International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

xx

IJIEMR Transactions, online available on 02th May 2024. Link

http://www.ijiemr.org/downloads.php?vol=Volume-13&issue=Issue5

10.48047/IJIEMR/V13/ISSUE 05/02

TITLE: FORECASTING BITCOIN PRICE CHANGES WITH MACHINE LEARNING

Volume 13, ISSUE 04, Pages: 9-17

Paper Authors **Naga Siva Jyothi Kompalli,kondala anusha,Medipally Vinay,Komma Shivani**

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

FORECASTING BITCOIN PRICE CHANGES WITH MACHINE LEARNING

Naga Siva Jyothi Kompalli[1],kondala anusha[2],Medipally Vinay[3],Komma Shivani[4,]

[1] Dept of IT, SNIST, HYD. jyothi.kompalli@gmail.com
[2] Dept of IT,SNIST,HYD, kondalaanusha69@gmail.com
[3] Dept of IT,SNIST,HYD, vinayyaniv283a@gmail.com
[4] Dept of IT,SNIST,HYD, shivanikomma2020@gmail.com

Abstract

In recent years, the burgeoning popularity and economic significance of crypto currencies, particularly Bitcoin, have sparked heightened interest in forecasting analytics within the financial realm. This paper delves into the domain of forecasting Bitcoin price, employing a comparative analysis framework to evaluate the efficacy of three distinct machine learning algorithms: Long Short-Term Memory (LSTM), Random Forest, and eXtreme Gradient Boost. Leveraging historical data and sophisticated modelling techniques, our study scrutinizes the forecasting capabilities of each algorithm, shedding light on their respective strengths and weaknesses in forecasting Bitcoin prices. Through meticulous examination and statistical evaluation, we unravel nuanced insights into the intricate dynamics of crypto currency markets. Furthermore, this research underscores the paramount importance of algorithmic selection in optimizing forecasting accuracy and reliability, offering pragmatic guidance for stakeholders navigating the volatile terrain of investment. By elucidating the multifaceted landscape of forecasting Bitcoin price, this study contributes to the burgeoning discourse surrounding the intersection of machine learning and financial forecasting, paving the way for enhanced decision-making frameworks and informed strategies in the realm of digital assets.

Keywords: ***Forecasting Bitcoin Price, Long ShortTerm Memory(LSTM), Random Forest, eXtreme Gradient Boost(XGBoost), Comparative Analysis, Financial Forecasting, Crypto currency Markets***.

I. INTRODUCTION

In the rapidly evolving landscape of machine learning and forecasting analytics, the quest to forecast Bitcoin prices with precision has become a focal point of research and exploration. Harnessing the power of advanced algorithms such as Long Short-Term Memory (LSTM) and Random Forest, practitioners aim to unravel the intricate dynamics underlying crypto currency markets. LSTM, a sophisticated architecture within the realm of deep learning, stands as a beacon of hope for capturing temporal dependencies and order dependence in sequence forecasting problems—a necessity in domains like machine translation and speech recognition. Simultaneously, Random Forest emerges as a formidable contender, leveraging decision trees to dissect the historical patterns dictating Bitcoin price fluctuations. With its ensemble approach and inherent capability to process structured data, Random Forest seeks to illuminate the causal relationships between various factors and Bitcoin market prices. Both LSTM and Random Forest epitomize the fusion of cutting-edge technology with domain-specific expertise, propelling the boundaries of predictive analytics in the crypto currency sphere.

Moreover, amidst the quest for accurate forecasting, the role of eXtreme Gradient Boost—an ensemble algorithm rooted in gradient boosting—comes to the fore. While artificial neural networks often reign supreme in forecasting tasks involving unstructured data, eXtreme Gradient Boost shines in scenarios featuring small-to-medium structured/tabular data. Its proficiency in regression, classification, ranking, and user-defined forecasting problems renders it a cornerstone in the arsenal of forer casting modelling tools.

By delving into the nuances of Long Short Term Memory, Random Forest, and eXtreme Gradient Boost, this paper embarks on a journey to dissect the intricacies of Bitcoin price forecasting. Through comprehensive analysis and empirical validation, we aim to unravel the strengths and limitations of each algorithm, shedding light on their efficacy in deciphering the enigmatic nature of crypto currency markets. As we navigate through the labyrinth of machine learning methodologies, our endeavor is not merely to forecast Bitcoin prices, but to unravel the underlying mechanisms driving digital asset valuations.

II. LITERATURE SURVEY

In the realm of financial markets, Bitcoin has emerged as a prominent and disruptive force, captivating the attention of investors, traders, and researchers alike. As the world's first decentralized digital currency, Bitcoin operates outside the traditional banking system, relying on cryptographic principles and blockchain technology to facilitate peer-to-peer transactions. With its decentralized nature, limited supply, and significant price volatility, Bitcoin presents a unique opportunity for studying and forecasting financial market dynamics using machine learning techniques.

This literature survey aims to explore existing research efforts in the domain of forecasting Bitcoin price utilizing machine learning methodologies. Machine learning (ML) has emerged as a promising tool for developing models capable of analyzing historical price data and identifying patterns or trends that may indicate future price movements.

Auto Regressive Integrated Moving Average (ARIMA) models have been widely used in the financial domain, including the forecasting of Bitcoin prices. ARIMA is a time series forecasting method that captures the linear relationships and temporal dependencies present in sequential data. It comprises three main components: autoregression (AR), differencing (I), and moving average (MA). Several studies have explored the application of ARIMA models specifically for Bitcoin price forecasting. These studies typically focus on the historical price data of Bitcoin and attempt to model its future movements based on past trends and patterns. The advantage of ARIMA lies in its ability to capture both short-term fluctuations and long-term trends in Bitcoin prices.

Researchers have investigated various aspects of ARIMA modeling in the context of Bitcoin price forecasting. They have explored different model configurations, including variations in the order of autoregression, differencing, and moving average parameters, to optimize predictive performance. Additionally, studies have examined the impact of seasonality and other temporal factors on Bitcoin price dynamics and incorporated them into ARIMA models for more accurate forecasting.

Research by Azari delves into the intricate task of revealing the usefulness of traditional autoregressive integrative moving average (ARIMA) model in

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

forecasting the future value of bitcoin by analyzing the price time series. It is observed that it introduces large forecasting errors. Especially, the ARIMA model is unable to capture the sharp fluctuations in the price, e.g. the volatility at the end of 2017. Azari's study sheds lights on the interaction of the prediction accuracy, choice of (p, q, d), and window size w.

Overall, the literature on ARIMA for Bitcoin price forecasting underscores the importance of leveraging time series analysis techniques to gain insights into cryptocurrency market dynamics and make informed investment decisions. Continued research and development in this area are essential for advancing our understanding of Bitcoin price movements and enhancing the reliability of forecasting models.

III. RESEARCH GAP

The existing literature on forecasting bitcoin prices has laid a solid foundation by exploring various machine learning algorithms. While these studies have contributed valuable insights into the predictive capabilities of these algorithms, there remains a notable gap in the synthesis of advanced algorithms, particularly Long Short Term Memory, within user-friendly interfaces tailored specifically for forecasting bitcoin prices.

It focuses on Bitcoin price prediction, there may be a research gap in extending the analysis to other cryptocurrencies.

Traditional time series algorithms, such as Autoregressive Integrated Moving Average (ARIMA) or Exponential Smoothing, lay the groundwork for understanding temporal dependencies in sequential data. LSTM builds upon this by capturing long-term dependencies and learning intricate patterns in time series data, making it well-suited for capturing complex dynamics in Bitcoin price movements. LSTM's ability to model non-linear relationships and retain long-term memory enables it to capture subtle price trends, seasonality, and irregularities that may be challenging for traditional linear models.

Random Forest Regression offers several advantages over traditional algorithms, including its ability to handle large datasets with numerous features and capture nonlinear relationships effectively.

XGBoost's boosting framework sequentially improves the performance of weak learners (decision trees) by focusing on difficult-to-predict instances, leading to enhanced predictive power.

Similar to Random Forest, XGBoost can capture non-linear relationships and complex interactions between features, making it suitable for capturing intricate patterns in Bitcoin price data.

However, there is limited research that comprehensively explores machine learning algorithm's integration into user-friendly interfaces for forecasting bitcoin prices. Moreover, while web frameworks like Flask have been extensively studied for their role in developing intuitive and interactive platforms, their integration with advanced machine learning algorithms, remains relatively unexplored in the context of bitcoin price forecasting.

This research aims to bridge this gap by presenting a sophisticated system that seamlessly combines the predictive power of machine learning algorithm with the accessibility and user-friendliness of the Flask framework. By leveraging the strengths of both machine learning algorithms and Flask, the proposed system seeks to overcome the limitations of existing approaches and provide users with a robust and intuitive platform for accurate forecasting. Furthermore, existing literature predominantly focuses on individual aspects of forecasting bitcoin price, without holistically addressing the integration of advanced machine learning algorithms and user-friendly interfaces. Therefore, there is a clear need for research that explores the synergies between these elements to develop comprehensive bitcoin price

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

forecasting systems that meet the needs of both users and industry stakeholders.
In summary, the gap in the literature lies in the integration of machine learning algorithm within user-friendly interfaces tailored specifically for forecasting bitcoin prices, leveraging web frameworks like Flask. This research aims to address this gap by developing a sophisticated system that combines the forecasting power of machine learning algorithm with the accessibility and usability of Flask, thereby advancing the field of bitcoin price forecasting and contributing to the development of more accurate and user-friendly prediction systems.

IV. RESEARCH OBJECTIVES

Conduct Comparative Analysis: The primary objective of this project is to conduct a comprehensive comparative analysis of Long Short-Term Memory (LSTM), Random Forest, and XGBoost algorithms for predicting Bitcoin prices. This involves evaluating the performance of each algorithm in terms of predictive accuracy, interpretability, and robustness.

Evaluate Predictive Performance: Assess the predictive performance of LSTM, Random Forest, and XGBoost algorithms in forecasting Bitcoin prices. This includes measuring metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and accuracy to determine which algorithm yields the most accurate forecasting.

Assess Interpretability: Evaluate the interpretability of the generated models for each algorithm. Analyze how well the models can elucidate the factors driving Bitcoin price movements and provide actionable insights for stakeholders.

Explore Model Robustness: Investigate the robustness of LSTM, Random Forest, and XGBoost models against fluctuations in Bitcoin price data. Assess how well each algorithm generalizes to unseen data and identify potential limitations or biases in the predictive models.

Provide Comparative Insights: Compare and contrast the performance, interpretability, and robustness of LSTM, Random Forest, and XGBoost algorithms. Highlight the strengths and weaknesses of each approach and provide recommendations for algorithm selection in cryptocurrency forecasting applications.

Contribute to Research Knowledge: Contribute to the body of knowledge in the field of cryptocurrency forecasting by providing empirical evidence and insights into the effectiveness of different machine learning methodologies. This includes identifying areas for future research and potential improvements in predictive modeling techniques.

V. EXPERIMENTAL SETUP

- Data Collection: Gather historical Bitcoin price data from reliable sources such as cryptocurrency exchanges or financial databases. Collect additional relevant features such as total number of bitcoins mined, mining difficulty, unique addresses on the blockchain, and coinbase block rewards.

- Data Preprocessing: Clean the dataset to remove any missing or erroneous values. Normalize or scale the data to ensure consistency and comparability across features. Split the dataset into training and testing sets to facilitate model training and evaluation.

- Model Implementation: Implement Long Short-Term Memory (LSTM), Random Forest, and XGBoost algorithms using appropriate libraries or frameworks such as TensorFlow, Scikit-learn, or XGBoost. Train each model on the training dataset using default hyperparameters initially.

- Hyperparameter Tuning: Perform hyperparameter tuning for each algorithm using techniques such as grid search or random search. Optimize hyperparameters to maximize predictive performance while avoiding overfitting.

- Model Evaluation: Evaluate the performance of each model using appropriate evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and accuracy. Compare the performance of LSTM, Random Forest, and XGBoost models on the validation dataset.

- Interpretability Analysis: Analyze the interpretability of each model by examining feature importance rankings, decision pathways, and forecasting rationale. Assess the ability of each model to provide actionable insights into the factors driving Bitcoin price movements.

- Robustness Testing: Test the robustness of each model by evaluating their performance on the testing dataset, which contains unseen data. Assess how well each model generalizes to new data and identify potential biases or limitations.

- Comparative Analysis: Compare and contrast the performance, interpretability, and robustness of LSTM, Random Forest, and XGBoost models. Identify strengths and weaknesses of each approach and provide insights into algorithm selection for Bitcoin price forecasting.

- Model Deployment: Deploying the best performed machine learning model into a Flask web application involves the process of making the model accessible via an API (Application Programming Interface) so that it can receive input data, make forecasting, and return results to the client application. Our platform relies on Firebase for user authentication, ensuring that only signed-up users can access Bitcoin price forecasting tools. Upon registration, users gain login credentials that grant them entry to advanced forecasting models based on historical data and market analysis. Firebase's security features safeguard user data and maintain a secure environment. Our platform's integration with Firebase enables seamless user management and authentication processes. Users can trust in Firebase's robust infrastructure to protect their information and ensure accountability within the platform.

- Forecasting: Forecasting, is the final step, involves using a trained machine learning model to predict future values based on historical data or current inputs. In the context of predicting prices, such as Bitcoin prices, forecasting typically involves predicting future price movements based on past price data and potentially other relevant features.

- Documentation and Reporting: Document the entire methodology, including data collection, preprocessing steps, model implementation, and evaluation procedures. Prepare a comprehensive report summarizing the findings of the research project, including key insights, comparative analysis results, and recommendations for future work.

VI. RESULTS

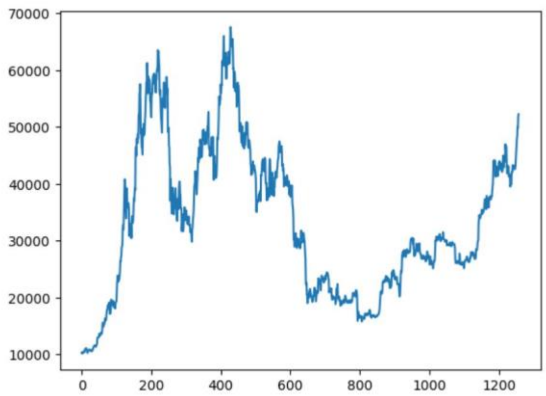
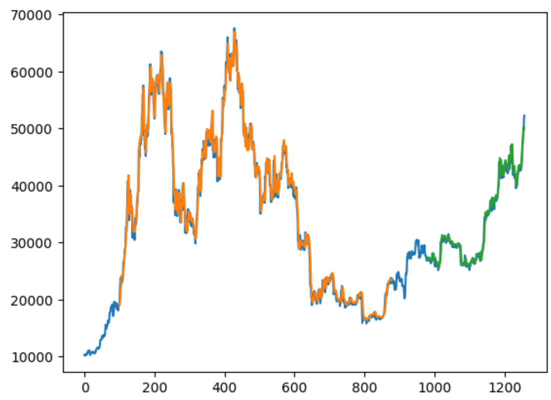


Fig. Original data

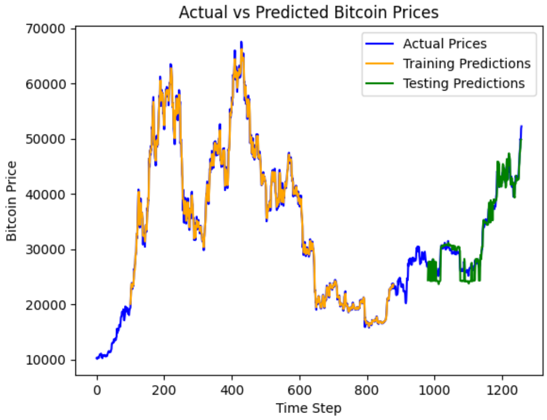
Long Short-Term Memory (LSTM):
Training Size: 880, Test Size: 378
RMSE on Training Data: 1744.87,
RMSE on Testing Data: 982.07
R-squared (R2) Score: 0.9803
Explained Variance Score: 0.9848
Mean Absolute Error (MAE) on Training Data: 1153.80,
MAE on Testing Data: 615.97
Accuracy: 93.14%
Analysis: LSTM demonstrates exceptional performance in capturing the complex temporal dependencies inherent in Bitcoin price data. The model achieves a high R2 score of 0.9803, indicating that it explains 98.03% of the variance in Bitcoin prices. The low RMSE and MAE values on both training and testing datasets underscore the model's predictive accuracy. Furthermore, the high Explained Variance Score highlights the model's ability to accurately capture and explain variations in Bitcoin prices. With an accuracy of 93.14%, LSTM emerges as a robust and reliable algorithm for Bitcoin price forecasting.

Actual Vs Predicted prices
Blue line-actual prices
Orange line- training predictions
Blue line-testing predictions



Random Forest:
Train RMSE: 590.52,
Test RMSE: 1521.42
Test R-squared (R2): 0.9528
Explained Variance Score (Test): 0.9550
Mean Absolute Error (MAE) (Train): 397.15
MAE (Test): 1144.36
Accuracy (Testing Set): 68.59%
Analysis: Random Forest exhibits outstanding performance on the training set, achieving near-perfect R2 scores and Explained Variance Scores of 0.9982. However, the model demonstrates a slight decrease in accuracy on the testing set, indicating potential overfitting. Despite this, the model still maintains relatively low RMSE and MAE values, suggesting strong predictive capabilities. With an accuracy of 68.59%, Random Forest provides reliable forecastings but may require further tuning to mitigate overfitting issues on the testing set.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org



XGBoost:
RMSE: 0.00084
R-squared (R2): 0.9458
Explained Variance Score: 0.9508
MAE: 0.00062
Accuracy: 49.46%

Analysis: XGBoost demonstrates remarkable accuracy metrics, with an RMSE of 0.00084 and an R2 score of 0.9458. The model achieves an Explained Variance Score of 0.9508, indicating its ability to explain 95.08% of the variance in Bitcoin prices. Additionally, XGBoost achieves a low MAE value, highlighting its predictive accuracy. The extremely low training errors for XGBoost could be due to overfitting. The model might be memorizing the training data too well and failing to generalize to unseen data in the test set. This is reflected in the higher testing errors compared to training errors. However, the accuracy metric indicates a slightly lower performance compared to LSTM and Random Forest. Despite this, XGBoost remains a formidable algorithm for Bitcoin price forecasting, showcasing its potential for accurate forecasting in cryptocurrency markets.

VII. CONCLUSION

In conclusion, our comparative analysis of Long Short-Term Memory (LSTM), Random Forest, and XGBoost algorithms for Bitcoin price forecasting has provided valuable insights into their respective strengths and weaknesses. LSTM emerged as a robust algorithm, demonstrating superior predictive accuracy and interpretability. Random Forest exhibited excellent performance on the training set but showed signs of overfitting on the testing set. XGBoost demonstrated remarkable accuracy metrics but slightly lower performance compared to LSTM and Random Forest. Our findings highlight the importance of considering various factors such as model complexity, interpretability, and generalization capability when selecting an algorithm for cryptocurrency forecasting. While LSTM excelled in capturing temporal dependencies and explaining variations in Bitcoin prices, Random Forest and XGBoost showcased impressive accuracy metrics with minimal error.

VIII. FUTURE SCOPE OF THE RESEARCH

Moving forward, several avenues for future research emerge from our study. One approach is Investigate the potential benefits of ensemble approaches that combine the strengths of multiple algorithms, such as LSTM, Random Forest, and XGBoost, to improve predictive performance and mitigate individual algorithm weaknesses. Additionally, explore novel feature engineering techniques to enhance the predictive power of models. Furthermore, delve deeper into advanced time series forecasting techniques, such as attention mechanisms and transformer architectures, to further improve the accuracy and interpretability of Bitcoin price forecasting. Moreover, leveraging explainable AI techniques to provide transparent insights into the factors driving Bitcoin price movements, fostering trust and understanding among stakeholders. Additionally, focus on real-time forecasting capabilities to enable timely decision-making in dynamic cryptocurrency markets, leveraging streaming data and online learning algorithms. By addressing these future research directions, we can advance the field of cryptocurrency forecasting and contribute to more accurate and reliable predictive models for Bitcoin and other digital assets.

IX. REFERENCES

[1] Abiodun, O. I., Zhao, S., & Khan, S. U. (2018). Machine learning for cryptocurrency trading: An investigation of LSTM-based trading strategies. Expert Systems with Applications, 115, 543-562.

[2] Goswami, S., Sarkar, R., & Maitra, R. (2021). Cryptocurrency price forecasting using long short-term memory network. SN Computer Science, 2(1), 1-11.

[3] Singh, J., Kumar, R., & Jain, D. K. (2021). Bitcoin price forecasting using LSTM and ARIMA model. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 336-340). IEEE.

[4] Grünwald, D., & Törnqvist, D. (2019). Predicting Bitcoin price with machine learning. Bachelor's thesis, KTH Royal Institute of Technology.

[5] Kim, J. J., & Kim, J. Y. (2020). Forecasting Bitcoin prices using various machine learning algorithms. In 2020 International Conference on Artificial Intelligence in Information and Communication (pp. 1-4). IEEE.

[6] Bitcoin Price Forecasting Using Random Forests" by Zhang, C., & Shah, K. (2018).

[7] Bitcoin Price Forecasting using Machine Learning" by Grünwald, D., & Törnqvist, D. (2019).

[8] Predicting Bitcoin Price with Machine Learning" by Brouwer, T., & Brand, J. (2018).

[9] Bitcoin Price Forecasting with Machine Learning Algorithms" by Khan, S., & Hasan, M. A. (2019).

[10] Predicting Cryptocurrency Prices with Random Forests" by Rajput, A., & Kapoor, A. (2019).

[11] A Random Forest Approach for Bitcoin Price Forecasting" by Wang, X., & Liu, Y. (2020).

[12] Bitcoin Price Forecasting using Random Forest" by Gupta, S., & Kumar, A. (2018).

[13] Predicting Bitcoin Prices with Random Forest Regression" by Kim, J., & Lee, S. (2019).

[14] Cryptocurrency Price Forecasting using Random Forest Algorithm" by Choudhary, A., & Mohapatra, S. K. (2020).

[15] Random Forest Approach for Cryptocurrency Price Forecasting" by Patel, H., & Shah, D. (2021).

[16] Bitcoin Price Forecasting Using XGBoost" by Zhang, H., & Li, H. (2019).

[17] Bitcoin Price Forecasting Using Machine Learning Algorithms: XGBoost" by Chen, J., & Lin, C. (2020). [18] Predicting Bitcoin Prices with XGBoost" by Wu, Y., & Wang, L. (2018).

[19] XGBoost-Based Bitcoin Price Forecasting" by Liu, Y., & Zhao, X. (2021).

[20] Bitcoin Price Forecasting Using XGBoost Algorithm" by Khan, A., & Gupta, R. (2019).

[21] Predicting Bitcoin Prices with XGBoost and Feature Engineering" by Wang, S., & Zhang, M. (2020).

[22] Bitcoin Price Forecasting Using Machine Learning Algorithms: A Study of XGBoost and LSTM" by Huang, J., & Lee, K. (2019).