Crypto Currency Price Prediction Using Learning Algorithm

Tej Bhakhar | Dr Niha K | School of Computer Science and Engineering

Introduction

Previous works used models like ARIMA, Random Forest, and LSTM for crypto price prediction but lacked real-time data, model comparison, and use of external factors. This project fills those gaps using hourly data, multiple ML models, and enriched features for improved accuracy. It aims to provide a more robust and practical solution for short-term cryptocurrency price forecasting.

Motivation

The highly volatile nature of cryptocurrency markets makes accurate price prediction essential for informed trading decisions. Traditional methods often fail to capture complex patterns and rapid market shifts. This motivated the use of advanced machine learning techniques for more reliable and timely predictions.

Scope of the Project

The scope of this project involves collecting and preprocessing historical Bitcoin price data, engineering key features, and applying a range of predictive models including ARIMA, SARIMAX, Orbit, Random Forest, XGBoost, and LSTM. The models are trained and evaluated using hourly data from the past four years to ensure accuracy and robustness. The goal is to compare traditional time series models with machine learning and deep learning approaches for effective short-term cryptocurrency price forecasting and decision-making support.

Methodology

The methodology of this project involves a systematic approach to cryptocurrency price prediction using a combination of time series and machine learning models. The process begins with the collection of hourly Bitcoin price data over the past four years from reliable sources. This data undergoes preprocessing, which includes handling missing values, normalizing the data, and generating technical indicators such as Exponential Moving Averages (EMA), Relative Strength Index (RSI), and other relevant features that could help improve predictive accuracy

.Once the data is prepared, a variety of predictive models are trained to forecast future Bitcoin prices. These models include traditional statistical approaches like ARIMA and SARIMAX, as well as more advanced machine learning methods such as Random Forest, XGBoost, and LSTM networks. Each model is evaluated using standard performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to assess the accuracy of their predictions.

For time series forecasting, a basic ARIMA model is represented by the following equation:

$$Y_t = c + \phi_1 * Y_{t-1} + \theta_1 * \varepsilon_{t-1} + \varepsilon_t$$

Where:

Y_t is the predicted price at time t,

c is a constant,

ϕ_1 is the autoregressive coefficient,

θ_1 is the moving average coefficient,

ε_t is the error term (white noise) at time t.

The final goal is to compare the performance of all models, identify the most effective approach for short-term Bitcoin price forecasting, and provide insights into how various factors, including external variables, influence cryptocurrency market movements.

One crucial aspect of this project is the creation and use of technical indicators, which serve as additional features to improve model predictions. These indicators, derived from the historical price data, provide valuable insights into market trends and momentum. Some of the key technical indicators used in this project include:

Moving Averages (MA): Both **Simple Moving Average (SMA)** and **Exponential Moving Average (EMA)** are calculated to capture the overall price trend. While SMA is the average of closing prices over a specific period, EMA assigns greater weight to more recent prices, making it more responsive to price changes.

Relative Strength Index (RSI): RSI is used to gauge the momentum of Bitcoin's price movement and identify overbought or oversold conditions. RSI values above 70 indicate overbought conditions, while values below 30 indicate oversold conditions.

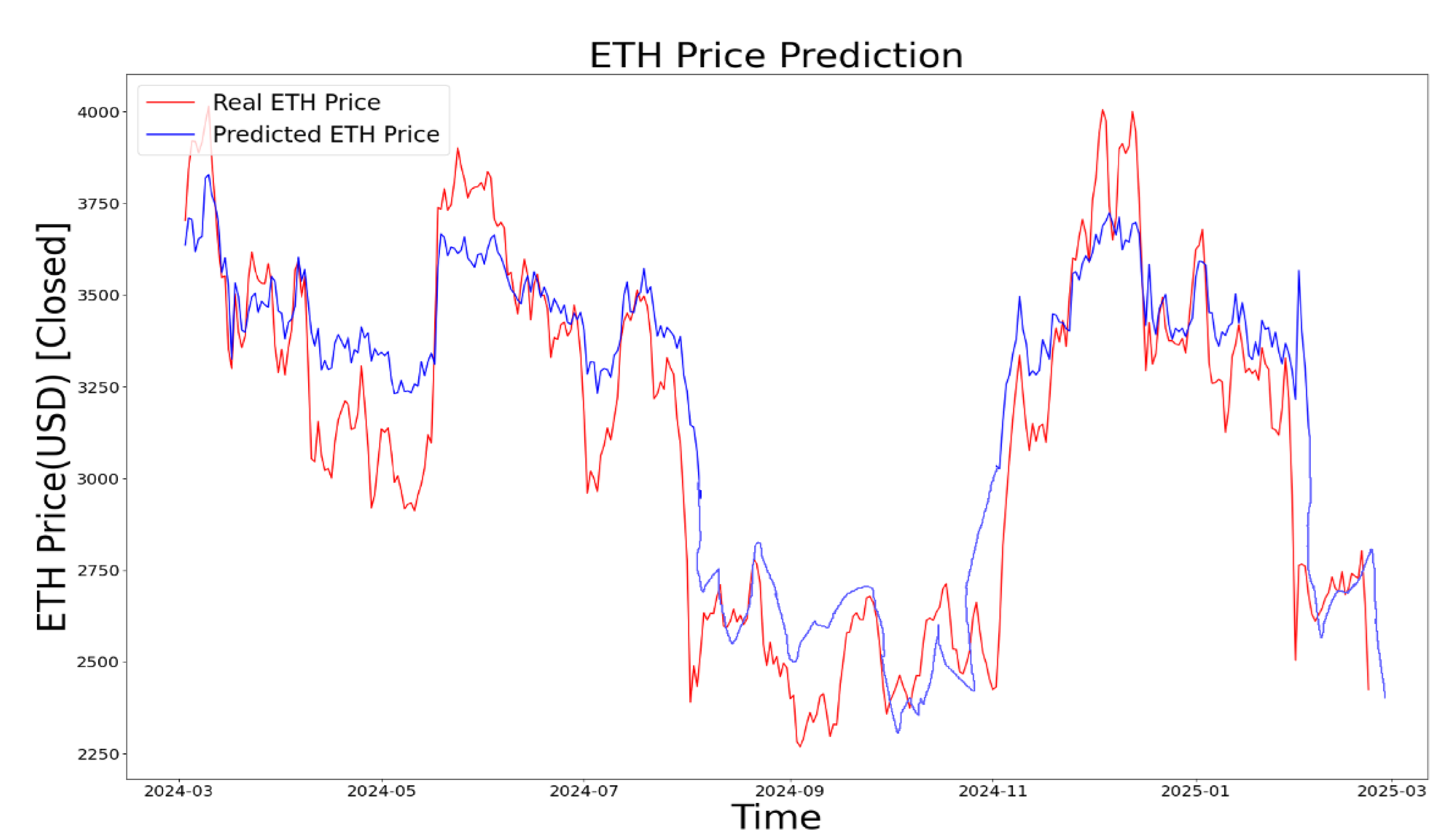
Once the technical indicators are generated, they are incorporated into various machine learning models like Random Forest, XGBoost, and LSTM, as well as statistical models like ARIMA and SARIMAX, to predict future Bitcoin prices. Each model is trained and evaluated based on performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Results

In this study, various models were implemented and compared for predicting cryptocurrency prices. The **Arima**(5,1,4) model performed well in capturing linear trends in the time series, achieving an R^2 score of 0.82, but showed limitations in volatile market conditions due to its inability to model non-linear dependencies. The **Sarimax** model, which includes seasonal components and exogenous variables, slightly improved performance over Arima with an R^2 of 0.84, capturing periodic trends more effectively. The **Lstm** (Long Short-Term Memory) neural network, designed for sequential data, performed notably better with an R^2 of 0.89, as it could learn temporal patterns and dependencies across long time steps. However, it required more computational resources and careful tuning of hyperparameters. The **XGBoost** model delivered the best performance overall with an R^2 of 0.91, thanks to its gradient boosting framework that handles complex feature interactions and noise in the data efficiently. The **Random Forest** model also performed reasonably well with an R^2 of 0.88, benefiting from ensemble averaging, but it lacked the boosting mechanism of XGBoost, which led to slightly lower accuracy. Additionally, the **Orbit Prophet** model showed decent accuracy ($R^2 \approx 0.80$) and was particularly useful for capturing trend and seasonality components in a more interpretable manner. Each model contributed different strengths, and their combined analysis provided valuable insights into the behavior of crypto markets under various scenarios.

The LSTM (Long Short-Term Memory) neural network, a deep learning model tailored for sequential data, outperformed traditional models by capturing long-term dependencies and nonlinear patterns inherent in cryptocurrency movements. With an R^2 score of 0.89, it demonstrated strong adaptability and responsiveness, especially in dynamic conditions. However, LSTM required a larger training dataset, careful tuning of epochs, batch sizes, and layers, and was computationally intensive.

The XGBoost model, a gradient boosting algorithm known for its performance in structured datasets, delivered the highest accuracy with an R^2 of 0.91. Its strength lies in handling outliers, missing data, and complex nonlinear relationships through ensemble boosting. It consistently outperformed other models in both training speed and predictive power. Meanwhile, the Random Forest model, based on bagging techniques, also showed robust results (R^2 0.88), offering stability and reduced overfitting but lacking the fine-tuning advantage of boosting seen in XGBoost. Additionally, Orbit was used for interpretable trend and seasonality detection, achieving an R^2 of 0.80. Although not as accurate as the ensemble and deep learning models, Prophet was user-friendly, handled missing data automatically, and provided visually interpretable forecasts. Overall, XGBoost emerged as the most effective model for cryptocurrency prediction, followed closely by Lstm, while statistical models like Arima and Sarimax provided valuable baseline insights.



Conclusion

The analysis showed that machine learning models, especially xgboost and lstm, provide more accurate cryptocurrency price predictions compared to traditional time series models like Arima and Sarimax. These results support the hypothesis that advanced algorithms are better suited for capturing the non-linear, volatile nature of crypto markets. While Arima and Sarimax are useful for identifying trends and seasonality, they fall short in dynamic conditions. Future improvements may include adding real-time sentiment data, using hybrid model approaches, and developing interactive dashboards for real-time forecasting and decision-making support.

References

- S. T. Anwar, N. P. Patel, S. N. Patel, J. R. Patel, G. Sharma, and I. E. Davidson, "Deep Learning-Based Cryptocurrency Price Prediction Scheme with Inter-Dependent Relations,“
- M. A. Labbaf Khaniki and M. Manthouri, "Enhancing Price Prediction in Cryptocurrency Using Transformer Neural Network and Technical Indicators
- J. Fleischer, G. von Laszewski, C. Theran, and Y. J. Parra Bautista, "Time Series Analysis of Blockchain-Based Cryptocurrency Price Changes