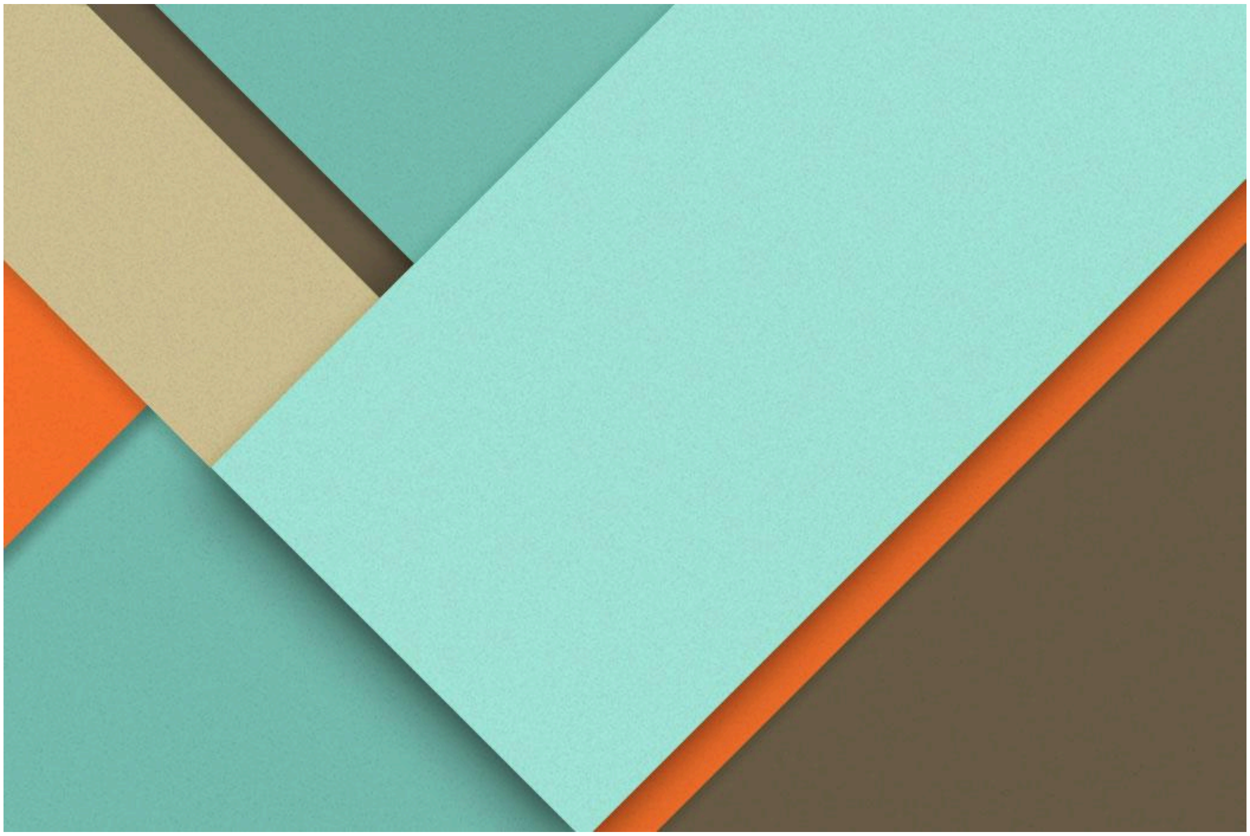


BitPredict - A Bitcoin Prediction App

Group Assignment

Group 1

206036P	MSV Gimhan
206001F	M.M Aabidh
206066G	M.Lakshitha
206116L	K.Senevirathne
206064A	K.A.R.P.Kumarasingha

Table of Content

Introduction

This research aims to use Long Short-Term Memory (LSTM) neural networks to construct and assess prediction models for predicting the closing prices of Bitcoin. As one of the most well-known and volatile cryptocurrencies, Bitcoin offers traders and analysts a lot of chances as well as difficulties. Reliable price prediction models can improve trading tactics, maximize profits, and more skillfully control risks. To identify the best model for predicting Bitcoin values, we examine the application and evaluation of three different LSTM models, each with a different architecture and set of features, in this study. A thorough understanding of each model's performance is provided by the evaluation criteria, which include Mean Squared Error (MSE), Mean Absolute Error (MAE), Sharpe Ratio, Total Return, Standard Deviation of Returns, and Trading Return.

Dataset Selection

The dataset contains records of Bitcoin market trading information spanning multiple years. It is essential for understanding Bitcoin's price movements, trading volumes, and other metrics crucial for analyzing and developing trading strategies in the cryptocurrency market. Understanding and analyzing this dataset is essential for developing effective trading strategies and gaining insights into Bitcoin's market dynamics and trends. The dataset analyzed in this project encompasses financial data from May 21, 2015, to May 21, 2024. It contains 3289 entries with the following columns

```
Top rows:

         Date        Open        High         Low       Close   Adj Close  \
0  2015-05-21  234.016006  236.242004  233.835007  235.343994  235.343994
1  2015-05-22  235.320999  240.968994  235.059998  240.348007  240.348007
2  2015-05-23  240.285995  241.024994  238.690994  238.871994  238.871994
3  2015-05-24  238.975998  241.977997  238.811005  240.953003  240.953003
4  2015-05-25  240.927002  241.020996  236.636993  237.110001  237.110001

     Volume
0  15108900
1  27003000
2  14605000
3  11508000
4  14423900


Bottom rows:

            Date          Open          High           Low         Close  \
3284  2024-05-17  65231.296875  67459.460938  65119.316406  67051.875000
3285  2024-05-18  67066.210938  67387.328125  66663.500000  66940.804688
3286  2024-05-19  66937.929688  67694.296875  65937.179688  66278.367188
3287  2024-05-20  66278.742188  71483.562500  66086.171875  71448.195313
3288  2024-05-21  71427.992188  71785.109375  70773.343750  71124.765625

         Adj Close       Volume
3284  67051.875000  28031279310
3285  66940.804688  16712277406
3286  66278.367188  19249094538
3287  71448.195313  43850655717
3288  71124.765625  55098576896
```

Detailed description of variables:

- Date: Represents the date of the recorded data point.
- Price: The closing price of Bitcoin on the given date.
- Open: The opening price of Bitcoin on the same date.
- High: Reflects the highest price of Bitcoin reached during the trading period.
- Low: Represents the lowest price of Bitcoin recorded within the same trading period.
- Volume: Represents Bitcoin's trading volume denoted in units such as thousands (K) or millions (M).

Data Preprocessing

1. Filtering Data

The dataset was filtered to retain only entries from May 21, 2022. It is a reduced dataset with 732 rows. It allows us to focus the analysis on more recent data. It may be more relevant for developing trading strategies in the current market environment. Since its launch, the dynamics of the Bitcoin market have changed significantly. Earlier data is potentially less reflective of current market conditions and trading patterns. By setting a lower bound on the data, it is prioritized to recent observations that capture Bitcoin's more contemporary market behavior, trends, and volatility. This decision helps ensure that the analysis and models are based on up-to-date information that better aligns with today's trading landscape..

2. Converting Columns to Numeric

Converting columns to numeric data types is essential for quantitative analysis and modeling. Numeric data types allow us to perform mathematical operations, calculate statistical measures, train machine learning models, and enable us to extract meaningful insights, and develop predictive trading strategies. Moreover, it facilitates more robust and accurate analyses of Bitcoin market data.

3. Removing Duplicates

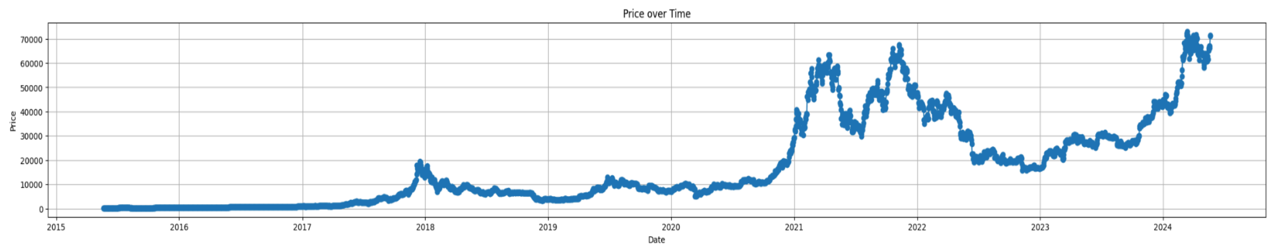
Identifying and removing duplicate rows in the dataset helps maintain data integrity and accuracy in the analysis. Duplicate rows can skew statistical measures, introduce biases in modeling results, and lead to misleading conclusions. The dataset was checked for any duplicate observations to ensure that each data point is unique and contributes only once to our analysis. Eliminating duplicates can prevent redundancy and ensure that our analysis is based on a clean and representative data set by enhancing the reliability and validity of our findings.

4. Cleaning Data Format

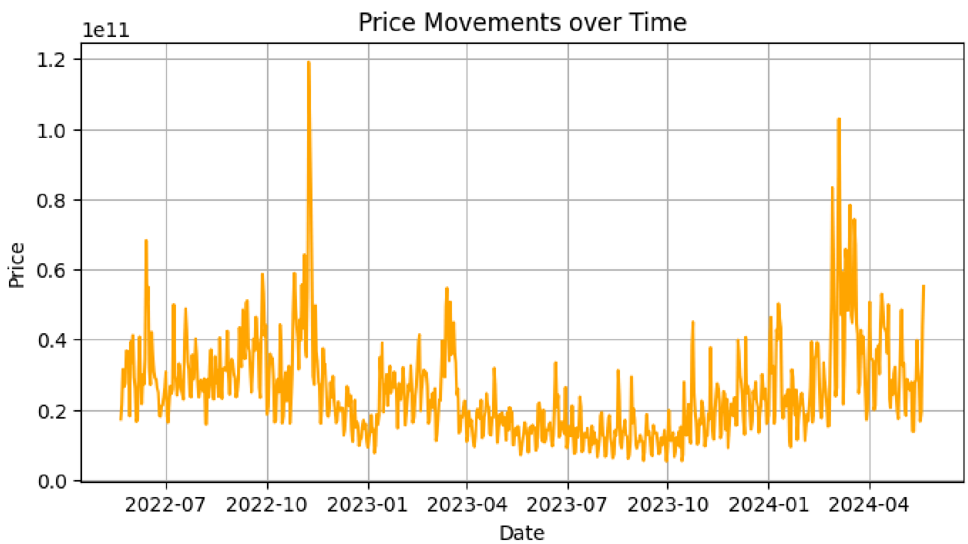
Converted the Date column to Datetime format. The Date column in the dataset was initially in a string format. It is not suitable for time-series analysis and plotting. It was necessary to convert this column to a date-time format to facilitate various date-related operations. for example, filtering and plotting.

Data Understanding

Bitcoin price movement

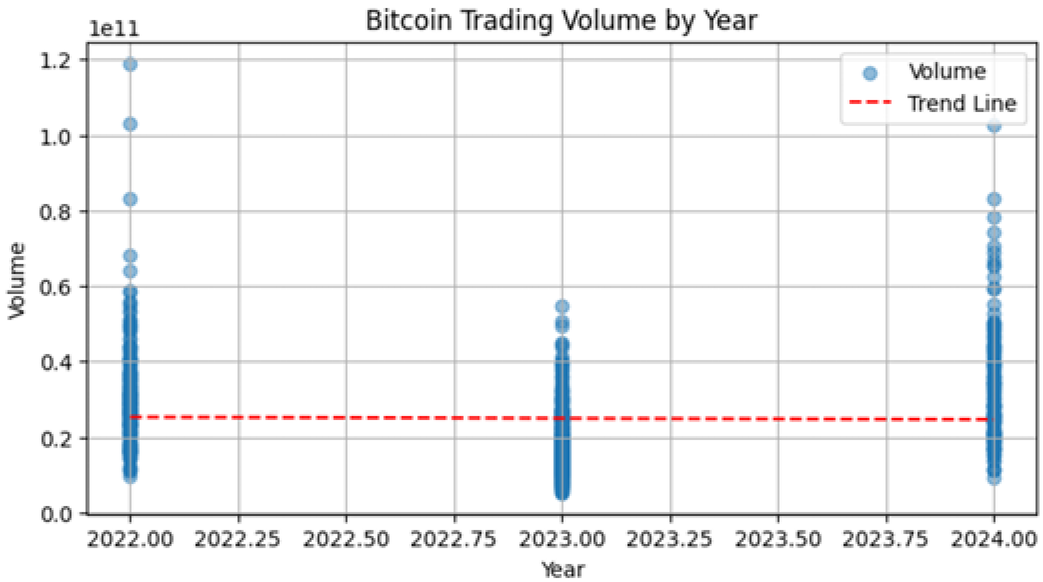


The line graphs show the value of Bitcoin from 2015 to 2024 captures the significant fluctuations in the cryptocurrency's trajectory. At first, Bitcoin's value was low, and it stayed pretty steady between 2012 and about 2017. But in 2017, there was a turning point when the value of Bitcoin increased significantly, indicating the start of its explosive rise. The graph then displays a sequence of highs and lows that illustrate the erratic nature of Bitcoin's value and its intrinsic volatility. The graph is noteworthy because it shows notable peaks, especially in late 2017 to early 2018 and late 2020 to early 2021, when Bitcoin peaked. These high points correspond to times when the cryptocurrency's general growth trajectory experienced significant appreciation. Nevertheless, the graph also highlights how volatile Bitcoin can be, showing downturns in addition to spikes in value. All things considered, the graph bears witness to the dynamic value evolution of Bitcoin, which has been shaped by a wide range of factors such as investor behavior, market mood, legislative changes, and technological improvements.



The "Price Movements over Time" graph shows how the price of Bitcoin changed between July 2022 and April 2024. During this period, there has been a notable fluctuation in the value of Bitcoin, with regular unexpected spikes and losses. Though the regular fluctuations highlight the cryptocurrency's volatility.

Bitcoin Volume Trading By Year

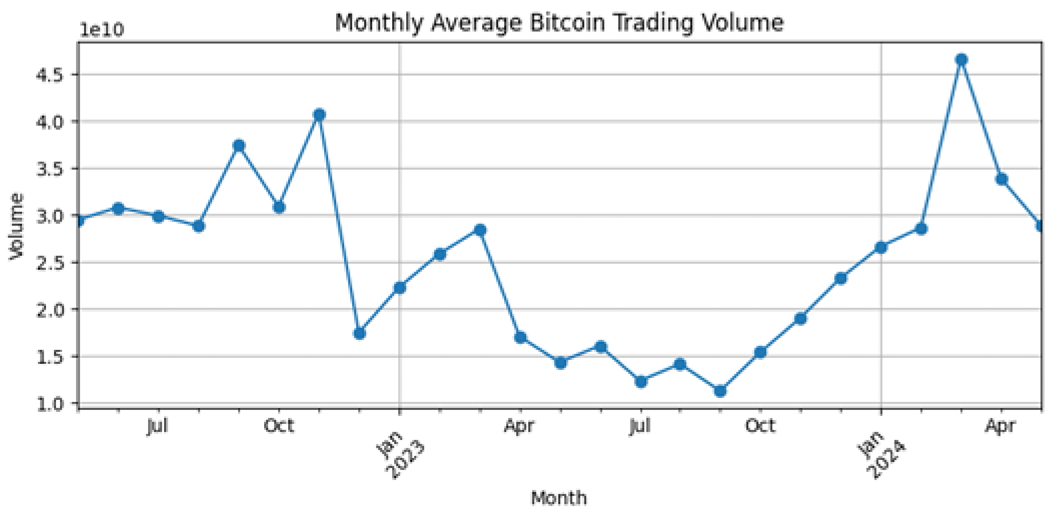


The graph illustrates the volume of Bitcoin trading from 2022 to 2024. Each blue dot signifies the trading volume for a specific date within the dataset. The red dashed line represents the trend line, which shows the overall trend in trading volume over the given period.

The trading volume shows a wide range of values. Most trading volumes are clustered below 0.4 on the y-axis with some extreme values reaching up to around 1.2. The red dashed trend line is almost flat and indicates a very slight increase in trading volume over the years. This suggests that despite the volatility and high trading volumes on certain days the overall average trading volume has remained largely constant with a slight upward trend.

There are significant increases in trading volume in 2022 and 2024. These rises could indicate periods of high trading activity. It could be due to major market events or news that significantly impacted Bitcoin's price or investor sentiment. There are several outliers with extremely high trading volumes in 2022. The majority of the trading volumes in 2022 are clustered below the 0.2 mark. It indicates that the average trading volume was relatively low, except for the few outliers.

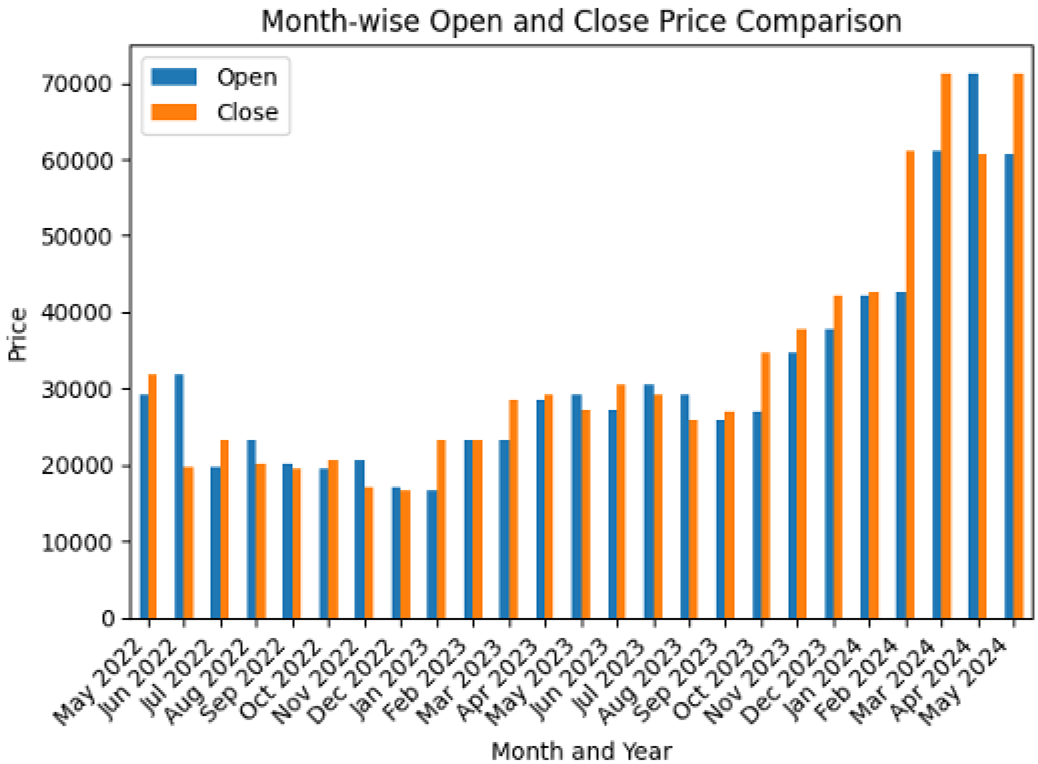
The year 2023 shows a more consistent pattern with fewer extreme outliers compared to 2022. The trading volume in 2024 starts with lower values but shows a significant increase towards the end of the graph. The general trend suggests that average trading volumes are increasing slightly year over year.



The chart displays the average Bitcoin trading volume by month in 2023 and 2024. The trend shows a fluctuation over time. The Peak is in March 2024. That peak could be due to specific market events, news, or regulatory changes that caused increased trading activity during this month.

Following the peak in March, there is a sharp decline in trading volume. And September 2024 shows the lowest average bitcoin trading volume. And there is a sharp increase after September, and it comes to a peak in March.

Month-wise Open and Close Price comparison



The chart shows that both open and closed prices follow a similar trend throughout the year and indicate a consistent relationship between the opening and closing prices of Bitcoin.

March and May 2024 exhibit the highest average prices for both open and closed values. The lowest prices are observed in December 2022 for both open and close prices.

The differences between the open and close prices for each month are relatively small. This suggests that on average Bitcoin prices do not change significantly within a single day. However, the slight variations that do exist can still be significant for day traders.

Overall, with peaks and troughs all year long, the average prices of Bitcoin generally follow cyclical patterns. Market mood, regulatory updates, macroeconomic variables, and technical advancements all impact these fluctuations.

Feature Engineering

Feature engineering is a crucial step in preparing data for predictive modeling. It involves creating new features or transforming existing ones to enhance the model's predictive power. For the Bitcoin price prediction project, we applied several techniques to our dataset. These techniques include adding time-based indicators, calculating technical indicators, and creating lag features. Each step aims to capture different patterns in the data that might affect Bitcoin prices.

Adding Day of the Week and Month Columns

As we found in the EDA sections Months and day of weeks have identical patterns with the bitcoin price. So in order to understand the cyclical patterns in Bitcoin prices, we added two new columns: 'Day_of_Week' and 'Month'. The 'Day_of_Week' column tells us the day of the week, with values from 0 (Monday) to 6 (Sunday). The 'Month' column indicates the month of the year, with values from 1 (January) to 12 (December). These features help us capture weekly and monthly trends, which are important because Bitcoin prices often show recurring patterns based on the day of the week or the time of the year as we found in the EDA section.

Adding Exponential Moving Average (EMA)

We calculated the 7-day exponential moving average (EMA) for the 'Price' and added it as a new feature. The EMA gives more weight to recent prices, making it more responsive to new information than the simple moving average. Including the 7-day EMA helps the model detect short-term trends and smooth out price fluctuations, potentially improving its ability to predict future prices.

Calculating Ichimoku Cloud Lines

The Ichimoku Cloud is a very powerful indicator for trading. Most of the investors believe that it can capture most of the hidden information inside the prices. Such as support and resistance levels, identify trend direction, and gauge momentum. We calculated the following components:

- Conversion Line (Tenkan-sen): The average of the 9-period high and low.
- Base Line (Kijun-sen): The average of the 26-period high and low.
- Leading Span A (Senkou Span A): The average of the Conversion Line and Base Line, shifted 26 periods ahead.

- Leading Span B (Senkou Span B): The average of the 52-period high and low, shifted 26 periods ahead.
- Lagging Span (Chikou Span): The current closing price shifted 26 periods back.

These components give a comprehensive view of the price action and help identify potential trend reversals, support, and resistance levels.

Creating Lag Features

We created lag features for all columns except date . For each of these columns, we generated lag features from 1 to 7 days. Lag features capture the historical dependencies in the data. By including past values, the model can learn from the historical context, which can improve its predictive accuracy.

Creating Target Variable

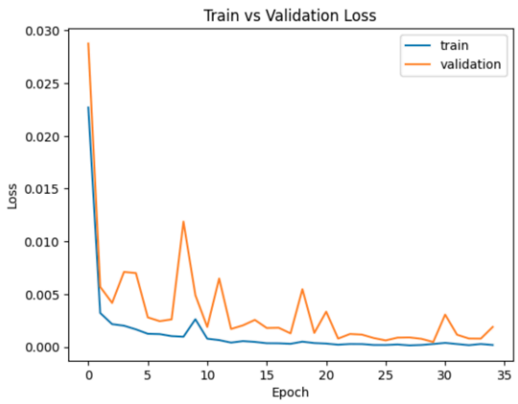
Instead of depending on Closing price as target variable, we created the target variable 'Tomorrow_Price' by shifting the 'Price' column one period ahead. This allows the model to learn the relationship between the current features and the price on the following day. By predicting the next day's price, we can provide actionable insights for trading decisions.

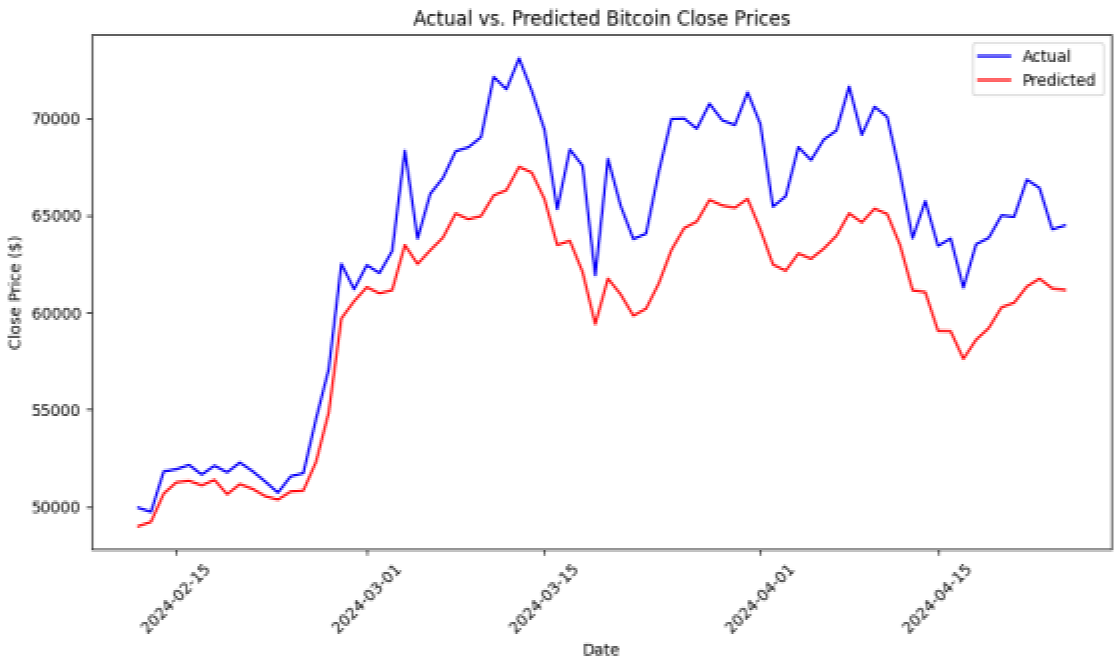
Model Development

In our quest to develop accurate predictive models for Bitcoin close prices, we explored three distinct architectures harnessing the power of deep learning. Each architecture was meticulously crafted to address specific challenges in forecasting, with the aim of improving prediction accuracy and robustness.

Model 01 - LSTM with Simple Architecture:

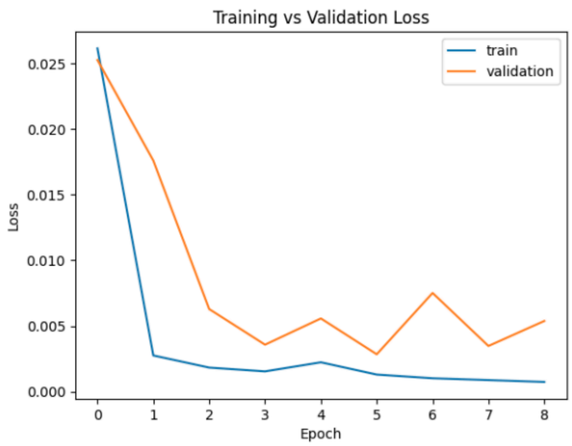
Our journey began with a simple yet effective LSTM architecture. This model served as our baseline, providing a solid foundation for comparison with more sophisticated designs. By incorporating dropout layers to combat overfitting and utilizing mean squared error as the loss function, we trained the model to discern patterns and relationships in the data. Through iterative refinement and validation, we achieved promising predictive performance.

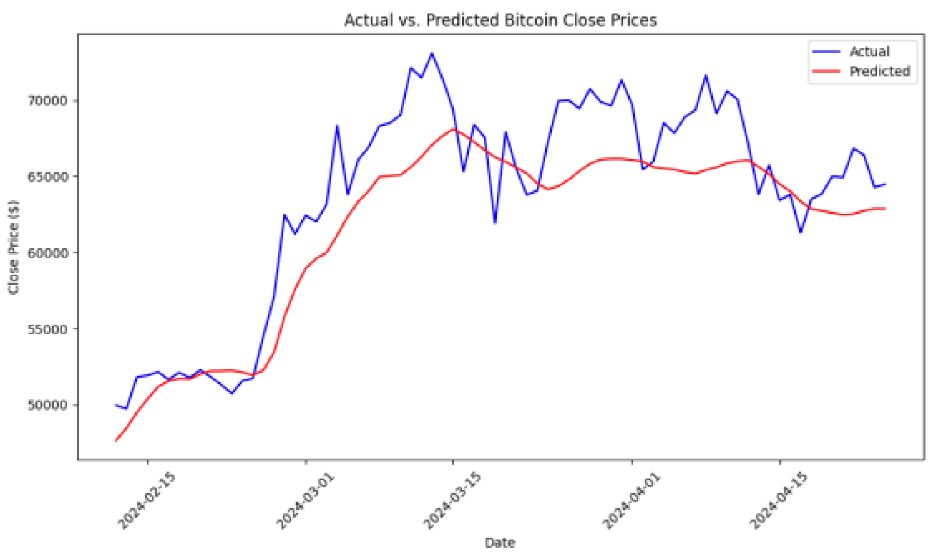




Model 02 - LSTM with Bidirectional Layers:

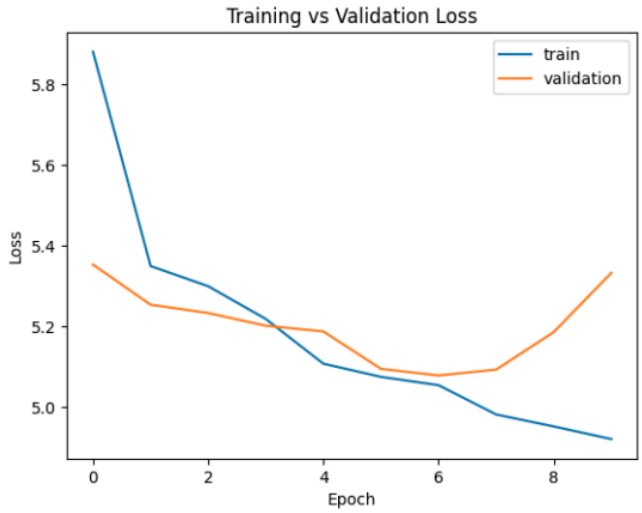
Building upon the success of our baseline model, we introduced bidirectional LSTM layers to capture temporal dependencies more comprehensively. This architectural enhancement allowed the model to consider both past and future information when making predictions, thereby enriching its understanding of the underlying data dynamics. Additionally, we implemented a dynamic learning rate schedule to optimize the training process, further enhancing the model's adaptability and performance.

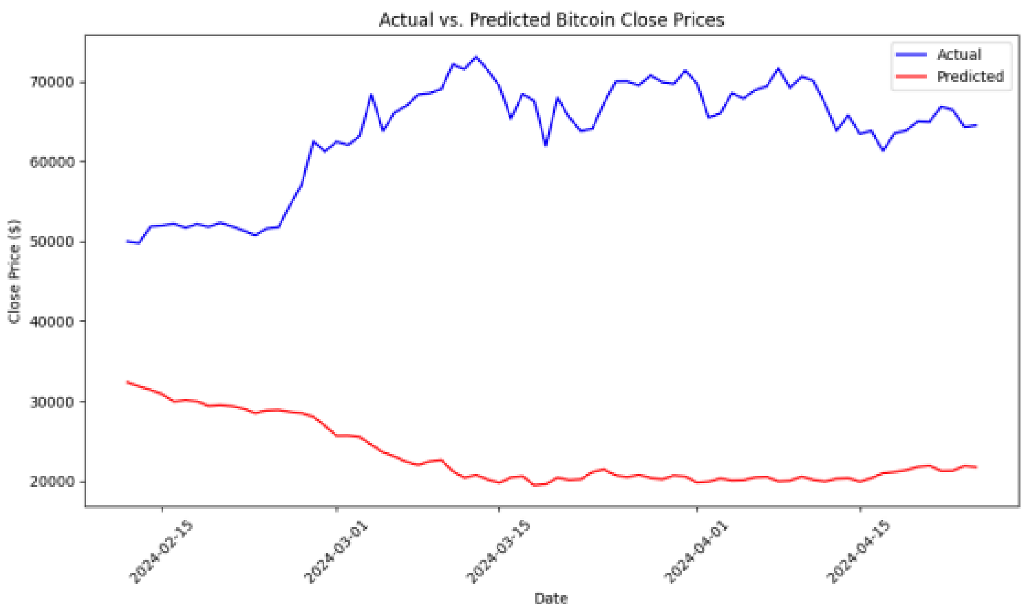




Model 03 - LSTM with Batch Normalization and L2 Regularization:

Recognizing the significance of stabilizing training dynamics and reducing overfitting, our third model integrated batch normalization and L2 regularization. These techniques aimed to mitigate internal covariate shift and penalize large parameter values, promoting smoother convergence and improved generalization. By addressing these common challenges in deep learning, we sought to elevate the model's predictive capabilities and resilience to unseen data.





In our exploration of model architectures, we discovered that simplicity often yields superior results. Despite the allure of complexity, it was the humble Model 01 that emerged as the frontrunner, showcasing remarkable predictive accuracy. This finding underscores the timeless principle of Occam's razor in machine learning – that simpler models are often more effective. Moving forward, we will continue to embrace simplicity while judiciously incorporating advanced techniques to push the boundaries of predictive modeling.

Model Selection

To identify the best-performing model, we evaluated three distinct LSTM models: LSTM Model 1, LSTM Model 2, and LSTM Model 3. The evaluation was based on several critical metrics including Mean Squared Error (MSE), Mean Absolute Error (MAE), Sharpe Ratio, Total Return, Standard Deviation of Returns, and Trading Return.

Upon analyzing these metrics, LSTM Model 2 emerged as the superior model. It demonstrated the lowest MSE and MAE, which indicates higher accuracy in predicting Bitcoin prices. Additionally, LSTM Model 2 achieved the highest Sharpe Ratio (0.3797), signifying a better risk-adjusted return. This model also posted a significant Total Return (0.3194) and had the lowest Standard Deviation of Returns (0.0102), reflecting its stability and lower volatility.

Furthermore, LSTM Model 2 showed a positive Trading Return (0.4049), indicating its profitability when applied to a simulated trading strategy. In contrast, LSTM Model 3, despite its complexity, performed poorly across all metrics, particularly in terms of profitability and stability.

Model Evaluation

To validate the performance of LSTM Model 2, we conducted an evaluation on a separate test dataset, which had not been seen by the model during training. This step was crucial to assess the model's generalization ability and its robustness in real-world scenarios.

LSTM Model 2 continued to perform well on the test dataset, maintaining profitability and stability. The trading strategy implemented based on LSTM Model 2's predictions yielded positive returns, confirming the model's capability to generate profits. The consistent performance in terms of Sharpe Ratio and lower volatility of returns indicated that LSTM Model 2 could effectively manage risk while achieving favorable returns.

Additionally, the analysis of variability in profitability suggested that LSTM Model 2 could adapt to different market conditions without significant degradation in performance. This robustness is essential for a trading model intended for dynamic and often volatile cryptocurrency markets.

Conclusion

In order to anticipate Bitcoin closing prices, we created and evaluated three LSTM models in this study. We paid particular attention to each model's accuracy, profitability, stability, and robustness. The LSTM Model 2 was found to be the best model after a thorough review using a variety of criteria. It showed the lowest MSE and MAE, the highest Sharpe Ratio, large total and trading returns, and the least amount of volatility. The model's resilience and generalization were validated by its constant performance on an independent test dataset, which makes it a trustworthy tool for maximizing trading strategies in the erratic bitcoin market. The results highlight the promise of sophisticated LSTM architectures for financial forecasting, providing insightful information to traders and analysts who want to improve their decision-making and attain profitable trading results.