

An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model

Hari Krishnan Andi

Centre for Postgraduate Studies, Asia Metropolitan University, Malaysia

E-mail: hari.andi@amu.edu.my

Abstract

In recent years, there has been an increase in demand for machine learning and AI-assisted trading. To extract abnormal profits from the bitcoin market, the machine learning and artificial intelligence (AI) assisted trading process has been used. Each day, the data gets saved for the specified amount of time. These approaches produce great results when integrated with cutting-edge algorithms. The results of algorithms and architectural structures drive the development of cryptocurrency market. The unprecedented increase in market capitalization has enabled the cryptocurrency to flourish in 2017. Currently, the market accommodates totally 1500 cryptocurrencies, all of which are actively trading. It is always possible to mine the cryptocurrency and use it to pay for online purchases. The proposed research study is more focused on leveraging the accurate forecast of bitcoin prices via the normalization of a particular dataset. With the use of LSTM machine learning, this dataset has been trained to deploy a more accurate forecast of the bitcoin price. Furthermore, this research work has evaluated different machine learning methods and found that the suggested work delivers better results. Based on the resultant findings, the accuracy, recall, precision, and sensitivity of the test has been calculated.

Keywords: LSTM, Bitcoin prediction

1. Introduction

Bitcoin refers to virtual money, which is widely utilized for both transaction and investment purpose. Bitcoin is a decentralized currency, which implies that it is not owned by a single person or group. Bitcoins are simple to use since they are not attached to any country. Using a bitcoin exchange is the best way to invest in bitcoins. Individuals can buy and sell bitcoins by using a variety of currencies [1-5].

As of January 2017, 170 hedge funds have been launched in cryptocurrencies for driving up the demand for the bitcoin in both trading and hedging future. These many conspiracy theories have been advanced to provide a theory about the causes of high volatility, and further these ideas have also been used to support the idea that cryptocurrency values will continue to fluctuate in the future. Another way to look at this is to indulge in automated bitcoin trading [6-8]. Figure 1 shows the perspective view of BTC price prediction.

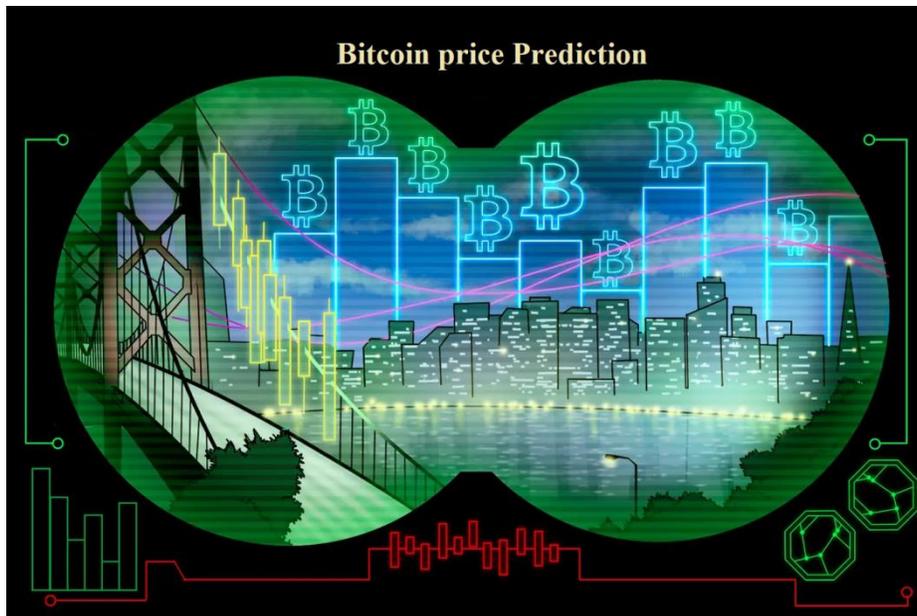


Figure 1. Bitcoin Price Prediction

To forecast BTC values, the machine learning and neural network utilizes numerical historical data. A recurrent neural network is an artificial neural network with directed graph nodes and connections that are constructed progressively, similar to synapses in the real brain. LSTM is an artificial RNN architecture that is commonly used in deep learning in addition to analysing single data points, which integrates the entire dataset [9, 10].

Virtual currency is the recently evolved worldwide phenomenon. Thus, it maintains a consistent identity, structure, and function. On the other hand, it is increasingly recognized as a superior financial medium with significant potential as time progresses. The development of Bitcoin was intended to reduce the use of third parties like banks, credit cards, and governments, and decrease transaction time and money transfer costs. Figure 2 shows the original data for last 5 years BTC price from registered website source, which is mentioned in the below figure.



Figure 2. Original data for last 5 years BTC price in USD (Source: www.chartoasis.com)

Bitcoin is among the virtual currencies with a considerable future ahead of it. Most cryptocurrencies, especially the most popular ones are mostly bitcoin clones. Because of this, it gained a lot of interest, and there were several papers published by utilizing both statistical and machine learning techniques. Statistics is a collection of many techniques that have been

developed over time to provide data summaries and quantify various features of a location, such as a specific set of observations [11-13]. To better comprehend ML algorithms, a firm knowledge should be gained on statistical techniques. While statistical techniques operate within the process of obtaining some relevant information by properly analyzing the dataset, wherein ML looks for patterns in the dataset and attempts to conclude exactly as humans would.

2. Organization of the Research

The remaining part of the research article is organized as follows: section 3 provides past research works on the prediction of bitcoin prices in the market. Section 4 discusses about the novel proposed work in a machine learning model to predict the price of bitcoin. Section 5 shows the performance of the proposed model along with its discussion. The conclusion and the future progress of the proposed work will be given in section 6.

3. Preliminaries

It is possible to build a time series dataset for Bitcoin by using different choices. Theoretically, the bitcoin dataset has a granular temporal period. Thus, various periods provide distinct datasets that may be gathered. Another feature of the bitcoin ecosystem is that all transactions are transparent to everyone. Researchers may leverage the dataset's causal connection, in addition to existing blockchain characteristics such as volume to incorporate new features in the blockchain. Cryptocurrency market participants are those that analyze the influence of networks on the market. Cryptocurrency exchange rates were used to calculate it. As a result, bitcoin is considered as the market leader. Consistent network effects were taken into account, since they provide stronger evidence [14, 15]. The researchers discuss about a well-known machine learning library and also about the one with the highest number of users. This machine learning tool is very helpful for developing suitable algorithms. The authors discuss about the library's simplicity and

efficacy to explain the benefits of Sci-Kit Learn. It explains how the library is integrated into the Python environment, as well as dealing with the implementation challenges that developers encounter while using this tool. For implementing the machine learning method, a number of additional libraries are required, including Keras and TensorFlow [16].

The main aim of the author is to develop a machine learning method (which uses random forests and generalized linear models) to predict the future price of bitcoin. The acquired dataset includes 25 different characteristics over a span of five years. While forecasting the future price, it was shown that they had achieved an accuracy of about 50 to 55 percent. Data was run for ten minutes [17].

The writers have done a great job of explaining the advantages of tools. Keras is an implementation of TensorFlow in the Python environment, which allows it to operate in the back-end. It was built with the intent of expediting neural network research. Convolutional networks and recurrent networks are supported by this library [18].

Using the sci-kit learns and Keras libraries, the author has predicted the value of bitcoin in advance. At the completion of the comparison, the best findings were that the R-square was high and the mean squared error was low [19].

4. Proposed Methodology

4.1 Data Pre-Processing

To begin with the history of the Bitcoin currency, it is distinct from the price of the coin market. Additionally, the information present on the blockchain is added to the database. Figure 3 shows the overall proposed framework. When it comes to blockchain scalability, we see no end in sight, which means the data should not be included. On the other side, a higher number of records often

mean either more exchange activity or more customers entering the system [20]. Furthermore, to get the conclusion's information, the PyTrends module is used to obtain the term "Bitcoin."

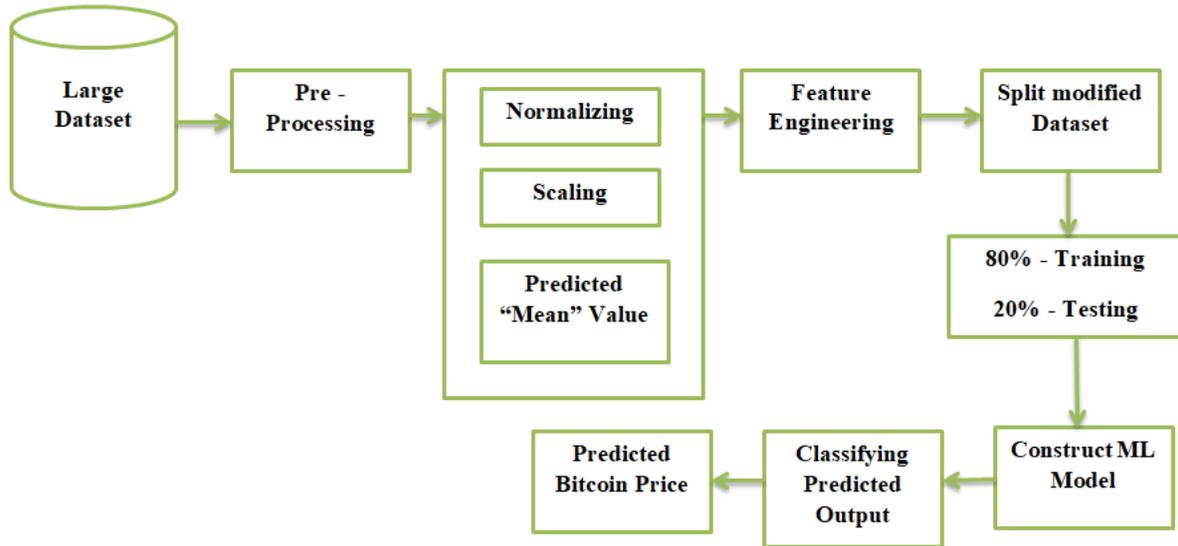


Figure 3. Overall Proposed Framework

4.1.1 Data collecting and cleaning

Based on the collected data, it includes relevant columns, such as volume, close, open, high costs, and market capitalization. "NaN" values are substituted with the meaning of a specific attribute. Next, all the datasets are merged into one, resulting in one overall dataset. Data from 2014 to the present is passed into the box.

4.1.2 Data normalization

Selecting a time series normalizing method is not a simple task. There is a significant amount of heterogeneous data to be analysed. As a result, substantial gradient updates will occur

throughout this phase by limiting the network convergence. Learning is made easier by converting the data into the format shown below.

Step 1

Maintain a value range of 0 to 1. More or less the same type and range of values should be used for all parameters.

Step 2

Every information input is scaled between 0 and 1, which is referred to as Min-Max Scaling for data normalization.

Step 3

Limits the range of numbers to between -1 and 1 and it normalizes the mean to a nearer value. Replacement parameters are mean of 0 and standard deviation of 1.

In order for our project to work, we utilize average scaling and go from 0 to 1. The average or mean scaling techniques perform well on these big datasets.

4.2 Time Series Data

The most important thing to remember about time series is that they are always a sequence of integers that increment with time. In contrast to its auto encoder representation, LSTM for series prediction works as a supervised set of rules. To divide the dataset into inputs and outputs, first, it should be separated into two halves [21, 22]. LSTM is also the finest concerning classical and statistical linear models since it is capable of handling varied data. By using the previous data, LSTM will predict the future data. To begin, it is highly required to determine how many prior days will be available for each prediction. Windows sizes are also available in this range. The predictions will vary depending on the values that are taken into consideration. In addition to

finding the patterns, a narrow window avoids cutting the patterns, which might develop over time. The data resulting from the processing has the window length and prediction range. When creating an LSTM network, the output size for the network is also determined by the prediction range.

4.3 Implementation of LSTM

Feed-forward networks rely on the fact that they avoid retaining any memory. In other words, each input is entirely separated from the other. Since we have a time series, where previous Bitcoin transaction data is required, we ought to maintain some of that data to have an accurate prediction of the future. The proposed framework can provide this design, in addition to the output, which includes a self-directing loop. Thus, as input, the window we provide is produced in a series. Gradients shrink or grow in size as the time step grows, and the result is disappearing or bursting gradients.

5. Results & Discussion

This project uses TensorFlow, Keras, Pandas, and Numpy for analysis. The Tensorflow and Keras are used for the Deep Learning backend, which means the front-end is dedicated to neural networks in the form of Tensorflow and Keras.

Table 1. Prediction for BTC by proposed model

Item / Year wise (Price in USD)	Year 2020	Current (2021)	Year 2025	Year 2030	Year 2035
Bitcoin	45000	47,780	97,837	1,68,129	3,01,253
Bitcoin cash	45005	47800	97,750	1,66,111	2,98,000

While working with data-related activities, Pandas are used for data manipulation and operations, Numpy for matrix and vector operations, and Scikit-learn for mean normalization. We

are displaying the data as charts in figure 5. Table 1 contained our accurate prediction as on dated (September, 2021) by our proposed model.

Split into training and test data

At this point, this stage is considered as one of the most important stages for Bitcoin. It is highly interested in predicting the values. For the testing purpose, the data from June 2017 will be used. The total number of 31942 training set values and the total number of 15734 test set values are determined. Input and output characteristics are seen in both training and test sets. Table 2 shows the computed performance metrics for the proposed algorithm.

Table 2. Overall Performance of Various Algorithms

S.No	Model	Accuracy	Precision	Recall	Sensitivity
1	Linear regression	89.1%	90.12%	89.98%	86.97%
2	Lasso Algorithm	91.4%	92.34%	93.19%	87.95%
3	Proposed logistic regression with LSTM	97.2%	98.32%	98.22%	95.35%

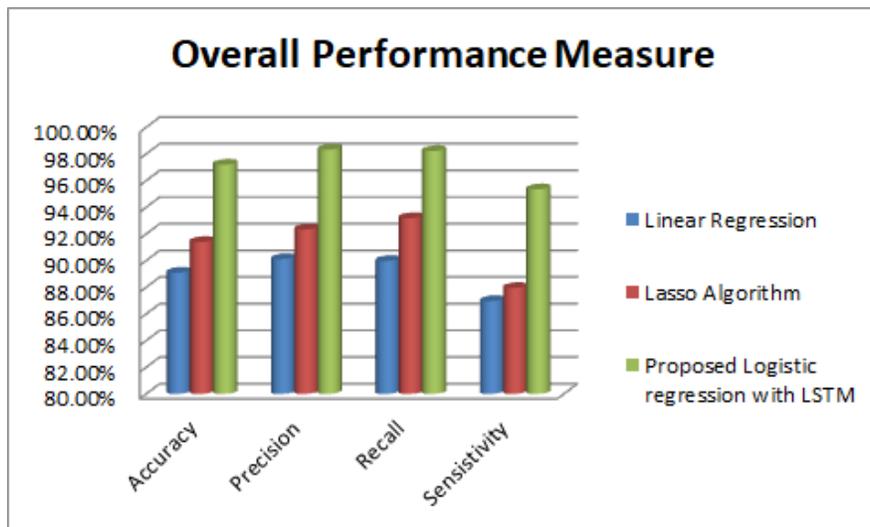


Figure 5. Performance Metrics for Various Algorithms

As a result, it has been found that when the training set is a large, it is more susceptible to the overfitting issues [23, 24]. The use of neural networks has helped researchers to get a deeper grasp of bitcoin and LSTM architectural design. The LSTM operates quickly and effectively, to produce the most possible precise results. Figure 5 shows the performance measure graph by analyzing various algorithms. It is showing that the proposed algorithm is more superior to other traditional algorithm in the large dataset normalization process.

6. Conclusion

As a result, the proposed algorithm demonstrates a better algorithm for predicting the bitcoin price based on the present global stock market. This research work has successfully attempted to solve the overfitting issue via proper large data management, as shown in the suggested work section. The proposed LSTM, which is based on the normalization method has successfully predicted the bitcoin price. Researchers have tried to use theories and algorithms to forecast the price of bitcoins, but the majority of models have proved to be ineffective due to overfitting and errors that occur due to huge datasets. Using a large dataset with LSTM, we can forecast the future value of bitcoin.

References

- [1] Dhanya, N. M. "An Empirical Evaluation of Bitcoin Price Prediction Using Time Series Analysis and Roll Over." In *Inventive Communication and Computational Technologies*, pp. 327-339. Springer, Singapore, 2021.

- [2] Farokhmanesh, F., & Sadeghi, M. T. (2019). Deep Feature Selection using an Enhanced Sparse Group Lasso Algorithm. 2019 27th Iranian Conference on Electrical Engineering (ICEE).
- [3] Sathesh, A. "Assessment of Environmental and Energy Performance Criteria for Street Lighting Tenders using Decision Support System." *Journal of Electronics and Informatics* 2, no. 2: 72-79.
- [4] D. Shah and K. Zhang, "Bayesian regression and Bitcoin," in 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2015, pp. 409-415.
- [5] Karunakaran, P., and Yasir Babiker Hamdan. "Early Prediction of Autism Spectrum Disorder by Computational Approaches to fMRI Analysis with Early Learning Technique." *Journal of Artificial Intelligence* 2, no. 04 (2020): 207-216.
- [6] "Project Based Learning: Predicting Bitcoin Prices using Deep Learning" S. Yogeshwaran ; Piyush Maheshwari; Maninder Jeet Kaur ; Amity University Dubai Dubai, UAE; IEEE 2019
- [7] Smys, S., and Wang Haoxiang. "Naïve Bayes and Entropy based Analysis and Classification of Humans and Chat Bots." *Journal of ISMAC* 3, no. 01 (2021): 40-49.
- [8] F. Andrade de Oliveira, L. Enrique Zúñiga and M. de Azevedo Reis; C. Neri Nobre, "The use of artificial neural networks in the analysis and prediction of stock prices," in IEEE International Conference on Systems, Man, and Cybernetics, 2011, pp. 2151-2155.
- [9] Chen, Joy Iong Zong, and P. Hengjinda. "Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method-A Comparative Study." *Journal of Artificial Intelligence* 3, no. 01 (2021): 17-33.
- [10] D. Shah and K. Zhang, "Bayesian regression and Bitcoin," in 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2015, pp. 409-415.
- [11] Chen, Joy Iong-Zong, and Kong-Long Lai. "Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert." *Journal of Artificial Intelligence* 3, no. 02 (2021): 101-112.

- [12] Huisu Jang and Jaewook Lee, "An Empirical Study on Modelling and Prediction of Bitcoin Prices with Bayesian Neural Networks based on Blockchain Information," in IEEE Early Access Articles, 2017, vol. 99, pp. 1-1.
- [13] G. Neil and H. Halaburda (2016) "Can we predict the winner in a market with network effects? Competition in cryptocurrency market," Games, vol.7 no.3, pp. 16.
- [14] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, and G. Varoquaux (2013) "API design for machine learning software: experiences from the scikitlearn project," arXiv preprint arXiv:1309.0238.
- [15] Tripathi, Milan. "Sentiment Analysis of Nepali COVID19 Tweets Using NB, SVM AND LSTM." Journal of Artificial Intelligence 3, no. 03 (2021): 151-168.
- [16] Madan, S. Saluja and A. Zhao (2015) "Automated Bitcoin Trading via Machine Learning Algorithms," URL: <http://cs229.stanford.edu/proj2015/>, vol. 20, pp. 1–5.
- [17] C. François (2015) "Keras: Deep learning library for theano and tensorflow," URL: <https://keras.io/k>.
- [18] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G S. Corrado, A. Davis, J. Dean, M. Devin, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng (2016) "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467.
- [19] Adam, Edriss Eisa Babikir. "Evaluation of Fingerprint Liveness Detection by Machine Learning Approach-A Systematic View." Journal of ISMAC 3, no. 01 (2021): 16-30.
- [20] F. Andrade de Oliveira, L. Enrique Z̃arate and M. de Azevedo Reis; C. Neri Nobre, "The use of artificial neural networks in the analysis and prediction of stock prices," in IEEE International Conference on Systems, Man, and Cybernetics, 2011, pp. 2151-2155.

- [21] Sivaganesan, D. "Novel Influence Maximization Algorithm for Social Network Behavior Management." Journal of ISMAC 3, no. 01 (2021): 60-68.
- [22] Issac Madan, Shaurya Saluja, Aojia Zhao, "Automated Bitcoin Trading via Machine Learning Algorithms", Department of Computer Science, Stanford University, Stanford, 2015.
- [23] Manoharan, Samuel, and Narain Ponraj. "Analysis of Complex Non-Linear Environment Exploration in Speech Recognition by Hybrid Learning Technique." Journal of Innovative Image Processing (JIIP) 2, no. 04 (2020): 202-209.
- [24] Brian Vockathaler, "The Bitcoin Boom: An In Depth Analysis of The Price Of Bitcoins", Thesis, University Of Ottawa, Ontario, Canada, June 2017.
- [25] Dhaya, R. "Flawless Identification of Fusarium Oxysporum in Tomato Plant Leaves by Machine Learning Algorithm." Journal of Innovative Image Processing (JIIP) 2, no. 04 (2020): 194-201.
- [26] Huisu Jang and Jaewook Lee, "An Empirical Study on Modelling and Prediction of Bitcoin Prices with Bayesian Neural Networks based on Blockchain Information," in IEEE Early Access Articles, 2017.

Author's biography

Hari Krishnan Andi is presently working as the director, Centre for Postgraduate Studies, Asia Metropolitan University, Malaysia. His major area of research includes emotional intelligence, data mining, soft skills, business management, psychological development, and mentoring techniques.