

Paper_Haerul.pdf

by

Submission date: 21-Jul-2020 02:46PM (UTC+0700)

Submission ID: 1360321076

File name: Paper_Haerul.pdf (175.53K)

Word count: 2575

Character count: 13240

Data mining for cryptocurrencies price prediction

Haerul Fatah^{1*}, Recha Abriana Anggraini², Deddy Supriadi³, Melisa Winda Pertiwi⁴, Ai Ilah Warnilah⁵, Nurul Ichsan⁶

Program Studi Sistem Informasi Universitas Bina Sarana Informatika^{1,2,3,5,6}
Program Studi Sistem Informasi STMIK Nusa Mandiri⁴

E-mail: haerul.hef@bsi.ac.id

Abstract. Electronic payments is something that is currently in high demand by investors today, but transactions are often constrained due to various problems, especially from third parties. For this reason, cryptocurrency emerged, which is one of the solutions for conducting electronic payment transactions. Some types of cryptocurrency that are most in demand by investors are bitcoin, ethereum, and ripple. The fluctuation value of cryptocurrency is very difficult to predict so that investors often experience losses when making transactions. This study aims to predict cryptocurrency prices such as bitcoin, ethereum and ripple using data mining algorithms. The data mining algorithm used in this prediction process is K-NN, Neural Network, SVM, Linear Regression, Random Forest and Decision Tree. Data mining modeling is done by dividing the dataset into each type of commodity and then analyzed using each algorithm. The results of this study indicate that the accuracy value obtained from some data mining algorithms is good enough to predict cryptocurrency prices

1. Introduction

Non-cash payment is a form of technology application in banking that is being promoted at the moment. Bank Indonesia divides 5 types of non-cash payment instruments, namely cards, checks, crossed checks, debit notes, and electronic money. Of course, of the five types of payment instruments, electronic money is the most needed option now [1]. Electronic money is now an option that is starting to be used by many people, especially entrepreneurs, business people and investors, this is because they think that electronic money will replace physical money in the future. With electronic payments for transactions, problems often arise due to electronic or virtual money both in terms of currency and others.

Cryptocurrency is one of the virtual money that has its currency unit that can be used to buy products or services online and offline [2]. Cryptocurrency emerged as an answer to the current payment system constraints that rely heavily on third parties. Cryptocurrency can be bought or sold using real or virtual currencies according to certain exchange rates in the same way someone would exchange real currencies [3]. One popular type of cryptocurrency is Bitcoin [4]. Bitcoin operates online a decentralized network without dependence on any government or legal entity because it is very dependent on peer-to-peer networks and cryptography to maintain its integrity [5].

The value of cryptocurrency such as bitcoin varies as much as stocks, the fluctuation is also very difficult to estimate because it is very fast, causing some people who invest with bitcoin to have difficulty knowing the right time to sell or buy it to get the maximum profit. The price of Bitcoin does not depend on business events or government intervention, unlike market stock.

So, to estimate value, it is necessary to utilize technology to predict cryptocurrency prices such as bitcoin.

Technological developments in this era, especially in the big data problem, make it easier for humans to classify or predict things based on data obtained by analyzing using various data mining and machine learning methods. Data mining is defined as "a non-trivial process of identifying data patterns that are valid, new, potentially useful, and ultimately understandable" [6]. Data mining techniques are applied in Indonesia in many domains, some well-known examples including valuation credit, weather prediction, and applications in the medical sector such as the selection of the best in vitro fertilization embryos [7]. Therefore, in this study, we analyzed cryptocurrency transaction data using several data mining methods to find the best price predictions for can make a profit when doing bitcoin transactions and other commodities, with this research we can help many interested parties in predicting the price of bitcoin in making predictions. .

2. Method

Stages of research conducted in predicting cryptocurrency prices are:

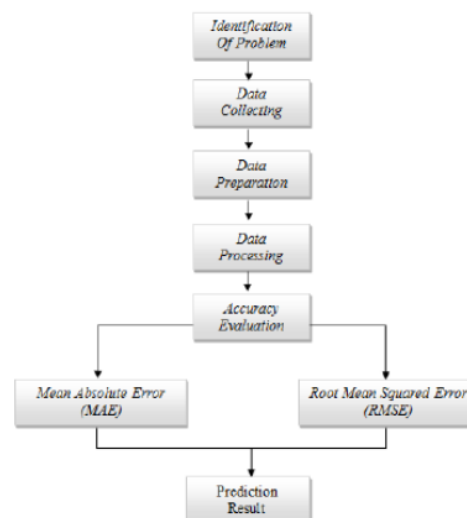


Figure 1. Stages of research conducted in predicting cryptocurrency prices

In Figure 1, explain the stages of the research carried out, consisting of 5 stages as follows:

2.1. Identification of problem

Problem identification is done by reading the literature relating to cryptocurrency and studying previous research on cryptocurrency data analysis to find this research topic.

2.2. Data collecting

The cryptocurrency data collection is done by scrapping transaction history data for the past 1 year from a website that is commonly used for transactions namely coinmarketcap.com

2.3. Data preparation

The initial stage before data processing is done using WEKA tools to find out the results of cryptocurrency price predictions from several datasets obtained during the data collection process. Before it is processed using several data mining algorithms to find out the price prediction results, the data preprocessing stage is carried out. Processing to eliminate noise in the data and to balance the data so that data processing is easier to do and the results obtained are also more relevant.

2.4. Data processing

Data processing is performed by implementing data mining algorithms used in the processing of this cryptocurrency dataset, including K-Nearest Neighbor, Support Vector Machine, Random Forest, Decision Tree, Neural Network, and Linear Regression. From each of these algorithms will be obtained the accuracy value in predicting cryptocurrency prices and can be known what method has the best accuracy in predicting cryptocurrency prices.

2.5. Accuracy evaluation

Evaluation of the accuracy value in this study is not only seen from the amount of accuracy obtained from a method but also seen from the value of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) generated by each method. In the prediction model, accuracy is said to be best if the value of MAE and RMSE is 0.

Novelty of this research is that we make predictions on three commodities of cryptocurrency namely in the form of bitcoin, ethereum and ripple and retrieve data from coinmarketcap.com and make predictions using several data mining algorithms such as K-NN, Neural Network, SVM, Linear Regression, Random Forest and Decision Tree where there have never been other studies that predict cryptocurrency prices with the data and algorithms we use in this study.

3. Related Work

Research on the prediction of bitcoin prices has been done by Rathana et al [8] in his research explained that the purpose of this work is to obtain the accuracy of bitcoin price predictions using different machine learning algorithms and compare their accuracy. Experimental results were compared for decision tree algorithms and regression techniques.

Cryptocurrency trading is now becoming popular in the field of investment so that its market is similar to foreign exchange and stocks. However, because of its volatility, predictions are needed in order to get the profit as desired. Previous research has been conducted by Radityo et al [9] this study studies various ANN methods to predict the market value of one of the most widely used cryptocurrencies, Bitcoin. The ANN method is used to develop a model to predict the value of a Bitcoin cap the next day (next day prediction). This study compared four ANN methods, namely backpropagation neural network (BPNN), neural network genetic algorithm (GANN), backpropagation neural network genetic algorithm (GABPNN), and neuroevolution from enhancing topology (NEAT). The method is evaluated based on accuracy and complexity. Experimental results show that BPNN is the best method with MAPE $1.998 \pm 0.038\%$ and training time 347 ± 63 seconds.

Research Mittal et al [10] focuses on the prediction of cryptocurrency market prices based on historical trends. For their research, they try to understand and identify the daily trends in the cryptocurrency market that analyze features related to cryptocurrency prices. The dataset used consists of more than nine features related to cryptocurrency prices recorded every day for a period of 6 months. This study applies several machine learning algorithms predicting changes in the daily price of a cryptocurrency.

4. Result and Discussion

4.1. Dataset

This study aims to predict the price of a cryptocurrency by looking at the value of a close at a certain time by looking at the history of transactions made. The dataset used in this study is the transaction history of some of the most popular cryptocurrency commodities taken from the coinmarketcap.com website. The dataset consists of 3 types of the most popular cryptocurrency commodity datasets namely bitcoin, ethereum and ripple. The data taken is the transaction history of each commodity for 1 year. The dataset attributes consist of 5 attributes namely date, open, high, low and close with a total of 181 data records. Data collection for the dataset is done by scraping the coinmarketcap.com website. Table 1 shows the explanation of the attributes of each dataset.

Table 1. Cryptocurrency Dataset

Attribute	Description
Date	Transaction date of cryptocurrency
Open	Opening price/cryptocurrency initial price at a certain time
High	The highest price of the opening price
Low	The lowest price of the opening price
Close	The closing price/cryptocurrency final price of a certain time

4.2. Data Preparation

The initial stage carried out in this research is the preprocessing stage. The preprocessing stage is carried out to eliminate noise in data such as missing values and others and to simplify data. Because the data taken from the website is too large and complicated, it needs to be normalized to balance the data. In this stage, the normalization of each dataset is carried out to make it easier to process data. The normalization phase is carried out by changing the maximum value and minimum data value to 1 and 0. Table 2 displays the results of statistical data after normalization.

4.3. Data Processing

The processed dataset consists of 3 datasets namely bitcoin, ethereum and ripple datasets with a total of 181 data each. The dataset is processed using several data mining algorithms such as K-NN, Neural Network, Decision Tree, SVM, Random Forest and Linear Regression with cross-validation validation and K-fold validation methods. The results of experiments with 10 folds are shown in table 3 below:

Table 3 shows the experimental results of each cryptocurrency dataset using 6 data mining algorithms with cross-validation validation, the number of fold = 10. From table 3 it can be seen that the best accuracy of each dataset is obtained when the dataset is processed using the Support Vector Machine algorithm (SVM) because it has the closest MAE and RMSE values to zero. While the lowest accuracy for cryptocurrency price predictions for each dataset is obtained when processed with a linear regression algorithm.

Table 4 shows the experimental results of each cryptocurrency dataset using 6 data mining algorithms with cross-validation, the number of fold = 5. From these results, it can be seen that the best accuracy for cryptocurrency price predictions is obtained from data processed by

Table 2. Statistical datasets after normalized

Cryptocurrency	Value	Result
Bitcoin	Minimum	0
	Maximum	1
	Mean	0.28
	StdDev	0.201
Ethereum	Minimum	0
	Maximum	1
	Mean	0.341
	StdDev	0.23
Ripple	Minimum	0
	Maximum	1
	Mean	0.164
	StdDev	0.185

Table 3. 10 fold cross-validation experimental results

DataSet	K-N		Neural Network		Decision Tree		SVM		Random Forest		Linear Regression	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Bitcoin	0.022	0.0353	0.0288	0.0409	0.1514	0.2017	0.0212	0.029	0.1234	0.1678	0.1438	0.1907
Ethereum	0.0224	0.0348	0.0476	0.0899	0.1889	0.2303	0.0227	0.0313	0.1469	0.1869	0.1813	0.2205
Ripple	0.0163	0.0354	0.0246	0.0363	0.1207	0.1856	0.0158	0.0278	0.1027	0.1543	0.1085	0.1694

Table 4. Results of experiments with 5 fold validation

DataSet	K-N		Neural Network		Decision Tree		SVM		Random Forest		Linear Regression	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Bitcoin	0.0208	0.0311	0.0372	0.0453	0.151	0.2011	0.0216	0.0294	0.1256	0.1703	0.1441	0.1911
Ethereum	0.0239	0.0377	0.0345	0.0464	0.1886	0.2299	0.0234	0.032	0.1447	0.1838	0.1819	0.2211
Ripple	0.0168	0.0359	0.0224	0.0351	0.1203	0.1851	0.0162	0.0276	0.1002	0.149	0.1203	0.1851

the Support Vector Machine (SVM) method. The lowest accuracy for each dataset is obtained when processed by the decision tree method.

5. Conclusion

Cryptocurrency has a variety of commodities, the most popular of which are bitcoin, ethereum, and ripple. Each of these commodities has fluctuating values that are very difficult to guess so data analysis is needed to predict commodity prices so that investors can obtain maximum profits when making transactions with these commodities, prediction of commodity prices can be done with the help of data mining.

This research succeeded in providing information in the form of the best accuracy value of cryptocurrency price predictions after processing the transaction history data of each cryptocurrency commodity using 6 data mining methods. The best accuracy results are obtained when the data is processed by the SVM method with cross validation. For the future research to get price predictions of cryptocurrency better and more accurate, in future studies an increase in the number of datasets and the addition of validation methods can be done or research using other algorithms that have not been used in this study.

References

- [1] F. Mulyanto, 2015 Pemanfaatan cryptocurrency sebagai penerapan mata uang Rupiah kedalam bentuk digital menggunakan teknologi bitcoin, Indones. J. Netw. Secur., vol. 4, no. 4, pp. 19–26, doi:

10.1123/ijns.v4i4.1364.

- [2] A. S. Khan and P. Augustine, 2019 Predictive analytics in cryptocurrency using neural networks: A comparative study, *Int. J. Recent Technol. Eng.*, vol. 7, no. 6, pp. 425–429.
- [3] J. Bohr and M. Bashir, 2014 Who Uses Bitcoin?, *PST 2014 Proc.*, pp. 94–101.
- [4] E. Dourado and J. Brito, 2020 The New Palgrave Dictionary of Economics, *New Palgrave Dict. Econ.*, pp. 1–9, doi: 10.1057/978-1-349-95121-5.
- [5] E. Sin and L. Wang, 2018 Bitcoin price prediction using ensembles of neural networks, *ICNC-FSKD 2017 - 13th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov.*, pp. 666–671, doi: 10.1109/FSKD.2017.8393351.
- [6] P. Ristoski and H. Paulheim, 2016 Semantic Web in data mining and knowledge discovery: A comprehensive survey, *J. Web Semant.*, vol. 36, pp. 1–22, doi: 10.1016/j.websem.2016.01.001.
- [7] J. Moeyersoms, E. Junqué De Fortuny, K. Dejaeger, B. Baesens, and D. Martens, 2015 Comprehensible software fault and effort prediction: A data mining approach, *J. Syst. Softw.*, vol. 100, pp. 80–90, doi: 10.1016/j.jss.2014.10.032.
- [8] K. Rathan, S. V. Sai, and T. S. Manikanta, 2019 Crypto-currency price prediction using decision tree and regression techniques, *Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019*, vol. 2019-April, no. Icoei, pp. 190–194, doi: 10.1109/icoei.2019.8862585.
- [9] A. Radityo, 2017 Prediction of Bitcoin Exchange Rate to American Dollar Using Artificial Neural Network Methods.
- [10] R. Mittal, S. Arora, and M. P. S. Bhatia, 2018 Automated cryptocurrencies prices prediction using machine learning, *Div. Comput. Eng. Netaji Subhas Inst. Technol. India*, vol. 8, pp. 2229–6956, doi: 10.21917/ijsc.2018.0245.

ORIGINALITY REPORT

13%

SIMILARITY INDEX

9%

INTERNET SOURCES

10%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

1

qoribmunajat.github.io

Internet Source

4%

2

files.eric.ed.gov

Internet Source

2%

3

Algifanri Maulana, Anggia Dasa Putri, Yulia.
"Development of digital currency technology",
Journal of Physics: Conference Series, 2019

Publication

2%

4

lirias.kuleuven.be

Internet Source

1%

5

Karunya Rathan, Somarouthu Venkat Sai,
Tubati Sai Manikanta. "Crypto-Currency price
prediction using Decision Tree and Regression
techniques", 2019 3rd International Conference
on Trends in Electronics and Informatics
(ICOEI), 2019

Publication

1%

6

Submitted to De Montfort University

Student Paper

1%

7

brage.bibsys.no

Internet Source

1%

8

www.cs.essex.ac.uk

Internet Source

1%

9

Submitted to Heriot-Watt University

Student Paper

1%

10

www.mdpi.com

Internet Source

<1%

Exclude quotes On

Exclude bibliography On

Exclude matches < 10 words