University of Essex
Department of Mathematical Sciences

MA981: DISSERTATION

Bitcoin Price and Volatility Forecasting Using LSTM, ARIMA and GARCH: Exploring Cryptocurrency Dynamics

Huma Nazneen
2204435

Supervisor: Wong Siaw Tze, Jackie

August 24, 2023
Colchester

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Cryptocurrencies have brought about a paradigm shift in the financial landscape by introducing a decentralized and secure digital form of currency, effectively challenging the traditional banking systems. The rise of cryptocurrencies like Bitcoin, Ethereum, and a plethora of others has sparked a surge of interest among investors, traders, and researchers eager to navigate the exciting yet volatile world of digital assets. In this ever-evolving landscape, one of the pivotal challenges lies in accurately predicting the price movements of cryptocurrencies, as these predictions hold the potential to significantly impact investment decisions and trading strategies. The machine learning algorithms have created a new hope for facing this type of challenge. These algorithms has the ability to make full use of computational methods and data analysis and has created an optimistic possibility of forecasting the prices of cryptocurrencies. Various creative methods can now be used and explored by researchers and data scientists to predict cryptocurrency price more accurately by analysing the historical data. Machine learning algorithms are able to spot complex and difficult patterns and trends of data which are not noticed by humans, which ultimately help in price prediction of the cryptocurrency. A combination of computational power and data analytics has helped in the development of complex predictive models which can produce useful insights. Consequently these models have high influence on the decision making capabilities of traders and investors making them take the most practical and rational decisions in the unpredictable and rapidly changing cryptocurrency market. The ongoing development in field of machine learning and ability to access different and large data-set has resulted in taking more precise and trustworthy predictions about price movement. Researchers continuously improve and optimize various models in order to advance in the field of price prediction. This research seeks to explore the price prediction algorithms of cryptocurrency with the help of cutting edge machine learning techniques. In this research, the researcher hopes to develop a consistent and effective predictive model which will help traders, investors and researchers in making profitable decisions using historical data, spotting patterns and using strong algorithms. The objective is to advance knowledge of the intricate dynamics of cryptocurrency price fluctuations and provide insightful information that may help investors and risk managers manage their investments more effectively.

1.1 Research Motivation and Objectives

The potential advantages it could provide to market participants and stakeholders is what drives research into the prediction of cryptocurrency prices. Successful prediction models can aid investors in decision-making, trading strategy optimization, and risk mitigation. Accurate forecasts can also help cryptocurrency projects and businesses foresee market trends, prepare investment plans, and adjust to the constantly shifting the environment of digital assets. The objectives of this project include:

1. Exploring the historical price data of various cryptocurrencies and understanding their market dynamics.
2. Performing rigorous Exploratory Data Analysis to identify patterns, trends, and anomalies within the cryptocurrency market.
3. Applying machine learning algorithms like time series analysis, and deep learning, to build accurate prediction models for bitcoin price and volatility.
4. Evaluating the performance of the developed models using appropriate metrics.

1.2 Key Contribution

The primary contributions of this dissertation revolve around the development and evaluation of novel approaches to prediction of cryptocurrency prices. The author aims to provide comprehensive analysis and experimentation to evaluate the effectiveness of different techniques and methodologies in forecasting bitcoin prices. The results obtained will provide insights into the viability of using advanced machine learning algorithms for cryptocurrency price prediction. The author develops and tests various models, analysing their performance on cryptocurrency price data. By conducting rigorous experimentation and benchmarking against established methods, the dissertation elucidates the advantages and limitations of novel techniques for this task. This makes important contributions towards advancing the state-of-the-art in applying machine learning to financial forecasting in the cryptocurrency domain. The researcher conducts a thorough review of the relevant academic literature and discusses earlier research on forecasting cryptocurrency prices and trends using a variety of different machine learning techniques in chapter 1. In Chapter 2, specifics regarding the data sources are covered along with a preliminary exploratory analysis performed on the raw data and a comparison of cryptocurrencies based on various criteria. The researcher continues in Chapter 3 by describing the project's methodology and emphasizing the machine learning algorithms that were employed to predict the price and volatility of bitcoin. This also covers the presentation and analysis of the results. Future research possibilities are suggested and discussed in chapter 4. The conclusion of this study is contained in the final chapter 5.

1.3 Relevant Studies and Research

Cryptocurrency is a type of digital or virtual money that runs without the help of any central authorities and uses encryption for security. To keep track of transactions and control the issuance of new units, it uses blockchain technology, a decentralised and distributed ledger system [1]. Without the use of intermediaries, peer-to-peer transactions are made possible by cryptocurrencies, which also provide security and transparency through cryptographic methods. Due of their scarcity, they generate value and future appreciation. The market is still extremely volatile and speculative even though it has expanded beyond digital payments to encompass uses like decentralised finance and non-fungible tokens. The technology that underpins the operation of cryptocurrencies like Bitcoin is called blockchain. Blockchain technology is a decentralised, distributed digital ledger that securely, openly, and irreversibly records transactions across numerous computers or nodes. The public continues to have faith in the current financial system despite its many flaws, mostly because it is so dependent on laws and other legal agreements. However, trust breaches in the past have resulted in large financial losses, as demonstrated by the dot-com boom in the 1990s and the real estate bubble in 2008 [20]. With the publication of the Bitcoin whitepaper in 2008 by an unidentified individual or group known as Satoshi Nakamoto, the blockchain concept was first proposed. Due to its potential to revolutionise many industries by offering safe, open, and effective methods for transactions and data management, blockchain technology has attracted a lot of attention. Figure 1.1 presents a graphical depiction of the global popularity trends of blockchain and cryptocurrency starting from the year 2010. The data used to create this figure comes from Google Trends. The graph displays a noticeable surge in the popularity of cryptocurrency during two distinct periods: in 2018 and again in 2020. However, from Nov 2022, there has been a clear downside trend in its popularity. In 2018, the surge in interest may have been influenced by significant market growth, media attention, and the emergence of new cryptocurrencies. Similarly, the increased popularity in 2020 could be attributed to various factors, including increased institutional involvement, mainstream media coverage, and growing adoption of cryptocurrencies as a form of investment and payment. Trading and investing in cryptocurrencies had become a safe haven for many individuals and organisations as a result of global economic uncertainties during and post Covid-19 pandemic. The downward trend from late 2022 could be due to factors such as regulatory concerns, market volatility, and fluctuations in cryptocurrency prices, leading to reduced public interest and search queries.

As per Khedr et al. (2021), the main objective of implementing a digital monetary system is to address the challenges posed by inflation and negative yields for consumers, ultimately aiming to offer improved financial stability. By adopting such a system, it is expected to enhance various aspects of monetary transactions, including convenience, speed, and cost-effectiveness, which in turn can lead to significant economic benefits. Although it is still a relatively new technology, continuing work is being done to address issues with scalability, energy consumption, and interoperability to make it more useful for broad adoption. Other difficulties with blockchain, according to Scott Likens, Global AI Lead at PwC US, are its sophisticated technology, regulatory implications, implementation difficulties, and rival platforms. Iwamura et al. (2014) emphasises in their study how the competition between Bitcoin and other cryptocurrencies benefits the market by promoting technological and security advancements. The competitive environment encourages developers and researchers

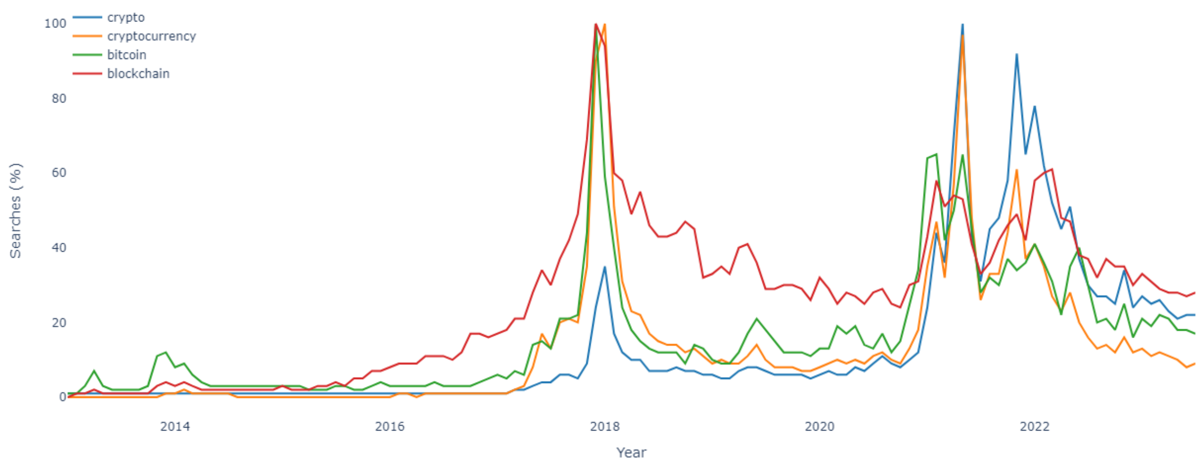


Figure 1.1: Global Popularity Trends of Blockchain, Cryptocurrency and Bitcoin

to innovate, introducing new features and improving their respective digital assets. The industry's overall technological environment is improved by this desire for advancement, which results in more advanced systems. Additionally, the competition encourages a focus on security measures, strengthening the system's resistance to online threats. The availability of numerous cryptocurrencies also gives users a variety of options, allowing for various use cases and drawing in a larger user base.

When analysing the cryptocurrency market, it's essential to recognize that changes in market dynamics can be influenced not only by internal factors specific to the crypto industry but also by external macroeconomic factors. The recent COVID-19 pandemic significantly impacted the global financial environment, leading to increased uncertainty in the cryptocurrency market. The interplay between macroeconomic events and the crypto market has contributed to continued volatility and unpredictability. To effectively navigate such a turbulent market, it is crucial to consider the influence of the broader financial system on cryptocurrency price movements. By considering both internal and external factors, analysts and investors can capture a more comprehensive picture of the market and make informed decisions amidst the ongoing uncertainties [3]. The cryptocurrency market's tremendous volatility makes it difficult to estimate market risk merely based on technical indications. The price of cryptocurrencies is notorious for fluctuating quickly and unpredictably, frequently due to a variety of reasons such as market sentiment, legislative changes, macroeconomic events, and media attention. Technical indicators can offer helpful insights into previous pricing patterns and trends, but Park et al. (2023), noted that they may not fully represent the complexity of the market and cannot forecast abrupt or unexpected shifts. According to Cohen, G (2021), recommending a "buy and hold" strategy for long-term investors might sometimes produce more accurate outcomes than trying to forecast short-term market fluctuations. The "buy and hold" strategy is investing in assets with the belief that, despite short-term changes, their value will rise over time. This strategy may provide better risk management and possibly more lucrative results for some investors than trying to time the market using technical indicators in highly volatile circumstances. The use of machine learning (ML) techniques, including Artificial Neural Networks (ANNs), Support Vector Machines (SVM), and Random Forest (RF), has been applied to predict Bitcoin prices due to their ability to analyse complex non-linear relationships among multiple variables [5], [6], [7], [8]. With their strong representational capabilities, Deep Learning (DL) techniques have demonstrated promising results in financial time-series forecasting [9]. The Long Short-Term Memory (LSTM) model has outperformed other DL models in terms of performance because it is particularly good at learning long-range patterns using recurrent and gate mechanisms [11],[10]. The LSTM model needs a lot of parameters to manage the high dimensionality, however as the number of input variables rises, this could result in overfitting issues. Cho et al. (2014), presented the Gated Recurrent Unit (GRU) as a simplified variant of LSTM, lowering the parameters of the gate mechanism, to overcome this problem. GRU still needs more parameters than certain ML models, despite learning more quickly and being more stable at preventing overfitting [12]. Pasak and Jayadi (2023) present a comparative study on the predictive performance of ARIMA and LSTM models for cryptocurrency price forecasting. The outcome of the study was that the ability to accurately predict cryptocurrency prices is crucial for informing investment decisions in this volatile market. The authors make a valuable contribution by benchmarking two popular forecasting techniques - the statistical ARIMA model and the deep learning LSTM model.

In the paper released in 2023 by Zhong et al. proposed a novel methodology for forecasting cryptocurrency price trends using a hybrid model that combines Long Short-Term Memory (LSTM) and Relational Graph Attention Network (ReGAT) techniques. This research recognises the rising demand for cryptocurrencies as alternative investments and the intricacy of their price fluctuations, which makes precise forecasting essential for traders and investors. The authors suggested that the deep relationships between cryptocurrencies in a network are frequently missed by current techniques. In order to overcome this drawback, LSTM-ReGAT adopts a network-centric viewpoint, wherein the connections between cryptocurrencies are represented as a graph structure, enabling more reliable prediction models. According to the literature, LSTM-ReGAT greatly advances the field of predicting cryptocurrency price trends by including ReGAT's attention mechanism, which enables the model to concentrate on key nodes in the network. This method combines the temporal pattern recognition capabilities of LSTM, enabling a thorough knowledge of the interaction between individual cryptocurrency user activity and their collective impact on price movements. The effectiveness of the LSTM-ReGAT model is evaluated by the authors against a number of cutting-edge prediction techniques using real-world cryptocurrency data. The literature review indicates that the proposed approach outperforms traditional models and exhibits superior accuracy in forecasting cryptocurrency price trends, demonstrating its potential value in aiding investment decision-making and risk management strategies. In their work, Chen et al. (2021), proposed a two-stage hybrid technique for predicting BTC prices. They used a combination of Artificial Neural Networks (ANN) and Random Forest (RF) in the initial stage to pinpoint the key characteristics that have a significant impact on price prediction. They moved on to the second stage after identifying these key features, where they used a Long Short-Term Memory (LSTM) model with the features they had chosen to produce more precise and reliable predictions for BTC prices. Their hybrid approach aimed to improve the precision and efficacy of BTC price forecasting by sequentially utilizing the strengths of various models.

Seabe et al. (2023), looked into the application of deep learning techniques to forecast cryptocurrency prices in their study .The importance of cryptocurrencies on the financial markets has made it crucial for traders and investors to make accurate price predictions. Three popular deep learning models-the Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and bidirectional LSTM models are looked at in the study. These models are recognised for their ability to identify temporal correlations and trends in time series data when it comes to forecasting bitcoin values. The authors of this study set out to fill a major knowledge gap in the area by offering insightful information on the efficacy and potential of deep learning techniques in predicting Bitcoin values. The researchers did admit that these techniques had their limitations, too. They emphasised that a number of intricate aspects, including market sentiment, regulatory changes, macroeconomic variables, and geopolitical events, have a significant impact on Bitcoin's price and may not be fully captured by deep learning models like LSTMs, GRUs, and Bi-LSTMs. These models are also vulnerable to overfitting, particularly when trained on small datasets, which can lead to decreased accuracy when used with new data. These algorithms face difficulties in reliably identifying underlying trends due to the considerable noise and volatility around cryptocurrency pricing, further complicating their prediction ability. Despite these drawbacks, the study offered crucial insights into the possible uses of deep learning in the world of cryptocurrencies. In their most recent publication, Park, and Yang (2023) present a unique machine learning-based Bitcoin trading system that combines AdaBoost and Long Short-Term Memory (LSTM) while also considering market volatility. Their study aims to develop an intelligent trading system that can thrive in the unpredictable cryptocurrency market. The key innovation lies in the combination of LSTM, a deep learning method proficient at modelling sequential data, with AdaBoost, which creates a strong ensemble model from multiple weak learners. Additionally, the authors implement a mechanism to adapt trading tactics based on real-time market turbulence indicators, recognizing the unpredictability of cryptocurrency markets. The work integrates LSTM and AdaBoost with market turbulence knowledge, considerably advancing the creation of intelligent Bitcoin trading systems. Chowdhury et al. (2020) also employed an ensemble model incorporating ANN, KNN, and gradient boosted trees to forecast price changes for nine different cryptocurrencies. They found that the ensemble learning strategy gave the most accurate predictions and outperformed individual techniques. The ensemble strategy for predicting bitcoin prices has proven to be a powerful tool, improving accuracy and dependability of the results by using the advantages of many models. Derbentsev et al. (2021) forecasted the prices of three different cryptocurrencies: BTC, ETH, and XRP using an ensemble model made up of a Gradient Boosting Machine (GBM) and Random Forest (RF). When they calculated the Mean Absolute Percentage Error (MAPE) for these projections, they discovered that the ensemble model's MAPE values ranged from 0.92 to 2.61 percent. These results demonstrate that the ensemble model showed good accuracy in its price predictions for the chosen coins, indicating a promising future for forecasting cryptocurrency prices.

A unique hybrid model is put out in the study by Garca-Medina and Aguayo-Moreno (2023) for predicting volatility in bitcoin portfolios. Predicting cryptocurrency volatility has become more crucial for risk management as cryptocurrencies gain acceptance as investment assets. Traditional forecasting methods, however, have difficulties due to the bitcoin market's high volatility and dynamic character. The authors create a hybrid model that combines Generalised Autoregressive Conditional Heteroskedasticity (GARCH) and Long Short-Term Memory (LSTM) to close this gap. Utilising the combined strengths of LSTM and GARCH constitutes the fundamental innovation. According to the authors, LSTM excels in identifying temporal connections and patterns in time series data. Long-term historical contexts can be memorised via recurrent neural network design. On the other hand, GARCH is designed for modelling time-varying volatility and volatility clustering. The hybrid model can take use of LSTM's capacity to learn from cryptocurrency price histories while also enabling GARCH to concentrate on volatility dynamics by combining these two methodologies. In comparison to more traditional approaches like historical simulation, exponential smoothing, and solo LSTM and GARCH models, the LSTM-GARCH hybrid model performs better.

In their study, Khedr et al. (2021), analysed a collection of related scientific papers published between 2010 and 2020. The survey findings highlight a growing emphasis among researchers on employing Machine Learning (ML) models for cryptocurrency price prediction. The significance of ML and Deep Learning (DL) approaches is evident as they play a pivotal role in advancing this field of research. The findings indicate that the accuracy of price prediction greatly relies on the selection of input attributes and the ML technique employed. Additionally, the effectiveness of an ML algorithm is heavily influenced by the nature of the problem being addressed and the quality and complexity of the training data set. Most of the studies in this field have focused on utilizing daily and interval-based price data for prediction purposes. However, some researchers have also explored the impact of socio-economic factors, user trends, and macro attributes on cryptocurrency price prediction. This indicates the growing interest in understanding the broader factors that may influence cryptocurrency prices beyond just historical price data. As this field continues to evolve, it is evident that ML techniques hold significant promise in enhancing the accuracy and efficiency of cryptocurrency price prediction models.

CHAPTER 2

Data Source and Data Overview

This chapter presents the details of the data sources and their corresponding features along with a description of the initial and in-depth approaches used for data exploration. The study of time series qualities, including volatility and trend dynamics and more, is the main objective. The chapter proceeds to elucidate the preliminary and advanced techniques employed for the exploration of the time series data. Statistical, computational, and visual methodologies are employed to uncover patterns, trends, and volatility shifts, among other significant properties. The chapter lays the groundwork for following investigations and supports the larger research goals by examining basic time series features.

2.1 Process diagram

The Process Diagram shown in Figure 2.1 visually presents the various steps taken in this study to predict cryptocurrency price and volatility trend. The preceding sequence of steps exemplifies the essential path taken for data analysis within the vast domain of machine learning. These steps, which must be carefully planned in order to extract meaningful insights from complex datasets [2]. Each stage is critical in solving the complexities of data and harnessing its potential to inform decision-making and power predictive models. It shows how different tasks are connected to achieve a bigger objective. Starting with Data Extraction, the study collects tickers of different cryptocurrencies from Yahoo Finance. These tickers are like keys to get more information. Then, the study looks at all this information together in the Combined Data Set Analysis. This helps find common patterns and trends among all the cryptocurrencies. The study also looks at people's feelings about the market using the Fear and Greed Index Analysis. This shows how emotions affect trading. Another step is to look at top 5 cryptocurrency closely to perform detailed exploratory data analysis. This helps understand what makes each one different and how they perform. Using special machine learning methods, the study tries to guess the price of Bitcoin in the Bitcoin Price Prediction step. This can help with decision-making. Similarly, the study tries to guess how much the Bitcoin price might change in the Bitcoin Volatility Prediction. The figure ties all these steps together, it's like a roadmap of the study's journey, from collecting data to making predictions. The study examines the application of machine learning techniques to analyse and forecast fluctuations in Bitcoin's price and value over a period. The primary focus lies in identifying and understanding patterns and trends within the Bitcoin market, while deliberately excluding the influence of external factors.

2.2 Data Source

Several open datasets that were available via the World Wide Web's global reach were analysed by the researcher as part of this study. Datasets from websites like Coin Market Cap, Coin Codex, and Yahoo Finance underwent careful examination. This investigation's goal was to evaluate the viability, difficulty, and dependability of the data extraction processes connected to each source. After careful consideration, it was decided that the best and most appropriate option for locating the necessary data was Yahoo Finance. This choice was made after giving numerous aspects, such as data reliability, extraction complexity, and dependability. The availability of a dedicated Python library for the Yahoo Finance API significantly mitigates the complexities associated with data extraction. This library offers a structured and user-friendly interface that expedites the retrieval of data elements as stipulated by the research objectives. By leveraging this library, the researcher can efficiently gather the requisite data points without delving into intricate data collection procedures.

A structured data extraction process was used to gain a thorough understanding of the cryptocurrency landscape. An extensive list of ticker symbols for different cryptocurrencies was first extracted from Yahoo Finance. Each cryptocurrency is identified specifically by its ticker symbol, which facilitates quick data retrieval. Then, the emphasis shifted to the

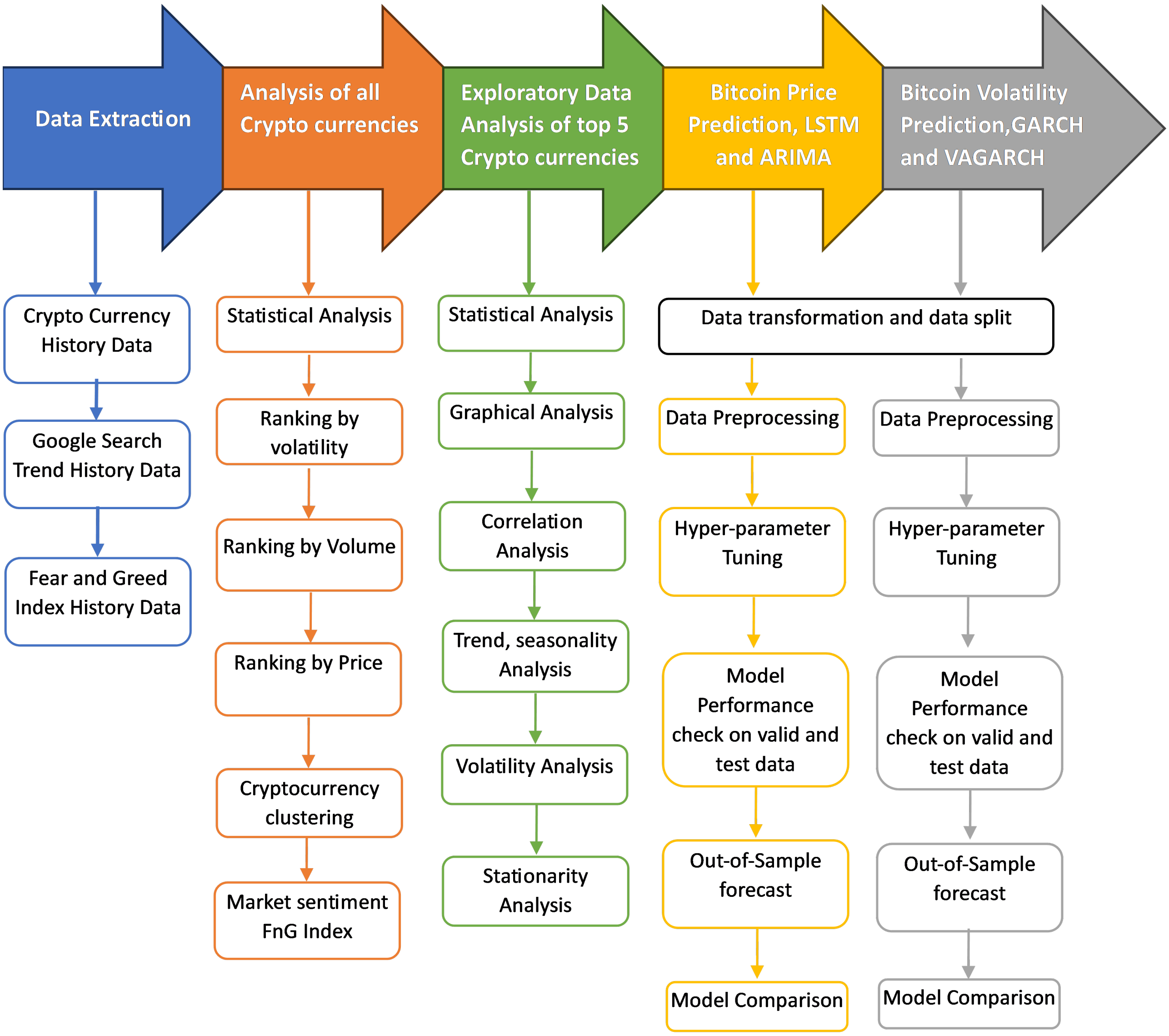


Figure 2.1: Process Diagram

extraction of OHLCV data for each cryptocurrency, which includes Open, High, Low, Close, and Volume metrics. The calculation of price percentage change was carried out for each individual cryptocurrency during the data retrieval process. These calculated values were then systematically stored in conjunction with the OHLCV data for further analysis. A daily interval was selected for data collection in accordance with responsible data management. This interval strikes a judicious balance between acquiring informative granularity and managing the volume of data. An additional strategic consideration in the data extraction process was the setting of the "period" parameter for each cryptocurrency. Opting for the "max" setting ensured the inclusion of the entire historical dataset available for every cryptocurrency.

With the help of the Python library pytrend, an analysis of Google search activity was done to measure the popularity of cryptocurrencies. Specifically in the context of investing and trading, the Fear and Greed Index is a tool used to gauge the sentiment and emotional state of the financial market. To determine whether fear or greed is the main motivator for market participants, it considers a number of variables and indicators. The index is typically made up of a variety of metrics, including safe-haven demand, investor surveys, put/call ratios, stock market volatility, and more. These metrics offer perceptions into the general psychology of traders and investors. The relationship between the Fear and Greed index and digital currencies was also examined.

2.3 Initial Data Explore - Comparative analysis

This section presents the results of the initial data analysis performed on historical cryptocurrencies datafiles. To gain insights into the behaviours and trends of various cryptocurrencies.

2.3.1 Statistical Data Analysis

The dataset encompassing historical information regarding various cryptocurrencies was subjected to statistical examination. The objective was to establish a comprehensive understanding of the data's characteristics. As part of this process, a consolidated table was formulated to facilitate a comparative evaluation of key statistical metrics such as counts, means, min, max, standard deviations, and quantiles of the different cryptocurrencies. The output, presented as in the Table 2.1, illustrates a subset of the sample statistics derived from the consolidated table.

Statistics	BTC-USD	ETH-USD	USDT-USD	BNB-USD	XRP-USD	USDC-USD
count	3252	2103	2103	2103	2103	1770
mean	13788.56	1196.42	1.00144	161.98	0.51949	1.00213
std	16013.17	1140.70	0.00542	178.69	0.34653	0.00552
min	178.10	84.31	0.96664	1.51036	0.13963	0.97012
25%	772.38	224.86	0.99999	15.07	0.29654	0.99993
50%	7818.47	707.05	1.00040	30.06	0.40624	1.00013
75%	20744.22	1846.17	1.00210	305.71	0.63968	1.00172
max	67566.83	4812.09	1.07788	675.68	3.37781	1.04403

Table 2.1: Statistics of Selected Cryptocurrencies

The researcher examines various cryptocurrencies by comparing their volatility, average traded volume, and mean price. Volatility is gauged through percentage changes in prices, providing insights into how much prices fluctuate over time. Average traded volume reflects the level of market activity, while recent prices offer a snapshot of current values.

2.3.2 Ranking cryptocurrencies by its volatility in price

Cryptocurrencies have been categorized according to their volatility, which is computed as the percentage change in their daily prices over the preceding 30-day and 7-day periods. In Figure 2.2 and 2.3, it's evident that 'worldcoin' exhibited the highest volatility, showing the greatest percentage change in price. Following closely was 'wrapped kava,' which also displayed significant volatility with values of 69.52 and 3.19, respectively. However, the landscape shifted dramatically when considering the volatility of the past week. 'Wrapped kava' emerged as the most volatile, registering a value of 0.37. Other coins like 'rollbit coin,' 'liverpeer,' and 'pepe' also entered the list of high volatility, with values ranging from 0.3 to 0.2. This shift highlights the unpredictable nature of crypto currency price, demonstrating its highly uncertain behaviours.

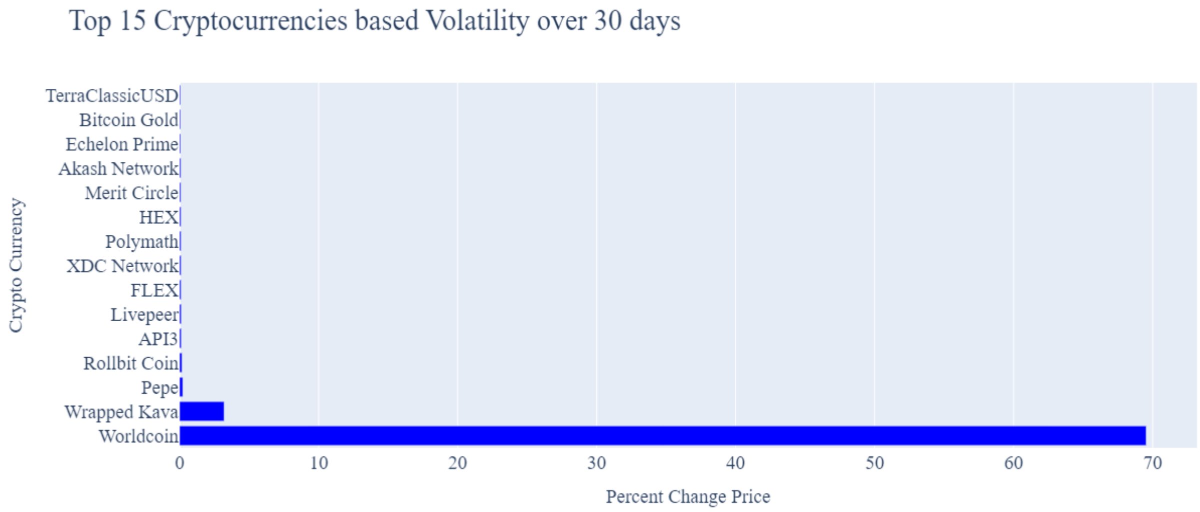


Figure 2.2: Ranking by Volatility over a 30-Day Timespan

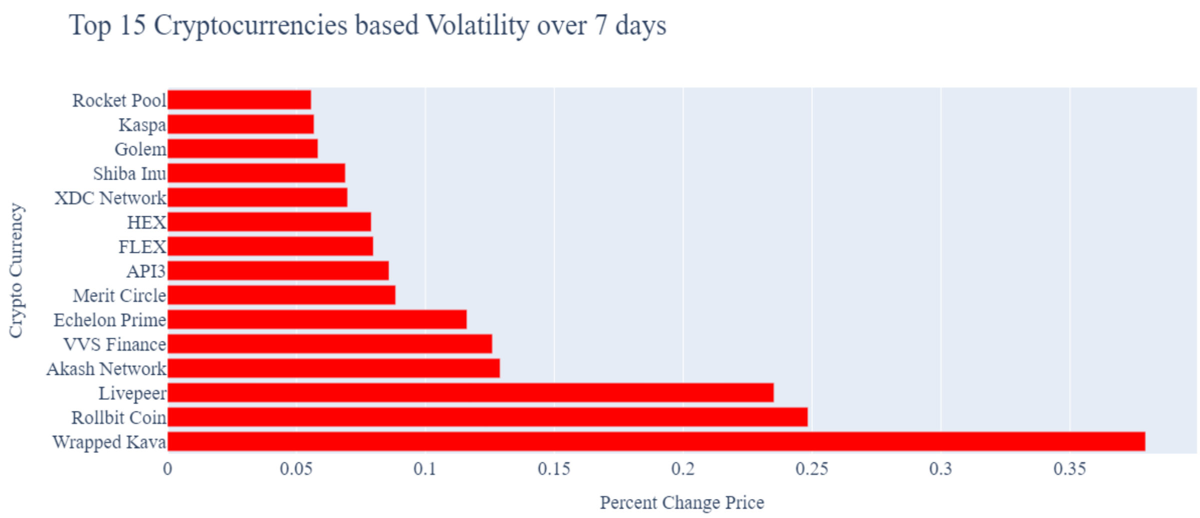


Figure 2.3: Ranking by Volatility over a 7-Day Timespan

2.3.3 Ranking cryptocurrencies by average traded volume

In contrast to the listing of cryptocurrencies based on their volatility, the compilation of the top 15 coins being traded extensively comprises well-known and widely recognized cryptocurrencies. This observation signifies that a significant portion of traders and investors tend to allocate their funds to these leading coins, presumably due to their perceived safety and stability. These statistics are visually represented in Figure 2 which shows that there is not significant change in the list based on average traded volume over 30days and 7 days.

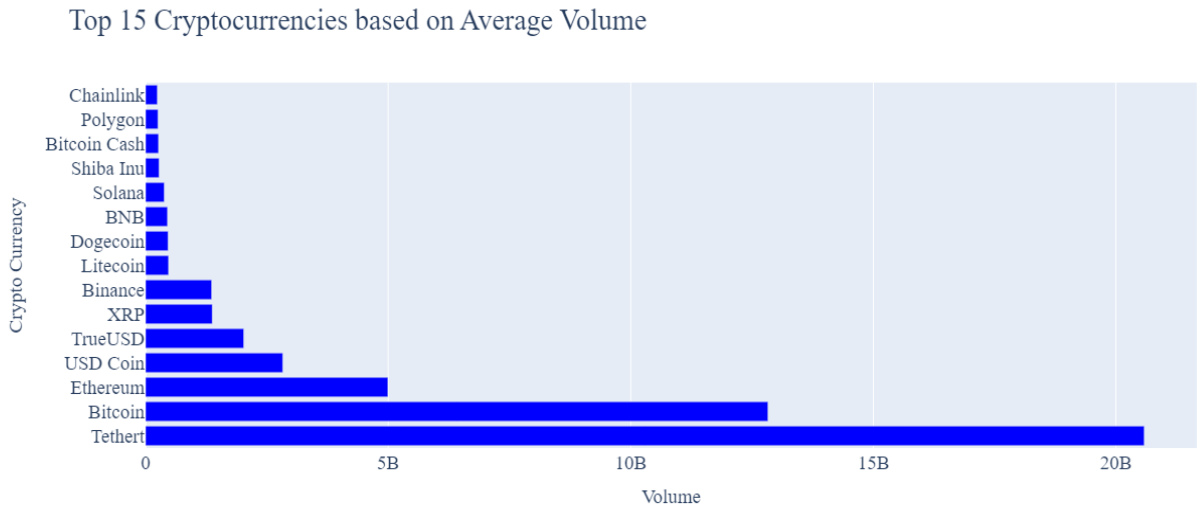


Figure 2.4: Ranking by average traded volume over a 30-Day timespan

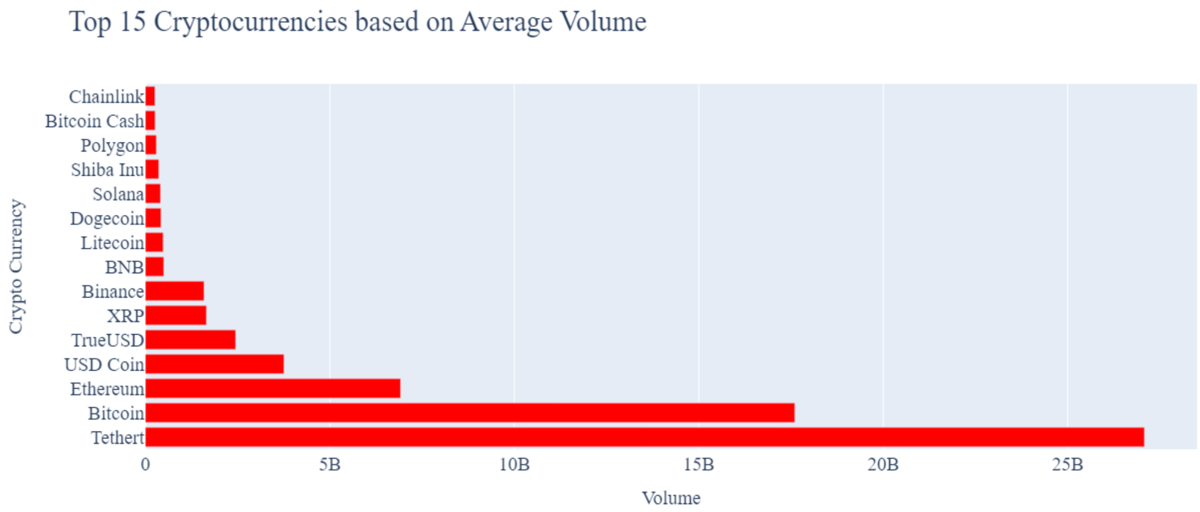


Figure 2.5: Ranking by average traded volume over a 7-Day timespan

2.3.4 Ranking cryptocurrencies by its mean price over different timespan

The average prices of various cryptocurrencies have been computed across different time-frames: one year, six months, one month, and seven days. This analysis aimed to discern the top-performing currencies based on their prices. During this evaluation, it was noted that while the prices of individual coins fluctuated over time, the list of the top 15 coins remained relatively consistent. This indicates that certain cryptocurrencies have maintained their prominence and desirability among investors, irrespective of the observed changes in their prices. The Figures 2.6, 2.7, 2.8, and 2.9 show top 15 coins based on the average price over timespan of one year, six months, one month and one week respectively.

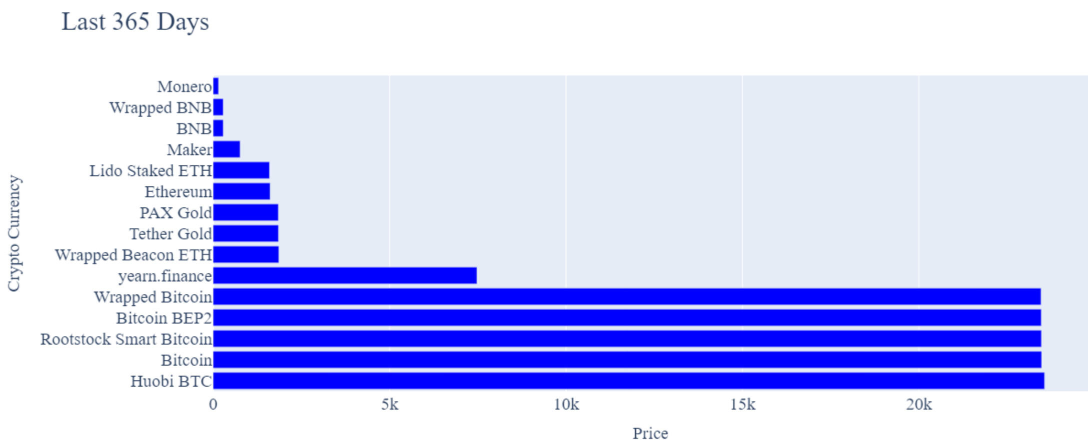


Figure 2.6: Ranking by Mean Adj Close over a 365-Day Timespan

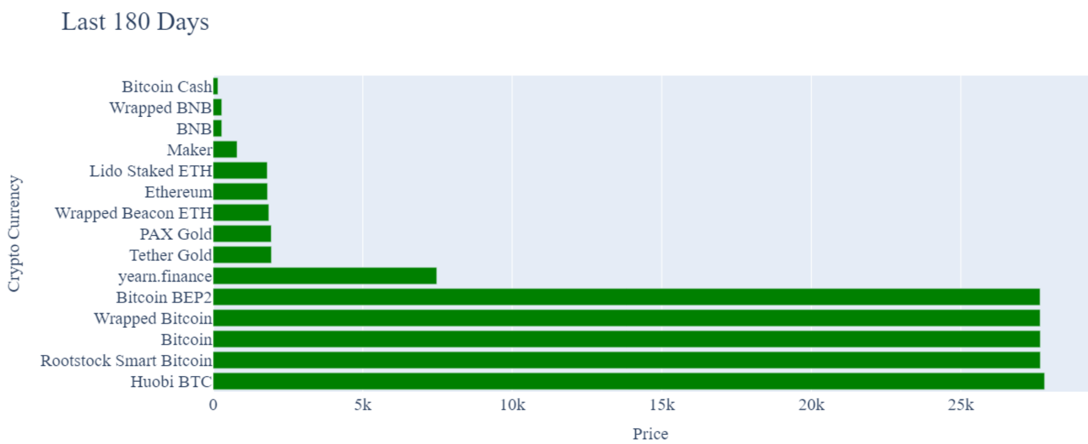


Figure 2.7: Ranking by Mean Adj Close over a 180-Day Timespan

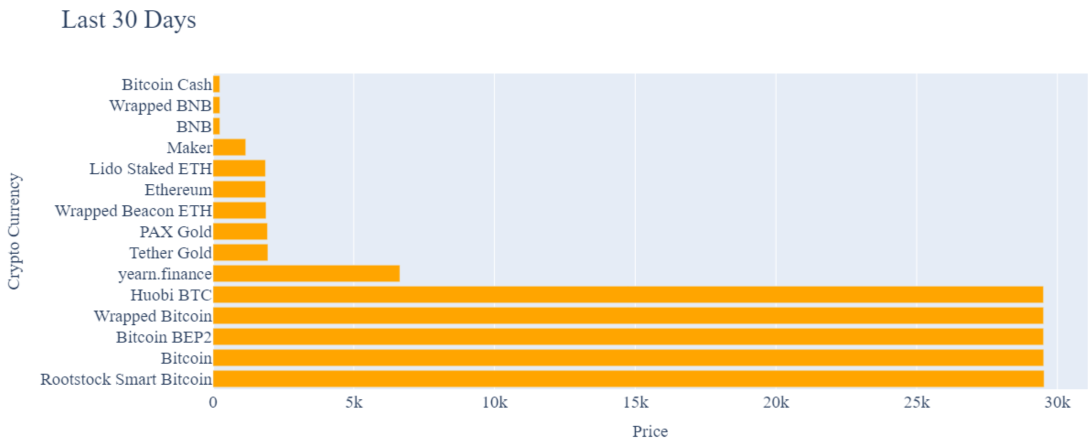


Figure 2.8: Ranking by Mean Adj Close over a 30-Day Timespan

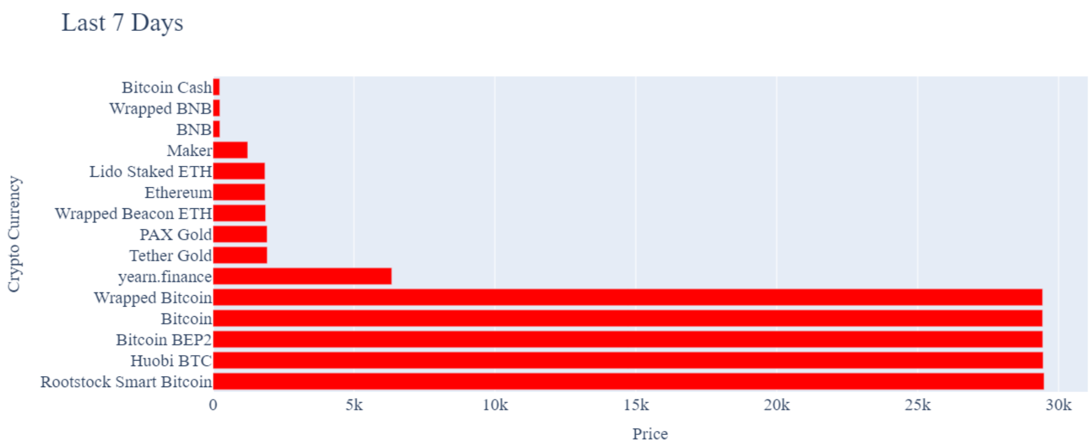


Figure 2.9: Ranking by Mean Adj Close over a 7-Day Timespan

In conclusion, the analysis of cryptocurrency data has revealed the intricate interplay between volatility, trading volume, and average prices. The examination of volatile periods underscores the unpredictable nature of certain cryptocurrencies, with 'worldcoin' and 'wrapped kava' as prime examples. The constancy of the top coins in terms of trading volume highlights the preference for established and trusted options. Meanwhile, the stability in the list of top-performing coins based on average prices suggests sustained investor interest despite fluctuating market conditions. This complicated understanding emphasizes the nuanced dynamics of the cryptocurrency landscape, where factors such as speculation, market sentiment, and fundamental value intersect to shape market trends and participant behaviours.

2.3.5 Clustering cryptocurrencies

A list of the top 50 cryptocurrencies was created, and their relationships were studied by looking at how they move together. This was done using coefficient of correlation, which helps to understand if these cryptocurrencies tend to go up or down in value together. After calculating these relationships, a "dendrogram" was constructed utilizing the spatial distances between the correlation coefficients. The diagram depicted in the Figure 2.10 illustrates the degree of interconnectedness among these cryptocurrencies. The diagram revealed that a significant number of these cryptocurrencies exhibit strong connections and relationships with one another. When the value of some cryptocurrencies rises, other cryptocurrencies frequently follow suit at the same time. However, a correlation does not prove a cause. Beneath these patterns lies a complex interplay of external factors, exerting their influence and setting up synchronized movements across selected cryptocurrencies.

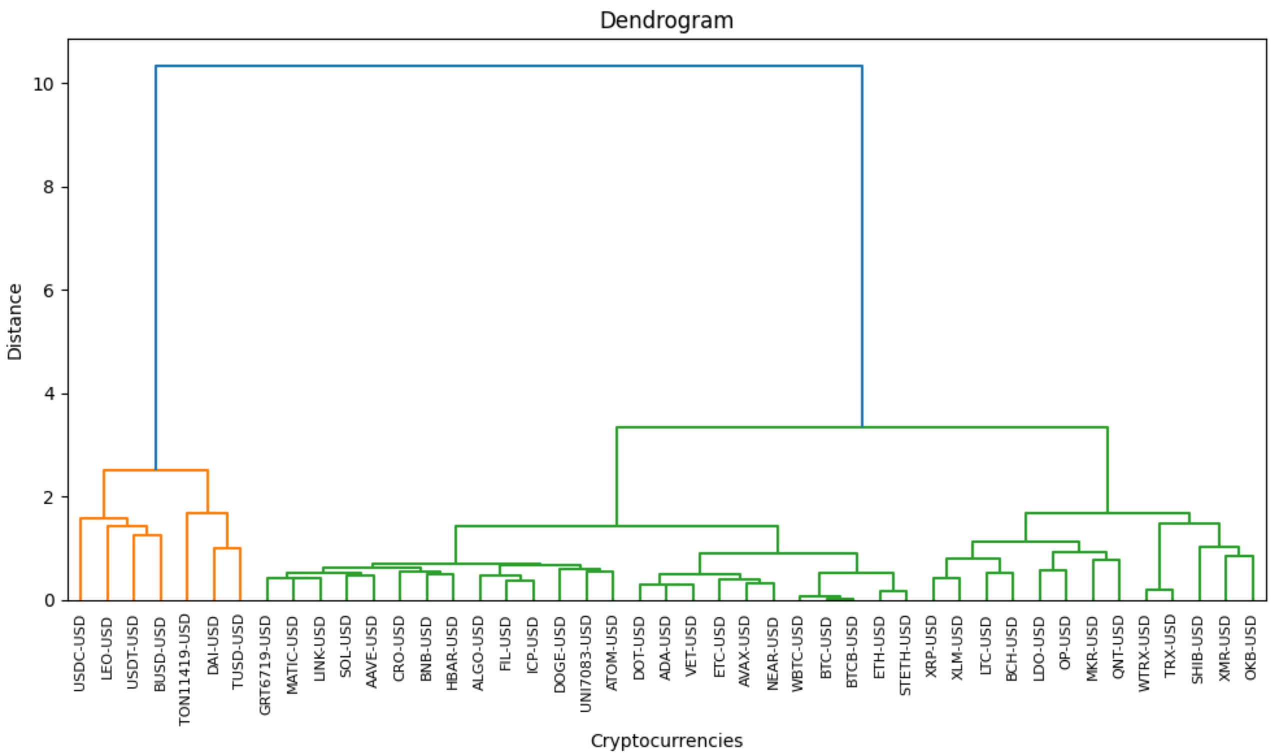


Figure 2.10: Visualizing Cryptocurrency Clustering through Dendrogram

2.3.6 Market sentiment analysis - Fear and Greed Index

The Fear and Greed Index is a popular tool used to gauge the general sentiment and emotions of investors and traders in the financial markets. It attempts to quantify the prevailing levels of fear and greed that are driving market behaviours at a given point in time. The index is derived from a combination of various market indicators and data points to provide an overall assessment of market sentiment. The price of top 5 cryptocurrency, how much they are being traded, and how much their prices are changing are put on the same scale and then compared with the Fear and Greed Index to see if these factors are related to the level of fear or greed in the market. The outcome of this work is shown in the figure 2.11 and 2.12.

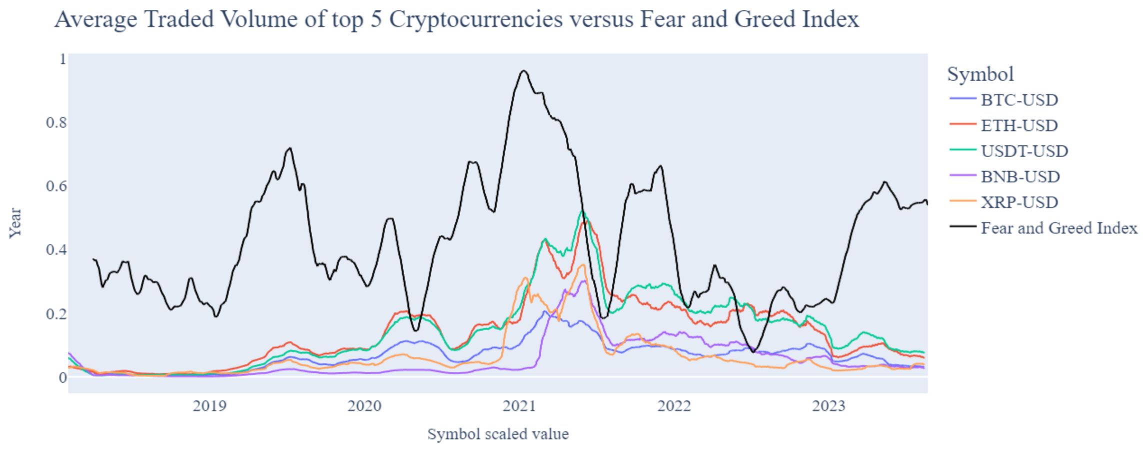


Figure 2.11: Average Traded Volume of Top 5 Cryptocurrencies versus Fear and Greed (FnG) Index.

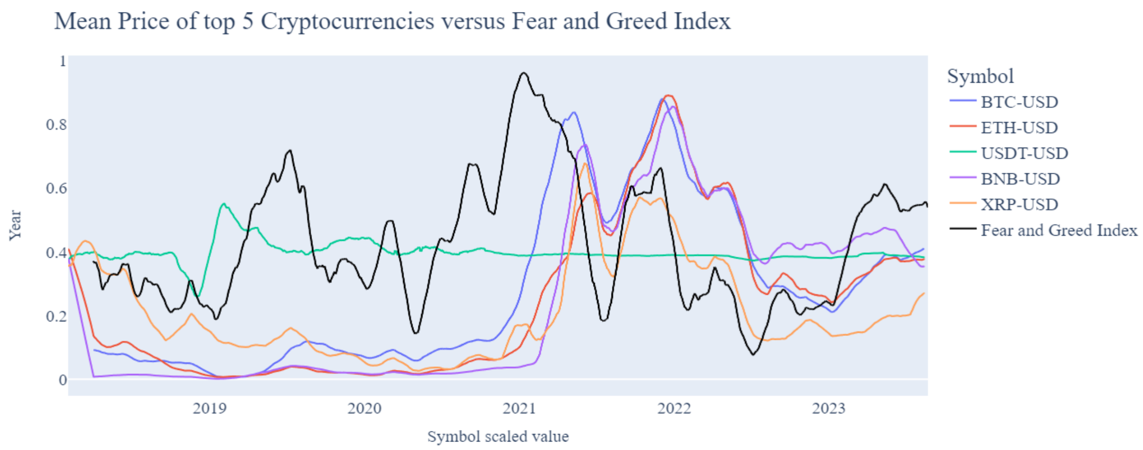


Figure 2.12: Comparison of Rolling Mean Prices of Top 5 Cryptocurrencies and the Fear and Greed (FnG) Index.

2.4 Exploratory data analysis on bitcoin

The dataset containing historical information about various cryptocurrencies has been divided into separate datasets for each individual cryptocurrency. A list of the top 5 cryptocurrencies, as ranked by Yahoo Finance was compiled . A meticulous examination of the data was undertaken, encompassing aspects such as data samples, missing values, and duplicates. During this scrutiny, it was observed that no missing or duplicated values were present within these currencies. However, instances of outliers in the price data were identified. These outliers were attributed to abrupt price surges, a plausible justification for their presence. A versatile code has been developed for conducting detailed exploratory data analysis (EDA) on the historical data of each specific coin. For this report, Bitcoin is chosen as the primary currency for analysis, and all subsequent procedures will be elaborated using Bitcoin as the reference currency.

2.4.1 Graphical Data Analysis

After a preliminary statistical analysis on Bitcoin historical data, it is evident that the dataset comprises 3240 entries spanning from September 17, 2014, to july 31, 2023. The highest recorded price within this period is 67566. To gain deeper insights into the distribution of the coin's price data, histograms were employed for visualization purposes. The histograms highlight that most recorded prices remained below ten thousand, with occasional instances of surpassing sixty thousand. This visual representation is clearly depicted in Figure 2.13 and 2.14.

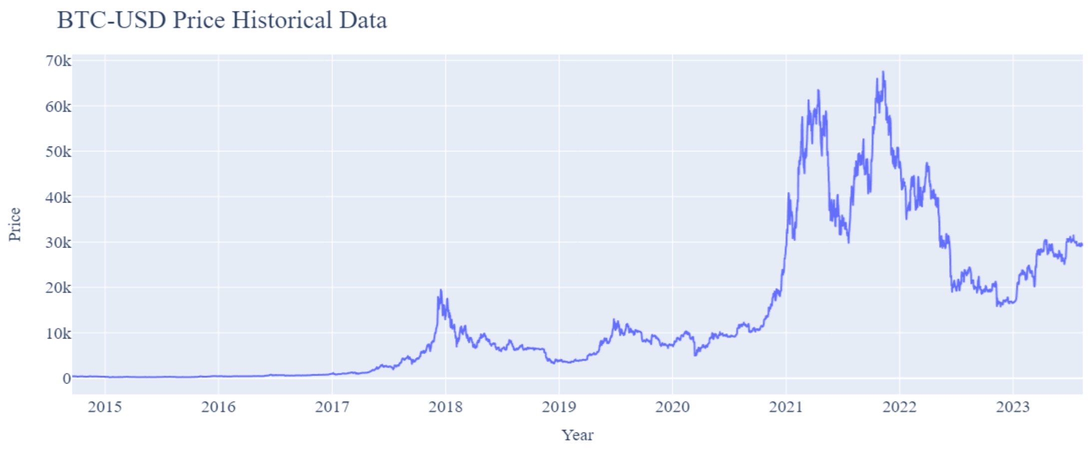


Figure 2.13: The price of Bitcoin from September 2014 to July 2023.

Figure 2.14: Distribution of Bitcoin Prices from September 2014 to July 2023.

2.4.2 Correlation Analysis

The correlation between OHLCV (Open, High, Low, Close, Volume) data and the percentage change in derived values of price is examined. The resulting correlation values are visualized in the heatmap plot depicted in Figure 2.15. Notably, the correlation coefficients for Open, High, Low, Close, and Adjusted Close exhibit a degree of similarity, indicating minimal fluctuations within a given day. Nevertheless, it is worth highlighting that the relationship between traded volume and the price of Bitcoin also carries significance. This implies that fluctuations in trading volume could potentially influence Bitcoin's price movements.

Figure 2.15: Bitcoin: Correlation across All Columns

2.4.3 Volatility Analysis

Volatility is a statistical measure of how much the price or value of an asset deviates from its average or mean. In the context of financial markets, higher volatility indicates larger price swings, while lower volatility suggests more stable and consistent price movement. Figure 2.16 shows log return of the bitcoin. In this study, volatility is assessed through two distinct methods. The first method involves computing the standard deviation of percentage changes in price, while the second method involves calculating the standard deviation of log returns for the Bitcoin price. The formulas used to calculate percentage change and log return of the bitcoin price are shown in equation 2.4.1 and 2.4.2.

$$\text{Percentage Change} = \left(\frac{\text{price}_t - \text{price}_{t-1}}{\text{price}_{t-1}}\right) \times 100\% \tag{2.4.1}$$

$$\text{Log Return} = \log\left(\frac{\text{price}_t}{\text{price}_{t-1}}\right) \tag{2.4.2}$$

Where:

$$\text{price}_t : \text{Price at time } t$$
$$\text{price}_{t-1} : \text{Price at time } t-1$$

The volatility of Bitcoin was calculated to be 0.0374 using the first method and 0.03778 using the second method. This value indicates that the price of Bitcoin exhibits a moderate level of volatility during the given period.

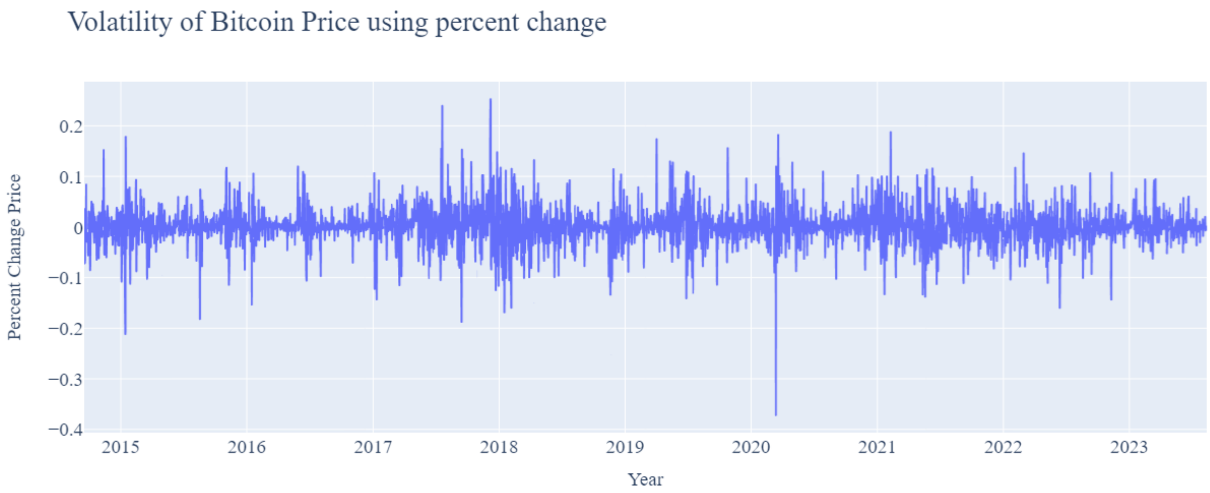


Figure 2.16: Exploring Bitcoin Price Volatility Over Time

2.4.4 Trend and Seasonality Analysis

Trend and seasonality are two important components in time series analysis to break down and understand the underlying patterns and variations within a time series that changes over time. The trend represents the long-term movement of the data. It captures the overall upward or downward movement over an extended period. Trends show the general pattern or tendency of the data to increase or decrease over time. Trends can be linear (constant rate of change) or nonlinear (changing rate of change). Linear trend line is given by the formula 2.4.3.The price of Bitcoin exhibited an upward trend during the period from 2015 to 2023. This trend is visually depicted as a linear upward trajectory, as evident in Figure 2.17. Seasonality refers to recurring and predictable patterns or fluctuations that occur at regular intervals within a time series data. These patterns can be influenced by factors like holidays, weather changes, or other calendar-based events. Seasonality repeats over specific periods, such as days, weeks, months, or years. Bitcoin demonstrates a certain level of seasonality that was identified through the STL - Seasonal and Trend decomposition using Loess process. To mitigate the effects of this seasonality, a Box-Cox transformation was applied. Seasonal decomposition of original and transformed bitcoin price is visualised in the figure **??**. The Box-Cox transformation is defined by the formula in 2.4.4

$$y_t = a + bt \tag{2.4.3}$$

Where:

y_t : is the observed value at time t

a : is the intercept

b : slope of the trend line

t : represents time

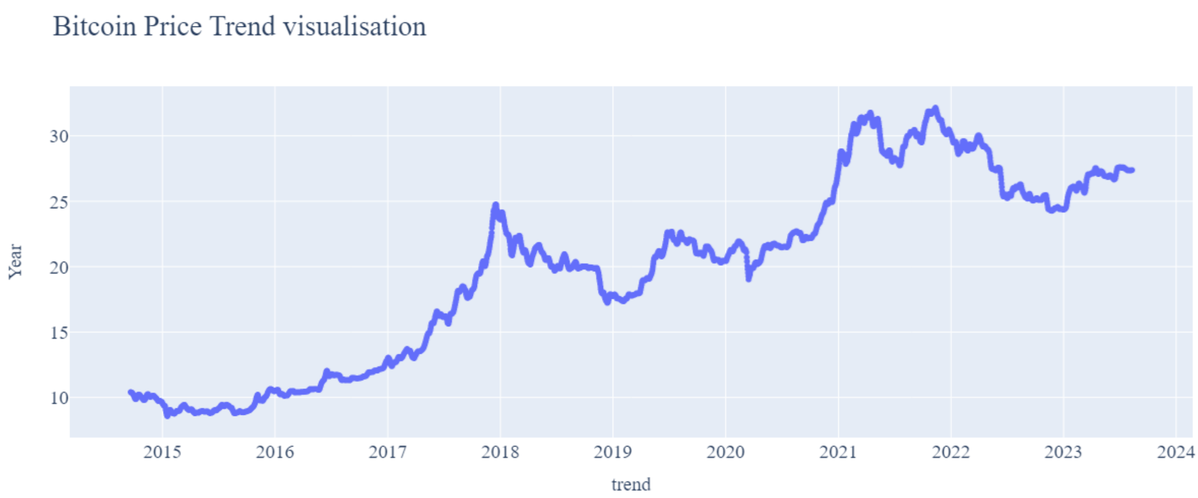


Figure 2.17: Bitcoin Price Trend Over Time

$$f(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(x) & \text{if } \lambda = 0 \end{cases} \tag{2.4.4}$$

Where:

x : is the original data point

λ : is the transformation parameter

Figure 2.18: Seasonal Decomposition of Bitcoin's Adjusted Price

Figure 2.19: Seasonal Decomposition of First Order Differentiated Adjusted Price of Bitcoin

2.4.5 Stationarity Analysis

This section aimed to assess the stationarity of cryptocurrency price data using the Augmented Dickey-Fuller (ADF) test. Stationarity is a fundamental concept in time series analysis, where the statistical properties of a dataset remain consistent over time. Non-stationary data can exhibit trends, cycles, or seasonality, making accurate analysis challenging. The test performed at 95% confidence level and T-statistic, along with the p-value, was calculated. Null and alternative hypotheses are given by:

$$\text{Null Hypothesis } (H_0) : \text{The time series has a unit root (non-stationary)}$$
$$\text{Alternative Hypothesis } (H_1) : \text{The time series is stationary.}$$

Test	**Value**
Dickey-Fuller Test on Adj Close Data	
Test Statistic	-1.560438
P-Value	0.503430
Dickey-Fuller Test on Period 1 Differenced Data	
Test Statistic	-9.555227
P-Value	0.000000

Table 2.2: Exploring Bitcoin's Price Stability: Dickey-Fuller Test

The initial Bitcoin price data exhibits non-stationarity, indicated by the p-value that is less than 0.05. Upon applying the first order differentiation to the time series data of adjusted close price leads to the achievement of stationarity. This transformation is evident from the p-value, which reached an extremely low value of 0.00. Additionally, the test statistic shifted significantly from -1.560438 to -9.555227. This substantial change in the test statistic provides strong evidence of the successful mitigation of non-stationarity in the Bitcoin price data after the application of the first differentiation.

Chapter 3

Methodology

This chapter is dedicated to the study to delves into the methodologies applied to forecast Bitcoin price and volatility through the utilization of various modelling techniques. This chapter elucidates the predictive modelling methodologies used for Bitcoin price prediction, encompassing Bi-directional LSTM, ARIMA, and volatility prediction through GARCH models. Within each section, the sequence unfolds as it explains the preprocessing steps for the data, the fine-tuning of hyper-parameters, comprehensive performance analysis, and the examination of out-of-sample forecasting capabilities intrinsic to each respective model.

3.1 Data Transformations and Data Splitting

The input feature selected for the prediction of price and volatility in the following section of this study was the adjusted close price of Bitcoin. The historical data of Bitcoin's adjusted close price underwent scaling transformations. A normalization transformation was applied to scale the data using the Min-Max Scaler technique. This technique scales the data to a range [0, 1]) to prevent any feature from dominating others due to differences in their magnitudes. This transformation ensures that the data is uniformly distributed within the chosen range. The formula for Min-Max scaling is:

$$x_{\text{scaled}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{3.1.1}$$

Where:

$$
\begin{aligned}
x_{\text{scaled}} &: \text{is the scaled value} \\
x &: \text{is the original value} \\
x_{\text{min}} &: \text{is the minimum value of the feature} \\
x_{\text{max}} &: \text{is the maximum value of the feature}
\end{aligned}
$$

The transformed data then divided into two datasets: the training set and test set, with a distribution ratio of 95:5. Again training set was divided into training set and validation set in ratio of 90:10 by setting the validation split to 0.1 while fitting LSTM model and it was manually divided into training and validation for ARIMA and GARCH models. A larger portion of the data is allocated to the training set as this facilitates the machine learning model in capturing price trends more effectively. The decision to allocate a smaller proportion of data to the test and validation sets is driven by several considerations. Firstly, time-series prediction models excel at forecasting data for the immediate future rather than the distant future. Their efficacy tends to diminish as the prediction horizon extends. Consequently, a reduced amount of data is reserved for predictions during the validation phases. This ensures that the predictive performance assessment closely aligns with the models' intended operational scope, maintaining a realistic and practically relevant evaluation. Therefore, data split ratio is designed to strike a balance between optimal learning opportunities for the model and a realistic assessment of its predictive prowess, ensuring that it is well-equipped to make informed forecasts for the near future while maintaining a robust performance evaluation framework. The model was designed to predict a 7-day forecast for the upcoming periods. These predicted values were subsequently evaluated once the actual values are released on the yahoo finance.

3.2 Performance Metrics for Model Evaluation

Time series data, due to its inherent temporal structure, non-constant variance, seasonality, and trends, presents unique challenges for prediction and evaluation. Traditional evaluation metrics may not fully capture the characteristics of time series forecasting. As a result, specialized metrics was used to address these specific challenges. The performance metrics employed for evaluating time series value and volatility predictions are as follows:

Price Prediction Metrics

- **Mean Absolute Percentage Error (MAPE)**:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

- **Mean Absolute Error (MAE)**:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Root Mean Square Error (RMSE)**:

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

- **Forecast Bias**:

$$Forecast\ Bias = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

Where:

n is the number of data points.

y_i is the actual volatility value.

$\hat{y}_i$ is the predicted volatility value.

$\bar{y}$ is the mean of actual volatility values.

$\bar{\hat{y}}$ is the mean of predicted volatility values.

Volatility Prediction Metrics

- **Root Mean Square Error (RMSE)**:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_i - \hat{v}_i)^2}$$

- **Mean Absolute Error (MAE)**:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|v_i - \hat{v}_i|$$

- **Correlation Coefficient**:

$$\text{Correlation Coefficient} = \frac{\sum_{i=1}^{n}(v_i - \bar{v})(\hat{v}_i - \bar{\hat{v}})}{\sqrt{\sum_{i=1}^{n}(v_i - \bar{v})^2 \sum_{i=1}^{n}(\hat{v}_i - \bar{\hat{v}})^2}}$$

- **Percentage of Correct Sign Predictions**:

$$\text{Percentage of Correct Sign Predictions} = \frac{\text{Number of Correct Sign Predictions}}{n} \times 100$$

Where:

n is the number of data points.

y_i is the actual volatility value.

$\hat{y}_i$ is the predicted volatility value.

$\bar{y}$ is the mean of actual volatility values.

$\bar{\hat{y}}$ is the mean of predicted volatility values.

3.3 Bitcoin Price Prediction using Bi-directional LSTM Model

The Long Short-Term Memory (LSTM) architecture falls under the category of Recurrent Neural Network (RNN) models, designed to effectively capture, and retain long-term dependencies within sequential data. LSTM networks find notable application in tasks like language modelling and time series prediction, given their capacity to maintain information over extended sequences. Unlike traditional feedforward neural networks, LSTM networks incorporate memory cells that retain information, rendering them well-suited for tasks requiring understanding of sequential context. This property, along with their ability to selectively update information, makes LSTM models an optimal choice for predicting Bitcoin prices in this study. The capacity to handle long sequences, a distinguishing feature of LSTMs, aligns well with the time-dependent nature of financial data. The LSTM architecture also addresses challenges posed by vanishing gradients, a common issue in conventional RNNs. LSTM cells consist of several key components that allow them to process and store information over sequences.

Input Gate (i_t): Controls the flow of new information into the cell state.

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i)$$

Forget Gate (f_t): Controls which information from the previous cell state should be forgotten.

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f)$$

Output Gate (o_t): Determines what information from the cell state should be output.

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o)$$

Candidate Cell State (g_t): Creates a candidate vector of new information to be added to the cell state.

$$g_t = \tanh(W_{gx}x_t + W_{gh}h_{t-1} + b_g)$$

Updated Cell State (c_t): Combines the forget gate, input gate, and candidate cell state to update the cell state.

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

Updated Hidden State (h_t): Produces the new hidden state using the updated cell state.

$$h_t = o_t \odot \tanh(c_t)$$

Where:

x_t is the input at time step t.

h_{t-1} is the previous hidden state.

W and b are the weight matrices and bias terms for different gates and candidates.

σ is the sigmoid activation function.

$\odot$ represents element-wise multiplication.

$\tanh$ is the hyperbolic tangent activation function.

3.3.1 Data Preprocessing

Data preprocessing is crucial task for all kind of machine learning algorithms. LSTM model is designed for processing sequential data therefore the steps performed in data pre such as time series or text. To utilize LSTM (Long Short-Term Memory) effectively, the input data needs to be structured as a series of sequential data, with corresponding output labels representing the subsequent values in the series. To achieve this, a data reshaping and preprocessing step was performed using a sliding window approach with a window size of 60. As a result of this preprocessing, the input data took the shape of (3017, 60, 1), while the output data shape was (3017, 1). This methodology allowed the LSTM model to understand and learn patterns within the sequential data for accurate forecasting.

3.3.2 Model Training and Tuning: Strategies and Insights

The LSTM model, structured according to the architecture detailed in Table 3.1, was developed. Its hyperparameters underwent a process of optimization through iterative experimentation. Ultimately, the optimized values for the dropout rate within the Dropout layer were determined to be 0.1. The batch size was set at 256, and the training process spanned across 100 epochs. This careful calibration aimed to achieve the best possible performance and predictive accuracy for the model. The input layer defines the input shape for the sequence of data points. In this case, the model is configured to anticipate sequences comprising 60 data points. Each of these points encompasses a solitary feature, "Adjusted Close". Bidirectional LSTM was used for its ability to capture both past and future context of the input sequence. The layer has 120 units, which means it can learn intricate patterns and dependencies within the historical data. Dropout is a regularization technique used to prevent overfitting. Third layer Bidirectional LSTM was added to continuing to capture and learn temporal relationships in the data and similarly dropout layer to avoid overfitting. The Dense layer with a single unit serves as the output layer of the model. It maps the extracted features and temporal patterns learned by the previous layers into a single predicted value.

Layer	Output Shape	Param #
InputLayer	[(None, 60, 1)]	0
Bidirectional LSTM	(None, 60, 120)	29,760
Dropout	(None, 60, 120)	0
Bidirectional LSTM	(None, 60, 120)	86,880
Dropout	(None, 60, 120)	0
Bidirectional LSTM	(None, 120)	86,880
Dense	(None, 1)	121
Activation	(None, 1)	0

Table 3.1: LSTM Model Architecture

Following the model's training process, an examination of the train-validation loss was conducted. The results are depicted in Figure 3.1. The visualization in this figure reveals a decrease in the training loss as the training epochs progress. This decline indicates that the model's ability to predict the target values improved over time as it learned from the training data. Figure 3.2 illustrates the extent to which the LSTM model learned the underlying patterns and trends during its training phase. The outcome displayed by the fitted data appears to be satisfactory, suggesting that the model successfully learned and replicated the intricate patterns present in the training dataset.

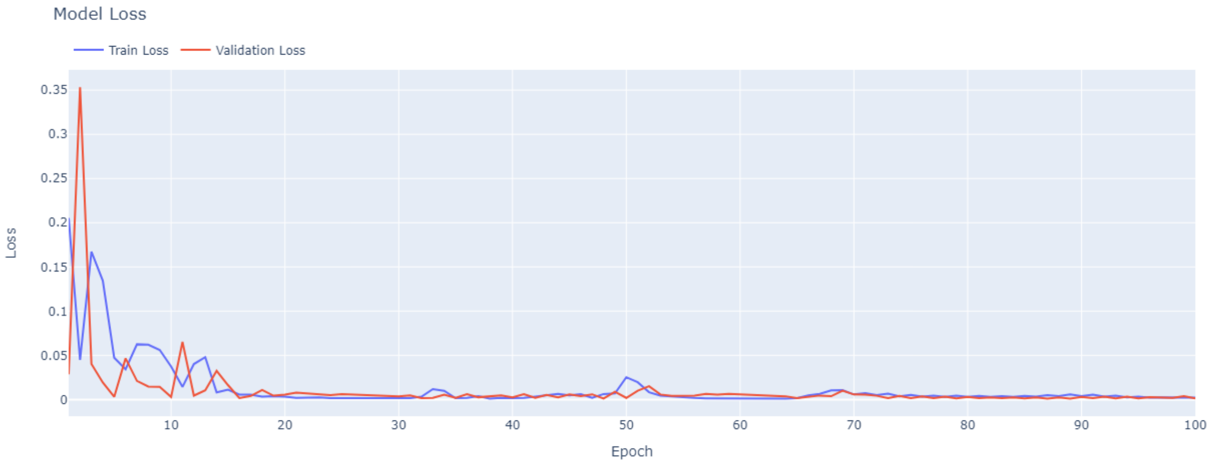


Figure 3.1: LSTM Training vs. Validation Loss visualisation

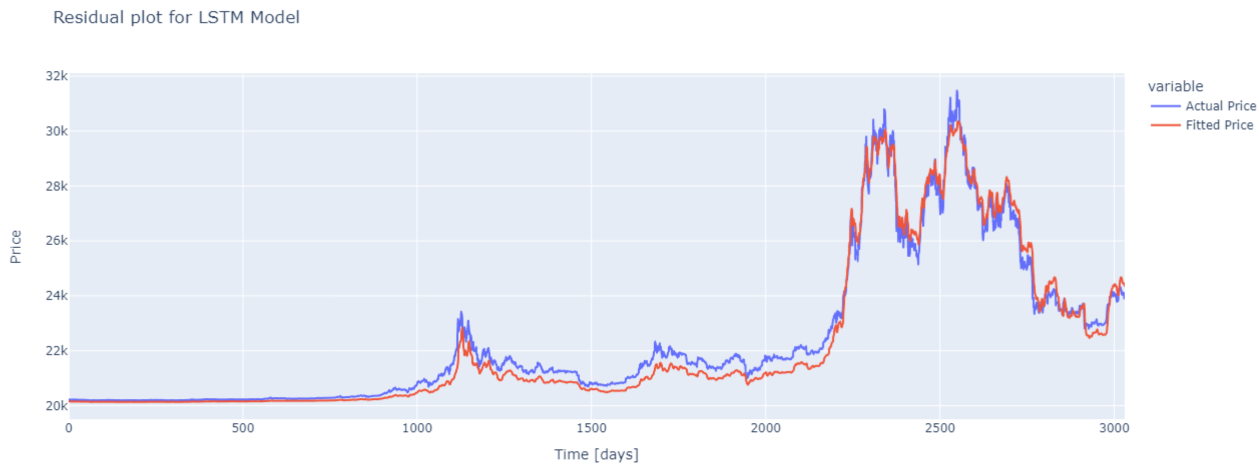


Figure 3.2: Contrasting Original Bitcoin Prices with LSTM Fitted Data

3.3.3 Performance Analysis

After the model had been trained using the training data, it underwent a testing phase where predictions were generated for the test set. This test set, comprising 101 distinct values, served as an independent standard to measure the model's performance. To visualize the results, the actual values were plotted against the model's predicted values. This visualization is represented in figure 3.3.

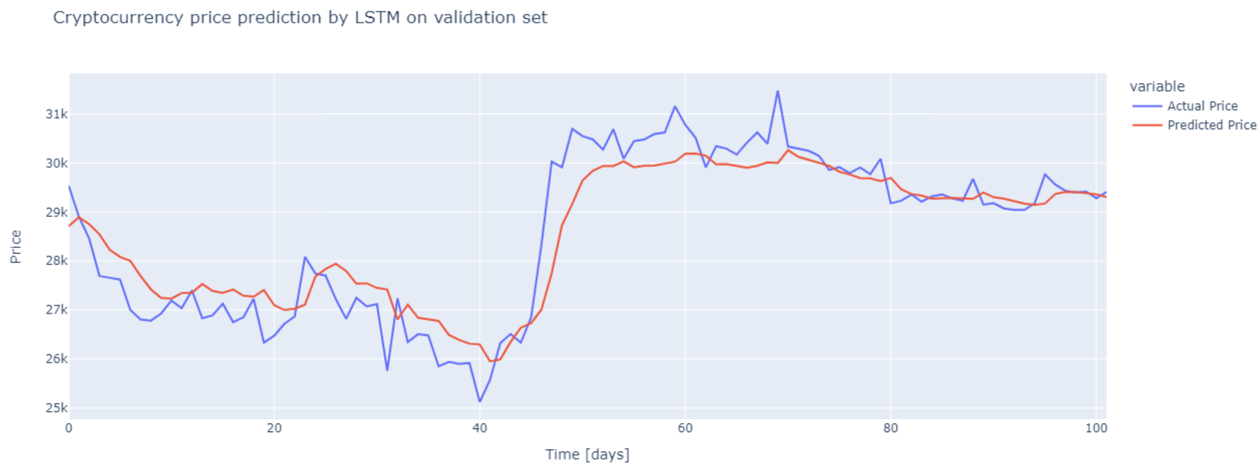


Figure 3.3: Model Performance: Visualizing LSTM Results on Test Data

Metric	Value
Mean Absolute Percentage Error (MAPE)	1.98
Mean Absolute Error (MAE)	572.63
Root Mean Squared Error (RMSE)	740.88
Explained Variance Score	0.828
Median Absolute Error	466.57
Symmetric Mean Absolute Percentage Error (SMAPE)	2.00
Forecast Bias	-296.42

Table 3.2: Performance Metrics of LSTM Model on Test Set

The performance metrics shown in table 3.2 demonstrate the LSTM model's predictive efficacy. The forecasts from the model differ from actual values by about \$572.63 on average, according to the Mean Absolute Error (MAE). The Root Mean Squared Error (RMSE) of 740.88 illustrates the square root of the average squared differences, demonstrating the size of the overall prediction error. With an Explained Variance Score of 0.828, the model effectively accounts for 82.8 percent of the variability in the data, demonstrating its capacity to identify underlying patterns. The model's strong predictive abilities are reaffirmed by the median absolute error of 466.57.11 and the low mean absolute percentage error (MAPE) and symmetric mean absolute percentage error (SMAPE) values of 1.98% and 2.00% respectively. Furthermore, a tendency toward slight underestimation in predictions is highlighted by the negative Forecast Bias of -296.42. Following the tuning of performance metrics, the obtained results matched the anticipated outcomes. Consequently, the decision was made to train the LSTM model using the complete dataset. This approach was adopted to enable the model to forecast values beyond the data it had been initially exposed to, providing a more comprehensive assessment of its predictive capabilities in real-world scenarios.

3.4 Bitcoin Price Prediction using ARIMA Model

The selection of the second model for Bitcoin price prediction was the ARIMA model. Its effectiveness in capturing a wide range of temporal patterns and trends renders it a versatile instrument applicable across various domains. The ARIMA (AutoRegressive Integrated Moving Average) model operates by disentangling and comprehending the underlying patterns within time series data, enabling accurate predictions. It synthesizes three fundamental components: AutoRegressive (AR), Integrated (I), and Moving Average (MA), each contributing distinct insights. The AR facet establishes predictive relationships by considering past values, while the Integrated facet involves differencing to achieve stationarity, mitigating trends and seasonal effects. The MA element centres on the interplay between forecast errors and subsequent values. These components amalgamate into a unified ARIMA equation, symbolized as ARIMA(p, d, q), where 'p' signifies AR order, 'd' represents differencing, and 'q' signifies MA order.

- **AutoRegressive (AR) Component**

  The AR component focuses on the relationship between a data point and its past values. The assumption of AR model is that the future value of a time series can be predicted based on its previous values. The "order" of the AR component specifies how many past values are used for prediction. Each past value is multiplied by a corresponding coefficient and added together, along with a white noise error, to predict the next value. White noise is a term used to describe a particular kind of random signal or collection of data points that exhibit characteristics of randomness but have no obvious pattern or structure. In terms, it is a set of data in which there is no correlation between successive values and each value is independently and identically distributed. AR(p) is represented by :

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t \tag{3.4.1}$$

Where:

x_t is the value of the time series at time t,

c is a constant term,

$\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients,

ε_t is the white noise error term at time t with distribution $\varepsilon_t \sim N(0, \sigma^2)$.

- **Moving Average (MA) Component**

  The MA component examines the relationship between a data point and its forecast errors. The assumption of MA model is that the future value of a time series is influenced by past forecast errors. The "order" of the MA component, denoted as MA(q), determines how many past errors are considered for prediction. Each past error term is multiplied by a coefficient and summed to forecast the next value. MA(q) is represented by :

$$x_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{3.4.2}$$

Where:

x_t is the value of the time series at time t,

c is a constant term,

ε_t is the white noise error term at time t, with distribution $\varepsilon_t \sim N(0, \sigma^2)$.

$\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients.

- **Integrated (I) Component**

The Integrated component involves differencing the time series data to make it stationary. Stationarity is crucial for many time series models, as it stabilizes the mean and variance of the data, making it easier to analyse and forecast. Differencing entails subtracting the current value from the previous value, leading to a new series of differences. This process is repeated until the series becomes stationary or exhibits no strong trend or seasonality., I(d) represented by :

$$\Delta x_t = x_t - x_{t-1} \tag{3.4.3}$$

Where:

Δx_t represents the differenced series at time t,

x_t is the value of the time series at time t,

x_{t-1} is the value of the time series at time $t-1$.

- **Final ARIMA Model**

  Combining the AR, I, and MA components, the general ARIMA model is denoted as ARIMA(p, d, q):

$$\Delta x_t = c + \phi_1 \Delta x_{t-1} + \phi_2 \Delta x_{t-2} + \ldots + \phi_p \Delta x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{3.4.4}$$

Where:

Δx_t represents the differenced series at time t,

c is a constant term,

$\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients,

ε_t is the white noise error term at time $t, \varepsilon_t \sim N(0, \sigma^2)$

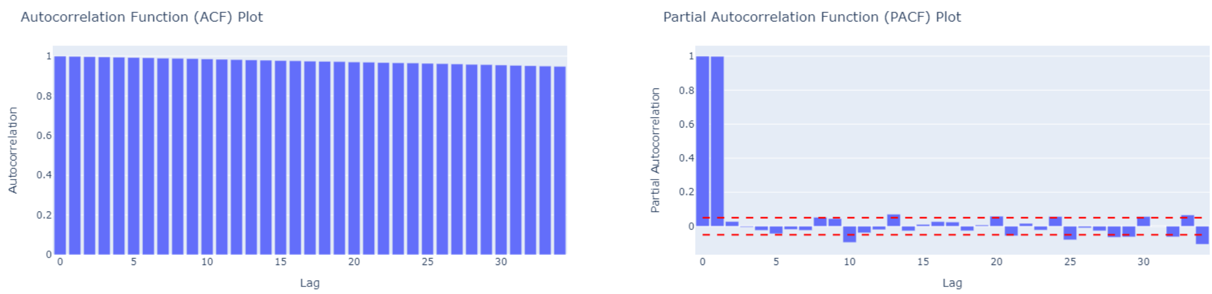
$\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients.

3.4.1 Data Preprocessing

ARIMA models were constructed utilizing univariate time series data extracted from the "Adj Close" prices of Bitcoin. The data was configured to operate on a daily frequency prior to executing the model fitting and fine-tuning processes. The training data which constitute 95% of full data was then divided into validation set in ratio of 90:10 to use in model tuning.

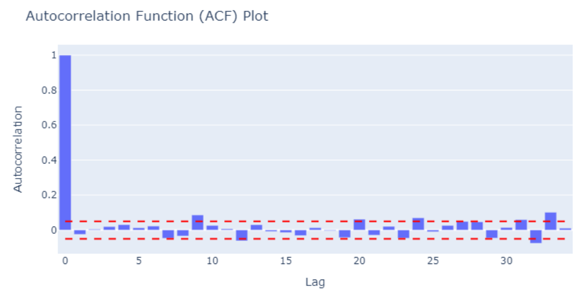
3.4.2 Model Training and Tuning Strategies

The ARIMA model needs three hyper parameters (p,d,q). The value of d was determined as 1 as shown in the table 2.2 the result of ADF test which shows that after one time differentiation of bitcoin price the time series reached stationarity. To find the order of the autoregressive (AR) and Moving average (MA) components, this study used the Partial Autocorrelation Function (PACF) and Auto-correlation plots. The ACF shows the correlation between the current value and lagged values of the time series. The PACF helps identify the direct relationship between the current value of a time series and its lagged values, while controlling for the effects of intermediate lags. It specifically shows the correlation between the current value and its lag after removing the effects of shorter lags. The figures 3.6 and **??** represent the PACF and ACF plot of Ajd close price along with first degree of differentiation and second degree of differentiation. The plots clearly indicate that employing a second-degree of differentiation results in excessive differentiation. Thus, the appropriate choice is to use a first-degree differentiation, which has been confirmed once again. Both the ACF and PACF plots highlight that the correlation coefficient for the 9th lag is 0.09, surpassing the established significance level of 0.05.

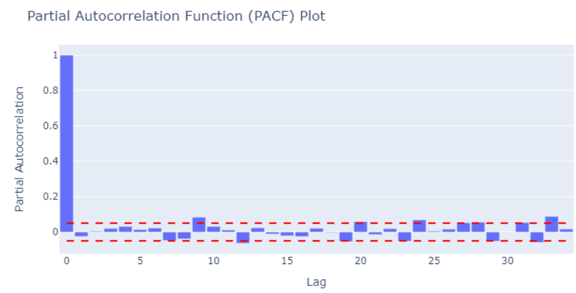


(a) Auto-correlation plot (b) Partial Auto-correlation plot

Figure 3.4: ACF and PACF Plots of Adj Close Values

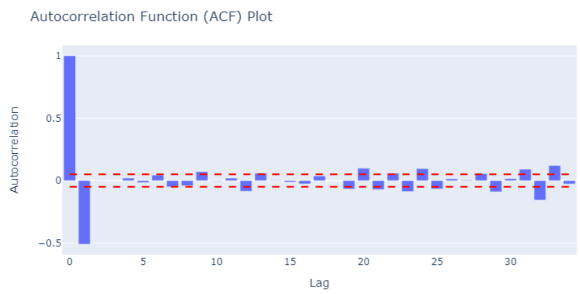


(a) Auto-correlation plot

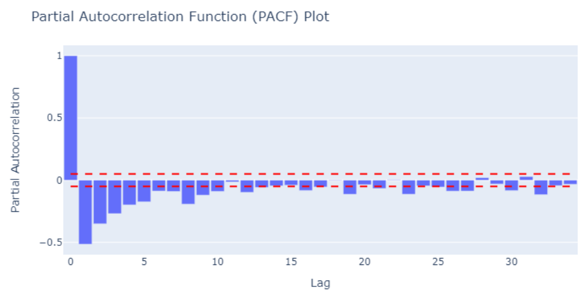


(b) Partial Auto-correlation plot

Figure 3.5: CF and PACF Plots of First-Order Differentiated Adj Close



(a) Auto-correlation plot



(b) Partial Auto-correlation plot

Figure 3.6: CF and PACF Plots of Second-Order Differentiated Adj Close

Consequently, a script was created to assess all potential combinations within the range of (0,9) for the values of p and q. Following this, the model summaries were scrutinized manually to examine the coefficients, their magnitudes, and the AIC value. This analysis led to the compilation of a set of possible parameter combinations. However, the model with the least AIC value did not converge during the optimization process, rendering it unreliable for prediction. As a result, this model was excluded from further consideration. Subsequently, the remaining list of prospective parameter combinations underwent validation by applying rolling predictions for the initial 10 rows of the validation set. The outcomes were then organized in Table 3.3, which showcases a subset of the potential parameter orders along with their corresponding AIC, MAPE, and MAE values.After a meticulous evaluation of the trade-offs encompassing significant coefficients, p-values, AIC, BIC, and MAPE, it was ascertained that the model characterized by hyperparameters (0,1,9) stood as the optimal choice for forecasting Bitcoin prices. This determination was underpinned by its superior performance in terms of having the lowest MAPE, MAE and AIC value, and a balanced model complexity that retained all significant coefficients.

ORDER	AIC	MAE	MAPE
(0, 1, 9)	45207.277	1331.00	3.84
(1, 1, 9)	45206.272	1337.15	3.86
(3, 1, 9)	45198.487	1341.72	3.86
(2, 1, 7)	45198.684	1343.86	3.86
(3, 1, 8)	45199.197	1341.46	3.86
(1, 1, 3)	45234.280	1360.42	3.94
(1, 1, 8)	45221.950	1351.25	3.91
(1, 1, 2)	45233.470	1362.19	3.95
(0, 1, 4)	45232.629	1353.14	3.92

Table 3.3: ARIMA Model Orders Performance Metrics for Evaluation and Comparison

3.4.3 Performance Analysis

The performance of the optimized ARIMA models was assessed initially on a subset of the validation set. Subsequently, the models were validated on the complete validation dataset using a rolling prediction approach, and the results were compared against the ground truth values. Similarly, model was evaluated on the unseen data using test set in rolling prediction way. Table 3.4 presents the performance metrics for the ARIMA model, while figure **??** provides a visual representation of the predicted values alongside the ground truth data of validation set and test set. The comparison between the two sets shows that the model performs relatively improved on the test set in terms of accuracy and bias as indicated by the lower values of MAE, MSE, RMSE, MAPE, and SMAPE.

Figure 3.7: ARIMA Model Predictions vs. Actual Validation Data

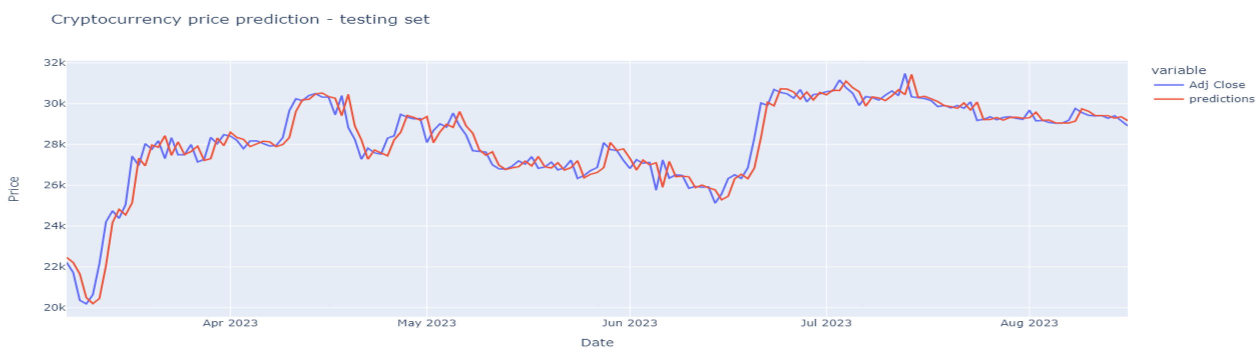


Figure 3.8: ARIMA Model Predictions vs. Actual Test Data

Metrics	**Validation Set**	**Test Set**
Number of observations	308	162
Mean Absolute Error (MAE)	501.8446	434.8645
Mean Squared Error (MSE)	606671.9572	380100.7395
Root Mean Squared Error (RMSE)	778.8915	616.5231
Mean Absolute Percentage Error (MAPE)	2.1653	1.6048
Symmetric Mean Absolute Percentage Error (SMAPE)	2.1523	1.6101
Forecast Bias	45.9333	-27.0016

Table 3.4: ARIMA model Performance Metrics

3.5 Model Comparision and Forecasting Out-of-Sample

After thoroughly comparing the outcomes of both models, it can be concluded that both models perform competitively in terms of their results. Notably, the ARIMA model demonstrates slightly better performance metrics on the test set. To forecast values for 7 days ahead, both the ARIMA and LSTM models were utilized. This was achieved using a rolling prediction method, where the forecast for one day was incorporated into the training set before retraining the model. The Figure 3.9 visually presents the out-of-sample forecasted values produced by these models and actual price of the bitcoin. Table **??** contains the vlaues predicted by the models and actual price of the bitcoin during that period. The table 3.6 show the performance of both the models on out-of-sample data. The comparison of the two model's performance metrics indicates that the ARIMA model generally outperformed the LSTM model in terms of MAE, MSE, RMSE, MAPE, and SMAPE. These results suggest that, based on the given dataset and evaluation metrics, the ARIMA model is more accurate in predicting cryptocurrency prices.

Date	ARIMA Forecast	LSTM Forecast	Actual Close
2023-08-01	29308.15	29435.45	29675.73
2023-08-02	29218.54	29502.55	29151.96
2023-08-03	29220.79	29512.80	29178.68
2023-08-04	29230.05	29510.75	29074.09
2023-08-05	29211.44	29585.31	29042.13
2023-08-06	29222.27	29573.05	29041.86
2023-08-07	29230.73	29575.51	29180.58

Table 3.5: Comparison of Forecasted Values and Actual Observations

Model	MAE	MSE	RMSE	MAPE	SMAPE	Forecast Bias
LSTM	223.58	55376.20	235.32	0.76%	0.76%	134.78
ARIMA	147.44	32768.05	181.02	0.50%	0.50%	42.42

Table 3.6: Performance Metrics of Models on Forecasted Data

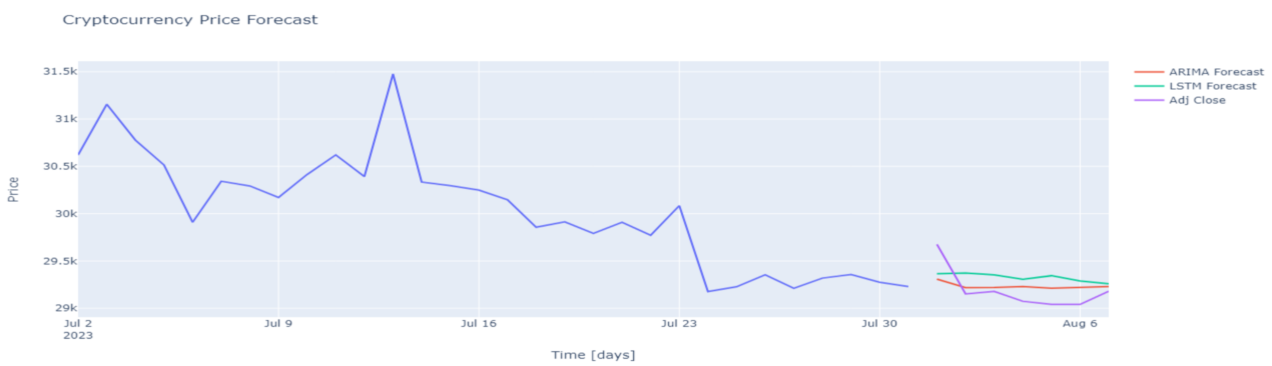


Figure 3.9: Bitcoin Out-of-Sample Forecast: Predicted vs. Actual Prices

3.6 Bitcoin Volatility Prediction using GARCH model

After successfully predicting the price of Bitcoin, the subsequent task pursued in this study was to forecast the volatility of Bitcoin's price. This endeavor holds substantial interest due to its relevance across diverse domains. The primary significance lies in its profound influence on investment and trading strategies. Traders and investors lean on volatility predictions to make informed decisions, leveraging opportunities derived from heightened volatility while also factoring in the reliability of lower volatility conditions. In addressing this challenge, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model emerged as a method of choice. The GARCH model was selected for predicting Bitcoin's volatility due to its ability to capture the intricate dynamics of volatility changes over time. The way this model handles time-varying volatility fits the characteristics of cryptocurrency markets, which are known for their erratic fluctuations. This study used the GARCH model to gain important understanding of Bitcoin's volatility trends. Since cryptocurrency markets are inherently volatile, this method makes it especially suitable for them because it enables the model to adapt to changing market conditions. It is represented in by equation 3.6.3. The GARCH model captures the conditional variance at each time step as a function of past squared residuals (ϵ_{t-i}^2) and past conditional variances (σ_{t-j}^2), weighted by the coefficients α_i and β_j respectively. The AVGARCH model makes the assumption that the distribution of the standardized residual follows a standard normal distribution, characterized by a Gaussian distribution having a mean of 0 and a variance of 1.

$$y_t = \mu_t + \epsilon_t \tag{3.6.1}$$

$$\epsilon_t = \sigma_t z_t \tag{3.6.2}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{3.6.3}$$

Where:

y_t the observed value at time t.

μ_t the mean value at time t.

ϵ_t the standardized residual (shock or innovation) at time t.

σ_t^2 the conditional variance at time t.

z_t a white noise error term with zero mean and unit variance.

p the order of the autoregressive part.

q the order of the moving average part.

ω the constant term in the variance equation.

α_i the coefficients of the autoregressive terms for the squared residuals.

β_j the coefficients of the moving average terms for the conditional variance.

3.6.1 Preprocessing and model tuning

This section aims to design a model that account for fluctuations in volatility as time progresses, and this ability is of significant importance when analysing financial data that evolves over time. In the conducted study, the derived column "Percentage Change Price" was employed as an input feature to effectively capture the volatility present in the Bitcoin price. Figure 3.10 visually presents the variations in the percentage change of the Bitcoin price, along with its corresponding distribution pattern. Notably, a slight negative skewness of -0.142015 was noted within this column. This observation suggests that the "Percentage Change Price" column holds promising potential for training a GARCH model aimed at predicting volatility. Volatility in the percent change price column was calculated by standard deviation, formula used is displayed in the equation 3.6.4.

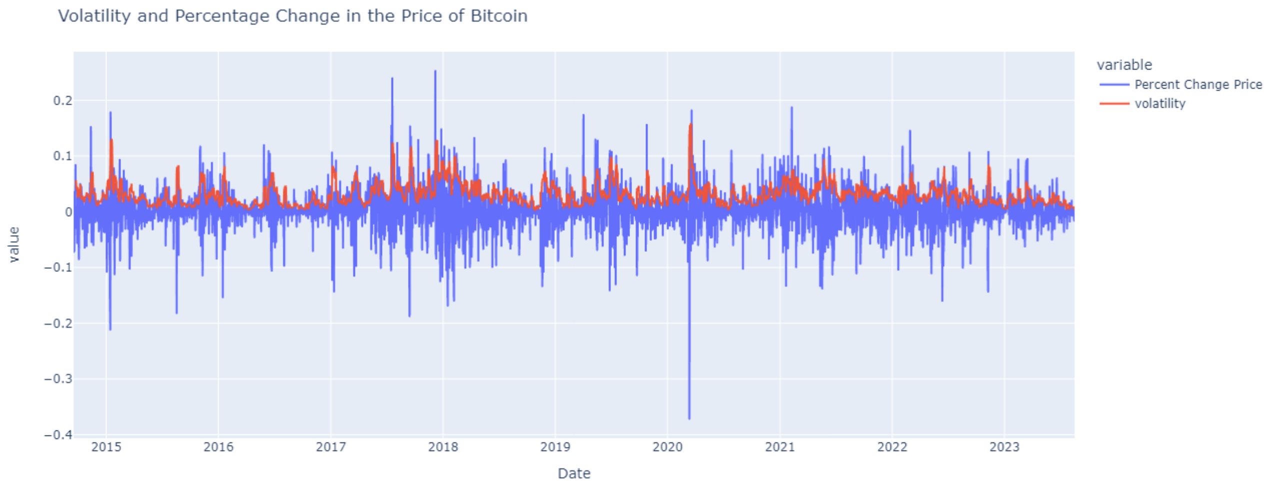


Figure 3.10: Visualising the Volatility: Exploring Bitcoin Price Changes and Standard Deviation

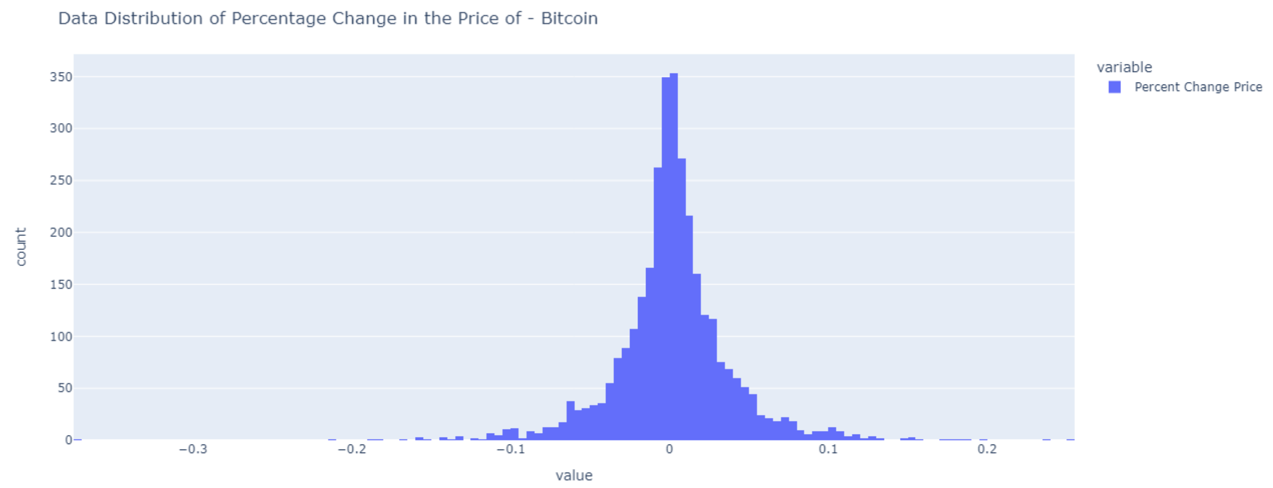


Figure 3.11: Visual Analysis of the Distribution of Bitcoin Price Percent Changes

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3.6.4}$$

where:

σ is the standard deviation,

n is the number of data points,

x_i represents each individual data point,

$\bar{x}$ is the mean of the data points.

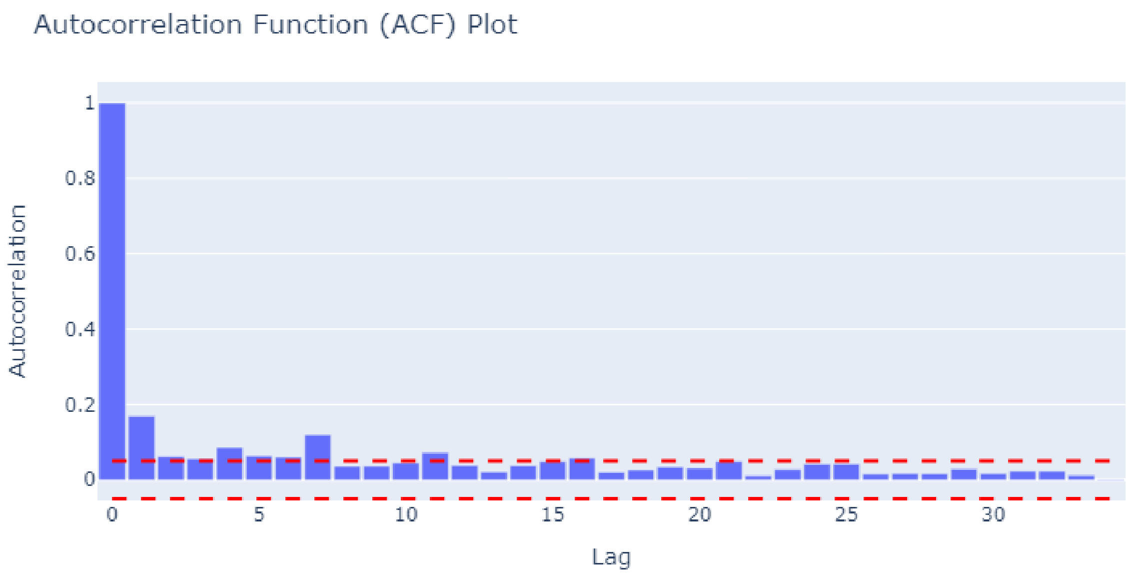


Figure 3.12: ACF Plot to Estimate Optimal (p,q) Parameters

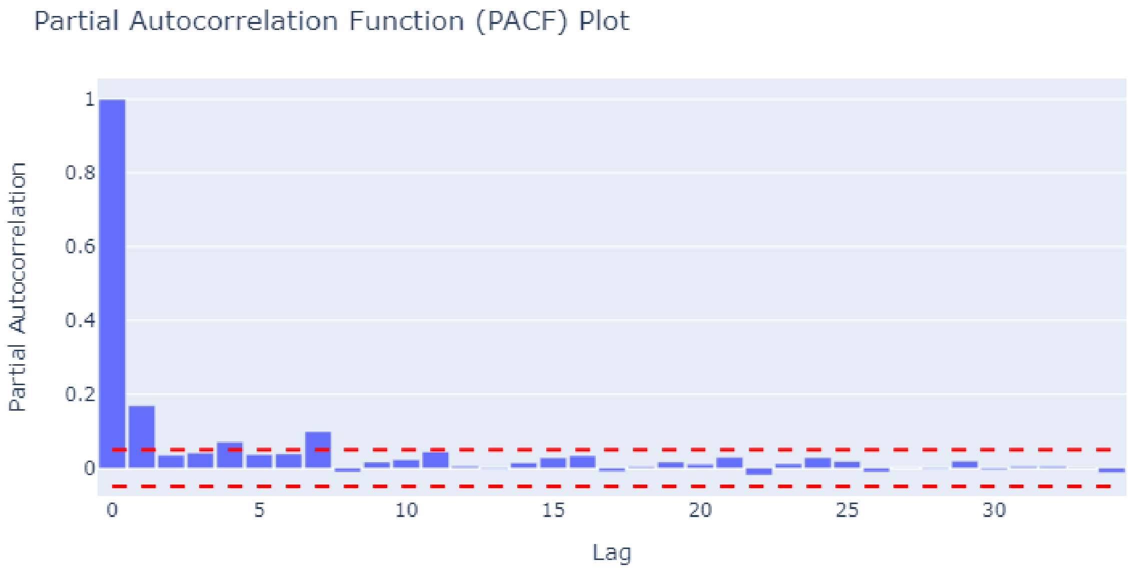


Figure 3.13: PACF Plot to Estimate Optimal (p,q) Parameters

The parameters (p,q) of the GARCH model, which represent the Order of Autoregressive Terms and the Order of Moving Average Terms, are determined through the analysis of the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. After identifying potential (p,q) values, a systematic procedure is applied to examine all possible combinations of (p,q) values on the validation dataset. This process helps in identifying the optimal hyper-parameter values. The table 3.7 contains the subset of p and q combination sorted in order of decreasing Correlation Coefficient values. after performing trad-off between model complexity, AIC, MAE, Correlation Coefficient and model coefficients and its p-values the optimised parameter was decided as p=3 and q=1.

3.6.2 Model evaluation on Test set

The ultimate model was subsequently trained on both the training and validation datasets combined. Its performance was then assessed using the test set. The figure 3.14 plots the actual volatility and the predicted volatility in the percentage change of bitcoin price. The evaluated model's performance is shown in table 3.8

p	q	AIC	MAE	Correlation Coefficient
5	1	-11931.7638	0.0086	0.8818
3	1	-11872.6815	0.0084	0.8817
4	1	-11933.7638	0.0085	0.8854
3	2	-11919.8800	0.0090	0.8645
2	2	-11921.7484	0.0090	0.8634
6	2	-11905.9738	0.0087	0.8599

Table 3.7: Shortlisted 10 Model Performance Metrics

Figure 3.14: Visualising Forecasted Bitcoin Return Volatility using GARCH(3,1) on Test Data

Metric	Value
Mean Absolute Error (MAE)	0.0087
Root Mean Squared Error (MSE)	0.00010
Root Mean Squared Error (RMSE)	0.01060
Correlation coefficient	0.7707
Percentage Correct sign	100.0%

Table 3.8: GARCH Model Evaluation Metrics on Test Data

3.6.3 A variation of GARCH model - VAGARCH

The residual of fitted GARCH model was tested for its distribution using visual approach and Anderson test. The Figures 3.15 and 3.16 represent the model residual and its distribution plot respectively. The test statistics of Anderson test was 56.35 which show that residual is not normally distributed. Hence, to further improve the volatility prediction accuracy different variation of GARCH model was used apart from the fitted GARCH(3,1) model. The next model implemented was VAGARCH which stands for Asymmetric Volatility Generalized Autoregressive Conditional Heteroskedasticity with 'dist' parameter set to "generalized error". After performing the model hyper-parameter tuning using validation the optimised value of (p,q) was (1,1). the model summary of optimised AVGARCH model is presented in Table **??** which shows that the coefficients are significant. The model's performance of validation set, and test set are shown in Table 3.9. The actual volatility versus predicted volatility in test and validation dataset are depicted in the figure

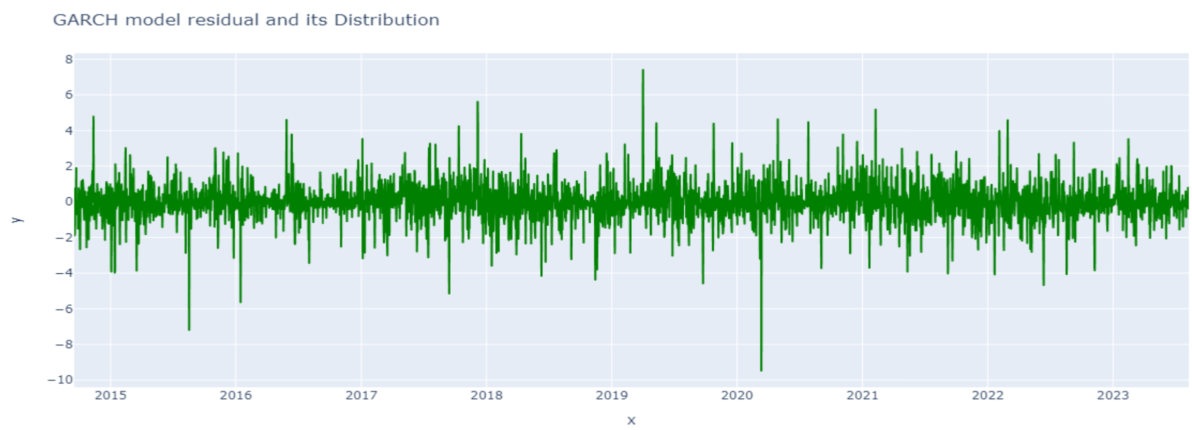


Figure 3.15: GARCH(3,1) Model Fit: Standard Residuals for Volatility Assessment

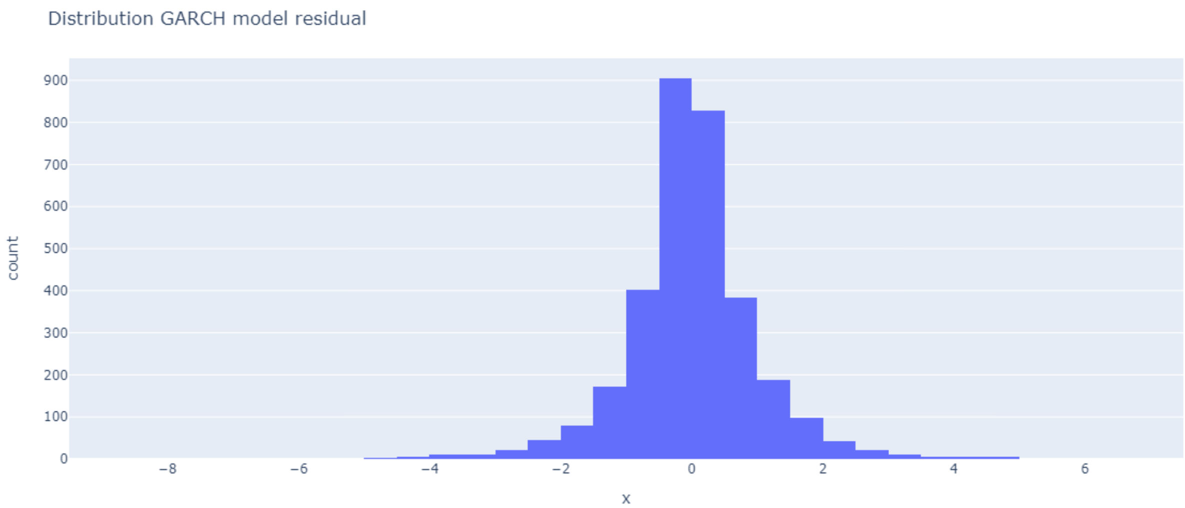


Figure 3.16: Distribution of Standard Residuals in GARCH(3,1) Volatility Model

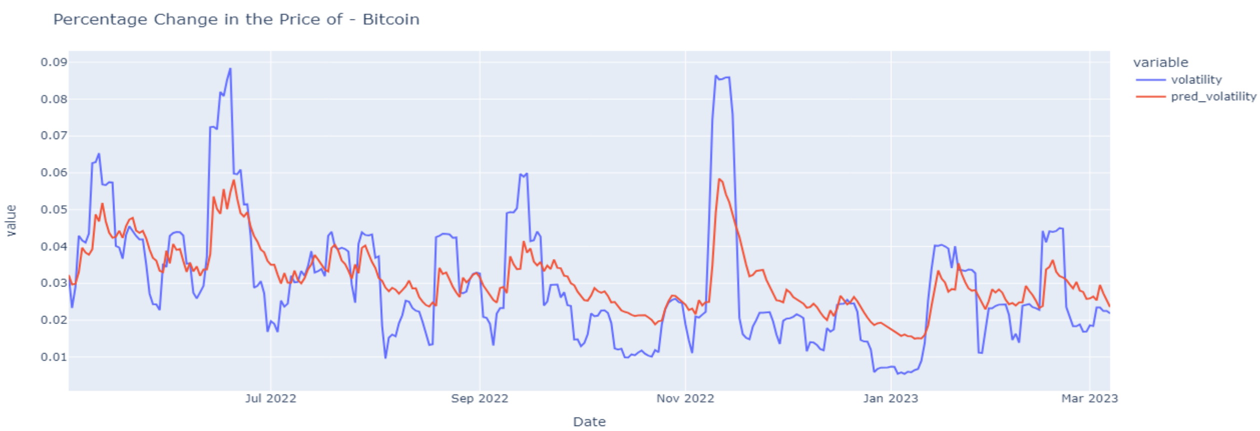


Figure 3.17: AVGARCH Model Performance on Validation Set

Figure 3.18: AVGARCH Model Performance on Test Set:

	MAE	MSE	RMSE	Correlation Coefficient
Validation Performance	0.00859	0.00012	0.01109	0.8327
Test Performance	0.0078	0.00008549	0.0094	0.68

Table 3.9: AVGARCH Model Performance Metrics on Validation and Test sets

3.7 Out-of-Sample Volatility prediction Model comparison

Both the models were used to forecast the volatility in future for 7 days. Each model's outcome was plotted in the figure 3.19. These values were then tested with real time data and results are discussed in Ressult section of this paper. Based on the presented metrics, the AVGARCH (Asymmetric GARCH) model exhibits enhanced predictive capabilities in comparison to the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model. The AVGARCH model achieves superior performance across multiple facets, including lower Mean Absolute Error (MAE) and Mean Squared Error (MSE), as well as a smaller Root Mean Squared Error (RMSE). Moreover, the AVGARCH model displays a higher correlation coefficient, indicating a stronger alignment between predicted and actual values. Both models achieve a perfect percentage score for correct sign prediction. These findings collectively underscore the AVGARCH model's heightened accuracy and efficacy in capturing the dynamics of the data, affirming its potential for more accurate price movement predictions.

Date	AVGARCH Forecast	GARCH Forecast	Real Volatility
2023-08-01	0.015593	0.022513	0.006676
2023-08-02	0.016289	0.023062	0.009894
2023-08-03	0.016945	0.024154	0.009749
2023-08-04	0.017551	0.025643	0.009661
2023-08-05	0.018125	0.026824	0.009602
2023-08-06	0.018784	0.027847	0.009598
2023-08-07	0.019357	0.028764	0.009844

Table 3.10: Volatility Forecast Comparison

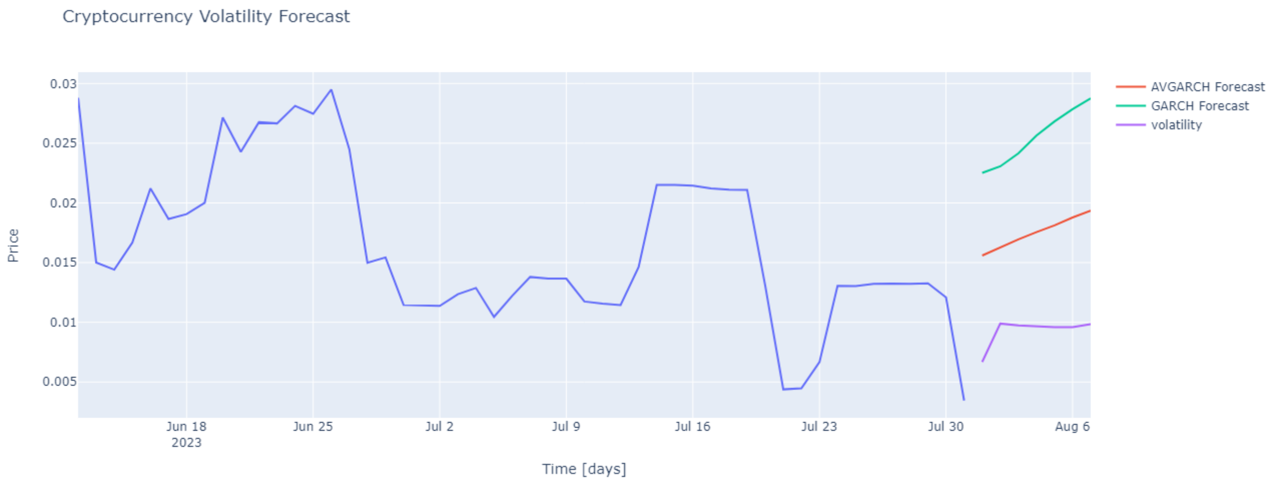


Figure 3.19: Comparing Out-of-Sample Volatility Forecasts of GARCH, AVGARCH, and Actual Volatility for Bitcoin Price

Model	MAE	MSE	RMSE	Correlation Coefficient	Percentage Correct Sign
GARCH	0.01627	0.00027	0.01638	0.52236	100.0%
AVGARCH	0.00821	6.8852e-05	0.00827	0.5999	100.0%

Table 3.11: GARCH and AVGARCH Model Performance Metrics for Out-of-Sample Data

CHAPTER 4

Future Work

The current study has shed light on how to forecast volatility and price of cryptocurrencies using sophisticated time series models. There are a number of opportunities for additional exploration and improvement, even though the current research has produced encouraging results. Utilizing the results of cryptocurrency clustering to enable the effective prediction of multiple cryptocurrencies is a promising direction for future research. With this method, it is possible to use the clustering findings from the training data to forecast the prices of a number of cryptocurrencies that have a lot in common. It becomes possible to use a single predictive model for each cluster by grouping cryptocurrencies with comparable market behavior and price patterns, which significantly reduces the complexity involved in maintaining separate models for each cryptocurrency. Within the scope of this study, a same set of hyperparameters was employed to deploy the LSTM model for Ethereum price prediction, considering its placement within the same cluster as Bitcoin. Encouragingly, the outcomes obtained were notably promising. Nevertheless, it is worth noting that additional refinement and meticulous tuning remain imperative, especially when venturing into the domain of volatility prediction. This dynamic area calls for further optimization to ensure the precision and reliability of the model's predictions. Implementing a procedure that uses the knowledge gained from clustering to speed up the prediction process is the proposed task. To find groups of cryptocurrencies behaving similarly, an extensive analysis of the clustering results would be done first. A predictive model could be trained specifically for each cluster using the representative cryptocurrency for that cluster. The historical data of the selected representative would be used to inform the model's training, and its predictions would be applied to the entire cluster. This strategy may result in fewer distinct predictive models being needed, which would make maintenance easier and increase efficiency. By addressing the challenge of predicting multiple cryptocurrencies while minimizing complexity, this task holds the potential to offer valuable contributions to the field of cryptocurrency.

Through the combination of ensemble machine learning methods and time series models, predictive accuracy may be improved. This strategy can produce more accurate predictions by combining the advantages of various models and addressing the weaknesses of each one separately. Examples of ensemble methods include bagging, boosting, and stacking techniques. Additionally, the addition of outside variables has potential for improving predictions. The model may be able to capture the impact of real-world events on cryptocurrency prices and volatility by exploring the integration of macroeconomic indicators, social media sentiment analysis, and other pertinent external data, leading to more thorough and accurate predictions. Exploring hybrid models, which combine conventional time series methods with machine learning strategies, can take advantage of both paradigms' strengths and potentially improve prediction accuracy.

It is suggested that a user-friendly web interface be developed to improve the usability and applicability of the research findings. Real-time forecasts, historical price visualizations, cryptocurrency clustering, and model comparison capabilities could all be offered by such an interface. Users would be able to choose and contrast various prediction models, giving them the power to decide based on each model's unique characteristics and capabilities and accommodating a range of user preferences. Additionally, transparency is improved and users are helped in determining the reliability of predictions by directly displaying model accuracy metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Correlation Coefficient on the web interface. It's also advisable to carry out robustness testing across various market conditions and timeframes to ensure model consistency and stability, validating the models' efficacy across various scenarios.

CHAPTER 5

Conclusions

In this study, historical data of cryptocurrencies was examined to uncover the intricate trends and patterns. The complex dynamics supporting the market for cryptocurrencies were revealed. The findings indicated that the cryptocurrency market exhibited high volatility, with the potential for drastic changes even within short timeframes. It was found that certain similarities existed among diverse cryptocurrencies, leading to analogous trends in their characteristics. These similarities arose from external market conditions that drove comparable price trends and volatility across these currencies. Upon analyzing and comparing the top 50 currencies ranked by Yahoo Finance, machine learning models were employed to analyze bitcoin's price and volatility. The adjusted closing price of bitcoin was used to train Bi-LSTM and GARCH models. The hyperparameters of the Bi-LSTM model were adjusted manually, and precautions were taken to ward off overfitting through the utilization of a validation set. Similarly, parameters for the ARIMA model were estimated and tuned. Both models exhibited competitive performance on the test datasets. The Bi-LSTM model achieved a Mean Absolute Percentage Error (MAPE) of 1.98, while the ARIMA model achieved a MAPE of 1.6 on the test data. The Bi-LSTM model displayed a mean absolute error of 572.63, whereas the ARIMA model exhibited an error of 434.86. Likewise, in terms of Root Mean Squared Error (RMSE), the ARIMA model outperformed the Bi-LSTM model. Significant differences in forecast bias were observed,the BI-LSTM prediction demonstrated a bias of -296, while the ARIMA prediction revealed a bias of -27.00. These values of performance matrices are attributed to the inherent characteristics and methodologies of the two models. The

Bi-LSTM model excelled in capturing intricate patterns when ample available data, whereas the ARIMA model proved proficient at short-term predictions. Both models demonstrated exceptional performance in predicting out-of-sample Bitcoin prices, achieving results that are remarkably close to the state-of-the-art benchmarks. The LSTM model achieved an impressive Mean Absolute Percentage Error (MAPE) value of 0.76 percent, while the ARIMA model achieved an even more remarkable MAPE value of 0.50 percent. These outcomes proofs the effectiveness and accuracy of both models in forecasting Bitcoin prices. The study also encompassed the application of rolling prediction, which led the ARIMA model to outperform the Bi-LSTM model both in the test set and in out-of-sample forecasts. Another facet of the study involved predicting bitcoin price volatility using the GARCH and AVGARCH models. These models demonstrated satisfactory performance in predicting volatility. The Mean Absolute Error (MAE) values for the GARCH and AVGARCH models on the test data were recorded as 0.0087 and 0.0078, respectively. The AVGARCH model displayed a slightly enhanced performance in out-of-sample predictions compared to the GARCH model, as indicated by its MAE of 0.0082, while the GARCH model achieved a higher MAE of 0.016. Despite the achievements of this study, there remains ample room for further enhancing the tuned models. New models could be fitted and tested to attain greater precision. The inclusion of external factors in the training data is anticipated to enhance model performance. Exploring hybrid models and ensemble methods also holds promise for future investigations.

CHAPTER 6

Appendix

6.1 Code and Source Data detail

The source code developed for the execution of this research can be readily accessed on the GitHub platform at the following link:
https://github.com/HumaNazneen/University-of-essex.

Datasource: The dataset is derived from real-time data acquired through the Yahoo Finance API, facilitated by the yfinance library in Python.

Code Execution and Environment: The implementation of the code involved in processing the dataset was meticulously carried out within the context of a Google Colab environment. Leveraging the connectivity with Google Drive and harnessing the computational power of a GPU are integral prerequisites for executing the code effectively.

Additionally, an alternate version of the code has been thoughtfully provided at the same path under the name 'CryptoCurrency_prediction_withoutGPU.ipynb'. This alternative version can be employed when access to GPU resources is not available, ensuring a seamless execution.

6.2 Correlation plot of top 20 cryptocurrencies

The correlation of top 50 cryptocurrency was analyse and below is the ordered correlation plot.

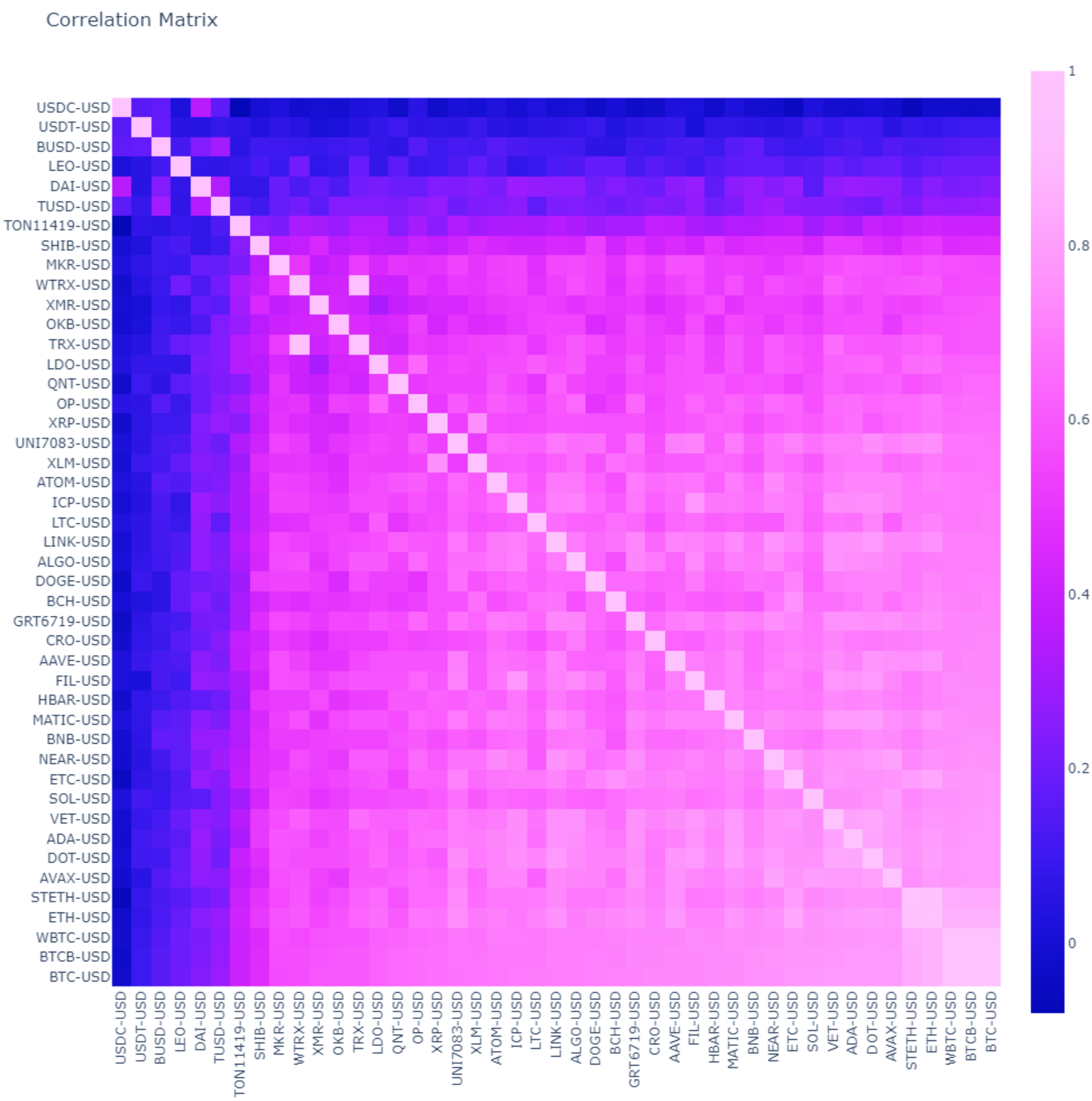


Figure 6.1: Comparing Out-of-Sample Volatility Forecasts of GARCH, AVGARCH, and Actual Volatility for Bitcoin Price

6.3 ARIMA(0,1,9) prediction result

The visual comparision between the original adjusted closing price of Bitcoin and the long-term prediction results derived from the ARIMA model, while excluding the utilization of rolling predictions, is elegantly showcased in Figure 6.2. In addition, Figure 6.3 provides a clear representation of the ARIMA model's predictive capability over a shorter span of

10 days. The insights drawn from these plots collectively underscore the proficiency of the ARIMA model in short-term forecasting, while also revealing its limitations in extending that accuracy to long-term predictions.

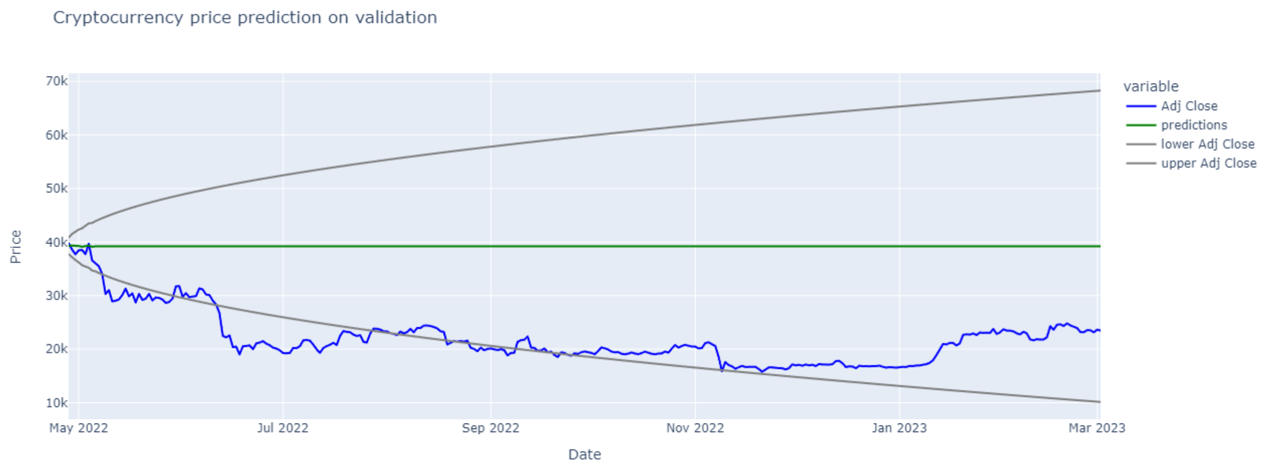


Figure 6.2: ARIMA model prediction for long term without rolling prediction

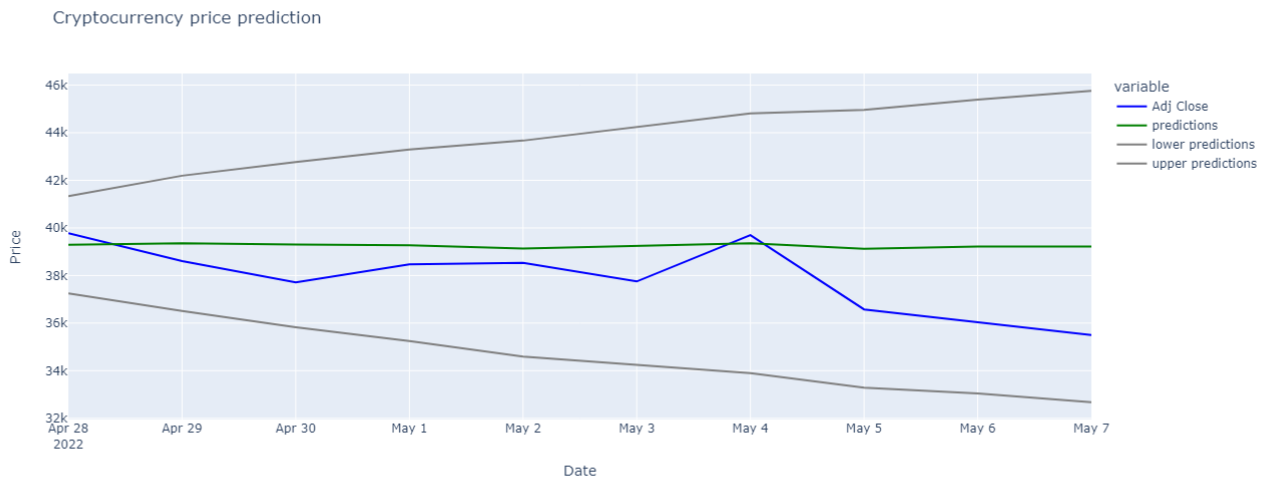


Figure 6.3: ARIMA model prediction for short term without rolling prediction

Bibliography

[1] Satoshi Nakamoto, 2008. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review.*

[2] Ott, R.L. and Longnecker, M.T., 2015. An introduction to statistical methods and data analysis. *Cengage Learning*.

[3] S. Park and J.S. Yang, 2023. Intelligent cryptocurrency trading system using integrated AdaBoost-LSTM with market turbulence knowledge. *Applied Soft Computing*, p.110568.

[4] G. Cohen, 2021. Optimizing Algorithmic Strategies for Trading Bitcoin. *Computational Economics*, 57, 639-654.

[5] M.C. Kristjanpoller, 2018. Minutolo, A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. *Expert Systems with Applications*, 109, 1-11.

[6] Y. Peng, P.H.M. Albuquerque, J.M.C. de SÃ¡, A.J.A. Padula, M.R. Montenegro, 2018. The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. *Expert Systems with Applications*, 97, 177-192.

[7] W. Chen, M. Jiang, W.-G. Zhang, Z. Chen, 2021. A novel graph convolutional feature based convolutional neural network for stock trend prediction. *Information Sciences*, 556, 67-94.

[8] M. CastÃ¡n-Lascorz, P. JimÃ©nez-Herrera, A. Troncoso, G. Asencio-CortÃ©s, 2022. A new hybrid method for predicting univariate and multivariate time series based on pattern forecasting. *Information Sciences*, 586, 611-627.

[9] S. Park and J.-S. Yang, 2022. Interpretable deep learning LSTM model for intelligent economic decision-making. *Knowledge-Based Systems*, 248, 108907.

[10] M.M. Patel, S. Tanwar, R. Gupta, N. Kumar, 2020. A deep learning-based cryptocurrency price prediction scheme for financial institutions. *Journal of Information Security and Applications*, 55, 102583.

[11] Z. Zhang, H.-N. Dai, J. Zhou, S.K. Mondal, M.M. GarcÃa, H. Wang, 2021. Forecasting cryptocurrency price using convolutional neural networks with weighted and attentive memory channels. *Expert Systems with Applications*, 115378.

[12] C. Xiao, N. Chen, C. Hu, K. Wang, J. Gong, Z. Chen, 2019. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sensing of Environment*, 233, 111358.

[13] S. Basco, 2014. Globalization and financial development: A model of the Dot-Com and the Housing Bubbles. *Journal of International Economics*, 92, 78-94.

[14] M. Iwamura, Y. Kitamura, T. Matsumoto, 2014. Is bitcoin the only cryptocurrency in the town? Economics of cryptocurrency and Friedrich A. Hayek. *SSRN Electronic Journal*.

[15] R. Chowdhury, M.A. Rahman, M.S. Rahman, M. Mahdy, 2020. An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning. *Physica A: Statistical Mechanics and its Applications*, 551, 124569.

[16] V. Derbentsev, V. Babenko, K. Khrustalev, H. Obruch, S. Khrustalova, 2021. Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices. *International Journal of Engineering*, 34, 140-148.

[17] W. Chen, H. Xu, L. Jia, Y. Gao, 2021. Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting*, 37, 28-43.

[18] P.L. Seabe, C.R.B. Moutsinga, E. Pindza, 2023. Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: a deep learning approach. *Fractal and Fractional*, 7(2), 203.

[19] A.M. Khedr, I. Arif, M. ElBannany, S.M. Alhashmi, M. Sreedharan, 2021. Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. *Intelligent Systems in Accounting, Finance and Management*, 28(1), 3-34.

[20] S. Basco, 2014. Globalization and financial development: A model of the Dot-Com and the Housing Bubbles. *Journal of International Economics*, 92, 78-94.

[21] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

[22] S. Pasak, R. Jayadi, 2023. Investment Decision on Cryptocurrency: Comparing Prediction Performance Using ARIMA and LSTM. *Journal of Information Systems and Informatics*, 5(2), 407-427.

[23] C. Zhong, W. Du, W. Xu, Q. Huang, Y. Zhao, M. Wang, 2023. LSTM-ReGAT: A network-centric approach for cryptocurrency price trend prediction. *Decision Support Systems*, 169, 113955.

[24] A. GarcÃa-Medina, E. Aguayo-Moreno, 2023. LSTM-GARCH Hybrid Model for the Prediction of Volatility in Cryptocurrency Portfolios. *Computational Economics*, 1-32.