Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13418
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparison of Performance of K-Nearest Neighbors and Neural Network Algorithm in Bitcoin Price Prediction

Eko Aziz Apriadi[1)*], Sriyanto[2)], Sri Lestari[3)], Suhendro Yusuf Irianto[4)]
[1)2)3)4)]Fakultas Ilmu Komputer, Insitut Informatika dan Bisnis Darmajaya, Lampung, Indonesia
[1)]ekoazizapriadi04@gmail.com, [2)]sriyanto@darmajaya.ac.id, [3)]srilestari@mail.darmajaya.ac.id, [4)]suhendroyusuf@darmajaya.ac.id

Submitted :Jan 21, 2024 | **Accepted** : Feb 12, 2024 | **Published** : Apr 1, 2024

Abstract: This research evaluates and compares the performance of two prediction methods, namely K-Nearest Neighbors (K-NN) and Neural Network, in the context of Bitcoin price prediction. Historical Bitcoin price data is used as input to train and test both algorithms. Experimental results show that the K-NN algorithm produces a Root Mean Square Error (RSME) of 389,770 and a Mean Absolute Error (MAE) of 89,261, while the Neural Network has an RSME of 614,825 and an MAE of 284,190. Performance comparison analysis shows that, on this dataset, K-NN has better performance in predicting Bitcoin prices compared to Neural Network. These findings provide important insights for the selection of crypto asset price prediction models, especially Bitcoin, in financial and investment environments

Keywords: K-Nearest Neighbors (K-NN), Neural Network, Bitcoin Price Prediction, Root Mean Square Error (RSME), Mean Absolute Error (MAE)

INTRODUCTION

Bitcoin is a digital currency that was introduced in 2009 and is used as an online payment tool. Bitcoin has high volatility, meaning its value can change quickly, even within hours or minutes. Since its inception, the price of Bitcoin has experienced a significant increase and attracted investors from all over the world. However, Bitcoin's high price volatility also poses risks for investors. Bitcoin price prediction has become a topic of interest for many researchers and investors. In this context, Bitcoin price predictions are very important for investors who want to make smart investment decisions. In previous research, the ARIMA model (*Autoregressive Integrated Moving Average*) is a popular statistical model for predicting time series. This model combines three components: autoregression (AR), moving average (MA) component, and differentiation (Ley 25.632, 2002). In the context of Bitcoin price prediction, the ARIMA model can be used to extract trends, seasonal patterns and random fluctuations in historical data and provide future price predictions (Wahyu & Hendrik, 2023). In previous research cited (Luxmana & Oktafiyani, 2022) Regression Analysis: This algorithm tries to build relationships mathematical relationship between independent variables (e.g., historical prices, trading volume, etc.). with Bitcoin prices. Through regression analysis, the model can try to predict the future price of Bitcoin based on historical data. However, the weakness of this approach is that it does not consider external factors that might influence prices. Therefore, a lot of research has been done to develop accurate Bitcoin price prediction models. One of the approaches used is the K-NN algorithm and Neural Network.

Currently, Bitcoin has become a popular digital currency and is a type of investment that is popular with many people around the world. Due to its highly volatile price, Bitcoin price predictions are important for investors to decide when to buy or sell Bitcoin. Several previous studies have been carried out to predict Bitcoin prices using techniques such as statistical analysis and mathematical models. However, prediction accuracy remains a major challenge, especially due to the high volatility in Bitcoin prices. In this research, using a comparison of the performance of the k-nearest neighbors algorithm and neural network in predicting Bitcoin prices, then at the end of the research, it is hoped that this model will be able to provide predictions of Bitcoin selling prices**.**

LITERATURE REVIEW

Bitcoin or abbreviated as "BTC" is a digital currency, which is not issued by institutions, organizations or governments under its regulations. Bitcoin utilizes a peer-to-peer network as a distribution medium using advanced cryptographic protocols. First proposed by Satoshi Nakamoto in 2008, bitcoin software was created and started running in 2009 (Iii & Teori, n.d.). A bitcoin (electronic coin) value is a series of digital signatures. All Bitcoin

* Corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

transactions are posted in blocks to an open ledger known as the Blockchain to be verified by miners using cryptographic proofs. This verification occurs in a trustless system and with no intermediaries required to pass funds from sender to recipient. Bitcoin offers new opportunities for prediction due to its relative youth and the resulting volatility. Additionally, it is unique in relation to traditional fiat currencies in terms of its open nature. Unlike traditional currencies and assets, bitcoin is portable, divisible, and immutable. Bitcoin increases the efficiency of the system and allows the provision of financial services at much lower costs, giving users more power and freedom. In comparison, there is no complete data regarding cash transactions or money in fiat currency circulation. The famous efficient market hypothesis (James A.F Stoner, 1988). indicates the price of an asset such as a currency reflects all available information, and as a result trades at its fair value. Although there is a wealth of data available related to Bitcoin and its network, not all market participants will utilize all this information effectively and the consequences may not be reflected in the price (Penelitian, 2008).

Data mining is a term used to describe the discovery of knowledge in databases. Data mining is a process that uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and related knowledge from various large databases (Nengsih, 2019). Data mining is the analysis of reviewing data sets to find unexpected relationships and summarize data in a different way than before, which is understandable and useful for the data owner.

In general, data mining model measurements refer to three criteria: Accuracy, Reliability and Usefulness. A balance between the three is necessary because an accurate model is not necessarily reliable, and one that is reliable or accurate is not necessarily useful. The methods used in data mining are learning methods (supervised learning) and methods without learning (Unsupervised learning). Learning methods include the roles of estimation, prediction, classification and association while non-learning methods include clustering (Anggada Maulana, 2018).

Data mining is a term used to describe the discovery of knowledge in databases. Data mining is a process that uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and related knowledge from various large databases (Nengsih, 2019). Data mining is the analysis of reviewing data sets to find unexpected relationships and summarize data in a different way than before, which is understandable and useful for the data owner.

In general, data mining model measurements refer to three criteria: Accuracy, Reliability and Usefulness. A balance between the three is necessary because an accurate model is not necessarily reliable, and one that is reliable or accurate is not necessarily useful. The methods used in data mining are learning methods (supervised learning) and methods without learning (Unsupervised learning). Learning methods include the roles of estimation, prediction, classification and association while non-learning methods include clustering (Anggada Maulana, 2018).

The KNN algorithm is an algorithm that is widely used in classification. KNN is an algorithm that is simple to implement but produces good accuracy. One of the weaknesses of this algorithm is in determining the k value, if the k value is too large it will make the classification results unclear or blurry, whereas if the k value is too small or for example k=1 will cause the classification results to feel stiff because there are no choices. Therefore, research is needed to determine a good k value (Khairiyah, 2020).

Neural Network is a category of Soft Computing science. Neural Networks actually adopt the ability of the human brain which is able to provide stimulation, carry out processes, and provide output. Output is obtained from variations in stimulation and processes that occur in the human brain. Human ability to process information is the result of the complexity of processes in the brain. For example, what happens to children is that they are able to learn to perform recognition even though they do not know what algorithm to use. The extraordinary computing power of the human brain is an advantage in scientific studies.

The functions of a Neural Network include:

1. Pattern classification
2. Mapping the pattern obtained from the input into a new pattern at the output
3. Store the pattern to be recalled
4. Map out similar patterns
5. Problem optimization
6. Prediction

The development of Neural Network science has been around since 1943 when Warren McCulloch and Walter Pitts introduced the first neural network model calculations. They combine several simple processing units together to provide an overall increase in computing power. This was continued in research carried out by Rosenblatt in 1950, where he succeeded in discovering a two-layer network, which was called a perceptron. Perceptron allows for specific learning classification work by adding weights to each inter-network connection (Syafarina, 2016).

Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of our model's predictions that are correct. Formally, accuracy has the following definition (Perbandingan et al., 2018) :

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13418
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$\text{Akurasi} = \frac{\text{Number of Correect Prediction}}{\text{Total Number of Prediction}}$$

For binary classification, accuracy can also be calculated in positive and negative terms as follows:

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Where :
TP : True Positif
TN : True Negatif
FP : False Positif
FN : False Negatif

METHOD

In carrying out analysis and looking for data patterns to be used as a dataset to make research easier and able to run systematically and fulfill the desired objectives, a flow is created in the research stages which will be carried out as follows:

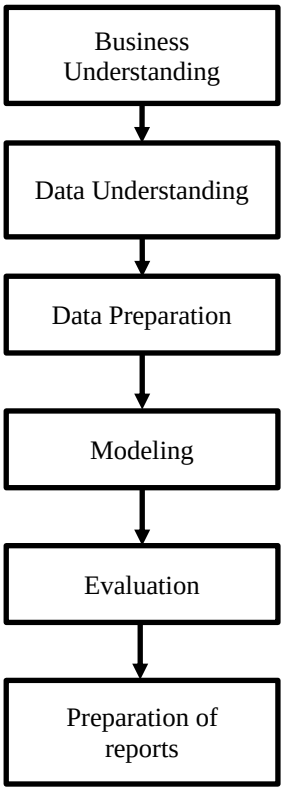


Figure 3.1 Flow in Research Stages

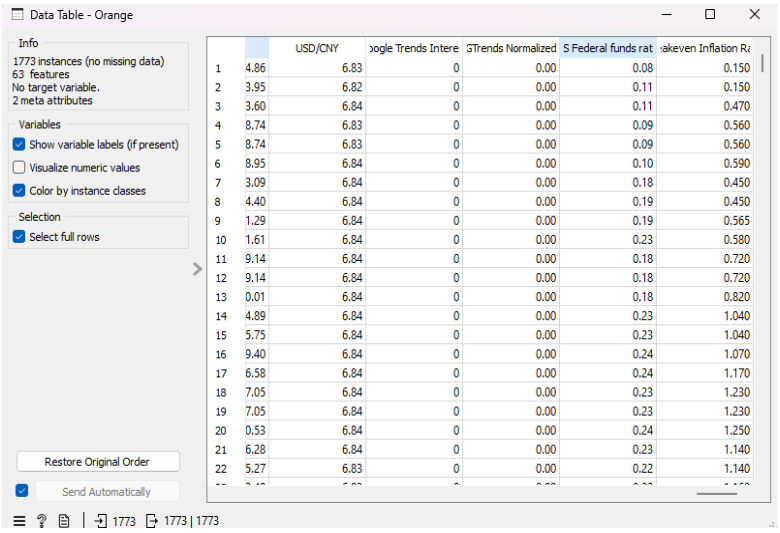
RESULT

In this section, the researcher will explain the results of the research obtained. Researchers can also use images, tables, and curves to explain the results of the study. These results should present the raw data or the results after applying the techniques outlined in the methods section. The results are simply results; they do not conclude.

The data used in this research is a dataset from Kaggle, namely Bitcoin data. The number of data records is 1773 data. This data set is used to predict bitcoin. Bitcoin data is in Figure 4.1 below:

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13418
e-ISSN : 2541-2019
p-ISSN : 2541-044X



		USD/CNY	ogle Trends Intere	GTrends Normalized	S Federal funds rat	akeven Inflation Ra
1	4.86	6.83	0	0.00	0.08	0.150
2	3.95	6.82	0	0.00	0.11	0.150
3	3.60	6.84	0	0.00	0.11	0.470
4	8.74	6.83	0	0.00	0.09	0.560
5	8.74	6.83	0	0.00	0.09	0.560
6	8.95	6.84	0	0.00	0.10	0.590
7	3.09	6.84	0	0.00	0.18	0.450
8	4.40	6.84	0	0.00	0.19	0.450
9	1.29	6.84	0	0.00	0.19	0.565
10	1.61	6.84	0	0.00	0.23	0.580
11	9.14	6.84	0	0.00	0.18	0.720
12	9.14	6.84	0	0.00	0.18	0.720
13	0.01	6.84	0	0.00	0.18	0.820
14	4.89	6.84	0	0.00	0.23	1.040
15	5.75	6.84	0	0.00	0.23	1.040
16	9.40	6.84	0	0.00	0.24	1.070
17	6.58	6.84	0	0.00	0.24	1.170
18	7.05	6.84	0	0.00	0.23	1.230
19	7.05	6.84	0	0.00	0.23	1.230
20	0.53	6.84	0	0.00	0.24	1.250
21	6.28	6.84	0	0.00	0.23	1.140
22	5.27	6.83	0	0.00	0.22	1.140

Figure 4.1 Bitcoin dataset

The application of the data in orange for Bitcoin predictions using the K-NN algorithm and Neural Network is shown in Figure 4.2 below:

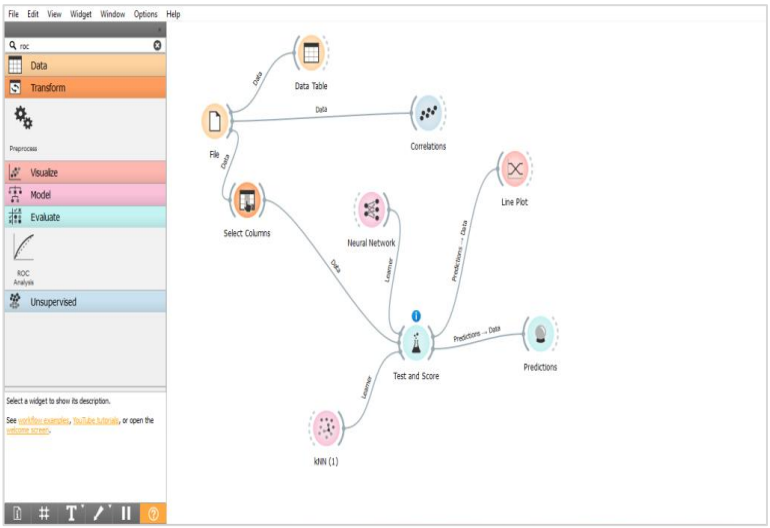


Figure 4.2 Application of Bitcoin Prediction Data Using the K-NN Algorithm and Neural Network on Orange

In Figure 4.2. It is a process of evaluating the comparative results of the prediction models being tested. The process of calculating the success of the prediction model in the Orange tool can use Test and Score.

DISCUSSIONS

Data that has gone through the preprocessing stage is then tested to obtain the best prediction model. The prediction model that has been tested and evaluated using a collection of test data on the Orange application where 1 attribute is the target, obtained simulation results as in Figure 4.3

* Corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 2, April 2024
DOI : https://doi.org/10.33395/sinkron.v8i2.13418
e-ISSN : 2541-2019
p-ISSN : 2541-044X

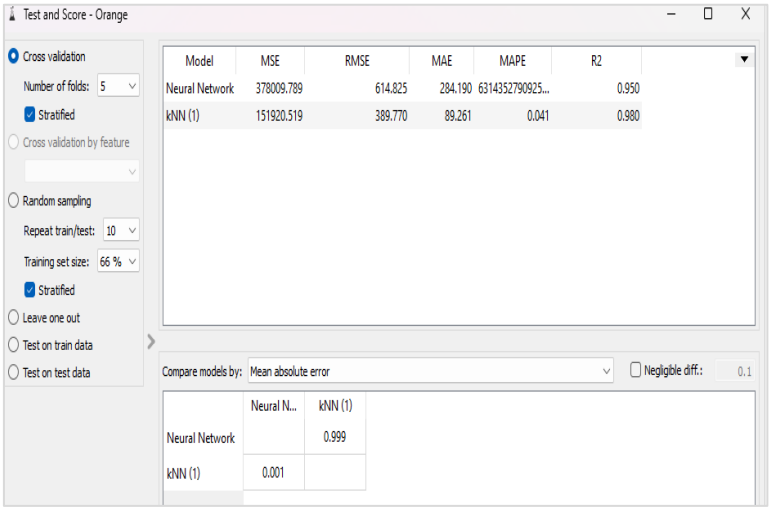


Figure 4.3 Test and Score Results

Figure 4.3 is a display of the Test and Score widget in the Orange application, providing a visual view of the prediction model evaluation process in this research. The testing process is carried out by applying the K-Fold Cross Validation method (K=5), which can be set and configured via the Test and Score widget as shown in the picture. The K-Fold Cross Validation method is used to ensure the reliability and generalization of the model by dividing the dataset into 5 subsets or "folds," where each fold is used as test data once, and the iteration is repeated K times.

The Test and Score widget also provides evaluation results which include a comparison between two prediction methods, namely K-NN (K-Nearest Neighbors) and Neural Network. Evaluation is carried out by taking into account a number of performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2). This research specifically focuses attention on the MAE and RMSE values as indicators of prediction model performance.

The evaluation results show significant differences between the two Bitcoin price prediction models. K-NN obtained an MAE value of 89,261 and RMSE of 389,770, while Neural Network had a higher value. These two values reflect the level of accuracy and precision of model predictions. With lower MAE and RMSE values, K-NN is considered better at predicting Bitcoin price movements compared to Neural Network. This conclusion indicates that, based on the evaluation results, K-NN is the best method for Bitcoin price prediction in the context of this research. The experimental results of this research using the KNN and NN algorithms are shown in the table below:

Table. 4.1 Comparison of KNN and NN Experimental Results:

Algorithm	RSME	MAE
K-NN	389.770	89.261
Neural Network	614.825	284.190

CONCLUSION

The results of implementing Orange Data Mining for Bitcoin price prediction can be seen in the Test and Score widget, K-NN obtained an MAE value of 89,261 and RMSE of 389,770, while the Neural Network had a higher value. These two values reflect the level of accuracy and precision of model predictions. Based on the RMSE and MAE values, it can be concluded that with lower MAE and RMSE values, K-NN is considered better in predicting Bitcoin price movements compared to Neural Network. This conclusion indicates that, based on the evaluation results, K-NN is the best method for Bitcoin price prediction in the research context.

REFERENCES

Anggada Maulana. (2018). Konsep Dasar Data Mining. *Konsep Data Mining*, *1*, 1–16.

* Corresponding author

Iii, B. A. B., & Teori, L. (n.d.). *Landaasan teori cryptocurrancy*. 9–41.

James A.F Stoner, 2018 : 41). (1988). Landasan Teori اديدج. *Dasar-Dasar Ilmu Politik*, *13*, 17–39.

Khairiyah, A. (2020). *Klasifikasi Jenis Buah Jambu Berdasarkan Daun Menggunakan Metode Principal Component Analysis*. 18–18.

Ley 25.632. (2002). *済無No Title No Title No Title*. 10–66.

Luxmana, D. B., & Oktafiyani, M. (2022). Analisis Fundamental Cryptocurrency Terhadap Fluktuasi Harga Pada Masa Pandemi. *Dinamika Akuntansi Keuangan Dan Perbankan*, *11*(1), 41–52. https://doi.org/10.35315/dakp.v11i1.8952

Nengsih, W. (2019). Analisa Akurasi Permodelan Supervised Dan Unsupervised Learning Menggunakan Data Mining. *Sebatik*, *23*(2), 285–291. https://doi.org/10.46984/sebatik.v23i2.771

Penelitian, L. B. (2008). *Bab I* ◌ ◌. 1–120.

Perbandingan, A., Naive, A., Mendapatkan, M. U., Desa, B. D., Pasca, P., Magister, S., Komputer, I., Tinggi, S., Informatika, M., Komputer, D. A. N., & Mandiri, N. (2018). *Analisis perbandingan algoritma naive bayes , random forest dan c.45 dalam klasifikasi kelayakan masyarakat untuk mendapatkan bantuan dana desa*.

Syafarina, G. A. (2016). Penerapan Algoritma Neural Network Dalam Menentukan Prioritas Pengembangan Jalan Di Provinsi Kalimantan Selatan. *Technologia : Jurnal Ilmiah*, *7*(2), 80–88. https://doi.org/10.31602/tji.v7i2.618

Wahyu, F., & Hendrik, B. (2023). Perbandingan Algoritma Time Series Dan Fuzzy Inference System Dalam Analisis Data Deret Waktu. *Jurnal Penelitian Teknologi Informasi Dan Sains*, *1*(3), 16–24. https://doi.org/10.54066/jptis.v1i3.711