı özniteliklerle zenginleştirilerek eğitim için



Şekil 2.2. Tarihe göre bitcoin fiyatında 26.08.2020-11.06.2022 arası eğitim verisi ve 11.06.2022-09.05.2023 arası test verileri gösterilmiştir.

kullanılmıştır. Algoritmaların başarımları test olarak ayrılan 12 Haziran 2022 - 9 Mayıs 2023 tarihleri aralığındaki 330 günlük verilerle test edilmiştir. Verinin tahmin başarımlarını ölçmek için sadece kapanış verileri kullanılmış, tahminler için ise OHLCV verileri kullanılmıştır.

## 2.7. Veri Ön İşleme

Verilerin yetersizliği veya eksikliği doğru algoritmayı seçmede de farklı bir problem meydana getirebilir. Bu adımda veri temizleme [116], veri birleştirme [117], veri dönüşümü [117], veri azaltma yöntemleri [118] kullanılarak veri analize hazır hale getirilir. Bu işlemler oluşacak modelin başarımlarını etkileyebilir. Veri kümesi üzerinde yapılan farklı veri ön işlemleri farklı sonuçlara sebep olabilir [66]. Verilerin yetersiz olduğu durumlar için veri çoğaltma [119] teknikleri literatürde mevcuttur. Çalışma için yatay eksenindeki farklı boyutlara dayalı indikatör olarak adlandırılan hesaplama teknikleri ile öz nitelik çoğaltma yapılmıştır. Bu hesaplama teknikleri bir sonraki başlık altında detaylı olarak açıklanmıştır.

## 2.8. İndikatör hesaplamalarıyla veri kümesi zenginleştirilmesi

Finans piyasası terimi olarak "indikatör", fiyat hareketlerini farklı biçimlerde analiz ederek piyasa hareketlerini yorumlama aracı olarak kullanılan matematiksel



Tablo 2.2. Çalışmada kullanılan ortalama tabanlı indikatörler.

Parametre	Açıklama
CloseMA5	Son 5 serinin kapanış Hareketli Ortalaması(MA)
CloseMA10	Son 10 serinin kapanış fiyatlarının MA'sı
CloseMA20	Son 20 serinin kapanış fiyatlarının MA'sı
CloseMA60	Son 60 serinin kapanış fiyatlarının MA'sı
CloseMA120	Son 120 serinin kapanış fiyatlarının MA'sı
CloseMA200	Son 200 serinin kapanış fiyatlarının MA'sı
CloseEMA8	Son 8 serinin kapanış fiyatlarının EMA'sı
CloseEMA13	Son 13 serinin kapanış fiyatlarının EMA'sı
CloseEMA14	Son 14 serinin kapanış fiyatlarının EMA'sı
CloseEMA15	Son 15 serinin kapanış fiyatlarının EMA'sı
CloseEMA21	Son 21 serinin kapanış fiyatlarının EMA'sı
CloseEMA30	Son 30 serinin kapanış fiyatlarının EMA'sı
VolumeMA5	Son 5 serinin hacimlerinin MA'sı
VolumeMA10	Son 10 serinin hacimlerinin MA'sı
VolumeMA20	Son 20 serinin hacimlerinin MA'sı
VolumeMA60	Son 60 serinin hacimlerinin MA'sı
VolumeMA120	Son 120 serinin hacimlerinin MA'sı
VolumeMA200	Son 200 serinin hacimlerinin MA'sı
VolumeEMA5	Son 5 serinin hacimlerinin EMA'sı
VolumeEMA10	Son 10 serinin hacimlerinin EMA'sı
VolumeEMA20	Son 20 serinin hacimlerinin EMA'sı
VolumeEMA60	Son 60 serinin hacimlerinin EMA'sı
VolumeEMA120	Son 120 serinin hacimlerinin EMA'sı
VolumeEMA200	Son 200 serinin hacimlerinin EMA'sı
CloseEMA5	Son 5 serinin kapanış fiyatlarının EMA'sı
CloseEMA10	Son 10 serinin kapanış fiyatlarının EMA'sı
CloseEMA20	Son 20 serinin kapanış fiyatlarının EMA'sı
CloseEMA60	Son 60 serinin kapanış fiyatlarının EMA'sı
CloseEMA120	Son 120 serinin kapanış fiyatlarının EMA'sı
CloseEMA200	Son 200 serinin kapanış fiyatlarının EMA'sı
CloseDEMA5	Son 5 serinin kapanış fiyatlarının DEMA'sı
CloseDEMA10	Son 10 serinin kapanış fiyatlarının DEMA'sı
CloseDEMA20	Son 20 serinin kapanış fiyatlarının DEMA'sı
CloseDEMA60	Son 60 serinin kapanış fiyatlarının DEMA'sı
CloseDEMA120	Son 120 serinin kapanış fiyatlarının DEMA'sı

hesaplama ve göstergelerdir. Toplanan veri kümesi sıklıkla kullanılan teknik analiz indikatörleri ile çok boyutlu hale getirilmiştir.

Tablo 2.2'de Hareketli Ortalama(MA), Üssel Hareketli Ortalama(EMA), Çift Üssel Hareketli Ortalama(DEMA) tabanlı indikatörler, Tablo 2.3'de trend indikatörleri gösterilmiştir. Tablo 2.4'de hesaplanan oransal parametreler gösterilmiştir.

Tablo 2.3. Çalışmada kullanılan trend tabanlı indikatörler.

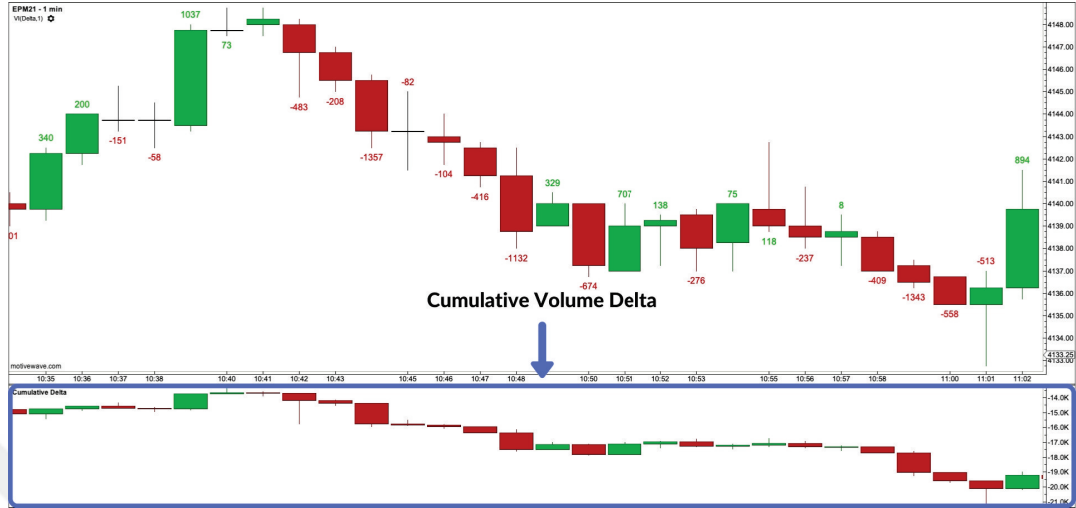
Parametre	Açıklama
CVD	Birikimli Hacim Farkı
ADX14	14 günlük ortalama doğrusal hareket endeksi
MACD	Hareketli ortalamanın yakınsaması-ıraksaması
RSI	Göreceli güç endeksi
BBANDS-U	Bollinger Bandları Üst
BBANDS-M	Bollinger Bandları Orta
BBANDS-L	Bollinger Bandları Alt
AD	Birikim ve dağıtım indikatörü
TR	Gerçek aralık
ATR	Ortalama Gerçek Aralık
MFI	Para Akış Endeksi
KDJ-K	KDJ osilatör göstergesi K parametresi
KDJ-D	KDJ osilatör göstergesi D parametresi
KDJ-J	KDJ osilatör göstergesi J parametresi

Hareketli ortalama (MA), belirli bir süre boyunca bir finansal varlığın ortalama fiyatını gösteren gecikmeli bir fiyat tabanlı göstergedir. MA basit hareketli ortalama (SMA) anlamında kullanılmaktadır. Peryotlardaki fiyatların toplamının periyot sayısına bölümü ile hesaplanmaktadır. Üstel hareketli ortalama (EMA) ise üç adımda hesaplanır:

- Basit Hareketli Ortalama (SMA) = (Toplam Peryot Fiyatları / Peryot Sayısı)
- Çarpan =  $2 / (\text{Peryot Sayısı} + 1)$
- (İlk EMA için, EMA(önceki gün) yerine önceki gün SMA kullanılır)  
EMA = Kapamış - EMA(önceki gün) x çarpan + EMA(önceki gün)

$DEMA=2*EMA-EMA(EMA)$  formülü ile hesaplanmaktadır.

Birikimli hacim farkı (CVD) bir gün boyunca her mumdan alınan fark (delta) toplamıdır. Bir varlık için likit hacim girişini ölçen indikatördür. CVD indikatöründeki "delta", mevcut piyasa hareketindeki eğilimi gösterir [120]. Şekil 2.3'de CVD indikatörü mum çubuklarında gösterilmiştir. Hesaplaması yapılan çubuğun yeşil renkte yükselen veya kırmızı renkte azalan olduğunun hesaplanması için kendisinden önce gelen mum çubuklarının hacim farklarının toplamından fark çıkarılır [120]. Artış veya azalış önceki farkların toplamı ile hesaplanır.



Şekil 2.3. Birikimli Hacim Farkı indikatörünü mum çubuklarıyla gösterilmiştir [5].

14 günlük ortalama doğrusal hareket indisi (ADXR14) eşitlik (2.6)'de formülü verilen ADX göstergesinin 14 günlük ortalamasını ifade eder. ADX fiyat hareketinin gücünü ölçmek için kullanılan bir göstergedir. ADX değeri, 0 ila 100 arasında bir değer alır. Yüksek bir ADX değeri, güçlü bir trendin varlığını gösterirken, düşük bir ADX değeri, trendin zayıf veya yok olduğunu gösterebilir. ADXR, ADX'nin pürüzsüzleştirilmiş bir versiyonudur. ADXR, ADX değerlerinin belirli bir süre boyunca ortalamasını alır ve daha pürüzsüz bir eğri oluşturur. Bu, fiyat hareketinin gücünü daha istikrarlı bir şekilde göstermeye yardımcı olur. Örneğin 14 günlük ADXR, son 14 günlük ADX değerlerinin ortalamasını hesaplar. Bu, trendin gücünü daha uzun bir dönemde değerlendirmek için kullanılabilir ve daha pürüzsüz bir gösterge sunar. Eşitlik (2.7)'de ADX göstergesinin 14 günlük ortalaması formülü gösterilmektedir.

$$ADX = \frac{1}{N} \sum_{i=1}^N \left( \frac{|DI_{i+1} - DI_i|}{TR_i} \right) \times 100 \quad (2.6)$$

$$ADXR_{14} = \frac{1}{14} \sum_{i=1}^{14} ADX_i \quad (2.7)$$

Hareketli ortalamanın yakınsaması-ıraksaması (MACD) indikatörü hesaplanmasında genellikle 12 günlük bir EMA ile 26 günlük bir EMA arasındaki fark kullanılır. Bu fark MACD hattını oluşturur. Ardından, MACD hattının 9 günlük bir EMA'sı alınır ve bu sinyal hattı olarak adlandırılır. Eşitlik (2.8)'de MACD formülü ve

eşitlik (2.10)'de sinyal hattı formülü verilmiştir. MACD ve sinyal hattı arasındaki kesişmeler, finansal yorumlarda alım-satım sinyallerini gösterdiği kabul edilmektedir. Örneğin, MACD hattının sinyal hattını yukarı yönde kesmesi bir alım sinyali olarak yorumlanabilir, aşağı yönde kesmesi ise bir satım sinyali olarak yorumlanabilir.

$$MACD = EMA_{12} - EMA_{26} \quad (2.8)$$

$$SignalLine = EMA_9(MACD) \quad (2.9)$$

Göreceli güç endeksi (RSI), teknik analizde yaygın olarak kullanılan başka bir göstergedir. RSI bir varlığın aşırı alım veya aşırı satım koşullarını belirlemek ve fiyatın gücünü ölçmek için kullanılır. RSI belirli bir dönemdeki yükseliş ve düşüşlerin karşılaştırılmasıyla hesaplanır. Genellikle 14 günlük bir dönem kullanılır. RSI değeri 0 ila 100 arasında bir değer alır. 70 ve üzeri bir RSI değeri aşırı alımı, 30 ve altında RSI değeri ise aşırı satımı ifade edebilir.

$$RSI = 100 - \frac{100}{1 + \frac{\text{Ortalama Yukarı Hareketli Kapanışlar}}{\text{Ortalama Aşağı Hareketli Kapanışlar}}} \quad (2.10)$$

Bollinger bantları (BBANDS), teknik analizde fiyatın volatilitisini ve potansiyel aşırı alım ve aşırı satım koşullarını belirlemek için kullanılan bir göstergedir. Bollinger bantları, orta, üst ve alt bant olmak üzere üç bileşenden oluşur. Orta bant, genellikle 20 günlük bir basit hareketli ortalama (SMA) olarak belirlenir. Üst bant ve alt bant ise orta banda belli bir standart sapma değeri eklenerek veya çıkarılarak oluşturulur. Standart sapma değeri, genellikle 20 günlük bir dönemdeki fiyat hareketlerinin volatilitisini ölçmek için kullanılır. Bollinger orta bandı eşitlik (2.11)'de, üst bandı eşitlik (2.12)'de, alt bandı eşitlik (2.13)'de gösterilmektedir. Burada; SMA, belirli bir dönemdeki fiyatların basit hareketli ortalamasıdır ve genellikle 20 günlük periyot kullanılır. Standart Sapma, belirli bir dönemdeki fiyat hareketlerinin standart sapmasıdır. Genellikle 2 olarak alınan  $k$  bant genişliği faktörünü temsil eder. Bollinger bantları, fiyatın üst veya alt bantlarına yaklaştığında aşırı alım veya aşırı satım koşullarını gösterir. Ayrıca, bantların genişlemesi veya sıkışması, fiyat hareketlerinin değişimini yansıtabilir.

$$OrtaBant = SMA \quad (2.11)$$

$$UstBant = OrtaBant + (k \times StandartSapma) \quad (2.12)$$

$$AltBant = OrtaBant - (k \times StandartSapma) \quad (2.13)$$

Birikim ve dağıtım (AD) indikatörü fiyat hareketlerini ve işlem hacmini birleştirerek eğilimleri değerlendirmek ve potansiyel dönüş noktalarını belirlemek için kullanılan bir göstergedir. Bu gösterge, özellikle hacim analizinde kullanılan bir araçtır. AD indikatörü, fiyatın yükseldiği günlerde alım hacmini ve fiyatın düştüğü günlerde satış hacmini dikkate alır. Göstergenin değeri, işlem hacmi ve fiyat hareketleri arasındaki ilişkiyi göstermektedir. Pozitif bir değer birikimi, negatif bir değer ise dağıtımı temsil eder. Eşitlik (2.14)'de AD indikatörü formülü verilmiştir.

$$\text{Birikim ve Dağıtım} = \text{Önceki Birikim ve Dağıtım} + \text{Mevcut Birikim ve Dağıtım} \quad (2.14)$$

Mevcut Birikim ve Dağıtım, eşitlik (2.15)'deki formülle hesaplanmaktadır.

$$\frac{(\text{Kapanış} - \text{Düşük}) - (\text{Yüksek} - \text{Kapanış})}{\text{Yüksek} - \text{Düşük}} \times \text{İşlem Hacmi} \quad (2.15)$$

Gerçek aralık (TR) indikatörü, bir dönemdeki fiyat hareketlerinin yüksek-düşük farkı, önceki günün kapanış fiyatı ile en yüksek veya en düşük fiyat arasındaki farkın maksimum değeri olarak hesaplanır. Eşitlik (2.16)'de TR formülü verilmiştir. ATR ise genellikle 14 günlük dönemdeki fiyat hareketlerinin ortalama gerçek aralığını temsil eder, fiyatın volatilitisini ölçmek ve olası trend değişikliklerini belirlemek için kullanılan bir göstergedir. Eşitlik (2.17)'de ATR formülü verilmiştir

$$\max((\text{Yüksek} - \text{Düşük}), |(\text{Yüksek} - \text{Önceki Kapanış})|, |(\text{Düşük} - \text{Önceki Kapanış})|) \quad (2.16)$$

$$ATR = \frac{1}{n} \sum_{i=1}^n TR_i \quad (2.17)$$

Para akış endeksi (MFI), bir finansal varlığın fiyatı ve işlem hacmi birlikte kullanarak şu aşamalarla hesaplanır: Tipik fiyatı (TF) hesapla: Her bir işlem periyodu için, yüksek (high), düşük (low) ve kapanış (close) fiyatları kullanılarak tipik fiyat hesaplanır. Tipik fiyat, eşitlik (2.18)'deki gibi hesaplanır.

$$TF = (\text{Yüksek} + \text{Düşük} + \text{Kapanış})/3 \quad (2.18)$$

Her bir işlem periyodu için, tipik fiyat ve işlem hacmi kullanılarak para akışı (PA) hesaplanır. PA, eşitlik (2.19) şeklinde hesaplanır. Pozitif para akışı, tipik fiyatın arttığı günlerde gerçekleşen işlemlerin toplamını, negatif para akışı ise tipik fiyatın azaldığı günlerde gerçekleşen işlemlerin toplamını temsil eder.

$$PA = TF * \text{Hacim} \quad (2.19)$$

Belirli bir periyot boyunca pozitif ve negatif para akışlarının oranı olan para akışı oranı (PAO) formülü eşitlik (2.20)'de verilmiştir.

$$PAO = \frac{\text{Toplam Pozitif PA}}{\text{Toplam Negatif PA}} \quad (2.20)$$

Para Akışı Oranının bir fonksiyonu olarak hesaplanan para akışı indeksi (MFI) eşitlik (2.21)'de verilmiştir.

$$MFI = 100 - \left( \frac{100}{1 + PAO} \right) \quad (2.21)$$

KDJ osilatörü (KDJ oscillator), teknik analizde kullanılan bir momentum göstergesidir. Finansal varlıkların aşırı alım ve aşırı satım durumlarını belirlemek ve fiyat trendlerinin değişim noktalarını tahmin etmek için kullanılır. KDJ göstergesi, 3 ana bileşen üzerinden hesaplanır [?]:

- (K) 5 günlük Stokastik(eşitlik (2.22)'de gösterilmiştir.),
- (D) K'nın 3 günlük Hareketli Ortalaması
- (J) 3K-2D farkının gösterimidir.

$$K = 100(\text{Kapanış} - \text{Düşük}_5)/(\text{Yüksek}_5 - \text{Düşük}_5) \quad (2.22)$$

Tablo 2.4. Çalışmada kullanılan oransal parametreler.

Parametre	Açıklama
OpenPriceChangeRates	Açılış fiyatının bir öncekine oranı
HighPriceChangeRates	En yüksek fiyatın bir öncekine oranı
LowPriceChangeRates	En düşük fiyatın bir öncekine oranı
ClosePriceChangeRates	Kapanış fiyatının bir öncekine oranı
HighDivOpen	En yüksek fiyatın açılış fiyatına oranı
LowDivOpen	En düşük fiyatın açılış fiyatına oranı
CloseDivOpen	Kapanış fiyatının açılış fiyatına oranı

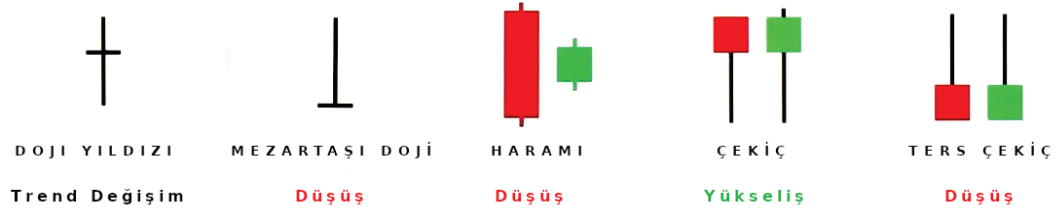
Tablo 2.5. Çalışmada kullanılan mumlar.

Parametre	Açıklama
CDLDOJISTAR	Doji yıldızı
CDLGRAVESTONEDOJI	Mezartaşı dojisi
CDLHAMMER	Çekiç mumu
CDLINVERTEDHAMMER	Ters çekiç mumu
CDLHARAMI	Harami mumu

$%D = %K$ 'nın 3 günlük basit hareketli ortalamasıdır. Grafik üzerinde görülen 80-100 arası bölge aşırı alım, 0-20 arası bölge ise aşırı satım bölgesidir.  $%K$  eğrisinin  $%D$  eğrisini aşağı yönlü kesmesi durumunda satış sinyali, yukarı yönlü kesmesi durumunda alım sinyali vermektedir. [?]

Oranlar bir varlığın fiyatının belirli bir süre içinde nasıl değiştiğini anlamak için kullanılabilir. Alıcıların veya satıcıların belirli bir süre boyunca daha güçlü olduğunu veya bir eğilimin ne kadar güçlü olduğunu göstermek amacıyla kullanılabilir. Örneğin, bir hisse senedi açılışta düşük bir fiyattan işlem görüyor ve kapanışta yüksek bir fiyata ulaşıyorsa, açılışın kapanışa oranı yüksek olacaktır. Bu, alıcıların hisse senedine talep gösterdiğini ve fiyatın artış eğiliminde olduğunu gösterebilir. Tam tersi durumda, düşük bir açılışın ardından kapanışta daha düşük bir fiyatla sonuçlanan bir hisse senedi için açılışın kapanışa oranı düşük olacaktır ve satıcıların baskın olduğunu veya fiyatın düşme eğiliminde olduğunu gösterebilir. Tablo 2.4'de veri kümesine özellik olarak eklenen değişik fiyatların birbirine olan oranları verilmektedir.

Tablo 2.5'de veri zenginleştirilmesinde kullanılan mumlar ve tahmin edilen trend yönleri gösterilmiştir. Genel olarak kendisinden önce gelen uzun gövdeli bir mumdan fiyat aralığı ile ayrılmış çok kısa gövdeli mumlara yıldız denir. Açılış ve kapanış



Şekil 2.4. Mumlar ve sonraki periyotlarda tahmin edilen trend yönleri gösterilmiştir.

fiyatları aynı olanlar ise doji olarak adlandırılırlar [121]. Fiyatlar açılıştan sonra önemli miktarda önce yukarı hareket etmiş fakat kapanış yine açılış fiyatına yakın bir seviyede olduğunda mezartaşı dojisi oluşmaktadır. İkinci mumun gövdesinin, ilk mumun gövdesi içinde kaldığı formasyona harami adı verilmektedir ve  $\{-100, 0, 100\}$  değerlerinden birini alır. Kısa bir gövdeye sahip ve bu gövde işlem aralığının üstündeyse, uzun bir alt gölgesi var ve üst gölge yoksa oluşan çekiç mumu düşüş trendinin geç safhalarında ortaya çıkması durumunda önemli ölçüde etkilidir. Ters çekiç mumu, çekiç mumunun tam tersi olan bu mumun alt kuyruğu yoktur [121].

Şekil 2.4'de veri zenginleştirilmesinde kullanılan mumlar ve sonraki periyotlarda tahmin edilen trend yönleri gösterilmiştir. Mumlar, tekli, ikili, üçlü şekilde farklı formasyonlarla birlikte değerlendirilebilirler [121]. Bu ve benzeri birçok gösterge hesabı finansal piyasalarda uzmanlar tarafından kullanılmakla birlikte hiçbir zaman kesinlik arz etmemektedir.



### 3. BÖLÜM

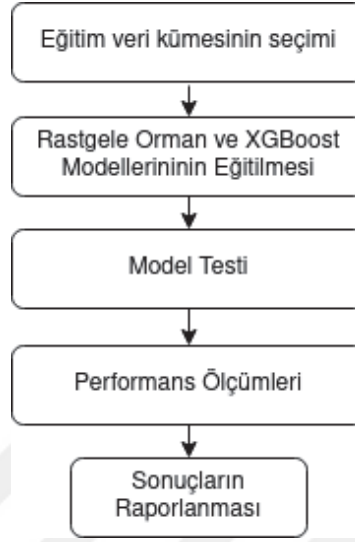
#### 3.DENEYLER

Algoritmaların, gecikme hiperparametresinin değiştirilerek kapanış fiyatlarının tahminlemedeki başarımları karşılaştırılacaktır. Bitcoin fiyatını tahmin etmek için RF ve XGBoost yöntemlerine dayalı regresyon yöntemi kullanılmıştır. Yapılan deneyler doğrultusunda iki farklı algoritmanın parametreleri bazında deneyler yapılarak performans karşılaştırması yapılmıştır. Deneyler aşağıda belirtilen işlem adımları doğrultusunda gerçekleştirilmiştir.

1. Belirlenen aralıklar çerçevesinde kripto borsası üzerinden OHLCV verileri indirilmiştir.
2. Bağımsız değişkenlerin zaman serisi analizlerinde kullanılan indikatörleri ile veri zenginleştirilmesi yapılmıştır ve bir sonraki zaman dilimindeki kapanış fiyatı bağımlı değişken olarak belirlenmiştir.
3. Veriyi eğitim ve test setleri olarak ayrılmıştır.
  - Verinin 2/3'si eğitim seti olarak kullanılmış.
  - Geriye kalan 1/3 lük kısmı testlerde kullanılmıştır.
4. Denetimli makine öğrenmesi için hedef değişken olarak Bitcoin'in ertesi günkü kapanış fiyatı belirlenmiştir.

#### 3.1. Algoritma parametre analizi ve performans ölçümleri

Literatürde yayınlanan benzer çalışmalardaki performanslarından dolayı regresyon modeli olarak özellikle RF ve XGBoost yöntemleri karşılaştırılması yapılmıştır. Binance borsasında yapılan manipulatif yönlendirmeler karşısında algoritmaların tahmin performansları gözlemlenmiştir. Veri kümesinin eğitim olarak kullanılan alt kümesi ile modeller eğitilmiştir. Daha sonra her iki modelin performansı test verileri kullanılarak değerlendirilmiştir. Test kümesi üzerinde değer tahmini yapılmıştır.



Şekil 3.1. Makine öğrenmesi eğitim ve test aşamasında yapılan işlemler gösterilmiştir.

Daha sonra tahminlerle gerçek değerler arasındaki hata payı hesaplanarak hataların karelerinin karekök ortalaması (RMSE) ölçülmüştür. Tahminler ve hata ölçümleri, modelin parametreleri ile birlikte değerlendirilmiştir. Şekil 3.1’te makine öğrenmesi eğitim ve test aşamasında yapılan işlem adımları gösterilmiştir.

Deneyler için python programlama dili 3.10.6 sürümü kullanılmıştır. Veri kümesi Binance’den indirilmiştir. Ta-lib kütüphanesi ile indikatörler elde edilerek veri kümesi zenginleştirilmiştir. Sklearn kütüphanesi makina öğrenmesinde kullanılacak olan RF ve XGBoost algoritması için kullanılmıştır. Skforecast kütüphanesinden RandomForestRegressor ve GradientBoostingRegressor tahminleme için kullanılmıştır. Matplotlib ile grafik çizimleri yapılmıştır.

RF parametreleri Tablo 3.1’de gösterilmiş ve sırasıyla aşağıda açıklanmıştır.

- bootstrap(yerleştirme): örneklemelerin yerine koyarak alınması gerektiğini belirler. Varsayılan değeri True’dur.
- ccp alpha(alfa2): Ağacın yığınlık kesimlerinde bir parametre, daha büyük bir ccp alpha değeri, daha küçük alt dallara sahip ağaçlara neden olur ve aşırı uydurmaları önleyebilir.
- criterion(kriter): Ağacın bölünmesi için bir kriter belirler.

Tablo 3.1. RF Tahmin Modeli parametreleri

Parametreler	RandomForestRegressor()
yerleştirme	Var
alfa2	0.0
kriter	MSE
maks derinlik	Yok
maks özellik	1.0
maks yaprak	Yok
maks örnek	Yok
min düşüş eşiği	0.0
yaprak min örnek	1
min yaprak bölümü	2
yaprak min ağırlık çeşitliliği	0.0
n tahminci	100
n iş parçası	Yok
Doğruluk değerlendirme	Yok
rastgele sayı üretici	Yok
çıktı	0
devam özelliği	Yok

- max depth(maks derinlik): Ağacın maksimum derinliğini belirler. Varsayılan değeri None'dur ve ağacın tamamen büyümesine izin verir.
- max features(maks özellik): Ağaç oluşturulurken maksimum özellik sayısını belirler. Varsayılan değeri 1'dir, yani tüm özellikler kullanılır.
- max leaf nodes(maks yaprak): Ağacın maksimum yaprak sayısını belirler. Varsayılan değeri None'dur ve yaprak sayısı sınırlanmaz.
- max samples(maks örnek): Ağacın her bir düğümünde kullanılacak örneklem sayısını belirler. Varsayılan değeri None'dur ve tüm örnekler kullanılır.
- min impurity decrease(min düşüş eşiği): Bölünmelerdeki düşüşü hesaplamak için bir eşik değeri belirler.
- min samples leaf(yaprak min örnek): Bir yaprakta minimum örnek sayısını belirler. Varsayılan değeri 1'dir.
- min samples split(min yaprak bölüm): Bölünme yapılırken minimum örnek sayısını belirler. Varsayılan değeri 2'dir.

- min weight fraction leaf(yaprak min ağırlık çeşitliliği): Bir yapraktaki minimum ağırlık fraksiyonunu belirler. Varsayılan değeri 0'dır.
- n estimators(tahminci): Oluşturulacak ağaç sayısını belirler. Varsayılan değeri 100'dür.
- n jobs(iş parçası): Eşzamanlı işlemler için kullanılacak CPU sayısını belirler. Varsayılan değeri None'dur.
- oob score(doğrulama dilimi): Out-of-bag (OOB) doğruluğunu hesaplamak için bir bayrak. Varsayılan değeri False'dur.
- random state(rastgele sayı üretici): Rastgele sayı üreticisi için bir başlangıç değeri belirler.
- verbose(çıkıtı): Modelin eğitimi sırasında ne kadar çıktı alınacağını belirler.
- warm start(devam özelliği): Eğitim sürecine bir önceki eğitim sonuçlarından başlanarak devam etme özelliğini belirler. Varsayılan değeri False'dur.

XGBoost tahmin modeli parametreleri Tablo 3.2'de gösterilmiş ve aşağıda açıklanmıştır.

- alpha(alfa): Daha büyük bir alpha değeri, modelin daha basit bir hale getirilmesine ve aşırı uydurmaların önlenmesine yardımcı olabilir.
- ccp alpha(alfa2): Ağacın yoğunluk kesimlerinde bir parametre, daha büyük bir ccp alpha değeri, daha küçük alt dallara sahip ağaçlara neden olur ve aşırı uydurmaları önleyebilir.
- criterion(kriter): Ağacın bölünmesi için bir kriter belirler, Friedman MSE, MSE ve MAE olabilir.
- init(baslangic): İlk tahmin olarak kullanılacak öğrenicinin belirlenmesi. learning rate(öğrenme oranı): Her iterasyonda kullanılan öğrenme oranını belirler.
- loss(kayıp): Kayıp fonksiyonunu belirler. Bu parametre, modelin hangi hatayı en aza indirmeye çalışacağını belirler. (parametre loss="quantile" şeklinde regresyon modeline gönderilmiştir.)

Tablo 3.2. XGBoost tahmin modeli parametreleri

Parametreler	GradientBoostingRegressor (kayıp='dilimsel')
alfa	0.05
alfa2	0.0
kriter	MSE
başlangıç	Yok
öğrenme oranı	0.1
kayıp	dilimsel
maks derinlik	3
maks özellik	Yok
maks yaprak	Yok
min düşüş eşiği	0.0
yaprak min örnek	1
min yaprak bölümü	2
yaprak min ağırlık çeşitliliği	0.0
n tahminci	100
iterasyon kesici	Yok
rastgele sayı üretici	Yok
örneklem	1.0
tolerans	0.0001
doğrulama dilimi	0.1
çıktı	0
devam özelliği	Yok

- max depth(maks derinlik): Ağacın maksimum derinliğini belirler.
- max features(maks özellik): Ağaç oluşturulurken maksimum özellik sayısını belirler.
- max leaf nodes(maks yaprak): Ağacın maksimum yaprak sayısını belirler.
- min impurity decrease(min düşüş eşiği): Bölünmelerdeki düşüşü hesaplamak için bir eşik değeri belirler.
- min samples leaf(yaprak min örnek): Bir yaprakta minimum örnek sayısını belirler.
- min samples split(min yaprak bölüm): Bölünme yapılırken minimum örnek sayısını belirler.
- min weight fraction leaf(yaprak min ağırlık çeşitliliği): Bir yapraktaki minimum ağırlık fraksiyonunu belirler.

- `n_estimator`(tahminci): Oluşturulacak ağaç sayısını belirler.
- `n_iter_no_change`(iterasyon kesici): İterasyonları durdurmak için bir eşik değerini belirler.
- `random_state`(rastgele sayı üretici): Rastgele sayı üreticisi için bir başlangıç değeri belirler.
- `subsample`(örnekleme): Her bir ağaç için kullanılacak veri örnekleme oranını belirler.
- `tol`(tolerans): Modelin eğitiminde kullanılan tolerans eşiği.
- `validation_fraction`(doğrulama dilimi): Veri kümesinin ne kadarının doğrulama için kullanılacağını belirler.
- `verbose`(çıktı): Modelin eğitimi sırasında ne kadar çıktı alınacağını belirler.
- `warm_start`(devam özelliği): Eğitim sürecine bir önceki eğitim sonuçlarından başlanarak devam etme özelliğini belirler.

Otomatik regresyonlu tahminci (`forecasterAutoreg`) ve ortam ayarları Tablo 3.3'de gösterilmiştir. `RFRegressor`, rastgele orman regresyon modeli kısaltmasıdır. `GBRegressor`, gradyan aşırı arttırma regresyon modeli kısaltmasıdır.

Gecikme olarak da adlandırılan lag, bir zaman serisinin kendi önceki değerlerine olan bağımlılığını belirler ve otomatik regresyon modellerinde, lag parametresi, kaç önceki gözlemin modele dahil edileceğini belirler. Örneğin,  $AR(1)$  modelinde, sadece bir önceki gözlem değeri, mevcut gözlemi tahmin etmek için kullanılır.  $AR(2)$  modelinde, hem bir önceki hem de iki önceki gözlem değeri, tahmin için kullanılır. Genel olarak,  $AR(p)$  modelinde,  $p$  tane önceki gözlem değeri, tahmin için kullanılır.

Lag parametresi, modelin doğruluğunu ve performansını etkileyen önemli bir parametredir ve uygun bir lag parametresi seçimi, modelin doğru tahminler yapabilmesi için önemlidir.

`forecasterAutoReg` tahmincinin aldığı parametre değerleri aşağıda açıklanmıştır.

Tablo 3.3. ForecasterAutoreg parametre deęerleri.

Parametreler	RFRegressor()	GBRegressor(alfa=0.05)
Gecikme	1-180	1-180
Baęlı deęişken dnştrc	Yok	Yok
Baęımsız deęişkenler dnştrc	Yok	Yok
Pencere boyutu	1-180	1-180
Aęırlık fonksiyonu dahil edildi	Hayır	Hayır
Baęımsız deęişkenler dahil edildi	Hayır	Hayır
Baęımsız deęişken tr	Yok	Yok
Baęımsız deęişkenlerin isimleri	Yok	Yok
Eęitim aralıęı	327 - 982	327 - 982
Eęitim indeks tr	Aralık indeks	Aralık indeks
Eęitim indeks frekansı	1	1
Skforecast srm	0.6.0	0.6.0
Python srm	3.10.6	3.10.6

- Gecikme: Modelde kullanılan gecikme sayısı kendisinden ne kadar nceki verinin etkisi olduęunu belirtmektedir.
- Baęlı deęişken dnştrc: Hedef deęişkenin (y) dnştrcs. Bu parametre, hedef deęişkenin dnştrlmş olabileceęini ve modele dnştrlmş hedef deęişkenin saęlandıęını belirtir.
- Baęımsız deęişken dnştrcs: Baęımsız deęişkenlerin (exog) dnştrcs.
- baęımsız deęişkenlerin dnştrlmş olabileceęini ve modele dnştrlmş baęımsız deęişkenlerin saęlandıęını belirtir.
- Pencere boyutu: Modelde kullanılan kaydırma penceresinin boyutu. Burada, 23 adım pencere boyutu belirtiliyor.
- Aęırlık fonksiyonu dahil edildi: Modelde aęırlık fonksiyonunun kullanılıp kullanılmayacaęını belirtir.
- Baęımsız deęişkenler dahil edildi: Modelde baęımsız deęişkenlerin kullanılıp kullanılmayacaęını belirtir.
- Baęımsız deęişken tr: Kullanılan baęımsız deęişken trn belirtir.
- Baęımsız deęişkenlerin isimleri: Kullanılan baęımsız deęişkenlerin isimlerini belirtir.

- Eğitim aralığı: Modelin eğitildiği veri aralığını belirtir.
- Eğitim indeks türü: Eğitim verilerinde kullanılan indeks türünü belirtir.
- Eğitim indeks frekansı: Eğitim verilerinde kullanılan indeks frekansını belirtir.

Deneylerde kullanılan makine öğrenme algoritmaları ve tahmin için kullanılan regresyonlar belirtildiği gibi belirlenmiştir.

Deneyler lag hiper parametresi üzerinden yapılmıştır. Lag 1 olarak seçilmiş ve eğitim verisi eğitim için kullanılmıştır. Test verisi sadece kapanış verisi üzerinden testlerde kullanılmıştır. Burada bir günlük tahminde bulunmak için önceki günün OHLCV verisi gerekmektedir. Bu durumda ikinci günün kapanış fiyatını tahmin etmek için birinci günün OHLCV verisi de gerekecektir.

Deneyler pencereleme ile yapılmıştır. Lag değeri 1 ile 180 arasında farklı değerler verilerek ayrı ayrı deneyler yapılmıştır. Lag değerine göre tahminlerde sırasıyla 1'den 180 güne kadar önceki günler birlikte modele verilerek sonraki güne ait 330 günlük OHLCV verileri tahmin edilmiştir. 330 günlük tahminler pencere kaydırma yöntemiyle sırasıyla tahmin edilmiştir. Bu şekilde lag 180'e kadar her bir pencere boyutu ayrı ayrı değerlendirilmiştir. Performans kriterleri olan RMSE değeri için 330 günün RMSE değerlerinin ortalaması alınmıştır.



## 4. BÖLÜM

### SONUÇ

#### 4.1. Bulgular ve Sonuçlar

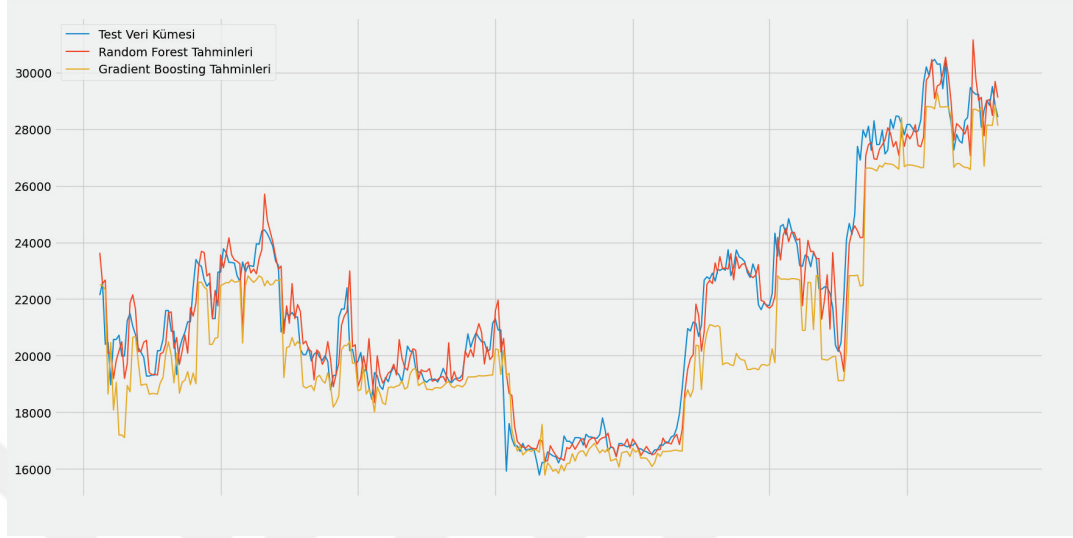
Lag değeri 1 olarak alındığında XGBOOST, RMSE değeri 1560 olarak çıkmaktadır. Buna karşılık RF, RMSE değeri 832 bulunmuştur. Şekil 4.1.'de verilen grafik üzerinde lag parametresi 1 olan RF ve XGBOOST algoritmalarının sonuçları gösterilmiştir. Lag parametresi değeri 180 olarak alındığında XGBOOST RMSE değeri 3258 olarak çıkmaktadır. Buna karşılık RF RMSE değeri 850 bulunmuştur. Şekil 4.2.'de verilen grafik üzerinde lag parametre değeri 180 olan RF ve XGBOOST algoritmalarının sonuçları gösterilmiştir.

Lag değeri 23 olarak alındığında XGBOOST RMSE değeri 1513 olarak çıkmaktadır. Buna karşılık RF RMSE değeri 730 bulunmuştur. Bütün gecikme parametrelerine ait RF ve XGBoost modellerinin elde ettiği RMSE değerleri Tablo 4.1.'de verilmiştir. Yapılan deneylerde fiyat tahmini için kendisinden önceki değişkenlere az veya aşırı bağlı yapılan tahminler daha yüksek hatalı tahminlerde bulunduğunu göstermiştir. XGBOOST ve RF algoritmaları değişen lag değerleri ile birlikte karşılaştırıldığında RF algoritmasının daha küçük hatalarla tahminleme yaptığı görülmüştür. Şekil 4.3.'de lag parametresi değişimine göre performans analizi gösterilmektedir. Buna göre en uygun olan lag değeri 23 olarak bulunmuştur ve Şekil 4.4.'de RMSE değerlerinin ortalaması verilmiştir.

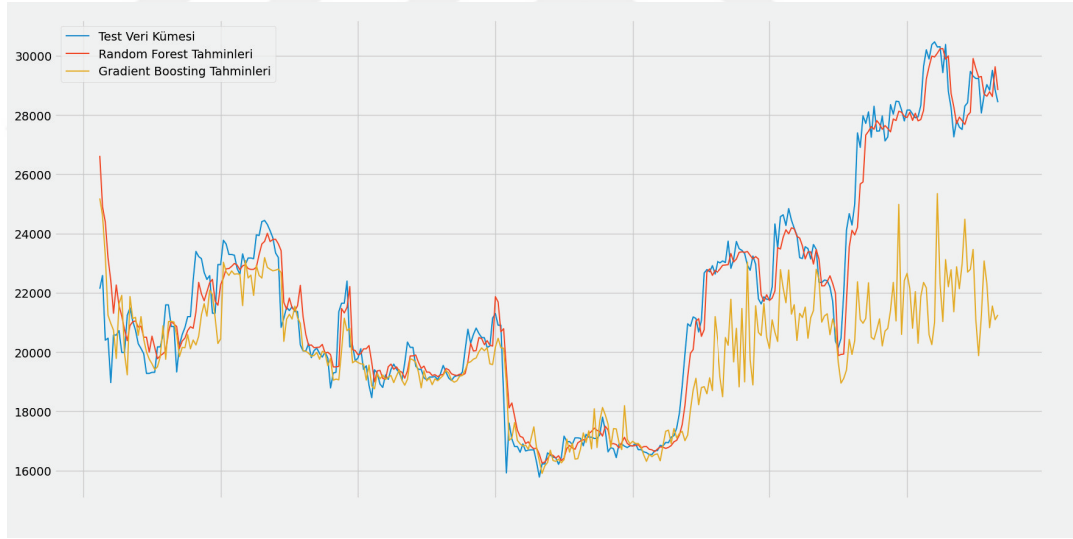
RF algoritmasının RMSE değerinin, XGBOOST algoritmasına göre düşük olması nedeniyle RF algoritması ve lag hiperparametresi en küçük RMSE değerini göstermesi nedeniyle 23 olarak seçilmiştir. 330 günlük Bitcoin fiyat tahminlemesi tekrar yapılmış ve günlük işlemler ile kazanç hesaplaması yapılmıştır. Tahmin edilen fiyatların gerçekleşmesinin sınıflandırılması için 4.2 tablosu hazırlanmıştır;

Tablo 4.1. Algoritma ve lag çiftinin RMSE etkisi.

Lag	RF	GB	Lag	RF	GB	Lag	RF	GB	Lag	RF	GB
1	832	1560	46	745	1667	91	765	1814	136	842	2803
2	797	1542	47	746	1727	92	773	1845	137	852	2808
3	779	1799	48	743	1744	93	780	1839	138	845	2917
4	769	1778	49	758	1745	94	783	1753	139	833	2833
5	770	1698	50	761	1775	95	774	1801	140	827	2934
6	768	1862	51	750	1737	96	770	1734	141	849	2863
7	760	1764	52	749	1836	97	774	1898	142	859	2743
8	751	1615	53	761	1849	98	770	1845	143	840	2857
9	763	1546	54	747	1705	99	775	1909	144	849	2858
10	768	1662	55	757	1808	100	769	1930	145	855	2950
11	762	1666	56	760	1848	101	790	2266	146	860	3015
12	757	1494	57	767	1734	102	784	2305	147	854	2954
13	757	1550	58	756	1737	103	783	2302	148	846	2936
14	757	1476	59	751	1737	104	772	2474	149	858	2985
15	741	1469	60	757	1809	105	774	2388	150	849	2946
16	745	1484	61	755	1870	106	780	2444	151	856	2869
17	738	1592	62	755	1780	107	779	2559	152	872	2949
18	744	1579	63	753	1739	108	775	2647	153	869	2996
19	738	1622	64	761	1663	109	782	2738	154	855	2951
20	740	1556	65	742	1809	110	792	2547	155	840	2957
21	738	1636	66	735	1763	111	782	2665	156	842	2999
22	748	1584	67	755	1722	112	800	2598	157	838	3047
<b>23</b>	<b>730</b>	1513	68	758	1727	113	774	2650	158	835	3067
24	740	1545	69	757	1798	114	777	2632	159	846	3022
25	738	1594	70	756	1687	115	784	2641	160	859	3082
26	736	1576	71	752	1740	116	768	2468	161	850	3036
27	735	1658	72	751	1695	117	792	2552	162	848	3129
28	734	1652	73	761	1752	118	791	2495	163	839	3030
29	731	1661	74	757	1689	119	787	2603	164	860	3073
30	740	1696	75	762	1767	120	794	2558	165	843	3104
31	744	1755	76	756	1753	121	794	2674	166	868	3065
32	732	1815	77	748	1691	122	778	2744	167	841	3118
33	744	1784	78	770	1726	123	794	2541	168	851	3160
34	732	1760	79	754	1663	124	792	2705	169	863	3174
35	747	1716	80	757	1715	125	803	2774	170	850	3215
36	739	1822	81	764	1790	126	813	2640	171	845	3191
37	743	1710	82	759	1852	127	839	2803	172	844	3215
38	744	1689	83	758	1820	128	829	2753	173	853	3193
39	737	1883	84	773	1768	129	844	2772	174	845	3207
40	747	1775	85	762	1948	130	844	2682	175	853	3169
41	751	1767	86	757	1835	131	855	2825	176	844	3223
42	735	1785	87	758	1898	132	825	2828	177	859	3210
43	750	1691	88	755	1879	133	831	2708	178	852	3217
44	758	1730	89	778	2083	134	849	2905	179	859	3237
45	735	1732	90	770	1973	135	835	2761	180	850	3258

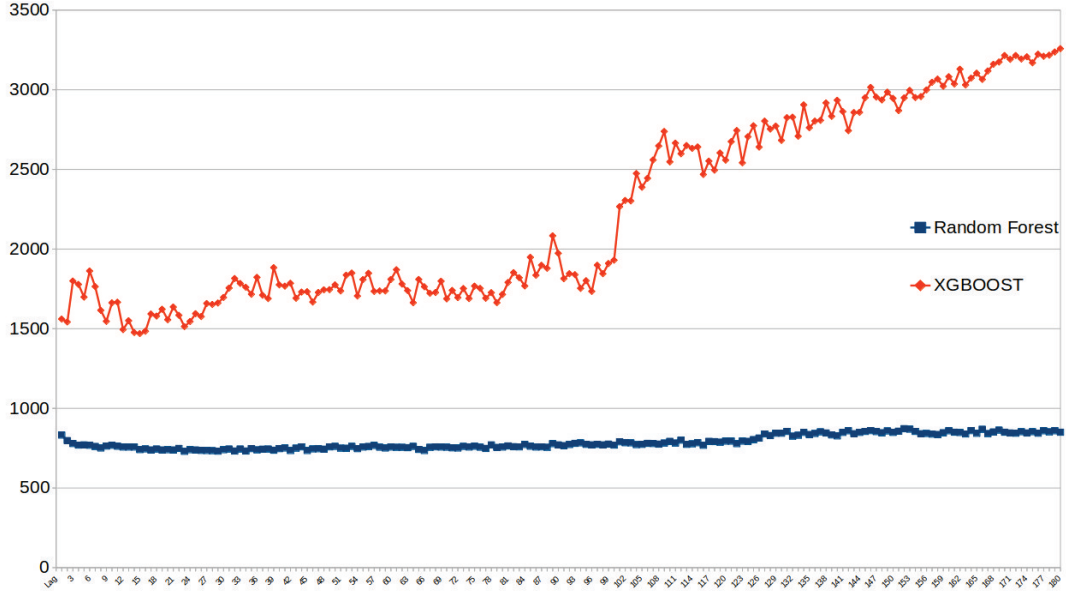


Şekil 4.1. RF (rmse:832) ve XGBOOST (rmse:1560) algoritmalarının lag 1'e göre tahminleri karşılaştırılmıştır.

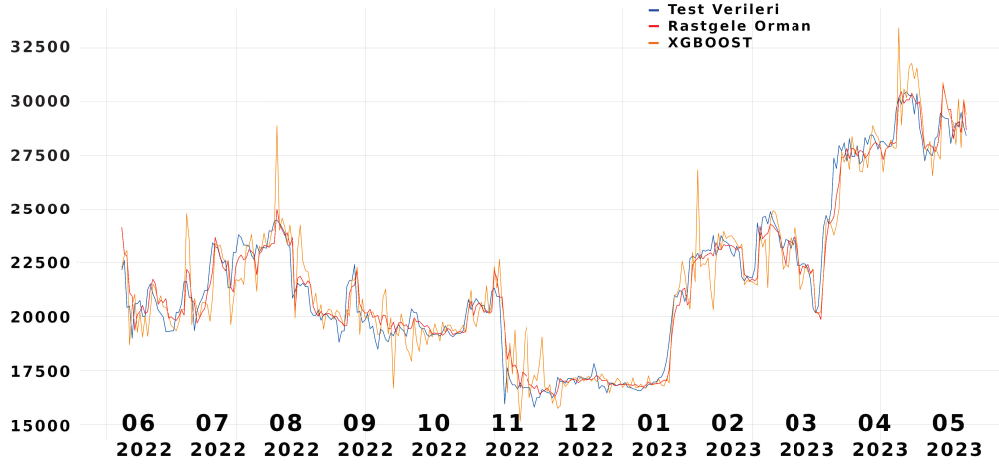


Şekil 4.2. RF (rmse:850) ve XGBOOST (rmse:3258) algoritmalarının lag 180'e göre tahminleri karşılaştırılmıştır.

Negatif yönlerin de önemli olduğundan dolayı fiyattan uzaklaşması hata payı olarak gösterilmiştir. Hata payı düşük tahmin ve yüksek kapanışlarda negatif değerler almaktadır. Fiyatın çıkacağı tahmin edildiğinde tahmin yönü 1, fiyatın düşeceği tahmin edildiğinde tahmin yönü 0 olarak gösterilmiştir. Gerçekleşen sütunu fiyat çıkmışsa 1, fiyat düştüyse 0, fiyat ve gerçekleşen aynı ise başarı sütunu 1 şeklinde verilmiştir.



Şekil 4.3. RF ve XGBOOST lag performans analizi yapılmıştır.



Şekil 4.4. RF (rmse:730) ve XGBOOST (rmse:1513) algoritmalarının lag 23'e göre tahminleri karşılaştırılmıştır.

## 4.2. Çıkarımlar

XGBoost algoritmasına oranla RF algoritması daha iyi sonuçlar vermiştir. Fiyat tahminlemesi için RF algoritmasının parametreleri değiştirilerek daha iyi başarımlar elde edilmiştir. Çalışmanın sonucu olarak tahminleme konusunda doğru algoritma ve doğru parametreler başarımları yüksek sonuçlar verecektir.

RF tabanlı 23 gecikmeli model kullanılarak toplam 330 test verisine 261 doğru

Tablo 4.2. RF lag23 başarımlı hesaplaması yapılmıştır. (Aylık ortalamalar şeklinde özet tablodur.)

RF Tahmini	Kapanış	Hata Payı	Tahmin Yönü	Gerçekleşen	BAŞARI
21288,61	20461,57	41725,73	0,67	0,37	0,43
22227,28	22767,59	-540,31	0,37	0,43	0,80
21398,96	21154,04	15534,97	0,53	0,47	0,67
19750,61	19568,89	8954,43	0,50	0,43	0,80
19811,81	19686,88	124,94	0,50	0,47	0,83
16862,41	16725,70	7402,67	0,63	0,47	0,77
16936,85	16869,20	5701,24	0,73	0,47	0,53
21681,25	22050,15	19083,72	0,53	0,57	0,90
22869,65	22911,88	-39255,87	0,30	0,30	0,87
26392,19	26974,73	-582,54	0,37	0,57	0,80
29030,00	29055,33	-22160,78	0,46	0,54	0,79
21659,06	21656,91	3271,66	0,51	0,46	0,74

yönlü tahminleme yapılarak yüzde 79 sınıflandırma başarımlı elde edilmiştir. Günlük açılan işlemlerle test süreci başlangıcında 100 doları olan bir cüzdan tahmin yönüne göre komisyonşuz şekilde 132 adet işlemlerle cüzdan bakiyesini 124 dolara çıkardığı görülmüştür. Diğer taraftan karşılaştırma amacıyla ilk gün 100 dolarlık Bitcoin alındığı ve hiç bir işlem yapılmadan tutulduğu varsayımına göre aynı dönemde cüzdan bakiyesi 108 dolarda kalmıştır.

### 4.3. Tartışma ve Öneriler

Deneylerde günlük periyot içinde elde edilen veri kümesi ile 657 günlük OHLCV verisi ile modeller eğitilmiş ve 330 günlük OHLCV verisi üzerinde test edilmiştir. RF ve XGBoost temelli regresyon yöntemleri ile sonraki güne ait fiyat tahminleri yapılmıştır. Farklı pencere boyutları ile algoritmaların performansları değerlendirilmiş ve 1 ile 180 arasında lag parametresi kullanılarak doğrulukları karşılaştırılmıştır. Farklı lag parametrelerinden en düşük hata değerini veren parametrenin 23 olduğu tespit edilmiştir. Elde edilen tahmin modeli alım-satım işlemleri için kullanıldığı simüle edilmiş ve başta 100 olan bakiye değeri 132'ye çıkmıştır.

Gelecek çalışmalarda OHLCV verisi üzerinde diğer sınıflandırma tabanlı makine öğrenme yöntemleri denenerek daha başarılı sonuçlar elde edilmeye çalışılacaktır. Derin öğrenme tabanlı yöntemler ile de tahminler yapılacaktır. Ağırlıklı ortalama

gibi regresyon indikatörleri haricinde trend deęişimine sebep olacak durumları farkedem indikatörlerin kullanılması ve regresyon harici dięer algoritmaların kullanılması tahminleme başarımlını artırabilir. İndikatörler arasındaki ilişkiler tahmin doęruluęunda etkisi analiz çıkarılabilir. Temel analiz de insanların sadece sayısal grafik tabanlı olmayan deęerlendirme yöntemi olduęu için hibrit tahmin modeli geliştirilebilecektir.



**KAYNAKLAR**

1. Can, M., 2009. İşletmelerde zaman serileri analizi ile tahmin, Doktora Tezi, İstanbul Üniversitesi.
2. Coinmarketcap, 2023, Bitcoin price, (Web Sayfası: <https://coinmarketcap.com/tr/currencies/bitcoin/>), Erişim tarihi: 16/5/2023.
3. Turhost, 2023, Makine öğrenmesi (machine learning) nedir?, (Web Sayfası: <https://turhost.com/blog/makine-ogrenmesi-machine-learning-nedir>) Erişim tarihi: 16/5/2023.
4. Coinmarketcap, 2023, Borsaların Hacim Sıralamaları, (Web Sayfası: <https://coinmarketcap.com/tr/rankings/exchanges>), Erişim tarihi: 16/5/2023.
5. Jumpstarttrading, 2023, Birikimli hacim farkı indikatörü(CVD), (Web Sayfası: <https://www.jumpstarttrading.com/volume-delta/>), Erişim tarihi: 16/5/2023.
6. Akgül, I., 1994. Zaman serisi analizi ve öngörü modelleri, **Öneri Dergisi**, 1(1):52-69.
7. Barokas, L., 2022. Essays on Credit Swap Default Market: Time Series Predictions with Machine Learning, Doktora Tezi, Marmara Üniversitesi (Turkey).
8. Alpaydin, E., 2004. Introduction to Machine Learning, The MIT Press.
9. Arslan, K., 2020. Eğitimde yapay zekâ ve uygulamaları, **Batı Anadolu Eğitim Bilimleri Dergisi**, 11(1):71-88.
10. Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system, **Decentralized business review**, 21260.
11. Yermack, D., 2013. Is Bitcoin a real currency? An economic appraisal (No. w19747), **National Bureau of Economic Research**, 36(2):843-850.

12. Kristoufek, L., 2015. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis, **PloS one**, **10**(4):e0123923.
13. Vishwanathan, S. and Murty, M.N., 2002. SSVM: a simple SVM algorithm, *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, volume 3, 2393–2398, IEEE.
14. Dunn, P.K., Smyth, G.K., Dunn, P.K., and Smyth, G.K., 2018. Chapter 9: Models for proportions: binomial GLMs, **Generalized Linear Models With Examples in R**, 333–369.
15. Rigatti, S.J., 2017. Random forest, **Journal of Insurance Medicine**, **47**(1):31–39.
16. Şen, Z., 2004. Yapay sinir ağları, Su Vakfı.
17. Medsker, L.R. and Jain, L., 2001. Recurrent neural networks, **Design and Applications**, **5**:64–67.
18. LaValley, M.P., 2008. Logistic regression, **Circulation**, **117**(18):2395–2399.
19. McNally, S., 2016. Predicting the price of Bitcoin using Machine Learning, Doktora Tezi, Dublin, National College of Ireland.
20. Indera, N., Yassin, I., Zabidi, A., and Rizman, Z., 2017. Non-linear autoregressive with exogeneous input (NARX) Bitcoin price prediction model using PSO-optimized parameters and moving average technical indicators, **Journal of fundamental and applied sciences**, **9**(3S):791–808.
21. Greaves, A. and Au, B., 2015. Using the bitcoin transaction graph to predict the price of bitcoin, **URL: [http://snap.stanford.edu/class/cs224w-2015/projects\\_2015/Using\\_the\\_Bitcoin\\_Transaction\\_Graph\\_to\\_Predict\\_the\\_Price\\_of\\_Bitcoin.pdf](http://snap.stanford.edu/class/cs224w-2015/projects_2015/Using_the_Bitcoin_Transaction_Graph_to_Predict_the_Price_of_Bitcoin.pdf)**, **8**:416–443.
22. Mern, J., Anderson, S., and Poothokaran, J., 2017. Using bitcoin ledger network data to predict the price of bitcoin, **cs229. stanford. edu**.



23. Kim, Y.B., Kim, J.G., Kim, W., Im, J.H., Kim, T.H., Kang, S.J., and Kim, C.H., 2016. Predicting fluctuations in cryptocurrency transactions based on user comments and replies, **PloS one**, **11**(8):e0161197.
24. Stenqvist, E. and Lönnö, J., 2017, Predicting Bitcoin price fluctuation with Twitter sentiment analysis.
25. Chen, Z., Li, C., and Sun, W., 2020. Bitcoin price prediction using machine learning: An approach to sample dimension engineering, **Journal of Computational and Applied Mathematics**, **365**:112395.
26. Yassin, I., Khalid, M.A., Herman, S., Pasya, I., Wahab, N.A., and Awang, Z., 2017. Multi-layer perceptron (MLP)-based nonlinear auto-regressive with exogenous inputs (NARX) stock forecasting model, **Int. J. Adv. Sci. Eng. Inf. Technol**, **7**(3):1098–1103.
27. Tadvi, A.A., 2018. Bitcoin price prediction, **International Journal of Academic Research and Development**.
28. Sakiz, B. and Kutlugün, E., 2018. Bitcoin price forecast via blockchain technology and artificial intelligence algorithms, *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 1–4, IEEE.
29. Minka, T., 2000. Bayesian linear regression, Technical report, Citeseer.
30. Shah, D. and Zhang, K., 2014. Bayesian regression and Bitcoin, *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 409–414.
31. Shumway, R.H., Stoffer, D.S., Shumway, R.H., and Stoffer, D.S., 2017. ARIMA models, **Time Series Analysis and Its Applications: With R Examples**, 75–163.
32. Amjad, M. and Shah, D., 2017. Trading Bitcoin and Online Time Series Prediction, *O. Anava, A. Khaleghi, M. Cuturi, V. Kuznetsov, and A. Rakhlin, Eds.; Proceedings of the Time Series Workshop at NIPS 2016*,

volume 55 of Proceedings of Machine Learning Research, 1–15, PMLR, Barcelona, Spain.

33. Amjad, U., Jilani, T.A., Tariq, H., and Hussain, A., 2018. A quantum based evolutionary algorithm for stock index and bitcoin price forecasting, **International Journal of Advanced Computer Science and Applications**, **9(9)**:123–132.
34. Guo, G., Wang, H., Bell, D., Bi, Y., and Greer, K., 2003. KNN model-based approach in classification, *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, 986–996, Springer.
35. Şengül, Z., 2022. Makine öğrenmesi algoritmalarını kullanarak bitcoin fiyat tahmini, Yüksek Lisans Tezi, Trakya Üniversitesi Sosyal Bilimler Enstitüsü.
36. Afşar, M., Özdemir, B.K., and Arslan, B.G., 2012. Para Teorisi, **faiz oranlarının anlamı, ölçümü ve belirlenmesi**, 46–71.
37. Şıklar, İ., 2009. Para Teorisi ve Politikası, Anadolu Üniversitesi.
38. Fidan, M., Dilek, S., and Esev, A., 2019. Dünden bugüne paranın tarihi ve Türkiye’de kağıt para kullanımı, **Kilis 7 Aralık Üniversitesi Sosyal Bilimler Dergisi**, **9(18)**:141 – 162.
39. Alpayo, H., 2018. Bitcoin’den selfcoin’e kripto para, **Uluslararası Bilimsel Araştırmalar Dergisi (IBAD)**, **3(2)**:411 – 428.
40. İnci, S. and Alpen, I., 2018. Bitcoin devrimi: Degisen dünya ekonomisinde kripto para sistemi, blockchain, altcoinler, Elma Yayınevi.
41. İslim, B., 2021. Bitcoin’in fiyat dinamiklerinin belirleyicileri: Bir zaman seri yaklaşımı, Yüksek Lisans Tezi, Kırklareli Üniversitesi.
42. Gökhan, Ü. and Uluçol, Ç., 2020. Blok zinciri teknolojisi, **Bilişim Teknolojileri Dergisi**, **13(2)**:167–175.

43. Acar, O., 2009. Türkiye’de dış ticaret ve dış ticaretin finansmanı durum değerlendirmesi, **TEPAV Raporu**.
44. Sobti, R. and Geetha, G., 2012. Cryptographic hash functions: a review, **International Journal of Computer Science Issues (IJCSI)**, **9(2)**:461.
45. Rachmawati, D., Tarigan, J., and Ginting, A., 2018. A comparative study of Message Digest 5 (MD5) and SHA256 algorithm, *Journal of Physics: Conference Series*, volume 978, 012116, IOP Publishing.
46. BTC scan org block scanner, 2023, <https://btcscan.org/block/000000000000000000000000000000000034dc521e2b9d0654ef4d6d74d5ad86265ae138dfd0a61>, erişim Tarihi: 22 Eylül 2023.
47. Meynkhart, A., 2019. Fair market value of bitcoin: Halving effect, **Investment Management and Financial Innovations**, **16(4)**:72–85.
48. Gökhan, Ü. and Uluyol, Ç., 2020. Blok zinciri teknolojisi, **Bilişim Teknolojileri Dergisi**, **13(2)**:167–175.
49. Sedlmeir, J., Buhl, H.U., Fridgen, G., and Keller, R., 2021. Recent developments in blockchain technology and their impact on energy consumption, **arXiv preprint arXiv:2102.07886**.
50. Digiconomist.net, 2023, Bitcoin Energy Consumption, (Web Sayfası: <https://digiconomist.net/bitcoin-energy-consumption>), Erişim tarihi: 16/5/2023.
51. Gündüz, F., Birogul, S., and Kose, U., 2023. Proof of optimum (PoO): Consensus model based on fairness and efficiency in blockchain, **Applied Sciences**, **13**:10149.
52. Özdemir, V. ve Yavuz, H., 2021. Yeni uluslararası iktisadi düzende dolarsızlaşma eğilimleri, **Başkent Üniversitesi Ticari Bilimler Fakültesi Dergisi**, **5(1)**:64–76.

53. de Vries, A., Gallersdörfer, U., Klaaßen, L., and Stoll, C., 2021. The true costs of digital currencies: Exploring impact beyond energy use, **One Earth**, 4(6):786–789.
54. Arf, C., 1959. Makine düşünebilir mi ve nasıl düşünebilir, **Atatürk Üniversitesi-Üniversite Çalışmalarını Muhite Yayıma ve Halk Eğitimi Yayınları Konferanslar Serisi**, (1):91–103.
55. Alpaydin, E., 2004, Introduction to Machine Learning (Adaptive Computation and Machine Learning).
56. Kırca, S., 2021. Derin öğrenme yöntemi ile araç ve plaka tanıma, Yüksek Lisans Tezi, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü.
57. Ögücü, M.O., 2006. Yapay sinir ağları ile sistem tanıma, Doktora Tezi, Fen Bilimleri Enstitüsü.
58. Cakir, M., 2005. Firma başarısızlığının dinamiklerinin belirlenmesinde makina öğrenmesi teknikleri: Ampirik uygulamalar ve karşılaştırmalı analiz.
59. Ferhat, Ç. Elektrik motorlarının arıza tespitinin otomatik olarak yapılmasında kullanılan yapay öğrenme yöntemleri, **Dicle Üniversitesi Fen Bilimleri Enstitüsü Dergisi**, 6(2):95–103.
60. Kaynar, O., Görmez, Y., Yıldız, M., and Albayrak, A., 2016. Makine öğrenmesi yöntemleri ile duygu Analizi, *International Artificial Intelligence and Data Processing Symposium (IDAP'16)*, volume 17, 17–18.
61. Şahin, C., 2015. Factored reinforcement learning using extended sequence trees, Yüksek Lisans Tezi, Middle East Technical University.
62. Onan, A. and Korukoğlu, S., 2016. Makine öğrenmesi yöntemlerinin görüş madenciliğinde kullanılması üzerine bir literatür araştırması., **Pamukkale University Journal of Engineering Sciences**, 22(2).
63. AltınYavuz, A. and Yavuz, H.S., 2021. Denetimli makine öğrenme yöntemleri ile yüzey su kalitesinin sınıflandırılması, **Biyoloji Bilimleri Araştırma Dergisi**, 14(2):142–155.

64. Chao, W.L., 2011. Machine learning tutorial, **Digital Image and Signal Processing**.
65. Kızılkaya, Y.M. and Oğuzlar, A., 2018. Bazı Denetimli Öğrenme Algoritmalarının R Programlama Dili ile Kıyaslanması, **Karadeniz Uluslararası Bilimsel Dergi**, **37(37)**:90–98.
66. Coşkun, C. and Baykal, A., 2011. Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması, **Akademik Bilişim**, **11**:51–58.
67. Irmak, S., 2009. Veri madenciliği yöntemleri ve sağlık sektörü veritabanlarında bilgi keşfi: tanımlayıcı ve kestirimci model uygulamaları.
68. Akmaz, D. Çok katmanlı algılayıcı algoritması, korelasyon tabanlı özellik seçme yöntemi ve eğri uydurma tekniği ile türkiye’de toplam elektrik tüketiminin tahmin edilmesi, **Fırat Üniversitesi Mühendislik Bilimleri Dergisi**, **34(2)**:677–686.
69. Şafak, Ç. Fonksiyonel veri analizi ve sağlık alanına özgü uygulama, Yüksek Lisans Tezi, Sağlık Bilimleri Enstitüsü.
70. Ataseven, B., 2013. Yapay sinir ağları ile öngörü modellemesi, **Öneri Dergisi**, **10(39)**:101–115.
71. Susto, G.A., Cenedese, A., and Terzi, M., 2018. Time-series classification methods: Review and applications to power systems data, **Big data application in power systems**, 179–220.
72. Sezer, Ö.B., 2018. Zaman serisi verilerinin derin yapay sinir ağları ile analizi ve eniyilemesi: Finansal tahmin algoritmaları.
73. Fırat, U., 2006. Kaotik zaman serilerinin yapay sinir ağlarıyla kestirimi: Deprem verisi durumu.
74. Kaynar, O., Tuna, M.F., Görmez, Y., and Deveci, M.A., 2017. Makine öğrenmesi yöntemleriyle müşteri kaybı analizi, **Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi**, **18(1)**:1–14.

75. Özata, M. and Aslan, Ş., 2004. Klinik karar destek sistemleri ve örnek uygulamalar, **Kocatepe Tıp Dergisi**, **5**(1).
76. Caliskan, A., Cil, Z.A., Badem, H., and Karaboga, D., 2020. Regression-based neuro-fuzzy network trained by ABC algorithm for high-density impulse noise elimination, **IEEE Transactions on Fuzzy Systems**, **28**(6):1084–1095.
77. Altaş, İ.H., 1999. Bulanık Mantık: Bulanıklılık Kavramı, **Enerji, Elektrik, Elektromekanik-3e**, **62**:80–85.
78. Yurtsal, A. and Kaynar, O. Sezgisel algoritmalar yardımıyla ders programı optimizasyonu, **Uluslararası Ekonomi ve Yenilik Dergisi**, **8**(1):1–18.
79. Aydoğan, E., 2014. Genetik programlama kullanılarak mobil zararlı yazılımların otomatik olarak üretilmesi.
80. Köylü, F., Çelik, M., and Karaboğa, D., 2013. Performance analysis of ABCMiner algorithm with different objective functions, *2013 21st Signal Processing and Communications Applications Conference (SIU)*, 1–5, IEEE.
81. Greene, D.P. and Smith, S.F., 1993. Competition-based induction of decision models from examples, **Machine Learning**, **13**:229–257.
82. Andraeo, R.V., Dorizzi, B., and Boudy, J., 2006. ECG signal analysis through hidden Markov models, **IEEE Transactions on Biomedical engineering**, **53**(8):1541–1549.
83. Kavzoğlu, T. and Çölkesen, İ., 2010. Karar ağaçları ile uydu görüntülerinin sınıflandırılması, **Harita Teknolojileri Elektronik Dergisi**, **2**(1):36–45.
84. Ozer, H.S., Demir, C., and Bilgin, G., 2019. Classification of cell types on histopathological images using local binary patterns, *2019 Medical Technologies Congress (TIPTEKNO)*, 1–4, IEEE.
85. Breiman, L., 2001. Random forests, **Machine learning**, **45**:5–32.

86. O'Brien, R. and Ishwaran, H., 2019. A random forests quantile classifier for class imbalanced data, **Pattern recognition**, **90**:232–249.
87. Ercan, U. et al., 2021. İnternette alıřveriř yapan hanelerin rastgele orman yöntemiyle tahmin edilmesi, **Kafkas Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, **12**(24):728–752.
88. Tařkın, G.G.E. and Çağatan, H., 2005. Veri madenciliğinde karar ağaçları ve bir satıř analizi uygulaması, **Eskiřehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi**, **6**(2):221–239.
89. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., et al., 2015. Xgboost: extreme gradient boosting, **R package version 0.4-2**, **1**(4):1–4.
90. Kılıç, S., 2013. Doğrusal regresyon analizi, **Journal of Mood Disorders**, **3**(2):90–92.
91. Alpar, C., 2016. Spor saęlık Ve eęitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik güvenilirlik.
92. Gabrali, D. and Aslan, Z., 2020. G"üneř enerjisi potansiyelinin çoklu lineer regresyon ve yapay sinir ağları ile modellenmesi, **AURUM Journal of Engineering Systems and Architecture**, **4**(1):23–36.
93. Damodar, N.G., 2004. Basic Econometrics, McGraw-Hill, New York, 4th edition.
94. Özen, N.S., Saraç, S., and Koyuncu, M., 2021. COVID-19 vakalarının makine öğrenmesi algoritmaları ile tahmini: Amerika Birleşik Devletleri örneęi, **Avrupa Bilim ve Teknoloji Dergisi**, (22):134–139.
95. Şenel, S. and Alatlı, B., 2014. Lojistik regresyon analizinin kullanıldıęı makaleler üzerine bir inceleme, **Journal of Measurement and Evaluation in Education and Psychology**, **5**(1):35–52.
96. Arseven, B. and ÇINAR, S.M. Dünya dıřı ıřınımlarla iyileřtirilmiş ARIMA, ridge regresyon ve lasso regresyon yöntemlerinin saatlik ıřınım tahmininde

kullanılması, **Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi**, **12**(3).

97. Çelik, Ş., 2021. Otomobil sahipliğini etkileyen faktörlerin farklı regresyon modelleri ile incelenmesi, **Journal of Original Studies**, **2**(2):79–85.
98. Deniz, Ö., 2005. Poisson regresyon analizi, **İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi**, **4**(7):59–72.
99. Işıl, M. and Bulut, Y.M., 2023. Gamma regresyon modelinde yanlı tahmin edicilerin karşılaştırılması, **Matematik ve Fen Alanında Uluslararası Araştırmalar IX**, 113.
100. Borchani, H., Varando, G., Bielza, C., and Larranaga, P., 2015. A survey on multi-output regression, **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, **5**(5):216–233.
101. Toprak, Ş., 2023. Makine öğrenmesi yöntemlerini kullanarak bir petrokimya firmasının hisse senedi fiyat tahmini= Stock price prediction of a petrochemical company using machine learning methods, Yüksek Lisans Tezi, Sakarya Üniversitesi.
102. Han, C., Phillips, P.C., and Sul, D., 2017. Lag length selection in panel autoregression, **Econometric Reviews**, **36**(1-3):225–240.
103. Abdi, H. et al., 2007. The method of least squares, **Encyclopedia of measurement and statistics**, **1**:530–532.
104. Seyyarer, E., Ayata, F., Uçkan, T., and Karci, A., 2020. Derin öğrenmede kullanılan optimizasyon algoritmalarının uygulanması ve kıyaslanması, **Computer Science**, **5**(2):90–98.
105. Şahinler, S., 2000. En küçük kareler yöntemi ile doğrusal regresyon modeli oluşturmanın temel prensipleri, **Mustafa Kemal Üniversitesi Ziraat Fakültesi Dergisi**, **5**(1-2):57–73.
106. Doğan, M., 2016. Lasso tahminlerinin genetik algoritma yaklaşımı ile bulunması, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü.



107. Aslan, Y., Yavasca, S., and Yasar, C., 2011. Long term Electric Peak load forecasting of Kutahya using different approaches, **International Journal on Technical and Physical Problems of Engineering**, **3**(2):87–91.
108. Chai, T. and Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE), **Geoscientific Model Development Discussions**, **7**(1):1525–1534.
109. Sertdemir, Y., 2022. Makine Öğrenme Yöntemlerinin Başarılarını Değerlendirirken Hangi Performans Ölçütleri Kullanılmalıdır, **23. Ulusal ve 6. Uluslararası Biyoistatistik Kongresi**, 38.
110. Van Rossum, G., Drake, F.L., et al., 1995. Python reference manual, Centrum voor Wiskunde en Informatica Amsterdam.
111. Python, W., 2021. Python, **Python Releases Wind**, **24**.
112. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python, **Journal of Machine Learning Research**, **12**:2825–2830.
113. Garita, M. and Garita, M., 2021. Using Stock Market Data in Python, **Applied Quantitative Finance: Using Python for Financial Analysis**, 71–83.
114. Ari, N. and Ustazhanov, M., 2014. Matplotlib in python, *2014 11th International Conference on Electronics, Computer and Computation (ICECCO)*, 1–6, IEEE.
115. Alexander, C., Heck, D.F., and Kaeck, A., 2022. The role of binance in bitcoin volatility transmission, **Applied Mathematical Finance**, **29**(1):1–32.
116. Çınar, M., 2003. Veri ambarlarında verilerin temizlenmesi, Doktora Tezi, Fen Bilimleri Enstitüsü.
117. Oğuzlar, A., 2003. Veri ön işleme, **Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, (21).

118. Budak, H., 2018. Özellik seçim yöntemleri ve yeni bir yaklaşım, **Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi**, **22**:21–31.
119. Torgo, L., Ribeiro, R.P., Pfahringer, B., and Branco, P., 2013. Smote for regression, *Progress in Artificial Intelligence: 16th Portuguese Conference on Artificial Intelligence, EPIA 2013, Angra do Heroísmo, Azores, Portugal, September 9-12, 2013. Proceedings 16*, 378–389, Springer.
120. Mercer, G., 2012, Volume and divergence, (Web Sayfası: <http://traders.com/>), Erişim tarihi: 16/5/2023.
121. Zaman, M., 2006. Finansal araçların alım satım kararında teknik analiz ve mum grafikler, Doktora Tezi, Marmara Üniversitesi (Türkiye).

## ÖZGEÇMİŞ

### KİŞİSEL BİLGİLER

Adı, Soyadı : Mustafa ÜLKER

Uyruğu : Türkiye (T.C.)

### EĞİTİM

Yüksek Lisans, Erciyes Üniversitesi, Kayseri, Bilgisayar Mühendisliği, 2023

Lisans, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, Bilgisayar Mühendisliği,  
2006

Lise, Nuh Mehmet Baldöktü Anadolu Lisesi, Kayseri, 2002

### İŞ DENEYİMİ

Bilgi İşlem Uzmanı, Kocasinan Belediyesi, Kayseri (devam)

### YABANCI DİL

İngilizce, Fransızca

### YAYIN

F. Köylü, M. Ülker, "Makine Öğrenmesine Dayalı Bitcoin Zaman Serisi Analizi ve Fiyatı Tahmini", 2023 Innovations in Intelligent Systems and Applications Conference (ASYU), Sivas, Turkey, 2023, pp. 1-5.