

Application of Double Seasonal Autoregressive Integrated Moving Average (DSARIMA) for Stock Forecasting

Mega Silfiani¹, Farida Nur Hayati², Muhammad Azka³

^{1,2}Department of Statistics, Institut Teknologi Kalimantan

³Department of Actuarial Science, Institut Teknologi Kalimantan

megasilfiani@lecturer.itk.ac.id¹, farida.nur@lecturer.itk.ac.id², muhammad.azka@lecturer.itk.ac.id³

Submitted 2 Mei 2023 Revised 24 Juni 2023 Accepted 27 Juni 2023

Abstract

Background: Stock price forecasting assists investors to anticipate risks and opportunities in making prudent investments and maximizing returns.

Objective: This study aims to identify the most accurate model for stock forecasting.

Methods: This paper utilized the daily closing stock price of Unilever Indonesia, Tbk (UNVR) from January 1, 2018 to July 31, 2022. Double Seasonal Autoregressive Integrated Moving Average (DSARIMA), was utilized in this study. Mean Absolute Scaled Error (MASE) and Median Absolute Percentage Error (MdAPE) are used to compare forecasting accuracy.

Results: Following conducting each model, we assessed that the best models are DSARIMAX (0,1,[4]) ([3],1,1)⁵(1,1,0)²⁵³, regarding MASE and MdAPE corresponding to approximately 1.423 and 0.111. The scope of this study has limitations to a test set for one-month forecast periods.

Conclusion: As stock prices rise, investors require precise forecasts. Models of forecasting must perform well. This analysis shows how the DSARIMA generate forecasts stock prices more accurately. This investigation evaluated the closing stock price of UNVR. Both MASE and MdAPE assess prediction. After analyzing each model, DSARIMAX (0,1,[4]) ([3],1,1)⁵(1,1,0)²⁵³ has the lowest MASE and MdAPE values, 1.423 and 0.111, respectively. The procedure lasted one month. Research may combine forecasts and improve their accuracy.

Keywords : Double SARIMA, Forecasting, MASE, MdAPE, Stock.

Abstrak

Latar Belakang: Peramalan harga saham dapat membantu investor untuk mengantisipasi risiko dan peluang dalam investasi dan memaksimalkan pengembalian.

Tujuan: Tujuan dari penelitian ini adalah untuk mendapatkan model terbaik untuk peramalan saham.

Metode: Penelitian ini menerapkan harga saham penutupan harian Unilever Indonesia, Tbk (UNVR), dari 1 Januari 2018 hingga 31 Juli 2022 dengan metode double seasonal autoregressive integrated moving average (DSARIMA). Mean Absolute Scaled Error (MASE) and Median Absolute Percentage Error (MdAPE) digunakan untuk membandingkan akurasi peramalan.

Hasil: Setelah dilakukan peramalan menggunakan masing-masing model, model terbaik yang kami dapatkan adalah DSARIMAX (0,1,[4]) ([3],1,1)⁵(1,1,0)²⁵³ yang memiliki MASE dan MdAPE masing-masing sebesar 1,423 dan 0,111. Penelitian ini dibatasi testing data adalah satu bulan.

Kesimpulan: Ketika harga saham naik, investor membutuhkan ramalan yang tepat. Model peramalan harus berkinerja baik. Analisis ini menunjukkan bagaimana DSARIMA menghasilkan peramalan harga saham yang lebih akurat. Investigasi ini mengevaluasi harga penutupan saham UNVR. Baik MASE dan MdAPE menilai prediksi. Setelah dilakukan analisis terhadap masing-masing model, DSARIMAX (0,1,[4]) ([3],1,1)⁵(1,1,0)²⁵³ memiliki nilai MASE dan MdAPE terkecil, masing-masing sebesar 1,423 dan 0,111. Penelitian dapat menggabungkan peramalan dan meningkatkan keakuratannya.

Kata kunci: Double SARIMA, Forecasting, MASE, MdAPE, Stock.

INTRODUCTION

Stocks play a significant role in the modern financial environment, providing as an essential channel for investment, asset creation, and business growth. Individuals have the potential to accumulate capital through the purchase of stocks. They enable investors to gain benefits from appreciation in capital and dividends, and also engage in the accomplishments of businesses. International capital flows are facilitated by stock markets, which contribute to global integration. They facilitate the participation of foreign investors in the expansion of international economies, thereby nurturing economic interdependence.

The market's interest in Unilever (UNVR) stock is intense. This is because UNVR stock have a favorable reputation for quality and the products are readily accessible. In addition, UNVR stocks are defensive (movement-stable) stocks. In recent years, the stock price of UNVR has declined. The decline was attributable to a decline in home and personal care (HPC) segment earnings. Despite the fact that this segment will account for over 67% of total annual revenue in 2021, it will continue to decline. In addition, COVID 19 and market competition have an impact on the business. Stock prices fluctuate erratically on a daily basis, necessitating investors to conduct technical analysis to reduce their risk exposure. Forecasting future fluctuations in is a strategy for mitigating this risk.

The importance of stock forecasting in current volatile financial markets cannot be overstated. Forecasting future price movements and trends enables investors and traders to make informed decisions. It

provides insightful information that facilitates the making of informed investment decisions. Therefore, it can acquire an increased comprehension of the behaviour of stocks, leading them to reach more thoughtful conclusions. It assists investors by discovering prospective victors, preventing poor investments, and optimising portfolio composition. Furthermore, it plays an important role in decision-making activities (Khoiriyah and Cahyani, 2022)

Forecasting of future stock price can be done by time series analysis (Rifai, 2019). Time series analysis is an approach used to forecast future values based on historical data (Silfiani and Lembang, 2023). The autoregressive integrated moving average (ARIMA) model is one of the most common and well-known time series analyses and its enhanced applications have provided excellent accuracy in forecasting in a wide range of domains (Li, Wu, and Liu, 2023). Numerous fields that applied ARIMA in forecasting are unemployment rate (Yamacli and Yamacli, 2023), oil production (Ning, Kazemi and Tahmasebi, 2022), daily reservoir inflow (Gupta and Kumar, 2022) and stock price (Hayati and Ulama (2016) and Putri and Setiawan, (2015)). ARIMA is frequently employed because its application and interpretation are simple (Perone, 2020). Furthermore, ARIMA model have some extensions that accomodates seasonality in the time series data.

The Double Seasonal Autoregressive Integrated Moving Average (DSARIMA) is an extension of the ARIMA model that incorporates two seasonal periods. Several studies on the DSARIMA model, especially on electrical load data, have been conducted (Mohamed, et al., 2010 and Dinata et al., 2020). The findings of the investigation indicate that the double seasonal ARIMA model is a good model for forecasting (Mohamed, et al., 2010). Furthermore,

Modo, Soeprijanto and Suhartono (2018) also applied DSARIMA for Electrical Power Demand Forecasting. The best model for forecasting the data is the DSARIMA $([1,2,7,16,18,35,46],1,[1,3,13,21,27,46]) (1,1,1)^{48}(0,0,1)^{336}$ with MAPE is 2.06%.

Stock market prices change over time and sometimes have seasonal patterns (Rahmadianto, Lesmana and Budiarti, 2022). Therefore, this study applied double seasonal autoregressive integrated moving average approach in UNVR stock price in order to forecast future stock price changes.

METHOD

Research Design

The research design in this research follows as Figure 1.

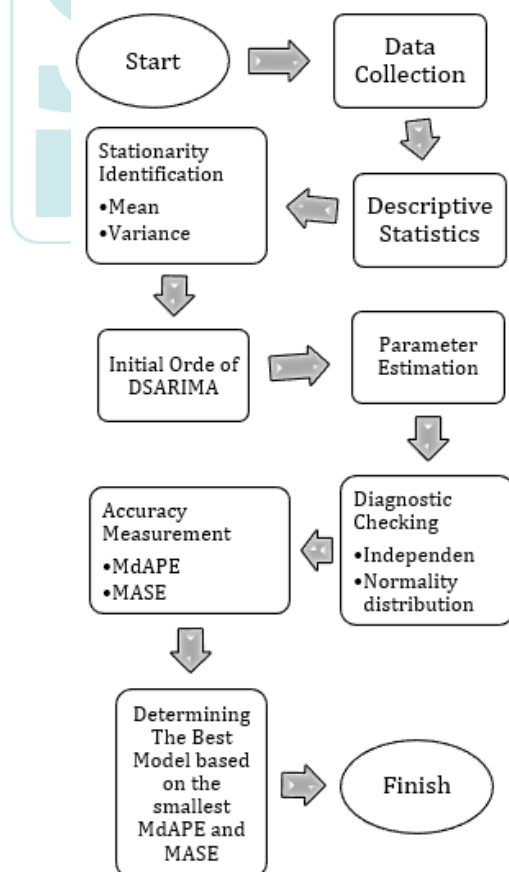


Figure 1. Research Design

Research Subjects

The dataset in this research are UNVR, a stock from PT Unilever Indonesia Tbk. It produces, sells, and distributes household goods in Indonesia. The business provides soaps, detergents, dairy products, ice cream, savoury snacks, soy sauce, cosmetics, tea, and fruit juices. Unilever Indonesia Tbk is a subsidiary by Unilever PLC.

Data Analysis Technique

We applied DSARIMA is a variation of the autoregressive integrated moving average (ARIMA) to construct forecasting model. In addition, we calculated mean absolute scaled error (MASE) and median absolute percentage error (MdAPE) to evaluate forecasting measurement accuracy. The explanation about DSARIMA and forecasting measurement accuracy as follows as:

DSARIMA is a variation of the Autoregressive Integrated Moving Average (ARIMA) that includes two seasonal periods. The seasonal model's orders imply that there are of two seasonal changes in the observed data. In general, DSARIMA model denotes by (1) (Dinata, et al., 2020).

$$Z_t = \frac{\theta_q(B)\theta_{Q_1}(B^{S_1})\theta_{Q_2}(B^{S_2})}{\phi_p(B)\phi_{P_1}(B^{S_1})\phi_{P_2}(B^{S_2})(1-B)^d(1-B^{S_1})^{D_1}(1-B^{S_2})^{D_2}} \tag{1}$$

where:

$\phi_{P_1}(B^{S_1}) = 1 - \phi_1 B^{S_1} - \phi_2 B^{2S_1} - \dots - \phi_{P_1} B^{P_1 S_1}$ is a component of first seasonal autoregressive, $\phi_{P_2}(B^{S_2}) = 1 - \phi_1 B^{S_2} - \phi_2 B^{2S_2} - \dots - \phi_{P_2} B^{P_2 S_2}$ is a component of second seasonal autoregressive $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ is a component of nonseasonal autoregressive, $(1 - B)^d$ is a nonseasonal differencing, $(1 - B^{S_1})^{D_1}$ is a first seasonal differencing, $(1 - B^{S_2})^{D_2}$ is a second seasonal differencing, $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ is a component of nonseasonal moving average, Z_t is a closing stock price at t-th period, $\theta_{Q_1}(B^{S_1}) = 1 - \theta_1 B^{S_1} - \theta_2 B^{2S_1} - \dots - \theta_{Q_1} B^{Q_1 S_1}$ is component of first seasonal moving average and $\theta_{Q_2}(B^{S_2}) = 1 - \theta_1 B^{S_2} - \theta_2 B^{2S_2} - \dots - \theta_{Q_2} B^{Q_2 S_2}$ is a component of the second seasonal moving average.

The same techniques as non-seasonal ARIMA apply to the construction of DSARIMA. Box and Jenkins created ARIMA using four processes, namely identification, estimate, diagnostic testing, and forecasting (Hyndman and Koehler, 2006). Through model identification, data series features including such seasonality and stationarity are discovered (Bowerman, O'Connell and Koehler, 2005). Applying model estimation, we may estimate the DSARIMA parameters. Thirdly, diagnostic checking is used to examine the white noise and normality distributions of residuals. The authors have considered outliers in the DSARIMA model if the residuals do not follow to a normal distribution. In time series, Additive Outliers (AO) and Level Shifts are two types of outliers (LS). In general, the DSARIMA model with outliers follows (2) (Bowerman, O'Connell and Koehler, 2005).

$$Z_t = \sum_{j=1}^k \omega_j v_j(B) I_j^{(T_j)} + \frac{\theta_q(B)\theta_{Q_1}(B^{S_1})\theta_{Q_2}(B^{S_2})}{\phi_p(B)\phi_{P_1}(B^{S_1})\phi_{P_2}(B^{S_2})(1-B)^d(1-B^{S_1})^{D_1}(1-B^{S_2})^{D_2}} N_t \quad (2)$$

where $I_j^{(T_j)}$ is an outlier at period T_j , and $v_j(B)=1$ is for AO, $v_j(B) = \frac{1}{(1-B)}$ is for LS. If seasonal double SARIMA satisfies all of the assumptions, then we can develop a forecasting.

To measure accuracy performance of the forecasting model, we applied MASE and MdAPE. The MASE and MdAPE can be expressed as (3) and (4) as follows:

$$MASE = \left(\frac{\sum_{j=1}^J |Z_j - \hat{Z}_j|}{\frac{1}{T-1} \sum_{t=2}^{T-1} (Z_t - Z_{t-1})} \right) \quad (3)$$

$$MdAPE = med \left(\frac{|Z_t - \hat{Z}_t|}{|Z_t|} \right) \quad (4)$$

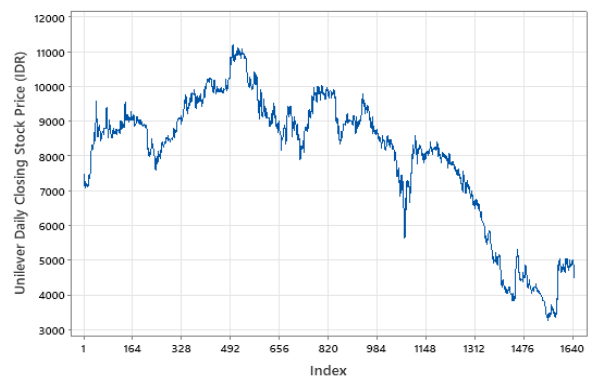
RESULT AND DISCUSSION

Unilever (UNVR) closing prices on weekdays are gathered. Each week, we have five data points. The central tendency and dispersion of the dataset are determined using the mean and standard deviation. According to Table 1, the largest and smallest annual averages for UNVR from 1 January 2018 to 29 July 2022 are IDR 9447.5 and IDR 4205, respectively. In 2018 and 2021, the largest and smallest standard deviations were IDR 850.7 and 1123.7, respectively. Table 1 also indicates that 253 observations are made each year.

The construction of a DSARIMA requires four phases. These processes include identification, estimation, diagnosis, and forecasting. Stationarity and seasonality were detected throughout the identifying process.

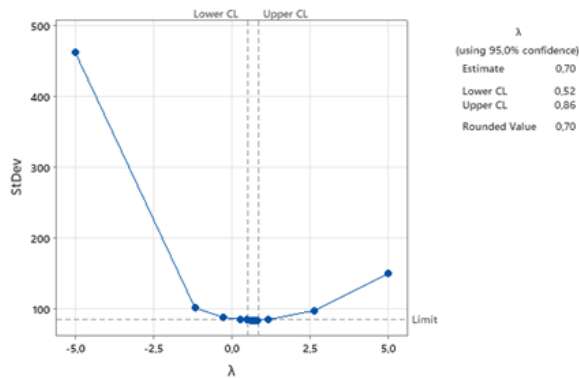
Table 1. Descriptive statistics of Unilever daily closing stock price.

Year	N	Mean	StDev
2018	261	9447.5	850.7
2019	258	9157.8	516.2
2020	242	7815.2	508.3
2021	247	5369.2	1123.7
2022	137	4205	582.4

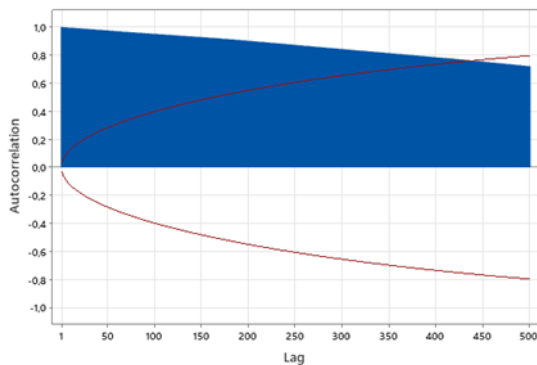


(a)

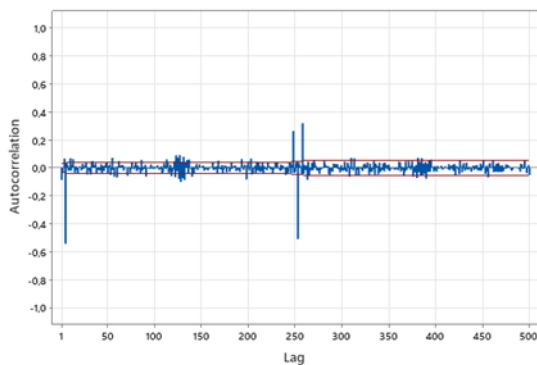
Application of Double Seasonal Autoregressive Integrated Moving....



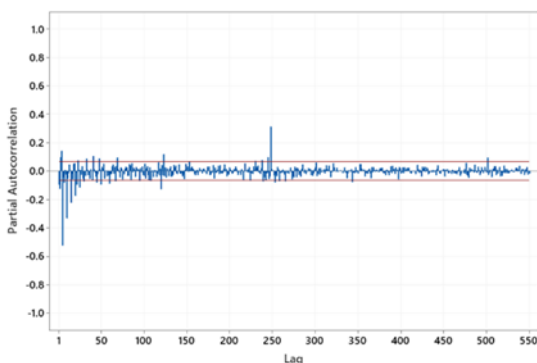
(b)



(c)



(d)



(e)

Figure 2. (a) Time series plot, (b) Box-cox transformation plot, (c) Autocorrelation, (d) Autocorrelation after differencing $d=1$, $d_1=5$ and $d_2=253$ and (e) Autocorrelation after differencing $d=1$, $d_1=5$ and $d_2=253$ of UNVR

Given the nonstationarity and seasonality of the time series data depicted in Figure 2a, UNVR is currently demonstrating a declining trend. On the basis of data features, it is anticipated that UNVR has two types of seasonality, namely weekly and annual seasonality, with 5 (5 days in a week) and 253 days, respectively (253 days in a year). In addition to examining the mean stationarity, we also analyzed the variance stationarity. We found that neither the mean nor the standard deviation of the UNVR was constant (Figure 2. b). Figure 2 demonstrates that the autocorrelation function lacks a lag, indicating that UNVR is a nonstationary process (c). To solve this issue, we created the nonseasonal differencing of $d=1$ and the double seasonal differencing of $D_1=5$ and $D_2=253$.

After establishing that the data fulfill the stationarity constraints for mean and variance, the pattern of the Autocorrelation Function (ACF) and Partial Autocorrelation Function will be employed to determine the order of DSARIMA. It is possible to construct orders for DSARIMA utilizing the seasonal ARIMA calculation approach based on ACF and PACF. The procedure proposed by Bowerman, O'Connell, and Koehler (2005) would be used to achieve the same outcome.

Figure 2 (d) show that lag of ACF cut off in lag 1, 2, 3, 4, 5, 122, 123, 127, 128, 248, 253, and 258. In addition, Figure 2 (e) show that lag of PACF cut off in lag 1, 2, 3, 4, 5, 10, 15, 120, 123, and 248. We construct the initial models DSARIMA (4,1,0)(3,1,0)⁵(1,1,0)²⁵³, DSARIMA (0,1,4)(0,1,1)⁵(0,1,1)²⁵³ and DSARIMA (4,1,0)(0,1,1)⁵(0,1,1)²⁵³ based on the ACF and PACF patterns in Figure 2 (d) and Figure 2 (d).

The estimate phase determines the values of the DSARIMA parameters and examines the significance of each parameter

Application of Double Seasonal Autoregressive Integrated Moving....

using the t-test. We used conditional least squares to estimate the DSARIMA parameters, and we set the significance level (α) at 5% to determine whether or not the parameters were significant. If one or more parameters do not pass the t-statistic test, the model must be reconstructed until each parameter passes the test. The subsequent step is the diagnostic testing phase. It is used to examine white noise and the normal distribution of DSARIMA's residual. We applied the Ljung-Box test and the Kolmogorov-Smirnov distribution test with a significance threshold (α) of 5% in order to examine the white noise and

normality distributions of residuals.

Consider that the residual does not fulfil the white noise assumption until lag 60. In such a circumstance, the model must be reconstructed, and the estimation process must be repeated until all parameters and the white noise criterion are fulfilled. If the residual distribution is not normal during this phase, additional actions such as outlier detection will be required. Frequent outliers prevent the residual from satisfying the requirements of the normal distribution. In order to confirm that the residuals have a normal distribution, it is necessary to include the outliers into the model.

Table 2. DSARIMA model of UNVR

Model	Number of Outlier		Significant Parameter	Diagnostic checking	
	Level Shift	Additional Outlier		White Noise Test	Normality test
DSARIMAX (0,1,[4]) ([3],1,1) ⁵ (1,1,0) ²⁵³	18	13	Satisfied	Satisfied	Satisfied
DSARIMAX ([19],1,[4]) (2,1,1) ⁵ (1,1,0) ²⁵³	17	13	Satisfied	Satisfied	Satisfied

Table 2 shows that there are two models of DSARIMA that satisfied all the test, i.e., DSARIMAX (0,1,[4])([3],1,1)⁵(1,1,0)²⁵³ and DSARIMAX ([19],1,[4])(2,1,1)⁵(1,1,0)²⁵³. The letter 'X' in each of these two models indicates that there are outliers in the data. The DSARIMAX (0,1,[4]) ([3],1,1)⁵(1,1,0)²⁵³ has 31 outliers, which is 18 level shift outliers and 13 additional outliers. Meanwhile DSARIMAX ([19],1,[4])(2,1,1)⁵(1,1,0)²⁵³ has 30 outliers, which is 17 level shift outliers and 13 additional outliers.

(0,1,[4])([3],1,1)⁵(1,1,0)²⁵³ demonstrates the most successful performance in terms of prediction. It is the most promising model since it has the lowest value for all metrics of accuracy prediction. Meanwhile, the model with the lowest predictive performance is DSARIMAX ([19],1,[4])(2,1,1)⁵(1,1,0)²⁵³.

CONCLUSION

As the stock market and stock prices gain greater significance, investors are going to need to construct forecasting models that are precise. It is crucial to have forecasting models that have a high-performance forecasting. This study aims to determine which DSARIMA form best forecasts stock prices. This investigation used UNVR's daily closing stock price from January 1, 2018 to July 31, 2022. MASE and MdAPE can be utilized to assess predictive performance. After evaluating each model, we discovered that DSARIMAX (0,1,[4])([3],1,1)⁵(1,1,0)²⁵³ had the best

Table 3. Summary DSARIMA of UNVR

Model	Forecasting Accuracy Measurement	
	MASE	MdAPE
DSARIMAX (0,1,[4])([3],1,1) ⁵ (1,1,0) ²⁵³	1.423	0.111
DSARIMAX ([19],1,[4])(2,1,1) ⁵ (1,1,0) ²⁵³	1.485	0.119

According to Table 3, the DSARIMAX

MASE and MdAPE values, 1.423 and 0.111. This study utilizes a test set with one-month projected horizons. Further research may involve combining or hybridizing forecasts to boost forecasting performance.

BIBLIOGRAPHY

- Bowerman, B. L., O'Connell, R.T.O., Koehler, A. B. (2005). "Forecasting, Time series, and Regression: an Applied Approach 4th ed", USA: Brook/Cole, Thomson Learning, Inc.
- Dinata, S.A.W., Azka, M., Faisal, M., Suhartono, Yendra, R. and Gamal, M. D. H. (2020). "Short-Term load forecasting double seasonal ARIMA methods: An evaluation based on Mahakam-East Kalimantan data", AIP Conference Proceedings 2268, 020004, <https://doi.org/10.1063/5.0017643>.
- Gupta, A., and Kumar, A. (2022). "Two-step daily reservoir inflow prediction using ARIMA-machine learning and ensemble models," *Journal of Hydro-environment Research*, 45, 39-52, <https://doi.org/10.1016/j.jher.2022.10.002>.
- Hayati, F. N. and Ulama, B. S. S. (2016) "Peramalan Harga Saham Jakarta Islamic Index Menggunakan Metode Vector Autoregressive," *Jurnal Sains dan Seni ITS*, 2337-3520. <http://dx.doi.org/10.12962/j23373520.v5i2.16931>
- Handani, S. and Astawinetu, E. D. (2020). *Teori Portofolio dan Pasar Modal Indonesia*, Scopindo Media Pustaka.
- Hyndman, R.J. and Koehler, A.B. (2006) "Another look at measures of forecast accuracy", *International Journal of Forecasting*, 22 (4), 679-688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Khoiriyah, N., and Cahyani, N. (2022). "Peramalan Banyaknya Pasien Rawat Jalan dengan Menggunakan Metode Brown's Double Exponential Smoothing," *Jurnal Statistika Dan Komputasi*, 1(1), 23-30, <https://doi.org/10.32665/statkom.v1i1.451>.
- Li, Y., Wu, K., and Liu, J. (2023). "Self-paced ARIMA for robust time series prediction," *Knowledge-Based Systems*, 269, 110489, <https://doi.org/10.1016/j.knosys.2023.110489>.
- Mohamed, N., Ahmad, M. H., Ismail, Z. and Suhartono. (2010). "Double Seasonal ARIMA Model for Forecasting Load Demand," *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, 217-231. <https://doi.org/10.11113/matematika.v26.n.565>
- Mado, I., Soeprijanto, A., and Suhartono, (2018). "Applying of Double Seasonal ARIMA Model for Electrical Power Demand Forecasting at PT. PLN Gresik Indonesia," *International Journal of Electrical and Computer Engineering (IJECE)*, 4892-4901. <http://doi.org/10.11591/ijece.v8i6.pp4892-4901>
- Ning, Y., Kazemi, H., Tahmasebi, P. (2022). "A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet," *Computers & Geosciences*, 164, 105126, <https://doi.org/10.1016/j.cageo.2022.105126>.
- Paningrum, D. (2022). *Buku Referensi Investasi Pasar Modal*, Kediri: Lembaga Chakra Brahmanda Lentera.

- Perone, G. (2020). "An ARIMA Model to Forecast the Spread and the Final Size of COVID-2019 Epidemic in Italy," HEDG - Health Econometrics and Data Group Working Paper Series, University of York. <https://doi.org/10.48550/arXiv.2004.00382>
- Putri, R. N and Setiawan. (2015). "Peramalan Indeks Harga Saham Perusahaan Finansial LQ45 Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) dan Vector Autoregressive (VAR)," Jurnal Sains dan Seni ITS, vol. 4, (2015). <http://dx.doi.org/10.12962/j23373520.v4i2.11162>
- Rahmadianto, B., Lesmana, D.C., Budiarti, R. (2022). Prediksi Harga Saham BBCA dan BMRI dengan Model Seasonal ARIMA", Undergraduate Thesis, Actuarial, IPB University.
- Rifai, N.A.K. (2019). "Pendekatan Regresi Nonparametrik dengan Fungsi Kernel untuk Indeks Harga Saham Gabungan," Statistika, 19(1), 53-61. <https://doi.org/10.29313/jstat.v19i1.4775>
- Silfiani, M., and Lembang, G.R. (2023). "Perbandingan Peramalan Jumlah Kasus Kecelakaan Lalu Lintas Kota Balikpapan dengan Linear Trend Analysis dan Double Exponential Smoothing," EQUIVA JOURNAL 1 (1), 14-18. <https://doi.org/10.35718/equiva.v1i1.757>
- Yamacli, D.S., Yamacli, S. (2023). "Estimation of the unemployment rate in Turkey: A comparison of the ARIMA and machine learning models including Covid-19 pandemic periods," Heliyon, 9(1),