Bitcoin Price Forecast with Time Series

Zeyu Chen
https://github.com/zeyuleochen/Time-Series-Project.git

Problem Statement

Introduction:

- The increasing popularity and widespread adoption of cryptocurrencies, particularly Bitcoin, have generated significant interest in understanding and predicting their price movements. Accurate forecasting of Bitcoin prices can assist investors, traders, and financial institutions in making informed decisions, mitigating risks, and optimizing their investment strategies. Time series analysis provides a powerful framework to analyze historical price patterns, identify trends, and develop forecasting models.

Problem Description:

- The objective of this project is to develop a robust time series forecasting model that can predict future Bitcoin prices. The model should utilize historical price data and relevant factors affecting Bitcoin prices to generate reliable predictions.

Data Source and Content

1. The dataset has one csv file for each cryptocurrency, the main focus of this project is Bitcoin.
2. Price history is available on a daily basis from April 28, 2013 to July 6, 2021.
    - Use the data until June 26, 2021 as training and the remaining 10 entries as test set.
3. The structure of data source:
    - Date : date of observation
    - Open : Opening price on the given day
    - High : Highest price on the given day
    - Low : Lowest price on the given day
    - Close : Closing price on the given day
    - Volume : Volume of transactions on the given day
    - Market Cap : Market capitalization in USD

Source: https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory?select=coin_Bitcoin.csv

Data Assumptions

1. **Data completeness**: It is assumed that the historical price data used for analysis is complete and does not contain significant gaps or missing values. Inaccurate or missing data can impact the quality of the forecasting models and their ability to capture the underlying patterns in the Bitcoin price series.
2. **Data consistency**: It is assumed that the Bitcoin price data used for analysis is consistent across different sources and exchanges. Inconsistencies or discrepancies in the price data from various sources can introduce biases and inaccuracies in the forecasting models.
3. **Efficient market hypothesis**: It is assumed that the Bitcoin market is efficient, meaning that historical price data contains all relevant information needed for forecasting, and no additional external factors or events are required to be considered explicitly.
4. **Non-stationarity**: The assumption of non-stationarity in Bitcoin price data is based on the recognition of a presumed long-term increasing trend observed from 2013 onwards.

Data Properties

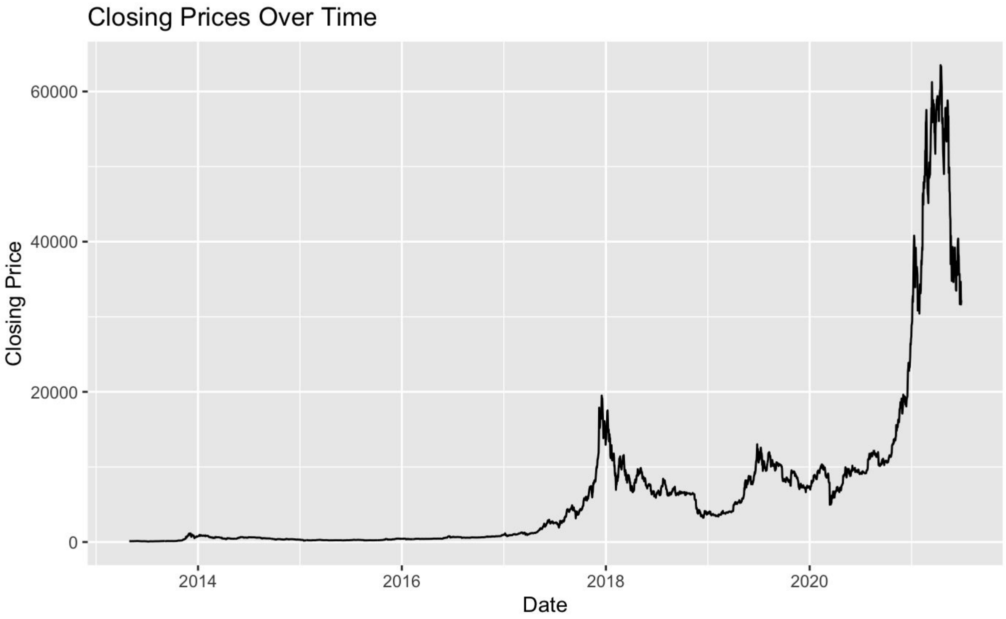
For the project, I chose to model the Daily Close Price of Bitcoin due to its consistency in terms of time. The time difference between each data point is uniform.

1. **Volatility**: Bitcoin close prices tend to exhibit high volatility, with significant price fluctuations occurring over short periods of time.
2. **Trend**: Bitcoin prices have demonstrated a notable long-term upward trajectory, characterized by a nonlinear pattern. Particularly, since 2020, there has been a significant acceleration in the rate of increase, indicating a more rapid upward trend in Bitcoin prices. However, this upward trend was interrupted by a significant decrease in prices since 2021, reflecting a sharp reversal in the previously established trend.
3. **Autocorrelation** is observed in Bitcoin price data, indicating a dependence between current and past price observations. This suggests that past price movements can influence future price movements, highlighting the presence of temporal patterns and potential predictive power in the data.
4. **Correlation with Trade Volume**: Market volume is often positively correlated with Bitcoin price movements, indicating higher trading activity during price increases or decreases.

Data Properties: Volatility

The volatility of Bitcoin refers to the extent of price fluctuations and uncertainty Bitcoin is known for its high volatility, characterized by significant price swings over short periods.

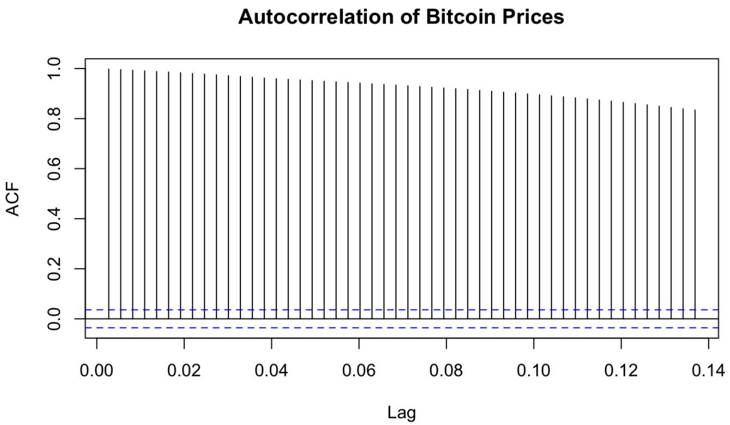
This volatility can be attributed to various factors such as market demand, investor sentiment.

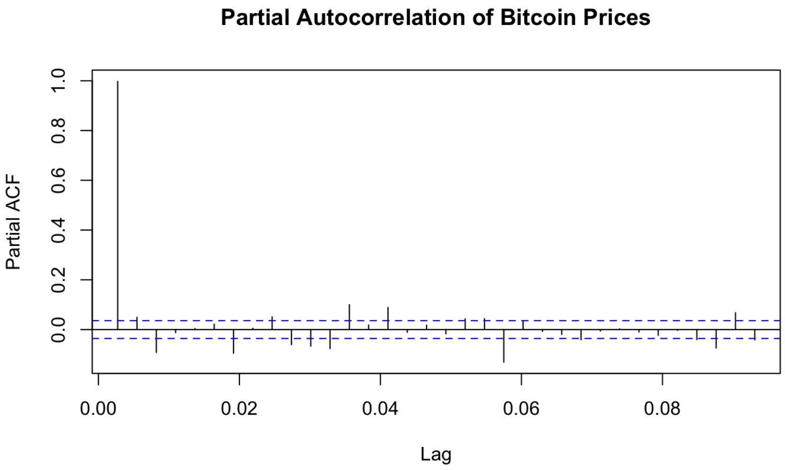


Data Properties: Autocorrelations

Autocorrelation of the time series suggests that the current values of Bitcoin prices are related to their past values. This implies that the price of Bitcoin today is influenced by its previous prices.

Past values can influence the price of Bitcoin due to the presence of market memory and behavioral biases. Traders and investors often consider historical price patterns, trends, and support/resistance levels when making decisions, leading to a feedback loop where past values impact the current price.



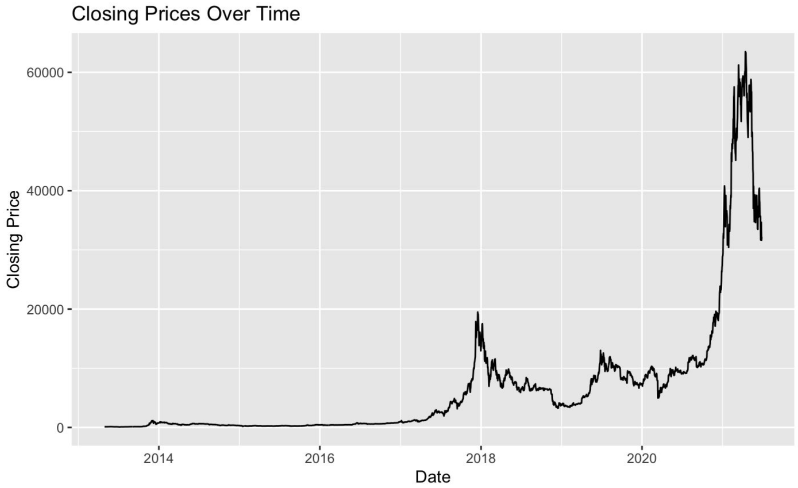


Data Properties: Stationality

The presence of this trend component indicates that the data does not exhibit stationarity characteristics.

- ADF Test: The p-value of 0.7165 obtained from the ADF test indicates that we fail to reject the null hypothesis of non-stationarity. This suggests the presence of a trend or other non-stationary components.
- KPSS Test for Level Stationarity: The p-value of 0.01 from the KPSS test further confirms the rejection of the null hypothesis of trend stationarity. This provides additional evidence supporting the existence of a trend.

Therefore, based on the results of both tests, it can be concluded that the data is likely non-stationary, with a discernible trend.

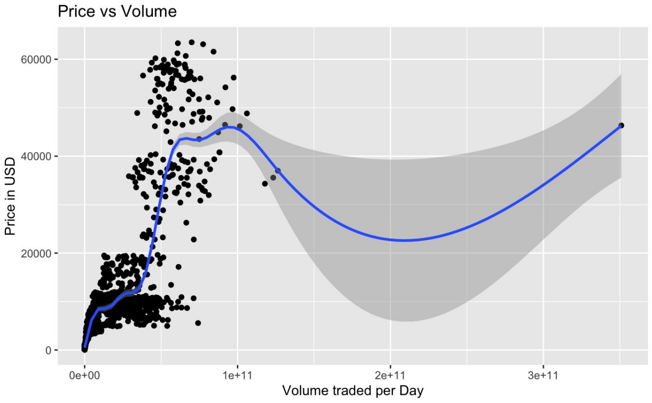


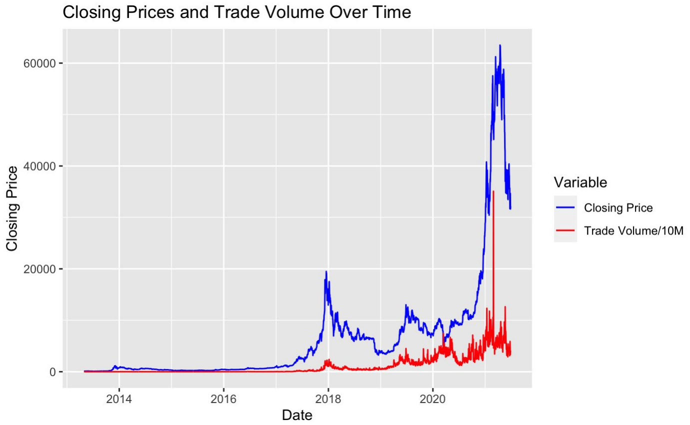
Test	Dickey_Fuller	Lag_order	KPSS_Level	Truncation_lag	p_value
ADF Test	-1.6737	14	NA	NA	0.7165
KPSS Test	NA	NA	13.363	9	0.0100

Data Properties: Correlations with Trade Volume

The relationship between trade volume and price in Bitcoin suggests that higher trade volume often coincides with significant price movements. Increased trade volume indicates higher market participation and liquidity.

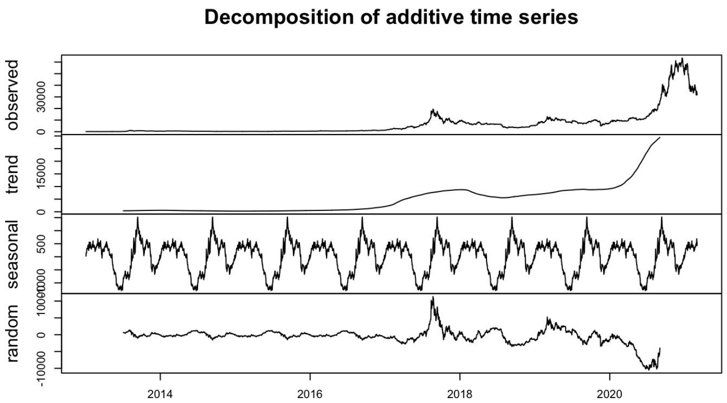
The trade volume of Bitcoin peaks at the highest point of price, reflecting increased market activity and liquidity during periods of significant price appreciation.

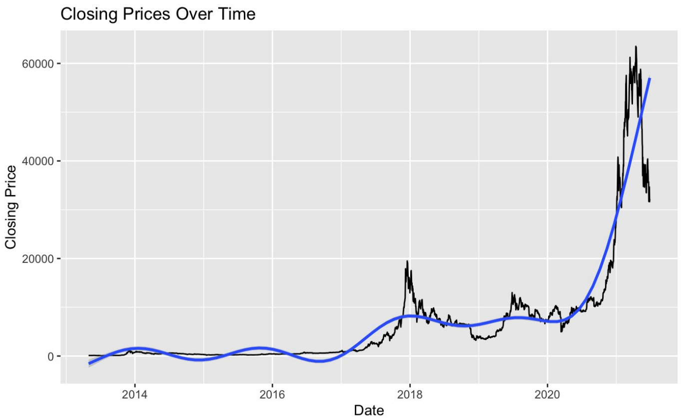




Data Processing: Trend

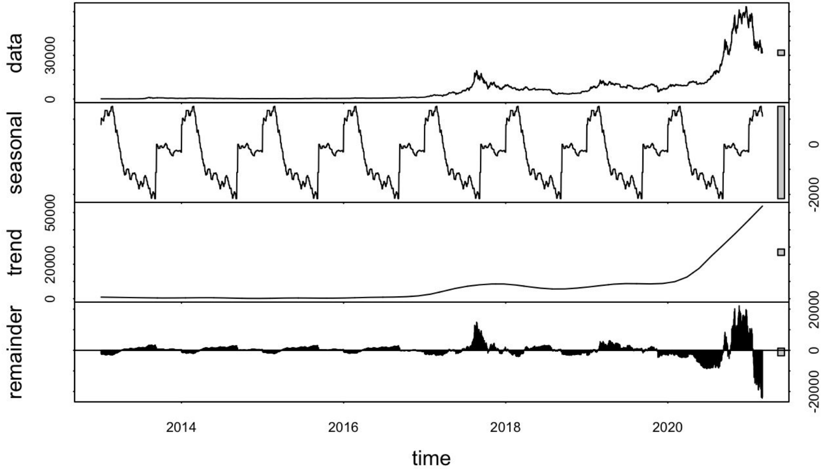
- Detecting Trend in Bitcoin Price
  - Observe the historical Bitcoin price data and identify the presence of a general nonlinear increasing trend which accelerates after 2020.
- Decomposing trend
  - Trend: nonlinear upwards
  - Residuals: Remaining irregular pattern before 2018 but trend starts to resurface after 2018.
  - Decomposition fails to catch the sudden drop from 2021.





Data Processing: Seasonality

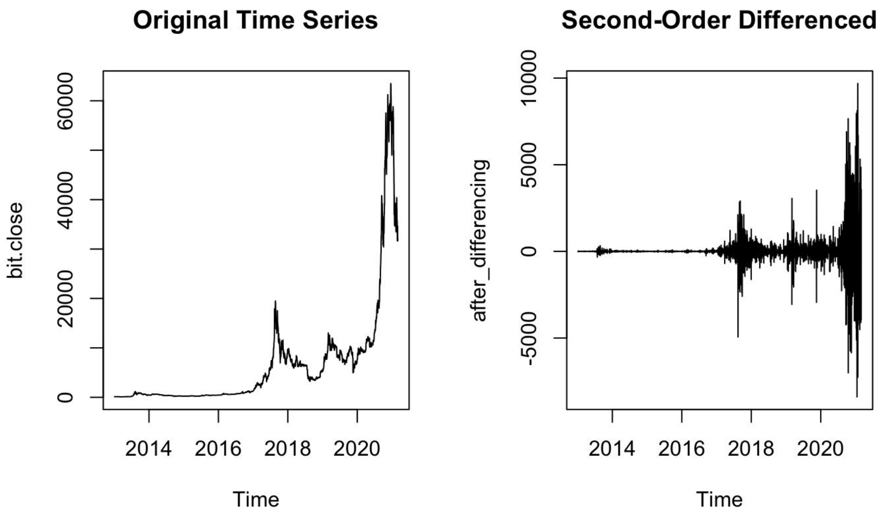
- Detecting Seasonality in Bitcoin Price
    - Observe the historical Bitcoin price data and identify 7.5 seasonality cycles displayed with the time window between 2013 and 2021.
    - Bitcoin price data demonstrates a yearly cycle.
- Decomposing Seasonality
    - STL (Seasonal and Trend decomposition using Loess) fails to capture the sudden drop of price in 2021.
    - Residuals of the decomposition still display patterns after 2018



Data Processing: Transformation

To address the presence of both trend and seasonality in the time series data, we perform second-order differencing. This process involves differencing the data twice to eliminate the effects of these components. The resulting differenced series appears more stationary, indicating the removal of long-term trends and seasonal patterns.

However, it is important to note that despite the differencing, residual patterns or irregularities may still persist, suggesting the presence of other underlying dynamics in the data.



Proposed Model: Bayesian Structural Time Series (BSTS)

- The proposed model aims to utilize the Bayesian Structural Time Series (BSTS) framework for accurate and dynamic modeling of Bitcoin price data.
- BSTS offers a powerful statistical approach that can capture the complex patterns, trends, and seasonality inherent in Bitcoin price movements.
- BSTS is more sensitive to sudden changes in the time series data, that provides insights into the sudden drop in bitcoin prices we observed since 2021 that STL decomposition fails to detect.
- BSTS can account for the residual variation we observed in the data that weren’t fully decomposed in STL.

Trade-offs

- Poor choices of hyperparameters such as the number of MCMC samples we used to inference the parameters can cause Overfitting.
- Modeling Bitcoin prices using BSTS can be computationally demanding, especially when dealing with a large number of observations or high-frequency data.

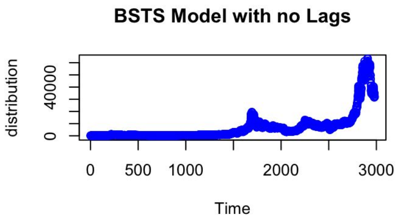
Feature Engineering

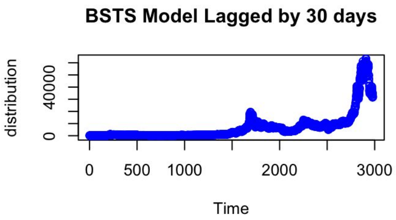
1. Lagged Close Price by 30 days as Predictor:
    - Lagged price is incorporated as a predictor variable in the BSTS model to capture temporal dependencies and autocorrelation. Specifically, we include the lagged value of the target variable (Bitcoin price) with a lag of 30 days to enable the model to forecast based on past values.
2. Exclusion of Highly Correlated Variables:
    - Other price variables (open, high, low) are excluded from the model due to their high correlation with the target variable.
3. AddSemilocalLinearTrend():
    - The AddSemilocalLinearTrend component is introduced to account for the overall trend in the Bitcoin price data.
    - This feature engineering technique helps capture the long-term trend and allows the model to adapt and adjust the trend component over time.
4. AddSeasonal():
    - To account for seasonal patterns in the Bitcoin price data, we incorporate the AddSeasonal() function.
    - This feature engineering step helps capture recurring patterns and seasonality, which can significantly influence the price fluctuations.

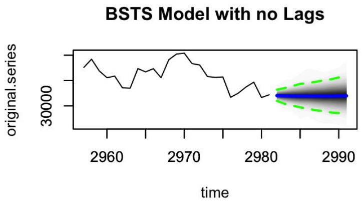
Results and Learning

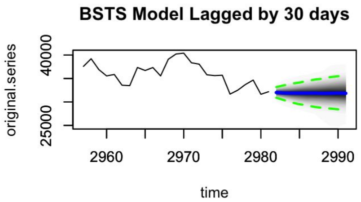
We compared three different BSTS models: one with a 30 days lagged target variable, another without any lagged variables, and a third model incorporating all other price variables.

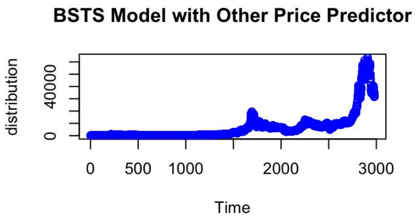
The results showed that all models captured the trend of the bitcoin data with high level of similarity. However, the model incorporates other price variables may present overfitting from the ten days forecast.

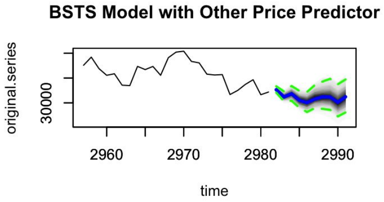








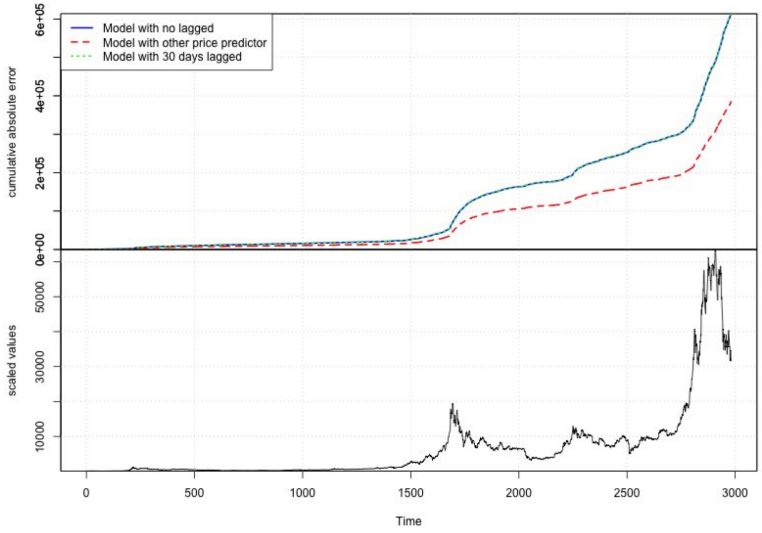




Results and Learning

Based on the Mean Absolute Percentage Error (MAPE), which measures the relative accuracy of the forecasts, the simple model without the lagged variable has a MAPE of 7.382076, the model with 30 days lagged target variable has a MAPE of 7.60796. The model that incorporates all prices yields the lowest cumulative absolute error while the MAPE is the highest, suggests a high likelihood of overfitting.

Lower MAPE values indicate better accuracy, and in this case, the simple BSTS model with only historical data of Bitcoin Close Price without the lagged variable is the best model that provides the most accurate forecasts.



Model	ME	RMSE	MAE	MPE	MAPE
Lagged Model	2640.151	2726.827	2640.151	7.607960	7.607960
No Lagged Model	2562.328	2653.427	2562.328	7.382076	7.382076
Model with Other Price	3417.034	3481.587	3417.034	9.880903	9.880903

Limitations & Future Works

Limitations:

- **Limited trials of hyperparameters** restrict the exploration of the full hyperparameter space, potentially leading to suboptimal model configurations and less accurate forecasts.
- **Better data segmentation** should be considered, since the bitcoin prices experienced violent increase and fluctuation after 2020, the inclusion of data before that can provide inaccuracies.
- **Focuses solely on the closing price** and does not consider other price indicators such as the open, high, and low prices. Ignoring these additional price variables may result in overlooking daily patterns.

Future Works:

- **Time-related Features**: Creating features that capture specific time-related patterns can be useful. For example, including binary indicators for weekends or holidays, or creating features that capture the time of day (e.g., morning, afternoon, evening) can account for temporal variations in Bitcoin prices.
- **Fourier Terms**: Introducing Fourier terms to model periodic patterns and seasonality in the data can be beneficial. Fourier terms allow capturing cyclical components with different frequencies, such as daily, weekly, or yearly patterns.
- **Advanced Model consideration**: Prophet for Bitcoin price forecasting can be advantageous due to its ability to handle multiple sources of uncertainty, incorporate seasonality, and capture non-linear trends and changes in volatility.