

**MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ**

**MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE  
BITCOIN TREND DÖNÜŞLERİNİN TAHMİN EDİLMESİ**

**YÜKSEK LİSANS TEZİ**

**Sergül ÜRGENÇ**

**İstatistik Anabilim Dalı**

**İstatistik Programı**

**Tez Danışmanı: Prof. Dr. Barış AŞIKGİL**

..... tarafından hazırlanan ..... adlı bu tezin ..... tezi olarak uygun olduğunu onaylarım.

.....  
Tez Yöneticisi

Bu çalışma, jürimiz tarafından ..... Anabilim Dalında ..... tezi olarak kabul edilmiştir.

Başkan : \_\_\_\_\_

Üye : \_\_\_\_\_

Üye : \_\_\_\_\_

Üye : \_\_\_\_\_

Üye : \_\_\_\_\_

Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi ..... Enstitüsü tez yazım kurallarına uygundur.

Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım klavuzuna uygun olarak hazırladığım bu tez çalışmada;

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel etik kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- ücret karşılığı başka kişilere yazdırmadığımı (dikte etme dışında), uygulamalarımı yaptırmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

....

*Üniversite hayatım boyunca bana yol gösteren değerli hocalarım Prof. Dr. Barış AŞIKGİL, Prof. Dr. Ayça ÇAKMAK PEHLİVANLI ve Dr. Öğr. Üyesi Elif Özge ÖZDAMAR'a, başta Ümit SİLE olmak üzere güvenini ve desteğini her zaman hissettiğim JFORCE ailesine, kariyer yolumda bilgi birimini asla esirgemeyen Umut ŞATIR GÜRBÜZ'e büyük desteklerinden dolayı teşekkür ederim.*

# MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE BITCOIN TREND DÖNÜŞLERİNİN TAHMİN EDİLMESİ

## ÖZET

Son yıllarda Bitcoin, kripto para birimi piyasasında en popüler dijital varlık haline gelmiştir. Fiyatları, hızla artan yatırımcı ilgisi nedeniyle son derece dalgalıdır ve bu durum fiyat hareketlerini öngörmeyi zorlaştırmaktadır. Bu bağlamda, Bitcoin fiyat yönünün tahmini, yatırımcıların yatırım stratejilerini belirlemelerine yardımcı olabilmektedir. Finansal piyasalarda fiyat yönünün tahmininde ve alım satım kararlarında makine öğrenimi teknikleri yaygın olarak kullanılmaktadır. Zamana bağlı veriler ve finansal indikatörler kullanılarak gelecekteki fiyat hareketlerini tahmin etmek için makine öğrenmesi modelleri geliştirilebilmektedir. Bitcoin fiyat hareketleri, kripto para birimi piyasasında çok sayıda faktörün etkisi altında olabilmektedir. Bu nedenle, makine öğrenmesi tekniklerinin doğruluğunun ve başarısının belirlenmesi önemlidir.

Bu çalışmanın amacı, Bitcoin fiyat hareketleri üzerindeki trend dönüşlerini tahmin etmek için makine öğrenmesi tekniklerini kullanmak ve bu tekniklerin başarı oranlarını karşılaştırmaktır. Bu amaçla, çalışmada trend yönünün değiştiği noktalara odaklanılmıştır. Fiyat yönünün hangi noktada değişeceği tespit edilerek, 'al', 'sat' ve 'bekle' olmak üzere üç sınıflı tahminleme çalışması yapılmıştır. Bu amaçla, 'binance.com' üzerinden 1 Eylül 2017 ile 1 Nisan 2023 tarihleri arasında alınan günlük veriler ile öncelikle fiyatın açılış (open), kapanış (close), günün en yüksek değeri (high), günün en düşük değeri (low) ve Bitcoin ticaret hacmi (volume) bilgileri kullanılarak trend tespiti yapılmış ve finansal göstergeler hesaplanmıştır. Veri inceleme aşamasından sonra bağımsız değişkenlerin çoklu bağlantı problemine yol açmaması amacıyla temel bileşenler analizi ile boyut küçültme çalışması yapılmıştır. 'al', 'sat' ve 'bekle' sınıfları alt örnekleme (under sampling) yöntemi ile dengelenerek, Aşırı Gradyan Arttırma (XGB), Rastgele Orman (RF), Rastgele Ağaçlar (RT), Yapay Sinir Ağları (ANN), Destek Vektör Makineleri (SVM), Bayes Ağları (BN) ve K En Yakın Komşu (KNN) modelleri geliştirilmiş ve sonuçlar,

kesinlik (precision), duyarlılık (recall), özgülük (specificity), F1 puanı (F1 score) ve doğruluk (accuracy) başarı kriterlerine göre değerlendirilmiştir. ANN, SVM, BN ve KNN modellerinin ağaç tabanlı algoritmalara göre çok daha düşük performans gösterdiği görüldüğünden, çalışmada sadece XGB, RF ve RT model sonuçlarının karşılaştırması verilmiştir. Çalışma sonucunda, XGB modelinin diğer modellere göre daha yüksek başarıya sahip olduğu görülmüştür.

**Anahtar Kelimeler:** Bitcoin trend tahmini, kripto para fiyat analizi, makine öğrenmesi, sınıflandırma

# **PREDICTING BITCOIN TRENDS REVERSALS WITH MACHINE LEARNING METHODS**

## **ABSTRACT**

In recent years, Bitcoin has become the most popular digital asset in the cryptocurrency market. Its prices are highly volatile due to rapidly increasing investor interest, making it difficult to predict price movements. In this context, predicting the direction of Bitcoin prices can assist investors in determining their investment strategies. Machine learning techniques are widely used in predicting price directions and making trading decisions in financial markets. Machine learning models can be developed using time-dependent data and financial indicators to forecast future price movements. Bitcoin price movements can be influenced by numerous factors in the cryptocurrency market. Therefore, determining the accuracy and success of machine learning techniques is crucial.

The aim of this study is to predict trend reversals in Bitcoin price movements by using machine learning techniques and compare the success rates of these techniques. For this purpose, the study focuses on points where the trend changes. A three-class prediction study is conducted in order to classify the price direction as 'buy', 'sell', or 'hold'. Daily data collected from 'binance.com' between September 1, 2017, and April 1, 2023, including the open, close, high, low prices and Bitcoin trading volume are used to identify trends and calculate financial indicators. After the data examination stage, dimension reduction is performed by using principal component analysis to avoid multicollinearity issues of independent variables. The 'buy', 'sell', and 'hold' classes are balanced through under-sampling. Extreme Gradient Boosting (XGB), Random Forest (RF), Random Tree (RT), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Bayesian Networks (BN), and K-Nearest Neighbors (KNN) models are developed. The results are evaluated by using precision, recall, specificity, F1 score and accuracy metrics. Since ANN, SVM, BN, and KNN models show significantly lower performance compared

to tree-based algorithms, the comparison in the study focuses on the results of XGB, RF, and RT models. The study concludes that the XGB model exhibits higher success compared to other models.

**Keywords:** Bitcoin trend prediction, cryptocurrency price analysis, machine learning, classification



## İÇİNDEKİLER

### Sayfa

<b>ÖZET</b> .....	<b>vi</b>
<b>ABSTRACT</b> .....	<b>viii</b>
<b>İÇİNDEKİLER</b> .....	<b>x</b>
<b>KISALTMALAR</b> .....	<b>xii</b>
<b>ÇİZELGE LİSTESİ</b> .....	<b>xiii</b>
<b>ŞEKİL LİSTESİ</b> .....	<b>xiv</b>
<b>1. GİRİŞ</b> .....	<b>15</b>
<b>2. KRİPTO PARA VE TEKNİK GÖSTERGELER</b> .....	<b>20</b>
2.1 Trend Göstergeleri .....	21
2.1.1 Basit hareketli ortalama .....	21
2.1.2 Üstel hareketli ortalama .....	22
2.1.3 Hareketli ortalamanın yakınsama ve ıraksaması.....	22
2.2 Momentum Göstergeleri .....	23
2.2.1 Göreceli güç endeksi .....	24
2.2.2 Stokastik osilatör .....	24
2.3 Volatilité Göstergeleri .....	25
2.3.1 Volatilité endeksi .....	25
2.3.2 Bollinger bantları .....	25
2.4 Hacim Göstergeleri .....	26
2.4.1 Denge hacmi .....	26
<b>3. YÖNTEMLER</b> .....	<b>27</b>
3.1 Makine Öğrenmesi .....	27
3.2.1 Aşırı gradyan arttırma .....	29
3.2.2 Rastgele orman.....	30
3.2.3 Rastgele ağaçlar .....	30
3.2.4 Yapay sinir ağları .....	31
3.2.5 Destek vektör makineleri .....	31
3.2.6 K en yakın komşu .....	32
3.2.7 Bayes ağları.....	32
3.2 Özellik Seçimi .....	33
3.3 Temel Bileşenler Analizi.....	34
3.4 Yeniden Örnekleme .....	35
3.4.1 Alt örnekleme.....	35
3.4.2 Aşırı örnekleme.....	35
3.5 Sınıflandırma Performans Ölçüleri .....	36
3.5.1 Doğruluk .....	36
3.5.2 Kesinlik .....	36
3.5.3 Duyarlılık .....	37
3.5.4 F1 puanı .....	37
3.5.5 Seçicilik.....	37

3.5.6 Alıcı çalışma karakteristiđi eğrisi .....	37
<b>4. UYGULAMA.....</b>	<b>38</b>
4.1 Veri Kümesi .....	38
4.2 Özellik Seçimi .....	39
4.3 Temel Bileşenler Analizi.....	40
4.4 Veri Dengeleme .....	40
4.5 Model Eğitimi ve Sonuçları .....	42
<b>5. SONUÇ .....</b>	<b>46</b>
<b>KAYNAKLAR .....</b>	<b>48</b>
<b>EKLER .....</b>	<b>51</b>
<b>ÖZGEÇMİŞ.....</b>	<b>Error! Bookmark not defined.</b>

## KISALTMALAR

<b>XGB</b>	: Aşırı Gradyan Arttırma modeli (XGBoost/eXtreme Gradient Boosting)
<b>RF</b>	: Rastgele Orman (Random Forest)
<b>RT</b>	: Rastgele Ağaçlar (Random Trees)
<b>ANN</b>	: Yapay Sinir Ağları (Artificial Neural Networks)
<b>SVM</b>	: Destek Vektör Makineleri (Support Vector Machines)
<b>BN</b>	: Bayes Ağları (Bayesian Networks)
<b>LR</b>	: Lojistik Regresyon (Logistic Regression)
<b>GB</b>	: Gradyan Arttırma (Gradient Boosting)
<b>NB</b>	: Naive Bayes
<b>ARIMA</b>	: Otoregresif Bütünleşik Hareketli Ortalama (Autoregressive Integrated Moving Average)
<b>VAR</b>	: Vektör Otoregresyon (Vector Autoregression)
<b>LSTM</b>	: Uzun Kısa Dönemli Bellek (Long Short-Term Memory)
<b>RNN</b>	: Yinelemeli Sinir Ağları (Recurrent Neural Network)
<b>CNN</b>	: Evrişimli Sinir Ağları (Convolutional Neural Network)
<b>DNN</b>	: Derin Sinir Ağları (Deep Neural Network)
<b>GRU</b>	: Geçitli Tekrarlayan Birim (Gated Recurrent Unit)
<b>MLP</b>	: Çok Katmanlı Algılayıcı (Multilayer Perceptron)
<b>Res-Net</b>	: Artık Ağlar (Residual Network)
<b>PCA</b>	: Temel Bileşenler Analizi (Principal Component Analysis)
<b>BTC</b>	: Bitcoin
<b>RSI</b>	: Göreceli Güç Endeksi (Relative Strength Index)
<b>MACD</b>	: Hareketli Ortalama Yakınsama ve Iraksama (Moving Average Convergence Divergence)
<b>SMA</b>	: Basit Hareketli Ortalama (Simple Moving Average)
<b>EMA</b>	: Üstel Hareketli Ortalama (Exponential Moving Average)
<b>STO</b>	: Stokastik Osilatör (Stochastic Oscillator)
<b>VIX</b>	: Volatilite Endeksi (Volatility Index)
<b>OBV</b>	: Denge Hacmi (On Balanced Volume)
<b>ACC</b>	: Alıcı Çalışma Karakteristiği Eğrisi (Receiver Operating Characteristic)

## ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 4.1 : Bağımlı değişken kuralları .....	38
Çizelge 4.2 : Bağımlı değişkenin sınıf dağılımı .....	41
Çizelge 4.3 : Tüm sınıfların eşit olduğu dağılım .....	41
Çizelge 4.4 : Bekle sınıfının yüksek olduğu dağılım .....	41
Çizelge 4.5 : RF modelinin hiperparametreleri .....	42
Çizelge 4.6 : XGB modelinin hiperparametreleri.....	42
Çizelge 4.7 : RT modelinin hiperparametreleri .....	43
Çizelge A.1 : Özellikler ve veri sözlüğü.....	51
Çizelge C.1 : Hiperparametre açıklamaları .....	55

## ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 4.1 : Bağımsız değişken listesi. ....	39
Şekil 4.2 : Temel bileşenler analizi varyans tablosu. ....	40
Şekil 4.3 : Birinci veri kümesinin eğitim ve test sonuçları.....	43
Şekil 4.4 : İkinci veri kümesinin eğitim ve test sonuçları .....	44
Şekil B.1 : Bileşen matrisi .....	54

## 1. GİRİŞ

Verinin işlenmesi ve analiz edilmesi, günümüzde birçok sektörde önemli bir rol oynamaktadır. Finansal piyasalar da bu sektörlerden biridir ve son yıllarda giderek artan dijitalleşme ile birlikte, veri analizi ve makine öğrenmesi teknikleri finansal piyasaların yönetiminde daha sık kullanılmaktadır. Kripto para birimleri de finansal piyasalardaki dijitalleşme trendinin bir parçasıdır. Özellikle Bitcoin, son yıllarda popülerliği artan kripto para birimleri arasında öne çıkmaktadır. Ancak, Bitcoin fiyatları son derece dalgalı olduğundan yatırımcılar Bitcoin'in trend dönüşlerini tahmin etmek için farklı yöntemler aramaktadırlar.

Literatür incelendiğinde, kripto para üzerine giderek artan bir araştırma alanı olduğu görülmektedir. BTC fiyat yönünün tahmin edilmesinde sınıflandırma çalışmaları regresyon çalışmalarına göre çok daha az sayıdadır. Sınıflandırma çalışmaları, çoğunlukla fiyat yönünün bir sonraki periyot için tahmin edilmesi ile sınırlı kalmakta ve trend yönü ve dönüşümlerini nadiren kapsamaktadır. Yapılan çalışmalarda, LSTM ve RNN gibi derin öğrenme modellerinin ve takviyeli öğrenme yönteminin yaygın olarak kullanıldığı görülmektedir.

İnce'nin (2019) yapmış olduğu yüksek lisans tez çalışmasında, tek yönlü işlem komisyon oranı kullanılarak oluşturulan 'al', 'sat' ve 'bekle' sınıflarını açılış, kapanış, en yüksek, en düşük ve hacim verilerine karşı test etmek amacı ile özellik kümesinde toplamda 156 adet teknik gösterge, matematiksel dönüşüm ve finansal kalıp kullanılmıştır. Çalışmada, DNN, LSTM ve GRU modelleri geliştirilmiş ve başarıları kıyaslanmıştır. Tüm derin öğrenme modellerinin, teknik göstergeler kullanılarak geliştirildiğinde daha başarılı sonuç verdiği görülmüştür. LSTM, %56.33'lük genel bir doğrulukla üç model arasında en başarılısı olmuştur.

Ji ve arkadaşlarının (2019), BTC fiyatını ve fiyatın bir önceki güne göre iniş veya çıkış yapacağını tahminlemek amacıyla blok zinciri (blockchain) verilerinden yararlanılmıştır. Ortalama blok boyutu, blok zinciri boyutu, tahmini işlem hacmi, onay süresi ve madencilerin geliri gibi özelliklerden yararlanılarak DNN, LSTM,

CNN, Res-Net modelleri ve bunların kombinasyonları geliştirilmiştir. BTC fiyat tahmini için LSTM, sınıflandırma tahmini için DNN modeli daha başarılı sonuç vermiştir. Ek olarak çalışmada, karlılık açısından sınıf tahmininin fiyat tahmininden daha etkili olduğu görülmüştür.

Qiang ve Shen'in (2021) yaptığı çalışmada, dakika seviyesindeki teknik göstergeler kullanılarak BTC fiyatlarının yüksek frekanslı trend tahminlerini bir adım önde gerçekleştirmek için sınıflandırma algoritmaları kullanılmıştır. Bu çalışma, geçmiş verilerden sinyaller yakalayıp yatırımcılar için en uygun BTC ticaret stratejilerini oluşturmayı amaçladığından, tahmin doğruluğuna ek olarak finans sektöründe yaygın olarak kullanılan net varlık değeri (NAV) ve sharpe oranı (SR) metrikleri kullanılmıştır. Bu bağlamda, fiyat iniş ve çıkışlarını tahmin etmek amacıyla CNN ve LSTM içeren hibrit bir derin öğrenme modeli geliştirilmiştir. NAV ve sharpe oranları farklı hiperparametrelerle hesaplanarak elde edilen hibrit model ile %53 doğruluk elde edilmiştir.

Cavalli ve Amoretti'nin (2021) BTC fiyatının iniş ve çıkışlarını günlük olarak tahmin etmek amacıyla yaptıkları çalışmada sosyal medya verileri, blok zinciri işlem verileri ve finansal göstergelerden yararlanılmıştır. Tek boyutlu evrişimli sinir ağı (1D CNN) modeli geliştirilerek sonucu CNN, LSTM ve standart makine öğrenimi algoritmaları ile karşılaştırılmıştır. 1D CNN modeli %74.2 doğruluk oranında yüksek bir performans göstermiştir.

A. Monsalve ve arkadaşları (2020), yaptıkları çalışmada MLP yerine CNN modelinin uygunluğu araştırılmıştır. Çalışmanın deneysel bölümü, yüksek frekansta teknik analiz kullanarak altı popüler kripto para biriminin (Bitcoin, Dash, Ethereum, Litecoin, Monero ve Ripple) Dolar para birimi karşısında bir sonraki dakikada değerinin artıp artmayacağını tahmin etmek için farklı sinir ağı mimarilerinin performansını karşılaştırmaktadır. RSI, MACD ve SMA gibi toplamda 18 teknik gösterge ile geliştirilen CNN modelleri MLP modellerine göre önemli ölçüde daha iyi performans göstermiştir.

Livieris ve arkadaşlarının (2020) Bitcoin, Ethereum ve Ripple fiyatlarını ve fiyat iniş-çıkış zamanlarını tahmin etmek amacıyla hem regresyon hem sınıflandırma yöntemlerinden yararlanmışlardır. Çalışmada geliştirilen, LSTM, Çift Yönlü LSTM (BiLSTM), CNN modelleri en yaygın kullanılan çeşitli topluluk öğrenmesi (ensemble) stratejileri ile birleştirilmiştir. Önerilen topluluk modelleri kullanılan

derin öğrenme modellerinin kombinasyonlarını içermektedir. Sonuçlar, derin öğrenme modellerinin topluluk öğrenme yöntemi ile birleştirilmesinin, tek başına bir derin öğrenme modeline göre tahmin doğruluğunu arttırdığını göstermiştir.

Cohen'in (2020) yaptığı çalışmada, BTC fiyat trendlerini tahmin etmek için kullanılan iki farklı yöntemin yetenekleri incelenmiştir. En iyi tahmin kombinasyonlarını bulmak için parçacık sürü optimizasyonu (PSO) kullanılmıştır. Minimum ve maksimum gerileme, maksimum karlı işlemlerin yüzdesi, kar faktörü ve net kar ölçümleri kullanılarak doğrusal regresyon ve Darvas kutusu yöntemleri ile tahminleme yapılmıştır. Bu fiyat tahmininden yola çıkarak fiyatın iniş ve çıkışları hesaplanmıştır. Sonuçlar, BTC fiyat değişimlerinin etkin piyasa hipotezini takip etmediğini ve hem Darvas kutusu hem de doğrusal regresyon yöntemlerinin, BTC'nin fiyat trendlerini tahmin etmede yardımcı olabileceğini göstermiştir.

Chen ve arkadaşları (2017), artan ve azalan Ethereum fiyatını tahmin etmek amacıyla LR, NB, SVM, RF, RNN, ANN ve ARIMA modellerini karşılaştırmışlardır. Tüm modellerin %50'nin üstünde başarıya sahip olduğu görülmüştür. En iyi performansı %61.17 doğruluk oranı ile ARIMA modeli vermiştir.

Akyıldırım ve arkadaşları (2021), kripto para birimleri fiyatlarının iniş ve çıkışlarını günlük olarak tahmin etmeyi amaçladıkları çalışmada, SVM, LR, ANN ve RF modelleri geliştirilmiş ve performansları karşılaştırılmıştır. En likit on iki kripto para birimi, günlük ve dakika bazında kullanılarak, geçmiş fiyat bilgileri ve teknik göstergelerin kullanıldığı makine öğrenimi sınıflandırma algoritmaları ile tahmin edilebilirliği analiz edilmiştir. Dört modelin ortalama sınıflandırma doğruluğu, tüm kripto para birimleri ve zaman dilimleri için %50 eşiğinin üzerindedir. SVM, %55–65 ile en iyi ve tutarlı sonuçları vermiştir.

Valencia ve arkadaşlarının (2019) yaptığı çalışmada, Bitcoin, Ethereum, Ripple ve Litecoin gibi kripto para piyasalarının fiyat hareketini tahmin etmek için yaygın olarak kullanılan makine öğrenimi araçları ve sosyal medya verilerinin kullanımını önerilmiştir. Twitter verileri ve piyasa verileri hem ayrı ayrı hem de birlikte kullanılarak model başarıları karşılaştırılmıştır. Fiyat iniş ve çıkışlarını tahmin etmek amacıyla ANN, SVM ve RF modelleri geliştirilmiştir. ANN modeli diğer modellere göre daha iyi performans göstermiştir ve Twitter verilerinin tek başına belirli kripto paraların tahmin edilmesinde kullanılabileceği sonucuna ulaşılmıştır.



Kwon ve arkadaşları (2019), Bitcoin, Ethereum, Ripple, Bitcoin Cash, Litecoin, Dash ve Ethereum Classic kripto para birimlerinin fiyat yönünü tahmin etmek için yaptıkları çalışmada, sadece açılış, kapanış, en yüksek fiyat, en düşük fiyat ve hacim değişkenleri kullanılmıştır. Bu amaçla, LSTM ve GB modelleri geliştirilmiştir. Sonuç olarak, LSTM modeli ile GB modeline göre yaklaşık %7'lik bir performans artışı elde edilmiştir.

Atçeken'in (2021) tez çalışmasında, BTC fiyatının 24 saat içerisinde farklı eşik değerlerden fazla yükselip yükselmeyeceğinin tahmin edilmesi amaçlanmıştır. Eşik değerleri için farklı yüzdelerde (%2, %3, %5, %10) kar hedefleri konulmuştur. Finansal göstergeler, destek ve direnç seviyeleri, en yüksek ve düşük değerler, açılış ve kapanış fiyatları kullanılarak regresyon ve sınıflandırma çalışmaları yapılmıştır. Regresyon için KNN, SVM, RF, ANN ve VAR modelleri, sınıflandırma için LR, NB, KNN, RF, SVM ve ANN modelleri geliştirilmiştir. RF, bu çalışmada kullanılan diğer modellere göre regresyon ve sınıflandırmada daha iyi sonuçlar vermiştir. En doğru performansı gösteren kar hedefinin %5 olduğu görülmüştür.

Shintate ve Pichl'in (2019) yaptıkları çalışmada, durağan olmayan kripto para birimi serisi için derin öğrenme tabanlı bir trend tahmin sınıflandırma yapısı olan rastgele örnekleme yöntemi önerilmiştir. Okcoin piyasasındaki istikrarsız BTC fiyatları üzerinde iki klasik temel yöntem yaklaşımlarının performansı karşılaştırılmış ve modelin sınıf dengesizliği sorununu azalttığı bulunmuştur. Dakikalık fiyat serisinde, yükseliş, alçalış ve sabit olma durumları için üç sınıflı tahminleme çalışması yapılmıştır. Yanıt yüzey metodolojisine (RSM) dayalı kar oranları LSTM modeline dayalı olanlardan daha iyi sonuçlar vermiştir. Durağanlık varsayımının oldukça uygun olduğu alanda iyi bir performans göstermektedir. Çok küçük ölçekli modellerle deneyler yapılmış ve yöntemin etkisini ayırt etmek için MLP ve LSTM modelleri ile karşılaştırarak RSM üstünlüğü doğrulanmıştır.

Bu çalışmanın amacı, bir sonraki gün için alım, satım veya bekleme kararlarının BTC fiyatının bugünün kapanış zamanında makine öğrenmesi yöntemleri ile verilmesidir. Çalışmada, BTC fiyat trendi tahmininde makine öğrenmesi tekniklerinin ve teknik göstergelerin etkinliği incelenmektedir. Bu amaçla günlük BTC verileri kullanılarak BTC fiyatına etki eden faktörler incelenmiş ve makine öğrenmesi teknikleriyle tahminleme modelleri geliştirilmiştir. Modeller, 'al', 'sat' ve 'bekle' olmak üzere üç sınıflı tahminleme yapacak şekilde eğitilmiştir. Finansal

göstergeler hesaplanarak veri kümesi oluşturulmuştur. Veri boyutu küçültülmüş ve sınıflar dengelenmiştir. Eğitilen modellerin performansı farklı başarı kriterleri kullanılarak değerlendirilmiştir.

## 2. KRİPTO PARA VE TEKNİK GÖSTERGELER

Günümüzün en popüler finansal varlıklarından biri olan kripto para birimleri, son yıllarda hızla yayılmış ve yatırımcıların ilgisini çekmiştir. Kripto para birimleri, dijital bir para birimi olarak adlandırılan ve çevrimiçi işlemler için kullanılan bir tür para birimidir. Blok zinciri adı verilen teknoloji ile desteklenen kripto para birimleri, merkezi olmayan bir yapıya sahiptir ve devletler ya da bankalar tarafından kontrol edilmez.

Bitcoin, en bilinen kripto para birimi olmakla birlikte en büyük özelliklerinden biri, merkezi bir otoritenin olmamasıdır. Bitcoin, açık kaynaklı bir yazılım olan Bitcoin ağı üzerinde çalışır ve kullanıcılar arasında doğrudan para transferi yapılmasına olanak tanır. Bununla birlikte, blok zinciri teknolojisi, tüm işlemlerin güvenli ve değiştirilemez bir şekilde kaydedilmesini sağlar. Diğer para birimleri gibi yatırım seçeneği olarak düşünülebilir, ancak risklerinin farkında olmak ve yatırım kararlarına özen göstermek önemlidir.

Teknik göstergeler, yatırım yapmak için kullanılan önemli araçlardan biridir. Bir hisse senedi veya diğer finansal varlıkların geçmiş performanslarından yararlanarak, fiyat eğilimlerini ve potansiyel trendleri ölçen finansal hesaplamalardır. Teknik göstergeler, yatırımcılara belirli bir varlığın fiyatının ne yöne gidebileceği hakkında fikir verir. Bu nedenle, yatırımcılar genellikle teknik göstergeleri kullanarak gelecekteki fiyat hareketlerini tahmin etmeye çalışırlar. Bu göstergeler, bir varlığın fiyat hareketi hakkında bilgi sağlar, ancak her zaman doğru olmayabilir. Göstergeler, piyasalardaki herhangi bir gelişmeye karşı kırılgan olabilirler ve yatırımcılar, yalnızca teknik göstergeleri kullanarak yatırım yapmak yerine temel analiz ve diğer araçları da kullanarak daha doğru sonuçlar elde edebilirler (İnce, 2019).

Teknik göstergelerin kullanımı finansal analiz ve yatırım dünyasında oldukça yaygındır. Teknik analiz, bir varlığın geçmiş fiyat hareketlerine dayanarak gelecekteki fiyat hareketlerini tahmin etmeyi amaçlayan bir analiz yöntemidir. Teknik göstergeler, bu analiz yönteminin bir parçası olarak kullanılan matematiksel hesaplamalar veya grafiksel gösterimlerdir. Bu sayede, bir varlığın fiyat

hareketlerindeki belirli trendleri veya kalıpları göstererek yatırımcılara gelecekteki fiyat hareketlerini tahmin etme imkanı sağlar. Teknik göstergeler, yatırımcılara hem alım hem de satım sinyalleri verebilir ve yatırımcıların daha bilinçli kararlar vermelerine yardımcı olabilir (Rockefeller, 2011).

## **2.1 Trend Göstergeleri**

Trend, belirli bir varlığın fiyat hareketlerindeki genel yönü ifade eder. Bu yön, fiyatların belirli bir süre boyunca yükseldiği veya düştüğü bir eğilimi ifade eder. Trendler genellikle, uzun vadeli yatırım stratejileri için kullanılan bir analiz aracıdır ve yatırımcılara bir varlığın fiyat hareketlerinin gelecekteki yönü hakkında fikir verir.

Yükselen trend, fiyatların genellikle belirli bir süre boyunca arttığı bir piyasa koşulunu ifade eder. Bu süre boyunca fiyatlar, önceki zirvelerin üzerine çıkarak yükselir ve yatırımcılar genellikle fiyatların yükselmeye devam edeceği beklentisiyle pozisyon alırlar.

Düşen trend ise, fiyatların genellikle belirli bir süre boyunca azaldığı bir piyasa koşulunu ifade eder. Bu süre boyunca fiyatlar, önceki diplerin altına düşerek azalır ve yatırımcılar genellikle fiyatların düşmeye devam edeceği beklentisiyle pozisyon alırlar. Trendler, teknik analizde birçok gösterge ve araçla birlikte kullanılır (Rockefeller, 2011).

### **2.1.1 Basit hareketli ortalama**

Hareketli ortalamalar, fiyatların belirli bir zaman dilimi boyunca ortalama değerini hesaplayan göstergelerdir. Bu gösterge, fiyatların belirli bir süre boyunca yönünü belirlemek için kullanılır. Kısa vadeli hareketli ortalamalar (5, 10, 20 günlük) daha hızlı hareket ederken, uzun vadeli hareketli ortalamalar (50, 100, 200 günlük) daha yavaş hareket eder. Basit hareketli ortalama, kapanışı kullanır. Bunun nedeni, kapanışın dönem hareketinin ve duyarlılığının özetini ifade etmesidir.

Hareketli ortalama, fiyatları takip etmesinin ve fiyat değişimlerini düz bir çizgiden daha iyi temsil etmesinin yanı sıra ara sıra meydana gelen düzensiz fiyatları da yumuşatır. Çapraz geçiş kuralı (crossover rules), fiyatın hareketli ortalama çizgisinin üzerine çıktığı noktada satın alınması ve altına düştüğü noktada satılması gerektiğini belirten bir sinyaldir. Geçiş kuralı düşük seviyelerden alıp yüksek seviyelerden

satmayı sağlayarak kar oluşturur (Gallegos-Erazo, 2022). Basit hareketli ortalamanın formülü denklem 2.1’de verilmiştir:

$$SMA = \sum_i^N Kapanış / N \quad (2.1)$$

Burada kapanış, kapanış fiyatını ve N, periyot sayısını ifade etmektedir.

### 2.1.2 Üstel hareketli ortalama

Üstel hareketli ortalama, teknik analizde sıklıkla kullanılan bir göstergedir. SMA göstergesine benzer, ancak fiyat hareketlerinin son fiyatlara göre daha fazla ağırlıklandırıldığı bir hesaplama yöntemi kullanır. EMA’nın ilk hesaplaması, SMA’nın ilk hesaplaması gibi yapılır. İlk EMA değeri, veri serisinin ilk periyotundaki SMA değerine eşittir. Sonraki EMA değeri, denklem 2.2’deki denklem kullanılarak hesaplanır (Chong, Ng ve Liew, 2014).

$$EMA = (K \times (Fiyat - EMA_{t-1})) + EMA_{t-1} \quad (2.2)$$

Burada fiyat, hareketli ortalaması hesaplanacak olan fiyat verisini,  $EMA_{t-1}$ , bir önceki EMA değerini ve K, düzeltme faktörünü temsil etmektedir. K, genellikle  $2 / (\text{periyot sayısı} + 1)$  şeklinde hesaplanır.

### 2.1.3 Hareketli ortalamanın yakınsama ve ıraksaması

MACD, fiyat hareketlerindeki trendleri belirlemek ve momentumu ölçmek için hareketli ortalamaların kullanıldığı bir göstergedir. Bu nedenle, fiyatların trendini takip eden bir gösterge olarak kabul edilir. Ancak aynı zamanda, fiyatların hızını ve ivmesini ölçerek momentumu da hesaplar. Bu gösterge, kısa vadeli ve uzun vadeli hareketli ortalamalar arasındaki farkı hesaplar ve bu farkın hareketli ortalaması alınır. MACD, alım satım sinyalleri için de kullanılır. Teknik analizde hem trend hem de momentum göstergesi olarak kabul edilir. Bu nedenle, birçok yatırımcı tarafından her iki kategoriye de dahil edilir. MACD, hareketli ortalamaların birbirleriyle ilişkisini ölçen bir göstergedir. MACD hesaplaması için genellikle 9, 12 ve 26 periyotluk üstel hareketli ortalamalar kullanılır (Chong, Ng ve Liew, 2014). MACD ve MACD sinyal hattı hesabı sırasıyla denklem 2.3 ve 2.4’te gösterilmiştir.

$$MACD = EMA_{t-12} - EMA_{t-26} \quad (2.3)$$

$$\text{Sinyal Hattı} = EMA(MACD)_{t-9} \quad (2.4)$$

Burada,  $EMA_{t-12}$ , fiyatın 12 periyotluk üstel hareketli ortalamasını,  $EMA_{t-26}$ , fiyatın 26 periyotluk üstel hareketli ortalamasını ve  $EMA(MACD)_{t-9}$ , MACD'nin 9 periyotluk üstel hareketli ortalaması ifade etmektedir. MACD histogramı, MACD çizgisinin  $EMA(9)$ 'a olan farkını gösterir. Histogramın yukarı yönlü bir hareketi, MACD çizgisinin sinyal hattını yukarı kesmesiyle oluşur. Histogramın aşağı yönlü bir hareketi ise MACD çizgisinin sinyal hattını aşağı kesmesiyle oluşur.

MACD göstergesi, trendi takip etmek ve alım-satım sinyalleri üretmek için kullanılır. Özellikle, MACD çizgisinin sinyal hattını yukarı yönlü kesmesi alış sinyali olarak kabul edilir. Tam tersi, MACD çizgisinin sinyal hattını aşağı yönlü kesmesi satış sinyali olarak kabul edilir (Gallegos-Erazo, 2022).

## 2.2 Momentum Göstergeleri

Momentum, fiyatın değişim hızının gücüdür. Momentum bir dizi teknik göstergelyi kapsayan genel bir terimdir, ancak birçok momentum göstergesi isimlerinde momentum kelimesini kullanmaz. Momentum ve değişim oranı terimleri teknik analizde birbirinin yerine kullanılır. Momentum öncü bir göstergedir ve hareketli ortalamalar gibi trend izleyen göstergelerin doğasında bulunan gecikmeyi dengelemek için kullanılabilir. Momentum göstergeleri, piyasa yana doğru hareket ederken yaklaşmakta olan bir hareketi tespit etmede başarılıdır ve belirli momentum önlemleri, bir menkul kıymetin ne zaman aşırı alındığını veya aşırı satıldığını anlamak için merkezi öneme sahiptir (Rockefeller, 2011).

Momentum göstergesi, mevcut fiyatın  $x$  gün önceki fiyata göre değişimini hesaplar ve sonuçları bir çizgi grafiği veya histogram olarak gösterir. Pozitif bir momentum göstergesi, fiyatların yükselmekte olduğunu gösterirken, negatif bir momentum göstergesi, fiyatların düşmekte olduğunu gösterir. Momentum hesabı denklem 2.5'te verilmiştir:

$$\text{Momentum} = \text{Kapanış}_t - \text{Kapanış}_{t-x} \quad (2.5)$$

Burada  $\text{Kapanış}_t$ , bugünün kapanış fiyatını ve  $\text{Kapanış}_{t-x}$ ,  $x$  gün önceki kapanış fiyatını ifade etmektedir.

### 2.2.1 Göreceli güç endeksi

RSI, bir varlığın aşırı alım veya aşırı satım koşullarını belirlemek için kullanılan bir momentum göstergesidir. Bu gösterge, belirli bir zaman dilimi boyunca fiyat değişimlerinin büyüklüğünü ve hızını ölçer. RSI, genellikle 14 günlük bir periyot kullanılır ve 0-100 aralığında bir değer üretir. 70 ve üstü değerlerin aşırı alım, 30 ve altındaki değerlerin ise aşırı satım koşullarını gösterdiği kabul edilir. Bu nedenle, RSI göstergesi, bir varlığın fiyatının düzelme veya düşüş eğiliminde olabileceği noktaları belirlemek için kullanılabilir (Chong, Ng ve Liew, 2014). RS ve RSI hesabı sırasıyla denklem 2.6 ve 2.7’de verilmiştir:

$$RS = \text{Ort}_y / \text{Ort}_d \quad (2.6)$$

$$RSI = 100 - (100 / (1 + RS)) \quad (2.7)$$

Burada  $\text{Ort}_y$ , yükselen günlerin kapanış fiyatının ortalamasını ve  $\text{Ort}_d$ , düşen günlerin kapanış fiyatının ortalamasını ifade etmektedir.

### 2.2.2 Stokastik osilatör

STO, bir varlığın kapanış fiyatını mevcut fiyat aralığına göre karşılaştırarak aşırı alım veya aşırı satım koşullarını belirleyen bir momentum göstergesidir. 0 ile 100 arasında bir aralıkta hareket eder. STO göstergesi, kapanış fiyatlarının nispeten düşük olduğu dönemlerde varlığın aşırı satım koşullarında olduğunu ve nispeten yüksek olduğu dönemlerde varlığın aşırı alım koşullarında olduğunu belirleyebilir. %K, %D’nin üzerine çıktığında, bu bir satın alma sinyali ve tam tersi bir satış sinyali olarak kabul edilir. (Rockefeller, 2011). STO hesaplanırken genellikle 14 günlük periyotlar tercih edilir fakat piyasa koşullarına, yatırım stratejisine ve yatırımcının tercihlerine göre değişebilir. Seçilen dönem boyunca en düşük değer ve en yüksek değer hesaplanır. STO formülü denklem 2.8’de verilmiştir:

$$\%K = (\text{Kapanış} - \text{Dip}) / (\text{Tepe} - \text{Dip}) \times 100 \quad (2.8)$$

Burada kapanış, mevcut kapanış fiyatını, dip, belirlenen periyottaki en düşük değeri, tepe, belirlenen periyottaki en yüksek değeri ve %D, %K’nın basit hareketli ortalamasını ifade etmektedir.

## 2.3 Volatilite Göstergeleri

Volatilite göstergeleri, finansal piyasalarda bir varlığın fiyatındaki dalgalanmaların şiddetini ölçmek için kullanılan göstergelerdir. Volatilite, belirli bir süre boyunca en düşük ve en yüksek fiyatlar arasındaki toplam hareket veya fiyatlardaki oynaklığın bir ölçüsüdür. Yüksek volatilite, genellikle fiyatların hızlı bir şekilde değiştiği anlamına gelir. Yani, volatilite ne kadar yüksek olursa risk de o kadar yüksek olur. Volatilitedeki bir değişiklik, ileride beklenen fiyat aralığında bir değişiklik anlamına gelir (Rockefeller, 2011). Piyasa oynaklığını ölçmenin bir yolu, x gün sayısı boyunca maksimum fiyat değişimini yakalamaktır. Bu dönem boyunca gerçekleşen en yüksek ve en düşük fiyat arasındaki fark alınarak hesaplanmaktadır.

### 2.3.1 Volatilite endeksi

Piyasanın ve kalabalığın stabil olmadığı durumlarda fiyatların anormal olacağını varsayan bir göstergedir. VIX yüksek olduğunda piyasanın düşebileceği düşünülür. VIX düşük olduğunda ise piyasa stabildir. VIX anormal derecede yüksek veya anormal derecede düşük olduğunda piyasanın riskli olduğu düşünülür (Rockefeller, 2011).

Volatilite endeksi genellikle, belirli bir periyot için fiyat değişimlerinin standart sapmasını veya varyansını hesaplayarak elde edilir. Varyans, veri değerlerinin aritmetik ortalamadan farkının karelerinin toplamının veri sayısının bir eksiğine bölümüdür. Standart sapma ise varyansın karekökü alınarak hesaplanır. Varyansın matematiksel formülü denklem 2.9'da verilmiştir:

$$\text{Var}(X) = (\Sigma(x - \bar{x})^2) / (n - 1) \quad (2.9)$$

Burada x, her bir veri noktasını,  $\bar{x}$ , veri kümesinin aritmetik ortalamasını ve n, veri sayısını ifade etmektedir.

### 2.3.2 Bollinger bantları

Bollinger bantları en popüler volatilite ölçülerinden biridir. Fiyatın her iki tarafına da hareketli ortalamanın iki standart sapmasından oluşan birer bant çizilerek oluşturulur. Fiyat hareketleri, genellikle bantların içinde kalır, ancak sıkı bir aralıkta kalırken bir trendin sonuna yaklaştığında bantların dışına çıkabilir. Aynı şekilde, volatilite arttığında bantlar genişleyebilir ve düşük volatilitede bantlar daralabilir. Bollinger Bantları, fiyatların aşırı alım veya satım seviyelerine ulaştığını göstermek için de



kullanılabilir. Fiyatlar üst banttan çıkarsa, aşırı alım sinyali olarak yorumlanabilir, alt banttan çıkarsa, aşırı satım sinyali olarak yorumlanabilir.

Bollinger bant genişlikleri, Bollinger Bantlarının bir başka ölçüsüdür. Bollinger Bantları'nın üst ve alt bantları arasındaki mesafeyi hesaplar. Yüksek ve düşük oynaklık dönemlerini belirlemek için kullanılabilir. Bant genişlikleri azaltığında fiyatındaki oynaklık azalır. Bu, varlık fiyatının trende girdiğini ve konsolide olduğunu gösterebilir. Genişlik değeri arttığında, fiyattaki oynaklık artar. Bu, fiyatın hareketli olduğunu ve bir trend değişikliği sinyali verdiğini gösterebilir (Bollinger, 2002).

## **2.4 Hacim Göstergeleri**

Hacim, bir varlığın alınıp satıldığı miktarı ifade eder ve bir varlıkla ilgili işlem hacmi arttığında, fiyatında da değişiklikler görülebilir. Hacim göstergeleri, yatırımcılara hacimle ilgili bilgileri sunmak için çeşitli yollarla kullanılır. Bazı hacim göstergeleri, varlık fiyatındaki değişiklikleri onaylamak için kullanılırken, diğerleri varlık fiyatının gelecekteki yönü hakkında ipuçları sağlayabilir.

### **2.4.1 Denge hacmi**

OBV kümülatif hacmi temsil eden tek bir sayıdır. OBV'yi hesaplamak için, kapanışın bir önceki günden daha yüksek olduğu günlerde hacim eklenir ve fiyatın bir önceki günden daha düşük olduğu günlerde hacim çıkarılır. Bu gösterge, alıcı ve satıcıların varlık üzerindeki etkisini ölçmek için kullanılır. OBV, bir varlıkla ilgili hacim bilgisini hesaplar ve bunu varlığın fiyatındaki değişimlerle ilişkilendirir (Gallegos-Erazo, 2022).

### 3. YÖNTEMLER

#### 3.1 Makine Öğrenmesi

Makine öğrenmesi, bilgisayarların belirli bir veri kümesindeki örüntüleri otomatik olarak tanımlamasına, öğrenmesine ve bunları kullanarak gelecekteki verileri tahmin etmesine izin veren bir yapay zeka dalıdır. Makine öğrenmesi, büyük miktarda veri toplayabilen ve analiz edebilen teknolojilerin ortaya çıkmasıyla birlikte son yıllarda oldukça popüler hale gelmiştir. Bu teknolojiler, belirli bir veri kümesindeki örüntüleri otomatik olarak tanımlamak ve gelecekteki verileri tahmin etmek için kullanılabilir. Bununla birlikte, makine öğrenmesi algoritmalarının kullanımı, güçlü bir bilgisayar teknolojisi altyapısı ve uygun veri kaynakları gerektirmektedir.

Makine öğrenmesi, teknik analiz için kullanılan göstergeleri otomatik olarak analiz etmek ve öngörüler yapmak için kullanılabilir. Bu, yatırımcıların daha iyi kararlar vermelerine yardımcı olabilir ve ticaret stratejilerini otomatikleştirebilir. Örneğin, bir makine öğrenmesi algoritması, belirli bir teknik göstergeye dayalı olarak gelecekteki fiyat hareketlerini tahmin edebilir veya belirli bir piyasa koşulunda hangi teknik göstergelerin en iyi sonuçları verdiğini belirleyebilir. Bu tür analizler, daha büyük ve daha karmaşık veri kümelerinin analiz edilmesine izin verir ve yatırımcıların daha hızlı ve daha doğru kararlar almalarına yardımcı olur. Bununla birlikte, makine öğrenmesi algoritmaları, doğru şekilde eğitilmedikleri takdirde hatalı sonuçlar verebilir veya yanlış kararlar verilmesine yol açabilir. Bu nedenle, makine öğrenmesinin kullanımı dikkatli bir şekilde yapılmalı ve uygun eğitim ve test verileri kullanılmalıdır (Goodfellow, Bengio ve Courville, 2018).

Makine öğrenmesinde sınıflandırma ve regresyon olarak iki temel problemden bahsedilebilir:

- i. Sınıflandırma: Bir veri kümesindeki örnekleri belirli bir sınıfa atama işlemidir. Örneğin, bir resim veri kümesindeki her resmin bir hayvan resmi mi yoksa bitki resmi mi olduğunu belirlemek isteyebiliriz. Sınıflandırma problemleri genellikle etiketli veri kümelerinde çalışır, yani verilerin hangi

sınıfa ait olduđu önceden belirlenmiştir. Sınıflandırma algoritmaları, veri kümesindeki özelliklerin bir kombinasyonunu kullanarak, veri örneklerini belirli sınıflara doğru şekilde atama yapmaya çalışır.

- ii. Regresyon: Bir veri kümesindeki örneklerin bir çıktı değerini tahmin etme işlemidir. Regresyon problemleri sürekli bir çıktı değişkeni üzerinde çalışır. Örneğin, bir evin fiyatını, o evin özellikleri (odaların sayısı, konumu, vb.) kullanarak tahmin edebiliriz. Regresyon algoritmaları, bağımlı değişkenin bağımsız değişkenlere olan bağımlılığını hesaplayarak, gelecekteki veriler üzerinde tahminler yapabilmemizi sağlar.

Makine öğrenmesinde kullanılan yöntemler genellikle üç ana kategoriye ayrılır: denetimli öğrenme, denetimsiz öğrenme ve takviyeli öğrenme (Goodfellow, Bengio ve Courville, 2018).

- i. Denetimli Öğrenme: Bu yöntem, önceden etiketlenmiş verilerin kullanıldığı bir öğrenme yöntemidir. Bu veriler, öğrenme algoritmasının doğru çıktıyı öğrenmesine ve gelecekteki tahminleri yapmasına yardımcı olur. Denetimli öğrenme yöntemleri, sınıflandırma ve regresyon problemlerini çözmek için kullanılmaktadır.
- ii. Denetimsiz Öğrenme: Bu yöntem, etiketlenmemiş verilerin kullanıldığı bir öğrenme yöntemidir. Bu veriler, algoritmanın veriler arasındaki ilişkileri tanıması ve verileri gruplandırması için kullanılır. Denetimsiz öğrenme yöntemleri, veri kümesindeki yapıları ve desenleri keşfetmek için kullanılmaktadır.
- iii. Takviyeli Öğrenme: Bu yöntem, bir öğrenme algoritmasının, bir görevi yürütmek için verilen bir ödül veya ceza almasıyla eğitildiği bir öğrenme yöntemidir. Bu yöntem, bir algoritmanın belirli bir görevi nasıl en iyi şekilde yerine getireceğini öğrenmesine ve en uygun eylemleri seçmesine yardımcı olur. Takviyeli öğrenme yöntemleri, oyun teorisi, robotik ve otomatik ticaret gibi uygulamalarda kullanılmaktadır.

Yukarıda verilen yöntemlerin her biri, farklı veri tipleri ve öğrenme problemleri için uygundur. Makine öğrenmesi uygulamaları, genellikle bu yöntemlerin bir veya daha fazlasını kullanarak en iyi sonuçları elde etmek için bir araya getirilir (Goodfellow, Bengio ve Courville, 2018).

Çalışmada denetimli öğrenme yönteminden yararlanılmış olup kullanılan makine öğrenmesi algortimaları aşağıda açıklanmıştır:

### 3.2.1 Aşırı gradyan arttırma

XGBoost, Gradyan Arttırma algoritmasına dayalı bir grup karar ağacı tabanlı makine öğrenmesi algoritmasıdır. XGBoost'ta ilk adım, ilk tahmini (baz puanı) oluşturmaktır. Bu tahmin herhangi bir sayı olabilir, çünkü doğru sonuç, bir sonraki adımda yapılacak işlemlerle yaklaşılarak elde edilir. Bu sayının varsayılan değeri 0.5'tir. Bu tahminin ne kadar iyi yapıldığı, modelin hatalı tahminleri ile bir ağaç oluşturularak incelenir (Chen ve Guestrin, 2016). XGBoost sınıflandırma problemi aşağıdaki gibi formüle edilmektedir.

Oluşturulan ağacın her bir dalı için benzerlik skoru (similarity score) hesaplanarak verilerin dallarda ne kadar iyi gruplandığı tespit edilir. Benzerlik skoru denklem 3.1'de gösterildiği gibidir:

$$\text{Benzerlik Skoru} = (\sum \text{hata})^2 / (\sum_i^N [P(1 - P)] + \lambda) \quad (3.1)$$

Burada N, hatalı tahmin sayısı, P, olasılık yüzdesi ve  $\lambda$ , regülarizasyon parametresidir.

Hangi ağacın daha iyi tahmin yaptığını anlamak için kazanç skoru (gain) hesaplanır. Benzerlik skoru ile dallar değerlendirilirken, kazanç skoru ile bütün ağaç değerlendirilir. Kazanç skoru denklem 3.2'de gösterildiği gibidir.

$$\text{Kazanç} = S_{\text{sol}} + S_{\text{sağ}} - S_{\text{düğüm}} \quad (3.2)$$

Burada  $S_{\text{sol}}$ , sol dal benzerlik skoru,  $S_{\text{sağ}}$ , sağ dal benzerlik skoru ve  $S_{\text{düğüm}}$ , önceki düğümün benzerlik skorudur.

En başarılı ağaca karar verildikten sonra budama işlemi başlar. Her bir dal için bir örtme (cover) değeri hesaplanır. Örtme değeri kazanç skorundan düşük ise dal budanır, eksi halde kırılmaya devam eder. Örtme değeri denklem 3.3'te gösterildiği gibidir:

$$\text{Örtme Değeri} = \sum_i^N [P(1 - P)] \quad (3.3)$$

Model hazır olduğunda çıktı değerleri,  $\text{Log(odds)} = \log(P/1-P)$  ve  $P = e^{\text{log(odds)}} / (1 + e^{\text{log(odds)})}$  olmak üzere denklem 3.4'te gösterildiği gibi hesaplanmaktadır:

$$\text{Çıktı Değeri} = \frac{\sum \text{hata}}{\sum_i^N [P(1-P)] + \lambda} \quad (3.4)$$

Burada N, hatalı tahmin sayısını, P, olasılık yüzdesini ve hata, yanlış tahmin edilen kayıtları ifade etmektedir.

### 3.2.2 Rastgele orman

RF, karmaşık sorunlara çözüm sağlamak için birçok zayıf sınıflandırıcıyı birleştiren topluluk öğrenmesini kullanan bir tekniktir. Her bir karar ağacı, tahmin edilen sınıf için oy kullanır veya gelen tahminlerin ortalaması alınır. Rastgele ormanların temel fikri, ağaç geliştirme sırasında sadece bir öznitelik alt kümesini içeren topluluk öğrenmesinden yararlanmasıdır. (Sekhar, Minal ve Madhu 2016).

Bir veri kümesinde bölünme yapılacak öznitelik seçilirken Gini endeksinden yararlanılır. Gini endeksi, bölünmenin mümkün olan en saf şekilde gerçekleşmesini sağlar. En düşük Gini değeri en düşük safsızlık anlamına gelir. Bölünmenin safsızlığını ölçmek için kullanılan bir diğer metrik entropidir. Matematiksel olarak Gini endeksi ve entropi denklem 3.5 ve 3.6'daki gibi ifade edilir:

$$\begin{aligned} \text{Gini} &= 1 - \sum_i^N (p_i)^2 \\ &= 1 - [(p_+)^2 + (p_-)^2] \end{aligned} \quad (3.5)$$

Burada  $p_+$ , pozitif sınıfların olasılık değeri ve  $p_-$ , negatif sınıfların olasılık değeridir.

$$E = -p_{(+)} \log p_{(+)} - p_{(-)} \log p_{(-)} \quad (3.6)$$

### 3.2.3 Rastgele ağaçlar

RT, makine öğrenmesi ve karar ağaçları alanında bir algoritmadır. Etiketli bir eğitim verisi kümesinden bir karar ağacı oluşturma yöntemidir. Algoritma, bir rastgele özellik temel alınarak veriyi daha küçük alt kümelerine iteratif olarak böler ve bir durdurma kriterine (örneğin, alt kümedeki minimum veri noktaları sayısı veya maksimum ağaç derinliği) ulaşıncaya kadar devam eder. Ağacın her yaprak düğümü, temsil ettiği veri alt kümesindeki sınıf dağılımına dayalı bir tahmin oluşturur. Rastgele ağaçlar, sınıflandırma ve regresyon dahil çeşitli görevler için etkili olduğu kanıtlanmıştır (Boothun ve Jayabalan, 2018).

Ağaç başarısının ölçümünde kural başarıları, ağaç başarıları ve ilginçlik (interestingness) ölçüsü kullanılır. İlginçlik ölçüsü denklem 3.7'deki gibi ifade edilmektedir:

$$I_{\text{index}(t)} = P(A_t) * P(B_t) * [P(B_t | A_t) + P(\bar{B}_t | \bar{A}_t)] \quad (3.7)$$

Burada,  $P(A_t)$ , ağaç doğruluğunu,  $P(B_t)$ , kural doğruluğunu,  $P(B_t | A_t)$ , hem ağaçlar hem de düğüm tarafından yapılan doğru tahminleri ve  $P(\bar{B}_t | \bar{A}_t)$ , hem ağaçlar hem de düğüm tarafından yapılan yanlış tahminleri ifade etmektedir.

### 3.2.4 Yapay sinir ağları

Bu model, beyindeki sinir hücrelerinin çalışma biçiminden esinlenerek geliştirilmiştir. Yapay sinir ağları, birçok katmandan oluşur ve genellikle çok katmanlı algılayıcılar olarak adlandırılır. İleri beslemeli sinir ağları, yapay sinir ağının en basit türüdür. Bilgi, girdi katmanından, gizli katmanlar aracılığıyla, sadece ileri yönde çıktı katmanına doğru gider. Nöronlar birbirleriyle etkileşime girer ve bilgi paylaşılır. Bu ağ türünde döngü yoktur (Goodfellow, Bengio ve Courville, 2018). N değişkenli bir veri kümesi için sinir ağı formülü 3.8'deki gibi oluşturulmaktadır:

$$\begin{aligned} f(x) &= b + w_1x_1 + w_2x_2 + \dots + w_nx_n \\ &= b + W^T X \end{aligned} \quad (3.8)$$

Burada  $w$ , ağırlıkları,  $W$ , ağırlık matrisini,  $x$ , bağımsız değişkeni ve  $X$ , bağımsız değişken matrisini ifade etmektedir.

### 3.2.5 Destek vektör makineleri

SVM, özellikle sınıflandırma amaçlı kullanılan bir makine öğrenmesi algoritmasıdır. Temel prensibi, sınıflar arasındaki sınırları en iyi şekilde belirleyen bir hiperdüzlem oluşturarak verileri bu düzleme yansıtmaktır. SVM, sınıfların doğru şekilde ayrılabilmesi için en uygun hiperdüzlemi oluşturmak için veri noktalarının bir kısmını kullanır. Bu veri noktaları destek vektörleri olarak adlandırılır ve hiperdüzlemi belirlemek için kullanılan en önemli verilerdir. Diğer veriler, bu destek vektörlerine göre sınıflandırılır. (Huang, Cai, Pacheco, Narrandes, Wang ve Xu, 2018). SVM sınıflandırıcı denklemini 3.9'deki gibidir:

$$\text{Her } i = 1, 2, \dots, n \text{ için } y_i(w^T x_i + b) \geq 1 - \epsilon_i, \quad \epsilon_i \geq 0$$

$$\min_{w,b,\epsilon_2} \frac{1}{2} \|w\|^2 + C \sum_i^n \epsilon_i \quad (3.9)$$

Burada C, ceza parametresini, w, normalleştirme vektörünü ve b, düzeltme parametresini ifade etmektedir.

### 3.2.6 K en yakın komşu

KNN, giriş örneğinin k en yakın komşusunun çoğunluk sınıfına dayalı olarak parametrik olmayan sınıflandırmasına izin verir. Algoritma, benzerlik ölçüsü kullanarak tüm mevcut durumları depolar ve yeni durumları sınıflandırır. Bu durum, uzaklık ölçütü ile ölçülen en yakın k komşusunun en yaygın sınıfına atanır. Bu uzaklık ölçütü, Öklid uzaklığı, Manhattan uzaklığı, Minkowski uzaklığı vb. olabilir. KNN modeli, sınıflandırma ve regresyon problemleri için kullanılabilir. Bir sınıflandırma probleminde KNN algoritması, yeni örneğin sınıf etiketini kullanır, regresyonda ise k en yakın komşuların ortalama veya medyanını kullanır (Saru ve Subashree, 2019). KNN algoritmasının çalışma prensibi ve gidiş yolu aşağıda verilmiştir:

- i. Belirli bir k komşu sayısı seçilir
- ii. Test örneği için en yakın k örneği bulunur
- iii. Bulunan k örneklerinin sınıf etiketleri kontrol edilir
- iv. Test örneğinin sınıf etiketi için, k örnekleri arasındaki en yaygın olan etiket seçilir

$x_1 = (x_{11}, x_{12} \dots x_{1n})$  ve  $x_2 = (x_{21}, x_{22} \dots x_{2n})$  olmak üzere, Öklid, Manhattan ve Minkowski uzaklıklarının matematiksel hesaplamaları sırasıyla, denklem 3.10, 3.11 ve 3.12'deki gibidir:

$$U_{oklid} = \sqrt{\sum_i^n (x_{1i} - x_{2i})^2} \quad (3.10)$$

$$U_{manhattan} = \sum_i^n |x_{1i} - x_{2i}| \quad (3.11)$$

$$U_{minkowski} = \sqrt[r]{\sum_i^n (x_{1i} - x_{2i})^r} \quad (3.12)$$

### 3.2.7 Bayes ağları

Bu algoritma, Bayes teoremini temel alır ve olasılık dağılımlarını kullanarak sınıflandırma yapar. Bayes ağları birbirleriyle etkileşimde bulunan bağımsız değişkenler arasındaki bağımlılıkları ve belirli bir sonucun meydana gelme olasılığını

hesaplamak için kullanılır. Böylece koşullu olasılık dağılımını kullanarak özellikler arasındaki ilişkileri modeller. Bayes ağlarının öğrenme süreci, ağın yapısını ve parametrelerini öğrenmekten oluşur. Yapı, ağdaki değişkenler arasındaki bağımlılıkları ifade eder. Parametreler ise koşullu olasılık dağılımındaki değerleri ifade eder (Korb ve Nicholson, 2010). Bayes ağlarının temel formülü denklem 3.13'te verilmiştir:

$$P(A|B) = \frac{P(B)P(B|A)}{P(A)} \quad (3.13)$$

Burada  $P(A)$ , A'nın olasılığını,  $P(B)$ , B'nin olasılığını,  $P(A|B)$ , A bilindiğinde B'nin koşullu olasılığını ve  $P(B|A)$ , B bilindiğinde A'nın koşullu olasılığını ifade etmektedir.

### 3.2 Özellik Seçimi

Özellik seçimi, makine öğrenmesi ve veri madenciliği gibi disiplinlerde, veri kümelerindeki öznitelikleri (değişkenleri) seçmek veya çıkarmak için kullanılan bir yöntemdir. Bu yöntem, modelin performansını arttırmak ve gereksiz veya zararlı özniteliklerin etkisini azaltmak için kullanılır.

Veri kümeleri genellikle çok sayıda öznitelik içerir ve bu öznitelikler arasında sık sık yüksek bir korelasyon vardır. Bu durumda, tüm öznitelikleri kullanmak yerine, özellikle de veri kümesindeki, gürültülü veya öngörücü olmayan öznitelikleri çıkarmak, modelin performansını artırabilir ve aşırı öğrenmeyi önleyebilir.

Özellik seçimi, filtre yöntemleri ve sarmal yöntemler olarak temelde iki ana kategoriye ayrılır. Filtreleme yöntemleri, özniteliklerin kendi aralarındaki ilişkilerini analiz eder ve korelasyon veya olasılık gibi belirli bir özellik kriterine dayanarak öznitelikleri seçer veya çıkarır. Sarmal yöntemler ise, öznitelikleri modele dahil etmenin etkisini doğrudan ölçer ve en uygun öznitelik kümesini belirler. En sık kullanılan yöntemlerden Pearson korelasyonu ve olabilirlik (likelihood) istatistiklerinin matematiksel ifadeleri sırasıyla denklem 3.14 ve 3.15'te verilmiştir.

Pearson korelasyon katsayısı ( $r$ ), -1 ile +1 arasında değişen bir değer alır. 1, tam pozitif korelasyonu gösterirken, -1 tam negatif korelasyonu gösterir ve 0, iki değişken arasında herhangi bir ilişki olmadığını gösterir (Benesty, Chen ve Huang, 2008).



$$r = E[xy] / \sigma_x \sigma_y \quad (3.14)$$

Burada  $x$ , bağımsız değişkeni,  $y$ , bağımlı değişkeni,  $E[xy]$ , beklenen değeri,  $\sigma_x$ ,  $x$ 'in varyansını ve  $\sigma_y$ ,  $y$ 'nin varyansını ifade etmektedir.

Olabilirlik, veri kümesindeki gözlemlerin belirli bir parametre kümesi tarafından nasıl açıklandığını ifade eden bir olasılık fonksiyonudur. Bu fonksiyon, parametrelerin belirli bir değer kümesi altında veri kümesinin olasılığını hesaplamaktadır. Olabilirlik oranı, iki hipotez arasındaki orandır ve denklem 3.16'daki gibi (Fan, Zhang ve Zhang, 2001) hesaplanmaktadır:

$$L(\theta|x) = f(x|\theta) \quad (3.15)$$

Burada  $x$ , veri kümesindeki gözlemleri,  $L(\theta|x)$ , gözlemlerin belirli bir parametre seti " $\theta$ " altında gözlemlenme olasılığını ve  $f(x|\theta)$ ,  $\theta$  altında gözlemlerin olasılığını ifade etmektedir.

$$LR = P(x | H_0) / P(x | H_1) \quad (3.16)$$

Burada  $P$ , Likelihood olmak üzere hipotez aşağıdaki gibi ifade edilir:

$H_0$ : Bağımlı değişken üzerinde etkisi olmadığı söylenebilir.

$H_1$ : Bağımlı değişken üzerinde etkisi olduğu söylenebilir.

### 3.3 Temel Bileşenler Analizi

PCA, veri kümesinin boyutunu küçültmek ve değişkenler arasındaki ilişkileri anlamak için kullanılan çok değişkenli istatistiksel analiz yöntemlerinden biridir. PCA, veri setindeki bağımsız değişkenlerin birbirleriyle yüksek oranda ilişkili olduğu durumlarda sıklıkla kullanılır. Bu yöntem, değişkenler arasındaki korelasyonu azaltır ve daha az sayıda bileşenin değişkenleri yüksek varyans ile açıklamasına olanak tanır. Böylece çoklu bağlantı sorunu olan veri kümelerinde değişkenleri matematiksel olarak birbirinden bağımsız hale getirir.

PCA sırasında elde edilen bileşenlerin yorumlanması ve daha iyi anlaşılması için genellikle rotasyon adı verilen bir işlem yapılır. Rotasyon, bileşenlerin yönünü ve değişkenlerle olan ilişkilerini değiştirir, böylece bileşenler daha kolay yorumlanabilir hale gelir. En yaygın kullanılan iki rotasyon türü Varimax ve Promax rotasyonudur. Varimax rotasyonu, bileşenlerin değişkenlerle arasındaki korelasyonu azaltırken, bileşenler arasındaki korelasyonu artırmayı amaçlar. Promax rotasyonu ise,

bileşenler arasındaki korelasyonu korurken, bileşenlerin değişkenlerle olan ilişkisini değiştirir (Abdi ve Williams, 2010).

PCA hesaplaması kovaryans matrisinin hesaplanması ve özdeğerlerin hesaplanması olmak üzere iki adımdan oluşur. Kovaryans matrisi ve özdeğer hesabının matematiksel ifadeleri sırasıyla denklem 3.17 ve 3.18’de verildiği gibidir:

$$\text{kov}(X, Y) = \frac{1}{n} \sum_i^n (x - \bar{x})(y - \bar{y}) \quad (3.17)$$

Burada  $\text{kov}(X, Y)$ ,  $X$  ve  $Y$  değişkenleri arasındaki kovaryansı,  $n$ , gözlem sayısını,  $\bar{x}$ ,  $x$  değişkenin ortalamasını ve  $\bar{y}$ ,  $y$  değişkenin ortalamasını ifade etmektedir.

$$\det(A - \lambda I) = 0 \quad (3.18)$$

Burada  $A$ , kovaryans matrisini,  $I$ , birim matrisi ve  $\lambda$ , özdeğeri ifade etmektedir.

### 3.4 Yeniden Örnekleme

Makine öğrenmesinde yeniden örnekleme, sınıflandırma problemlerinde sınıflar arasındaki dengesizlik nedeniyle oluşan performans sorunlarını çözmeye yardımcı olan bir yaklaşımdır. Eğitim veri kümesinde sınıflar arasındaki dengesizlik, modelin bir sınıftan diğerine geçiş yapmasını zorlaştırır ve sonuç olarak daha az sayıda örneğe sahip olan sınıfın doğruluğunu düşürür ve baskın olan sınıfın aşırı öğrenilmesine neden olur (Kotsiantis, Kanellopoulos ve Pintelas, 2006). Bu nedenle çeşitli yeniden örnekleme yöntemleri önerilmiştir.

#### 3.4.1 Alt örnekleme

Eldeki veri kümesindeki baskın olan sınıftan, rastgele bir şekilde belirli sayıda örneği çıkararak veri kümesinin boyutunu azaltan bir veri dengeleme yöntemidir. Kümeleme ve örnekler arasındaki uzaklıklar kullanılarak da alt örnekleme yapılabilmektedir. Bu yöntemde veri kaybı söz konusudur fakat sentetik veri üretiminin önüne geçildiğinden sıklıkla kullanılmaktadır.

#### 3.4.2 Aşırı örnekleme

Veri kümesinde daha az sayıda temsil edilen sınıfın örneklerinin rastgele kopyalanması, varyasyonlarının yapılması veya sentetik örneklerin oluşturulması gibi birkaç farklı yöntemle gerçekleştirilebilir. Bu yöntemlerden her biri, özgün veri

kümesinde az temsil edilen sınıfların örneklerinin sayısını artırır ve böylece makine öğrenmesi modelinin az temsil edilen sınıfları daha iyi öğrenmesini sağlamaktadır.

### 3.5 Sınıflandırma Performans Ölçüleri

Sınıflandırma problemlerinde hata matrisinden yola çıkarak bir çok performans ölçüsü hesaplanabilmektedir. Hata matrisi, sınıflandırma modellerinin performansını ölçmek amacıyla sıklıkla kullanılmaktadır. Matris, tahmin edilen ve gerçek değerlerin 4 farklı kombinasyonunu içermektedir:

Gerçek Pozitifler (TP) : Gerçek değeri 1 ve tahmin edilen değer de 1 olduğu örneklerdir.

Gerçek Negatifler (TN) : Gerçek değeri 0 ve tahmin edilen değer de 0 olduğu örneklerdir.

Yanlış Pozitifler (FP) : Gerçek değeri 0 ancak tahmin edilen değer 1 olduğu örneklerdir.

Yanlış Negatifler (FN) : Gerçek değeri 1 ancak tahmin edilen değer 0 olduğu örneklerdir.

Hata matrisi ile Doğruluk (Accuracy), Duyarlılık (Recall), Kesinlik (Precision), Seçicilik (Specificity), F Puanı (F Score) ve ACC (Alıcı Çalışma Karakteristiği Eğrisi) gibi ölçüler de hesaplanabilmektedir (Grandini, Bağlı ve Visani, 2020; Luque, Carrasco, Martin ve de las Heras, 2019).

#### 3.5.1 Doğruluk

Doğru tahminlerin (TP+TN) tüm tahminlere oranıdır. Esas köşegenin toplama oranı olarak da ifade edilebilir. Doğruluk oranının hesaplanması denklem 3.19'de verilmiştir:

$$\text{Doğruluk} = (TP+TN) / (TP+TN+FN+FP) \quad (3.19)$$

#### 3.5.2 Kesinlik

Pozitif olarak doğru tahmin edilenlerin (TP) tüm pozitif tahminlere (FP+TP) oranıdır. Kesinlik oranının hesaplanması eşitlik 3.20'de verilmiştir:

$$\text{Kesinlik} = TP / (FP+TP) \quad (3.20)$$

### 3.5.3 Duyarlılık

Pozitif olarak tahmin edilenlerin (TP) tüm pozitif gözlenenlere (TP+FN) oranıdır. Hassasiyet olarak da isimlendirilmektedir. Modelin pozitifleri doğru tahmin edebilme konusundaki etkinliđi olarak da ifade edilebilir. Duyarlılık oranının hesaplanması denklem 3.21’de verilmiştir:

$$\text{Duyarlılık} = \text{TP} / \text{TP} + \text{FN} \quad (3.21)$$

### 3.5.4 F1 puanı

Kesinlik ve duyarlılıđı dengelemek için F<sub>1</sub> puanı performans ölçüsünden yararlanılır. F<sub>1</sub> puanı kesinlik ve duyarlılıđın harmonik ortalamasıdır. Matematiksel olarak denklem 3.22’de ifade edilmiştir:

$$F_1 = 2 * (\text{kesinlik} * \text{duyarlılık}) / (\text{kesinlik} + \text{duyarlılık}) \quad (3.22)$$

### 3.5.5 Seçicilik

Negatif olarak doğru tahmin edilenlerin (TN) tüm negatif tahminlere (TN+FP) oranıdır. Modelin negatifleri doğru tahmin edebilme konusundaki etkinliđi olarak da ifade edilebilir. Seçiciliđin matematiksel ifadesi denklem 3.23’te verilmiştir:

$$\text{Seçicilik} = \text{TN} / \text{TN} + \text{FP} \quad (3.23)$$

### 3.5.6 Alıcı çalışma karakteristiđi eğrisi

ACC, sınıflandırma modelinin performansını değerlendirmek için kullanılan bir gösterimdir. Bu eğri, doğru pozitifler (TP) ve yanlış pozitifler (FP) arasındaki ilişkiyi gösterir. Eğrinin altında kalan alan, sınıflandırma modelinin performansını ölçmek için kullanılan bir ölçüttür ve 1’e yaklaştıkça modelin performansı artar. ACC eğrisi, sınıflandırma modelinin performansını değerlendirmenin yanı sıra, sınıflandırma probleminde kullanılacak en uygun olasılık eşik değerini belirlemek için de kullanılabilir. Eğri, ticari amaçlara uygun bir olasılık eşiđi seçmek için kullanılabilir.

## 4. UYGULAMA

### 4.1 Veri Kümesi

Çalışmada 'binance.com' üzerinden alınan günlük veriler ile çalışılmıştır. 1 Eylül 2017 ile 1 Nisan 2023 tarihleri arasında olmak üzere yaklaşık 2000 gözlem ile çalışılan güncel veri seti, BTC'nin açılış, kapanış, en yüksek fiyat, en düşük fiyat ve hacim bilgilerini içermektedir. Bu bilgiler kullanılarak üretilen farklı alanlarla birlikte toplam 90 özellik ile çalışılmıştır. Veri kümesi ve veri sözlüğü ekler bölümünde Çizelge A.1'de sunulmuştur. Çalışmada, IBM SPSS Modeller 18.2 sürümü kullanılmıştır.

Günün en yüksek ve en düşük fiyatları bir önceki güne göre daha yüksek ise artan fiyat çubuğu, düşük ise düşen fiyat çubuğu olarak adlandırılmıştır. En az üç gün artan çubuk gerçekleşmiş ise yükselen trend olarak kabul edilmiştir. Aynı şekilde en az üç gün düşen fiyat çubuğu gözlemlenmiş ise düşen trend olarak kabul edilmiştir. Bir gözlem yükselen trend (1) veya düşen trend (-1) değilse sabit trend (0) olarak değerlendirilmiştir. Bu bağlamda, al, sat ve bekle kararları için oluşturulan bağımlı değişkenin kuralları Çizelge 4.1'de verilmiştir:

**Çizelge 4.1:** Bağımlı değişken kuralları.

Bir Önceki Günün Trendi	Bugünün Trendi	Bağımlı Değişken
0	1	Al
-1	1	Al
0	-1	Sat
1	-1	Sat
1	0	Bekle
-1	0	Bekle
0	0	Bekle
1	1	Bekle
-1	-1	Bekle

Çizelge 4.1'de görüldüğü gibi trend yükselişe geçtiğinde alışı kararı, düşüşe geçtiğinde satışı kararı, sabit kaldığında ise herhangi bir aksiyon alınmayacak şekilde veri etiketleme çalışması yapılmıştır.

## 4.2 Özellik Seçimi

Olabilirlik oran istatistiği kullanılarak uygulanan özellik seçimi sonucunda 40 bağımsız değişkenin kullanılmasına karar verilmiştir. Kullanılacak bağımsız değişkenlerin listesi Şekil 4.1’de verilmiştir:

	Rank	Field	Measurement	Importance	Value
✓	1	UP_BAR	Flag	Important	1.0
✓	2	K_PERCENT	Continuous	Important	1.0
✓	3	UP_BAR_COUNT	Continuous	Important	1.0
✓	4	D_PERCENT	Continuous	Important	1.0
✓	5	DOWN_BAR	Flag	Important	1.0
✓	6	DOWN_BAR_COUNT	Continuous	Important	1.0
✓	7	CO_BUY	Flag	Important	1.0
✓	8	LOWER_LOW	Flag	Important	1.0
✓	9	LOWER_HIGH	Flag	Important	1.0
✓	10	HIGHER_HIGH	Flag	Important	1.0
✓	11	CO_SELL	Flag	Important	1.0
✓	12	HIGHER_LOW	Flag	Important	1.0
✓	13	DOWN_TREND_FLAG	Flag	Important	1.0
✓	14	RSI_5	Continuous	Important	1.0
✓	15	DOWN_SMA5	Flag	Important	1.0
✓	16	UP_SMA5	Flag	Important	1.0
✓	17	UP_TREND_FLAG	Flag	Important	1.0
✓	18	WHITE_BODY	Flag	Important	1.0
✓	19	DOWN_DAY	Flag	Important	1.0
✓	20	BLACK_BODY	Flag	Important	1.0
✓	21	UP_DAY	Flag	Important	1.0
✓	22	RSI_BUY	Flag	Important	1.0
✓	23	RSI_14	Continuous	Important	1.0
✓	24	MOMENTUM	Continuous	Important	1.0
✓	25	DOWN_VOLUME_DEGREE	Continuous	Important	1.0
✓	26	RSI_SELL	Flag	Important	1.0
✓	27	CON_DIV	Continuous	Important	0.989
✓	28	UP_VOLUME_DEGREE	Continuous	Important	0.948
✓	29	TICK_MARKS	Continuous	Important	0.935
✓	30	SHADOW_TOP	Continuous	Important	0.896
✓	31	MACD	Continuous	Important	0.856
✓	32	SDEV_RANGE	Continuous	Marginal	0.509
✓	33	OBV	Continuous	Marginal	0.445
✓	34	BOLLINGER_RANGE	Continuous	Marginal	0.377
✓	35	DAYLIGHT	Continuous	Marginal	0.29
✓	36	BAR_RANGE	Continuous	Marginal	0.258
✓	37	BOLLINGER_BAND_TOP	Continuous	Marginal	0.256
✓	38	SMA_5	Continuous	Marginal	0.252
✓	39	BOLLINGER_BAND_BOTTOM	Continuous	Marginal	0.248
✓	40	MAX_MOVE	Continuous	Marginal	0.246

Şekil 4.1: Bağımsız değişken listesi

Şekil 4.1’de field sütunu bağımsız değişkenleri, measurement sütunu ölçüm seviyelerini, importance sütunu önem durumunu, value sütunu ise bağımsız değişkenlerin bağımlı değişken ile olan ilişki gücünü göstermektedir. Burada, ölçüm seviyesi flag olan değişkenler ikili (binary) değişkenleri, continuous olanlar ise sürekli değişkenleri belirtmektedir. Olabilirlik oran istatistiği parametreleri, ilişki gücü 70 ve üzeri ise yüksek ilişkili (important), 20 ve 70 arasında ise orta ilişkili

(marginal) olacak şekilde tanımlanmıştır. İlişki gücü düşük olan değişkenler çalışmada kullanılmayacağı için Şekil 4.1’ de verilmemiştir.

### 4.3 Temel Bileşenler Analizi

Bağımsız değişken sayısının fazla olmasından dolayı çoklu bağlantı problemini önlemek amacıyla temel bileşenler analizi kullanılarak veri kümesinin boyutu küçültülmüştür. Bileşenlerin değişkenleri tam olarak ayıramaması sebebi ile RSI\_14 değişkeni veri kümesinden çıkarılmış ve varimax dönüşümü yapılmıştır. Analiz sonucu oluşturulan ilk 10 bileşen bağımsız değişkenlerin toplam varyansını %81 oranında açıklamaktadır. Toplam açıklanan varyans tablosu Şekil 4.2’de verilmiştir. Bileşen matrisi ekler bölümünde Şekil B.1’de sunulmuştur.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.148	23.456	23.456	9.148	23.456	23.456	7.900	20.255	20.255
2	7.928	20.328	43.784	7.928	20.328	43.784	4.324	11.088	31.343
3	3.226	8.273	52.057	3.226	8.273	52.057	3.771	9.669	41.013
4	2.543	6.520	58.577	2.543	6.520	58.577	3.681	9.437	50.450
5	2.006	5.142	63.719	2.006	5.142	63.719	2.896	7.427	57.877
6	1.666	4.272	67.991	1.666	4.272	67.991	2.642	6.775	64.651
7	1.570	4.026	72.018	1.570	4.026	72.018	2.050	5.257	69.908
8	1.297	3.326	75.344	1.297	3.326	75.344	1.518	3.892	73.800
9	1.234	3.163	78.507	1.234	3.163	78.507	1.462	3.750	77.549
10	1.042	2.671	81.178	1.042	2.671	81.178	1.415	3.629	81.178
11	.967	2.479	83.657						
12	.820	2.103	85.760						
13	.786	2.015	87.775						
14	.702	1.799	89.575						
15	.646	1.657	91.232						

Şekil 4.2: Temel bileşenler analizi varyans tablosu.

Temel bileşenler analizi varyans tablosu çıktılarını 39. bileşene kadar ilerlemektedir. Ancak, Şekil 4.2’de görüleceği gibi açıklanan varyansın kümülatif toplamı 10. bileşende %81.178’e ulaşmıştır. Bu nedenle, çalışmaya 10 bileşen ile devam edilmiştir.

### 4.4 Veri Dengeleme

Bağımlı değişkenin sınıfları arasında yüksek oranda dengesizlik mevcuttur. Bu problem, makine öğrenmesi modelinin tahmin sonuçlarında tutarsızlık yaratmaması amacıyla rastgele alt örnekleme yöntemi ile giderilmiştir. Bağımlı değişkenin sınıf dağılımı Çizelge 4.2’de verilmiştir:

**Çizelge 4.2:** Bağımlı değişkenin sınıf dağılımı.

Sınıf	Gözlem Sayısı	Oran
Al	192	%9.47
Sat	187	%9.23
Bekle	1648	%81.3

Çizelge 4.2’de görüldüğü gibi ‘al’ ve ‘sat’ sınıflarının veri kümesindeki oranları yaklaşık %9 iken, ‘bekle’ sınıfının oranı %81.3’tür. Sınıflar arası dengesizliğin çok yüksek olması sebebi ile makine öğrenmesi algoritmalarının bu veri kümesi ile eğitilmesi yanlı ve tutarsız tahminler üretmesine neden olacaktır. Bu problemin çözümü için iki farklı alt örnekleme çalışması yapılarak sonuçlar karşılaştırılmıştır. Birinci ve ikinci veri kümelerinin sınıf dağılımları sırasıyla Çizelge 4.3 ve Çizelge 4.4’te verilmiştir:

**Çizelge 4.3:** Tüm sınıfların eşit olduğu dağılım.

Sınıf	Gözlem Sayısı	Oran
Al	192	%33.51
Sat	187	%32.64
Bekle	194	%32.86

**Çizelge 4.4:** Bekle sınıfının yüksek olduğu dağılım.

Sınıf	Gözlem Sayısı	Oran
Al	192	%26.82
Sat	187	%26.12
Bekle	337	%47.07

Birinci veri kümesi (Çizelge 4.3), tüm sınıfların kayıt sayılarının birbirine çok yakın olduğu alt örnekleme çalışması sonucunda elde edilmiştir. Birinci veri kümesinde eğitilen modeller ‘bekle’ sınıfını yeteri kadar öğrenemediği için farklı oranlarda tekrar örnekleme çalışması yapılmıştır. İkinci örnekleme çalışmasında veri kümesi, ‘bekle’ sınıfının orijinal veri kümesinde olduğu gibi diğer sınıflara göre daha yüksek oranda gözleme sahip olacak şekilde hazırlanmıştır. İkinci veri kümesinde (Çizelge 4.4), ‘al’ ve ‘sat’ sınıflarının veri kümesindeki oranları yaklaşık %26 iken, ‘bekle’ sınıfının oranı %47.07’dir. Sonuç olarak iki farklı veri kümesi elde edilmiştir.



#### 4.5 Model Eğitimi ve Sonuçları

Veri seti, %80 eğitim, %20 test kümesi olmak üzere rastgele ayrılmıştır. RF, XGB ve RT modelleri, iki veri kümesi için aynı hiperparametreler ile eğitilmiştir. Model sonuçları 5 katlı çapraz doğrulama yapılarak test edilmiştir. Kullanılan hiperparametreler, RF, XGB ve RT modelleri için sırasıyla, Çizelge 4.5, Çizelge 4.6 ve Çizelge 4.7’de verilmiştir. Hiperparametrelerin detaylı açıklamaları ekler bölümünde Çizelge C.1’de sunulmuştur.

**Çizelge 4.5:** RF modelinin hiperparametreleri.

Hiperparametre	
bootstrap	TRUE
ccp_alpha	0
criterion	gini
max_depth	10
max_features	auto
min_impurity_decrease	0
min_samples_leaf	1
min_samples_split	2
min_weight_fraction_leaf	0
n_estimators	10
oob_score	FALSE
verbose	0
warm_start	FALSE
number_of_trees_to_build	10
learning_rate	0.01
max_iteration	1000
max_evaluation	300

**Çizelge 4.6:** XGB modelinin hiperparametreleri.

Hiperparametre	
tree_method	auto
number_boost_round	10
max_depth	6
min_child_weight	1
max_delta_step	0
objective	multi
stopping_rounds	10
evaluation_data_ratio	0.3
sub_sample	1
eta	0.3
gamma	0
colsample_by_tree	1
colsample_by_level	1
lambda	1
alpha	0
scale_pos_weight	1

**Çizelge 4.7:** RT modelinin hiperparametreleri.

Hiperparametre	
number_of_models_to_build	100
sample_size	1
max_number_of_nodes	10000
max_tree_depth	10
min_child_node_size	5
max_percentage_of_missing_values	70
exclude_fields_with_a_single_category_majority_over(%)	95
max_number_of_field_categories	49
min_field_variation	0.05
number_of_bins	10
number_of_interesting_rules_to_report	50

Yukarıdaki hiperparametreler kullanılarak birinci ve ikinci veri kümeleri ile eğitilen modellerin sonuçları sırasıyla, Şekil 4.3 ve Şekil 4.4'te verilmiştir:

XGB												
TEST	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk	EĞİTİM	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk	
BEKLE	0.74	0.43	0.93	0.55	0.77	BEKLE	0.81	0.52	0.94	0.63	0.80	
SAT	0.77	0.87	0.89	0.82	0.88	SAT	0.74	0.90	0.86	0.82	0.87	
AL	0.73	0.92	0.85	0.81	0.87	AL	0.75	0.91	0.84	0.82	0.87	
RF												
TEST	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk	EĞİTİM	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk	
BEKLE	0.83	0.51	0.94	0.63	0.77	BEKLE	0.78	0.49	0.93	0.60	0.78	
SAT	0.76	0.91	0.87	0.83	0.88	SAT	0.78	0.89	0.87	0.83	0.88	
AL	0.79	0.94	0.88	0.85	0.90	AL	0.77	0.92	0.85	0.83	0.87	
RT												
TEST	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk	EĞİTİM	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk	
BEKLE	0.77	0.40	0.94	0.52	0.76	BEKLE	0.70	0.39	0.91	0.50	0.73	
SAT	0.72	0.91	0.80	0.80	0.84	SAT	0.70	0.88	0.81	0.78	0.84	
AL	0.75	0.82	0.86	0.78	0.85	AL	0.70	0.85	0.83	0.77	0.83	

**Şekil 4.3:** Birinci veri kümesinin eğitim ve test sonuçları.

Şekil 4.3'te XGB modeli incelendiğinde, 'al' ve 'sat' sınıflarının tüm başarı metrikleri eğitim ve test veri kümelerinde %70'in üzerindedir. Ancak, 'bekle' sınıfının duyarlılık oranının test kümesinde %43 olması, modelin bu sınıfı diğer sınıflar kadar iyi öğrenemediğini göstermektedir. F1 puanının test kümesinde %55 olması, 'bekle' sınıfının dengeli tahminlenemediğini göstermektedir. Eğitim kümesi incelendiğinde de benzer sonuçlara ulaşılmaktadır. Aynı şekilde RF ve RT modellerinin eğitim ve test sonuçları incelendiğinde 'bekle' sınıfının yeteri kadar iyi tahmin edilemediği görülmektedir.

XGB												
TEST	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk		EĞİTİM	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk
BEKLE	0.79	0.78	0.85	0.78	0.82		BEKLE	0.82	0.78	0.85	0.80	0.82
SAT	0.74	0.83	0.89	0.78	0.87		SAT	0.76	0.86	0.90	0.80	0.89
AL	0.79	0.78	0.91	0.78	0.87		AL	0.82	0.80	0.93	0.81	0.90
RF												
TEST	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk		EĞİTİM	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk
BEKLE	0.81	0.66	0.87	0.72	0.77		BEKLE	0.88	0.69	0.92	0.77	0.81
SAT	0.71	0.91	0.87	0.80	0.88		SAT	0.77	0.92	0.91	0.84	0.91
AL	0.75	0.83	0.90	0.79	0.88		AL	0.75	0.88	0.88	0.81	0.88
RT												
TEST	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk		EĞİTİM	Kesinlik	Duyarlılık	Seçicilik	F1	Doğruluk
BEKLE	0.82	0.61	0.88	0.70	0.75		BEKLE	0.86	0.61	0.91	0.71	0.77
SAT	0.62	0.86	0.83	0.73	0.84		SAT	0.68	0.85	0.85	0.75	0.85
AL	0.60	0.85	0.83	0.70	0.84		AL	0.71	0.84	0.86	0.77	0.85

**Şekil 4.4:** İkinci veri kümesinin eğitim ve test sonuçları.

Şekil 4.4 incelendiğinde, XGB modelinin tüm başarı metrikleri %70'in üzerindedir. Test ve eğitim kümelerindeki sonuçlar birbirine çok yakındır. Bu durum, modelin tutarlı bir şekilde eğitildiği anlamına gelmektedir. Birinci veri kümesi ile eğitilen XGB modelinin duyarlılık ve F1 puanının yeterli bulunmamasının aksine burada, 'al', 'sat' ve 'bekle' sınıfları için test kümesindeki duyarlılık oranları sırasıyla %78, %83 ve %78'dir. Seçicilik oranları incelendiğinde ise sırasıyla %91, %89 ve %85 olduğu görülmektedir. Bu sonuçlar, modelin tüm sınıflar için başarılı bir şekilde eğitildiğini göstermektedir. F1 puanının tüm sınıflar için %78 olması, model tahminlerinde sınıflar arasındaki dengesizlik probleminin çözüldüğü anlamına gelmektedir. RF modeli incelendiğinde, 'bekle' sınıfının duyarlılık oranının test kümesinde %66, eğitim kümesinde ise %69 olduğu görülmektedir. Ancak, F1 puanının test ve eğitim kümelerinde sırasıyla %72 ve %77 olması kesinlik ve duyarlılık arasındaki dengeyi sağlayabilmiş olduğunu göstermektedir. Aynı şekilde, 'bekle' sınıfı için RT modelinin duyarlılık oranı düşük olsa da F1 puanı kabul edilebilir seviyededir. XGB modelinin doğruluk oranlarına bakıldığında sırasıyla, test kümesi için %87, %87, %82, eğitim kümesi için %90, %89, %82 olmak üzere diğer modellerin doğruluk oranlarından çoğunlukla daha yüksektir. Tüm bu sonuçlar karşılaştırıldığında, XGB modelinin RF ve RT modellerine göre yüksek performansa sahip olduğu sonucuna varılmaktadır.

Sonuç olarak, Şekil 4.3 ve Şekil 4.4'deki başarı ölçüleri karşılaştırıldığında 'bekle' sınıfının diğer sınıflara oranla daha fazla sayıda gözlem içeren ikinci veri kümesi ile eğitilen modellerin birinci veri kümesi ile eğitilen modellere göre daha başarılı sonuç verdiği görülmektedir. İkinci veri kümesinin model sonuçları incelendiğinde (Şekil 4.4) eğitim ve test kümelerinin tutarlılığı açısından XGB modeli diğer modellere göre daha yüksek performans göstermektedir. Aynı şekilde duyarlılık oranları

incelendiğinde XGB modelinin pozitif gözlemleri doğru tahmin edebilme etkinliğinin RF ve RT modellerine göre daha yüksek seviyede olduğu görülmektedir. Modelin tüm başarı ölçüleri eğitim ve test kümelerine göre incelendiğinde ikinci veri kümesi ile eğitilen XGB modelinin kabul edilebilir ve tutarlı bir model olduğu sonucuna varılmaktadır.

## 5. SONUÇ

Bu çalışmada, BTC kripto para biriminin trend dönüşlerinin tahmin edilmesi için sınıflandırma algoritmalarından yararlanılmıştır. Çalışmada, 'binance.com' üzerinden alınan iki buçuk yılı kapsayan günlük güncel veri ve bu verilerden türetilen yeni özellikler ile çalışılmıştır. Temel bileşenler analizi ile veri kümesinin boyutu 10 bileşene düşürülmüş ve çoklu bağlantı sorununun önüne geçilmiştir. Alt örnekleme yönteminden yararlanılarak farklı sınıf oranlarına sahip iki farklı veri kümesi oluşturulmuştur. Veri kümeleri rastgele %20 test, %80 eğitim olmak üzere ayrılmıştır.

Tahmin modelleri, XGB, RF, RT, SVM, KNN, BN ve ANN algoritmaları kullanılarak geliştirilmiştir. SVM, KNN, BN ve ANN modelleri ağaç tabanlı topluluk öğrenmesi algoritmalarına göre çok düşük performans gösterdiğinden çalışmaya eklenmemiştir. XGB, RF ve RT algoritmaları kullanılarak 'al', 'sat', 'bekle' olmak üzere 3 sınıflı tahmin modelleri geliştirilmiştir. Sonuçlar, kesinlik, duyarlılık, seçicilik, F1 puanı ve doğruluk metriklerine göre karşılaştırılmıştır. 'Bekle' sınıfının diğer sınıflara göre daha fazla sayıda gözlem içeren ikinci veri kümesi ile geliştirilen modellerin daha başarılı sonuç verdiği görülmüştür. Test ve eğitim kümelerindeki başarı metrikleri karşılaştırıldığında, XGB modelinin RF ve RT modellerine göre daha yüksek performans gösterdiği sonucuna varılmıştır.

Model sonuçlarına göre, BTC fiyatının trend değişimlerine etki eden en önemli faktörlerin, fiyat çubuklarının konumu, RSI, EMA, SMA ve hacim bilgilerinden türetilen değişkenler olduğu görülmüştür. Ancak, bu değişkenler trendin sabit olduğu durumları yeteri kadar açıklayamamıştır. Bu sorun, yeniden örnekleme çalışmasında 'bekle' sınıfının gözlem sayısı artırılarak çözülmüştür. Sonuç olarak, tüm sınıflar eldeki değişkenler kullanılarak yeterli doğruluk ile tahmin edilebilmiştir.

Sonuçlar literatür taramasına göre karşılaştırıldığında, RF ve XGB gibi topluluk öğrenme modelleri diğer çalışmalarda olduğu gibi bu çalışmada da üstün bir performans göstermiştir. Literatürdeki çalışmaların model başarıları %50-%60

arasında iken bu alıřmada sadece trend donüşlerine odaklanılarak %80 ve üzeri doğrulukla modeller geliştirilebilmiştir.

## KAYNAKLAR

- Abdi, H., ve Williams, L. J.** (2010). Principal component analysis. *WIREs Computational Statistics*, 2(4), 433–459.
- Akyildirim, E., Goncu, A., ve Sensoy, A.** (2021). Prediction of cryptocurrency returns using machine learning. *Annals of Operations Research*, 297(1), 3–36, doi: 10.1002/wics.101.
- Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A., ve Quintana, D.** (2020). Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. *Expert Systems with Applications*, 149, 113250, doi:10.1016/j.eswa.2020.113250.
- Atçeken, M. A.** (2021). *Trading Strategy Based Classification on Cryptocurrency Price Prediction*, (Master of Science). TED University, Applied Data Science, Ankara.
- Benesty, J., Chen, J., ve Huang, Y.** (2008). On the Importance of the Pearson Correlation Coefficient in Noise Reduction. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(4), 757–765, doi:10.1109/TASL.2008.919072.
- Binance. (t.y.). Binance:** The World's Leading Cryptocurrency Exchange. <https://www.binance.com/>
- Bollinger, J. A.** (2002). *Bollinger on Bollinger Bands* (1st ed.). New York: McGraw-Hill.
- Boodhun, N., ve Jayabalan, M.** (2018). Risk prediction in life insurance industry using supervised learning algorithms. *Complex & Intelligent Systems*, 4(2), 145–154.
- Cavalli, S., ve Amoretti, M.** (2021). CNN-based multivariate data analysis for bitcoin trend prediction. *Applied Soft Computing*, 101, 107065, doi: 10.1016/j.asoc.2020.107065.

- Chen, M., Narwal, N., ve Schultz, M.** (2017). *Predicting Price Changes in Ethereum*, <http://cs229.stanford.edu/proj2017/final-reports/5244039.pdf>
- Chen, T., ve Guestrin, C.** (2016). *XGBoost: A Scalable Tree Boosting System*, <https://arxiv.org/abs/1603.02754>
- Chong, T. T.-L., Ng, W.-K., ve Liew, V. K.-S.** (2014). Revisiting the Performance of MACD and RSI Oscillators. *Journal of Risk and Financial Management*, 7(1), (1-12), doi:10.3390/jrfm7010001
- Cohen, G.** (2020). Forecasting Bitcoin Trends Using Algorithmic Learning Systems. *Entropy*, 22(8), 838, doi:10.3390/e22080838
- Fan, J., Zhang, C., ve Zhang, J.** (2001). Generalized Likelihood Ratio Statistics and Wilks Phenomenon. *The Annals of Statistics*, 29(1), 153–193, doi: 10.1214/aos/996986505.
- Gallegos-Eraza, F.** (2022). *Technical Indicator for a Better Intraday Understanding of Uptrends or Downtrends in the Financial Markets using Volume Transactions as a Trigger*, ICAIW 2022: Workshops at the 5th International Conference on Applied Informatics, Arequipa, Peru.
- Goodfellow, I., Bengio, Y., ve Courville, A.** (2018). *Deep Learning* (1st ed.). The MIT Press.
- Grandini, M., Bagli, E., ve Visani, G.** (2020). Metrics for Multi-Class Classification: An Overview, <https://arxiv.org/abs/2008.05756>, doi:10.48550/arXiv.2008.05756.
- Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., ve Xu, W.** (2018). Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics & Proteomics*, 15(1), 41–51.
- İnce, N. T.** (2019). *Predicting the Bitcoin Trend Using Technical Indicators for Deep Learning Algorithmic Features*, (Master of Arts). Boğaziçi University, Social Sciences, İstanbul.
- Ji, S., Kim, J., ve Im, H.** (2019). A Comparative Study of Bitcoin Price Prediction Using Deep Learning. *Mathematics*, 7(10),898, doi: 10.3390/math7100898.
- Korb, K. B., ve Nicholson, A. E.** (2010). *Bayesian Artificial Intelligence*. CRC Press.
- Kotsiantis, S., Kanellopoulos, D., ve Pintelas, P.** (2006). *Handling imbalanced datasets: A review*, GESTS International Transactions on Computer Science and Engineering, Vol.30.



- Kwon, D.-H., Kim, J.-B., Heo, J.-S., Kim, C.-M., ve Han, Y.-H.** (2019). Time Series Classification of Cryptocurrency Price Trend Based on a Recurrent LSTM Neural Network. *Journal of Information Processing Systems*, 15(3), 694–706, doi: 10.3745/JIPS.03.0120.
- Livieris, I. E., Pintelas, E., Stavroyiannis, S., ve Pintelas, P.** (2020). Ensemble Deep Learning Models for Forecasting Cryptocurrency Time-Series. *Algorithms*, 13(5), 121, doi: 10.3390/a13050121.
- Luque, A., Carrasco, A., Martín, A., ve de las Heras, A.** (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216–231, doi:10.1016/j.patcog.2019.02.023
- Qiang, Z., ve Shen, J.** (2021). *Bitcoin High-Frequency Trend Prediction with Convolutional and Recurrent Neural Networks*, [http://cs230.stanford.edu/projects\\_winter\\_2021/reports/70308950.pdf](http://cs230.stanford.edu/projects_winter_2021/reports/70308950.pdf)
- Rockefeller, B.** (2011). *Technical analysis for dummies (2nd ed.)*. John Wiley & Sons.
- Saru, S., ve Subashree, S.** (2019). Analysis and Prediction of Diabetes Using Machine Learning. *Applied Computing Ejournal*, Vol.5, Issue.4.
- Sekhar, Ch. R., Minal ve Madhu, E.** (2016). Mode Choice Analysis Using Random Forrest Decision Trees. *Transportation Research Procedia*, 17, 644–652, doi: 10.1016/j.trpro.2016.11.119.
- Shintate, T., ve Pichl, L.** (2019). Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning. *Journal of Risk and Financial Management*, 12(1), 17, doi: 10.3390/jrfm12010017.
- Valencia, F., Gómez-Espinosa, A., ve Valdés-Aguirre, B.** (2019). Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. *Entropy*, 21(6), 589, doi: 10.3390/e21060589.

## EKLER

### Ek A: Veri Kümesi

**Çizelge A.1: Özellikler ve veri sözlüğü.**

Özellik Sıra No	Özellik Adı	Özellik Ölçüsü	Açıklama /Hesaplama
1	LOW	sürekli	Günün en düşük değeri
2	HIGH	sürekli	Günün en yüksek değeri
3	CLOSE	sürekli	Günün kapanış değeri
4	OPEN	sürekli	Günün açılış değeri
5	CLOSE_PRV1	sürekli	t-1 kapanış değeri
6	OPEN_PRV1	sürekli	t-1 açılış değeri
7	HIGH_PRV1	sürekli	t-1 high değeri
8	LOW_PRV1	sürekli	t-1 low değeri
9	LOWER_HIGH	flag	HIGH < HIGH_PRV1
10	LOWER_LOW	flag	LOW < LOW_PRV1
11	HIGHER_HIGH	flag	HIGH > HIGH_PRV1
12	HIGHER_LOW	flag	LOW > LOW_PRV1
13	HIGH_DIFF_PERCENT	sürekli	(HIGH - HIGH_PRV1) / HIGH
14	SIG_HIGHER_HIGH	flag	HIGHER_HIGH = 1 and HIGH_DIFF_PERCENT >= 0.05
15	ROW_ID	ID	@INDEX, 1'den başlayan tekil numara
16	UP_BAR	flag	HIGHER_HIGH = 1 and HIGHER_LOW = 1, high ve low bir önceki günden daha yüksek
17	UP_BAR_COUNT	sürekli	Ardışık up barların kümülatif toplamı
18	UP_TREND_FLAG	flag	Yükselen trend
19	DOWN_BAR	flag	LOWER_HIGH= 1 and LOWER_LOW = 1, high ve low bir önceki günden daha düşük
20	DOWN_BAR_COUNT	sürekli	Ardışık down barların kümülatif toplamı
21	DOWN_TREND_FLAG	flag	Düşen trend
22	STABLE	flag	Trend yok
23	UP_DAY	flag	Kapanış bir önceki günden daha yüksek
24	DOWN_DAY	flag	Kapanış bir önceki günden daha düşük
25	TICK_MARKS	sürekli	Açılış ve kapanış arasındaki mutlak fark abs(OPEN-CLOSE)
26	DOJI	flag	Açılış ve kapanış eşit veya çok yakın
27	SHADOW_TOP	sürekli	Üst gölgenin uzunluğu
28	SHADOW_BOTTOM	sürekli	Alt gölgenin uzunluğu
29	SHAVEN_TOP	flag	Üst gölge yok veya çok kısa
30	SHAVEN_BOTTOM	flag	Alt gölge yok veya çok kısa
31	WHITE_BODY	flag	CLOSE > OPEN olan günler
32	BLACK_BODY	flag	CLOSE < OPEN olan günler
33	RISING_WINDOW	flag	@OFFSET(WHITE_BODY,1)=1 and @OFFSET(WHITE_BODY,2)=1 and WHITE_BODY=1 and @OFFSET(OPEN,1) > @OFFSET(CLOSE,2) and OPEN >= @OFFSET(OPEN,1) @OFFSET(WHITE_BODY,1)=1 and @OFFSET(WHITE_BODY,2)=1 and WHITE_BODY=1 and @OFFSET(CLOSE,12 < @OFFSET(CLOSE,1) and CLOSE > @OFFSET(CLOSE,1) and @OFFSET(OPEN,2) < @OFFSET(OPEN,1) and OPEN > @OFFSET(OPEN,1) @OFFSET(BLACK_BODY,1)and @OFFSET(BLACK_BODY,2)=1 and BLACK_BODY=1 and
34	THREE_WHITE_SOLDIERS	flag	@OFFSET(CLOSE,2) > @OFFSET(CLOSE,1) and CLOSE < @OFFSET(CLOSE,1) and @OFFSET(OPEN,2) > @OFFSET(OPEN,1)
35	THREE_BLACK_CROWS	flag	@OFFSET(CLOSE,2) > @OFFSET(CLOSE,1) and CLOSE < @OFFSET(CLOSE,1) and @OFFSET(OPEN,2) > @OFFSET(OPEN,1)

			and
			OPEN < @OFFSET(OPEN,1)
36	WHITE_BODY_LONG	flag	uzun beyaz gövde
37	BLACK_BODY_LONG	flag	uzun siyah gövde
38	BAR_RANGE (Trading Range)		HIGH ve LOW arasındaki fark
39	BAR_RANGE_DIFF	sürekli	BAR_RANGE_CAT - @OFFSET(BAR_RANGE_CAT,1)
40	LONG_BAR	flag	Bir önceki güne göre çok daha uzun olan çubuklar
41	INSIDE_DAY	flag	HIGH < HIGH_PRV1 and LOW > LOW_PRV1
42	OUTSIDE_DAY	flag	HIGH > HIGH_LAG1 and LOW < LOW_LAG1
43	SPIKE	flag	VOLUME_BTC >= @OFFSET(VOLUME_MEAN_3,1) * 2
44	GAP	flag	Ardışık iki çubuk arasındaki boşluk
45	UP_VOLUME	flag	VOLUME_BTC > @OFFSET(VOLUME_BTC,1)
46	UP_VOLUME_FLAG	flag	En az son 3 gün up volume olmuş
47	UP_VOLUME_DEGREE	sürekli	Ardışık up volume flag günlerinin kümülatif toplamı
48	DOWN_VOLUME	flag	VOLUME_BTC < @OFFSET(VOLUME_BTC,1)
49	DOWN_VOLUME_FLAG	flag	En az son 3 gün down volume olmuş
50	DOWN_VOLUME_DEGREE	sürekli	Ardışık down volume flag günlerinin kümülatif toplamı
51	OBV	sürekli	On balanced volume
52	UP_RANGE	flag	BAR_RANGE > @OFFSET(BAR_RANGE,1)
53	UP_RANGE_FLAG	flag	En az 3 gün up range olmuş
54	DOWN_RANGE	flag	BAR_RANGE < @OFFSET(BAR_RANGE,1)
55	DOWN_RANGE_FLAG	flag	En az 3 gün down range olmuş
56	UP_RANGE_HIGH_VOLUME	flag	UP_RANGE_FLAG = 1 and UP_VOLUME_FLAG=1
57	DOWN_RANGE_DOWN_VOLUME	flag	DOWN_RANGE_FLAG = 1 and DOWN_VOLUME_FLAG=1
58	UP_RANGE_HIGHER_CLOSE	flag	UP_RANGE_FLAG = 1 and UP_DAY=1
59	UP_RANGE_LOWER_CLOSE	flag	UP_RANGE_FLAG = 1 and DOWN_DAY=1
60	DOWN_RANGE_HIGHER_CLOSE	flag	DOWN_RANGE_FLAG = 1 and UP_DAY =1
61	DOWN_RANGE_LOWER_CLOSE	flag	DOWN_RANGE_FLAG = 1 and DOWN_DAY =1
62	MAX_MOVE	sürekli	@MAX(HIGH,5)-@MIN(LOW,5)
63	SDEV_RANGE (VIX)	sürekli	@SDEV(BAR_RANGE,5)
64	BOLLINGER_BAND_TOP	sürekli	Bollinger bandının üst sınırı
65	BOLLINGER_BAND_BOTTOM	sürekli	Bollinger bandının alt sınırı
66	BOLLINGER_RANGE	sürekli	BOLLINGER_BAND_TOP - BOLLINGER_BAND_BOTTOM
67	CO_BUY	flag	CLOSE > SMA_5 (Crossover), kapanış fiyatı 3 adım geriden alındı
68	CO_SELL	flag	CLOSE < SMA_5 (Crossover), kapanış fiyatı 3 adım geriden alındı
69	SMA_n	sürekli	Kapanışın n'li basit hareketli ortalaması
70	UP_SMA5	flag	SMA_5 > @OFFSET(SMA_5,1) and DOWN_TREND_FLAG=1
71	DOWN_SMA5	flag	SMA_5 < @OFFSET(SMA_5,1) and UP_TREND_FLAG=1
72	SMA_BUY	flag	SMA_5 > SMA_20 and @OFFSET(SMA_5,1)<= @OFFSET(SMA_20,1)
73	SMA_SELL	flag	SMA_5 < SMA_20 and @OFFSET(SMA_5,1)>= @OFFSET(SMA_20,1)
74	DAYLIGHT (Open Space)	sürekli	abs(SMA_5 - SMA_20)
75	CON_DIV	flag	SMA_5 - SMA_20 (yakınsama ve iraksama ölçüsü)
76	EMA_n	sürekli	Kapanışın n'li üssel hareketli ortalaması
77	MACD	sürekli	moving average convergence-divergence
78	MOMENTUM	sürekli	CLOSE-@OFFSET(CLOSE,12)
79	RSI_5	sürekli	5'li basit hareketli ortalamaya göre hesaplanan relative strength index
80	RSI_14	sürekli	14'lü basit hareketli ortalamaya göre hesaplanan relative strength index
81	RSI_5_SELL	flag	RSI_5 > 70
82	RSI_14_SELL	flag	RSI_14 > 70

83	RSI_5_BUY	flag	RSI_5 < 30
84	RSI_14_BUY	flag	RSI_14 < 30
85	RSI_SELL	flag	5 ve 14 ün birleşimi
86	RSI_BUY	flag	5 ve 14 ün birleşimi
87	LOW_Min5	sürekli	@MIN(LOW,5)
88	HIGH_Max5	sürekli	@MAX(HIGH,5)
89	K_PERCENT	sürekli	((CLOSE-LOW_Min5) / (HIGH_Max5-LOW_Min5))*100
90	D_PERCENT	sürekli	K_PERCENT'in 3'lü basit hareketli ortalaması
91	TARGET	kategorik	al, sat, bekle olmak üzere üç sınıflı bağımlı değişken
75	CON_DIV	flag	SMA_5 - SMA_20 (yakınsama ve ıraksama ölçüsü)
76	EMA_n	sürekli	Kapanışın n'li üssel hareketli ortalaması
77	MACD	sürekli	moving average convergence-divergence
78	MOMENTUM	sürekli	CLOSE-@OFFSET(CLOSE,12)
79	RSI_5	sürekli	5'li basit hareketli ortalamaya göre hesaplanan relative strength index
80	RSI_14	sürekli	14'lü basit hareketli ortalamaya göre hesaplanan relative strength index
81	RSI_5_SELL	flag	RSI_5 > 70
82	RSI_14_SELL	flag	RSI_14 > 70
83	RSI_5_BUY	flag	RSI_5 < 30
84	RSI_14_BUY	flag	RSI_14 < 30
85	RSI_SELL	flag	5 ve 14 ün birleşimi
86	RSI_BUY	flag	5 ve 14 ün birleşimi
87	LOW_Min5	sürekli	@MIN(LOW,5)
88	HIGH_Max5	sürekli	@MAX(HIGH,5)
89	K_PERCENT	sürekli	((CLOSE-LOW_Min5) / (HIGH_Max5-LOW_Min5))*100
90	D_PERCENT	sürekli	K_PERCENT'in 3'lü basit hareketli ortalaması
91	TARGET	kategorik	al, sat, bekle olmak üzere üç sınıflı bağımlı değişken

## Ek B: Temel Bileşenler Analizi

	Component									
	1	2	3	4	5	6	7	8	9	10
LOWER_HIGH				-.884						
LOWER_LOW			.840							
HIGHER_HIGH				.883						
HIGHER_LOW			-.839							
UP_BAR			-.348	.806						
UP_BAR_COUNT				.714					.340	
UP_TREND_FLAG				.361				.694	.304	
DOWN_BAR			.790	-.401						
DOWN_BAR_COUNT			.729							
DOWN_TREND_FLAG			.613					-.318		.399
UP_DAY		.968								
DOWN_DAY		-.968								
TICK_MARKS	.697									
SHADOW_TOP	.680									
WHITE_BODY		.969								
BLACK_BODY		-.969								
BAR_RANGE	.893									
UP_VOLUME_DEGREE										-.566
DOWN_VOLUME_DEGREE										.558
OBV	.794									
MAX_MOVE	.944									
SDEV_RANGE	.860									
BOLLINGER_BAND_TOP	.926									
BOLLINGER_BAND_BOTTOM	.907									
BOLLINGER_RANGE	.894									
SMA_5	.917									
CO_BUY							-.955			
CO_SELL							.955			
UP_SMA5								-.345		.655
DOWN_SMA5								.860		
DAYLIGHT	.734									
CON_DIV					.963					
MACD					.889					
MOMENTUM					.908					
RSI_5							.838			
RSI_SELL							.301		.643	
RSI_BUY							-.867			
K_PERCENT		.339	-.339	.330		.469			.315	
D_PERCENT					.324	.520	.338		.416	

Extraction Method: Principal Component Analysis.  
 Rotation Method: Varimax with Kaiser Normalization. <sup>a</sup>  
 a. Rotation converged in 17 iterations.

Şekil B.1: Bileşen matrisi

## Ek C: Model Eğitimi

**Çizelge C.1: Hiperparametre açıklamaları.**

Model	Hiperparametre Adı	Açıklama
RF	bootstrap	Rastgele Altörneklem Yöntemi kullanılarak veri kümesinde bir ağaç oluşturulurken kullanılacak örneklem miktarını belirleyen bir parametredir.
RF	ccp_alpha	Karar ağacında yapılacak olan kesme noktasını belirleyen bir parametredir. Bu parametre ne kadar yüksekse, ağaçta kesilme noktası o kadar yüksek olur.
RF	criterion	Ağaç oluşturulurken kullanılacak bölme kriterlerini belirleyen bir parametredir. Örneğin, "gini" veya "entropy" gibi kriterler kullanılabilir.
RF	max_depth	Oluşturulan ağacın maksimum derinliğini belirleyen bir parametredir. Her düğümdeki bölünmelerde kullanılacak özellik sayısını belirleyen bir parametredir. "max_features='auto'" parametresi, her düğümdeki bölünmelerde kullanılacak olan özellik sayısını otomatik olarak belirlemek için kullanılır.
RF	max_features	Bölünme sonrası düşük kaliteye sahip olan düğümlerin kesilmesi için kullanılan bir parametredir.
RF	min_impurity_decrease	Bir yaprak düğümünde en az kaç örneklem olması gerektiğini belirleyen bir parametredir.
RF	min_samples_leaf	Bir düğümün iki alt düğüme bölünmesi için gerekli olan minimum örneklem sayısını belirleyen bir parametredir.
RF	min_samples_split	Bir yaprak düğümünde en az kaç ağırlık fraksiyonu olması gerektiğini belirleyen bir parametredir.
RF	min_weight_fraction_leaf	Oluşturulacak olan ağaç sayısını belirleyen bir parametredir.
RF	n_estimators	Oluşturulan modelin out-of-bag verileri üzerindeki doğruluk oranını hesaplamak için kullanılan bir parametredir.
RF	oob_score	Eğitim sırasında çıktının ne kadar ayrıntılı olacağını belirleyen bir parametredir. "verbose=0" sıfır olduğunda, çıktı gösterilmezken, "verbose=1" olduğunda, çıktı eğitim ilerlemesi hakkında bilgi verir. Yeni bir model eğitime başlamadan önce önceki eğitimden kaldığı yerden devam etmek için kullanılan bir parametredir. Bu, daha kısa eğitim süreleri sağlayabilir.
RF	verbose	Random Forests algoritmasında, oluşturulacak olan ağaç sayısını belirleyen bir parametredir.
RF	warm_start	Gradient Boosting algoritmasında, her aşamada kullanılacak olan önceki ağaçların katkısının ne kadar olacağını belirleyen bir parametredir. Daha yüksek bir öğrenme oranı, daha hızlı öğrenme ve daha yüksek varyansla sonuçlanabilirken, daha düşük bir öğrenme oranı daha yavaş öğrenme ve daha düşük varyansla sonuçlanabilir.
RF	number_of_trees_to_build	Gradient Boosting algoritmasında, kullanılacak olan maksimum ağaç sayısını belirleyen bir parametredir.
RF	learning_rate	7 Gradient Boosting algoritmasında, eğitim işlemi sırasında kullanılacak olan maksimum değerlendirme sayısını belirleyen bir parametredir. Bu, aşırı uyum sorunlarını önlemek için kullanılabilir.
RF	max_iteration	XGBoost kütüphanesinde, ağaçların oluşturulması için kullanılan yöntemi belirleyen bir parametredir. Bu parametre "auto", "exact", "approx", "hist" gibi değerler alabilir.
RF	max_evaluation	XGBoost algoritmasında, oluşturulacak olan ağaç sayısını belirleyen bir parametredir.
XGB	tree_method	Ağaçların maksimum derinliğini belirleyen bir parametredir. Bu parametre, aşırı öğrenme problemlerini önlemek için kullanılır.
XGB	number_boost_round	Ağacın bir düğümde bölünmesi için gereken minimum örnek sayısını belirleyen bir parametredir. Bu, düğümlerin daha iyi bir şekilde geliştirilmesine yardımcı olabilir.
XGB	max_depth	Adım büyüklüğünü sınırlamak için kullanılan bir parametredir. Bu parametre, aşırı değerlerin neden olduğu problemleri önleyebilir.
XGB	min_child_weight	XGBoost algoritmasında, optimize edilecek olan amaç fonksiyonunu belirleyen bir parametredir. Bu parametre, sınıflandırma veya regresyon gibi farklı görevler için farklı amaç fonksiyonları içerebilir.
XGB	max_delta_step	Eğitim işlemi sırasında, ağaç oluşturma işleminin durması gerektiğini belirleyen bir parametredir. Bu, eğitim süresini kısaltabilir ve aşırı uyum sorunlarını önleyebilir.
XGB	objective	Eğitim işlemi sırasında kullanılan verilerin yüzdesini belirleyen bir parametredir. Bu, eğitim süresini kısaltabilir ve modelin daha iyi bir şekilde gelişmesine yardımcı olabilir.
XGB	stopping_rounds	Ağaçların oluşturulması sırasında kullanılacak olan alt örneklem
XGB	evaluation_data_ratio	
XGB	sub_sample	

		oranını belirleyen bir parametredir. Bu, aşırı uyum problemlerini önlemeye yardımcı olabilir.
XGB	eta	Ağaçların oluşturulması sırasında kullanılacak olan öğrenme oranını belirleyen bir parametredir. Bu, ağaçların daha iyi bir şekilde genelleşmesine yardımcı olabilir.
XGB	gamma	Ağaçların oluşturulması sırasında kullanılan minimum düşüş değerini belirleyen bir parametredir. Bu, ağaçların daha iyi bir şekilde genelleşmesine yardımcı olabilir.
XGB	colsample_by_tree	Ağaçların oluşturulması sırasında kullanılacak olan özelliklerin yüzdesini belirleyen bir parametredir.
XGB	colsample_by_level	Ağaç seviyelerinde kullanılacak olan özelliklerin yüzdesini belirleyen bir parametredir.
XGB	lambda	Ağacın oluşturulması sırasında kullanılan L2 düzenlemesi katsayısını belirleyen bir parametredir. Bu, aşırı uyum problemlerini önleyebilir.
XGB	alpha	Ağacın oluşturulması sırasında kullanılan L1 düzenlemesi katsayısını belirleyen bir parametredir. Bu, aşırı uyum problemlerini önleyebilir.
XGB	scale_pos_weight	Dengesiz veri kümelerinde sınıflandırma yaparken kullanılan bir parametredir. Bu parametre, olumsuz etkileri azaltmak için pozitif sınıf örneklerinin ağırlığını artırabilir.
RT	number_of_models_to_build	Random Trees algoritması, birden fazla karar ağacı kullanarak bir model oluşturur. Bu parametre, oluşturulacak karar ağacı sayısını belirler.
RT	sample_size	Random Trees algoritması, her bir karar ağacı için rastgele örneklemeler oluşturur. Bu parametre, her örneklemede kullanılacak örneklem büyüklüğünü belirler.
RT	max_number_of_nodes	Karar ağaçları, dallanmaları ile oluşan bir yapıdır ve bu parametre, oluşturulacak karar ağaçlarında maksimum düğüm sayısını belirler.
RT	max_tree_depth	Karar ağaçları, bir örneğin sonucunu belirlemek için kullanılan birden fazla karar düğümünden oluşur. Bu parametre, oluşturulacak karar ağacının maksimum derinliğini belirler.
RT	min_child_node_size	Bu parametre, karar ağacı yapısında minimum düğüm boyutunu belirler.
RT	max_percentage_of_missing_values	Karar ağacı oluşturulurken, eksik verilerin nasıl ele alınacağını belirler. Bu parametre, maksimum eksik veri yüzdesini belirler.
RT	exclude_fields_with_a_single_category_majority_over(%)	Bu parametre, veri kümesindeki özellikleri belirler. Bu parametre, yalnızca belirtilen yüzdenin üzerinde bir sınıf oranına sahip olan özellikleri hariç tutar.
RT	max_number_of_field_categories	Bu parametre, veri kümesindeki özelliklerin maksimum sınıf sayısını belirler.
RT	min_field_variation	Bu parametre, veri kümesindeki özelliklerin minimum değişkenliğini belirler.
RT	number_of_bins	Karar ağacı oluşturulurken, sayısal verilerin nasıl bölüneceğini belirler. Bu parametre, sayısal özellikleri bölmek için kullanılacak maksimum bölme sayısını belirler.
RT	number_of_interesting_rules_to_report	Bu parametre, oluşturulan karar ağacından rapor edilecek maksimum kural sayısını belirler.
RF	bootstrap	Rastgele Altörneklem Yöntemi kullanılarak veri kümesinde bir ağaç oluşturulurken kullanılacak örneklem miktarını belirleyen bir parametredir.
RF	ccp_alpha	Karar ağacında yapılacak olan kesme noktasını belirleyen bir parametredir. Bu parametre ne kadar yüksekse, ağaçta kesilme noktası o kadar yüksek olur.
RF	criterion	Ağaç oluşturulurken kullanılacak bölme kriterlerini belirleyen bir parametredir. Örneğin, "gini" veya "entropy" gibi kriterler kullanılabilir.