Journal of Information Systems and Informatics
Vol. 5, No. 2, June 2023 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v5i2.473 Published By DRPM-UBD

Investment Decision on Cryptocurrency: Comparing Prediction Performance Using ARIMA and LSTM

Svend Pasak[1], Riyanto Jayadi[2]

[1,2]Information Sytem Management Department, BINUS Graduate Program, Master of Information System, Bina Nusantara University, West Jakarta 11480, Indonesia
Email: [1]svend.pasak@binus.ac.id, [2]riyanto.jayadi@binus.edu

Abstract

The increasing popularity of cryptocurrencies as a means of financial inclusion for investment and trade has become a major concern for individuals seeking to benefit from the cryptocurrency market. This study aims to provide insights for cryptocurrency investors, financial sector professionals, and academics by utilizing machine learning techniques such as ARIMA and LSTM to compare the accuracy of modeling performance on datasets predicting the prices of five cryptocurrencies, namely Bitcoin, Ethereum, Binance Coin, Tether, and Cardano. Data was obtained by downloading from the Yahoo Finance website using Jupyter notebook. The LSTM method outperformed the ARIMA method, achieving a lower MAPE value of less than 10 percent and effectively capturing price movements, providing valuable information for decision-making.

Keywords: Cryptocurrency, forecast, accuracy, ARIMA, LSTM

1. INTRODUCTION

Digital currency, also known as cryptocurrency, is a cryptographic technology that has emerged as a decentralized medium of exchange and transfer, allowing users to transact without the need for intermediaries, thereby ensuring user privacy [1], [2]. The global development of cryptocurrencies over the past five years has been significant, primarily driven by millennial users. The COVID-19 pandemic has further propelled the use of cryptocurrencies for trading, investment, and mining activities. As the cryptocurrency market matures, it is gaining increasing interest from both finance professionals and the social media community [3].

Investors must pay attention to not only the hype surrounding cryptocurrencies but also obtain in-depth knowledge and clear information about digital assets to be purchased, technological advancements associated with the use of cryptocurrency tokens, and investment time horizons. Investors can limit the risks of their investments by making informed decisions and avoiding potential significant losses. The growing value of investments and the number of investors

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

can also be attributed to the role of social media platforms such as Twitter, which is used by influencers to disseminate news and issues related to cryptocurrencies [4]. The excessive hype is often driven by influencers promoting schemes to generate quick wealth through cryptocurrency investments, leading to a Fear of Missing Out (FOMO) among investors. Additionally, cryptocurrency prices can become volatile due to tweets about energy issues related to cryptocurrency mining.

In this study, we examine three proposed research questions, which are as follows: Can machine learning models provide an overview of cryptocurrency asset investing trends? What machine learning methods are used to predict the trend direction and the value of cryptocurrency assets? How does the accuracy of the machine model provide predictive results on cryptocurrency assets? Additionally, between 2018-2022, many crypto assets faced high volatility due to the Covid-19 pandemic [6], including the leading crypto assets, Bitcoin and Ethereum, which experienced a sharp decline. Therefore, machine learning models can predict and forecast these volatile price movements and identify the direction of price trends, making them an excellent tool for knowing cryptocurrency prices [7].

The main contribution of this research is to advise investors on the importance of dynamic cryptocurrency price prediction through machine learning models. With a high level of accuracy, investors can succeed in the cryptocurrency market, and academics increasingly use machine learning in an academic environment. Researchers can further improve their modeling to obtain a picture of data and forecasting from time series data. This study proposes two machine learning algorithms, ARIMA (Auto Regressive Integrated Moving Average) and LSTM (Long Short-Term Memory), which can provide price predictions for five cryptocurrencies.

Several studies have been conducted on cryptocurrency forecasting and the application of machine learning models. In one such study, researchers compared the effectiveness of different predictive models, including K-Nearest Neighbor, Gradient Boosted Trees, Neural Net, and Ensemble, for predicting the prices of multiple cryptocurrencies and 30 cryptocurrency indices [8]. Another study used Multi-Layer Perceptron models, Radial Basis Function Neural Networks, Convolutional Neural Networks, and LSTM to compare accuracy performance on several cryptocurrencies [9]. Additionally, other studies predicted the prices of various cryptocurrencies, such as Monero and Litecoin, using LSTM and Gated Recurrent Units [10], while Linear Regression and Support Vector Machine methods were used to research the cryptocurrency [11], Ether.

Further research was conducted using different methods, including ARIMA, Auto-Regressive Fractionally Integrated Moving Average, and Detrended Fluctuation Analysis, to explore time series price dynamics for several

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

cryptocurrencies [12]. In another study, six deep learning models, including Convolutional Neural Networks and Stacked Long Short-Term Memory, were presented for predicting the price of Ethereum [13]. Several machine learning models were also compared for high-frequency trading on Bitcoin, including Support Vector Regression, Gaussian Poisson Regression, and Regression Tree [14]. Moreover, several machine learning models were used to predict Bitcoin prices based on daily prices and high-frequency prices, including Logistic Regression, Linear Discriminant Analysis, Random Forest, XG Boost, Quadratic Discriminant Analysis, SVM, and LSTM [2], [15], [16], [17].

This study focuses on ARIMA and LSTM models, which use the closing price of five cryptocurrency assets, namely Bitcoin, Ethereum, Binance Coin, Tether, and Cardano, and evaluates the accuracy of the models using MAPE and RMSE metrics. As the use of cryptocurrencies continues to grow, these studies provide valuable insights into the application of machine learning models for predicting cryptocurrency prices and offer a foundation for further research in this field.

2. METHODS

The main focus of this research is to predict the value of five cryptocurrencies, using datasets obtained from the Yahoo Finance website. These datasets consist of a total of 1328 rows and seven columns, which include Date, Open, High, Low, Close, Adj. Close, and Volume. The data interval spans from November 9, 2017, to June 28, 2021. To prepare the data for analysis, a pre-processing step was performed, which involved cleaning the data to remove any missing values. The data was then visualized to better understand the patterns and trends present in the data.

To achieve the research objectives, two methods were employed: ARIMA and LSTM. The closing price datasets for each of the five cryptocurrencies (Bitcoin, Ethereum, Binance Coin, Tether, and Cardano) were used in the analysis. Figures 1 through 5 present the visualizations of the closing price datasets for each of the five cryptocurrencies. These visualizations serve as an important reference point for the subsequent analysis using ARIMA and LSTM models. The accuracy of the models will be evaluated using two metrics: Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE).

2.1. ARIMA

In this study, the ARIMA method, also known as Box-Jenkins, was chosen for analysis. This method was selected due to its flexibility in following data patterns. The time series is represented as a vector set x(t), where t theoretically denotes the elapsed time and x(t) is considered a random variable. Measurements were taken

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

in a chronologically arranged time series for the event [18],[19]. The concept of stationarity can be understood as a statistical form of equilibrium. Stationary processes have statistical properties such as mean and variance that are independent of time, making them useful for future forecasting.

To apply the ARIMA method, five steps are followed: Autoregressive, Integrated, and Moving Average. The AR part is used to model the correlation between observations at different points in time. The I part is used to make the time series stationary, and the MA part is used to model the dependency between an observation and a residual error from a moving average model applied to lagged observations. These steps help in identifying patterns and trends in the time series data and create a helpful time series model for future forecasting.

Figure 1. Bitcoin

Figure 2. Ethereum

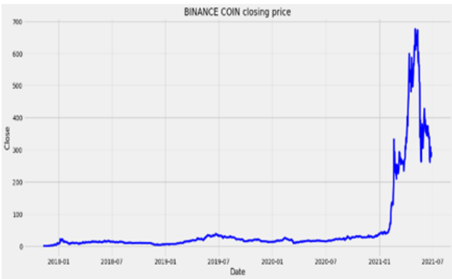

Figure 3. Binance Coin

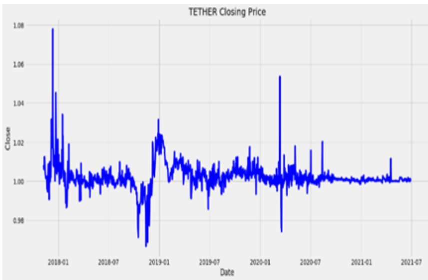

Figure 4. Tether

Figure 5. Cardano

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2.1.1. Autoregressive (AR)

According to [18], in the AR model (p order), the future value of a variable is measured as a linear combination of past observations and random error along with a constant. Mathematical equation (1) is the AR model in question.

$$y_t=C+\O_1 y_{t-1}+\O_2 y_{t-2}+\cdots+\O_p y_{t-p}+\varepsilon_t \quad (1)$$

2.1.2. Integrated (I)

The value I in the model refers to the integration level of the variables. Integrated variables can be changed to be stationary through the differentiation process (differencing). The ARIMA composition is modeled on autocorrelation circumstances and can be used to model stationary or non-stationary time series. A stationary series is one in which the expected values and variances do not change over time, the values of the series itself do not deviate from their initial values, and the covariance values for two observations depend on the distance between them, not the time of origin.

2.1.3. Moving Average (MA)

Like the AR (p) model, predicting by paying attention to the value of the previous series, the mathematical equation (2) MA model (order q) uses the past error as an explanatory variable.

$$Y_t=C+\varepsilon_t-\O_1\varepsilon_{t-2}-\ldots-\O_q\varepsilon_{t-q} \quad (2)$$

C is a constant or average, is a parameter of the model, and q is an order moving average.

2.1.4. Model Autoregressive moving integrated average (ARIMA)

Non-seasonal ARIMA models are generally denoted with p,d and q [20]. AR and MA can be combined effectively to form a valuable and standard time series class model. If the series is stationary, then the ARIMA model can be expressed as follows:

$$y_t=\sum_{i=1}^{p} a_i y_t\text{-}i+\ \sum_{j=1}^{q}\delta j\varepsilon t\text{-}j \quad (3)$$

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2.1.5. Autocorrelation dan Partial Autocorrelation Function (ACF and PACF)

In this study, the ACF and PACF were analyzed to determine a suitable model for specific time series data. These statistical measures reflect the relationship between observations in the time series. ACF and PACF plots were created by plotting the correlation coefficient against successive lags. The ACF plot displays the observed correlations with the lag values, with the x-axis showing the lag and the y-axis showing the correlation coefficient, ranging from minus one for negative correlations to one for positive correlations. The PACF plot compiles the correlations for surveillance with lag values not accounted for by previous lag surveillance.

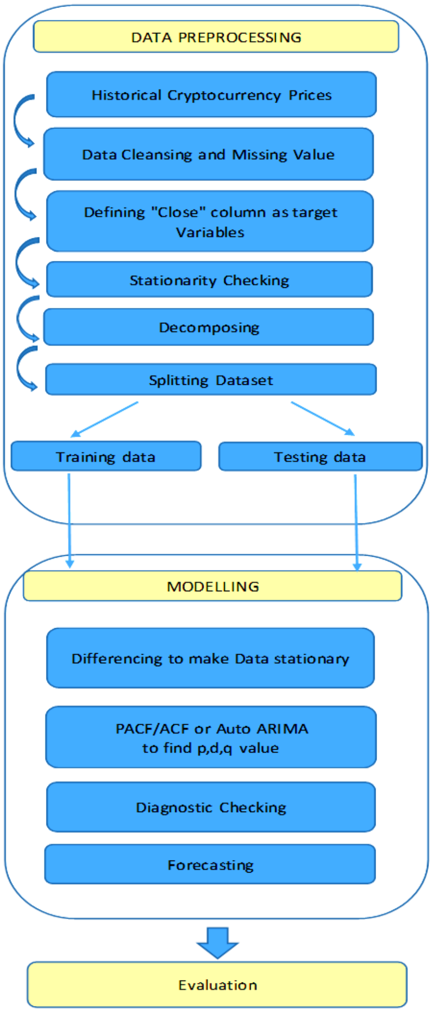

Figure 6. ARIMA Framework structure

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

To select the appropriate model, the ACF and PACF were examined using the Augmented-Dickey-Fuller statistical tests. The selected parameters were approximated in the second step, which involved approximation and testing. The selection of the approved model was based on an analysis of various criteria, including model parameter significance, metrics of error, and information criteria such as Akaike Information Criteria and Bayesian Information Criteria. The next step involved diagnostic testing. If the residual model was a noise process, and there were no significant ACF or PACF values of model residuals, the model could be continued for forecasting. If not, the approximation and testing phases had to be repeated, and another model had to be selected. In the ARIMA method, a framework consisting of preprocessing, modeling, and evaluation stages was proposed, as shown in Figure 6.

2.2. LSTM

Long Short-Term Memory (LSTM) cells are a type of recurrent neural network (RNN) [22], which has the ability to capture long-term dependencies [23]. LSTM is also highly effective in predicting the value of time series data based on historical data [24]. In a typical RNN, small weights are repeatedly multiplied over several steps, and the gradient decreases asymptotically to zero, known as the gradient problem. As depicted in Figure 7, LSTM cells generally consist of memory blocks, referred to as cells, that are connected via layers. The cell information is contained in the cell state (Ct) and hidden state (ht) and regulated by mechanisms known as gates via the sigmoid and Tanh activation functions. The sigmoid function/layer outputs a number between 0 and 1, with 0 indicating that nothing went through, and 1 implying that everything went through. LSTM can add or remove information from the cell state. Generally, as an input, the gate takes the hidden state of the previous time step ht−1 and the current input Xt and multiplies the input totally by a matrix of weights, W, and bias to be enhanced to the product. There are three primary gates, namely the input gate (it), the forgetting gate (ft), and the output gate (ot), where the input gate states whether the input may enter or not, and the forgetting gate is responsible for deleting information that is not important. The output gate determines what information will be generated [25]. The equation is as follows:

$$i_t = \sigma_g(w_i x_t + v_i h_{t-1} + b_i) \tag{4}$$
$$f_t = \sigma_g\left(w_f x_t + v_f h_{t-1} + b_f\right) \tag{5}$$
$$o_t = \sigma_c(w_c x_t + v_c h_{t-1} + b_c) \tag{6}$$

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

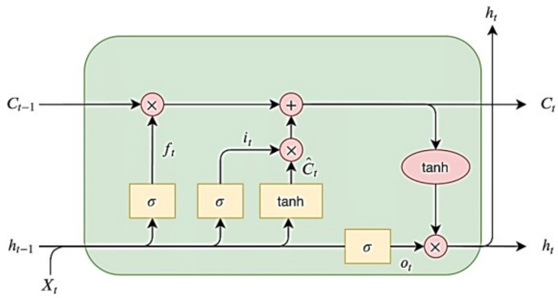

Figure 7. LSTM Cell

Where Xt = The input vector at time t. Ht-1 = The hidden state from the previous time step. Ct-1 = The previous state memory. Ht = The current hidden state. Ct = The current state memory. [*] = Represents the element of multiplication function. [+] = Represents the element of addition function.

In each LSTM module, the input consists of Xt (recent input), Ht-1, and Ct-1, and the output comprises Ht and Ct. The input gate allows only a certain number of recent input states to pass through, while the forget gate determines the number of previous states that are allowed to be forgotten. Meanwhile, the output gate controls the number of internal states that should be exposed to the cell for the next time step and higher layer. The stages utilized for predictions with the LSTM method in this study are outlined as follow.

a) Identify the input and output components.
b) Normalize the data.
c) Allocate data for training, validation, and testing/prediction.
d) Experiment with the number of nodes in the hidden layer and the delay time.
e) Train the model.
f) Validate the model.
g) Generate forecasts.
h) Denormalize the data.
i) Divide the data into training, validation, and testing sets. The training data is utilized to learn unknown patterns, while the validation data guarantees that the created network is appropriate. The testing data is utilized for forecasting using the trained and tested model. Refer to Figure 8 for the resulting model.

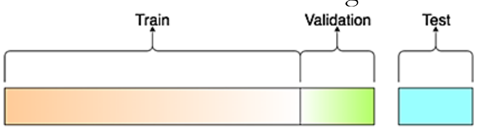

Figure 8. Visualization of split

Then the LSTM method is configured to get the lowest RMSE and MAPE models to get high accuracy values.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Yi-\hat{Yi})^2} \quad (7)$$

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Yi-\hat{Yi}}{Yi}\right) \quad (8)$$

Where:
n : Amount of Data, $Y\hat{i}$: Predicted Value, and Yi : Actual Value

In addition, for the LSTM method, we propose a framework which consists of data pre-processing, modeling, and evaluation in Figure 9.

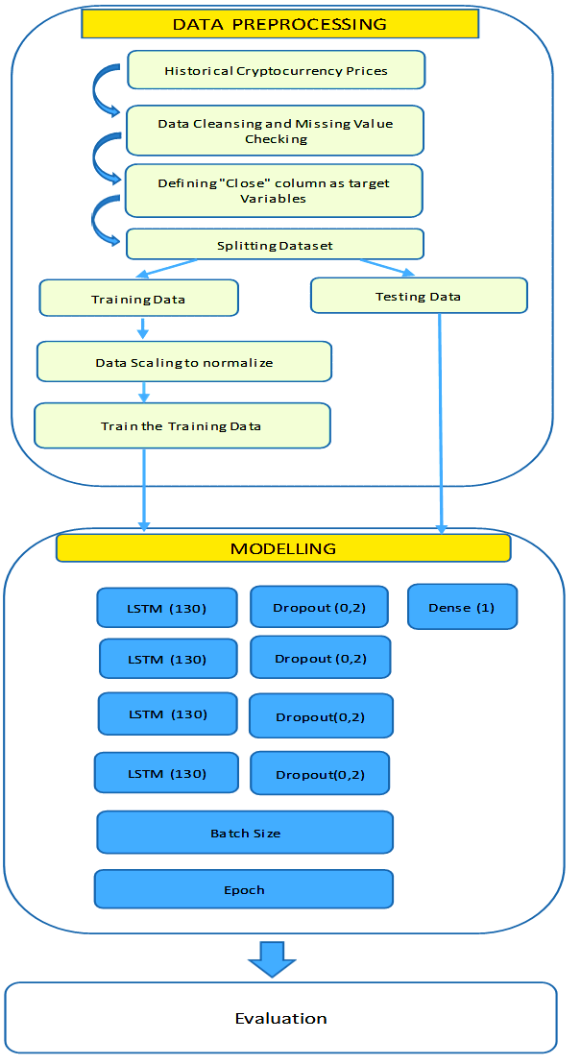

Figure 9. LSTM Framework structure

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

3. RESULTS AND DISCUSSION

3.1 ARIMA

The first step of the ARIMA model involves converting the data into a time series using a time series function. This allows us to check the stationarity of the data by examining a visualization of balance statistics. At this stage, we tested the data stationarity using the ADF test. Next, the processed data was plotted on a graph of rolling statistics, which shows the rolling average (mean) as a brown line, the rolling standard deviation as a blue line, and the original time series data as an orange line.

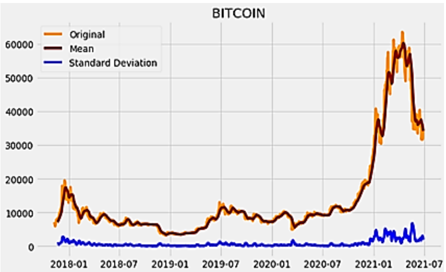

Figure 10. ADF test Bitcoin

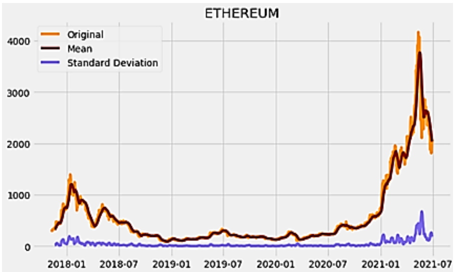

Figure 11. ADF test Ethereum

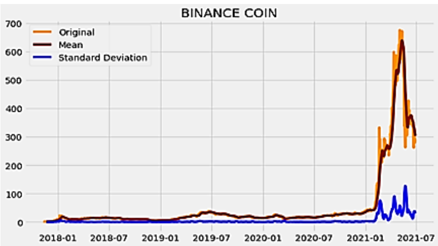

Figure 12. ADF test Binance Coin

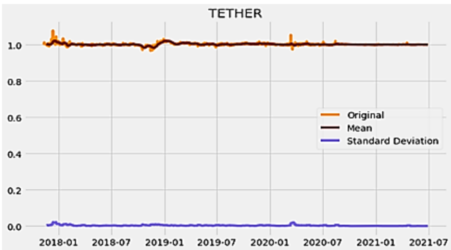

Figure 13. ADF test Tether

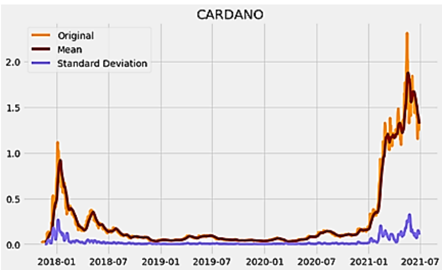

Figure 14. ADF test Cardano

The ARIMA method performs best when the data exhibits a consistent pattern in the time series. If the data tends to increase or decrease and has a seasonal pattern, then it is not stationary. The rolling statistics and graphs in this study reveal a p-value of > 0.05, which means we accept H0 (null hypothesis) of non-stationary data, as shown in Table 1. Figures 10 through 14 demonstrate that the five cryptocurrencies under investigation do not have a seasonal pattern but exhibit an upward trend, with the exception of Tether, which remains constant. Based on these findings, we can conclude that the data is in good condition.

Table 1. Result of ADF Test

Test	BITCOIN	ETHEREUM	BINANCE COIN	TETHER	CARDANO
Test Statistics	-0.497589	-0.215785	-0.972823	-4.796.710	-0.364814
p-value	0.892486	0.936614	0.763047	0.000055	0.915810
No. of lags used	23	21	23	14	22
Number of observations used	1304	1306	1304	1313	1305
critical value (1%)	-3.435	-3.435	-3.435	-3.435	-3.435
critical value (5%)	-2.863	-2.863	-2.863	-2.863	-2.863
critical value (10%)	-2.567	-2.567	-2.567	-2.567	-2.567

The results of the ADF test showed that the null hypothesis (H0) could not be rejected. To address the magnitude and uptrend in the series, we applied the logarithmic function to the time series data. This transformation allowed us to obtain the rolling average of the series by taking input over the past 12 months and providing the average consumption value at each point in the series. We then partitioned the logged time series data into two sets: 85% as training data and 15% as testing data. To identify and isolate seasonality and trend, we utilized the decompose process, which we visualized in figures 15 to 19. By doing so, we gained a more thorough understanding of the underlying patterns in the data, enabling us to generate more precise predictions. This step was critical in ensuring that our model accurately captured the key patterns in the data and produced reliable forecasts.

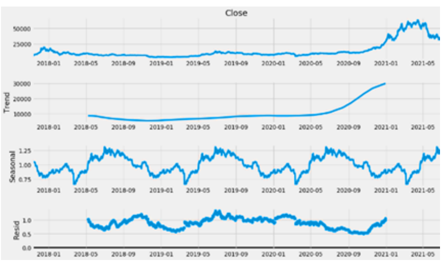

Figure 15. Bitcoin Decompose

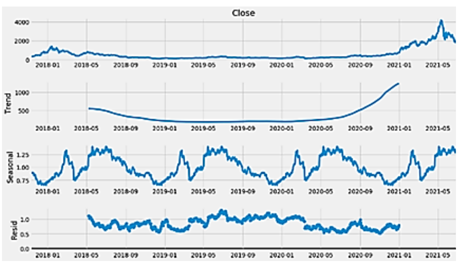

Figure 16. Ethereum Decompo

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

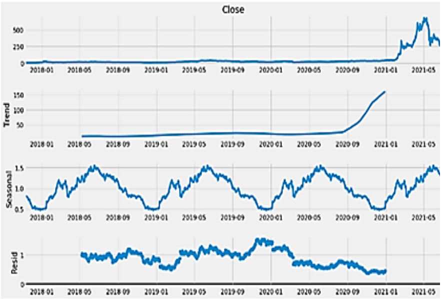

Figure 17. Binance Coin Decompo

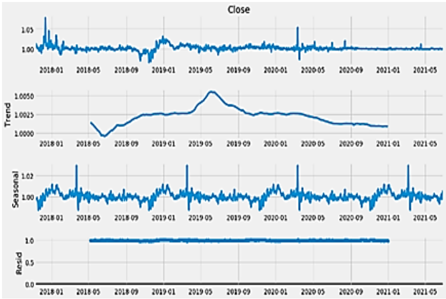

Figure 18. Tether Decompose

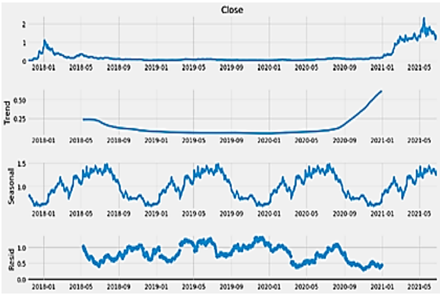

Figure 19. Cardano Decompose

The next step in our analysis involved differencing the logged time series data. This process enabled us to determine the I (integrated) value for the ARIMA model and assess whether the data series was stationary. Specifically, we examined the p-value and looked for values below 0.05. For Tether, we found that further differentiation was unnecessary as the data was already stationary with a p-value of <0.05. The results of our analysis are presented in figures 20 to 23, which illustrate stationarity and p-values across the four remaining cryptocurrencies. Additionally, we summarized the p-value results in Table 2. By carefully assessing the stationarity of our data, we ensured that our model was robust and would yield accurate predictions.

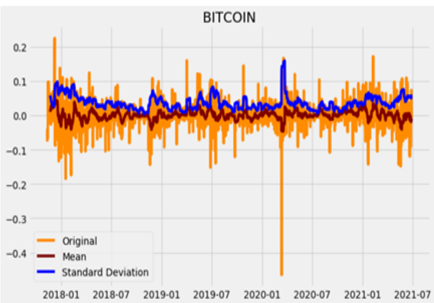

Figure 20. Stationarity and Differencing Bitcoin

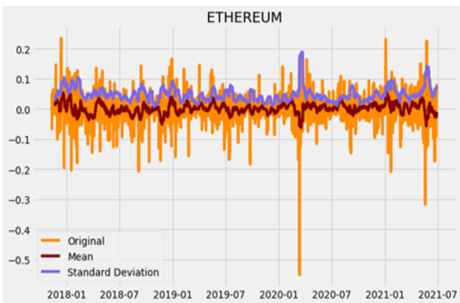

Figure 21. Stationarity and Differencing Ethereum

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

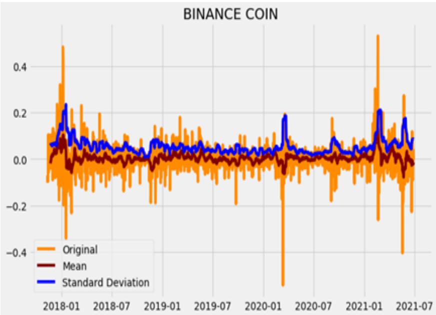

Figure 22. Stationarity and Differencing Binance Coin

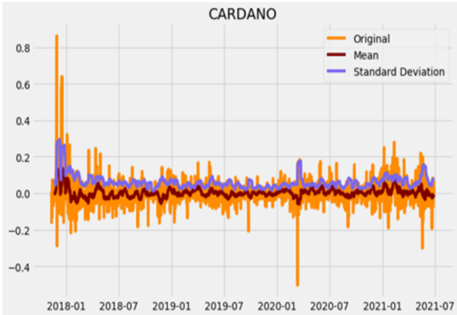

Figure 23. Stationarity and Differencing Cardano

Table 2. Stationarity and Differencing

Test	BITCOIN	ETHEREUM	BINANCE COIN	TETHER	CARDANO
Test Statistics	-24.952.067	-1,08E+07	-9,56E+06	Tether is already in stationary	-6.402.837
p-value	0.000000	2,56E-13	2,38E-10		1,98E-02
No. of lags used	1	9	9		1,80E+07
Number of observations used	1325	1,32E+09	1,32E+09		1,31E+09
critical value (1%)	-3.435	-3.435	-3.435		-3.435
critical value (5%)	-2.863	-2.863	-2.863		-2.863
critical value (10%)	-2.567	-2.567	-2.567		-2.567

Furthermore, the PACF value is determined, namely autoregressive (AR), and the ACF value, which is the moving average (MA), is the value of p, q. In Figure 24 to Figure 28 plots, a blue zone represents the 95% confidence area. It is a threshold level of significance, anything within the blue zone is statistically close to zero, and anything outside the blue zone is statistically non-zero.

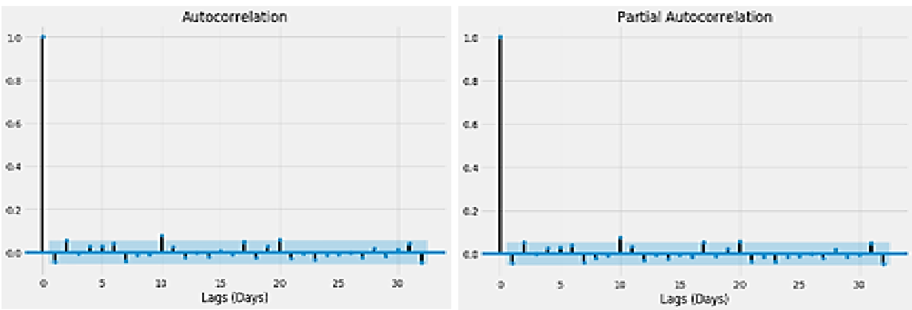

Figure 24. AR and MA BITCOIN value

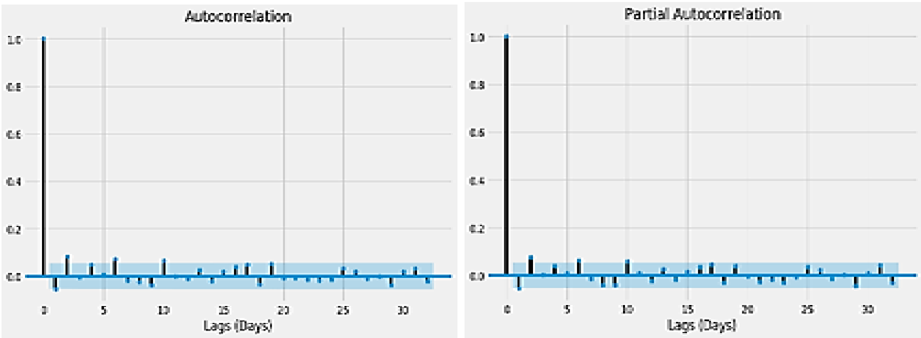

Figure 25. AR and MA ETHEREUM value

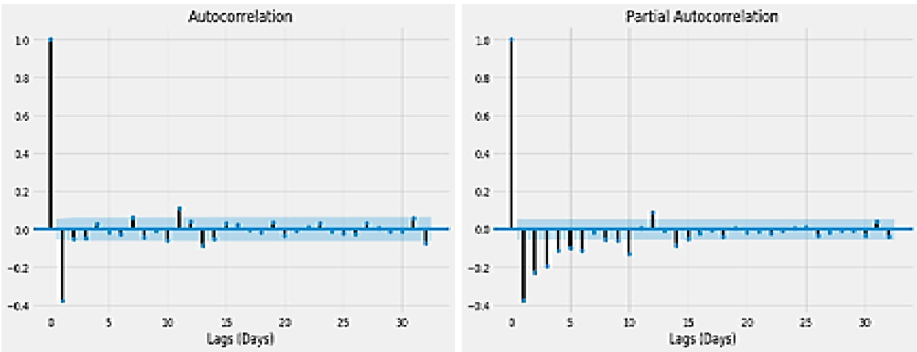

Figure 26. AR and MA TETHER value

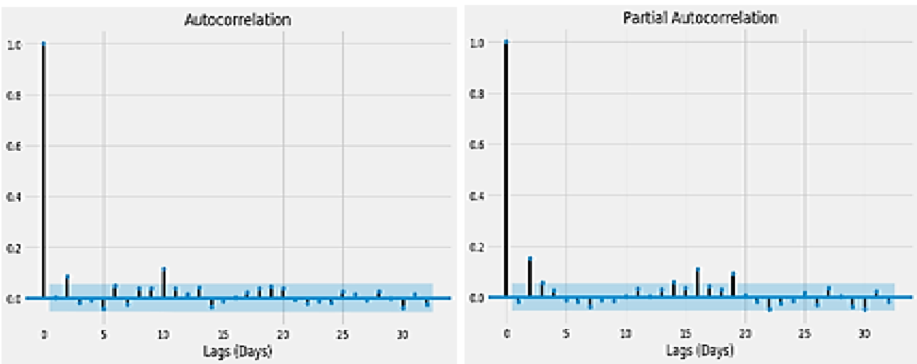

Figure 27. AR and MA BINANCE COIN value

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

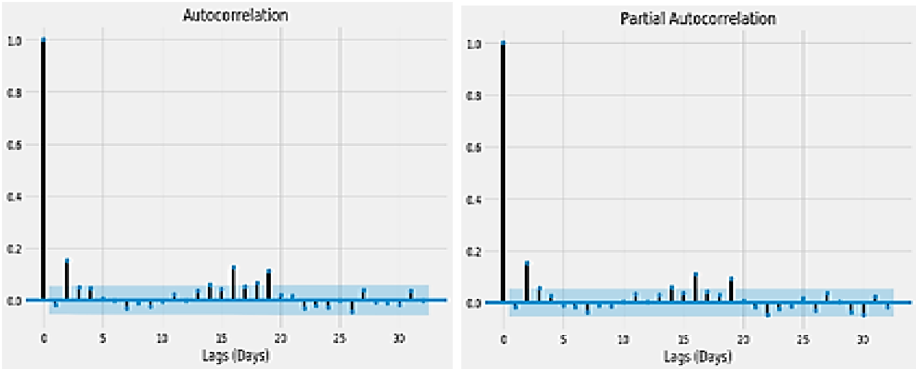

Figure 28. AR and MA CARDANO value

After analyzing the ACF and PACF plots, the best p, d, and q values can be determined using the Autoarima or ARIMA function. The lag values from figures 24 to 28 are used to identify significant lines that come out from the blue area, which are taken as the values for p and q. The value of d is set to one because the four cryptocurrencies have already been differentiated once to achieve stationarity, except for Tether.

To determine the best p, d, and q values, we select the smallest Akaike Information Criterion (AIC) value, which corresponds to the number of differencing (one-time differencing, except for Tether). Once the best p, d, and q values are identified, we proceed to model each cryptocurrency by inputting the order values proposed in Table 1.

The dataset of 1328 is divided into 1129 for training data and 199 for test data. The model is trained by integrating the ARIMA function, training data, and the proposed order. Once the model is trained, it is used as a predictive model by incorporating the forecast function and test data. Figures 29 to 33 are the visualizations of the prediction models generated.

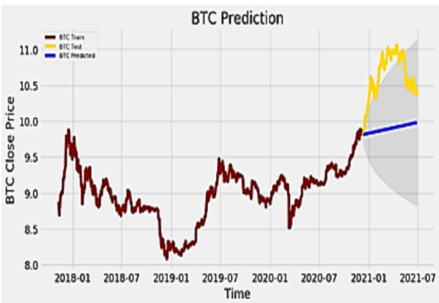

Figure 29. BITCOIN-ARIMA price Prediction

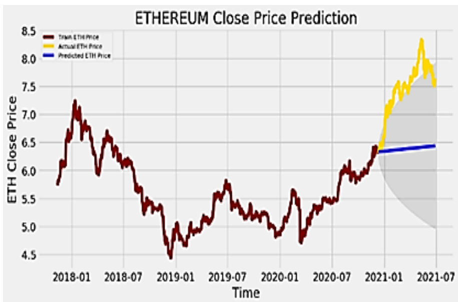

Figure 30. ETHEREUM-ARIMA price Prediction

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

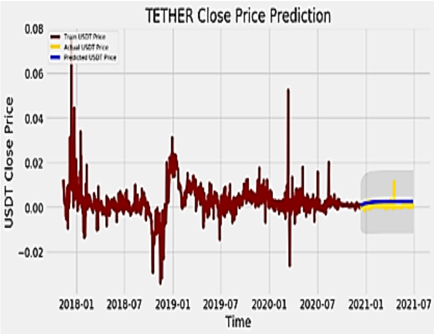

Figure 31. TETHER-ARIMA price Prediction

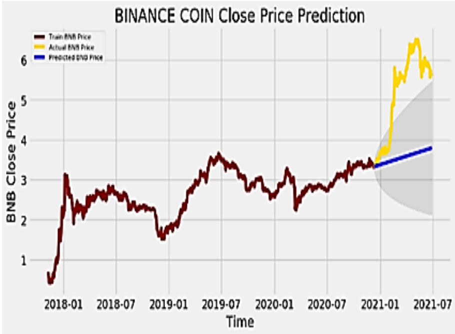

Figure 32. BINANCE COIN-ARIMA price Prediction

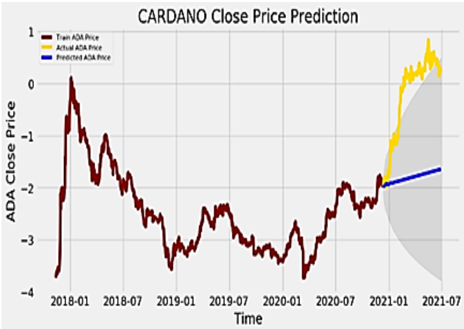

Figure 33. CARDANO-ARIMA price Prediction

Table 3. The Result of Arima Method

Crypto Assets	p,d,q proposed	AIC	Accuracy	
			MAPE (%)	RMSE
BTC	0, 1, 0	-3.959.170	6.952	0.797
ETH	2, 1, 0	-3.474.642	14.141	1.166
USDT	2, 0, 1	-8.742.304	1.806.809	0.001
BNB	2, 1, 2	-3.175.687	28.510	1.872
ADA	3, 1, 3	-2.750.110	763.557	1.740

Table 3 shows the ARIMA result accuracy metric values displayed in MAPE and RMSE units. The lowest MAPE value is 6.9 percent, and the rest is above 10 and above 1000 percent.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

3.2 LSTM

To predict cryptocurrency assets using the LSTM method, the training data is divided into 85% for training and 15% for validation and test data. The first step is to normalize the data scale with MinMaxscaler feature, creating as many training data objects as training data, determining the scale of the training data, and turning the data into three-dimensional data.

The next stage involves building the model, consisting of four LSTM layers with 130 neurons each, four layers of the Dropout layer 0.2, and one Dense layer. The iteration function is applied as a repetition function by studying every ten initial training data (0 to the ninth index) and one data (index ten) used as a prediction target. The model is compiled using the Adam optimizer and trained with a batch size of 700 and 588 epochs. After modeling with training data, the same steps are repeated for test data, creating as many test data objects as the number of test data, determining the scale of the test data, and turning the data into three-dimensional data. Figures 34 to 38 visualize the prediction model using test data, producing MAPE and RMSE accuracy metric values. The green line represents the training data, the blue line represents.

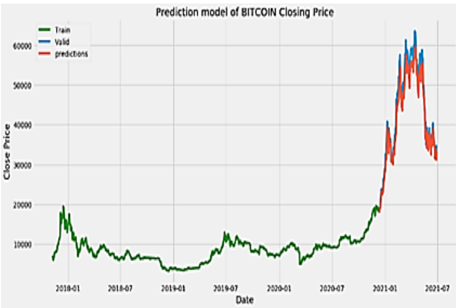

Figure 34. BITCOIN-LSTM price Prediction

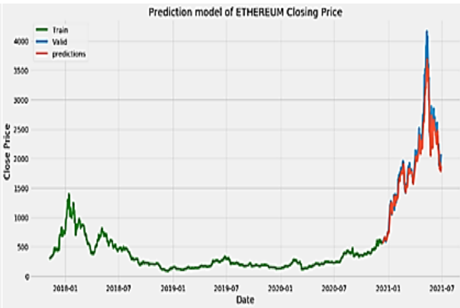

Figure 35. ETHEREUM-LSTM price Prediction

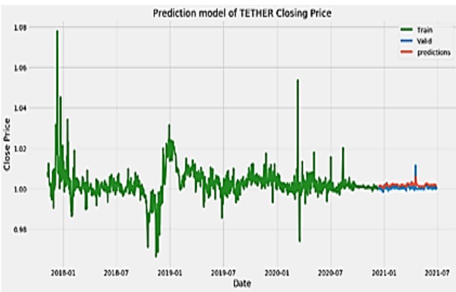

Figure 36. TETHER-LSTM price Prediction

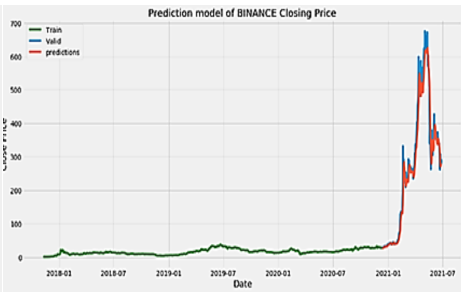

Figure 37. BINANCE COIN-LSTM price Prediction

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

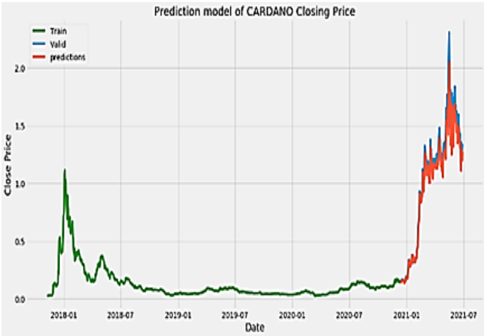

Figure 38. CARDANO-LSTM price Prediction

Table 4. The Result of LSTM Method

		Cryptocurrency Assets				
		BTC	**ETH**	**USDT**	**BNB**	**ADA**
	Training data	1129	1129	1129	1129	1129
	Testing data	199	199	199	199	199
	Batch Size	700	700	700	700	700
	Epoch	588	588	588	588	588
4 Layers	**LSTM** (Neurons)	130	130	130	130	130
4 Layers	**Dropout**	0,2	0,2	0,2	0,2	0,2
1 Layers	**Dense**	1	1	1	1	1
Accuracy	**MAPE (%)**	5,202	5,63	0,073	9,956	6,457
	RMSE	2879,051	182,907	0,001	41,427	0,111

Table 4 presents the accuracy metric performance of the LSTM method, which demonstrates a superior accuracy level compared to ARIMA, as reflected in the smaller MAPE values of less than 10 percent for all five cryptocurrencies. The visualization of the results indicates that the LSTM method is capable of accurately tracking the fluctuations of cryptocurrency prices, while ARIMA appears as a straight line, failing to capture the intricacies of the data.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

4. CONCLUSION

This study aimed to compare the accuracy of the ARIMA and LSTM methods in predicting time series data patterns. The results of the models indicate that the LSTM method outperforms the ARIMA method in terms of accuracy metrics and visualization. While the ARIMA method captures volatility and shows trends and seasonality, the LSTM method can read data movements and track fluctuations in cryptocurrency prices more effectively. Moreover, the LSTM method allows for adjustments to the number of neurons, batch size, and epoch values in both the training and test data sections to improve accuracy. Future research could further enhance accuracy by increasing these parameters in each layer, given the high volatility in cryptocurrencies. Overall, this study contributes to the understanding of the effectiveness of different methods for time series prediction and provides insights for future research in this field.

REFERENCES

[1] N. Gonzálvez-Gallego and M. C. Pérez-Cárceles, "Cryptocurrencies and illicit practices: The role of governance," *Econ. Anal. Policy*, vol. 72, pp. 203–212, 2021, doi: 10.1016/j.eap.2021.08.003.

[2] Ferdiansyah, S. H. Othman, R. Z. M. Radzi, D. Stiawan, and T. Sutikno, "Hybrid gated recurrent unit bidirectional-long short-term memory model to improve cryptocurrency prediction accuracy," *IAES Int. J. Artif. Intell.*, vol. 12, no. 1, pp. 251–261, 2023, doi: 10.11591/ijai.v12.i1.pp251-261.

[3] M. Ortu, S. Vacca, G. Destefanis, and C. Conversano, "Cryptocurrency ecosystems and social media environments: An empirical analysis through Hawkes' models and natural language processing," *Mach. Learn. with Appl.*, vol. 7, no. November 2021, p. 100229, 2022, doi: 10.1016/j.mlwa.2021.100229.

[4] C. C. Whitlock, "Cryptocurrency scam warning: 'Fomo' leads to £63m being lost through social media fraud," May 26, 2021.

[5] M. Del Castillo and S. Ehrlich, "Crypto's Great Reset: How Digital Asset Investors Will Recover From The Market's $1 Trillion Meltdown," May 2022.

[6] A. Ugolini, J. C. Reboredo, and W. Mensi, "safe-haven assets," vol. 53, no. November 2022, 2023.

[7] A. A. Oyedele, A. O. Ajayi, L. O. Oyedele, S. A. Bello, and K. O. Jimoh, "Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction," *Expert Systems with Applications*, vol. 213. 2023, doi: 10.1016/j.eswa.2022.119233.

[8] R. Chowdhury, M. A. Rahman, M. S. Rahman, and M. R. C. Mahdy, "Predicting and forecasting the price of constituents and index of cryptocurrency using machine learning," *arXiv*, pp. 1–38, 2019.

[9] S. Alonso-Monsalve, A. L. Suárez-Cetrulo, A. Cervantes, and D. Quintana,

"Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators," *Expert Syst. Appl.*, vol. 149, p. 113250, 2020, doi: 10.1016/j.eswa.2020.113250.

[10] M. M. Patel, S. Tanwar, R. Gupta, and N. Kumar, "A Deep Learning-based Cryptocurrency Price Prediction Scheme for Financial Institutions," *J. Inf. Secur. Appl.*, vol. 55, no. June, 2020, doi: 10.1016/j.jisa.2020.102583.

[11] 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.

[12] S. A. David, C. M. C. Inacio, R. Nunes, and J. A. T. Machado, "Fractional and fractal processes applied to cryptocurrencies price series," *J. Adv. Res.*, vol. 32, pp. 85–98, 2021, doi: 10.1016/j.jare.2020.12.012.

[13] N. A. Hitam, A. R. Ismail, and F. Saeed, "An Optimized Support Vector Machine (SVM) based on Particle Swarm Optimization (PSO) for Cryptocurrency Forecasting," *Procedia Comput. Sci.*, vol. 163, pp. 427–433, 2019, doi: 10.1016/j.procs.2019.12.125.

[14] T. Zoumpekas, E. Houstis, and M. Vavalis, "ETH analysis and predictions utilizing deep learning," *Expert Syst. Appl.*, vol. 162, no. March, p. 113866, 2020, doi: 10.1016/j.eswa.2020.113866.

[15] S. Lahmiri and S. Bekiros, "Intelligent forecasting with machine learning trading systems in chaotic intraday Bitcoin market," *Chaos, Solitons and Fractals*, vol. 133, 2020, doi: 10.1016/j.chaos.2020.109641.

[16] T. A. Borges and R. F. Neves, "Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods," *Appl. Soft Comput. J.*, vol. 90, p. 106187, 2020, doi: 10.1016/j.asoc.2020.106187.

[17] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," *J. Comput. Appl. Math.*, vol. 365, p. 112395, 2020, doi: 10.1016/j.cam.2019.112395.

[18] J. Sun, "Forecasting COVID-19 pandemic in Alberta, Canada using modified ARIMA models," *Comput. Methods Programs Biomed. Updat.*, vol. 1, no. April, p. 100029, 2021, doi: 10.1016/j.cmpbup.2021.100029.

[19] Q. Yang, J. Wang, H. Ma, and X. Wang, "Research on COVID-19 based on ARIMA modelΔ—Taking Hubei, China as an example to see the epidemic in Italy," *J. Infect. Public Health*, vol. 13, no. 10, pp. 1415–1418, 2020, doi: 10.1016/j.jiph.2020.06.019.

[20] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Appl. Soft Comput. J.*, vol. 90, p. 106181, 2020, doi: 10.1016/j.asoc.2020.106181.

[21] D. Kobiela, D. Krefta, W. Król, and P. Weichbroth, "ARIMA vs LSTM on NASDAQ stock exchange data," *Procedia Computer Science*, vol. 207. pp. 3830–3839, 2022, doi: 10.1016/j.procs.2022.09.445.

[22] A. Yadav, C. K. Jha, and A. Sharan, "Optimizing LSTM for time series

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

prediction in Indian stock market," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 2091–2100, 2020, doi: 10.1016/j.procs.2020.03.257.

[23] J. Zhang, Y. Zeng, and B. Starly, "Recurrent neural networks with long term temporal dependencies in machine tool wear diagnosis and prognosis," no. February, 2021, doi: https://doi.org/10.1007/s42452-021-04427-5.

[24] G. Ardiyansyah, Ferdiansyah;, and U. Ependi, "View of Deep Learning Model Analysis and Web-Based Implementation of Cryptocurrency Prediction," *J. Inf. Syst. Informatics*, vol. 4, 2022, doi: 10.51519/journalisi.v4i4.365.

[25] S. Mirzaei, J. L. Kang, and K. Y. Chu, "A comparative study on long short-term memory and gated recurrent unit neural networks in fault diagnosis for chemical processes using visualization," *J. Taiwan Inst. Chem. Eng.*, vol. 130, 2022, doi: 10.1016/j.jtice.2021.08.016.