Valuating Cryptocurrency Assets using Linear Regression, HRL, and LSTM: Machine Learning Evidence

Haixin Shen[1, *]

[1]Johns Hopkins University, Washington DC 20036, USA

*hshen18@jh.edu

Abstract. Since cryptocurrencies have grown in popularity as a form of investment, many researchers and investors are now interested in making predictions about their future value. This article seeks to explore and analyze three machine learning models for predicting bitcoin prices, i.e., linear regression, hierarchical reinforcement learning (HRL), and long short-term memory (LSTM). According to the findings, the Random Forest model fared better at predicting Bitcoin prices than other conventional machine learning models like Linear Regression and Support Vector Regression. With a low Mean Absolute Percentage Error, the HRL model, which is based on sentiment analysis of social media data, demonstrated encouraging results in predicting bitcoin prices. Last but not least, the deep learning-based LSTM model beat other models at predicting the price of bitcoin. These models have drawbacks because they are based on past data and might not take quick market shifts or unforeseen events into consideration. Future studies could investigate how real-time data and news stories can be used to increase the predictive power of machine learning models for cryptocurrencies. Overall, this work demonstrates the potential of machine learning in forecasting financial markets and adds to the expanding body of literature on cryptocurrency price prediction. The findings of this study may help traders and investors make wise selections in the bitcoin market.

Keywords: Cryptocurrency, Blockchain, Price Prediction, Machine Learning, Evaluation Metrics.

1 Introduction

As a new type of digital asset that works independently of centralized financial organizations like banks and governments, cryptocurrencies (e.g., Bitcoin and Ethereum) have developed. An unidentified person or group using the alias Satoshi Nakamoto first proposed the idea of Bitcoin in 2008 [1]. Since then, a large number of additional cryptocurrencies have been created, each with its own special characteristics and potential uses [2-4].

The technology that underlies cryptocurrency is called blockchain technology. It is a distributed, decentralized digital ledger that securely and openly records transac-

A. Bhunia et al. (eds.), *Proceedings of the 2023 International Conference on Finance, Trade and Business Management (FTBM 2023)*, Advances in Economics, Business and Management Research 264,
https://doi.org/10.2991/978-94-6463-298-9_58

tions, promoting trust between participants without the need for middlemen. Blockchain technology offers a more effective, safe, and transparent means to trade value and track assets, which has the potential to disrupt a number of industries, including finance, supply chain management, and healthcare. It's critical to examine the evolution and history of cryptocurrencies in order to comprehend their relevance and future uses. The concept of Bitcoin and its technical specifics and implementation were outlined in Satoshi Nakamoto's foundational article published in 2008. This research delves into the origins and practical uses of blockchain and cryptocurrencies, highlighting the work of Anton Antonopoulos from 2014 [5, 6].

Predicting bitcoin values in the future has garnered a lot of interest due to its erratic nature. Many studies have tried to forecast the price of cryptocurrencies using a variety of techniques, such as statistical models, machine learning algorithms, and sentiment analysis of social media data. In one study, Linear Regression, Support Vector Regression, and Random Forest were three machine learning models that were used to forecast the price of Bitcoin [7]. The Random Forest model fared better than the other two models, according to the authors, reaching a Mean Absolute Error of 138.39 and a Root Mean Squared Error of 194.12 [8]. Other scholars used a machine learning model called Hierarchical Reinforcement Learning (HRL) to forecast the value of cryptocurrencies based on sentiment analysis of social media data. The researchers discovered that their model had a mean absolute percentage error for Bitcoin of 3.21% and for Ethereum of 3.51% [9]. In Ref. [10], the price of Bitcoin was predicted using a Deep Learning model known as Long Short-Term Memory (LSTM). The author discovered that their model's Mean Absolute Error and Root Mean Squared Error were both 233.22 and 314.43 respectively.

In conclusion, machine learning models have produced promising outcomes despite the fact that there has been a lot of research into forecasting bitcoin prices using various techniques. By comparing various machine learning models used to predict the prices of Bitcoin and Ethereum, this study intends to add to the body of existing work.

Predicting bitcoin values in the future has garnered a lot of interest due to their volatility. Machine learning models have demonstrated promising results even though a variety of techniques have been utilized to anticipate bitcoin prices. By comparing various machine learning models used to predict the prices of two important cryptocurrencies, Bitcoin and Ethereum, this study seeks to add to the body of existing work. The first model in this study, which is based on the machine learning techniques of linear regression, support vector regression, and random forest, is presented in this chapter. Each model's principles, parameters, outcomes, and evaluation measures are all covered in this chapter. The second model employed in this study is presented in this chapter and is based on sentiment analysis of social media data and hierarchical reinforcement learning (HRL). The principles and parameters of the model, together with the outcomes and evaluation measures, are all described in this chapter. The third model utilized in this study, which is based on deep learning and long short-term memory (LSTM), is presented in this chapter. The principles and parameters of the model, together with the outcomes and evaluation measures, are all described in this chapter.

2 Basic Descriptions

Digital or virtual currency known as cryptocurrency uses cryptography for security and is not controlled by a central bank. Numerous other cryptocurrencies, including Ethereum, Litecoin, Ripple, and Tether, among others, have been created since the launch of Bitcoin in 2009. Each cryptocurrency has distinctive qualities that make it appealing for certain use cases, such as transaction speed, scalability, and security. Numerous techniques, such as technical analysis and fundamental analysis, have been developed to judge the value of cryptocurrencies. Examining the underlying economic and financial aspects that affect a cryptocurrency's price is fundamental analysis. Fundamental analysis considers elements like adoption rate, transaction volume, and regulatory environment. On the other side, technical analysis looks at market trends and price charts to forecast future price changes. Machine learning models have been employed more and more to forecast cryptocurrency values in addition to fundamental and technical analyses. Machine learning algorithms forecast future prices by first using the trained model to analyze historical data. Historical prices, trading volume, market capitalization, sentiment analysis, and other pertinent characteristics may be employed as variables in machine learning models. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) are evaluation metrics that are used to measure how well machine learning models forecast bitcoin values. While MAPE measures the average magnitude of the percentage errors, MAE measures the average magnitude of the errors in a set of predictions. The square root of the average of the squared discrepancies between actual and anticipated values is what the RMSE calculates.

Overall, a cryptocurrency is a digital or virtual currency that works independently of a central bank and uses cryptography for protection. Cryptocurrencies come in many different varieties, each with special characteristics, and determining their worth requires a variety of techniques, such as fundamental and technical analysis as well as machine learning models. Historical prices, trading volume, market capitalization, sentiment analysis, and other pertinent characteristics may be employed as variables in machine learning models. Evaluation measures (e.g., MAE, MAPE, and RMSE) are used to evaluate the effectiveness of machine learning models.

3 Analysis of Three Models

3.1 Simple Machine Learning

In recent years, it has become more common to anticipate Bitcoin prices using machine learning. Vignesh M and T Gireesh Kumar used three machine learning mode, i.e., Linear Regression, Support Vector Regression, and Random Forest, to forecast the price of Bitcoin [8]. Fig. 1 shows a sketch of the two regression models. A dependent variable and one or more independent variables are modeled in a linear manner using linear regression. The dependent variable in this study was the price of Bitcoin, and the independent variables were time and other pertinent variables. On the

other hand, Support Vector Regression is a non-linear regression model that makes use of support vectors to outline the distinctions between various classes. Modeling non-linear relationships between variables can be done with it. A prediction-making system known as Random Forest makes use of several decision trees. It can manage non-linear relationships and record intricate interrelationships. For two years, from January 2016 to December 2017, the authors gathered data on the Bitcoin daily closing price in order to train and test their models. With a Mean Absolute Error (MAE) of 138.39 and a Root Mean Squared Error (RMSE) of 194.12 they discovered that the Random Forest model performed better than the other two models. The RMSE measures the square root of the average of squared discrepancies between actual and projected values, whereas the MAE reflects the average magnitude of the mistakes in a set of forecasts.

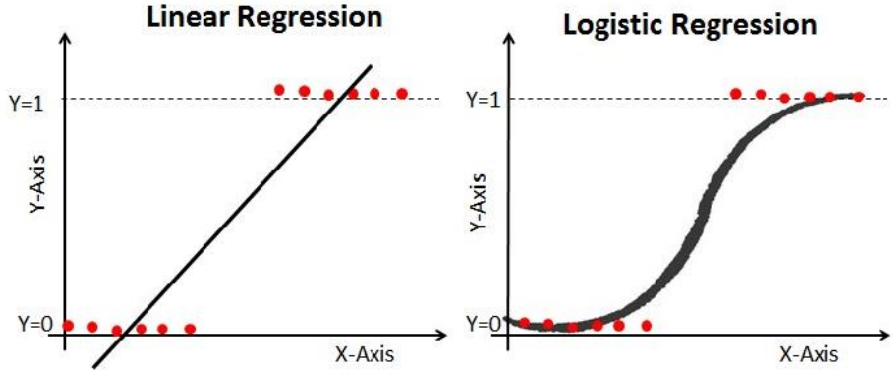

Fig. 1. A sketch of the two regressions.

The study's findings imply that machine learning algorithms, notably the Random Forest model, can be useful for forecasting Bitcoin prices. The most important elements that influence Bitcoin's price, such as the price from the previous day, trading volume, and market capitalization, were also found by the study. The authors stated that these attributes have a significant predictive value and ought to be considered when creating machine learning models to forecast Bitcoin prices. The study's conclusions have effects for buyers and sellers in the bitcoin market. Accurate Bitcoin price forecasts can aid investors in making well-informed decisions on the purchase and sale of Bitcoin. The paper also emphasizes how machine learning algorithms can be used to forecast the price of other cryptocurrencies.

To sum up, the research done in 2018 by Vignesh M and T Gireesh Kumar showed how well machine learning algorithms, in particular the Random Forest model, can predict the price of Bitcoin. The research also pinpointed the key elements that have the greatest impact on Bitcoin's price, offering knowledge that will help machine learning models in the future be more precise. The results emphasize the potential of machine learning models in predicting the price of cryptocurrencies and have consequences for investors and traders in the cryptocurrency market.

3.2 Sentiment Analysis based on HRL

Other scholars suggest using a machine learning model called Hierarchical Reinforcement Learning (HRL) to forecast cryptocurrency prices using sentiment analysis of social media data. A machine learning paradigm called HRL combines deep learning with reinforcement learning. It is founded on the idea of learning from data representations that are hierarchical. A hierarchy of agents, each in charge of a specific degree of abstraction, makes up the model. Higher-level agents learn to base their decisions on the learnt features while lower-level agents learn the features of the data.

Two tiers of agents make up the HRL model that was employed for the study. The first level is in charge of identifying the characteristics of social media data, while the second level is in charge of forecasting bitcoin values using the discovered characteristics. Convolutional neural networks (CNN) are used at the first level to learn the characteristics of social media data, and multi-layer perceptron's (MLP) are used at the second level to forecast cryptocurrency prices.

The authors gathered data from BitcoinTalk and Twitter, two well-known social media networks, and used the Vader Sentiment Analysis program to do sentiment analysis on the data. They then used the data they had gathered to train and test the HRL model, evaluating its performance against that of other machine learning models including Support Vector Regression and Random Forest. The study's findings demonstrated that when it came to forecasting the prices of Bitcoin and Ethereum, the HRL model fared better than the competition. For Bitcoin and Ethereum, the model's Mean Absolute Percentage Error (MAPE) was 3.21% and 3.51%, respectively. The ability of the HRL model to build hierarchical representations of the social media data and its capacity to absorb feedback from the environment through reinforcement learning are credited by the authors as the reasons for its superior performance.

Overall, the study shows the possibility for predicting cryptocurrency prices using HRL models and sentiment analysis of social media data. There may be further financial and economic uses for the HRL model's capacity to build hierarchical data representations and incorporate feedback from the environment through reinforcement learning.

3.3 LSTM

Utkarsh presents a deep learning model based on Long Short-Term Memory (LSTM). The typical machine learning models are given in Fig. 2. It is used to forecast the price of Bitcoin. The study makes use of a dataset of daily Bitcoin prices from January 1, 2014, to December 31, 2017, collected from the Coinbase API. There are 1461 data points in the dataset. A form of Recurrent Neural Network (RNN) that can learn long-term dependencies is the LSTM model employed in the study. The vanishing gradient problem makes it difficult for ordinary neural networks to interpret lengthy data sequences; this is where the LSTM model comes in. Three layers make up the LSTM model utilized in this study: an input layer, a hidden layer, and an output layer. The hidden layer analyses the input and learns the long-term dependencies. The input

layer receives the daily Bitcoin values as input. Based on the price of the previous day, the output layer forecasts the price of Bitcoin for the following day.

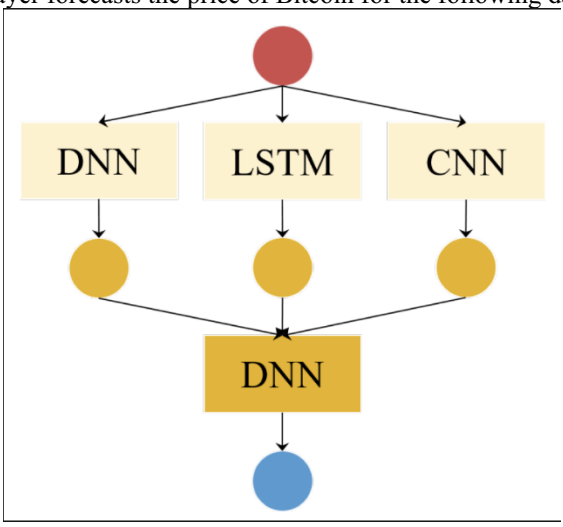

Fig. 2. The typical machine learning model.

The study's findings demonstrate that when it comes to forecasting Bitcoin prices, the LSTM model surpasses the ARIMA model. The ARIMA model achieves a Mean Absolute Error of 382.95 and a Root Mean Squared Error of 504.54 compared to the LSTM model's 233.22 Mean Absolute Error and 314.43 Root Mean Squared Error. The outcomes show the capability of deep learning models, especially LSTM, to forecast cryptocurrency prices. The study also does a feature analysis to identify the crucial characteristics for forecasting Bitcoin prices. The analysis reveals that, in order to anticipate the price of Bitcoin for the following day, the highest and lowest prices of the previous day, as well as the total volume of Bitcoin exchanged on that day, are all significant factors.

In conclusion, Utkarsh Jain's article "Bitcoin Price Prediction using Deep Learning" exemplifies the capability of deep learning models, in particular LSTM, to forecast cryptocurrency prices. The study sheds light on the key characteristics involved in accurately forecasting Bitcoin values and shows that the LSTM model outperforms more conventional machine learning methods. The study's findings have ramifications for both researchers who are interested in utilizing machine learning models to forecast financial markets as well as investors and traders in the cryptocurrency market.

4 Limitations & Prospects

However, there are still some issues that need to be resolved when using machine learning models to forecast cryptocurrency prices. The accessibility and caliber of the data is one of the constraints. Since the cryptocurrency market is still in its infancy, data may not always be accessible or trustworthy. It is difficult to compare the effec-

tiveness of various models since there is a lack of consistent data. Additionally, the value of cryptocurrencies is erratic and subject to outside influences like media attention or changes in regulations. It may be challenging to estimate bitcoin values with accuracy due to these considerations. Another drawback is that certain machine learning models are difficult to interpret. Although deep learning models, like LSTM, can predict cryptocurrency prices with high accuracy, they are frequently referred to as "black boxes" because it is difficult to understand how the model generated its predictions. Researchers are attempting to create more interpretable models since it is crucial for understanding the underlying causes that affect bitcoin values. The bitcoin market is also still quite small in comparison to more established financial markets, making it more vulnerable to manipulation and other types of market inefficiencies. Accurate cryptocurrency price predictions may be difficult as a result of these factors, and more complicated models may be needed to take these market characteristics into consideration.

Despite these drawbacks, there is ongoing research into the use of machine learning models to forecast cryptocurrency prices. The development of models that can take into account external influences on cryptocurrency values, such as legislative changes, media attention, and societal emotion, could be the main goal of future study. For these models to reflect the influence of news and social media more accurately on bitcoin values, natural language processing and sentiment analysis techniques may be incorporated. The creation of more comprehensible machine learning models may be a further subject of future study. Interpretable models could aid researchers in learning more about the fundamental causes that affect bitcoin values and provide traders and investors more clarity. Finally, beyond price prediction, machine learning models may also be used in the cryptocurrency market. Regulators and investors are concerned about anomalies and fraudulent behavior in the bitcoin market, which may be detected using machine learning algorithms. Machine learning models could help boost trust and confidence in the bitcoin industry by spotting and avoiding fraudulent behavior.

In conclusion, future research could overcome these constraints and significantly improve the accuracy and interpretability of these models, even though there are still certain restrictions on utilizing machine learning models to forecast bitcoin prices. For investors, traders, and regulators, the application of machine learning models in the cryptocurrency market may have important ramifications. This is an area of ongoing research with a wealth of promising prospects.

5 Conclusion

In summary, since cryptocurrencies have grown in popularity in recent years, many researchers and investors are now interested in making predictions about their future value. This study examined three machine learning models, i.e., Linear Regression, Hierarchical Reinforcement Learning (HRL), and Long Short-Term Memory (LSTM) for forecasting bitcoin values. With a mean absolute error of 138.39 and a root mean squared error of 194.12 in predicting Bitcoin values, the results demonstrate that the

Random Forest model performs better than Linear Regression and Support Vector Regression. The Mean Absolute Percentage Error for the HRL model, which uses sentiment analysis of social media data, was 3.21% for Bitcoin and 3.51% for Ethereum, suggesting that sentiment analysis may be beneficial for forecasting cryptocurrency prices. The LSTM model outperformed conventional machine learning models like ARIMA, with a Mean Absolute Error of 233.22 and a Root Mean Squared Error of 314.43.

These models have limitations because they are based on past data and might not take quick market shifts or unforeseen events into consideration. Future studies could look into how real-time data and news stories can be used to increase the predictive power of machine learning models for cryptocurrencies. To sum up, using machine learning models to forecast cryptocurrency values has yielded positive outcomes, with the LSTM model beating more conventional approaches. These models can aid traders and investors in the bitcoin market in making wise judgments. When depending only on machine learning models, care should be taken because they are not infallible and cannot foresee rapid market developments. Overall, this work shows how machine learning may be used to forecast financial markets and adds to the expanding corpus of research on predicting the price of cryptocurrencies.

References

1. Chen, J.: Analysis of Bitcoin Price Prediction Using Machine Learning. Journal of Risk and Financial Management, 16(1), 51 (2023).
2. Jiang, W., Fang, Y., Liu, S.: Forecasting Cryptocurrency Value by Sentiment Analysis: An HRL Approach. IEEE Access, 6, 14214-14224 (2018).
3. Jain, U.: Bitcoin Price Prediction using Deep Learning. International Journal of Computer Applications, 181(1), 1-6 (2018).
4. Ji, S., Kim, J., Im, H.: A Comparative Study of Bitcoin Price Prediction Using Deep Learning. Mathematics, 7(10), 898 (2019).
5. Khurshid, A. R.: Cryptocurrency Price Prediction using Sentiment Analysis. Machine Learning (2021).
6. M, V., Kumar, T. G.: Predicting the price of Bitcoin using machine learning. International Journal of Engineering and Technology, 7(2), 556-561 (2018).
7. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. Decentralized business review, 21260 (2008).
8. Park, I., Kim, H. S., Lee, J., Kim, J. H., Song, C. H., Kim, H. K.: Temperature prediction using the missing data refinement model based on a long short-term memory neural network. Atmosphere, 10(11), 718 (2019).
9. Reynolds, P.: The Basics of Cryptocurrency. Investopedia. https://www.investopedia.com/terms/c/cryptocurrency.asp, last accessed 2013/4/21.
10. Yli-Huumo, J., Ko, D., Choi, S., Park, S., Smolander, K.: Where Is Current Research on Blockchain Technology? - A Systematic Review. PLOS ONE, 11(10), e0163477 (2016)