Received: 23 January 2023 | Revised: 22 February 2023 | Accepted: 23 March 2023 | Published online: 24 March 2023

Journal of Data Science and Intelligent Systems
2024, Vol. 2(2) 100–112
DOI: 10.47852/bonviewJDSIS3202677

BON VIEW PUBLISHING

RESEARCH ARTICLE

Measuring the Performance of Machine Learning Forecasting Models to Support Bitcoin Investment Decisions

Wipawan Buathong[1], **Piroj Sieng-EK**[1] **and Pita Jarupunphol**[1,*]

[1]*Department of Digital Technology, Phuket Rajabhat University, Thailand*

Abstract: This research proposed machine learning forecasting models to support bitcoin investment decisions based on bitcoin price and trade volume from 2019 to 2021. The moving average crossovers of 5, 30, and 90 daily closing prices and their variances were inputs loaded into decision tree, random forest, and extreme gradient boosting (XGBoost) techniques to forecast bitcoin investment strategies, including market trends, actions, and holding amounts. The research also measured the models' performance based on accuracy, precision, recall, F1-score, and area under the curve-receiver operating characteristics (AUC-ROC). The results indicated that the XGBoost is the most efficient model: (1) trend (0.930 accuracy, 0.930 precision, 0.930 recall, 0.929 F1-score, and 0.983 AUC-ROC); (2) action (0.985 accuracy, 0.985 precision, 0.985 recall, 0.985 F1-score, and 0.998 AUC-ROC); and (3) amount (0.987 accuracy, 0.987 precision, 0.987 recall, 0.987 F1-score, and 0.997 AUC-ROC). The random forest achieved the second most efficient model, while the decision tree provided the lowest forecasting results. Since the bitcoin investment market in 2022 is significantly different from the previous two years due to several negative factors, the research further validated the models' performance with an unseen data set comprising 275 days of bitcoin market prices from January 1 to October 2, 2022. All the models suggested that investors hold with half the investment consistent with the investment market in 2022. Furthermore, although the decision tree and XGBoost models forecasted the investment trend for most days as up, the random forest forecasted the trend as sideway, consistent with the 2022 trend.

Keywords: bitcoin, cryptocurrency, decision tree, machine learning, moving average crossover

1. Introduction

According to Peters and Schutte (2022), trading with emotion, no entry and exit strategies, no limit order, and too many technical signals are the most common traders' mistakes in any investment. Nowadays, digital assets have played a crucial role in different perspectives, such as initial coin offering (ICO), fundraising through public offerings, and a medium for trading on the exchange. The weekly digital asset market summary report of the Securities and Exchange Commission provides information that the digital asset market had a total market capitalization of $2.38 trillion in December 2021. The approximate daily turnover is 70 billion US dollars. This market capitalization comes from bitcoin; 40% of the total investor accounts for approximately two million (The Securities and Exchange Commission, 2021). Therefore, Thailand has enacted the Emergency Decree on the Business of Digital Assets BE 2561 to regulate acts and procedures to protect investors and for transparency of digital assets investment Rat. This decree states that digital assets such as cryptocurrencies and digital tokens must be clarified and classified according to the nature of use and holders' rights.

Bitcoin, developed in 2009, is the first cryptocurrency and becoming more popular. This popularity has caused the price of bitcoin to skyrocket over time, leading to the creation of different types of cryptocurrency. Prominent characteristics of bitcoin include safety, reliability, and popularity. Users can create an anonymous account on distributed network technology, which is independent, unregulated by any agency, transparent, verifiable, and lower fees than other transactions, quick transactions, and hedges against inflation risks (Peng et al., 2022).

Nevertheless, there are some disadvantages of bitcoin. For instance, bitcoin is widely used as a medium for illegal business operations. There is also a risk of loss when a transaction fails and cannot be recovered. The most significant disadvantage of bitcoin is its very high price volatility compared to other assets such as stocks, gold, and bonds (Doumenis et al., 2021; Shahvari, 2022). During the COVID-19 outbreak, the bitcoin increase rate is much higher than other assets. However, bitcoin price volatility is still problematic for investors, who consider bitcoin a speculative asset rather than an investment asset. Therefore, this research applies machine learning techniques and moving average (MA)

*__Corresponding author:__ Pita Jarupunphol, Department of Digital Technology, Phuket Rajabhat University, Thailand. Email: p.jarupunphol@pkru.ac.th

crossovers based on short term (5 days), medium term (30 days), and long term (90 days) to analyze and forecast the bitcoin price movement. This research utilizes the daily data of bitcoin trading volume, opening price, closing price, highest price, and lowest price between October 1, 2018, and December 31, 2021. The forecasting model will logically inform investors of three indicators: (1) market trend (uptrend, downtrend, and sideway); (2) action (buy, sell, and hold); and (3) amount (all in, none, and half or neutral). In addition, the forecasting model also logically advises investors to adjust their bitcoin holding amounts according to the bitcoin price situation by balancing stability and growth in the invested asset due to the volatility factor of the bitcoin price at a certain period. The model emphasizes MA crossover indicators without utilizing many technical signals, consistent with the suggestions of Peters and Schutte (2022).

2. Literature Review

Cryptocurrency is a type of digital currency, an electronic unit depending upon cryptography and blockchain technology to validate transactions, control the creation of additional monetary units, and confirm the correctness of the transfer of assets. Several leading studies have found that cryptocurrencies can help prevent or mitigate investment portfolio risks. For example, according to Akhtaruzzaman et al. (2021), the US bitcoin investment data between 2011 and 2018 implied that bitcoin could avoid or minimize risks in industrial investment portfolios because the bitcoin price does not move entirely in the same direction as other assets. In addition, Urquhart and Zhang (2019) pointed out that bitcoin can hedge and diversify trading between cryptocurrencies.

Nevertheless, there are some disadvantages of bitcoin. For instance, bitcoin is misused as a medium for illegal business operations. There is also a risk of loss when a transaction fails and cannot recover. The most significant disadvantage of bitcoin is its very high price volatility compared to other assets (Gazali, 2019). Therefore, the bitcoin actual price will never be stable since it depends on the buyer and seller's satisfaction. Different types of cryptocurrency have been created as a medium for exchanging goods, services, and other digital assets, including digital tokens divided into (1) investment tokens and (2) utility tokens. An investment token is an electronic unit created to determine joint investment rights, such as revenue-sharing rights. Investment profit is issued and offered for sale through the ICO process. However, the utility token is an electronic unit created to determine the right to receive a specific good or service issued and offered for sale through the ICO process. In Thailand, cryptocurrencies were initially under the Digital Asset Business Act, which is only a medium for trading on the exchange only. As for the digital tokens granted to holders, the value of the digital tokens depends on the correct value.

Several asset forecasting approaches involve basic information about the asset, such as its share value, corporate profit, expectations, and demand for the investment. This type of data analysis is a "fundamental analysis," whereas "technical analysis" is another kind of analysis that is concerned only with the fluctuation of price and time series, considering the historical price, the current price, and the number of currently trading assets (Murphy, 1999). Since digital assets do not have intrinsic value derived from operating profits or data from other investments, the research emphasizes only technical analysis. Technical analysis of assets arises from forex trading driven by supply and demand as the basis for price fluctuations, initially proposed around the 17th century. In the late 19th century, technical analysis became more popular with the Dow Theory, introduced by Charles Dow, founder and editor of The Wall Street Journal. Three essential technical analysis components include (1) the price summary that fully reflects the information for each period; (2) the price moves in a trend; and (3) the price pattern repeatedly occurs due to investor behavior. There are three types of price trends: uptrend, downtrend, and sideways. Technical analysis can be done directly by learning the chart pattern or through indicators such as price averages, converting to other indices providing a picture of different market perspectives.

2.1. Moving average

MA is an average calculation using the asset's price historically for a specified period. For example, the 5-day asset price will be calculated using the formula of the desired average type to understand the 5-day MA value. A single average cannot provide sufficient information to analyze a digital asset. Therefore, several consecutive averages are determined and displayed as a line along with the price chart. The benefit of using MAs is that indicator representations are smoothed through the MAs by spreading anomalies out of the data. MA can help track if the price trend is near its end or entering a new direction, identifying buy and sell signals more clearly than analyzing the price chart directly. The asset's daily trading closing price is mainly preferred to calculate the MA. The most popular MAs in the stock analysis are (1) simple moving average (SMA) is the most straightforward calculation where all data are weighted equally; (2) linear weighted moving average (LWMA) is a calculation that gives more weight to the latest data than the previous data; and (3) exponential moving average (EMA) is a more complex weighted calculation. The smoothing constant adjusts the next average to be more accurate (Murphy, 1999; Zakamulin, 2017). Equations (1)–(3) represent calculations of SMA, LWMA, and EMA, where P is the price, n is the number of days or periods, W is the weight of each period, and EMA is the exponential moving average.

$$SMA = \frac{P_1 + P_2 + P_3 \ldots + P_n}{n} \tag{1}$$

$$LWMA = \frac{[\![(P]\!]_n * W_1) + [\![(P]\!]_{n-1} * W_2) + [\![(P]\!]_{n-2} * W_3)}{\sum W} \tag{2}$$

$$EMA_{Today} = \left(Value_{Today} * \left(\frac{Smoothing}{1 + Days} \right) \right) + EMA_{Yesterday} * \left(1 - \left(\frac{Smoothing}{1 + Days} \right) \right) \tag{3}$$

2.2. MA crossover

MA crossover is a technical analysis method based on two or more MAs of different historical periods to analyze the market and price trends. The EMA is the most commonly used indicator because it responds faster to price changes than other averages. "Golden cross" and "death cross" are two types of MA crossovers. While the golden cross signals a bull market when the short-term MA crosses above the long-term MA, the death cross signals a bear market when the short-term MA crosses the long-term MA. The MA crossover should collaborate with a high volume of three-line averages to understand the market's current short-term, medium-term, and long-term trends. The long-term average indicates a direction, while the short-term average indicates momentum.

2.3. Decision tree learning

Decision tree learning, developed by John Ross Quinlan in 1986, is a machine learning classification method based on statistics from previously grouped training data sets (Ogheneovo & Nlerum, 2020). The system learns to identify patterns and relationships from the features of the input data set in each instance by creating rules and using these rules to decide how to group the test data sets. A decision tree comprises three components: the leaf, the branch, and the root. The working principle of a decision tree is the selection of the most defining factors or features of the data to generate the tree root to separate data classes with the least amount of mixing. Such attributes are based on the entropy and gain for each feature (Bhatia, 2019). Given that the data have many S items, the data class consists of positive and negative classes. Equation (4) shows the mixing degree of S data, where S is the total learning data set, P is the number of positive classes learned, and N is the number of negative classes learned.

$$Entropy(S) = \frac{-P}{P+N} log_2 \frac{P}{P+N} - \frac{N}{P+N} log_2 \frac{N}{P+N} \quad (4)$$

If the data contain more than two classes, the entropy formula can be displayed in Equation (5), where pi is the proportion of data in class i relative to the total data, C is the total number of classes, and N is the number of negative classes learned. If the entropy is 0, the data are organized. On the other hand, the data are unorganized if the entropy is 1.

$$Entropy(S) = \sum_{i=1}^{C} - p_i log_2 p_i \quad (5)$$

Gini index is a famous formula used to measure the degree of data compromise (Rokach & Maimon, 2015). If the Gini index is small, then the information is less mixed. If the value is 1, the data are mixed and unorganized. The criterion for deciding on a root node is to test each feature acting as the root node and find the gain ratio, a value that tells which attributes to serve as the root node. Equation (6) represents the Gini index formula, where p_i is the proportion of data in class i relative to the total data.

$$Gini(S) = 1 - \sum_{i=1}^{k} p_i^2 \quad (6)$$

Information gain calculates each feature weight, which can only be applicable with nominal attributes (Bramer, 2016). Information gain measures the data disorder before and after the data are divided according to the class. If efficiency is improved, information gain is high. Information gain utilizes entropy to measure the difference or scatter of information. If the information is different or scattered, the entropy is high. On the other hand, if the data are very similar or the scatter is small, the entropy will be low (Witten et al., 2016). Equations (7) and (8) represent calculations of the information gain, where (1) S is the data before attribute A is utilized as a separator; (2) |SV| is the number of items whose attribute value is A; (3) |S| is the total number of entries in the data set before the discrimination; (4) Entropy(S) is the disorderly value of the learning data set S; and (5) Entropy(SV) is the disorderly value of the data distinguished by attribute A with a value of v.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Vaue(A)} \frac{|S_V|}{|S|} Entropy(S_V) \quad (7)$$

$$Gain(S, A) = Gini(S) - \sum_{v \in Value(A)} \frac{|S_V|}{|S|} \left(1 - \sum_{i=1}^{k} p_i^2\right) \quad (8)$$

The criterion for deciding on a root node is to test each feature acting as the root node and find the "gain ratio," identifying which attributes to serve as the root node. The measurement unit is bits derived from computations based on Information Theory, where a data's information value depends on the data's probability. The feature with the most gain ratio will be selected as the root node as shown in Equations (9)–(11), where (1) T represents the set of the learned data set; (2) x represents the attribute chosen as the classifier; (3) $info(T)$ is a function that specifies the amount of data required to be able to identify the necessary features; (4) $|T|$ is the total number of data in the learning data set; (5) $freq(C_j, T)$ is the frequency at which data in T appear as class C_j; (6) $|(info)|x(T)$ or entropy is a function that specifies the amount of data required to classify a data class using the x attribute as a validator to extract the data; (7) i is the number of possible values of the x attribute; and (8) $|T_i|$ is the number of data values $x = i$.

$$Gain(x) = info(T) - info_x(T) \quad (9)$$

$$info(T) = -\sum_{j=1 \text{ to } k} \left[\frac{freq(C_j, T)}{|T|}\right] * log_2 \left[\frac{freq(C_j, T)}{|T|}\right] bits \quad (10)$$

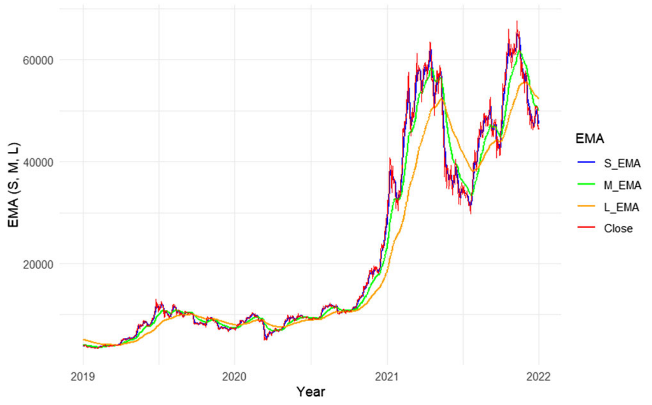
$$info(T) = \sum_{j=1 \text{ to } n} \frac{|T_j|}{|T|} * info(T_i) bits \quad (11)$$

Random forest is a machine learning process developed from decision trees by modeling multiple sub-decision trees (Rokach & Maimon, 2015; Sheppard, 2017; Youngberg, 2021). Each tree receives different information as a subset of the aggregate data and forecasts the value according to a computed model. XGBoost is another machine learning process using a forecast error analysis method and takes the results to adjust the model for more accurate forecasts (Wade & Glynn, 2020). Please note that XGBoost is a highly efficient ensemble learning technique that encompasses three main methods: bagging, boosting, and stacking. Of these methods, gradient boosting algorithm is particularly noteworthy due to its sequential learning process, where previous learners are utilized to continuously reduce errors and improve accuracy. Unlike bagging, gradient boosting can produce a lower bias, but it is also more prone to overfitting. The XGBoost algorithm builds upon the gradient boosting method, incorporating optimization features such as regularization to mitigate overfitting and sparse aware handling for missing values, resulting in a robust and professional model.

2.4. Conceptual framework

Successful investors should be able to analyze market conditions and price trends for a range of assets and adjust the proportion of their investments to suit the ever-changing situation. Machine learning techniques can achieve this analysis ability by learning from past events and information that affects current assets. As such, the researchers used machine learning to study the changes in bitcoin asset prices from 2019 to 2021, a total of 3 years of bitcoin market forecast using 3 EMA MAs for

Figure 1
The relationship between the closing price and the EMA from 2019 to 2021



analysis as displayed in Figure 1, where S_EMA represents the weekly short-term average (5 days), M_EMA represents the monthly mid-term average (30 days), and L_EMA represents the quarterly long-term average (90 days).

In addition, there are three different values derived from the MAs: (1) SM_DIF (the difference between the short-term and medium-term averages [M_EMA – S_EMA]); (2) SL_DIF (the difference between the short-term and long-term average [L_EMA – S_EMA]), and ML_DIF (the difference between the medium-term average and the long-term averages [L_EMA – M_EMA]). The differences are then combined with the preliminary daily data, namely the open price (open), the closing price (close), the highest price (high), the lowest price (low), and volume (volume). In this case, 11 features are derived and analyzed using the decision tree to forecast three market strategies: trend (uptrend, downtrend, sideways), actions (buy, sell, hold), and amount (all, half, none).

The trend in the market can be determined by the alignment of the three EMAs.

(1) A robust upward trend is indicated when the short-term EMA is positioned above the medium-term EMA, and the medium-term EMA is positioned above the long-term EMA. Under these conditions, the recommended action would be to initiate a purchase and invest the entire available amount (action = buy, amount = all).
(2) An upward trend is indicated when the short-term MA crosses below the medium-term MA, and the latter is above the long-term MA. Under these circumstances, it is recommended to initiate a buying position with a moderate investment amount, equivalent to half the total allocation (action = buy, amount = half).
(3) In instances where the market trend exhibits a sideways movement or lacks a discernible trend, it is advisable to adopt a wait-and-observe approach and maintain a portion of the investment (action = hold, amount = half).
(4) A downward trend is indicated when the short-term MA is positioned above the medium-term MA and the medium-term MA is located below the long-term MA. In such a scenario, the recommended action would be to sell, with the recommended quantity being half (action = sell, amount = half).
(5) A robust downward trend is indicated when the short-term trend line falls below the medium-term MA, and the medium-term MA is in turn below the long-term MA. In this case, the recommended action is to sell, with the recommended amount being none (action = sell, amount = none).

2.5. Related works

Numerous studies have been conducted on the prediction of bitcoin utilizing a variety of statistical and machine learning techniques. Consideration was given to a range of factors in the forecasting process. For example, Mudassir et al. (2020) experimented with bitcoin price forecasting based on bitcoin price data from 2013 to 2020. Various indices include the weighted average of prices over 3, 7, and 30 up to 90 days, daily transactions, mining volume, and mining yield. The forecast reached 62% accuracy over 90 days with machine learning techniques and increased accuracy to 65% one day in advance (Chen et al., 2020). In addition, Li et al. (2019) collected social media conversations about cryptocurrencies from Twitter every hour for 3.5 weeks. The authors applied MA indicators and machine learning techniques to understand the bitcoin price movements. As a result, there was a significant correlation of 81% between the expressed emotions and the current trends in digital asset prices.

Basher and Sadorsky (2022) combined tree-based machine learning classifiers and conventional logit econometric models. They posited that random forest algorithms were capable of more accurate predictions of bitcoin and gold prices compared to logit models. The authors further claimed that using bagging and random forest techniques could result in an accuracy rate ranging from 75% to 80% for a 5-day prediction of bitcoin prices and 85% for predictions spanning 10–20 days.

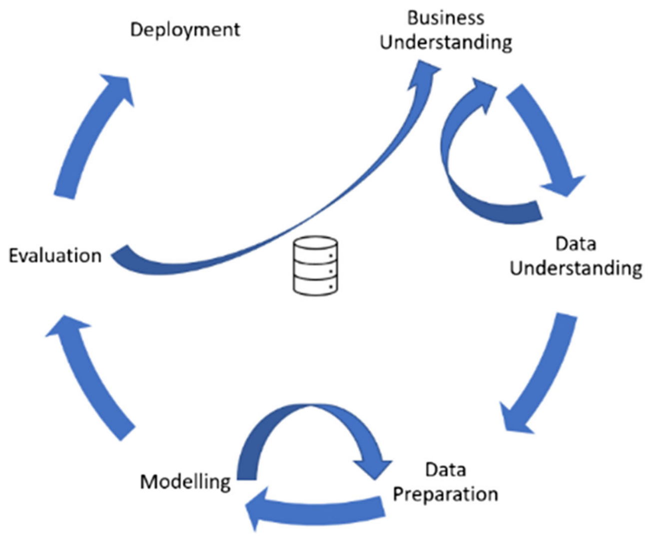
According to Rathore et al. (2022), forecasting the price of bitcoin is challenging due to its dynamic and volatile nature, which is influenced by various factors such as seasonal changes. The authors noted that traditional models such as Long Short-Term Memory Neural Network (LSTM) and Autoregressive Integrated Moving Average (ARIMA) struggle to effectively predict the price of bitcoin due to these complexities. To address this issue, the authors proposed using Fbprophet as a more efficient model for predicting the price of bitcoin. This model was designed to be more rigorous and eliminate the limitations and inaccuracies generated by Long Short-Term Memory Neural Network (LSTM) and Autoregressive Integrated Moving Average (ARIMA) models. The authors concluded that the Fbprophet model exhibits a low difference between predicted and actual values, making it a viable option for real-world applications.

Jaquart et al. (2021) explored the feasibility of using various machine learning models to predict the future movements of the bitcoin market across prediction horizons that spanned from 1 to 60 min. The authors found that recurrent neural networks and gradient-boosting classifiers were effective for the prediction tasks. A range of features, including technical, blockchain-based, sentiment/interest-based, and asset-based features, were utilized in the analysis. The findings indicate that technical features remained the most impactful for most models, followed by a selection of blockchain-based and sentiment/interest-based features. The authors concluded that the predictability of the bitcoin market increases as the prediction horizon expands.

3. Research Methodology

This research methodology follows the "Cross-Industry Standard Process for Data Mining (CRISP-DM)," developed in 1996 with four agencies, DaimlerChrysler, SPSS, NCR, and OHRA (Chapman et al., 2000; Shearer, 2000). This process consists of six steps, shown in Figure 2 (Chapman et al., 2000).

Figure 2
Data analysis process with CRISP-DM



3.1. Business understanding

The first step in the CRISP-DM emphasizes understanding the business or problem, converting the issue into a data analysis process, and then planning to analyze the data. This research aims to study the bitcoin price transformation with a high rate of change that helps diagnose and forecast bitcoin price trends. Inexperienced investors, who may easily make wrong decisions, are advised to understand the current market conditions and trends. In addition, this research also suggests strategies and appropriate holding volumes for stability, reducing risk in investment assets.

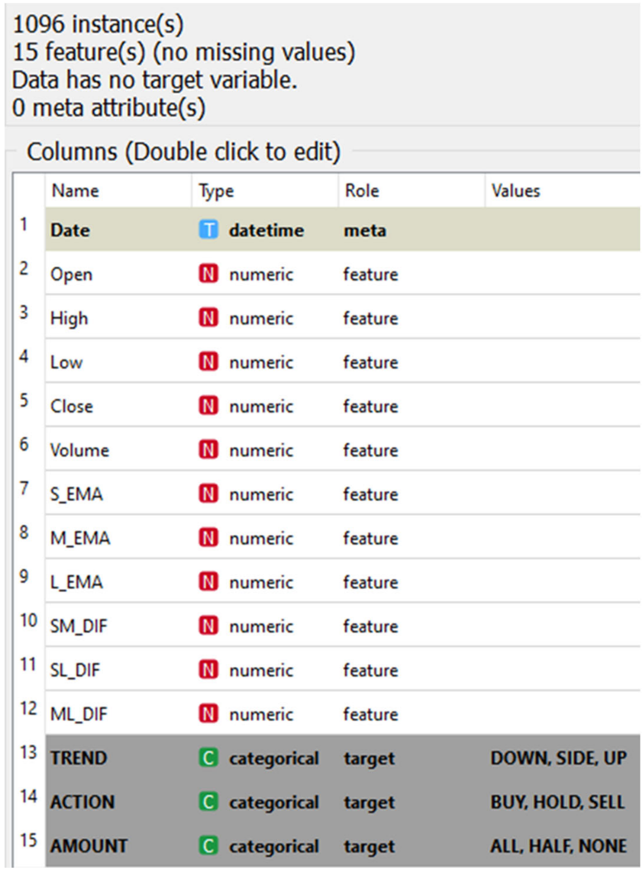
3.2. Data understanding

This step involves collecting the data required for analysis. First, the daily bitcoin price data from 2018 to 2021 were downloaded from the website of Yahoo Finance using the R programming language with the "quantmod" (Quantitative Financial Modeling Framework) and XTS (eXtensible Time Series) packages. Next, the EMAs of 5 days, 30 days, and 90 days and the differences between the MAs are calculated. Finally, the calculated data are transformed into a table containing Date, Open, High, Low, Close, Adjusted, and Volume columns.

3.3. Data preparation

This step involves several data transformation steps, from the screening of the essential data to the transformation of the data. First, parts that are not required are removed from the data. As this research utilizes daily price data of the bitcoin price between the years 2019 to 2021 for analysis, a minimum of 30 days of data is necessary for calculating the long-term average as of January 1, 2019. Once the calculation of the long-term average has been completed, the information is deemed redundant and is subsequently removed. After checking the transformed data, the daily closing price (Close) is used to calculate S_EMA, M_EMA, and L_EMA at 5 days, 30 days and 90 days, respectively, filtering the data between 2019 and 2021 only, and then the difference is calculated between the MAs SM_DIF(M_EMA – S_EMA), SL_DIF(L_EMA – S_EMA), and ML_DIF(L_EMA – M_EMA). Next, the daily data classes are defined for analyzing three strategies: (1) trends (uptrend, downtrend, sideways) in the TREND column; (2) actions (buy, sell, hold) in the ACTION column; and (3) amounts (all, half, none) into the AMOUNT column. These goals are based on the opening price, closing price, high price, lowest price, and daily trading volume. There are 1,096 instances and 11 features, as shown in Figure 3.

Figure 3
Data preparation



1096 instance(s)
15 feature(s) (no missing values)
Data has no target variable.
0 meta attribute(s)

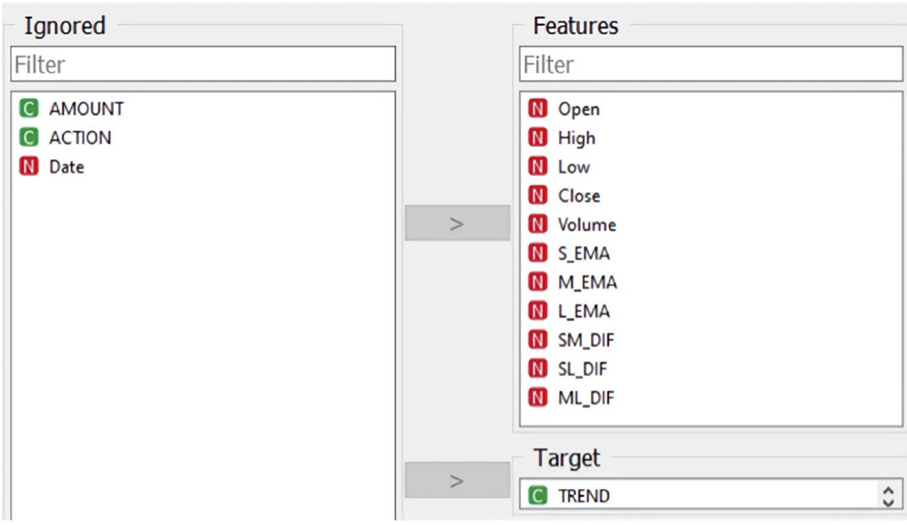
Columns (Double click to edit)

	Name	Type	Role	Values
1	Date	datetime	meta	
2	Open	numeric	feature	
3	High	numeric	feature	
4	Low	numeric	feature	
5	Close	numeric	feature	
6	Volume	numeric	feature	
7	S_EMA	numeric	feature	
8	M_EMA	numeric	feature	
9	L_EMA	numeric	feature	
10	SM_DIF	numeric	feature	
11	SL_DIF	numeric	feature	
12	ML_DIF	numeric	feature	
13	TREND	categorical	target	DOWN, SIDE, UP
14	ACTION	categorical	target	BUY, HOLD, SELL
15	AMOUNT	categorical	target	ALL, HALF, NONE

3.4. Modeling

This step generates a model based on the decision trees using the Orange Data Mining tool developed by the University of Ljubljana, Slovenia (Demšar et al., 2013). Orange Data Mining also provides tools for importing, modifying, and selecting the data to be analyzed. The analysis can create forecasting models based on supervised, unsupervised, and time series in a single window. The analysis begins by importing the data in the provided CSV format. Then, the columns are selected to connect to the data file, set a target to TREND, and select the 11 features. Three types of decision tree-based algorithms are chosen for forecasting: decision tree, random forest, and XGBoost. The decision tree algorithm utilizes the decision tree principle to select the most significant features of the data with the highest discriminatory power to establish the roots of the tree and minimize contamination of data classes. The random forest, on the other hand, is an ensemble learning technique or "bagging" approach that optimizes the

Figure 4
Defining attributes to create a trend-target forecaster (TREND) model



machine learning process using multiple decision trees constructed from the entire data set. Lastly, XGBoost is a boosting ensemble learning method that is optimized using the gradient boosting algorithm. The Test and Score, Confusion Matrix, and ROC Analysis widgets join the data and models in Figure 4. These modeling processes are also required for ACTION and AMOUNT targets.

3.5. Model evaluation

Before deployment, it is crucial to thoroughly evaluate the model to determine if it is appropriately qualified and to assess whether it is susceptible to overfitting or underfitting. The research conducted a thorough evaluation of multiple methods for assessing model performance, including 5-Fold Cross Validation, 10-Fold Cross Validation, and 70:30 Percentage Split. The results of the experimentation indicated that the 10-Fold Cross Validation approach provided the best performance for the given data set. Based on these findings, this method is the most appropriate choice and is therefore presented as the recommended approach in this study. The model performance measurements can utilize all learned data or the "Self Consistency Test," which will get an accuracy value that is high or equal to 100%, dividing the data into training data and testing data sets. After that, the performance is measured using 10-Fold Cross Validation divided into 10 equal parts, each randomized for distribution. The first test is performed using the first data set as the test data. The second data set to the tenth is the learning data set for measuring forecast accuracy. Then, the second model was tested using the second set of data as test data. The first and third data sets to the tenth are learning data sets. These processes will repeat this 10 times. The forecast performance is the average of all accuracy. The decision tree performance is based on the confusion matrix comparing the validity of the predicted class to the actual event class. As shown in Figure 5(a) and (b), a confusion matrix is a table that is used to evaluate the performance of a classification model. It is typically arranged as a 2 × 2 in F or an N × N table, where N is the number of classes in the classification problem. The rows of the table represent the actual class labels, while the columns represent the predicted class labels. For example, there are components where predicted class is the expected value of actual class: (1) true positive (TP) means the forecast is "true," and the value is "true"; (2) false positive (FP) means the forecast is "true," but the value is "false"; (3) false negative (FN) means the forecast is "false," but the value is "true"; and (4) true negative (TN) means the forecast is "false,", and the value is "false."

Figure 5
Confusion Matrix for measuring the model's performance: (a) A 2*2 Confusion Matrix (b) A 1..n Confusion Matrix

(a)

		Predicted Class	
		Positive	Negative
Actual Class	Positive	**TP**	**FN**
	Negative	**FP**	**TN**

(b)

		Predicted Class			
		C_1	C_2	…	C_N
Actual Class	C_1	$C_{1,1}$	**FP**	…	$C_{1,N}$
	C_2	**FN**	**TP**	…	**FN**
	…	…	…	…	…
	C_N	$C_{N,1}$	**FP**	…	$C_{N,N}$

There are six criteria to measure the models' performance.

(1) Accuracy is the value at which the forecast model is divided by all values, which can be calculated using Equation (12).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{12}$$

(2) Precision is the value at which the model accurately predicts whether the class in question is divided by the model's value as both true and false, which can be calculated using Equation (13).

$$Precision = \frac{TP}{(TP + FP)} \tag{13}$$

(3) Recall is the value at which the model accurately predicts the class under consideration divided by the valid event values of all considered classes, which can be calculated in Equation (14).

$$Recall = \frac{TP}{(TP + FN)} \tag{14}$$

(4) The F1-Score, which is calculated as the harmonic mean of precision and recall, is utilized to assess the overall efficiency of the model, as described by Equation (15).

$$F1 = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right) \tag{15}$$

(5) Specificity, also known as the TN rate, represents the accuracy of the forecast model in correctly identifying instances that belong to the non-divisible class. It is calculated as the ratio of TNs to the sum of TNs and FPs, as described in Equation (16).

$$Specificity = \frac{TN}{(TN + FP)} \tag{16}$$

(6) area under the curve-receiver operating characteristics (AUC-ROC) is a correlation curve metric to visualize the performance metrics in the case of multiclass classification. ROC is a graph showing the relationship between a TP rate in a vertical direction and a FP rate in a horizontal direction, represented by 1 specificity. The model performance is highly efficient when the AUC value approaches one because the model can accurately predict "true" or "false."

3.6. Deployment

The model is ready to deploy when the model performance metrics achieve satisfactory results. The model can analyze trends and provide recommendations for decision-making in adjusting investment asset ratios. There are several ways to implement the model. For example, the model may be stored as an object (pickle object when modeled in Python or RDS object when built-in R) and then used in a forecasting program with new attributes. In addition, the model may create rules and use them as a guide for forecasting, or the model may be connected to other systems via an application programming interface.

4. Results and Discussion

This section describes the results of the decision tree-based machine learning models derived from three algorithms: decision tree, random forest, and XGBoost, in forecasting bitcoin price trends. First, the MA crossover method includes measuring the accuracy of the investor advice. Then, the values classified in the three targets, including TREND, ACTION, and AMOUNT, were compared. As a result, 1,096 transactions between 2019 and 2021 were classified into eight indicators: opening, closing, high, low, daily trading volumes, and short-term, medium-term, and long-term MAs. They were later utilized to define three data classes: (1) TREND (375 UP, 584 SIDE, and 137 DOWN); (2) ACTION (114 BUY, 954 HOLD, and 28 SELL); and (3) AMOUNT (69 ALL, 899 HALF, and 128 NONE). Finally, the 11 features mentioned previously were utilized for three machine learning models, targeting TREND, ACTION, and AMOUNT.

4.1. Trend

According to Figure 6, there were 375 true events in UP class. The forecasting results toward trend of the models are (1) decision tree (312 true, 63 false); (2) random forest (322 true, 48 false); and (3) XGBoost (340 true, 35 false). In terms of SIDE class, there were 584 true events in total with the forecasting results: (1) decision tree (534 true, 50 false); (2) random forest (553 true, 31 false); and (3) XGBoost (562 true, 22 false). Considering the total 137 true events in the DOWN class, the decision tree, the random forest, and the XGBoost could forecast true events at 109, 112, and 117, respectively. The AUC-ROC curve represents the forecast performance of the UP, SIDE, and DOWN in Figure 6(d), (e), and (f).

The models' performances toward TREND are (1) accuracy [decision tree (0.871), random forest (0.905), and XGBoost (0.930)]; (2) precision [decision tree (0.871), random forest (0.899), and XGBoost (0.930)]; (3) recall [decision tree (0.871), random forest (0.905), and XGBoost (0.930)]; (4) F1-scores [decision tree (0.871), random forest (0.904), and XGBoost (0.929)]; and (5) AUC-ROC [decision tree (0.896), random forest (0.972), and XGBoost (0.983)]. Figure 7 illustrates the models' forecast efficiency toward TREND targets.

4.2. Action

Figure 8 shows 114 true events in the BUY class. The forecasting results of the models are (1) decision tree (99 true, 15 false); (2) random forest (103 true, 11 false); and (3) XGBoost (107 true, 7 false). While there are 954 true events in the HOLD class, the results are (1) decision tree (946 true, 8 false); (2) random forest (945 true, 9 false events); and (3) XGBoost (949 true, 5 false). For the SELL class, the decision tree, random forest, and the XGBoost could forecast true events at 23, 21, and 24, respectively. The forecast performance values of the BUY, HOLD, and SELL classes can be represented by the AUC-ROC curve in Figure 8(d), (e), and (f).

When the forecast results of ACTION targets were taken with different decision tree processes, the models' performances are (1) accuracy [decision tree (0.974), random forest (0.975), and XGBoost (0.985)]; (2) precision [decision tree (0.974), random

Figure 6
The confusion matrix and the AUC-ROC for forecasting TREND targets: (a) decision tree, (b) random forest, (c) XGBoost, (d) UP, (e) SIDE, and (f) DOWN

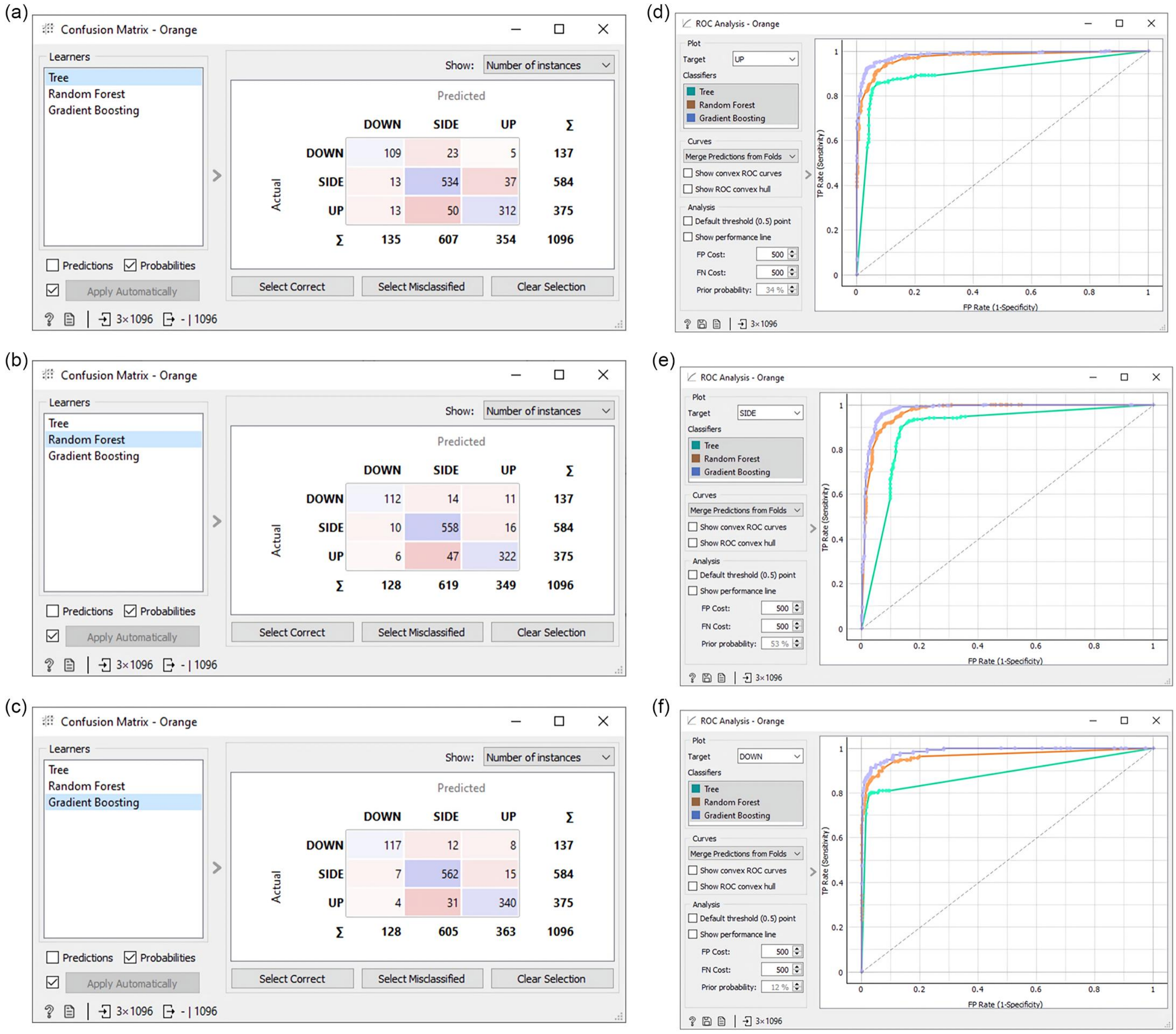
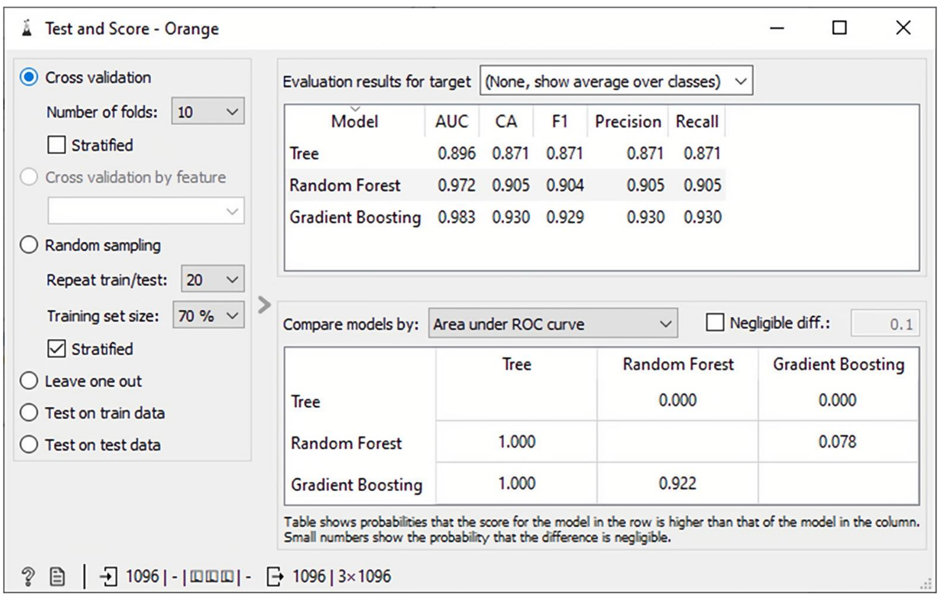


Figure 7
Comparison of the forecast performance values of the decision tree, random forest, and XGBoost in TREND targets



Model	AUC	CA	F1	Precision	Recall
Tree	0.896	0.871	0.871	0.871	0.871
Random Forest	0.972	0.905	0.904	0.905	0.905
Gradient Boosting	0.983	0.930	0.929	0.930	0.930

	Tree	Random Forest	Gradient Boosting
Tree		0.000	0.000
Random Forest	1.000		0.078
Gradient Boosting	1.000	0.922	

Figure 8
The confusion matrix and the AUC-ROC for forecasting ACTION targets: (a) decision tree, (b) random forest, (c) XGBoost, (d) BUY, (e) HOLD, and (f) SELL

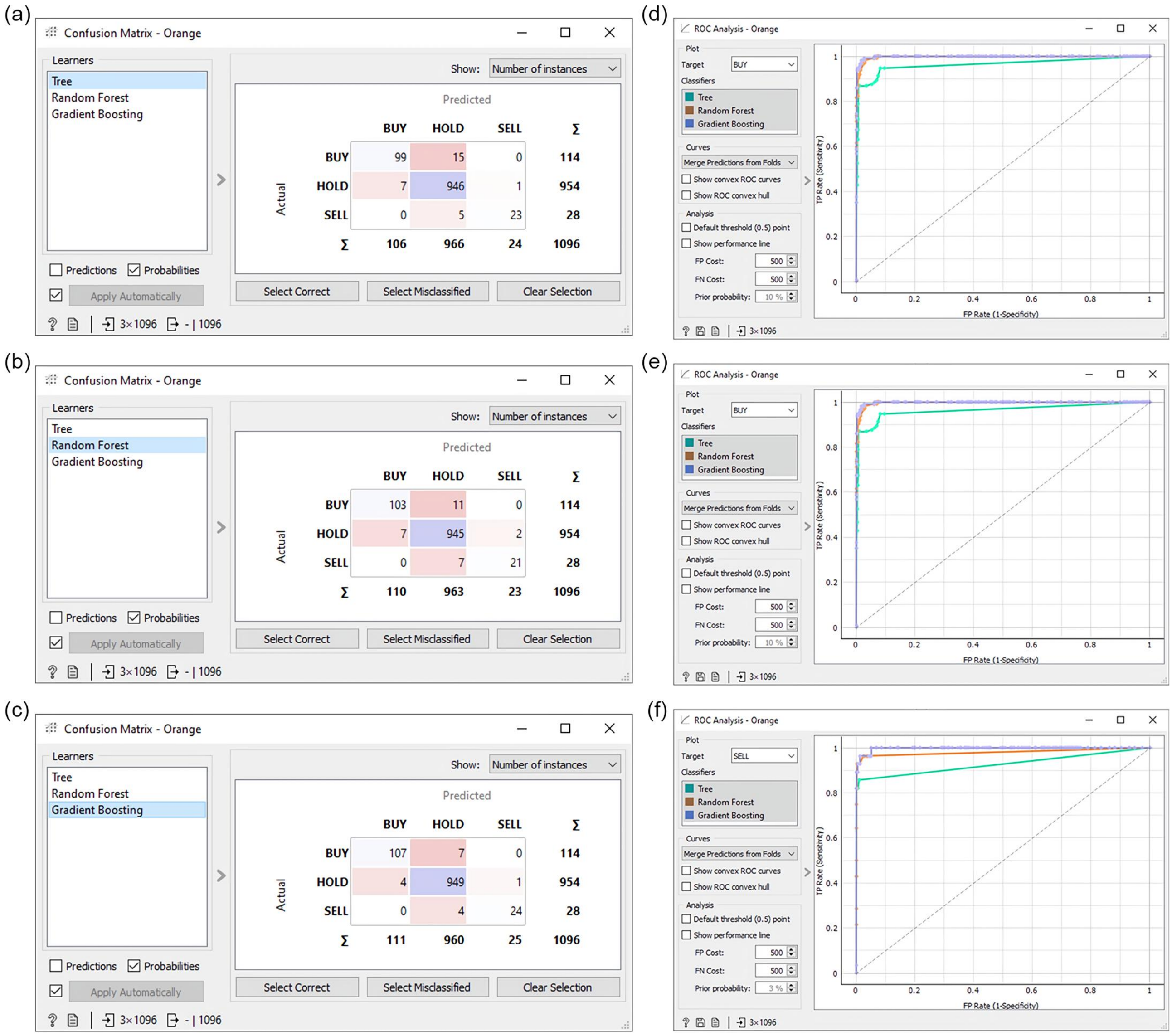
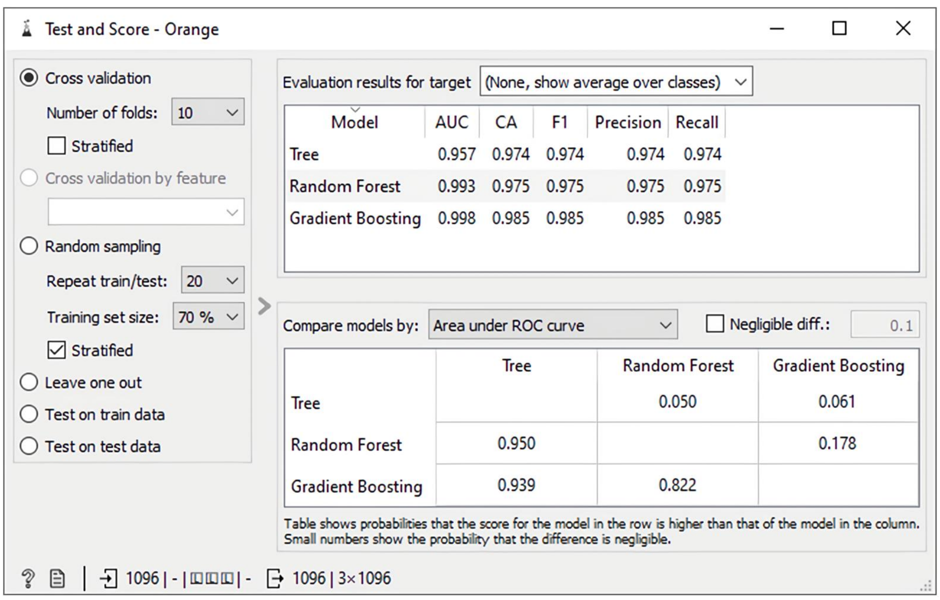


Figure 9
Comparison of the forecast performance values of decision tree, random forest, and XGBoost processes in ACTION targets



Model	AUC	CA	F1	Precision	Recall
Tree	0.957	0.974	0.974	0.974	0.974
Random Forest	0.993	0.975	0.975	0.975	0.975
Gradient Boosting	0.998	0.985	0.985	0.985	0.985

	Tree	Random Forest	Gradient Boosting
Tree		0.050	0.061
Random Forest	0.950		0.178
Gradient Boosting	0.939	0.822	

forest (0.975), and XGBoost (0.985)]; (3) recall precision [decision tree (0.974), random forest (0.975), and XGBoost (0.985)]; (4) F1-scores [decision tree (0.974), random forest (0.975), and XGBoost (0.985)]; and (5) AUC-ROC [decision tree (0.957), random forest (0.993), and XGBoost (0.998)]. Figure 9 illustrates the models' forecast efficiency toward ACTION targets.

4.3. Amount

According to Figure 10, there are 69 true events in the ALL class with the forecasting results: (1) decision tree (59 true, 10 false); (2) random forest (60 true, 9 false); and (3) XGBoost (61 true, 8 false). Besides, there is a total of 899 true events in the HALF class: (1) decision tree (887 true, 12 false); (2) random forest (896 true, 3 false); and (3) XGBoost (895 true, 4 false). Therefore, from a total of 128 true events in the NONE class, the decision tree, the random forest, and the XGBoost could predict true events at 123, 124, and 126, respectively. The forecast performance values of the ALL, HALF, and NONE classes are represented by the AUC-ROC curve in Figure 10(d), (e), and (f).

For the AMOUNT target, the measurement results are (1) accuracy [decision tree (0.975), random forest (0.985), and XGBoost (0.987)]; (2) precision [decision tree (0.975), random forest (0.985), and XGBoost (0.987)]; (3) recall [decision tree (0.975), random forest (0.985), and XGBoost (0.987)]; (4) F1-scores [decision tree (0.975), random forest (0.985), and XGBoost (0.987)]; and (5) AUC-ROC [decision tree (0.980), random forest (0.996), and XGBoost (0.997). Figure 11 shows the models' forecast efficiency toward AMOUNT targets.

Figure 10
The confusion matrix and the AUC-ROC for forecasting AMOUNT targets: (a) decision tree, (b) random forest, (c) XGBoost, (d) ALL, (e) HALF, and (f) NONE

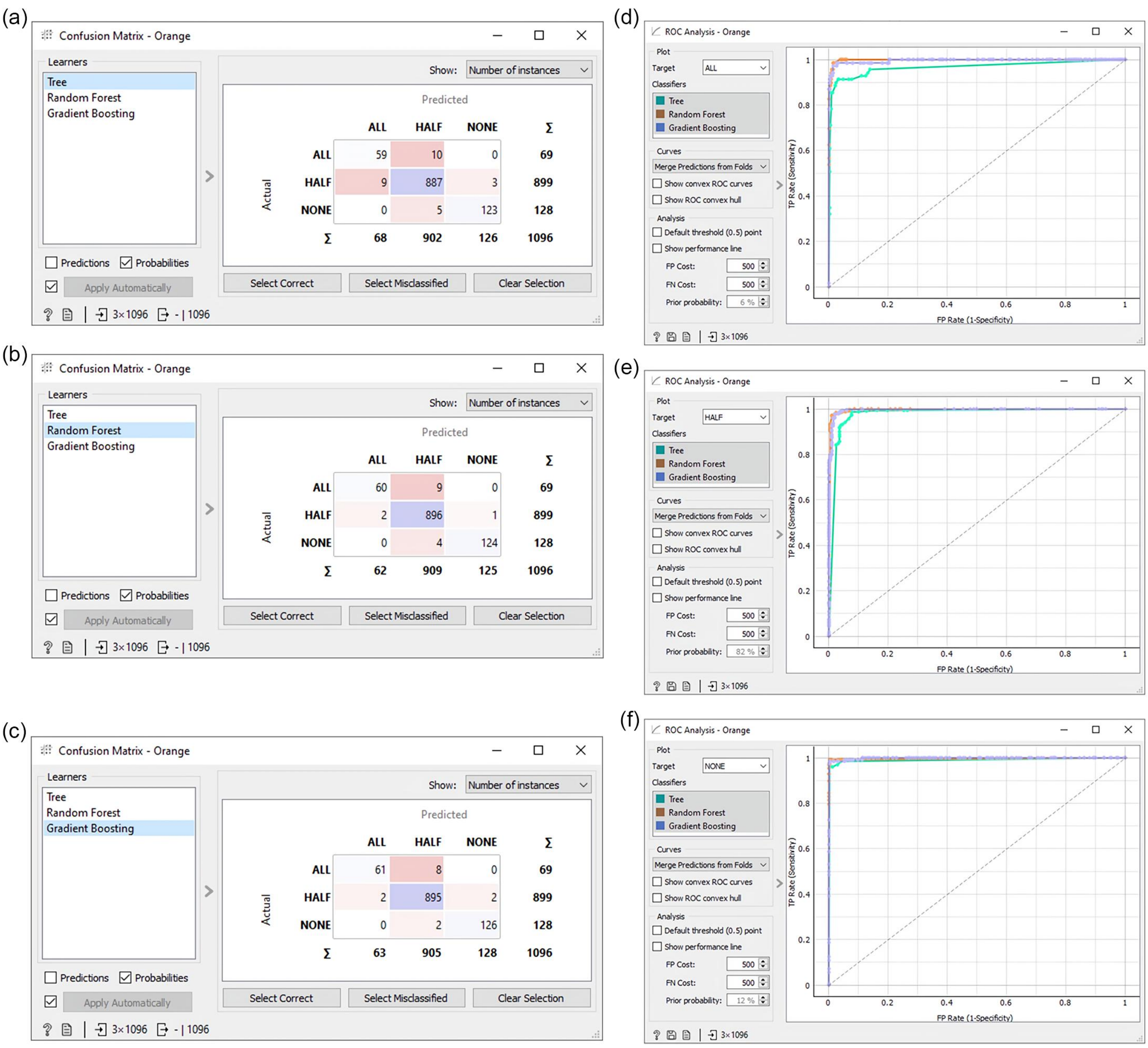


Figure 11
Comparison of the forecast performance values of decision tree, random forest, and XGBoost processes in AMOUNT targets

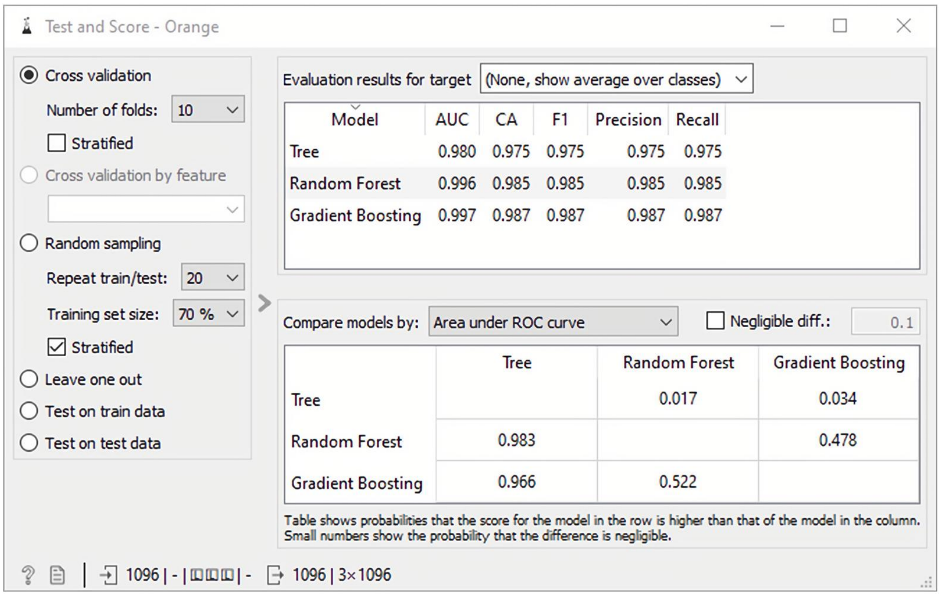


Table 1 summarizes all the models' predictive results based on accuracy, precision, recall, F1-score, and AUC-ROC. The results indicate that the XGBoost provides the most efficient prediction for all three targets. The random forest is the second most efficient, and the decision tree is the least efficient.

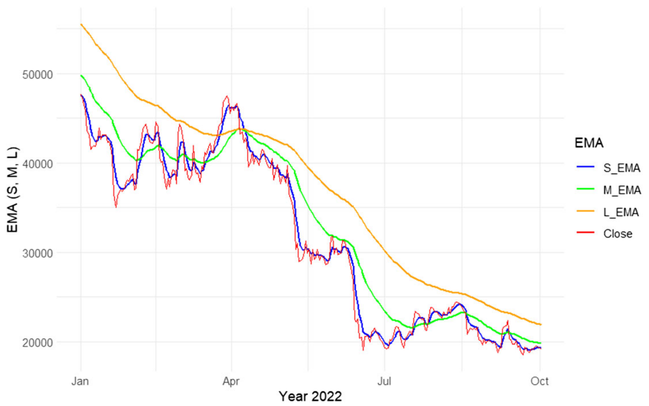
Table 1
The performance of three decision tree models on the three targets

Target	Model	CA	Precision	Recall	F1	AUC
TREND (UP, SIDE, DOWN)	*Decision tree*	0.871	0.871	0.871	0.871	0.896
	Random forest	0.905	0.905	0.905	0.904	0.972
	XGBoost	**0.930**	**0.930**	**0.930**	**0.929**	**0.983**
ACTION (BUY, HOLD, SELL)	*Decision tree*	0.974	0.974	0.974	0.974	0.957
	Random forest	0.975	0.975	0.975	0.975	0.993
	XGBoost	**0.985**	**0.985**	**0.985**	**0.985**	**0.998**
AMOUNT (ALL, HALF, NONE)	*Decision tree*	0.975	0.975	0.975	0.975	0.980
	Random forest	0.985	0.985	0.985	0.985	0.996
	XGBoost	**0.987**	**0.987**	**0.987**	**0.987**	**0.997**

4.4. Testing the model against an unseen data set

The investment situation in 2022 became more complicated when investors had to be more cautious about several factors affecting the investment market. Not only did the crypto market go into a bear market in 2022, but other events might also affect investor confidence. For instance, the war between Russia and Ukraine. Stocks and crypto markets continued to decline, and investors' expected bullish scenarios remained ambiguous. Besides, the situation worsened during the Luna crisis, where prices have

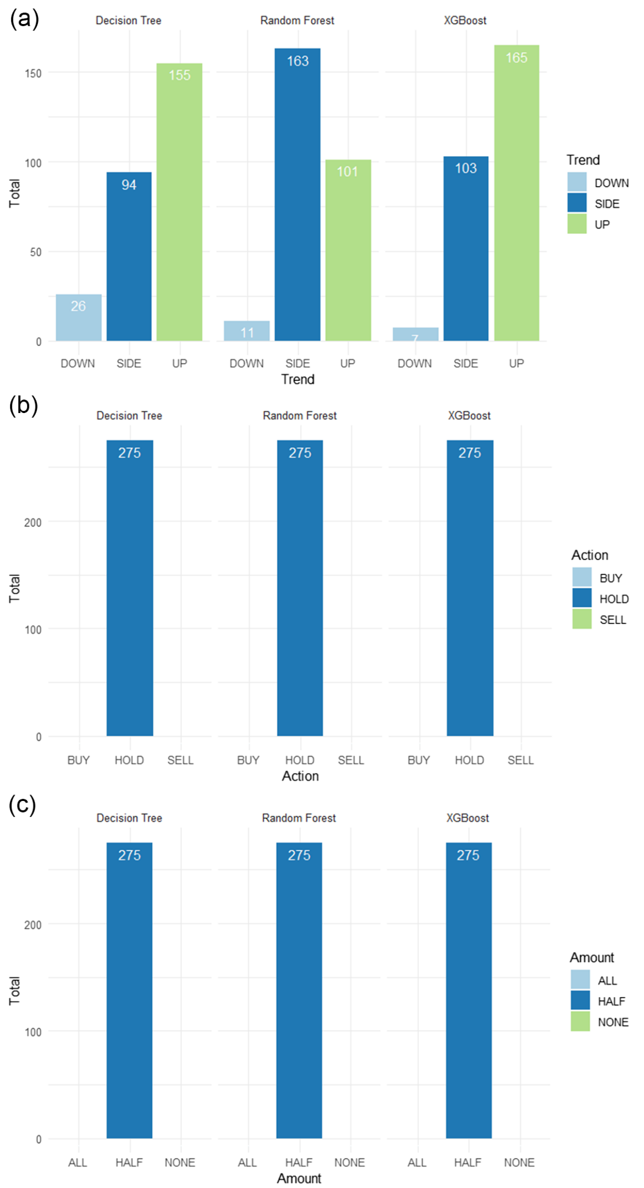
Figure 12
An unseen data set represents a 275-day bitcoin price in 2022



dropped more than 99% since early May, with UST losing its $1 peg, causing investors to panic. The crisis also affected Celsius Network, a prominent crypto lending company, which deposited about $500 million in Anchor, Terra's UST Stablecoin lending platform. In addition, the potential bankruptcy of cryptocurrency hedge fund 'Three Arrows Capital (3AC)' also suffered heavy losses from its investment in Luna Gat (Olatunji, n. d.). Due to such situations, validating the bitcoin investment recommended model using bitcoin trading price data in 2022, it would be appropriate to review the daily bitcoin investment recommendations from the models. For example, is it credible to make investment decisions consistent with bitcoin's market forces? The data comprised 275 days of bitcoin market prices from January 1 to October 2, 2022. After that, it examined how the recommended daily strategy investments consisted of trend, action, and amount. Figure 12 shows the bitcoin market trend in 2022.

Figure 13(a), (b), and (c) shows that the three models provided similar forecasting results in trend, action, and amount. However, they differed in trend, with the decision tree (up = 165 days,

Figure 13
The model forecasting results: (a) trend, (b) action, (c) amount, and (d) strategies



(d)

Description: df [27 x 4]

Strategies <chr>	Decision Tree <int>	Random Forest <int>	XGBoost <int>
SIDE|BUY|HALF	155	101	165
DOWN|BUY|HALF	0	0	0
UP|HOLD|HALF	0	0	0
SIDE|HOLD|HALF	94	163	103
DOWN|HOLD|HALF	0	0	0
UP|SELL|HALF	0	0	0
SIDE|SELL|HALF	26	11	7
DOWN|SELL|HALF	0	0	0
UP|BUY|NONE	0	0	0
SIDE|BUY|NONE	0	0	0

11-20 of 27 rows Previous 1 2 3 Next

sideway = 94 days) and the XGBoost (up = 165 days, sideway = 103 days), while the random forest forecasts 163 days as sideway and 101 days as uptrend. Nevertheless, all three models yielded the same analysis results for action (hold = 275 days) and amount (half = 275 days). Since all three elements under the strategy consist of three types of forecasting, there will be a total of 27 strategies to support bitcoin investment decisions based on the principle of classification ($3 \times 3 \times 3$). The analysis results in Figure 13(d) show that all three models recommend 3 of the 27 strategies suitable for investment in 2022, that is, SIDE|BUY|HALF, SIDE|HOLD|HALF, and SIDE|SELL|HALF. Among these three strategies, SIDE|BUY|HALF is mainly recommended by the XGBoost, the decision tree, and the random forest at 165, 155, and 101, respectively. For SIDE|HOLD|HALF, there are 163 days for the random forest, while there are 103 and 94 days for the XGBoost and the decision tree. Although there are only a few recommendations for SIDE|SELL|HALF from the three models, the decision tree provided the highest recommendations at 26 days, and the random forest and the XGBoost recommend 11 and 7 days.

While the direction of bitcoin moves into a long-term bear market in 2022, it is surprising that the data provided by the forecasts of the three models will go up. The reason may be that the pattern of data on the learned direction of the models from the previous two-year data is significantly different from the data in 2022. Thus, all three models predict that the direction of prices will increase after further decreases continually. However, all three models advise investors not to buy or sell but hold the cash first. If the investment is necessary, all-in investments are prohibited, but only half the investment. Therefore, the three models' predictions toward "action" and "amount" are appropriate for the investment situation in 2022.

5. Conclusion

The research proposed three decision-tree-based machine learning models to forecast bitcoin based on three EMAs of 5, 30, and 90 days with three targets: (1) TREND (uptrend, downtrend, and sideways); (2) ACTION (buy, hold, and sell); and (3) AMOUNT (all, half, and none). The number of true events in the confusion matrix indicated that the XGBoost provided the most accurate forecasting results across all targets and classes, followed by the random forest and the decision tree. The article also shows that MAs can help forecast bitcoin price trends. The intersection of MAs, coupled with the machine learning process of decision trees, can forecast market conditions.

Furthermore, machine learning methods can facilitate investors to make more accurate decisions. In this case, mistakes caused by inexperience or emotional inclinations can diminish by providing stability in asset investment and management. It may be necessary to adjust the MA to match the asset's price behavior if applying this method to analyze other assets. While technical price analysis provides several indices to predict the market direction, other factors in the fundamental analysis might help investors manage their assets, for example, halving, mining rate, acceptance or resistance to foreign commodities, inflation, and interest adjustments on stable assets.

The value of bitcoin might be challenging to predict due to the many variables involved. Besides, cryptocurrency trading is in a negative direction, like any other asset investment market, due to economic and social conditions and the effects of conflicts between superpowers. Therefore, the current market condition is influenced by several uncertain factors. Nevertheless, cryptocurrency is also attractive to investors who believe that these technologies are desirable options for investment and that the trading situation will return to normal when the situation improves in a better direction. The benefits of cryptocurrency and blockchain still have the potential to transform not only global finance but also politics. In addition, while the primary aim of the investigation into the bitcoin prediction model is to optimize profit, inconsistent predictions at different levels may result in varying degrees of loss. To aid in the selection of the most appropriate model developing metrics that effectively measure the potential profit gains across different methods might be required.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in [The Securities and Exchange Commission] at https://www.sec.or.th/TH/Pages/WEEKLYREPORT-2564-12.aspx; in [Yahoo Finance] at https://finance.yahoo.com/

References

Akhtaruzzaman, M., Boubaker, S., & Sensoy, A. (2021). Financial contagion during Covid-19 crisis. *Finance Research Letters*, *38*, 101604. https://doi.org/10.1016/j.frl.2020.101604

Basher, S. A., & Sadorsky, P. (2022). Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility? *Machine Learning with Applications*, *9*, 100355. https://doi.org/10.1016/j.mlwa.2022.100355

Bhatia, P. (2019). *Data mining and data warehousing: Principles and practical techniques*. UK: Cambridge University Press.

Bramer, M. (2016). *Principles of data mining*. USA: Springer.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. *SPSS Inc, 9*(13), 1–73.

Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics, 365*, 112395. https://doi.org/10.1016/j.cam.2019.112395

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., . . . , & Zupan, B. (2013). Orange: Data mining toolbox in Python. *Journal of Machine Learning Research*, *14*(1), 2349–2353.

Doumenis, Y., Izadi, J., Dhamdhere, P., Katsikas, E., & Koufopoulos, D. (2021). A critical analysis of volatility surprise in Bitcoin cryptocurrency and other financial assets. *Risks*, *9*(11), 207. https://doi.org/10.3390/risks9110207

Gazali, H. M. (2019). Bitcoin investment behaviour: A pilot study. *International Journal on Perceptive and Cognitive Computing*, *5*(2), 81–86. https://doi.org/10.31436/ijpcc.v5i2.97

Jaquart, P., Dann, D., & Weinhardt, C. (2021). Short-term Bitcoin market prediction via machine learning. *The Journal of Finance and Data Science*, *7*, 45–66. https://doi.org/10.1016/j.jfds.2021.03.001

Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R., & Fu, F. (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *Frontiers in Physics*, *7*, 98. https://doi.org/10.3389/fphy.2019.00098

Rokach, L., & Maimon, O. Z. (2015). *Data mining with decision trees: Theory and applications*. Singapore: World Scientific Publishing.

Mudassir, M., Bennbaia, S., Unal, D., & Hammoudeh, M. (2020). Time-series forecasting of Bitcoin prices using high-dimensional features: A machine learning approach. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-020-05129-6

Murphy, J. J. (1999). *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. UK: The Penguin Group.

Ogheneovo, E. E., & Nlerum, P. A. (2020). Iterative dichotomizer 3 (ID3) decision tree: A machine learning algorithm for data classification and predictive analysis. *International Journal of Advanced Engineering Research and Science*, *7*(4), 514–521. https://dx.doi.org/10.22161/ijaers.74.60

Okanurak, W., Nakarat, W., & Koohasaneh, S. (2018). *Emergency decree on the business of digital assets*. Retrieved from: https://www.sec.or.th/EN/Documents/EnforcementIntroduction/digitalasset_decree_2561_EN.pdf

Olatunji, M. (n. d.). *Following Terra (LUNA's) crash, Celsius Network and 3AC take a nose-dive*. Retrieved from: https://www.gate.io/blog_detail/1228

Peng, A., Ang, S. L., & Lim, C. Y. (2022). Automated cryptocurrency trading bot implementing DRL. *Pertanika Journal of Science & Technology*, *30*(4), 2683–2705.

Peters, J., & Schutte, J. R. (2022). *Crypto trading for ambitious beginners: A practical guide to profitable trading in Bitcoin and other cryptocurrencies*. Netherlands: Odyssea Publishing.

Rathore, R. K., Mishra, D., Mehra, P. S., Pal, O., Hashim, A. S., Shapii, A., . . . , & Shutaywi, M. (2022). Real-world model for Bitcoin price prediction. *Information Processing & Management*, *59*(4), 102968. https://doi.org/10.1016/j.ipm.2022.102968

Shahvari, N. (2022). The relation between gold price movement and Bitcoin investment sentiment. *International Journal of Management, Accounting and Economics*, *9*(9), 566–582.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 13–22.

Sheppard, C. (2017). *Tree-based machine learning algorithms: Decision trees, random forests, and boosting*. CreateSpace USA: Independent Publishing Platform.

Urquhart, A., & Zhang, H. (2019). Is Bitcoin a hedge or safe haven for currencies? An intraday analysis. *International Review of Financial Analysis*, *63*, 49–57. https://doi.org/10.1016/j.irfa.2019.02.009

Wade, C., & Glynn, K. (2020). *Hands-on gradient boosting with XGBoost and Scikit-learn: Perform accessible machine learning and extreme gradient boosting with Python*. UK: Packt Publishing.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining practical machine learning tools and techniques*. USA: Morgan Kaufmann Publishers.

Youngberg, C. (2021). *Machine learning for beginners book: Decision trees and random forests work: Classification machine learning algorithms*. USA: Kindle Direct Publishing.

Zakamulin, V. (2017). *Market timing with moving averages: The anatomy and performance of trading rules*. USA: Springer International Publishing.

How to Cite: Buathong, W., Sieng-EK, P., & Jarupunphol, P. (2024). Measuring the Performance of Machine Learning Forecasting Models to Support Bitcoin Investment Decisions. *Journal of Data Science and Intelligent Systems*, *2*(2), 100–112. https://doi.org/10.47852/bonviewJDSIS3202677