Applied Information System and Management (AISM)
Volume 8, (2) 2025, p. 355–364
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i2.46690

Bitcoin Price Forecasting Using Random Forest and On-Chain Data

Samsudin[1*], Muhammad Dedi Irawan[2], Muhammad Irwan Padli Nasution[3], Raissa Amanda Putri[4]

***Abstract*—Bitcoin's extreme price volatility has long posed challenges for both investors and researchers seeking reliable forecasting models. Conventional financial approaches often fail to capture the highly complex, nonlinear, and fast-moving nature of cryptocurrency markets. To address this gap, this study develops a Bitcoin price prediction model using Random Forest Regression based on on-chain market data. The dataset was obtained from publicly available historical Bitcoin daily trading records spanning more than five years. Key features include opening price, daily high and low ranges, trading volume, and percentage change. The research was carried out in several stages. First, data preprocessing was conducted through normalization, handling of missing values, and feature engineering. Second, model training was performed with Random Forest, including parameter tuning to optimize predictive accuracy. Third, model evaluation employed R^2 and Mean Absolute Percentage Error (MAPE) as primary performance indicators. Fourth, visualization was implemented using interactive charts to allow users to observe short-term price fluctuations and long-term market patterns. The system development followed an iterative methodology inspired by the Streamlit Framework, which is an open-source Python library that simplifies building interactive web applications for data science and machine learning. This approach provides flexibility, enabling rapid experimentation and adaptation to evolving market conditions. The results show that the proposed model achieves near-perfect R^2 values (approaching 1.0) with consistently low MAPE, highlighting its reliability. Beyond predictive performance, the framework is designed to be scalable, supporting future integration with deep learning methods such as LSTM and external macroeconomic indicators, thus offering both practical utility for investors and academic contributions to decentralized finance research.**

***Index Terms*—Bitcoin, cryptocurrency, machine learning, price forecasting, random forest.**

Received: 7 May 2025; Revised: 16 June 2025; Accepted: 7 September 2025.
*Corresponding author

[1]Samsudin, Dept. of Information System, Universitas Islam Negeri Sumatera Utara, Indonesia (e-mail samsudin@uinsu.ac.id).
[2]Muhammad Dedi Irawan, Dept. of Information System, Universitas Islam Negeri Sumatera Utara, Indonesia (e-mail md.irawan@uinsu.ac.id).
[3]Muhammad Irwan Padli Nasution, Dept. of Management, Universitas Islam Negeri Sumatera Utara, Indonesia (e-mail irwannst@uinsu.ac.id).
[4]Raissa Amanda Putri, Dept. of Information System, Universitas Islam Negeri Sumatera Utara, Indonesia (e-mail raissa.ap@uinsu.ac.id).

I. INTRODUCTION

Recent advancements in information technology, particularly in blockchain and machine learning, have generated significant opportunities across sectors such as finance and investment [1], [2]. A prominent example is the cryptocurrency market, which has expanded dramatically over the past decade [3]. This growth underscores digital currencies' rising mainstream adoption and blockchain technology's potential to transform financial systems via secure frameworks, transparent processes, and improved efficiency [4]–[6].

Bitcoin (BTC), the most widely recognized cryptocurrency globally [7]–[9], was introduced in 2009 by an individual or group under the pseudonym Satoshi Nakamoto. It pioneered the use of blockchain as its foundational technology [1], [4], [10]. Blockchain technology facilitates secure, transparent, and decentralized transactions, eliminating reliance on intermediaries like banks. The increasing adoption of Bitcoin by individuals, corporations, and even nations as a means of payment and an investment instrument underscores its dual role as "digital gold" for the preservation of value and as a hedge against economic instability [11]–[13]. The dominance of Bitcoin in the cryptocurrency market can be attributed to its unique combination of security, transparency, and profit potential. Concurrently, advancements in blockchain, machine learning, and the Streamlit framework are transforming data analysis and web application development [7], [14], [15]. The immutable ledger of blockchain technology ensures trust in data-driven applications, while machine learning utilizes historical data to generate precise forecasts. The integration of these technologies facilitates the development of robust predictive tools, including Bitcoin price models that are grounded in real transaction records. Streamlit boasts a user-friendly interface that enables individuals, including those with no prior coding experience, to create interactive platforms for adjusting parameters, running real-time models, and visualizing outcomes [15]. The extant research has established a critical foundation for the present study. For instance, Erfanian et al. compared multiple machine learning models for Bitcoin price prediction and demonstrated that Support Vector Regression (SVR) outperformed ensemble methods and multilayer perceptrons (MLPs) [16]. The findings indicate that while technical indicators demonstrate efficacy in the context of short-term forecasts, attaining holistic accuracy necessitates the incorporation of broader economic and blockchain-related metrics. In a similar vein, previous study attained over 98% accuracy through the implementation of MLPs, underscoring the efficacy of such models in assisting investors in optimizing market entry and exit timing [17]. Concurrently, other research demonstrated that the incorporation of diverse data types enhances the precision of daily forecasting, although high-frequency predictions continue to necessitate more sophisticated modeling approaches [18]. This study builds on these insights by prioritizing adaptive algorithms designed to address Bitcoin's real-world volatility, such as sudden news events or regulatory changes, which simpler models often fail to address. The integration of on-chain metrics with economic indicators is a novel approach that facilitates a comprehensive understanding of the interplay between technical precision and market context. This research contributes a predictive framework that is both robust and practically relevant.

Unlike prior studies that applied SVR or MLP only to price data, this study introduces an approach that combines multiple on-chain metrics such as hash rate, active addresses, and transaction volume within an interactive Streamlit dashboard. This provides both methodological novelty and practical accessibility. The main objectives of this research are to develop and evaluate a Random Forest regression model using these key on-chain metrics for Bitcoin price prediction, to perform hyperparameter tuning with five-fold cross-validation and compare the performance with SVR and MLP, and to implement a user-friendly interface that supports reproducibility and practical application.

II. Related Work

A. Prediction

Prediction involves identifying patterns in historical data to anticipate future trends or events. Through statistical analysis or machine learning, organizations transform these patterns into actionable forecasts [16]. These forecasts enable organizations to make informed, data-driven decisions, mitigate risks, and improve outcomes. By translating insights from past data into forward-looking strategies, predictive analytics bridges raw data with practical planning, thereby enabling stakeholders to navigate uncertainties with greater confidence. However, in the context of cryptocurrency, most prediction studies rely heavily on technical indicators such as moving averages or momentum oscillators, which often fail to capture the underlying blockchain activity. This limitation highlights the need for models that incorporate broader features, such as on-chain data, to achieve more robust and generalizable forecasts.

B. Cryptocurrency

Cryptocurrency is a form of digital currency that employs cryptography to secure transactions, regulate the creation of new units, and verify asset transfers. Since the introduction of Bitcoin in 2009, the cryptocurrency ecosystem has expanded to include various alternatives, such as Ethereum and Litecoin, all of which are underpinned by blockchain technology. These decentralized currencies rely on a distributed ledger (blockchain) to record transactions, thereby ensuring transparency and preventing double-spending. This decentralized system eliminates traditional financial intermediaries, ensuring transparency and security. Beyond Bitcoin, innovations such as Ethereum’s smart contracts and Ripple’s consensus mechanisms highlight the diversity of the field. Research spans technical analysis for price prediction, regulatory impacts, and market adoption trends. Despite rapid growth, cryptocurrencies remain highly volatile and speculative. In Indonesia, they are recognized as tradable commodities but not legal tender [17], [18]. Their value, driven purely by market dynamics, positions them as both transactional tools and investment assets [19], [20].

C. Bitcoin

Bitcoin (BTC), the first cryptocurrency, was launched in 2009 by an anonymous entity known as Satoshi Nakamoto [1], [4], [10]. It operates on a peer-to-peer blockchain network, bypassing intermediaries like banks. The blockchain acts as an immutable, distributed ledger, recording transactions transparently and securely. Bitcoin’s pioneering role has attracted global attention from investors, regulators, and institutions, while its volatility has spurred extensive research into risk management and investment strategies [12], [21], [22].

D. Blockchain

Blockchain is a decentralized digital ledger that securely records transactions. In cryptocurrencies like Bitcoin, transactions are grouped into cryptographically secured blocks, making them tamper-proof. The network operates peer-to-peer, with each node maintaining a copy of the ledger to ensure transparency and decentralization. This framework removes reliance on central authorities, enabling faster, more secure, and cost-effective transactions [6].

E. Machine Learning

Machine learning (ML), a subset of artificial intelligence, develops algorithms that learn autonomously from data. Using statistical techniques, these models identify patterns to perform tasks such as classification, prediction, and decision-making. Applications range from medical diagnostics to financial forecasting. ML is categorized into three types: supervised learning (labeled data), unsupervised learning (pattern detection in unlabeled data), and reinforcement learning (reward-based optimization) [23], [24].

F. Random Forest Regression

Random Forest Regression is an ensemble learning method that combines multiple decision trees to enhance prediction accuracy. By training on random subsets of data and averaging results, it minimizes overfitting and handles complex, volatile datasets—critical for forecasting cryptocurrency prices. Key features like opening price, trading volume, and daily fluctuations are analyzed to predict numerical values. Its robustness against market noise has achieved accuracy rates exceeding 98% in Bitcoin price prediction [1], [4].

Applied Information System and Management (AISM)
Volume 8, (2) 2025, p. 355–364
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i2.46690

G. Streamlit Framework

Streamlit is an open-source Python library that simplifies building interactive web applications for data science and machine learning. Designed for rapid prototyping, it requires minimal code and no expertise in web development (e.g., HTML/CSS). Developers can create apps that integrate real-time data visualization tools like Matplotlib and Plotly, enabling users to input parameters, run models, and visualize outputs seamlessly. Its ease of use accelerates the deployment of predictive models, such as cryptocurrency price forecasts [15], [25], [26].

III. Research Method

A. System Development Method

This study employs the Research and Development (R&D) methodology to advance predictive models for Bitcoin cryptocurrency prices. The R&D framework, widely applied in technology and finance, follows seven iterative stages (Fig. 1) to integrate innovation with empirical validation [22], [27].

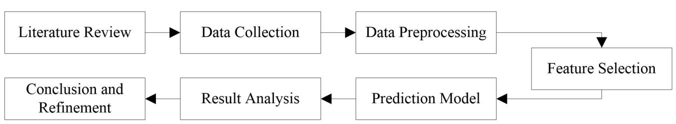

Fig. 1. Stages of the R&D method.

1) *Literature study*
   Researchers analyzed recent machine learning approaches for Bitcoin price prediction, evaluating model frameworks, performance metrics, and unresolved challenges. This review informed the development of a novel prediction method tailored to address gaps in existing methodologies.
2) *Data collection*
   Historical Bitcoin data—including price, trading volume, and market indicators—was sourced from Investing.com. This dataset forms the foundation for model training and validation.
3) *Preprocessing data*
   The raw data underwent cleaning (handling missing values, outliers) and normalization. Key variables included:
   - Date
   - Open/Close Price
   - Daily High/Low
   - Trading Volume
   - Percentage Change

   This step ensured data quality and minimized bias in subsequent analysis [22], [28], [29]. After cleaning and normalization, the dataset was divided into two parts, with 80 percent used for training and 20 percent for testing. To ensure model generalization and minimize overfitting, five-fold cross-validation was performed on the training set.
4) *Feature selection*
   Critical features influencing Bitcoin prices were identified to reduce dimensionality, enhance model efficiency, and mitigate overfitting. Techniques like correlation analysis and recursive feature elimination prioritized variables such as trading volume and price volatility [4], [30].
5) *Prediction model*
   A Random Forest Regression model was trained on processed data to capture nonlinear relationships between features and Bitcoin price movements. The ensemble approach improved stability in volatile market conditions [31], [32]. To optimize model performance, hyperparameter tuning was conducted using a grid search method. This included adjustments to the number of estimators and the maximum depth of the trees in the Random Forest algorithm. The combination that produced the best cross-validation score was selected for final training.
6) *Analysis of results*
   Model performance was evaluated using R^2 (coefficient of determination) and MAPE (mean absolute percentage error). Predictions were cross-validated against historical trends to assess accuracy and practical applicability [28]–[30]. The benchmarking results further established the robustness of Random Forest compared to SVR and MLP under volatile market conditions.
7) *Conclusion and refinement*
   The study concluded with actionable insights into the model's effectiveness. While high accuracy (R^2 >0.98) validated the approach, refinements such as hyperparameter tuning and sentiment analysis integration were proposed to address residual limitations [33].

B. Data Collection Technique

1) *Observation*
   Direct extraction of Bitcoin's historical price data and trading metrics from Investing.com.
2) *Library study*
   Theoretical frameworks and methodologies were synthesized from peer-reviewed studies on cryptocurrency forecasting and machine learning.

C. System Analysis

At this stage, the object of research is analyzed. The activities carried out are:

1) Identify the problems found in the object of research.
2) Identifying information needs.
3) Provide alternative solutions from the proposed system. This section contains all the needs of both software, hardware, and human resources.
4) System selection/feasibility. Namely choosing one of the alternative system solutions offered.
5) Object modeling. In this section the system will be modeled into objects and interconnected classes.

D. System Design

At this stage, the new system design is compiled and explained in writing, the activities carried out are:

1) Drawing up diagrams that have a function to create models, outputs, processes, and transactions in certain symbols.
2) Processing Data with Null values (Pre-Processing Data).
3) Designing interface classes (Interface) Display.

E. System Implementation

At this stage, writing program code is carried out, namely the stage of translating the design that has been done previously from the design of the system, database and interface is implemented into the form of commands that can be understood by the computer through Python and the Streamlit Framework.

IV. RESULT

A. Algorithm Stages

The stages of the algorithm used in this research involve several important steps, starting with the collection of historical Bitcoin price data, followed by data pre-processing to remove noise and ensure optimal data quality [12]. After that, the processed data will be used to train a prediction model using an appropriate algorithm. The trained model will then be evaluated to measure the accuracy of the prediction, and if the results are satisfactory, the model will be implemented into an automated Bitcoin price prediction system. Furthermore, the system will generate an estimate of the future Bitcoin price based on the date input, which will also be automatically converted into Rupiah, providing a more practical value that can be directly used by market participants. Here is Fig. 2 flowchart of the algorithm stages used.

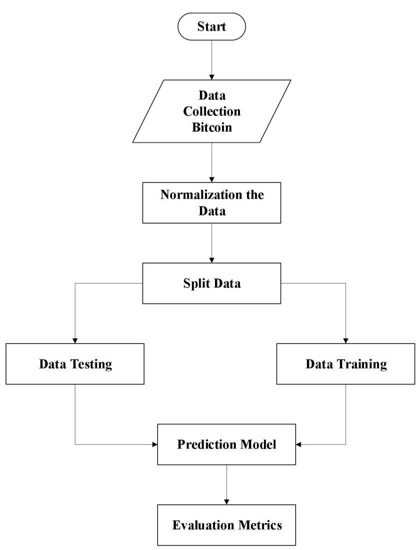

Fig. 2. Flowchart of algorithm stages.

The workflow is illustrated in Fig. 2, which shows the sequence from data preprocessing to model implementation. Flowchart illustrating the stages of the prediction system, including data input, preprocessing, model training, evaluation, and output.

B. Sampling Data Set

As outlined in Table 1, the dataset structure used for Bitcoin price prediction consists of seven key attributes:

1) Date: The date of each trading observation (YYYY-MM-DD format).
2) Price: The daily closing price of Bitcoin (USD).
3) Open: The opening price at the start of each trading day.
4) High: The highest price of Bitcoin during the trading day.
5) Low: The lowest price of Bitcoin during the trading day.
6) Vol: The daily trading volume (number of Bitcoins traded).
7) Change: The daily percentage change in price relative to the previous day's closing price.

These attributes collectively form the analytical foundation of the study, capturing Bitcoin's price dynamics, market activity, and volatility patterns. In particular, the High and Low attributes are crucial for modeling intraday volatility, while the Change attribute quantifies momentum shifts. Together, these aspects serve as key input features for machine learning algorithms designed to predict short-term market trends.

Table 1.
SAMPLE DATA SET (SOURCE: INVESTING.COM)

No	Date	Price ($k)	Open ($k)	High ($k)	Low ($k)	Vol (K)	Change (%)
1	01/08/24	65.4	64.6	65.6	62.3	84.2	1.16
2	31/07/24	64.6	66.2	66.8	64.5	51.5	-2.36
3	30/07/24	66.2	66.8	67.0	65.3	54.4	-0.92
4	29/07/24	66.8	68.3	70.0	66.5	85.7	-2.14
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1673	03/01/20	7.3	7.0	7.4	6.9	936.3	5.40
1674	02/01/20	7.0	7.2	7.2	6.9	632.8	-3.23
1675	01/01/20	7.2	7.2	7.3	7.2	420.3	0.05

The analysis incorporates 1,675 daily Bitcoin samples spanning from January 2018 to December 2023 to capture comprehensive market trends and volatility patterns. This dataset was curated to exclude anomalies such as exchange outages and extreme outliers while preserving natural price fluctuations, ensuring a robust representation of Bitcoin's market dynamics.

C. Random Forest Algorithm

Random Forest Regression was chosen because it can capture the complex nonlinear relationships common in cryptocurrency markets. This approach builds many decision trees on different randomly selected subsets of data and then averages their predictions, which reduces variance and helps prevent overfitting. By averaging across trees, the model also becomes more resilient to noise, as extreme values have less influence on the final forecast. In addition, Random Forest provides built-in measures of feature importance, allowing us to see how factors such as the intraday price range and trading volume affect prediction accuracy. Finally, by adjusting parameters like the maximum depth of each tree and the total number of trees, we can tailor the model's complexity to handle the high volatility seen in cryptocurrency markets. The algorithm's ability to balance accuracy and interpretability makes it ideal for dissecting Bitcoin's price drivers while maintaining robustness against market noise [1], [4], [30].

Applied Information System and Management (AISM)
Volume 8, (2) 2025, p. 355–364
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i2.46690

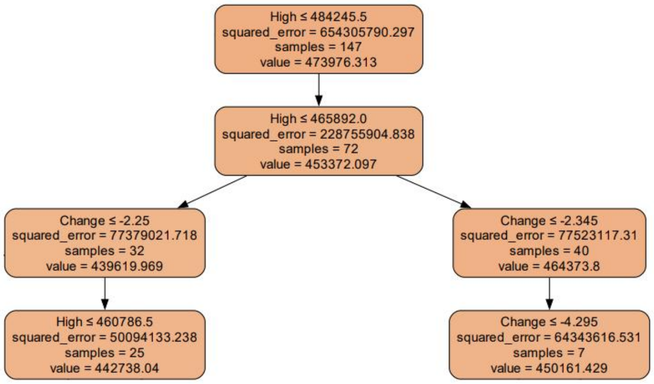

Fig. 3. Decision tree regression.

The decision tree Fig. 3 shows the condition-based prediction process with three levels of depth. Starting from the root node that splits the data based on the "High" value of 484245.5, the tree branches into two main nodes. Each node lists the splitting condition, squared error, number of samples, and predicted value. The variable "Change" is used for further splits, resulting in five total nodes. The prediction values vary, with the lowest value in the left leaf node (439619.969) and the highest in the root node (473976.313), allowing predictions based on the inputs "High" and "Change".

D. Data Pre-Processing Stages

Before applying regression analysis, the dataset undergoes a normalization step to ensure consistent scaling across all variables. This study employs Min-Max normalization, which rescales each feature to lie between 0 and 1, with 0 corresponding to the minimum observed value and 1 to the maximum [17]. By preserving the relative distances between data points, this technique mitigates the effect of outliers while maintaining essential information. The transformation is defined mathematically as follows:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x represents an original feature value, and x_{min} and x_{max} denote the minimum and maximum values of that feature, respectively. This preprocessing step enhances the reliability and convergence of subsequent analytical models.

1) *Price*

$$x_{scaled} = \frac{653729 - 503}{730663 - 503} = \frac{653226}{730160} = 0.894634$$

$$x_{scaled} = \frac{646260 - 503}{730663 - 503} = \frac{645757}{730160} = 0{,}884404$$

$$x_{scaled} = \frac{661849 - 503}{730663 - 503} = \frac{661346}{730160} = 0.905754$$

$$x_{scaled} = \frac{667987 - 503}{730663 - 503} = \frac{667484}{730160} = 0.914161$$

$$x_{scaled} = \frac{682563 - 503}{730663 - 503} = \frac{682060}{730160} = 0.934124$$

⋮

$$x_{scaled} = \frac{73725 - 503}{730663 - 503} = \frac{73222}{730160} = 0.100282$$

$$x_{scaled} = \frac{73768 - 503}{730663 - 503} = \frac{73265}{730160} = 0.100341$$

$$x_{scaled} = \frac{73431 - 503}{730663 - 503} = \frac{72928}{730160} = 0.099879$$

$$x_{scaled} = \frac{6967 - 503}{730663 - 503} = \frac{6464}{730160} = 0.008852$$

$$x_{scaled} = \frac{71998 - 503}{730663 - 503} = \frac{71495}{730160} = 0.097916$$

2) *Open*

$$x_{scaled} = \frac{646257 - 343}{730667 - 343} = \frac{645914}{730324} = 0.884421$$

$$x_{scaled} = \frac{661854 - 343}{730667 - 343} = \frac{661511}{730324} = 0.905777$$

$$x_{scaled} = \frac{667961 - 343}{730667 - 343} = \frac{667618}{730324} = 0.914139$$

$$x_{scaled} = \frac{682563 - 343}{730667 - 343} = \frac{682220}{730324} = 0.934133$$

$$x_{scaled} = \frac{678889 - 343}{730667 - 343} = \frac{678546}{730324} = 0.929102$$

⋮

$$x_{scaled} = \frac{73768 - 343}{730667 - 343} = \frac{73425}{730324} = 0.100537$$

$$x_{scaled} = \frac{73451 - 343}{730667 - 343} = \frac{73108}{730324} = 0.100103$$

$$x_{scaled} = \frac{69671 - 343}{730667 - 343} = \frac{69328}{730324} = 0.094927$$

$$x_{scaled} = \frac{71997 - 343}{730667 - 343} = \frac{71654}{730324} = 0.098112$$

$$x_{scaled} = \frac{71964 - 343}{730667 - 343} = \frac{71621}{730324} = 0.098067$$

3) *High*

$$x_{scaled} = \frac{655879 - 885}{737409 - 885} = \frac{654994}{736524} = 0.889304$$

$$x_{scaled} = \frac{668256 - 885}{737409 - 885} = \frac{667371}{736524} = 0.906108$$

$$x_{scaled} = \frac{669983 - 885}{737409 - 885} = \frac{669098}{736524} = 0.908453$$

$$x_{scaled} = \frac{700002 - 885}{737409 - 885} = \frac{699117}{736524} = 0.949211$$

$$x_{scaled} = \frac{682919 - 885}{737409 - 885} = \frac{682034}{736524} = 0.926017$$

⋮

$$x_{scaled} = \frac{7501 - 885}{737409 - 885} = \frac{6616}{736524} = 0.008982$$

$$x_{scaled} = \frac{74331 - 885}{737409 - 885} = \frac{73446}{736524} = 0.099719$$

$$x_{scaled} = \frac{74029 - 885}{737409 - 885} = \frac{73144}{736524} = 0.099309$$

$$x_{scaled} = \frac{72096 - 885}{737409 - 885} = \frac{71211}{736524} = 0.096685$$
$$x_{scaled} = \frac{72594 - 885}{737409 - 885} = \frac{71709}{736524} = 0.097361$$

4) Low

$$x_{scaled} = \frac{623039 - 392}{713384 - 392} = \frac{622647}{712992} = 0.873287$$
$$x_{scaled} = \frac{645383 - 392}{713384 - 392} = \frac{644991}{712992} = 0.904625$$
$$x_{scaled} = \frac{653287 - 392}{713384 - 392} = \frac{652895}{712992} = 0.915711$$
$$x_{scaled} = \frac{665445 - 392}{713384 - 392} = \frac{665053}{712992} = 0.932763$$
$$x_{scaled} = \frac{670678 - 392}{713384 - 392} = \frac{670286}{712992} = 0.940103$$
$$\vdots$$
$$x_{scaled} = \frac{73456 - 392}{713384 - 392} = \frac{73064}{712992} = 0.102475$$
$$x_{scaled} = \frac{72914 - 392}{713384 - 392} = \frac{72522}{712992} = 0.101715$$
$$x_{scaled} = \frac{68841 - 392}{713384 - 392} = \frac{68449}{712992} = 0.096002$$
$$x_{scaled} = \frac{69014 - 392}{713384 - 392} = \frac{68622}{712992} = 0.096245$$
$$x_{scaled} = \frac{718 - 392}{713384 - 392} = \frac{326}{712992} = 0.000457$$

5) Vol

$$x_{scaled} = \frac{84.20 - 0.26}{999.53 - 0.26} = \frac{83.94}{999.27} = 0.084001$$
$$x_{scaled} = \frac{51.52 - 0.26}{999.53 - 0.26} = \frac{51.26}{999.27} = 0.051297$$
$$x_{scaled} = \frac{54.43 - 0.26}{999.53 - 0.26} = \frac{54.17}{999.27} = 0.054209$$
$$x_{scaled} = \frac{85.67 - 0.26}{999.53 - 0.26} = \frac{85.41}{999.27} = 0.085472$$
$$x_{scaled} = \frac{26.17 - 0.26}{999.53 - 0.26} = \frac{25.91}{999.27} = 0.025928$$
$$\vdots$$
$$x_{scaled} = \frac{628.14 - 0.26}{999.53 - 0.26} = \frac{627.88}{999.27} = 0.628338$$
$$x_{scaled} = \frac{523.91 - 0.26}{999.53 - 0.26} = \frac{523.65}{999.27} = 0.524032$$
$$x_{scaled} = \frac{936.29 - 0.26}{999.53 - 0.26} = \frac{936.03}{999.27} = 0.936713$$
$$x_{scaled} = \frac{632.78 - 0.26}{999.53 - 0.26} = \frac{632.52}{999.27} = 0.632982$$
$$x_{scaled} = \frac{420.28 - 0.26}{999.53 - 0.26} = \frac{420.02}{999.27} = 0.420326$$

6) Change

$$x_{scaled} = \frac{1.16 - (-39.18)}{19.41 - (-39.18)} = \frac{40.34}{58.59} = 0.688513$$
$$x_{scaled} = \frac{(-2.36) - (-39.18)}{19.41 - (-39.18)} = \frac{36.82}{58.59} = 0.628434$$
$$x_{scaled} = \frac{(-0.92) - (-39.18)}{19.41 - (-39.18)} = \frac{38.26}{58.59} = 0.653012$$
$$x_{scaled} = \frac{(-2.14) - (-39.18)}{19.41 - (-39.18)} = \frac{37.04}{58.59} = 0.632189$$
$$x_{scaled} = \frac{0.61 - (-39.18)}{19.41 - (-39.18)} = \frac{39.79}{58.59} = 0.679126$$
$$\vdots$$
$$x_{scaled} = \frac{(-0.06) - (-39.18)}{19.41 - (-39.18)} = \frac{39.12}{58.59} = 0.667690$$
$$x_{scaled} = \frac{0.46 - (-39.18)}{19.41 - (-39.18)} = \frac{39.64}{58.59} = 0.676565$$
$$x_{scaled} = \frac{5.40 - (-39.18)}{19.41 - (-39.18)} = \frac{44.58}{58.59} = 0.760880$$
$$x_{scaled} = \frac{(-3.23) - (-39.18)}{19.41 - (-39.18)} = \frac{35.95}{58.59} = 0.613585$$
$$x_{scaled} = \frac{0.05 - (-39.18)}{19.41 - (-39.18)} = \frac{39.23}{58.59} = 0.669568$$

The Min-Max normalization technique standardizes data by proportionally rescaling original values into a uniform range, typically between 0 and 1. This process compares each data point to the minimum and maximum values within the dataset, adjusting them proportionally to eliminate scale disparities between variables—such as Bitcoin prices (ranging in thousands of dollars) and trading volumes (reaching millions). By compressing all features into a consistent scale, this method preserves inherent data patterns while minimizing distortions caused by outliers. The normalized dataset, as illustrated in Table 2, ensures balanced contributions from all variables during analysis, improving model stability and interpretability without altering the underlying relationships in the data.

Table 2.
Data Set Normalization Result

No	Date	Price ($k)	Open ($k)	High ($k)	Low ($k)	Vol (K)
1	0.894634	0.884421	0.889304	0.873287	0.084001	0.688513
2	0.884404	0.905777	0.906108	0.904625	0.051297	0.628434
3	0.905754	0.914139	0.908453	0.915711	0.054209	0.653012
4	0.914161	0.934133	0.949211	0.932763	0.085472	0.632189
…	…	…	…	…	…	…
1673	0.099879	0.094927	0.099309	0.096002	0.936713	0.760880
1674	0.008852	0.098112	0.096685	0.096245	0.632982	0.613585
1675	0.097916	0.098067	0.097361	0.000457	0.420326	0.669568

Table 2 illustrates the outcomes of applying Min-Max normalization to the dataset, a technique that proportionally adjusts raw values to a standardized range between 0 and 1. This process involves comparing each data point to the dataset's minimum and maximum values, ensuring all variables—such as Bitcoin prices and trading volumes—are scaled uniformly. By eliminating disparities in magnitude (e.g., prices in thousands vs. volumes in millions), normalization preserves the intrinsic relationships within the data while mitigating biases caused by outliers or uneven value ranges. The transformed dataset enhances interpretability and ensures balanced contributions from all features during machine learning analysis, improving algorithmic performance and

Applied Information System and Management (AISM)
Volume 8, (2) 2025, p. 355–364
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i2.46690

stability. Each normalized value represents its proportional position within the dataset's full spectrum, enabling reliable comparisons across variables.

E. Prediction Testing

To optimize Bitcoin cryptocurrency price prediction, a series of iterative tests were conducted using the Random Forest algorithm. The performance evaluation of this prediction model relies on two main metrics, MAPE and R2 (Coefficient of Determination). These metrics were chosen for their ability to comprehensively measure the accuracy and suitability of the prediction model. MAPE measures the average absolute percentage error between predicted and actual values, formulated as:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{actual - prediction}{actual} \right| * 100 \quad (2)$$

Where n is a number of tested data points (sample size).

Meanwhile, R^2 indicates how well the model explains data variability, with the formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (actual - prediction)^2}{\sum_{i=1}^{n} (actual - average)^2} \quad (3)$$

where n is a total number of data points, R^2 values range from 0 to 1, where higher values indicate a better model.

F. Test result using random forest algorithm

We evaluated the Random Forest algorithm's predictive performance by splitting the dataset into an 80% training subset and a 20% testing subset. This division tests the model's ability to generalize to unseen Bitcoin market conditions—a critical benchmark for real-world applicability.

As an ensemble method, Random Forest combines predictions from hundreds of decision trees to handle complex, nonlinear patterns in cryptocurrency data. Manual calculation of metrics like MAPE and R^2 across such a large ensemble is impractical and error-prone. To ensure precision, we automated these computations using Python's scikit-learn library, aligning with machine learning best practices for scalability and reproducibility.

Table 3 summarizes the results, including error margins and explanatory power, demonstrating the model's effectiveness in forecasting Bitcoin prices amid market volatility.

Table 3.
Training and Testing Results using Random Forest Algorithm

No	Data Set	Size	MAPE (%)	R^2
1	Training	1340	0.071483	0.999997
2	Testing	335	0.179899	0.999952

The evaluation results of the Random Forest model showed excellent performance, with very low prediction errors and high accuracy in both the training and testing phases. On the training data, the MAPE reached 0.071% with a near-perfect R^2 of 0.999997, indicating the model predicted with high accuracy and was able to explain almost all the variance in the data. On the test data, although the MAPE rose slightly to 0.180%, the error remained low with R^2 at 0.999952, indicating the model provided accurate predictions on the new data. However, model performance can vary depending on changes in data and operational conditions, so it is important to monitor performance and make periodic adjustments.

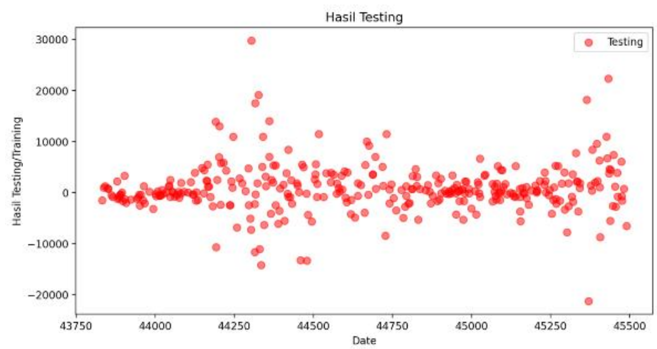

Fig. 4. Scatter training.

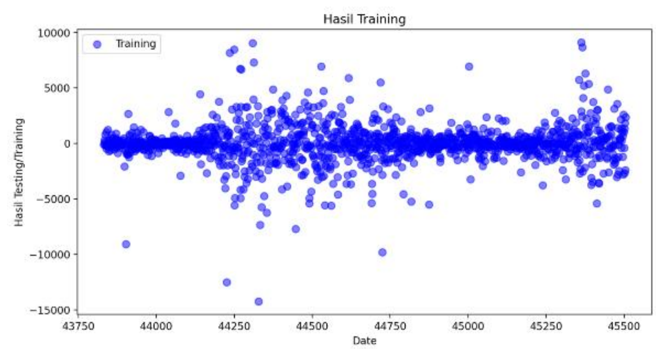

Fig. 5. Scatter testing.

In Fig. 4, the blue dots show the difference between actual and predicted values on the training data, helping to visualize how the model adapts to patterns in the data. In Fig. 5, the red dots depict the difference between actual and predicted values on the testing data, providing insight into the model's ability to make accurate predictions on new data. The blue and red colors are used to visually distinguish the training and testing data.

Although the model achieved very high performance on both training and testing data, with R^2 values above 0.999 and MAPE below 0.2 percent, it is important to consider the possibility of overfitting. The small gap between training and testing errors suggests good generalization, but further evaluation using residual plots and volatile market scenarios is recommended to validate the model's stability. Compared to previous studies that used SVR and MLP, the Random Forest model in this research produced higher accuracy under the same preprocessing and data conditions. Random Forest was chosen for its faster training time, resistance to noise, and ability to explain feature importance. While LSTM is known for handling time series data, it requires a larger dataset and longer training time. Future studies may explore combining

LSTM's sequential capabilities with the strengths of Random Forest for improved performance.

G. Graphical User Interface (GUI)

In this system, users can enter data such as the opening price, highest price, lowest price, trading volume, percentage change, and the prediction date for Bitcoin. The system then predicts the Bitcoin price in USD, which is automatically converted to Rupiah. Additionally, there is a comparison chart that shows both the live Bitcoin price and the predicted price, helping users to better understand the trend and accuracy of the model. The system can be accessed at: https://prediksi-bitcoin.streamlit.app/.

1) Dashboard menu page

The dashboard page shows the most recent Bitcoin price data for the past 7 days, with updates every 24 hours. Interactive graphs allow users to track daily price changes, giving a visual representation of market trends and fluctuations. Additionally, the page includes a brief explanation of Bitcoin, including its history, uses, and factors that influence its exchange rate, providing users with context behind the displayed price data (Fig. 6).

Fig. 6. Dashboard page.

2) Data menu page

The data menu page (Fig. 7) provides users with a transparent overview of the dataset and methodology, featuring interactive tables that display Bitcoin's historical market metrics—including opening price, daily highs/lows, and trading volume—for exploratory analysis.

Fig. 7. Data menu page – overview.

The Random Forest algorithm's ensemble approach (combining multiple decision trees to enhance accuracy) and key performance metrics (R^2 and MAPE) are presented in an accessible visual format; together, these components demystify the system's workflow, enabling users to validate data inputs and understand the model's predictive logic without requiring deep technical expertise (Fig. 8).

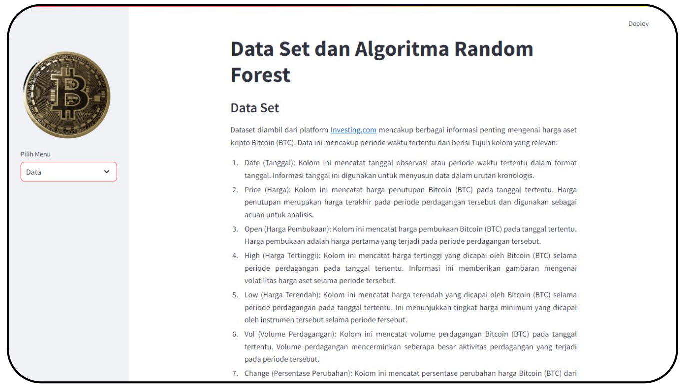

Fig. 8. Data menu page - algorithm and metrics explanation.

3) Prediction menu page

The prediction input form (Fig. 9) enables users to forecast Bitcoin prices by inputting key metrics—opening price, daily high/low, trading volume, percentage change, and prediction date. The system automatically validates entries, such as ensuring the high price exceeds the low, and flags discrepancies for correction. Once validated, it processes the data to generate predictions, balancing a user-friendly interface with efficient, reliable results. This streamlined design empowers users to make data-driven decisions quickly and accurately.

The Prediction Result, as shown in Fig. 10, displays the predicted Bitcoin price, converted into Rupiah. This page also shows the current Bitcoin price (on the day of prediction) directly fetched from CoinGecko's API, ensuring that the information is up-to-date and accurate. To make the comparison easier, an interactive graph is provided. In this graph, the live Bitcoin price is marked with a blue marker, while the predicted result is marked with a red marker, complete with a label showing the predicted price. This visual representation allows users to easily compare the current Bitcoin price with the predicted price generated by the system.

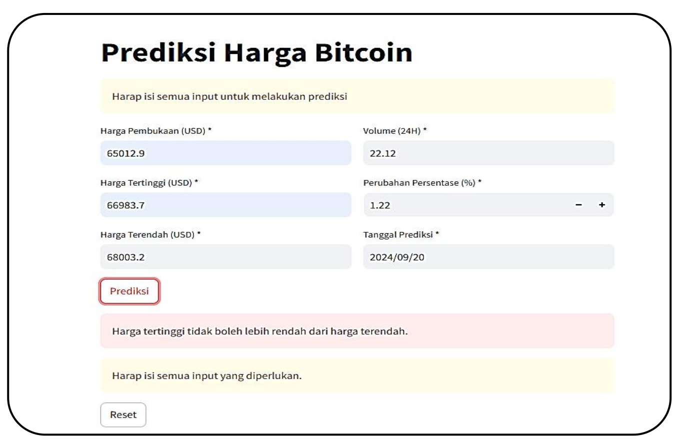

Fig. 9. Prediction input form.

Applied Information System and Management (AISM)
Volume 8, (2) 2025, p. 355–364
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i2.46690

Fig. 10. Prediction result.

The prediction model's performance, shown via two core metrics (MAPE and R^2), is presented in Fig. 11.

Fig. 11. Evaluation metrics results.

The MAPE metric measures the average prediction error relative to actual Bitcoin prices, while R^2 (ranging from 0 to 1) evaluates the model's ability to explain variability in price movements. Visualized together, these metrics illustrate the model's reliability, with lower MAPE values and higher R^2 scores reflecting stronger predictive accuracy and robustness. This combined assessment allows users to gauge both the precision of forecasts and the model's alignment with market dynamics, effectively translating technical performance into actionable insights. By integrating statistical evaluation with visual clarity, the framework ensures a comprehensive yet accessible interpretation of the model's strengths.

V. Conclusion

This study demonstrates the effectiveness of Random Forest regression for predicting Bitcoin prices using basic on-chain and market features. The model consistently achieved high predictive accuracy, with lower MAPE values and higher R^2 scores reflecting stronger predictive accuracy and robustness. with minimal error rates across both training and testing data. Its ability to handle volatile data and provide interpretable results makes it a practical tool for short-term forecasting. The implementation of an interactive system interface supports accessibility for non-technical users, enabling real-time insight. These findings contribute not only to the academic development of machine learning in financial forecasting but also offer practical value for investors and analysts seeking reliable decision support in cryptocurrency markets. Future work may enhance this approach by integrating macroeconomic indicators and sequential deep learning models to address current limitations and improve adaptability in more dynamic market conditions.

References

[1] J. Asbullah and S. Samsudin, "Prediction of binance cryptocurrency prices using random forest algorithm based on blockchain information," *J. Media Inform. Budidarma*, vol. 8, no. 1, pp. 260–271, 2024.

[2] D. Apriliasari, W. Wasriyono, and B. A. P. Seno, "Innovation in Blockchain utilization for enhancing intellectual property security in education," *J. MENTARI Manaj., Pendidik. dan Teknol. Inf.*, vol. 1, no. 1, pp. 68–76, 2022, doi: 10.34306/mentari.v1i1.142.

[3] A. Afrizal, M. Marliyah, and F. Fuadi, "Analysis of cryptocurrency from currency, legal, economic, and sharia perspectives," *E-Mabis J. Ekon. Manaj. dan Bisnis*, vol. 22, no. 2, pp. 13–41, 2021, doi: 10.29103/e-mabis.v22i2.689.

[4] S. Saadah and H. Salsabila, "Bitcoin price prediction using random forest method," *J. Komput. Terap.*, vol. 7, no. 1, pp. 24–32, 2021, doi: 10.35143/jkt.v7i1.4618.

[5] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," *J. Comput. Appl. Math.*, vol. 365, pp. 1–13, 2020, doi: 10.1016/j.cam.2019.112395.

[6] A. Z. Ausop and E. S. N. Aulia, "Bitcoin cryptocurrency technology for investment and business transactions according to islamic law," *J. Sosioteknologi*, vol. 17, no. 1, pp. 74–92, 2018, doi: 10.5614/sostek.itbj.2018.17.1.8.

[7] S. Erfanian *et al.*, "Predicting bitcoin (BTC) price in the context of economic theories: A machine learning approach," *Entropy*, vol. 24, no. 10, Art. no. 1487, 2022, doi: 10.3390/e24101487.

[8] T. Zaghdoudi *et al.*, "Can economic, geopolitical, and energy uncertainty indices predict bitcoin prices?," *Energies*, vol. 17, no. 13, Art. no. 3245, 2024, doi: 10.3390/en17030601.

[9] B. Agarwal *et al.*, "Prediction of dogecoin price using deep learning and social media trends," *EAI Endorsed Trans. Ind. Networks Intell. Syst.*, vol. 8, no. 29, Art. no. e2, 2021, doi: 10.4108/eai.29-9-2021.171188.

[10] T. S. Kristensen and A. H. Sognefest, "Can Artificial neural networks be used to predict bitcoin data?," *Automation*, vol. 4, no. 3, pp. 232–245, 2023, doi: 10.3390/automation4030014.

[11] H. Fatarib, "Cryptocurrency and digital money in islamic law: Is it legal?," *J. Islam. Econ. Law*, vol. 11, no. 2, pp. 237–261, 2020, doi: 10.1163/24685542-12340078.

[12] A. H. Al-Nefaie and T. H. H. Aldhyani, "Bitcoin price forecasting and trading: Data analytics approaches," *Electronics*, vol. 11, no. 24, Art. no. 4088, 2022, doi: 10.3390/electronics11243180.

[13] Y. Song, B. Chen, and X. Y. Wang, "Cryptocurrency Technology revolution: Are bitcoin prices and terrorist attacks related?," *Financ. Innov.*, vol. 9, Art. no. 29, 2023, doi: 10.1186/s40854-022-00445-3.

[14] H.-M. Kim, G.-W. Bock, and G. Lee, "Predicting ethereum prices with machine learning based on blockchain information," *Expert Syst. Appl.*, vol. 184, Art. no. 115480, 2021, doi: 10.1016/j.eswa.2021.115480.

[15] M. F. N. Syahbani and N. G. Ramadhan, "Yoga movement classification using convolutional neural network model with streamlit framework," *J. Media Inform. Budidarma*, vol. 7, no. 1, pp. 509–519, 2023, doi: 10.30865/mib.v7i1.5520.

[16] P. A. Raharja, "Ethereum price prediction using vector autoregressive method," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 3, no. 2, pp. 71–79, 2021.

[17] V. Derbentseva *et al.*, "Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices," *Int. J. Eng. Trans. A Basics*, vol. 34, no. 1, pp. 140–148, 2021, doi: 10.5829/ije.2021.34.01a.16.

[18] A. P. Putri, R. R. Yusnita, and R. A. F. Mochtar, "Cryptocurrency portfolio selection: Bitcoin, yearn.finance, dogecoin, binance coin,

cardano, and monero," *Bisnis-Net J. Ekon. dan Bisnis*, vol. 6, no. 1, pp. 127–141, 2023, doi: 10.46576/bn.v6i1.3249.

[19] A. S. Prasetyo and R. E. Latumahina, "Legality of cryptocurrency as an investment tool in indonesia," *Bur. J. Indones. J. Law Soc. Gov.*, vol. 3, no. 1, pp. 204–214, 2023.

[20] A. Halim and R. Amalia, "Application of data mining using k-means method to determine public interest in cryptocurrency based on age," *Log. J. Ilmu Komput. dan Pendidik.*, vol. 2, no. 2, pp. 421–426, 2024.

[21] M. K. Abdalhammed *et al*., "Application of deep learning to predict crypto currency prices and their relationship to market adequacy (applied research bitcoin as an example)," *Financ. Theory Pract.*, vol. 26, no. 4, pp. 95–108, 2022, doi: 10.26794/2587-5671-2022-26-4-95-108.

[22] M. F. Rizkilloh and S. Widiyanesti, "Cryptocurrency price prediction using long short term memory (LSTM) algorithm," *J. RESTI*, vol. 6, no. 1, pp. 25–31, 2023, doi: 10.29207/resti.v6i1.3630.

[23] R. Faizal, B. D. Setiawan, and I. Cholissodin, "Cryptocurrency value prediction using extreme learning machine (ELM) algorithm," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 4226–4233, 2019.

[24] A. D. Sidik and A. Ansawarman, "Motor vehicle quantity prediction using machine learning," *Formosa J. Multidiscip. Res.*, vol. 1, no. 3, pp. 559–568, 2022, doi: 10.55927/fjmr.v1i3.745.

[25] A. Putranto, N. L. Azizah, and A. I. Ratna Ika, "Heart disease prediction system using SVM method and streamlit framework," *J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 4, no. 2, pp. 442–452, 2023.

[26] Y. Akkem, B. S. Kumar, and A. Varanasi, "Streamlit application for advanced ensemble learning methods in crop recommendation systems—a review and implementation," *Indian J. Sci. Technol.*, vol. 16, no. 48, pp. 4688–4702, 2023, doi: 10.17485/IJST/v16i48.2850.

[27] M. A. Salim and Y. Anistyasari, "Development of an online essay exam assessment application using nazief and adriani algorithm with cosine similarity method," *J. IT-EDU*, vol. 2, no. 1, pp. 126–135, 2017, doi: 10.26740/it-edu.v2i1.21338.

[28] M. Sobiri, "Optimization of Cryptocurrency Prediction Using Deep Learning Approach," *JSAI (J. Sci. Appl. Informatics)*, vol. 6, no. 2, pp. 197–204, 2023, doi: 10.36085/jsai.v6i2.5288.

[29] K. Kasliono, N. Candraningrum, and K. Sari, "Modeling ethereum price prediction (attributes: open, high, and low) using extreme learning machine algorithm," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 95−103, 2023, doi: 10.47065/bits.v5i1.3567.

[30] M. N. Pangestu, M. Jajuli, and U. Enri, "Prediction of NVIDIA graphics card prices based on cryptocurrency price influence using support vector regression," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 17, pp. 280–287, 2022, doi: 10.5281/zenodo.7076540.

[31] M. K. Anam and D. A. Jakaria, "Cryptocurrency price prediction system using regression method," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 2, pp. 467–479, 2023, doi: 10.35957/jatisi.v10i2.4787.

[32] M. A. Maliki, I. Cholissodin, and N. Yudistira, "Prediction of cryptocurrency bitcoin price movements against indonesian rupiah using LSTM algorithm," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 7, pp. 3259–3268, 2022.

[33] A. H. A. Othman, S. Kassim, R. B. Rosman, and N. H. B. Redzuan, "Prediction accuracy improvement for bitcoin market prices based on symmetric volatility information using artificial neural network approach," *J. Revenue Pricing Manag.*, vol. 19, no. 5, pp. 314–330, 2020, doi: 10.1057/s41272-020-00254-2.