Forecasting Weekly Stock Price of PT. Bank Central Asia Tbk (BBCA) Using ARIMA Box-Jenkins Method

Dheeya Raja Noorlesmana

Study Program of Actuarial Science, School of Business, President University, Bekasi, 17550, Indonesia

Corresponding author: dheeya.noorlesmana@student.president.ac.id

Abstract**—** The dynamic fluctuation of stock prices in financial markets highlights the importance of accurately forecasting future price movements, particularly for investors and financial institutions making strategic decisions. This study focuses on forecasting the weekly stock price of PT Bank Central Asia Tbk (BBCA) using the Autoregressive Integrated Moving Average (ARIMA) method based on the Box-Jenkins approach. The dataset consists of weekly closing stock prices from January 7, 2024, to December 29, 2024, totaling 51 observations obtained from Investing.com. Following data preprocessing, including stationarity testing and model selection through ACF and PACF analysis, several ARIMA models were evaluated. The ARIMA(2,2,0) model was identified as the best-fitting model based on error metrics, achieving a Mean Absolute Percentage Error (MAPE) of 2.29%. This result demonstrates a high level of forecasting accuracy, providing valuable insights for investment planning and financial risk management. The findings confirm that ARIMA remains a reliable and practical method for short-term stock price forecasting in dynamic financial environments.

Keywords**—** Stock Price; ARIMA; Time Series; BBCA; Forecasting.

I. INTRODUCTION

Stock price forecasting is an important aspect of financial analysis, especially in the context of investment planning and risk management. In dynamic financial markets, predicting future stock price movements helps investors and institutions make informed decisions. One of Indonesia's largest banks, PT Bank Central Asia Tbk (BBCA), is a benchmark for the financial sector, and its stock performance is closely followed by market participants. Given its economic influence and trading volume, developing a reliable forecasting model for BBCA stock is relevant and valuable.

Several studies have applied time series models for stock price forecasting. Among them, the Autoregressive Integrated Moving Average (ARIMA) model, which is based on the Box-Jenkins methodology, remains one of the most frequently used due to its effectiveness in modeling linear trends in time series data. However, despite the increasing use of machine learning approaches, ARIMA still has practical importance, especially for short- term forecasts with relatively limited data sets.

Previous researchs have been conducted on stock price forecasting, especially for PT Bank Central Asia (BBCA), employing various methodologies ranging from traditional statistical models to advanced machine learning techniques. Wardhani and Yudhanegara (2024) used ARIMA Box-Jenkins, forecasting stock price of PT. Aneka Tambang Tbk (ANTM) from January 1, 2022 to January 27, 2024 and get the ARIMA (3,1,0) with a MAPE value of 7.68% for last 4 weeks the data [1]. Saputra (2024) used ARIMA Box-Jenkins, forecasting stock price of PT. GoTo Gojek Tokopedia Tbk (GOTO.JK) from April 11, 2022 to February 27, 2023 get the ARIMA (2,1,2) with a MAPE value of 7.85% for the next 10 weeks [2]. Saputra and Pangestika and Josaphat (2025) used a combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) to predict BBCA's stock price and found CNN to perform better than LSTM [3]. Wathani et al. (2023) utilized LSTM to predict stock trends and demonstrated that LSTM can be a useful tool for stock price prediction when parameters are set optimally [4]. Abdillah et al. (2021) focused on the ARIMA model, forecasting BBCA stock prices with a MAPE value of 1.95%, demonstrating the efficacy of ARIMA in time series prediction for financial data [5]. In another study, Abidha and Ahdika (2023) employed an ARIMA-GARCH hybrid model to forecast BBCA stock prices and reported a MAPE of 0.961, illustrating that hybrid models can improve forecasting accuracy [6]. Furthermore, Rahmawati et al. (2024) compared the ARIMA model with the TSR-ARIMA hybrid model and found the latter to be more accurate in forecasting BBCA stock prices [7]. To fill the gap, this study applies the Box-Jenkins ARIMA method to forecast the weekly stock price of PT Bank Central Asia Tbk (BBCA) using data from January 7, 2024 to December 29, 2024. The methodology used includes data preprocessing, stationarity test, model specification through ACF and PACF analysis, and evaluation based on error metrics such as MSE, RMSE, MAE, and MAPE. Among the tested models, ARIMA (2,2,0) emerged as the best performing model with the lowest MAPE of 2,29%, indicating high accuracy in short-term forecasting.

The purpose of this study is to demonstrate the effectiveness of ARIMA in forecasting BBCA weekly stock prices using real market data. The findings are expected to provide valuable insights for investors, financial analysts, and risk managers in understanding price movements and improving investment strategies.

II. LITERATURE REVIEW

A. Time Series Analysis

A statistical technique called time series analysis is used to examine data points that are gathered or recorded at predetermined intervals of time, such daily, weekly, monthly, or yearly. In order to anticipate future values based on past observations, this approach is usually used to find patterns, trends, or seasonal impacts in data. Depending on the structure and goal of the data, this method may be applied to both short-term and long-term forecasting. It is frequently used in domains including finance, economics, and weather prediction. Because time series analysis offers a methodical approach to modeling the behavior of data over time, it is very helpful in financial forecasting, such as stock price prediction. ARIMA, or Autoregressive Integrated Moving Average, is a popular model for time series forecasting, developed through the Box-Jenkins methodology [8], which combines autoregressive, integrated, and moving average components. Although some experts argue that stock prices behave randomly and are difficult to predict [9], other studies have shown that ARIMA can still effectively model and forecast linear patterns in stock movements [10]. Moreover, despite the emergence of advanced machine learning techniques, traditional statistical models such as ARIMA remain reliable and widely used, especially for short- term forecasting [11].

B. Stationarity and Differencing

Stationarity refers to the property of a time series in which its statistical characteristics (such as mean, variance, and autocorrelation) remain constant over time. In time series modeling techniques like ARIMA, which presumes a Table data structure, stationary time series are essential. But a lot of time series in the actual world are not stationary. often exhibiting trends or seasonal patterns. To overcome this, differencing is used. This technique transforms the data by subtracting the current observation from the previous observation. The differencing process can be presented using the following formula as a basis for understanding how differencing stabilizes the time series for ARIMA modeling.

$$W_t = \Delta^d Y_t$$

First-order differencing: $W_t = \Delta Y_t = Y_t - Y_{t-1}$

If the series remains non-stationary after first-order differencing, the process can be repeated (second-order differencing, etc.) until stationarity is achieved.

Second-order differencing: $W_t = \Delta^2 Y_t = \Delta(\Delta Y_t) = Y_t - 2Y_{t-1} + Y_{t-2}$

Third-order differencing: $W_t = \Delta^3 Y_t = \Delta(\Delta^2 Y_t) = \Delta(Y_t - 2Y_{t-1} + Y_{t-2})$

Differencing helps eliminate trends and stabilize the variance, making the time series more suitable for modeling and forecasting.

C. Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF)

A diagnostic tool that calculates the correlation between a time series and its lag values is called the Autocorrelation Function (ACF). It sheds light on how long patterns last across time. Long-term dependencies are shown by the ACF's sluggish fall in non-stationary series and its quick reduction in stationary series [12]. The ACF is also useful for identifying the order of the Moving Average (MA) component in ARIMA models Complementing the ACF, the Partial Autocorrelation Function (PACF) measures the direct correlation between a time series and its lagged values, after removing the effects of intervening lags. It is particularly valuable in identifying the order of the Autoregressive (AR) component [13]. A PACF plot that cuts off sharply after a certain lag suggests the appropriate AR order. ACF and PACF plots play a critical role in model identification and validation. Their patterns guide the selection of AR and MA terms and help determine whether differencing is necessary to achieve stationarity. This combined diagnostic approach forms the basis for effective ARIMA modeling.

D. Error Estimation

Error estimation is an important for evaluating the performance and predictive accuracy of time series models, various error estimation metrics are used. Among the most commonly applied are Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics assess the deviation between actual observations and model forecasts. MAE measures the average absolute error, providing a straightforward interpretation of forecast accuracy. MSE and RMSE penalize larger errors more heavily due to the squaring operation, making them sensitive to outliers. MAPE measures forecast error as a percentage, which is useful for interpretability across different scales. Selecting appropriate error metrics is essential for comparing models and validating their effectiveness in real-world applications [14].
Here i the error estimation formula,

Mean Absolute Error (MAE)

$$\mathbf{MAE} = \frac{1}{n}\sum_{t=1}^{n}|A_t - \hat{F}_t|$$

Mean Squared Error (MSE)

$$\mathbf{MSE} = \frac{1}{n}\sum_{t=1}^{n}(A_t - \hat{F}_t)^2$$

Root Mean Squared Error (RMSE)

$$\mathbf{RMSE} = \sqrt{\mathbf{MSE}}$$

Mean Absolute Percentage Error (MAPE)

$$\mathbf{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

with the descriptions:

- n represent the number of observation or total number of actual and forecasted value pairs being compared.
- A_t is the actual data or actual value.
- F_t is predicted data or forecasted value.

MAPE expresses prediction accuracy as a percentage, making it easy to interpret and compare across models and datasets. MAPE values are often categorised based on interpretation:

- Less than 10% is considered very precise.
- 10% - 20% great.
- 20% - 50% justified.
- More than 50% indicates inadequate performance in forecasting.

These thresholds help practitioners quickly assess whether a forecasting model is very justified [15].
However, MAPE has its limitations. It becomes a problem when the actual value is close to zero, as this can lead to a very high or undefined percentage error. Despite these drawbacks, MAPE remains widely used due to its intuitive interpretation and ease of application.

III. METHODOLOGY

A. Box-Jenkins Method

The Box-Jenkins methodology, basicly referred to as the ARIMA (Autoregressive Integrated Moving Average) model. George Box and Gwilym Jenkins developed the Box-Jenkins approach, sometimes known as the ARIMA (Autoregressive Integrated Moving Average) model. In time series (especially forecasting), it is one of the most used methods, particularly for financial and economic data, such as stock prices. To represent the structure and patterns of time series data, the model consists of three parts; Autoregression (AR), Differencing (I), and Moving Average (MA).

The time series must be stationary—that is, its statistical characteristics, including mean and variance, must not change over time—in order to use an ARIMA model. It is necessary to use differencing to convert non-stationary data into a stationary time series.

The ARIMA model consists of the following components:

1. Autoregressive (AR)

   The Autoregressive (AR) model assumes that the current value of the time series (Y_t) is influenced by its previous values. The general form of an AR model of order p is expressed as:

   $$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

   Where:

   - ϕ_p is AR coefficient at lag p.
   - e_t is white noise (error term) at time t.

2. Differencing (I)

   The differencing approach subtracts the time series's past values from its present values in order to transform a non-stationary time series into a stationary one. This makes the data appropriate for ARIMA modeling by removing seasonality or trends. The general form of an differencing model of d is expresed as:

   $$W_t = \Delta^d Y_t$$

   First-order differencing: $W_t = \Delta Y_t = Y_t - Y_{t-1}$

   If the series remains non-stationary after first-order differencing, the process can be repeated (second-order differencing, etc.) until stationarity is achieved.

   Second-order differencing: $W_t = \Delta^2 Y_t = \Delta(\Delta Y_t) = Y_t - 2Y_{t-1} + Y_{t-2}$

   Third-order differencing: $W_t = \Delta^3 Y_t = \Delta(\Delta^2 Y_t) = \Delta(Y_t - 2Y_{t-1} + Y_{t-2})$

   Differencing helps eliminate trends and stabilize the variance, making the time series more suitable for modeling and forecasting.

3. Moving Average (MA)

   The Moving Average model assumes that the current value of the series is a linear function of past error terms. The general form of an MA model of order q is defined as:

   $$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

   Where:

   - θ_q is MA coefficient at lag q.
   - e_t is white noise (error term) at time t.

4. Autoregressive Integrated Moving Average (ARIMA)

   Autoregressive (AR) model and Moving Average (MA) equation can be combined and written as the ARMA (p,q):

   $$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Forecasting nonstationary data can be done with the ARIMA (p,d,q) method by combining the AR and MA methods and using differencing $W_t = \Delta^d Y_t$. ARIMA model can be represented by:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Where $W_t = \Delta^d Y_t$ is differenced time series after applying differencing d times.

5. Stationary Check using Augmented Dickey Fuller (ADF) test

The ADF test is used to determine whether a time series is stationary. Stationarity is crucial for time series modeling (like ARIMA), where the statistical properties such as mean and variance must remain constant over time. The ADF test extends the basic Dickey-Fuller test by including lagged differences to account for autocorrelation in the residuals.

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-1} + e_t$$

- ΔY_t is the first difference of the time series: $Y_t - Y_{t-1}$. This transformation helps to remove trends and stabilize the mean.
- β_1 is constant (intercept) term. It represents the mean level of the differenced series.
- $\beta_2 t$ is time trend term (optional). Used if the series has a deterministic trend.
- δY_{t-1} is the lagged level of the series. This is the key term being tested. If δ=0, then a unit root is present (non-stationary).
- $\alpha_i \sum_{i=1}^{m} \Delta Y_{t-1}$ is lagged differences of the dependent variable, added to control for autocorrelation. **The number of lags m is usually selected based on criteria like AIC or BIC.**
- e_t is the error term (white noise), which is assumed to be independently and identically distributed (i.i.d.).

6. Shapiro-Wilk Test

The Shapiro-Wilk test is used to determine whether a given sample comes from a normally distributed population. It is widely used in statistical modeling to validate the assumption of normality, especially for residuals in regression or time series analysis.

$$W_t = \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

- n is sample size.
- a_i is precomputed constants (based on expected values and covariances of order statistics of a normal distribution).
- x_i is The ordered sample values, from smallest to largest.
- $\bar{x}$ is the sample maen.

A **W value** close to 1 indicates the sample likely comes from a normal distribution.
A small **W** and small **p-value** (typically < 0.05) suggest the data deviates from normality, so you reject the null hypothesis.

7. Ljung-Box Test

The Ljung-Box test is used to check whether a time series or more commonly, the residuals from a time series model (e.g., ARIMA) contain significant autocorrelation at multiple lags. It helps assess if the model has captured all the structure in the data.

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\hat{p}_k^2}{n-k}$$

- Q is ljung-Box test statistic (approximately follows a chi-square distribution).
- n is number of observations.
- $p^{\wedge}{}_k{}^2$ is sample autocorrelation at lag k.
- h is the number of lags to test.
- $\sum_{k=1}^{h}$ is summation from lag 1 to lag h.

B. Forecasting Procedure Using ARIMA

The process of forecasting time series data using ARIMA models involves several systematic steps to ensure the accuracy and reliability of the model. These steps are as follows:

1. Data Preparation

   At first, historical data relevant to the forecasting objective should be collected and organized. The data set should be in a consistent time format, and any missing data or extreme outliers should be handled appropriately to avoid distortions in model training.

2. Stationarity Check

   A basic hypothesis in the Arima model is that the data must be repaired. This means that statistical attributes such as average and variance will not be changed over time. Dickey-Fuller testing (ADF) is applied to evaluate the standing of the time series. If the test results indicate without the structure, the difference is applied to the data until it becomes still standing.

3. Model specification (Determining p,d,q)

   Once the data is stationary, the next step is to identify the appropriate values for the ARIMA parameters:
   - p represent autoregressive order.
   - d represent the number of differencing.
   - q represent the moving average order.

   These values can be estimated by analyzing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. The ACF helps identify the value of q, while the PACF helps determine p.

4. Parameter Estimation

   Following the selection of the preliminary model structure, statistical tools or programming languages like R or Python are used to estimate the model parameters. The greatest likelihood approach is typically used to estimate the parameters. The data is then fitted to the model.

5. Residual Analysis

   Once the model is fitted, residual diagnostics are performed to evaluate whether the assumptions of the model are met. Residuals should ideally follow a normal distribution, have a mean of zero, and show no autocorrelation. This can be checked using:
   - Shapiro test for normality
   - Ljung-Box test for autocorrelation, if the p-values of these tests are greater than 0.05, it indicates that the residuals meet the required assumptions.

6. Forecasting

   After model authentication, it can be used to create forecasts for future stages. The accuracy of the forecast can also be assessed by performance measurements such as the average (MAE), average square error (RMSE) or the absolute percentage error (MAPE).

IV. RESULT AND DISCUSSION

A. Data Preparation

The data used in this analysis is PT Bank Central Asia Tbk (BBCA) weekly stock price data from January 07, 2024 to December 29, 2024 with a total of 51 data. Data shown in Table 1 and Figure 1 as graph. The data is taken from *Investing.com* website and is processed using R Studio.

Table 1
Data of The Stock Price of PT. Bank Central Asia Tbk

Date	Close	Date	Close	Date	Close
07/01/2024	9700	12/05/2024	9750	08/09/2024	10425
14/01/2024	9625	19/05/2024	9425	15/09/2024	10775
21/01/2024	9350	26/05/2024	9250	22/09/2024	10650
28/01/2024	9700	02/06/2024	9325	29/09/2024	10475
04/02/2024	9700	09/06/2024	9200	06/10/2024	10375
11/02/2024	9950	16/06/2024	9600	13/10/2024	10750
18/02/2024	9825	23/06/2024	9925	20/10/2024	10750
25/02/2024	9825	30/06/2024	9950	27/10/2024	10425
03/03/2024	10150	07/07/2024	10075	03/11/2024	10075
10/03/2024	10150	14/07/2024	10125	10/11/2024	10175
17/03/2024	10100	21/07/2024	10350	17/11/2024	9850
24/03/2024	10075	28/07/2024	10200	24/11/2024	10000
31/03/2024	9825	04/08/2024	10150	01/12/2024	10075
14/04/2024	9475	11/08/2024	10325	08/12/2024	10050
21/04/2024	9625	18/08/2024	10325	15/12/2024	9650
28/04/2024	9850	25/08/2024	10325	22/12/2024	9800
05/05/2024	9375	01/09/2024	10300	29/12/2024	9850

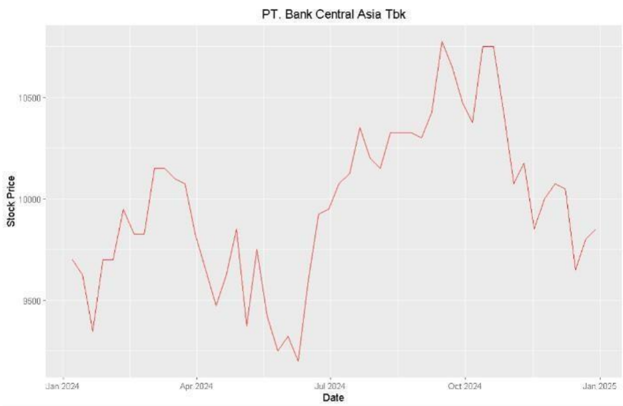

Figure. 1 Graph of PT. Bank Central Asia Tbk Stock Price

Before conducting a stationary check, the data must first be exported into the **R software**. The following are some of the libraries needed to install ini R software, this library is used for forecasting in this study:

- library(readxl)
  This library used to reads our data (in excel sheet) into a data frame.
- library(forecast)
  This library used to create a forecast based on the arima model.
- library(tseries)
  This library used for time series analysis and statistical testing to check stationarity.
- library(lmtest)
  This library used for autocorrelation test.
- library(ggplot2)
  This library used to create elegant plots or visualizations.
- library(metrics)
  This library used to calculate performance metrics for prediction models.
- library(kableExtra)
  This library used to create the base table in R software.
- library(scales)
  This library used to create date in line of the graph.

B. Stationary Check

Before processing data, data must be confirmed to stand still using the increasing Funt Dickey. The test method in R Studio. If the data has a value of P below 0.05, the data can be considered to stand still. In Figure 2, it has been shown that the graphics of Central Asia -Central Asia Bank, the action process with the value of p collected from the Dicker test increased by 0.7399. This value exceeds the significance of 0.05, which indicates that data is not standing still.

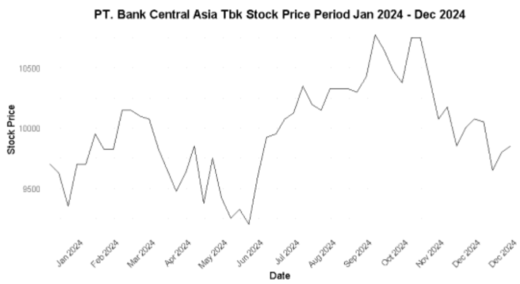

Figure. 2 Plot of PT Bank Central Asia Tbk Stock Price

To make the data stationary, differencing is required using the "diff(data)" code in R Studio. In Figure 3, it is shown the plot of secoon difference of PT Aneka Tambang Tbk Stock Price after performing first differencing with p-value obtained from the Augmented Dickey-Fuller Test is 0,05023. The p-value after first differencing is more than 0.05 indicates that we must performing the second differencing. The p-value from the Augmented Dickey-Fuller after performing second differencing is 0,01. The p-value after second differencing is less than 0.05 indicates that the data is stationary and order d used for the ARIMA model is 2.

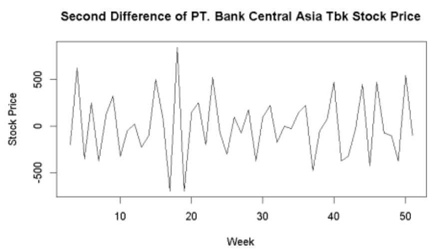

Figure. 3 Plot of Second Difference PT Bank Central Asia Tbk Stock Price

C. Model Specification

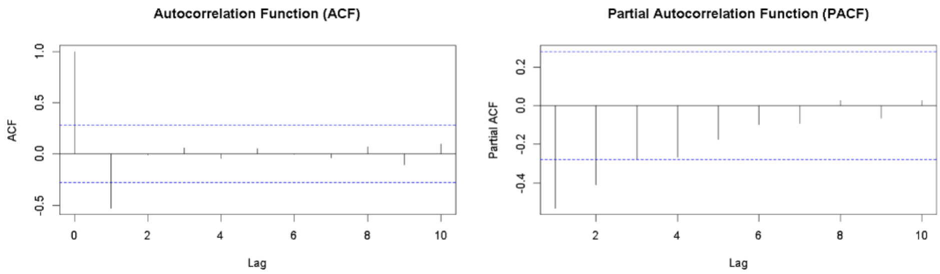

Figure. 4 Plot of ACF and PACF Second Difference PT Bank Central Asia Tbk Stock Price

Based on the information provided by Figure 4, it is known that PACF cuts off at a lag time of 2 and ACF cuts off at a lag time of 1. Therefore, the order of p is determined as 2 and order of q is determined as 1. Table 2 shows that the ARIMA model is organized on the basis of these parameters.

Table 2 shows the ARIMA model arranged based on these parameters

Table 2
ARIMA Model Specification

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

D. Parameter Estimation

The parameters of the Autoregressive (AR) formula are denoted as φ with order p. While the parameters of the Moving Average (MA) formula are denoted as θ with order q. Table 3 shows that by using the summary(model) code, the value of each parameter as well as Log Likelihood and Akaike Information Criterion (AIC) can be obtained.

Table 3
Parameter Estimation for ARIMA Model

Model ARIMA	AR1	AR2	MA1	Log Likelihood	AIC
ARIMA (0, 2, 0)				-355,03	712,05
ARIMA (0, 2, 1)			-1,0000	-336,63	677,25
ARIMA (1, 2, 0)	-0,5255			-346,98	697,95
ARIMA (1, 2, 1)	-0,1298		-1,0000	-336,21	678,42
ARIMA (2, 2, 0)	-0,7476	-0,4316		-342,44	690,88
ARIMA (2, 2, 1)	-0,1473	-0,0963	-1,0000	-335,99	679,98

E. Residual Analysis

Determining the best ARIMA model is performed using the Shapiro Test and the Ljung-Box Test. Through these tests, the ARIMA model with a p-value that exceeds 0.05 will be selected. Table 4 shows the results of the Shapiro Test and Ljung-Box Test and the ARIMA models that satisfy the condition. Models that passed the residual test will be forecasted and the error of each model will be calculated.

Table 4
Result of The Residual Analysis

Model ARIMA	Shapiro Test	Ljung-Box Test	BIC	AIC	Result
ARIMA (0, 2, 0)	0,6897289	0,000486825	713,9458	712,0540	Not Passed
ARIMA (0, 2, 1)	0,5916246	0,501403640	681,0365	677,2529	Passed
ARIMA (1, 2, 0)	0,2485085	0,007410723	701,7338	697,9503	Not Passed
ARIMA (1, 2, 1)	0,3692105	0,784624018	684,0957	678,4203	Passed
ARIMA (2, 2, 0)	0,4969921	0,220141403	696,5519	690,8764	Passed
ARIMA (2, 2, 1)	0,2201687	0,996756459	687,5432	679,9759	Passed

F. Forecast

Table 5
Forecasting Value of ARIMA (0,2,1)

Date	Actual Data	Predicted Data	Lower Bound 80%	Upper Bound 80%	Lower Bound 95%	Upper Bound 95%
05/01/2025	9725	9853	9560,23	10145,77	9405,25	10300,76
12/01/2025	9900	9856	9437,92	10274,08	9216,60	10495,40
19/01/2025	9350	9859	9342,06	10375,94	9068,41	10649,60
26/01/2025	9450	9862	9259,48	10464,52	8940,53	10783,48

Table 6
Forecasting Value of ARIMA (1,2,1)

Date	Actual Data	Predicted Data	Lower Bound 80%	Upper Bound 80%	Lower Bound 95%	Upper Bound 95%
05/01/2025	9725	9846,98	9554,37	10139,59	9399,47	10294,50
12/01/2025	9900	9850,84	9459,18	10242,50	9251,85	10449,84
19/01/2025	9350	9853,81	9378,03	10329,60	9126,17	10581,46
26/01/2025	9450	9856,90	9307,68	10406,12	9016,94	10696,86

Table 7
Forecasting Value of ARIMA (2,2,0)

Date	Actual Data	Predicted Data	Lower Bound 80%	Upper Bound 80%	Lower Bound 95%	Upper Bound 95%
05/01/2025	9725	9737,39	9396,66	10078,11	9216,29	10258,48
12/01/2025	9900	9789,50	9243,44	10335,57	8954,37	10624,63
19/01/2025	9350	9788,65	9009,27	10568,03	8596,69	10980,61
26/01/2025	9450	9756,30	8665,62	10846,98	8088,25	11424,35

Table 8
Forecasting Value of ARIMA (2,2,1)

Date	Actual Data	Predicted Data	Lower Bound 80%	Upper Bound 80%	Lower Bound 95%	Upper Bound 95%
05/01/2025	9725	9832,33	9538,49	10126,17	9382,94	10281,72
12/01/2025	9900	9834,25	9444,33	10224,17	9237,92	10430,59
19/01/2025	9350	9839,81	9382,38	10297,24	9140,23	10539,39
26/01/2025	9450	9842,95	9321,23	10364,66	9045,05	10640,84

Based on forecasting value on Table 5, Table 6, Table 7, and Table 8 the best model for forecasting using ARIMA method can be determined by comparing the error among models as shown below.

Table 9
Error Measured from ARIMA Models

Model ARIMA	MSE	RMSE	MAE	MAPE
ARIMA (0, 2, 1)	111786,25	334,34	273,25	2,89%
ARIMA (1, 2, 1)	109172,82	330,41	270,46	2,86%
ARIMA (2, 2, 0)	74649,32	273,22	216,96	2,29%
ARIMA (2, 2, 1)	102541,58	320,22	263,96	2,79%

Based on the error metrics comparison across all tested ARIMA models as shown in Table 9, the **ARIMA (2,2,0)** model emerges as the most suitable for forecasting PT Bank Central Asia Tbk's stock prices. This model achieved the lowest error values across multiple metrics, including:

- MSE (Mean Squared Error): 74649,32 (lowest among all models)
- RMSE (Root Mean Squared Error): 273,22
- MAE (Mean Absolute Error): 216,96
- MAPE (Mean Absolute Percentage Error): 2,29%

The model's superior performance is particularly evident in its MAPE value of 2,29%, which indicates that the average forecast error is only about 2,71% of the actual stock price values. This level of accuracy suggests strong predictive capability for practical investment decision-making. Refers to Table 3, the best model that is ARIMA (2,2,0) can be expressed in the form of equation as shown below.

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + e_t$$

Let

$$W_t = \Delta^2 Y_t = \Delta(\Delta Y_t) = Y_t - 2Y_{t-1} + Y_{t-2}$$

Then the equation become:

$$Y_t - 2Y_{t-1} + Y_{t-2} = \phi_1(Y_{t-1} - 2Y_{t-2} + Y_{t-3}) + \phi_2(Y_{t-2} - 2Y_{t-3} + Y_{t-4}) + e_t$$

$$Y_t = 2Y_{t-1} - Y_{t-2} + (-0{,}7476)(Y_{t-1} - 2Y_{t-2} + Y_{t-3}) + (-0{,}4316)(Y_{t-2} - 2Y_{t-3} + Y_{t-4}) + e_t$$

$$Y_t = 1{,}2524Y_{t-1} + 0{,}0630Y_{t-2} + 0{,}1156Y_{t-3} - 0{,}4316Y_{t-4} + e_t$$

The accuracy of this model, demonstrated by a Mean Absolute Percentage Error (MAPE) of 2,29%.The model's strong performance, combined with its relative simplicity (requiring only two MA parameters), makes it particularly suitable for operational forecasting. Investors and analysts can utilize this model to generate reliable short-term price forecasts while maintaining awareness that all predictions carry inherent uncertainty, as reflected in the model's confidence intervals.

Figure 5 below shows the visualization plot of forecasting from ARIMA (2,2,0) of PT. Bank Central Asia (BBCA) Tbk stock price forecasting from January 5, 2025 to January 26, 2025.

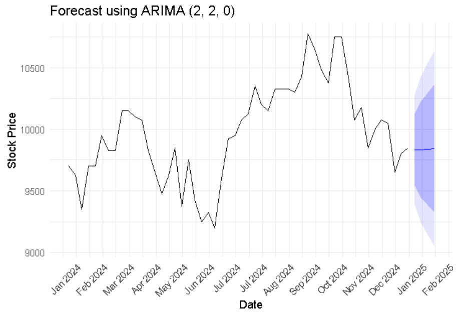

Figure. 5 Forecasting Plot from ARIMA (2,2,0)

A forecasting vs actual graph or plot is a visualization used to compare the predicted values (forecast) with the actual values (actual) of data over a period of time. It is very useful for evaluating the performance of prediction models, such as time series, regression, or machine learning models. Figure 6 shows the forecasting vs actual of ARIMA (2,2,0).

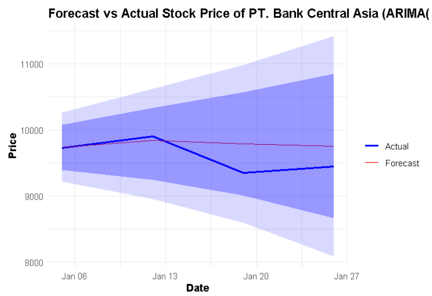

Figure. 6 Forecasting vs Actual from ARIMA (2,2,0)

V. CONCLUSION

Stock price forecasting has been done for PT Bank Central Asia Tbk (BBCA) for the next 4 weeks from January 5, 2025 to January 26, 2025 with original stock price data from January 7, 2024 to December 29, 2024 with total 51 data obtained from *investing.com* website. Using Box-Jenkins and processed using R Studio, it was found that the ARIMA (2,2,0) model passed the residual test and has the smallest error with the MAPE 2,29% to be used to forecast stock price and can be expressed in thr folowing equation.

$$Y_t = 1{,}2524Y_{t-1} + 0{,}0630Y_{t-2} + 0{,}1156Y_{t-3} - 0{,}4316Y_{t-4} + e_t$$

However, it is important to note that while the ARIMA model provides adequate results, there are limitations that should be considered. One such limitation is its ability to handle unexpected market fluctuations. The model relies solely on historical data and does not account for external variables that may influence stock prices. As a next step, it would be worthwhile to consider using other models such as ARIMAX, if there are significant external variables that could impact stock prices, in order to improve the accuracy of the forecast.

REFERENCES

[1] A. I. S. Wardhani dan M. R. Yudhanegara, "Forecasting Weekly Stock Price of PT. Aneka Tambang Tbk (ANTM) Using ARIMA Box-Jenkins Method," *Journal of Actuarial, Finance, and Risk Management*, vol. 3, no. 2, pp. 1–10, 2024. [Online].Tersedia: http://e-journal.president.ac.id/presunivojs/index.php/JAFRM/article/view/5564ojs/index.php/JAFRM/article/view/5564

[2] J. Saputra, “Forecasting Stock Price of PT. Goto Gojek Tokopedia Tbk (GOTO.JK) using ARIMA Box-Jenkins Method,” *Indonesian Journal of Applied Mathematics and Statistics*, 1(2), 79-88. Availabel: https://idjams.com/index.php/jms/article/view/14

[3] Z. Pangestika and B. P. Josaphat, "Predicting Stock Price Using Convolutional Neural Network and Long Short Term Memory (Case Study: Stock of BBCA)," *J. Indones. Math. Soc.*, vol. 31, no. 1, pp. 1512, 2025. [Online]. Available: https://doi.org/10.22342/jims.v31i1.1512

[4] M. N. Wathani, Kusrini, and K. Kusnawi, "Prediksi Tren Pergerakan Harga Saham PT Bank Central Asia Tbk, Dengan Menggunakan Algoritma Long Shot Term Memory (LSTM)," *Infotek: Jurnal Informatika dan Teknologi*, vol. 6, no. 2, pp. 502–512, 2023. [Online]. Available: https://doi.org/10.29408/jit.v6i2.19824

[5] M. F. P. Abdillah, E. Zukhronah, and Respatiwulan, "Peramalan Harga Saham PT. Bank Central Asia Tbk Menggunakan Metode Auto Regressive Integrated Moving Average (ARIMA)," *Prosiding SNAST*, 2021. [Online]. Available: https://ejournal.akprind.ac.id/index.php/snast/article/view/3345

[6] H. L. Abidha and A. Ahdika, "Unlocking market insight: Forecasting PT Bank Central Asia Tbk stock prices with ARIMA-GARCH analysis," *Desimal: Jurnal Matematika*, 2023. [Online]. Available: https://ejournal.radenintan.ac.id/index.php/desimal/article/view/22150

[7] R. Rahmawati, E. Zukhronah, and W. Sulandari, "Perbandingan Peramalan Harga Saham PT Bank Central Asia Tbk Menggunakan Model ARIMA dan Hibrida TSR-ARIMA," *Jurnal Gaussian*, vol. 13, no. 2, pp. 289–299, 2024. [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/article/view/43293

[8] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 1994.

[9] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[10] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for time series forecasting," *Neurocomputing*, vol. 74, no. 3, pp. 3528–3539, 2010.

[11] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. Melbourne, Australia: OTexts, 2018. [Online]. Available: https://otexts.com/fpp2/

[12] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 3rd ed. New York, NY, USA: Springer, 2016.

[13] C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th ed. Boca Raton, FL, USA: CRC Press, 2003.

[14] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: Methods and Applications*, 3rd ed. New York, NY, USA: Wiley, 1998.
[15] M. del R. C. Estrada, M. E. G. Camarillo, M. E. S. Parraguirre, M. E. G. Castillo, E. M. Juárez, and M. J. C. Gómez, "Evaluation of Several Error Measures Applied to the Sales Forecast System of Chemicals Supply Enterprises," International Journal of Business Administration, vol. 11, no. 4, pp. 39–51, 2020. Available: https://doi.org/10.5430/ijba.v11n4p39