T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KRİPTO PARALARIN GELECEKTEKİ TAHMİNİ

Elif Dilasa KURT

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Prof. Dr. Banu DİRİ

Eş-Danışman
Öğr. Gör. Dr. Ahmet ELBİR

Temmuz, 2024

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KRİPTO PARALARIN GELECEKTEKİ TAHMİNİ

Elif Dilasa KURT tarafından hazırlanan tez çalışması 11.07.2024 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Banu DİRİ
Yıldız Teknik Üniversitesi
Danışman

Öğr. Gör. Dr. Ahmet ELBİR
Yıldız Teknik Üniversitesi
Eş-Danışman

Jüri Üyeleri

Prof. Dr. Banu DİRİ, Danışman
Yıldız Teknik Üniversitesi _____

Prof.Dr. Mehmet Sıddık AKTAŞ, Üye
Yıldız Teknik Üniversitesi _____

Prof.Dr. Nizamettin AYDIN, Üye
İstanbul Teknik Üniversitesi _____

Danışmanım Prof. Dr. Banu DİRİ sorumluluğunda tarafımca hazırlanan MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KRİPTO PARALARIN GELECEKTEKİ TAHMİNİ başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Elif Dilasa KURT

İmza

Aileme ve dostlarıma ithafen

TEŞEKKÜR

Bu tez çalışmam süresince desteğini ve yardımını esirgemeyen birçok kişi ve kuruma teşekkürlerimi sunarım. Öncelikle, tez çalışmam boyunca bana rehberlik eden ve değerli katkılarıyla sürecin her aşamasında destek olan danışmanlarım Banu Diri ve Ahmet Elbir'e sonsuz teşekkür ederim. Kendileri, akademik yolculuğum boyunca bana ilham kaynağı olmuştur.

Tezimin hazırlanmasında bana yardımcı olan Yıldız Teknik Üniversitesi öğretim üyelerine, arkadaşlarıma ve aileme de teşekkürü borç bilirim. Özellikle, tezimin tamamlanmasında bana destek olan aileme, gösterdikleri sabır ve anlayış için minnettarım.

Ayrıca, çalışmam boyunca bana veri sağlayan ve çeşitli kaynaklardan yararlanmamı sağlayan Novacon Bilişim Danışmanlık Şirketi'ne de teşekkür ederim. Bu tez, onların katkıları olmadan bu kadar kapsamlı ve detaylı olamazdı.

Son olarak, çalışmam boyunca bana motivasyon veren tüm dostlarıma ve sevdiklerime teşekkür ederim. Her biri, bu sürecin bir parçası olmuştur ve başarıma katkıda bulunmuştur.

Elif Dilasa KURT

İÇİNDEKİLER

SİMGE LİSTESİ vii

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ ix

TABLO LİSTESİ x

ÖZET xi

ABSTRACT xiii

1 GİRİŞ 1

1.1 Tezin Amacı ve Önemi 1

1.2 Problem Tanımı 2

1.3 Çalışmanın Kapsamı ve Sınırları 2

1.4 Literatür Taraması 3

2 ÖNERİLEN YÖNTEM 7

2.1 Zaman Serisi Modelleri 7

2.1.1 AutoRegressive Integrated Moving Average - ARIMA . . . 7

2.1.2 Long Short-Term Memory - LSTM 9

2.2 Gradyan Artırma Yöntemleri 10

2.2.1 XGBoost 10

2.2.2 LightGBM 12

2.2.3 CatBoost 15

3 MODEL GELİŞTİRME VE EĞİTİM 17

3.1 Veri Toplama ve Hazırlama 18

3.1.1 Veri Seti Özellikleri ve İçeriği 18

3.1.2 Kripto Para Birimleri ve Zaman Dilimleri 19

3.1.3 Analiz ve Uygulamalar 20

3.1.4 Etik Kurallar 20

3.2 Veri Ön İşleme Teknikleri 21

4 DENEYSEL SONUÇLAR **22**
4.1 Zaman Dilimine Göre Değerlendirme 22
4.2 Kapanış Değerlerine Göre Değerlendirme 22

5 SONUÇ **31**
5.1 Sonuç 31
5.2 Gelecek Çalışmalar ve Öneriler 32

KAYNAKÇA **33**

TEZDEN ÜRETİLMİŞ YAYINLAR **35**

SİMGE LİSTESİ

T	Ağacın yaprak sayısı
γ	Düzenleme parametresi
Ω	Düzenleme terimi
l	Eğitim hatasını ölçen bir fonksiyonu
w_j	j. yaprağın ağırlığı
λ	Regülarizasyon parametresi

KISALTMA LİSTESİ

ADAUSDT	Cardano / TetherUS
ANN	Artificial Neural Network
ARIMA	AutoRegressive Integrated Moving Average
BNBUSDT	Binance Coin / TetherUS
BTCUSDT	Bitcoin / TetherUS
CatBoost	Categorical Boosting
CNN	Convolutional Neural Network
DASHUSDT	Dash / TetherUS
dk	Dakika
DOGEUSDT	Dogecoin / TetherUS
ETHUSDT	Ethereum / TetherUS
GBM	Gradient Boosting Machine
GRU	Gated Recurrent Unit
LightGBM	Light Gradient Boosting Machine
LR	Logistic Regression
LSTM	Long Short-Term Memory
LTCUSDT	Litecoin / TetherUS
MSE	Mean Squared Error
NEOUSDT	Neo / TetherUS
RF	Random Forest
RMSE	Root Mean Squared Error
TRXUSDT	TRON / TetherUS
XGBoost	Extreme Gradient Boosting
XRPUSDT	Ripple / TetherUS

ŞEKİL LİSTESİ

Şekil 2.1 LSTM mimarisi [20] 10
Şekil 2.2 XGboost: Yaprak tabanlı ağaç büyüme stratejisi [9] 11
Şekil 2.3 LightGBM: Seviye tabanlı ağaç büyüme stratejisi [9] 13
Şekil 3.1 İşlem adımları 17
Şekil 3.2 Bir saatlik kripto para birimlerinin tarihe göre kapanış değerleri . 20
Şekil 4.1 15 Dakika-TRXUSDT-CATBOOST-RMSE 23
Şekil 4.2 15 Dakika-TRXUSDT-CATBOOST-TEST 24
Şekil 4.3 15 Dakika-DOGEUSDT-CATBOOST-RMSE 25
Şekil 4.4 15 Dakika-DOGEUSDT-CATBOOST-TEST 25
Şekil 4.5 15 Dakika-XRPUSDT-CATBOOST-RMSE 26
Şekil 4.6 15 Dakika-XRPUSDT-CATBOOST-TEST 26
Şekil 4.7 15 Dakika-ADAUSDT-CATBOOST-RMSE 27
Şekil 4.8 15 Dakika-ADAUSDT-CATBOOST-TEST 27
Şekil 4.9 15 Dakika-BTCUSDT-CATBOOST-RMSE 28
Şekil 4.10 15 Dakika-BTCUSDT-CATBOOST-TEST 28

TABLO LİSTESİ

Tablo 2.1 LightGBM, XGBoost ve CatBoost Modellerinin Karşılaştırılması 16
Tablo 4.1 15 Dakikalık RMSE Sonuçları 29
Tablo 4.2 1 Saatlik RMSE Sonuçları 30
Tablo 4.3 4 Saatlik RMSE Sonuçları 30

ÖZET

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KRİPTO PARALARIN GELECEKTEKİ TAHMİNİ

Elif Dilasa KURT

Bilgisayar Mühendisliği Anabilim Dalı
YÜKSEK LİSANS Tezi

Danışman: Prof. Dr. Banu DİRİ
Eş-Danışman: Öğr. Gör. Dr. Ahmet ELBİR

Günümüzde her türden veri, makine öğrenmesi algoritmalarıyla işlenmekte ve bu verilerden yararlanarak tahmin yapacak modeller oluşturulmaktadır. Bu algoritmaların ürettiği modeller sayesinde geleceğe yönelik tahminler yapmak, her sektör için büyük bir önem taşır. Özellikle ekonomi ve finans sektörü, geleceğe yönelik tahminlerin yapılmasının kritik olduğu alanlardan biridir. Finans sektöründe son dönemde dikkat çeken kripto para piyasaları, kripto paralar başta olmak üzere birçok finansal unsurun ticaretinin yapıldığı, dünya genelinde yaygın olarak kullanılan finansal varlık sınıflarından birisidir. Kripto para piyasalarının sosyal medya etkileşimleri, son dakika haberleri, savaş veya salgın gibi olağanüstü durumlar ve spekülatif davranışlar gibi çeşitli faktörler nedeniyle yüksek dalgalanma ve karmaşıklık göstermesi, bu kripto para birimlerinin fiyat tahmininde düşük hata payıyla sonuç almak açısından çeşitli zorluklar yaratmaktadır. Bu faktörler ve zorluklar altında, yapay zekâ ve makine öğrenimi yöntemlerinin kullanımı, günümüzde giderek daha fazla önem kazanmaktadır. Özellikle kısa vadeli tahminler yaparak bu finansal araçlardaki yatırımlardan zarar etmemek için bu yöntemlerin kullanımı çoğu finans sektörü kullanıcısının ortak hedefi olmuştur. Bu çalışmada, çeşitli kripto paraların fiyat tahminleri için ARIMA, LSTM, XGBoost, LightGBM ve CatBoost gibi beş farklı makine öğrenmesi yöntemi kullanılmıştır. Deneysel çalışmalar; 15 dakika, 1 saat ve 4 saatlik zaman dilimlerinde ve on farklı kripto para birimi (XRPUSDT, TRXUSDT, NEOUSDT, LTCUSDT, ETHUSDT, DOGEUSDT, DASHUSDT, BTCUSDT,

BNBUSDT ve ADAUSDT) için gerçekleştirilmiştir. Elde edilen deneysel sonuçlar değerlendirildiğinde, kök ortalama kare hatası (RMSE) metriği açısından CatBoost algoritması en iyi performansı göstermiştir. Ek olarak, düşük kapanış değerlerine sahip kripto paraların daha iyi tahmin sonuçları verdiği gözlemlenmiştir. Bu sonuçlar, kripto para piyasalarında etkili fiyat tahminleri yapabilmek için makine öğrenimi yöntemlerinin ne kadar kullanışlı olduğunu ortaya koymaktadır. Bu bağlamda, kripto para yatırımcıları ve finans sektörü uzmanları için geleneksel analiz yöntemlerine ek olarak bu yöntemleri kullanmak daha hızlı ve daha bilinçli bir analiz yapabilmek mümkün olabilecektir. Bu türdeki çalışmalardan elde edilen deneyim bu alanda gelecekteki yapılacak çalışmalara öncülük edebilecektir.

Anahtar Kelimeler: kripto para birimleri, makine öğrenimi, fiyat tahmini, zaman serisi analizi.

ABSTRACT

FUTURE PREDICTION OF CRYPTOCURRENCIES USING MACHINE LEARNING METHODS

Elif Dilasa KURT

Department of Computer Engineering
Master of Science Thesis

Supervisor: Prof. Dr. Banu DİRİ
Co-supervisor: "Lecturer Dr. Ahmet ELBİR

Today, all kinds of data are processed with machine learning algorithms and models are created to make predictions using these data. Making predictions for the future thanks to the models produced by these algorithms is of great importance for every sector. Especially the finance sector is one of the areas where it is critical to make predictions for the future. Cryptocurrency markets, which have recently attracted attention in the finance sector, are one of the widely used financial asset classes worldwide, where many financial elements, especially cryptocurrencies, are traded. The high volatility and complexity of cryptocurrency markets due to various factors such as social media interactions, breaking news, extraordinary situations such as war or pandemic, and speculative behaviours create various challenges in terms of getting results with a low margin of error in price prediction of these cryptocurrencies. Under these factors and challenges, the use of artificial intelligence and machine learning methods is becoming increasingly important today. The use of these methods has been a common goal of most financial sector users, especially in order to avoid losses from investments in these financial instruments by making short-term forecasts. In this study, five different machine learning methods such as ARIMA, LSTM, XGBoost, LightGBM and CatBoost have been employed for price predictions of various cryptocurrencies. Experimental studies have been conducted for ten different cryptocurrencies (XRPUSDT, TRXUSDT, NEOUSDT, LTCUSDT,

ETHUSDT, DOGEUSDT, DASHUSDT, BTCUSDT, BNBUSDT and ADAUSDT) in 15 minutes, 1 hour and 4 hours time periods. When the experimental results are evaluated, the CatBoost algorithm performed the best in terms of the root mean square error (RMSE) metric. Additionally, it has been observed that cryptocurrencies with low closing values gave better prediction results. These results reveal how useful machine learning methods are for making effective price predictions in cryptocurrency markets. In this context, it will be possible for cryptocurrency investors and financial sector experts to use these methods in addition to traditional analysis methods to make a faster and more informed analysis. The experience gained from this type of studies can lead to future studies in this field.

Keywords: cryptocurrency, machine learning, price prediction, financial markets, time series analysis.

1
GİRİŞ

1.1 Tezin Amacı ve Önemi

Bu tez çalışmasının temel amacı, mevcut makine öğrenmesi yöntemlerini kullanarak kripto para fiyatlarının tahmin edilmesine yönelik mevcut modelleri incelemek, değerlendirmek ve karşılaştırarak daha iyi sonuçlar vermeleri için bu modelleri geliştirmektir.

Kripto para piyasaları, yüksek volatilite ve belirsizliklerle karakterizedir ve geleneksel finansal araçlara kıyasla daha az öngörülebilir bir yapıya sahiptir. Bu nedenle, doğru tahmin yöntemlerine olan ihtiyaç artmaktadır. Makine öğrenmesi, büyük veri setlerini işleyebilme ve karmaşık örüntüleri tespit edebilme yetenekleriyle, kripto para fiyat tahminlerinde etkili bir araç olarak öne çıkmaktadır.

Bu çalışmada, çeşitli makine öğrenmesi teknikleri kullanılarak kripto para fiyatlarının tahmin edilebilirliği incelenmiş ve en başarılı modeller belirlenmiştir. Ayrıca, bu modellerin performansları farklı piyasa koşullarında ve on farklı kripto para birimleri üzerinde değerlendirilmiştir.

Bu tez çalışması, kripto para piyasalarının hızla büyümesi ve artan yatırımcı ilgisi nedeniyle önemlidir. Geleneksel finansal analiz yöntemleri, kripto para piyasalarının dinamik ve değişken yapısını tam olarak kavrayamamaktadır. Makine öğrenmesi yöntemlerinin kullanılması, kripto para tahmininde yeni bir perspektif sunmaktadır. Bu çalışma, yenilikçi bir yaklaşım, yatırımcılara rehberlik, akademik katkı, pratik uygulamalar ve veri bilimi ile finans alanlarının bütünleşmesi açısından önemlidir. Sonuç olarak, bu tez çalışması, kripto para tahmininin mümkün olup olmadığını araştırarak, hem akademik dünyaya hem de finans sektörüne değerli katkılar sağlamayı amaçlamaktadır.

1.2 Problem Tanımı

Kripto para piyasaları, son yıllarda hızla büyüyen ve önem kazanan bir finansal varlık sınıfı olmuştur. Ancak bu piyasalar, yüksek volatilite, düşük likidite ve manipülasyon riskleri gibi özelliklerle de öne çıkmaktadır. Geleneksel finansal araçlardan farklı olarak kripto paraların fiyat hareketleri, büyük ölçüde haberlerden, sosyal medya etkileşimlerinden ve spekülatif davranışlardan etkilenebilmektedir. Bu durum, kripto para piyasalarında fiyat tahmini yapmayı zorlaştıran faktörlerin başında gelmektedir.

Mevcut finansal piyasa analiz yöntemleri genellikle kripto para piyasalarının dinamik yapısını tam olarak kavrayamamakta ve bu piyasalarda doğru tahminlerde bulunmayı güçleştirmektedir. Dolayısıyla, yatırımcıların ve finansal analistlerin doğru kararlar alabilmesi için daha öngörülebilir ve etkili tahmin yöntemlerine olan ihtiyaç artmaktadır.

Bu bağlamda, makine öğrenmesi yöntemlerinin kullanımı, kripto para fiyatlarının tahmin edilmesinde potansiyel bir çözüm olarak öne çıkmaktadır. Makine öğrenmesi, büyük veri setlerini işleyebilme, karmaşık örüntüleri tespit edebilme ve zaman içindeki değişen faktörleri göz önünde bulundurma gibi yetenekleriyle, kripto para piyasalarındaki belirsizlikleri azaltmada ve daha doğru tahminler yapmada önemli bir rol oynayabilir.

Bu tez çalışması, mevcut makine öğrenmesi tekniklerini indikatörlerle birlikte kullanarak kripto para fiyat tahmini üzerine odaklanacak ve farklı modellerin performanslarını karşılaştırarak bu alandaki bilgi birikimini artırmayı amaçlamıştır. Ayrıca, geliştirilecek tahmin modellerinin pratik uygulamaları ve finansal karar vericiler için sağladığı potansiyel faydaları da değerlendirilmiştir.

1.3 Çalışmanın Kapsamı ve Sınırları

Bu çalışma, makine öğrenmesi yöntemlerinin kripto para piyasalarındaki fiyat hareketlerini tahmin etmede potansiyel bir araç olarak kullanılabilirliğini incelemektedir. Özellikle, ARIMA ve LSTM gibi zaman serisi modelleri ile Extreme Gradient Boosting (XGBoost), LightGBM ve CatBoost gibi gradient boosting algoritmaları üzerine odaklanmıştır.

Çalışmada, üç farklı zaman diliminde (15 dakika, 1 saat ve 4 saat) işlem gören on farklı kripto para birimi (XRPUSDT, TRXUSDT, NEOUSDT, LTCUSDT, ETHUSDT, DOGEUSDT, DASHUSDT, BTCUSDT, BNBUSDT ve ADAUSDT) üzerinde analizler yapılmıştır. Her bir kripto para birimi için, teknik analiz

indikatörleri kullanılarak zaman serisi verileri hazırlanmış ve beş farklı makine öğrenmesi algoritmasıyla fiyat tahminleri gerçekleştirilmiştir.

Performans değerlendirmesi için Root Mean Squared Error (RMSE) ölçütü kullanılmış ve her bir algoritmanın tahmin sonuçları karşılaştırılarak analiz edilmiştir. Bu değerlendirme, algoritmaların kripto para fiyatlarını tahmin etmedeki etkinliğini belirlemeye yönelik kapsamlı bir inceleme sunmaktadır.

1.4 Literatür Taraması

Kripto para tahmini, yüksek volatilite ve karmaşıklık sebebiyle finansal piyasalarda önemli bir araştırma alanıdır. Bu tarama, kripto para tahmininde kullanılan makine öğrenmesi yöntemlerine dair mevcut literatürü özetlemektedir.

Saad ve arkadaşlarının 2020 yılındaki çalışmasında, geniş bir kullanıcı tabanına, yaygın popülerliğe ve yüksek piyasa değerine sahip olan Bitcoin ve Ethereum kripto para birimlerini incelemek amacıyla yapılmıştır. Nisan 2016 ile Mayıs 2018 arasında Blockchain API'si aracılığıyla elde edilen veri seti, cüzdan sayısı, harcanmamış işlem çıktıları (UTXO'lar), mempool boyutu, blok boyutu, ortalama onay süresi, madencinin geliri, günlük işlem sayısı, blok başına işlem sayısı, benzersiz Bitcoin adresleri, toplam ağ hash oranı, ağ zorluğu, ücret, işlem başına ücret, sistematik toplam Bitcoinler, ticaret hacmi ve Bitcoin'in piyasa fiyatı gibi çeşitli özellikleri içermektedir. Fiyat tahmini için en kullanışlı özelliklerin belirlenmesi için, veri setindeki tüm özniteliklerin korelasyon matrisi hesaplanmıştır. Bu çalışmada, Doğrusal Regresyon (LR), Rastgele Orman (RF) ve Gradient Boosting algoritmaları kullanılarak bir analiz yapılmıştır. LR yöntemi, test verisinin %10'uyla en iyi performansı gösterirken, GRadient Boosting ve RF yöntemleri %5 test verisiyle en iyi sonuçları elde etmiştir [1].

Buyrukoğlu ve arkadaşları tarafından sunulan çalışmada, LSTM ve LSTM Ensemble modellerin performansı ve analizi incelenmiştir. LSTM Ensemble, AUDI, BTG ve TEL kripto para birimlerinde tekli LSTM'den daha iyi sonuçlar göstermiştir. Telcoin (TEL) için en iyi performansı gösteren model de LSTM Ensemble modelidir. Ancak, OKB (OKB) ve USDP kripto para birimlerinin analizinde tekli LSTM modeli daha iyi sonuçlar vermiştir [2].

Başka bir makalede, fiyat para birimleri ve kripto para birimleri arasındaki döviz kuru tahmininde VQA (Görsel ve Sözel Soru-Cevap) kavramını kullanarak ilişki ağlarını uyguladıkları belirtilmektedir. İlk olarak, fiyat özelliklerini çıkarmak için bir boyutlu bir CNN (Evrişimli Sinir Ağı) kullanılmıştır. Bu özellik haritaları, para

birimleri arasındaki ilişkileri keşfetmek için konvolüsyon işlemine tabi tutulmuştur. Ardından, soru anlama aşamasında, fiyat eğilimi ve volatilite gibi soruları belirlenip bu soruları modellemek için ikili dizilere dönüştürmüşlerdir. Cevap oluşturma bölümünde ise, fiyat özellik haritaları ve sorulara dayalı olarak bir ikili sınıflandırıcı kullanarak cevaplar üretilmiştir. Bu sınıflandırıcıyı eğitmek için ikili çapraz entropi kaybı fonksiyonunu kullanılmış olup performansı değerlendirmek için doğruluk oranını kullanmışlardır. Sonuç olarak ise kripto para birimlerinin tahmin performansının yaklaşık %65 doğruluk oranıyla fiyat para birimlerinden daha yüksek olduğunu göstermektedir[3].

En yaygın on iki kripto paranın fiyatlarının geçmiş fiyat bilgileri ve teknik verilerle ne kadar tahmin edilebilir olduğu bir başka çalışmada analiz edilmiştir. Bu analizde, destek vektör makineleri, lojistik regresyon, yapay sinir ağları ve rastgele ormanlar gibi sınıflandırma algoritmaları kullanılmıştır. Algoritmaların ortalama sınıflandırma doğruluğu, tüm kripto paralar ve zaman ölçekleri için sürekli olarak %50'nin üzerindedir. Günlük veya dakika düzeyindeki frekanslarda, destek vektör makineleri diğer algoritmalara kıyasla daha yüksek ve tutarlı tahmin doğruluğu sağlamaktadır [4].

Başka bir çalışmada, LSTM ve GRU gibi özel RNN yöntemleri kullanılarak günümüzde en yüksek piyasa değeri ve işlem hacmine sahip kripto paralar olan Bitcoin, Ethereum ve Ripple'ın 30 günlük fiyat tahmini yapılmıştır. Araştırmanın sonuçlarına göre, her iki modelde de en iyi tahmin sonucunu Bitcoin vermiştir. İkinci en iyi tahmin sonucu Ripple, üçüncü ise Ethereum için elde edilmiştir. Kullanılan yöntemler karşılaştırıldığında, MAPE performans ölçütüne göre en iyi tahmin sonucuna Bitcoin ve Ripple için GRU, Ethereum için ise LSTM modeli ile ulaşılmıştır [5].

Farah Muhammad Sakran Alamery (2023) tarafından gerçekleştirilen çalışmada Bitcoin fiyat tahmininde lineer regresyon, karar ağacı, rastgele orman, destek vektör regresyonu ve derin öğrenme modelleri (ANN ve CNN) gibi çeşitli makine öğrenmesi modellerini değerlendirmiştir. CNN modeli, derin öğrenme modelleri arasında en yüksek tahmin doğruluğunu sağlamış, rastgele orman modeli ise makine öğrenmesi modelleri arasında en iyi performansı göstermiştir [6].

H. Sebastião ve P. Godinho (2021), çeşitli makine öğrenmesi modellerini birleştirerek oluşturdukları "Ensemble 5" modelinin Ethereum ve Litecoin için yıllık Sharpe oranlarında %80.17 ve %91.35, yıllık getiri oranlarında ise sırasıyla %9.62 ve %5.73 olduğunu bulmuşlardır [7].

Dr. Alamelu Mangai Jothidurai ve arkadaşları 2023 yılındaki çalışmasında, Bitcoin

fiyat tahmininde LSTM modeli kullanmış ve sosyal medya verilerini kullanarak duygu analizi gerçekleştirmişlerdir. Çalışmaları, Bitcoin'i örnek olarak kullanarak kârı en üst düzeye çıkarmayı ve riski en aza indirmeyi amaçlayan, makine öğrenimi modellerini ve duyarlılık analizini kullanarak kripto para birimi fiyat tahmini için yeni bir yaklaşım önermektedir. Sonuçlar LSTM modelinin, kripto para fiyat tahmininde diğer modellere göre daha yüksek doğruluk sağladığını göstermiştir [8].

Xiaoke Song (2023), Bitcoin, Ethereum, Dogecoin ve Cardano gibi kripto paraların kapanış fiyatlarını tahmin etmek için XGBoost ve LightGBM gibi gradient boosting algoritmalarını kullanmıştır. Çalışma, LightGBM'nin tahmin doğruluğunda biraz daha iyi performans gösterdiğini bulmuştur [9].

Jie Wu ve arkadaşları (2022), XGBoost algoritmasını kullanarak 14 farklı kripto para piyasasının kısa vadeli getirilerini tahmin etmişlerdir. XGBoost algoritması, geleneksel makine öğrenmesi algoritmalarına kıyasla %12.5 ile %43.3 arasında daha yüksek tahmin performansı sergilemiştir [10].

Swati ve Mohan (2022) tarafından gerçekleştirilen araştırmada, AdaBoost, XGBoost ve CatBoost algoritmalarının Bitcoin ve Ripple'ın uzun vadeli tahminlerindeki performansları incelenmiştir. AdaBoost ve XGBoost, benzer tahmin etkinliği göstermiştir ve CatBoost'a kıyasla daha basit stratejilerle benzer sonuçlar elde edilmiştir [11].

V. Derbentsev ve arkadaşları (2021), Bitcoin, Ethereum ve Ripple gibi büyük kripto paraların kısa vadeli tahminleri için Random Forest ve Gradient Boosting Machine (GBM) yöntemlerini kullanmışlardır. GBM, kısa vadeli tahminlerde Random Forest'tan daha yüksek doğruluk sağlamıştır [12].

Tianyu Ray Li ve arkadaşları (2018), Twitter verilerini kullanarak ZClassic adlı bir kripto paranın fiyat dalgalanmalarını tahmin etmek için XGBoost algoritmasını kullanmışlardır. Model, test verileri ile %0.81 korelasyon göstermiştir [13].

Navmeen Latif ve arkadaşları (2023), Bitcoin fiyat tahmini için ARIMA ve LSTM modellerini karşılaştırmışlardır. İki farklı eğitim ve test örneği kullanılarak çapraz doğrulama yapılmıştır. İlk eğitim örneğinde ARIMA, LSTM'yi geride bırakırken, ikinci eğitim örneğinde LSTM, ARIMA'yı aşmıştır. Test örneklerinde ise LSTM, ARIMA'dan daha iyi performans göstermiştir [14].

Linxi Pan (2023), Bitcoin, Ethereum ve Dogecoin fiyat tahmini için ARIMA, Random Forest ve LSTM algoritmalarını karşılaştırmıştır. Sonuçlar, LSTM'nin en yüksek doğruluk oranına sahip olduğunu, ardından Random Forest ve ARIMA'nın

geldiğini göstermiştir [15].

Harshvardhan Sinha ve arkadaşları (2022), ARIMA ve LSTM modellerinin kombinasyonunu kullanarak Bitcoin fiyatlarını tahmin etmişlerdir. Hibrit model, tek başına kullanılan ARIMA ve LSTM modellerine kıyasla daha yüksek doğruluk sağlamıştır [16].

Ersin Sener ve I. Demir (2022), Bitcoin fiyatlarını tahmin etmek için Dalga Dönüşümü ile denoising yapılmış ARIMA ve LSTM modellerini kullanmışlardır. LSTM, RMSE metriğinde en iyi performansı göstermiştir [17].

Yiqing Hua (2020), ARIMA ve LSTM modellerini karşılaştırarak, LSTM'nin özellikle CPU üzerinden eğitim süresi uzun olsa da daha iyi performans gösterdiğini bulmuştur [18].

Svend Pasak ve Riyanto Jayadi (2023), Bitcoin, Ethereum, Binance Coin, Tether ve Cardano fiyatlarını tahmin etmek için ARIMA ve LSTM'yi karşılaştırmışlardır. LSTM, MAPE değeri %10'un altında kalarak ARIMA'dan daha iyi performans göstermiştir [19].

2
ÖNERİLEN YÖNTEM

Kripto para birimleri, belirli bir zaman aralığındaki verilerin sıralı bir şekilde temsil edildiği bir veri türüdür. Bu veri türü, zaman serisine dayalı modellerin kullanımını mantıklı kılmaktadır. Bu bağlamda, kripto para birimlerinin gelecekteki değerlerini tahmin etmek için ARIMA ve LSTM yöntemleri kullanılmıştır. Aynı zamanda finansal tahmin analizlerinde kullanılan gradiyent artırma (gradient boosting) yöntemleri, kripto para tahminin de kullanabileceği düşünülerek bu yöntemlerden olan XGBoost, LightGBM ve CatBoost modelleri üzerinde çalışmalar yapılmıştır.

2.1 Zaman Serisi Modelleri

2.1.1 AutoRegressive Integrated Moving Average - ARIMA

ARIMA (AutoRegressive Integrated Moving Average) modeli, zaman serisi verilerinin analizinde ve tahmin edilmesinde sıkça kullanılan bir istatistiksel modeldir. ARIMA modeli, otoregresyon (AR), farklanma (difference-I), ve hareketli ortalama (MA) bileşenlerini birleştirir. Otoregresyon bileşeni (AR), zaman serisinin kendi gecikmiş değerleriyle ilişkilendirilir. Farklanma bileşeni (I), zaman serisini durağan hale getirmek için uygulanır. Hareketli ortalama bileşeni (MA), modelin hata terimlerini geçmiş değerleriyle ilişkilendirir. Bir $ARIMA(p, d, q)$ modelinde p, d, ve q parametreleri sırasıyla otoregresyon, farklanma ve hareketli ortalama bileşenlerinin karmaşıklığını tanımlar.

ARIMA modelini kurarken, öncelikle zaman serisinin grafiksel analizi yapılır ve serinin durağan olup olmadığı incelenir. Durağanlık yoksa farklanma uygulanarak zaman serisi durağan hale getirilir. Ardından, modelin otoregresyon (p) ve hareketli ortalama (q) terimleri için en uygun parametre değerleri, otokorelasyon fonksiyonu (ACF) ve kısmi otokorelasyon fonksiyonu (PACF) grafikleri kullanılarak belirlenir. Son olarak, ARIMA modeli istatistiksel yazılımlar veya programlama dilleri kullanılarak kurulur ve modelin performansı çeşitli kriterlerle değerlendirilir. Belirlenen model üzerinden zaman serisinin gelecekteki değerleri tahmin edilir.

ARIMA Modelinin Bileşenleri

Otoregresyon (AR) Bileşeni:

- Bu bileşen, zaman serisinin kendi gecikmiş değerleriyle ilişkilendirildiği kısmıdır.
- p parametresi, modelin kullanacağı gecikmiş değerlerin sayısını ifade eder.
- Matematiksel olarak, $AR(p)$ modeli şu şekilde ifade edilir:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$$

  Burada Y_t zaman serisinin mevcut değeri, c sabit terim, ϕ_i otoregresyon katsayıları ve ϵ_t hata terimidir.

Farklanma (Integrated, I) Bileşeni:

- Bu bileşen, zaman serisini durağan hale getirmek için uygulanır.
- d parametresi, seriyi durağan hale getirmek için alınan farkların sayısını ifade eder.
- Matematiksel olarak, farklanmış seri şu şekilde ifade edilir:

$$Y'_t = Y_t - Y_{t-1}$$

  d kez farklama yapıldığında, farklanmış zaman serisi elde edilir.

Hareketli Ortalama (MA) Bileşeni:

- Bu bileşen, modelin hata terimlerinin geçmiş değerleriyle ilişkilendirildiği kısmıdır.
- q parametresi, modelin kullanacağı geçmiş hata terimlerinin sayısını ifade eder.
- Matematiksel olarak, $MA(q)$ modeli şu şekilde ifade edilir:

$$Y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

  Burada θ_i hareketli ortalama katsayılarıdır.

2.1.2 Long Short-Term Memory - LSTM

LSTM (Uzun Kısa Vadeli Bellek), derin öğrenme içinde kullanılan bir tür tekrarlayan sinir ağı yapısıdır. Özellikle zaman serileri ve dizisel verilerin (metin, ses gibi) işlenmesinde etkilidir. Standart RNN'lerin uzun vadeli bağımlılıkları öğrenmede zorlandığı sorunları çözer. Bu özelliği, özel hücre yapıları ve kapı mekanizmaları sayesinde sağlanır.

LSTM hücresi, üç ana kapıdan oluşur: unutma kapısı (forget gate), giriş kapısı (input gate) ve çıkış kapısı (output gate). Unutma kapısı, önceki hücre durumundan ne kadar bilginin saklanacağını belirler. Giriş kapısı, yeni bilgilerin hücre durumuna eklenmesini kontrol eder. Hücre durumu güncelleme adımında, önceki hücre durumu unutma ve yeni bilgilerle güncellenir. Çıkış kapısı ise hangi bilgilerin çıkış olarak kullanılacağını belirler.

LSTM'nin avantajları arasında uzun vadeli bağımlılıkları etkili bir şekilde öğrenebilmesi, kapı mekanizmaları ile hassas kontrol sağlaması ve çeşitli dizisel veri türlerini işleyebilmesi yer alır. Bu özellikleri sayesinde dil modelleme, makine çevirisi, konuşma tanıma gibi çeşitli uygulamalarda kullanılır.

Matematiksel olarak, LSTM'nin her adımı sigmoid ve tanh aktivasyonlarıyla tanımlanan denklemlerle özetlenebilir. Bu yapılar LSTM'nin karmaşık dizisel veri ilişkilerini etkili bir şekilde öğrenebilmesini sağlar.

Giriş Kapısı (Input Gate)

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2.1}$$

Unutma Kapısı (Forget Gate)

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2.2}$$

Çıkış Kapısı (Output Gate)

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2.3}$$

Hücre Durumu (Cell State)

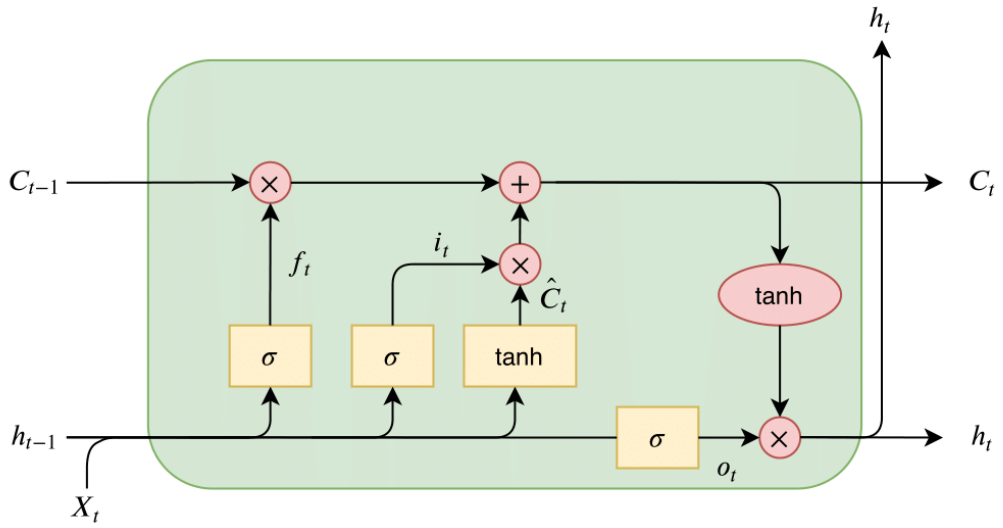
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{2.4}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{2.5}$$

Güncellenmiş Gizli Durum (Updated Hidden State)

$$h_t = o_t \cdot \tanh(C_t) \tag{2.6}$$

Burada, σ sigmoid fonksiyonunu, $\tanh$ ise hiperbolik tanjant fonksiyonunu temsil eder. W_i, W_f, W_o, W_C ağırlık matrisleri ve b_i, b_f, b_o, b_C bias terimleridir. x_t giriş vektörünü, h_t güncellenmiş gizli durumu, C_t hücre durumunu ve $\tilde{C}_t$ yeni hücre adayını temsil eder. Şekil 2.1 bu yöntemin model mimarisi verilmiştir.



Şekil 2.1 LSTM mimarisi [20]

2.2 Gradyan Artırma Yöntemleri

2.2.1 XGBoost

XGBoost (Extreme Gradient Boosting), makine öğreniminde sıklıkla kullanılan güçlü bir algoritmadır. Özellikle sınıflandırma ve regresyon problemlerinde yüksek performansı ve hızıyla dikkat çeker. XGBoost, gradyan artırma (gradient boosting) algoritmalarının bir uygulamasıdır, çeşitli optimizasyon teknikleriyle geliştirilmiştir ve üç temel ilkesi vardır.

Bunlardan ilki olan gradyan artırma, zayıf öğrenicilerin (genellikle karar ağaçları) ardışık olarak birleştirilmesiyle oluşturulan bir topluluk öğrenme (ensemble learning) tekniğidir. Her ağaç, bir önceki ağacın hatalarını düzeltmeye çalışır. Amaç, her yeni modelin hatanın gradyanını (türevini) takip ederek hatayı minimize etmesidir. Bir diğeri ise zayıf öğrencilerdir. XGBoost'ta kullanılan zayıf öğreniciler genellikle karar ağaçlarıdır. Her bir ağaç, bir önceki ağaçların hatalarını düzeltmek için eğitilir.

Ayrıca bu yöntemde birden fazla modelin bir araya getirilmesiyle daha güçlü ve

genel bir model oluşturulur. Bu teknik, bireysel modellerin hatalarını azaltarak genel performansı artırır. XGBoost, paralel işlemeyi destekler ve yüksek verimli hesaplamalar yapar. Bu özellik, özellikle büyük veri setlerinde model eğitimi süresini önemli ölçüde azaltır. Şekil 2.2 ile yaprak tabanlı ağaç büyümesi stratejisi görselleştirilmiştir.



Şekil 2.2 XGboost: Yaprak tabanlı ağaç büyüme stratejisi [9]

Aşırı uyum (overfitting) sorununu azaltmak için XGBoost, L1 (Lasso) ve L2 (Ridge) düzenleme yöntemlerini kullanır. Bu düzenleme terimleri, modelin karmaşıklığını kontrol ederek daha genel bir model elde edilmesini sağlar. XGBoost, ağaç yapılarını optimize etmek için çeşitli teknikler kullanır. Örneğin, greedy algoritması ve tree pruning gibi yöntemlerle en iyi bölünme noktalarını seçer ve gereksiz dalları budar. Ayrıca, XGBoost çeşitli hata fonksiyonları ile çalışabilir (örneğin, log-loss, mean squared error). Bu esneklik, farklı problem türlerine uyum sağlamasını sağlar.

Son olarak, XGBoost sınıf dengesizliği problemlerine karşı dayanıklıdır ve bu tür durumları ele almak için çeşitli parametreler sunar.

2.2.1.1 XGBoost'un Matematiksel Modeli

XGBoost'ta ilk adım, bir tahmin yapmaktır. Bu tahmin, daha sonraki adımlarda yapılan işlemlerle doğru sonuca ulaşacak şekilde iyileştirilir; genellikle varsayılan olarak 0.5 kullanılır.

Modelin doğruluğu, hatalı tahminler (artıklar) üzerinden değerlendirilir. Artıklar, gözlemlenen değer ile tahmin edilen değer arasındaki farklardır.

Veri setinde, tahmin edilmek istenen "median_income" değişkeni olarak belirlenmiştir. İlk tahmin "initial leaf" sütununda yer alırken, tahmin hatası "residual1" sütununda bulunur.

Gradient Boosting'te olduğu gibi, hataları tahmin eden bir karar ağacı oluşturulur. Bu aşamada amaç, hataları öğrenerek doğru tahmine yaklaşmaktır.

Her ağaç dalı için hesaplanan benzerlik skoru, verilerin ne kadar iyi gruplandığını gösterir.

$$\text{Benzerlik Skoru} = \frac{\begin{matrix}\text{Sol Dalın} \\ \text{Hata Karelerinin Toplamı}\end{matrix}}{\text{Sol Dalın Veri Sayısı} + \lambda} + \frac{\begin{matrix}\text{Sağ Dalın} \\ \text{Hata Karelerinin Toplamı}\end{matrix}}{\text{Sağ Dalın Veri Sayısı} + \lambda}$$

Benzerlik skorları hesaplandıktan sonra, daha iyi bir tahmin yapılıp yapılamayacağı değerlendirilir. Bu değerlendirme sırasında regülarizasyon parametresi kullanılır.

$$\mathrm{L}(y, \hat{y}) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Bu süreçte, her olası ağaç için benzerlik skorları hesaplanır ve hangi ağacın daha yüksek kazanç sağladığı belirlenir. Kazanç, sol ve sağ benzerlik skorlarının önceki ağaç benzerlik skorundan çıkarılmasıyla hesaplanır. En yüksek kazanç değerine sahip ağaç seçilir ve budama işlemi başlatılır. Budama için gamma adı verilen bir eşik değeri seçilir; bu değer, kazanç skoruna uygulanarak yalnızca yüksek kazançlı dalların ağaçta kalmasını sağlar ve aşırı öğrenme riskini azaltır. λ değeri varsayılan olarak 1'dir. Lambda'nın artmasıyla öğrenme zorlaşır ancak aşırı öğrenme riski azalır. Bu nedenle lambda, modelin hassasiyetini ayarlar.

Gradient Boosting'de olduğu gibi, tahmin yapılırken, ilk tahmin ve öğrenme oranı (learning rate) ile ilk ağaç kullanılarak ikinci tahmin yapılır. Bu süreç, hatalar düzeltilene ve belirlenen ağaç sayısına ulaşılıncaya kadar devam eder.

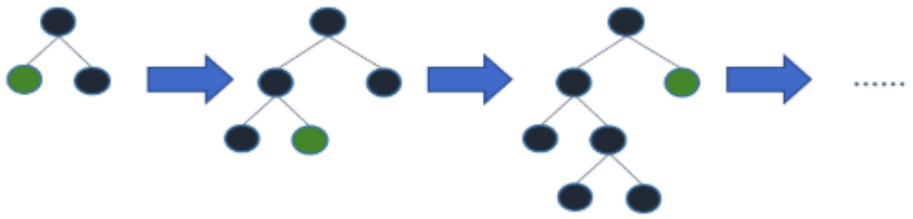
Düzenleme formülü:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

2.2.2 LightGBM

LightGBM (Light Gradient Boosting Machine), Microsoft tarafından geliştirilen ve gradyan artırma (gradient boosting) algoritmasını kullanan açık kaynaklı bir makine öğrenme kütüphanesidir. Özellikle büyük veri setleri ve yüksek boyutlu özelliklere sahip veri setleri üzerinde hızlı ve verimli çalışmasıyla bilinir. LightGBM, gradyan artırma çerçevesine çeşitli optimizasyonlar getirerek model eğitim süresini ve bellekteki ayak izini önemli ölçüde azaltır.

LightGBM, gradyan artırma karar ağaçları (GBDT - Gradient Boosting Decision Trees) algoritmasına dayalıdır. GBDT, zayıf öğrenicilerin (karar ağaçları) ardışık olarak birleştirildiği ve her ağacın bir önceki ağaçların hatalarını düzeltmeye çalıştığı bir topluluk öğrenme yöntemidir. LightGBM, bir dizi yenilikçi özellik sunarak makine öğrenimi modellerinin hızını ve verimliliğini artırır. İlk olarak, "Histogram Tabanlı Bölme" yöntemi sayesinde sürekli özellikleri hızla ayrık hale getirir ve histogramlar kullanarak bölme noktalarını efektif bir şekilde belirler. Bu özellik, eğitim süresini önemli ölçüde kısaltır. Ayrıca, bölme işlemi sırasında tüm özelliklerin olası bölme noktalarını değerlendirerek en yüksek bilgi kazanımına sahip olanı seçmesiyle "Bölme Kazanımı" sağlar. LightGBM, büyük ölçekli veri setleriyle çalışırken de etkinliğini korur. "Veri Küçültme" stratejisi, her adımda tüm veri yerine bir alt küme kullanarak hesaplamaları hızlandırır. Ayrıca, yüksek boyutlu özelliklere sahip veri setlerinde de verimli çalışabilir. Seviye-tabanlı Ağaç Büyüme yaklaşımıyla, LightGBM ağaç yapılarını derinleştirirken her seviyede tüm yaprakları genişletmek yerine, en fazla bilgi kazanımına sahip olan yaprağı genişleterek daha derin ve karmaşık modeller oluşturur. Bu yöntem doğruluğu artırırken overfitting riskini de beraberinde getirebilir, bu yüzden düzenleme teknikleri önem kazanır. Bu yöntem Şekil 2.3 ile gösterilmiştir.



Şekil 2.3 LightGBM: Seviye tabanlı ağaç büyüme stratejisi [9]

LightGBM aynı zamanda "Gelişmiş Düzenleme Teknikleri" sunar. L1 ve L2 düzenlemesi kullanarak model karmaşıklığını kontrol eder ve her iterasyonda bazı özellikleri rastgele seçerek overfitting riskini düşüren "Dropout" yöntemini uygular.

Son olarak, LightGBM paralel ve dağıtık hesaplamayı destekler, bu da büyük veri setleri üzerinde hızlı model eğitimi sağlar. Bu özellikler, LightGBM'in geniş veri setleriyle ve yüksek performans gerektiren uygulamalarla etkin bir şekilde başa çıkmasını sağlar..

2.2.2.1 LightGBM'nin Matematiksel Modeli

LightGBM, her iterasyonda yeni bir karar ağacı ekleyerek kayıp fonksiyonunu minimize etmeye çalışır. Genel bir LightGBM modelinin matematiksel ifadesi aşağıdaki gibidir:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i)$$

Burada:

- $\hat{y}_i$, tahmin edilen değerdir.
- K, ağaç sayısını ifade eder.
- f_k, k-inci karar ağacıdır.
- x_i, i-inci örneğin özellikleridir.

2.2.2.2 Kayıp Fonksiyonu ve Düzenleme

Kayıp fonksiyonu, modelin hatasını ölçen bir fonksiyondur. LightGBM'de kayıp fonksiyonu iki bileşenden oluşur: eğitim hatası (training loss) ve düzenleme terimi (regularization term).

$$\mathrm{L}(y, \hat{y}) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Burada:

- l, eğitim hatasını ölçen bir fonksiyondur (örneğin, log-loss veya mean squared error).
- Ω, düzenleme terimidir ve şu şekilde tanımlanır:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

Burada:

- T, ağacın yaprak sayısıdır.
- w_j, j-inci yaprağın ağırlığıdır.
- γ ve λ, düzenleme parametreleridir.

2.2.2.3 LightGBM'nin Eğitimi

LightGBM'nin eğitimi, her adımda yeni bir karar ağacı ekleyerek kayıp fonksiyonunu minimize etmeye dayanır. Bu süreç aşağıdaki adımlarla özetlenebilir:

1. Başlangıç Modeli: İlk model, genellikle tüm örneklerin ortalaması gibi basit bir tahminle başlar.

2. Ağaç Eklenmesi: Her adımda, mevcut hataları minimize etmeye çalışan yeni bir ağaç eklenir. Yeni ağacın her yaprağı için optimal ağırlıklar hesaplanır.

3. Güncelleme: Mevcut tahminler, yeni ağacın katkısıyla güncellenir. Bu işlem, belirli bir durma kriterine (örneğin, maksimum ağaç sayısı veya belirli bir hata eşiği) ulaşılana kadar tekrarlanır.

2.2.3 CatBoost

CatBoost (Categorical Boosting), Yandex tarafından geliştirilen ve özellikle kategorik verilerle çalışırken üstün performans gösteren bir gradyan artırma (gradient boosting) algoritmasıdır. CatBoost, hem kategorik hem de sayısal verilerle etkili bir şekilde çalışabilen güçlü ve esnek bir makine öğrenme kütüphanesidir. CatBoost, temel olarak gradyan artırma karar ağaçları (GBDT - Gradient Boosting Decision Trees) algoritmasına dayanır. GBDT, zayıf öğrenicilerin (karar ağaçları) ardışık olarak birleştirildiği ve her ağacın bir önceki ağaçların hatalarını düzeltmeye çalıştığı bir topluluk öğrenme yöntemidir.

CatBoost, veri hazırlığı sürecini önemli ölçüde kısaltarak dikkat çeker. Boş verilerle başa çıkabilir ve kategorik verilere özgü bir kodlama yöntemi uygular. CatBoost'un kategorik verilerle yüksek performans göstermesinin arkasındaki neden, kendine özgü bir kodlama metoduna sahip olmasıdır. Bu özellik sayesinde veri hazırlığı sırasında ek kodlama işlemine ihtiyaç duyulmaz, hatta kodlamanın yapılmaması önerilir. Bu durum öğrenme sürecinin hızını artırırken sonuçların kalitesini de etkiler. Ayrıca CatBoost, simetrik ağaçlar kurarak derin ağaçlar oluşturmadan yüksek tahmin doğruluğu sağlar ve aşırı öğrenme sorununu minimize eder. Aşırı öğrenme durumunda, CatBoost belirtilen parametreler (örneğin belirtilen ağaç sayısı) ulaşılmadan önce öğrenmeyi durdurur. Bu seçenek başlangıç parametrelerinden ayarlanabilir. Aşırı öğrenme riskini minimize etmek için CatBoost, benzersiz düzenleme yöntemleri kullanır ve "Ordered Boosting" ile "Random Permutations" gibi tekniklerle modelin genelleme yeteneğini güçlendirir.

CatBoost, GPU kullanarak öğrenme işlemini hızlandırabilir. Bu özellik sayesinde öğrenme süresi önemli ölçüde kısalır. GPU üzerinde çalışabilmesi için ise NVIDIA

Driver version 418 veya daha yeni bir ekran kartına ihtiyaç duyulur.

Öğrenme sırasında CatBoost, bellek kopyaları alır. Beklenmedik durumlar (örneğin bilgisayarın aniden kapanması) yaşandığında öğrenme süreci kaldığı yerden devam edebilir. Bellek kopyası alma durumu parametrelerle kontrol edilebilir ve büyük veri setleriyle çalışırken önemli bir özelliktir.

Kaydedilen modelin tekrar başlaması için, çalışmanın aynı dosyada ve aynı parametrelerle yeniden başlatılması gereklidir. Eksik veri işleme konusunda CatBoost, eksik verileri etkili bir şekilde işleyebilir ve tahminlerde kullanabilir. Ayrıca eksik veriler için özel işleme yöntemleri sunar, bu da modelin veri eksikliği durumlarında dahi güvenilir tahminler yapabilmesini sağlar.

Son olarak, CatBoost çeşitli veri kümeleri ve problemler üzerinde yüksek performans gösterir. Hem sınıflandırma hem de regresyon problemlerinde üstün başarı sağlamasıyla bilinir ve geniş uygulama alanlarında güvenilir sonuçlar sunar.

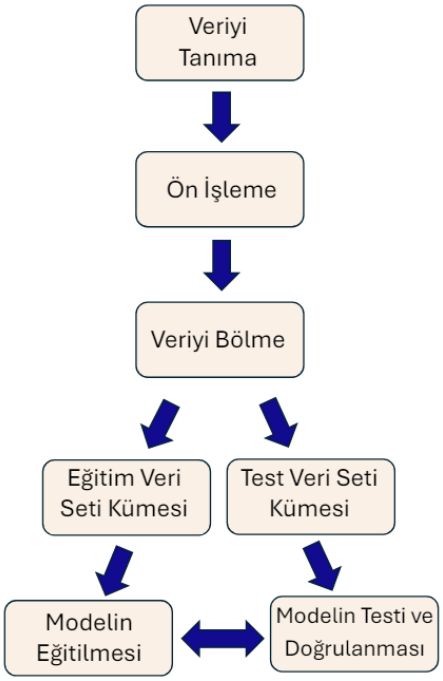
Tablo 2.1 LightGBM, XGBoost ve CatBoost Modellerinin Karşılaştırılması

Özellikler / Modeller	**LightGBM**	**XGBoost**	**CatBoost**
Kategori Değişken Desteği	Bazı sınırlı destek	Hayır	Evet
Dal Stratejisi	Yaprak-Tabanlı	Seviye-Tabanlı	Düzey Duyarlı Daldırma
Hız	Hızlı eğitim	Daha yavaş eğitim	Hızlı eğitim
Performans	Büyük veri setlerinde etkili	Yüksek performans, geniş kullanım	Rekabetçi performans, geniş kullanım
Geniş Topluluk Desteği	Evet	Evet	Evet
Veri Ön İşleme	Histogram tabanlı eğitim	Histogram tabanlı eğitim	Daha az veri ön işlem gerektirir
Rekabetçi Performans	Evet	Evet	Evet

Tablo 2.1, LightGBM, XGBoost ve CatBoost'un bazı önemli özelliklerini karşılaştırarak farklarını ortaya koyar. Her bir modelin kendine özgü avantajları ve kullanım senaryoları vardır; bu nedenle hangi modelin tercih edileceği, spesifik problemlere ve veri seti özelliklerine bağlı olarak değişebilir.

3
MODEL GELİŞTİRME VE EĞİTİM

Model geliştirme ve eğitim süreci, kripto fiyat tahminlemesi üzerinde etkili olan ARIMA, LSTM, XGBoost, LightGBM ve CatBoost gibi beş farklı makine öğrenimi modelini içerir. İlk olarak, veri seti %80 eğitim ve %20 test olarak bölündü. Eğitim için seçilen özellikler arasında yıl, ay, gün, saat, işlem hacmi ve çeşitli teknik göstergeler yer aldı. Her bir model, belirli parametrelerle özelleştirilerek eğitim veri seti üzerinde öğretildi. Modellerin doğruluğunu değerlendirmek için 5-katlı cross-validation kullanıldı ve her model için elde edilen RMSE değerleri hesaplandı. Bu değerler, modellerin performansını objektif bir şekilde karşılaştırmak için kullanıldı. Son olarak, eğitim sırasında elde edilen RMSE değerleri grafikle görselleştirilerek modellerin tahmin yetenekleri görsel olarak sunuldu. Bu süreç, kripto piyasalarındaki fiyat tahminlemesi için en uygun modelin belirlenmesinde önemli bir adım sağlar.



Şekil 3.1 İşlem adımları

3.1 Veri Toplama ve Hazırlama

Oradata firmasından tahsis edilen ve 3 yıllık bir dönemi kapsayan veri seti, çeşitli kripto para birimlerinin ticaret verilerini ve ilgili finansal göstergeleri içermektedir.

3.1.1 Veri Seti Özellikleri ve İçeriği

Bu veri seti, kripto para birimlerinin işlem zamanlarına, işlem aralıklarına ve çeşitli teknik göstergelere dayalı olarak detaylı bir analiz imkanı sunmaktadır.

3.1.1.1 Temel Özellikler

- **DATE_TIME (Tarih ve Saat Bilgisi)**: Her bir ticaret işleminin gerçekleştiği tarih ve saat bilgisini içerir.
- **INTERVAL_NAME (İşlem Aralığı Adı)**: İlgili ticaretin zaman aralığını tanımlar. Veriler, 15 dakikalık, 1 saatlik ve 4 saatlik zaman dilimlerine ayrılmıştır.
- **OPEN_VAL (Açılış Değeri)**: Belirli bir zaman diliminde bir kripto para biriminin açılış fiyatını gösterir.
- **HIGH_VAL (En Yüksek Değer)**: Belirli bir zaman diliminde bir kripto para biriminin ulaşabildiği en yüksek fiyatı gösterir.
- **LOW_VAL (En Düşük Değer)**: Belirli bir zaman diliminde bir kripto para biriminin ulaştığı en düşük fiyatı gösterir.
- **CLOSE_VAL (Kapanış Değeri)**: Belirli bir zaman diliminde bir kripto para biriminin kapanış fiyatını gösterir.
- **VOLUME_VAL (İşlem Hacmi)**: Belirli bir zaman diliminde alınıp satılan kripto para birimlerinin toplam miktarını ifade eder.

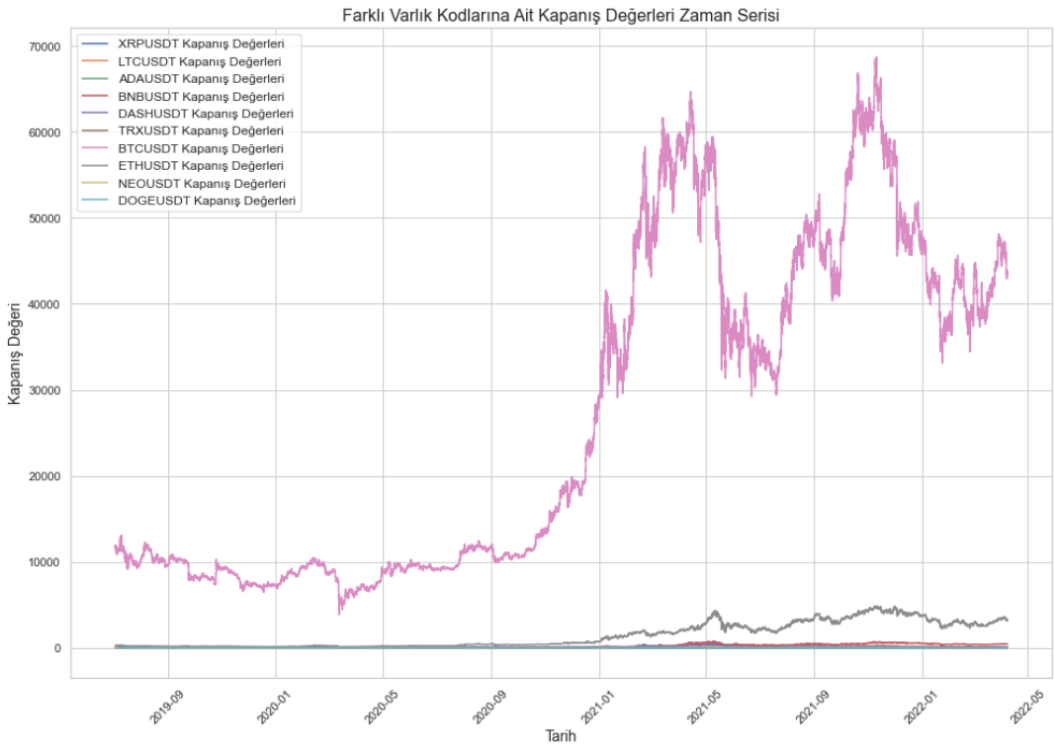
3.1.1.2 Teknik Göstergeler

- **SMA_50 (50 Günlük Basit Hareketli Ortalama)**: Son 50 günün kapanış fiyatlarının ortalamasını alarak hesaplanır ve kısa vadeli fiyat trendlerini belirlemede kullanılır.
- **SMA_200 (200 Günlük Basit Hareketli Ortalama)**: Son 200 günün kapanış fiyatlarının ortalamasını alarak hesaplanır ve uzun vadeli fiyat trendlerini belirlemede kullanılır.

- **RSI (Göreceli Güç Endeksi)**: 0 ile 100 arasında değer alır ve varlığın aşırı alım veya aşırı satım koşullarında olup olmadığını belirler. Genellikle 70 üzeri aşırı alım, 30 altı aşırı satım olarak değerlendirilir.
- **MACD (Hareketli Ortalama Yakınsama Iraksama)**: Kısa vadeli (genellikle 12 günlük) EMA (Üstel Hareketli Ortalama) ile uzun vadeli (genellikle 26 günlük) EMA arasındaki farkı gösterir.
- **MACD_SIG (MACD Sinyal Çizgisi)**: Genellikle 9 günlük EMA kullanılarak hesaplanır ve MACD'nin daha yumuşatılmış bir versiyonudur.
- **MACD_HIST (MACD Histogramı)**: MACD ile MACD sinyal hattı arasındaki farkı gösterir. Pozitif değerler yükseliş, negatif değerler düşüş sinyali olabilir.
- **VTXP_14 (Pozitif Vortex Göstergesi 14)**: Son 14 gün için hesaplanan pozitif trend gücünü gösterir.
- **VTXM_14 (Negatif Vortex Göstergesi 14)**: Son 14 gün için hesaplanan negatif trend gücünü gösterir.

3.1.2 Kripto Para Birimleri ve Zaman Dilimleri

Veri seti, LTCUSDT, ADAUSDT, BNBUSDT, DASHUSDT, TRXUSDT, BTCUSDT, ETHUSDT, NEOUSDT, XRPUSDT ve DOGEUSDT olmak üzere 10 farklı kripto para birimini kapsamaktadır. Veriler, 15 dakikalık (966,408 adet satır), 1 saatlik (242,580 adet satır) ve 4 saatlik (60,694 adet satır) zaman dilimlerinde gruplandırılmıştır. Şekil 3.2 de on farklı kripto paranın, bir saatlik kapanış değerleri verilmiştir.



Şekil 3.2 Bir saatlik kripto para birimlerinin tarihe göre kapanış değerleri

3.1.3 Analiz ve Uygulamalar

Bu geniş veri seti, kripto para birimlerinin ticaretine dair kapsamlı bilgi sunmakta ve çeşitli analiz yöntemleri ve modellerin geliştirilmesi için bir temel oluşturmaktadır. Özellikle, teknik göstergeler aracılığıyla trend analizi, aşırı alım/aşırı satım koşullarının tespiti ve hareketli ortalama geçişleri gibi stratejiler geliştirilebilir. Bu çalışmadan uygulanan stratejiler aşağıda verilmiştir.

- **SMA_Crossover**: Kısa vadeli SMA'nın (örneğin, 50 günlük) uzun vadeli SMA'yı (örneğin, 200 günlük) kesip geçmesi durumunda "altın haç" (bullish) veya "ölüm haçı" (bearish) gibi sinyaller analiz edilebilir.
- **Price_Diff**: Bir varlığın kapanış fiyatı ile bir önceki kapanış fiyatı arasındaki farkı ifade eder. Fiyatların yüzde değişimini gösterecek şekilde eklenmiştir.

3.1.4 Etik Kurallar

Veri seti, Oradata firmasından etik kurallara uygun bir şekilde tahsis edilmiştir. Bu veri setinin analizi ve kullanımı sırasında, tüm araştırma ve uygulama süreçlerinde etik kurallara uyulması büyük önem taşımaktadır. Bu veri seti, kripto para piyasalarının hareketlerini analiz etmek ve gelecekteki hareketleri tahmin etmek

için çok etkili bir kaynaktır. Sunulan veriler ve teknik göstergeler, yatırım stratejileri oluşturma ve risk yönetimi konusunda büyük fayda sağlamaktadır. Bu bağlamda, veri setinin doğru ve etkili bir şekilde analiz edilmesi, kripto para ticaretinin başarılı bir şekilde yönetilmesine katkı sağlayacaktır.

3.2 Veri Ön İşleme Teknikleri

Veri seti üzerinde yapılan ön işleme adımları şu şekildedir: İlk olarak, boş satır ve sütunlar temizlenmiştir. Ardından, tarih bazında sıralama yaparak veri seti düzenlenmiştir. Bu düzenleme, zaman serisi analizi yapabilmek ve verilerin doğru bir şekilde sıralanmış olmasını sağlamak amacıyla gerçekleştirilmiştir.

Özellik mühendisliği aşamasında, fiyat verilerinden yeni özellikler türetme işlemi gerçekleştirilerek iki önemli özellik eklenmiştir. İlk olarak, fiyatların yüzdelik değişimini gösteren `Price_Diff` özelliği oluşturulmuştur. Diğer bir özellik ise 50 günlük basit hareketli ortalama ile 200 günlük basit hareketli ortalama arasındaki farkı belirten `SMA_Crossover` özelliğidir. Bu özellik, hareketli ortalama crossover'larını temsil etmektedir.

Her bir kripto para birimi, zaman dilimine ve türlerine göre ayrılmış klasörlere konulmuştur. Bu, her kripto paranın zaman içindeki değişimlerini daha ayrıntılı bir şekilde inceleme ve analiz etme imkanı sağlamaktadır. Bu özelleştirilmiş klasörleme, kripto paraların değerlendirmesini ve karşılaştırılmasını kolaylaştırmaktadır.

Her bir kripto para biriminin kapanış fiyatı tahmini için model oluşturulurken, veri setinin etkili bir şekilde eğitilmesi ve doğrulanması amacıyla %80 eğitim ve %20 test ayrımı yapılmıştır. Bu ayrım, her bir kripto paranın geçmiş fiyat hareketlerini anlamak, modeli eğitmek ve ardından gelecekteki kapanış fiyatlarını doğrulamak için kullanışlı bir yaklaşım sağlamaktadır.

4
DENEYSEL SONUÇLAR

Kripto para birimi fiyat tahminleri, karmaşık ve dinamik bir konu olduğu için tek bir modelin tüm durumlar için en iyi performansı sağlaması zordur. Ancak, bu çalışma kapsamında yapılan analizlere göre, farklı zaman aralıklarında ve farklı kripto para birimleri için kullanılan ARIMA, LSTM, XGBoost, LightGBM ve CatBoost modellerinin performansı değerlendirilmiştir. Değerlendirmelere göre en iyi sonuçları CatBoost'un elde ettiği, onu takip eden sırasıyla LightGBM ve XGBoost yöntemleri olmuştur.

4.1 Zaman Dilimine Göre Değerlendirme

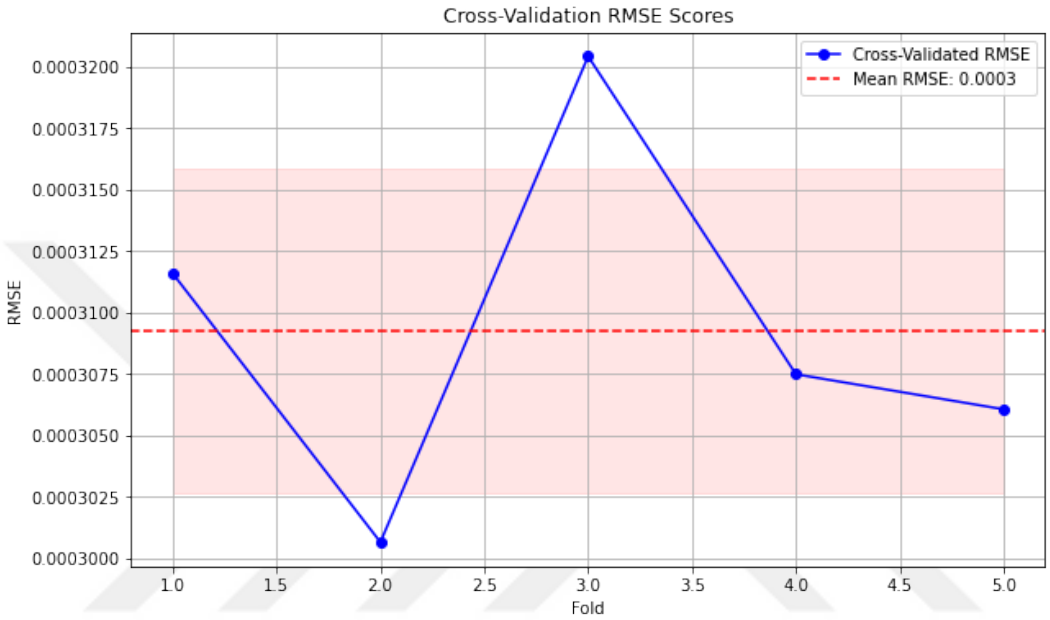
Zaman dilimi açısından yapılan gözlemlerde, üç farklı zaman dilimi arasında en iyi performansı 15 dakikalık veriler göstermiştir. Kısa zaman dilimlerinde fiyat hareketleri genellikle daha belirgin trendler sergilediği için, bu trendlerin modellere daha doğru tahminler yapmasına olanak tanımış olabilir. Ayrıca kısa zaman dilimlerindeki veriler, hızlı fiyat değişimlerini daha iyi yakalayabilir ve bu da modellerin doğruluğunu artırabilir. Ek olarak, daha fazla veri içermesi de daha iyi sonuçlar doğurmuş olabilir.

15 dakikalık veriler daha iyi sonuç vermesine karşın performans açısından zayıf kalmıştır. Özellikle LSTM veri çokluğundan dolayı fazla vakit almıştır ve hesaplanamamıştır, bu durum Tablo 4.1 ile gösterilmiştir.

4.2 Kapanış Değerlerine Göre Değerlendirme

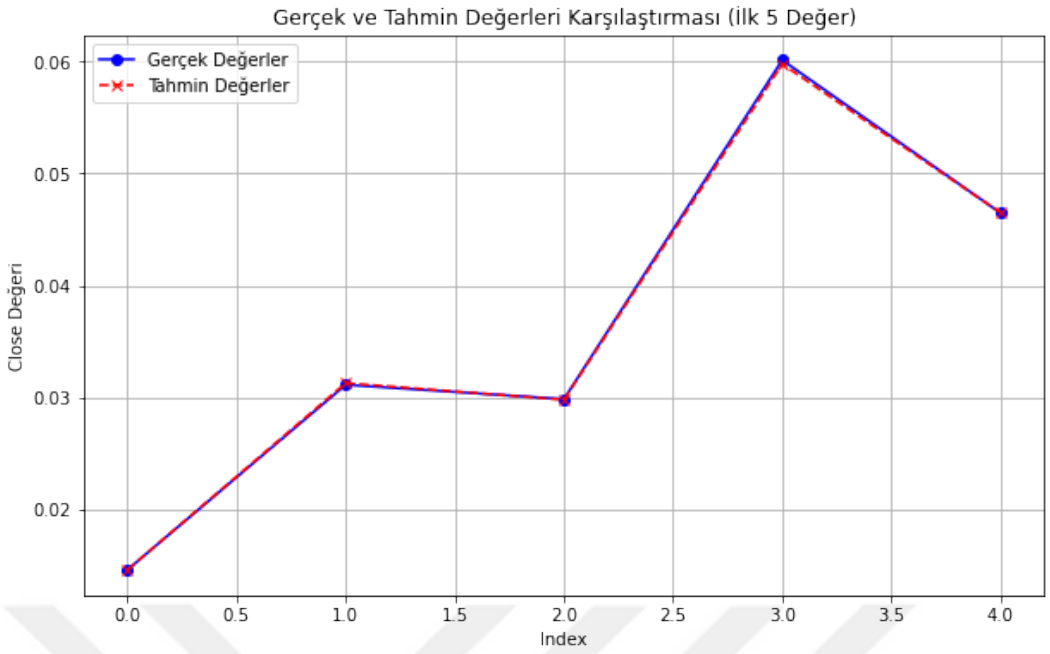
Düşük kapanış değerlerine sahip sırasıyla TRXUSDT, DOGEUSDT, XRPUSDT ve ADAUSDT gibi kripto para birimlerinin tahminlerinin daha doğru sonuçlar verdiği gözlemlenmiştir. Bu gözlem, düşük değerlere sahip olan kripto birimlerinin daha istikrarlı bir şekilde tahmin edilebildiğini düşündürmektedir. Diğer yandan, yüksek kapanış değerlerine sahip BTCUSDT ve ETHUSDT gibi kripto birimlerinin,

yüksek hata potansiyeline sahip olduğu saptanmıştır. Bu durum, yüksek değerlere sahip olmanın tahmin etmesi zorlaşan ve yanlış tahminlerin büyük hatalara sebep olabileceği bir ortamın oluşturduğunu göstermektedir. Yüksek değerlere sahip olan kripto birimlerinin piyasadaki diğer birimlere göre daha fazla dikkat gerektirdiğini ve bu birimlerin fiyat hareketlerinin daha karmaşık olduğu sonucuna varılmıştır. Model seçimi ve parametre ayarlamaları, özellikle yüksek değerli birimler için daha özenli bir şekilde gerçekleştirilmelidir.



Şekil 4.1 15 Dakika-TRXUSDT-CATBOOST-RMSE

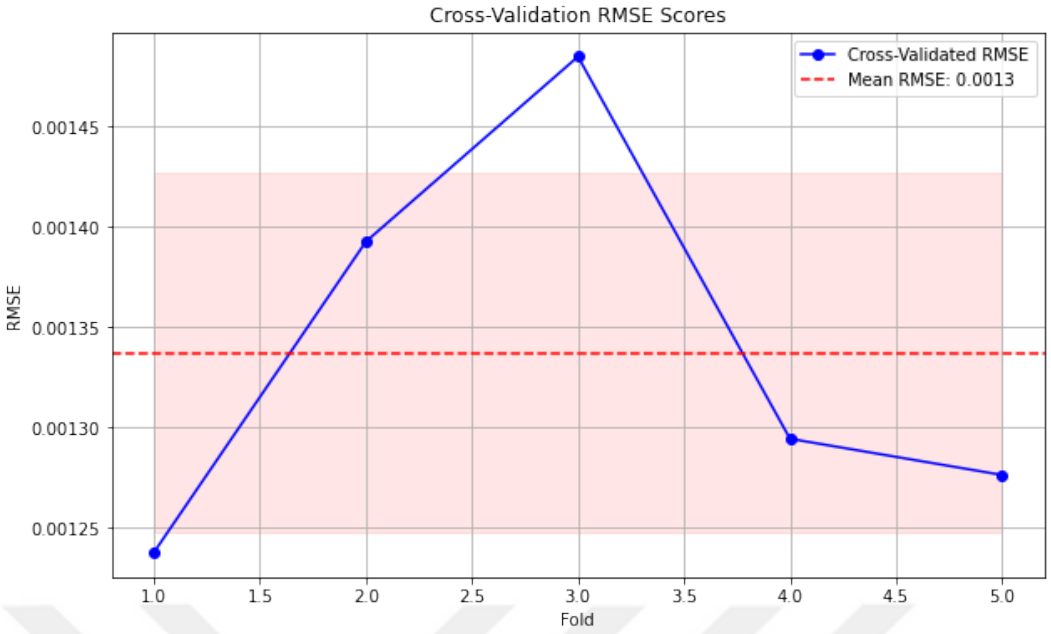
Şekil 4.1'de TRXUSDT'nin CatBoost modelindeki 15 dakikalık RMSE değerleri gösterilmiştir. Bu değerler modelin doğruluğunu değerlendirmek için önemli bir gösterge olup, TRXUSDT'nin diğer kripto paralara göre daha düşük RMSE değerlerine sahip olduğunu göstermektedir.



Şekil 4.2 15 Dakika-TRXUSDT-CATBOOST-TEST

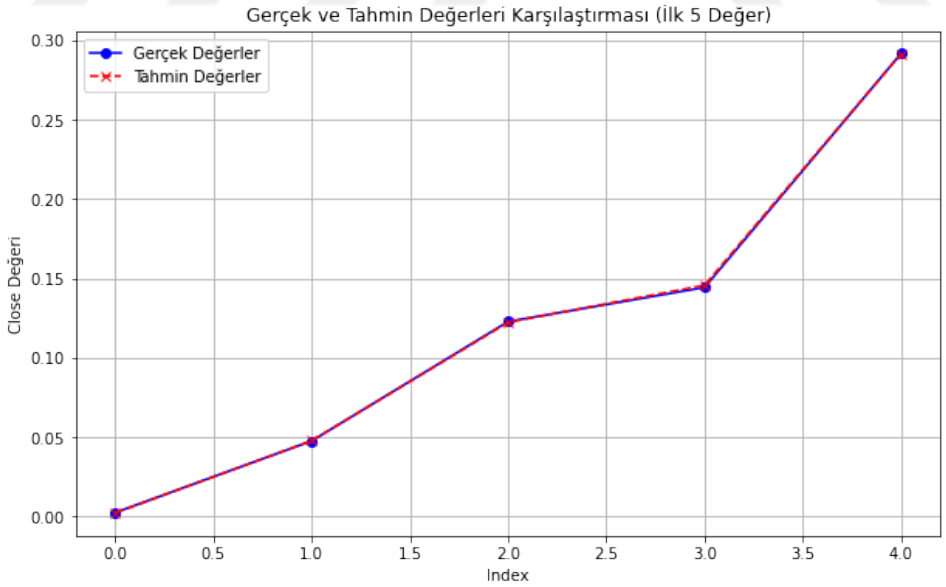
	Gerçek Değerler	Tahmin Değerler
0	0.01462	0.014608
1	0.03114	0.031292
2	0.02986	0.029798
3	0.06014	0.059794
4	0.04649	0.046509

Şekil 4.2'te TRXUSDT'nin 5 örnek üzerinde gerçek değerler ve tahmin değerleri gösterilmiştir. Yakından baktığımızda diğer para birimlerine kıyasla en düşük kapanış değerlerine sahip olduğunu görmekteyiz.



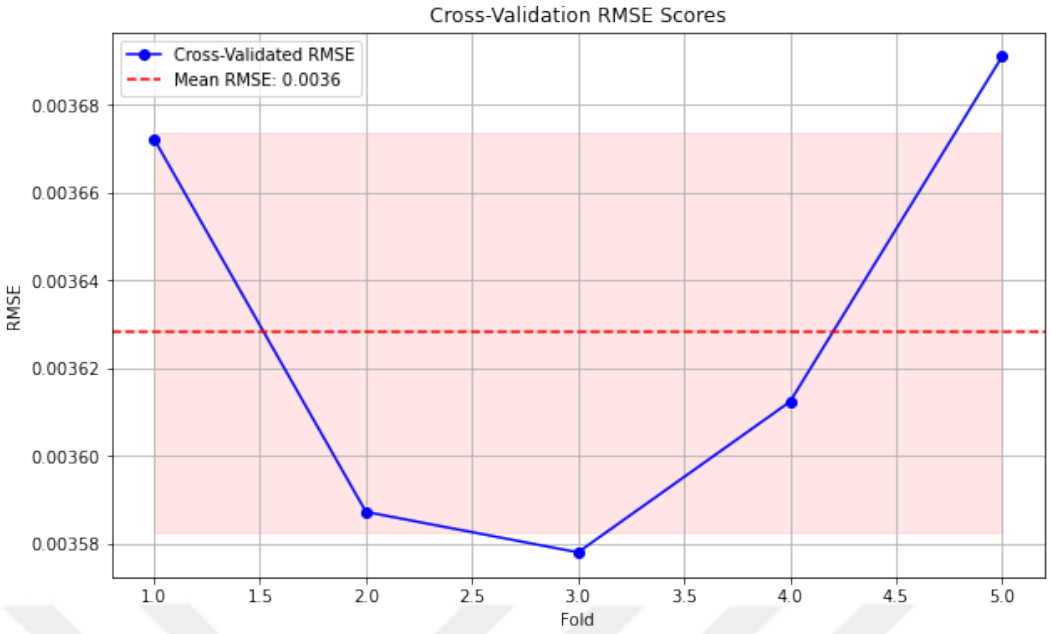
Şekil 4.3 15 Dakika-DOGEUSDT-CATBOOST-RMSE

Şekil 4.3'de DOGEUSDT'nin CatBoost modelindeki 15 dakikalık RMSE değerleri gösterilmiştir. Düşük kapanış değerlerine sahip olması nedeniyle DOGEUSDT'nin tahminleri de TRXUSDT gibi daha düşük RMSE değerlerine sahiptir.



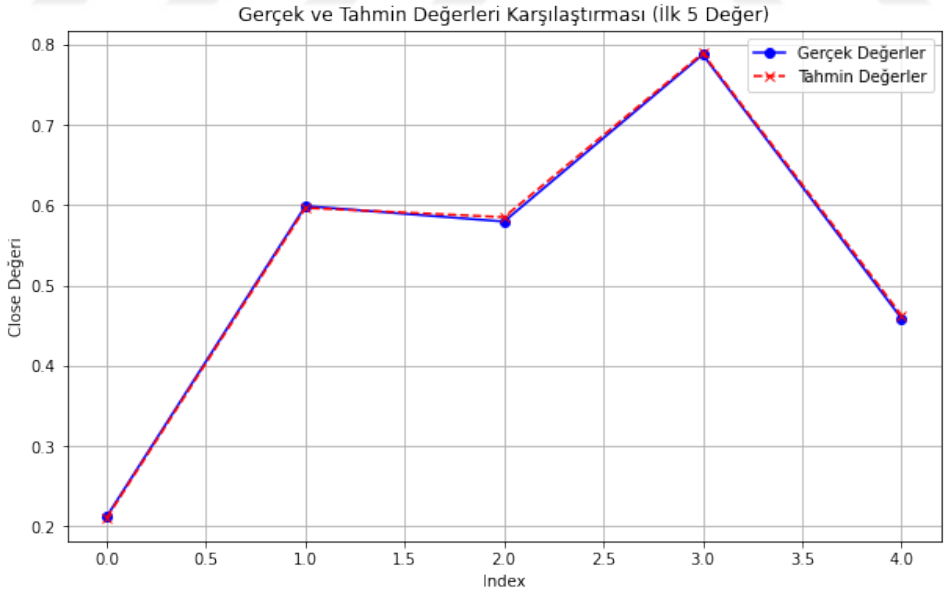
Şekil 4.4 15 Dakika-DOGEUSDT-CATBOOST-TEST

Şekil 4.4'te DOGEUSDT'nin CatBoost modelindeki test sonuçları gösterilmiştir. DOGEUSDT'nin düşük kapanış değerleri, tahminlerde daha az hata ile sonuçlanmaktadır.



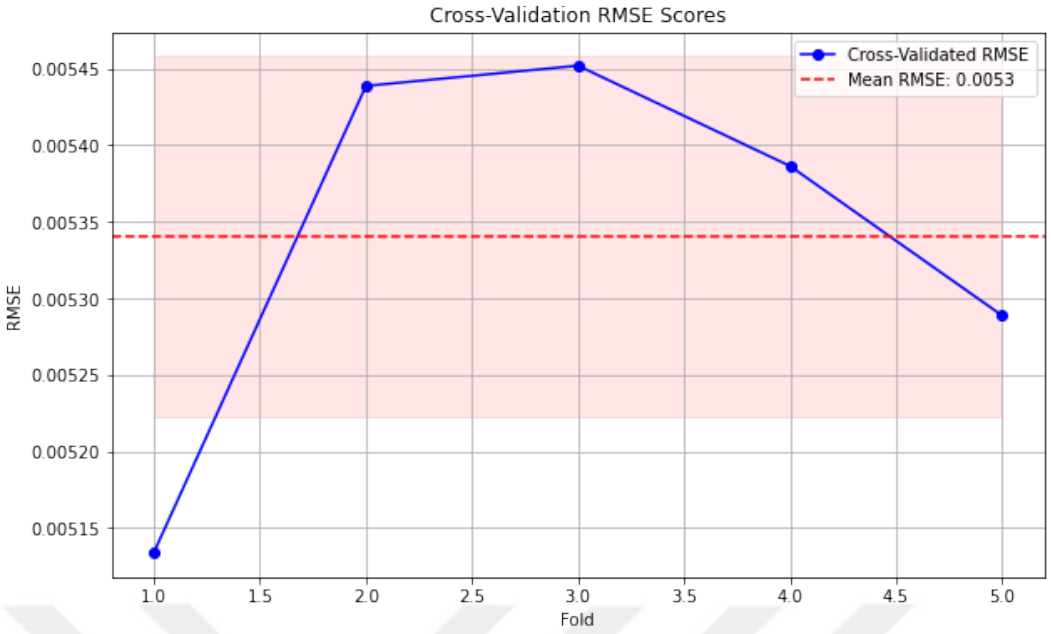
Şekil 4.5 15 Dakika-XRPUSDT-CATBOOST-RMSE

Şekil 4.5'de XRPUSDT'nin CatBoost modelindeki 15 dakikalık zaman serisi için 5 farklı RMSE değerleri gösterilmiştir. XRPUSDT'nin de düşük kapanış değerlerine sahip olduğu ve bu nedenle daha doğru tahminler sağladığı görülmektedir.



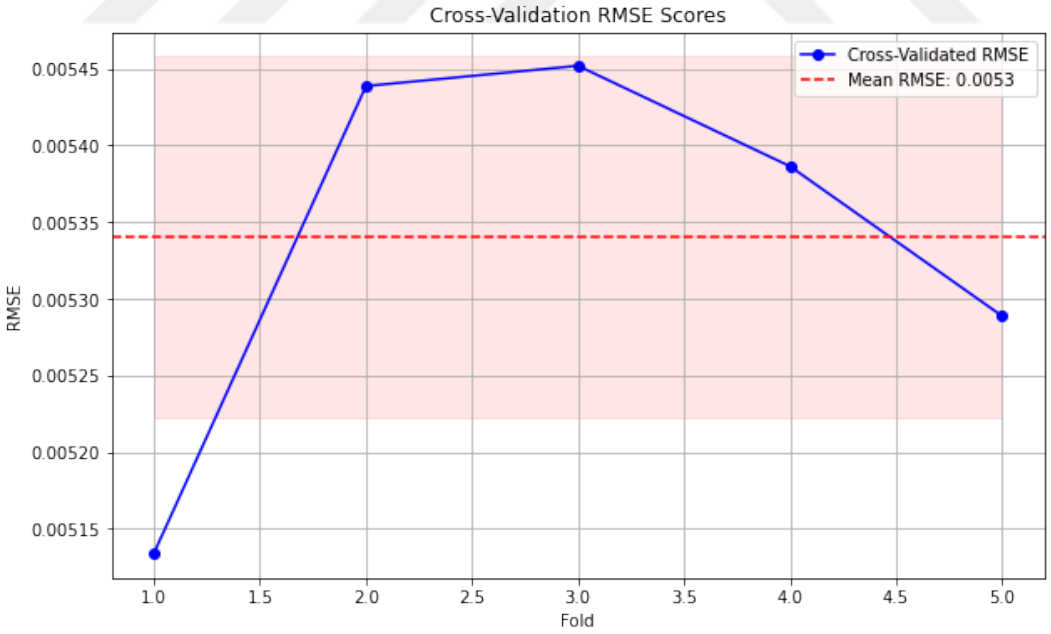
Şekil 4.6 15 Dakika-XRPUSDT-CATBOOST-TEST

Şekil 4.6'te XRPUSDT'nin CatBoost modelindeki test sonuçları gösterilmiştir. Gerçek değerler ve tahmin değerler arasındaki uyum, düşük kapanış değerlerinin getirdiği avantajı yansıtmaktadır.



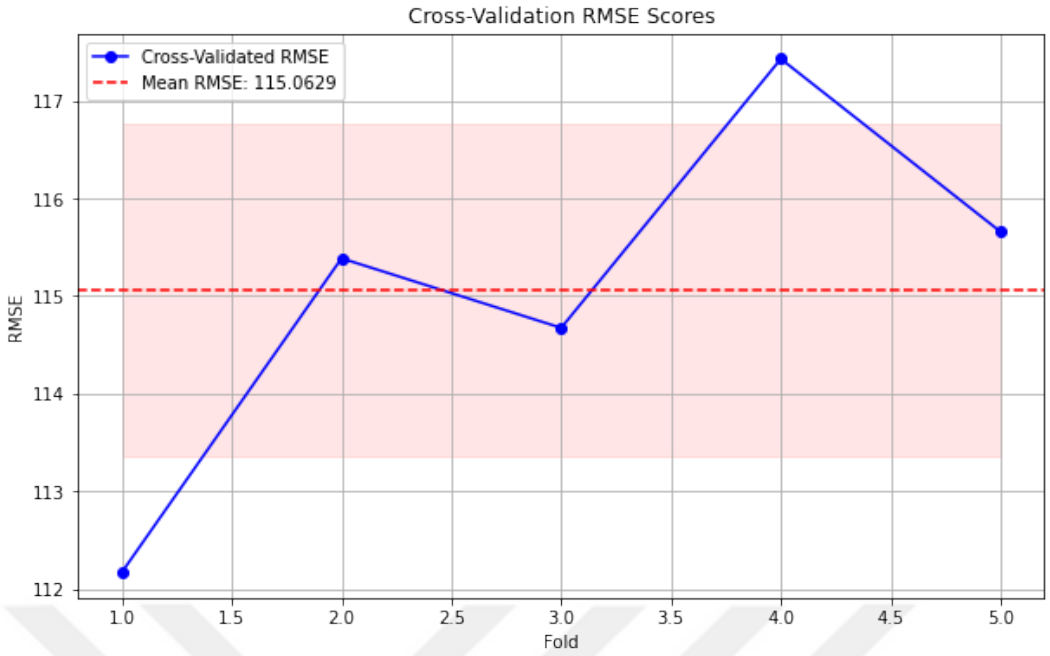
Şekil 4.7 15 Dakika-ADAUSDT-CATBOOST-RMSE

Şekil 4.7'de ADAUSDT'nin CatBoost modelindeki 15 dakikalık RMSE değerleri gösterilmiştir. ADAUSDT de düşük kapanış değerlerine sahip olduğu için diğer düşük kapanış değerli kripto paralar gibi daha düşük RMSE değerlerine sahiptir.



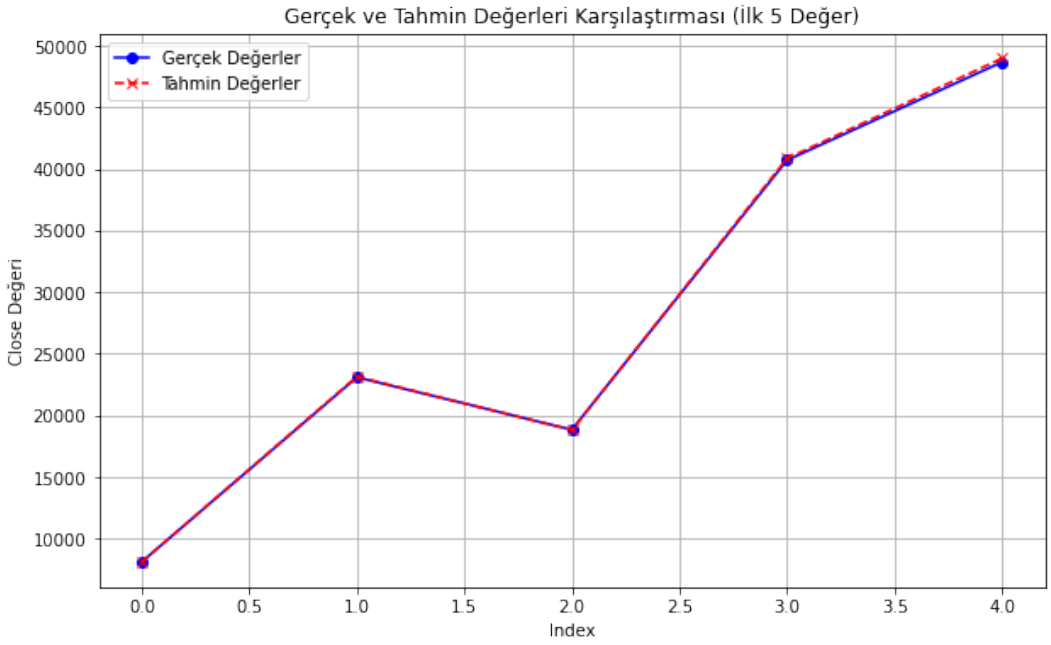
Şekil 4.8 15 Dakika-ADAUSDT-CATBOOST-TEST

Şekil 4.8'te ADAUSDT'nin CatBoost modelindeki test sonuçları gösterilmiştir. ADAUSDT'nin tahminleri, düşük kapanış değerleri nedeniyle daha doğru sonuçlar vermektedir.



Şekil 4.9 15 Dakika-BTCUSDT-CATBOOST-RMSE

Şekil 4.9'de BTCUSDT'nin CatBoost modelindeki 15 dakikalık RMSE değerleri gösterilmiştir. BTCUSDT'nin yüksek kapanış değerleri nedeniyle, RMSE değerlerinin diğer düşük kapanış değerli kripto paralara göre daha yüksek olduğu gözlemlenmektedir.



Şekil 4.10 15 Dakika-BTCUSDT-CATBOOST-TEST

	Gerçek Değerler	Tahmin Değerler
0	8113.12	8118.777835
1	23096.80	23153.491925
2	18844.78	18811.135275
3	40712.01	40853.818932
4	48686.94	48986.915207

Şekil 4.9'de BTCUSDT için CatBoost modelinin test aşamasındaki RMSE değerleri gösterilmiştir. Bu grafik, modelin doğruluğunu değerlendirmek için kritik öneme sahiptir ve BTCUSDT'nin diğer kripto paralara kıyasla daha yüksek RMSE değerlerine sahip olduğunu ortaya koymaktadır. Bu durum, BTCUSDT'nin fiyat tahminlerinde daha fazla belirsizlik yaşanabileceğini işaret etmektedir. Şekil 4.10'te BTCUSDT'nin kapanış değerlerini gözlemliyoruz. Tahminler genellikle yakın olsa da, büyük değer farklılıkları yatırımcıları potansiyel zarar riskiyle karşı karşıya bırakabilir.

Tablo 4.1 15 Dakikalık RMSE Sonuçları

15 Dakika	ARIMA	LSTM	XGBOOST	LightGBM	CatBoost
ADAUSDT	0.4669	-	0.0061	0.0070	0.0053
BNBUSDT	121.4563	-	1.4279	1.5572	1.2846
BTCUSDT	17902.0949	-	125.2618	134.8309	115.0629
DASHUSDT	108.1039	-	1.0427	1.1078	0.8635
DOGEUSDT	0.1355	-	0.0016	0.0018	0.0013
ETHUSDT	649.6170	-	9.6078	10.3098	8.4535
LTCUSDT	76.3161	-	0.8404	0.9089	0.6991
NEOUSDT	28.9082	-	0.2927	0.3236	0.2580
TRXUSDT	0.0351	-	0.0004	0.0004	0.0003
XRPUSDT	0.3147	-	0.0045	0.0047	0.0036

Tablo 4.2 1 Saatlik RMSE Sonuçları

1 Saat	ARIMA	LSTM	XGBOOST	LightGBM	CatBoost
ADAUSDT	0.4732	0.2643	0.0102	0.0108	0.0085
BNBUSDT	124.1818	67.4507	2.5501	2.5297	2.0776
BTCUSDT	15731.3731	30157.9666	213.8568	210.5296	176.6490
DASHUSDT	91.7706	23.4892	1.7487	1.7298	1.3310
DOGEUSDT	0.1003	0.0328	0.0029	0.0029	0.0023
ETHUSDT	635.5889	1104.6708	16.2879	15.7820	12.9527
LTCUSDT	77.0919	23.5390	1.5486	1.4163	1.1434
NEOUSDT	25.2926	6.6706	0.4875	0.4580	0.3849
TRXUSDT	0.0309	0.0035	0.0007	0.0006	0.0005
XRPUSDT	0.3117	0.0746	0.0073	0.0074	0.0058

Tablo 4.3 4 Saatlik RMSE Sonuçları

4 Saat	ARIMA	LSTM	XGBOOST	LightGBM	CatBoost
ADAUSDT	0.5074	0.0194	0.0185	0.0168	0.0159
BNBUSDT	122.1126	126.5485	4.7632	5.0591	3.9592
BTCUSDT	16425.1683	30833.8808	396.8241	364.7286	330.6277
DASHUSDT	90.1159	46.4292	3.0004	3.1685	2.1628
DOGEUSDT	0.1192	0.0513	0.0052	0.0052	0.0039
ETHUSDT	628.8836	1525.6243	29.1087	27.4407	23.4605
LTCUSDT	73.4668	42.7780	2.6240	2.4184	1.9745
NEOUSDT	23.2223	8.9385	0.9361	0.7899	0.7799
TRXUSDT	0.0279	0.0058	0.0013	0.0011	0.0009
XRPUSDT	0.2909	0.0742	0.0147	0.0131	0.0105

Yapılan deneysel çalışmalarda hem zaman dilimine göre değerlendirme hem de kapanış değerlerine göre değerlendirme kriterleri dikkate alınarak elde edilen sonuçlar Tablo 4.1, 4.2 ve 4.3 ile gösterilmiştir. Bu tablolara göre en başarılı sonuç RMSE kriteri ile 0,0003 olarak CatBoost algoritmasıyla elde edilmiştir.

5
SONUÇ

5.1 Sonuç

Kripto para birimi fiyat tahminleri, karmaşık ve dinamik bir konudur ve tek bir modelin tüm durumlar için en iyi performansı sağlaması zor olabilir. Bu çalışmada, farklı zaman aralıklarında ve çeşitli kripto para birimleri için ARIMA, LSTM, XGBoost, LightGBM ve CatBoost gibi beş farklı makine öğrenimi modelinin performansı değerlendirilmiştir.

Analizlerimize göre, CatBoost genel olarak en iyi performansı göstermiş, onu LightGBM ve XGBoost izlemiştir. Zaman dilimi açısından, 15 dakikalık verilerin diğer zaman aralıklarına göre daha iyi performans gösterdiği tespit edilmiştir. Kısa zaman dilimlerinde fiyat hareketleri genellikle daha belirgin trendler sergiler ve bu da modellerin daha doğru tahminler yapmasını sağlayabilir. Ancak, bu zaman dilimleri bazı modeller için performans zorlukları da getirebilir, özellikle de LSTM gibi derin öğrenme modelleri için.

Kapanış değerlerine göre yapılan değerlendirmede ise, düşük kapanış değerlerine sahip TRXUSDT, DOGEUSDT, XRPUSDT ve ADAUSDT gibi kripto para birimlerinin tahminlerinin daha doğru sonuçlar verdiği görülmüştür. Bu birimlerin düşük değerlere sahip olmasının, tahmin edilmelerini daha istikrarlı hale getirdiği düşünülmektedir. Yüksek kapanış değerlerine sahip birimlerin ise daha karmaşık fiyat hareketleri sergileyerek tahmin edilmelerinin zorluğu belirlenmiştir. Özellikle BTCUSDT ve ETHUSDT gibi yüksek değerli birimler için model seçimi ve parametre ayarlamalarının daha dikkatli yapılması gerekmektedir.

Bu çalışma, kripto para birimi fiyat tahminleri üzerine yapılan araştırmaların karmaşıklığını ve makine öğrenimi tekniklerinin potansiyelini vurgulamaktadır. CatBoost gibi gradient boosting yöntemlerinin genellikle diğerlerinden daha iyi performans gösterdiği bulgusu, gelecekteki çalışmalarda bu yöntemlerin daha derinlemesine incelenmesini teşvik etmektedir.

Analizlerimize dayanarak, düşük kapanış değerlerine sahip olan TRXUSDT, DOGEUSDT, XRPUSDT ve ADAUSDT gibi kripto para birimlerinin yatırım için daha uygun olabileceğini tespit ettik. Bu kripto paraların, daha düşük hata oranlarına sahip olmaları nedeniyle, daha istikrarlı tahminler sağladığı gözlemlenmiştir. Potansiyel yatırımcılar için, bu tür düşük kapanış değerlerine sahip kripto paralar, daha az riskli yatırım fırsatları sunabilir.

Diğer yandan, yüksek kapanış değerlerine sahip BTCUSDT ve ETHUSDT gibi kripto paralar, yüksek hata potansiyeline sahip olup, daha karmaşık fiyat hareketleri sergilemektedir. Bu nedenle, yüksek kapanış değerlerine sahip kripto paraların, daha dikkatli ve özenli bir yatırım stratejisi gerektirdiği sonucuna varılmıştır.

Genel olarak çalışmamız, kripto para birimi fiyat tahminlerini geliştirmek ve uygun yatırım fırsatlarını belirlemek için gradient boosting tekniklerinin etkili bir şekilde kullanılabileceğini göstermektedir. Potansiyel yatırımcılara, düşük kapanış değerlerine sahip kripto paralara odaklanarak daha güvenli ve öngörülebilir yatırımlar yapmaları önerilmektedir.

5.2 Gelecek Çalışmalar ve Öneriler

Gelecekte yapılacak çalışmalarda daha geniş bir model yelpazesi üzerindeki etkileri değerlendirmek daha kapsamlı bir analiz sunabilir. Ayrıca, zaman aralığına bağlı olarak modellerin adaptasyon yetenekleri üzerine odaklanan çalışmalar, kripto para birimi fiyat tahminleme alanındaki anlayışımızı daha da derinleştirebilir. Yüksek kapanış değerlerine ve yüksek hacme sahip kripto para birimlerine farklı bir yaklaşım sergilenerek daha uygun bir doğruluk ölçütü sonraki çalışmalarda kullanılabilir; yüzdelik olarak artma ve azalma miktarlarına göre tahminler yapılabilir.

Veri temizleme, öznitelik mühendisliği ve ölçeklendirme gibi ön işleme adımları daha detaylı bir şekilde incelenebilir ve farklı yöntemlerle denenebilir.

Farklı piyasa koşullarında (volatilite, likidite vb.) modellerin performansının nasıl değiştiği ayrıca incelenebilir.

Kripto para birimi fiyat tahminlerinin yatırım kararlarında nasıl kullanılabileceği ve risk yönetimi stratejilerinin nasıl geliştirilebileceği üzerine daha derinlemesine çalışmalar yapılabilir.

KAYNAKÇA

[1] M. Saad, "Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions," *IEEE Systems Journal*, 2020.

[2] S. Kervanci, A. K. A. Y. Fatih, "Review on bitcoin price prediction using machine learning and statistical methods," *Sakarya University Journal of Computer and Information Sciences*, vol. 3, no. 3, pp. 272–282, 2020.

[3] C. Chen, "Forecasting interaction of exchange rates between fiat currencies and cryptocurrencies based on deep relation networks," in *2019 IEEE International Conference on Agents (ICA)*, 2019.

[4] E. Akyildirim, A. Goncu, A. Sensoy, "Prediction of cryptocurrency returns using machine learning," *Annals of Operations Research*, vol. 297, pp. 3–36, 2021.

[5] E. Demirci, M. Karaatlı, "Kripto para fiyatlarının lstm ve gru modelleri ile tahmini," *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 2023.

[6] F. Alamery, "Cryptocurrency analysis using machine learning and deep learning approaches," *Journal of Computer & Electrical and Electronics Engineering Sciences*, 2023. doi: 10.51271/jceees-0007.

[7] H. Sebastião, P. Godinho, "Forecasting and trading cryptocurrencies with machine learning under changing market conditions," *Financial Innovation*, vol. 7, pp. 1–30, 2021.

[8] D. Jothidurai, P. Kanchan, R. Raj, "Cryptocurrency price prediction and forecasting market momentum using machine learning techniques," *International Journal for Research in Applied Science and Engineering Technology*, 2023. doi: 10.22214/ijraset.2023.55026.

[9] X. Song, "Predicting cryptocurrency investment suitability using machine learning techniques," *Applied and Computational Engineering*, 2023. doi: 10.54254/2755-2721/29/20230726.

[10] J. Wu, X. Guo, M. Fang, J. Zhang, "Short term return prediction of cryptocurrency based on xgboost algorithm," in *2022 International Conference on Big Data, Information and Computer Network (BDICN)*, 2022, pp. 39–42. doi: 10.1109/BDICN55575.2022.00015.

[11] S. Swati, A. Mohan, "Cryptocurrency value prediction with boosting models," in *2022 International Conference on Intelligent Innovations in Engineering and Technology (ICIIET)*, 2022, pp. 183–188. doi: 10.1109/ICIIET55458.2022.9967540.

[12] V. Derbentsev, V. Babenko, K. Khrustalev, H. Obruch, S. Khrustalova, "Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices," *International Journal of Engineering*, 2021. doi: 10.5829/IJE.2021.34.01A.16.

[13] T. Li, A. Chamrajnagar, X. Fong, N. Rizik, F. Fu, "Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model," *Frontiers in Physics*, 2018. doi: 10.3389/fphy.2019.00098.

[14] N. Latif, J. Selvam, M. Kapse, V. Sharma, V. Mahajan, "Comparative performance of lstm and arima for the short-term prediction of bitcoin prices," *Australasian Accounting, Business and Finance Journal*, 2023. doi: 10.14453/aabfj.v17i1.15.

[15] L. Pan, "Cryptocurrency price prediction based on arima, random forest and lstm algorithm," *BCP Business & Management*, 2023. doi: 10.54691/bcpbm.v38i.4313.

[16] H. Sinha, S. Keshavrao, P. Verma, "Comparative analysis of prediction models for cryptocurrency price prediction (may 2022)," *International Journal of Management and Humanities*, 2022. doi: 10.35940/ijmh.j1497.0681022.

[17] E. Sener, I. Demir, "Wavelet-denoised arima and lstm method to predict cryptocurrency price," in *2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2022, pp. 577–582. doi: 10.1109/ISMSIT56059.2022.9932727.

[18] Y. Hua, "Bitcoin price prediction using arima and lstm," *E3S Web of Conferences*, 2020. doi: 10.1051/e3sconf/202021801050.

[19] S. Pasak, R. Jayadi, "Investment decision on cryptocurrency: Comparing prediction performance using arima and lstm," *Journal of Information Systems and Informatics*, 2023. doi: 10.51519/journalisi.v5i2.473.

[20] Unknown, *Cryptocurrency price prediction lstm model*, Accessed: 2024-07-13, 2024. [Online]. Available: https://miro.medium.com/v2/resize:fit:1039/1*QiLYco0hB8EterWYTyyv4g.png.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. E. Dilasa Kurt, A. Elbir, B. Diri, *"Prediction of Cryptocurrencies Using Machine Learning Methods"* ,Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü GradColloquium'24, 2024.