

Forecasting Indonesia Stock Price Using Time Series Analysis and Machine Learning in R

¹*Fajar Dwi Wibowo, ¹Thanh-Tuan Dang, ¹Chia-Nan Wang.

¹Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan

*Correspondence: fajardwiwibowo@gmail.com

Abstract— This study investigated the appropriate model to predict 30 days ahead of Unilever Indonesia stock price and Telekomunikasi Indonesia stock price using time series analysis and machine learning in R, time series forecasting is a fun and interesting way to learn data science. The data is format Close Price. The goal of this project is to predict the future stock price of unilever indonesia and telekomunikasi indonesia using various predictive forecasting models and then analyze the various models. The dataset for unilever stocks is obtained from yahoo finance using Quantmod package in R. The final results that have been compared show that using the arima and neural network methods produces good accuracy values. Research and analysis of stock prices will help investors carry out investment is more accurate, investors can determine what steps will be taken, either buying a share or selling acquired shares the right step in taking an action. The data model used to predict close stock prices in this study unilever Indonesia using arima has an accuracy of 98.87%. and using neural network model has an of 98.92%. Telekomunikasi Indonesia using arima has an accuracy of 98.74%. and using neural network Model has an accuracy of 98.77% there are suggestions that can be given for further research and development. Trying to add to the existing historical data to be more complete so as to improve the accuracy of forecasting.

Keywords— Forecasting; Time Series; Arima; Neural Network.

I. INTRODUCTION

The prediction shows what will happen in a certain situation and is an input for the planning and decision-making process. Analysis and the right prediction can provide management and retrieval of reliable decisions for departments and investors. Research and analysis of stock prices will help investors carry out investment is more accurate, investors can determine what steps will be taken, either buying a share or selling acquired shares the right step in taking an action[1]. The stock market has proven to be an attractive target for investing unused money, much better than a bank account[2], [3]. In the stock market, investors show an interest in profit by investing a certain amount of money in the stock market. The stock market has shown the interest of investors due to the sophisticated applications by which predictions can lead to prosperous market forecasting.[4] A forecasting algorithm is an information process that seeks to predict future values based on past and present data. These historical data points are extracted and prepared trying to predict future values for a selected variety of the dataset[5]. In this project approach, we will focus on quantitative forecasting involving our variable to forecast (close price), statistical principles analysis and

advanced concepts applied to a given historical data [6]. The first is a weak stock market where historical market prices cannot be used for forecasting future behavior. The second is a semi-strong market where neither historical data nor public info is useful for forecasting. A strong stock market is one for which there is no historical, public or private data that can be used for the forecasting process. Subsequent studies refuted this claim and demonstrated that a powerful new artificial intelligence model can accurately predict market behavior and thus profits[7]. The prediction does not have to provide an answer surely events will happen, but rather try to find answers as close as possible will happen[8]. The capital market is a market for a variety of long-term tradable financial instruments, a means of funding for companies and other institutions, and as a vehicle for investment activities. The capital market has a very important role for the economy of a country[9]. The form of investment in the capital market is in the form of shares. Shares are a sign of an individual or party (business entity) capital participation in a company or limited liability company[10]. Investments in shares are considered to be very profitable in the future. However, before becoming an investor in the capital market there are many things to consider because being an investor will have its own advantages and risks[11].

II. METHODS

A. Database

In the first stage, literature study is a research activity in gathering information needed during the research work process and to propose solutions related to existing problems. Information gathering can be obtained through resource persons, books, and previous studies. The goal is that the author can know and understand the problems and become a theoretical basis in conducting studies and also become the basis for forecasting stock prices at Unilever Indonesia and Telekomunikasi Indonesia. In doing this research required data that supports and can be used in carrying out the research process of the. research. In this stage the data search will be carried out for this research. The data required is secondary data from the Yahoo Finance website[12]. This data is the share price data Unilever Indonesia and Telekomunikasi Indonesia, where the data is close price data for the period October 1, 2015 – October 1 2020. In this stage, the reliability and validity tests are carried out on the data used to determine whether the data obtained is suitable for use in forecasting.

TABEL 1. Data source information.

Stock Name	Stock Code	Stock Date
Unilever Indonesia Co., Ltd.	UNVR.JK	10/01/2015-10/01/2020
Telekomunikasi Indonesia Co., Ltd.	TLKM.JK	10/01/2015-10/01/2020

B. ARIMA Model Analysis Process.

Introducing ARIMA models comes from the assumption that we are not working with a non-stationary dataset series[9], [13]. That time-series datasets are stationary when their means, variance and autocovariance don't change during time. The majority of economic time series is not stationary, but differencing them determined number times makes them stationary[14]. With this previous operation we can apply arima models to any stock price. In general that a temporal set Y_t admits an integrated autoregressive representation with p , q and d moving average orders respectively[15]. Denote this forecasting model by ARIMA (p , d , q).

Based on AR (1) and MA (1), the following general forms will be obtained
ARMA(1,1)

$$y_t = \delta + \{\phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p}\} + \{\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}\} + \varepsilon_t$$

$$y_t = \delta + \sum_{i=0}^p \phi_i y_{t-i} + \sum_{j=0}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

If non-stationarity is added to the ARMA process mixture, then the general ARIMA AutoRegressive Integrated Moving Average model (p, d, q) is fulfilled. Simple equation for AR (1) MA (1) and differencing ARIMA are as follows:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

If non-stationarity is added to the ARMA process mixture, then the general ARIMA AutoRegressive Integrated Moving Average model (p, d, q) is fulfilled. Simple equation for AR (1) MA (1) and differencing ARIMA are as follows:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

C. Research Model with ANN.

Artificial Neural Network (ANN) is an intelligent system used to process information which is the development of a generalized mathematical model. ANN's working principles are inspired by the working principles of the human neural network. Scientists create mathematical algorithms that work in these neurons, so the name Artificial Neural Network(ANN)[16][17][18].

III. RESULTS AND DISCUSSION

A. Plot Identification

The data plot appears to contain a trend element, which means that the data is not stationary invariance. Then it can be revealed that the variant of Time series data is not constant and changes from one period to another and there is a possibility that it contains elements of heteroscedasticity. Therefore a stationary test can be done first.

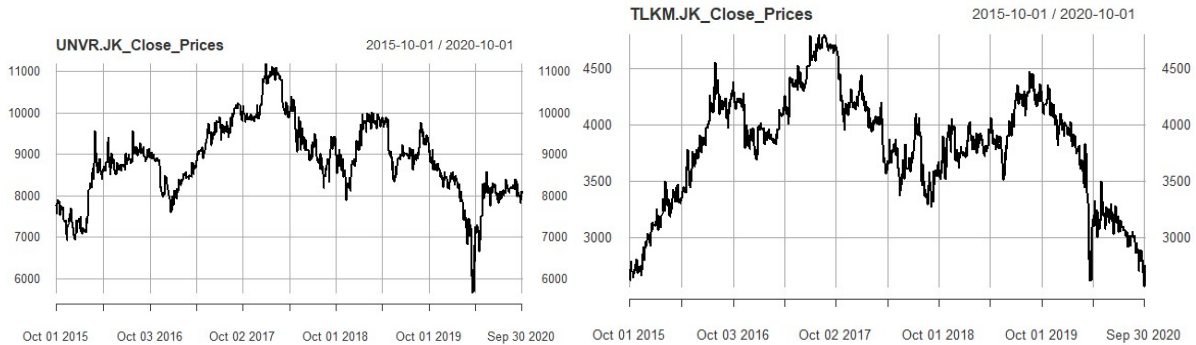


FIGURE 1. TIME SERIES PLOT OF UNILEVER AND TELEKOMUNIKASI INDONESIA FROM

B. Analyse ACF and PACF

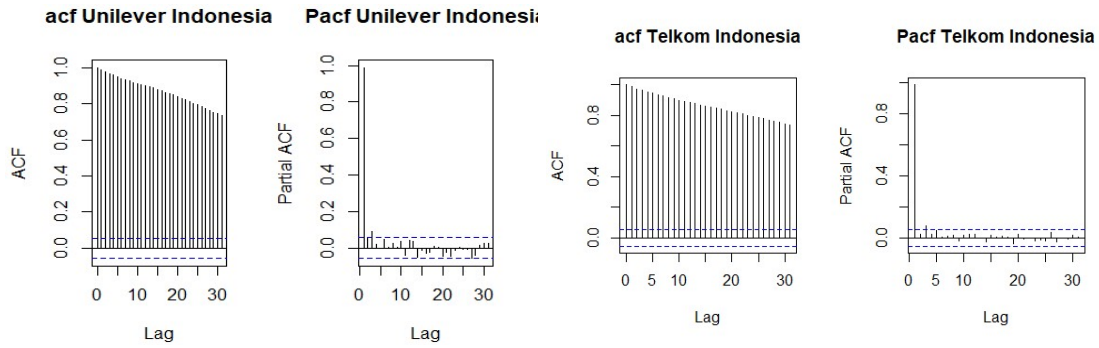


FIGURE 2. ACF AND PACF

C. Forecasting arima

Fitting our forecasting function integrated in the forecast package, we can now plot our forecast for the next 30 days.

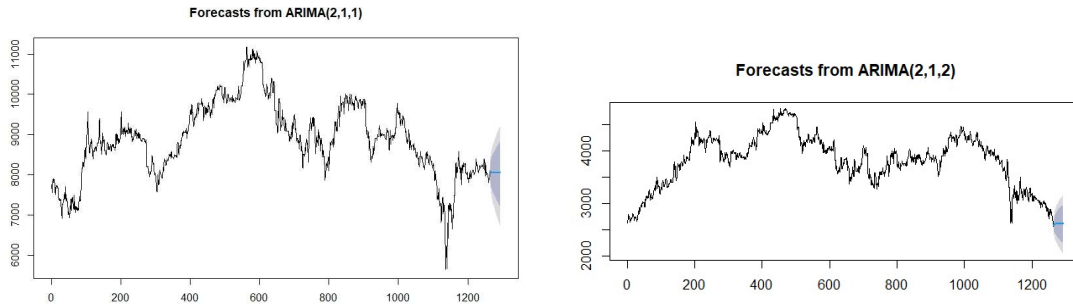


FIGURE 3. PREDICTION USING ARIMA

The blue portion of the graph represents the values generated by the forecast() function using the ARIMA(2,1,1) model for Unilever Indonesia and using the ARIMA(2,1,2) model for Telekomunikasi Indonesia

D. Artificial Neural Network

The first step is to determine the number of hidden layers for our neural network. Although, there is no specific method for calculating the number of hidden layers, the most common approach followed for timeseries forecasting is by calculating is using the formula.

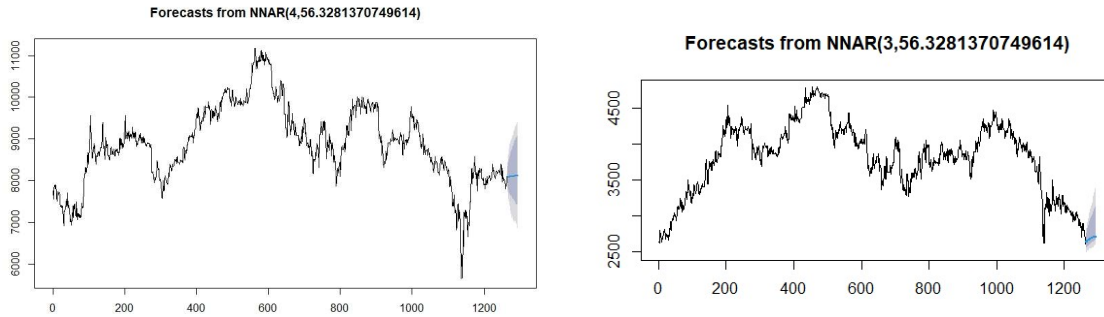


FIGURE 4. PREDICTION USING NEURAL NETWORK

After developing arima and neural network show in this section the main models result. For this approach we will start with the auto.arima accuracy values:

TABLE 2. Result Comparison arima and neural network

Stock Name	Model	RMSE	MAE	MAPE	Accuracy
Unilever Indonesia Co., Ltd.	arima	140.07	97.74	1.13	98.87%
	neural network	129.63	93.78	1.08	98.92%
Telekomunikasi Indonesia Co., Ltd.	arima	64.72	46.54	1.26	98.74%
	neural network	63.18	45.77	1.23	98.77%

From the above summary of model performance parameters, we see unilever indonesia using arima has an accuracy of 98.87%. and using Neural Network Model has an of 98.92%. telekomunikasi indonesia using arima has an accuracy of 98.74%. and using neural network model has an accuracy of 98.77%.

IV. CONCLUSION

This study investigated the appropriate model to predict 30 days ahead of unilever Indonesia stock price and telekomunikasi indonesia stock price using arima and neural network given same good result. This model was then used to make future predictions and produce its 80% and 95% prediction interval under the assumptions that errors are independently and normally distributed. There are so many factors that are at work in the makeup of the value of a stock that it is highly variable and this method is for learning and should be taken in a directional, not measurable manner. However, time series forecasting is a fun and interesting way to learn data science. This is not financial advice and it is recommended that a licensed financial advisor or similar be obtained for proper financial advice. The sole purpose here is for learning time series analysis and forecasting.

REFERENCES

- [1] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," *Expert Systems with Applications*, vol. 184, p. 115537, Dec. 2021, doi: 10.1016/J.ESWA.2021.115537.

- [2] K. Arias-Calluari, F. Alonso-Marroquin, M. N. Najafi, and M. Harré, “Methods for forecasting the effect of exogenous risks on stock markets,” *Physica A: Statistical Mechanics and its Applications*, vol. 568, p. 125587, Apr. 2021, doi: 10.1016/J.PHYSA.2020.125587.
- [3] D. Kumar, P. K. Sarangi, and R. Verma, “A systematic review of stock market prediction using machine learning and statistical techniques,” *Materials Today: Proceedings*, vol. 49, pp. 3187–3191, Jan. 2022, doi: 10.1016/J.MATPR.2020.11.399.
- [4] A. Sharma, D. Bhuriya, and U. Singh, “Survey of stock market prediction using machine learning approach,” in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, 2017, vol. 2, pp. 506–509. doi: 10.1109/ICECA.2017.8212715.
- [5] W. E. Zhang, R. Chang, M. Zhu, and J. Zuo, “Time Series Visualization and Forecasting from Australian Building and Construction Statistics,” *Applied Sciences*, vol. 12, no. 5, p. 2420, Feb. 2022, doi: 10.3390/app12052420.
- [6] P. J. & R. A. Davis. Brockwell, *Introduction to Time Series and Forecasting (2nd ed)*. New York: Springer, 2002.
- [7] F. Xu and V. Keelj, “Collective Sentiment Mining of Microblogs in 24-Hour Stock Price Movement Prediction,” in *2014 IEEE 16th Conference on Business Informatics*, 2014, vol. 2, pp. 60–67. doi: 10.1109/CBI.2014.37.
- [8] Z. Jian, X. Li, and Z. Zhu, “Sequential forecasting of downside extreme risk during overnight and daytime: Evidence from the Chinese Stock Market☆,” *Pacific-Basin Finance Journal*, vol. 64, p. 101454, Dec. 2020, doi: 10.1016/J.PACFIN.2020.101454.
- [9] E. Dave, A. Leonardo, M. Jeanice, and N. Hanafiah, “Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM,” *Procedia Computer Science*, vol. 179, pp. 480–487, Jan. 2021, doi: 10.1016/J.PROCS.2021.01.031.
- [10] W. Kristjanpoller R. and K. Michell V., “A stock market risk forecasting model through integration of switching regime, ANFIS and GARCH techniques,” *Applied Soft Computing*, vol. 67, pp. 106–116, Jun. 2018, doi: 10.1016/J.ASOC.2018.02.055.
- [11] S. P. Chatzis, V. Siakoulis, A. Petropoulos, E. Stavroulakis, and N. Vlachogiannakis, “Forecasting stock market crisis events using deep and statistical machine learning techniques,” *Expert Systems with Applications*, vol. 112, pp. 353–371, Dec. 2018, doi: 10.1016/J.ESWA.2018.06.032.
- [12] “Yahoo Finance,” <http://finance.yahoo.com>.
- [13] M. Matyjaszek, P. Riesgo Fernández, A. Krzemień, K. Wodarski, and G. Fidalgo Valverde, “Forecasting coking coal prices by means of ARIMA models and neural networks, considering the transgenic time series theory,” *Resources Policy*, vol. 61, pp. 283–292, Jun. 2019, doi: 10.1016/J.RESOURPOL.2019.02.017.
- [14] R. Efendi, N. Arbaiy, and M. M. Deris, “A new procedure in stock market forecasting based on fuzzy random auto-regression time series model,” *Information Sciences*, vol. 441, pp. 113–132, May 2018, doi: 10.1016/J.INS.2018.02.016.
- [15] Sima Siami, Namini, Neda Tavakoli, and Akbar Siami Namin, “A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM,” *CoRR*, vol. abs, no. 1911.09512, 2019.
- [16] J. Horák and T. Krulický, “Comparison of exponential time series alignment and time series alignment using artificial neural networks by example of prediction of future development of stock prices of a specific company,” *SHS Web Conf.*, vol. 61, 2019, [Online]. Available: <https://doi.org/10.1051/shsconf/20196101006>
- [17] L. J. Cao and F. E. H. Tay, “Support vector machine with adaptive parameters in financial time series forecasting,” *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1506–1518, 2003, doi: 10.1109/TNN.2003.820556.
- [18] S. Chen and H. He, “Stock Prediction Using Convolutional Neural Network,” *IOP Conference Series: Materials Science and Engineering*, vol. 435, p. 012026, 2018, doi: 10.1088/1757-899x/435/1/012026.