(REVIEW ARTICLE)

Analyzing the machine Learning methods to predict Bitcoin Pricing

R. Siva [1], Bharathi Subrahmanian [2, *] and Chaturya. K [2]

[1] Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, India.
[2] Final Year B.Tech (CSE – AI&ML) Students, Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, India.

Abstract

A cryptocurrency is a form of digital money that uses the blockchain technology and cryptography to protect the information about transactions and exchange made on the digital market. A cryptocurrency like Bitcoin consists of a big network that has many peers working on it and every peer has a record of the whole history that contains all the transactions that ever happened. Bitcoin is the most popular cryptocurrency. Bitcoin has attracted a lot of attention from individual and institutional investors. The purpose of this paper is to analyse the machine learning methods to predict Bitcoin pricing. Machine learning and its associated fields have made notable advances in recent years. Machine learning techniques is used in different areas of science particularly cryptocurrency price forecasting. Using this machine learning model, we can predict the price direction of Bitcoin. Machine learning methods have been demonstrated to be effective in predicting bitcoin prices. Few machine learning models to predict the Bitcoin price are Linear Regression, Logistic Regression, Bayesian Regression, Support Vendor Machine, Random Forest, Neural Network were discussed. Every method of machine learning has its own advantages and disadvantages, however from the literature analysis it is understood that the Artificial Neural Network and Support Vendor Machine have the highest effectiveness rate. Machine learning methods have higher prediction accuracy than parametric regression approaches.

Keyword: Machine learning; Bitcoin; Cryptocurrency; Prediction; Digital Money

1. Introduction

Cryptocurrency is a private system that enables trades between individuals without a central and intermediate agency (Erfanian, et al., 2022). The term cryptocurrency entered public usage with the surge of Bitcoin in 2008, a protocol aimed at enabling a network of people connected via peer-to-peer digital communications infrastructure to issue digital tokens and transfer them between themselves whilst securing the process through cryptography (Nakamoto, 2008). Cryptocurrency represents digital money that you can't feel, you can't hold it in your hands (Milutinović, 2018). The cryptocurrency, denoted by bitcoin or BTC, can be accepted as a payment for goods and services or bought either from other people or directly from exchanges/vending machines. These bitcoins can be transacted via software, apps, or various online platforms that provide wallets (Chuen, et al., 2017). Bitcoin, invented in 2008 to solve the inherent weakness of the trust-based model of transactions and initially defined as a purely peer-to-peer electronic cash system, has become an asset or commodity-like product traded in more than 16,000 markets around the world (Nakamoto, 2008). Bitcoin enabled by the blockchain technology and allows for peer-to-peer transactions secured by cryptography (Beck, et al., 2017).

* Corresponding author: Bharathi Subrahmanian

1.1. Purpose

Much research has predicted Bitcoin price in two ways: empirical analysis and analysis of robust machine learning algorithms. Machine learning algorithms have been widely applied to make accurate predictions. The purpose of this paper is to analyse the machine learning methods to predict Bitcoin pricing. Literature relating to price prediction of Bitcoin through machine learning methods were analysed and discussed.

2. Literature Review

One of the biggest inventions of technology in the modern world, that has attracted a lot of public attention, is the phenomenon of cryptocurrencies. This is the biggest technology invention in the last ten years. So, cryptocurrencies became very popular through a very short period (Milutinović, 2018). Bitcoin is primarily designed to eliminate the need of financial institutions or 'trusted third-party' entities. After the first cryptocurrency (Bitcoin) in 2009, there were a lot of other cryptocurrencies appeared on the market, but they were called the altcoins, as they represented the mix of Bitcoin alternatives. Bitcoin does not have a centralized system, no one can control it entirely, like in electronic banking systems (Milutinović, 2018). Cryptocurrencies are also considered to be a financial asset by many users, and a considerable amount of research that has been applied to financial markets can also be applied to this field.

Bitcoin does this by eliminating the possibility of fraud, increasing efficiencies, and providing objective proof-of-work to guarantee validity and security in any transaction (Nakamoto, 2008). The use of a public ledger as well as digital signatures allow for a secure and anonymous transaction without the need for trust, as the public network of nodes validates transactions through finding a consensus among most nodes. Thus far, the primary use cases for Bitcoin revolve around increasing efficiencies and eliminating unnecessary time and costs that arrive from using multiple trusted third parties to facilitate transactions (D'Alfonso, et al., 2016). In early 2009, Bitcoin (BTC) was valued for the first time at US$0.08, it touched $1110 in 2013. Due to high volatility and massive fluctuations in prices in cryptocurrencies, accurate price predictions are a complex and challenging task. That is mainly because the costs of cryptocurrency move unpredictably and chaotically. (Erfanian, et al., 2022).

The most important thing about cryptocurrencies, and especially the Bitcoin is that it can't be controlled by no server or any authority, it is completely safe and there is a bigger chance for the humanity to be wiped from the face of the Earth, than for a transaction or a user on this platform can be revealed (Milutinović, 2018). The blockchain technology used by cryptocurrency, such as Bitcoin, is an open distributed ledger that records transactions. This solves the double-spending problem and does not require a trusted third party (Chuen, et al., 2017). The idea was to create a digital cash system, that will work on the principle of peer-to-peer network and files would be shared like this. And that is how it became the cryptocurrency. So, there is no server that controls this platform, but every peer must own a list that has all the transactions so it can see if some transactions in the future are valid (Milutinović, 2018). There are more than thousand different cryptocurrencies that can be bought, but Bitcoin, Ethereum, Ripple, Litecoin, Monero, Ethereum Classic, NEM, Dash, IOAT, Waves, Augur are some of the most valuable that have the biggest capital on the market (Milutinović, 2018).

Simply machine Learning is a science that teaches machines how to learn new things from themselves. Machine learning is a subset of artificial intelligence that focuses on learning from the database to build intelligent computer systems (Yu, et al., 2021 & Erfanian, et al., 2022). Machine learning is a growing discipline of computing algorithms that are aimed to mimic human intelligence by learning from their surroundings. Machine learning algorithms are always changing and improving (Hilpisch, 2020). Machine learning and its associated fields have made notable advances in recent years (Jordan, M. I., & Mitchell, 2015). Some of these technological breakthroughs have led to the creation or improvement of products that are used by billions of people worldwide (Ashayer, 2019). Since the advent of machine learning research and its related technologies, many researchers have focused their efforts on applying these new techniques on financial markets (Ashayer, 2019).

In recent years, machine learning (ML) has altered the ways in which these predictions are generated (Rovira et al. 2017). ML techniques can extract complex patterns from vast stores of data (Rosenberg & Krist, 2021). Multiple ML algorithms (e.g., a combination of decision trees) that individually would perform less accurately are often combined to bolster prediction performance. These ML algorithms, called ensemble methods, have the potential to yield more accurate and timely predictions. Machine learning techniques may help bring in some methodology that will lead to better solutions to the problem. In the last several years, there has been an increasing interest in using machine learning techniques in different areas of science (Bertolini, et al., 2021 and Rosenberg & Krist, 2021), particularly cryptocurrency price forecasting (Ashayer, 2019).

3. Machine Learning Methods

Previous studies have predicted Bitcoin price in two ways: empirical analysis and analysis of robust machine learning algorithms. Machine learning algorithms have been widely applied to make accurate predictions in many areas (Chen, et al., 2020 and Nieto, et al., 2018). By learning the details of past instances, machine learning programs and models can be produced that make predictions based on training data. Such algorithms can be replicated for the Bitcoin market, even in the world of cryptocurrency, due to the higher liquidity and volatility caused by the T+0 trading rules (Yermack, 2015). There are several models to predict the price of Bitcoin, few of them were discussed in this paper.

3.1. Linear Regression

A Linear regression is a simple regression model that makes the relationship between two variables. One variable is considered an explanatory variable, and the other variable is considered as dependent (Ali & Shatabda, 2020). Linear Regression (LR) is a predictive model that formulates a line of best fit between a scalar dependent variable and multiple explanatory variables. Linear fit occurs by minimizing the mean squared error between the predicted and actual output (Greaves & Au, 2015).

3.2. Logistic Regression

Logistic Regression is a predictive regression model in which the dependent variable is categorical. In the simple case (as in our problem formulation) where there are only two categories, Logistic Regression uses Maximum Likelihood Estimation to formulate the probabilities in which Logistic Regression will take on a particular class, with an iterative algorithm such as Newton's method used to obtain the fitted model (Greaves & Au, 2015). It uses a logistic function to estimate probabilities of a categorical dependent variable, unlike linear regression which is suitable for continuous variables. Logistic regression uses Maximum Likelihood Estimation to formulate the probabilities (Ashayer, 2019). It used maximum likelihood estimation to formulate the probabilities in which Logistic Regression will take on a particular class, with an iterative algorithm (Poongodi, et al., 2020).

3.3. Bayesian Regression

In Bayesian regression, linear regression is formulated using probability distribution rather than point estimates. Therefore, the response is not estimated as a single value but is assumed to be drawn from a probability distribution. This approach is especially useful when the amount of data is limited, or some prior knowledge can be used in creating the model (Ashayer, 2019). Shah and Zhang (2014) used Bayesian regression in their study to predict the price variations of Bitcoin and create a profitable cryptocurrency trading strategy.

3.4. Support Vector Machine

Support vector regression (SVR) is an emerging nonlinear regression method based on statistical learning theory having a more stable solution than traditional neural network models (Erfanian, et al., 2022). A support vector machine is a kind of machine learning methodology that is applied in binary classification problems (Kumar, et al., 2016 & Chen, et al., 2020). SVM are non-probabilistic binary linear classifiers that are used for classification and regression analysis. SVMs are commonly used in text categorization, image classification, and handwriting recognition. SVMs have gained popularity for their impressive performance in classification (Poongodi, et al., 2020).

3.5. Random Forest

Tree-based ensemble machine learning classifiers are used to predict Bitcoin. Random forest models use an ensemble of decision trees for various tasks to obtain a better classification result and are a popular approach. The use of decision trees (Quinlan, 1986 & Geurts and Louppe, 2011) is one of the basic machine learning methods and is used to solve a wide range of problems in classification. It is basically several classifying decision tree on various samples of the dataset. It is like concept of averaging; it is used to improve the predictive accuracy of the Decision tree and prevent the model from overfitting (Vikram, et al., 2022).

3.6. Neural Network

Neural Network are a family of learning methods inspired by biological neural networks by modelling a system of interconnected neurons, which are tuned based on iterative learning. Feedforward neural networks connect a multidimensional input into one or more hidden layers of neurons before predicting an output. Dropout is used to prevent overfitting of the model. Hidden layers are modelled by affine transformation and final layers are modelled by SoftMax (Poongodi, et al., 2022 & Greaves and Au, 2015).

Bayesian neural networks (BNN) is a transformed Multilayer perceptron (MLP) which is a general term for ANNs in the fields of machine learning (Jang and Lee, 2017). The structure of a BNN is constructed with several processing units classified into three categories: an input layer, an output layer, and one or more hidden layer. Pichl and Kaizoji (2017) performed a volatility analysis on Bitcoin price time-series and used Multilayer Perceptron (MLP) to predict daily log returns. In their analysis, they used an MLP with two hidden layers and utilized the past 10-day moving window for daily log return sampling as their predictors. Multilayer Perceptron (MLP) is a class of feed-forward artificial neural network that has at least three layers of nodes. Each node in an MLP, except the input nodes, is a neuron that uses a nonlinear activation function to operate (Ashayer, 2019).

Jang and Lee (2017) use Bayesian ANNs to forecast daily Bitcoin prices. Blockchain information (trading volume, block size, transactions per block, hash rate, number of transactions, miner's revenue) and economic variables (stock prices of major exchanges, oil prices, VIX, gold prices) are used as features. Bayesian neural networks (BNN) is a transformed Multilayer perceptron (MLP) which is a general term for ANNs in the fields of machine learning. The networks have been successful in many applications such as image recognition, pattern recognition, natural language processing, and financial time series (Murphy, 2012). It becomes known that much effective to represent the complex time series than the conventional linear models, i.e. autoregressive and moving average, etc (Jang and Lee, 2017).

4. Discussion

A cryptocurrency is a digital currency that is managed through cryptography. Bitcoin is the most popular cryptocurrency. When placing orders in an exchange, traders can place market orders or limit orders (Kristensen and Sognefest, 2023). Cryptocurrencies were developed under Blockchain technology. In contrast with the traditional central authority systems wherein the sole control lies under one organization, Blockchain technology has a diversified approach (Erfanian, et al., 2022). Bitcoin has attracted a lot of attention from individual and institutional investors. Accurate Bitcoin price direction forecasts are important for determining the trend in Bitcoin prices and asset allocation (Basher and Sadorsky, 2022). Machine learning methods have been demonstrated to be effective in predicting bitcoin prices. Mallqui and Fernandes (2019) have used machine learning techniques to estimate the timing and the acceptance of cryptocurrencies as financial products to find more precise predictions for the Bitcoin conversion scale. Along these lines, in the first phase of the proposed philosophy, the most applicable characteristics of the Bitcoin price prediction were assessed with different feature selection techniques.

The goal of a linear regression model is to determine a linear model with loss function which can be used for prediction of the price of bitcoin based on the closing price (Ali & Shatabda, 2020). A linear regression model was used to predict the various cryptocurrency price using the open, low, and high cost. The experiment shows an accuracy of 99.3% (Ho, et al., 2021). The main advantage of the logistic regression model is due to its parsimony and speed of implementation. Due to its smaller number of parameters to be estimated, it is also less prone to the over-fitting problem compared to the artificial neural networks (Akyildirim, et al., 2021). Bayesian Regression strategy can nearly double the investment in a Bitcoin portfolio in less than 60 days when running against real trading data from cryptocurrency exchanges. In the Blockchain and cryptocurrency field, several researchers have applied (Support Vector Machine) SVMs for the purpose of predicting Bitcoin and other cryptocurrency prices (Greaves and Au, 2015, Ashayer, 2019 & Shah and Zhang, 2014). The reason for choosing SVR is that it is robust to outliers. The decision model can be easily updated. It has excellent generalization capacity with high prediction accuracy, and its implementation is straightforward (Erfanian, et al., 2022).

The decision to use tree-based ensemble machine learning methods (Random Forest) is based on research showing that these methods have high accuracy when predicting the bitcoin. The strong predictive power of tree-based machine learning models indicate that investors and policymakers alike should consider taking advantages of these models (Basher and Sadorsky, 2022). Ashayer, (2019) used linear regression to investigate the predictive power of Blockchain network-based features on the future price of Bitcoin. Using this machine learning model, they were able to predict the price direction of Bitcoin, one hour in the future, with 55% accuracy. Logistic Regression is a predictive regression model in which the dependent variable is categorical. Logistic Regression is typically seen as a robust baseline method for classification (Greaves and Au, 2015).

Adcock and Gradojevic (2019) use feedforward neural networks to predict bitcoin prices. Compared to other forecasting methods the feedforward neural network produced the most accurate point and density forecasts. The Bayesian ANN outperforms the benchmark linear regression model (Jang and Lee, 2017). ANN possess strong parallel mapping skills. Acknowledging that, we also take note of the biggest disadvantage of ANNs complexity increases exponentially with network size. One of the most important issues when constructing an ANN is to find a balance between precision and complexity (Kristensen and Sognefest, 2023).

For price direction expectations, it examined the behavior of ANN, SVM, and ensemble algorithms (built on RNN and the k-means clustering method). (Gupta and Nalavade, 2023). The Bayesian neural systems (BNN) employed in Jang and Lee (2017) analyse the time series of Bitcoin price, and here the prediction error of log volatility can be improved. The ANN, SVM, and ensemble algorithms (Mallqui and Fernandes 2019) have the highest effectiveness rate. (Othman, et al), 2020 suggest that Artificial Nural Network (ANN) is an effective and adequate model for correctly predicting Bitcoin market prices using symmetric volatility attributes with accuracy level of 92.15% against the actual price, whereas the low-price attribute is found to be the major promoter for Bitcoin price trend with percentage of 63%.

Pabuçcu et al. (2020) use Support Vector Machines (SVM), Artificial Neural Networks (ANN), the Naïve Bayes (NB), Random Forest (RF), and logit regression to predict Bitcoin prices. For classification, ANN has the highest predictive accuracy while NB has the lowest. For the continuous data, RF has the highest predictive accuracy and NB has the lowest. Lahmiri and Bekiros (2019) conclude that deep learning methods are computationally intensive but useful for modelling and forecasting the hidden nonlinear patterns in these cryptocurrencies.

5. Conclusion

Bitcoin is a successful cryptocurrency, and it has been extensively studied in fields of economics and computer science. Various machine learning price prediction models are being used for predicting the price of this extremely unpredictable crypto currency. Machine learning methods have higher prediction accuracy than parametric regression approaches. Different authors use different data sets and different methods, technical indicators are probably the most important features in predicting Bitcoin prices. There is no clear consensus on the impact of business cycle variables like interest rates, inflation, and market volatility for forecasting Bitcoin prices. The crypto market is volatile and influenced by social media and other external factors, data sets cannot be the only reason for forecasting. As technology advances, data can be collected, analysed, and practiced, resulting in better results for predicting the Bitcoin price.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1]   Adcock, R., & Gradojevic, N. (2019). Non-fundamental, non-parametric Bitcoin forecasting. Physica A: Statistical Mechanics and its Applications, 531, 121727.

[2]   Akyildirim, E., Goncu, A., & Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. Annals of Operations Research, 297, 3-36.

[3]   Ali, M., & Shatabda, S. (2020, November). A data selection methodology to train linear regression model to predict bitcoin price. In 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT) (pp. 330-335). IEEE.

[4]   Ashayer, A. (2019). Modeling and Prediction of Cryptocurrency Prices Using Machine Learning Techniques. East Carolina University.

[5]   Basher, S. A., & Sadorsky, P. (2022). Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?. Machine Learning with Applications, 9, 100355.

[6]   Beck, R., Avital, M., Rossi, M., & Thatcher, J. B. (2017). Blockchain technology in business and information systems research. Business & information systems engineering, 59, 381-384.

[7]   Bertolini, R., Finch, S. J., & Nehm, R. H. (2021). Testing the impact of novel assessment sources and machine learning methods on predictive outcome modeling in undergraduate biology. Journal of Science Education and Technology, 30, 193-209.

[8]   Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. Journal of Computational and Applied Mathematics, 365, 112395.

[9]   Chuen, D. L. K., Guo, L., & Wang, Y. (2017). Cryptocurrency: A new investment opportunity?. The journal of alternative investments, 20(3), 16-40.

[10] D'Alfonso, A., Langer, P., & Vandelis, Z. (2016). The future of cryptocurrency. Ryerson university, 1-25.

[11] Erfanian, S., Zhou, Y., Razzaq, A., Abbas, A., Safeer, A. A., & Li, T. (2022). Predicting Bitcoin (BTC) Price in the Context of Economic Theories: A Machine Learning Approach. Entropy, 24(10), 1487.

[12] Geurts, P., & Louppe, G. (2011, January). Learning to rank with extremely randomized trees. In Proceedings of the Learning to Rank Challenge (pp. 49-61). PMLR.

[13] Greaves, A., & Au, B. (2015). Using the bitcoin transaction graph to predict the price of bitcoin. No data, 8, 416-443.

[14] Gupta, R., & Nalavade, J. E. (2023). Metaheuristic Assisted Hybrid Classifier for Bitcoin Price Prediction. Cybernetics and Systems, 54(7), 1037-1061.

[15] Hilpisch, Y. (2020). Artificial Intelligence in Finance. O'Reilly Media.

[16] Ho, A., Vatambeti, R., & Ravichandran, S. K. (2021). Bitcoin Price Prediction Using Machine Learning and Artificial Neural Network Model. Indian Journal of Science and Technology, 14(27), 2300-2308.

[17] Jang, H., & Lee, J. (2017). An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. Ieee Access, 6, 5427-5437.

[18] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

[19] Kristensen, T. S., & Sognefest, A. H. (2023). Can Artificial Neural Networks Be Used to Predict Bitcoin Data?. Automation, 4(3), 232-245.

[20] Kumar, D., Meghwani, S. S., & Thakur, M. (2016). Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. Journal of Computational Science, 17, 1-13.

[21] Lahmiri, S., & Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solitons & Fractals, 118, 35-40.

[22] Mallqui, D. C., & Fernandes, R. A. (2019). Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. Applied Soft Computing, 75, 596-606.

[23] Milutinović, M. (2018). Cryptocurrency. Ekonomika-Journal of economic theory and practice and social issues, (1), 105-122.

[24] Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.

[25] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system, bitcoin.pdf

[26] Nieto, P. G., García-Gonzalo, E., Antón, J. Á., Suárez, V. G., Bayón, R. M., & Martín, F. M. (2018). A comparison of several machine learning techniques for the centerline segregation prediction in continuous cast steel slabs and evaluation of its performance. Journal of Computational and Applied Mathematics, 330, 877-895.

[27] Othman, A. H. A., Kassim, S., Rosman, R. B., & Redzuan, N. H. B. (2020). Prediction accuracy improvement for Bitcoin market prices based on symmetric volatility information using artificial neural network approach. Journal of Revenue and Pricing Management, 19, 314-330.

[28] Pabuçcu, H., Ongan, S., & Ongan, A. (2023). Forecasting the movements of Bitcoin prices: an application of machine learning algorithms. arXiv preprint arXiv:2303.04642.

[29] Pichl, L., & Kaizoji, T. (2017). Volatility analysis of bitcoin. Quantitative Finance and Economics, 1(4), 474-485.

[30] Poongodi, M., Vijayakumar, V., & Chilamkurti, N. (2020). Bitcoin price prediction using ARIMA model. International Journal of Internet Technology and Secured Transactions, 10(4), 396-406.

[31] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1, 81-106.

[32] Rosenberg, J. M., & Krist, C. (2021). Combining machine learning and qualitative methods to elaborate students' ideas about the generality of their model-based explanations. Journal of Science Education and Technology, 30, 255-267.

[33] Rovira, S., Puertas, E., & Igual, L. (2017). Data-driven system to predict academic grades and dropout. PLoS one, 12(2), e0171207.

[34] Shah, D., & Zhang, K. (2014, September). Bayesian regression and Bitcoin. In 2014 52nd annual Allerton conference on communication, control, and computing (Allerton) (pp. 409-414). IEEE.

[35] Vikram, K., Sivaraman, N., & Balamurugan, P. (2022). Crypto Currency Market Price Prediction Using Data Science Process. Int. J. Res. Appl. Sci. Eng. Technol, 10(2), 1451-1454.

[36] Yermack, D. (2015). Is Bitcoin a real currency? An economic appraisal. In Handbook of digital currency (pp. 31-43). Academic Press.

[37] Yu, Y., Rashidi, M., Samali, B., Yousefi, A. M., & Wang, W. (2021). Multi-image-feature-based hierarchical concrete crack identification framework using optimized SVM multi-classifiers and D–S fusion algorithm for bridge structures. Remote Sensing, 13(2), 240.