PREDICTING THE BITCOIN PRICE USING LINEAR REGRESSION OPTIMIZED WITH EXPONENTIAL SMOOTHING

Indah Suryani [1*)], Hani Harafani [2]

Informatika
Universitas Nusa Mandiri
www.nusamandiri.ac.id
indah.ihy@nusamandiri.ac.id [1*)], hani.hhf@nusamandiri.ac.id [2]

(*) Corresponding Author

Abstrak

Bitcoin merupakan salah satu mata uang kripto yang paling popular saat ini. Di dalam kondisi pandemic yang melanda dunia saat ini akibat Covid-19, maka bitcoin diharapkan dapat dijadikan sebagai sebuah investasi ketika tingkat ketidakpastian ekonomi sedang tinggi. Pada penelitian ini, data yang digunakan adalah data harga bitcoin yang termasuk ke dalam data deret waktu. Salah satu metode yang umum digunakan untuk prediksi dalam deret waktu adalah metode regresi linear. Untuk dapat mengembangkan hasil prediksi tersebut, digunakan teknik transformasi data menggunakan metode yang popular yaitu exponential smoothing. Pada metode exponential smoothing, dilakukan optimasi parameter alpha untuk dapat mendongkrak hasil prediksi dari regresi linear. Dan dari hasil ekperimen yang dilakukan, terbukti bahwa optimasi parameter alpha pada exponential smoothing mampu meningkatkan performa prediksi regresi linear dengan hasil perbandingan RMSE dengan uji t telah menghasilkan hasil perbedaan yang signifikan.

Kata kunci: Bitcoin; linear regresi; exponential smoothing

Abstract

Bitcoin is one of the most popular cryptocurrencies today. In the current pandemic conditions that hit the world due to Covid-19, bitcoin is expected to be used as an investment when the level of economic uncertainty is high. In this study, the data used is bitcoin price data which is included in time series data. One of the commonly used methods for prediction in time series is the linear regression method. To be able to develop the prediction results, a data transformation technique is used using the popular method, namely exponential smoothing. In the exponential smoothing method, optimization of the alpha parameter is carried out to be able to boost the prediction results from linear regression. And from the experimental results, it is evident that the optimization of the alpha parameter in exponential smoothing can improve the prediction performance of linear regression with the results of the comparison of RMSE with the t-test which has resulted in significant differences.

Keywords: bitcoin; linear regression; exponential smoothing

INTRODUCTION

Bitcoin having received increased levels of attention from the media and investors alike in recent years (Kalyvas et al., 2020). So making bitcoin one of the most popular among all cryptocurrencies. In line with (Jareño et al., 2020), they stated that In recent years, cryptocurrency markets have become much more popular, so cryptocurrencies may have moved to the category of investment assets. Much research on bitcoin has focused on the price discovery process and market efficiency in the bitcoin market (Tsang & Yang, 2020).

There are several bitcoin exchanges and the price difference between them is large and changes over time (Tsang & Yang, 2020). The global COVID-19 pandemic has disrupted normal business and affected sustainable economic development in many countries. However, it seems that the economic uncertainty following the COVID-19 containment measures is supporting the cryptocurrency market signal (Sarkodie et al., 2021). In line with (Kalyvas et al., 2020), their findings indicate that bitcoin may possess hedging

properties against economic uncertainty; therefore, it may be beneficial for investors to consider this cryptocurrency as an investment when economic uncertainty is high.

The linear regression model is representative of the most well-known family of regression models, this model consists of a linear function that underlies the class of hypotheses (Vercellis, 2009). Linear regression is a statistical technique that describes a linear relationship between two variables, namely the dependent variable and the independent variable (Aslanyan, 2021; Mondal & Rehena, 2020). Linear regression (LR) can be useful not only for discovering patterns in experimental data but also as a baseline for benchmarking and validating new analysis techniques (Zakeri et al., 2020), especially novel or unfamiliar ones. Linear regression is also one of the prediction methods in machine learning that is quite popular for researchers to develop, as done by (Huang & Hsieh, 2020), (Matiz & Barner, 2020), and (Patel & Kiran, 2019).

A relevant problem that often faced by practitioners regarding the dynamic nature of time series is the selection of a particular exponential smoothing model. For example, the choice between adopting a local linear trend and simple exponential smoothing is usually driven by the detection (or absence) of a trend in the data (Sbrana & Silvestrini, 2014). However, during the course of a business cycle, the trend dynamics of a series are sometimes not constant over time and may vary (Sbrana & Silvestrini, 2013). Data transformation can be in the form of smoothing, aggregation, generalization, normalization, and attribute construction or feature construction. One of the functions of the smoothing technique is to remove noise from the data. And exponential smoothing is one of those smoothing techniques (Han & Kamber, 2006). Another advantage of exponential smoothing is that it can consider trends and seasonal effects of the data so that it can produce estimates with simple formulas (Tratar, 2015). In addition, exponential smoothing also can beat many other advanced methods (Beaumont, 2014). Therefore, exponential smoothing is also widely used to develop time series prediction models, as was done by previous research in (Yager, 2013),(Koehler et al., 2012), and (Suryani, 2015).

Based on the literature, it is interesting in this study to be able to predict the price of bitcoin by developing a linear regression method which developed by transforming the data using exponential smoothing. Which in previous studies, efforts to improve performance with exponential smoothing used for the gold price dataset and were directly carried out on the Neural Network method without first comparing with other machine learning methods. While in this study, optimization with exponential smoothing carried out after comparing the RMSE values between the three machine learning methods and used to predict bitcoin prices.

RESEARCH METHODS

Types of research

This type of research is currently being carried out in the form of quantitative research in the form of experimental research.

Time and Place of Research

This research used secondary data from https://www.investing.com/crypto/bitcoin/historical-data. These data records collected from 01 march 2017 until 05 march 2021

Procedure

In this study, the dataset in the form of bitcoin closing prices was processed first with data pre-processing techniques such as set roles, normalize and windowing. The role set is used to define labels and id. Normalize is used to normalize the data and windowing is used to break the closing price attribute into 5 parts, namely 5 input data and 1 output data.

Modeling in this research is the optimization of the alpha parameter in exponential smoothing to improve performance on prediction results using linear regression as shown in Figure 1. The first thing to do is to process a dataset in the form of bitcoin closing prices with pre-processing techniques such as set roles, normalize and windowing. The role set is used to define labels and id. Normalize is used to normalize data using binary sigmoid activation function and windowing is used to break the closing price attribute into 5 parts, namely 5 input data and 1 output data.

Furthermore, exponential smoothing will used to optimize the performance of linear regression by optimizing its alpha parameter.

After that, the new data will be produced and then processed by linear regression method using 4 future selection options in the form of t-test, m5prime, Greedy, and iterative-test. The processing is carried out using the 10 fold cross-validation technique, namely by dividing the training and testing data. Then the RMSE value will obtained from each experiment carried out and then a comparison is made.
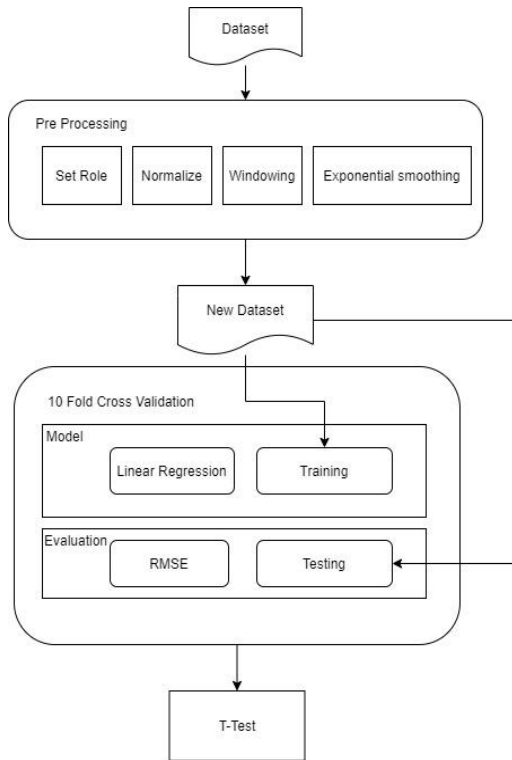
Figure 1. Proposed Method

Data, Instruments, and Data Collection Techniques

The data collected is in the form of historical data on bitcoin prices which includes the attributes of date, opening price, highest price, low price, closing price, volume_BTC, volume_currency, and weighted prices. The attributes used to be processed are only the attributes of the date and closing price. Which is contains 1.170 records.

Data analysis technique

The data used in this research is time-series data in the form of historical data from bitcoin prices. Wherefrom the bitcoin price data, only one price data attribute is used in the form of the closing price data only. As shown in Table 1. below.

Table 1. Samples of Bitcoin Prices Data

Date	Closing Price
05/03/2021	56.826
05/02/2021	57.016
05/01/2021	57.700
4/30/2021	56.803
4/29/2021	53.006
4/28/2021	54.456
4/27/2021	55.067
4/26/2021	53.297
4/25/2021	48.075
4/24/2021	50.955

Based on this data, there are two attributes, namely the date and closing price. Then made arrangements to determine the Id and Label attributes. We specify the date attribute as the Id attribute and the closing price attribute as the Label. Furthermore, the data normalization was carried out using the activation function of the binary sigmoid. Then the windowing technique is carried out because the data used is in the form of univariate data. After that, the data is ready to be processed using machine learning.

From the experiments conducted, the performance of several methods in machine learning was tested, namely using k-nearest neighbor, neural network, and linear regression.

Based on the RMSE results generated from the three methods, the method that produces the highest average RMSE is chosen and then optimized with exponential smoothing.

RESULTS AND DISCUSSION

Evaluation

The data that is ready to use after pre-processing is then predicted by experimenting with three methods. Then an evaluation will be made of the average RMSE value generated from each of these methods.

First, the experiment was carried out using the KNN method. In this experiment, the k parameter optimization was carried out on the KNN with 4 experimental samples. The results can be seen in Table 2 below. The average RMSE value generated from the KNN method is 0.608.

Table 2. Experiments Result Using KNN

No.	K	RMSE
1	0.7	0.5
2	0.5	0.48
3	0.3	0.478
4	0.1	0.974
Average		**0.608**

The next experiment is to use the neural network method. In this experiment, optimization was carried out on the learning rate and momentum parameters with 4 experiments. The result obtained is to get an average RMSE value of 0.497 as shown in Table 3 below.

Table 3. Experiments Result Using Neural Network

No.	LR	Mom	RMSE
1	0.01	0.9	0.507
2	0.001	0.9	0.435
3	0.01	0.5	0.454
4	0.001	0.5	0.59

Average	0.497

The third method that was tested is linear regression. This experiment was carried out with 4 different feature selections as listed in Table 4. The resulting average RMSE value was 0.451. And it turns out that this method produces the lowest RMSE value, which means that this method produces better predictive results.

Table 4. Experiments Result Using Linear Regression

No.	Feature selection	RMSE
1	t-test	0.451
2	m5prime	0.452
3	Greedy	0.449
4	iterative-t test	0451
Average		**0.451**

From the experiments that produced the best average RMSE value, then it was made to improve performance by using the exponential smoothing method. Then an experiment was carried out by optimizing the results of linear regression with exponential smoothing. The experiment was carried out by optimizing the alpha value in exponential smoothing with 4 feature selections in linear regression. And the results of these experiments can be seen in Table 4 below.

Table 5. Experiments Result Using Linear Regression + Exponential Smoothing

No.	Alpha	Feature selection	RMSE
1	0.5	t-test	0.229
2	0.3	t-test	0.157
3	0.1	t-test	0.175
4	0.5	m5prime	0.229
5	0.3	m5prime	0.157
6	0.1	m5prime	0.178
7	0.5	Greedy	0.229
8	0.3	Greedy	0.157
9	0.1	Greedy	0.178
10	0.5	iterative-t test	0.229
11	0.3	iterative-t test	0.157
12	0.1	iterative-t test	0.178
Average			**0.188**

By performing optimization using exponential smoothing in the linear regression method, it turns out that it can produce a lower RMSE value with an average RMSE value of 0.188. This value is generated over 12 experiments. And it can be seen that the choice of features does not affect increasing the RMSE value, while the optimization of the alpha value on exponential smoothing has a sufficiently good impact on the increase in the RMSE value.

Validation

To prove whether there is a difference, and how significant the difference is between the usual linear regression method and the proposed method, in the form of optimizing linear regression using exponential smoothing, so it validates with a T-test.

Table 6. RMSE Comparison Between LR and LR+ES Using T-test

	Variable 1	Variable 2
Mean	0.45075	0.18475
Variance	1.58333E-06	0.00095625
Observations	4	4
Pearson Correlation	-0.070674182	
Hypothesized Mean Difference	0	
df	3	
t Stat	17.14049415	
P(T<=t) one-tail	0.000216309	
t Critical one-tail	2.353363435	
P(T<=t) two-tail	0.000432618	
t Critical two-tail	3.182446305	

From the results of the T-test in Table 6, it produces a t-table value of 17.14049415, and t-count value of 3.182446305, then the t-table value is greater than the t-count value. This means that there is a difference in alias H1 is accepted and H0 is rejected. This difference also shows a significant value. This can be seen at the p-value which is less than 0.05 yaitu semester 0.000432618.

CONCLUSIONS AND SUGGESTIONS

Conclusion

Based on the experiments conducted, this study uses three machine learning methods, namely k-Nearest Neighbor, Neural Network and Linear Regression. From the three methods, it is known that the linear regression method shows the highest average RMSE value of 0.451. After that, efforts were made to improve linear regression performance with exponential smoothing. It is known that the optimization of the alpha parameter in exponential smoothing can provide an average RMSE value of 0.188 and can provide a significant difference in the classical linear regression method. Where in previous studies, efforts to increase performance with exponential smoothing were directly carried out on the Neural Network method without first comparing what machine learning method has a better RMSE. That study also used another dataset, which was the gold price dataset. Meanwhile, in this study, optimization with exponential smoothing was carried out after

knowing the method that produced the most superior RMSE for predict the bitcoin prices. And it can be concluded that exponential smoothing can improve the performance of linear regression to be able to predict bitcoin prices.

Suggestion

From the results of the research conducted, it turns out that exponential smoothing can provide increased performance in predictions using the linear regression method. So in future research, it is hoped that the use of exponential smoothing will be developed as a method in pre-processing data to improve the performance of other machine learning methods. Other experiments are also expected to be carried out with different datasets.

REFERENCES

Aslanyan, T. K. (2021). *Fundamentals Of Statistics For Data Scientists and Analysts*. Towards Data Science. https://towardsdatascience.com/fundament als-of-statistics-for-data-scientists-and-data-analysts-69d93a05aae7

Beaumont, A. N. (2014). Data transforms with exponential smoothing methods of forecasting. *International Journal of Forecasting*, *30*(4), 918–927. https://doi.org/10.1016/j.ijforecast.2014.03. 013

Han, J., & Kamber, M. (2006). Mining Stream, Time-Series and Sequence Data. In *Data Mining: Concepts and Techniques* (Vol. 54, pp. 468–473).

Huang, C. H., & Hsieh, S. H. (2020). Predicting BIM labor cost with random forest and simple linear regression. *Automation in Construction*, *118*(May), 103280. https://doi.org/10.1016/j.autcon.2020.1032 80

Jareño, F., González, M. de la O., Tolentino, M., & Sierra, K. (2020). Bitcoin and gold price returns: A quantile regression and NARDL analysis. *Resources Policy*, *67*(February). https://doi.org/10.1016/j.resourpol.2020.10 1666

Kalyvas, A., Papakyriakou, P., Sakkas, A., & Urquhart, A. (2020). What drives Bitcoin's price crash risk? *Economics Letters*, *191*(September 2011), 1–5. https://doi.org/10.1016/j.econlet.2019.1087 77

Koehler, A. B., Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). A study of outliers in the exponential smoothing approach to forecasting. *International Journal of Forecasting*, *28*(2), 477–484. https://doi.org/10.1016/j.ijforecast.2011.05. 001

Matiz, S., & Barner, K. E. (2020). Conformal prediction based active learning by linear regression optimization. *Neurocomputing*, *388*, 157–169. https://doi.org/10.1016/j.neucom.2020.01.0 18

Mondal, M. A., & Rehena, Z. (2020). Road Traffic Outlier Detection Technique based on Linear Regression. *Procedia Computer Science*, *171*(2019), 2547–2555. https://doi.org/10.1016/j.procs.2020.04.276

Patel, D. R., & Kiran, M. B. (2019). A non-contact approach for surface roughness prediction in CNC turning using a linear regression model. *Materials Today: Proceedings*, *26*(xxxx), 350–355. https://doi.org/10.1016/j.matpr.2019.12.02 9

Sarkodie, S. A., Ahmed, M. Y., & Owusu, P. A. (2021). COVID-19 pandemic improves market signals of cryptocurrencies–evidence from Bitcoin, Bitcoin Cash, Ethereum, and Litecoin. *Finance Research Letters*, *January*, 102049. https://doi.org/10.1016/j.frl.2021.102049

Sbrana, G., & Silvestrini, A. (2013). Forecasting aggregate demand: Analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework. *International Journal of Production Economics*, *146*(1), 185–198. https://doi.org/10.1016/j.ijpe.2013.06.022

Sbrana, G., & Silvestrini, A. (2014). *Random switching exponential smoothing and inventory forecasting*. https://www.bancaditalia.it/pubblicazioni/t emi-discussione/2014/2014-0971/en_tema_971.pdf

Suryani, I. (2015). Penerapan Exponential Smoothing untuk Transformasi Data dalam Meningkatkan Akurasi Neural Network pada Prediksi Harga Emas. *Journal of Intelligent Systems*, *1*(2), 67–75.

Tratar, L. F. (2015). Int . J . Production Economics. *Intern. Journal of Production Economics*, *161*, 64–73. https://doi.org/10.1016/j.ijpe.2014.11.019

Tsang, K. P., & Yang, Z. (2020). Price dispersion in bitcoin exchanges. *Economics Letters*, *194*, 109379. https://doi.org/10.1016/j.econlet.2020.1093 79

Vercellis, C. (2009). Business Intelligence: Data Mining and Optimization for Decision Making. In *Business Intelligence: Data Mining and Optimization for Decision Making*. https://doi.org/10.1002/9780470753866

Yager, R. R. (2013). Exponential smoothing with credibility weighted observations. *Information Sciences*, *252*, 96–105. https://doi.org/10.1016/j.ins.2013.07.008

Zakeri, Z., Mansfield, N., Sunderland, C., & Omurtag, A. (2020). Cross-validating models of continuous data from simulation and experiment by using linear regression and artificial neural networks. *Informatics in Medicine Unlocked*, *21*(July), 1–6. https://doi.org/10.1016/j.imu.2020.100457