BTC price Prediction and Analysis

1. Introduction

Bitcoin, the first and most well-known cryptocurrency, has demonstrated significant price volatility since its inception in 2009. This characteristic, while making Bitcoin an attractive investment and speculative opportunity, also presents unique challenges in predicting its future price movements. Accurate predictions can help investors manage risks and capitalize on potential gains. However, the complex and speculative nature of Bitcoin prices, influenced by various economic, political, and social factors, makes forecasting particularly challenging.

Recent advancements in machine learning, particularly in the field of deep learning, have opened up new avenues for financial forecasting. Among the various models available, Gated Recurrent Units (GRUs) have emerged as a powerful tool for time series analysis due to their efficiency in learning from sequential data. GRUs, a type of recurrent neural network (RNN), are well-suited for modeling financial time series like Bitcoin prices because they can capture temporal dependencies and non-linear patterns that are common in such data.

This project aims to apply GRU models to predict the future prices of Bitcoin using historical data. The focus will be on developing a predictive model that transforms the data into logarithmic returns to stabilize the variance and improve the model's performance. Furthermore, the project will explore the impact of different sequence lengths on the model's accuracy and training time, ensuring that the predictive model is both effective and efficient.

2. Exploratory Data Analysis (EDA)

The initial phase of our research involved an extensive exploratory data analysis (EDA) on the historical Bitcoin prices sourced from Yahoo Finance using the yfinance library. This step was crucial to understand the underlying patterns, trends, and anomalies in the data.

Figure 1: BTC Price

The average opening price of Bitcoin during the analyzed period stands at approximately $15,000, with the highest recorded opening price peaking at $73,000 and the lowest observed at $5,000. Notably, since the onset of 2021, there has been a significant surge in the opening price, with an increase of over $4,000, reflecting a continued upward trend. Similarly, the average closing price mirrors the opening at $15,000, with the highest closing at $73,000. From the beginning of 2021, the closing price has also experienced a substantial rise, surpassing $4,000, and showing a consistent upward trajectory.

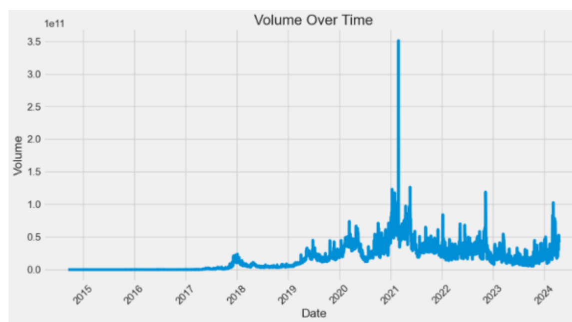


Figure 2: BTC Volume

The average trading volume during the period analyzed is approximately 45.03, with the highest recorded at 303.81 and the lowest at 11.46. Although 45 outliers were detected in the volume data, these are instances of significant spikes, indicating heightened trading activity.

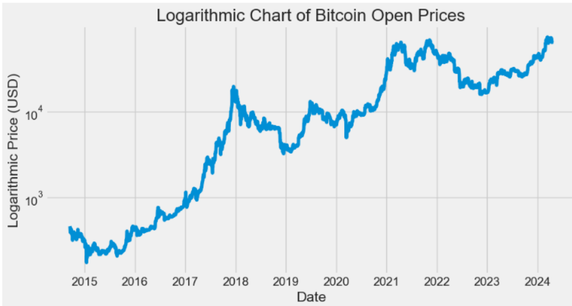


Figure 3: BTC LOG price

Upon transforming the Bitcoin price data into logarithmic scale (Log_Close), the visualization reveals a more pronounced linear trend. This transformation stabilizes the variance and makes the pattern of returns more homoscedastic, which is an essential assumption in many linear modeling techniques. The logarithmic transformation not only simplifies the patterns but also provides a clearer perspective where a linear regression line could potentially fit well. This indicates a more predictable behavior suitable for linear regression analysis and highlights the advantage of using logarithmic scale for financial time series that exhibit exponential growth or large variances.

The correlation analysis between various price metrics revealed a robust positive correlation of 0.99 between the opening and closing prices of Bitcoin, which suggests a strong tendency for the prices to close higher than they open on a given day, indicative of generally bullish daily price movements. This finding underscores the frequent alignment between opening and closing values, reinforcing the market's positive sentiment over the analyzed period. Conversely, the analysis identified a notable negative correlation between the trading volume and the opening price; this suggests that higher opening prices tend to coincide with lower trading volumes, possibly indicating investor hesitation at higher price levels. Additionally, a significant positive correlation between the opening price and the highest daily price was observed, indicating that days starting with higher prices tend to reach higher maximums, which aligns with upward market trends. The relationship between opening and closing prices appears symmetric, suggesting a balanced behavior between these two metrics throughout the trading day. Furthermore, a slight negative correlation was detected between trading volume and closing price, which might indicate that larger trading volumes are associated with days when the price closes lower than it opens, potentially reflecting higher selling pressure or profit-taking activities on those days.

Figure 4: BTC Halving

The visualization of Bitcoin price in relation to halving events presents a compelling narrative about Bitcoin's supply dynamics and their influence on price. As depicted in the chart, halving events, marked by the vertical dashed lines, have historically been followed by periods of substantial price increase. This pattern is consistent with the economic principle of supply and demand: as the rate of new Bitcoin creation halves, if demand remains steady or increases, the price is likely to rise.

In the period following each halving event, there is a noticeable appreciation in price, reflecting the market's anticipation of reduced supply growth. The first halving event, occurring in 2012 (not visible on the chart), set the stage for the subsequent cycles. The 2016 halving precedes a significant bull run that culminates in the late-2017 peak. Similarly, after the 2020 halving, the price trajectory shows an upward trend leading into 2021. The visualization illustrates these pivotal moments with clarity, showcasing the repeating cycles of Bitcoin's economic model.

It's important to note that while halving events can influence Bitcoin's price due to the supply shock, other market factors also play a significant role. This includes institutional adoption, market sentiment, regulatory news, and broader economic factors. However, the alignment of price surges post-halving offers insight into the market's cyclical nature and is an essential consideration for any predictive model focusing on long-term price trends.

3. Method

3.1 Data preprocessing

The study commenced with the retrieval of historical Bitcoin prices from Yahoo Finance. The dataset was cleansed to ensure accuracy, handling missing values and outliers. A logarithmic transformation was applied to the price data to stabilize the variance and normalize the distribution, allowing for a more effective application of statistical models.

Additionally, the dataset was split into a training set and a test set. The training set included data up to December 31, 2022, while the test set consisted of data from January 1, 2023, onwards. The MinMaxScaler from the scikit-learn library was used to scale the logarithmically transformed prices to a range between 0 and 1, fitting the scaler on the training data and applying it to both the training and test sets to prevent data leakage.

3.2 Model Architecture

A Gated Recurrent Unit (GRU) model, a variant of a recurrent neural network, was chosen due to its effectiveness in handling sequential data and its efficiency in capturing temporal dependencies. The GRU model architecture was designed with the following layers:

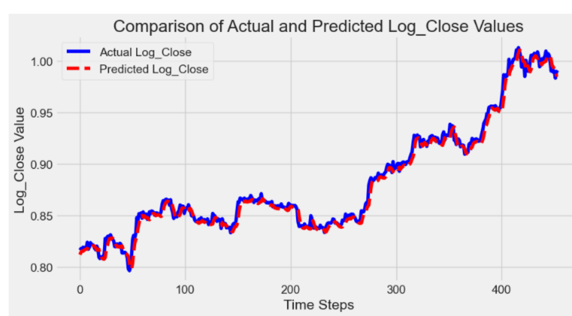
An input layer that accepts sequences of the scaled logarithmic prices. Two GRU layers with 50 units each and 'relu' activation functions, with the first GRU layer returning sequences to feed into the subsequent layer. A dense output layer with a single unit to predict the scaled logarithmic price of Bitcoin for the next time step. The model was compiled with the Adam optimizer and the mean squared error loss function, suitable for regression problems.

The training process involved feeding the scaled logarithmic price sequences into the GRU model. An early stopping mechanism was incorporated to halt training if the validation loss ceased to improve, preventing overfitting. The model was trained over multiple epochs with a batch size determined through experimentation to balance training time and convergence.

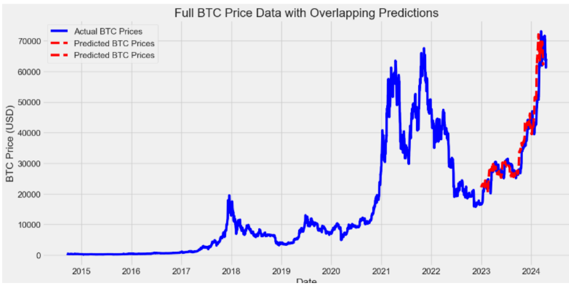
3.3 Model Evaluation

The model's performance was assessed by its ability to predict the scaled logarithmic prices in the test set. The predictions were inversely transformed to retrieve the original logarithmic price scale and further exponentiated to convert them back to the original price scale. The evaluation metrics included the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) on the test set, comparing the model's predictions with the actual Bitcoin prices.

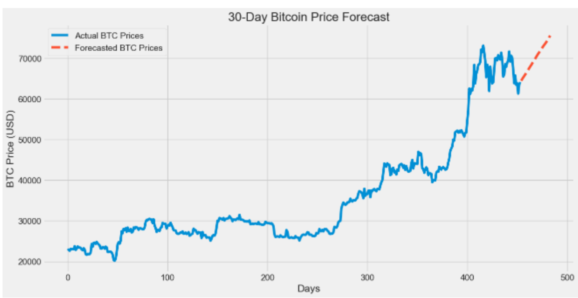
4. Result



The result of prediction using GRU shows pretty much accurate. The RMSE is 0.005 which is very low on error. This result of the price that transform to logarithmic.



The prediction after transforming back to the actual price also still shows accurate predict it. But there is a cache that the data being tested is already given actauly price. So basically it only predicts the next day successfully.



This is what happens when the model try to predict for the next 30 days. It is not able to capture the volatility of the market price, unlike the testing. The testing becomes a lagging price trend. The testing can predict accurately because it is given the actual previous price, and it just predicts the next price to be a little bit higher. This makes the price will follow what already happened. But if we just try to predict then next 30 days it will make only a straight line.

5. Conclusion

While the models demonstrated robust performance on historical test data, affirming their capability to learn from past trends up to 2022, they struggled with forecasting future values over a 30-day horizon. The forecasts generally showed only marginal increases from the last recorded prices, leading to projections that resembled a nearly straight line. This behavior illustrates the models' reliance on short-term historical patterns, potentially indicating an overfitting issue where the models excel on training data but fail to generalize to unseen data.

The challenge is exacerbated by Bitcoin's inherent volatility, which is influenced by a myriad of factors beyond historical prices. Such factors include market sentiment, global economic conditions, and sudden regulatory changes, which the current models do not account for. This reliance on recent trends without accommodating external variables limits the models' effectiveness in predicting downturns and rapid market shifts.

This project highlights both the potential and the limitations of using machine learning to forecast cryptocurrency prices. The findings underscore the need for a more holistic approach that incorporates a broader spectrum of data inputs, including external economic and sentiment indicators. Future improvements should focus on integrating ensemble methods to mitigate model biases, implementing continuous learning to adapt to new market behaviors, and enhancing feature engineering to capture deeper market dynamics.

Further research should also explore more advanced machine learning techniques that can dynamically adjust to new information and offer more nuanced interpretations of market movements. By extending the model's capabilities to better understand complex market forces, future iterations can aim to provide more accurate and actionable predictions, thus advancing the field of financial market forecasting in cryptocurrencies.