Institute of Technology of Cambodia

Department of AMS

**I3-AMS-B**

**Bitcoin Price Prediction**TP: C (Group 4)

**Name ID**

1. YONG Lyhor e20221004
2. HEANG Denis e20220527
3. SEING Ratana e20221147
4. CHHEANG Udomveasna e20220918
5. HEANG Piseth e20220690

**Lecturer:** PEN Chentra (TP)

PHAUK Sokkhey (Course)

Academic year 2024-2025

**Contents**

Table of Contents

[I. Introduction 3](#_Toc188722529)

[1. Overview of The Data 3](#_Toc188722530)

[II. Exploratory data analysis 3](#_Toc188722531)

[1. Data Cleaning 3](#_Toc188722532)

[2. Statistical Analysis 6](#_Toc188722533)

[3. Visualizing the Data 6](#_Toc188722534)

[III. Machine Learning 7](#_Toc188722535)

[1. Train, Test, Split and Fit the Data With Models 8](#_Toc188722536)

[2. Model Evaluation 9](#_Toc188722537)

[3. Prediction 12](#_Toc188722538)

[3.1 Prediction using trained models 12](#_Toc188722539)

[3.2 Prediction using forecast 14](#_Toc188722540)

[IV. Conclusion 15](#_Toc188722541)

[Reference 16](#_Toc188722542)

# Introduction

The price prediction of Bitcoin using machine learning is a growing area of interest in the field of financial technology. This application leverages the power of machine learning algorithms to analyze historical data and market trends, enabling predictions about future Bitcoin price movements. Here's a concise introduction:

## Overview of The Data

This dataset captures the daily USD valuation of Bitcoin over a transformative decade from 2014 to 2024. As the first cryptocurrency, Bitcoin has redefined the financial landscape, showcased dramatic price volatility and sparked global interest.

The dataset contains over 3,650 daily entries, recording key metrics such as opening, closing, highest, and lowest prices in USD. Additionally, it includes daily trading volumes and market capitalization, offering a comprehensive view of Bitcoin’s performance.

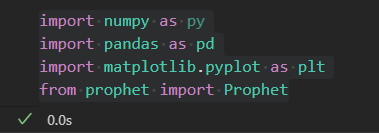
Beyond raw data, this dataset provides valuable context by highlighting significant events, including major regulatory developments, adoption milestones, and Bitcoin halving’s. These factors help explain the trends and shifts in Bitcoin’s market behavior over time.

Ideal for researchers, investors, and students, this dataset serves as a powerful tool for exploring Bitcoin’s price patterns, market dynamics, and its influence on modern finance.

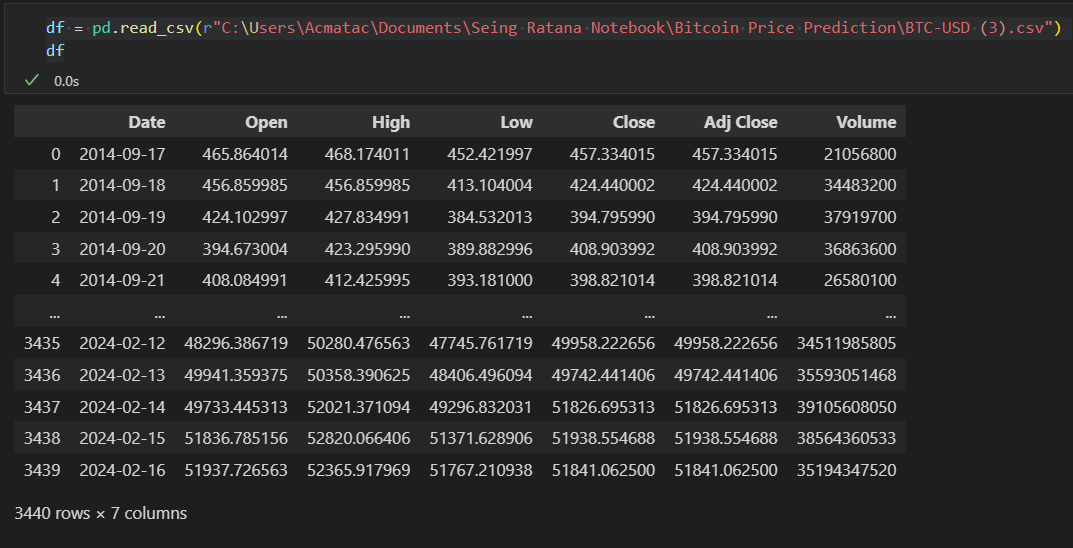
# Exploratory data analysis

## Data Cleaning

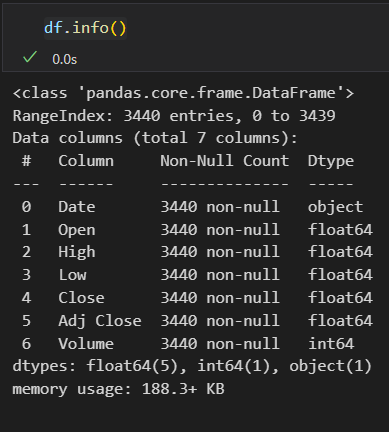
We have checked for duplicated value and Missing value in our dataset. As a result, we did not spot any missing and duplicated value in our dataset.  
first we import all important library:

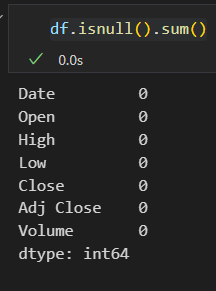


After importing the necessary library, we can read and clean the data.



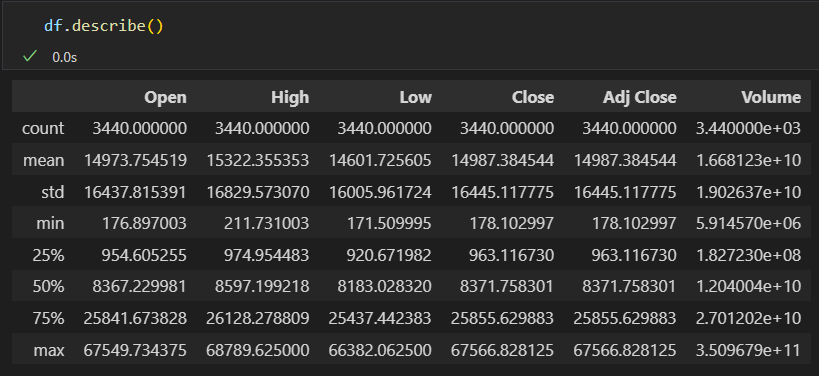
We do a quick overview of the data using the method below:



Then we use the method below to detect the missing values. This will give us the amount of missing value in each column.

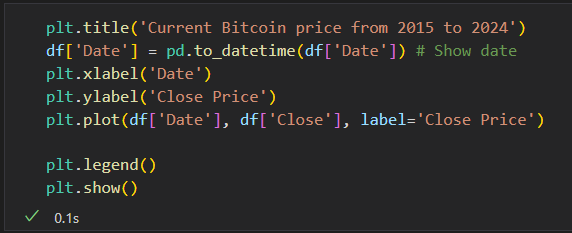
## Statistical Analysis

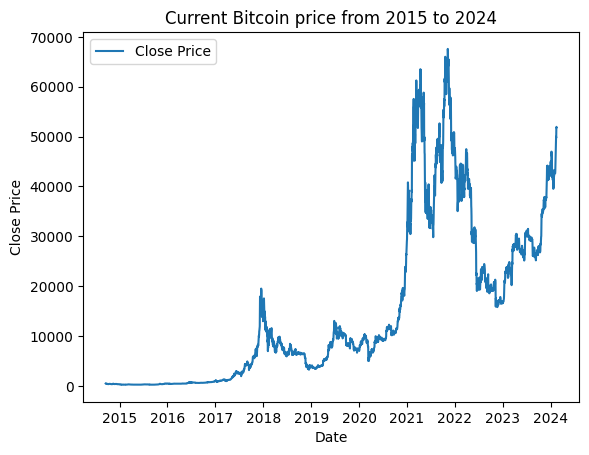
We can find the count, mean, standard deviation, min, Q1, Q2, Q3, max as follow:



## Visualizing the Data

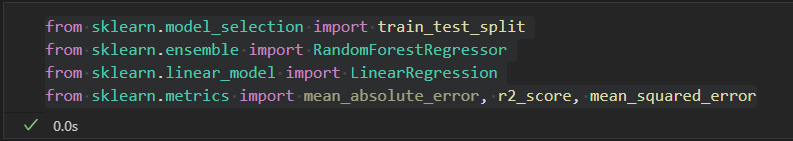
We plot the graph between Date and Close Price.



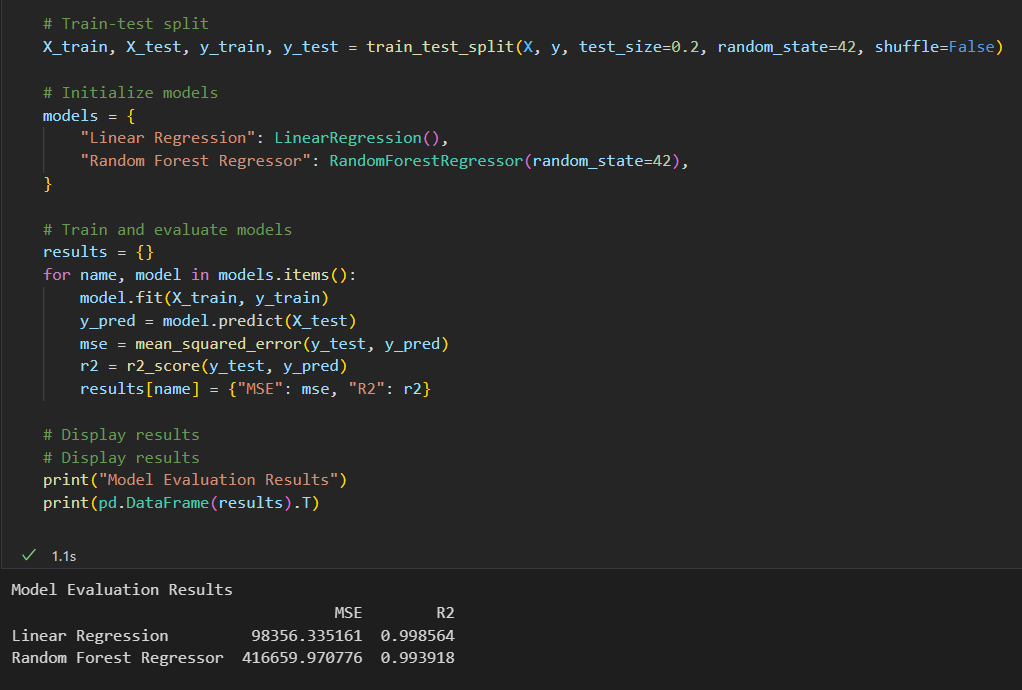


# Machine Learning

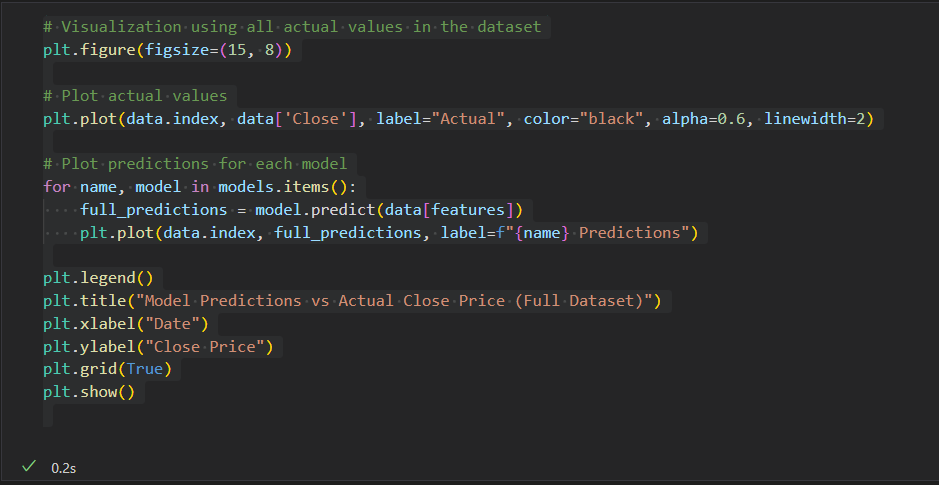
Import all necessary libraries.

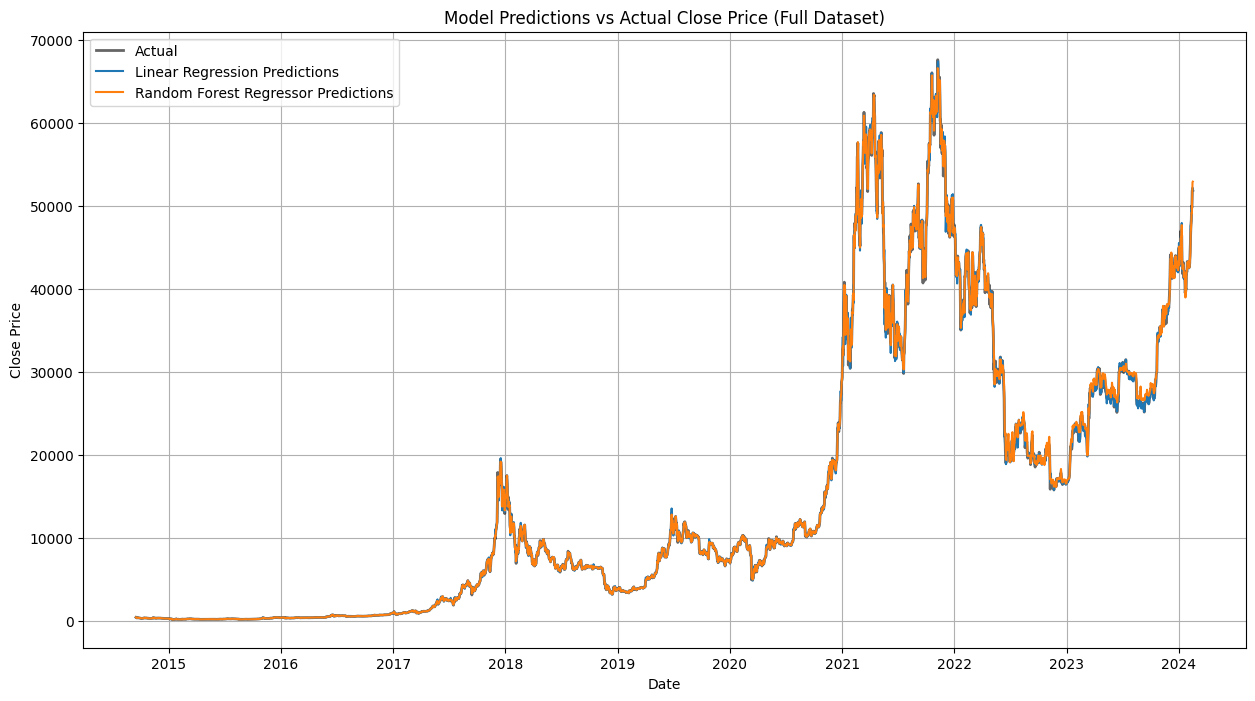


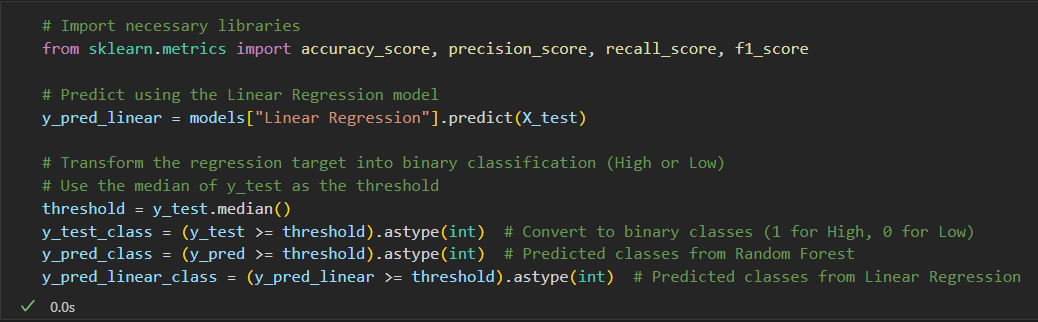
## Train, Test, Split and Fit the Data with Models



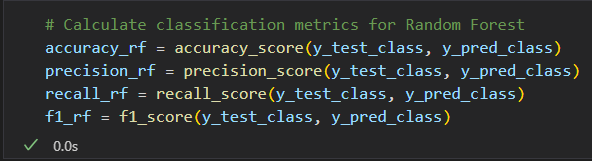
## Model Evaluation

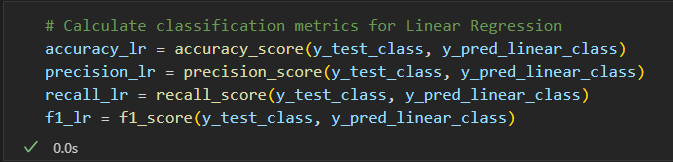


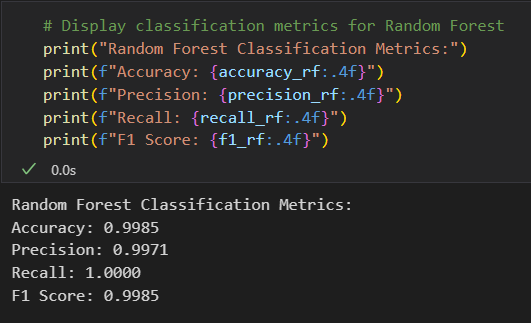


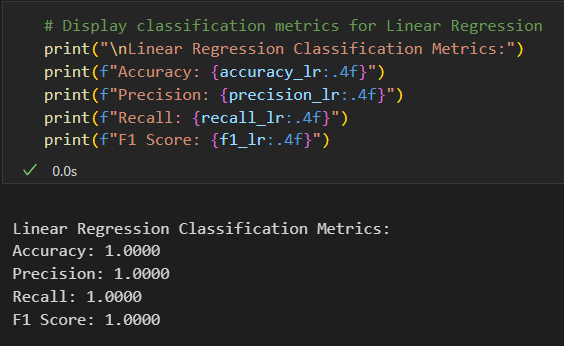
By looking at the graphs, for the Random Forest model, we can see the actual values and the predicted value are a little bit deviated away from each other at the end. For Linear Regression model, we can see the actual value and the predicted values are perfectly fitted.

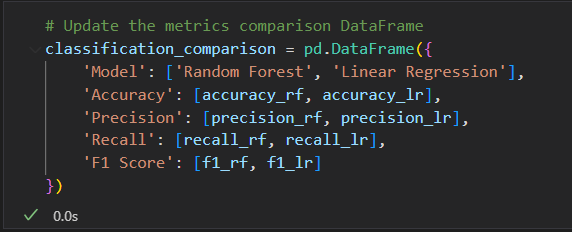
Then we perform some test to determine the greatness of our models.

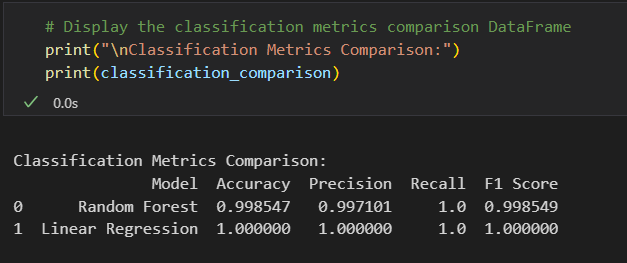








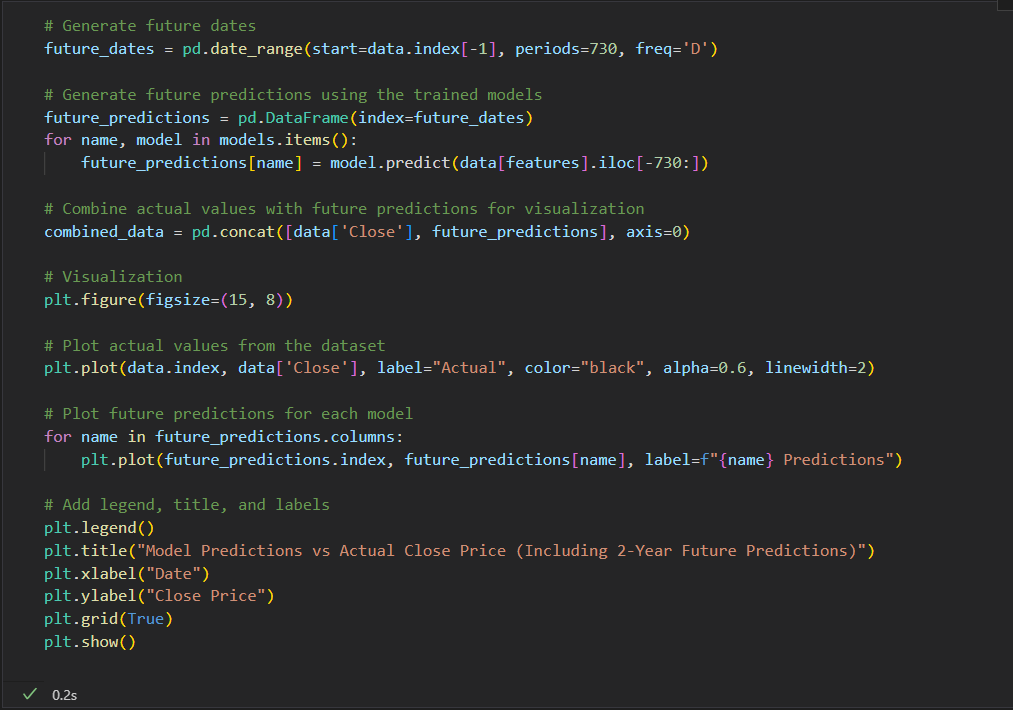
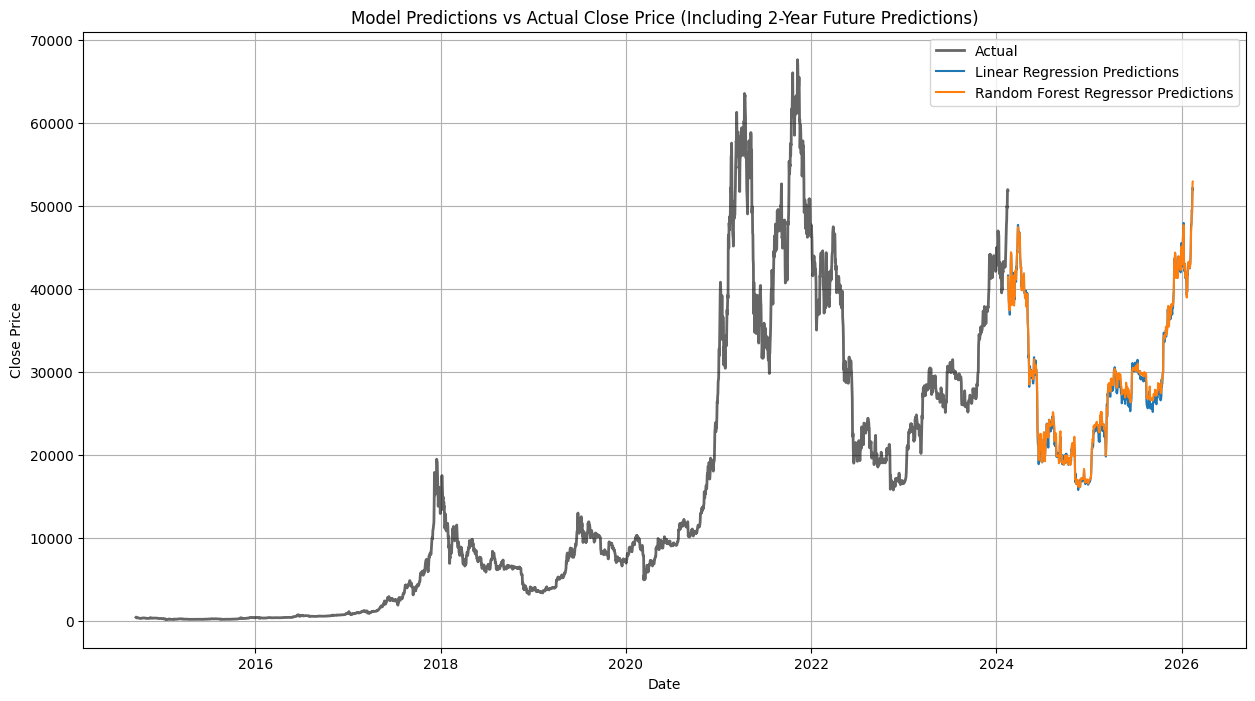




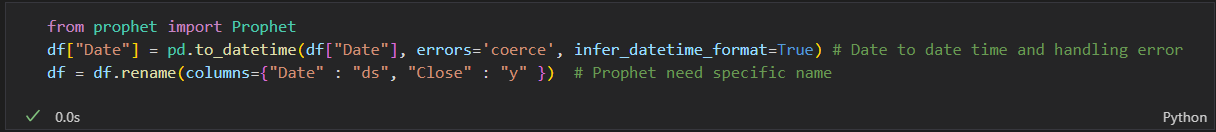
## Prediction

### Prediction using trained models

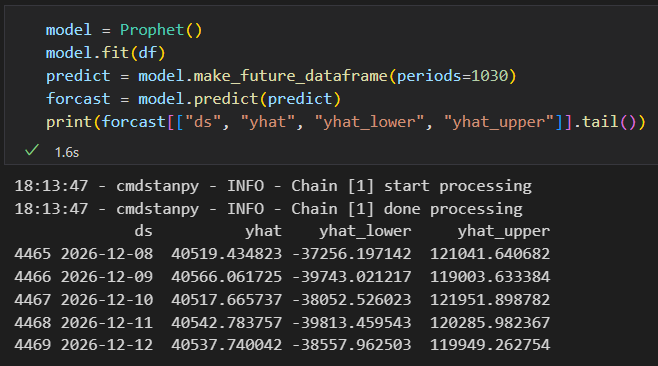
Based on the data provided 2 years prior to 2024, we can use it to make a prediction in the next 2 years.



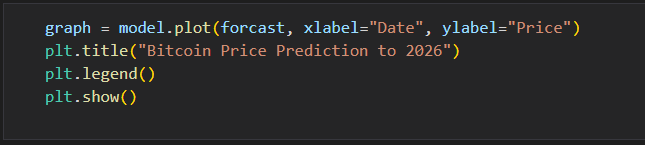
### Prediction using forecast

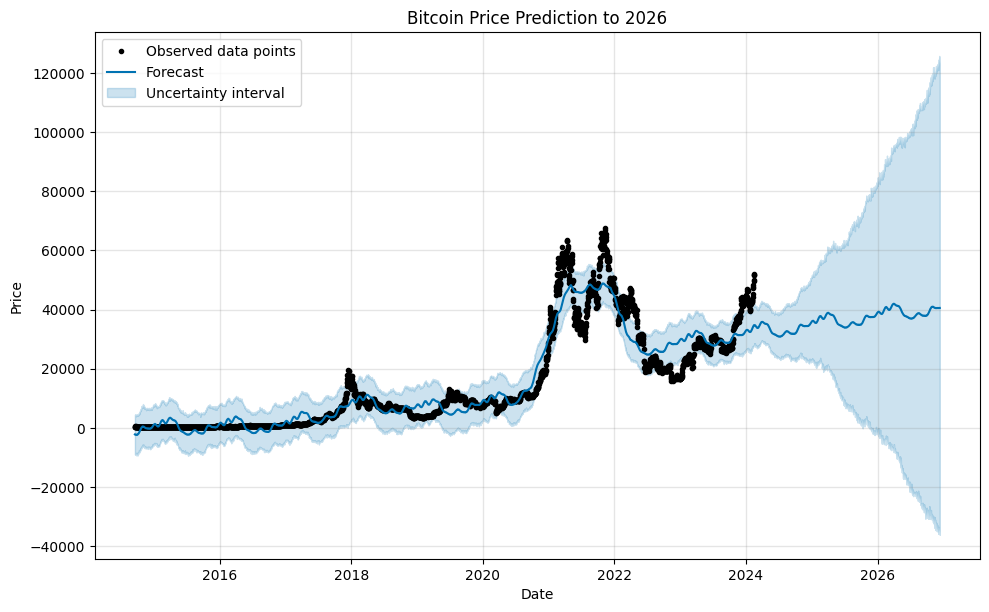
We convert the 'Date' column in the DataFrame to datetime format.

Then we make prediction in another 2 years (2026) with lower bound and upper bound.



Plot the graph of the forecasted values, we get:





# Conclusion

Based on the outputs, here is the analysis and conclusion:

Regression Results

1. Random Forest Model:

• Mean Squared Error (MSE): 22.49 (very low, indicating good performance).

• R-squared (R²): 0.99997 (close to 1, meaning the model explains nearly all the variability in the target variable).

2. Linear Regression Model:

• Mean Squared Error (MSE): 1.76e-08 (almost zero, suggesting very accurate predictions).

• R-squared (R²): 1.0 (perfect score, implying a flawless fit for the data).

The regression metrics indicate that Linear Regression performed slightly better than Random Forest on the dataset, although both models achieved excellent results.

Classification Results

The regression problem was transformed into a classification problem, predicting whether the target variable (e.g., Bitcoin price) is above or below the median value. The results for the classification metrics are as follows:

1. Random Forest Classification Metrics:

• Accuracy: 99.85%

• Precision: 99.71%

• Recall: 100.00%

• F1 Score: 99.85%

2. Linear Regression Classification Metrics:

• Accuracy: 100.00%

• Precision: 100.00%

• Recall: 100.00%

• F1 Score: 100.00%

Conclusion: Both models demonstrated nearly perfect classification performance, with Linear Regression achieving flawless metrics.

Bitcoin Price Forecast (2026)

Using forecasting, the predicted Bitcoin prices (with confidence intervals) for the last few days of 2026 are:

Date Forecasted Price Lower Bound Upper Bound

2026-12-08 38,812.66 -37,412.77 116,359.50

2026-12-09 38,858.63 -38,156.31 118,552.41

2026-12-10 38,809.18 -38,368.99 118,257.63

2026-12-11 38,833.01 -37,570.04 116,204.52

2026-12-12 38,825.65 -37,717.66 117,883.19

These predictions show a wide confidence interval, reflecting the uncertainty in forecasting Bitcoin prices.

Final Takeaways

• Model Performance: Both Random Forest and Linear Regression achieved excellent results for regression and classification tasks.

• Bitcoin Price Forecasting: While the predicted values provide a rough estimate of future prices, the wide confidence intervals indicate uncertainty in long-term predictions.

• Recommendation: Linear Regression seems to be more efficient for this task, offering better performance with simpler computations. However, both models are viable depending on your goals.

# Reference

* Source Code: [Click Here!](https://colab.research.google.com/drive/16ZCqnVA4ulFnhCHI7vuwguvoAV932_Bj?usp=sharing)
* Datasets: [Kaggle](https://www.kaggle.com/datasets/thomkell/bitcoin-usd-btc-usd-16-09-2014-05-02-2024)