

FORECASTING BITCOIN PRICE BASED ON BLOCKCHAIN INFORMATION USING LONG-SHORT TERM METHOD

Kinanti Dhea Larasati^{1*}, Arum Handini Primandari²

^{1,2}Statistics Department, Faculty of Mathematics and Natural Sciences, UII

**e-mail: 16611065@students.uii.ac.id*

ABSTRACT

Since its founding in 2008, Bitcoin (financial code: BTC) has emerged as a digital currency in market cap and continues to attract investors and policymakers' attention. In recent years, BTC has high price volatility, a substantial increase in 2016, followed by a significant decline in 2018. Unlike stock markets, BTC is open for 24x7 dan has no closing period. It means everyone can trade it for any time. However, this flexibility carries investment risk. This research attempts to forecast BTC's price by considering the blockchain's information to minimize the risk. We employ Long-Short Term Memory (LSTM), the artificial Recurrent Neural Network (RNN) architecture. Its model can avoid long-term problems. The data used is BTC's price and blockchain information data from August 4, 2018, to January 21, 2020. The model with 20 neurons and 500 epochs has the smallest MSE value. Then a prediction has an accuracy rate of 91.07%.

Keywords: *Bitcoin, Blockchain Information, Forecasting, Long-Short Term Memory.*

INTRODUCTION

Technology is presented as something new as it pushes for increasing change. In reality, information technology (IT) processes data, collects information, stores collected material, accumulates knowledge, and made communication easier, and plays an important role in many aspects of daily operations of business world nowadays. In response to technological changes, criminals and society are find the new ways to develop. Therefore, technology is introduced to many financial mechanisms like cryptocurrency (Bakar & Rosbi, 2017).

Bitcoin is one of cryptocurrency that founded in 2008. BTC has emerged as a digital currency in market cap and continues to attract investors and policymakers' attention. In recent years, BTC has high price volatility, a substantial increase in 2016, followed by a significant decline in 2018 (Al-Yahyaee, 2019). First bitcoin transaction has occurred in January 2009. More two years later, various reports estimate the circulation of Bitcoin can be more than 6.5 million with approximately 10.000 users. Unlike stock markets, Bitcoin and the others cryptocurrency can be traded in any time cause it has no closing period. However, this flexibility can carries risk. Forecast BTC's price by considering blockchain information can minimize the risk if its done with predict the Bitcoin price accurately.

The methods that can be used is Long-Short Term Memory (LSTM) which is a development form from Recurrent Neural Network (RNN). The common LSTM unit consist of cells, input gate, output gate and forget gate. Cells remember value during arbitrary time interval and the gates regulates the flow of information into an out of cell. The LSTM model filters information through a gate structure to maintain and update the state of the memory cells. Each memory cells has three sigmoid layers and one tanh layer (Qiu, Wang, & Zhou, 2020).

The LSTM model used in financial data is able to produce a highr coefficient determination with smallest mean square error. The LSTM models used include the LSTM model with wavelet denoising and Gated Recurrent Unit (GRU) neural network model (Qiu, Wang, & Zhou, 2020) (Fischer & Krauss, 2018) (Aldi, Jondri, & Aditsania, 2018). LSTM implementation on electricity load and hydrology (rainfall) data also has high accuracy (Kratzert, Klotz, Brenner, Schulz, & Herrnegger, 2018) (Zheng, Xu, Zhang, & Li, 2017).

LSTM can be used for univariate or multivariate time series data. In way to forecast the bitcoin price (univariate) which has high fluctuation, LSTM is able to compete ARIMA accuracy model (McNally, Roche, & Caton, 2018). However, in one case if an LSTM combined with an AR model, is able to compete the performance of an LSTM on its own (Wu, Lu, Ma, & Lu, 2018).

Blockchain information like difficulty level, average number of blocks, transaction per block, etc. can be used to forecast the bitcoin price. The bitcoin price is affected by macro financial indicators such as the stock index, gold price, etc. Because bitcoin is valued in currency (USD, GBP, JYP, etc.) then the exchange rate of currency give an affect to ups and down of the bitcoin prices (Jang & Lee, 2018) (Sriwiji & Primandari, 2019).

MATERIALS AND METHODS

1. Data Source

The type of data used in this study is secondary data. The data is obtained from several website like www.blockchain.com to get bitcoin blockchain information and www.investing.com to get historical data of bitcoin prices (in USD exchange rates).

2. Research Variables

The variables used for this research consisted of 26 variables. The operational definition of research variables is an explanation of each variable used in the study of the indicators that make it up. It can be seen in Table 3.1

Table 1. Operational Definition of Research Variables

No.	Variable	Categoric	Operational Definition
1.	Price (Y) & (X ₁)	Bitcoin price	Bitcoin price in USD currency
2.	Miners Revenue(X ₂)	Blockchain Information	Total coinbase block prize and transaction fees paid to the miners.
3.	Difficulty(X ₃)	Blockchain Information	A relative measure of how difficult to find new blocks.
4.	Hash Rate(X ₄)	Blockchain Information	Estimated number of Tera (trillion) hashes per second for all performing miners (market participants to solve hash problems to create blocks)
5.	Median Confirmation Time(X ₅)	Blockchain Information	Average time of each transaction to be received into the mined block and recorded to the ledger
6.	Average Block Size(X ₆)	Blockchain Information	24-hour average block size in MB (Megabytes)
7.	Total Number of a Unique Bitcoin(X ₇)	Blockchain Information	Total number of unique addresses used on the Bitcoin Blockchain
8.	Transaction per Block(X ₈)	Blockchain Information	Average number of transactions per block
9.	Confirmed Transaction(X ₉)	Blockchain Information	Number of confirmations validity of transactions per day.
10.	Cosh % per Transaction(X ₁₀)	Blockchain Information	Income miners as a percentage of the transaction volume.

3. Data Analysis Methods

The steps that will be carried out in this study are as follows:

1. Preprocessing data by looking the existence of missing value
2. Perform descriptive analysis of daily Bitcoin data
3. Conducting LSTM Multivariate analysis with the following steps:
 - Calculating the sigmoid and tanh values using the equation (Ma, 2015):

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (1)$$

$$\tanh(x) = 2\sigma(2x) - 1 \quad (2)$$

where x is input data and e is an constanta.

- Normalizing research data by converting actual data into values with range intervals [0,1] using min-max scaling (Aldi, Jondri, & Aditsania, 2018):

$$x' = \frac{x - \min_x}{\max_x - \min_x} \tag{3}$$

Where x' is a result of normalized data, x is the data to be normalized, \min_x is minimum value of the data, and \max_x is maximum value of the data.

- Split the data into training data and testing data.
- Determine neuron and epoch number.
- Forecasting data
- Perform denormalization data using formulas:

$$dn_i = ((n_i)(x_{max} - x_{min})) + (x_{min}) \tag{4}$$

where dn_i is a result of denormalized data, n_i is the data to be denormalization, x_{min} is minimum value of the data, and x_{max} is maximum value of the data.

4. Comparing the MSE, RMSE, and MAPE value from each experiment using formulas (Budiman, 2016) :

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \tag{6}$$

$$MAPE = \frac{\sum_{i=1}^n \frac{|\hat{Y}_i - Y_i|}{Y_i}}{n} * 100\% \tag{7}$$

dengan Y_i : actual value; \hat{Y}_i : predict value; n; amount of time predicted

Below is a flowchat of the analysis.

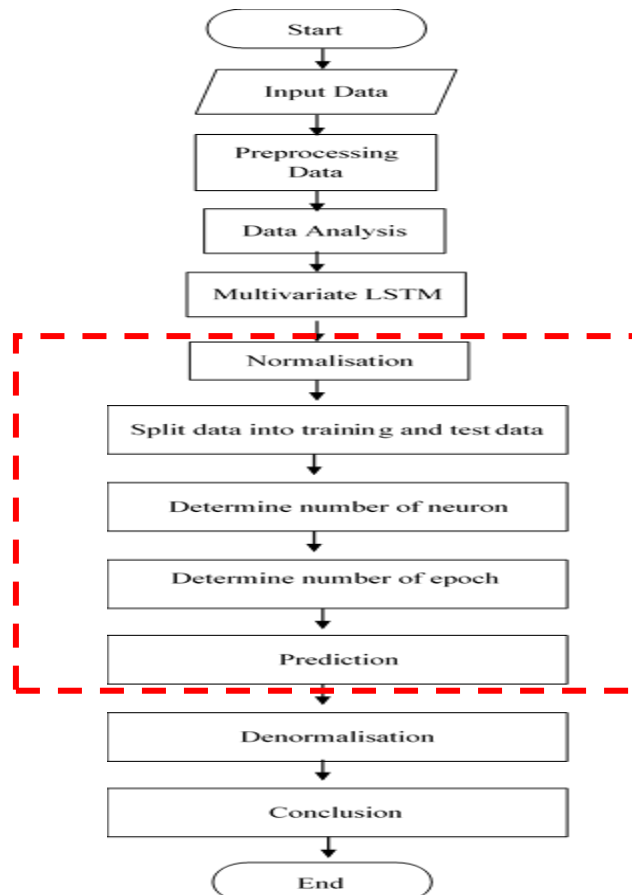


Figure 1. Data Analysis Flowchart Picture

RESULTS AND DISCUSSION

1. Descriptive Analysis

Descriptive analysis of daily Bitcoin data was carried out to find out the common description of Bitcoin on April 4th, 2018 to January 4th, 2020.

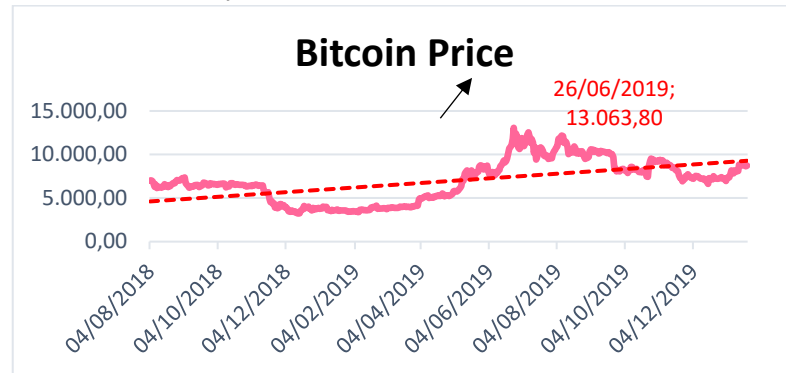


Figure 2. Bitcoin Price Plot

Based on Figure 2 it can be seen that Bitcoin price movement tend to increased or its trend is increasing. On June, 20th 2019, bitcoin reached the highest price i.e US\$ 13.063,80. Thi price is still low if compared to bitcoin price in 2017 cause it manage to reach \$19.533 per coin. The lowest bitcoin price was around US\$ 3.300 in early December 2018 a decrease of about 83% from the highest price (Farras, 2018).

2. Multivariate Long-Short Term Memory Analysis

The research data is split into training data and testing data with a ratio of 80% for training data or as much as 420 and 20% for testing data or as much as 107. Then, the network is formed with 10 input variables and 1 output layer with the number of neurons in the hidden layer that will be used for experiments are 10, 20, 30, 40, and 50. Number of epoch that will be used are 100, 500, and 1000. To find out the exact number of neurons and epoch it can be seen by the smallest loss value where loss value here means MSE value.

Parameters optimized using Adam's optimization. This optimization using bias correction technique. Theres no rules to determine number of neuron and epoch so the number obtained through simulation to get most optimal number in forecasting the time series data. The following table is the results with several numbers of neurons and epochs.

Table 2. Test Result with several Number of Neuron and Epoch

Neuron	Epoch	MSE
10	100	699199,312
	500	69395,375
	1000	87834,422
20	100	1992050,750
	500	68231,836
	1000	79453,750
30	100	646334,250
	500	70125,391
	1000	74488,625
40	100	701207,875
	500	69515,039
	1000	81914,812
50	100	999175,125
	500	70999,648
	1000	75650,141

Based on Table 2 it can be seen when compared to all number of neurons and epochs. the smallest MSE value is obtained by 20 neurons with 500 epochs (iterations). A bit of epoch can be effect to

network that formed was too general. It means the network ability to recognize the pattern is almost none. Meanwhile, too much epoch can effect to network experiencing an overfit condition or the network is too specific to training data.

3. Forecasting Data

Model that obtained from training process will be tested using testing data. To make sure the result of the model is good or not can be seen from the plot in Figure 3.

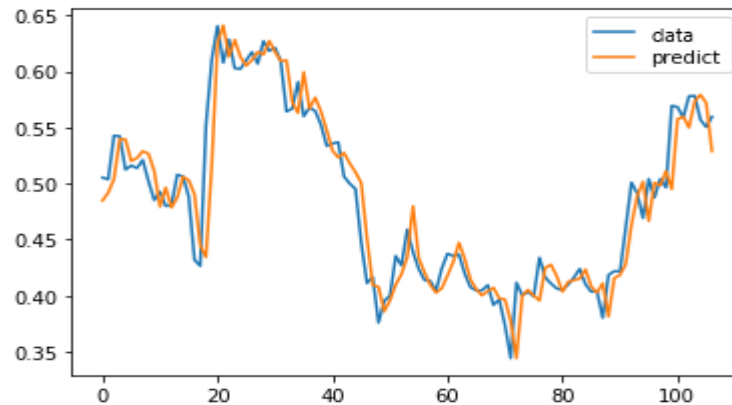


Figure 3. Comparison Plot between Actual dan Prediction Data

In Figure 3 we can see the pattern of predicted data follow the pattern of actual data. It means models that formed give a good forecasting (prediction) result. Actual data represented as blue line and prediction data represented as orange line.

Error value is measured by Mean Absolute Percentage Error (MAPE). To forecast the testing data, the LSTM model has MAPE 8.93%. It means, the model accuracy is 91.07%.

CONCLUSION

The movement of Bitcoin price tend to increased or its trend is increasing. The highest bitcoin price is reached On June, 20th 2019 i.e US\$ 13.063,80. Thi price is still low if compared to bitcoin price in 2017 cause it manage to reach \$19.533 per coin. The lowest bitcoin price was around US\$ 3.300 in early Desember 2018 a decrease of about 83% from the highest price.

In data preprocessing, dependent variable (bitcoin price) and independent variable (blockchain information) is normalized by mix-max scaling. Then, the data split into training data and testing data with a ratio 80% for training data and 20% for testing data. The optimal number of neuron and epoch was tuned doing simulation by looking the smallest MSE value. In this research we used 20 neurons and 500 epochs. Error value of LSTM model performance for testing data obtained 8.93% of MAPE. It means, the accuracy of the model is 91.07%.

REFERENCES

- Aldi, M. W., Jondri, & Aditsania, A. (2018). Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin. *e-Proceeding of Engineering*, 3551-3552. https://repository.telkomuniversity.ac.id/pustaka/files/144393/jurnal_eproc/analisis-dan-implementasi-long-short-term-memory-neural-network-untuk-prediksi-harga-bitcoin.pdf.
- Al-Yahyaee, K. H.-J. (2019). Can Uncertainty Indices Predict Bitcoin Prices? A Revisited Analysis Using Partial and Multivariate Wavelet Approaches. *North American Journal of Economics and Finance*, 48. <https://www.sciencedirect.com/science/article/pii/S1062940818306703>.
- Bakar, N. A., & Rosbi, S. (2017). Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction. *International Journal of Advanced Engineering Research and Science*, 130.
- Budiman, H. (2016). Analisis Dan Perbandingan Akurasi Model Prediksi Rentet Waktu Support Vector Machines Dengan Support Vector Machines Particle Swarm Optimization Untuk Arus Lalu

-
- Lintas Jangka Pendek. *SYSTEMIC*, 21-22.
<http://jurnalsaintek.uinsby.ac.id/index.php/SYSTEMIC/article/view/103>.
- Farras, B. (2018). *Harga Bitcoin Anjlok, Pengangguran Bertambah*. Jakarta: CNBC Indonesia.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669.
- Jang, H., & Lee, J. (2018). An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. *IEEE*, vol. 6, ISSN: 2169-3536, 5427-5437. <https://ieeexplore.ieee.org/document/8125674>.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 6005-6022.
- Ma, X. (2015). Long Short-Term Memory Neural Network for Traffic Speed Prediction Using Remote Microwave Sensor Data. *Transportation Research Part C*, 191. <https://www.sciencedirect.com/science/article/pii/S0968090X15000935>.
- McNally, S., Roche, J., & Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. *26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)* (pp. 339-343). Cambridge: IEEE.
- Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism. *PLOS ONE* 15(1): e0227222, 2-3. <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0227222>.
- Sriwijji, R., & Primandari, A. H. (2019). An Empirical Study in Forecasting Bitcoin Price Using Bayesian Regularization Neural Network. *Proceedings of the 1st International Conference on Statistics and Analytics* (pp. 350-361). Bogor: EAI.
- Wu, C.-H., Lu, C.-C., Ma, Y.-F., & Lu, R.-S. (2018). A New Forecasting Framework for Bitcoin Price with LSTM. *EEE International Conference on Data Mining Workshops (ICDMW)* (pp. 168-175). Singapore: IEEE.
- Zheng, J., Xu, C., Zhang, Z., & Li, X. (2017). Electric Load Forecasting in Smart Grid Using Long-Short-Term-Memory based Recurrent Neural Network. *IEEE*, 1-6. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7926112&isnumber=7926061>.