ARIMA Model and Holt Winters Seasonal Smoothing Accuracy for Stock Price Prediction

Agus Sofian Eka Hidayat[1], Putu Darmawan[2]

[1] *Universitas Presiden, Jababeka, Indonesia*

[2]*Universitas Prasetiya Mulya, Tanggerang, Indonesia*
**Corresponding author: agus.eka@president.ac.id*

Abstract— One way to generate capital is by investments in the capital market, with the expectation that this money will increase in tandem with an issuer's stock price. In the past few decades, technology for communication and information has developed incredibly quickly, followed by the introduction of 5G, which boosts data transfer rates. This study intends to compare the accuracy of SRIMA and Holt's Winter Seasonal Smoothing in predicting the stock price of TLKM as TLKM has shown the trend and seasonal effect yearly. The SARIMA $(\mathbf{1},\mathbf{1},\mathbf{1})\ (\mathbf{0},\mathbf{1},\mathbf{1})^{\mathbf{52}}$ the model outperforms Holt's Winter Seasonal Smoothing both additive and multiplicative in any aspect of MAPE, MAE, or RMSE.

Keywords— forecasting, time series, Holt's Winter Seasonal Smoothing, Seasonal ARIMA, TLKM

I. INTRODUCTION

The capital market is one way to get funding for government-owned and private companies, as well as an investment vehicle for investors who have these funds [1]. Because the capital market generally serves two roles, namely the economic function and the financial function, it plays a significant role in the economy of a nation. The capital market offers facilities that bring together two interests consisting of parties who have excess funds (investors) and parties who need funds (issuers) to carry out its function as an economic function. While in the financial function, the capital market provides opportunities and opportunities to obtain returns for the owners of funds, following the characteristics of the chosen investment. According to Rintana and Hania [3], the capital market is a market that trades securities as proof of ownership of a business company or ownership of capital to be invested by the agreement that has been made.

In the era of globalization, which is growing rapidly, it has an impact on the advancement of communication technology. In this study, researchers will predict the stock price movement of TLKM (PT Telkom Indonesia) which is one of the national telecommunications companies. PT Telkom Indonesia (Persero) Tbk is a State-Owned Enterprise that is engaged in information and communication technology services as well as Indonesian telecommunications networks. Telkom Indonesia shares began to join the Indonesian stock exchange (Initial Public Offering / IPO) on November 14, 1995, with the stock code TLKM [4].

The development of TLKM shares is interesting to study because communication technology will be more advanced coupled with the presence of 5G technology. There has not been much research on the effect of 5G technology on a stock issuer. However, based on research from Deutsche Bank, the time difference is very visible when viewed from the advantages and disadvantages of several industry groups related to 3G and 4G technology. The arrival of 5G itself is an interesting thing to discuss regarding the share price of communication technology issuers. Quoted from Trefis, several companies following 5G developments, such as Qualcomm (QCOM), Qorvo (QRVO), Skyworks Solutions (SWKS), Keysight Technologies (KEYS), American Tower (AMT), have experienced an increase in returns since 2019 by 30%.

The models used in this study are Holt's Winter Exponential Smoothing and the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. Both The SARIMA model and Holt's Winter Exponential Smoothing were chosen because they can handle trend and seasonal data well. In the data, there are several variables such as Date, Open, High, Low, Close, and Adj Close. In this case, the Close variable will be used for later forecasting. This study aims to predict the stock price of TLKM within a certain period so that the model obtained from the calculation results can be useful for investors in making decisions.

II. METHODE

Montgomery [5] states that forecasting is a prediction of an event in the future, and a time series is a time-oriented or chronological order of observations on the variables studied. Another opinion from Jason Brownlee [6], states that the time series is a sequence of observations made based on the sequence of time. In conducting time series forecasting, several steps must be taken such as [7] problem statement, ingestion data, exploration data, pre-processing and future engineering data, and model building, selection, and development.

A. SARIMA

Seasonal Method Autoregressive Integrated Moving Average (SARIMA) is one of the Time Series Forecasting models for data that contain seasonal patterns. SARIMA is used to overcome the Autoregressive Integrated Moving Average (ARIMA) where there are seasonal patterns in the time-series data [8]. The general form of SARIMA is ARIMA $(p,d,q)(P,D,Q)^S$ with p,d,q is the non-seasonal part of the model, and (P, D, Q) s is the seasonal part and s is the number of periods for each season [8]. The following is the general equation of the Seasonal ARIMA model [9][10]

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t = \phi_0 + \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{3}$$

with

$\Phi_P(B^s)$	: AR Seasonal
$\phi_p(B)$	: AR Non-Seasonal
$(1-B)^d$	: Differencing Non-Seasonal
$(1-B^s)^D$	: Differencing Non-Seasonal
$\theta_q(B)$	: MA Non-Seasonal
$\Theta_Q(B^s)$	: Mas Seasonal

B. Holt-Winter's Model

Holt-Winters's Model is used when the data shows the behavior of trend and seasonal. This method is also known as the exponential smoothing method approaches [11]. In the model-building stage, the thing that needs to be considered is the decomposition process. The decomposition procedure is used to explain trend and seasonality factors in time series data. Time series data that show a trend are time series data that are not stationary (nonstationary). Decomposition is generally divided into two types which are.

additive decomposition

$$Y_t = T_t + S_t + R_t \tag{4}$$

and Multiplicative Decomposition

$$Y_t = T_t.S_t.R_t \tag{5}$$

Additives are used when the seasonal variation of the data tends to be constant. Meanwhile, multiplicative is used when the seasonal variation of the time series data changes and is not constant.

additive decomposition

$$y_{t+h} = l_t + hb_t + s_{t+h-m(k+1)} \quad (6)$$

$$l_t \quad = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) \quad (7)$$

$$b_t \quad = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \quad (8)$$

$$s_t \quad = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \quad (9)$$

while the Multiplicative Model

$$y_{t+h} = (l_t + hb_t)s_{t+h-m(k+1)} \quad (10)$$

$$l_t = \alpha\left(\frac{y_t}{s_{t-m}}\right) + (1-\alpha)(l_{t-1} + b_{t-1}) \quad (11)$$

$$b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1} \quad (12)$$

$$s_t = \gamma\left(\frac{y_t}{l_{t-1} + b_{t-1}}\right) + (1-\gamma)s_{t-m} \quad (13)$$

with

l_t : component level

b_t : trend component

s_t : seasonal component

α, β^*, γ : smoothing parameter

m : seasonal frequency

C. Model Accuracy

Several methods are used to calculate the accuracy of the model. This is essential because we need to consider whether our model can inform about the actual data accurately. There are three methods used for calculating the accuracy of the model such as the Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) values. For

MAE

$$MAE = \frac{1}{n}\sum |y_i - \hat{y}_i| \quad (14)$$

for MSE

$$MSE = \frac{1}{n}\sum (y_i - \hat{y}_i)^2 \quad (15)$$

while RMSE

$$RMSE = \sqrt{\frac{1}{n}\sum (y_i - \hat{y}_i)^2} \quad (15)$$

III. RESULT AND DISCUSSION

D. SARIMA Process

To find out more about the distribution of TLKM stock prices, trends, and seasonal effects, then the decomposition method is carried out. The decomposition procedure is used to explain trend and seasonality factors in time series data. Time series data that show a trend are time series data that are not stationary. Based on the results of the decomposition plot in Figure 1, historical data can be seen from TLKM stock prices. In the trend section, the data tends to experience an up-and-sloping trend in 2018 and behind. In the seasonal section, the data has an annual seasonal pattern. And from the residual plot, the residuals from the data tend to be constant. From Figure 1, it can be seen from historical data from TLKM stock prices, that the data tends to experience an upward and sloping trend in 2018 and behind. Furthermore, the data has an annual seasonal pattern and from the residual plot, the residuals from the data tend to be constant.

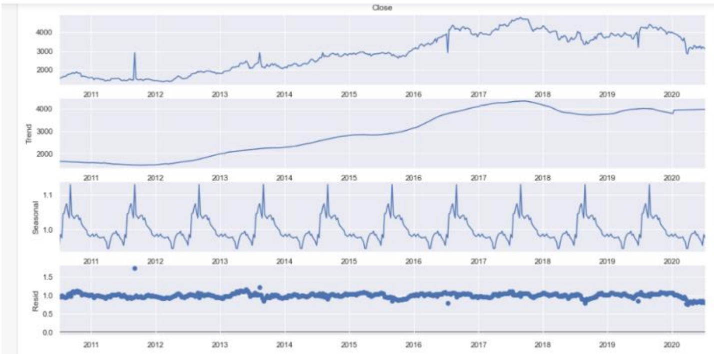

Figure. 1 Time Series Data Decomposition Plot

1) Stationary Test

The test was carried out using the Augmented Dicky-Fuller (ADF Test) test method on Python software. Stationary evaluation of time series data can be determined based on the p-value of the ADF test. The assumption of stationary time series data is fulfilled if the p-value is less than 5% or 0.05.

TABLE 1

ADF TEST

ADF Statistic	**P-*value***	***Critical Value***		
		1%	**5%**	**10%**
-1.381001	0.591350	-3.443	-2.867	-2.570

Based on the results from the table above where the p-value and the Autocorrelation function (ACF) and Partial autocorrelation function (PACF) plots, it can be concluded that the time series data is not stationary. Therefore, it is necessary to adjust using the differencing method

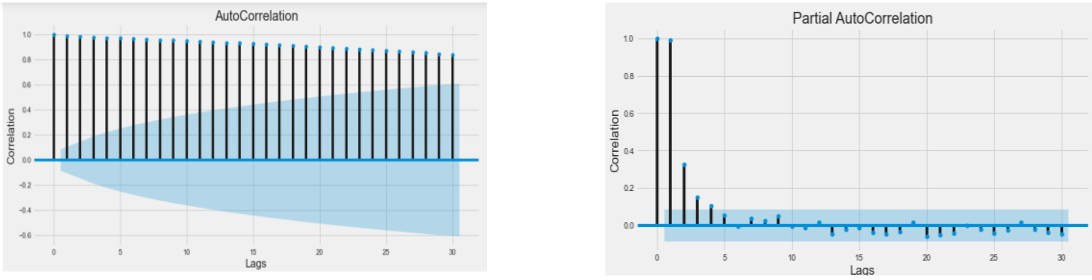

Figure. 2 ACF and PACF Plot

2) Differencing

In general, differencing is only done once, which is known as first-order differencing. However, if the data is still in a non-stationary condition, then proceed with second-order differencing, namely by differencing the previous data which has been transformed into first-order differencing. After first-order differencing, the data has a constant mean and variance. the ADF test shows that for critical values 1%, 5%, and 10% the $p\ value\ < 0.05$.

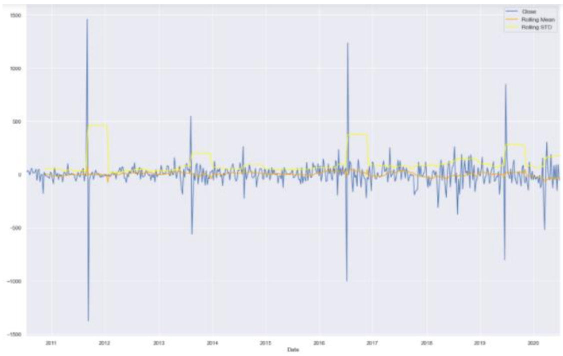

Figure. 3 Time Series After 1st Differencing

Further From the ACF and PACF plots, we can find out the order of AR and MA. The plot above shows spikes in lag 1 of the ACF plot which indicates the order of Moving Average (MA) A (1) and PACF also at lag 1 which indicates the order of Autoregressive AR (1). AR (1) and MA (1) become the base models for underfitting overfitting using the Grid Search method or auto Seasonal ARIMA where the p-value <5%.

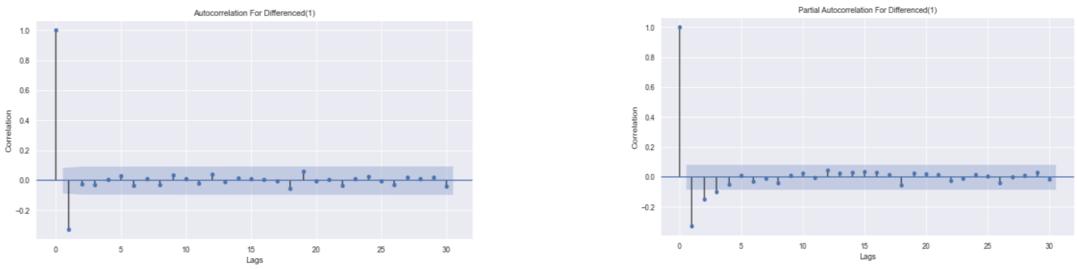

Figure. 2 ACF and PACF Plot After 1st Differencing

3) Underfitting and Overfitting Model using SARIMA

The results of underfitting and overfitting to find the order of the SARIMA model with the smallest AIC value indicate the significance of the model. Some of the results from 64 Iteration are given below. Based on these results, the parameter order of the Seasonal ARIMA model with the smallest AIC value is obtained, namely Seasonal ARIMA with parameters $(1, 1, 1)\ (0, 1, 1)^{52}$ and the minimum AIC 5350.19.

TABLE 2

UNDERFITTING AND OVERFITTING

Seasonal ARIMA (p, d, q) x (P, D, Q) 52						
AR(p)	**d**	**MA(q)**	**AR(P)**	**D**	**MA(Q)**	**AIC**
0	0	0	0	0	0	9871.18
0	0	0	0	0	1	8411.62
.	.	.	.	.	.	.
1	1	1	0	1	1	5350.19
.	.	.	.	.	.	.
1	1	1	1	1	1	5351.08

4) Summary of Model Selection

The plot in Figure 3 shows the residual diagnostics of the ARIMA Seasonal model $(1, 1, 1)$ x $(0, 1, 1)^{52}$. It can be seen from the standardized residual plot and histogram plot; the model residual fluctuates at an average value of 0 and is normally distributed with the average value is 0. The Q-Q plot of the model shows the distribution of residual values is located or coincides with the red line which indicates that the residual data is normally distributed. From the correlogram plot (see Figure 4), the residual model is not autocorrelated which is indicated by the absence of a line that crosses the blue area boundary

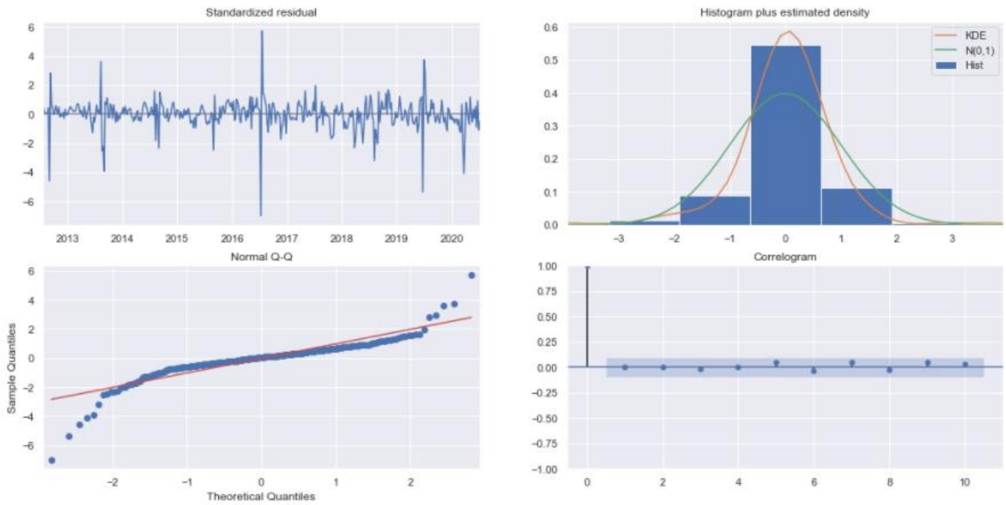

Figure. 3 Diagnostic Model

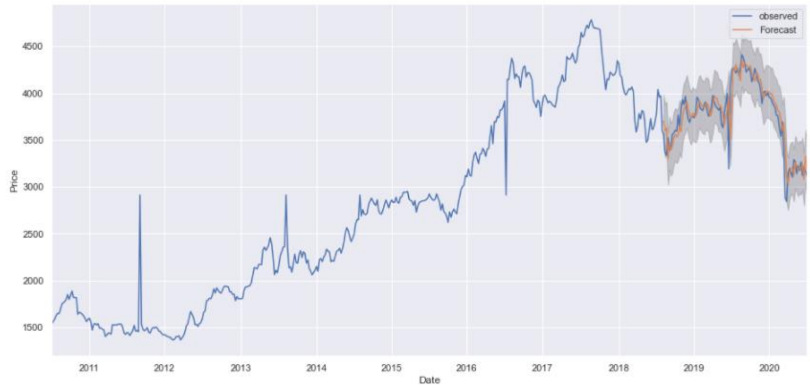

Figure. 4 Model Accuracy by Data Observed and Data Test Forecasting

Based on the results obtained from the calculation of model accuracy, the Mean absolute percentage error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) values. From this accuracy, by taking the limit of 11% or 0.11 for the limit of the MAPE value, it can be said that the accuracy of the model is very good because it is far below 11%.

TABLE 3

MODEL ACCURACY

MAPE	MAE	RMSE
0.03	107.7932	160.1193

E. Holt's Winter Seasonal Smoothing

5) Additive Model and Multiplicative Model

From the plot of the additive model, the additive model when compared to the SARIMA model, the results are worse because it does not manage to describe the test data well. Meanwhile, the multiplicative model produces forecasting test data that is almost the same as the additive model and when compared with the SARIMA model, this model looks less good at describing time series data.

Figure. 5 Additive Plot Model

Figure. 6 Multiplicative Plot Model

The accuracy model of additive and multiplicative is closed as it shows. However, the SARIMA model has the biggest accuracy among the others. Based on the results of the accuracy model described in table 4 it can be concluded that the SARIMA model has the best accuracy among the three models with a MAPE value of 0.03 or 3%. Then the SARIMA model will be used for forecasting.

TABLE 4

COMPARISON OF ACCURACY OF SARIMA AND HOLT'S

	SARIMA	HWSS	
		Additive	**Multiplicative**
MAPE	0.03	0.094	0.093
MAE	107.7932	318.649	318.24
RMSE	160.1193	483.66	476.84

F. Forecasting

Based on the results of the TLKM stock forecasting plot above, for the next 2 years from July 2020, the TLKM stock price is sideways in the range of Rp 3,000 – Rp 3,500, with a 95% CI. The plot above is a plot of the actual data of TLKM's stock price. When compared with the forecasting plot of the SARIMA model, both show identical results, namely sideways at Rp 3,000 – Rp 3,500. So, it can be said that the SARIMA $(1,1,1)\ (0,1,1)^{52}$ the model has been able to describe the actual data well.

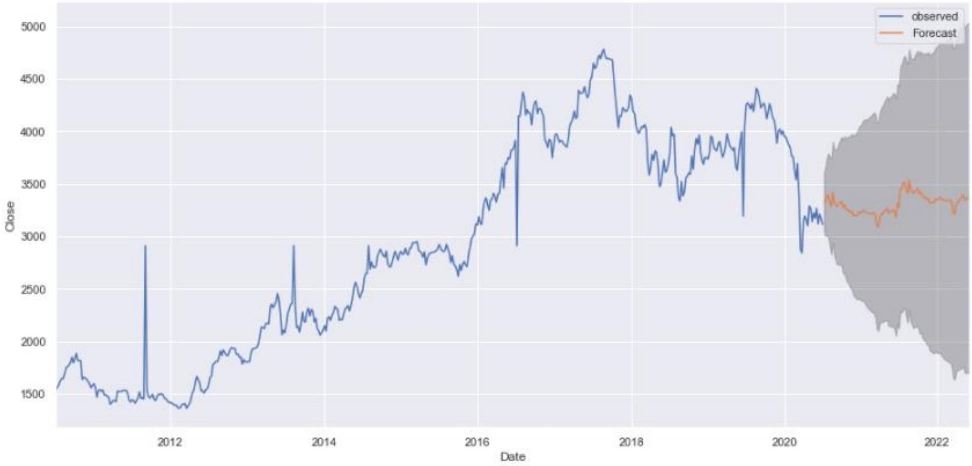

Figure. 6 Forecasting using SARIMA $(1,1,1) \times (0,1,1)^{52}$

IV. CONCLUSION

The forecasting of the closing price of TLKM from July 2010 to July 2020 shows that there are increasing trends but start to decline in 2018. Further, it shows that seasonal effects appear yearly. Forecasting result, using both SARIMA and Holt's Winter Seasonal shows that SARIMA $(1,1,1)\ (0,1,1)^{52}$ model outperforms Holt's Winter Seasonal Smoothing in both additive and multiplicative with a 3% MAPE. For, MAE and RMSE also show that the SARIMA model has better model prediction compared to Holt's. Results from the forecasting of SARIMA $(1,1,1)\ (0,1,1)^{52}$ model, the stock price of TLKM will move sideways for the next two years starting in July 2020 in the region of Rp. 3,000 to Rp. 3,500. When compared to actual data from TLKM stock prices, the SARIMA $(1,1,1)(0,1,1)^{52}$ forecasting model's output is showing identical results.

REFERENCES

[1] OJK. "Pasar Modal Seri Literasi Perguruan Tinggi, Otoritas Jasa Keuangan dan Industri Jasa Keuangan", Indonesia, 2016
[2] Mufidah, Eva. "Analisis laba, Arus kas operasi dan nilai buku ekuitas terhadap harga saham." *Eksis: Jurnal Riset Ekonomi dan Bisnis* 12, no. 1 (2017).
[3] Herman, Rintana, and Hania Rahma. "Mengenal dan Memahami Pasar Modal." (2012).
[4] Telkom Indonesia. "Profil dan Riwayat Singkat. Retrieved from telkom.co.id: https://www.telkom.co.id/sites/about-telkom/id_ID/page/profil-dan-riwayat-singkat-22 (2021, April 23)
[5] Montgomery, Douglas C., Cheryl L. Jennings, and Murat Kulahci. Introduction to time series analysis and forecasting. John Wiley & Sons, 2015.
[6] Brownlee, Jason. Introduction to time series forecasting with Python: how to prepare data and develop models to predict the future. Machine Learning Mastery, 2017.
[7] Lazzeri, Francesca. *Machine learning for time series forecasting with Python*. John Wiley & Sons, 2020.
[8] Dubey, Ashutosh Kumar, Abhishek Kumar, Vicente García-Díaz, Arpit Kumar Sharma, and Kishan Kanhaiya. "Study and analysis of SARIMA and LSTM in forecasting time series data." Sustainable Energy Technologies and Assessments 47 (2021): 101474.
[9] Box, George EP, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
[10] Kibunja, Hellen W., John M. Kihoro, George O. Orwa, and Walter O. Yodah. "Forecasting precipitation using SARIMA Model: A case study of Mt. Kenya Region." (2014).

[11] Kalekar, Prajakta S. "Time series forecasting using holt-winters exponential smoothing." *Kanwal Rekhi school of information Technology* 4329008, no. 13 (2004): 1-13.