Cryptocurrency Price Prediction Using Machine Learning Techniques

Divyanshu Prince[1], Aniket Baranwal[2], Deepak Kumar[3], Nitin Singh[4], Dr Sreenivasa B C[5]

[1234]*UG Student, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru, Karnataka, India*

[5]*Associate Professor, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru, Karnataka, India*

Abstract: **Cryptocurrencies have emerged as a disruptive innovation in the financial sector, attracting widespread interest from investors, researchers, and regulators. However, predicting their price remains a significant challenge due to their high volatility and sensitivity to market sentiment. This paper explores the use of machine learning techniques for forecasting cryptocurrency prices, focusing primarily on Bitcoin and Ethereum. We implement and compare models such as Linear Regression, Long Short-Term Memory (LSTM), and Facebook Prophet, evaluating their performance based on historical price data. Our findings indicate that deep learning models like LSTM outperform traditional approaches in capturing temporal dependencies, offering better predictive accuracy.**

Keywords: Cryptocurrency, Price Prediction, Machine Learning, LSTM, Time Series Forecasting, Bitcoin, Ethereum

I. INTRODUCTION

The rise of cryptocurrencies has transformed digital finance by introducing decentralized, blockchain-based assets that operate outside traditional banking systems. Bitcoin, the first and most well-known cryptocurrency, has inspired the creation of thousands of other digital currencies. Due to their speculative nature and significant price fluctuations, accurately forecasting cryptocurrency prices is of great interest to investors, traders, and researchers. Traditional financial models often fall short due to the nonlinear and non-stationary behavior of these markets. Consequently, machine learning approaches have gained prominence for their ability to model complex patterns and make data-driven predictions..

II. LITERATURE SURVEY

Several studies have explored various methods to forecast cryptocurrency prices. McNally et al. (2018) employed LSTM networks for Bitcoin prediction and demonstrated the model's ability to capture temporal dependencies. Jang and Lee (2017) combined blockchain information with deep learning models to improve forecast accuracy. On the other hand, Patel et al. (2015) explored traditional regression and tree-based models for stock prediction, which have been adapted for crypto. Recently, Prophet, a forecasting tool developed by Facebook, has shown promise due to its simplicity and interpretability. Despite progress, challenges remain due to the influence of social sentiment, regulatory announcements, and technical limitations in feature selection.

III. METHODOLOGY

This section outlines the complete process of data collection, preprocessing, model implementation, and evaluation. The comparison of Linear Regression, Long Short-Term Memory (LSTM), and Facebook Prophet models provides a comprehensive analysis of their suitability for cryptocurrency price prediction.

I Data Collection

The data for both Bitcoin (BTC) and Ethereum (ETH) was sourced from CoinMarketCap, a widely trusted platform for historical cryptocurrency data. The dataset covers a period of more than eight years, from January 1, 2017, to March 31, 2025, capturing critical market movements, including several cycles of bull and bear markets. The data includes essential market features like:

- Open: The first price recorded at the start of the day.
- Close: The last price recorded at the end of the day.
- High: The highest price achieved during the day.
- Low: The lowest price achieved during the day.

- Volume: The total trading volume for the cryptocurrency on a given day.

These time-series data points form the foundation for the subsequent analysis and model development. Additional features, such as moving averages and technical indicators, were derived from the raw data to enhance the models' ability to capture market behavior.

II Data Preprocessing

Effective data preprocessing is crucial in preparing the data for model training. Several steps were implemented to handle the raw data and convert it into a usable format for machine learning:

- Data Cleaning: Any missing, outlier, or erroneous data entries were identified and dealt with. For instance, when large gaps in data occurred (e.g., from exchange downtime), those dates were either removed or interpolated using linear methods to maintain data consistency.
- Missing Value Treatment: Missing price points and volume data were handled using forward fill for small gaps (assuming the last known value is representative), or linear interpolation when larger gaps were present to maintain the continuity of the time series.
- Normalization: Given the wide range of values in the dataset (especially in cryptocurrency prices), normalization was performed to scale the data within a specific range. Min-Max Scaling was used to ensure that features like price and volume were transformed to a similar scale to prevent any feature from dominating the learning process, especially for the LSTM model.
- Supervised Learning Conversion: Since LSTM and Prophet models work with sequences, the dataset was reshaped into a supervised learning format where the past n days' data (lag features) were used to predict the price on the $n+1$ day. This was done for both BTC and ETH to make the models handle the time-series forecasting task effectively.

III Feature Engineering

The feature engineering process included the creation of various inputs designed to capture the market's trend and volatility. These features were selected based on technical analysis principles and their relevance in cryptocurrency forecasting:

- Price Indicators:

- Moving Averages (MA): Calculated for various window sizes (e.g., 7-day, 30-day, 200-day) to capture short-term and long-term trends.
- Exponential Moving Average (EMA): Given more weight to recent prices, it was useful for capturing immediate market sentiment.
- Technical Indicators:
- Relative Strength Index (RSI): Used to measure overbought or oversold conditions in the market, helping identify potential reversal points.
- Moving Average Convergence Divergence (MACD): Helped to capture trends in price momentum and signal crossovers between fast and slow-moving averages.
- Lagged Variables:
- Lag Features: The closing price from the previous n days (e.g., 1-day lag, 7-day lag, etc.) was used as input to forecast the next day's closing price.
- Volatility Measures:
- Average True Range (ATR): Used to gauge the volatility of the cryptocurrency market, which is particularly important given the inherent price fluctuations in cryptocurrencies.

These features provided the models with more granular insights into the market dynamics and allowed them to capture both price trends and volatility patterns effectively.

IV. MODEL IMPLEMENTATION

The implementation of the three models followed these key steps:

- Linear Regression: Linear Regression was implemented as a benchmark model due to its simplicity and ease of interpretability. This model assumes a linear relationship between the target (price) and the features (independent variables). Although this model has limited capacity for capturing complex patterns, it serves as a valuable baseline for comparison.
- LSTM (Long Short-Term Memory): The LSTM model, a type of recurrent neural network, was designed to capture temporal dependencies in the data. Since cryptocurrency prices are highly influenced by historical patterns and trends, the LSTM model was chosen due to its ability to maintain long-term memory over sequences.

The LSTM architecture was configured with:

- 1-2 hidden layers with 50-100 neurons each.
- Dropout regularization (0.2) to prevent overfitting.
- Adam optimizer with a learning rate of 0.001.
- Batch size of 32 and epochs set to 50 to balance training time and model convergence.

- Facebook Prophet:

Prophet is designed specifically for time-series forecasting and is capable of capturing seasonality and trend shifts without the need for manual feature engineering. The model's strengths include its ability to:
- Handle missing data naturally.
- Account for holidays and special events.
- Automatically model seasonality (daily, weekly, yearly).
- Prophet was set with:
  - Automatic seasonalities for daily, weekly, and yearly trends.
  - Custom holidays were not added in this experiment, assuming no special events influenced the price.
  - Standard settings for uncertainty intervals (95%) were used.

Each model was trained on a rolling window of historical data, with the last 20% of the data used as a test set for evaluating the forecast accuracy.

Evaluation Metrics

The evaluation of model performance was based on three common metrics, chosen to assess both the accuracy and robustness of the predictions:
- Mean Absolute Error (MAE): MAE measures the average of the absolute errors between the predicted and actual values. It is a simple, interpretable metric for assessing the accuracy of the model.
- Root Mean Square Error (RMSE): RMSE is sensitive to large errors due to the squared term, making it a good indicator of how well the model handles outliers or volatile price fluctuations.
- Mean Absolute Percentage Error (MAPE): MAPE offers a percentage-based error measurement, which is particularly useful in evaluating the model's prediction accuracy relative to actual values, making it easier to interpret performance across different cryptocurrencies.

By using these metrics, the study provides a comprehensive view of how well each model performs and how they handle the unique characteristics of cryptocurrency price data.

V. RESULTS ANALYSIS AND DISCUSSION

This section evaluates the performance of the three forecasting models—Linear Regression, LSTM, and Facebook Prophet—using the metrics Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Linear Regression Model Performance:
- MAE and RMSE were significantly higher compared to the other models, indicating that Linear Regression struggled to capture the complex, non-linear patterns in cryptocurrency price movements.
- MAPE was also high, reflecting the model's inability to predict sharp price fluctuations or capture sudden market shifts effectively.

Discussion: Linear Regression performed poorly due to its simplicity and inability to handle the volatility of cryptocurrency prices. Its linear assumptions are insufficient for predicting the price dynamics of cryptocurrencies, which are influenced by numerous non-linear factors.

Long Short-Term Memory (LSTM) Model Performance:
- MAE and RMSE were significantly improved compared to Linear Regression, particularly in capturing long-term trends.
- MAPE was lower than Linear Regression, though still higher during periods of high volatility.

Discussion: The LSTM model showed stronger performance, especially for long-term price trends. However, it struggled to predict large price movements and sudden market shocks. The model benefits from its ability to capture temporal dependencies but could improve by incorporating additional external features.

Facebook Prophet Model Performance:
- MAE, RMSE, and MAPE were the lowest among all models, particularly for Ethereum, demonstrating Prophet's ability to capture trends and seasonality effectively.

- For Bitcoin, Prophet's performance was slightly weaker, likely due to Bitcoin's higher volatility.

Discussion: Prophet outperformed both Linear Regression and LSTM by capturing seasonal patterns and trends. While it performed best for Ethereum, it had limitations in forecasting Bitcoin's more volatile movements, where short-term shocks played a significant role.

Model Comparison
- Linear Regression: Lags behind in performance due to its inability to handle volatility.
- LSTM: Improved performance, capturing long-term trends but still sensitive to sudden price changes.
- Prophet: Best overall performer, excelling in capturing seasonal trends, especially for Ethereum, though less effective for Bitcoin's volatility.

In conclusion, Prophet provided the best results overall, particularly for Ethereum, while LSTM performed better than Linear Regression but still faced challenges during periods of high volatility.

ACKNOWLEDGMENT

REFERENCES

[1]. K.-K. R. Choo, "Cryptocurrency and virtual currency: Corruption and money laundering/terrorism financing risks?" in *Handbook of Digital Currency*, Elsevier, pp. 283-307, 2015.

[2]. R. Chowdhury, M. A. Rahman, M. S. Rahman, and M. R. C. Mahdy, "An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning", *Phys. A Stat. Mech. its Appl.*, vol. 551, pp. 124569, 2020.

[3]. E. Akyildirim, A. Goncu, and A. Sensoy, "Prediction of cryptocurrency returns using machine learning", *Ann. Oper. Res.*, vol. 297, no. 1, pp. 3-36, 2021.

[4]. Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering", *J. Comput. Appl. Math.*, vol. 365, pp. 112395, 2020.

[5]. H. Sebastião and P. Godinho, "Forecasting and trading cryptocurrencies with machine learning under changing market conditions", *Financ. Innov.*, vol. 7, no. 1, pp. 1-30, 2021.

[6]. P. Jay, V. Kalariya, P. Parmar, S. Tanwar, N. Kumar, and M. Alazab, "Stochastic neural networks for cryptocurrency price prediction", *IEEE Access*, vol. 8, pp. 82804-82818, 2020.