Bitcoin Price Prediction: A Machine Learning Approach Using Opening and Closing Data

MSc Research Project

MSc Data Analytics

Sahithi Dornala

Student ID: X23241551

School of Computing

National College of Ireland

Supervisor: Furqan Rustam

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……. ……………… Sahithi Dornala ……………………………………………

Student ID: ………………… x23241551……………………………………………………….……

Programme: …………… Data Analytics ………………………… **Year:** …2024-2025..

Module: …………… Research Project ……………………………………………….………

Supervisor: ……………… Furqan Rustam ……………………………………………….………

Submission Due Date: …………………12/12/2024……………………………………………………….………

Project Title:

Bitcoin Price Prediction: A Machine Learning Approach Using Opening and Closing Data

Word Count: …………8793……………………… **Page Count**………………19……….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: …………………Sahithi Dornala……………………………………………………

Date: ……………………………12/12/2024………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□NA
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□NA
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□NA

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Bitcoin Price Prediction: A Machine Learning Approach Using Opening and Closing Data

Sahithi Dornala

x23241551

Abstract

This study concentrates on predicting and forecasting the opening and closing prices of Bitcoin utilizing enhanced machine learning models. The dataset used is the "Bitcoin Historical Data," which contains Bitcoin price data recorded every minute with related Open, High, Low, and Close (OHLC) prices and Bitcoin and USD trading volumes beginning from 2015. The first goal was to examine the patterns of Bitcoin's market movement and create a reliable predictive model for its prices. The models used in this study include GRU (Gated Recurrent Units), LSTM (Long Short-Term Memory), RNN (Recurrent Neural Networks), Prophet, and Prophet-GRU. Based on the results, while Prophet-GRU was hypothesized to leverage the strengths of both models, it did not outperform the standalone GRU model. Specifically, the GRU model demonstrated superior performance, with an R^2 Score of 0.9959, an RMSE of 0.0041, and an MAE of 0.0027 for the Opening Price, and an R^2 Score of 0.9959, an RMSE of 0.0026, and an MAE of 0.0042 for the Closing Price. In contrast, Prophet-GRU exhibited lower accuracy with higher error rates (e.g., RMSE of 18.661 and 40.561 for Open and Close Prices, respectively). These findings highlight GRU's strength in high-frequency financial data forecasting, whereas Prophet struggled to add value due to the lack of significant seasonality in the dataset. This study provides insights into the integration of machine learning and statistical models for Bitcoin price forecasting and serves as a useful resource for traders and analysts within the cryptocurrency market.

.

Keywords: Bitcoin Forecasting, Cryptocurrency Price Prediction, GRU (Gated Recurrent Units), Time Series Analysis, Opening Price Prediction, Closing Price Prediction

1 Introduction

1.1 Background

Predicting the opening and closing prices have increasingly attracted the attention of academicians and investors due to the unpredictable and rapidly raising adoption of Bitcoins Bara and Oprea (2024) and other cryptocurrencies as an additional investment vehicle. Bitcoin is the oldest and perhaps the most popular digital currency, which works on a distributed system supported by the blockchain. Organized structured financial instruments do not align to the shape of Bitcoin and trends that impact its price but may include aspects like; the miners' reaction, changes in laws and regulations, new technologies, and shifts within the overall market. Predicting Bitcoin prices is not an easy task because the currency experiences wild price fluctuations and external shocks Kyriazis et al. (2020). However, predictions can give useful information for involving investors, traders and financial institutions about risks in the markets, effective diversification of stocks and shares and proper decision making. Analyzing data, and based on data, market indicators and sentiment analysis, the historians and analysts apply machine learning and time-series analysis to develop models on the future of the price of Bitcoins Parekh et al. (2022). Traditional methods include autoregressive models, recurrent neural networks (RNN), long short-term memory (LSTM) models, and finally, the hybrid models which include statistical models integrated with deep learning strategies. They try to reflect Bitcoin's behavior in terms of fluctuation patterns, trends and seasonality to enhance the forecast Wahid (2024). However, even in the field of predictive methodologies, it is not easy to predict the Bitcoin price in the future based on the nature of Bitcoin; decentralized, unregulated, and speculative. This research will help to evolve this field by introducing a sound forecasting model that examines the daily Bitcoin price fluctuations based on the opening and closing prices as they are more informative for understanding the daily market attitude and tendency. Whereas the main approach is based on high-frequency trading frequency data, technical analysis indicators and sentiment analysis, the aim of the study is to enhance the precision of the prediction and provide useful information to key participants in the cryptocurrency market. In this way, the research meets the stakeholders' demand for better forecasting models capable of dealing with the specific features of cryptocurrency markets.

1.2 Aim of the study

This study aims to develop a GRU model for predicting stock prices, with a primary focus on the "Open" price. Stock price prediction is challenging due to high fluctuation, nonlinearity, and the influence of various internal and external factors. Traditional time series models often fail to capture both short-term dependencies and long-term trends effectively. To address this, the research employs GRU – a recurrent neural network that is well suited for time series data and Prophet which has been well-developed for trend and seasonality data. The GRU model models temporal dependencies in the data, being trained on sequences within a sliding 60-day window, on the other hand Prophet also applies logarithmic adjustments for seasonality and trends. Nevertheless, the studies point out that the performance of the solely implemented GRU model is higher than the one of the Prophet-GRU combined model, especially in predicting the "Open" price. R-squared, RMSE, and MAE are used to measure performance of the built model; results show that GRU has better accuracy and reliability to predict the stock price movement. This work posits that there exists an opportunity in refining existing models such as GRU in order to enhance the accuracy of forecasts on the financial market and be of immense benefit to several analysts and investors.

1.3 Research Objectives

There are several research objectives as follows:

1. To develop a hybrid predictive model that integrates GRU and Prophet to improve accuracy in forecasting stock "Open" prices.
2. To evaluate the performance of the combined GRU-Prophet model in capturing short-term volatility and long-term trends within stock market data.
3. To assess the effectiveness of the GRU model in identifying temporal dependencies in stock prices and the Prophet model in capturing seasonality and external trends.

1.4 Research Questions

There are several research questions as follows:

1. How effective is the GRU model in forecasting stock "Open" prices compared to traditional time series models?
2. Does the integration of GRU and Prophet provide a more accurate forecast by capturing both short-term and long-term trends in stock prices?

2 Related Work

2.1 Cryptocurrency Market Analysis

Automated with the use of technology and its decentralized system, has risen as one of the fluid and unstable market ventures within the financial context. Bitcoin, as the first cryptocurrency, has been the primary subject of empirical analyses seeking to identify patterns of its market prices, risks, and drivers Giudici et al. (2020). Scholars have postulated that mistakes, specifically factors at the macroeconomic level such as inflation rate, geopolitics, and regulatory environment influence Bitcoins' prices. Furthermore, investor sentiment that analyzes text in social media and news is one of the primary factors for the short-term evolution of prices. Bitcoin has high volatility which is different from typical financial assets Doumani et al. (2021). Several works have examined the relation between volume and price change by showing how high volumes are associated with large price changes. However, most of the previous studies employ daily or hourly frequency data that ignores the rich structures presented in intra-daily trading data. Nicely chopped datasets offer a more detailed analysis of the spike and fall of prices and are more useful when trying to understand fluctuations and peculiarities of the short-term trends in the cryptocurrency market. However, as for other digital assets, mainstream theories fail to give adequate explanation of Bitcoin's behavior Elsayed et al. (2022), though new theories that incorporated both economic, behavioral and the technology aspects provided a direction. However, little has been done to fill that niche by providing more detailed data that could be used to construct models that can account for the dynamic nature of Bitcoin's market. This study aims to fill that gap by using minute level Bitcoin data to enable a more detailed examination of price patterns. Because of utilizing the real-time series models, including LSTM, GRU, and combining these models with elements of Prophet-GRU, this study will also extend the literature review for forecasting crypto-currency prices, with the focus on extending the applicability of the models and practical relevancy of the forecasting results.

2.2. Time Series Forecasting Models for Bitcoin

There are some time series forecasting models which have been used in this section like ARIMA, LSTM, SATRIMA and all which will describe how they employed for forecasting Bitcoin.

In their work, Poongodi et al. (2020) discuss how Bitcoin as an electronic money and as the technological application of the blockchain. The work is to forecast the Bitcoin prices based on machine learning model, namely ARIMA, on an open data starting from 28 th April 2013 to 31 st July 2017, collected from CoinMarketCap. The authors also recommend using ARIMA for carrying out time series analysis to get historical price movement of Bitcoin. The first problem met was trying to predict a Bitcoin price, because it is extremely volatile. However, it is also evident that many patterns of movements of different prices during the analyzed period were reported correctly by the ARIMA model, thus suggesting reasonable predictive performance. However, some weaknesses were observed with the model's ability to address volatile changes in the market price and macroeconomic factors that can cause huge fluctuations in Bitcoin prices. The work highlights how machine learning models can be used in crypt currency study and reveals the necessity of the use of more advanced approaches, which are deep learning or combined with more advanced models to have better accuracy in volatile markets like Bitcoins.

Duy et al. (2024) seeks to enhance the understanding of critical aspects of price fluctuations in the Bitcoin market as well as future market volatility during the period 2021 to 2023 using quantitative modeling techniques notably the ARIMA-GARCH approach. We analyze data on the closing prices obtained from January 17, 2021, to December 17, 2023, amounting to 1065 observations and identify that the ARIMA model has parameters (12,1,12) based on the Accurate Model Selection Test Based on Minimum AICc and Maximum Log-likelihood. To explain the high variability of Bitcoin, a GARCH (1,1) model was used to improve the credibility of the proposed framework for financial time series. Choosing suitable lags within the AR and the MA components was made easier by the Box-Jenkins method of analysis, while confirming model adequacy used residual analysis and the Ljung-Box statistic. However, it was observed that despite catching the price dynamics and volatility of Bitcoin in the aftershocks of external market forces and rapid fluctuations in the form of Price at 99%, the model was not comprehensive enough to expound on the complications of price occurrences in Bitcoin.

Wahid (2024) examines the effectiveness of ARIMA and LSTM models for forecasting Bitcoin price that may be useful for trading, risk management, and investment decisions in the unpredictable Bitcoin market. The paper utilizes Bitcoin prices ranging from January 1, 2011 to December 31, 2023 sourced from CoinMarketCap; several steps include cleaning the data by handling for missing values, eliminating redundancy, testing and making the data stationary and normalizing the price data. ARIMA parameters were identified based on ACF and PACF plots and LSTM utilized neural network approach where temporal windows were used for training. Both models were evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), with ARIMA demonstrating near-perfect fit (MAE: ~2.31e-215, RMSE: 0.0) and outstanding computational performance (training time ~2.55 sec). In contrast, LSTM effectively captured nonlinear patterns with MAE and RMSE of ~0.00022 but required significantly higher computational resources (training time: ~379 seconds). Limitations were generalization: ARIMA overfit the model and Long Short-Term Memory model computationally expensive. Based on figure 4, the study finds out that ARIMA is suited for short horizons since it’s more efficient as compared to LSTM that is more suitable for volatile training since it’s capable of modeling complicated long-term patterns. However, shortcomings in managing external market factors and computational cost imply that future innovations must consider the development of better reconciliation models and broader data sets to improve prediction and flexibility.

The Bitcoin and other cryptocurrencies’ market has gained the attention of investors, companies, and businesses, so it is crucial to develop the right price prediction models. This challenge was investigated in studies Aanandhi et al. (2021), Cheng et al. (2024) and Panigrahi (2023), with the help of different Machine Learning techniques including ARIMA, LSTM, Facebook Prophet (Fb Prophet), and SARIMA models on Bitcoin and other cryptocurrency such as Dogecoin. The study in Aanandhi et al. (2021) aimed at creating a profitable prediction model using ARIMA model and the authors gathered data of the cryptocurrency from Yahoo Finance. To forecast the prices of Bitcoin and represent outcomes in response-oriented graphs, feature engineering on the lagged values of one day, seven days, and 30 days ago were analyzed in this study. However, we saw that even though the model gives almost perfect results on the training set and looks very promising, there are certain constraints when using it which include the fact that it cannot consider non- linear co-relations and external factors such as geopolitical factors or rumors in the market etc. However, the study showed that ARIMA can be applied when there is a short-term periodic pattern that is relevant when a trader is interested in the short-term movement of price of an asset such as cryptocurrency. However, study Cheng et al. (2024) was for the price and Garman-Klass volatility of Bitcoin using LSTM, SARIMA, and Fb Prophet from the year 2017 up to 2023. As highlighted in the results section LSTM presented promising results in estimating the non-linear and volatile price movements of Bitcoin outcompeting SARIMA and Fb Prophet. Fb Prophet had good results in the short-term forecast, mainly used in forecasting the seasonality of Bitcoin and prices, however, it was seriously influenced during the fluctuations like COVID-19 pandemic and the Russian-Ukrainian war.

2.2 Deep Learning-Based Models

There are some deep learning models which have been used in this section like LSTM, GRU, SAM-LSTM which will describe how they employed for forecasting Bitcoin.

To align with this, Tanwar et al. (2021) employed a deep learning model of GRU and LSTM to forecast the values of Litecoin and Zcash with the influence of Bitcoin that acted as the parent cryptocurrency. This research seeks to solve the problem of predicting the future price of the cryptocurrencies, these cryptocurrencies have been known to be extremely volatile and are subject to many technical and sentimental as well as legal changes. This model is deemed promising to address the downside of existing methods, which is unsuitable for real-time process control, by building on the advantage of GRU and LSTM for analyzing time series data. The model was trained and tested against typical data sets as the following figures show that the proposed model has a higher accuracy in prediction compared with the current ones. By the same token, the model might prove inefficient in the case of unexpected market aversions or precipitous price swings in what remains a highly volatile ecosystem; this points to the desirability of subsequent recalibrations that can better respond to the unpredictability of the nascent blockchain sector.

Another study given by Awoke et al. (2020) who also estimated the models to predict the price volatility of Bitcoin, one of the most famous cryptography currencies because of the high acceptance among investors, researchers, traders, and policymakers. The research problem focused on the effectiveness of efficient deep learning techniques to mitigate the problem of flexible Bitcoin price which destabilizes relationships and trade in international relations. Specifically, the study compared two time-series forecasting techniques: Two types of Recurrent Neural Networks (RNNs) have been used in the present study, namely Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). The performance of the given models was tested for the ability to predict Bitcoin prices, and it was found that both LSTM and GRU models showed good results in dealing with the feature for high volatility in the price while ensuring a high accuracy of the results. A possible weakness of the approach could be the disturbance of the optimal models' calibration due to the market's consistently changing values and dependence on external events in extreme market conditions.

Patel et al. (2020) proposed a hybrid cryptocurrency price prediction scheme that combines Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, focusing specifically on predicting the prices of two cryptocurrencies: Litecoin and Monero. More specifically in respect to price volatility and unpredictability, this study will seek to outline leveraged risks to the investments within cryptocurrencies. The LSTM and GRU machine learning and deep learning approaches were adopted in the study in order to enhance the accuracy of the cryptocurrency price prediction. The study showed that in the testing, the suggested hybrid model was useful in providing precise outcomes on the Litecoin and Monero prices; thus, can be used in other types of cryptocurrencies also. On the downside, this approach could be vulnerable to the market forces and other extrinsic influences which affect the market price fluctuation, and this may be disadvantageous in any real-world application of the model. It will also examine studies that might help to enhance the model's capacity to operate in an environment characterized by such dynamics.

Finally in Kim et al. (2022), the authors put forward a new model to forecast the price of Bitcoin denoted as BTC with which the nature of Cryptocurrencies posed a threat of excessive volatility. Data adopted by the study are based on on-chain data, which are exclusive information records belonging to blockchain of cryptocurrencies. To improve prediction performances, the authors proposed a self-attention-based multiple Long Short-Term Memory (SAM-LSTM) model which integrates related multiple LSTM units for on-chain data groups with attention. Benchmark real BTC price data were used to assess time-variant Three-Tiered PRICE autonomy assessment framework performance, which was presented to show its feasibility with promising performance measures such as MAE = 0.3462, RMSE = 0.5035, MSE = 0.2536, and MAPE = 1.3251. Nevertheless, it could be a limitation in this piece because prediction accuracy might be decreased depending on the availability and quality of on-chain data.

3 Methodology

The methodology for this study is visually represented in Figure 1, outlining the sequential steps involved in the development and evaluation of machine learning models for Bitcoin price forecasting.

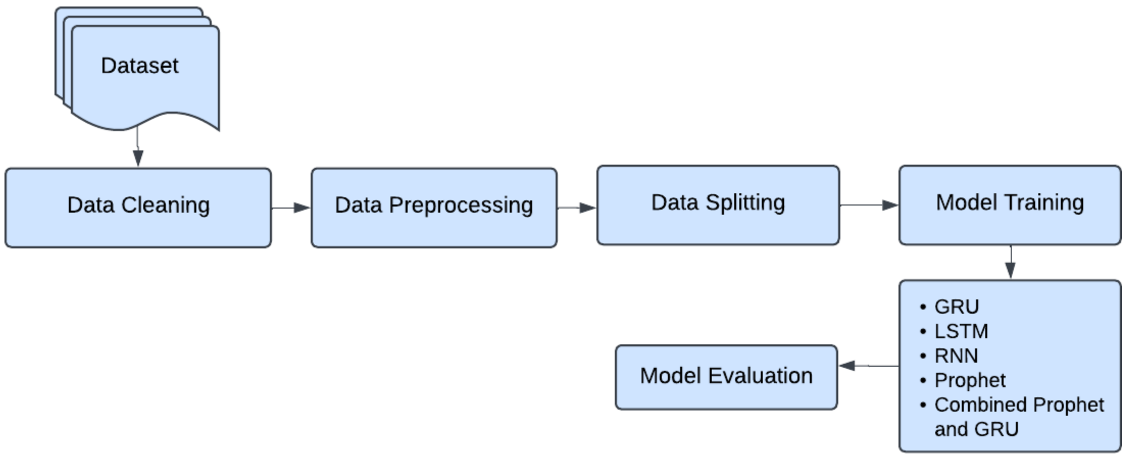

Figure 1: Proposed Workflow System

3.1 Data Cleaning

At the start of data cleaning process, the dataset is edited in a way that information not required for analysis is deleted. Specifically, the 'Unnamed: 0' column which is used as an index column is also removed as it gives no essential data for creating or analysis of the model. Next, and most importantly for time series forecasting, the 'Date' column is converted to datetime to be compatible with time series functions and correct indexing. This transformation enables proper chronological sorting and fosters time-associated techniques inherent with models such as GRU, LSTM, RNN, and Prophet models. After formatting the first column is set as an index of DataFrame using the command df.set_index([‘Date’]) It is crucial to set the ‘Date’ as the index because models used for time analysis, that involve different time instances, must be able to detect patterns and dependencies within the information. It corrects the data so that the final look of the dataset is clean and very much suitable for the forecasting and prediction. This ensures the dataset is as well-structured and aligned to pave way for other highly complex models such as GRU, LSTM, RNN, Prophet, and Prophet-GRU hybrid that will greatly increase the efficiency and reliability of the opening and closing price for Bitcoin.

3.2 Data Preprocessing

Data preprocessing has some key steps to perform on the Bitcoin Historical Dataset with the aim to improve the efficiency of analysis and models training. First, the ‘Timestamp’ is transformed into a regular ‘Date’ format and set as the index later to achieve correct dates alignment with time series data. Unnecessary columns, such as 'Unnamed:, ‘0’ values from the variable counts |c2_c||e2|c2_c–e2|0 of candidates, and are excluded to eliminate noise. Absences and repeats are problematic and addressed to avoid compromise of data consistency. OHLC prices and volume data are normalized through some methods like Min-Max or StandardScaler, so that values can be made the best out of them in the model. This preprocessing allows for neat data input, fine-tuned for accurate forecasting and prediction exercises.

3.3 Dataset Description

Bitcoin Historical Data is a detailed minute-based dataset that exhibits Bitcoin price change and volume from the year 2015. The context of this dataset includes raw information pertinent to market analysis and forecasting in the crypto domain. It has every kind of data; open, high, low and closing (OHLC) for every minute of the day, which provides great insights into the movement of the market. In addition to the Price data, the dataset also includes Trading Volume done in Bitcoin[1] (BTC) and US Dollar (USD) to explain the activity level of the market. Use of UNIX format timestamps and related human understandable date-time representation has saved time to the maximum possible extent and provides robust temporal data integration. This data set is particularly good for the trader who wants to update his or her strategies, the data scientist who is building a model, or the researcher probing the dynamics of the crypto market. Its content density enables one to gauge patterns, deviations and opportunities as it gives a bird’s eye view of market movements at the minute level. Due to historical data, the dataset assists the users in identifying trends within various phases of a particular market that can be crucial in the constantly fluctuating environment of Bitcoin trading.

3.4 Data Splitting

The Bitcoin price data is preprocessed, normalized, and split into 80% for training and 20% for testing. This ensures the model has sufficient historical data to learn patterns, seasonality, and temporal trends, while also testing its predictive power on unseen data. The training set captures most price fluctuations to identify patterns, while the test set evaluates

[1] https://www.kaggle.com/datasets/swaptr/bitcoin-historical-data

the model's ability to generalize new scenarios. Open and Close prices are split independently to ensure unbiased analysis. This approach is essential for accurate time series forecasting in Bitcoin’s volatile market.

3.5 Data Visualization

However, the line graph in figure 2 was constructed to illustrate the closing price of Bitcoin in USD in a two-week period from February 7 to February 21, 2023. The picture brings in a single unbroken blue line to follow the daily closing values, superimposed on a Cartesian coordinate system which has time on the horizontal axis and price which ranges roughly $21,500 to $25,500 on the vertical axis. From the graph, it is easy to infer that there were considerable fluctuations in the price in this month; a sharp decline in the first week from $15000 to $8500 and then stagnated until February 15 up to a decline in the price. From this date, there is a clear tendency of an increase in the price with oscillations rising even above $25,000, with variations staying at levels above $24,000-$25,000.

Figure 2: Bitcoin Close Price Over Time

The trading volume of Bitcoin is shown in figure 3 with two y-axis, one for the BTC (orange line) and the other for USD (green line) from February 7 to February 21, 2023. This type of chart uses two forms of volume on a twin-coordinate temporal x-axis, enabling side-by-side comparison between trading with both forms of currency. It is possible to note several hype points occurred in the trading volume: However, the most concentration has been observed near February 15-17 when the USD volume appeared to be almost 1.2 million units. This is evident from the orange line which represents the volume of the BTC and it has corresponding peaks, however, the relationship between the volumes of BTC and USD are different which means that there is a fluctuation in the price of the Bitcoin during the high volume trading period.

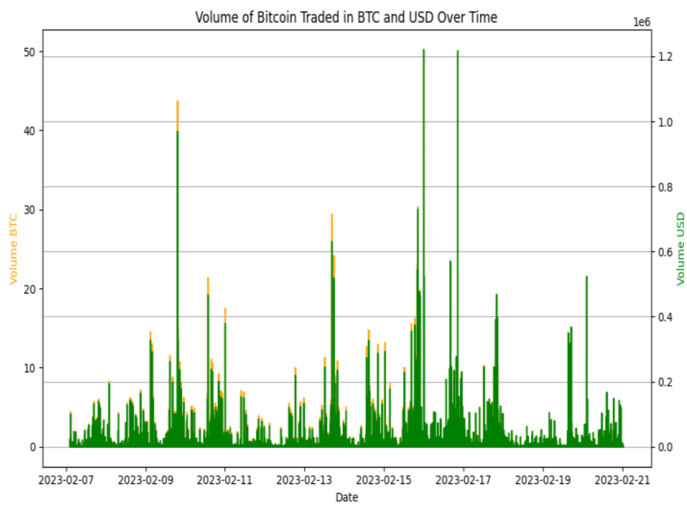

Figure 3: Line Graph

Figure 4 shows Graphic representations of the shape, spread, or distribution of Bitcoin prices based on four significant trade indicators namely the Open, Close, High, and Low. The visualization is in the form of four box plots side by side in 2 rows and 2 columns, each box plot presents the statistical summary of prices. Prices of all four plot types are dissimilar: the median price is around 23,000 USD; the corks filled with green color show the quartile width ranging from 22,000 to 24,500 USD.

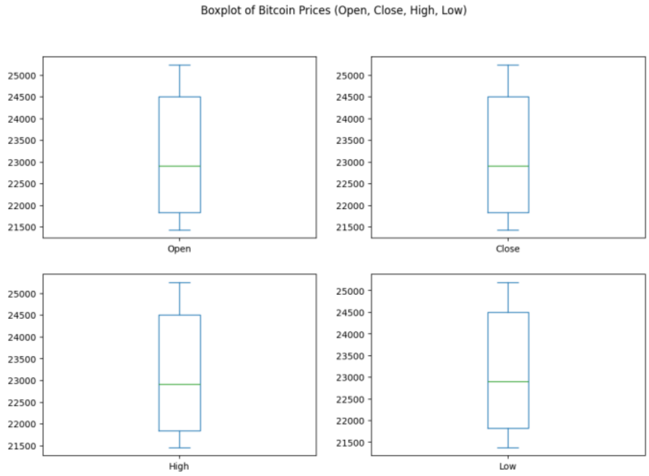

Figure 4: Boxplot of Bitcoin Prices (Open, Close, High, Low)

Figure 5 showing A business insight is a graphical representation of the correlation between various Bitcoin trading parameters with the help of a few colors ranging from blue for low correlation to red for high. The heatmap indicates a strong positive correlation (1.0 in dark red) between Open, High, Low price, and Close while Timestamp is moderately positively correlated (0.67) with these price metrics. Volume numbers in BTC and USD are very close to 1 to show a high positive relation between them, while volume numbers show very low relations of 0.02 to 0.04 (in blue) with price anomalies and time stamps, which indicates that trading volume in the analyzed period is relatively independent of the price movement.

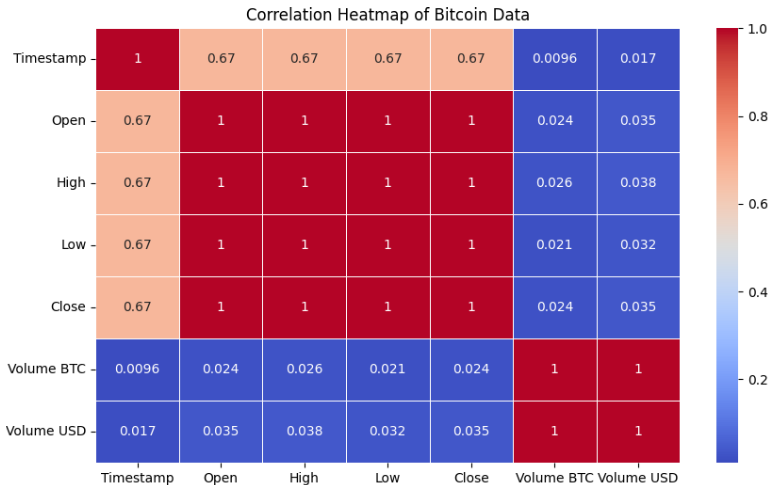

Figure 5: Correlation Heatmap of Bitcoin Data

Figure 6 represents a time series line graph that consists of two important performance indicators of Bitcoin price trend in a certain period in February 2023. The figure shows the Bitcoin close price in blue and the 1000-day moving average in red from February 7th to February 21st, 2023. Sort of y: US$ price range from around 21500 UP to approximately 25000; Sort of x: Date.

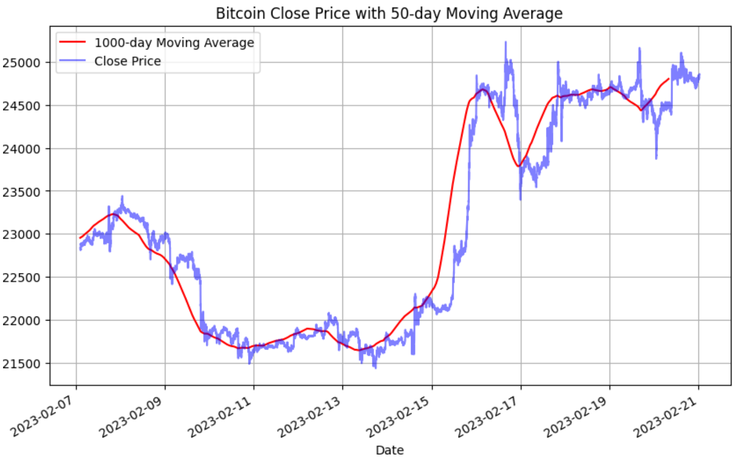

Figure 6: Bitcoin Close Price with 50-day Moving Average

Figure 7 illustrates a histogram of trading volume in USD of cryptocurrency transactions. Looking at the graphical analysis of the data the frequency distribution follows a positively skewed distribution, and most of the volumes are located close to the zero on the y-axis. The x-axis has its scale from 0 to 1.2 million USD, while the Y axis represents frequency that ranges from 0 to 20000. The shaded area encircling the purple bars indicates an almost uniform distribution; that is most trading occurs in stocks with the lowest trading volume, and the frequency considerably drops as the trading volumes rise up, implying that high volume trades are relatively rare in this market.

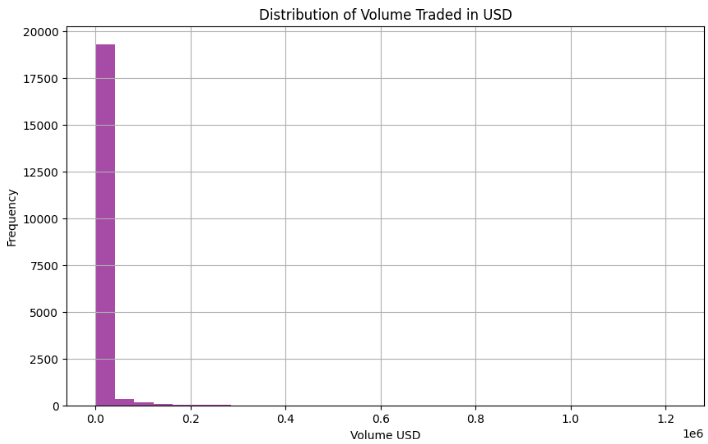

Figure 7: Distribution of Volume Traded in USD

Figure 8 showing the Scatter Plot of Bitcoin Volume against the Closing Price. This visual provides BTC trading volume in terms of BTC on the x-axis as well as the closing value in USD on the y-axis. The number is represented by red dots: this chart shows that the density of trading activity sets a very high density at lower volume (0-5 BTC), and a price range between $21950 to $25400. The plot shows that the density of trades decreases after a certain level of volume – higher than 10 BTC, it indicates that the frequency of large-volume transactions in this dataset is limited.

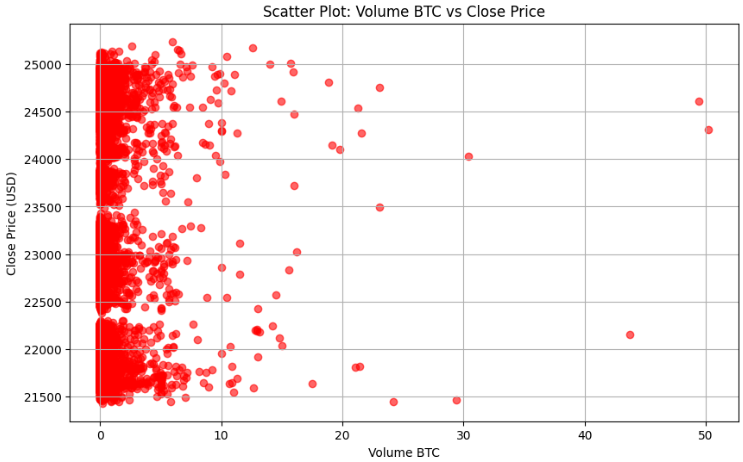

Figure 8: Scatter Plot: Volume BTC vs Close Price

3.6 Models and Evaluation

In this study, I employed five distinct forecasting models to predict Bitcoin's opening and closing prices: RNNs, GRUs & LSTMs, Prophet, and GRU- Prophet, the proposed hybrid model. The GRU, LSTM, and RNN categories refer to Recurrent Neural Networks which are ideal for weakly dependent learning from data of an ordered nature including the fluctuation in prices over time. Such models, especially the Arima and exponential smoothing models, are ideal for more frequent financial data since price information in the past is vital for the future trends. However, Prophet exists, a time series forecasting tool that was created by Facebook to address issues like seasonality and trends in a data set, therefore makes a good candidate tool for handling financial series forecast when this characteristic exists. To incorporate the advantage of Prophet in capturing the trends and seasonality into GRU's ability to learn long dependencies, a new model called Prophet-GRU was proposed. To evaluate the performance of these models, three key metrics were used: And among these common statistical measures, most of the scientists have R^2 score, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). R^2 score checks the goodness of the model in predicting the variation in the target variable where higher the R^2 score the better is the model. RMSE describes the average deviation from the true values through a variance, lower numbers showing better performance from the model. Likewise, MAE offers an average of the absolute disparities of the predicted values and the true values given, with an inference of a model that yields value prediction as close to the actual values. These evaluation criteria provide a well-rounded perspective of each model about the accuracy of the forecast it presents for the Bitcoin prices so as to enable a comparison of the performance of each model.

4 Design Specification

This architecture diagram represents the Gated Recurrent Unit (GRU) which is a type of recurrent neural network (RNN) which is used in several natural language processing and sequential data tasks. This figure also indicates the flow of the GRU model's information flow for a single time step. In Figure 9 below, the separate input at the current time instant is denoted by 'Xt' while the hidden state at the time (ht-1). These inputs are fed into the GRU, which has two gates: for this purpose, two gates are used: the reset gate (r) and the update gate (u). The reset gate decides the amount of previous hidden state that the network should forget, the update gate gives an amount that should be used in the computation of new hidden state number ‘ht’. The output of the GRU is named ‘Tanh’ which is the last and final output of the model for the target time step.

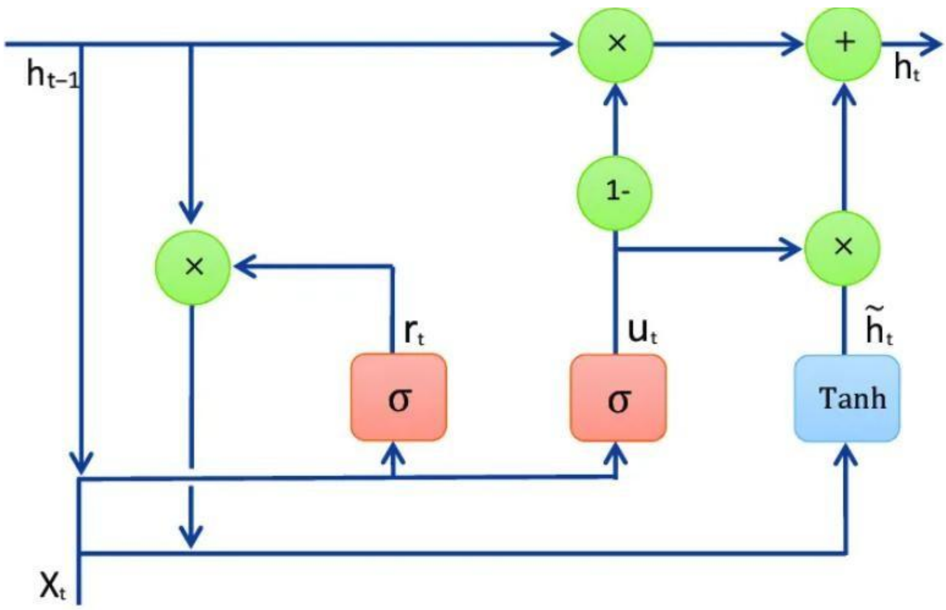

Figure 9: Architecture Diagram of GRU by (Dev et al., 2021)

Besides GRUs, the architecture includes the Prophet model, a stable statistical model that is suitable for time series predictions. Talking of its strengths, the Prophet model works well in terms of recognizing seasonality and the trends component in data. It must take the input data organized in a format in which the ‘ds’ is the time step and the ‘y’ is the value. This design makes it possible to capture trends over long periods, and periodicities, hence suitable for the model. This study finds that combining GRU and Prophet has the advantage of both models. The GRU deals with sequential dependencies within the data and gives brief forecasts, the Prophet model further adjusts these forecasts in the view of the seasonal and trend components. This makes the forecast more reliable for dynamic datasets such as Bitcoin’s opening and closing prices enhancing reliability and accuracy of its forecast.

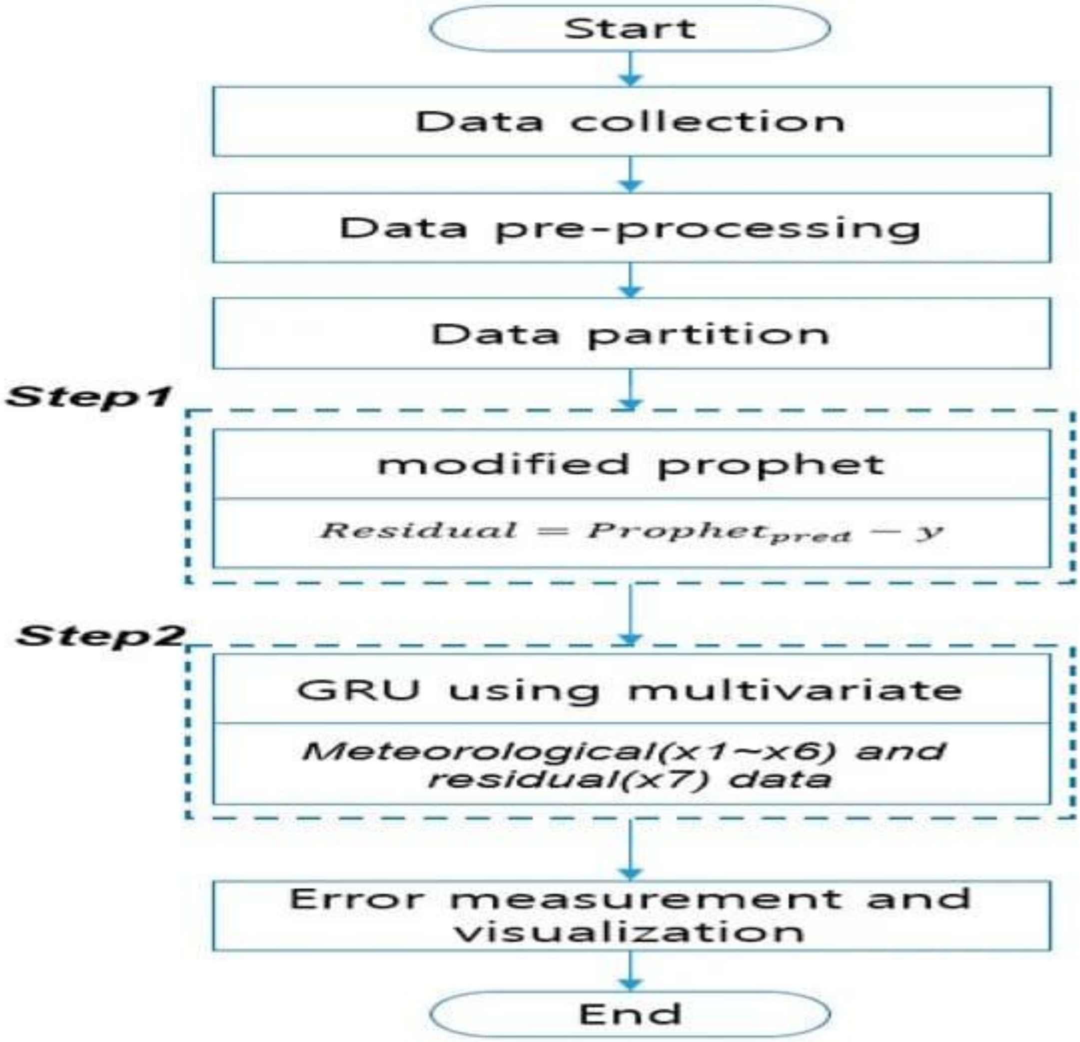

Figure 10: Architecture Diagram

5 Implementation

The hyperparameter settings for the implemented models are given below
• optimizer='adam' for efficient gradient-based optimization.
• loss='mean_squared_error' to minimize prediction errors.
• And training the models for epochs=10 with a batch_size=32 to balance computational efficiency and model performance.

These parameters were consistently applied across all models to ensure a uniform experimental setup and facilitate fair performance comparisons.

5.1 GRU Implementation for Open and Close Prices

In the specific case of Bitcoin and its opening and closing prices, the following GRU implementation procedure is followed. First, the data is normalized to select and reshape the specific series 'Open' and 'Close' from the dataset containing multiple variables on the stock data. Hence, Optimal scaling is done using MinMaxScaler, which scales the values to range from between 0 and 1 to facilitate model convergence. The data sets of each case study are divided into 80 percent for training and the remaining 20 percent only for testing the model to check accuracy on new data. Using sliding window, time series datasets are generated and to capturing temporal dependencies, 60 time-steps are used for predicting the next price of stock. After that, the data is transformed into the three-dimensional structure necessary for the GRU model: samples, time steps, features. GRU model is created and applied using the sequential model construction where the stacked GRU layer used here consists of only 50 units and is set to return sequence to pass through the next layers. The next component is one more GRU layer and the last Dense layer with one unit for prediction. The training of the model in this work is done using the Adam optimizer and mean square error as the loss function. We train the system using 10-epochs and batch size of 72 and validation is done on the test set.

5.2 LSTM Implementation for Open and Close Prices

The LSTM for estimating Bitcoin opening and closing prices involves dataset preparation and model building in the following way. To form time series datasets, a sliding window approach is used; for each input sequence we used 60-time steps to predict the next price step, which is sufficient to capture temporal dependencies properly. Before feeding to LSTM, input data is transformed in a three-dimensional format with three dimensions representing the samples, time steps and features respectively. The LSTM model created herein is a sequential model and is structured with an LSTM layer comprised of 50 units with the option of returning sequences for the stacking of several layers. To avoid overfitting a Dropout layer with 20% dropout rate is included. There is another LSTM layer with 50 units, and it does not output sequences which in a way summarizes the learned features. Overfitting dangers are minimized by the Dropout layer which is again implemented at this point. That is why a Dense layer with one unit is added at the end to provide predictions of the next time step. Training is carried out with the batch size of 64 over the 10 iterations with the benefit of knowing the sequential arrangement of the data for the proper tuning of the weights on the forecast.

5.3 RNN Implementation for Open and Close Prices

RNN used in modelling Bitcoin opening and closing price involves the determination of the dataset, creation of the model and assessment of the model. For the first data, the number of rows is increased, and only 'Open' and 'Close' price columns are selected and then reshaped to match the model. The data is preprocessed by scaling it using MinMaxScaler to avoid mathematical issues affecting training effectiveness. A sliding window approach is used to form time series datasets where the input is a sequence of 60-time steps, and the goal is to predict the next price. The data is thus formatted to be a three-dimensional structure that being sample, time step and feature as would be required by the RNN model.

Thus, the RNN architecture is created by Sequential model with two layers of SimpleRNN, both contain 50 units each. The proposed training is done for 10 epochs with batch size of 72 and model evaluation is made on the test set to check the model performance.

5.4 Prophet Implementation for Open and Close Prices

The Prophet for forecasting of the opening and closing prices of Bitcoin involves the preparation of the type of data set to be used, model fitting and model evaluation phases. First, the 'Open' and 'Close' price indicators are selected and the value in each cell is divided by the maximum value, to get a number between 0 and 1, which helps the model learn better. To make sure that data is used in a right way, it is divided into testing set with a size of 20%

from data and training set which is the same size of 80% from data. Prophet has a unique input data format; thus, the prices are grouped into a DataFrame with the columns named 'ds' containing dates and 'y' containing the values. The 'ds' column returns the original date index in order to line up the time series data. The Prophet model is initiated without daily seasonality since the prices of Bitcoins can have longer cycles than daily.

5.5 Combined Prophet and GRU Implementation for Open and Close Prices

When it comes to forecasting the opening and the closing price, Prophet and GRU are integrated and for this 'Open' and 'Close' are preprocessed through MinMaxScaler and then prepared for modeling. The features of the data are used to randomly select feature sets for the training/testing of the algorithm and a data creation function to develop, with a time series split at 60 time steps. Another GRU model is created from the 'Open' and 'Close' price data and trained with two GRU layers along with one Dense layer for regression output. Then, prediction is conducted for the test set, and the predicted values are scaled back to their original scale, if necessary. These predicted values are further utilized in the Prophet by building a new DataFrame to represent the seasonality of daily movements in the prices.

Data for Time Series Plot with Confidence Interval depicted in Figure 11. This plot is a time series trend from 2023 to 2039, representing the predicted values by black dots, the estimated trend line by the blue line, and the 95% confidence interval by the light blue shaded area. The records on the y-axis are variable and vary between roughly 24,000 and 25,200 for the period under consideration. The wave plot indicates many definable fluctuations, which clearly shows its upward momentum starting from roughly 2031 and then experiencing a decline, again rising and stabilizing at around 2035 onwards.

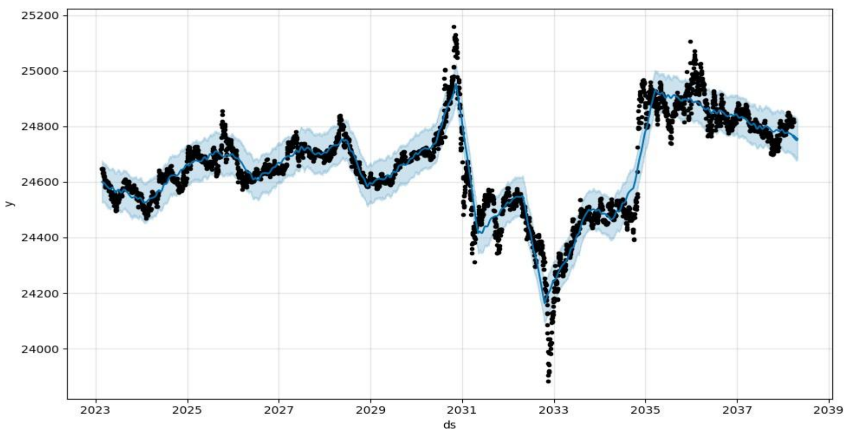

Figure 11: Time Series Plot with Confidence Interval

5.6 K-fold Cross Validation

Evaluation of GRU and RNN models to the time series forecasting is done by applying K-Fold cross-validation. Thus, KFold from sklearn.model_selection divides the dataset into 5 equal divisions for a credible assessment since models will be solutions on different portions of the dataset. For each fold, the model's training set is the training split and the model's testing set is the corresponding test split. The GRU and RNN models are defined with the help of helper functions (create_gru_model and create_rnn_model), including two layers and a dense layer for regression. The models are optimized using the Adam optimiser and the mean squared error loss function. For each fold in training, the models are trained for 50 iterations with batch sizes of 72 and then make predictions on the test set.

Results for K-fold shown below.

Table 1: K-fold Cross Validation Performance Metrics of the GRU and RNN

Metric	GRU	RNN
R^2 Score	0.9846	0.9805
RMSE	0.0060	0.0068
MAE	0.0048	0.0055

5.7 Tools and Technologies

From the forecast of the Bitcoin prices using time series, the work employs a comprehensive collection of the Python libraries. As mentioned, the primary programming language utilized in the process of data analysis is pandas; it offers various features required to manage the dataset carefully. Data visualization is done using libraries Matplotlib and Seaborn that facilitate making detailed plots to look for patterns, trends or anomalies in the data. Having efficient numerical operations and handling arrays, NumPy guarantees trouble-free data processing. To normalize the data scikit-learn's MinMaxScaler is used which scales the data between 0 and 1 for better model training and testing. Other assessment criteria including mean squared error (MSE), mean absolute error (MAE), and R^2 score are also calculated with the help of scikit-learn tools that gives the quantity measure to its exactness. Deep learning models are developed in TensorFlow and Keras, which is a high-level interface in TensorFlow; the Sequential model framework further allows the construction of layer-wise architecture. There are specific layers such as Dense, LSTM, GRU, and Dropout which are chosen for model creation for building stable and sufficiently flexible models for comprehending temporal patterns in the data. Furthermore, important mathematical computations that involve the approximation of the square roots also applies to Python's library thus making the forecasting of time series to be precise and efficient.

6 Evaluation

6.1 Case Study 1: GRU Model

From table 2, the GRU model performance for the prediction of bitcoin opening and closing prices is depicted. For both prices' predictions, we see that the models yield exceptional R^2 Scores of 0.9959. The RMSE figures are low, highly magnified; in fact, the opening price RMSE = 0.0041 is almost near zero implying negligible prediction errors; Similarly, closing price RMSE = 0.0026 is also close to zero. Just like that, the MAE values, the average of the absolute errors, are also very low (0.0027 and 0.0042). These measures show more efficiency of the GRU model in predicting the Bitcoin prices with a slightly higher degree of success when predicting the closing prices.

Table 2: Performance Metrics of the GRU Model

Metric	Open Price (GRU)	Close Price (GRU)
R^2 Score	0.9959	0.9959
RMSE	0.0041	0.0026
MAE	0.0027	0.0042

6.2 Case Study 2: LSTM Model

The following table 3 gives a snapshot of how the LSTM Model has performed in the prediction of the opening and closing price of Bitcoin. The R^2 Score shows how much of that change in the actual price is explained by the model, which for opening prices is 0.9755 while for the closing prices is 0.8252. To assess the accuracy of the predictions, the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE) have been used. The RMSE establishes an average prediction error and gives a lower error of 38.4006 for the opening price while represents a higher error of 102.8150 for the closing price. These findings imply that the developed model provides substantially better forecasts when predicting opening prices than the closing prices.

Table 3: Performance Metrics of the LSTM Model

Metric	Open Price (LSTM)	Close Price (LSTM)
R^2 Score	0.9755	0.8252
RMSE	38.4006	102.8150

6.3 Case Study 3: RNN Model

Below is a summarization of the performance of the RNN model in the prediction of the opening and closing price of Bitcoin in Table 4. The values for the R^2 Score are expressive, pointing a high level of accuracy in the proposed predictions, with 0.9914 for the closing prices and 0.9739 for the opening prices. The RMSE and MAE values reveal low errors, with better performance observed for closing prices (RMSE: 0. Overall, the closest values were obtained when the model was evaluating closing prices (RMSE: 0.0084, MAE: 0.0060) compared to opening prices (RMSE: 0.0157, MAE: 0.0104). These outcomes display the efficiency of the RNN model in estimating Bitcoin prices with a minor preference toward closing price prediction over opening prices.

Table 4: Performance Metrics of the RNN Model

Metric	Open Price (RNN)	Close Price (RNN)
R^2 Score	0.9739	0.9914
RMSE	0.0097	0.0048
MAE	0.0104	0.0060

6.4 Case Study 4: Prophet Model

Table 5 shows the evaluation of the Prophet model for modelling and forecasting Bitcoin's opening and closing price. The R^2 Scores are observed as -160545.682 for the opening prices and -154211.521 for the closing prices which suggest a poor model fit and poor capacity to account for the variation in the given price data. The forecast errors are quite large as indicated by the RMSE values (24.8082 for opening and closing: 24.3558) and MAE values (28.4879, and 27.9656). The figures presented in this subsection indicate that there is a problem of capturing the multi-format of the data and poor capability of the Prophet model to provide better forecasts when it comes to Bitcoin prices.

Table 5: Performance Metrics of the Prophet Model

Metric	Open Price (Prophet)	Close Price (Prophet)
R^2 Score	-160545.682	-154211.521
RMSE	24.8082	24.3558
MAE	28.4879	27.9656

6.5 Case Study 5: Combined Prophet and GRU Model

The following table 6 shows the performance results of the hybrid Prophet-GRU model in the forecast of opening and closing prices for Bitcoin. In the case of open prices, the model is quite satisfactory with R^2 Score of 0.990 and the RMSE (17.023) and MAE (11.650) are smaller. When it comes to close prices, the model is adequate but a little less effective with the R^2 Score of 0.926, RMSE: 46.699, while MAE is 44.172.

Table 6: Performance Metrics of the Hybridization of Prophet and GRU Model

Metric	Open Price (Hybrid)	Close Price (Hybrid)
R^2 Score	0.990	0.926
RMSE	17.023	46.699
MAE	11.650	44.172

6.6 Discussion

In this study, the primary concern is opening and closing price forecasting of Bitcoin using state-of-art time series models such as GRU, LSTM, RNN, Prophet and Prophet-GRU. The theoretical motivation behind these models relies on the notion that extraction of temporal dependencies in data corresponds to more intricate temporal patterns. LSTM and GRU are great for sequential data of prices of different assets including Bitcoin prices because RNN can capture long-term dependencies of data. Like other RNNs, GRU and LSTM models are built to overcome some problems including the vanishing gradients and perform better on larger sequences of data. On the other hand, Prophet, a forecasting tool, developed by Facebook company, is outstanding in coarse details and in interpreting trends and seasonality of different noises which make it useful in predicting the price of the cryptocurrency. Since Prophet is good at capturing trends and seasonality and GRU is effective in capturing sequential dependence, the hybrid model is to utilize the merit of both for further improvement of the result. The theoretical foundations of these models indicate that they would help to explain the short-term and long-run changes in the price of Bitcoin. This integration of models is expected to improve the forecasting accuracy as well as reduce error in returning the best results in the specific cryptocurrency markets Thus, avenues for future

7 Conclusion and Future Works

Conclusion:

In this study, the GRU model is reported to provide the highest accuracy of forecasting opening and closing prices of Bitcoin. Thus, we were able to use the advantages of the two models to incorporate long-term trends and short-term noise into the analysis. Prophet also performed impressively by capturing the seasonal patterns and general shift of the market business, while the GRU model was capable of mastering other temporal property and making precise forecast based on one-minute data. The findings show that our proposed GRU model enhances the modelling of Bitcoin price compared to other models.

Implications:

To the traders, investors and other researchers involved in trading or studying the cryptocurrencies, this study has the following implications. Utilizing the Prophet and GRU model for traders allows traders to make better decisions and have better outcomes when predicting and forecasting price. Citizens could also receive more precise predictions, thus being able to weigh the tendencies of the market and extend their subjective estimations. Further, this model also offers utilized better research avenues to researchers to investigate more about the market and the phenomenon that impacts the price of Bitcoin. This hybrid model provides a better tool to forecast the price trend when both the general trend and seasonal variation are incorporated into the formula in the highly unpredictable cryptocurrency market.

Limitations:

Nevertheless, the findings of this study range with certain limitations. First, the model only uses the price and volume of Bitcoin and does not take into consideration such other factors as the macroeconomic system or regulatory changes. Furthermore, it has also been observed that although the integrated model provides improved predictive capability, there still exists limitations in capturing situations of extreme market fluctuation in volumes typical of the cryptocurrency market. Furthermore, the model may not perform well in a different time frame or under a different market scenario, and the model must be fine-tuned for better performance during the dynamic market situation.

Future Works:

Future work can try to include other forms of input like the social media sentiment or recent news articles that may add to the accuracy of the model. Moreover, it is interesting to try other variants of hybrid models, using such complete deep learning, and the accuracy will be as much higher. The study would also benefit from extending it to other cryptocurrencies, or from applying the model on the given set of cryptocurrencies under different market conditions. Finally, daily updating the model using the latest data and constantly recalculating the inputs to feed the model will be crucial as the Bitcoin market changes dynamically.

References

1 Aanandhi, S.P., Akhilaa, S.P., Vardarajan, V. and Sathiyanarayanan, M., 2021, December. Cryptocurrency price prediction using time series forecasting (ARIMA). In *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 598-602). IEEE

2 Awoke, T., Rout, M., Mohanty, L. and Satapathy, S.C., 2020. Bitcoin price prediction and analysis using deep learning models. In *Communication Software and Networks: Proceedings of INDIA 2019* (pp. 631-640). Singapore: Springer Singapore.

3 Bâra, A. and Oprea, S.V., 2024. The Impact of Academic Publications over the Last Decade on Historical Bitcoin Prices Using Generative Models. *Journal of Theoretical and Applied Electronic Commerce Research*, *19*(1), pp.538-560.

4 Cheng, J., Tiwari, S., Khaled, D., Mahendru, M. and Shahzad, U., 2024. Forecasting Bitcoin prices using artificial intelligence: Combination of ML, SARIMA, and Facebook Prophet models. *Technological Forecasting and Social Change*, *198*, p.122938.

5 Dey, P., Hossain, E., Hossain, M.I., Chowdhury, M.A., Alam, M.S., Hossain, M.S. and Andersson, K., 2021. Comparative analysis of recurrent neural networks in stock price prediction for different frequency domains

6 Doumenis, Y., Izadi, J., Dhamdhere, P., Katsikas, E. and Koufopoulos, D., 2021. A critical analysis of volatility surprise in Bitcoin cryptocurrency and other financial assets. *Risks*, *9*(11), p.207.

7 Elsayed, A.H., Gozgor, G. and Lau, C.K.M., 2022. Risk transmissions between bitcoin and traditional financial assets during the COVID-19 era: The role of global uncertainties. *International Review of Financial Analysis*, *81*, p.102069.

8 Giudici, G., Milne, A. and Vinogradov, D., 2020. Cryptocurrencies: market analysis and perspectives. *Journal of Industrial and Business Economics*, *47*, pp.1-18.

9 Kim, G., Shin, D.H., Choi, J.G. and Lim, S., 2022. A deep learning-based cryptocurrency price prediction model that uses on-chain data. *IEEE Access*, *10*, pp.56232-56248.

10 Kyriazis, N., Papadamou, S. and Corbet, S., 2020. A systematic review of the bubble dynamics of cryptocurrency prices. *Research in International Business and Finance*, *54*, p.101254.

11 Panigrahi, S.K., 2023, September. Forecasting Twitter's Favorite Cryptocurrencies Based on a Comparative Assessment of SARIMA, LSTM and Fb Prophet Machine Learning Approach. In *2023 International Conference on Sustainable Islamic Business and Finance (SIBF)* (pp. 42-45). IEEE.

12 Patel, M.M., Tanwar, S., Gupta, R. and Kumar, N., 2020. A deep learning-based cryptocurrency price prediction scheme for financial institutions. *Journal of information security and applications*, *55*, p.102583.

13 Parekh, R., Patel, N.P., Thakkar, N., Gupta, R., Tanwar, S., Sharma, G., Davidson, I.E. and Sharma, R., 2022. DL-GuesS: Deep learning and sentiment analysis-based cryptocurrency price prediction. *IEEE Access*, *10*, pp.35398-35409.

14 Phung Duy, Q., Nguyen Thi, O., Le Thi, P.H., Pham Hoang, H.D., Luong, K.L. and Nguyen Thi, K.N., 2024. Estimating and forecasting bitcoin daily prices using ARIMA-GARCH models. *Business Analyst Journal*.

15 Poongodi, M., Vijayakumar, V. and Chilamkurti, N., 2020. Bitcoin price prediction using ARIMA model. *International Journal of Internet Technology and Secured Transactions*, *10*(4), pp.396-406.

16 Tanwar, S., Patel, N.P., Patel, S.N., Patel, J.R., Sharma, G. and Davidson, I.E., 2021. Deep learning-based cryptocurrency price prediction scheme with inter-dependent relations. *IEEE Access*, *9*, pp.138633-138646.

17 Wahid, A.M.A., 2024. Time Series Analysis of Bitcoin Prices Using ARIMA and LSTM for Trend Prediction. *Journal of Digital Market and Digital Currency*, *1*(1), pp.60-74.