Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 1 Issue 2 December, 2024, pp. 79-88
https://idjams.com/index.php/jms/index

Forecasting Stock Price of PT. GoTo Gojek Tokopedia Tbk (GOTO.JK) using ARIMA Box-Jenkins Method

Julian Saputra

Study Program of Actuarial Science, School of Business, President University, Bekasi, 17550, Indonesia
Corresponding author: julian.saputra@student.president.ac.id

Copyright: ©2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License-(CC-BY-SA) (https://creativecommons.org/licenses/by-sa/4.0/)

Article Info

Article history:
Received: October, 2024
Revised: October, 2024
Accepted: November, 2024
Available online: December 2024

Abstract

PT. GoTo Gojek Tokopedia Tbk, formed through the merger of Indonesian tech giants Gojek and Tokopedia, is a leading digital technology company. It offers diverse services including logistics, financial services, e-commerce, and food delivery. Operating both domestically and internationally, it continues to innovate itsproducts and services, playing a pivotal role in shaping the digital economy landscape in Indonesia and beyond. Will use ARIMA (*p,d,q*) technical analysis to predict the stock price of The forecasting of stock price of PT. GoTo Gojek Tokopedia Tbk for the next 10 weeks from March 06, 2023 to May 8, 2023. The researchers used PT. GoTo Gojek Tokopedia Tbk historical stock price data from April 11, 2022 to February 27, 2023. Based on the test result of all the ARIMA Models, the best model that the researcher got is Model 3 ARIMA (2,1,2). This model is taken as the best model because the model has the smallest MSE Valueof 106.36830; the RMSE Value of 0.07845252; the MAE value of 8.07106; and thesmallest MAPE Value of 7.84525%. It shows that that the stock price is predicted torise for the next 10 weeks.

Keywords: ARIMA; Forecasting; Gojek Tokopedia; Stationary; Time Series Analysis

I. Introduction

The currency market is one of the most dynamic sectors of the economy. One common investment tool used in the currency market is the stock price, which represents the ownership stake in a company. This makes it the single best way for investors to participate in the growth and development of the aforementioned company. Over the last several years, Indonesia's technology and startup industries have experienced significant growth. Companies like as Gojek and Tokopedia, which formerly operated independently, were forced to merge to become PT. GoTo Gojek Tokopedia. Operating in many sectors such as ride-hailing, e-commerce, and digital payments, this company has become the leading player in the Indonesian technological ecosystem. As a result, the price of PT. GoTo Gojek Tokopedia stocks has a significant dampening effect on the Indonesian stock market [1].

The standard of living for individuals has greatly increased along with the economy's ongoing rise. As a result, more people are increasingly investing their spare money in stocks, viewing them as a tool for boosting wealth [2]. The analysis of stock prices is one method used to predict the movement of stock prices over time. One method that may be used in this analysis is the ARIMA method (Autoregressive Integrated Moving Average). This method is a time series method based on the assumption that the observed data has a characteristic polynomial that can be used to predict future values. Because it uses the data from the previous period to produce accurate short term estimation, this method works well for this purpose. While ARIMA is quite effective in short term estimation, its effectiveness in long term estimation is only somewhat good [2]. Based on this, the study will be conducted every 10 weeks to determine the price of GOTO.JK stocks.

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 1 Issue 2 December, 2024, pp. 79-88
https://idjams.com/index.php/jms/index

The ARIMA models are commonly known as Box-Jenkins models since Box and Jenkins were the ones who initially popularized the ARIMA technique. Box and Tiao [3] discussed the generic transfer function model used in the ARIMA technique [3]. The distinction lies in the fact that the ARIMA Model first transforms nonstationary data into stationary data before analyzing it. The ARIMA model is commonly used for predicting linear time series data [4]. As a discrete time series model, stock prices may be regarded as non-random values and their trends can be evaluated and hence predicted. Stock forecasting has several purposes. Efficient decision-making for investors and financial professionals can be facilitated by a system that can discern which firms are performing well in the volatile stock market and which ones are not [4]. As a result, the main objective of this work is to forecaststock prices through the use of ARIMA to convert the time series and the estimated error value to assessthe effectiveness of the performance model. The theoretical frameworks of autoregressive (AR) and moving average (MA) models were developed by Yule et al. [5].

The motivation for utilizing time series forecasting, particularly the ARIMA model, lies in its ability to effectively capture historical data patterns for short-term predictions. This approach is well-suited for the 10-week forecast intervals in this study. Unlike predictive distribution, which relies on probabilistic models and assumptions, ARIMA provides a data-driven method that offers practical insights for dynamic stock markets. While its accuracy may diminish over longer periods, ARIMA remains a robust tool for reliable short-term forecasting, aligning with the objectives of this research.

II. Method

A. Time Series

Time series analysis is a special way of analyzing the sequence of data points collected over a certain time interval. In time series analysis, the analyst records data points at consistent intervals over a period of time rather than simply recording data points intermittently or randomly. One of the time series analysis models is the Autoregressive Integrated Moving Average (ARIMA) [6].

B. Stochastic Process

As a model for the observed time series, the random variable sequence $\{Y_t: t = 0, \pm 1, \pm 2, \dots\}$ is referred to as. A mathematical model that provides a quantitative explanation of natural occurrences is known as a stochastic process or stochastic model. Greek words for "random or chance" are the source of the word "stochastic." Among the most significant distinctions between deterministic and stochastic models is that the former can predict a set of possible outcomes weighted by their probabilities, or probabilities, while the latter can only give a probability of an outcome from a given set of circumstances. To achieve the most realistic stochastic model, additional analytical and computational techniques are utilized. These include the computation of probability histograms, estimation of extinction probability, numerical simulation of sample trajectories, and study of moments such as the distribution's mean and variance. The mean function for a stochastic process $\{Y_t: t = 0, \pm 1, \pm 2, \dots\}$, is defined by

$$\mu_t = E(Y_t) \text{ for } t = 0, \pm 1, \pm 2, \dots, \tag{1}$$

μ_t just represents the process's anticipated value at time t. In general terms, μ_t can vary at any given moment. Autocovariance function $\gamma_{t,s}$ is defined as

$$\gamma_{t,s} = cov(\gamma_t, \gamma_s) E(\gamma_t, \mu_t) \tag{2}$$
$$\gamma_{t,s} = E[(\gamma_t, \mu_t)(\gamma_s, \mu_s)]. \tag{3}$$

Autocorrelation function $\rho_{t,s}$ is defined as

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 1 Issue 2 December, 2024, pp. 79-88
https://idjams.com/index.php/jms/index

$$\rho_{t,s} = corr(Y_t, Y_s), \text{for } t, s = 0, \pm 1, \pm 2, \ldots \quad (4)$$

$$corr(Y_t, Y_s) = \frac{cov(Y_t, Y_s)}{\sqrt{\text{Var}(Y_t)\,\text{Var}(Y_s)}}. \quad (5)$$

C. Autoregressive (AR)

Simply put, autoregressive processes are regressions on themselves. More specifically, Equation 6 is saisfied by a pth-order autoregresive process $\{Y_t\}$ [7].

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + e_t \quad (6)$$

D. Moving Average (MA)

A moving average process is what we get when only a finite number of the ψ-weights are nonzero. In this scenario, we tweak notation slightly and write [8].

$$Y_t = e_t - \theta_1 e_{t-1} + \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (7)$$

E. Autoregressive Integrated Moving Average (ARIMA)

In fact, it is common to find that economic data is non-stationary, necessitating its modification by differentiating it in order to obtain stationary data. Subtraction of the value in one period from the value in the preceding period is used to make the distinction. The transformation's already-stationary data, not the initial data, is what the ARIMA model uses as its input. The differencing procedure is repeated several times, as represented by d. The ARIMA model is commonly represented by the notation ARIMA (p, d, q), indicating that it makes use of q residual lags, d levels of differentiation, and p dependent lag values. Flowchart at Figure 1 shows a summary of all the stages involved in predicting using the ARIMA model.

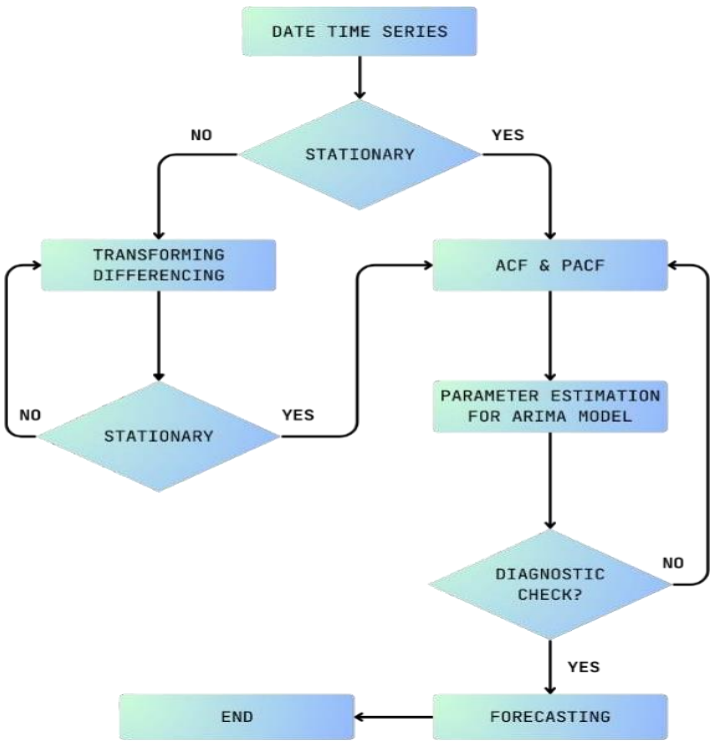


Figure 1. ARIMA Box-Jenkins Flowchart

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 1 Issue 2 December, 2024, pp. 79-88
https://idjams.com/index.php/jms/index

F. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

The covariance and correlation between x_t and x_{t+k} using the same process in time series analysis. For the time series $x(t), t = 0,1,2, \ldots$ Autocovariance γ_k, is defined by

$$\gamma_k = cov(x_t, x_{t+k}) = E[(x_t - \mu)(x_{t+k} - \mu)]. \tag{8}$$

The autocorrelation function (ACF), ρ_k, can be described as

$$\rho_k = \frac{cov(x_t, x_{t+k})}{\sqrt{var(x_t)}\sqrt{var(x_{t+k})}} = \frac{\gamma_k}{\gamma_0}. \tag{9}$$

The partial autocorrelation function (ACF), $\Phi_{k,k}$, can be described as

$$\Phi_{k,k} = \frac{\rho_k - \sum_{j=1}^{k-1} \Phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \Phi_{k-1,\mathrm{j}}\rho_j} \tag{10}$$

,

where $\Phi_{k,j} = \Phi_{k,-1.j} - \Phi_{kk}\Phi_{k-1,k-j}$ for $j = 1,2,3, \ldots., k-1$

G. Forecasting

The process of forecasting involves making predictions about the future values of a time series using past observations. In order to forecast future values of the time series, it entails first utilizing statistical techniques to find patterns and trends in previous data. Providing accurate and trustworthy projections of future values for planning, decision-making, and other uses is the aim of forecasting. It is a critical decision-making tool in many fields, including economics, finance, and business [9].

H. Estimated Error-Value

The term Estimated Error Value (EEV) is used in measurement and metrology to describe the uncertainty surrounding a measurement result. The EEV is a standard deviation or confidence interval that estimates the difference between the measured and real value of a quantity.

III. Result and Discussion

In this study, the stock prices of PT. GoTo Gojek Tokopedia Tbk (GOTO.JK) were analyzed for the period from April 11, 2022, to May 8, 2023. A thorough examination of the data distribution provides essential insights into the characteristics of the stock prices prior to modeling with the ARIMA approach.

A. Data Preparation

The data used in this study is PT. GoTo Gojek Tokopedia Tbk (GOTO.JK) weekly stock price period April 11, 2022 to May 8, 2023.

Table 1. PT. GoTo Gojek Tokopedia, April 11, 2022 – May 8, 2023

Data	Close	Date	Close	Date	Close
4/11/2022	376	8/22/2022	318	1/2/2023	95
4/18/2022	340	8/29/2022	284	1/9/2023	105
4/25/2022	272	9/5/2022	282	1/16/2023	114
5/2/2022	272	9/12/2022	248	1/23/2023	117
5/9/2022	194	9/19/2022	264	1/30/2023	124
5/16/2022	304	9/26/2022	246	2/6/2023	105
5/23/2022	312	10/3/2022	244	2/13/2023	125

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 1 Issue 2 December, 2024, pp. 79-88
https://idjams.com/index.php/jms/index

Data	Close	Date	Close	Date	Close
5/30/2022	354	10/10/2022	202	2/20/2023	120
6/6/2022	386	10/17/2022	200	2/27/2023	118
6/13/2022	366	10/24/2022	189	3/6/2023	125
6/20/2022	392	10/31/2022	200	3/13/2023	106
6/27/2022	370	11/7/2022	210	3/20/2023	115
7/4/2022	350	11/14/2022	222	3/27/2023	109
7/11/2022	320	11/21/2022	185	4/3/2023	101
7/18/2022	308	11/28/2022	132	4/10/2023	92
7/25/2022	296	12/5/2022	93	4/17/2023	95
8/1/2022	296	12/12/2022	96	4/24/2023	104
8/8/2022	292	12/19/2022	86	5/1/2023	106
8/15/2022	334	12/26/2022	91	5/8/2023	106

B. Stationary Check

Prior to using the data, we must use R studio and the Augmented Dicky Fuller Test technique to validatethe data for stationarity. It indicates that the data is not yet stationary if the p-value is higher than 0.05. The p-value for it is 0.08537, as shown in Figure 2 below. Thus, we may conclude that it is not stationary.

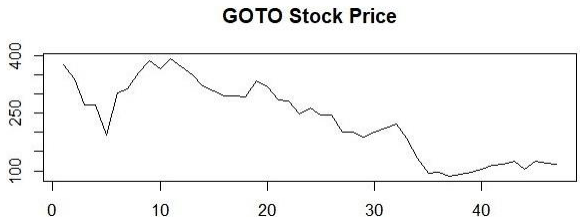


Figure 2. Before Differencing

Given that the p-value is higher than 0.05, we must perform a data differencing in order to obtain the stationary value. We apply another Augmented Dicky Fuller Test on the data. As seen in Figure 3, we have a pvalue of 0.01, which is still less than 0.05. Therefore, we may assume that it is stationary.

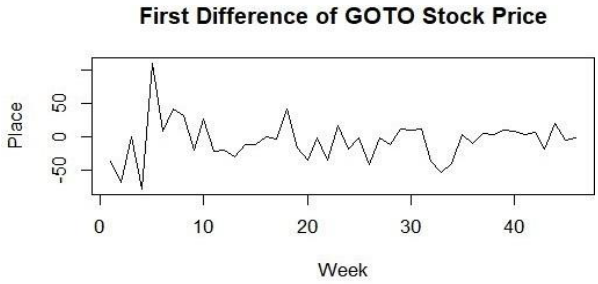


Figure 3. After First Differencing

C. Model Specification

We may obtain the specific model for our data by identifying the p and q values, which indicate the order of AR and MA, respectively. Figures 4 and 5 below provide the values for p and q.

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 1 Issue 2 December, 2024, pp. 79-88
https://idjams.com/index.php/jms/index

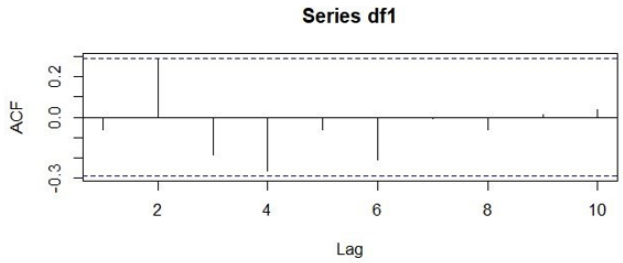


Figure 4. PACF Plot

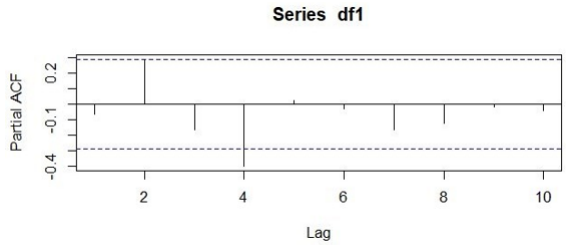


Figure 5. ACF Plot

Figure 4 shows that the PACF cut off occurs at lag time 2, which leads us to conclude that $p = 2$. The value of q is 2, as we can infer from Figure 5 that the cutoff for ACF is at lag time 2. Because we went through the difference procedure twice in the previous section before reaching the stationary, the value of d is 2. Table 2 presents the nine ARIMA models that we obtain with the ARIMA (2,1,2).

Table 2. ARIMA Model Specification

No	Model ARIMA	p	d	q
1	ARIMA (2,1,0)	2	1	0
2	ARIMA (2,1,1)	2	1	1
3	ARIMA (2,1,2)	2	1	1
4	ARIMA (1,1,0)	1	1	0
5	ARIMA (1,1,1)	1	1	1
6	ARIMA (1,1,2)	1	1	2
7	ARIMA (0,1,0)	0	1	0
8	ARIMA (0,1,1)	0	1	1
9	ARIMA (0,1,2)	0	1	2

D. Parameter Estimation

Based on Table 3 we know that ARIMA (2,1,2) has the smallest MSE value and ARIMA (0,1,2) has the smallest AIC value.

Table 3. Parameter Estimation for ARIMA Model

ARIMA Model	Coefficient of Estimation Result						
	AR1	AR2	MA1	MA2	MSE	Log likelihood	AIC
ARIMA (2,1,0)	-0.0032	0.3317			841.1508	-220.78	447.56
ARIMA (2,1,1)	-0.1241	0.3359	0.1375		835.4262	-220.63	449.26

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 1 Issue 2 December, 2024, pp. 79-88
https://idjams.com/index.php/jms/index

ARIMA Model	Coefficient of Estimation Result						
	AR1	AR2	MA1	MA2	MSE	Log likelihood	AIC
ARIMA (2,1,2)	-1.2049	0.7041	1.1680	0.9999	682.3681	-218.06	446.12
ARIMA (1,1,0)	-0.0327				938.8114	-223.19	450.38
ARIMA (1,1,1)	0.45060		0.4110		938.1013	-223.18	452.35
ARIMA (1,1,2)	-0.1768		0.2145	0.5720	738.6931	-218.08	444.16
ARIMA (0,1,0)					939.8326	-223.22	448.43
ARIMA (0,1,1)			-0.0206		939.1988	-223.2	450.40
ARIMA (0,1,2)			0.103	0.5521	749.7159	-218.38	442.77

E. Residual Analysis

The Shapiro test and the Ljung-Box test can be performed with R-Studio. A statistical technique calledthe Shapiro-Wilk test is used to determine whether or not the data under consideration is regularly distributed. The Shapiro test accepts the null hypothesis, which asserts that the population is normally distributed, if the p-value is higher than 0.05. If the p-value is less than or equal to 0.05, the null hypothesis is rejected, supporting the alternative hypothesis that the population is not regularly distributed. Table 4 shows the test results for each of the models. The table indicates that although model 9 ARIMA (0,1,2) has a low AIC value, it fails the Shapiro and Ljung-Box tests. On the other hand, model 3 ARIMA (2,1,2) pass the test for Shapiro and Ljung-Box Test.

Table 4. Parameter Estimation for ARIMA Model

Model ARIMA	Shapiro Test	Ljung-Box Test	AIC	Result
ARIMA (2,1,0)	0.001686	0.9329	447.56	Not Passed
ARIMA (2,1,1)	0.002376	0.8826	449.26	Not Passed
ARIMA (2,1,2)	0.1058	0.2665	446.12	Passed
ARIMA (1,1,0)	0.01056	0.8556	450.38	Not Passed
ARIMA (1,1,1)	0.003762	0.4308	452.35	Not Passed
ARIMA (1,1,2)	0.005886	0.773	444.16	Not Passed
ARIMA (0,1,0)	0.006913	0.6335	448.43	Not Passed
ARIMA (0,1,1)	0.009029	0.7704	450.4	Not Passed
ARIMA (0,1,2)	0.002043	0.4973	442.77	Not Passed

The analysis of residuals in Model 3 is crucial for assessing the model's fit to the data. Through QQ plots, ACF Plot, and PACF plot and to assess the residuals' normality and serial correlation. These analyses provide critical insights into model reliability and predictive accuracy, as shown in Figures 6 and 7.

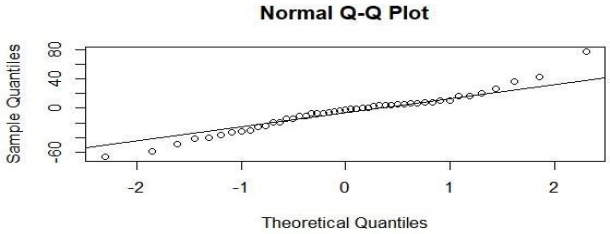


Figure 6: Normal QQ-Plot of ARIMA (2,1,2)

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 1 Issue 2 December, 2024, pp. 79-88
https://idjams.com/index.php/jms/index

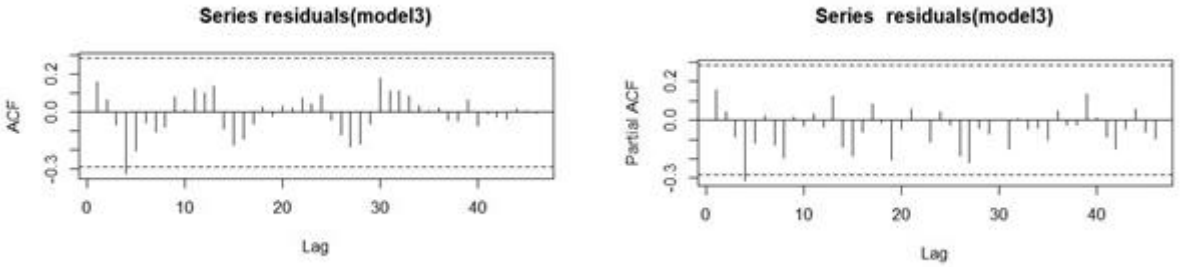


Figure 7. Plot of ACF and PACF of ARIMA (2,1,2)

F. Forecasting

Forecasting the Stock price of PT. GoTo Gojek Tokopedia Tbk for the next 10 weeks, using ARIMA (2,1,2) are presented Figure 4 and Figure 5.

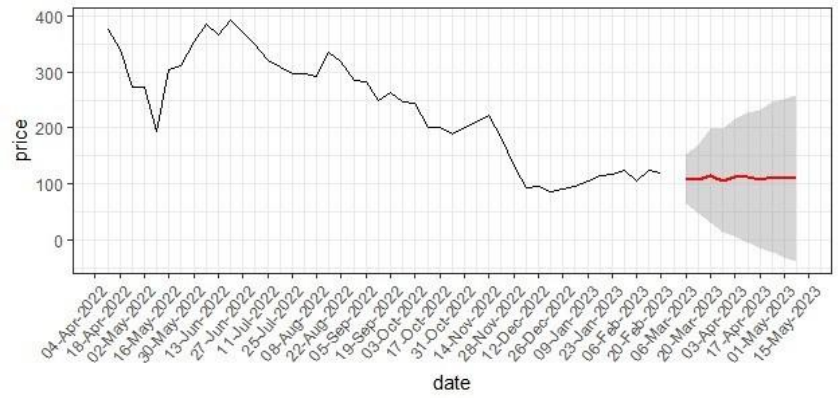


Figure 8. Forecast of ARIMA (2,1,2)



Figure 9: Forecast VS Actual Stock Price of PT. GoTo Gojek Tokopedia

Table 5. Error Measures for Forecasting Using ARIMA (2,1,2)

	ARIMA (2,1,2)
MSE	106.368301
RMSE	0.07845252
MAE	8.07106
MAPE	7.84525%

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 1 Issue 2 December, 2024, pp. 79-88
https://idjams.com/index.php/jms/index

From Table 5, we can see that the error value of ARIMA (2,1,2) has the smallest value overall. Therefore, we choose ARIMA (2,1,2) as our best model. From Table 4, we can express the model in the form of Equation 11.

$$Y_t = (\Phi + 2)Y_{t-1} + (\Phi - 1)Y_{t-2} + e_t - \theta e_{t-1} - \theta e_{t-2}. \quad (11)$$

Tabel 6. Forecasting Value of the ARIMA (2,2,2)

Date	Lower Bound	Upper Bound	Predicted Data	Actual Data
March 6, 2023	63.501759	152.0329	107.7674	125
March 13, 2023	47.189956	168.7303	107.9601	106
March 20, 2023	32.136103	197.7296	114.9329	115
March 27, 2023	14.2270663	198.5645	106.3958	109
April 3, 2023	7.299932	216.2448	111.7724	101
April 10, 2023	-4.4303282	227.0409	111.3053	92
April 17, 2023	-15.362607	231.5272	108.0823	95
April 24, 2023	-21.1864602	245.7736	112.2945	104
May 1, 2023	-31.749520	250.7268	109.4886	106
May 8, 2023	-38.700140	258.5071	109.9035	106

Table 6 shows the estimated and real values of ARIMA (2,1,2), as well as the lower and upper bounds for ARIMA (2,1,2) after 10 weeks.

IV. Conclusion

The stock price estimate for PT. GoTo Gojek Tokopedia Tbk for the next ten weeks, March 6, 2023 toMay 8, 2022. The researchers utilized Yahoo Finance to gather historical stock price data for PT. GoTo Gojek Tokopedia Tbk from April 11, 2022 to February 27, 2022. To find the optimum ARIMA model for the data in this study, the researcher used the ARIMA Box-Jenkins Method. With a MAPE of 895.10% and all the requirements satisfied, ARIMA (2,1,2) is the best model we can find. Changing the ϕ and θ with the AR and MA values in Table 5 for ARIMA (2,1,2) yields the Equation 12.

$$Y_t = -1.2049Y_{t-1} + 0.7041Y_{t-2} + e_t - 1.1680e_{t-1} - 0.9999e_{t-2}. \quad (12)$$

Future research should consider investigating alternative forecasting techniques to further enhance the understanding and prediction of PT. GoTo Gojek Tokopedia Tbk's stock prices. Specifically, exploringmethodologies such as the Exponential Smoothing State Space Model (ETS) or Long Short-Term Memory (LSTM) neural networks could provide valuable insights. By conducting comparative analysesbetween these advanced techniques and the ARIMA model, researchers can assess their relative performance in terms of forecasting accuracy and reliability. Such investigations will contribute significantly to the body of knowledge on stock price forecasting and may yield improved predictive frameworks.

Acknowledgment

I would like to thank President University for supporting the research.

References

[1] H. Wiguna, Y. Nugraha, R. F. Rizka, A. Andika, J. I. Kanggrawan, A. L. Suherman, "Data-Based Policy: Analysis and Prediction of the Spread of COVID-19 in Jakarta using the Autoregressive Integrated Moving Average (ARIMA) Method," *J. Sist. Cerdas*, vol. 3, no. 2, pp. 74-83, 2020, DOI: 10.37396/jsc.v3i2.76

[2] E. Tan and I. Astuti, "Autoregressive Integrated Moving Average Method for Forecasting Sales."

EKOMABIS: J. Ekonomi. Manajemen. Bisnis., vol. 1, no. 2, pp. 149–158, 2020.

[3] Box, George EP, and George C. Tiao. "Intervention Analysis with Applications to Economic and Environmental Problems," *J. Am. Stat. Assoc.*, vol. 70, no. 349, pp. 70-79, 1975.

[4] P. Mondal, L. Shit, and S. Goswami, "Study of Effectiveness of Time SeriesModeling (ARIMA) in Forecasting Stock Prices," *Int. J. Comput. Sci., Eng. Appl.*, vol. 4, no. 2 pp.13-29, 2014, DOI: 10.5121/ijcsea.2014.4202

[5] S. Chen, X. Lan, Y. Hu, Q. Liu, and Y. Deng, "The Time Series forecasting: from the Aspect of Network," *arXiv:1403.1713 Physic*, pp. 1-5, 2014.

[6] E. A. Widodo, E. S. Nugraha, and D. A. Hamzah, "Forecasting PT Bank Mandiri Tbk Stock Price Using ARIMA," in *Symposium on Data Science*, 2022.

[7] J. D. Cryer and K. S. Chan, *"Models for Non-stationary Time Series in Time series analysis: with Applications in R,*" 2nd ed.Springer Science & Business Media, p. 66, 2008.

[8] J. D. Cryer and K. S. Chan, *"Models for Non-stationary Time Series in Time Series Analysis: with Applications in R,"* 2nd ed., Springer Science & Business Media, p. 57, 2008.

[9] S. Prabhakaran, "Arima Model - Complete Guide to Time Series Forecasting in Yython: ML+," Machine Learning Plus," *Fuzzy Syst. Soft Comput.*, vol 7. No. 2, pp. 954-963, 2022.