International Conference on ATLAS (Advanced Technologies, Learning Algorithms, and Systems)
ATLAS

Implementation of the Random Forest Algorithm in Predicting Bitcoin (BTC) Prices

Steven Owen[b], Muhammad Siddik[b*], Wilda Susanti[b*], Gustientiedina[b*]

[b]Department of Computer Science, Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia

Article History
Received
X Month 202Y
Received in revised form
XA Month 202Y
Accepted
XB Month 202Y
Published Online
XC Month 202Z

*Corresponding author
siddik@lecturer.pelitaindonesia.ac.id
*Please provide an **official organization email** of the corresponding author*

Abstract
The fluctuating price of Bitcoin is widely used as a means of generating profit. The instability of Bitcoin price movements causes fluctuations, which traders utilize as a trading strategy. The purpose of this research is to analyze and predict Bitcoin prices to determine buying and selling strategies. The Random Forest algorithm is a machine learning method that combines multiple decision trees for classification and prediction with high accuracy. The classified data consists of Bitcoin price data from January 2024 to November 2024. The results of this study show that the accuracy level of the Random Forest algorithm using a 60:40 and 80:20 data split achieves an accuracy of 50.75% and 47.76%, respectively. The most suitable data split for Bitcoin price prediction is 60:40 and 80:20. This is because a balanced data split can influence the Random Forest algorithm's calculations, maintaining accuracy in predicting Bitcoin prices, thereby helping investors predict Bitcoin prices for buying and selling strategies.

Keywords: Classification, Decision Tree, Random Forest, Prediction, Bitcoin

1.0 INTRODUCTION

The development of information technology today significantly influences the world of investment. Information Technology is a technology used to process data, including processing, obtaining, organizing, storing, and manipulating data in various ways to produce high-quality information, i.e., information that is relevant, accurate, and timely, used for personal, business, and government purposes, and is strategic for decision-making (Cecep Abdul Cholik, 2021). The development of information technology has led to many changes, not only for individuals but also for companies, which are required to adapt to the ongoing technological advancements (Bangun Pasaribu & Wilda Susanti, 2021).

According to Pramita & Hendrayana (2021), investment is an activity that companies use to grow wealth (acceleration of wealth) through the distribution of investment returns (such as interest, royalties, dividends, and rent), for the appreciation of investment value, or for other benefits for the investing company, such as benefits obtained through trade relationships. As more knowledge is gained about investment, whether through learning or capital market socialization, individuals are increasingly motivated to invest in the capital market (Nadila et al., 2023).

The emergence of new companies can be one of the driving factors for the growing number of investors, especially in cryptocurrency stock investments. Investment is an effort made by an individual or business entity to gain profit from the money they have (Naufal Hasani, 2022). According to Aulia (2019), Bitcoin prices fluctuate rapidly each year, with a significant decline in 2018, where the Bitcoin price dropped by USD 6,955.27 compared to the 2017 price of USD 19,497.40. According to Putra & Robiyanto, (2021), cryptocurrency has experienced significant growth in recent years. For example, the price of Bitcoin at the end of 2017 was around IDR 180,386,628, and by the end of 2020, Bitcoin's price increased to IDR 403,315,071.

The high frequency of the crypto market is likely influenced by the active transactions of Bitcoin's price, as there is considerable interest from investors in Bitcoin. In general, changes in investor interest in a digital asset are reflected in Bitcoin price fluctuations in the crypto market. In Indonesia, cryptocurrency has gained government attention and has become subject to regulations. For the government, the main challenge in this policy is to create adequate and harmonious regulations that can guide economic activities (Rohman, 2021).

The cryptocurrency system has attracted the attention of various investors, one of which is Bitcoin. Bitcoin is a form of payment that uses cryptography or special encryption algorithms to control the management and creation of Bitcoin itself (Purnomo et al., 2022). Bitcoin is also one of the most famous and widely traded cryptocurrencies in the world. It was created in 2009 by an anonymous individual or group using the name Satoshi Nakamoto. Since its launch, Bitcoin's price has been highly volatile and difficult to predict, but the price of Bitcoin tends to rise frequently. In one day, the price of Bitcoin can rise dramatically or fall, although not as much as it increases. Bitcoin was initially launched with no value at all (Gaol Lumban Humphrey, 2022). Bitcoin is favored because of its transparent system, where transactions can be viewed by anyone, making Bitcoin a promising investment opportunity for investors (Aulia, 2019). This opportunity for significant price changes often becomes the main focus in market analysis due to price fluctuations that reflect market sentiment and quick changes in investor interest.

Bitcoin's volatile price makes it widely used as a means to seek profit. This price instability is leveraged by traders to engage in trading strategies focused on price speculation for both short-term and long-term profits. Therefore, analysis and prediction are needed to understand these price movements, supported by scientific methods to increase accuracy in predicting future price movements (Al Fajri, 2023). Bitcoin price prediction can be done through various methods, as seen in previous research on Bitcoin price prediction (Indriyanti et al., 2022).

In the study by Saadah & Salsabila, (2021) titled "Bitcoin Price Prediction Using the Random Forest Method," the goal was to analyze Bitcoin price fluctuations during the Covid-19 pandemic. The method used was Random Forest, beginning with data preprocessing on Bitcoin and selecting three attributes: Low, High, and Open, followed by the prediction of Bitcoin prices during the Covid-19 pandemic from January 1, 2019, to October 13, 2019. The study found that the Random Forest algorithm, using the Low, High, and Price attributes, successfully predicted Bitcoin prices with good performance, achieving a MAPE of 1.50%, or an accuracy of around 98% using random data.

Sujjada et al., (2024) conducted a study titled "Bitcoin Price Prediction Using Long Short-Term Memory Algorithm." This study aimed to address the issue of long-term dependencies in price movements that cannot be handled by traditional RNN models. It used the Long Short-Term Memory (LSTM) algorithm, which has the ability to "remember" information over a long period, thus recognizing patterns and trends. The dataset used was Bitcoin data from December 12, 2020, to April 14, 2024. The study found that the RMSE value for training data was 17,318.40 and for test data was 27,921.84, while the MAPE value for training data was 3.24% and for test data was 5.36%. This indicates that the RMSE and MAPE values for the training data were relatively small due to Bitcoin's wide price range.

Based on previous research, the fluctuating price movements of Bitcoin make predicting Bitcoin's price a challenge for investors and market analysts. To predict Bitcoin's volatile price, analysis and prediction using data mining learning algorithms, such as the Random Forest algorithm, which has proven effective in classifying and predicting complex data, is necessary.

Thus, individuals who invest in Bitcoin at market prices can easily predict the potential future price movements. In this study, the author aims to develop the Random Forest algorithm, which can accurately predict the possible future prices of Bitcoin. The main objective of the proposed method is to classify predictions using the Random Forest algorithm for cryptocurrencies, including Bitcoin, with high accuracy.

2.0 LITERATURE REVIEW

Bitcoin Price Movement

Bitcoin is a form of payment that uses cryptography or special encryption algorithms to control the management and creation of Bitcoin itself (Purnomo et al., 2022). Bitcoin is also one of the most famous and widely traded cryptocurrencies in the world. Bitcoin was created in 2009 by an anonymous individual or group using the name Satoshi Nakamoto. Since its launch, Bitcoin's price has been highly volatile and difficult to predict, although its price tends to increase frequently. In a single day, Bitcoin's price can rise dramatically or fall, although not as much as its rise. When Bitcoin was first launched, it had no value at all (Gaol Lumban Humphrey, 2022).

Decision Tree

A decision tree is a classification method that uses a tree structure representation, where each node represents an attribute, each branch represents the value of that attribute, and the leaves represent the class (Partogi & Pasaribu, 2021). This algorithm consists of a collection of decision nodes connected by branches, moving downward from the root node until ending at a leaf node. In a decision tree, there are three types of nodes: Root Node, Internal Node, and Leaf Node (Dewi et al., 2021). According to Permana et al. (2021), one advantage of the decision tree is its flexibility, which can enhance the quality of the decisions made. However, a drawback of this algorithm is the potential for overlap when working with data that has a large number of classes and criteria.

Random Forest

Random Forest is an extension of the Decision Tree method that utilizes multiple Decision Trees, where each tree is trained using individual samples and each attribute is split based on a randomly selected subset of attributes (Supriyadi et al., 2020). Each decision tree is trained using a subset of data selected randomly through Bootstrap Sampling. Bootstrap Sampling is a method of random sampling from the original dataset with replacement. In Bootstrap Sampling, random samples are taken from the source dataset using a rough sampling method, which involves selecting random samples followed by resampling using the bagging technique. These results are then averaged to produce a stronger overall outcome. This is the primary goal of Random Forest (Jollyta et al., 2023). Bagging helps improve unstable classification algorithms such as Decision Trees, which can experience significant changes in prediction, by stabilizing them through aggregation (Saputri et al., 2022).

3.0 METHODOLOGY

In this study, the research method used is the Random Forest algorithm. The workflow of the research process is illustrated in Figure 1 below.

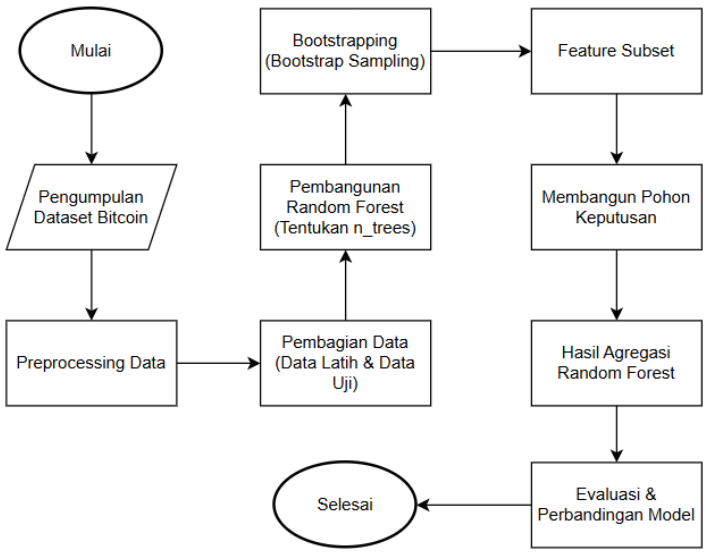

Figure 1. Research Methodology

The stages of the research methodology begin with collecting the Bitcoin dataset, which will serve as the input data for processing. This is followed by the data preprocessing stage, which includes Data Exploration, Data Reduction, and Label Encoding. Afterward, the data is split into training and testing datasets, after which the Random Forest model is built by determining the number of decision trees to be created. Once the desired number of decision trees is determined, bootstrapping is performed, which involves randomly sampling subsets from the dataset over multiple iterations and with a certain number of variables. Then, feature selection is done by choosing the variables suitable for forming a decision tree, calculating Entropy and Information Gain as criteria for selecting the variables that will form the decision tree according to the specified number of trees. Afterward, all the decision trees formed based on the predetermined number of trees are aggregated by counting the number of correct predictions according to the defined classification goal. Finally, model evaluation and comparison are conducted using a confusion matrix that includes Accuracy, Recall, Precision, and F1-Score to test the performance of the built Random Forest model.

4.0 RESULTS AND DISCUSSION

This research was conducted using a Bitcoin dataset obtained from the website investing.com. The data covers the period from January 2024 to November 2024 and consists of 335 records. The prediction process will be carried out using Visual Studio Code for programming.

	Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Vol.	Perubahan%
0	01/01/2024	683.907.008	663.249.984	684.998.976	662.004.992	0,03K	3,11%
1	02/01/2024	701.203.968	683.907.008	712.491.008	683.907.008	0,10K	2,53%
2	03/01/2024	669.945.024	701.203.968	709.000.000	660.000.000	0,10K	-4,46%
3	04/01/2024	686.304.000	669.945.024	694.000.000	666.001.984	0,05K	2,44%
4	05/01/2024	686.377.984	686.304.000	689.580.032	667.000.000	0,06K	0,01%

Figure 2. Bitcoin Dataset

Figures 3 and 4 show the results of data pre-processing on the Bitcoin dataset, where no missing values or duplicate values were found—indicating that there are no empty entries and no duplicate records. After checking for missing values, the dataset was split into variable X and variable Y with ratios of 60:40, 70:30, and 80:20. The pre-processing stage involved data transformation to clean combinations of letters and numbers, feature engineering by categorizing data into Low, Medium, and High, and label encoding to convert categorical values into numerical format for further analysis. The resulting pre-processed data is then used to evaluate model performance. The following table presents the accuracy, precision, recall, and F1-score results of the Random Forest algorithm across the three data split ratios: 60:40, 70:30, and 80:20.

	Closing	Opening	Highest	Lowest	Volume
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
..	...	...	...	...	...
330	1	1	1	1	0
331	1	1	1	1	0
332	1	1	1	1	0
333	1	1	1	1	0
334	1	1	1	1	0

Figure 3. Preprocessing Results of Variable X

```
[335 rows x 5 columns]
0      Turun
1       Naik
2      Turun
3       Naik
4      Turun
       ...
330    Turun
331     Naik
332    Turun
333     Naik
334     Naik
Name: Change%, Length: 335, dtype: object
```

Figure 4. Preprocessing Results of Variable y

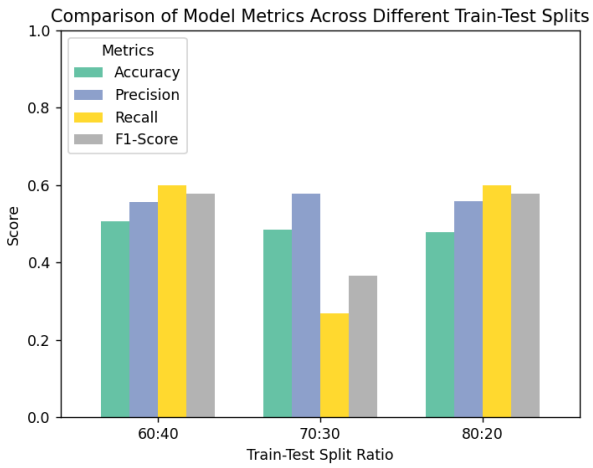

Figure 5. Testing Results for Each Data Split Ratio

Table 1. Accuracy, Precision, Recall, and F1-Score Results

Comparison	Accuracy	Precision	Recall	F1-Score
60:40	50,75%	55,56%	60,00%	57,69%
70:30	48,51%	57,69%	26,79%	36,59%
80:20	47,76%	55,81%	60,00%	57,83%

Source: Steven owen (2025)

Figure 5 and Table 1 present the results of the evaluation matrix used to assess the performance of the developed model. As shown, the 60:40 data split achieved an accuracy of 50.75%, precision of 55.56%, recall of 60.00%, and F1-score of 57.69%. The 80:20 split achieved an accuracy of 47.76%, precision of 55.81%, recall of 60.00%, and F1-score of 57.83%, making it highly suitable for predicting Bitcoin prices compared to the 70:30 split, which resulted in an accuracy of 48.51%, precision of 57.69%, recall of 26.79%, and F1-score of 36.59%.

Table 2. Comparison of Accuracy Results with Other Algorithms

Algorithms	Accuracy
Random Forest (Classification)	50,75%
Long ShortTerm Memory	5,36%
Moving Average	25%
Recurrect Neural Network	0,11%
Random Forest (Regression)	98,50%

Source: Steven owen (2025)

Based on Table 2, it is shown that the Random Forest (Classification) algorithm achieved an accuracy of 50.75%, the Long Short-Term Memory algorithm achieved 5.36%, the Moving Average algorithm achieved 25%, the Recurrent Neural Network algorithm achieved 0.11%, and the Random Forest (Regression) achieved 98.50%. These accuracy results indicate that the Random Forest algorithm provides better accuracy compared to Long Short-Term Memory, Moving Average, and Recurrent Neural Network algorithms. This is because the Random Forest algorithm combines multiple decision trees, which enhances overall accuracy, and it is also capable of handling overfitting issues effectively.

5.0 CONCLUSION

Based on the test results using the historical Bitcoin dataset, the key elements that significantly influence Bitcoin value prediction in this model are the opening and closing prices, followed by the highest and lowest prices, trading volume, while news and market sentiment also play a major role, ranking second in importance. This results in more accurate Bitcoin value predictions. When compared to the Moving Average, Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN) algorithms, the Random Forest algorithm has proven to be more accurate, offering more stable accuracy values. The strength of the Random Forest algorithm in predicting Bitcoin trading values is supported by the use of training/testing data ratios of 60:40 and 80:20, which yield balanced precision, accuracy, recall, and F1-score values.

References

Al Fajri, M. (2023). Penerapan K-Means Clustering Dalam Memprediksi Mata Uang Cryptocurrency Untuk Mengetahui Pergerakan Kenaikan, Penurunan, Dan Sideways Dalam Harga Bitcoin Pada Blockchain Binance. *Education Sains Technology Mathematic (EDUSTEM)*, *1*(1), 146–159.

Aulia, A. S. (2019). Analisis Volatility Spillover Harga Bitcoin Dengan Harga Altcoin Tahun 2013-2018. *JASa (Jurnal Akuntansi, Audit Dan Sistem Informasi Akuntansi)*, *3*(2), 183–194. http://journal.unla.ac.id/index.php/jasa/article/view/942

Bangun Pasaribu, & Wilda Susanti. (2021). Sistem Informasi Pengajuan RancanganUsulan Penelitian Menggunakan PHP Native dan Bot Telegram. *Jurnal Mahasiswa Aplikasi Teknologi Komputer Dan Informasi*, *3*(1), 29–38.

Cecep Abdul Cholik. (2021). PERKEMBANGAN TEKNOLOGI INFORMASI KOMUNIKASI / ICT DALAM BERBAGAI BIDANG. *Jurnal Fakultas Teknik*, *2*(2), 39–46.

Dewi, I. G. A. M. P., Parwita, W. G. S., & Setiawan, I. M. D. (2021). Algoritma Decision Tree untuk Klasifikasi Calon

Debitur LPD Desa Adat Anggungan. *Jurnal Krisnadana*, *1*(1), 23–36. https://doi.org/10.58982/krisnadana.v1i1.79

Gaol Lumban Humphrey, I. (2022). Faktor-Faktor Yang Mempengaruhi Pergerakan Harga Bitcoin Tahun 2018. In *Universitas Brawijaya*. https://jimfeb.ub.ac.id/index.php/jimfeb/article/view/8142

Indriyanti, I., Ichsan, N., Fatah, H., Wahyuni, T., & Ermawati, E. (2022). Implementasi Orange Data Mining Untuk Prediksi Harga Bitcoin. *Jurnal Responsif : Riset Sains Dan Informatika*, *4*(2), 118–125. https://doi.org/10.51977/jti.v4i2.762

Jollyta, D., Hajjah, A., Haerani, E., & Siddik, M. (2023). *Algoritma Klasifikasi untuk Pemula Solusi Python dan RapidMiner*. Deepublish. https://books.google.co.id/books?id=y84TEQAAQBAJ

Nadila, D., Silfia, Hidayaty, D. E., & Mulyadi, D. (2023). Pemahaman investasi, motivasi investasi dan minat investasi di pasar modal. *Jurnal Pijar Studi Manajemen Dan Bisnis*, *1*(2), 104–109. https://e-journal.naureendigition.com/index.php/pmb

Naufal Hasani, M. (2022). ANALISIS CRYPTOCURRENCY SEBAGAI ALAT ALTERNATIF DALAM BERINVESTASI DI INDONESIA PADA MATA UANG DIGITAL BITCOIN. *Jurnal Ilmiah Ekonomi Bisnis*, *8*(2), 329–344. http://ejournal.stiepancasetia.ac.id/index.php/jiebJilid

Partogi, Y., & Pasaribu, A. (2021). Perancangan Metode Decision Tree Terhadap Sistem Perpustakaan STMIK Kuwera. *Jurnal Sistem Informasi Dan Teknologi (SINTEK)*, *1*(2), 20–26. https://doi.org/10.56995/sintek.v1i2.4

Permana, A. P., Ainiyah, K., & Holle, K. F. H. (2021). Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *6*(3), 178–188. https://doi.org/10.14421/jiska.2021.6.3.178-188

Pramita, K. D., & Hendrayana, K. D. (2021). Perlindungan Hukum Terhadap Investor Sebagai Konsumen dalam Investasi Online. *Jurnal Pacta Sunt Servanda*, *2*(1), 1–7.

Purnomo, E., Rahmawati, I. R., & Holis, M. N. (2022). PENGARUH PANDEMI COVID 19 TERHADAP HARGA SAHAM PADA JAKARTA STOCK EXCHANGE COMPOSITE (JKSE) DAN BITCOIN PER RUPIAH INDONESIA The Effect Of The Covid 19 Pandemic On Stock Prices In Jakarta Stock Exchange Composite (Jkse) And Bitcoin Per Indonesian Rupiah. *Nusantara Hasana Journal*, *2*(2), 318–332.

Putra, N. D. N., & Robiyanto, R. (2021). Korelasi Dinamis Pergerakan Cryptocurrency Dan Indeks Harga Saham Sektoral Di Bursa Efek Indonesia. *Management and Accounting Expose*, *4*(1), 35–44. https://doi.org/10.36441/mae.v4i1.275

Rohman, M. N. (2021). Tinjauan Yuridis Normatif Terhadap Regulasi Mata Uang Kripto (Crypto Currency) di Indonesia. *Jurnal Supremasi*, *11*(2), 1–10. https://doi.org/10.35457/supremasi.v11i2.1284

Saadah, S., & Salsabila, H. (2021). Prediksi Harga Bitcoin Menggunakan Metode Random Forest. *Jurnal Komputer Terapan*, *7*(1), 24–32.

Saputri, N. D., Khalid, K., & Rolliawati, D. (2022). Komparasi penerapan metode Bagging dan Adaboost pada Algoritma C4. 5 untuk prediksi Penyakit Stroke. *SISTEMASI: Jurnal Sistem Informasi*, *11*(3), 567–577. http://sistemasi.ftik.unisi.ac.id

Sujjada, A., Sembiring, F., & Febriansyah. (2024). Prediksi Harga Bitcoin Menggunakan Algoritma Long ShortTerm Memory. *Jurnal Inovtek Polbeng*, *9*(1), 450–459. https://doi.org/10.36802/jnanaloka.2022.v3-no2-69-76

Supriyadi, R., Gata, W., Maulidah, N., & Fauzi, A. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. *E-Bisnis : Jurnal Ilmiah Ekonomi Dan Bisnis*, *13*(2), 67–75. https://doi.org/10.51903/e-bisnis.v13i2.247