

Crypto Forecasting Project: Bitcoin Time Series Prediction

Doaa FATHALLAH



Université Hassan 1er
Academic Year 2025-2026

Contents

1	Introduction	3
2	Dataset	3
2.1	Source	3
2.2	Description	3
2.3	Size and Target	3
3	Data Preprocessing	3
3.1	Phase 1: Acquisition & Simulation	3
3.2	Phase 2: Data Wrangling / Cleaning	3
3.3	Phase 3: Feature Engineering	4
4	Exploratory Data Analysis (EDA)	4
4.1	Visualizations	4
4.2	Observations	6
5	Modeling	6
5.1	Train/Test Split	6
5.2	Models Tested	6
5.3	Hyperparameter Tuning	6
5.4	Evaluation Metrics	6
6	Results	7
6.1	Model Performance	7
6.2	Plots	7
6.3	Observations	8

7 Conclusion 8

8 References 8

1 Introduction

Cryptocurrency markets are highly volatile and predicting future prices is critical for investors and analysts. The goal of this project is to forecast Bitcoin's future value using historical data and to analyze which features most influence price movements.

Problem Type: Time Series Forecasting

Machine Learning Approach: Regression / Predictive Modeling

2 Dataset

2.1 Source

The dataset was obtained from Kaggle: Bitcoin Historical Data.

2.2 Description

Column	Description
Date	Trading date
Open	Opening price of Bitcoin
High	Highest price of the day
Low	Lowest price of the day
Close	Closing price
Volume	BTC traded volume

Table 1: Dataset columns and their description

2.3 Size and Target

The dataset contains XXX rows and YYY columns. The target variable is `Close` price or log returns derived from it.

3 Data Preprocessing

3.1 Phase 1: Acquisition & Simulation

The dataset is loaded from Google Drive in Colab. To simulate real-world data noise, 5% of numeric values are replaced with missing values (NaN).

3.2 Phase 2: Data Wrangling / Cleaning

Missing values are imputed using the mean strategy. Any duplicates or inconsistencies are corrected.

3.3 Phase 3: Feature Engineering

New features such as log returns are created:

$$\text{log_return}_t = \log(\text{Close}_t) - \log(\text{Close}_{t-1})$$

Additional technical indicators (optional) can be added to enhance model performance.

4 Exploratory Data Analysis (EDA)

4.1 Visualizations



Figure 1: Bitcoin closing price evolution over time.

Log_Returns_Distribution.png

Figure 2: Distribution of log returns.

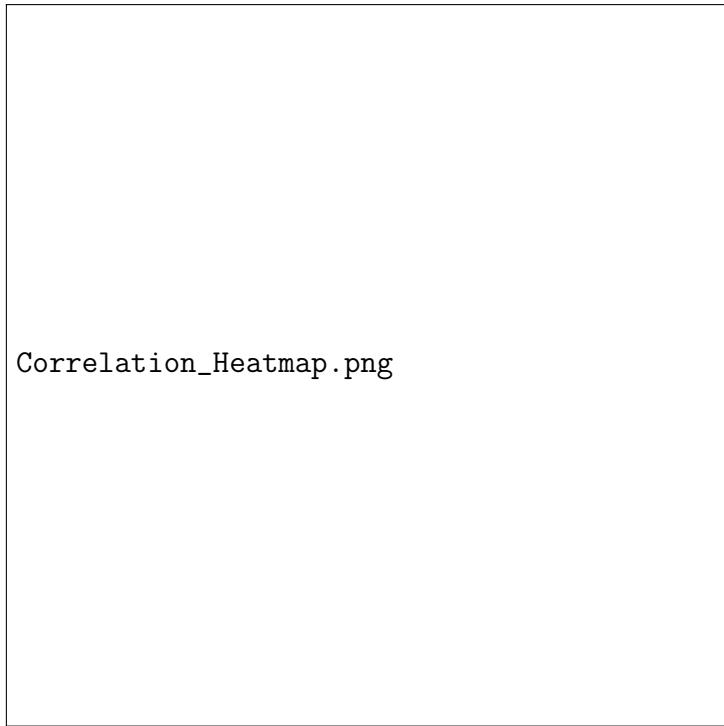


Figure 3: Correlation between numeric features.

4.2 Observations

- Strong trends and high volatility in prices.
- Open, High, Low, Close prices are highly correlated (expected).
- Volume has moderate influence on price changes.

5 Modeling

5.1 Train/Test Split

Data is split chronologically: last 180 days used as the test set.

5.2 Models Tested

- LightGBM Regressor
- Optional: ARIMA, Random Forest Regressor

5.3 Hyperparameter Tuning

- Default parameters used (GridSearchCV optional)

5.4 Evaluation Metrics

- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- R-squared (R^2)

6 Results

6.1 Model Performance

Test RMSE: XXX

Test MAE: XXX

R² Score: XXX

6.2 Plots

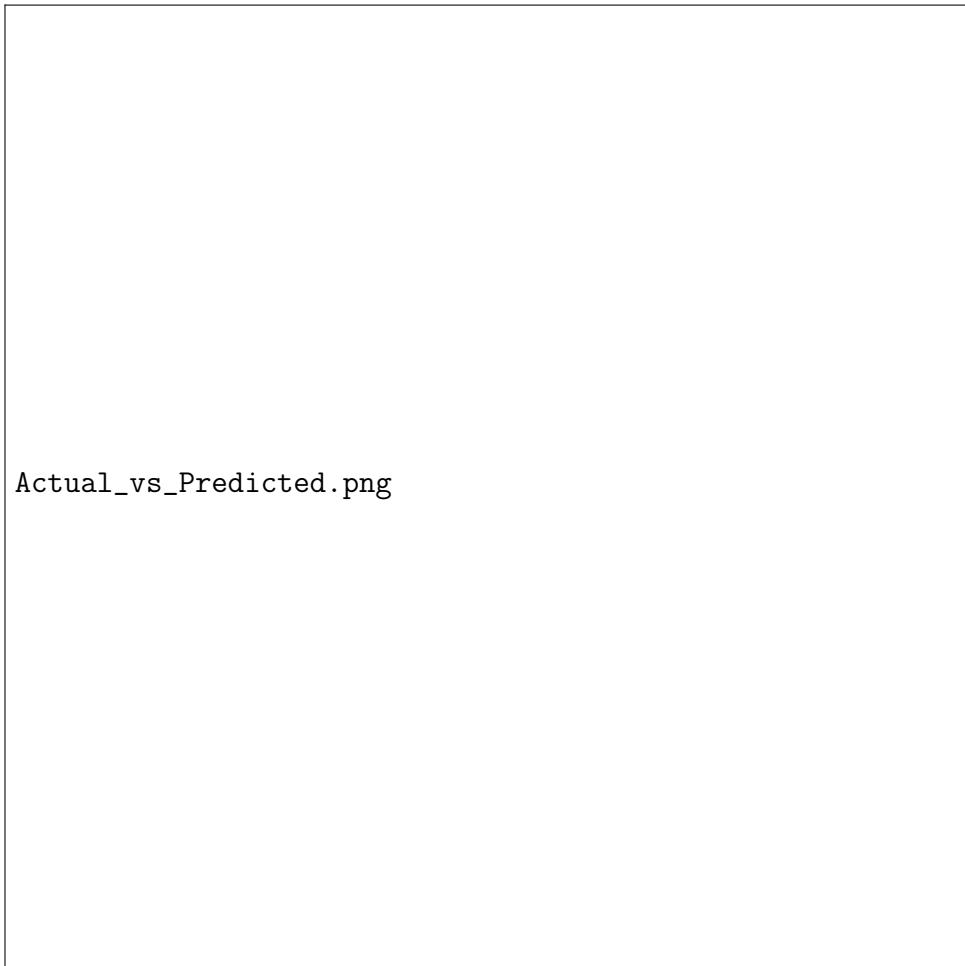
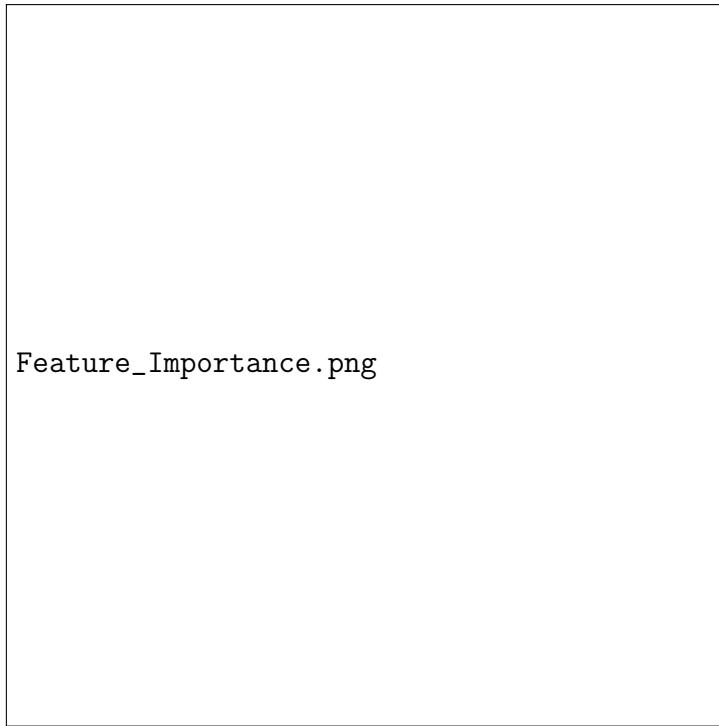


Figure 4: Actual vs Predicted Bitcoin log returns.



Feature_Importance.png

Figure 5: Feature importance chart.

6.3 Observations

- The model captures general trends but may not predict sudden spikes/drops accurately.
- Volume and Open/Close prices are the most influential features.

7 Conclusion

Summary: The predictive model follows all phases, providing insights into Bitcoin price dynamics.

Limitations: - Only historical prices used, no external factors like news sentiment.
- High volatility and sudden market shocks are difficult to predict.

Future Work: - Incorporate technical indicators or sentiment analysis from social media/news. - Use advanced models such as LSTM, Prophet, or ensemble methods.

8 References

- Kaggle Bitcoin Historical Data: <https://www.kaggle.com/datasets>
- Python libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, LightGBM