IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.17, No.1, January 2023, pp. 33~44
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.80776

Improved LSTM Method for Predicting Cryptocurrency Price Using Short-Term Data

Risna Sari[1], Kusrini Kusrini*[2], Tonny Hidayat[3], Theofanis Orphanoudakis[4]
[1,2,3] Dept. of Information Technology, Universitas Amikom Yogyakarta, Indonesia
[4]School of Science and Technology, Hellenic Open University, Patras, Greece
e-mail: [1]risnasari@students.amikom.ac.id, ***[2]kusrini@amikom.ac.id**, [3]tonny@amikom.ac.id , [4]fanis@eap.gr

Abstrak

Seiring dengan berkembangnya cryptocurrency, tidak dipungkiri bahwa harga dari crypto tidaklah stabil. Salah satu faktor yang mempengaruhinya yaitu meningkatnya volume transaksi yang menarik minat peneliti untuk melakukan penelitian dalam mengembangkan metode prediksi harga koin dari cryptocurrency. Hasil prediksi dipengaruhi oleh penggunaan metode, algoritma hingga jumlah data. Pada penelitian ini akan dilakukan pemodelan prediksi dengan menggunakan metode LSTM dan data jangka pendek. Penelitian ini akan melakukan 2 percoba menggunakan metode LSTM sederhana dan memanfaatkan multivariate time series dengan LSTM. Hasilnya diperoleh nilai terkecil prediksi menggunakan skenario pembangian alokasi data 80/20, inputan layer LSTM = 360, Epoch = 500 yang dilakukan yaitu koin Solana dengan RMSE = 0.111, R2 = 0.9962. Dapat disimpulkan bahwa penggunaan data jangka pendek dapat digunakan dalam pembuatan pemodelan prediktif, namun perhatian khusus perlu diberikan pada karakteristik dataset yang digunakan dan metodologi pemodelan, dan diharapkan hasil penelitian ini dapat dapat dimanfaatkan dalam penelitian selanjutnya

Kata kunci*—Cryptocurrency, Prediksi, Long short-term memory (LSTM), Short-term data*

Abstract

As cryptocurrencies develop, it cannot be denied that crypto prices are volatile. One of the influencing factors is the increasing volume of transactions which attracts the interest of researchers to conduct research in developing coin price predictions from cryptocurrencies. The method, algorithm and amount of data affect the prediction results. In this study, prediction modelling will be carried out using the LSTM method and short-term data. This study will conduct two experiments using the simple LSTM method and utilising multivariate time series with LSTM. The smallest predicted value is obtained using an 80/20 data allocation distribution scenario, input layer LSTM = 360, Epoch = 500, a Solana coin with RMSE = 0.111, R2 = 0.9962. It can be interpreted that short-term data can be used in making predictive models. Still, special attention needs to be paid to the characteristics of the dataset used and the modelling methodology, and it is hoped that the results of this study can be used in further research.

Keywords— *Cryptocurrency, Prediction, Long Short-term Memory (LSTM), Short-term data*

1. INTRODUCTION

The emergence of the Industry 4.0 era in the technological realm has affected all aspects of human life. One can feel the presence of a revolution impacting a consumer society with social, cultural, and economic changes. IoT-based systems, Robotics, and Cloud systems contribute to the realisation of the Industry 4.0 revolution and a new digital economy following this revolution.

The digital economy that is felt in everyday life is characterised by transaction processes that are carried out not only traditionally but also digitally processes, also known as digital payments. Traditionally implemented financial transaction systems rely on third parties,

Received December 30th,2022; Revised January 19th, 2023; Accepted January 30th, 2023

such as banks, to process transactions in a particular format. The third party acts as an intermediary for exchanging money, which makes this scenario work well for financial transactions. However, it is also possible that customers do not trust this non-transparent and flexible system [1].

People trust these third parties because of their accountability and predictability status [2]. However, its presence causes people to lose control over ownership of their data due to the monopoly system they use. Based on the trust of more than 6 billion people, the annual transaction value of money reaches 200 trillion [3]. Following this fact, a question arises as to how people can trust a financial monopoly system implemented without guaranteeing security and access of third parties to personal data and transaction processes that are not transparent to the unfortunate event of a failed transaction. In 2008, Satoshi Nakamoto responded to the unrest by bringing significant changes. The invention of the bitcoin, which utilises the blockchain distributed ledger technology, became the forerunner of the first coins used as currency in cryptocurrency systems [4] by utilising a peer-to-peer (P2P) transfer system that allows users to transfer digital money through public networks without a third party or intermediary [5]. Cryptocurrency is a virtual currency that is used in the financial system. The security of this virtual currency is ensured by using cryptography, which, combined with the distributed ledger technology, makes it impossible to counterfeit or double spend.

Cryptocurrencies (also shortly called crypto) are founded on the principle of decentralised control compared to standard currencies, which depend on a banking system [6]. However, crypto itself is generally unstable due to the volume or number of transactions and the occurrence of significant changes [7] due to one of them, namely the issue or public trend that occurs. This is the reason why an increasing number of researchers are interested in conducting research related to the instability of cryptocurrency, including predicting the price of coins in cryptocurrency by utilising machine and deep learning algorithms, including RNN, LSTM [8], ARIMA, GRU [9], LR and SVM [10], [11] and Decision Tree based Regression approach [12]. Apart from using algorithms, some researchers also use historical data, news, public announcements, and the latest trends that affect the rise and fall of coin prices [13] [14].

The amount of data affects the results of a prediction. Soejoeti said that a time series analysis requires a minimum of 50 time series data [15]. Previous research used one-month historical data using the autoregressive vector method and a crypto coin, namely Ethereum. The metric used for evaluation by researchers is RMSE, and the results of their research with an opening price of RMSE 890.29, the highest price is 930.50, the lowest price is 1164.12, and a closing price is 978.37 [16].

The authors in [17] used ten months of historical data using the random forest method and a crypto coin, namely Bitcoin. Evaluation of the method used by the researchers, namely MAPE with the scenario of dividing the data carried out, resulted in MAPE values of 9.14% for a 70:30 split, 80.16% for a 50:50 split, and 63.05%for a 30:70 split. Other researchers use the LSTM and ARIMA methods to predict the price of crypto coins, namely EOS, BTC, ETH, and DOGE. Vulnerable data were collected for the period November 9, 2017 – June 30, 2022. The error evaluation used by researchers is RMSE with the predicted results of each coin, namely for EOS 0.119, for BTC 1334,755, for ETH 117,655, and for the DOGE case 0.007 for LSTM and EOS 0.436, BTC 1718.339, ETH 136.605 and DOGE 0.025 for ARIMA [18].

Based on previous research, it is not only the amount of data that affects the prediction results but also the scenario used and also the crypto coins that will be predicted. The main contribution of the work that will be presented in this study is how the sharing of historical short-term data can increase the use of methods to predict specific cryptocurrency prices.

2. METHODS

Research methods should incorporate a design model or workflow that is used to carry out the research. In this case, this research workflow followed the steps presented in the research [18]. Figure 1 presents the flow of this research.

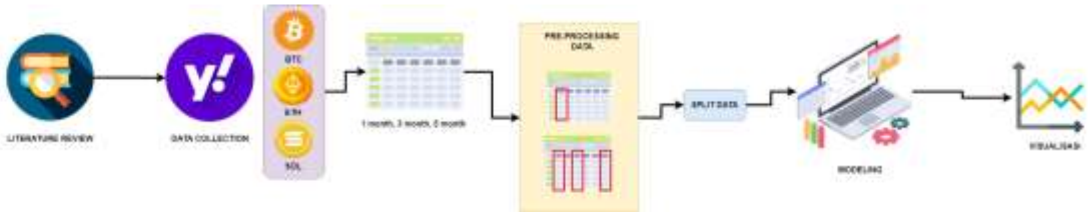

Figure 1 Workflow

2.1 Literature Review

The first step is the review and analysis of literature, including related work on similar subjects. Reviews include a survey of scientific articles, books, or relevant sources to our field of research [19] and results in a summary of the main findings.

2.2 Data Collection

The next step is to collect cryptocurrency data from global portals that provide analysis, financial market news, and historical crypto data (detailed intermediate steps are analysed in Figure 2.

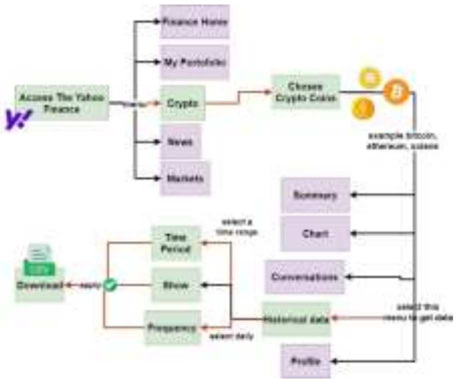

Figure 2 Data collection

The process of getting data starts with access to finance.yahoo.com and selecting the crypto menu. Then we select the desired coin and select historical data. In the time period menu, the interested researcher can select the desired time frame and change the daily frequency. Relevant data can then be exported in CSV format.

The datasets used in this research are Bitcoin (BTC), Ethereum (ETH), and Solana (SOL), as shown in Table 1, where the obtained data are separated by scenario. The data obtained in CSV format for the three selected periods and for all three coins resulted in 9 CSV datasets in total.

Table 1 Types Of Cryptocurrencies

Coin	Month	Range
Bitcoins, Ethereum, Solana	1-month	1 - 30 June 2021
	3-months	June 1, 2021 - August 31, 2021
	5-months	June 1, 2021 – October 2021

For this research, we used Bitcoin and Ethereum crypto coins as a reference that could be compared against results published in the literature over the last five years, also focusing on these most popular coins to predict crypto prices. To also allow a comparison to one of the newest coins gaining popularity and entering into the top 20 coins, the Solana crypto is also included in our study. Figure 3, Figure 4, and Figure 5 present samples of the datasets used for modelling the prediction of cryptocurrency prices.

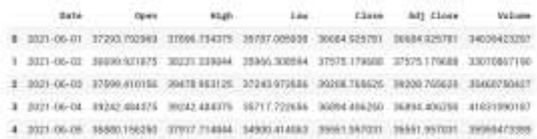

Figure 3 Bitcoin Data Samples

	Date	Open	High	Low	Close	Adj Close	Volume
0	2021-06-01	2707.580547	2738.737540	2531.180848	2633.518311	2633.518311	27363223080
1	2021-06-02	2634.456058	2801.702334	2555.401367	2706.129000	2706.129000	27722267368
2	2021-06-03	2708.376331	2881.254883	2687.684326	2855.126465	2855.126465	30038207402
3	2021-06-04	2857.165527	2857.165527	2562.637451	2688.195068	2688.195068	34173841611
4	2021-06-05	2691.619385	2817.484863	2558.233643	2630.576904	2630.576904	30496672724

Figure 4 Ethereum Data Samples

	Date	Open	High	Low	Close	Adj Close	Volume
0	2021-06-01	32.770885	33.289917	29.880173	30.985260	30.985260	432055117
1	2021-06-02	30.935238	35.038155	30.432858	33.956318	33.956318	497567418
2	2021-06-03	34.018715	41.145370	33.792233	39.470905	39.470905	960056100
3	2021-06-04	39.581303	39.609863	34.310902	37.415070	37.415070	1056279506
4	2021-06-05	37.417277	42.822674	37.099110	39.585400	39.585400	1112425664

Figure 5 Solana Data Samples

As can be seen, the sample data consist of 7 variables, namely Date, Open, High, Low, Close, Adj Close, and Volume. This research later attempts to predict crypto prices based on the Open variable, which represents the price earlier on that date [17]. Then we try to combine variables Open, Close, and Volume as a scenario to predict whether the price opening, closing, and volume transaction will later produce remarkably different predictions.

2.3 Pre-processing

Pre-processing of data follows the data collection step utilising the Python programming language and Google collaboratory tools. Pre-processing is the stage in data mining where the data will be processed by eliminating data that lie outside selected ranges or data changed to a more understandable format [20]. Pre-processing is carried out, namely loading data, null checking whether there is a data column that is null or empty, then dropping data if there are null data and data normalisation. Data normalisation is done for appropriately scaling the range of data values since significant variances may not allow them to be used directly [1].

In this study, pre-processing was carried out by applying data normalisation as expressed in equation 1), which represents a min-max normalisation formula used to appropriately scale data in the range 0-1.

$$x_{normalized} = \frac{x_{original} - x_{min}}{x_{max} - x_{min}} \quad (1)$$

$X_{original}$ is the data to be normalised, and $X_{normalized}$ is data that has been normalised. X_{min} is the minimum value of all data, and X_{max} denotes the maximum value out of the whole dataset. In addition, the Min-max scaler is enabled to minimise errors that occur during data modelling [21].

2.4 Data Splitting

Next, data splitting into separate training and testing data sets are performed for using them later to build and evaluate the machine learning model. The amount of training data usually have a larger percentage in comparison to the testing data [22].

Table 2 Allocation Data

Training	Testing
60	40
70	30
75	25
80	20

The scenarios in this study aim to develop optimal data allocation for short-term data to predict the price of each cryptocurrency. Based on past results from our literature review, we decided to use the data allocation shown in Table 2. Later, training testing data will also be selected (developing python scripts) randomly.

2.5 Modeling

After splitting the data, Machine Learning modelling is carried out for our prediction using the Long Short Term Memory (LSTM) method.

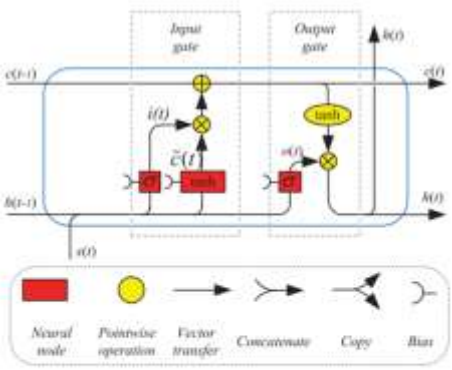

Figure 6 Long-Sort-Term Memory model architecture

LSTM is a variant of the containing RNN cell memory in which it is frequently used. LSTM itself was developed in the late 90s by Sepp Hochreiter and Jurgen Schmidhuber [23]. LSTM is known to have qualified capabilities in building predictive models. LSTM is a derivative of Recurrent Neural Network (RNN), a method designed to process data sequences [24]. Figure 6 shows the original architecture of the LSTM, which consists of an input gate (i_t) and an output gate (o_t). In the LSTM computing process, the calculation is carried out using the following formula;

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \tag{2}$$

$$\tilde{c}_t = tanh(W_{ch}h_{t-1} + W_{cx}x_t + b_c) \tag{3}$$

$$c_t = c_{t-1} + i_t . \tilde{c}_t \tag{4}$$

$$o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \tag{5}$$

$$h_t = o_t . tanh(c_t) \tag{6}$$

Where c_t on equation (3) denotes the cell state of LSTM. W_i, W_c, and W_o in equentions (2), (3), and (5) are the weights, and the operator '.' Donates the pointwise multiplication of two vectors. H_t is the output vector, b is the bias vector, σ is the sigmoid activation function. When updating the cell state, the input gate can decide what new information can be stored in the cell state, and the output gate decides what information can be output based on the cell state [25].

The main target of our method was to result in the smallest possible error value of predictions made and to increase the accuracy of the method used in prediction. For error evaluation, this study uses the Root Mean Square Error (RMSE). RMSE is the square root of variant residue that shows how close the observed data points are with respect to the evaluated values produced by the prediction method [26]. RMSE is often used for regression problems. Y_i in equation (7) denotes the value prediction, y denotes the actual value, and n is the amount of data.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}i - yi)^2}{n}} \tag{7}$$

In addition to RMSE, a test of the coefficient of determination was carried out to measure the extent to which the contribution of the independent variables in the regression model can explain the variation of the dependent variable, which can be assessed by R – Squared (R^2). The higher the value of R^2 close to number one, the better the prediction model of the proposed research model. The value of the coefficient of determination is between 0 and 1 [27]. $\hat{y}_i$ from equation (8) is the value prediction or point on the regression line, $\tilde{y}_i$ represents the average of all values, and y_i represents the actual value.

$$R^2 = \frac{SSR}{SST} = \frac{\Sigma(\hat{y}i - yi)^2}{\Sigma(yi - \bar{y})^2} \tag{8}$$

There are three categories of R^2 values that are strong, moderate and weak. 0.99 – 0.75 is included in the strong category. 0.74 – 0.50 is included in the moderate. 0.49 – 0.25 is the weak category. If R^2 = 0.982, it can be explained that around 98.2% of the variation in the value of variable Y can be explained by variations in the value of variable X [28].

2.6 Visualisation

The final stage is the data visualisation process in the form of graphs based on the results obtained from the cryptocurrency price prediction process. Visualisation is implemented using the plot function of the python programming language.

3. RESULTS AND DISCUSSION

This research uses as a baseline results from 7 previous studies published within the last five years and presents results with respect to cryptocurrency price prediction using machine learning and deep learning methods, as shown in Table 3 below.

Table 3 Literature Review

Author	Summary	Technique	Cryptocurrencies	Datasets
Sean McNally et al. (2018) [8]	Ascertain how accurately the direction of the bitcoin price in USD can be predicted with the price source of the bitcoin index	RNN, LSTM, ARIMA	Bitcoins	CoinDesk August 19 - 2013, to July 19, 2016
Muhammad Rizwan et al. (2019) [9]	Develop a model that predicts bitcoin prices using deep learning algorithms	RNN, LSTM, ARIMA, GRU	Bitcoins	August 19 - 2013, to July 19 2016
Poongodi M et al. (2020) [10]	Investigate how the price of the Ethereum cryptocurrency is predicted and what kinds of trends are found over time for this currency.	LR, SVM	Ethereum	-
Siti Saadah et al. (2021) [17]	Build a bitcoin price prediction system using a random forest algorithm.	Random Forest	Bitcoins	investing.Com January 1, 2019, to October 13, 2019
Pradana Ananda Raharja (2021) [16]	Perform analysis using the vector autoregressive approach for Ethereum coin currency time series data.	Vector Autoregressive	Ethereum	Yahoo Finance May 2021
Zeinab Shahbazi et al. (2021) [28]	Achieve better performance for cryptocurrency predictions with a small error rate. Researchers use a Reinforcement Learning approach integrated into the framework for predictions of Litecoin and Monero crypto coins.	Reinforcement Learning	Bitcoins, Monero	2016-2020
Deny Haryadi et al. (2022) [19]	Applying the SVR method to predict the closing price of a polka dot crypto coin.	SVR	Polka dot	Yahoo Finance August 20, 2020, to December 31, 2021

Machine learning itself is an underlying process of artificial intelligence that attempts to imitate human intelligence to make possible the prediction of future events based on past data. The authors in [29] showed that to find the effects of cryptocurrency on well-formed portfolios using the Modern Portfolio Theory approach, we can create an investment portfolio. The results show that inclusion in the portfolio increases its effectiveness by reducing the standard deviation and providing investors with an allocation of options. The authors in [30] use traditional SVM and linear regression methods to predict bitcoin value taking into account daily time series. The variable used is the closing price of bitcoin to create its predictive model.

It is worth noting that from the analysis of the previous research, we can deduce that it's not just the amount the variables they use in their prediction models but also the variations in the time series that researchers use differently. We can observe 1-month, 10-months, 17-months, 3-years, 4-years, 6-years, and 8-year intervals. Some studies use only one crypto coin as a reference for the methods used, while others study a number of crypto coins for the methods they use. An additional observation is that the most frequent method used in these last five years is based on the LSTM model architecture, and the crypto coins used most frequently are Bitcoin and Ethereum. In our study, we are interested in evaluating how short data can affect the performance of the prediction method. Specifically, we are interested in scenarios that can optimise performance in terms of error (to be minimised) and accuracy (to be maximised).

The prediction model in this study was developed using Google collaboratory tools and the python programming language. Next, modelling was carried out using the LSTM as provided by the Keras library using the initialisation parameters that are shown in the following **Error! Reference source not found.**.

Table 4 Modeling Parameters

No	Type	Information
1	Layers	2
2	Optimizer	Adam
3	Loss	MAE
4	Epoch	300, 500
5	Batch-size	100

Modelling using LSTM was carried out through 2 modelling experiments. The first is a model that uses an input layer filled with 100 neurons, 300 epochs, and batch-size 100 tested on Bitcoin, Ethereum, and Solana. The allocation of data used in this first experiment is 60/40, 70/30, 75/25, and 80/20.

Table 5 Results of the First Experiment.

Crypto	Allocation Data	RMSE		
		1-month	3-month	5-month
Bitcoin	75/25	1510.913	1598.894	2023.919
Ethereum	80/20	119,707	123,289	150,552
Solana	80/20	2,404	6,438	9,755

Table 5 presents the RMSE results from the first modelling of Bitcoin, Ethereum, and Solana with the data used 1-month, 3-months, and 5-months. The results show the use of the best data allocation with the lowest error value is 80/20 for Solana and Ethereum coins and 75/25 for Bitcoin. The error value generated for the Solana coin is more minor than Bitcoin, which has an error value in the thousands for 1-month data usage. It can be seen that the difference in the RMSE value generated by the three coins is relatively small, with the amount of data used being different.

The second experiment was carried out utilising Multivariate Time Series Forecasting. Multivariate Time Series Forecasting is a forecasting method where it is inherently assumed that the variables depend on each other [31]. In other words, this method uses more than one variable or changing criteria from time to time[32]. Using this, we expect the prediction results to be more accurate than univariate methods, i.e., based on only one variable [33]. In this second experiment, it is not only data allocation and short-term data that define the scenario. Here, we test the value of the neuron in its LSTM layer and the running epoch. First, the conversion process was carried out so that supervised learning could be carried out, entering the input sequence, casting the sequence, combining the sequence results, and checking for null data. Next, we determine the data variables to be used. Open, Close, and Volume from datasets will be used to test this data. Then we normalise the data with the min-max scaler and proceed with reframing the sequence data earlier.

Then we do the data allocation to build a prediction model using the LSTM method. In this case, the data allocation is the same as in the first experiment 60/40, 70/30, 75/25, and 80/20. The LSTM layers used here are 100, 128, 142, 182, and 360. The different inputs of the LSTM layer are intended to determine whether the input value will affect the results of the predictions made using this short-term data. There is no specific value or formula to determine the LSTM Layer input during modelling, and the researcher only estimates the possible numbers to produce the best deal. The epochs used are 300 and 500. Experiments were carried out on short-term data on Bitcoin, Ethereum, and Solana coins over a period of 1-month, 3-months, and 5-months.

Table 6 Results of RMSE and R^2 Bitcoin Second Experiment

Data	Data Allocations	LSTM layers	Epochs	RMSE	R^2
1-month	60/40	360	500	61,646	0.9984
3-months	80/20	360	500	49.02	0.9998
5-month		360	500	83.55	0.9997

Table 6 shows the three smallest results from the second experimental process with the resulting RMSE and R2 values to predict bitcoin prices using short-term data. The use of reasonable data allocation on short-term data for Bitcoin is 60/40 and 80/20 with the input layer LSTM 360 and epoch 500. The minor results for Bitcoin are the use of 3 months of data with a value of 49.02 and testing the coefficient of determination using R2 is 0.9998.

Figure 7 The results of the second bitcoin prediction experiment

Even though the resulting error is relatively high, the results of testing the coefficient of determination using R2 touch the number 0.999. Where according to the previous explanation that the higher the value of R2 close to number one means, the better the prediction model of the proposed research model. These results can also be seen in the visualisation in Figure 7, for the y-axis is the price of coins in USD, and the x-axis is the date, that the actual and predicted graphs are the same and balanced, which means this method can be used.

Table 7 Results of RMSE and R^2 Ethereum Second Experiment

Data	Data Allocations	LSTM layers	Epoch	RMSE	R^2
1-month	80/20	360	500	12,241	0.9905
3-months		360	500	3,618	0.9985
5-months		360	300	2,887	0.9999

Table 7 shows the three smallest results from the second experimental process with the resulting RMSE and R^2 values to predict Ethereum prices using short-term data. The data allocation distribution in the Ethereum experiment was the same for the slightest error value of 80/20 with LSTM 360 and 500 epoch inputs.

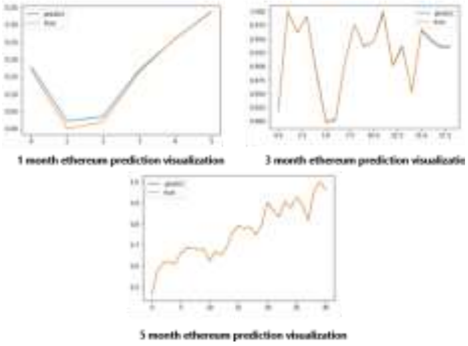

Figure 8 The results of the second Ethereum prediction experiment

The minor result for Ethereum price prediction uses 5-month data with a value of 2.887, and the test value of the coefficient of determination using R2 is 0.9999. The graph shown in Figure 8 shows that there is no line difference between the actual value and the predicted results from the modelling made.

Table 8 Results of RMSE and R^2 Solana Second Experiment

Data	Data Allocations	LSTM layers	Epoch	RMSE	R^2
1-month	80/20	360	500	0.111	0.9962
3-months	60/40	360	500	0.884	0.9984
5-months	75/25	128	300	0.318	0.9998

Table 8 shows the three smallest results from the second experimental process with the resulting RMSE and R2 values to predict Solana prices using short-term data. Different from the two previous crypto coins, the prediction results for the Solana coin are pretty varied for the use of data allocation, so the RMSE results are below number 1.

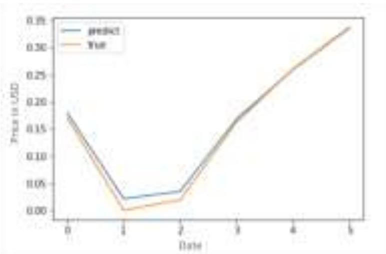

Figure 9 1-Month Solana Prediction Visualization

Starting with 1-month short-term data used on Solana crypto coin, Figure 9 shows the actual value and the results of the predictions made. The data allocation used is 80/20 with the number of LSTM layers inputted, which is 360, and the epoch used is 500 resulting in the smallest RMSE value of 0.111 with an R^2 value of 0.9962. From the visualisation results obtained, there is a prediction graph that is slightly above the actual value, but these results can still be used by considering the amount of error obtained from the prediction results.

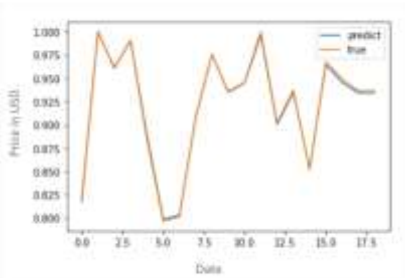

Figure 10 3-Month Solana Prediction Visualization

Then we used 3-month of short-term data on the Solana crypto coin. Figure 10 shows the actual value and results of the predictions made. The data allocation used is 60/40, with the number of LSTM input layers being 360 and the number of epochs used 500, resulting in the smallest RMSE value of 0.884 with an R^2 value of 0.9984. From the results of the visualisation obtained, there is a prediction graph that does not show a significant difference between the actual value and the results of the predictions made.

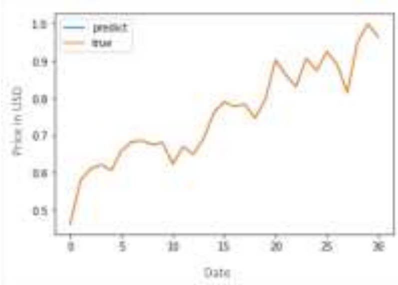

Figure 11 5-Month Solana Prediction Visualization

And lastly, 5-month short-term data on the crypto coin Solana have been used. Figure 11 shows the actual value and the results of the predictions made. The data allocation used is 75/25, with the number of input LSTM layers being 182 and the number of epochs used being 300 resulting in the smallest RMSE value of 0.318 with an R^2 value of 0.9997. From the results of the visualisation obtained, there is a prediction graph that does not show a significant difference between the actual value and the results of the predictions made.

The use of data in predicting cryptocurrency prices using the LSTM method will affect the results of the prediction model. This is emphasised more when using short-term data. Various evaluation values are obtained from the results of short-term data experiments to predict cryptocurrency prices using the LSTM method. In the first experiment using the simple LSTM method, it was able to make predictions on all crypto coins used with the smallest RMSE value in this first experiment, namely the RMSE of the Solana coin being 2,404 for the scenario of its application, namely 1-month data, 60/40 data allocation, and 300 epochs. Bitcoin generates an RMSE value of 1510,913 and Ethereum generates an RMSE of 119,707.

This second experiment uses a multivariate time series which produces a significant error. In particular, the multivariate time series method using LSTM on Bitcoin produces a value of RMSE = 49.02 with R2 = 0.9998. Furthermore, the resulting RMSE for Ethereum is 2.887 with R2 = 0.9999, and the RMSE for the Solana coin is 0.111 with R2 = 0.9962. Compared to published research results (Pradana Ananda Raharja, 2021) which used one month of Ethereum coin data in its opening data, resulted in an RMSE value of 890.29, there was a decrease in the error value in this study using 1-month of Ethereum data resulting in an RMSE of 12,241.

In addition to utilising Multivariate Time Series Forecasting, this research also studies scenarios with varying LSTM input layers, data allocation, and epoch values in the predictive modelling process. By observing quite a variety of results in each test, the method used can only be considered optimal for certain types of crypto coins. Several factors may be involved, especially the difference in the price of the coins used to predict the price of these cryptocurrencies. Bitcoin (BTC) and Ethereum (ETH) have quite high prices with high USD-level volatility, so high and low prices have a large enough gap to be studied using the method used.

4. CONCLUSIONS

Based on the results and previous discussion, it can be seen that using short-term data to predict cryptocurrency prices using the LSTM method can be done with two experiments, namely the first experiment using the LSTM method and the second using Multivariate Time Series Forecasting and LSTM. This study achieved the nominal error value for using short-term data in predicting using Solana coins, namely RMSE = 0.111 and R2 = 0.9962, with the distribution of the data allocation used being 80/20, the input layer LSTM = 360 and epoch = 500.

The author realises that the results obtained still need to be improved. Therefore, the consideration for further research is to consider the hidden layer in the model and the number of input layers used. In addition, future research also contemplates using coins with high and low transaction volumes, which will affect the prediction results.

REFERENCES

[1] M. M. Patel, S. Tanwar, R. Gupta, and N. Kumar, “A Deep Learning-based Cryptocurrency Price Prediction Scheme for Financial Institutions,” *J. Inf. Secure. Appl.*, vol. 55, no. August, p. 102583, 2020, doi: 10.1016/j.jisa.2020.102583.

[2] S. Tanwar, N. P. Patel, S. N. Patel, J. R. Patel, G. Sharma, and I. E. Davidson, “Deep Learning-Based Cryptocurrency Price Prediction Scheme with Inter-Dependent Relations,” *IEEE Access*, vol. 9, pp. 138633–138646, 2021, doi: 10.1109/ACCESS.2021.3117848.

[3] Capfemini, “Non-cash payments volume – World Payments Report.” https://worldpaymentsreport.com/non-cash-payments-volume-2/ (accessed September 29, 2022).

[4] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” August 31, 2009. https://git.dhimmel.com/bitcoin-whitepaper/ (accessed September 29, 2022).

[5] P. N. Sureshbhai, P. Bhattacharya, and S. Tanwar, “KaRuNa: A blockchain-based sentiment analysis framework for fraud cryptocurrency schemes,” *2020 IEEE Int. Conf. Commun. Work. ICC Work. 2020 - Proc.*, pp. 0–5, 2020, doi: 10.1109/ICCWorkshops49005.2020.9145151.

[6] A. M. Khedr, I. Arif, P. V. Pravija Raj, M. El-Bannany, S. M. Alhashmi, and M. Sreedharan, “Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey,” *Intell. Syst. Accounting, Financ. Manag.*, vol. 28, no. 1, pp. 3–34, 2021, doi: 10.1002/isaf.1488.

[7] R. Böhme, N. Christin, B. Edelman, and T. Moore, “Bitcoin: Economics, Technology, and Governance,” vol. 29, no. 2, pp. 213–238, 2015.

[8] S. McNally, J. Roche, and S. Caton, “Predicting the Price of Bitcoin Using Machine Learning,” *Proc. - 26th Euromicro Int. Conf. Parallel, Distrib. Network-Based Process. PDP 2018*, pp. 339–343, 2018, doi: 10.1109/PDP2018.2018.00060.

[9] M. Rizwan, S. Narejo, and M. Javed, “Bitcoin price prediction using Deep Learning Algorithm,” *MACS 2019 - 13th Int. Conf. Math. Actuar. Sci. Comput. Sci. Stat. Proc.*, pp. 1–7, 2019, doi: 10.1109/MACS48846.2019.9024772.

[10] M. Poongodi *et al.*, “Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system,” *Comput. Electr. Eng.*, vol. 81, p. 106527, 2020, doi: 10.1016/j.compeleceng.2019.106527.

[11] M. Ali and S. Shatabda, “A data selection methodology to train linear regression model to predict bitcoin price,” *2020 2nd Int. Conf. Adv. Inf. Commun. Technol. ICAICT 2020*, no. November 2020, pp. 330–335, 2020, doi: 10.1109/ICAICT51780.2020.9333525.

[12] A. Naghib-Moayed and R. Habibi, “Crypto-Currency Price Prediction with Decision Tree Based Regressions Approach,” *J. Algorithms Comput.*, vol. 52, no. 2, pp. 29–40, 2020, [Online]. Available: http://jac.ut.ac.ir.

[13] A. Vo, Q. Nguyen, and C. Ock, “Sentiment Analysis of News for Effective Cryptocurrency Price Prediction,” vol. 5, no. 2, 2019, doi: 10.18178/ijke.2019.5.2.116.

[14] R. Parekh *et al.*, “DL-GuesS: Deep Learning and Sentiment Analysis-Based Cryptocurrency Price Prediction,” *IEEE Access*, vol. 10, no. March, pp. 35398–35409, 2022, doi: 10.1109/ACCESS.2022.3163305.

[15] D. Setyo Dwitanto, “Analisis Runtun Waktu Untuk Meramalkan Jumlah Pasien Yang Berobat Di Puskesmas Blora Dengan Menggunakan Software Minitab 14,” UNIVERSITAS NEGERI SEMARANG, 2011.

[16] P. A. Raharja, “Prediksi Harga Ethereum Menggunakan Metode Vector Autoregressive,” *J. Informatics, Inf. Syst. ...*, vol. 8106, pp. 71–79, 2021, [Online]. Available: http://journal.ittelkom-pwt.ac.id/index.php/inista/article/view/285.

[17] S. Saadah and H. Salsabila, “Prediksi Harga Bitcoin Menggunakan Metode Random Forest (Studi Kasus: Data Acak Pada Awal Masa Pandemic Covid-19),” *J. Komput. Terap.*, vol. 7, no. Vol. 7 No. 1 (2021), pp. 24–32, 2021, doi: 10.35143/jkt.v7i1.4618.

[18] J. P. Fleischer, G. von Laszewski, C. Theran, and Y. J. P. Bautista, “Time Series Analysis of Cryptocurrency Prices Using Long Short-Term Memory,” *Algorithms*, vol. 15, no. 7, pp. 1–13, 2022, doi: 10.3390/a15070230.

[19] Bloomsburg University of Pennsylvania Library, “Definition A Literature review Literature review,” *Lit. Rev.*, no. November, pp. 33–37, 2015.

[20] Supripto, R. N. Rahmanita, and A. S. Kirana, “Teknik pre-processing dan classification dalam data science – Master of Industrial Enginering,” 2022. https://mie.binus.ac.id/2022/08/26/teknik-pre-processing-dan-classification-dalam-data-science/ (accessed December 14, 2022).

[21] M. Farryz Rizkilloh and S. Widiyanesti, “Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM),” *Resti*, vol. 1, no. 1, pp. 19–25, 2022.

[22] D. Haryadi, A. R. Hakim, D. M. U. Atmaja, and S. N. Yutia, “Implementation of Support Vector Regression for Polkadot Cryptocurrency Price Prediction,” *Int. J. Informatics Vis.*, vol. 6, no. 1–2, pp. 201–207, 2022, doi: 10.30630/joiv.6.1-2.945.

[23] H. Puspita and dkk, “Pengantar Teknologi Informasi - Google Books,” *Heri Utama*, 2022. https://www.google.co.id/books/edition/Pengantar_Teknologi_Informasi/43h8EAAAQBAJ?hl=en&gbpv=1&dq=LSTM+adalah&pg=PA197&printsec=frontcover (accessed Oct. 15, 2022).

[24] F. Faturohman, B. Irawan, and C. Setianingsih, “Analisis Sentimen Pada Bpjs Kesehatan Menggunakan Recurrent Neural Network,” *e-Proceeding Eng.*, vol. 7, no. 2, pp. 4545–4552, 2020.

[25] Y. Yu, “A Review of Recurrent Neural Networks: LSTM Celss and Network Architectures,” vol. 31, no. March, pp. 2709–2733, 2019, doi: 10.1162/NECO.

[26] K. Grace and Martin, “Measures of Model Fit for Linear Regression Models,” *theanalysisfactor*, 2013. https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/ (accessed December 14, 2022).

[27] I. Ghozali, “Aplikasi Analisis Multivariete IBM SPSS 23, Badan Penerbit Universitas Diponegoro, Semarang,” 2016. https://scholar.google.co.id/citations?view_op=view_citation&hl=id&user=K8g3CywAAAAJ&citation_for_view=K8g3CywAAAAJ:r0BpntZqJG4C (accessed Dec. 14, 2022).

[28] Harinaldi, *Prinsip-prinsip Statistik untuk Teknik dan Sains Fantastic Adventure : September 1939 Notes from a Literal Life Dødelig ideal Lullabies for Little Criminals Jesus Works Here : Leading Christians in Business Talk About How You Can Walk With Christ Through*, no. September 1939. 2005.

[29] Y. Andrianto, “The Effect of Cryptocurrency on Investment Portfolio Effectiveness,” *J. Financ. Account.*, vol. 5, no. 6, p. 229, 2017, doi: 10.11648/j.jfa.20170506.14.

[30] S. Karasu, A. Altan, Z. Saraç, and H. Ğ. Lu, “Prediction of Bitcoin prices with machine learning methods using time series data,” pp. 1–4, 2018.

[31] X. Chen, “A Multivariate Time Series Modeling and Forecasting Guide with Python Machine Learning Client for SAP HANA,” 2021. https://blogs.sap.com/2021/05/06/a-multivariate-time-series-modeling-and-forecasting-guide-with-python-machine-learning-client-for-sap-hana/ (accessed December 14, 2022).

[32] S. Zahara and Sugianto, “Peramalan Data Indeks Harga Konsumen Berbasis Time Series Multivariate Menggunakan Deep Learning,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 24–30, 2021, doi: 10.29207/resti.v5i1.2562.

[33] D. A. Lusia and A. Ambarwati, “Perbandingan Peramalan Univariat Dan Multivariat Arima Pada Indeks Harga Saham Gabungan,” *J. Stat. Univ. Muhammadiyah Semarang*, vol. 6, no. 2, 2018, [Online]. Available: https://jurnal.unimus.ac.id/index.php/statistik/article/view/4311.