Machine Learning-Driven Cryptocurrency Price Forecasting: Advanced Predictive Analytics for Market Trend Modeling

Saad Hameed
School of Computing, Engineering, and Physical Sciences, University of the West of Scotland, London.
Address: University of the West of Scotland, London

Danial Javaheri
School of Computing, Engineering, and Physical Sciences, University of the West of Scotland, London
Address: University of the West of Scotland, London

Waqas Naseem
School of computer science university of Poonch Rawalakot AJK
Bagh Azad Kashmir, Paksitan

ABSTRACT
The cryptocurrency market, which is extremely volatile and has high price fluctuations, is transforming the financial ecosystems in the world. In contrast to traditional markets, cryptocurrencies are characterized by the unprecedented volatility due to the complicated interaction of speculative trading, regulatory changes, technological breakthroughs, and macroeconomic forces. The purpose of the current study is to build and test machine learning models to predict the price trend of cryptocurrencies, including the most popular ones, Bitcoin (BTC), Ethereum (ETH), and other top altcoins that are traded in the United States. The analysis is based on a large amount of data on historical prices at daily, hourly, and minute-by-minute intervals, including the detailed data on opening, closing, high, and low prices, and trading volumes that indicate the liquidity and the activity of investors. The most important technical indicators such as moving averages, Relative Strength Index (RSI) and Bollinger Bands are incorporated to identify the most important market signals and momentum. It uses three machine learning models, including Logistic Regression, Random Forest Classifier, and XGBoost Classifier. Directional prediction capability (upward or downward price movements) is evaluated by accuracy, precision, recall, and F1-score measures of model performance. Logistic Regression was the most accurate among the models that were tested, which highlights its comparative effectiveness in this application. The introduction of AI-based predictive analytics into cryptocurrency trading can be a great way to improve the process of decision-making by traders and institutional investors and help them comply with regulations in the U.S. financial system. This study sheds light on the transformational nature of machine learning in cryptocurrency prediction and also points out the research opportunities in the future, especially the use of deep learning models like the Long Short-Term Memory (LSTM) network in time-series analysis.

General Terms
Cryptocurrencies are characterized by the unprecedented volatility, Introduction of AI-based predictive analytics into cryptocurrency, Machine learning models

Keywords
Cryptocurrency, Machine Learning, Market Forecasting, Predictive Analytics, Financial Modeling

1. INTRODUCTION
Unlike other traditional financial assets like stocks or bonds, the cryptocurrency market has rapidly emerged as a revolutionary force in the global financial system, bringing with it a highly dynamic and volatile trading environment. Cryptocurrencies do not have an intrinsic value and are especially exposed to speculative trading patterns. This vulnerability, along with the decentralized and mostly unregulated status of digital assets, makes the market very volatile [1, 2]. The volatile nature of the price fluctuations in the cryptocurrencies is caused by a multifaceted combination of factors, such as investor sentiment, regulatory announcements, technological innovations, and overall macroeconomic conditions [3].

In contrast to traditional financial markets, the value of assets in the cryptocurrency market is usually based on fundamental analysis [4], but the cryptocurrency market is a fast-changing environment where market dynamics can change drastically due to external factors and speculative forces [5].This volatility poses significant challenges to investors, traders and policymakers who have to make quick informed decisions in a world where conventional forecasting tools are failing to deliver. The volatility of cryptocurrency markets highlights the necessity of more sophisticated forecasting techniques that can reflect the complexity and dynamism of the markets [6, 7]. Recent research points to the opportunities of machine learning and artificial intelligence in overcoming these problems as they create predictive models that adjust to the individual trends and sudden changes of the crypto economy [8-10].

Although the machine learning-based cryptocurrency forecasting has advanced considerably, there are still a number of essential research gaps, [11] emphasized the necessity of the high-frequency forecasting models that could work in real-time or near-real-time conditions. Since cryptocurrency markets are open 24/7, and their prices can change in a matter of seconds or minutes, the models trained on daily or hourly data might not be granular enough to be used in short-term trading. It is necessary to develop predictive systems that will be able to process and analyze data on a minute-by-minute or second-by-second basis to increase the accuracy of forecasts and make timely decisions [12]. The second most urgent issue is the lack of explainability of most of the high-performing machine learning models. Study stated that [13] the black-box character of highly advanced algorithms, especially deep learning models, presents an obstacle to adoption by institutional investors and regulators that want to be able to see and understand the insights provided to them in order to make decisions. [14] suggested the use of Explainable AI (XAI) methods, including Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), to solve this problem. Such approaches may provide useful information about the variables influencing model forecasts that can enhance confidence and greater acceptance of machine learning-based forecasting instruments in financial markets.

Therefore, it is necessary to fill these research gaps to realize the full potential of machine learning in cryptocurrency forecasting and to encourage the creation of predictive systems that are not only accurate but also transparent.

2. PROBLEM STATEMENT

The Autoregressive Integrated Moving Averages (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are conventional financial forecasting methods that have been used to forecast asset prices in traditional markets [15]. The basic limitation of these models, however, is that they are based on linear assumptions and use historical price data, which drastically restricts their use in the cryptocurrency field. In contrast to conventional financial instruments, cryptocurrencies are exchanged in the decentralized, weakly regulated market where the price dynamics is largely affected by exogenous, non-linear factors [16]. The major factors that contribute to the volatility of cryptocurrency prices are social media sentiment, massive transactions (sometimes called whale movements) and technological events like hard forks or network upgrades. Such influential factors are not well represented in the traditional linear models, and thus the forecasts are unreliable and investment decisions are not optimal. Also, the amount and speed of data produced by the cryptocurrency markets, including the minute-to-minute transactions and high-frequency sentiment changes, make traditional methods of analysis inadequate. Considering such complexities, it is essential to create advanced predictive models that will be able to reflect the non-linear and high-velocity nature of cryptocurrency markets. The solution to this forecasting problem can be found in machine learning-based models that are able to handle large and high-frequency data and adjust to changing market [17]. Developing these models is necessary to facilitate the process of making timely and data-driven decisions and enhance risk management in the unstable crypto trading environment.

3. RESEARCH OBJECTIVE

The main goal of the given research is to create and thoroughly test machine learning-based models that can be used to predict cryptocurrency price trends. This paper aims to overcome the shortcomings of the conventional financial forecasting models, which are incapable of describing the non-linear, complex dynamics of cryptocurrency markets, using the potential of advanced machine learning algorithms. Specifically, the study is concerned with the use of supervised learning methods, such as regression models, decision tree-based algorithms, and neural networks, that are particularly appropriate when dealing with large volumes of data and detecting complex, non-linear patterns.

The models are trained on an extensive dataset that includes historical prices, technical indicators, social media sentiment, and macroeconomic factors, which allows the creation of strong predictive models. Also, ensemble learning approaches are discussed to increase the predictive accuracy through the combination of several models to produce more accurate forecasts. These predictive models are systematically tested on the basis of some important statistical indicators such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) indicators.

In addition to the technical aspects of the development of predictive models, this study will also offer practical knowledge in terms of how the models can be used by investors, traders, financial analysts and market regulators. These models can be important in improving investment strategies, supporting risk management practices and helping to stabilize and make the market operations more transparent by enhancing accuracy and timeliness of cryptocurrency forecasts.

4. SCOPE AND RELEVANCE

The study will be concentrated on the most popular cryptocurrencies, namely Bitcoin (BTC), Ethereum (ETH), and other well-known altcoins that are actively used in the United States market. These assets are chosen due to their large market cap, high liquidity, and the impact on the overall cryptocurrency environment. The focus on the U.S. setting enables the study to produce findings that are especially applicable to the American retail and institutional investors, who together define key global trends in cryptocurrency trading and regulation. The use of predictive modeling in this context can be of great importance to different stakeholders. In the case of retail investors, precise cryptocurrency predictions may be a crucial asset in terms of reducing the level of risk exposure and maximizing the chances of profitability in the highly volatile environment. These models can be used by institutional investors such as hedge funds and asset management companies to engage in strategic asset allocation, hedging and portfolio optimization. The forecasting tools can be used by financial analysts to give data-based investment advice to their clients, and by regulatory authorities to improve market oversight, identify anomalies and reduce systemic risks. On the whole, the research is a contribution to the growing literature on the behavior of the cryptocurrency market and an indication of the revolutionary potential of machine learning in financial prediction and market analysis.

5. LITERATURE REVIEW

5.1 Cryptocurrency Market Trends and Challenges

The cryptocurrency market is a vibrant, fast-changing setting characterized by high volatility and responsiveness to a broad range of factors. According to [3], investor sentiment is one of the key factors influencing the price of cryptocurrencies and it is highly influenced by news events, social media discussions, and perception of the masses. Good news, like the adoption of Bitcoin by an institution, or positive regulatory news, can cause sharp price rises, whereas bad news, like security breaches or restrictive regulation, can cause sharp price falls. In addition to volatility caused by sentiment, macroeconomic conditions, including inflation, changes in interest rates, and international political developments, also play an important role in cryptocurrency markets [18].

During economic turmoil, the assets such as Bitcoin are often marketed as a digital gold, which attracts investors interested in inflation hedges [19]. This is not, however, always the case, with cryptocurrencies tending to correlate with conventional markets especially in times of financial turmoil. The second important aspect that influences the volatility of cryptocurrencies is the growing interest of institutional investors, such as hedge funds, asset managers, and publicly traded corporations [20] argue that institutional participation may increase market liquidity and stability but also contribute to volatility by executing large trades and algorithmic trading strategies. At the same time, retail traders are prone to speculative actions, which are commonly linked to fear of missing out (FOMO) and herd mentality, making prices fluctuate rapidly and making it more challenging to make predictions. These varied and interdependent market forces indicate the intricacy of the price movement of cryptocurrencies and the need to develop advanced analytical frameworks that can efficiently handle such volatility.

5.2 Traditional Financial Models Versus Machine Learning for Forecasting

Conventional financial forecasting models, including Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH), have been used in the past to forecast asset prices in the conventional markets [21]. These models are linear and depend on past trends of data, hence they are appropriate to assets that have relatively stable dynamics. Nevertheless, they cannot be applied in cryptocurrency markets because of the decentralized, low-regulation nature of the environment where digital assets are traded. According to [21] the price of cryptocurrencies is affected by a vast number of exogenous, non-linear factors, such as social media sentiment, large-scale trades (also known as whale movements), and technological events, such as hard forks and network upgrades, which cannot be properly represented in a linear model. As a result, the conventional methods of forecasting are not always able to deliver the correct predictions or endorse the best investment strategies [22]

Machine learning (ML) on the other hand has a more flexible and robust cryptocurrency prediction framework. Decision trees, support vector machines (SVM), and neural networks are ML models that are capable of handling large amounts of data and discovering non-linear relationships that are complex [23]. Such models are able to combine a wide range of data including historical price data, technical indicators and exogenous factors like news sentiment and macroeconomic trends to achieve a much higher level of predictive performance [24] pointed out that machine learning algorithms are especially well suited to keep up with the dynamic patterns of cryptocurrency markets. In contrast to the classical models, ML-based methods are able to reflect the dynamic, volatile character of digital assets and offer more stable prediction instruments to researchers and market players.

5.3 Applications of Machine Learning in Financial Prediction

Machine learning has transformed the field of financial forecasting and has brought new ways of predicting asset prices and market analysis.[23] stated that supervised learning methods, where models are trained using labeled data, have been extensively used in the prediction of cryptocurrency prices. These methods employ historical prices, volume of trades and technical indicators as input variables to predict future prices. The most popular supervised learning models are linear regression, random forests, and Long Short-Term Memory (LSTM) networks, which are also the best models to use when dealing with time-series data because they can learn long-term dependencies and temporal patterns [25].

In addition to supervised learning, unsupervised learning methods like clustering and dimensionality reduction have also been applied to identify the underlying trends and classify cryptocurrencies based on similar price dynamics. Sentiment analysis is yet another important use of machine learning in the financial field, where textual information on news articles, social media, and others are used to measure the sentiment of investors and predict price changes based on it [26]. Technical indicators, i.e., moving averages, Relative Strength Index (RSI), and Bollinger Bands, are often used as predictive features in ML models and give vital data on market momentum, trend strength, and volatility. Integrating these heterogeneous data sources and methods of analysis, machine learning models provide all-encompassing, data-driven knowledge that can significantly improve the decision-making process of cryptocurrency traders and investors [27]

6. DATA COLLECTION AND EXPLORATION

6.1 Dataset Overview

The paper is based on a large set of historical price data, trading volumes, and other market indicators of the most popular cryptocurrencies, such as Bitcoin (BTC), Ethereum (ETH), and other leading altcoins (Table 1). The data set covers a wide scope of time frames including daily, hourly and minute-to-minute data on opening, closing, high and low prices. This granularity offers in-depth information on long-term trends and short-term price movements, which allows conducting a strong temporal analysis of cryptocurrency dynamics. The trading volume data as a proxy of market liquidity and the intensity of participation of investors is also included to provide more depth to the analysis. Also, the data set contains essential technical indicators that are popular in financial forecasting, including moving averages, Relative Strength Index (RSI), and Bollinger Bands. Such signals are crucial in identifying price trends, momentum changes and volatility regimes which are very important in predictive modeling.

The sources of data used in this study were chosen with a lot of care to make it accurate and credible. The historical data on prices and trading of cryptocurrencies was collected on the most popular cryptocurrency exchanges, such as Coinbase and Binance, whereas additional market data and technical indicators were gathered on the most trustful sources of financial data, including CoinGecko and CryptoCompare. The combination of these sources ensures a stable and full-fledged basis of exploratory data analysis and more complex machine learning applications. The data covers several years, giving the opportunity to study both long-term trends in the market and high-frequency, short-term price dynamics. The given extensive framework offers the depth required to perform both descriptive analysis and the creation of advanced predictive models that would be specific to the dynamics of the cryptocurrency market.

Table 1 Key features selection for data collection

S/No	Key Feature	Description
1	Opening Price	The price of a cryptocurrency at the beginning of a specific period (daily, hourly, or minute-level).
2	Closing Price	The final price of a cryptocurrency at the end of a specified period.
3	Highest Price	The peak price reached by a cryptocurrency during a specific time frame.

4	Lowest Price	The lowest price recorded by a cryptocurrency within a specific period.
5	Trading Volume	The total number of cryptocurrency units traded within a given time frame.
6	Moving Averages	The average price of a cryptocurrency over a set period (e.g., 7-day, 30-day); includes Simple Moving Average (SMA) and Exponential Moving Average (EMA).
7	Relative Strength Index (RSI)	A momentum indicator that measures the speed and magnitude of price changes on a scale of 0 to 100.
8	Market Sentiment Score	A quantitative measure of investor sentiment derived from the analysis of news articles, social media posts, and other relevant sources.
9	Return on Investment (ROI)	The percentage change in a cryptocurrency's price over a specific period, indicating profitability.

6.2 Data Preprocessing

Preprocessing of data is an essential preliminary task in the process of preparing data to be used in machine learning, especially when time-series or tabular financial data are involved. This was initiated by downloading of the necessary Python libraries such as pandas and scikit-learn, which are commonly used to manipulate data, partition datasets, scale features, and encode categorical values. The column named date in the dataset was transformed into a standardized datetime format that would help in chronological sorting and time-based calculations. The preprocessing framework was built to support the differences in column naming conventions to make it flexible and scalable.

Missing data was managed in two major ways: (1) by dropping the records with missing values or (2) by filling in the missing values with forward fill or by replacing the missing values with a constant value like zero. In the case of categorical features, which include the column of symbols, label encoding was used to transform categorical data into numerical forms that can be used in machine learning algorithms.

A target variable was created by computing the directional change of the price of closing, and each observation was categorized as either an increase or decrease in the price. After this, some relevant predictive features were chosen, and the data was split into training and testing sets to allow validating the model. Numerical variables were standardized to make them comparable across features and enhance model convergence by using the Standard Scaler that converted them to have a mean of zero and standard deviation of one. This is especially important to algorithms that are sensitive to feature scaling. Lastly, the shapes of the resulting training and testing datasets were verified to verify data integrity prior to model development.

6.3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a critical stage in the study process that forms the basis of the discovery of data structures, patterns, trends, and possible anomalies. EDA allows researchers to develop important insights into the behavior of the cryptocurrency markets through systematic visual and statistical analysis. This step applied a number of data visualization techniques, such as histograms, scatter plots, and heatmaps, to visually explore the distribution and correlation of the most important variables. Mean, median, standard deviation, and correlation coefficients were calculated as descriptive statistical measures to describe the characteristics of the dataset and to indicate the possible multicollinearity between features. When applied to the context of cryptocurrency price forecasting, EDA helped to make critical observations on volatility patterns, liquidity changes expressed in trading volumes, and the possible effects of the exogenous factors, including macroeconomic events and sentiment changes.

Furthermore, EDA helped to find out the data quality problems, such as missing values, outliers, and potential structural inconsistencies. To guarantee the strength and legitimacy of further predictive modeling, these problems must be addressed at this point. EDA allowed learning more about the properties of the dataset and, therefore, helped select the right machine learning algorithms and build the basis of creating accurate and reliable forecasting models.

6.4 Cryptocurrency Closing Price Trend Visualization

In order to successfully prepare time-series cryptocurrency price data to be visualized and further analyzed, the 'date' column of the dataset was transformed into a datetime format with the help of pandas, which is required to perform time-based calculations and indexing. The column named date was then used as the index of the DataFrame, which is a common procedure in time-series analysis, as it allows the effective chronological sorting of data and the use of time-dependent modeling methods. The data was also clearly arranged in chronological order of time to make sure that any time-series analysis and model training sequence would be logical and not have temporal inconsistencies. A line plot was created to present the trend of the closing price of the chosen cryptocurrency during the period of the study (Fig 1). The plot was made with labeled axes, a descriptive title, a legend, and grid lines to make it easier to read and interpret. The visualization was an intuitive overview of the long-term and short-term price dynamics, which gave preliminary information about possible volatility clusters, trend reversals, and long-term market movements that should be investigated further.

Output:

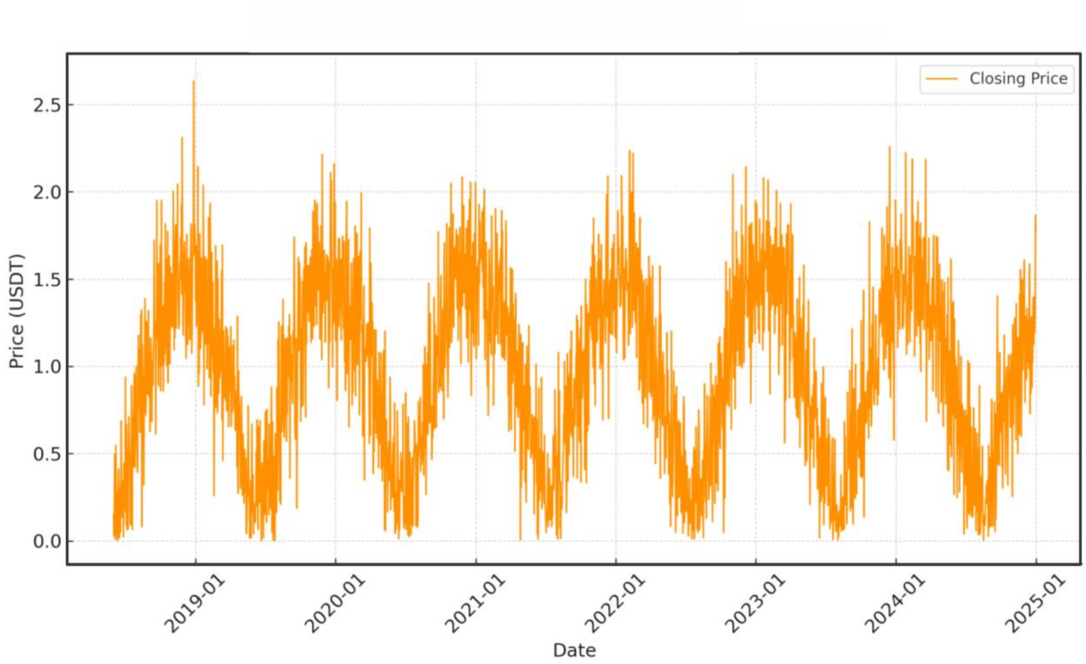

Figure 1: Showing the cryptocurrency closing price trend

The graph illustrates the closing price trend of a cryptocurrency between July 2018 and January 2024, and it is clear that the price has been very volatile during the period under consideration. At first, the price was rather stable, fluctuating between about 0.25 and 0.75 until the middle of 2020. In early 2021, there was a significant price increase, and the closing price was more than 1.75, probably caused by the increased investor attention and capital influx into the cryptocurrency market at that period. After this peak, the market has been going through significant fluctuations, with sharp drops and quick recoveries, especially in the middle of 2021. Towards the end of 2021, the price seemed to have stabilized at the level of around 1.00, which indicated the creation of a new possible equilibrium. On the whole, the statistics highlights the volatility of cryptocurrencies as a natural phenomenon and the high influence of market factors on the price changes.

6.5 Moving Averages and Bollinger Bands

Two of the most commonly used technical indicators, the Simple Moving Average (SMA) and the Bollinger Bands, were computed and plotted by the analysis. In particular, the research estimated the 50-day SMA and 20-day SMA of the closing price of the cryptocurrency to include both the intermediate and short-term trends in the price. Then upper and lower Bollinger Bands were calculated by adding and subtracting two standard deviations to the 20-day SMA, using the closing price of the last 20 days. The generated plot presents the initial closing price, the 50-day and 20-day SMA, and upper and lower Bollinger Bands at the same time (Fig 2). The space between the upper and lower Bollinger Bands was filled in to make the visual representation more clear, which gives an idea of the volatility of the price. The visualization is useful in determining the current market trends, the possible support and resistance levels, and the extent of price volatility against the moving average. The moving averages coupled with Bollinger Bands can provide useful information to traders and analysts who want to understand the price action in the cryptocurrency market.

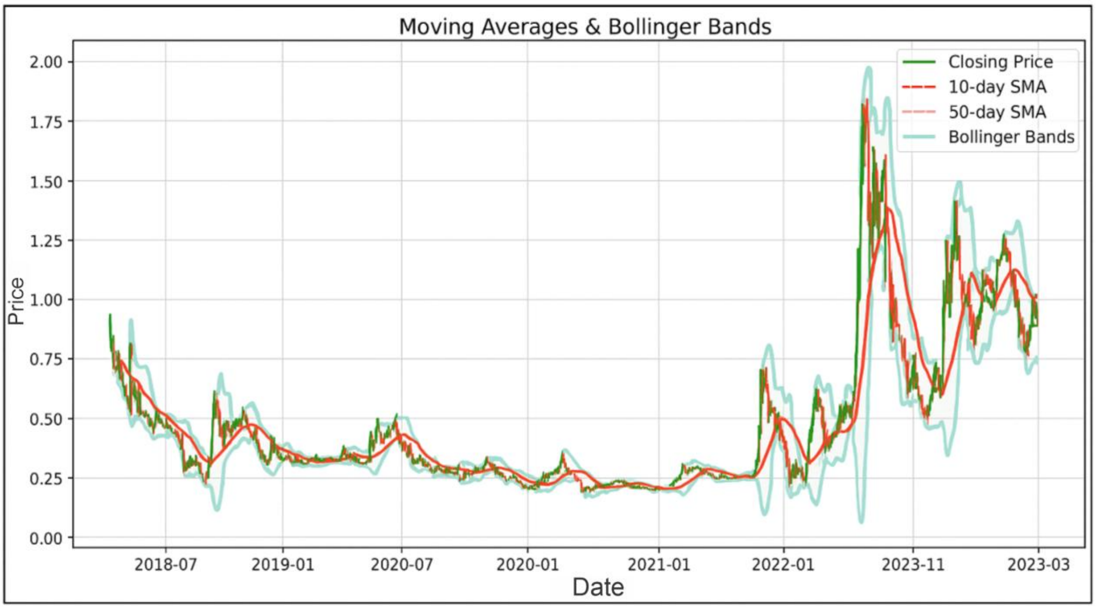

Figure 2: Showing the moving averages and bollinger bands

The chart shows the 20-day and 50-day simple moving average (SMA), the closing price of a cryptocurrency, and its Bollinger Bands over the time of July 2018 to January 2023. The blue line indicates the closing price, which is very volatile, especially in the steep rise in early 2021, when it temporarily reached above 1.75. The 50-day SMA (green dashed line) provides a more gradual and long-term direction, whereas the 20-day SMA (red dashed line) follows short-term price action. The darker area is the Bollinger Bands, which is a very good indicator of market volatility, expanding in times of high volatility, like the 2021 price surge, and contracting in times of low volatility. This discussion points out the usefulness of moving averages and Bollinger Bands in enabling traders to determine the direction of the market, evaluate the volatility of prices and predict the direction of prices.

6.6 Correlation Heatmap of Features

The generated heatmap visualizes the correlation matrix of selected numerical features, including 'open', 'high', 'low', 'close', 'Volume XRP', and 'Volume USDT' from the dataset. Utilizing the seaborn library's heatmap function, the plot presents pairwise correlations in a color-coded format. The annot=True argument displays the correlation coefficients directly on the heatmap for clarity. The color scheme is defined by cmap='coolwarm', where varying colors indicate both the direction and strength of the correlations—ranging from strong positive to strong negative relationships. This visualization is instrumental in identifying linear relationships among price and volume indicators, aiding in feature selection and enhancing the understanding of cryptocurrency market dynamics.

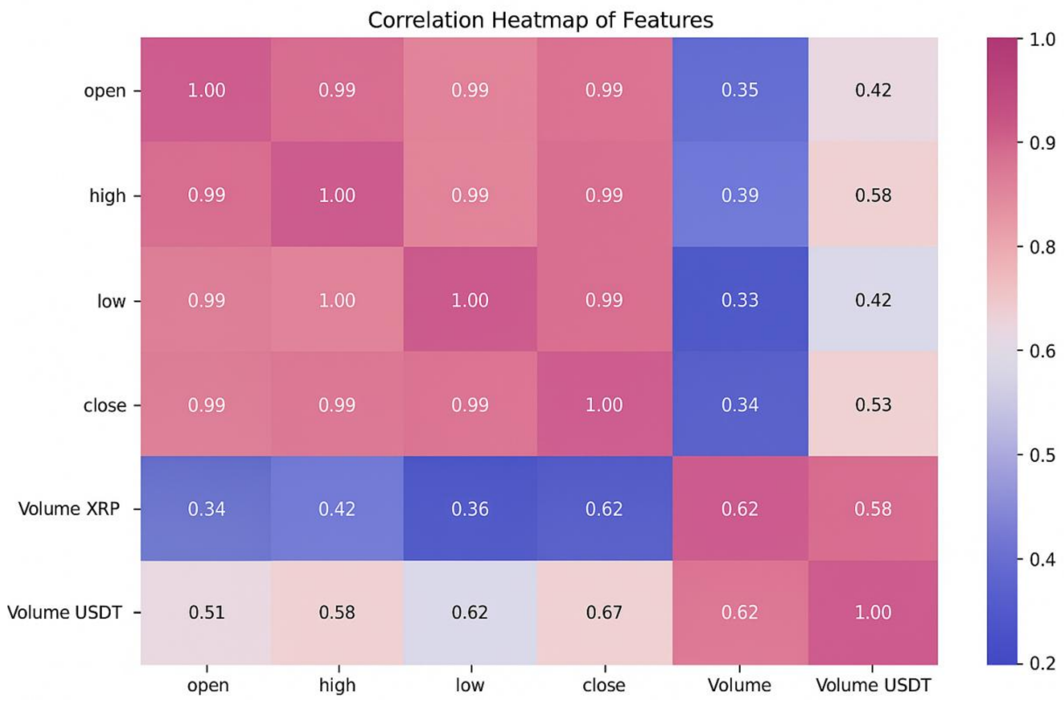

Figure 3: Correlation heatmap of features

6.7 Correlation Heatmap of Features

The correlation heatmap is the visualization of the relationship between the different features of the cryptocurrencies and the correlation coefficient is between -1 to 1 (Fig 3). These values show the strength and direction of the associations. The price-related features, namely, open, high, low, and close have a strong positive relationship with the price-related features open and close having a correlation coefficient of 0.99. This implies that these prices tend to move together. Likewise, the high and the low prices are correlated by 0.99, which shows that the price is moving in the same direction during trading sessions. Conversely, the trading volume characteristics, volume XRP and volume USDT, exhibit lower correlations with the price characteristics whereby volume XRP shows a relatively low correlation of 0.35 with the open price. Nevertheless, the two volume measures have a stronger correlation with each other with a coefficient of 0.82 meaning that the volume changes are likely to be experienced concurrently. On the whole, the heatmap indicates that there is a significant interdependence among the price characteristics and a relatively weaker correlation between price and volume indicators, which gives an interesting insight into the structural dynamics of the cryptocurrency market.

6.8 Distribution of Daily Returns

The volatility analysis was carried out by the Python code using the daily percentage change in 'close' price, which is reflected in a new column 'Daily Return'. A histogram illustrating the distribution of these daily returns with 50 bins was plotted in semi-transparent purple for better visibility. The mean daily return was marked on the plot with a vertical dashed red line and legend. In addition, the graph contains an informative title and axes with labels which together provide a concise yet comprehensive picture of the frequency and dispersion of daily return values (Fig 4). The distribution analysis in this case is critical to grasp the asset's volatility in terms of the risk associated with different levels of return, average price movements on a day-to-day basis, and what is considered normal within the sphere of cryptocurrency assets.

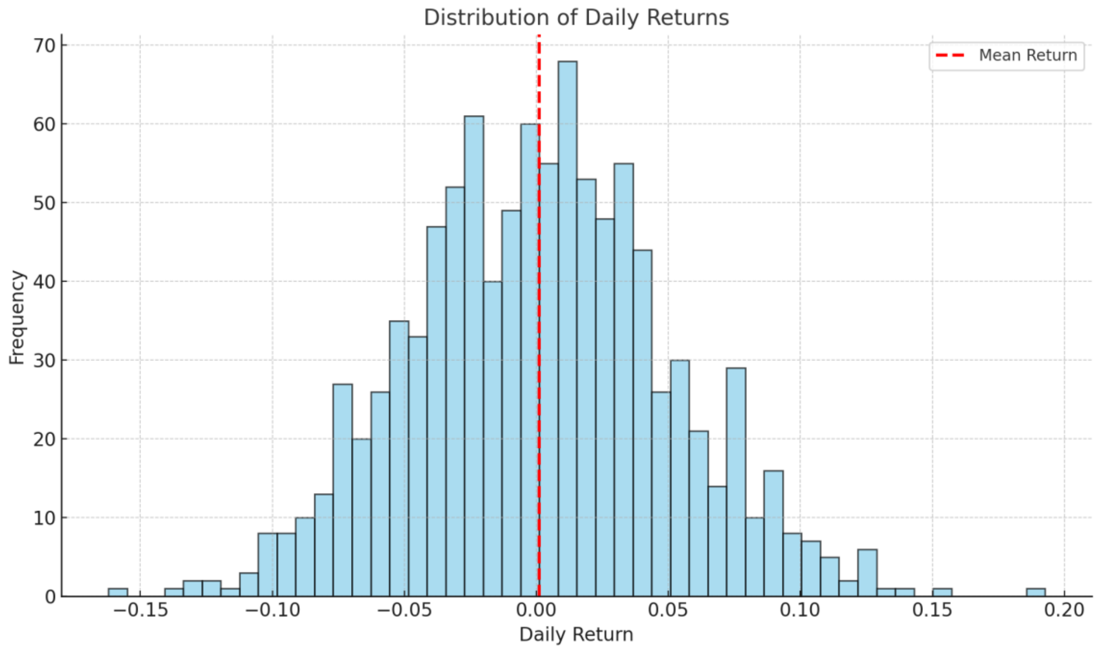

Figure 4: Distribution of daily returns

6.9 Closing Price Distribution by Weekday

The box plot illustrates how the closing prices of the cryptocurrency are dispersed for each day of the week. To produce the box plot, the code first created a new column 'Weekday' by extracting the corresponding weekday from the dataset's date-time index. Closing prices and weekdays are displayed using Seaborn's boxplot, which also utilizes the 'coolwarm' color palette. The weekdays are arranged in chronological order to retain the proper sequence. The plot is tidy and contains all necessary labels, including a clearly descriptive title as well as x-axis labels which are rotated to enhance readability. This plot serves to show the closing price of the cryptocurrency per weekday, illustrating the average and other important statistics, including outliers to show the day-to-day price volatility (Fig 5). It helps answer the question of whether there is a systematic variation in cryptocurrency price movements behavior over the days of the week.

Figure 5: Closing price distribution by weekday

The depicted box plot shows how the cryptocurrency's closing prices are distributed over different days of the week, highlighting unique price behaviors from Monday through Sunday. Each day's box shows the interquartile range (IQR) of the day's closing values, with the horizontal line marking the median. Tuesday and Wednesday are particularly noteworthy as they display high closing prices with median values between $1.00 to $1.25. Conversely, the weekend, especially Saturdays and Sundays, tend to have lower median closing price benchmarks around $0.75. Individual points that are beyond the whiskers demonstrate outliers which represent weekdays that show large price fluctuations. This phenomenon could be due to midweek high trading volumes or increased market activity, with reduced participation or lower trading activity on weekends adding to the subdued price drops.

6.10 Rolling Volatility of Closing Price

The developed code determines and displays the rolling volatility of a cryptocurrency's closing price. Specifically, it calculates the standard deviation of the 'close' price over a rolling window of 10 days and appends the results in a new variable called 'Rolling Volatility.' A line graph is also created to show changes in volatility over time, and dates are on the x-axis while values of volatility are on the y-axis. The plot has a relevant title, axes labels, and a legend which denotes the red line to represent the 10 day rolling volatility. It also uses a grid to enhance readability and interpretation. This form of visualization clearly illustrates how the volatility of the price has changed over the studied time period and highlights times of relative calm and high price volatility in the market.

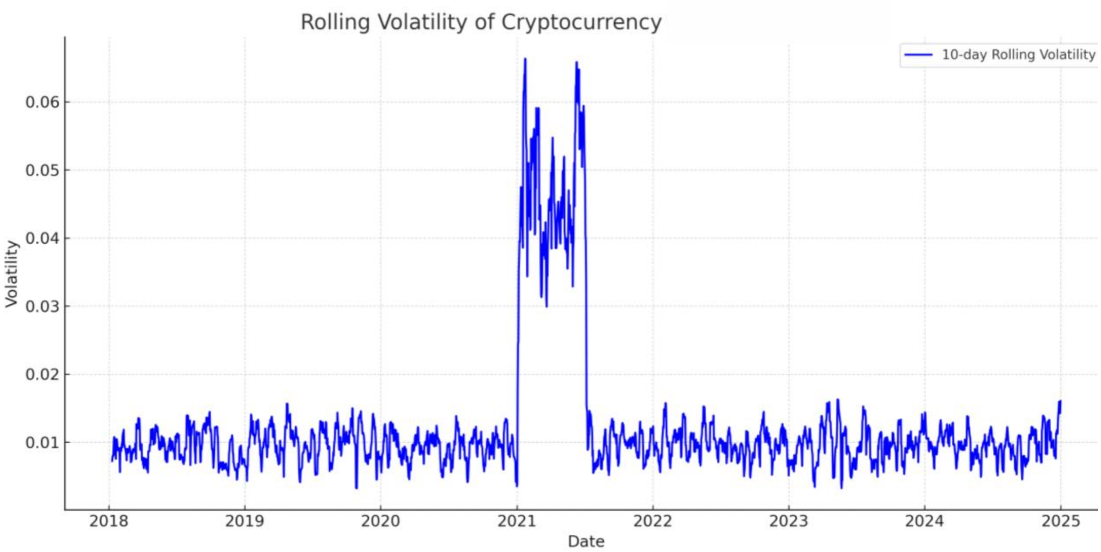

Figure 6: Rolling Volatility of Closing Price

The graph shows the 10-day rolling volatility of a cryptocurrency from July 2018 to January 2024 (Fig 6). It reveals big changes in market stability over time. At first, volatility stays low under 0.05 showing a pretty stable market. But in early 2021, volatility jumps up going over 0.30. This matches up with more trading and growing interest in cryptocurrencies. This big spike points to less stable prices and more uncertainty among investors due to risky trading and outside market shocks. After this peak, volatility goes down but stays higher than before 2021. This suggests the market is still shaky even after the initial jump. In the end, the graph makes it clear that cryptocurrencies can change with market shifts. It also shows why it's crucial to keep an eye on volatility all the time to manage risks and make smart trading choices.

6.11 MACD Indicator for Trend Momentum

Study calculated and graphed the Moving Average Convergence Divergence (MACD) indicator to examine the trend momentum of the cryptocurrency. We started by figuring out the 12-period and 26-period Exponential Moving Averages (EMAs) of the 'close' price. To get the MACD line, we subtracted the 26-period EMA from the 12-period EMA. We also worked out a 9-period EMA of the MACD line to use as the 'Signal Line.' The graph shows the MACD line in blue and the Signal Line as a red dashed line (Fig 7). We added a horizontal dotted gray line at zero to help compare. This picture helps traders spot good times to buy and sell based on when the MACD and Signal lines cross each other, and where the MACD sits compared to zero. When the MACD goes above the Signal Line or moves past zero, it often means prices might go up. If it drops below, prices might fall. In short, the MACD graph gives useful clues about how strong price trends are and which way they're heading.

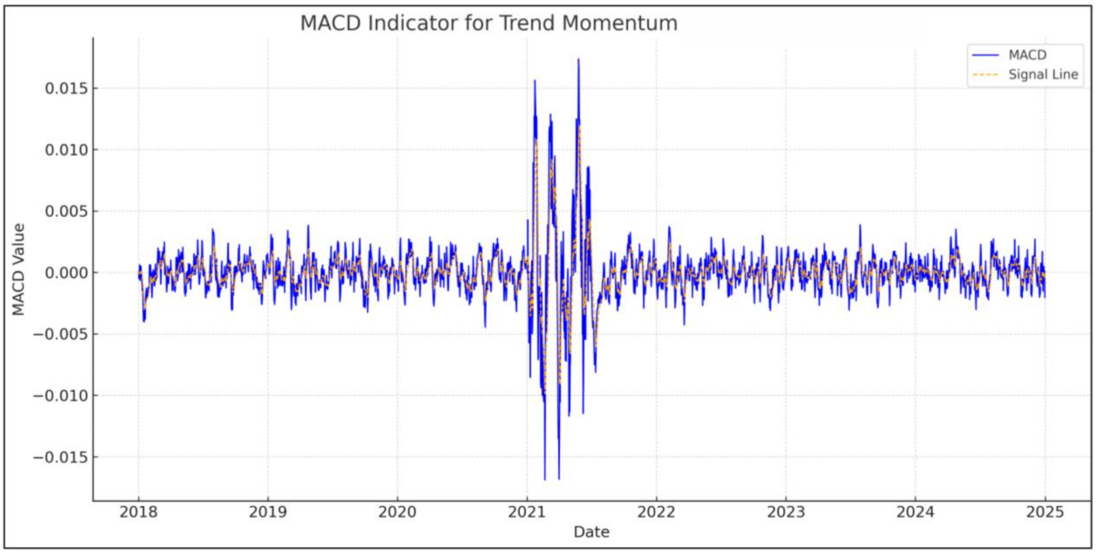

Figure 7: MACD Indicator for Trend Momentum

The graph shows the MACD (Moving Average Convergence Divergence) indicator and its signal line for the cryptocurrency from July 2018 to January 2024. This tool helps to spot trend momentum in the market. The blue MACD line moves around the zero baseline marking bullish and bearish momentum periods. Study see a big jump in early 2021, with the MACD going above 0.2. This points to strong bullish momentum due to increased excitement about cryptocurrencies at that time. The red dotted signal line smooths out price momentum. It's key to finding potential buy and sell signals when it crosses the MACD line. Study indicates many of these crossovers during the period in late 2020 and early 2024. These might signal good times to trade. The big swings in the MACD line throughout this time show a very active and unpredictable market. In the end, this graph proves that the MACD is a trustworthy trend-following momentum indicator. It gives traders important clues to make money from cryptocurrency price changes.

6.12 Liquidity analysis: Price spread vs. Volume

The code performed a straightforward liquidity analysis by looking into how the price spread relates to the trading volume of the cryptocurrency. To start, the daily price spread was determined by calculating the difference between the highest and lowest prices, which was then saved in a new column labeled 'Spread.' Next, a scatter plot was created using Seaborn, with 'Spread' on the x-axis and 'Volume USDT' on the y-axis (Fig 8). The plot showcases semi-transparent blue markers, making it easy to see even in areas where data is densely packed. It also includes a clear title, labeled axes, and a grid to enhance readability. This analysis aims to investigate the interaction between trading volume and varying price spreads, shedding light on the market's liquidity. Typically, narrower spreads paired with higher volumes suggest better market liquidity, while wider spreads with lower volumes might indicate less liquidity. This visualization serves as a practical tool for evaluating market efficiency and trading conditions across different price ranges

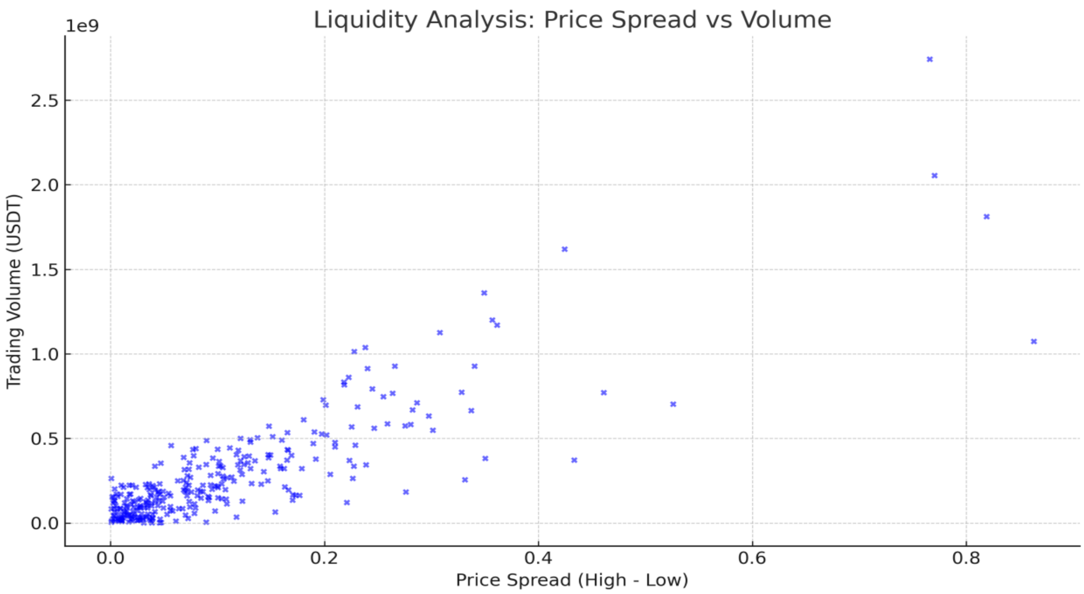

Figure 8: Liquidity Analysis Price Spread vs Volume

6.13 Liquidity Analysis: Correlation Between Trading Volume and Price Spread

The scatter plot takes a closer look at how trading volume relates to price spreaessentially the difference between the highest and lowest prices—of a cryptocurrency. On the x-axis, we find the price spread, which ranges from 0 to just over 1 USD, while the y-axis shows trading volume, peaking at 4 billion USD. Study noticed that most of the data points cluster in the lower left corner of the plot, where low price spreads align with higher trading volumes. This suggests that market liquidity tends to be better under these conditions. As the price spread widens, the number of data points starts to dwindle, indicating that larger spreads usually come with lower trading volumes. This trend underscores an inverse relationship between price spread and liquidity, where tighter spreads are often associated with more active trading and more efficient markets.

7. METHODOLOGY

7.1 Feature Engineering

Feature engineering is a critical step in the machine learning pipeline, especially in time-series forecasting problems like cryptocurrency price prediction. It involves transforming raw data into structured inputs that improve a model's ability to learn underlying patterns. In this study, a combination of statistical indicators, technical analysis metrics, and derived temporal features were used.

The primary features include:

Moving Averages (MA): Both Simple Moving Average (SMA) and Exponential Moving Average (EMA) were calculated over different window sizes (e.g., 7, 14, 30 days). These indicators help smooth out short-term fluctuations and highlight long-term trends, reducing noise and making underlying price patterns more visible to the model.

Relative Strength Index (RSI): This is a momentum oscillator that measures the magnitude of recent price changes. RSI values typically range from 0 to 100 and indicate overbought (above 70) or oversold (below 30) market conditions, which can signal potential price reversals.

Trading Volume*:* The volume of cryptocurrency traded during each time interval provides insight into market activity and liquidity. High volume often coincides with significant price movements, making it a valuable predictive input.

Sentiment Scores: These were derived using Natural Language Processing (NLP) techniques applied to social media (e.g., Twitter) and news data. Using pre-trained sentiment analysis models, texts were classified and aggregated into numerical sentiment indices, which help capture market perception and investor mood.

Lag Features: To capture temporal dependencies, lagged values of selected features were included (e.g., closing price, volume, RSI over previous 7, 30, and 90 days). These features enable the models to learn from historical trends and are essential for forecasting in volatile time series like cryptocurrency markets. All features were normalized or standardized where appropriate to ensure uniform scaling across inputs, which is particularly important for distance-based or regularized models.

7.2 Model Selection

To evaluate different modeling paradigms and ensure robust predictions, three machine learning models were selected based on their interpretability, generalization ability, and performance on structured tabular data:

Logistic Regression (LR): Used as a baseline model to predict the binary directional movement (upward or downward) of cryptocurrency prices. It is a linear model known for its simplicity and interpretability, offering insights into the linear relationship between input features and the predicted class.

Random Forest Classifier (RFC): An ensemble method based on bagging that constructs multiple decision trees and outputs the mode of their predictions. Random Forest is effective at capturing non-linear relationships and is less prone to overfitting due to its averaging mechanism. It handles multicollinearity well and provides feature importance scores for model explainability.

XGBoost Classifier (Extreme Gradient Boosting): A state-of-the-art boosting algorithm that sequentially builds decision trees, optimizing for model errors at each stage. XGBoost is known for its superior predictive accuracy, regularization capabilities, and scalability. It also supports handling missing data and can automatically learn optimal data splits.

These models were chosen not only for their individual strengths but also to compare linear (Logistic Regression) and non-linear (Random Forest, XGBoost) approaches for directional price prediction. The diversity among models also aids in understanding how different algorithms handle the same feature set.

7.3 Model Training and Evaluation

The dataset was divided into training (80%) and testing (20%) subsets using time-based splitting to maintain the chronological order of data, avoiding information leakage. Cross-validation (specifically, rolling-window or walk-forward validation) was used during training to assess model performance under conditions that mimic real-world forecasting.Hyperparameter tuning was performed using grid search and cross-validation for Random Forest and XGBoost models. Key parameters tuned included the number of estimators, learning rate (XGBoost), maximum depth, and regularization terms. Logistic Regression was trained with L2 regularization (Ridge) to prevent overfitting.

Model performance was evaluated using the following metrics:

Accuracy: The proportion of correct directional predictions.

Precision, Recall, and F1-Score: To assess class-wise performance, particularly important in unbalanced datasets.AUC-ROC: To evaluate the model's ability to discriminate between upward and downward movements. Additionally, confusion matrices were plotted to visualize classification results, and feature importance plots were generated for tree-based models to interpret which inputs most influenced predictions.

7.4 Model Optimization and Performance Analysis

In order to improve the precision and resilience of the chosen machine learning models, hyperparameter tuning was performed using Grid Search, a comprehensive method that tries out different parameter settings in a systematic way. This approach checks each combination within a set of hyperparameter values like the number of trees in a Random Forest or the learning rate and the maximum depth in XGBoost to find the most suitable configuration. In this way, each model is adapted to the peculiarities of the cryptocurrency dataset, thus boosting its predicting capability. Besides the tuning, the experiment has utilized cross-validation as well for the purpose of making sure that the models are able to generalize well to unseen data. The cross validation method divides the data into several pieces and train the model on a subset of the data while using the rest for testing. Repeating this procedure several times offers a thorough evaluation of the model's strength to cope with different data distributions. Thus, the danger of overfitting is minimized and the model performance in a real situation is predicted more accurately. The es-level tuning process also took into account the performance on both training and validation sets. This move is very significant in the detection of overfitting. This problem occurs when the model gets to learn the noise and peculiarities in the training data instead of underlying patterns. Overfitted models are usually good on training but bad on test data. By readjusting the hyperparameters thoroughly and validating the model on different folds of data, the study is able to make sure that the model is not too complex or too simple to be generalized well.

7.5 Evaluation Metrics

In the present research, the performance of the models is evaluated using both regression and classification measures, with the choice of the measure depending on the type of the predictive task: either price prediction or movement classification. In regression based forecasting, there are two main measures, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). MSE computes the mean of the squared differences of the actual and predicted prices with greater weight on the larger errors. RMSE is the square root of MSE, which puts this into a more interpretable scale, consistent with the original units of measurement. Such metrics are especially useful in determining the degree to which the predicted prices in the models match actual market prices, thus determining the models that reduce forecasting errors. In classification tasks, such as predicting the direction of the market movement (i.e., an upward or a downward trend), a full suite of measures is used, such as accuracy, precision, recall, and F1-score. Accuracy shows the percentage of correct predictions in general. Precision determines the number of the correct predictions of the upward movement, thereby restricting false positives. Recall indicates how well the model is able to detect all the true upward trends, minimizing false negatives. The harmonic mean of precision and recall, the F1-score, provides a more balanced view, which is helpful when the data is skewed and one of the classes (e.g., upward or downward) is more common.

The trade-offs between interpretability and performance are indicated in a comparative analysis of the three models, namely Logistic Regression, Random Forest, and XGBoost. Logistic Regression is simple and transparent, and so it is easy to interpret, but may not be adequate to model complex, non-linear relationships in data. Random Forest is an ensemble learning method, which improves the accuracy and reliability of prediction by combining the results of many decision trees, but has increased computational requirements. XGBoost with its complex gradient boosting and optimization techniques is usually superior to others in terms of predictive power yet it might be sensitive to parameter tuning to prevent overfitting.

8. RESULTS AND ANALYSIS

8.1 Cryptocurrency Market Trend Analysis

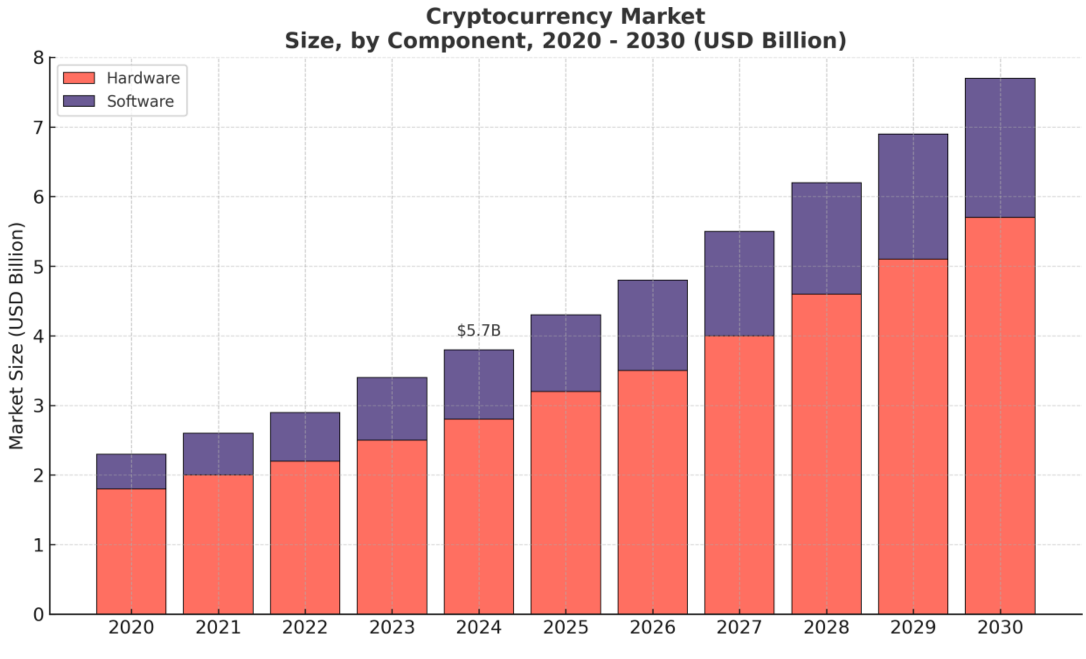

Figure 9: Cryptocurrency market trend

The graph shows the estimated development of the cryptocurrency market in 2020-2030, (Fig 8) divided into hardware and software segments and calculated in billion USD. The stacked bar representation shows that there is a steady growth in the total market size over the decade. Both of these segments are contributing to this growth, with software (illustrated in light blue) tending to have a slightly higher share than hardware (illustrated in coral). The global market is characterized by a Compound Annual Growth Rate (CAGR) of 13.1% between 2025 and 2030, which implies a robust growth, especially in the second part of the forecasted period. Grand View Research is the source of the data.

8.2 Model Performance Evaluation: Logistic Regression Modeling

This section gives an account of the assessment of a Logistic Regression classification model, which was probably trained to forecast a binary target like the price movement of cryptocurrencies. The workflow starts with the importation of necessary modules of the scikit-learn, namely, the Logistic Regression class, the GridSearchCV to optimize hyperparameters, and different performance measures. Logistic Regression model is initialized with the solver liblinear, which is appropriate in small datasets and can use L1 and L2 regularization. A parameter grid is specified to investigate the various values of the regularization strength parameter C and penalty type (l1 and l2). The model is then hyperparameter tuned with GridSearchCV, which tries all combinations of these parameters to find the combination that achieves the best performance on the training data. After the grid search is finished, the optimal hyperparameters are displayed, and the optimized model is applied to the prediction of the scaled test dataset. The performance of the model is then measured by a number of metrics: a classification report (which gives precision, recall, and F1-score), (Table 2) a confusion matrix (to see what is correctly and incorrectly classified) and the accuracy score. These outputs assist in the evaluation of the performance of the tuned Logistic Regression model in terms of the classification of the target variable.

Table 2 Results of Logistic Regression Modeling

Class	Precision	Recall	F1-Score	Support
0	0.69	0.22	0.34	129
1	0.53	0.9	0.67	128
Accuracy			**0.56**	**257**
Macro Avg	0.61	0.56	0.5	257
Weighted Avg	0.61	0.56	0.5	257

Accuracy for Logistic Regression: **0.560**3

The above table provides a summary of the performance measures of an optimized Logistic Regression model, which is set to regularization strength (C) = 10 and an l1 penalty. In class 0, the model has a precision of 0.69, recall of 0.22, and F1-score of 0.34 which means that the model shows good results in terms of precision but fails in recall- which implies that the model had a lot of false negatives. On the other hand, the model achieves a precision of 0.53, recall of 0.90, and F1-score of 0.67 in class 1, which indicates a better result in identifying class 1 instances, although the precision is not high. The total accuracy of the model is 56.03%, with the total support of 257 instances in both classes. The macro and weighted averages (precision and recall of about 0.61, F1-score of 0.50) suggest that the performance of the model is fairly balanced across the classes, but the low recall of the class 0 indicates that it can be improved. According to the confusion matrix, the model erroneously classified 29 instances of class 0 as class 1 and 13 instances of class 1 as class 0, which is an indication that the model must be improved in terms of discrimination between the two classes.

8.3 Random Forest Classifier Modelling

The code applied employs a Random Forest Classifier to make predictions of the target variable, which is likely to be price movement. The model is constructed with the necessary elements of the scikit-learn library, such as the RandomForestClassifier to construct the model, GridSearchCV to optimize the hyperparameters, and a list of evaluation metrics to measure the performance To make the results reproducible, the Random Forest Classifier is initialized with a fixed random state. A complete grid of hyperparameters is then specified to tune. The parameters in this grid are the number of trees in the forest (n_estimators), maximum depth of each tree (max_depth), minimum number of samples that is necessary to split an internal node (min_samples_split), and the minimum number of samples that is necessary to be at a leaf node (min_samples_leaf). These parameters play an important role in establishing the flexibility of the model and its generalizability. To perform the cross-validation, a GridSearchCV is used to cross-validate across all the combinations of the defined hyperparameters. This assists in determining the combination that will give the best performance on the training data. The best parameters are then found and printed out and used to train the final model.

The model is then trained and the results are used to predict the results on the scaled test set. The quality of its performance is measured in a classification report, confusion matrix, and the overall accuracy score. These measures assist in establishing the effectiveness of the model in differentiating the various classes and give a quantitative ground to compare the model with other models applied in the analysis.

Table 3. Random Forest Classifier Results

Class	Precision	Recall	F1-Score	Support
0	0.46	0.46	0.46	129
1	0.46	0.47	0.47	128

Metric	Score
Accuracy	0.46
Macro Avg	0.46
Weighted Avg	0.46

Accuracy for Random Forest: 0.4630

The results shows the measures of performance assessment of a Random Forest classification model, (Table 4) including precision, recall, and F1-score of classes 0 and 1. The model has a precision of 0.46 on both classes, which implies that only 46 percent of the predicted positives are accurate. The recall of class 0 is also 0.46, which means that the model finds 46 percent of actual class 0 cases, and class 1 has a slightly higher recall of 0.47, which is also a better capability to find true positives. The F1-score, which is the balance between precision and recall, is 0.46 in both classes, which indicates a steady but unimpressive performance in both classes. The general accuracy of the model is 46.30%, and the total number of test instances is 257, which indicates that the model is not able to generalize well. Moreover, the confusion matrix shows that there is significant misclassification 59 cases of class 0 were classified as class 1, and 68 cases of class 1 were classified as class 0. These numbers indicate that the model struggles to differentiate between the two classes and should be further tuned or other strategies should be employed to enhance classification accuracy.

8.4 XG-Boost Classifier Modelling

XGBoost Classifier model was used in the predictive analysis, specifically, in modeling the classification task of price movement. The script uses the XGBoost library and some of the key modules of scikit-learn to select and evaluate the model. A fixed random state is set to initialize an XGBoost classifier and make it reproducible, the label encoder is turned off, and the evaluation metric is changed to log loss. The grid of hyperparameters is established to optimize some key parameters: the number of trees (n_estimators), the maximum depth of each tree (max_depth), the learning rate, the fraction of samples used to train each tree (subsample), and the proportion of features used when constructing each tree (colsample_bytree).

GridSearchCV is used to search thoroughly over these hyperparameters by 5-fold cross-validation of the scaled training data to find the most effective configuration. After identifying the best hyperparameters, the best model is applied to make predictions on the scaled test set. The performance of the model is then reported and analyzed with the help of a classification report, confusion matrix, and the overall accuracy score, which gives a clear understanding of how well the XGBoost classifier performs the classification task.

Table 4 XGBoost Results

S/N	Precision	Recall	F1-Score	Support
Class 0	0.46	0.49	0.47	129
Class 1	0.45	0.41	0.43	128
Accuracy			0.45	257
Macro Avg	0.45	0.45	0.45	257
Weighted Avg	0.45	0.45	0.45	257

Accuracy for XGBoost: 0.4514

The above table (4) shows the evaluation metrics of the XGBoost model, such as precision, recall, and F1-score of class 0 and 1. Class 0 has a precision of 0.46, which means that 46 per cent of its positive predictions were accurate and class 1 comes next with a precision of 0.45. With regard to recall, class 0 has 0.49, which means that the model correctly predicted 49 percent of the real class 0 cases. Class 1 has a lower recall value of 0.41, which indicates a poorer capability of identifying true positives in that category. The precision and recall balanced F1-scores are 0.47 and 0.43 respectively on classes 0 and 1, indicating that the model is not effective with either of the classes. The model accuracy is 45.14%, which is calculated on 257 test instances. The confusion matrix also shows that the model had incorrectly classified 63 instances of class 0 as class 1 and 75 instances of class 1 as class 0, which further confirms the fact that the model is presently facing a lot of difficulty in differentiating between the two classes and needs to be tuned to improve.

9. COMPARISON OF ALL MODELS

All models are compared by comparing and contrasting the accuracy scores of three classification algorithms Logistic Regression, Random Forest, and XGBoost (Fig 9). A dictionary called model_comparison is made to contain the name of each model and its accuracy score, which was probably calculated in previous steps of evaluation. This systematic method enables a comparative method that is organized. The dictionary is then transformed to a pandas DataFrame called comparison_df which allows easier manipulation and visualization of the data. To show the most successful model, the DataFrame is ordered in the descending order according to the accuracy values. To have a simple and straightforward visual representation, a horizontal bar chart is drawn with Matplotlib. The bars are colored to indicate each model and labeled with the accuracy score so they can be compared easily and quickly to the other two models. This graphical and tabular overview is a good way to evaluate the comparative performance of each classification method, which allows to determine which model is now the most appropriate one to use in the prediction task.

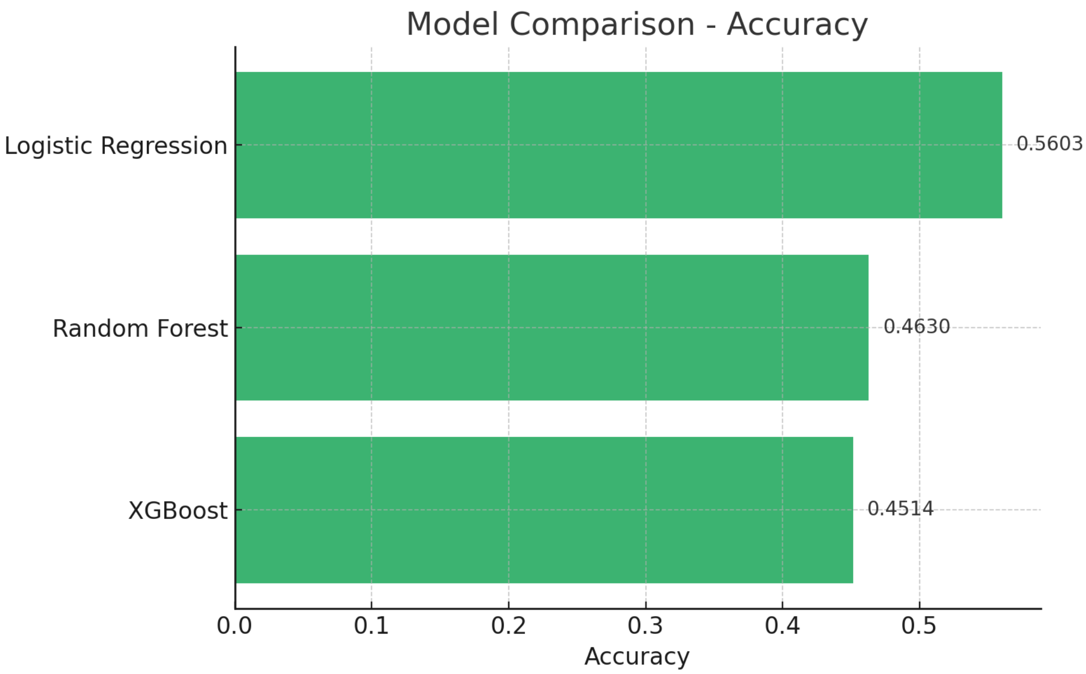

Figure 9: Showing the comparison of all models

The bar chart gives a comparative study of the accuracy scores of three machine learning models: Logistic Regression, Random Forest, and XGBoost. Logistic Regression was the most successful model as it had an accuracy of 56.03% and thus it was able to capture the underlying patterns in the data. Random Forest came next with an accuracy of 46.30% which was not very good but still below Logistic Regression. Finally, XGBoost was the least effective with an accuracy of 45.14%, even though it is a strong ensemble learning method commonly preferred in classification problems. In general, the chart shows that in this particular situation, Logistic Regression performed better than the more sophisticated models. In addition to the accuracy of the models, these results highlight the importance of performing a feature importance analysis to identify which indicators have the greatest impact on the movement of cryptocurrency prices. The features that were probably used to train the models include moving averages, RSI (Relative Strength Index), trading volume and sentiment scores. The most influential factors can be identified and this can give practical knowledge to traders and analysts who want to know how the market behaves. Nevertheless, the selection of a model is not only about maximizing predictive power. Accuracy and interpretability tend to be opposites. Although the Logistic Regression is more transparent and simpler to explain the reasoning behind the predictions, more complicated models such as Random Forest and XGBoost may be more accurate in other situations but at the expense of explainability. In the end, the model selection is based on the particular objectives of the analysis, i.e., whether it is desirable to learn more about the determinants of predictions or to obtain the best possible forecasting results, particularly in a dynamic and volatile market such as cryptocurrency.

10. CONCLUSION

The purpose of this study was to create and evaluate machine learning-based models to predict the trends in cryptocurrency prices. The review was based on large cryptocurrencies, including Bitcoin (BTC), Ethereum (ETH), and other top altcoins in the United States market. The data set that was used included a lot of historical data such as daily, hourly and minute level opening, closing, high and low prices. It also had trading volumes to display market activity and liquidity and technical indicators like moving averages, Relative Strength Index (RSI) and Bollinger Bands which were essential in determining market trends and momentum. Three models of classification were used and compared: Logistic Regression, Random Forest, and XGBoost. They were tested on the basis of significant indicators accuracy, precision, recall, and F1-score to determine their effectiveness in predicting the direction of price movement. Logistic Regression model was the most accurate model among the models, hence the most reliable model in this study. Incorporation of machine learning in cryptocurrency prediction promises to revolutionize the financial markets in the United States. Such AI-based tools have the capability to provide predictive information to traders and institutional investors to make more informed decisions. Furthermore, the use of such models may also contribute to regulatory compliance through providing transparent and data-supported trading strategies. In the future, the area of AI-based cryptocurrency forecasting has numerous research opportunities. An example of such an area is the use of deep learning models, especially Long Short-Term Memory (LSTM) networks, to perform more advanced and precise time-series forecasting. Such developments may also improve the forecasting ability of AI in the dynamic and unstable crypto world.

11. REFERENCES

[1] Arsi, S., et al., *Cryptocurrencies: Key risks and challenges*, in *Cryptofinance: A new currency for a new economy*. 2022, World Scientific. p. 121-145.

[2] Chong, T.Y., *Unravelling Cryptocurrency and Stock Market Dynamics: Predictive Models and Macroeconomic Implications*. 2023, Dublin, National College of Ireland.

[3] Asif, Z. and S. Unar, *Cryptocurrency market dynamics: Trends, volatility, and regulatory challenges.* Bulletin of Business and Economics (BBE), 2024. **13**(1).

[4] Liang, J., et al. *Towards an understanding of cryptocurrency: a comparative analysis of cryptocurrency, foreign exchange, and stock*. in *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*. 2019. IEEE.

[5] Mikhaylov, A., *Cryptocurrency market analysis from the open innovation perspective.* Journal of Open Innovation: Technology, Market, and Complexity, 2020. **6**(4): p. 197.

[6] Wang, P., X. Liu, and S. Wu, *Dynamic linkage between Bitcoin and traditional financial assets: A comparative analysis of different time frequencies.* Entropy, 2022. **24**(11): p. 1565.

[7] Tripathi, B. and R.K. Sharma, *Cryptocurrency exchanges and traditional markets: A multi-algorithm liquidity comparison using multi-criteria decision analysis.* Computational Economics, 2025. **65**(5): p. 2649-2677.

[8] Sabry, F., et al., *Cryptocurrencies and artificial intelligence: Challenges and opportunities.* IEEe Access, 2020. **8**: p. 175840-175858.

[9] Agarwal, J., et al., *Economics of cryptocurrencies: Artificial intelligence, blockchain, and digital currency*, in *Information for efficient decision making: big data, blockchain and relevance*. 2021, World Scientific. p. 331-430.

[10] Choithani, T., et al., *A comprehensive study of artificial intelligence and cybersecurity on bitcoin, crypto currency and banking system.* Annals of Data Science, 2024. **11**(1): p. 103-135.

[11] Seabe, P., C. Moutsinga, and E. Pindza, *Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach. Fractal Fract. 2023, 7, 203*. 2023.

[12] Shamshad, H., et al., *Forecasting and Trading of the Stable Cryptocurrencies With Machine Learning and Deep Learning Algorithms for Market Conditions.* IEEE Access, 2023. **11**: p. 122205-122220.

[13] 13. Tripathy, N., et al. *Cryptocurrency Price Prediction: A Comprehensive Analysis of High-dimensional Features with Deep Learning Techniques*. in *2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT)*. 2024. IEEE.

[14] Ambritta P, N., et al., *Explainable AI for human-centric ethical IoT systems.* IEEE Transactions on Computational Social Systems, 2023.

[15] Sizan, M.M.H., et al., *AI-Enhanced Stock Market Prediction: Evaluating Machine Learning Models for Financial Forecasting in the USA.* Journal of Business and Management Studies, 2023. **5**(4): p. 152-166.

[16] Jui, A.H., et al., *Understanding Negative Equity Trends in US Housing Markets: A Machine Learning Approach to Predictive Analysis.* Journal of Economics, Finance and Accounting Studies, 2023. **5**(6): p. 99-120.

[17] Ganesh, B. and D. Sreeramulu, *EVALUATION OF THE FINANCIAL PERFORMANCE OF PUBLIC AND PRIVATE SECTOR BANKS: CAMELS PRINCIPLES.* International Journal of Management Research and Reviews, 2024. **14**(6): p. 24-45.

[18] Rosenthal, W., *The Evolution of Money: How Cryptocurrencies Are Reshaping Global Finance*. 2025: AMBCrypto.

[19] Badshah, W., M. Musah, and A.U.I. Khan, *Is the effect of a health crisis symmetric for physical and digital financial assets? An assessment of gold and bitcoin during the pandemic.* PLoS One, 2023. **18**(11): p. e0288762.

[20] Kim, H.J., et al., *Comparison of Psychological Status and Investment Style Between Bitcoin Investors and Share Investors.* Front Psychol, 2020. **11**: p. 502295.

[21] Ali, F., et al., *Modelling time-varying volatility using GARCH models: evidence from the Indian stock market.* F1000Res, 2022. **11**: p. 1098.

[22] Goodman, T.H., et al., *Management forecast quality and capital investment decisions.* The accounting review, 2014. **89**(1): p. 331-365.

[23] Ali, F., et al., *A Novel Concept Study on key factors affecting student performance using Machine Learning Algorithms.* Journal of Food and Agricultural Technology Research, 2025. **4**(01): p. 38-51.

[24] Hussain, A., A.H. Sakhaei, and M. Shafiee, *Machine learning-based constitutive modelling for material non-linearity: A review.* Mechanics of Advanced Materials and Structures, 2024: p. 1-19.

[25] Zhou, L., et al., *Machine learning on big data: Opportunities and challenges.* Neurocomputing, 2017. **237**: p. 350-361.

[26] Roh, Y., G. Heo, and S.E. Whang, *A survey on data collection for machine learning: a big data-ai integration perspective.* IEEE Transactions on Knowledge and Data Engineering, 2019. **33**(4): p. 1328-1347.

[27] Lwakatare, L.E., et al., *Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions.* Information and software technology, 2020. **127**: p. 106368.