Forecasting PT Bank Mandiri Tbk Stock Price Using ARIMA Model

Edward Anthony Widodo[1*], Edwin Setiawan Nugraha[2], Dadang Amir Hamzah[3]

[1,2,3] Study Program of Actuarial Science, School of Business, President University, 17550, Indonesia
[*] Corresponding author: edward.widodo@student.president.ac.id

Abstract— Stocks are one of the most popular financial market instruments. Issuing Stocks is one of the company's choices when deciding to fund the company including in financial sector such as PT Bank Mandiri Tbk. On the other hand, stocks are an investment instrument that many investors choose because they are able to provide an attractive level of profit. In this work, The Autoregressive Integrated Moving Average (ARIMA) method is used to predict the Stock price of PT Bank Mandiri. The result show that ARIMA (1,2,1) is the best model to forecasting the Stock price of PT Bank Mandiri with AIC 1178.48. This forecasting is very useful as consideration for investors in making decisions on PT Bank Mandiri Stocks.

Keywords— ARIMA; PT Bank Mandiri; Forecasting; MAPE; Stock Price

I. INTRODUCTION

PT Bank Mandiri was established on October 2, 1998, as part of the banking restructuring program implemented by the Indonesian government. To this day, PT Bank Mandiri Tbk continues its more than 140 years of tradition of contributing to the banking world and the Indonesian economy [1]. Stocks are one of the most popular financial market instruments. Issuing Stocks is one of the company's choices when deciding to fund the company. On the other hand, stocks are an investment instrument that many investors choose because they are able to provide an attractive level of profit. Stocks can be defined as a sign of capital participation of a person or party (business entity) in a company or limited liability company. By including this capital, the party has a claim on the company's income, claims on company assets, and is entitled to attend the General Meeting of Stockholders (GMS) [2].

In previous works, many research used ARIMA model to forecast the stock price of financial industries including banking. In [3], Iqbal and Ningsih forecast stock price of PT BTPN Syariah by using ARIMA model and GARCH model with data from January 2019 to December 2019. Their result show that the best model is ARIMA (1,1,1) and GARCH (2,1). Lilipaly *et al*. [4] forecast the maximum and minimum stock price of PT. BRI. They showed the best model for maximum and minimum stock price is ARIMA (2,1,3) with the data of stock price of PT. BRI are January 3, 2014 to. October, 20, 2014 . In [5], there was a discussion about forecasting of PT Bank Mandiri Tbk stocks from January 1, 2014 to December 17, 2018. Their result show that the best model was ARIMA model (1, 1, 2) with AIC 14971.39. In this work, Stock Price of PT Bank Mandiri Tbk will be predicted by using ARIMA method with data from 3 April 2021 until 8 July 2021 [6].

II. METHOD

A. Introduction to the Time Series Model

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) [7]. The next sub sections in section of method refers to [7].

B. Stochastic Process

The sequence of random variables {Yt: t = 0, ±1, ±2, ±3,…} is called a stochastic process and serves as a model for the observed time series. A stochastic process or stochastic modeling is a mathematical model which is a quantitative description of natural phenomena. Stochastic itself means "random or chance" which comes from the Greek.

One of the most important differences between deterministic and stochastic models is that a deterministic model predicts a single outcome from a given set of circumstances with absolute certainty, whereas a stochastic model only gives a probability of an outcome and can predict a set of possible outcomes weighted by their probabilities, or probabilities The case for the most realistic stochastic model , other analytical and computational methods are applied – analysis of moments such as the mean and variance of the distribution, estimation of extinction probability, numerical simulation of sample paths, and calculation of probability histograms. For a stochastic process {Yt: t = 0, ±1, ±2, ±3,…}, the mean function is defined by

$$\mu_t = E(Y_t) \quad for\ t = 0, \pm1, \pm2, \ldots.. \tag{1}$$

μ_t is only the expected value of the process at time t. In general, μ_t can be different at any time point t. Autocovariance function, $\gamma_{t,s}$ defined as

$$\gamma_{t,s} = Cov(Y_t, Y_s)$$
$$\gamma_{t,s} = E(Y_t - \mu_t)(Y_s - \mu_s) \tag{2}$$

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

$$\rho_{t,s} = Corr(Y_t, Y_s) \quad for\ t, s = 0, \pm1, \pm2, \ldots$$
$$Corr(Y_t, Y_s) = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t)}\sqrt{Var(Y_s)}} \tag{3}$$

C. Stationary

A time series $\{X_t, t = 0, \pm1, \ldots\}$ is said to be strictly stationary if it has statistical properties similar to the "time shift" series $\{X_{t+k}, t = 0, \pm1, \ldots\}$, and if joint distribution $(X_1, X_2, X_3, \ldots X_n)$ and $(X_{1-k}, X_{2-k}, \ldots, X_{n-k})$ for every integer k and $n > 0$ and k is called the lag time. This time series is around a constant mean, constant variance, and constant covariance over time.

This condition applies to all point sets $t_0, t_1, t_2, \ldots t_3$. It must be valid when $n = 1$ or only one point in time. X_t and X_{t-k} will have the same marginal distribution for all t and k.

$$E(X_t) = E(X_{t-k}) \tag{4}$$
$$Var(X_t) = Var(X_{t-k}) \tag{5}$$

This mean $\mu_t = E(X_t)$ and $Var(X_t)$ for all values of t, k, and constant over time.

Since the condition holds for all point sets $t_0, t_1, t_2, \ldots t_3$. It must persist when n=2 or only one point in time. This implies that (X_t, X_s) and (X_{t-k}, X_{s-k}) also has the same stock distribution for all t, s, and k. Since the common division is the same, then

$$Cov(x_t, x_s) = (x_{t-k}, x_{s-k})\ For\ t, k, and\ s \tag{6}$$

for $k=s$

$$Cov(x_t, x_s) = (x_{t-s}, x_{s-s})$$
$$Cov(x_t, x_s) = (x_{t-s}, x_0) \tag{7}$$

for $k=t$

$$Cov(x_t, x_s) = (x_{t-s}, x_{s-s})$$
$$Cov(x_t, x_s) = (x_0, x_{s-t}) \tag{8}$$

then,

$$\gamma_{t,s} = Cov(x_t, x_s) = Cov(x_{t-s}, x_0) = \gamma_{t,s} = Cov(x_0, x_{|t-s|}) \tag{9}$$

Where k = |t-s| show that the covariance $x_t\ and\ x_s$ depending on the time difference |t-s| not the actual t and s. Thus, we can express the autocovariance function.

$$\gamma_k = Cov(x_t, x_{t-k})$$
$$\rho_k = Corr(x_t, x_{t-k}) = \frac{\gamma_k}{\gamma_0} \tag{10}$$

D. Autocorrelation Function (ACF), and Partial Autocorrelation Function (PACF)

In time series analysis, the covariance and correlation between x_t and x_{t+k} using the same process but making a time difference of k lags is called autocovariance (γ_k) and autocorrelation (ρ_k). For the time series x(t), t = 0.1,2,3,4,… autocovariance is defined by:

$$\gamma_k = cov(x_t, x_{t+k}) = E[(x_t - \mu)(x_{t+k}, \mu)] \tag{11}$$

The autocorrelation function (ACF) is as follows:

$$\rho_k = \frac{cov(x_t, x_{t+k})}{\sqrt{Var(x_t)}\sqrt{Var(x_{t+k})}} = \frac{\gamma_k}{\gamma_0} \tag{12}$$

The partial autocorrelation function (PACF) is as follows:

$$\Phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \Phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \Phi_{k-1,j}\rho_j} \tag{13}$$

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

E. Time Series Model

1) Autoregressive Process (AR)

The autoregressive model is a model that describes that the dependent variable is influenced by the dependent variable itself in the previous period and time. In general, the Autoregressive model has the following form,

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \emptyset_3 Y_{t-3} + \ldots\ldots + \emptyset_p Y_{t-p} + e_t \quad (14)$$

where Y_t is time series variable, $\emptyset_t$ is constant, e_t is residual at time *t*.

2) Moving Average (MA)

Combining the dependencies between observations and the residual error of the moving average model applied to the lagging observations. In general, the moving average model has the following form:

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3} + \ldots\ldots + \theta_p e_{t-q} \quad (15)$$

where Y_t is time series variable, θ_t is constant, e_t is residual at time *t*.

3) Autoregressive Moving Average (ARMA) Model

The autoregressive moving average model is a mixed model between the autoregressive (AR) model and the moving average (MA) model. The ARMA method is also often referred to as the Box-Jenkins method because it was developed by George Box and Gwilym Jenkins in 1976. The general form for the ARMA model is,

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots\ldots + \emptyset_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots\ldots + \theta_p e_{t-q} \quad (16)$$

4) Autoregressive Integrated Moving Average (ARIMA) Model

In practice, it is often found that economic data is non-stationary, so it is necessary to modify it, by differentiating it, to produce stationary data. The distinction is made by subtracting the value in a period from the value in the previous period. The data used as input to the ARIMA model is the data from the transformation that is already stationary, not the original data. Several times the differencing process is performed, denoted by d. The ARIMA model is usually denoted by ARIMA (p, d, q) which implies that the model uses p dependent lag values, d levels of differentiation process, and q residual lags. Summary of all steps in forecasting by using ARIMA model is illustrated by a flow chart at Figure 1.

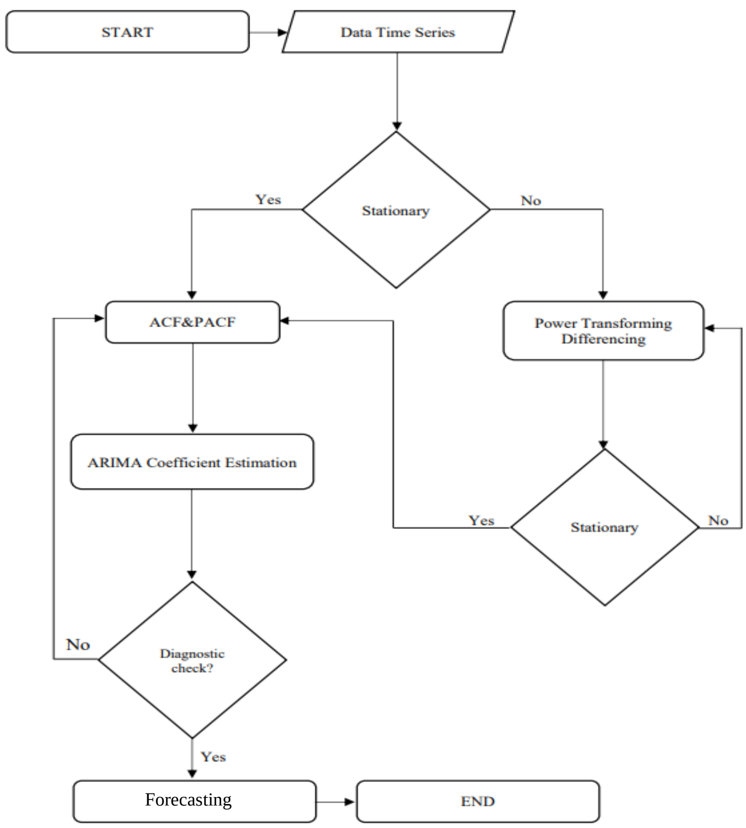

Figure 1. Flow chart of the forecasting method by using ARIMA model

III. RESULT and discussion

A. Data Preparation

We use data on the Stock price of PT Bank Mandiri from April 3, 2021 to July 8, 2021 [6]. The case data is contained in the table below [13]

TABLE 1
DATA ON THE SHARE PRICE OF PT BANK MANDIRI

Date	Price
03/04/2021	5,525
04/04/2021	5,800
05/04/2021	5,750
06/04/2021	5,900
07/04/2021	5,750
08/04/2021	5,800
09/04/2021	6,000
10/04/2021	6,100
11/04/2021	6,250
12/04/2021	6,125
13/04/2021	6,125
14/04/2021	6,000
15/04/2021	6,150
16/04/2021	6,200
17/04/2021	6,275
18/04/2021	6,275
19/04/2021	6,325
20/04/2021	6,275
21/04/2021	6,200
22/04/2021	6,200
23/04/2021	6,025
24/04/2021	6,175
25/04/2021	6,025
26/04/2021	5,975
27/04/2021	5,900
28/04/2021	5,700
29/04/2021	5,775
30/04/2021	5,900
01/05/2021	5,975
02/05/2021	5,950
03/05/2021	5,850
04/05/2021	5,775
05/05/2021	5,750
06/05/2021	5,900
07/05/2021	5,775
08/05/2021	5,800
09/05/2021	5,775
10/05/2021	5,700
11/05/2021	5,875
12/05/2021	5,875
13/05/2021	5,750
14/05/2021	5,925
15/05/2021	6,050
16/05/2021	5,975
17/05/2021	5,800
18/05/2021	5,825
19/05/2021	5,900
20/05/2021	5,750
21/05/2021	5,700
22/05/2021	5,675
23/05/2021	5,775
24/05/2021	5,825
25/05/2021	6,050
26/05/2021	5,975
27/05/2021	5,800
28/05/2021	5,800
29/05/2021	5,925
30/05/2021	6,000
31/05/2021	5,975
01/06/2021	6,050
02/06/2021	5,850
03/06/2021	5,900
04/06/2021	5,975
05/06/2021	5,825
06/06/2021	5,950
07/06/2021	5,825
08/06/2021	5,800
09/06/2021	5,950
10/06/2021	6,100
11/06/2021	6,000
12/06/2021	6,050
13/06/2021	6,125
14/06/2021	6,200
15/06/2021	6,250
16/06/2021	6,250
17/06/2021	6,350
18/06/2021	6,200
19/06/2021	6,150
20/06/2021	6,100
21/06/2021	6,125
22/06/2021	6,150
23/06/2021	6,050
24/06/2021	6,025
25/06/2021	5,975
26/06/2021	6,075
27/06/2021	6,075
28/06/2021	5,975
29/06/2021	5,900
30/06/2021	5,950
01/07/2021	6,000
02/07/2021	6,150
03/07/2021	6,100
04/07/2021	6,475
05/07/2021	6,425
06/07/2021	6,600
07/07/2021	6,700
08/07/2021	6,900

(Source: Investing.com, 2021)

B. Stationarity

Before we use the data, we need to check the stationary of this data first using method of Augmented Dickey-Fuller Test . This data will be called stationary if the p value is less than 0.05. In the first checking, he Augmented

Dickey-Fuller Test give the p value is 0.9395 > 0.05. This mean that the data is non stationary. The next step is differencing the data in order to transform the data become stationary. After the second differencing, the Augmented Dickey-Fuller Test give p value is 0.01 < 0.05. Thus, it can be concluded that the data already stationary. The plot of data before and after second differencing is shown in Figure 2 and Figure 3.

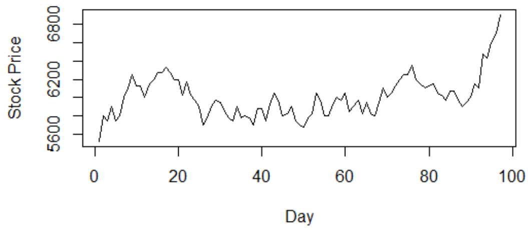

Figure 2. Before Differencing

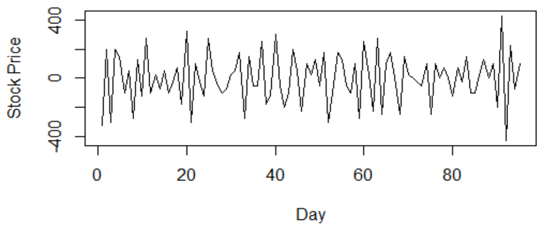

Figure 3. After Second Differencing

C. Model Specification

Model specification can be determined by p and q which represent of order AR and MA, respectively. The plot for ACF and PACF is presented by Figure 4 and Figure 5.

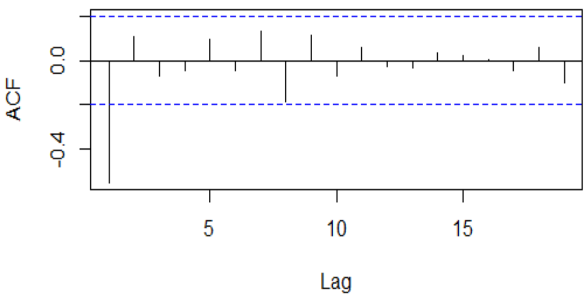

Figure 4. ACF plot

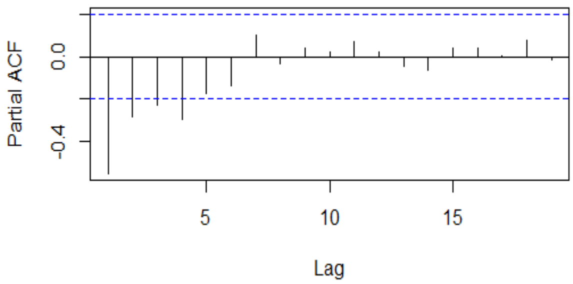

Figure 5. PACF plot

From the Figure 9, the cut of for ACF at lag time 1, thus the value of q is 1. In similar way, from Figure 9, the value of p is 4. In the previous section, the differencing process was carried out twice, so the value of d is 2. Hence, there are 10 specifications of ARIMA model as presented at Table 2.

TABLE 2
ARIMA MODEL SPECIFICATION

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

D. Parameter Estimation

The computation by help of R software, the parameter estimation including log Likehood as well as AIC are provided at Table 3.

TABEL 3
SUMMARY MODEL ARIMA (4, 2, 1)

Model ARIMA	Coefficient of Estimation Result						
	AR1	AR2	AR3	AR4	MA1	Log Likelihood	AIC
ARIMA (4,2,1)	-0.4163	-0.2670	-0.3200	-0.2104	-0.6408	-585.56	1181.12
ARIMA (3,2,1)	-0.2277	-0.0851	-0.1684		0.8334	-586.47	1180.95
ARIMA (2,2,1)	-0.1076	0.0455			-0.9833	-587.19	1180.39
ARIMA (1,2,1)	-0.1366				-0.9420	-587.24	1178.48
ARIMA (0,2,1)					-1.000	-587.83	1177.67
ARIMA (4,2,0)	-0.9323	-0.6963	-0.5787	-0.3651		-589.26	1186.51
ARIMA (3,2,0)	-0.8347	-0.507	-0.2954			-595.28	1196.56
ARIMA (2,2,0)	-0.7375	-0.2937				-599.29	1202.57
ARIMA (1,2,0)	-0.5735					-603.48	1208.95
ARIMA (0,2,0)						-621.72	1243.44

E. Residual Analysis

In this section, the candidate model will be selected by using Shapiro test and Ljung box test. Saphiro test is carried out to determine whether the residual follow normal distribution or not, while the another one is to determine whether the residual has auto correlation or not. The result of the both of the test for models from previous section is presented at Table 4. From the table, the best model is ARIMA (0,2,1) which has minimum AIC value. The ARIMA (1,2,1) will be consider as alternative best model due to this model has AIC value a little bit difference from AIC value ARIMA (0,2,1).

TABLE 4
RESULTS OF CHECKING SHAPIRO AND LJUNG TEST

No	Model ARIMA	Shapiro test	Ljung test	AIC	Result
1	ARIMA (4,2,1)	0.1937	0.8558	1181.12	Passed
2	ARIMA (3,2,1)	0.2858	0.8885	1180.95	Passed
3	ARIMA (2,2,1)	0.2725	0.7802	1180.39	Passed
4	ARIMA (1,2,1)	0.3292	0.7667	1178.48	passed
5	ARIMA (0,2,1)	0.3994	0.4326	1177.67	Passed
6	ARIMA (4,2,0)	0.5563	0.621	1186.51	Passed
7	ARIMA (3,2,0)	0.5669	0.478	1196.56	Passed
8	ARIMA (2,2,0)	0.38	0.5102	1202.57	Passed
9	ARIMA (1,2,0)	0.2156	0.1389	1208.95	Passed
10	ARIMA (0,2,0)	0.7734	2.975e-08	1243.44	Not passed

F. Forecasting

Forecasting the Stock price of PT Bank Mandiri for the next 10 days started from July 9, 2021 by using ARIMA (0, 2, 1) and ARIMA (1,2,1) are presented at Figure 6 and Figure 7, respectively.

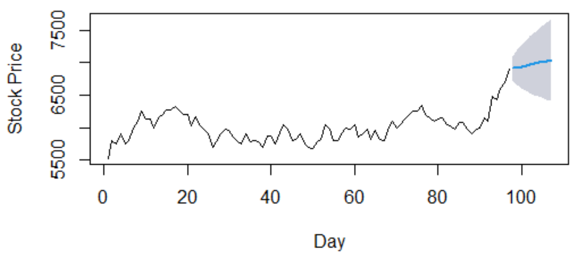

Figure 6. Forecast using ARIMA (0,2,1)

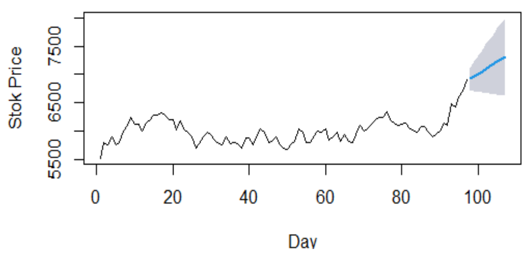

Figure 7. Forecast using ARIMA (1,2,1)

In these figures, the blue line indicates the prediction data, the black line indicates the actual data, and the gray are indicate the confident level area 90 %.

TABLE 5
FORECASTING VALUE OF THE ARIMA MODEL (0, 2, 1)

Tanggal	Forecast point	Actual point	Lower bound	Upper bound
9 Juli 2021	6,914.323	6,900	6,724.215	7,104.431
10 Juli 2021	6,928.646	7,000	6,658.411	7,198.882
11 Juli 2021	6,942.969	7,100	6,610.316	7,275.623
12 Juli 2021	6,957.292	7,175	6,571.242	7,343.343
13 Juli 2021	6,971.615	7,150	6,537.845	7,405.386
14 Juli 2021	6,985.939	7,175	6,508.420	7,463.457
15 Juli 2021	7,000.262	7,200	6,481.961	7,518.562
16 Juli 2021	7,014.585	7,175	6,457.815	7,571.355
17 Juli 2021	7,028.908	7,200	6,435.532	7,622.284
18 Juli 2021	7.043.231	7,125	6,414.786	7,671.676

TABLE 6
FORECASTING VALUE OF THE ARIMA MODEL (1, 2, 1)

Tanggal	Forecast point	Actual point	Lower bound	Upper bound
9 Juli 2021	6,921.344	6,900	6,731.293	7,111.395
10 Juli 2021	6,967.086	7,000	6,708.658	7,225.514
11 Juli 2021	7,009.496	7,100	6,689.817	7,329.175
12 Juli 2021	7,052.361	7,175	6,676.642	7,428.080
13 Juli 2021	7,095.164	7,150	6,666.116	7,524.213
14 Juli 2021	7,137.976	7,175	6,657.239	7,618.712
15 Juli 2021	7,180.786	7,200	6,649.343	7,712.229
16 Juli 2021	7,223.596	7,175	6,642.011	7,805.182
17 Juli 2021	7,266.407	7,200	6,634.959	7,897.855
18 Juli 2021	7,309.217	7,125	6,627.987	7,990.447

Table 5 and Table 6 indicate that the approximate value is relatively the same as the actual value and does not go outside the lower and upper limits. So, we can assume that this forecasting is going well. To get best model of the both of the mode, next comparison the error measures in forecasting using ARIMA (0,2,1) and ARIMA (1,2,1) are given at Table 7.

TABLE 7
ERROR MEASURES FOR FORECASTING USING ARIMA(0,2,1) AND ARIMA(1,2,1)

	ARIMA(0,2,1)	ARIMA (1,2,1)
MSE	24,650	7,022
RMSE	157.005	83.800
MAE	144.090	67.770
MAPE	2.014%	0.950%

Table 7 show that the error value of ARIMA (1,2,1) is smaller than ARIMA (0,2,1). Thus, the ARIMA (1,2,1) model is the best model. From Table 12, the model can be expressed in the form of equation below

$$Y_t = -0.1366Y_{t-1} + e_t - 0.9420e_{t-1} \tag{22}$$

IV. CONLUSION

In this work, the forecasting of PT Bank Mandiri Tbk Stock price start from 3 April 2021 until 8 July 2021 has been discussed. The result show that the best model is ARIMA (1, 2, 1) with equation that

$$Y_t = -0.1366Y_{t-1} + e_t - 0.9420e_{t-1}.$$

The error of the forecasting for next ten days using the model give MSE, RMSE, MAE and MAPE are 7,023, 67.77 an 0.950%, respectively. This model is very good because this model has very small error between actual data dan forecast data. This result is expected useful for the investor as their consideration before they are making a business decision related to PT Bank Mandiri Tbk.

REFERENCES

[1] PT Bank Mandiri, "Profil Perusahaan," Available: https://www.bankmandiri.co.id/profil-perusahaan

[2] PT Bursa Efek Indonesia, "Saham," Available: https://www.idx.co.id/produk/saham/

[3] M. Iqbal and N.W. Ningsih, "Prediksi harga saham harian PT. BTPN Syariah menggunakan model ARIMA dan model GARCH," *Jurnal Ilmiah Ekonomi Islam,* vol. 7, no. 3, pp. 1573-1580, 2021, doi:http://dx.doi.org/10.29040/jiei.v7i3.2795

[4] G.S. Lilipaly, D. Hatidja, and J.S. Kekenusa, "Prediksi harga saham PT. BRI menggunakan metode ARIMA," *Jurnal Ilmiah Sains*, vol. 14, no. 2, pp. 60-67, October 2014.

[5] E.D.N. Sari, "Peramalan harga saham perbankan dengan metode ARIMA Box Jenkins," Indonesia, Available: https://rstudio-pubs-static.s3.amazonaws.com/4532835d44bb0f19a54bceb1867a6840d99c1b.html.

[6] Investing.com, "Bank Mandiri Persero Tbk PT (BMRI)," Indonesia, Available: https://www.investing.com/equities/bank-mandiri-t-historical-data. Retrive on April 14, 2022.

[7] J.D. Cryer and K.S. Chan, *Time Series Analysis With Application in R Second Edition*, 2nd ed. Lowa City: Springer, 2008.