Abstract: This paper aims to develop an alternative algorithmic model with a strong predictive capability for forecasting the future price of Bitcoin. The research employs random forest regression and LSTM techniques while analysing the influential factors impacting bitcoin prices. While existing literature predominantly focuses on Bitcoin price prediction using the ARMA model and LSTM algorithm, this study explores the performance of random forest regression. While the Diebold-Mariano test does not conclusively establish the superior prediction accuracy of random forest regression over LSTM, the evaluation metrics, namely RMSE and MAPE, indicate that random forest regression outperforms LSTM in terms of prediction errors. Furthermore, by employing random forest regression, we identify the variable dynamics driving Bitcoin prices across different periods. From 2015 to 2018, the variables impacting Bitcoin prices include stock market indexes, oil price, and ETH price. However, since 2018, the critical variables have shifted to ETH price and the Japanese stock market index. Additionally, we analyse the relationship between prediction accuracy and the number of lagged explanatory variables incorporated in the model. The findings suggest that for accurately forecasting the next-day Bitcoin price, the model with a single-lag explanatory variable demonstrates the highest prediction accuracy.

Keywords: Bitcoin; Machine Learning; Random Forest; Linear Regression; Long Term Short Memory (LSTM)

Introduction

Bitcoin, a decentralized digital currency introduced in 2008 by the pseudonymous Satoshi Nakamoto, operates on a peer-to-peer electronic cash system. Its transactions are recorded on the blockchain, a public ledger accessible to anyone, enabling instant global transfers without the involvement of central banks or financial institutions. Over the years, Bitcoin has gained popularity as both a medium of exchange and a store of value. Despite its growing adoption, Bitcoin is known for its high volatility, posing challenges to its role as a reliable store of value and a stable currency. Between April 2015 and April 2022, Bitcoin exhibited a daily return rate standard deviation of 3.85%, which was 2.68 times higher than that of gold and 3.36 times higher than that of the S&P500 during the same period. The significant price fluctuations have raised concerns regarding Bitcoin's reliability for value storage and transactional purposes.

Effectively anticipating Bitcoin's trends to mitigate the risks associated with its volatility has become a complex task. Many researchers have attempted to predict Bitcoin's behavior by exploring its correlation with other commodities. However, previous studies investigating the relationship between Bitcoin and various comparison assets, such as gold, stock market indices, or crude oil prices have found weak correlations.

In past studies, another type of research direction to grasp the price trend of Bitcoin is to predict the price of Bitcoin in the future through AI algorithms and powerful computing power of computers. With the improvement of hardware performance in the 21st century, machine learning technology which has become a hot field of research. Primarily, machine learning has been used across a variety of areas such as that of stock markets, crude oil markets, gold markets and futures markets.

Prediction of Bitcoin by AI is mainly divided into two categories. The first category is the classification research of predicting the rise or fall of Bitcoin in the future. The error standard is DA and F1. The other category is regression research on predicting Bitcoin prices, while the corresponding errors are RMSE and MAPE. Due to the sharp fluctuations in the price of Bitcoin, only grasping the rise or fall of the price of Bitcoin in the future cannot help investors avoid risks. In contrast, getting the specific bitcoin price as a reference price is more useful.

Given the imperative to mitigate the price risk associated with Bitcoin, this research focuses on using machine learning algorithms, specifically random forest regression and LSTM models, to predict Bitcoin prices. The primary objective is to evaluate the performance of random forest regression in Bitcoin price prediction compared to the results obtained from LSTM. Random forest regression, unlike neural networks, provides transparency by revealing the importance of each explanatory variable in predicting Bitcoin prices through its weak learners' outputs.

Random forest has proven effective in predicting stock price direction. However, there is limited literature employing random forest regression in studying the cryptocurrency market. Existing research utilizing random forest regression, focuses on highly correlated OHLC (Open, High, Low, Close) prices and transaction volumes of Bitcoin as explanatory variables. Hence, the inclusion of explanatory variables from other domains holds significant research value. For this study, a total of 47 explanatory variables across eight categories were collected, including Bitcoin price variables, specific technical features of Bitcoin, other cryptocurrencies, commodities, market indices, foreign exchange rates, public attention, and dummy variables for the week. The aim is to verify the accuracy of random forest regression in predicting Bitcoin prices.

To compare the prediction accuracy of random forest regression, this research employs the LSTM algorithm of Recurrent Neural Networks (RNN) as a benchmark. Numerous studies have demonstrated that LSTM and GRU models outperform other models, including traditional time series models such as ARMA, in terms of prediction accuracy.

In addition to striving for a high-precision forecasting model, this study also conducts an in-depth analysis of the explanatory variables determining the importance of Bitcoin prices and explores the relationship between prediction accuracy and the lag of explanatory variables.

Literature Review

In recent years, machine learning has become a prominent approach for Bitcoin price prediction. This literature review outlines key findings and trends in Bitcoin prediction research, highlighting the significance of machine learning techniques. Researchers have primarily employed two types of models in Bitcoin prediction: classification models and regression models. Classification models aim to predict whether Bitcoin's price will rise or fall within a certain time frame, while regression models focus on predicting specific price values.

Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU): These recurrent neural network (RNN) models have gained popularity for their ability to capture temporal dependencies in time series data. Numerous studies have reported superior prediction accuracy using LSTM and GRU models, outperforming traditional time series models like ARMA.

Random Forest Regression: Although less explored in the cryptocurrency prediction domain, random forest regression has shown promise. Unlike neural networks, it offers transparency by revealing the importance of each explanatory variable.

To predict Bitcoin prices, researchers have employed a wide range of explanatory variables, including historical Bitcoin prices, technical indicators, data from other cryptocurrencies, market indices, macroeconomic indicators, and social media sentiment.

Data preprocessing is a crucial step in Bitcoin price prediction. Researchers often divide the data into training and testing sets and may perform additional tasks such as scaling, differencing, and handling missing values to prepare the data for modeling.

Comparative analyses have been conducted to assess the performance of different models and algorithms. Metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) are commonly used to evaluate prediction accuracy.

Researchers have acknowledged several challenges in Bitcoin prediction, including the difficulty of predicting extreme price fluctuations, the impact of market sentiment, and the evolving nature of the cryptocurrency market.

Future research in Bitcoin price prediction could explore the incorporation of additional data sources, such as on-chain transaction data, and the development of ensemble models combining the strengths of different algorithms.

Methodology

Machine learning, a vital component of artificial intelligence (AI), encompasses various branches such as supervised learning, unsupervised learning, and reinforcement learning, depending on the presence of a target variable. In this study, as the objective is to predict future Bitcoin prices, a regression function utilizing supervised learning is employed. The general execution logic of machine learning involves predefining an algorithm, generating a learner, and iteratively training the learner using training data to obtain a high-precision learner through validation. Finally, the trained learner is applied for evaluation and prediction using test data.

The implementation of both random forest regression and LSTM model training in this research is conducted using open-source machine learning libraries in Python. The sklearn library is utilized for random forest regression, while the keras library is employed for LSTM modeling. Data preprocessing and organization tasks are performed using the pandas library.

- *Data Collection:* The data used in this study was collected from various reliable sources to ensure the accuracy and reliability of the analysis. The dataset spans from March 31, 2015, to April 1, 2022, and consists of daily observations of Bitcoin prices and related variables.

  Data Sources

1. Investing.com: Historical data on global commodities, including crude oil prices.
2. Bitinfocharts.com: Bitcoin mining difficulty data.
3. Coinmetrics.io: Metrics related to Bitcoin transactions and network activity.
4. Other Data Sources: Additional data sources were consulted to collect information on cryptocurrency market indices, foreign exchange rates, and public attention metrics.

- Data Preprocessing: Before conducting any analysis, extensive data preprocessing was carried out to ensure the quality and consistency of the dataset. The following steps were taken:

1. Handling Missing Values: Missing data points were identified and either imputed using appropriate methods or removed, depending on the nature of the missing values.
2. Data Formatting: All data types were standardized and formatted consistently to ensure compatibility between variables.
3. Outlier Detection and Treatment: Outliers in the data were detected using statistical methods and addressed through appropriate means, such as winsorization or removal.
4. Feature Engineering: New variables were created, and existing ones were transformed to extract relevant features that could enhance the predictive capabilities of the models.

- *Model Selection :* Two machine learning algorithms were selected for predicting Bitcoin prices: Random Forest Regression and Long Short-Term Memory (LSTM) models.

Random Forest Regression

Random Forest Regression is an ensemble learning method that combines multiple decision trees to make predictions. It was chosen for its ability to provide transparent insights into variable importance.

1. Model Configuration: The random forest regression model was configured with 500 sub-regression trees and a maximum depth of 10 for each tree.

2. Variable Importance: The model's ability to determine the importance of each explanatory variable was leveraged to gain insights into the factors affecting Bitcoin prices.

LSTM Model

The LSTM model is a type of recurrent neural network (RNN) designed to capture temporal dependencies in time series data.

1. Model Architecture: The LSTM model was constructed with multiple layers of LSTM cells, with varying numbers of units per layer. Dropout layers were incorporated to prevent overfitting.
2. Hyperparameter Tuning: Extensive experimentation was conducted to identify the optimal hyperparameters, including the number of LSTM layers, units per layer, and dropout rates.

- *Data Splitting:* To evaluate the performance of the models, the dataset was split into training and testing sets. Approximately 85.70% of the data was used for training, while the remaining 14.30% was reserved for testing. Additionally, a validation set consisting of the last 10% of the training data was employed to monitor model performance during training.

- *Evaluation Metrics:* The following evaluation metrics were used to assess the accuracy and performance of the predictive models:

1. Root Mean Squared Error (RMSE): To measure the average prediction error.

2. Mean Absolute Percentage Error (MAPE): To quantify the accuracy of predictions relative to the actual values.

3. Directional Accuracy (DA): To assess whether the models correctly predicted the direction of Bitcoin price movement (i.e., increase or decrease).

- *Experimental Design :* The study's analysis was divided into two distinct periods: Period 1 (March 31, 2015, to September 30, 2018) and Period 2 (October 1, 2018, to April 1, 2022). Separate models were trained and tested for each period to account for potential changes in Bitcoin price dynamics over time.

Linear Regression:

Linear regression is indeed one of the most fundamental and widely used algorithms for forecasting and prediction. It is used to model the relationship between two or more variables by fitting a linear equation to the data.

In linear regression, the researcher aims to understand or establish a relationship between one dependent variable and one or more independent variables. The dependent variable is the variable of interest that we want to predict or explain, while the independent variables, also known as predictor variables, are used to predict or explain the variation in the dependent variable.

The dependent variable in linear regression must be continuous, meaning it can take any numerical value within a certain range. On the other hand, the independent variables can be either continuous or categorical (discrete) variables. Continuous independent variables are those that can take any numerical value, while categorical independent variables take on specific categories or levels.

To examine the relationship between continuous variables in linear regression, researchers often create scatter plots. Scatter plots visually display the relationship between two continuous variables by plotting their values on a graph, with one variable on the x-axis and the other on the y-axis. The pattern and direction of the plotted points can provide insights into the nature of the relationship, such as whether it is positive (increasing) or negative (decreasing).

Example of scatter plot is given in fig.1 below:

Fig.1: Scatter plot for dependent and independent values

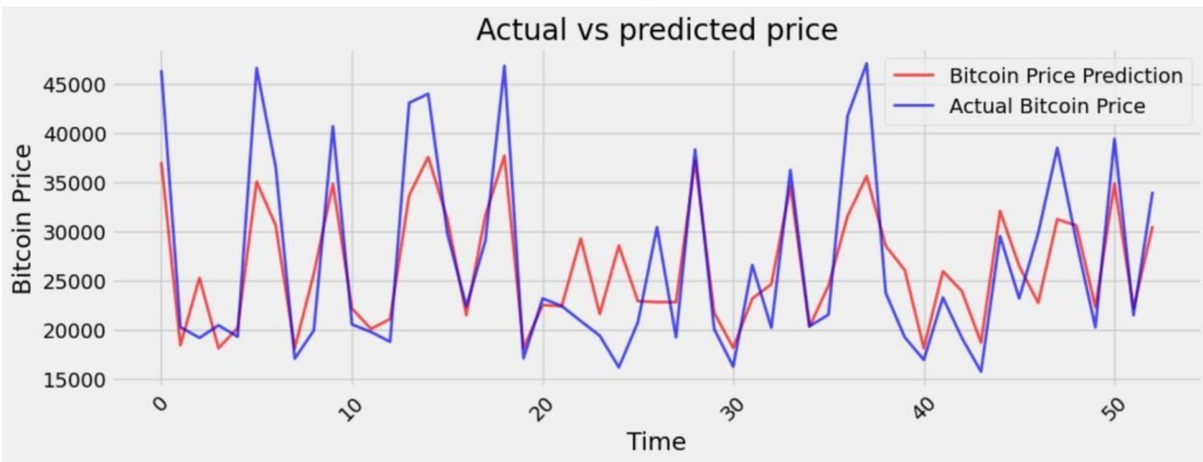


Fig 2: Actual Price vs Predicted Price

Random Forest

Random forest is an ensemble form of multiple regression trees. Its advantage is high explicability, but the predicted results are limited by the training samples. The principle of the regression tree is to divide the parent group into subgroups using an indicator of a certain variable, and the classification is based on making the average of the sum of squared residuals of each group the smallest. Regarding parameter settings, the maximum depth of a single sub-regression tree is 10, and the number of sub-regression trees in the random forest is 500 (Figure 1). I tested the maximum depth of the interval [min = 3, max = 20] and the number of sub-regression trees of the interval [min = 200, max = 1000], respectively. My further experiments show that after the maximum depth is greater than 10 or the number of sub-regression trees is greater than 500, the training data and the prediction error no longer changes.

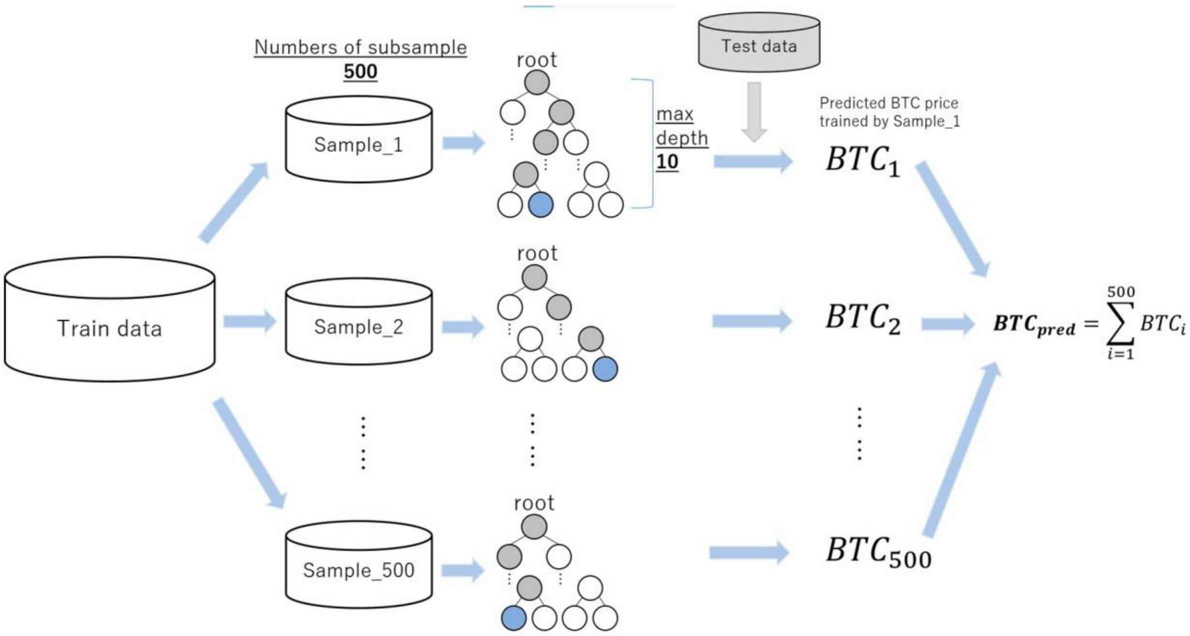


Fig 3 :Parameters and framework of random forest regression.

LSTM:

The LSTM (Long Short-Term Memory) model is a type of recurrent neural network (RNN) that addresses the issue of short memory in traditional RNNs. Unlike regular deep neural network (DNN) algorithms, the LSTM model not only generates an output value when data is inputted, but also modifies the model's parameters. This is achieved through the use of three activation functions: Forget Gate, Input Gate, and Output Gate, which allow the model to retain information from previous inputs.

In this study, the LSTM model is chosen due to its ability to address the short memory limitation of RNNs. By utilizing the output values of one LSTM layer as input to another layer, and incorporating the dropout layer, as mentioned in the literature, the LSTM model structure for this experiment is designed. The parameter settings for the dropout layer are tested within the range of [min = 10%, max = 50%]. It was observed that when the overall dropout value is small, the model tends to overlearn, performing well on training data but resulting in large prediction errors on validation data. Conversely, when the overall dropout value is too large, both training and validation errors become large. Furthermore, the experiment revealed that the prediction accuracy decreases when the dropout values are arranged in descending order compared to ascending order. The number of LSTM layers is tested within the range of [min = 2, max = 6], and the units in each layer are tested with values of [32, 64, 128, 256, 512]. The selection of these values aims to strike a balance between accuracy and the risk of overfitting. The activation function for each layer is set to "ReLU," which has been found to perform better than "sigmoid" and "tanh" functions. The specific values and framework of the LSTM model are presented below. The last 10% of the training data is designated as the validation data.

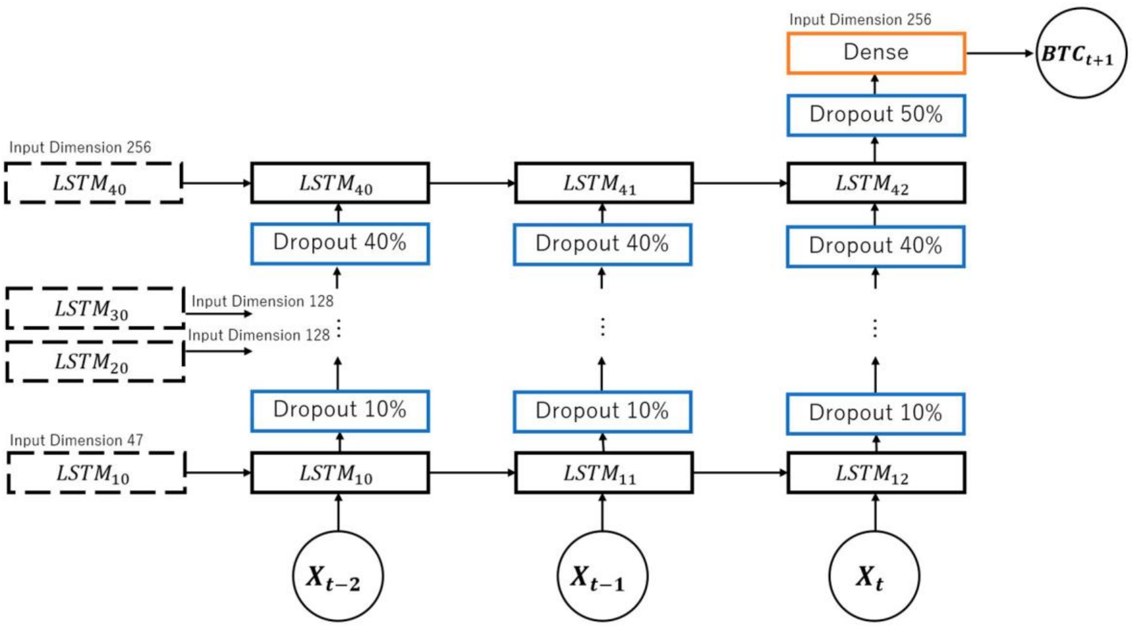
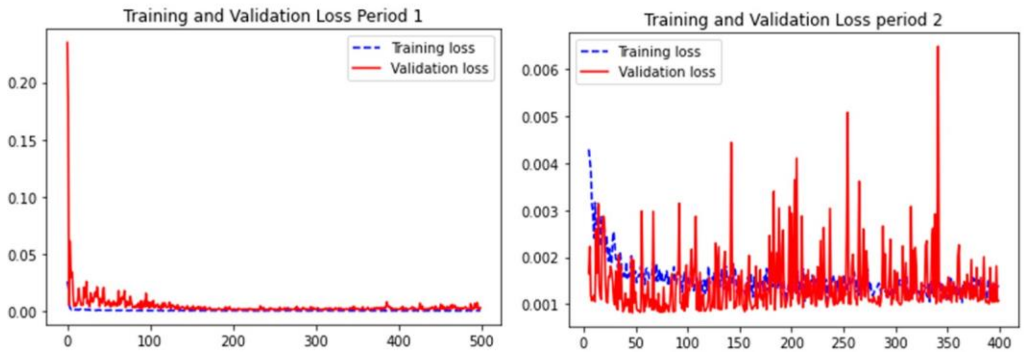


Fig4: Parameters and framework of LSTM

In addition to the model framework, another crucial hyperparameter in deep learning is the number of epochs. The number of epochs represents the number of passes made through the training data during the learning process. Increasing the number of epochs typically leads to a reduction in the prediction error of the training data. However, setting the number of epochs too high can result in overfitting issues.

To determine the appropriate number of epochs, the training and validation loss curves for Period 1 and Period 2 are analyzed. Figure 3 below showcases these curves. Based on the analysis, it is determined that setting the number of epochs to 250 for Period 1 and 75 for Period 2 yields satisfactory results. This balance ensures effective learning from the training data while mitigating the risk of overfitting.



Training and Validation Loss Curves for Period 1 and Period 2(Illustrative)

Data and Preprocessing:

The sample data used in this study span from March 31, 2015, to April 1, 2022, and consist of daily observations. The data for the research were sourced from various platforms, including yahoo finance, Coinmarketcap.com, investing.com, bitinfocharts.com, and coinmetrics.io.

The target variable in this experiment is the price of Bitcoin denominated in USD. A total of 47 explanatory variables are utilized to predict the future price of Bitcoin. These variables are categorized into eight groups:

(a) Bitcoin price variables

(b) Specific technical features of Bitcoin

(c) Other cryptocurrencies

(d) Commodities

(e) Market indices

(f) Foreign exchange rates

(g) Public attention

(h) Dummy variables representing the week

During the preprocessing stage, the collected data is organized and cleaned using the pandas library in Python. This includes handling missing values, data formatting, and ensuring consistency across variables. The processed data is then ready for further analysis and model training.

Preprocessing

The collected data for this research spans a total of 7 natural years, from March 31, 2015, to April 1, 2022. However, considering the specific characteristics of Bitcoin, which experienced price bubbles at the end of 2017 and 2021, and the fact that previous studies have typically used sample periods of no longer than 4 years, the total sample is divided into two distinct periods: Period 1 (from March 31, 2015, to September 30, 2018) and Period 2 (from October 1, 2018, to April 1, 2022). Independent research is conducted on each sub-sample, involving training models specific to each period and making predictions accordingly.

In machine learning, the process involves training initial samples using training data and then evaluating the trained models by substituting them into test data. Typically, training samples constitute 75% to 90% of the total dataset. For this study, the specific division of training and testing samples is presented in Table below . The last 10% of the training data is designated as the validation data, which helps assess the model's performance during the training process.

Table for Division of Training and Testing Samples

	Train Data	Test Data	Percentage of Train Data
Period 1	31 March 2015–31 March 2018	1 April 2018–30 September 2018	85.70%
Period 2	1 October 2018–30 September 2021	1 October 2021–1 April 2022	85.69%

Results

Results of Random Forest Regression

The trained random forest regression model is employed to predict the test samples of Period 1 and Period 2. The results, presented in Table 1 below, depict the comparison between the actual Bitcoin price (red line) and the price predicted by the random forest regression model (green dashed line).

	Period 1	*Period 2*
RMSE	*321.61*	*2096.24*
MAPE	*3.39%*	*3.29%*
DA	*51.93%*	*52.49*

Table 1: Error results for random forest regression

While the RMSE of Period 1 is considerably smaller than that of Period 2, it is important to note that the average price of Bitcoin in Period 1 is significantly lower than that in Period 2. Therefore, direct comparison of the RMSE values across different periods may not be meaningful. The MAPE and DA indicators in both periods exhibit close values, with slightly better prediction accuracy observed in Period 2. It is worth mentioning that during the early stage of the test interval in Period 2, the random forest regression algorithm exhibits poor prediction performance when the Bitcoin price exceeds $60,000. This can be attributed to the limited presence of training samples with Bitcoin prices surpassing $60,000 in the data of Period 2. This outcome accurately reflects the limitation of the random forest algorithm in predicting values outside the training sample range. However, regardless of Period 1 or Period 2, the random forest regression algorithm demonstrates excellent performance in predicting prices below $60,000, with the predicted price trend aligning closely with the actual price trend.

In addition to predictive analysis, the random forest algorithm also provides insights into the importance of each explanatory variable when predicting Bitcoin prices. This information is derived from the statistics of the occurrence frequency of boundary variables across all 500 sub-regression trees. The variable importance results are illustrated.

The significant increase in RMSE in Period 1 occurs when BTC_Open is removed, indicating the importance of including the previous day's OHLC (Open, High, Low, Close) Bitcoin prices when predicting the next Bitcoin price. The removal of BTC_Open corresponds to excluding all OHLC data from the previous day, suggesting that at least one of the current OHLC Bitcoin prices is necessary for accurate predictions.

In Period 2, three substantial increases in RMSE are observed when BTC_Open, ETH, and DOGE are removed individually. Although the frequency of DOGE appearing in the nodes of all sub-regression trees is lower than that of JP225 and S&P500, the sharp rise in RMSE after removing DOGE highlights its impact on prediction accuracy. These three significant changes in Period 2 are all related to cryptocurrency price variables, indicating an increased correlation between Bitcoin price and the cryptocurrency market after 2018.

Considering the importance of the variables in predicting Bitcoin prices, a comparison is made between the model using all variables and the model using only the important variables (BTC_Close, BTC_High, NASDAQ, and BTC_Low for Period 1; BTC_Close, BTC_High, BTC_Low, BTC_Open, ETH, and JP225 for Period 2). The results indicate that the model with all explanatory variables exhibits better prediction accuracy, with an RMSE that is 3% smaller compared to the model using only the important variables.

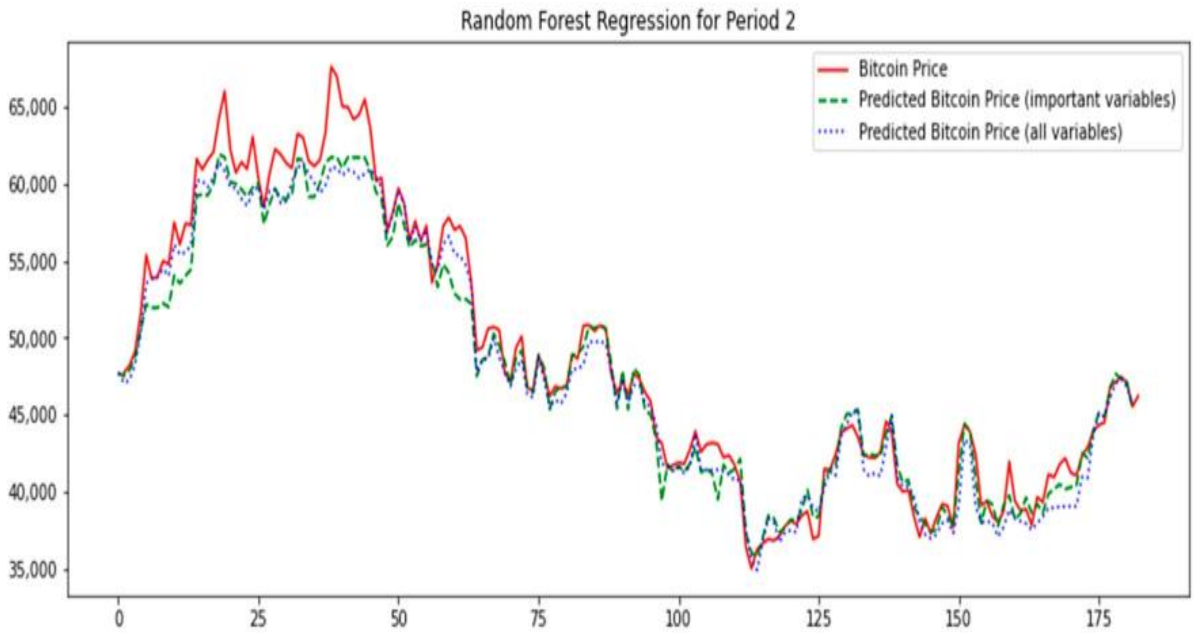


Fig: Random Forest Regression for model 2

Results of LSTM

In the LSTM model, I observed that including redundant explanatory variables during training leads to a decrease in model accuracy. The accuracy of the model, when trained with all 47 explanatory variables, is lower compared to the model that utilizes only a subset of variables, such as the lightweight model using only four Bitcoin price variables. Conversely, using too few explanatory variables also results in reduced prediction accuracy. For instance, when additional variables are added to the lightweight model with only the four Bitcoin price OHLC variables, the prediction accuracy improves. Therefore, extensive experimentation and exploration were conducted to determine the optimal set of explanatory variables for each period. It is worth noting that this issue does not arise in random forest regression, as it is an ensemble algorithm, and thus, the discussion is specific to LSTM.

Since the combination of explanatory variables directly impacts the prediction accuracy of the LSTM model, I referred to the importance rankings obtained from random forest regression to determine the respective sets of explanatory variables for Period 1 and Period 2. The variable sets for each period are specified.

As the learning outcomes of deep learning models are influenced by the randomly selected training samples from the dataset, there is inherent randomness in the experimental results. Therefore, when comparing model results, it is more meaningful to assess the average performance of multiple experiments rather than focusing on the accuracy of a single model. This approach aligns with the methodology employed by Liu et al. (2021) in their experiments.

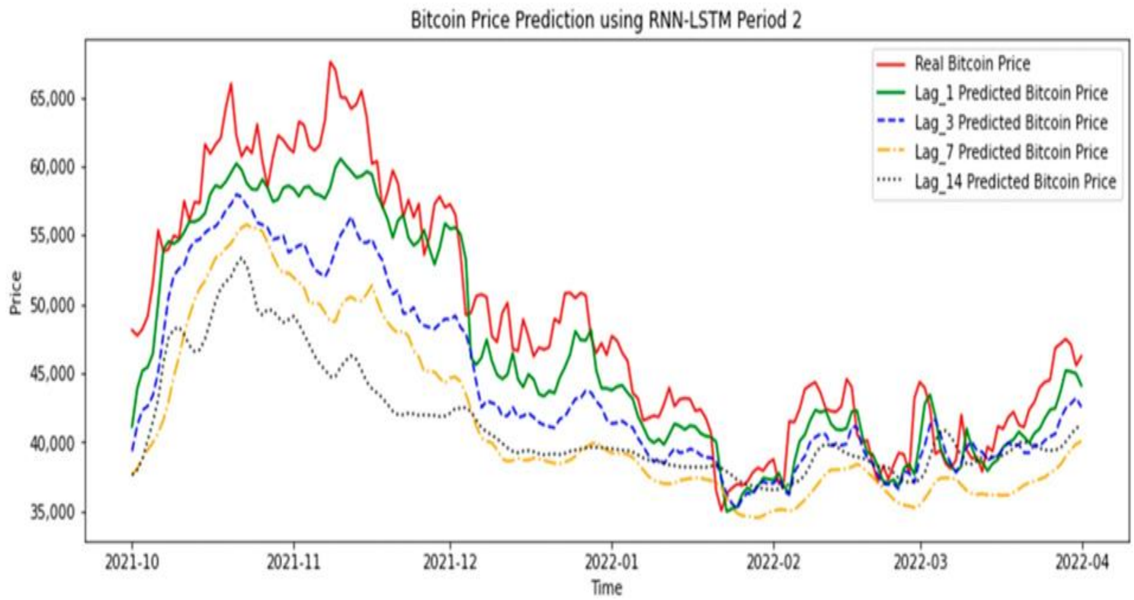


Fig.: <u>Bitcoin Price Prediction using RNN-LSTM Period 2</u>

Contributions

This research makes several contributions to the field of Bitcoin price prediction:

Prediction Accuracy: The random forest regression model demonstrates higher prediction accuracy, as indicated by smaller RMSE values compared to the LSTM algorithm. Although statistical tests such as the Diebold-Mariano (DM) test and Clark-West test do not reject the hypothesis that LSTM is superior to random forest regression at a significance level of α = 95%, multiple experimental results consistently show the superior prediction accuracy of random forest regression.

Identification of Variable Changes: The experimental results of the random forest regression model shed light on the factors influencing Bitcoin prices, particularly in the vicinity of 2018. Throughout the entire sample period, the OHLC prices of Bitcoin itself are found to be the most significant. However, in the Period 1 sample (April 2015 to October 2018), the importance of the U.S. stock markets (NASDAQ, DJI, and S&P500), which initially exhibit high importance, sharply decreases in the Period 2 sample (October 2018 to April 2022). Conversely, the importance of digital currency markets, such as ETH and DOGE, increases during Period 2.

Time Series Analysis: As an LSTM model designed for time series data analysis, control experiments involving the substitution of explanatory variables with different lags demonstrate that the highest prediction accuracy is achieved when using only the latest period of data. Random forest regression also corroborates this conclusion.

Overall, this research contributes by showcasing the superior prediction accuracy of random forest regression over LSTM, providing insights into the changing importance of variables influencing Bitcoin prices, and highlighting the significance of the latest data in achieving accurate predictions.

Conclusions

In this paper, we utilized eight categories of explanatory variables, comprising a total of 47 variables, to predict the price of Bitcoin on the next day. These variables encompassed various aspects such as Bitcoin price, technical features, other cryptocurrencies, commodities, market indices, foreign exchange rates, public attention, and weekly dummy variables. Our findings indicate that random forest regression outperforms LSTM in terms of price prediction accuracy. While LSTM has been widely recognized for its high accuracy in Bitcoin price prediction, our study demonstrates that random forest regression, an underexplored machine learning algorithm in previous literature, yields superior results.

However, it is important to note that random forest regression has limitations in predicting results that do not appear in the training samples. For instance, when the Bitcoin price reached new record highs, random forest regression could not provide higher price predictions than the previous historical highs. Nonetheless, as the transaction history of Bitcoin continues to expand, we anticipate that random forest regression will exhibit improved performance when the Bitcoin price stabilizes.

In a comparative analysis with other studies employing daily data for Bitcoin price prediction, the RMSE error of random forest regression in our experiment (0.017 in Period 1 and 0.035 in Period 2) is superior to the results of LSTM (0.045) and GRU (0.051) in Awoke et al.'s (2021) study but falls short of the SDAE model (0.009) in Liu et al.'s (2021) experiment. Comparing prediction accuracy across different Bitcoin price prediction experiments is challenging. Firstly, the presence of Bitcoin price bubble periods and the inclusion of test data within these periods significantly impact the results. For instance, the RMSE error of random forest regression in Period 2 of our study is twice that of Period 1. Secondly, the comparison of test errors is complicated by the use of different time units in different studies. Interestingly, in Awoke et al.'s (2021) experiment, the models with the highest accuracy incorporated a lag of seven periods, which differs from our conclusion that the optimal model only requires the latest explanatory variables.

The results of random forest regression also shed light on the explanatory variables that determine Bitcoin prices in different periods. During the first price bubble interval (April 2015 to October 2018), factors such as the previous period's Bitcoin price, US stock market indices (NASDAQ, DJI, and S&P500), oil prices, ETH price, and Bitcoin mining difficulty play significant roles in predicting the next-day price. In the second price bubble period (October 2018 to April 2022), the OHLC prices of Bitcoin from the previous day, ETH price, and Japan's JP225 index assume substantial importance. It is worth noting that random forest regression faces challenges in predicting Bitcoin prices exceeding $60,000 per coin, as it struggles to anticipate values that are not present in the training samples. However, the prediction accuracy remains strong for price ranges below $60,000.

Furthermore, our research findings indicate that increasing the number of past periods for the substituted explanatory variables leads to a decrease in prediction accuracy for both random forest regression and LSTM. The model with the highest accuracy is the one that solely incorporates explanatory variables from the most recent period. This conclusion aligns closely with the classic efficient market hypothesis.

In conclusion, this study demonstrates the effectiveness of random forest regression in predicting Bitcoin prices, outperforming the widely used LSTM algorithm. The analysis of explanatory variables highlights their importance in different periods, while considerations of model accuracy and the number of past periods provide valuable insights into the prediction process. Future research can explore incorporating additional variables and employing alternative machine learning techniques to further enhance Bitcoin price prediction accuracy.

References

- Aggarwal, Apoorva, Isha Gupta, Novesh Garg, and Anurag Goel. 2019. Deep Learning Approach to Determine the Impact of Socio Economic Factors on Bitcoin Price Prediction. Paper presented at 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, August 8–10.

- Akyildirim, Erdinc, Oguzhan Cepni, Shaen Corbet, and Gazi Salah Uddin. 2021. Forecasting mid-price movement of Bitcoin futures using machine learning. Annals of Operations Research 1–32. [CrossRef] [PubMed]

- Awoke, Temesgen, Minakhi Rout, Lipika Mohanty, and Suresh Chandra Satapathy. 2021. Bitcoin Price Prediction and Analysis Using Deep Learning Models. In Communication Software and Networks. Singapore: Springer, pp. 631–40.

- Basak, Suryoday, Saibal Kar, Snehanshu Saha, Luckyson Khaidem, and Sudeepa Roy Dey. 2019. Predicting the direction of stock market prices using tree-based classifiers. The North American Journal of Economics and Finance 47: 552–67. [CrossRef]

- Baur, Dirk G., and Lai Hoang. 2021. The Bitcoin gold correlation puzzle. Journal of Behavioral and Experimental Finance 32: 100561.[CrossRef]

- Baur, Dirk G., and Thomas Dimpfl. 2021. The volatility of Bitcoin and its role as a medium of exchange and a store of value. Empirical Economics 61: 2663–83. [CrossRef] [PubMed]

- Blake, R. 2019. An Econometric Analysis of the Relationship between Bitcoin & Gold. Available online: https://medium.com/@blake_ richardson/an-econometric-analysis-of-the-relationship-between-bitcoin-gold-2018-584b4c63a17 (accessed on 21 September 2022).

- Alahmari, S. A. (2019). Using machine learning ARIMA to predict the price of cyptocurrencies. *ISeCure—The ISC International Journal of Information Security*, 11(3), 139– 144.

- Akyildirim, E., Goncuy, A., & Sensoy, A. (2018). *Prediction of Cryptocurrency Returns Using Machine Learning.*

- V. Derbentsev, V. Babenko, K.I.R.I.L.L. Khrustalev, H. Obruch and S.O.F.I.I.A. Khrustalova, "Comparative performance of machine learning ensemble algorithms for forecasting cryptocurrency prices", *Int. J. Integr. Eng.*, vol. 34, no. 1, pp. 140-148, 2021.

- Sun Woong Kim, Prediction Performance of Altcoin Price using Machine Learning Algorithm, Journal of Digital Contents Society, 10.9728/dcs.2023.24.1.141, **24**, 1, (141-151), (2023)

- Garcia, D. & Schweitzer, F. (2015). Social signals and algorithmic trading of Bitcoin. *Royal Society Open Science* 2(9), 150288. Retrieved from https://dx.doi.org/10.1098/rsos.150288 10.1098/rsos.150288

- S. Abboushi, "Global Virtual Currency – Brief Overview.19(6), 10-118.," The Journal of Applied Business and Economics, vol 19, no. 6, 10 2017.

- Amjad, M., & Shah, D. (2017). Trading bitcoin and online time series prediction. *Proceedings of Machine Learning Research*, **55**, 1– 15.

- Anupriya, & Garg, S. (2018). Autoregressive integrated moving average model based prediction of bitcoin close price. *International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 473– 478.