**International Journal of Advances in Data and Information Systems**
Vol. 5, No. 2, October 2024, pp. 183~188
ISSN: 2721-3056, DOI: 10.59395/ijadis.v5i2.1338


# Comparative Analysis of Cryptocurrency Prediction based on Deep Learning, Decision Tree, Gradient Boosted Tree, Random Tree, and k-NN Model

**Sugeng Riyadi, Faisal Fahmi**
Department of Library and Information Science, Airlangga University, Indonesia

**Article Info**

***Article history:***

Received Aug 15, 2024
Revised Oct 02, 2024
Accepted Oct 30, 2024

***Keywords:***

Bitcoin
Machine Learning
Prediction
Cryptocurrency

**ABSTRACT**

Cryptocurrency being a digital or virtual currency that uses cryptography to secure transactions and control the creation of new units. Bitcoin, one of the most popular cryptocurrency, offers various advantages such as security, transparency, and efficiency. The value of Bitcoin can change over time, similar to the regular currencies, and the need to predict the value can be as important as those in the regular. The prediction can be done by multiple algorithms. The purpose of this research is to compare five algorithms in predicting bitcoin value based on Root Mean Squared Error (RMSE) and Squared Error ($R^2$). The five algorithms compared can model the prediction of changes in the bitcoin cryptocurrency, effectively. Based on the experiment, Random Forest outperformed the other algorithms based on its RMSE and $R^2$ result.



***Corresponding Author:***

Faisal Fahmi,
Department of Library and Information Science,
Airlangga University,
Jl. Airlangga No.4 - 6, Airlangga, Gubeng, Surabaya, East Java 60115, Indonesia.
Email: faisalfahmi@fisip.unair.ac.id

## 1. INTRODUCTION

Cryptocurrency is a digital or virtual currency that uses cryptography to secure transactions and control the creation of new units. National currencies have transformed throughout history from gold and silver coins to simple coins, paper money, and finally digital money, encompassing both physical and legal changes, and with the digital revolution, this currency evolution is inevitable [1]. Cryptocurrencies have no central authority, but rather are decentralized, managed by a distributed network of computers. Cryptocurrencies are cryptographic commodities designed to serve as a medium of exchange, using cryptography to encrypt transfers, monitor unit creation, and validate asset transfers [2]. Cryptocurrency has changed the way the world views financial systems and digital transactions. By offering a decentralized, secure, and transparent system, cryptocurrencies have great potential to transform various aspects of our lives. As the first fully decentralized digital currency, Bitcoin offers various advantages such as security, transparency, and efficiency. However, challenges such as volatility, regulation, and security must be overcome to ensure the long-term growth and stability of the cryptocurrency ecosystem. With ongoing technological developments, the future of cryptocurrencies looks full of new innovations and opportunities.

In cryptocurrency, the use of algorithmic trading is often used which refers to the use of computer applications with multiple algorithms to identify and execute trades with a speed and frequency that is not possible for human traders. Price prediction is the main difficulty in finding and

executing potentially profitable trades [3]. These predictions can use multiple algorithms to get close results according to machine calculations. Machine learning can be applied to predict data [4]. Machine learning is a sub-branch from artificial intelligence that permits software applications to be even more accurate in predicting data, as well as to forecast current performance and upgrade for future data [5]. There are several methods that can be used to predict things such as K-Nearest Neighbor to predict rainfall [6], Random Forest used to predict house prices [7], Deep Learning used to exploring sentiment trends about social media in Google Play Store review [8], Decision Tree used to determination of prospective construction service providers [9], Gradient Boosted Trees used to forecasting of hierarchical time series [10].

## 2. RESEARCH METHOD

### 2.1 Related Work

Several related studies have been conducted in an effort to compare better methods and techniques to predict a certain value. In this paper as a reference to determine the novelty of the research, we will review some of the research related to the comparison between algorithms and related to cryptocurrency that has been done before.

In this research, one of the methods used is deep learning which can be described as a class of Machine Learning algorithms that use multiple layers of nonlinear processing organized in a cascade for feature extraction and transformation [11]. Research using deep learning analysis was conducted to predict trip duration using New York taxi trip duration dataset with a deep learning approach, namely Long Short Term Memory Recurrent Neural Network (LSTM-RNN), by adjusting related parameters such as epoch, dropout value, and number of neurons [12]. There is another new method that we can use there is DenseTNT, is an anchor-free and end-to-end trajectory prediction method that directly generates a set of trajectories from dense candidate destinations[13]. The use of deep learning in predicting active cases of Covid-19 also opens up opportunities to develop better systems for predicting and managing future pandemics [14]. Other research shows The results of analysis, experimentation, and testing of rainfall and flood disaster data show that the proposed model, namely the Deep Neural Investigation Network (DNIN), is superior to the Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) [15].

The decision tree method as one of the research has been used to find out, determine, and get the accuracy value in providing an assessment of the feasibility of prospective construction service providers, as well as analyzing the variables that affect the feasibility assessment of prospective construction service providers [9]. The previous work used the Gradient Boosted Trees method to solve point and probabilistic forecasting problems by describing a blending methodology for machine learning models from the family of gradient boosted trees and neural networks that has been successfully applied in the recent M5 Competition in the Accuracy and Uncertainty track [10]. The Random Forest method is used to predict house prices based on its features with good performance, but the random forest method has drawbacks if it uses too many variables, the training process will take longer and feature selection tends to choose uninformative features [7]. By using the K-Nearest Neighbor method, machine learning can learn rainfall data patterns to predict rainfall data according to the criteria of average temperature, average wind speed, average humidity, and rainfall [6].

Research related to comparative analysis that has been done is by creating a forecasting model that aims to estimate rental prices and asset values using apartment characteristics, socio-economic variables, and crime rates using machine learning techniques including Random Forest, Decision Tree, and Gradient Boosting Machine algorithms [16]. There are other studies that predict the binding mode of flexible polypeptides to proteins is an important task that falls outside the domain of application of most small molecule and protein-protein docking tools [17].

### 2.2 Research Methods

After all the necessary data is collected, both primary data and secondary data, the next step is data analysis. The purpose of data analysis is to find unique patterns or the most effective results in predicting Cryptocurrency from the data collected by comparing the results of the five existing algorithms. The efficiency of the model is assessed using standard strategic indicators: Root Mean

Squared Error (RMSE) and squared error with lower values of these indicators indicating that the model is effective [18]. The results of Root Mean Squared Error (RMSE) and Squared error ($R^2$) will be used to determine the effectiveness of the algorithm. In the data analysis section, it is key to understand and explore the potential that exists in the five algorithms and the results of processing the available data set. This research uses the RapidMiner in comparing five algorithms. RapidMiner is selected in this paper due to its intuitive visual interface and ability to process data in various formats [19]. This series of research methods begins with data selection, data mining, then data processing. Attribute selection is an important step to ensure that the data used is suitable for the needs of the core research process, aiding subsequent testing and preparation of training data [20]. followed by the application of the model which is used for comparative analysis between five algorithms, and finally the results by looking at the performance of five algorithms based on Root Mean Squared Error (RMSE) and Squared error ($R^2$).

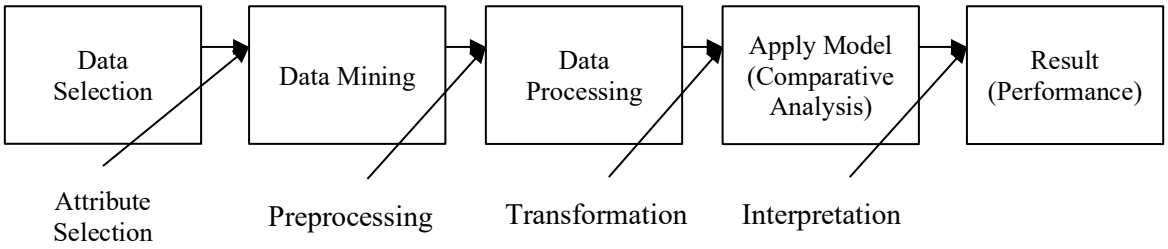


Figure 1. Models Used for Comparative Analysis in RapidMiner Studio

## 3. RESULTS AND DISCUSSION

In the results and discussion section, the results of the research are explained and at the same time a comprehensive discussion is given. The research results will be in the form of the value of the Root Mean Squared Error (RMSE) and the Squared error ($R^2$) value of the five algorithms. The values of Root Mean Squared Error (RMSE) and Squared error ($R^2$) are used for comparative analysis to find out which algorithm is better at predicting the value of bitcoin as a form of cryptocurrency prediction.

### 3.1. Data Selection

In the data selection stage, the dataset is obtained from the public website kaggle.com and consists of 2,652 records of daily price states (5 values per week) covering the time span from November 28, 2014 to March 01, 2022. The data is considered sufficient, since the dataset consists of >1,000 annotated samples and has a rich distribution [21]. In addition, the same period for all cryptocurrencies investigated in this study are considered. The following is Cryptocurrency data from the Bitcoin coin taken from www.kaggle.com which includes Date, Open, High, Low, Close, and Volume BTC as shown in table 1.

Table 1. Bitcoin Dataset

| date | open | high | low | close | Volume BTC |
|---|---|---|---|---|---|
| 2022-03-01 | 43221,71 | 43626,49 | 43185,48 | 43185,48 | 49,0062887 |
| 2022-02-28 | 37717,1 | 44256,08 | 37468,99 | 43178,98 | 3160,61807 |
| 2022-02-27 | 39146,66 | 39886,92 | 37015,74 | 37712,68 | 1701,817043 |
| ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... |
| 2014-11-29 | 376,42 | 386,6 | 372,25 | 376,72 | 2746157,05 |
| 2014-11-28 0 | 363,59 | 381,34 | 360,57 | 376,28 | 3220878,18 |

### 3.2. Data Mining

In this research, the data used comes from the public website www.kaggle.com with the search keyword "Bitcoin Cryptocurrency". Then, the data will go through a customization process to ensure that it can be effectively processed using RapidMiner with the five algorithm method to compare which one is better. This process ensures that the data is ready to be used for further exploration and proper model development. This customization step is necessary to ensure that the data used is appropriate and can be optimized for analysis using the five algorithms method.

### 3.3. Preprocessing Data

In this study, the first step in processing data using the RapidMiner tool is to set the attributes, eliminating missing values in the dataset used, and labels of each variable that must be done at the beginning. Then, the next step is to divide the data into two parts, namely for training and testing data using the split data feature. The setting used is 80:20 for training and testing. This procedure is performed without the use of windowing, making it more efficient and unaffected by the chosen window length [22]. Other studies have also shown that non-windowing methods are more effective in detecting outliers in time series data with large data lengths [23]. With a smaller number of outliers, the algorithm can learn more complex data patterns and produce more accurate predictions [24]. Moreover, the data used in this study is data with a temporal form so that the non-windowing method can be more effective. Certain methods can be more effective when the data has a temporal shape that allows a non-windowing approach [25]. These non-windowing methods can be particularly effective when the temporal shape of the data allows for it, such as in cases where the data exhibits strong periodic patterns or trends that can be captured through other means [26]. The five algorithms that will be compared in this study are operated simultaneously so that the next setting uses the multiply feature so that it can distribute data to each algorithm simultaneously.

### 3.4. Application of Comparative Analysis

In the application stage of comparative analysis, the order used starting from the top is Deep Learning, Decision Tree, Gradient Boosted Tree, Random Forest and finally k-NN. The use of the apply model feature is used to operate the existing algorithm model. The performance operator is used to determine performance criteria which include Root Mean Squared Error (RMSE) and Squared error ($R^2$).

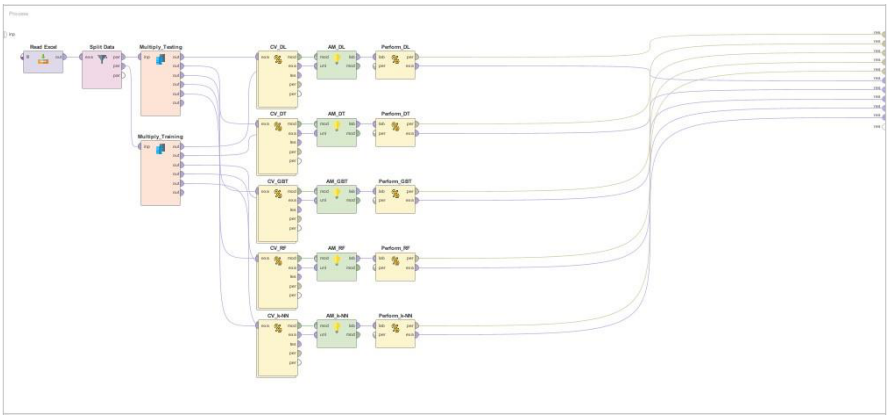


Figure 2. Models Used for Comparative Analysis in RapidMiner Studio

### 3.5. Evaluation Performance

Based on the results of testing five algorithms for prediction comparing them using cryptocurrency data using the RapidMiner application. The efficiency of the model is assessed using standard strategic indicators: Root Mean Squared Error (RMSE) and Squared error ($R^2$) with lower values of these indicators indicating that the model is effective [18]. The results of Root Mean Squared Error (RMSE) and Squared error ($R^2$) will be used to determine the effectiveness of the algorithm. The differences obtained from each algorithm will be drawn to a final conclusion as a result of the research. The drawback of this model is that the article lacks a description of how the model's performance is validated. There is no mention of cross-validation or separation strategies (e.g., K-fold cross-validation) used to ensure the accuracy of the model is unbiased. The shortcomings of this study can be used as a reference for other studies in optimizing the comparison of 5 algorithms in the prediction model.

### 3.6. Discussion

In this discussion section, the results of the process in RapidMiner Studio will be presented with the output in the form of a chart of the Bitcoin coin cryptocurrency, RMSE value, and Squared

error ($R^2$) value. Based on the table of the process results in RapidMiner, the close prediction value, while the close value on the table.

Table 2. RMSE and $R^2$ results of five algorithms in predicting bitcoin close value

| | Deep Learning | Decision Tree | Gradient Boosted Tree | Random Forest | k-NN |
|---|---|---|---|---|---|
| RMSE | 715.284 | 551.719 | 10072.705 | 434.470 | 1170.515 |
| $R^2$ | 511631.732 +/- 1310777.293 | 304394.229 +/- 1154398.546 | 101459386.001 +/- 216379502.704 | 188764.417 +/- 738084.053 | 1370106.171 +/- 7000863.972 |

In the Deep Learning algorithm, the RMSE value is 715.284 and the $R^2$ value is 511631.732 +/- 1310777.293. In the Decision Tree algorithm, the RMSE value is 551.719 and the $R^2$ value is 304394.229 +/- 1154398.546. In the Gradient Boosted Trees algorithm, the RMSE value is 10072.705 and the $R^2$ value is 101459386.001 +/- 216379502.704. In the Random Forest algorithm, the RMSE value is 434.470 and the $R^2$ value is 188764.417 +/- 738084.053. In the k-NN algorithm, the RMSE value is 1170.515 and the $R^2$ value is 1370106.171 +/- 7000863.972.

From the results obtained, each model used to predict cryptocurrency with each algorithm produces a chart graph that is almost in accordance with the existing close value. The RMSE and Squared error ($R^2$) results also show that Random forest is an algorithm with good results, followed by Decision Tree, Deep Learning, k-NN, and Gradient Boosted Tree. Random Forest can quickly process training data well in terms of predicting with RMSE results of 434,470. Decision Tree is 551,719, then deep learning is 715,284, then k-NN is 1170,515, and Gradient Boosted Tree is 10072,705.

## 4. CONCLUSION

Based on the results of this study, the five algorithms compared can model the prediction of changes in the bitcoin cryptocurrency. Random Forest can quickly process training data well in terms of predicting with RMSE results of 434,470. The Decision Tree is 551,719, then deep learning is 715,284, and k-NN is 1170,515. It was also found that the Gradient Boosted Tree is 10072,705 less if a comparative analysis is carried out with other methods. suggestions for further research are that this method can be applied to data with diverse variables and adjustments by focusing on similar methods so that varied results are obtained.

## REFERENCES


[1] J. K. Chahal, N. Bhatia, G. Singh, V. Gandhi, And P. Kaushal, "Cryptocurrency In Modern Finances," In *1st International Conference On Applied Data Science And Smart Systems, Adsss 2022*, 2023, P. 140001. Doi: https://doi.org/10.1063/5.0177802

[2] A. M. Shantaf, A. S. Mustafa, M. M. Hamdi, M. B. Abdulkareem, A. D. Salman, And Y. H. Sulaiman, "A Review Of Cryptocurrencies: Tether And Ethereum," In *2nd International Conference On Emerging Technology Trends In Internet Of Things And Computing, Tiotc 2022*, 2023, P. 020019. Doi: https://doi.org/10.1063/5.0188274

[3] R. Thinakaran, T. B. Kurniawan, And M. Batumalay, "Algorithmic Trading Strategy Development Using Machine Learning," *J. Eng. Sci. Technol.*, Vol. 18, No. 6, Pp. 22–31, 2023.

[4] R. Risanti, "Analisis Model Prediksi Cuaca Menggunakan Support Vector Machine, Gradient Boosting, Random Forest, Dan Decision Tree," 2024, Vol. Xii, Pp. 119–128. Doi: https://doi.org/10.21009/03.1201.Fa18

[5] N. S. Ja'afar, J. Mohamad, And S. Ismail, "Machine Learning For Property Price Prediction And Price Valuation: A Systematic Literature Review," *Plan. Malaysia*, Vol. 19, No. 3, Pp. 411–422, Oct. 2021, Doi: https://doi.org/10.21837/Pm.V19i17.1018

[6] D. M. Nanda And P. N. Pudjiantoro, Tacbir Hendro Sabrina, "Metode K-Nearest Neighbor (Knn) Dalam Memprediksi Curah Hujan Di Kota Bandung," *Snestik*, Pp. 387–393, 2022, Doi: Https://Doi.Org/10.31284/P.Snestik.2022.2750

[7] E. A. F. Elmuna, T. Chamidy, And F. Nugroho, "Optimization Of The Random Forest Method Using Principal Component Analysis To Predict House Prices," *Int. J. Adv. Data Inf. Syst.*, Vol. 4, No. 2, Pp. 155–166, Oct. 2023, Doi: https://doi.org/10.25008/Ijadis.V4i2.1290

[8] R. Eliviani And D. D. Wazaumi, "Exploring Sentiment Trends: Deep Learning Analysis Of Social Media Reviews On Google Play Store By Netizens," *Int. J. Adv. Data Inf. Syst.*, Vol. 5, No. 1, Pp. 62–70, Mar. 2024, Doi: https://doi.org/10.59395/Ijadis.V5i1.1318

[9] E. Yustina, M. A. Hariyadi, And C. Crysdian, "Recommendation Of Prospective Construction Service Providers In Government Procurement Using Decision Tree," *Int. J. Adv. Data Inf. Syst.*, Vol. 5, No. 1, Pp. 39–48, Mar. 2024, Doi: https://doi.org/10.59395/Ijadis.V5i1.1316

[10] I. Nasios And K. Vogklis, "Blending Gradient Boosted Trees And Neural Networks For Point And Probabilistic Forecasting Of Hierarchical Time Series," *Int. J. Forecast.*, Vol. 38, No. 4, Pp. 1448–1459, Oct. 2022, Doi: https://doi.org/10.1016/J.Ijforecast.2022.01.001

[11] M. H. Diponegoro, S. S. Kusumawardani, And I. Hidayah, "Tinjauan Pustaka Sistematis : Implementasi Metode Deep Learning Pada Prediksi Kinerja Murid ( Implementation Of Deep Learning Methods In Predicting Student Performance : A Systematic Literature Review )," Vol. 10, No. 2, Pp. 131–138, 2021.

[12] N. G. Ramadhan, Y. Setiya, R. Nur, And F. D. Adhinata, "Pendekatan Deep Learning Untuk Prediksi Durasi Perjalanan Deep Learning Approach For Trip Duration Prediction," Vol. 11, No. 2, Pp. 85–89, 2022, Doi: https://doi.org/10.34148/Teknika.V11i2.460

[13] J. Gu, C. Sun, And H. Zhao, "Densetnt: End-To-End Trajectory Prediction From Dense Goal Sets," *Proc. Ieee Int. Conf. Comput. Vis.*, Pp. 15283–15292, 2021, Doi: https://doi.org/10.1109/Iccv48922.2021.01502

[14] L. Syafa'ah And M. Lestandy, "Penerapan Deep Learning Untuk Prediksi Kasus Aktif Covid-19," *J. Sains Komput. Inform.*, Vol. 5, No. 1, Pp. 453–457, 2021.

[15] S. Sandiwarno, "Penerapan Machine Learning Untuk Prediksi Bencana Banjir," *J. Sist. Inf. Bisnis*, Vol. 14, No. 1, Pp. 62–76, 2024, Doi: https://doi.org/10.21456/Vol14iss1pp62-76

[16] D. Noorcahya, A. P. Rifai, A. Darmawan, And W. P. Sari, "Prediction Of Apartment Price Considering Socio Economic And Crime Rates Factors In Dki Jakarta," *Int. J. Adv. Data Inf. Syst.*, Vol. 4, No. 2, Pp. 145–154, Sep. 2023, Doi: https://doi.org/10.25008/Ijadis.V4i2.1294

[17] T. Chi, J. Li, L. G. Trigeorgis, And A. E. Tsekrekos, "Real Options Theory In International Business," *J. Int. Bus. Stud.*, Vol. 50, No. 4, Pp. 525–553, Jun. 2019, Doi: https://doi.org/10.1057/S41267-019-00222-Y

[18] M. Vijh, D. Chandola, V. A. Tikkiwal, And A. Kumar, "Sciencedirect Stock Closing Price Prediction Using Machine Learning Techniques," *Procedia Comput. Sci.*, Vol. 167, No. 2019, Pp. 599–606, 2020, Doi: https://doi.org/10.1016/J.Procs.2020.03.326

[19] [1]P. M. Hasugian and P. B. N. Simangunsong, "Apriori Algorithm Testing Using The RapidMiner Application," *J. Info Sains : Informatika dan Sains*, vol. 13, no. 01, pp. 33–40, Mar. 2023, Available: http://ejournal.seaninstitute.or.id/index.php/InfoSains

[20] Sulika, R. Kusumawati, and Y. M. Arif, "Classification of Students' Academic PerformanceUsing Neural Network and C4.5 Model," *Int. J. Adv. Data Inf. Syst.*, vol. 5, No. 1, Pp. 29-38, Apr. 2024, doi: https://doi.org/10.59395/ijadis.v5i1.1311

[21] A. Safonova, G. Ghazaryan, S. Stiller, M. Main-Knorn, C. Nendel, and M. Ryo, "Ten deep learning techniques to address small data problems with remote sensing," *Int. J. of Applied Earth Observation and Geoinformation*, vol. 125, p. 103569, Dec. 2023, doi: https://doi.org/10.1016/j.jag.2023.103569

[22] M. Ridwan, "Penentuan Panjang Optimal Data Deret Waktu Bebas Outlier dengan Menggunakan Metode Window Time," *repository ITS*, 2018, Accessed: Jul. 17, 2024. [Online]. Available: https://repository.its.ac.id/

[23] R. S. Aulia and R. M. Atok, "Penentuan Panjang Optimal Data Deret Waktu Bebas Outlier dengan Menggunakan Metode Window Time," *J. Sains dan Seni ITS*, vol. 6, no. 1, Mar. 2017, doi: https://doi.org/10.12962/j23373520.v6i1.22520

[24] M. R. Firmansyah, "Forecasting Data Deret Waktu dengan Pendekatan Neural Network," *J. Manajemen, Strategi Bisnis dan Kewirausahaan*, 2019, Accessed: Jul. 17, 2024. [Online]. Available: https://ojs.unud.ac.id/index.php/jmbk/VisitorStatistik

[25] M. Kvet and K. Matiaško, "Study on Effective Temporal Data Retrieval Leveraging Complex Indexed Architecture," *Appl. Sci.*, Nov. 2019, Accessed: Aug. 15, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/11/3/916

[26] Tao Li and Sheng Ma, "Mining temporal patterns without predefined time windows," Fourth IEEE International Conference on Data Mining (ICDM'04), Brighton, UK, 2004, pp. 451-454, doi: https://10.1109/ICDM.2004.10016.