Vol. 2 No. (2) (December 2024) 61-71

Journal of Data Insights

http://journalnew.unimus.ac.id/index.php/jodi

Forecasting Red Onion Prices in Riau Islands Using the Seasonal Autoregressive Integrated Moving Average (SARIMA) Method

Revika Inta Nur Kholifah[1], Ihsan Fathoni Amri[2], M.Al Haris[3], Nasyiatul Izzah[4], Miftakhiyah Fazza Baita[5], Siti Nurhalisa[6]

[1] *Department of statistics*, Universitas Muhammadiyah Semarang, Indonesia
[2] *Department of data science*, Universitas Muhammadiyah Semarang, Indonesia
[3] *Department of statistics*, Universitas Muhammadiyah Semarang, Indonesia
[4] *Department of statistics*, Universitas Muhammadiyah Semarang, Indonesia
[5] *Department of statistics*, Universitas Muhammadiyah Semarang, Indonesia
[6] *Department of statistics*, Universitas Muhammadiyah Semarang, Indonesia

DOI: https://doi.org/10.26714/jodi.v2i2.412

Article Information

Article History:
Submitted: 15th July 2024
Revised: 16th December 2024
Accepted: 31th December 2024

Keywords:
SARIMA; Red Onion Prices; Forecasting; MAPE.

Abstract

The price of shallots is one of the crucial commodities that affects economic stability and community welfare in the Riau Islands. The main factors influencing shallot production are seed variety, land, and weather. This study aims to forecast the price of shallots in the Riau Islands using the Seasonal Autoregressive Integrated Moving Average (SARIMA) method. The data used in this study is sourced from official data and covers a specific period to ensure the accuracy of the forecasting model. The SARIMA (0 1 1) (0 1 1)^5 model with the smallest AIC of 2211.59 was selected as the best model based on data analysis and model performance evaluation, with a Mean Absolute Percentage Error (MAPE) of 2.690835 percent, indicating that the model's ability to predict shallot prices in the Riau Islands is very accurate. The prediction results indicate that the price of shallots will decrease in the coming days according to the developed model. Based on these results, this forecast is expected to serve as a reference for the government and market participants in decision-making related to the production, distribution, and control of shallot prices in the Riau Islands.

✉ *Correspondence address:*
E-mail: ihsanfathoni@unimus.ac.id

e-ISSN: 2988 - 2109

INTRODUCTION

Shallots are a horticultural crop that plays an important role and is in high demand among the community. Shallots have significant economic value in Indonesia and are the third-largest cultivated commodity after chili and cabbage (Zulfa, Nhita, & Saepudin, 2019). Shallots are one of the most well-known and popular spices among Indonesians, primarily due to their ability to add a distinctive flavor to dishes. Additionally, shallots offer health benefits for the body. Given the extensive uses and benefits of shallots in daily life, the demand for this commodity continues to rise. This presents significant opportunities for marketing, both in domestic markets and for export, with the potential to boost national income and meet existing market needs (Fajriyanto, Syukur, & Supriyanto, 2017).

Based on retail food price data from the National Food Agency (Bapanas) on Monday, May 13, 2024, at 1:48 PM WIB, the price of shallots in the Riau Islands reached IDR 40,210 per kilogram. Last week, this price increased by IDR 2,190 (5.76%). Meanwhile, there was an increase of IDR 6,070 (17.78%) compared to 30 days prior. Shallot production in Riau Province reached 323 tons in 2023, up from 192 tons the previous year. This indicates that shallot production increased by 68.23% in 2023, yet shallot prices are still expected to rise significantly in the short term (weekly and monthly trends). This highlights the essential role of shallots in the daily lives of Indonesians (Jatmiko, Rahayu, & Darmawan, 2017).

However, the uncertain supply of shallots can lead to price fluctuations influenced by several factors (Zulfa, Nhita, & Saepudin, 2019). Changes in shallot prices are influenced by uncertainties in production caused by variability in harvest seasons and weather, the perishable nature of the commodity, and suboptimal management(Windhy, Suci, & Jamil, 2019). The instability of shallot prices impacts both producers and consumers. Producers face high risks due to price fluctuations during harvest, while consumers experience a decline in purchasing power when shallot prices are high. Therefore, price forecasting analysis is needed to predict the market prices that will occur (Rosyid, Viana, & Saputro, 2021)Additionally, the estimation is conducted to predict shallot production in order to avoid stock shortages, so that farmers do not incur losses due to a lack of shallot supply (Sunariadi, Intan, Novitasari, & Hariningsih, 2022).

Forecasting is a comprehensive process to obtain the necessary methods to estimate future values, which can then be used as input to achieve objectives (Lip, et al., 2020). In other words, the goal of forecasting is to generate predictions that can minimize prediction errors, which are typically measured by mean squared error, mean absolute error, and similar metrics (Pangestu,S. 2013) (Dimashanti & Sugiman, 2021). One method that can be used to forecast shallot prices is Seasonal Autoregressive Integrated Moving Average (SARIMA). Seasonal Autoregressive Integrated Moving Average (SARIMA) is a model developed from Autoregressive Integrated Moving Average (ARIMA) to analyze time series data that exhibit seasonal patterns (Durrah, Yulia, Parhusip, & Rusyana, 2018). In this study, the SARIMA model is used to forecast shallot market prices because it can produce low prediction errors. The best SARIMA model can be determined based on the lowest MAPE (Mean Absolute Percentage Error) value (Sunariadi, Intan, Novitasari, & Hariningsih, 2022). This study aims to develop more effective guidelines for planning shallot production and marketing in Riau. Therefore, this research is expected to observe changes in shallot prices in Riau over the next five days, from May 14 to May 18, 2024.

METHODS

2.1 Data Sources

This study is a quantitative research that uses secondary data obtained from the Databooks website(https://databoks.katadata.co.id/datapublish/2024/05/13/harga-bawang-merah-di-kepulauan-riau-sebulan-terakhir-naik-rp6070) this refers to the data on shallot prices in the Riau Islands. The data used is in the form of daily time series from January 1 to May 13, 2024, with a total of 134 days of data.

2.2 SARIMA *(Seasonal AutoRegressive Integrated Moving Average)*

The SARIMA (Seasonal AutoRegressive IntegrateISLd Moving Average) model is a time series model that combines autoregressive (AR) components, moving average (MA), and integration (differencing) to handle non-stationary data with seasonality. SARIMA is an extension of the ARIMA model that includes seasonal components to address recurring patterns in time series data that are periodic in nature (Rizki & Taqiyyuddin, 2021). The notation SARIMA (p, d, q) $(\mathrm{P}, \mathrm{D}, \mathrm{Q})^S$ contains parameters that define the dynamics of the model, both in the seasonal and non-seasonal components. The SARIMA equation can be written as follows :

$$\phi_p(B^S)\phi_p(B)\,(1 - B)^d\,(1 - B^S)^D Y_t = \theta_q(B)\ominus_Q(B^S)\varepsilon_t$$

With

p, P	: *Autoregressive*
d, D	: *Differencing*
q, Q	: *Moving Average*
$\phi_p(B)$	: *Non-seasonal Autoregressive order*
$\phi_p(B^S)$	: *Seasonal Autoregressive order*
$(1 - B)^d$	: *Non-seasonal differencing order*
$(1 - B^S)^D$	: *Seasonal differencing order*
$\theta_q(B)$	: *Non-seasonal Moving Average*
$\ominus_Q(B^S)$	: *Seasonal Moving Average*
Y_t	: *Actual data at time t*
ε_t	: *Error at period t*

2.3 Forecasting

Forecasting is the effort to predict future conditions based on the analysis of historical data (Rismawati & Darsyah, 2018). This forecasting approach is also useful for business or organization owners to estimate future product sales, making it easier for them to make decisions related to strategies for increasing or decreasing production (Ahmad, 2020). In general, forecasting is often done using time series data provided by the Central Statistics Agency (BPS) at the district, provincial, or national level, as this data influences government policy decisions. The accuracy of predictions will certainly result in different outcomes when presented with different techniques (Prabuningrat, Salma, Muharamah, Haris, & Nur, 2023). Various approach methods have been applied to measure the errors that arise in certain forecasting techniques. Almost all of these measures utilize some functions of the actual values and the forecasted results. One way to evaluate the results of forecasting techniques is by using a measure that contains information about the difference between the forecasted results and the actual demand that occurs in the field (Ahmad, 2020).

2.4 Stages of Data Analysis

The steps of analysis used to predict shallot prices in the Riau Islands using the SARIMA method are as follows :

1. Describe the data of red onions in the Riau Islands
2. Create a visualization of the red onion data in the Riau Islands
   Data visualization helps in identifying trends, seasonality, and other components of a time series. Plot the time series data to observe trend and seasonal patterns.
3. Data Stationarity
   Time series data must be stationary for SARIMA analysis. Stationarity is tested using statistical tests such as the Augmented Dickey-Fuller (ADF) test to check whether the data has a trend or seasonal patterns that need to be removed. If the data is non-stationary, differencing is required, which involves calculating the change or difference in observation values from one period to the next. After differencing, the data is re-examined to check if it is now stationary. If it is still non-stationary, differencing is performed again. If the data variance remains non-stationary, logarithmic transformation is applied (Waryanto & Wanti, 2019).
4. Model Identification
   The SARIMA model is identified using the notation (p,d,q), $(P,D,Q)^S$.
5. Parameter Estimation
   After the model is identified, the parameters of the model are estimated. This is typically done using the Maximum Likelihood Estimation (MLE) method.
6. Model Significance Test
   After estimating the model parameters, the next step is to perform the model significance test.
7. Model Diagnostics
   Check the model residuals to ensure that they are random and follow a normal distribution. Residual plots and statistical tests, such as the Ljung-Box test, are used for this. The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals are also examined to ensure that all information has been captured by the model. Model diagnostics include two tests: the normality test and the white noise test using the Ljung-Box test.
8. Model Evaluation
   Evaluate the model's performance using Mean Absolute Percentage Error (MAPE) to measure prediction accuracy.
9. Best Model Selection
   The best model is the one that has significant parameters, residuals that show white noise and a normal distribution, the lowest AIC value, and the lowest Mean Absolute Percentage Error, which is used as a measure of the model's goodness (Ruhiat & Effendi, 2018). Next, the equation of the best model can be written mathematically.
10. Forecasting
    Use the SARIMA model to predict future Red Onion prices. Then, the forecast results are visualized and compared with actual data to assess the model's accuracy.
    The following analysis steps in the form of a flowchart are presented in **Figure 1.**

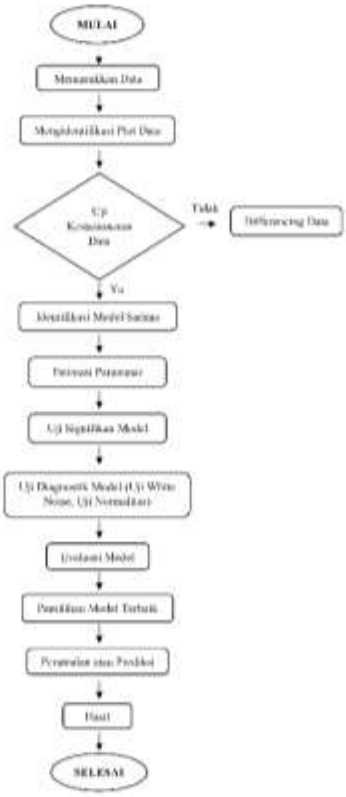


Figure 1. Flowchart for the steps in the SARIMA model analysis and forecasting

RESULTS AND DISCUSSION

3.1 Data Interpolation

Red onions are an essential daily necessity. They are widely considered a key ingredient in traditional Indonesian cuisine. However, the importance of red onions in daily life also makes price fluctuations a major issue. Therefore, it is necessary to record the daily prices of red onions so that they can be used for forecasting future prices. However, technical issues, weather conditions, natural disasters, and national holidays in the price recording process sometimes cause the loss of red onion price data for some periods. As a result, mathematical methods are often used to approach red onion price data to help recover the missing price data.

Interpolation is a numerical approach that performs calculations to estimate values between other values. On April 10 and 11, 2024, the red onion price data in the Riau Islands was missing or not recorded because those dates were national holidays, specifically the Eid al-Fitr holiday. Therefore, interpolation needs to be performed first to obtain the missing red onion price data for those dates.

3.2 Data Identification

After the data is interpolated, the next step, before modeling, is to display a plot of the red onion price data to understand the emerging patterns and identify the time series model that can be used to model the data characteristics. Below is the plot of the red onion price data using R Studio, shown in **Figure 2.**

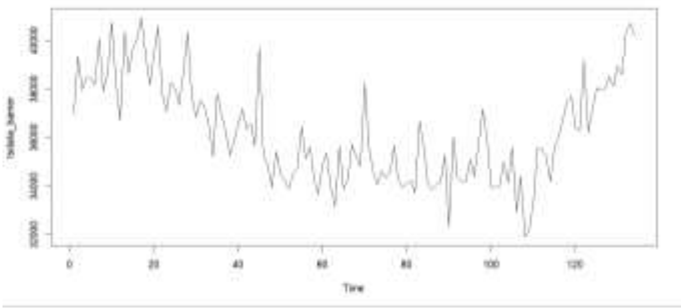
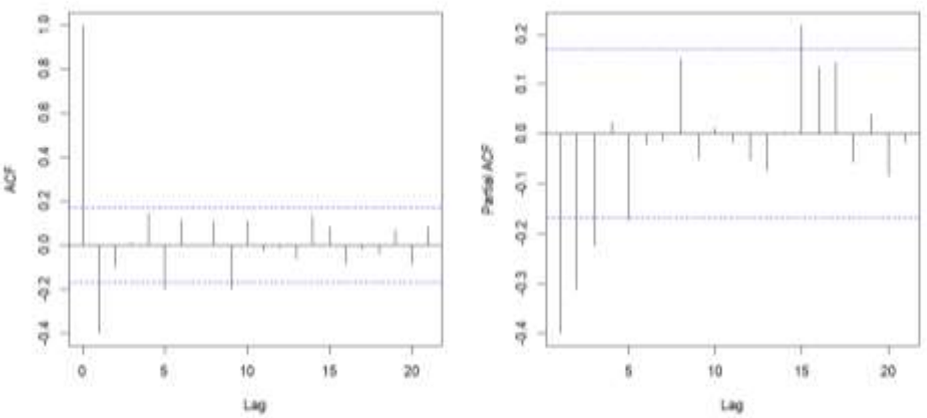


Figure 2. Time series data plot

3.3 Data Stationarity

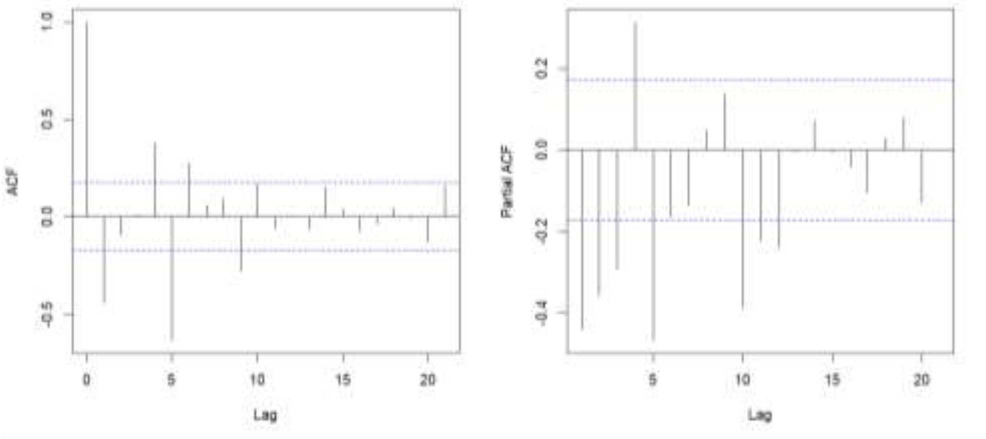
From Figure 2, it can be seen that the data is non-stationary and contains seasonal patterns. However, the stationarity of the data cannot be determined solely by using the plot. For greater accuracy, stationarity testing can be done using the Augmented Dickey-Fuller (ADF) test with a 5% significance level. In the ADF test using R Studio, a p-value of 0.99 was obtained, which means the data is not stationary because it is > 0.05. Since the data is not stationary, differencing needs to be performed.

After performing one differencing on the non-seasonal data (d=1), a p-value of 0.01 was obtained, which is less than 0.05, indicating that the data is now stationary in terms of mean. Next, to remove the seasonal effect, one differencing was performed on the seasonal data (D=1), resulting in a p-value of 0.01, which is also less than 0.05, indicating that the data is now stationary in terms of mean (the data no longer contains a unit root). The ACF and PACF plots for both non-seasonal and seasonal data are presented in **Figure 3** and **Figure 4.**



(a) (b)

Figure 3. (a). ACF plot for non-seasonal and (b). PACF plot for non-seasonal after differencing



(a) (b)

Figure 4. (a). ACF plot seasonal and (b). PACF plot seasonal after differencing

3.4 Model Identification

After the data is stationary, the next step is to identify the SARIMA model. This model identification is done by analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. From the ACF and PACF plots, the initial SARIMA model (p,d,q) (P,D,Q)S can be determined. Model identification can be seen from the number of lags that fall outside the significance limits.

a. ARIMA

The ARIMA model has the notation p,d,q, where order p represents the Autoregressive (AR) component, d represents differencing, and q represents the Moving Average (MA) component. For the non-seasonal model (p,d,q), the order p can be observed from the PACF plot, and the order q can be seen from the ACF plot. It is observed that the PACF plot decreases exponentially (dies down), indicating that order p is 0. The ACF plot decreases sharply (cut off) at the first lag, meaning that the lag used is lag 1. Therefore, the order q is 1. For d, since differencing was performed once at the beginning, order d is 1. Thus, the resulting ARIMA model is (0,1,1).

b. SARIMA

The SARIMA model has the notation $(p,d,q)(P,D,Q)^S$ where p is the Autoregressive (AR) order for the non-seasonal component, d is the differencing order for the non-seasonal component, and q is the Moving Average (MA) order for the non-seasonal component. Meanwhile, P is the Autoregressive (AR) order for the seasonal component, D is the differencing order for the seasonal component, and Q is the Moving Average (MA) order for the seasonal component. For the seasonal model identification (P,D,Q), order P can be observed from the seasonal PACF plot, and order Q can be seen from the seasonal ACF plot. It is observed that the PACF plot shows seasonality, evidenced by a cut-off at lags 5 and 10, indicating seasonality every 5 days. The plot decreases exponentially (dies down), meaning that order P is 0. The ACF plot decreases sharply (cut off) at lag 1, meaning that lag 1 is used. Therefore, order Q is 1. For d, since differencing was performed once at the beginning, order d is 1. Thus, the resulting SARIMA model is $(0,1,1)(0,1,1)^5$.

3.5 Model Significance Test

A good model is one that shows that the estimation of its parameters is significant.

a. ARIMA

After obtaining the model, the next step is to perform the significance test. The results of the ARIMA model significance test are presented in **Table 1.**

Table 1. ARIMA Significance Tests

Model	Coeft	Estimate	Pr (>|z|)	Significance	AIC
(0,1,1)	ma1	-0.647539	< 2.2e-16 ***	Significant	2277.76

b. SARIMA

After obtaining the model, the next step is to perform the significance test. The results of the SARIMA model significance test are presented in **Table 2.**

Table 2. SARIMA Significance Tests

Model	Coeft	Estimate	Pr (>|z|)	Significance	AIC
$(0,1,1)(0,1,1)^5$	ma1	-0.651138	< 2.2e-16 ***	Significant	2211.59
	sma1	-0.915583	< 2.2e-16 ***	Significant	

3.6 Model Diagnostic Test

The best model is diagnostically tested to ensure that it meets the assumptions of the residuals. There are two residual assumptions that are tested is the white noise test (residual independence) and the normality test of the residuals. Below is a comparison of the results of the white noise test in **Table 3.**

Hypothesis :

H0 : The data follows a White Noise distribution

H1 : The data does not follow a White Noise distribution

H0 is rejected if the p-value < α (0,05)

Table 3. White Noise Test

Model	P-value	White Noise
ARIMA (0,1,1)	0,756	White Noise
SARIMA $(0,1,1)^5$	0,7914	White Noise

The best model is diagnostically tested to ensure that it meets the assumptions of the residuals. There are two residual assumptions that are tested: the white noise test (residual independence) and the normality test of the residuals. Below is a comparison of the results from the residual normality test for the ARIMA and SARIMA models presented in **Table 4.**

Hypothesis :

H0 : The data follows a Normal distribution

H1 : The data does not follow a Normal distribution.

H0 is rejected if the p-value < α (0,05)

Table 4. Normality Test

Model	P-value	Normality
ARIMA (0,1,1)	0,224	Normal
SARIMA (0,1,1)[5]	0,7184	Normal

3.7 Model Evaluation

Model evaluation is conducted to determine which model can be considered the best based on its MAPE value. The model with the smallest MAPE and AIC values will be used for forecasting. The comparison results of MAPE and AIC are presented in **Table 5.**

Table 5. Results of the Comparison of MAPE and AIC

Model	MAPE (%)	AIC
ARIMA (0,1,1)	2.70548	2277.76
SARIMA (0,1,1)[5]	2. 690835	2211.59

Based on the MAPE values, it can be seen that the MAPE of SARIMA is smaller than that of ARIMA, which is 2.690835%. Therefore, the model that can be used for forecasting red onion prices is the SARIMA model $(0,1,1)(0,1,1)^5$.

3.8 Best Model Selection

From the diagnostic test and model evaluation using MAPE and AIC, the best model obtained is SARIMA $(0,1,1)(0,1,1)^5$. The model equation is written as follows :

$$(1 - B)^1 (1 - B^5)^1 Y_t = (1 + 0.651138)(1 + 0.915583^5)\varepsilon_t$$

3.9 Red Onion Price Forecasting

The model $(0,1,1)(0,1,1)^5$ is used to forecast the red onion prices in the Riau Islands. The following are the forecasted data for the next 5 days, from May 14 to May 18, 2024, presented in **Table 6.**

Table 6. Red Onion Price Forecast in the Riau Islands for the next 5 days

Date	Forecasted Price
May 14, 2024	40171.87
May 15, 2024	40019.43
May 16, 2024	40495.58
May 17, 2024	40292.08
May 18, 2024	40146.62

The plot of the forecasting results is presented in **Figure 6.**

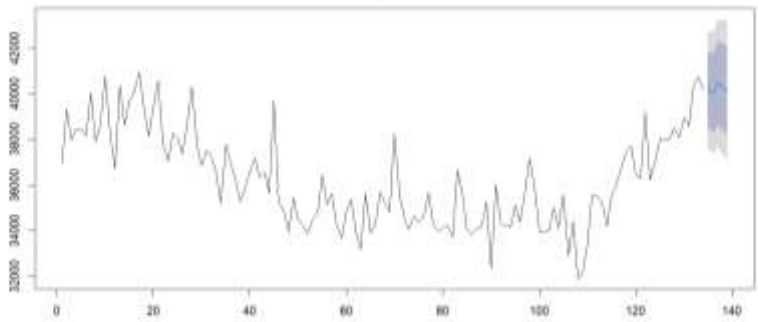


Figure 6. The plot of the red onion price forecast in the Riau Islands

CONCLUSION

The best model obtained is the SARIMA (0,1,1)(0,1,1)^5 with the model equation as follows:

$$Y_t = -0.651138_{\varepsilon_{t-1}} - 0.915583_{\varepsilon_{t-5}} + \varepsilon_t$$

Based on the research conducted to predict red onion prices in the Riau Islands, it was found that the Seasonal Autoregressive Integrated Moving Average (SARIMA) method is effective in forming the best prediction model. The model used is SARIMA (0,1,1)(0,1,1)^5, which has an AIC value of 2211.59 and a MAPE of 2.690835%. The prediction results from the SARIMA(0,1,1)(0,1,1)^5 model for the 5-day period, from May 14 to May 18, 2024, show a pattern similar to the actual data, although it experiences a decline.

REFERENCE

Ahmad, F. (2020). PENENTUAN METODE PERAMALAN PADA PRODUKSI PART NEW GRANADA BOWL ST Di PT.X. *JISI: Jurnal Integrasi Sistem Industri*, *7*(1), 31. https://doi.org/10.24853/jisi.7.1.31-39

Fajriyanto, Syukur, A., & Supriyanto, C. (2017). Optimasi Prediksi Tingkat Produksi Bawang Merah Nasional Menggunakan Metode Backpropagation Neural Network Berbasis Algoritma Genetika. *Jurnal Teknologi Informasi*, *13*, 115–124.

Harum Prabuningrat, S., Khoirunnafisa Salma, N., Wahyu Muharamah, P., Al Haris, M., & Saifuddin Nur, M. (2023). *Peramalan Indeks Harga Konsumen Kota Semarang dengan Metode Autoregressive Integrated Moving Average e-ISSN* (Vol. 1, Issue 1). http://journalnew.unimus.ac.id/index.php/jodi

Inka Durrah, F., Prihartina Parhusip, T., Rusyana, A., Statistika, J., Syiah Kuala, U., & Aceh, B. (2018). Peramalan Jumlah Penumpang Pesawat Di Bandara Sultan Iskandar Muda Dengan Metode SARIMA (Seasonal Autoregressive Integrated Moving Average). In *Journal of Data Analysis* (Vol. 1, Issue 1).

Jatmiko, Y. A., Rahayu, R. L., & Darmawan, G. (2017). Perbandingan Keakuratan Hasil Peramalan Produksi Bawang Merah Metode Holt-Winters Dengan Singular Spectrum Analysis (Ssa). *Jurnal Matematika "MANTIK," 3*(1), 13. https://doi.org/10.15642/mantik.2017.3.1.13-24

Model, P., Jenkins, B. O. X., Dalam, A., Konsumen, H., Merah, B., & Provinsi, D. I. (2021). *TENGAH. 13*(April), 29–37.

Mohd Lip, N., Jumery, N. S., Ahmad Termizi, F. A., Mulyadi, N. A., Anuar, N., & Ithnin, H. (2020). Forecasting International Tourist Arrivals in Malaysia Using Sarima and Holt-Winters Model. *Journal of Tourism, Hospitality and Environment Management*, *5*(18), 41–51. https://doi.org/10.35631/jthem.518004

Rismawanti, Y., & Darsyah, M. Y. (2018). Perbandingan Peramalan Metode Moving Average dan Exponential Smoothing Holt Winter Untuk Menentukan Peramalan Inflasi di Indonesia. *Prosiding Seminar Nasional Mahasiswa Unimus*, *1*(1), 330–335.

Rizki Dimashanti, A. (2021). PRISMA, Prosiding Seminar Nasional Matematika Peramalan Indeks Harga Konsumen Kota Semarang Menggunakan SARIMA Berbantuan Software Minitab. *PRISMA*, *4*, 565–576. https://journal.unnes.ac.id/sju/index.php/prisma/

Rizki, M. I., & Taqiyyuddin, T. A. (2021). Penerapan Model SARIMA untuk Memprediksi Tingkat Inflasi di Indonesia. *Jurnal Sains Matematika Dan Statistika*, *7*(2). https://doi.org/10.24014/jsms.v7i2.13168

Ruhiat, D., & Effendi, A. (2018). Pengaruh Faktor Musiman Pada Pemodelan Deret Waktu Untuk Peramalan Debit Sungai Dengan Metode Sarima. *Teorema*, *2*(2), 117. https://doi.org/10.25157/.v2i2.1075

Sunariadi, N. M., Keumala Intan, P., Candra, D., Novitasari, R., Hariningsih, Y., Matematika, P., Sunan, U., Surabaya, A., & Kabupatennganjuk, B. (2022). PREDIKSI PRODUKSI BAWANG MERAH DI KABUPATEN NGANJUK DENGAN METODE SEASONAL ARIMA (SARIMA). *Jurnal Pendidikan Matematika Dan Matematika*, *6*(1), 49–60. https://doi.org/10.36526/tr.v%vi%i.1672

Waryanto, H., & Wanti, D. A. (2019). Prediksi Penjualan Seragam Sekolah Dengan Menggunakan Metode Arima. *Jurnal Statistika Dan Matematika*, *1*(1), 88–102.

Windhy, A. M., Suci, Y. T., & Jamil, A. S. (2019). Analisis Peramalan Harga Bawang Merah Nasional Dengan Pendekatan Model Arima. *Seminar Nasional Pembangunan Pertanian Berkelanjutan Berbasis Sumber Daya Lokal*, 591–604.

Zulfa, H., Nhita, F., & Saepudin, D. (2019). Hibridisasi Algoritma Classification and Regression Tree (CART) dan Artificial Neural Network (ANN) untuk Prediksi Harga Bawang Merah di Kabupaten Bandung. *E-Proceeding of Engineering*, *6*(2), 9776–9784.