Project ID: 2024/CSE(DS)/GID=06

BITCOIN PRICE PREDICTION USING MACHINE LEARNING

A PROJECT REPORT
Submitted By

RISHAV RAJ SINGH
(2000271540044)
RAUSHAN KUMAR
(2000271540043)
ADITYA KUMAR
(2000271540007)
GAURAV SINGH
(2100271549002)

Submitted in partial fulfillment of the requirements for the degree of Bachelor of Technology in Computer Science and Engineering(DATA SCIENCE)

to

Department of Computer Science & Engineering
AJAY KUMAR GARG ENGINEERING COLLEGE, GHAZIABAD
DR. APJ ABDUL KALAM TECHNICAL UNIVERSITY, LUCKNOW

May 22, 2024

Declaration

We hereby declare that the work presented in this report entitled "Bitcoin Price Prediction Using Machine Learning", was carried out by us. We have not submitted the matter embodied in this report for the award of any other degree or diploma of any other University or Institute. I have given due credit to the original authors / sources for all the words, ideas, diagrams, graphics, computer programs, experiments, results, that are not my original contribution. I have used quotation marks to identify verbatim sentences and given credit to the original authors / sources.

I affirm that no portion of my work is plagiarized and the experiments and results reported in the report are not manipulated. In the event of a complaint of plagiarism and the manipulation of the experiments and results, I shall be fully responsible and answerable.

Name : RISHAV RAJ SINGH
Roll No. : 2000271540044

Name : RAUSHAN KUMAR
Roll No. : 2000271540043

Name : ADITYA KUMAR
Roll No. : 2000271540007

Name : GAURAV SINGH
Roll No. : 2100271549002

Certificate

This is to certify that the report entitled **BITCOIN PRICE PREDICITON USING MACHINE LEARNING** submitted by **Rishav Raj Singh (2000271540044)**, **Raushan Kumar (2000271540043)**, **Aditya Kumar(2000271540007)** and **Gaurav Singh(2100271549002)** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in CSE(DS) is a bonafide record of the project work carried out by him/her under my/our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Mr. Sharvan Kumar
Assistant Professor
Dept. of Computer Science & Engineering
AKG Engineering College

Dr.Avdhesh Gupta
Professor Incharge (CSE DS)
Dept. of Computer Science & Engineering
AKG Engineering College

Place: Ghaziabad
Date:

Acknowledgements

We would like to express my heartfelt gratitude to all those who have played a crucial role in the completion of this report on the **Bitcoin Price Prediction using Machine Learning**. This project has been a journey of growth and learning and I am deeply appreciative of the support and encouragement we have received. First and foremost, we extend our sincere thanks to my report advisor **Mr. Sharvan Kumar** for their unwavering guidance, insightful feedback and continuous support throughout the research process. His expertise has been instrumental in shaping the direction of this Research. A special appreciation goes to the dedicated team of developers and researchers who contributed to the design and implementation of the Bitcoin Price Prediction using Machine Learning. Their commitment to excellence and tireless efforts have resulted in a robust and effective system. We are grateful to Department of Computer Science and engineering for providing the necessary resources and facilities that facilitated the successful execution of this research. We would also like to express my thanks to my friends and family for their unwavering support, encouragement and understanding during the demanding phases of report writing. their belief in us has been a driving force. Lastly, we want to acknowledge the countless individuals who, directly or indirectly, contributed to the completion of this project. Our collective efforts have left an indelible mark on this research. Thank you all for being an integral part of this journey. your support has been invaluable and We are truly grateful.

Abstract

We select the subject of our research as cryptocurrency Price prediction as it is difficult to counterfeit because of this security feature and a defining feature of cryptocurrency and arguably its most endearing allure, is its organic nature;it is not issued by any central authority, rendering it theoretically immune to government interference or manipulation.

In the recent years,in the terms of market capitalization by value and volume bitcoin is one of the most important cryptocurrency so this is the primary reason for us to choose it as our subject for research.Another reason for selection of bitcoin is based on its volatile nature which affects the financial marketers which costs them in monetary value.The price of the bitcoin is fluctuated because of the various reasons such as Company level factors and External factors which include industry shifts,government regulations.The previous forecasting methods uses the machine learning algorithm for prediction of the bitcoin price based on the historical data and market voices related to the bitcoin.It also does not set the upper and lower boundary for fluctuations in the price of bitcoin. Therefore the main objective of our research is to forecast the price of bitcoin using various machine learning algorithms like LSTM,Fb Prophet, ARIMAX and XG-BOOST.Out of the four models we will compare the Rmse and Mae and among these four algorithm whichever perform better will be selected as the base model.In this research we set the upper, lower limit and price predicted value of the bitcoin so that the people can judge the fluctuation in the price of the bitcoin.The upper limit can reduce the risk factor when the bitcoin price exceeds it.The lower limit can reduce the sudden crash in the financial market when it exceeds it limit.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

In this chapter, we discuss cryptocurrencies, why it is necessary to predict the price of cryptocurrencies and its importance and the challenges associated with cryptocurrencies. It also highlights factors that influence the value of cryptocurrencies. Then, present the problem and research objectives to convey the purpose of the report.

1.2 Cryptocurrency

Cryptocurrency is a digital or virtual form of currency that uses cryptography for security. Unlike traditional currencies issued by governments (fiat money), cryptocurrencies operate on decentralized networks based on blockchain technology.

1.2.1 Key Characteristics of Cryptocurrencies:

- Decentralization: Most cryptocurrencies are decentralized networks based on blockchain technology. This means they are not controlled by any single entity, like a central bank or government.

- Cryptography: Cryptocurrencies use cryptographic techniques to secure transactions and control the creation of new units. This ensures the integrity and security of the system.

- Blockchain Technology: The blockchain is a public, distributed ledger that records all transactions made with a cryptocurrency. It is immutable, meaning once data is recorded, it cannot be changed without altering all subsequent blocks, ensuring transparency and security.

1.2.2 Popular Cryptocurrencies:

- Bitcoin (BTC): The first and most well-known cryptocurrency, created by an anonymous person or group of people using the pseudonym Satoshi Nakamoto in 2008. Bitcoin is often referred to as digital gold due to its limited supply (21 million coins).

- Ethereum (ETH): A blockchain platform that enables the creation of smart contracts and decentralized applications (dApps). Ethereum introduced the concept of programmable blockchain.

- Ripple (XRP): Designed primarily for digital payment processing and remittances, Ripple aims to enable real-time cross-border payments with low fees.Litecoin (LTC): Created by Charlie Lee, Litecoin is a peer-to-peer cryptocurrency that offers faster transaction times compared to Bitcoin.

- Cardano (ADA): A blockchain platform focused on providing a more secure and scalable infrastructure for the development of dApps and smart contracts, with an emphasis on academic research and peer-reviewed development.

1.2.3 Use Cases of Cryptocurrencies

- Digital Payments: Cryptocurrencies can be used for online transactions, providing an alternative to traditional payment methods like credit cards or bank transfers.

- Investment: Many people invest in cryptocurrencies, hoping their value will increase over time. This has led to the rise of cryptocurrency exchanges and trading platforms.

- Decentralized Finance (DeFi): DeFi platforms aim to recreate traditional financial services, such as lending, borrowing and trading, using blockchain technology without intermediaries.

- Smart Contracts: These are self-executing contracts with the terms directly written into code, facilitating, verifying, or enforcing the negotiation or performance of a contract.

1.2.4 Risks and Challenges:

- Volatility: Cryptocurrency prices are highly volatile, which can lead to significant gains or losses in a short period.

- Regulatory Uncertainty: The legal status of cryptocurrencies varies by country and regulatory frameworks are still evolving, leading to uncertainty.

- Security Concerns: While blockchain technology is secure, cryptocurrency exchanges and wallets can be susceptible to hacks and fraud.

- Scalability: Many cryptocurrencies face challenges with scalability, affecting the speed and cost of transactions as network usage grows.

- Environmental Impact: Some cryptocurrencies, particularly Bitcoin, require substantial energy for mining, raising concerns about their environmental footprint.

1.2.5 Conclusion:

Cryptocurrencies represent a transformative innovation in the financial sector, offering new possibilities for digital transactions, decentralized finance and asset management. However, they also present challenges and risks that need to be managed carefully. As technology and regulatory landscapes evolve, the role of cryptocurrencies in the global economy will likely continue to grow and diversify.

Machine learning makes price guessing more accurate. Old ways of setting prices weren't very exact and relied more on guesswork than facts. This often led businesses in the wrong direction. ML can help you make more money. It shows you how prices in your industry change over a year. For example, if you know a supplier usually raises prices in October, you can buy more in September before the prices go up. This saves you money and increases your profit. Using AI for pricing means you'll work faster, more efficiently and get the prices right, no matter what happens in the market. The next chapter talks about different machine learning and deep learning models, which are smart computer methods that help make even better predictions about Bitcoin prices based on social media and past price trends.

1.3 Objective:

The main objective of this research is to forecast the price of Bitcoin using various machine learning algorithms, including Long Short-Term Memory (LSTM), Facebook Prophet (Fb Prophet), Autoregressive Integrated Moving Average with Explanatory Variables (ARIMAX) and Extreme Gradient Boosting (XGBoost). By comparing the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) of these models, the research aims to identify the most accurate algorithm for Bitcoin price prediction. Additionally, the research sets upper and lower boundaries for Bitcoin price fluctuations to help mitigate risk and prevent sudden market crashes, thereby providing valuable insights for financial market participants.

Chapter 2

Cryptocurrencies Visualization

2.1 Overview

In this chapter, we will try to visualize different cryptocurrency and their different attributes such as market share, volume along with their high, low , close, open price on daily, weekly, monthly, yearly basis.

2.2 Visualization

The COVID-19 has changed the dynamics of the investors related to investment and trading and the people move their traditional portfolio to cryptocurrency. After the growth of interest of common people towards investment in different field rather than only in Stock and gold market.

So we have split the visualization in two parts where the first parts is from year(2015-2020) and the second one will be from year (2020-2023) and the frequency of data be will on daily basis.The dataset is prepared from the historical data of different cryptocurrencies from yahoo finance and kaggle.Some of the dataset photo are attached in further chapter of data collection.

The various types of visualisations include various graph, charts ,plots and many more is used to visualise different pattern and trends so that which helps in classification to implementation feature engineering, cleaning of data.Thus we will understand the use of each plot,chart,map and their characteristic in the following subsections:

2.2.1 Pie chart

Pie charts are vital in data visualization as they offer a straightforward and intuitive method for comparing proportions, allowing for the easy comprehension of complex data at a single glance. Their ability to depict parts of a whole in a clear manner makes them indispensable in decision-making processes that require visual data comparison, ensuring that information is communicated effectively and efficiently to all stakeholders involved.

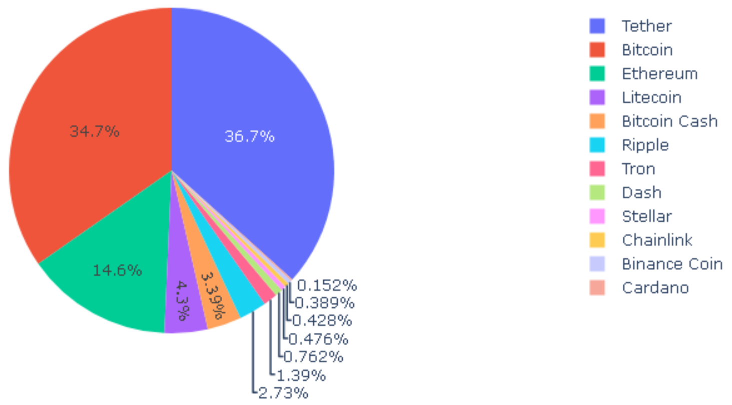


Figure 2.1: Pie chart for year (2015-2020)

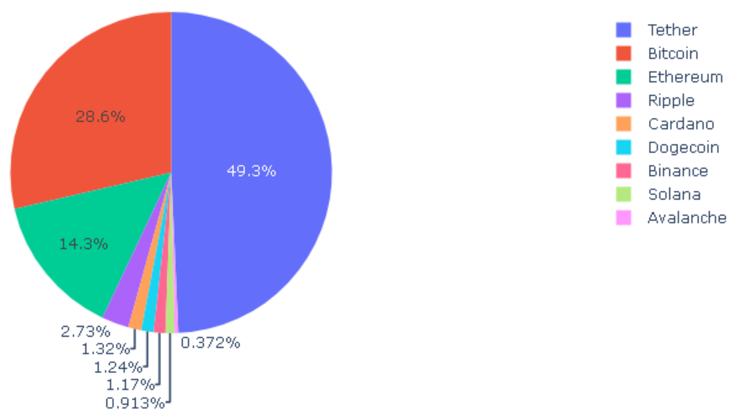


Figure 2.2: Pie chart for year(2020-2023)

The given two pie chart clearly depicts the fact that tether and bitcoin still hold the major chunk of market share in Figure:2.1 & Figure:2.2 .

2.2.2 Scatter plot

A scatter plot is a graphical representation that shows the relationship between two continuous variables, making it simpler to spot patterns, trends and clusters within the data. These plots are instrumental in revealing the data's underlying structure and pinpointing outliers—data points that diverge from the established trend, potentially signaling anomalies or errors. For creating scatter plots in python, libraries such as matplotlib, seaborn and plotly are available, offering extensive features for detailed and interactive visual representations.

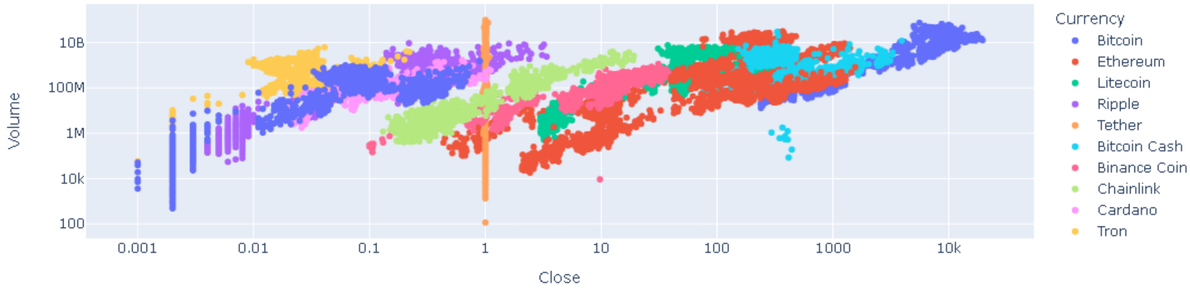


Figure 2.3: Scatter plot for years(2015-2020) between Volume and Close

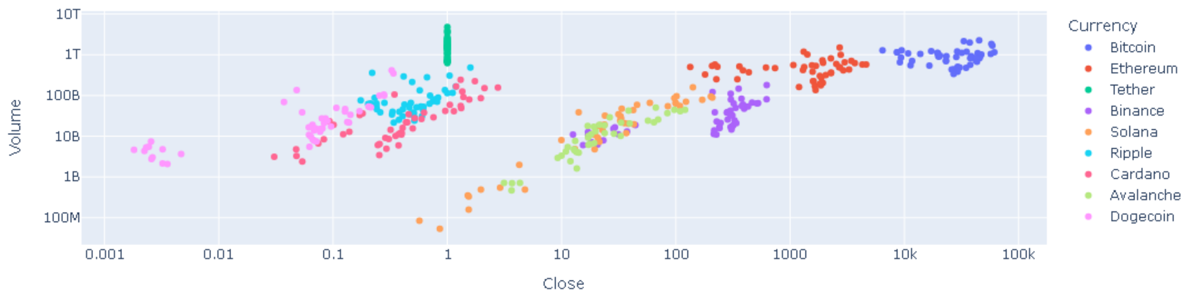


Figure 2.4: Scatter plot for years(2020-2023) between Volume and Close

From the above two scatter plot it is clearly visible that the COVID-19 has affected the price of bitcoin drastically.Till the 2020 the closing price of bitcoin price is in range of 10K USD with volume in 10B range, while post 2020 bitcoin price surges to 100k and the volume in 1T range.It clearly overshadowed the second best individually in volume and closing price on daily basis.

2.2.3 Box plot

Box Plots,also termed as whisker plots, are essential tools in statistics for visualizing data distribution and variability. They encapsulate the core measures of a dataset minimum, first quarterly, median,upper fence,lower fence, third quarterly and maximum allowing for a quick assessment of central tendency, spread and outliers. Available in Python's matplotlib, seaborn and pandas, among other programming libraries, Box Plots facilitate comparisons across datasets, making them invaluable in exploratory data analysis and various professional fields like finance and research.

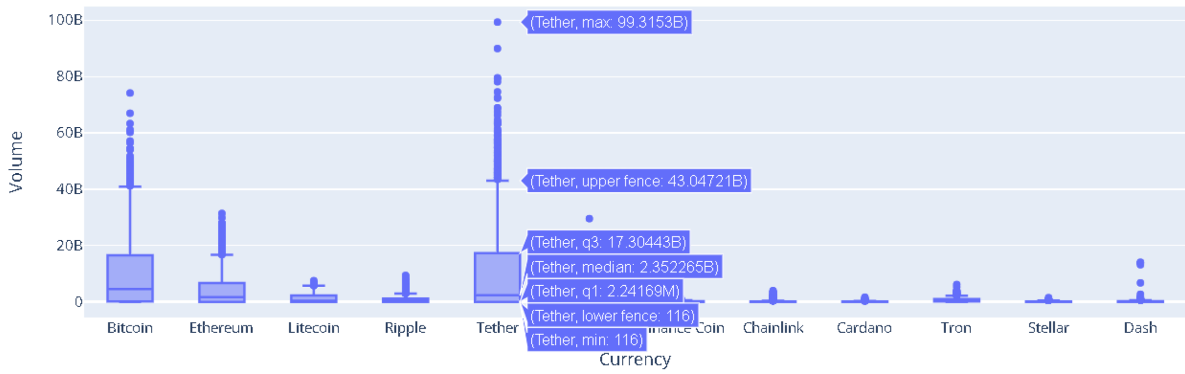


Figure 2.5: Box plot for years(2015-2020) between Volume and Currency

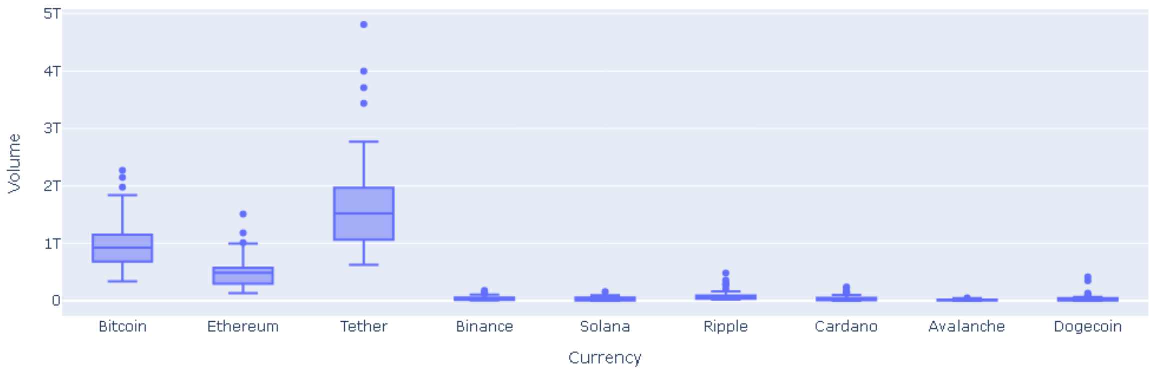


Figure 2.6: Box plot for years(2020-2023) between Volume and Currency

In above Figure:2.5 we can see the importance of box plot into which a single plot describes so much details about the dataset.

2.2.4 Candlestick chart

Candlestick charts are a staple in financial analysis, providing a visual representation of price movements within a specific time frame. Each 'candle' in the chart reflects the opening, high, low and closing prices of a security, commodity, or currency. The color coding of the candles helps traders quickly ascertain market trends and potential reversals.In Python, candlestick charts can be created using libraries such as matplotlib, seaborn and plotly. These libraries offer a range of customization options, allowing analysts to tailor the charts to their specific needs.

The use cases for candlestick charts extend beyond financial markets. They can be applied to any dataset that varies over time, such as weather patterns or stock prices. For example, a trading firm might track hourly stock prices to inform trading decisions, while a meteorological organization could use them to visualize temperature changes over days.Candlestick charts are thus a versatile tool for various data analysis applications, offering clear insights into temporal data trends.

Figure 2.7: Candlestick chart with range slider

2.2.5 Heat Map

Heat maps are a crucial visualization tool in data analysis, offering a color-coded representation of complex data sets to reveal patterns, correlations and trends at a glance. They are particularly useful in areas like geography for mapping density, in finance for correlation matrices and in science for gene expression levels. In Python, heat maps can be created using libraries such as seaborn,matplotlib, plotly and folium which provides a high-level interface for drawing attractive and informative statistical graphics. Heat maps serve as an essential tool for analysts and researchers to communicate complex information effectively.

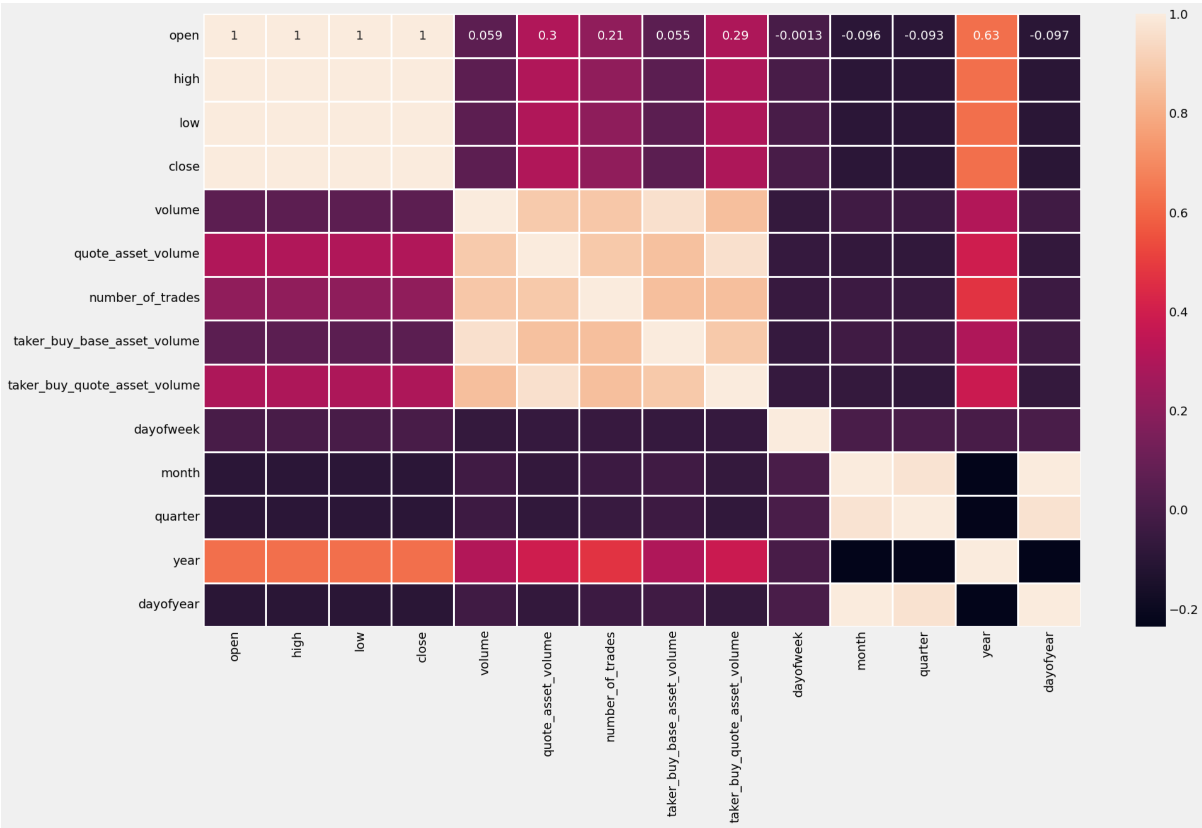


Figure 2.8: Heat map

The Figure:2.9heatmap chart that represents the correlation between different variables related to Bitcoin trading where each square indicates the correlation coefficient between two variables, with values ranging from 1 (indicating a perfect positive correlation) to -1 (indicating a perfect negative correlation). The color gradient from dark purple to dark red reflects the strength of the correlation, with dark red representing a strong positive correlation and dark purple indicating a low or negative correlation.

The variables included in this heatmap typically involve trading metrics like 'open', 'high', 'low', 'close' and 'volume', as well as other factors that could influence trading patterns, such as 'dayofweek' and 'dayofyear'. This kind of visualization is particularly useful for traders or analysts who are looking to understand how different aspects of Bitcoin trading are interrelated, which can aid in developing informed trading strategies or analyzing market behavior. For example, a strong positive correlation between 'close' and 'high' prices might suggest that on days when Bitcoin starts strong, it often closes strong as well. Conversely, a lack of correlation between 'volume' and price change could indicate that trading volume isn't always a reliable predictor of price movements.

Such insights are valuable in the cryptocurrency market, where volatility is common and understanding the interplay between different factors can lead to better decision-making.

2.2.6 Pair plot

The Figure:2.9 is a matrix of scatter plots, which is a comprehensive visualization tool used to analyze the relationships between multiple variables in a dataset. In the context of Bitcoin, each scatter plot within the matrix likely represents the relationship between different trading metrics such as 'open', 'high', 'low', 'close' and 'volume'. These metrics correspond to the opening, highest, lowest and closing prices of Bitcoin, as well as the trading volume within specific time periods.

This type of visualization is particularly useful for identifying patterns, trends and potential correlations between the different aspects of Bitcoin's trading data. For instance, by examining the scatter plot that compares ‘open’ and ‘close’ prices, one could assess whether there is a consistent trend in Bitcoin’s price movement within a day. Similarly, the plot correlating ‘volume’ and ‘close’ prices might reveal if higher trading volumes are associated with significant price changes at closing.

Such insights gleaned from the scatter plot matrix can be invaluable for traders and analysts in making informed decisions, as they provide a deeper understanding of market dynamics and the interplay between various trading indicators. It’s a strategic tool for cryptocurrency market analysis, helping to forecast future price movements based on historical data.

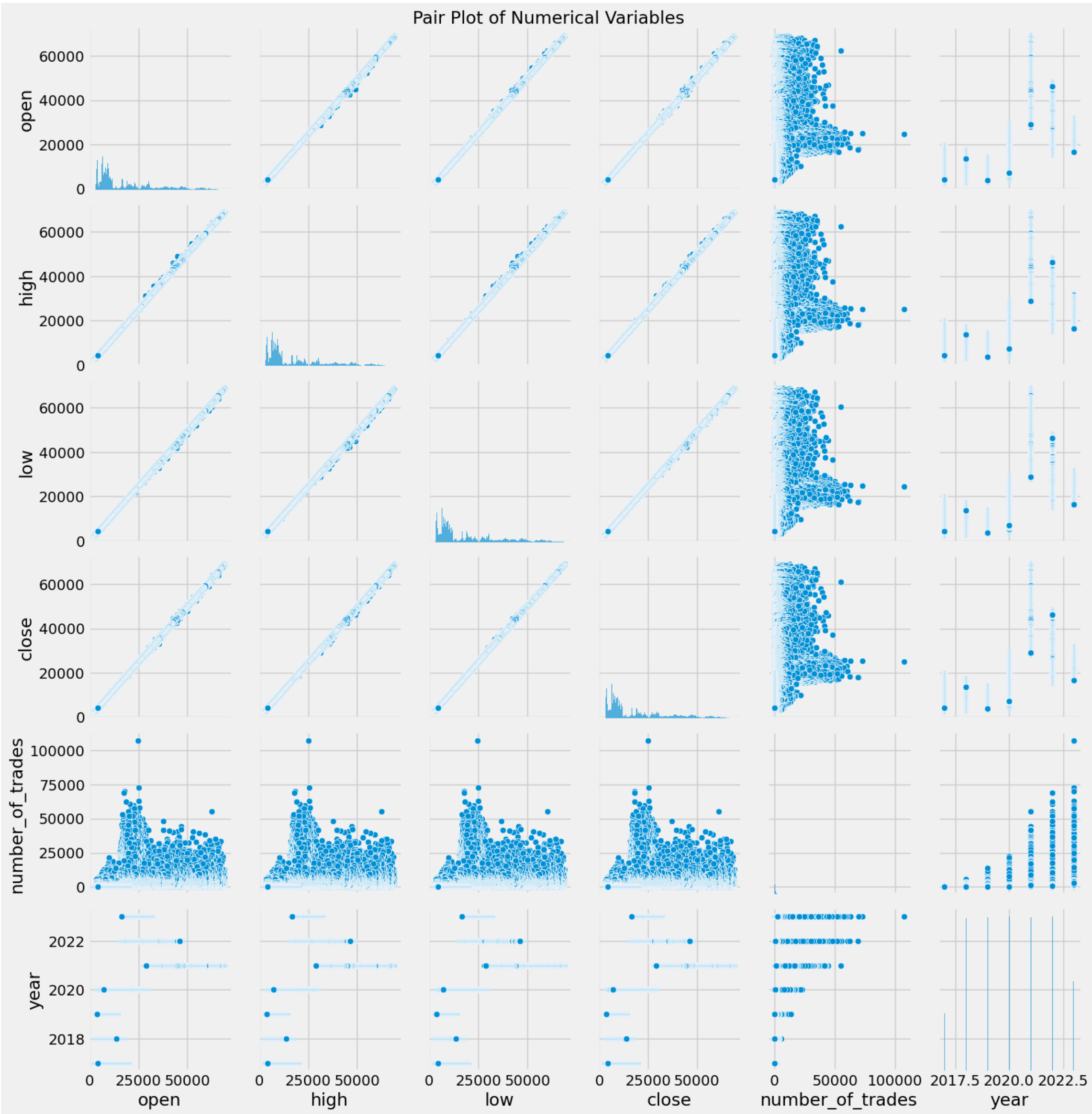


Figure 2.9: Pair plot(matrix of scatter plot)

2.3 Summary

The chapter provides a comprehensive overview of various visualization techniques used to analyze cryptocurrencies, emphasizing their market share, volume and price movements. The visualizations are divided into two periods: pre- and post-COVID-19, highlighting the shift in investment trends towards cryptocurrencies.

Pie charts demonstrate the dominant market share held by Bitcoin and Tether, while scatter plots reveal the significant impact of COVID-19 on Bitcoin's price and trading volume, with a marked increase post-2020.

Box plots offer a detailed view of the distribution of trading volumes across different cryptocurrencies, showcasing the variability and outliers within the data.

Candlestick charts provide insights into the price action of cryptocurrencies over time, with color-coded candles indicating market trends and potential reversals.

Heat maps illustrate the correlation between various trading metrics, such as opening and closing prices, which is crucial for understanding market behavior and developing trading strategies.

Pair plots (scatter plot matrices) allow for the examination of relationships between multiple trading metrics, aiding in the identification of trends and correlations that inform trading decisions.

In summary, the chapter underscores the importance of visual analytics in cryptocurrency trading, with Bitcoin emerging as a preferred choice for the project due to its significant price advantage and popularity among traders.

This is evidenced by the patterns observed in the latter visualizations can also be seen in below Figure:2.10

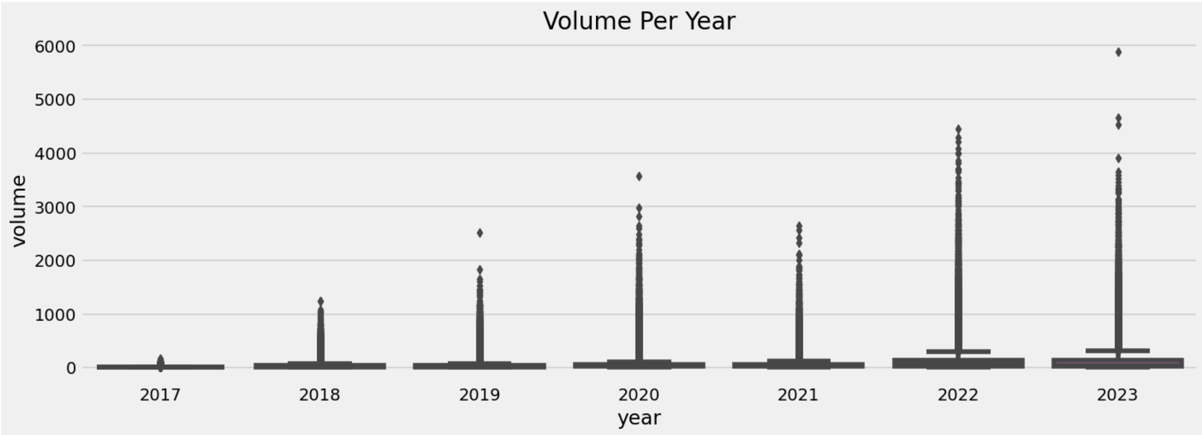


Figure 2.10: Volume Bar

The chart indicates a significant growth in volume over time, with the year 2023 showcasing the highest volume. This reflects the increasing interest and trading activity in Bitcoin, suggesting that more people are engaging with this cryptocurrency as time progresses. The rising trend depicted in the chart aligns with the broader context of Bitcoin's growing prominence in the financial world as a leading digital asset.

The data visualized in Figure:2.11 could be of particular interest to investors, analysts and researchers who track the cryptocurrency's performance and its potential future trajectory.Thus we selected Bitcoin as our digital cryptocurrency for our work.

Bitcoin Halving Event	Price on Halving Day	Price 150 days later
2012	$12.35	$127
2016	$650.53	$758.81
2020	$8,821.42	$10,943

Figure 2.11: Bitcoin growth Chart

Chapter 3

Literature Review

Junwei Chen [21]

- Year: (2023)
- Objective: Analyzing and evaluating the effectiveness of various machine learning techniques in predicting Bitcoin prices.
- Advantages: Comprehensive comparison of different machine learning models; enhanced robustness through the use of multiple financial indicators.
- Limitations: Inherent volatility and unpredictability of Bitcoin prices; findings may not generalize well to future market conditions.
- Summary: Junwei Chen's study focuses on analyzing and evaluating the effectiveness of various machine learning techniques for predicting Bitcoin prices.The research offers a comprehensive comparison of different models and improves robustness by incorporating multiple financial indicators.Despite these strengths, the study acknowledges significant challenges, including the inherent volatility of Bitcoin prices and the potential for findings to be less applicable to future market conditions, highlighting the need for ongoing adaptation and model refinement.

Harsh Parikh, Nisarg Panchal and Ankit Sharma [22]

- Year: (2023)

- Objective: Forecasting cryptocurrency prices.
- Advantages: Provides a comparative performance analysis of different models.
- Limitations:Predictive Limitations and Data Quality.
- Summary: Their study focuses on forecasting cryptocurrency prices, offering a comparative performance analysis of various models.While their research provides valuable insights, it also highlights limitations concerning predictive accuracy and data quality.This underscores the ongoing challenges in predicting cryptocurrency prices, particularly in the context of fluctuating market dynamics and data integrity issues.

Amirzadeh, Rasoul, Asef Nazari and Dhananjay Thiruvady [23]

- Year: (2022)
- Objective: To survey the application of artificial intelligence techniques in cryptocurrency markets.
- Advantages: AI can process vast and complex datasets at high speed, uncovering patterns and insights for more effective trading strategies.
- Limitations: High market volatility and limited historical data challenge the consistency and adaptability of AI models.
- Summary: The surveys the use of artificial intelligence in cryptocurrency markets, exploring various AI techniques such as machine learning and deep learning for market analysis and trading.It highlights the advantages of AI in processing large datasets quickly and uncovering complex patterns that can enhance trading strategies.However, it also points out significant limitations, including high market volatility and the relatively short history of cryptocurrency data, which pose challenges to the effectiveness and consistency of AI models.Despite these challenges, the study underscores the transformative potential of AI in the evolving field of cryptocurrency trading.

Linxi Pan [24]

- Year: (2022)
- Objective: Predicting cryptocurrency prices.
- Advantages: Comparative analysis of prediction models.
- Limitations: Despite the diverse set of algorithms utilized, the study may face challenges related to the inherent volatility and unpredictability of cryptocurrency markets, potentially affecting the accuracy of the predictions.
- Summary: Linxi Pan's research aims to predict cryptocurrency prices through a comparative analysis of prediction models.While the study provides valuable insights into the performance of different models, it notes that ARIMA (AutoRegressive Integrated Moving Average) may struggle to accurately predict prices in highly volatile cryptocurrency markets.This highlights the importance of considering the specific characteristics of cryptocurrency data when selecting prediction models.

Sahil Sejwal, Kartik Aggarwal, Soumya Ranjan Nayak [25]

- Year: (2023)
- Objective: Time series analysis of cryptocurrency.
- Advantages: Better results than ARIMA.
- Limitations: ARIMAX's performance is highly dependent on the external variables used.
- Summary: The study focuses on conducting time series analysis of cryptocurrency data.It highlights that their approach yields superior results compared to ARIMA .However, the research notes that the performance of their ARIMAX model heavily relies on the selection and quality of external variables.This underscores the importance of

careful consideration and selection of variables in cryptocurrency time series analysis.

Kanksha and Harjit Singh [27]

- Year: (2021)
- Objective: Predicting Bitcoin and Litecoin prices.
- Advantages: The study leverages the FB Prophet model, which is specifically designed for time series forecasting and has shown effectiveness in capturing patterns and trends in cryptocurrency price data.
- Limitations: Model may still encounter challenges in accurately predicting cryptocurrency prices due to the inherent volatility and unpredictability of the market, as well as potential limitations in the model's ability to handle extreme fluctuations.
- Summary: In their ,they aim to predict the prices of Bitcoin and Litecoin.The research demonstrates that their model provides accurate predictions, showcasing its effectiveness for these cryptocurrencies.However, a notable limitation is that the model may not adequately capture sudden market anomalies, which can lead to unexpected deviations in price forecasts.This highlights the need for incorporating mechanisms to handle abrupt market changes for more robust predictive performance.

Khiewngamdee, C., Chanaim, S.[28]

- Year: (2023)
- Objective: To investigate whether incorporating cryptocurrency data can enhance the forecasting accuracy of exchange rate returns.
- Advantages: The integration of cryptocurrency data improves forecasting performance, potentially providing more accurate predictions

for exchange rate returns.

- Limitation: The complexity and volatility of cryptocurrency markets may introduce additional challenges and unpredictability into the forecasting models.
- Summary: They investigate the impact of cryptocurrency on forecasting exchange rate returns.Their study examines whether the inclusion of cryptocurrency data enhances forecasting accuracy the research explores the potential of cryptocurrency to improve predictive models in the context of exchange rate dynamics.By analyzing relationship, the authors contribute valuable insights into the intersection of cryptocurrency and traditional financial markets

Hegde, Gayatri [29]

- Year: (2023)
- Objective: To predict Bitcoin prices utilizing deep learning techniques, contributing to the field of cryptocurrency forecasting.
- Advantages: The study harnesses the power of deep learning, which can effectively capture complex patterns in Bitcoin price movements, potentially leading to more accurate predictions.
- Limitation: The reliance on historical data and the inherent volatility of cryptocurrency markets may pose challenges to the model's predictive accuracy in real-time scenarios.
- Summary: It employs deep learning techniques to predict Bitcoin prices, demonstrating the capability of neural networks to identify complex patterns in market data.The study highlights the potential for improved forecasting accuracy, though it also notes challenges related to Bitcoin's volatility and reliance on historical data.

Chapter 4

Algorithm Used

4.1 LSTM

LSTMs are an advanced type of recurrent neural network (RNN) architecture designed to handle sequential data with long-term dependencies more effectively than conventional RNNs.

4.1.1 The Problem with Conventional RNNs:

Regular RNNs struggle with storing information for extended periods. They fail to handle "long-term dependencies," where referencing information from a distant past is necessary to predict the current output. Additionally, RNNs suffer from issues like vanishing and exploding gradients during training.

4.1.2 Structure of an LSTM:

LSTMs address these problems by introducing a more sophisticated cell structure. Unlike RNNs, LSTMs have a gated unit (cell) that interacts with other layers to produce the output and maintain a cell state.
An LSTM cell consists of:

- **Cell State C-t:** Represents the memory of the network. It can store information over long periods.
- **Hidden State(h-t):** The output of the cell, which carries relevant information to the next time step.

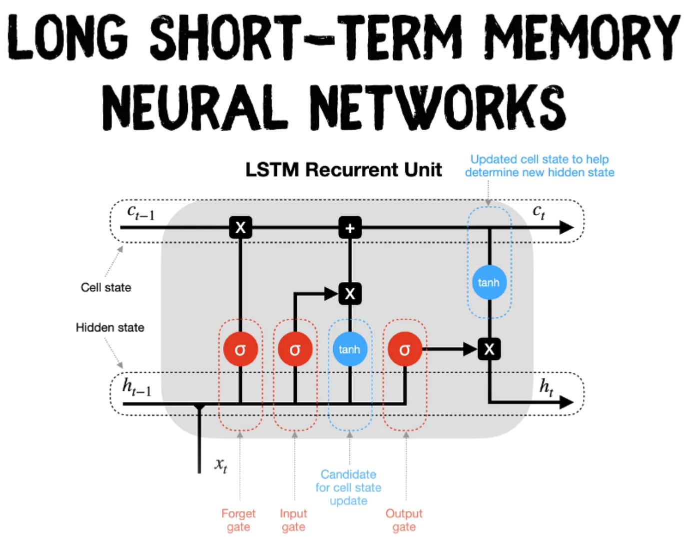


Figure 4.1: LSTM Model Architecture

- **Gates:** These control the flow of information within the cell.

  1. *Forget Gate:* Determines which parts of the cell state to retain or forget. It multiplies the old memory (C-(t-1)) by a vector close to 0 to forget irrelevant information.
  2. *Input Gate:* Updates the cell state with new information.
  3. *Output Gate:* Produces the final output (h-t) based on the cell state.
  4. *Remember Gate (optional):* Some LSTM variants include this gate to enhance memory retention.

4.1.3 Applications of LSTMs:

- Natural language processing (NLP) for text generation, sentiment analysis and language modeling.
- Time-series prediction (e.g., stock prices, weather forecasting).
- Speech recognition.
- Machine translation.
- Music composition.

4.2 ARIMAX

The ARIMAX model is an extension of the ARIMA model that includes exogenous variables (X), which are external factors influencing the time series data. It combines autoregressive (AR) terms, differencing (I) for stationarity and moving averages (MA) with these external variables to improve forecasting accuracy. The model is particularly useful when the data is influenced by factors outside of the time series itself, such as economic indicators or seasonal effects. By incorporating these exogenous variables, the ARIMAX model provides a more comprehensive analysis and can lead to more accurate predictions.Elaborated explanation of the ARIMAX model's components is in the following points:

- **Autoregressive (AR) Component:** This part of the model captures the relationship between an observation and a certain number of lagged observations. The autoregressive term is denoted by AR(p), where 'p' is the number of lag observations included in the model.

- **Integrated (I) Component:** This component is concerned with making the time series stationary by differencing the data (d) times. Stationarity is a crucial aspect of time series analysis, as it ensures that the properties of the series do not change over time.

- **Moving Average (MA) Component:** The MA part models the relationship between an observation and a residual error from a moving average model applied to lagged observations. It is denoted by MA(q), where 'q' is the number of lagged forecast errors in the prediction equation.

- **Exogenous Variables (X):** These are the external factors or variables that are not part of the time series but can influence its behavior. In the ARIMAX model, these variables are included to account for the impact they may have on the dependent variable. The inclusion of exogenous variables allows the model to capture effects that are not explained by the past values of the time series alone.

The ARIMAX model is particularly beneficial in scenarios where the time series is believed to be influenced by external factors, such as economic indicators, weather conditions, or policy changes. By including these exogenous variables, the ARIMAX model aims to provide a more accurate and comprehensive understanding of the underlying process that generates the time series data, leading to improved forecasting performance. The model is versatile and can be tailored to fit the specific characteristics of the data, making it a powerful tool for time series analysis in various fields such as finance, economics and environmental studies.

4.3 XGBOOST

XGBoost, which stands for eXtreme Gradient Boosting, is an advanced implementation of gradient boosting algorithms. It is designed for speed and performance and is widely used in machine learning for various supervised learning tasks such as regression and classification.Below points are detailed explanation of the XGBoost model:

Ensemble Learning:

XGBoost is an ensemble learning method that combines the predictions of several base learners, typically decision trees, to produce a more accurate and robust model.

Gradient Boosting Framework:

It operates within the gradient boosting framework, where new models are created that predict the residuals or errors of prior models and then added together to make the final prediction.

Regularization:

XGBoost includes regularization terms in its objective function to control over-fitting, which improves the model's performance on unseen data.

Handling Missing Values:

The algorithm has an efficient way of handling missing data. Unlike other algorithms, XGBoost doesn't require imputation and can handle missing values internally.

Tree Pruning:

XGBoost uses a depth-first approach and prunes trees backward. It splits up to the max depth specified and then starts pruning the tree backwards if the split does not improve the overall gain.

Parallel Processing:

It is parallelizable, meaning it can utilize multiple CPU cores for faster computation and can handle large datasets efficiently.

Flexibility:

XGBoost allows users to define custom optimization objectives and evaluation criteria, adding a layer of flexibility to the model.

Cross-validation:

The model can perform cross-validation at each iteration of the boosting process, allowing for the assessment of the quality of the model.

XGBoost has gained popularity in machine learning competitions like Kaggle for its performance and speed. It's a powerful tool that can be applied to a wide range of problems, from predictive modeling to complex regression and classification tasks. The model's ability to manage large datasets, along with its high degree of customizability, makes it a go-to algorithm for many data scientists.

4.4 Fb Prophet

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend and typically handles outliers well.

4.4.1 Core Functionality of Prophet:

Additive Model:

Prophet employs an additive model that is suitable for fitting non-linear trends which are influenced by seasonal patterns like yearly, weekly and daily cycles, as well as holiday effects.

Optimal for Seasonal Data:

It is highly effective for time series with pronounced seasonal fluctuations and benefits from having multiple seasons of historical data for analysis.

Robustness:

The model is resilient to missing data, shifts in trends and is typically good at managing outliers.

4.4.2 Key Features of Prophet:

Accuracy and Speed:

Known for generating reliable forecasts swiftly, Prophet is a go-to tool for planning and goal setting. It often exceeds the performance of other forecasting methods and uses Stan for quick model fitting.

Automation:

Prophet provides a fully automatic experience in forecasting, capable of handling messy data without requiring manual input.

Tunability:

The tool is flexible, allowing users to adjust forecasts. It has human-interpretable parameters that facilitate the integration of domain expertise into fine-tuning the predictions.

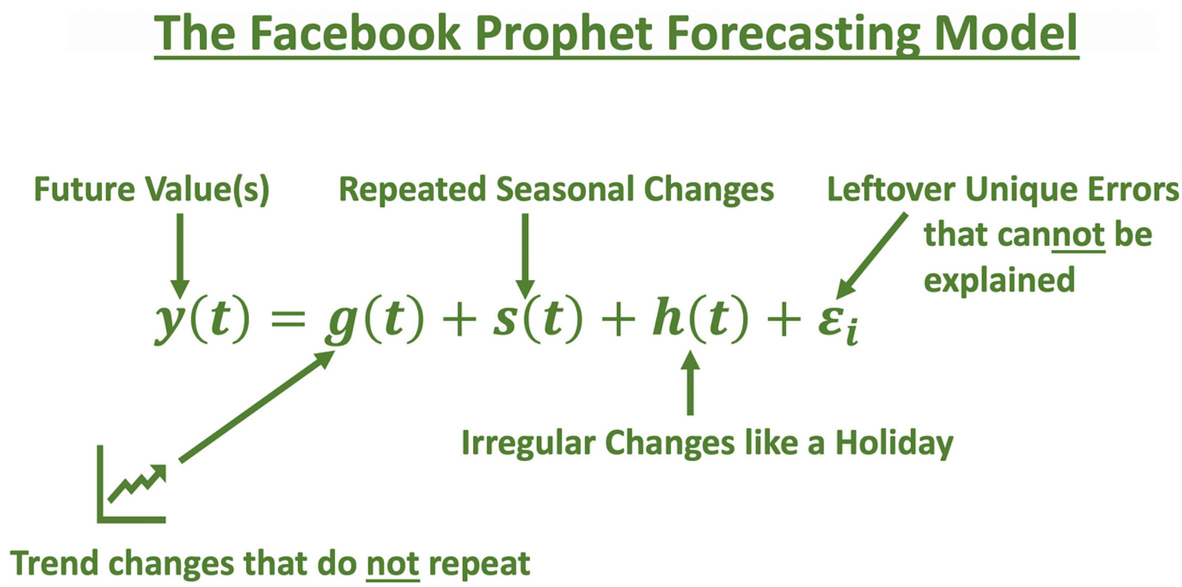


Figure 4.2: Fb Prophet Model equation

In essence, Prophet stands out as a user-friendly and efficient tool for forecasting time series data, appreciated for its adaptability and the control it offers to users for customizing forecasts with its holiday effect.

Chapter 5

Proposed Methodology

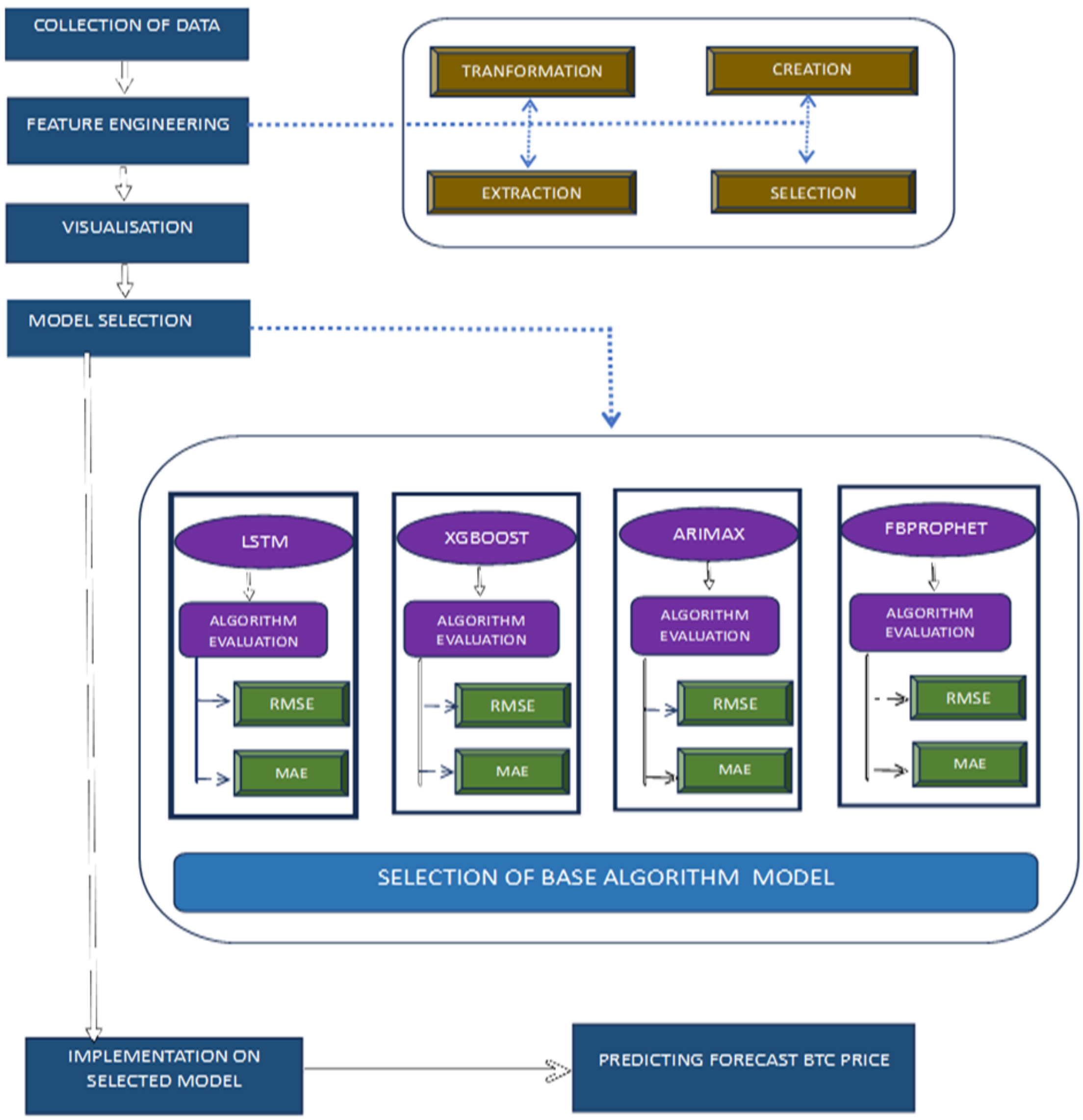


Figure 5.1: Proposed Methodology

Introduction:

This project focuses on developing a predictive model for Bitcoin (BTC) prices using various machine learning algorithms. The goal is to create a robust and accurate tool that can aid cryptocurrency investors in making informed decisions. The project involves multiple stages: data collection, feature engineering, visualization, model selection, algorithm evaluation and implementation of the selected model for price prediction.

Data Collection:

Data collection is the foundational step in building the predictive model. It involves gathering historical price data for Bitcoin, along with relevant economic indicators and market sentiment data. These data points are crucial for training and testing the predictive models.

Feature Engineering:

Feature engineering transforms raw data into meaningful features that can enhance the model's predictive power. This process includes:

- Transformation: Adjusting the data scale and distribution to improve model performance.
- Extraction: Deriving new features from the existing data, such as moving averages and volatility indices.
- Creation: Generating new features that capture complex relationships in the data.
- Selection: Choosing the most relevant features for the model, reducing dimensionality and improving efficiency.

Visualization:

Visualization helps in understanding the data and the relationships between different features. It involves creating plots and charts to identify trends, patterns and anomalies in the dataset. Effective visualization aids in better feature selection and model tuning.

Model Selection:

The model selection phase involves choosing the best algorithm(s) for predicting Bitcoin prices. The project evaluates four different models:

- Long Short-Term Memory (LSTM): A type of recurrent neural network suited for time-series prediction.
- Extreme Gradient Boosting (XGBoost): A powerful and efficient implementation of gradient boosting for supervised learning.
- AutoRegressive Integrated Moving Average with Exogenous variables (ARIMAX): A statistical model for time-series analysis.
- Facebook Prophet (FBProphet): A tool for producing high-quality forecasts for time-series data that has daily observations.

Algorithm Evaluation:

Each model undergoes rigorous evaluation using two primary metrics:

- Root Mean Square Error (RMSE): Measures the average magnitude of the errors between predicted and actual values.
- Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual values.

The evaluation helps in identifying the model that offers the best performance in terms of accuracy and reliability.

Selection of Base Algorithm Model:

Based on the evaluation metrics (RMSE and MAE), the best-performing algorithm is selected as the base model for implementation. This model is expected to provide the most accurate and reliable predictions for Bitcoin prices.

Implementation and Prediction:

Once the base model is selected, it is implemented to predict future Bitcoin prices. This involves integrating the model with real-time data feeds and deploying it in a user-friendly application. The final tool aims to provide timely and accurate price forecasts, assisting investors in making better investment decisions.

Conclusion:

This project outlines a comprehensive approach to predicting Bitcoin prices using advanced machine learning techniques. By incorporating data collection, feature engineering, visualization, model selection and rigorous evaluation, the project aims to develop a sophisticated predictive model. Future enhancements could include expanding the model to other cryptocurrencies, integrating more complex algorithms and developing an accessible platform for a wider audience.

Chapter 6

Requirement Analysis

6.1 Python

Python is a versatile and widely-used high-level programming language known for its simplicity and readability. Some key points about Python:

Ease of Use:

Python syntax is clean and easy to understand, making it ideal for beginners and experienced programmers alike.

Interpreted Language:

Python is an interpreted language, meaning it executes code line by line, which helps in debugging and dynamic typing.

Versatility:

Python supports multiple programming paradigms, including procedural, object-oriented and functional programming.

Extensive Libraries and Frameworks:

Python has a rich ecosystem of libraries and frameworks for various applications, such as:

- Web Development: Django, Flask

- Data Science and Machine Learning: NumPy, pandas, scikit-learn, TensorFlow, PyTorch
- Automation and Scripting: BeautifulSoup, Selenium

Cross-Platform:

Python runs on various platforms, including Windows, macOS and Linux, making it highly portable.

Community Support:

Python has a large and active community, providing extensive documentation, tutorials and third-party modules.

Integration Capabilities:

Python integrates well with other languages and technologies, such as C/C++ (via ctypes, Cython), Java (via Jython) and .NET (via IronPython).

Popular Use Cases

Python is used in web development, data analysis, machine learning, artificial intelligence, automation, scientific computing and more.
Overall, Python's combination of simplicity, versatility and powerful libraries makes it a popular choice for a wide range of programming tasks.

6.2 Google Colab

Google Colab, short for "Colaboratory," is a free, cloud-based Jupyter notebook environment provided by Google. It is designed for machine learning, data analysis and Python programming. Some of the key features and benefits of using Google Colab:

Cloud-Based:

- *No Installation Required:* Colab runs in the cloud, so you don't need to install any software locally.
- *Access from Anywhere:* You can access your notebooks from any device with internet access.

Free GPU and TPU:

- *Hardware Acceleratio*n: Colab provides free access to GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units), which can significantly speed up machine learning and deep learning tasks.

Integration with Google Drive:

- *Save and Load Files:* You can save your notebooks directly to Google Drive and load datasets from Drive.
- *Collaboration:* Share notebooks easily with others, allowing for real-time collaboration similar to Google Docs.

Pre-installed Libraries:

- *Extensive Libraries:* Colab comes with many popular Python libraries pre-installed, such as NumPy, pandas, TensorFlow, Keras, PyTorch, scikit-learn and more.

Easy Import of Libraries:

- *Additional Libraries:* You can easily install additional libraries using !pip install library-name.

Control Integration:

- *GitHub:* Colab allows you to open notebooks directly from GitHub repositories.

Free and Pro Versions:

- *Free Tier:* Offers a generous amount of compute resources for free.
- *Colab Pro:* Provides enhanced resources, including more powerful GPUs, longer runtimes and priority access for a monthly fee.

6.3 Jupyter Notebook

Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations and narrative text. It is widely used in data science, scientific computing and machine learning. Some of the key features and benefits are:

6.3.1 Key Feature

Interactive Code Execution:

- *Code Cells:* Write and execute code in an interactive manner.
- *Instant Feedback:* View the output of your code immediately below the code cell.

Visualizations:

- *Integrated Plotting:* Create and display visualizations using libraries like Matplotlib, Seaborn and Plotly directly within the notebook.

Versatile Language Support:

- *Multi-Language Support:* Jupyter supports over 40 programming languages, including Python, R, Julia and Scala.

Notebook Extensions:

- Custom Functionality: Extend the functionality of Jupyter Notebooks with various extensions, such as table of contents, code folding and spell checker.

Data Science and Machine Learning:

- Integration with Libraries: Seamlessly integrate with popular data science libraries like NumPy, pandas, TensorFlow and scikit-learn.

Sharing and Collaboration:

- *Export Options:* Export notebooks to various formats, including HTML, PDF and slides.
- *Version Control:* Use Git to manage and version control your notebooks.

6.3.2 Benefits

1. Ease of Use:Jupyter Notebooks provide an intuitive interface for writing and running code interactively.
2. Documentation and Visualization: Combine code with rich text and visualizations for comprehensive documentation and analysis.
3. Reproducibility: Maintain a complete record of code, results and explanations in a single document.

6.4 Kaggle

Kaggle is an online platform that hosts data science competitions and provides a vast repository of datasets, notebooks and educational resources. It is widely used by data scientists, machine learning practitioners and researchers to collaborate, learn and showcase their skills.Some of the key features and benefits of using Kaggle:

6.4.1 Key Features

Competitions:

- *Data Science Competitions:* Participate in competitions to solve real-world problems, ranging from predicting housing prices to detecting fraudulent transactions.

- *Prizes and Recognition:* Win monetary prizes, medals and recognition in the data science community.

Datasets:

- *Open Datasets:* Access a vast collection of publicly available datasets across various domains such as healthcare, finance and social sciences.
- *Data Exploration:* Use Kaggle's integrated tools to explore and analyze datasets directly on the platform.

Notebooks:

- *Jupyter Notebooks:* Create, share and collaborate on Jupyter Notebooks in the cloud without any setup.
- *Kernels:* Run code on Kaggle's servers, leveraging free GPU and TPU resources for computation-heavy tasks.

Community:

- *Discussion Forums:* Engage with a community of data scientists and machine learning enthusiasts to ask questions, share knowledge and collaborate on projects.
- *Competitions and Collaborations:* Work together with other participants in competitions and community projects.

Integration with Google Cloud:

- *Seamless Integration:* Easily integrate Kaggle with Google Cloud services for more advanced data storage and computing needs.

6.4.2 Benefits

1. Learning and Growth: Kaggle provides an excellent platform for learning new skills and improving existing ones through practice and

community interaction.

2. Networking: Connect with other data scientists and practitioners, gaining insights and feedback on your work.

3. Recognition: Build your profile and gain recognition in the data science community through competition participation and high-quality contributions.

4. Resources: Access a wealth of datasets and example projects to inspire and support your data science journey.

6.5 HTML and CSS

HTML (HyperText Markup Language) is used to structure content on the web. A basic HTML document starts with <!DOCTYPE html> to declare it as an HTML5 document. The document is enclosed in <html> tags. Inside, the ¡head¿ section contains metadata, such as the document's title (<title>) and links to stylesheets (<link>). The ¡body¿ section contains the content displayed on the webpage, including headings (<h1> to <h6>), paragraphs (<p>), navigation (<nav>), sections (<section>) and footers (<footer>)

CSS (Cascading Style Sheets) is used to style HTML elements. Styles are defined in a separate .css file and linked to the HTML document using the ¡link¿ tag. CSS rules consist of selectors and declarations. Selectors target HTML elements and declarations define the styles, which include properties and values. For example, body font-family: Arial; sets the font for the entire document to Arial.

In practice, an HTML file might include a header with navigation links, a main content area divided into sections and a footer. The corresponding CSS file would style these elements by setting fonts, colors, layout and other visual properties.

Together, HTML and CSS allow you to create well-structured, visually appealing web pages.

6.6 Streamlit

Streamlit is an open-source Python library that makes it easy to create and share beautiful, custom web apps for machine learning and data science. Here's a brief overview and an example to get you started.

6.6.1 Overview

Streamlit allows you to turn data scripts into interactive web applications in just a few lines of code. It's particularly useful for creating quick prototypes, dashboards and data exploration tools without needing extensive knowledge of web development.

6.6.2 Key Features

Widgets:

Streamlit provides a wide range of widgets to interact with your data, such as sliders, buttons and text inputs.

Layouts:

You can easily arrange your components using columns and expandable sections.

Real-time updates:

Streamlit apps update in real-time as you interact with them or modify the underlying code.

Integration with popular libraries:

Streamlit seamlessly integrates with popular Python libraries like Pandas, NumPy, Matplotlib, Plotly and more.

Streamlit makes it straightforward to create interactive, data-driven web applications with minimal code. It's a powerful tool for data scientists and developers who need to quickly prototype and share their work.

6.7 Pygwalker

6.7.1 Overview

Pygwalker (Python binding for Graphic Walker) is a tool that helps data scientists and analysts explore data interactively using Python. It provides a simple way to create interactive visualizations, enabling users to gain insights from their data without extensive coding.

6.7.2 Key Features

Interactive Data Exploration:

- Pygwalker allows users to interact with their data directly within Jupyter notebooks or other Python environments.
- Users can create, modify and customize various types of visualizations, such as bar charts, scatter plots, line charts and more.

Ease of Use:

- Designed to be user-friendly, Pygwalker requires minimal code to produce sophisticated visualizations.
- The API is intuitive, making it accessible even to those who may not have extensive programming or data visualization experience.

Integration with Popular Libraries:

- Pygwalker integrates seamlessly with popular Python data manipulation libraries such as Pandas.
- This integration allows users to leverage the power of these libraries for data cleaning and preprocessing before visualizing their data with Pygwalker.

Customizable Visualizations:

- Users can customize their visualizations through various configuration options.
- These customizations include setting colors, adjusting axis labels and adding titles, among other options.

6.7.3 Benefits

1. Rapid Prototyping: Quickly create visualizations to understand data and identify trends.
2. Exploratory Data Analysis: Easily explore data through interactive features to uncover insights.
3. Presentation Ready: Generate visualizations that are ready to be shared in reports or presentations.

6.8 Summary

To get started with the latest version and environment setup, it's essential to ensure that Python is installed on our system, as it forms the backbone of our analytical tasks. For cloud-based development, Google Colab offers a free and powerful environment, providing an easy setup with pre-installed libraries. Additionally, Jupyter Notebook is a vital tool due to its interactive nature, offering features that facilitate data visualization and sharing, significantly benefiting the data science workflow. Kaggle complements these tools by providing key features like datasets and competitions, enhancing practical learning and application. For web development and dashboard creation, proficiency in HTML and CSS is required, while Streamlit simplifies the creation of interactive web apps for data science projects. Lastly, Pygwalker is highlighted for its capability to create interactive visualizations, enhancing data exploration and analysis.

Chapter 7

Design and Implementation

7.1 Overview

Our Project is divided into two task, first one is to work on proposed methodology and find one of the best algorithm among the given above and further proceed to second section to build and dashboard of cryptocurrency where we will forecast the data produced with the selected algorithm to achieve our objective of showing the upper, lower and predicted value with the help of different visualization tools.

7.1.1 Adressing some common question

Why We Selected the Daily Dataset and Closing Price for Analysis ?

Lag plots, as illustrated in Figure:7.1 are a tool used to check for autocorrelation in time series data, helping to identify patterns and dependencies across different time intervals. In our project, we chose to use a daily dataset due to several reasons. Firstly, daily data provides a balance between granularity and noise reduction. Minute and hourly data can be overly granular, containing a lot of noise which can obscure meaningful patterns. Secondly, daily patterns in cryptocurrency markets are often more stable and reflective of broader trends compared to intra-day fluctuations. Additionally, daily data is more computationally efficient to process

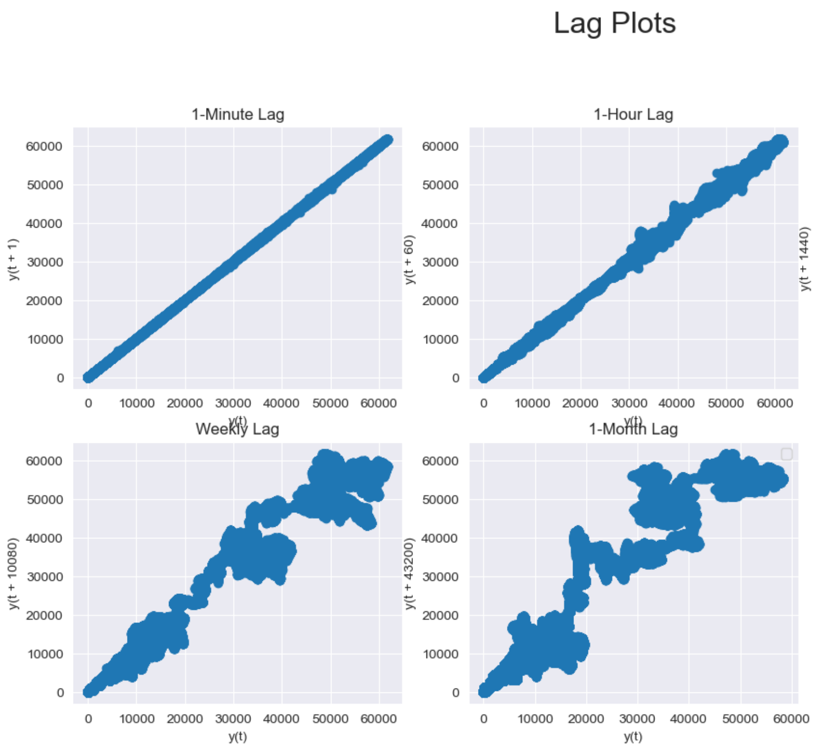


Figure 7.1: Lag Plot

and analyze compared to minute or hourly data, especially for long-term studies.

The closing price of Bitcoin is selected for our analysis because it represents the final consensus of investors and encapsulates the day's overall market sentiment. It is a comprehensive metric that includes all trading activity up until the market closes, thus providing a reliable indicator of market trends and investor behavior. While using closing prices means potentially missing out on intra-day volatility, it simplifies the analysis by focusing on a single, well-understood data point. Depending on the specific requirements of our analysis, other price points such as open, high, or low prices could also be considered, but the closing price is often the most informative for end-of-day reflections.

7.2 Implementing Proposed Methodology

7.2.1 Collection of data:

We have taken the data from yahoo finance and Kaggle as it is very authentic and accurate data on which we can rely our research. Yahoo finance contains the up to date data of various cryptocurrency. We have taken the dataset in different format like on hourly, weekly, Monthly and yearly basis.

7.2.2 Feature Engineering:

Feature engineering involves modifying or creating new features from existing data to improve machine learning model performance. This includes selecting relevant features, scaling them to similar ranges, encoding categorical variables, handling missing values, creating new features, reducing dimensionality and selecting the most informative features. Each step aims to enhance the quality and relevance of the features used by the model, ultimately leading to better predictions and insights.

- Creation-Feature creation involves generating new features from existing ones or external data sources like Kaggle and yahoo finance.
- Transformation- Feature transformation involves altering the distribution or scale of features to meet the assumptions of the machine learning algorithms or to improve model performance.Feature like liner interplotation is used.
- Extraction- Feature extraction involves deriving new features from existing ones through techniques like rolling feature.
- Selection- Feature selection involves choosing the most relevant features for modeling while discarding irrelevant or redundant ones. It varies along with the algorithm requirement time to time.

7.2.3 Visualisation:

Visualization in Bitcoin price prediction research involves using charts, graphs and other visual tools to analyze historical price data, identify trends, correlations and patterns, explore relationships between variables, evaluate model performance, interpret predictions and understand market sentiment.

In time series analysis, autocorrelation and partial autocorrelation plots play a crucial role in understanding the underlying structure of a time series.

Autocorrelation (ACF):

- *Definition:* Autocorrelation measures the correlation between a time series and a delayed version of itself. It quantifies how related a value at a given time is to its previous values (lags).

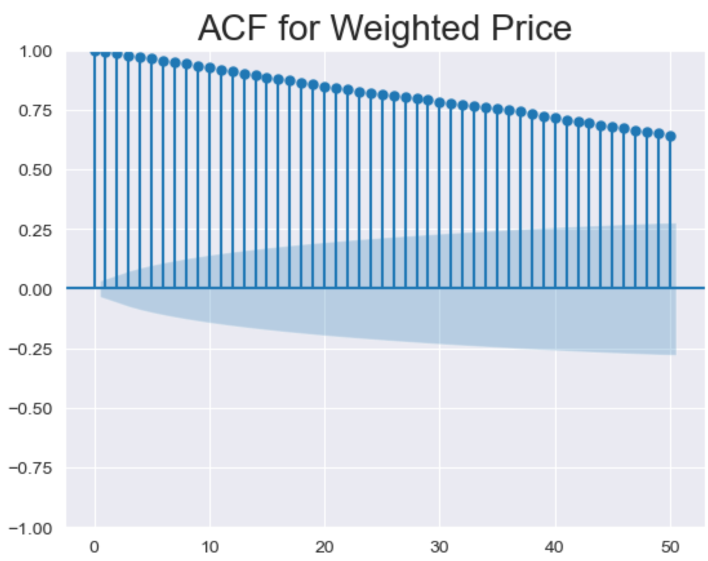


Figure 7.2: ACF Plot

- *ACF Plot:* The ACF plot as shown in Figure:7.2 shows the correlation coefficient against different lags. A lag corresponds to a certain point in time after which we observe the first value in the time series. The blue bars on the ACF plot represent error bands and anything within these bars is not statistically significant. A coefficient of 0 means no relationship between the variables.

- *Interpretation:* If the ACF plot exponentially decays after a lag spike (e.g., at lag 2), it suggests that the time series may be stationary1. However, it's essential to confirm this with statistical tests.

Partial Autocorrelation (PACF):

- *Definition:* The PACF captures the correlation between two variables after controlling for the effects of other variables. It helps identify the direct relationship between a value at a specific lag and the current value, excluding the influence of intermediate lags.

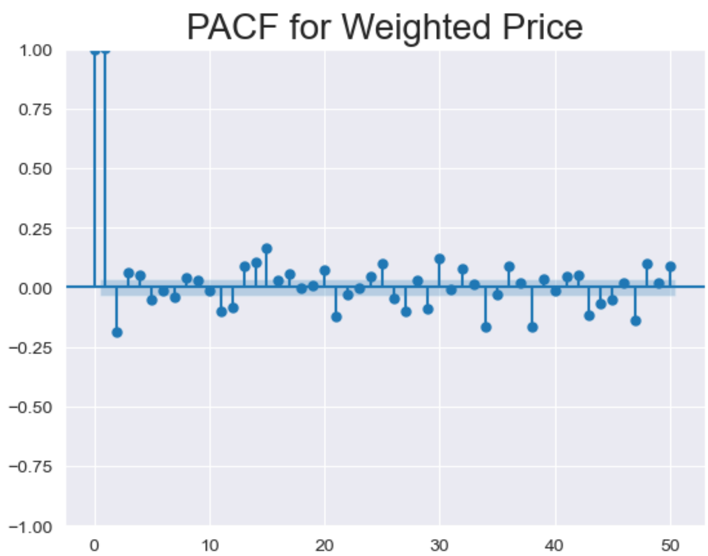


Figure 7.3: PACF Plot

- *PACF Plot:* The PACF plot shown in Figure:7.3 the partial correlation coefficient against different lags. It often cuts off after lag 0, indicating the direct relationship between the current value and its immediate lag.

- *Interpretation:* For instance, if the PACF plot cuts off after lag 1, it suggests an autoregressive (AR) process with lag 12.

Stationarity and ACF/PACF:

- *Stationarity:* A stationary time series has constant mean, variance and autocorrelation over time. ACF and PACF plots can help assess stationarity.
- *ACF and PACF for Stationarity:* In a stationary time series, the ACF and PACF values should be close to zero and hover around 0 at different lags3. If they exhibit this behavior, it supports the hypothesis of stationarity.
- *Statistical Tests:* To confirm stationarity, perform tests like the Augmented Dickey-Fuller (ADF) test we have performed.

7.2.4 Model Selection:

In the model selection phase of a Bitcoin price prediction project, We choose and develop predictive models using machine learning algorithms. We preprocess features, train models, evaluate performance, interpret results, refine models as needed and ultimately select the best-performing model for price forecasting.

We consider various machine learning algorithms such as linear regression, support vector machines (SVM),ARIMAX,FBPROPHET,XGBOOST, decision trees, random forests, gradient boosting machines (GBM) and deep learning models like recurrent neural networks (RNNs) or long short-term memory networks (LSTMs).

LSTM-

Long Short-Term Memory (LSTM) algorithms are popular for their ability to capture temporal dependencies and patterns in time series data, making them well-suited for predicting Bitcoin prices.
LSTM is a type of recurrent neural network (RNN) designed to address the vanishing gradient problem in traditional RNNs, which struggle with

retaining information over long sequences. LSTM incorporates memory cells and gating mechanisms to selectively remember or forget information over time, allowing it to capture long-term dependencies in sequential data.

Layer (type)	**Output Shape**	**Param #**
lstm (LSTM)	(None, 100, 50)	10,400
dropout (Dropout)	(None, 100, 50)	0
lstm_1 (LSTM)	(None, 100, 50)	20,200
dropout_1 (Dropout)	(None, 100, 50)	0
lstm_2 (LSTM)	(None, 100, 50)	20,200
dropout_2 (Dropout)	(None, 100, 50)	0
lstm_3 (LSTM)	(None, 50)	20,200
dropout_3 (Dropout)	(None, 50)	0
dense (Dense)	(None, 1)	51
Total params: 71,051 (277.54 KB)		
Trainable params: 71,051 (277.54 KB)		
Non-trainable params: 0 (0.00 B)		

Table 7.1: Model: "LSTM Summary"

- *LSTM Layers:*These are recurrent neural network (RNN) layers designed to capture long-term dependencies in sequential data. They are particularly useful for tasks like time series prediction and natural language processing.We have used 4 layers.

- *Dropout Layers:* Positioned between LSTM layers, dropout layers help prevent overfitting by randomly deactivating a fraction of input units during training. This improves the model's robustness.

- *Dense Layer:* The final layer in the model, which produces the output. It is fully connected, meaning each neuron receives input from all neurons in the previous layer.

- *Total Parameters:* The model has a total of 71,051 trainable parameters, indicating its complexity and capacity to learn from data.

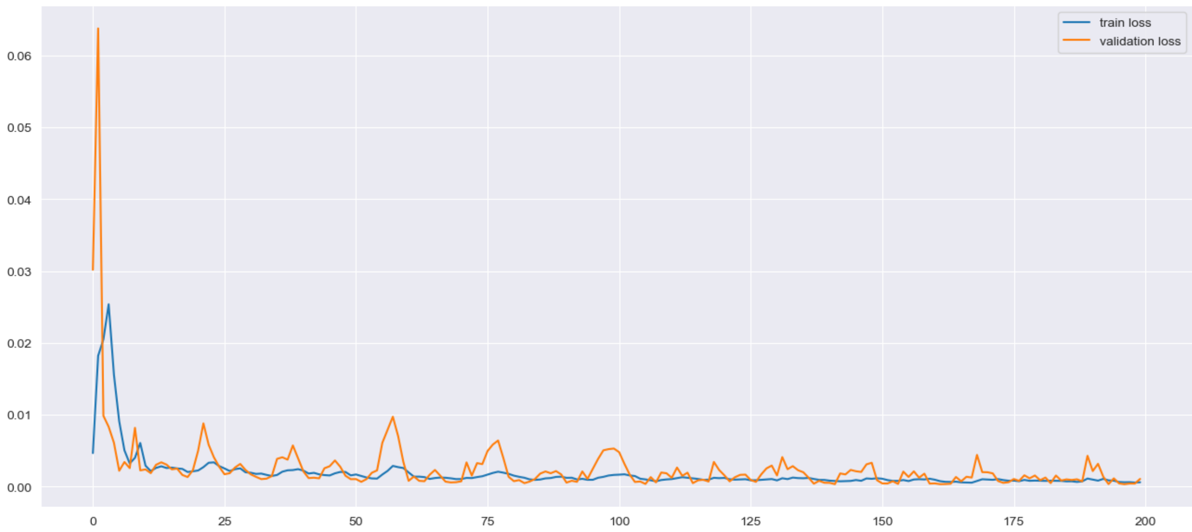


Figure 7.4: Validation Loss

- Training Loss:

  1. The training loss assesses how well a deep learning model fits the training data. It quantifies the error produced by the model on the training set.
  2. It is measured after each batch during training and is usually visualized by plotting a curve of the training loss.

- Validation Loss:

  1. The validation loss evaluates the model's performance on a separate validation set. This set is distinct from the training data and is used to validate the model's generalization ability.
  2. Similar to the training loss, the validation loss is calculated by summing up the errors for each example in the validation set.
  3. It is measured after each epoch (complete pass through the training data) and provides insights into whether the model needs further tuning or adjustments.
  4. A rising validation loss may indicate overfitting, where the model performs well on the training data but poorly on unseen data.

XGBOOST-

XGBoost (Extreme Gradient Boosting) is a popular machine learning algorithm used for supervised learning tasks, including regression and classification.
XGBoost is a favored algorithm for building predictive models due to its ability to handle diverse data types efficiently. We preprocess features like historical prices, market sentiment and trading volume, then train the model to predict future Bitcoin prices. XGBoost's strength lies in its fast training speed, accurate predictions and feature importance analysis, allowing researchers to identify key factors driving price movements and make informed forecasts for trading or investment decisions.

ARIMAX-

We utilize the ARIMAX algorithm to forecast Bitcoin prices while considering external factors that could influence price movements. By incorporating exogenous variables into the forecasting process, ARIMAX enables us to capture the impact of external factors on Bitcoin price movements, potentially leading to more accurate predictions compared to traditional time series models that only consider historical price data.

FB PROPHET-

The Facebook Prophet algorithm is a popular forecasting tool due to its simplicity and effectiveness. the Facebook Prophet algorithm is a valuable tool for Bitcoin price prediction research due to its ability to handle complex seasonal patterns, non-linear trends and uncertainty estimation. Its ease of use and flexibility make it a popular choice among researchers and data scientists in the cryptocurrency domain.

We calculate the RMSE and MAE for all the model made from the algorithm and out of all four which gave the minimum RMSE and MAE for will be selected as the base model for our research.

7.2.5 Implementation of Selected Model:

After the selection of base model we will take the recent dataset from year 2015 because if we take all the available dataset it will create a distortion in our finding, as the past there is multiple variation in the price of bitcoin because of various reasons like sudden increase in the interest of investor and multiple crashes in the stock market and many more reasons. After the collection of the dataset we will further do the feature engineering, visualization and model building(splitting for training and testing).

7.2.6 Forecast Prediction of Bitcoin Price:

We will forecast the the BTC price on the latest model and we will transfer the forecasted data to csv file and later in to excel or google sheets. Because it will make it easy for the normal people to visualize the forecasted result.

We will set the upper, lower limit and price predicted value of the bitcoin so that the people can judge the fluctuation in the price of the bitcoin. The upper limit can reduce the risk factor when the bitcoin price excedes it. The lower limit can reduce the sudden crash in the financial market when it excedes it limit.

7.3 Design of Cryptocurrency Dashboard

Libraray and Module used:

1. **streamlit:** For creating the web app and its interactive features.
2. **streamlit.components.v1:** To integrate custom HTML components.
3. **datetime:** To handle date and time operations.
4. **pandas:** For data manipulation and analysis.
5. **yfinance:** To fetch historical market data from Yahoo Finance.

6. **prophet:** For making forecasts using the Prophet time series forecasting procedure.
7. **prophet.plot:** To plot the forecasts with Plotly.
8. **plotly.graph objs:** For creating interactive plots
9. **pygwalker:** A library for data exploration and visualization.
10. **pygwalker.api.streamlit:** To integrate PyGWalker with Streamlit.
11. **streamlit echarts**: To create charts using Echarts of streamlit.

- *Libraries and Setup:* The code imports necessary libraries, sets up the Streamlit page configuration and defines a function to create download links for data.
-*Data and Parameters:* It defines a start date and uses the current date to fetch historical cryptocurrency data from Yahoo Finance. The sidebar allows users to select a cryptocurrency and the prediction frequency.
-*Data Loading and Display:* The 'load data' function downloads the data, which is then displayed as a raw candlestick chart.
- *Forecasting:* The application uses Prophet to forecast future prices based on the historical closing prices. It allows users to view the forecasted data and plots.
- *Download Links:* Users can download the complete forecast data, future forecast data, or an optimized version of the forecast data.
- *Visualization*: The application provides a section for visualizing the forecast using different chart types, such as line, bar, or candlestick charts, with the help of ECharts.
- *User Data Visualization:* Users can upload their own CSV or Excel files to visualize and explore their datasets within the application.

Overall, the application is designed to provide an interactive experience for users to predict and visualize cryptocurrency prices, with the ability to analyze their own data through Pyqwalker.

Chapter 8

Result and Finding

In our Bitcoin price prediction project, we compared ARIMAX, LSTM, XGBoost and FB Prophet models. The FB Prophet model, developed by Facebook, has shown superior performance in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), common metrics for evaluating forecasting models. FB Prophet is an additive model that fits non-linear trends with yearly, weekly and daily seasonality, plus holiday effects. It's particularly effective with time series that have strong seasonal effects and several seasons of historical data. FB Prophet has accurately captured the trend and seasonality of Bitcoin prices, including the peaks in late 2017/early 2018 and early 2021. This accuracy in capturing Bitcoin's volatility is likely why FB Prophet has lower RMSE and MAE values compared to the other models. Raw plot generated by fb prophet model where Black Dots are the initial historical data points or observed value over time. Dark Blue Line is the predicted forecasted values and Light Blue Shaded Area represents the uncertainty intervals or confidence interval around the forecast. This type of visualization is interesting and relevant as it provides insights into trends, patterns, or predictions over time which can be crucial for analysis in various fields such as finance, science, or economics.

At last, the selection of FB Prophet for our project seems justified given its superior performance. Hence, we further used it for prediction with different visualisation tools which give us significantly help in understanding

Figure 8.1: Fb Prophet Predicted

the price fluctuation better than the previous traditional methods. Some of the plots are shown below with the upper and lower limit generated by our models and thus helps in knowing predicted value and the limits to invest in the market. Also, the dataset provided with future price can be used as the raw dataset for advanced visualisation tools like PyGwalker, MS excel.

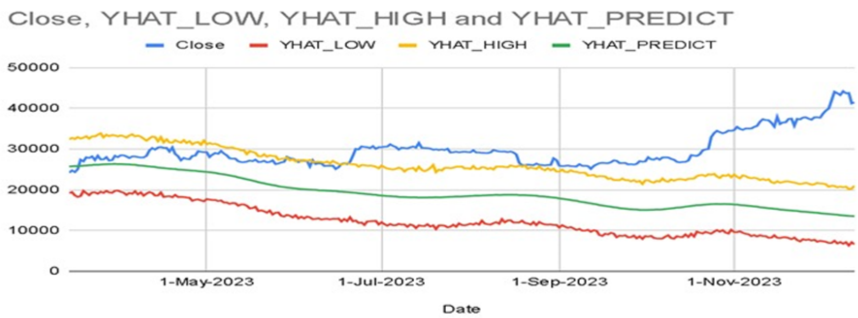


Figure 8.2: Predicted line graph

The plot Figure:8.3 depicts the fact that fb prophet outperforms other 3 algorithm significantly.

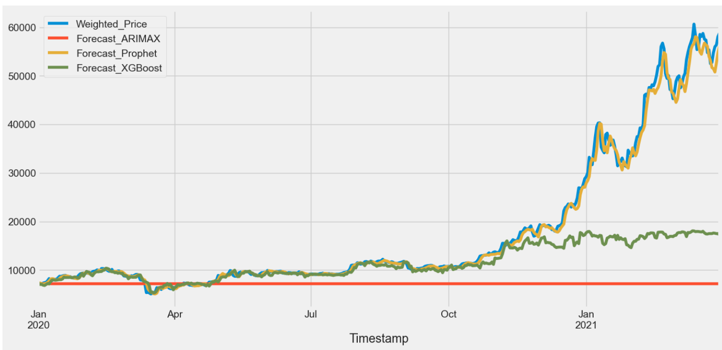


Figure 8.3: Algorithm comparison plot

Chapter 9

Conclusion

The research project on Bitcoin price prediction has provided significant insights into the volatile nature of cryptocurrency markets, particularly Bitcoin. By leveraging various machine learning algorithms such as LSTM, ARIMAX, XGBOOST and FB Prophet, the study has developed a robust predictive model for Bitcoin prices.

The model's ability to set upper and lower limits for price fluctuations is a novel approach that can aid investors in risk management and decision-making processes. This feature can mitigate the risk factor when the Bitcoin price exceeds the upper limit and prevent a sudden market crash when it exceeds the lower limit.

Moreover, the research has shed light on Bitcoin's fundamental aspects, its internal architecture and the multitude of factors influencing its price dynamics. This comprehensive analysis empowers stakeholders with actionable intelligence for navigating the volatile cryptocurrency markets.

The primary objective of developing a predictive model for Bitcoin prices has been achieved. The model caters to diverse needs such as investment decision-making, risk management, algorithmic trading, portfolio optimization, market analysis, financial planning, cryptocurrency exchange operations, regulatory compliance and risk assessment in trading.

The research project has successfully developed a robust predictive model for Bitcoin price prediction using machine learning algorithms like LSTM, ARIMAX, XGBOOST and FB Prophet. This model, which sets upper and lower limits for price fluctuations, aids in risk management and decision-making. It offers insights into Bitcoin's volatility, its fundamental aspects and the factors influencing its price. Catering to diverse needs such as investment decision-making, risk management and more, the model serves as a reliable guide for stakeholders in the cryptocurrency domain. The research underscores the potential of machine learning in forecasting Bitcoin prices and emphasizes the importance of continuous innovation in this rapidly evolving field.

Chapter 10

Future Scope

The future scope of this project offers several promising directions for expansion and enhancement:

Broadening Cryptocurrency Coverage:

- *Predicting Prices of Other Cryptocurrencies:* Extend the model to include predictions for a wider range of cryptocurrencies beyond Bitcoin, such as Ethereum, Ripple, Litecoin and more. This would create a more comprehensive tool for investors seeking to diversify their portfolios.

Incorporating Advanced Machine Learning Techniques:

- *Enhanced Algorithms*: Integrate more sophisticated machine learning algorithms, such as deep learning models, ensemble methods, or reinforcement learning, to improve the accuracy and robustness of price predictions.

- *Real-Time Data Integration:* Incorporate real-time data feeds to ensure that the model's predictions are as up-to-date and relevant as possible. This would involve streaming data on market prices, trading volumes and other relevant metrics.

Analyzing the Impact of Global Events:

- *Incorporating Economic Indicators:* Examine the influence of global economic factors such as oil and gold prices on cryptocurrency markets. These factors often correlate with market movements and can provide additional predictive power.

- *News and Sentiment Analysis:* Utilize Natural Language Processing (NLP) techniques to analyze news articles, social media posts and other text sources. This can help in understanding the sentiment and public perception around cryptocurrencies, which significantly impact market trends.

User-Friendly Application Development:

- *Developing an Accessible Platform:* Create a user-friendly application or web platform that leverages the predictive model. This platform should be designed with intuitive interfaces and features that cater to both novice and experienced investors.

- *Democratizing Investment Tools:* By making sophisticated prediction tools accessible to a broader audience, the project could contribute to the democratization of cryptocurrency investment, fostering greater market participation and potentially enhancing market stability.

Educational Resources and Community Engagement:

- *Creating Educational Content:* Develop educational resources to help users understand how to use the prediction tools effectively and make informed investment decisions.

- *Engaging with the Community:* Establish a community forum or support network where users can share insights, provide feedback and collaborate on improving the predictive models.

Chapter 11

Snapshots

Forecast and Parameter selection:with upper,lower and predicted price of top 10 cryptocurrency(Bitcoin,Binance) illustrate Figure:11.1 with a user friendly feature of selecting on the frequency of daily, weekly,monthly ,yearly shown in Figure:11.2 upto 10 years. Also a candlestick chart is shown from the 2015 to till date to analyze the previous trends if required.

Download: Also we have given the feature to download the forecasted data with initial dataset from 2015 to till date or else only individual forecasted data with and without filter of separate limits shown in below Figure:11.4. It enhances the user easiness so to implement in their visualization tool like excel, google sheet and print the dataset to analyse it futuristic. For the new users who are not well known to that all external tool we have already given plot in one of the most line chart which itself with clearly shown upper ,lower and predicted value for the user selected parameter(Cryptocurrency,frequency and time).

USER DATA VISUALIZATION: After all if user still wants to plot and check the downloaded dataset from our website or else from external sources (upto 200MB which can be further extended to 100GB) , we have also given one of the currently advanced visualization tool (PyG walker) which encapsulates with various chart and plot to analyze it .

Traffic Report: Various people have visited our website in initial phase of hosting which clearly shown in analytics report Figure:11.3

Select dataset for prediction

BTC-USD

Select frequency for prediction

Daily

Refresh Data

Download Section

Click on the link below to download the data.

Choose a file to download:

- Complete Forecast Data
- Future Forecast Data
- Optimized Forecast Data

Download Future Forecast Data

Cryptocurrencies Price Predictor

FORECASTING USER DATA VISUALIZATIC

Our innovative prediction tool empowers users to chart the potential trajectory of top cryptocurren monthly, and yearly basis, extending up to a visionary 10-year outlook

Select a cryptocurrency and the frequency of prediction from the dropdown menus below.

Select Parameters

Number of periods for prediction:

213

1

```
Loading data... done!
```

Raw data

	Date	Open	High	Low	Close	Adj Close	Volume
3,417	2024-05-10 00:00:00	63,055.1914	63,446.7422	60,208.7813	60,792.7773	60,792.7773	27,804,9
3,418	2024-05-11 00:00:00	60,793.3555	61,451.1523	60,492.625	60,793.7109	60,793.7109	13,842,2
3,419	2024-05-12 00:00:00	60,793.5039	61,818.1563	60,632.6016	61,448.3945	61,448.3945	13,800,4
3,420	2024-05-13 00:00:00	61,451.2188	63,422.6602	60,769.8398	62,901.4492	62,901.4492	27,889,1
3,421	2024-05-14 00:00:00	62,900.7734	63,092.125	61,123.7656	61,552.7891	61,552.7891	28,186,2

Raw Data Candlestick Chart

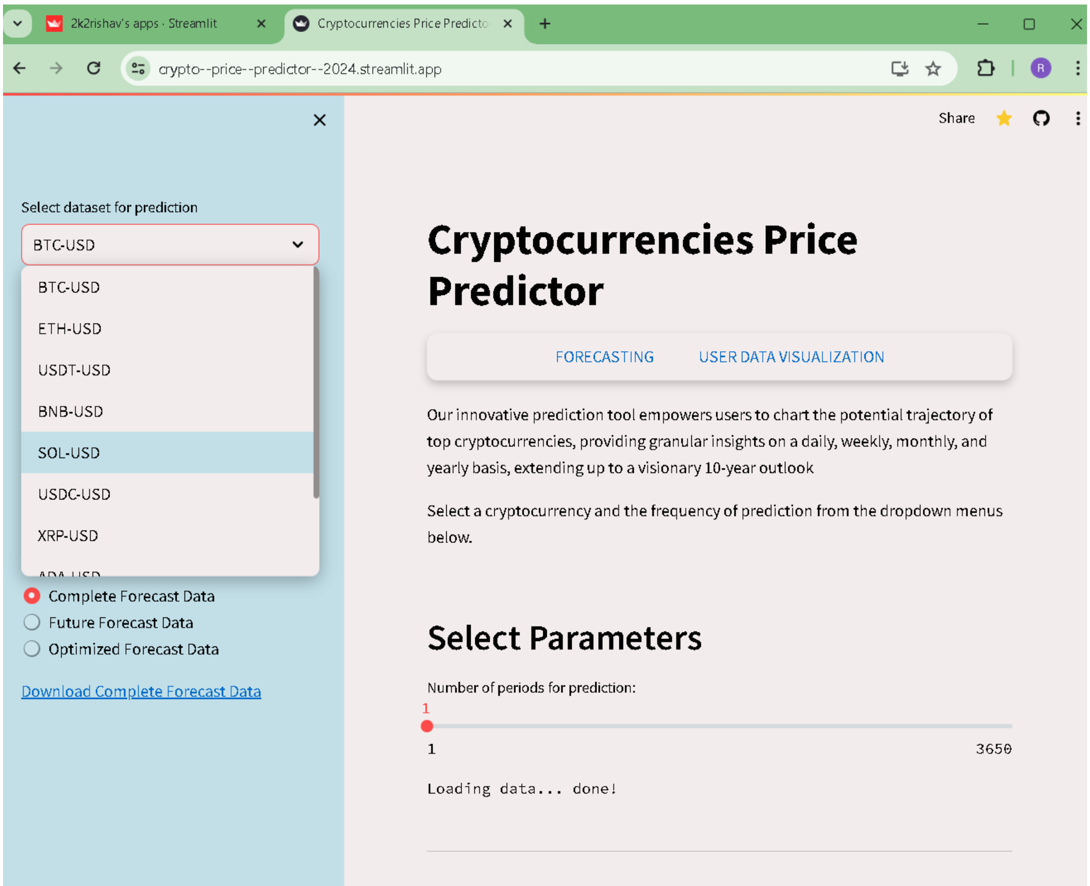


Figure 11.1: Currency Selection

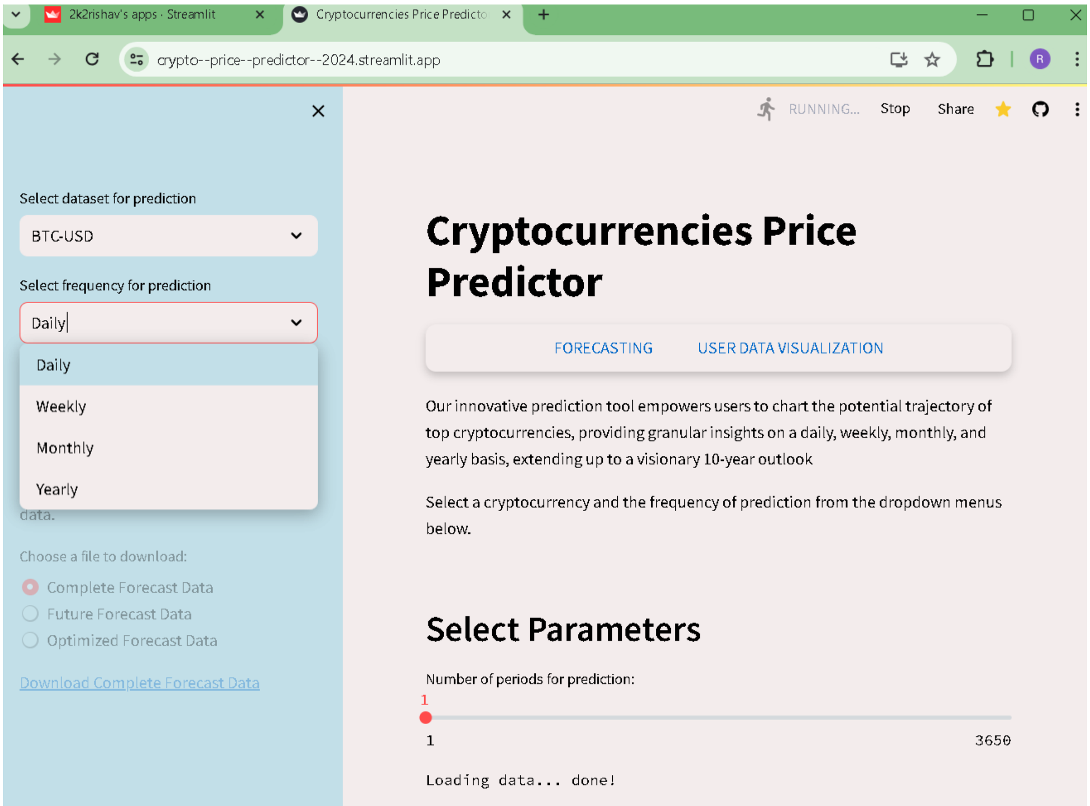


Figure 11.2: Frequency Selection

Cryptocurrencies Price Predictor

FORECASTING USER DATA VISUALIZATION

Our innovative prediction tool empowers users to chart the potential trajectory of top cryptocurrencies, providing granular insights on a daily, weekly, 10-year outlook

Select a cryptocurrency and the frequency of prediction from the dropdown menus below.

Select Parameters

Number of periods for prediction:

213

1

```
Loading data... done!
```

Raw data

	Date	Open	High	Low	Close	Adj Close	Volume
3,417	2024-05-10 00:00:00	63,055.1914	63,446.7422	60,208.7813	60,792.7773	60,792.7773	27,804,954,694
3,418	2024-05-11 00:00:00	60,793.3555	61,451.1523	60,492.625	60,793.7109	60,793.7109	13,842,272,968
3,419	2024-05-12 00:00:00	60,793.5039	61,818.1563	60,632.6016	61,448.3945	61,448.3945	13,800,459,405
3,420	2024-05-13 00:00:00	61,451.2188	63,422.6602	60,769.8398	62,901.4492	62,901.4492	27,889,181,179
3,421	2024-05-14 00:00:00	62,900.7734	63,092.125	61,123.7656	61,552.7891	61,552.7891	28,186,271,527

Raw Data Candlestick Chart

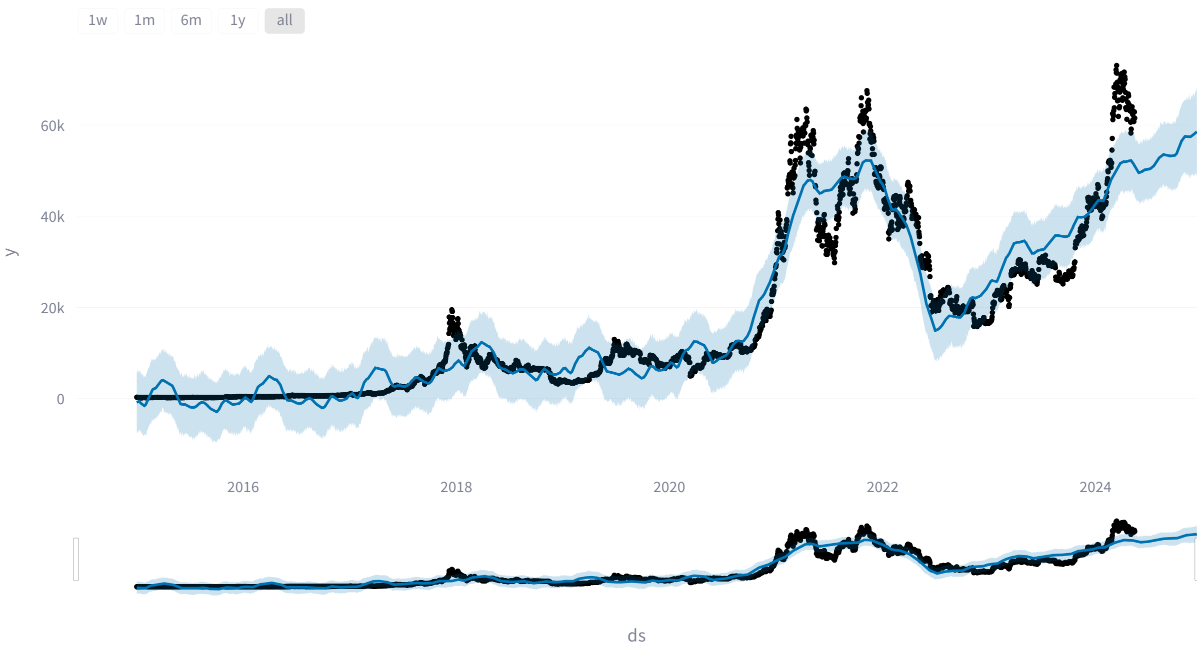
FORECASTING

The Forecasting section of the application leverages to predict future cryptocurrency prices based on historical data. Users can select from a range of
daily, weekly, monthly, or yearly. The forecast includes a plot of predicted values along with upper and lower limits, providing a visual representation o

Forecasted Data

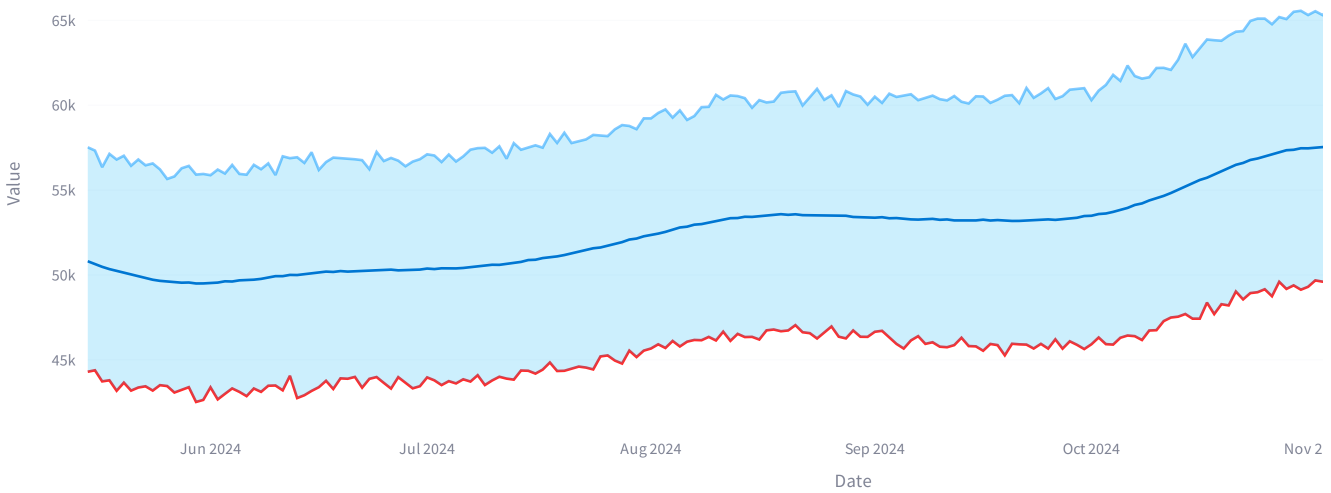
	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive_terms_lower	additive_terms_upp
3,630	2024-12-09 00:00:00	59,682.3315	49,062.0412	67,306.1128	53,751.7535	65,061.1181	-1,318.3086	-1,318.3086	-1,318.30
3,631	2024-12-10 00:00:00	59,730.7341	49,324.551	67,234.487	53,753.2165	65,227.603	-1,357.5194	-1,357.5194	-1,357.51
3,632	2024-12-11 00:00:00	59,779.1367	48,865.7891	68,245.0364	53,754.6796	65,276.8812	-1,318.9837	-1,318.9837	-1,318.98
3,633	2024-12-12 00:00:00	59,827.5393	49,387.1756	67,665.701	53,756.1426	65,365.5738	-1,355.9841	-1,355.9841	-1,355.98
3,634	2024-12-13 00:00:00	59,875.9419	49,272.9561	67,805.7088	53,752.7266	65,475.6714	-1,337.8666	-1,337.8666	-1,337.86

Forecasted plot for 213 D

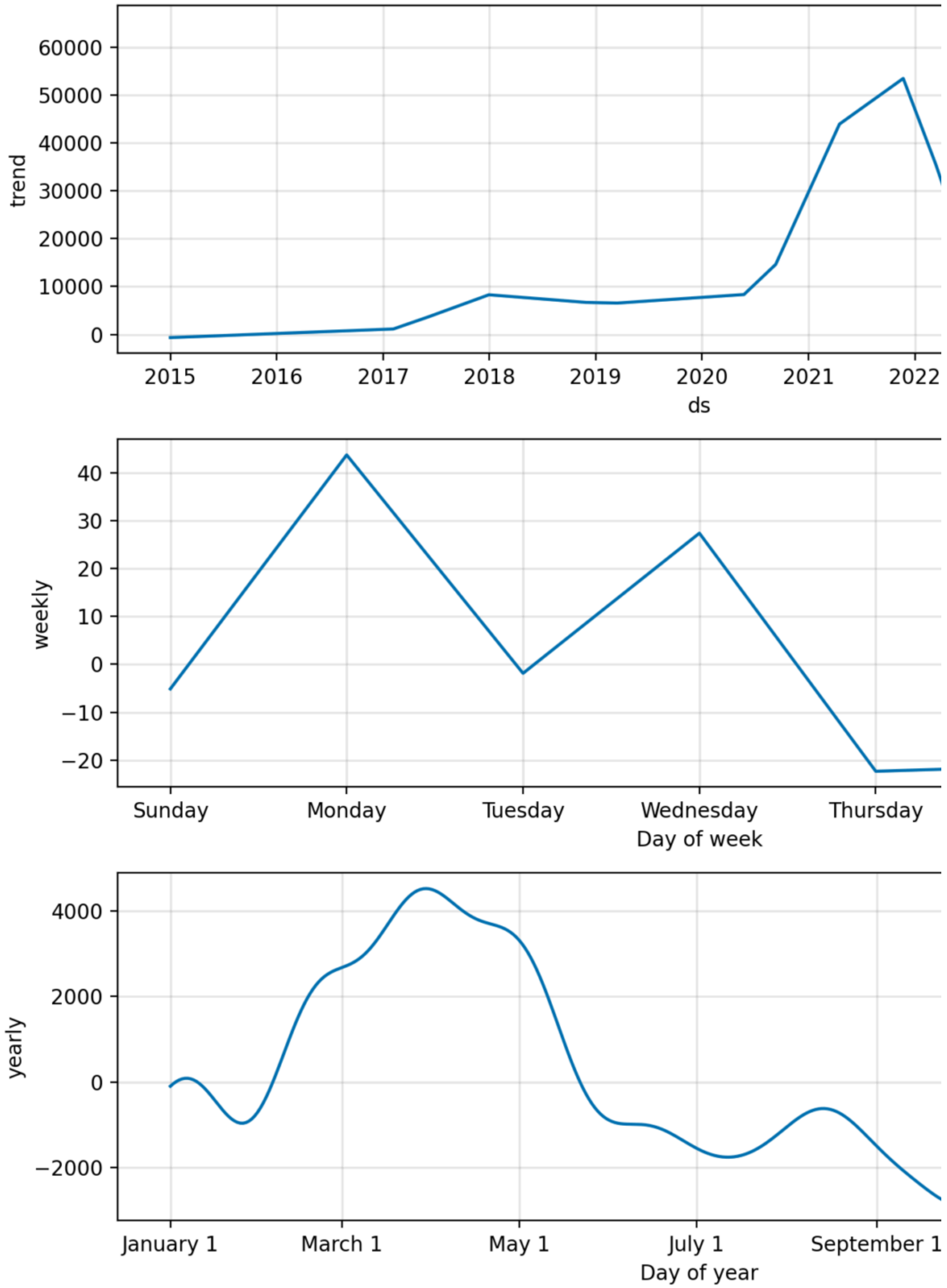


Forecasted plot for 213 D from Today

Forecasted line chart



FORECASTED COMPONENTS



USER DATA VISUALIZATION

This section offers users the ability to upload their own datasets in CSV or Excel format. Once uploaded, it will be allowing users to delve into their data
to analyze their personal or business-related data within the same application.

Upload your CSV or Excel file. (200MB max)

Drag and drop file here

Limit 200MB per file • CSV, XLSX

forecast (6).csv 1.1MB

Explore your dataset

Data | Visualization

Chart 1 ⋮ + New

Field List

- ds
- Measure names

- # Unnamed: 0
- # trend
- # yhat_lower
- # yhat_upper
- # trend_lower
- # trend_upper
- # additive_terms
- # additive_terms_lower
- # additive_terms_upper
- # weekly
- # weekly_lower
- # weekly_upper
- # yearly
- # yearly_lower
- # yearly_upper
- # multiplicative_terms
- # multiplicative_terms_lower
- # multiplicative_terms_upper
- # yhat
- # Row count
- # Measure values

Filters

Color

Drop Field Here

Opacity

Drop Field Here

Size

Drop Field Here

Shape

Drop Field Here

Details

Drop Field Here

X-Axis

Y-Axis

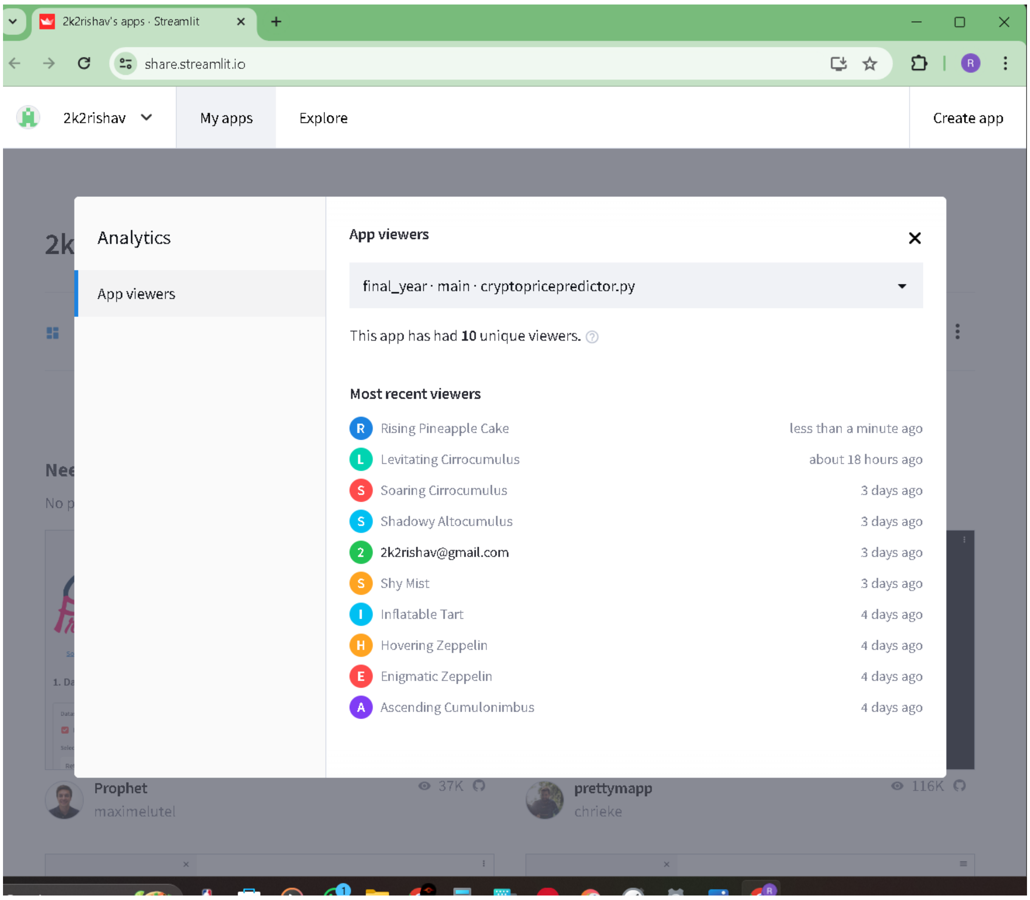


Figure 11.3: Web Analytics report

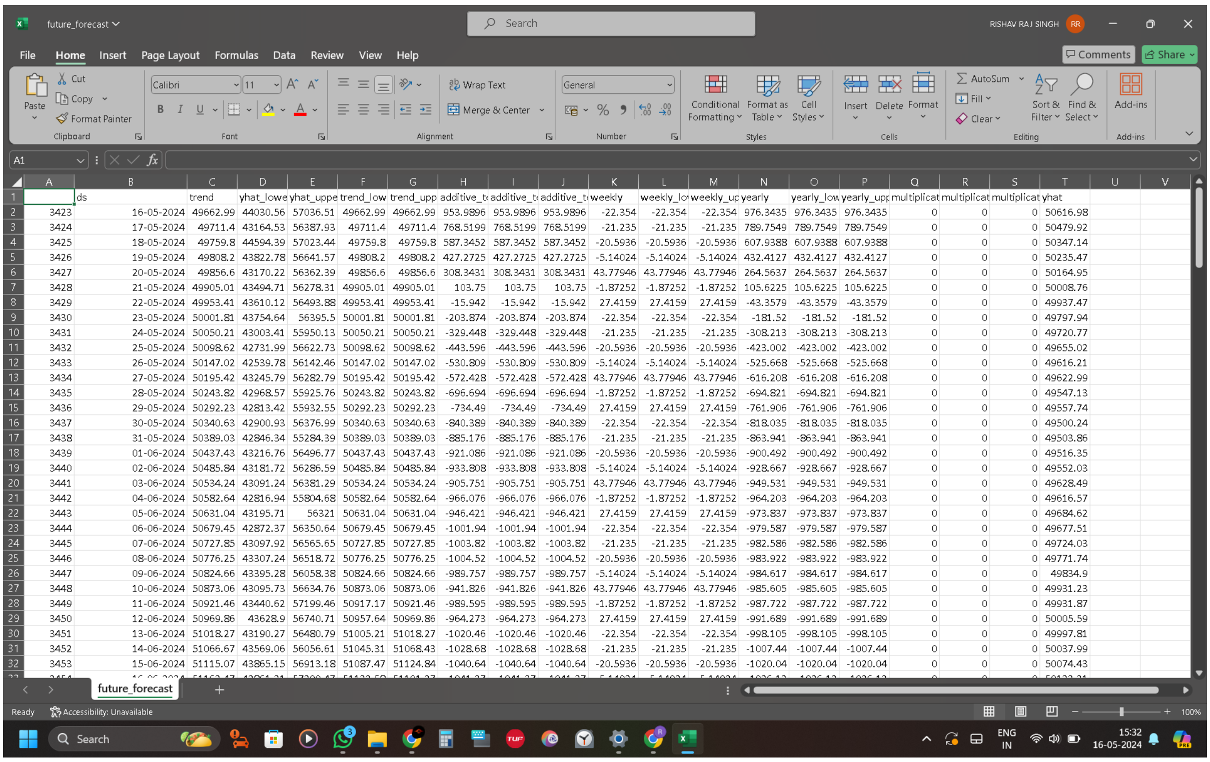


Figure 11.4: Forecast CSV downloads from download

Bibliography

[1] nakamoto2008 S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008. [Online]. Available: https://bitcoin.org/en/bitcoin-paper.

[2] szabo2008 N. Szabo, “Bit gold,” Website/Blog, 2008.

[3] roondiwala2017 M. Roondiwala, H. Patel and S. Varma, “Predicting stock prices using LSTM,” *International Journal of Science and Research (IJSR)*, vol. 6, no. 4, pp. 1754–1756, 2017. Doi:https://www.researchgate.net/publication/327967988_Predicting_Stock_Prices_Using_LSTM.

[4] R. Mittal, S. Arora and M. P. S. Bhatia, “AUTOMATED CRYPTOCURRENCIES PRICES PREDICTION USING MACHINE LEARNING,” 2018. Doi: https://doi.org/10.21917/IJSC.2018.0245.

[5] C.-H. Wu, C.-C. Lu, Y.-F. Ma and R.-S. Lu, “A New Forecasting Framework for Bitcoin Price with LSTM,” in 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 2018, pp. 168–175.DOI:https://doi.org/10.1109/ICDMW.2018.00032

[6] S. McNally, J. Roche and S. Caton, ”Predicting the Price of Bitcoin Using Machine Learning,” 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), Cambridge, UK, 2018, pp. 339-343, Doi:https://doi.org/10.1109/PDP2018.2018.00060

[7] Marcell Tamás Kurbucz, Predicting the price of Bitcoin by the most frequent edges of its transaction network,Economics Letters,Volume 184,,2019,,108655,,ISSN 0165-1765, doi:https://doi.org/10.1016/j.econlet.2019.108655.

[8] B. U. (BTC-USD) and C.-C. C. in USD, "Yahoo Finance," 2024. [Online]. Available: https://finance.yahoo.com/quote/BTC-USD

[9] Tschorsch, Florian; Scheuermann, Björn (15 May 2015). "Bitcoin and Beyond: A Technical Survey of Decentralized Digital Currencies". Available pdf: http://eprint.iacr.org/2015/464.pdf

[10] Frisby, Dominic (2014) "Who is Satoshi Nakamoto?" In Bitcoin: The Future of Money?, p. 85-149. Unbound. ISBN 1783520779

[11] "The Bitcoin Fee Market". 7 March 2017. our transaction growth of nearly 3x [...] Many of the businesses we've signed up over the years have started using BitPay for B2B supply chain payments. Available[online]:https://medium.com/@spair/the-bitcoin-fee-market-4df1857d12b7

[12] Colibasanu, Antonia. "Here's why Russia is opening the door to cryptocurrencies".Available [Online]: https://www.businessinsider.com/why-russia-legalized-cryptocurrencies-2017-5?IR=T

[13] Yingjie Zhu, Jiageng Ma, Fangqing Gu, Jie Wang, Zhijuan Li, Youyao Zhang, Jiani Xu, Yifan Li, Yiwen Wang and Xiangqun Yang, "",2023, Mathematics — Free Full-Text — Price Prediction of Bitcoin Based on Adaptive Feature Selection and Model Optimization (mdpi.com).Available:https://www.mdpi.com/2227-7390/11/6/1335

[14] Facebook Prophet model, Available: https://facebook.github.io/prophet/docs/quick_start.html.

[15] Sean J. Taylor, Ben Letham , “Prophet: forecasting at scale”, Prophet: forecasting at scale - Meta Research — Meta Research (facebook.com). AVvailable;https://research.facebook.com/blog/2017/2/prophet-forecasting-at-scale/

[16] O. Al–Qudah, H. Zureigat, A. Al–Badarneh and S. Abu–Soud, ”Prediction of Cryptocurrencies Prices Using Long Short Term Memory and Technical Indicators,” 2023 14th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2023, pp. 01-06, Doi:https://doi.org/10.1109/ICICS60529.2023.10330471.

[17] Blockchain Info.https://www.blockchain.com/explorer

[18] Alahmari, S. A. (2019). ’Using Machine Learning ARIMA to Predict the Price of Cryptocurrencies’, The ISC International Journal of Information Security, 11(3), pp. 139-144., Doi: https://doi.org/10.22042/isecure.2019.11.0.18

[19] Catania, Leopoldo and Stefano Grassi. ”Modelling crypto-currencies financial time-series.” Available:https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3475957_code2040251.pdf?abstractid=3028486

[20] Mazed, Mashtura. “Stock price prediction using time series data”, Diss. BracUniversity, 2019.Available https://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/12818/14201037_CSE.pdf?sequence=1&isAllowed=y

[21] Chen, Junwei. 2023. ”Analysis of Bitcoin Price Prediction Using Machine Learning” Journal of Risk and Financial Management 16, no. 1: 51. https://doi.org/10.3390/jrfm16010051

[22] Parikh, H., Panchal, N., Sharma, A. (2023). Cryptocurrency Price Prediction Using Machine Learning. In: Pati, B., Panigrahi, C.R.,

Mohapatra, P., Li, KC. (eds) Proceedings of the 6th International Conference on Advance Computing and Intelligent Engineering. Lecture Notes in Networks and Systems, vol 428. Springer, Singapore.Available:https://doi.org/10.1007/978-981-19-2225-1_25

[23] Amirzadeh, Rasoul, Asef Nazari and Dhananjay Thiruvady. 2022. "Applying Artificial Intelligence in Cryptocurrency Markets: A Survey" Algorithms 15, no. 11: 428.Available:https://doi.org/10.3390/a15110428

[24] Pan, Linxi. (2023). Cryptocurrency Price Prediction Based on ARIMA, Random Forest and LSTM Algorithm. BCP Business & Management. 38. 3396-3404. 10.54691/bcpbm.v38i.4313. Available:https://www.researchgate.net/publication/369430933_Cryptocurrency_Price_Prediction_Based_on_ARIMA_Random_Forest_and_LSTM_Algorithm

[25] Sejwal, S., Aggarwal, K., Nayak, S.R., Awotunde, J.B. (2023). Time Series Analysis of Cryptocurrency: Factors and Its Prospective. In: Dhar, S., Do, DT., Sur, S.N., Liu, H.CM. (eds) Advances in Communication, Devices and Networking. Lecture Notes in Electrical Engineering, vol 902. Springer, Singapore.Avaialable [online]:https://doi.org/10.1007/978-981-19-2004-2_22.

[26] Qiu, Mingyue and Yu Song. "Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model." PloS one vol. 11,5 e0155133. 19 May. 2016, Doi: https://doi.org/10.1371/journal.pone.0155133

[27] Kanksha and Harjit Singh, "Cryptocurrency Price Prediction Using FB Prophet Model", https://www.eurekaselect.com/chapter/15670. 2023, Available:https://www.eurekaselect.com/chapter/15670

[28] Khiewngamdee, C., Chanaim, S. (2023). Does Cryptocurrency Improve Forecasting Performance of Exchange Rate Returns?. In:

Huynh, VN., Le, B., Honda, K., Inuiguchi, M., Kohda, Y. (eds) Integrated Uncertainty in Knowledge Modelling and Decision Making. IUKM 2023. Lecture Notes in Computer Science(), vol 14375. Springer, Cham.Available [Online]:https://doi.org/10.1007/978-3-031-46775-2_25.

[29] Hegde, Gayatri. (2023). Bitcoin Price Prediction using Deep Learning. International Journal for Research in Applied Science and Engineering Technology. 11. 1780-1788. 10.22214/ijraset.2023.50481.https://www.researchgate.net/publication/370412294_Bitcoin_Price_Prediction_using_Deep_Learning

[30] Figure:4.2 Available[online]: https://medium.com/@foxworthy_8036/the-facebook-prophet-prediction-model-\and-product-analytics-a1db05fbe454

[31] De Gooijer, J. G. and Hyndman, "25 Years of Time Series Forecasting", International Journal of Forecasting 2006 Doi: https://doi.org/10.1016/j.ijforecast.2006.01.001.

[32] Pygwalker , https://pypi.org/project/pygwalker/ website

Appendix A

Datasets and Data Preprocessing

A.1 Datasets

- **Source of Data:** The datasets used in this study were sourced from Yahoo Finance and Kaggle. These platforms provide reliable and up-to-date data on various cryptocurrencies, including Bitcoin.
- **Data Format:** The data was collected in different time intervals: hourly, daily, weekly, monthly, and yearly formats, to ensure comprehensive coverage of Bitcoin price trends over time.

A.2 Data Preprocessing

- **Data Cleaning:**
  - Removed any missing or null values to ensure consistency.
  - Applied data transformation techniques like normalization and scaling to standardize the data.
- **Feature Engineering:**
  - Creation: Generated new features from existing ones, such as moving averages and rolling mean.
  - Transformation: Used linear interpolation to handle missing values and smooth out the data.

- Extraction: Derived new features using rolling statistics to capture short-term and long-term trends.
- Selection: Selected the most relevant features based on their correlation with the target variable (Bitcoin price).

A.3 Data Visualization

- **Purpose:** Visualization helps in understanding the underlying patterns and trends in the data.
- **Tools Used:** Utilized Python libraries such as Matplotlib, Seaborn, and Plotly for creating various plots and graphs.
- **Types of Plots:**
  - Time Series Plots to show Bitcoin price over time.
  - Autocorrelation and Partial Autocorrelation Plots to understand the dependencies in the time series data.
  - Lag Plots to identify patterns and autocorrelations at different time lags.

Appendix B

Model Implementation Details

B.1 Model Selection

- **Chosen Models:**
  1. Long Short-Term Memory (LSTM)
  2. Extreme Gradient Boosting (XGBoost)
  3. AutoRegressive Integrated Moving Average with Exogenous variables (ARIMAX)
  4. Facebook Prophet (FB Prophet)
- **Evaluation Metrics:**
  1. Root Mean Square Error (RMSE)
  2. Mean Absolute Error (MAE)

B.2 Model Training and Evaluation

- **Training Process:**
  1. Split the data into training and testing sets.
  2. Used cross-validation to ensure model robustness.
  3. Hyperparameter tuning to optimize model performance.
- **Evaluation:**

1. Compared the models based on RMSE and MAE.
2. FB Prophet model was selected as the best-performing model due to its superior accuracy in capturing trends and seasonality in Bitcoin prices.

B.3 Implementation of FB Prophet

- **Forecasting Method:**
  1. Utilized FB Prophet to model daily, weekly, and yearly seasonality along with holiday effects.
  2. Integrated real-time data feeds for continuous updates.
- **Visualization:**
  1. Plotted the forecasted values alongside the historical data to visualize the model's accuracy.
  2. Created upper and lower prediction intervals to account for uncertainty in predictions.

B.4 Deployment

- **Tools and Technologies:**
  1. Streamlit for building an interactive web application.
  2. Pandas for data manipulation and analysis.
  3. Plotly for creating interactive plots.
- **Features:**
  1. Real-time forecasting and visualization of Bitcoin prices.
  2. User-friendly interface for exploring historical and predicted data.
  3. Option to download forecasted data in CSV format for further analysis.