UJM 11 (2) 141-152

UNNES Journal of Mathematics

http://journal.unnes.ac.id/sju/index.php/ujm

OPTIMAL MODELING OF STOCK PRICE FORECASTING USING LONG SHORT-TERM MEMORY AND AUTO ARIMA METHODS

Jannatun Khustia Lubis✉**, Iqbal Kharisudin**

Universitas Negeri Semarang, Kampus Sekaran Gunungpati, Semarang 50229, Indonesia

Gedung D7 Lt 1, Kampus Sekaran Gunungpati, Semarang 50229, Indonesia

Article Info

Article History:

Received Agustus 2022

Accepted Oktober 2022

Published November 2022

Keywords:

Long Short Term Memory, Automatic Autoregressive Integrated Moving Average.

Abstract

This study aims to determine the stock price forecasting of EXCL, ISAT, and TLKM using Long Short Term Memory (LSTM) and auto ARIMA. Stock closing price data from January 1, 2015, to December 31, 2020, is divided into two, 80% for training data and 20% for testing data. Data is processed using each method separately. The stages of research on the LSTM method are data preprocessing, building models, model selection, and forecasting. We conducted experiments to see the effect of variations on the parameters of the LSTM method, that is, the number of epochs and the number of neurons. The stages of research on the auto ARIMA method are data preparation, building a model, and forecasting. The results of the modeling of the two methods were compared with the MSE and RMSE accuracy metrics. In some cases, the two methods compete quite tightly, especially in data that can be transformed into a linear model, Auto ARIMA's capability can be better than LSTM. In our case study, it was found that for EXCL and TLKM stocks the auto ARIMA method was better than the LSTM method, while for ISAT stocks, the LSTM method was found to be better than the auto ARIMA method.

How to cite:

Lubis, J. K., Kharisudin, I. 2022. Optimal Modeling of Stock Price Forecasting Using Long Short-Term Memory and Auto Arima Methods. *UNNES Journal of Mathematics*. 11(2) 141-152.

✉Correspondence address:

E-mail: jannatunkl@students.unnes.ac.id

p-ISSN 2252-6943

e-ISSN 2460-5859

INTRODUCTION

The stock market is one of the main indicators of economic conditions (Ho, Darman, & Musa, 2021). Stock price forecasting is the process of determining the future value of a company's shares or other financial instruments traded on the stock exchange. Forecasting stock price movements play an important role in financial markets. Investors who accurately predict the direction of stock price movements can take optimal strategies in trading that yield large profits. On the other hand, investors who miscalculate the price will suffer losses. Therefore, investors who seek maximum profit are eager to find a method that can predict stock prices correctly (Liu, 2022).

Stock price data is a category of time series data. Time series is a type of data that is observed sequentially from time to time (Saeed, 2022). There are several ways to model time series data, namely stationary models, non-stationary models, linear models, nonlinear models, univariate models, and multivariate models. Autoregressive Integrated Moving Average (ARIMA) is a stationary model. One of the nonlinear models is Long Short Term Memory (LSTM) (Ito, Iima, & Kitamura, 2022).

Research conducted by Sethia & Raut (2018) shows that the LSTM model is proven to outperform other models in predicting the stock price of the S&P 500 company with an R^2 value of 0.9486 and a return 400% greater than the hold and wait strategy. Abdillah, Zukhronah, & Respatiwulan (2021) obtained that the ARIMA model (1,1,1) is the best ARIMA model for predicting stock prices with a MAPE value on out-sample data of 1.95%.

Research conducted by Liu (2022) shows that the ARIMA model can be used to predict short-term stock prices such as one day while the multi-step LSTM model is suitable for predicting long-term stock price trends such as a week. In Shankar's research (2022), the MAPE LSTM value is lower than ARIMA and UCM in predicting the stock price of Tata Motors India Ltd. The results of research by Ho, Darman, & Musa (2021) show that LSTM can produce more than 90% accuracy in predicting stock prices during the pandemic.

In forecasting the NEPSE index using the auto ARIMA method, the ARIMA model has a high potential to anticipate short-term market changes, which may be beneficial for traders or short-term investors (Saeed, 2022). In the research of Vani, Varma, Lakshman, & Krishna (2019), the results show that the ARIMA and LSTM models have strong potential to forecast stocks. The ARIMA model has an accuracy value of 98.87% in forecasting Unilever Indonesia stock prices (Wibowo, Dang, & Wang, 2022).

This research is entitled ‘Optimal Modeling of Stock Prices Forecasting Using Long Short Term Memory and Auto ARIMA Methods’. The purpose of this study was to determine stock price data forecasting using the LSTM and auto ARIMA methods and to compare the accuracy of the two.

Long Short Term Memory (LSTM) is a type of RNN. LSTM was introduced by Hochreiter & Schmidhuber in 1997. LSTM is an RNN architecture designed specifically to address the problem of RNN, because LSTM can capture context-specific temporal dependencies for long periods. LSTM can read, write, and selectively forget information so that only important information remains in memory cells (Ito et al., 2022). Each LSTM neuron is a memory cell that can store other information, that is, it maintains its cell state. While neurons in a normal RNN only take the previous hidden state and input the current to issue a new hidden state, LSTM neurons also take up the old cell state and issue a new cell state (Brownlee, 2017).

LSTM has four layers, that is forget gate (1), input gate (2), new cell state candidate (3), and output gate (4) in the iteration model architecture which looks like Figure 1 (Winata, 2018).

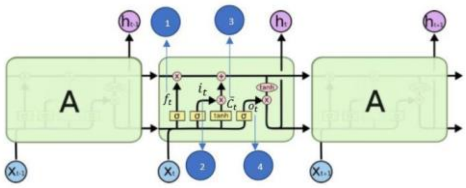

Figure 1. Looping Model with Four Layers in LSTM

Langkah The first step is that the LSTM determines the data information to be removed from the cell state. This decision is made by the sigmoid layer named "layer forget gate" with the following formulation.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

The second step is to determine the data (input gate) that will be stored in the cell state with the following formulation.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tan h\,(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

The third step is to update the old cell state, C_{t-1}, into a new cell state with the following formula.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

The fourth step aims to decide the output results with the following formula.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tan h\,(C_t) \quad (6)$$

where f_t is the forget gate, σ is the sigmoid function, W_f is the weight value for the forget gate, h_{t-1} is the output value before the t-order, x_t is the input value of the t-order, b_f is the bias value at forget gate, i_t is the input gate, b_i is the bias value at the input gate, $\tilde{C}_t$ is the new value that can be added to the cell state, $\tan h$ is the $\tan h$ function, W_c is the weight value for the cell state, b_c is the bias value in the cell state, C_t is the cell state, C_{t-1} is the cell state before the t-order, o_t is the output gate, W_o is the weight value for the output gate, b_o is the bias value at the output gate, and h_t is the output value of order-t.

In the LSTM method, there is a normalization and denormalization process which is formulated as follows.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (7)$$

$$x = x'(x_{max} - x_{min}) + x_{min} \quad (8)$$

where x' is the data after normalization, x is the actual, x_{min} is the minimum value of the whole data, and x_{max} is the maximum value of the whole data.

ARIMA is a time series data modelling and forecasting method developed by Box and Jenkins (1976). ARIMA is a method known for its strength and flexibility. Due to its strength and flexibility, ARIMA is a complex technique, complicated, not easy to use, and requires a lot of experience. Auto ARIMA is an ARIMA method to determine the order value of the model automatically (Jamil & Akbar, 2017). Auto ARIMA automatically handles the parameters of the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots, automatically determines the appropriate p, d and q

values and returns the best ARIMA model (Ito et al., 2022).

The ARIMA method works well if the time series data used are statistically dependent or related to each other. The form that shows the mixed model of the autoregressive moving average is as follows.

$$Z_t = \mu' + \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \quad (9)$$

where Z_t the forecast value at-t, Z_{t-p} is the data value at $t-p$, μ' is a constant value, ϕ_p is the p-th parameter autoregressive at-p, a_t is the error value at-t, θ_q is the q-th parameter moving average, and a_{t-q} is the error value at $t-q$.

In the ARIMA method, there is a white noise test using the L-Jung Box test and a normality test using the Jarque-Bera (Jarque & Bera, 1987). L-Jung Box test statistics (Brockwell & Davis, 2016) and Jarque-Bera are formulated as follows.

$$Q = n(n+2)\sum_{k=1}^{K}\frac{\hat{\rho}_k^2}{n-k} \quad (10)$$

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right) \quad (11)$$

where Q is the L-Jung Box, $\hat{\rho}_k$ is autocorrelation residual at the k-th lag, n is the number of data, K is the maximum number of lags, JB is Jarque-Bera, S is Skewness, K is Kurtosis. The significant level used is $\alpha = 0{,}05$. Residual data is not white noise if the $p-value < \alpha$. Residual data is not normally distributed if the $p-value < \alpha$.

The performance of the model in this study is indicated by the Akaike Information Criterion (AIC), coefficient of determination (R^2), Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The AIC equation is stated as follows (Box, Jenkins, Reinsel, & Ljung, 2016).

$$AIC = \ln\hat{\sigma}^2 + \frac{2}{n}r \quad (12)$$

where ln is the natural logarithm, $\hat{\sigma}^2$ is the RSS divided by the number of observations, n is the number of observations (residuals), and r is the number of parameters (including constants).

The equation R^2 is expressed as follows (Korstanje, 2021).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \quad (13)$$

where y_i is the i-th actual data, $\hat{y}_i$ is the i-th prediction data, $\bar{y}_i$ is the average of the data. The MSE, MAE, RMSE, and MAPE equations are expressed as follows.

$$MSE = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n} \quad (14)$$

$$MAE = \frac{\sum_{i=1}^{n}|x_i - y_i|}{n} \quad (15)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad (16)$$

$$MAPE = 100\% \times \frac{\sum_{i=1}^{n}\left|\frac{x_i - y_i}{y_i}\right|}{n} \quad (17)$$

Where x_i is the i-th prediction data, y_i is the i-th actual data, and n is the number of samples. The model with the lowest AIC value is the best. The closer to the value of 1, the better the value of R^2. The lowest value of MSE, MAE, RMSE, and MAPE is the best value.

METHOD

The data used are the closing price of the shares of the company PT XL Axiata Tbk with the code EXCL.JK, PT Indosat Tbk with the code ISAT.JK, and PT Telkom Indonesia Tbk with the code TLKM.JK in daily frequency with a time span of January 1, 2015, to December 31, 2020. Data was obtained from the website www.finance.yahoo.com. The data is inputted and checked to obtain the mean (average stock price), standard deviation (deviation of the stock price from the average), min (lowest stock price), max (highest stock price), lots of missing data, and plot data. In the next step, the data is analyzed using the LSTM and auto ARIMA methods, and the result visualization is shown.

The stages of the LSTM method are data preprocessing, model building, model selection, and forecasting. Stages of data preprocessing data include the distribution of training data by 80% and testing data by 20%, data normalization, determination of input and output patterns. In the build model, the values of two parameters that are the focus of research can be changed, namely epochs and neurons. The selected epochs in this study consisted of 10, 50, 100, and 200. The number of selected neurons in each hidden layer consisted of 50, 100, 150, and 200. In the evaluation of the model, a training loss value plot was used to view the model description. In model selection, the performance of the training data model and testing data is obtained. The best LSTM model was chosen based on the R^2 MSE, MAE, RMSE, and MAPE in the testing data. After the best LSTM model is determined, the data is denormalized to obtain forecasting data with true values.

The research steps use the auto ARIMA method, namely, data preparation, building models, and forecasting. The data preparation stage includes dividing the data into two, namely 80% training data and 20% testing data, and getting the ADF value from the training data. In the build model, the training data is processed using the auto ARIMA function in the AR(1)-AR(5) and MA(1)-MA(5) ranges, so that the best model will be obtained automatically based on the AIC value. The best model is applied to training and testing data to be evaluated. Next, the data testing is predicted with the selected model.

In the result visualization stage, the actual data, the forecasted data using the LSTM method, and the forecasting data using the auto ARIMA method are compared based on accuracy metrics in the form of R^2 and RMSE values, and the results of the three data plots in one figure will be shown.

RESULT AND DISCUSSION

The amount of data used by each company is 1505 data, starting on January 2, 2015, and ending on December 30, 2020. There is no missing value in the data. In EXCL stock data, the lowest share price is Rp. 1,410/share, with a standard deviation value which indicates the deviation of the stock price from the average of 623.3091. The highest share price is Rp 5,129,2319/share, while the mean share price is Rp 3,035,7977/share. For ISAT stock data, the lowest mean share price is Rp 4256.7641 with a standard deviation of 1,440.2789. The lowest share price is Rp 1,200/share and the highest share price is Rp 7,500/share. TLKM's highest share price was Rp. 4,800/share, with a standard deviation of 564.1894. The lowest share price in that period was Rp 2,560/share, while the mean share price was at Rp 3,664.7907/share. Here is a data plot of three stocks.

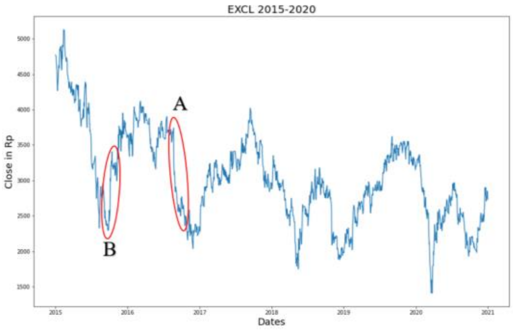

Figure 2. EXCL Stock Plot

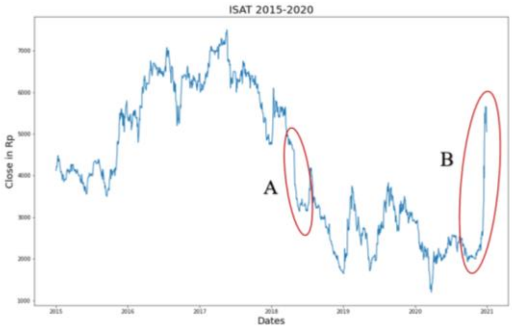

Figure 3. ISAT Stock Plot

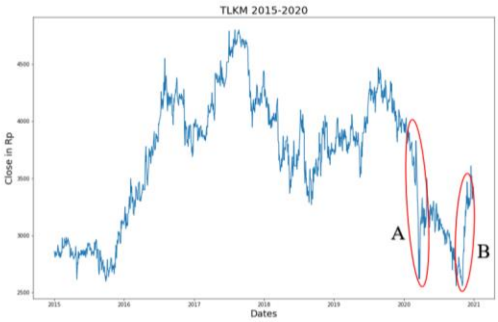

Figure 4. TLKM Stock Plot

Based on the three plots, the movement of the three stock prices is said to be volatile, experiencing an increasing trend and a decreasing trend. Event A is an event that the stock price experiences a downward trend, while event B is an event that the stock price experiences an upward trend. The training data consists of 1204 data, January 02, 2015 - October 04, 2019. The testing data consists of 301 data, from October 07 - December 30, 2020. The plot of the normalized training data and testing data can be shown as follows.

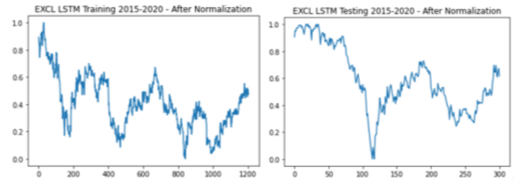

Figure 5. EXCL Stock Plot After Normalization

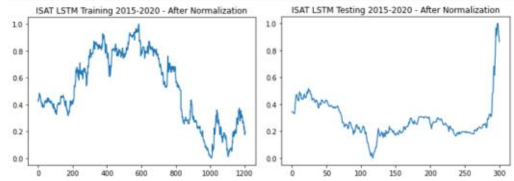

Figure 6. ISAT Stock Plot After Normalization

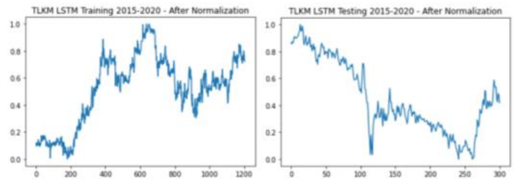

Figure 7. TLKM Stock Plot After Normalization

There is no standard rule in determining the optimal LSTM architecture to be applied to the system. So the search for LSTM architecture and parameters is done by trial and error. In this study, the LSTM network architecture consists of 1 input layer, 3 hidden layers, and 1 output layer. The predefined parameters are a batch size of 32, a sequence length of 100, and the Adam optimizer. The parameters that are used as the focus of this study are the parameters of the number of epochs and the number of neurons. The following is the loss training data value.

Table 1. Value of Stock Training Data Loss

Epoch	*Neuron*	*Loss* EXCL	ISAT	TLKM
10	50	0,0022	0,002	0,0029
	100	0,0019	0,0017	0,0026
	150	0,0021	0,0017	0,0022
	200	0,0016	0,0014	0,0023
50	50	0,00069014	0,000594	0,001100
	100	0,00068977	0,000462	0,000850
	150	0,00070927	0,000464	0,000968
	200	0,00066203	0,000436	0,000918
100	50	0,00064314	0,000450	0,000841
	100	**0,00061716**	0,000393	0,000993
	150	0,00062835	0,000388	0,000869
	200	0,00064685	0,000416	0,000821
200	50	0,00063619	**0,000341**	0,000845
	100	0,00072206	0,000392	0,000805
	150	0,00063033	0,000362	0,000851
	200	0,00062336	0,000383	**0,000790**

In the training data, the loss value of the model's performance is measured using the MSE metric. The smaller the value of the metric, the better the model. Based on the table, in the EXCL stock, the lowest loss value is 0,00061716 found in the model with epoch 100 and neuron 100. The lowest loss value for ISAT stock is 0,00034109 found in the model with epoch 200 and neuron 50. The lowest loss value for TLKM stock is 0,00078981 is found in the model with 200 epochs and 50 neurons. The following is a plot of the lowest training loss value for each stock.

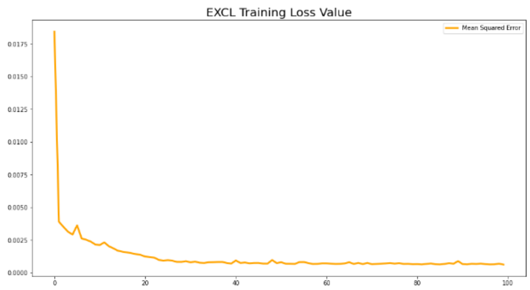

Figure 8. EXCL Stock Lowest Loss Value Training Plot

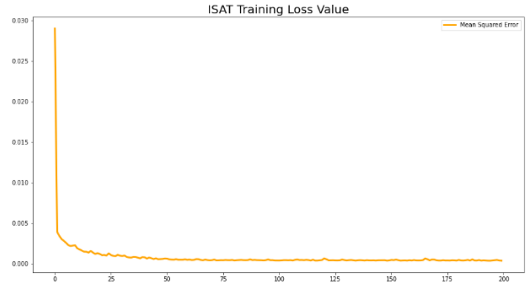

Figure 9. ISAT Stock Lowest Loss Value Training Plot

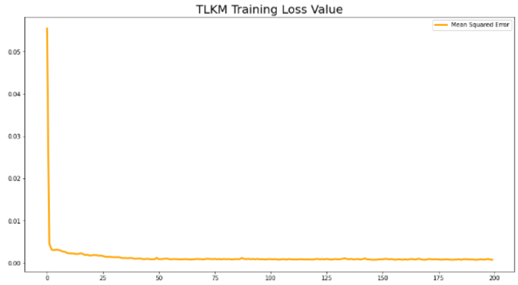

Figure 10. TLKM Stock Lowest Loss Value Training Plot

The next stage is model selection. Model selection is based on model performance. The following are the results of the performance of the LSTM model on training data and testing data.

Table 2. EXCL Stock Training Data Model Performance Results

Epoch	*Neuron*	R^2	*MSE*	*MAE*	*RMSE*	*MAPE*
10	50	-1,5002	718904,7655	811,4741	847,8825	0,2639
	100	-1,6085	750022,8145	828,5583	866,0386	0,2692
	150	-1,6914	773864,2344	842,3533	879,6955	0,2739
	200	-1,6350	757638,9639	834,3716	870,4246	0,2713
50	50	-1,3731	682338,8246	799,5860	826,0380	0,2618
	100	**-1,2667**	**651761,6986**	**781,0278**	**807,3176**	**0,2556**
	150	-1,5858	743507,4839	830,5900	862,2688	0,2708
	200	-1,2888	658100,4829	784,2549	811,2339	0,2565
100	50	-1,5284	726992,5536	825,0778	852,6386	0,2700
	100	-1,2980	660744,9958	785,1935	812,8622	0,2566
	150	-1,3119	664744,3096	788,3989	815,3185	0,2579
	200	-1,3097	664129,7002	788,3765	814,9415	0,2580
200	50	-1,4697	717890,4459	818,2620	847,2842	0,2673
	100	-1,3982	689549,8305	803,4029	830,3914	0,2629
	150	-1,2962	660233,5610	785,4830	812,5476	0,2569
	200	-1,4494	704293,3822	811,6155	839,2219	0,2654

Table 3. EXCL Stock Testing Data Model Performance Results

Epoch	*Neuron*	R^2	MSE	MAE	RMSE	MAPE
10	50	0,7473	25849,2862	120,7400	160,7771	0,0542
	100	0,7856	21932,7665	113,5803	148,0972	0,0500
	150	0,7470	25877,9104	124,6183	160,8661	0,0546
	200	0,8275	17647,9329	103,2634	132,8455	0,0448
50	50	0,9233	7848,5775	66,4853	88,5922	0,0290
	100	0,9181	8380,7094	72,0681	91,5462	0,0316
	150	0,9054	9677,5090	75,2054	98,3743	0,0320
	200	0,9218	8000,5358	69,5025	89,4457	0,0305
100	50	0,9168	8513,8145	67,8330	92,2703	0,0292

	100	0,9243	7739,1198	67,5242	87,9723	0,0295
	150	0,9230	7879,1604	67,0369	88,7646	0,0293
	200	0,9231	7867,0450	67,5217	88,6964	0,0295
200	50	0,9180	8387,0605	67,8585	91,5809	0,0293
	100	**0,9244**	**7731,8867**	**65,3989**	**87,9311**	**0,0285**
	150	0,9240	7778,2930	67,5133	88,1946	0,0295
	200	0,9218	8001,0474	65,9832	89,4486	0,0287

Table 4. ISAT Stock Training Data Model Performance Results

Epoch	*Neuron*	R^2	MSE	MAE	RMSE	MAPE
10	50	0,4300	1496438,7048	1141,6169	1223,2901	0,2385
	100	0,3751	1640703,6969	1200,9921	1280,8996	0,2520
	150	0,4725	1384939,3924	**1095,3047**	1176,8345	0,2881
	200	0,4462	1454065,3100	1136,0317	1205,8463	0,2399
50	50	0,3911	1598795,9347	1198,9154	1264,4350	0,2537
	100	0,3472	1714025,3984	1239,8114	1309,2079	0,2616
	150	0,4142	1538074,8477	1173,1669	1240,1915	0,2472
	200	**0,4889**	**1341878,7581**	1102,6390	**1158,3949**	**0,2349**
100	50	0,4299	1496691,6401	1161,1512	1223,3935	0,2459
	100	0,3944	1589954,3365	1200,3238	1260,9339	0,2553
	150	0,4426	1463437,8799	1139,4788	1209,7264	0,2385
	200	0,4150	1535978,4299	1175,9967	1239,3460	0,2489
200	50	0,4096	1549984,8842	1180,6282	1244,9839	0,2496
	100	0,4332	1488161,7976	1159,1779	1219,9024	0,2459
	150	0,4104	1548096,5762	1174,0349	1244,2253	0,2464
	200	0,3611	1677495,5829	1227,9043	1295,1817	0,2595

Table 5. ISAT Stock Testing Data Model Performance Results

Epoch	*Neuron*	R^2	MSE	MAE	RMSE	MAPE
10	50	0,6356	164182,1773	206,5357	405,1940	0,0826
	100	0,6690	149147,3974	197,6854	386,1961	0,0785
	150	0,6956	137164,7766	198,5814	370,3576	0,0816
	200	0,8046	88059,3671	163,1863	296,7480	0,0665
50	50	0,9309	31147,8511	103,4382	176,4875	0,0433
	100	0,9415	26383,5674	95,6374	162,4302	0,0405
	150	0,9454	24615,6714	92,0693	156,8938	0,0396
	200	0,9501	22497,6084	94,4158	149,9920	0,0400
100	50	0,9545	20508,1642	82,8675	143,2067	0,0344
	100	0,9509	22119,7908	85,6963	148,7272	0,0353
	150	0,9514	21903,5797	97,5179	147,9986	0,0414
	200	**0,9561**	**19791,6358**	**78,6629**	**140,6827**	**0,0322**
200	50	0,9549	20311,4093	79,2991	142,5181	0,0324
	100	0,9540	20738,0128	80,5866	144,0070	0,0326
	150	0,9547	20432,7853	84,1896	142,9433	0,0349
	200	0,9510	22082,2475	82,4872	148,6010	0,0336

Table 6. TLKM Stock Training Data Model Performance Results

Epoch	*Neuron*	R^2	MSE	MAE	RMSE	MAPE
10	50	0,6959	80518,9922	251,0566	283,7587	0,0630
	100	0,5496	119261,7886	312,6960	345,3430	0,0787
	150	0,6080	103780,0503	291,0551	322,1491	0,0732
	200	0,7352	70118,0229	235,3441	264,7981	0,0591
50	50	0,6051	104565,4958	298,7975	323,3659	0,0754
	100	0,6881	82576,9951	264,2548	287,3621	0,0666
	150	0,6578	90605,1296	277,7156	301,0069	0,0700
	200	0,6222	100018,1453	291,3294	316,2565	0,0734
100	50	**0,7648**	**62287,1622**	**229,1120**	**249,5740**	**0,0578**
	100	0,6658	88477,2721	272,9591	297,4513	0,0687
	150	0,6883	82533,9993	265,0725	287,2873	0,0669
	200	0,6743	86245,0864	270,5226	293,6751	0,0682
200	50	0,6726	86696,9653	270,2764	294,4435	0,0680
	100	0,6516	92241,1640	279,3788	303,7123	0,0703
	150	0,7262	72481,7037	248,8122	269,2243	0,0629
	200	0,7050	78100,9983	255,2163	279,4656	0,0641

Table 7. TLKM Stock Testing Data Model Performance Results

Epoch	*Neuron*	R^2	MSE	MAE	RMSE	MAPE
10	50	0,4721	34024,3284	128,7357	184,4568	0,0425
	100	0,5560	28617,0492	113,9858	169,1657	0,0372
	150	0,5848	26764,6887	113,8589	163,5992	0,0372
	200	0,5952	26092,1699	121,7668	161,5307	0,0400
50	50	0,8230	11410,1889	77,6602	106,8185	0,0252
	100	0,8573	9200,6588	73,0450	95,9201	0,0239
	150	0,8515	9571,5688	72,8726	97,8344	0,0238
	200	0,8668	8585,4551	68,9060	92,6577	0,0225
100	50	0,8521	9534,7933	76,2721	97,6463	0,0250
	100	0,8717	8271,0529	68,7337	90,9508	0,0225
	150	0,8779	7871,4239	65,9458	88,7210	0,0215
	200	0,8784	7835,1694	65,9514	88,5165	0,0215
200	50	0,8808	7684,9498	65,4947	87,6638	0,0214
	100	**0,8828**	**7554,9269**	**64,4258**	**86,9191**	**0,0210**
	150	0,8758	8006,7087	66,3482	89,4802	0,0216
	200	0,8755	8025,6590	67,8545	89,5860	0,0222

The best model in EXCL stock is the model with 50 epochs and 100 neurons for training data and the model with 200 epochs and 100 neurons for testing data which are marked with the best R^2, MSE, MAE, RMSE, and MAPE values. The best model is determined based on the value of the testing data, so the best LSTM model for EXCL stock is the model with 200 epochs and 100 neurons.

In ISAT stock training data, the best values of R^2, MSE, RMSE, and MAPE are found in models with 50 epochs and 200 neurons. The best MAE values best in training data are in models with 10 epochs and 150 neurons. In ISAT stock testing data The best values of R^2, MSE, MAE, RMSE, and MAPE are found in the model with 100 epochs and 200 neurons, so the best LSTM model for ISAT stock is the model with 100 epochs and 200 neurons.

The model with 100 epochs and 50 neurons is the best model for TLKM stock training data. The best model for testing TLKM stock data is a model with 200 epochs and 100 neurons is the best model, characterized by the best R^2, MSE, MAE, RMSE, and MAPE values. Based on the results of model performance on data testing, the best LSTM model for TLKM stocks is the model with 200 epochs and 100 neurons.

In the auto ARIMA method, the resulting ADF *p-value* is 0,0389 for EXCL stock, 0,661 for ISAT stock, and 0,3392 for TLKM stock. Based on the *p-value* of ADF, EXCL stock has a value of less than 0.05 and it can be concluded that the data is stationary.

ISAT and TLKM stocks can be concluded that the data is not stationary. In ARIMA, data that is not stationary will be differencing. However, in this study, auto ARIMA was carried out, so that no differencing will be carried out.

Table 8. AIC Results With Auto ARIMA Method

ARIMA	AIC		
	EXCL	ISAT	TLKM
(0,1,0)	**14090,0940**	**14578,1370**	13247,7090
(0,1,1)	14092,8800	14581,2270	13247,2390
(1,1,0)	14092,8970	14581,2270	13246,4630
(1,1,1)	14094,8560	14581,3670	13231,3760
(1,1,2)	-	-	13230,7440
(2,1,0)	-	-	13238,1130
(2,1,1)	-	-	**13230,7280**
(2,1,2)	-	-	13232,6920
(3,1,0)	-	-	13236,7450
(3,1,1)	-	-	13232,7220
(3,1,2)	-	-	13234,3120

Based on Table 8 auto ARIMA results, the ARIMA model (0,1,0) is the best model for EXCL stock, the ARIMA model (0,1,0) is the best model for ISAT shares, and the ARIMA model (2,1,1) be the best model for TLKM stock.

In the EXCL stock training data with the ARIMA(0,1,0) model, the results show that the AIC value is 14.090,094, the *p-value* of the coefficient is 0, the *p-value* $(Q) = 0{,}32$, and the *p-value* of (JB)=0. In the data testing, EXCL stock with ARIMA(0,1,0) model has an AIC value of 3.479,043. The *p-value* of the coefficient is 0, so the model is significant. Based on the value of *p-value* (Q) =0,29> α, so that the residual data in the model is declared white noise or the data is independent. The result of the *p-value* (JB) shows 0 < α, so it is stated that the residuals are not normally distributed. Thus, the assumption of normality is not met.

In the ISAT stock training data with the ARIMA(0,1,0) model, the results show that the AIC value is 14.578,137, the *p-value* of the coefficient is 0, *p-value* $(Q) = 0{,}38$, *p-value* $(JB) = 0$. In the ISAT stock testing data with the ARIMA(0,1,0) model, it has an AIC value of 3.767,603. The *p-value* of the coefficient is 0 < 0,05, so the model is concluded to be significant. Based on the *p-value* (Q) =0,67> α, the residual data in the model is declared white noise or the data is independent. The result of the *p-value* (JB) shows 0 < α , so it is stated that the residuals are not normally distributed. Thus, the assumption of normality is not met.

In TLKM stock training data with ARIMA model (2,1,1) the results show that the AIC value is 13230,728, *p-value* (Q) =0,98, *p-value* (JB)= 0. In the testing data, TLKM stock with the ARIMA model (2,1,1) has an AIC value of 3457,193. The AIC value in the testing data is lower than the AIC value in the training data. T The p-value of the coefficient shows 0 < 0,05, so the model is significant. Based on the value of *p-value* (Q) =0,96 > α, the residual data in the model is declared white noise or the data is independent. The result of the *p-value* (JB) shows 0 < α, so it is stated that the residuals are not normally distributed.

For forecasting using the LSTM method, denormalized data testing is used to predict, while the auto ARIMA method does not require a denormalization process. The following are the forecasting results for each stock on February 28, 2020, to December 30, 2020.

Table 9. EXCL Stock Forecast Results

Date	Close	Predictions LSTM	Predictions ARIMA
28-02-2020	2590	2438,728	2460
02-03-2020	2520	2591,122	2590
03-03-2020	2500	2505,359	2520
04-03-2020	2540	2478,003	2500
05-03-2020	2480	2528,38	2540

...	...	...	...
22-12-2020	2760	2882,424	2900
23-12-2020	2710	2750,292	2760
28-12-2020	2810	2702,726	2710
29-12-2020	2850	2815,633	2810
30-12-2020	2730	2861,51	2850

Table 10. ISAT Stock Forecast Results

Date	Close	Predictions LSTM	Predictions ARIMA
28-02-2020	2590	2438,728	2460
02-03-2020	2520	2591,122	2590
03-03-2020	2500	2505,359	2520
04-03-2020	2540	2478,003	2500
05-03-2020	2480	2528,38	2540
...	...	...	...
22-12-2020	2760	2882,424	2900
23-12-2020	2710	2750,292	2760
28-12-2020	2810	2702,726	2710
29-12-2020	2850	2815,633	2810
30-12-2020	2730	2861,51	2850

Table 11. TLKM Stock Forecast Results

Date	Close	Predictions LSTM	Predictions ARIMA
28-03-2020	3490	3473,679	3484,323
02-03-2020	3440	3491,992	3490,058
03-03-2020	3620	3442,521	3434,395
04-03-2020	3830	3608,323	3639,16
05-03-2020	3830	3794,767	3793,834
...	...	...	...
22-12-2020	3360	3506,32	3512,101
23-12-2020	3320	3357,289	3356,599
28-12-2020	3430	3328,662	3349,968
29-12-2020	3420	3434,771	3424,613
30-12-2020	3310	3418,471	3402,347

The plot of forecasting results for each stock is as follows.

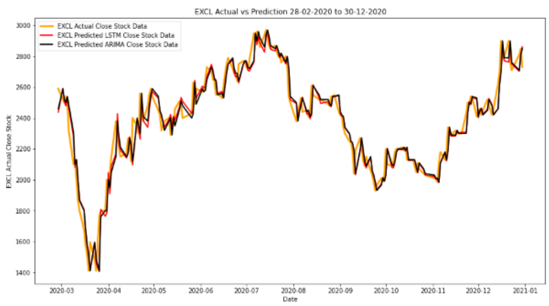

Figure 11. EXCL Stock Forecast Results Plot

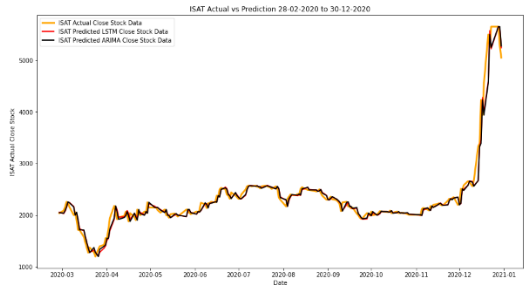

Figure 12. ISAT Stock Forecast Results Plot

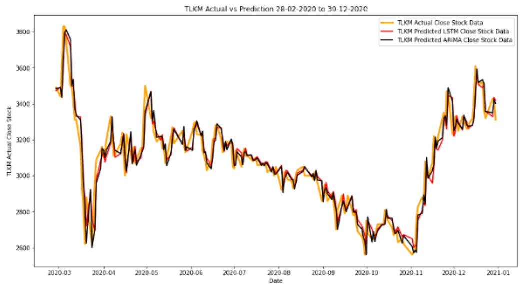

Figure 13. TLKM Stock Forecast Results Plot

In the comparison plot of Telkomsel stock data forecasting results using the LSTM and ARIMA methods, the orange plot represents the actual stock data. The red plot represents forecasting data using the LSTM method. In Figure 13, the black plot is the forecast data using the ARIMA method. The MSE and RMSE values of each share are as follows

Table 12. R^2 and RMSE Values

Saham	Metode	$\boldsymbol{R^2}$	RMSE
EXCL	LSTM	0,9244	87,9311
	ARIMA	**0,9259**	**87,0466**
ISAT	LSTM	**0,9561**	**140,6827**
	ARIMA	0,9550	142,4126
TLKM	LSTM	0,8828	86,9191
	ARIMA	**0,8885**	**84,7917**

Based on Table 12, the best R^2 and RMSE values are found in the auto ARIMA method for EXCL and TLKM stocks, and the LSTM method for ISAT stocks. Forecasting using the auto ARIMA method is better than the LSTM method on EXCL and TLKM stocks, and forecasting using the LSTM method is better than the auto ARIMA method on ISAT stocks.

CLOSING

In this study, time series modelling experiments were carried out in the case of stock data forecasting using the LSTM and auto ARIMA models. LSTM modeling with several variations of epoch parameters and many neurons, can give better results. In some cases, especially on data that can be approximated by a stationary model, the auto ARIMA model is better. Based on the experiments in our case, it was found that the LSTM method with 200 epochs and 100 neurons is the best model for EXCL and TLKM stocks, and the model with 100 epochs and 200 neurons is the best LSTM model for ISAT stocks. In the auto ARIMA method, the ARIMA model (0,1,0) is the best model for EXCL and ISAT shares, and the ARIMA model (2,1,1) is the best model for TLKM shares. Based on the results of the forecasting, it is concluded that the forecasting using the auto ARIMA method is better than the LSTM method on EXCL and TLKM stocks, and forecasting using the LSTM method is better than the auto ARIMA method on ISAT stocks. For further research, it is recommended to add parameters to the research focus on the LSTM method.

BIBLIOGRAPHY

Abdillah, M. F. P., Zukhronah, E., & Respatiwulan. (2021). Peramalan Harga Saham PT . Bank Central Asia Tbk Menggunakan Metode Auto Regresive Integrated Moving Average (ARIMA). *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2021*, 110–117.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis Forecasting And Control* (Fifth). Hoboken: John Wiley & Sons, Inc.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (third). https://doi.org/10.2307/2965440

Brownlee, J. (2017). *Long Short-Term Memory Networks With Python* (1.0, Vol. 1). Machine Learning Mastery.

Ho, M. K., Darman, H., & Musa, S. (2021). Stock Price Prediction Using ARIMA, Neural Network and LSTM Models. *Journal of Physics: Conference Series*, *1988*(1), 1–11. https://doi.org/10.1088/1742-6596/1988/1/012041

Ito, K., Iima, H., & Kitamura, Y. (2022). LSTM Forecasting Foreign Exchange Rates Using Limit Order Book. *Finance Research Letters*, *47*(PA), 1–6. https://doi.org/10.1016/j.frl.2021.102517

Jamil, M. S., & Akbar, S. (2017). Taxi Passenger Hotspot Prediction using Automatic ARIMA Model. *2017 3rd International Conference on Science in Information Technology (ICSITech)*, 23–28. https://doi.org/10.1109/ICSITech.2017.8257080

Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals.

International Statistical Review, *55*(2), 163–172.

Korstanje, J. (2021). Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR. In *Advanced Forecasting with Python*. https://doi.org/10.1007/978-1-4842-7150-6

Liu, P. (2022). Time Series Forecasting Based on ARIMA and LSTM. *Proceedings of the 2022 2nd International Conference on Enterprise Management and Economic Development (ICEMED 2022)*, *656*(Icemed), 1203–1208. https://doi.org/10.2991/aebmr.k.220603.195

Saeed, W. (2022). Frequency-Based Ensemble Forecasting Model for Time S eries Forecasting. *Computational and Applied Mathematics*, *41*(66), 1–17. https://doi.org/10.1007/s40314-022-01765-x

Sethia, A., & Raut, P. (2018). Application of LSTM, GRU and ICA for stock price prediction. *Information and Communication Technology for Intelligent Systems*, *2*, 479–487. https://doi.org/10.1007/978-981-13-1747-7_46

Shankar, P. (2022). Stock Price Prediction Using LSTM, ARIMA, and UCM. *Journal of Development Economics and Management Research Studies*, *09*(11), 55–66. https://doi.org/10.53422/jdms

Vani, V., Varma, N., Lakshman, & Krishna, G. (2019). Stocks Prediction Using Auto Arima & LSTM. *International Journal of Engineering and Advanced Technology (IJEAT)*, *8*(4), 1766–1769.

Wibowo, F. D., Dang, T.-T., & Wang, C.-N. (2022). Forecasting Indonesia Stock Price Using Time Series Analysis and Machine Learning in R. *Indonesian Scholars Scientific Summit Taiwan Proceeding*, *4*, 103–108. https://doi.org/10.52162/4.2022166

Winata, W. (2018). Prakira Suhu Udara Rata-Rata Kota-Kota Besar Dunia Menggunakan Metode Long Short-Term Memory (Universitas Sumatera Utara). Retrieved from http://repositori.usu.ac.id/handle/123456789/7341