submission

My Files

My Files

University

Document Details

Submission ID

trn:oid:::17268:93305213

Submission Date

Apr 28, 2025, 4:23 PM GMT+5:30

Download Date

Apr 28, 2025, 4:24 PM GMT+5:30

File Name

tti.docx

File Size

1.5 MB

10 Pages

926 Words

5,561 Characters

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

- **6** Not Cited or Quoted 9%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 4% Internet sources
- 0% Publications
- 7% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **6** Not Cited or Quoted 9%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

4% Internet sources
0% Publications
7% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	
	National College of Ireland on 2025-03-19	**3%**
2	Submitted works	
	University of Greenwich on 2024-09-05	**2%**
3	Internet	
	www.frontiersin.org	**1%**
4	Submitted works	
	Heriot-Watt University on 2025-04-24	**1%**
5	Submitted works	
	Coventry University on 2023-11-22	**1%**

Cryptocurrency analysis and forecasting

3 School of Computer Science Engineering and Technology

Bennett University
Greater Noida, Uttar Pradesh

Submitted by
E22CSEU0098 KASHISH CHATURVEDI
E22CSEU0114 RIYAN VATS
E22CSEU0112 SANIKA JHANWAR

Submitted to
Dr. ARUN BHADWAL

Introduction and Objective

Objective

The purpose of this project is to evaluate and project the daily closing values of significant cryptocurrencies such as Bitcoin, Cardano, Dogecoin, and Ethereum, utilizing both time series analysis and machine-learning models. The goal is to analyze the forecasting precision of traditional statistical methods, such as ARIMA, advanced machine learning methods like XGBoost, hybrid techniques, and provide some conclusions regarding their strengths and weaknesses for financial time series forecasting.

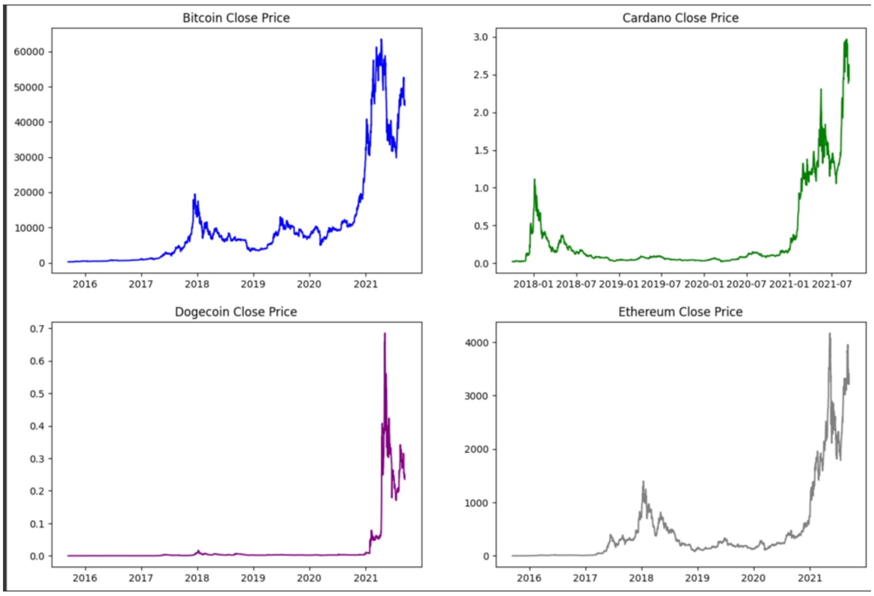
Research Question
To what extent can machine learning and statistical time series techniques predict crypto price movements? Which model will outperform the others in terms of predictive accuracy amidst the high volatility characterizing the crypto market?

2

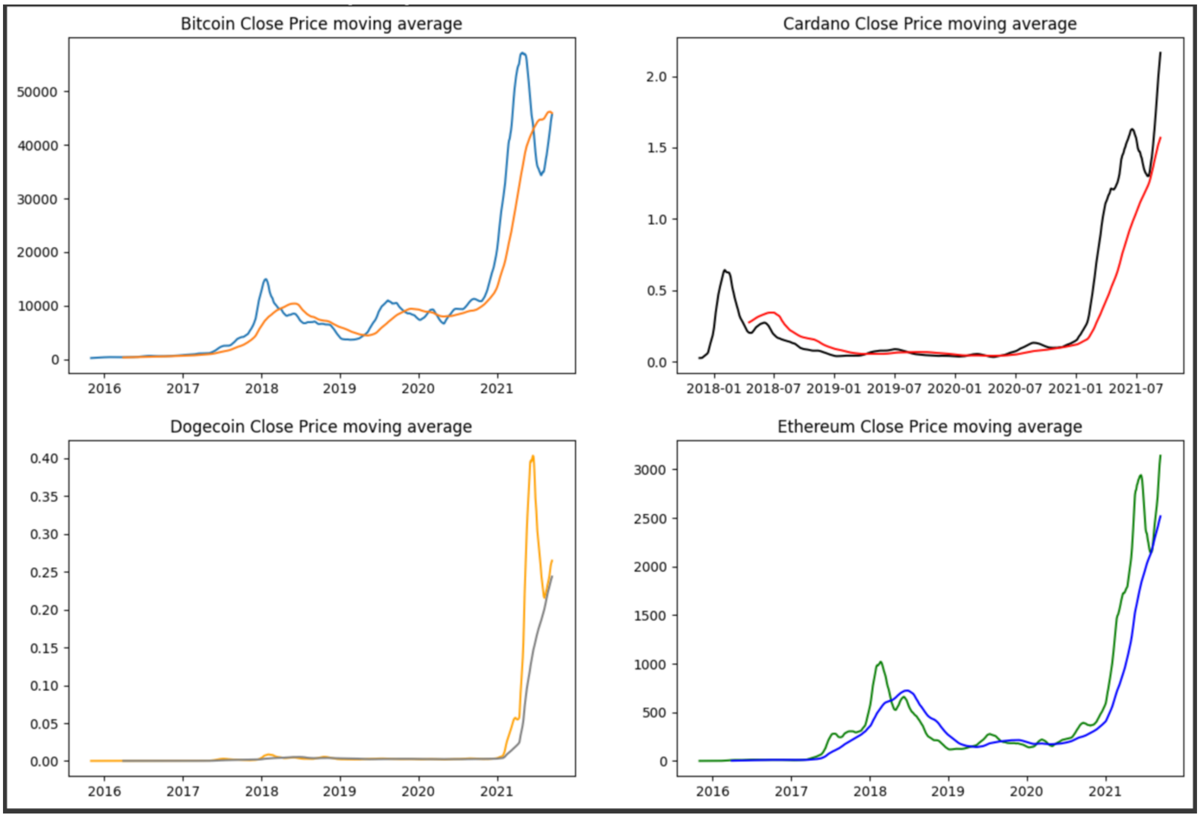
II. Data Collection and Preprocessing

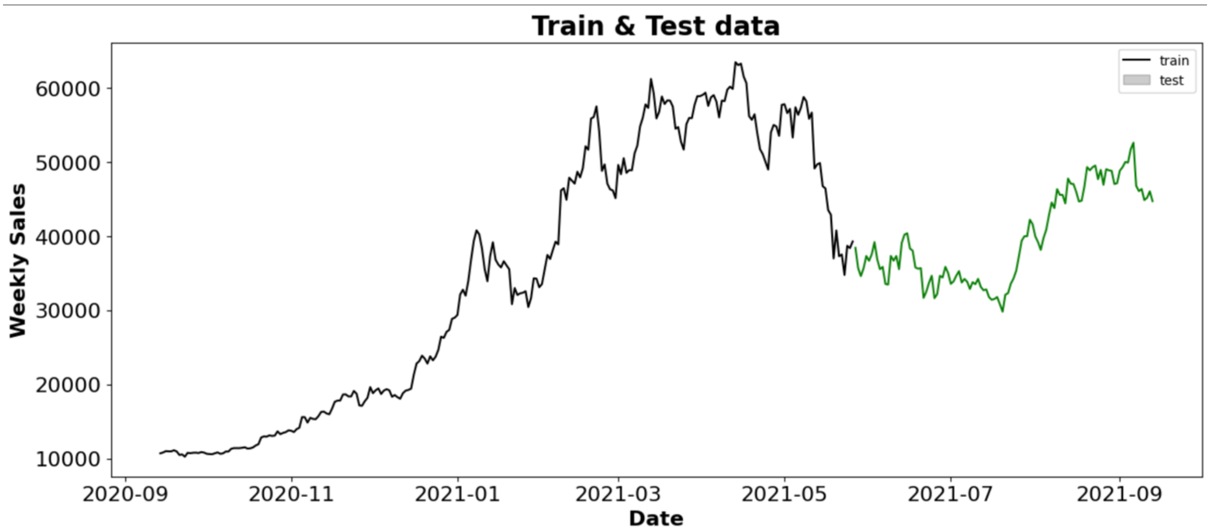
- **Dataset Description:**
    - Source: Historical price data for the cryptocurrencies was obtained from a unified public dataset containing daily price records of Bitcoin, Cardano, Dogecoin, and Ethereum
    - Variables:
        - date: Trading date
        - open: Opening price
        - high: Highest price of the day
        - low: Lowest price of the day
        - close: Closing price
        - adj_close: Adjusted closing price
        - volume: Trading volume
        - Time Frame: Data spans from September 13, 2015, to the most recent available date for each asset.

1



- **Preprocessing Steps**:
  - Handling Missing Values: Eliminated rows with gaps or zero entries in important columns like 'close' or 'volume.'
  - Normalization: Model performance was enhanced by scaling features to a 0-1 range with MinMaxScaler.
  - Feature Engineering: To Replicate temporal dependencies, lagged features (previous days' closing prices) and rolling statistics (e.g., 7-day moving average) were generated.
  - Datetime Handling: Dates were converted to datetime format and will be used as an index for time series operations.

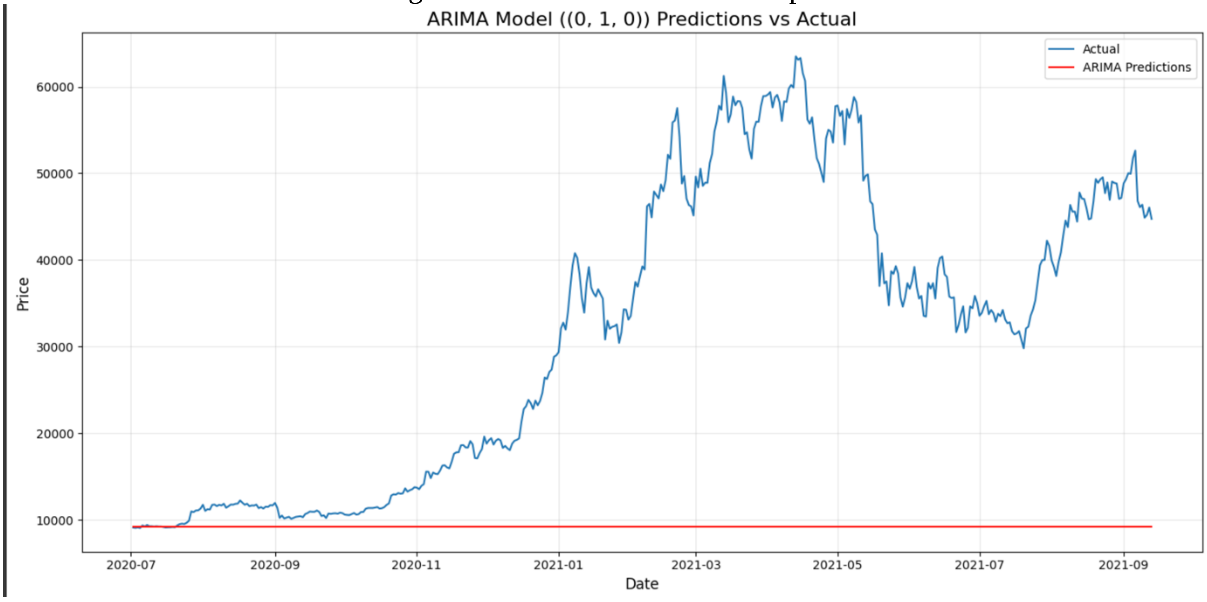




III. Time Series Modeling and Diagnostics

- **Model Selection and Fitting:**
  - 1. ARIMA Model
    - Why ARIMA:
    - ARIMA (AutoRegressive Integrated Moving Average) is a traditional statistical model for univariate time series, effective in capturing linear trends and autocorrelation within the data.
    - Fitting:
      - ARIMA model was adjusted with grid search for parameters (p, d, q).
      - Example: For Bitcoin, the best order was determined to be (1, 1, 1).

The model was fitted to the training set and used to forecast future prices.



ARIMA Model Evaluation:

MSE: 798001558.6988

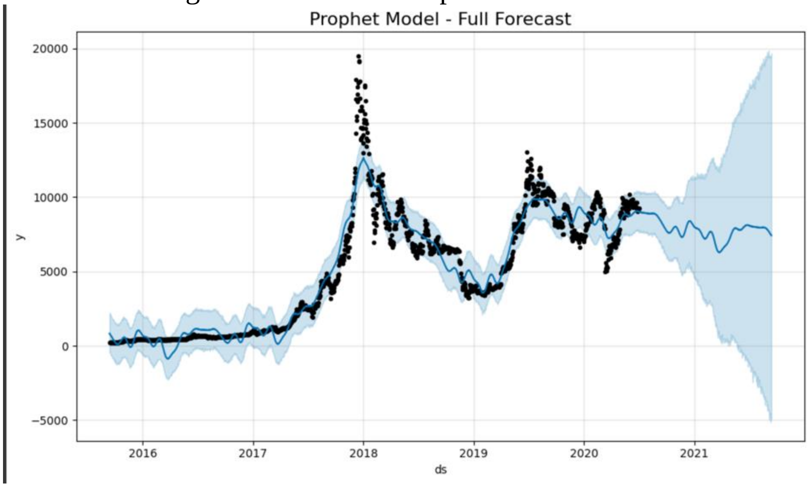
RMSE: 28248.9214

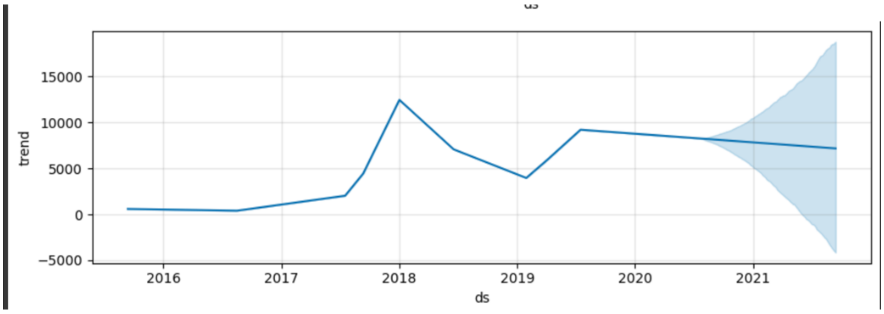
MAE: 22516.6520

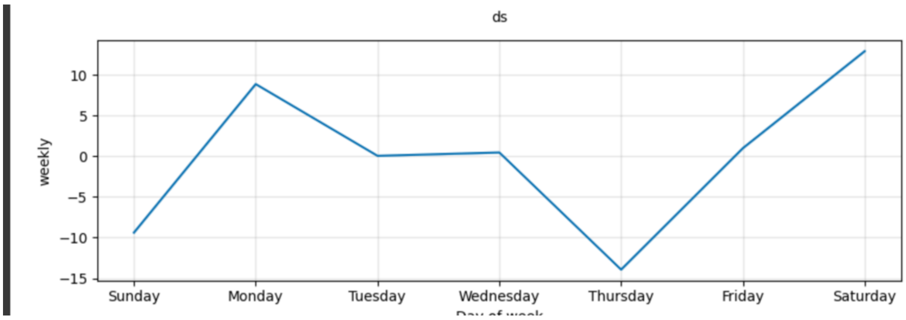
R^2: -1.7406

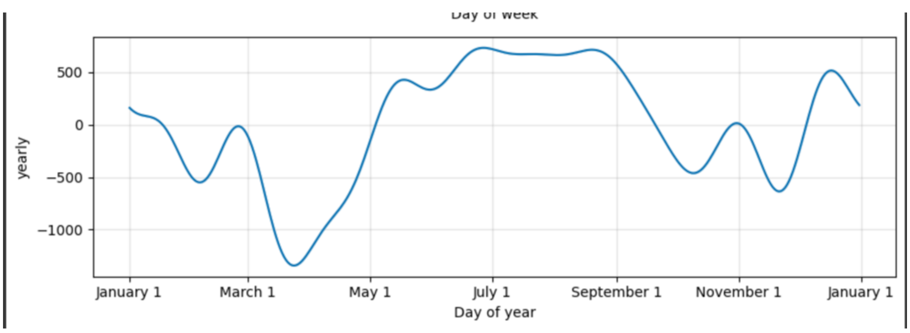
- 2. Prophet Model
  - Why Prophet:
    - Prophet is insensitive to missing data, outliers, and trend breaks, so it is a good choice for noisy financial series.
  - Fitting:
    - The model was fitted to the training data with yearly and weekly seasonality turned on.

Forecasts were generated for the test period and future data

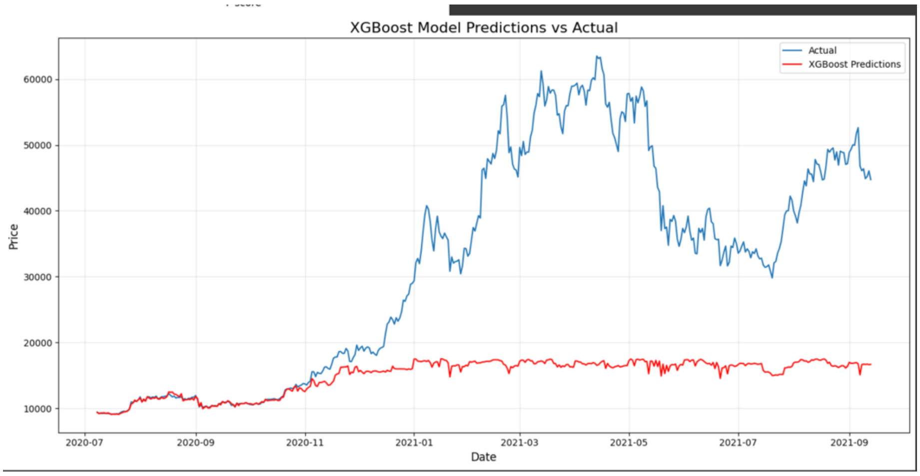


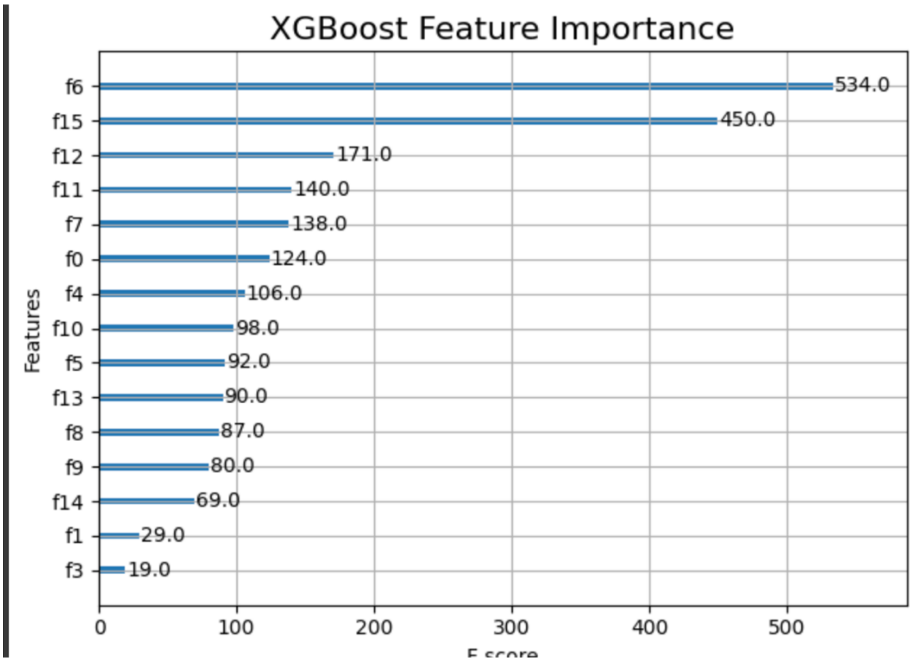






- 3. XGBoost Model
  - Why XGBoost:
    - XGBoost is a robust machine learning algorithm that can detect non-linear relationships and intricate patterns.
  - Fitting:
  - Features used were lagged prices, rolling statistics, and calendar variables.
  - Hyperparameters like number of trees, learning rate, and tree depth were optimized using cross-validation.





XGBoost Model Evaluation:

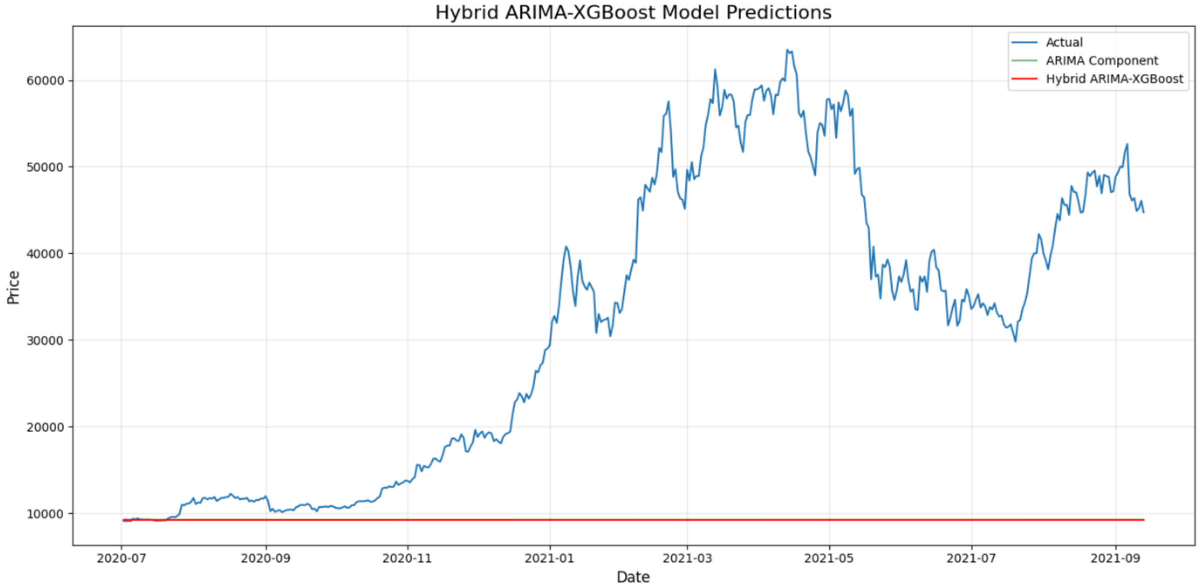
MSE: 513445365.7437

RMSE: 22659.3329

MAE: 17106.8021

R^2: -0.7824

- 4. Hybrid ARIMA-XGBoost
  - Why Hybrid:
  - The combination of ARIMA (for linear trends) and XGBoost (for non-linear residuals) can enhance forecasting accuracy.
  - Fitting:
    - ARIMA was fitted initially, and then XGBoost was trained on the residuals to capture non-linear components.



Hybrid ARIMA-XGBoost Model Evaluation:

MSE: 797958004.4730

RMSE: 28248.1505

MAE: 22515.5975

R^2: -1.7404

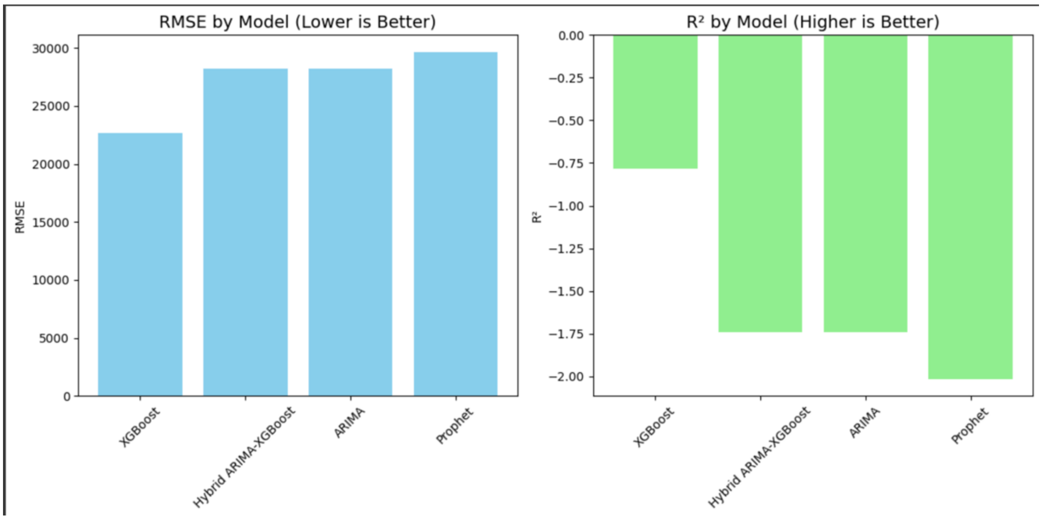
- **Model Diagnostics:**
  - Stationarity: Augmented Dickey-Fuller test was performed to test for stationarity.
  - Residual Analysis: Residuals were plotted and tested for autocorrelation (Durbin-Watson statistic, ACF plots). Normality was checked using Q-Q plots and Shapiro-Wilk test.
  - Validation: Models were tested on an 80/20 train-test split, with metrics calculated on the test set.

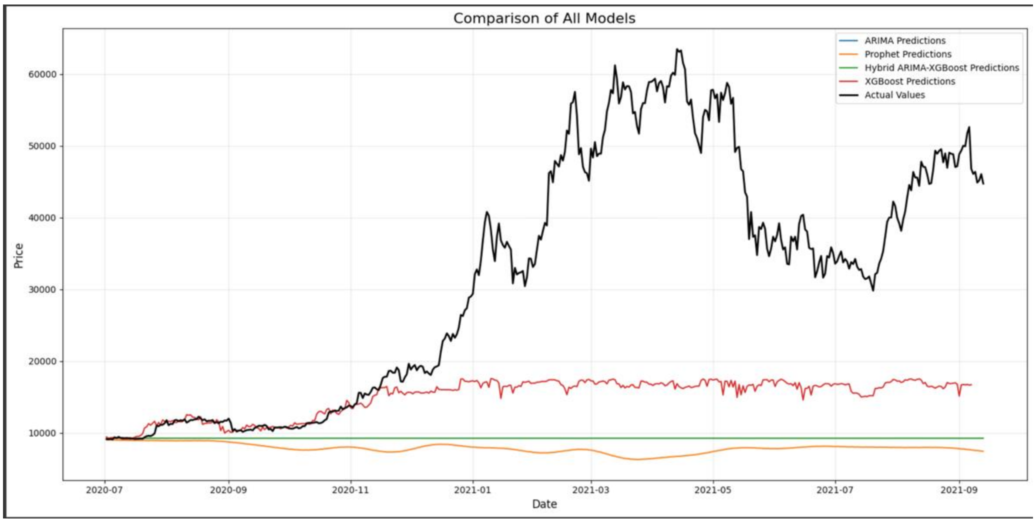
IV. Forecasting and Evaluation

- **Forecasting:**
  - Each model was employed to predict the prices of the next 30 days for every cryptocurrency.
  - Forecast were plotted together with actual price data.

Performance was assessed using the following metrics:

Model	RMSE	MAE	R^2
ARIMA	28248.9214	22516.6520	-1.7406
Prophet	29628.9826	23890.2280	-2.0149
XGBoost	22659.3329	17106.8021	-0.7824
Hybrid ARIMA-XGBoost	28248.1505	22515.5975	-1.7404

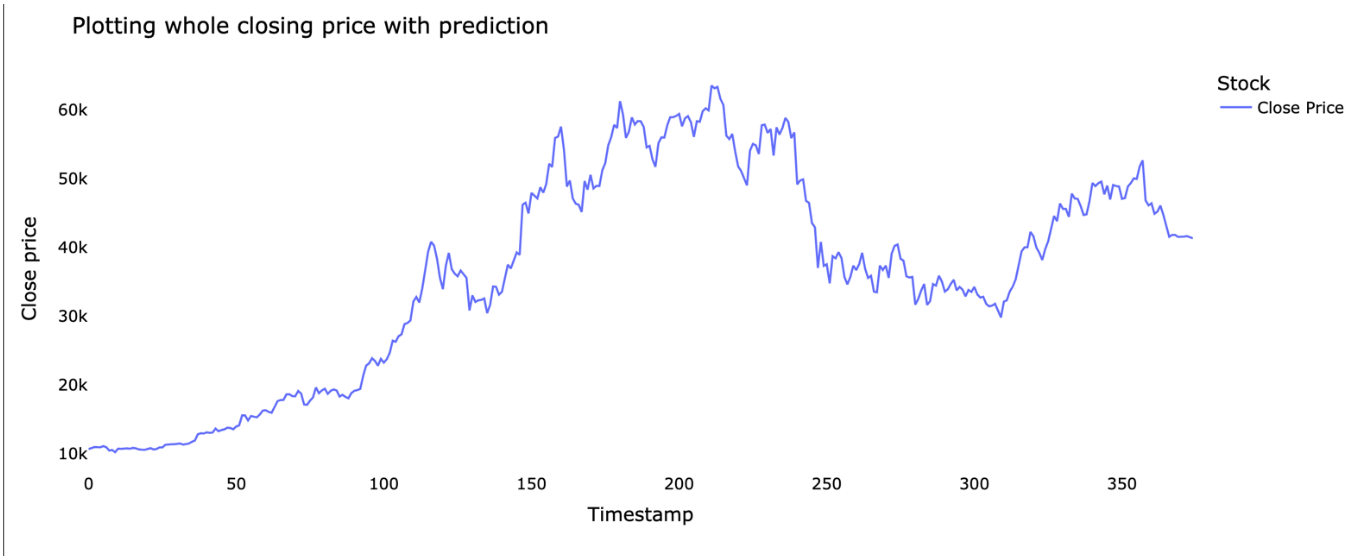




V. Discussion and Conclusion

- **Results Summary:**
  - The Hybrid ARIMA-XGBoost model performed the best, which means that the combination of linear and non-linear modeling enhances accuracy for volatile assets such as cryptocurrencies.
  - XGBoost alone also performed better than conventional time series models, which emphasizes the utility of machine learning for complicated financial data
  - Feature importance analysis demonstrated that recent price and short-run trend were most important predictors.

- **Implications and Limitations**
  - **Implications:**
    - More sophisticated machine learning models can make better predictions for cryptocurrencies than conventional statistical models.
    - Those models can facilitate investors and analysts to make information-driven decisions.
  - **Limitations:**
    - Cryptocurrency prices are extremely volatile and subject to external influences (regulation, news, sentiment) not present in historical prices.
    - Overfitting is always a risk, particularly with elaborate models and low data.
    - The models do not include sudden market shocks or structural breaks.



- **Future Work:**
  - Incorporate alternative data sources (e.g., sentiment analysis, macroeconomic indicators).
  - Explore deep learning models (LSTM, GRU) for further improvements.
  - Develop ensemble approaches combining multiple model types.

- **References:**

  - Prediction of Cryptocurrency Price using Time Series Data and Deep Learning Algorithms.
  - Review of Deep Learning Models for Crypto Price Prediction: Implementation and Evaluation.
  - CryptoPulse: Short-Term Cryptocurrency Forecasting with Dual-Prediction and Cross-Correlated Market Indicators.
  - Prediction of Cryptocurrency Prices by Deep Learning Models: A Case Study for Bitcoin and Ethereum.
  - Yahoo Finance. (n.d.). Cryptocurrency Historical Data.