



Sales Prediction on the Diamond Cell Counter Using Autoregressive Integrated Moving Average (ARIMA) Method

Kristin Rahayu ¹, Putri Taqwa Prasetyaningrum²

¹Information Systems Study Program, Mercubuana University, Yogyakarta, Indonesia

²Lecturer of Information Systems Department, Mercubuana University, Yogyakarta, Indonesia

Email: ¹kristinrahayu07@gmail.com, ²putri@mercubuana-yogya.ac.id

Abstract

Diamond Cell is a specialized retailer that offers a diverse range of smartphone accessories, electronic credits, and internet vouchers from different providers, each with varying active periods. However, the uncertainty surrounding internet voucher sales transactions often leaves counter owners hesitant to increase their stock due to the short active period of the vouchers. This leads to frequent customer dissatisfaction as the internet vouchers run out, resulting in lost sales opportunities. To address this issue, this study aimed to predict voucher sales for the upcoming month to serve as a reference for calculating the stock of voucher supply. The Auto-regressive Integrated Moving Average (ARIMA) method was used based on voucher sales data from November 2022 to January 2023. Out of the three tentative models obtained, only one proved suitable for use. The best ARIMA model was the (2,1,0) model, with a MAD value of 29.65, an MSE value of 2409.95, and a MAPE value of 23.3%. Based on the February voucher sales, the stock level can remain the same as the previous period since the sales were stable. The findings of this study can help Diamond Cell counter owners make more informed decisions about stocking internet vouchers, resulting in better customer satisfaction and reduced likelihood of losses.

Keywords: Forecasting, ARIMA, Minitab, J-ISI

1. INTRODUCTION

The importance of management control in business cannot be overstated. It makes it easier to develop and manage various aspects of a business, particularly stock, financial reporting, and inventory control. The calculation of stock levels is particularly influential since inadequate stock can affect transactions, while excessive stock can lead to losses. Diamond Cell is one of the businesses that sells internet vouchers from various active periods and providers. On a daily basis, the Diamond Cell counter engages in non-fixed voucher sales transactions, and this sales uncertainty often causes counter owners to hesitate when determining the stock of each internet voucher, given their short active period. Consequently,



mistakes often occur during the voucher stocking process, which can lead to consumer disappointment when they run out of voucher stock. These frequent errors can cause losses for Diamond Cell counters.

Forecasting is a process of predicting future conditions by analyzing and calculating past conditions. Predicting future sales involves estimating the potential number of sales, as well as the market area to be captured in the future [1]. Forecasting data can be obtained through historical data projections using mathematical methods (quantitative), or subjective human considerations (qualitative) [2]. Depending on the time frame, forecasting can be categorized into three types: short-term (less than 3 months - 1 year), medium-term (less than 3 years), and long-term (more than 3 years) [3]. Time series data can be used to forecast using several models, including Exponential Smoothing, Moving Average, and ARIMA [4]. The Exponential Smoothing method is more suitable for forecasting stationary data, while the Moving Average model is more accurate when used for data that does not have seasonal and trend elements.

The ARIMA Box-Jenkins method is widely regarded as the best iterative model for stationary time series data [5]. This method has numerous advantages, including high accuracy, flexibility in capturing data patterns, and the ability to accurately forecast multiple variables. Furthermore, it is relatively simple, quick, and inexpensive to implement, as it only requires historical data for forecasting [6]. By using the ARIMA method, it is hoped that more accurate forecasting data can be generated, which can be used as a reference for calculating voucher stock and predicting consumer demand in the future. The use of Minitab software can facilitate the forecasting calculations, resulting in more precise and accurate results [7].

In Alit Fajar Kurniawan's (2021) study, the ARIMA method with the Carmer matrix was used to forecast the number of home sales in the Bojongsoang area during Christmas time. The best ARIMA model was found to be ARIMA (1,1,1), and the predicted number of home sales in the area from June to December showed a stable trend [8]. Similarly, Ensiwi Mursih aimed to predict the number of unemployed individuals in South Sumatra Province for the next ten years using an ARIMA (2,1,1) model with unemployment data from 2002 to 2016. The predicted results showed a decrease in the number of unemployed compared to the previous period [9]. Another study conducted by Mohammad Syaiful Pradana (2020) forecasted the value of NTP in the following year and identified the agricultural sub-sector that needs to be regulated by the Lamongan district government. The results showed that food crops with an NTP value were quite low, with an average of ≥ 100 per month for 3 years and experienced the highest decrease in NTP in 2019, which was 10.25% [10].

Lafnidita Farosanti (2022) conducted research to predict the number of microbiological sales to Denpasar in the next period, using ARIMA model testing (4,2,1) that yielded an RMSE value of 4.129(0.04129) [11]. Ida Bagus Bayu's (2022) study aimed to predict the sale of heml in Good Store stores in the future, using primary data from September 21, 2021, to December 21, 2021. The best ARIMA model was found to be (1,0,1) with a p-value smaller than 0.5 [12]. In the present study, we will apply the ARIMA method to predict voucher sales for the next 1-month period, specifically February 2013. The aim is to use this forecast as a reference material for ordering vouchers and accurately predict consumer demand.

2. METHODS

2.1 Data Description

The data used in this study is data taken from the voucher sales report at the Diamond Cell counter in November 2022 - January 2023.

2.2 Analysis Methods

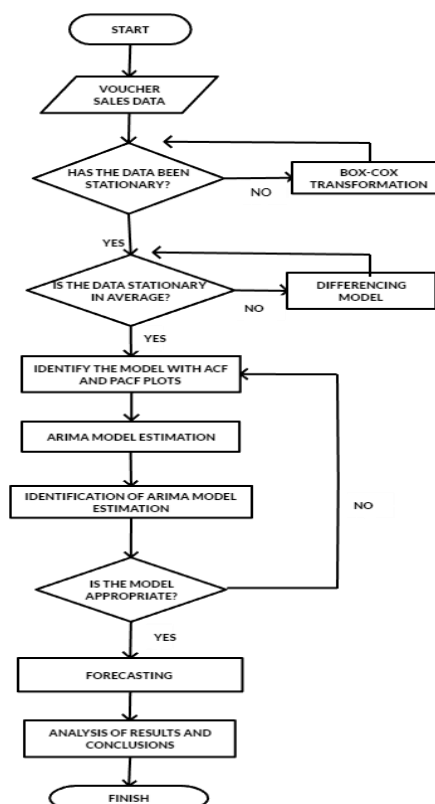


Figure 1. The path of research

2.2.1 Data preparation

The data used in forecasting is 3 months starting from November 2022 - January 2023.

2.2.2 Model Identification

The first step in identifying the model is to determine the variability of the data being used. This step aims to determine whether the data is stationary with respect to its variance and whether it is stationary in its mean. Data is said to be stationary with respect to its variance if the value of $\lambda=1$. If the data is not stationary, a transformation process is necessary. Testing the accuracy of the data with regards to its mean can be done by examining the lag on the ACF plot that is not patterned and contains seasonal elements. If the data is not stationary, then differencing is necessary.

2.2.3 Parameter Estimation

The objective of parameter estimation is to determine the values of the parameters in the ARIMA model. These estimated parameters will then be tested to determine their significance in the model. Once the model is estimated, a significance test will be conducted to determine whether the hypothesized parameters are significant or not. The process of parameter estimation is performed to obtain the best model with the highest accuracy rate and the smallest error value.

2.2.4 Diagnosis Checking

The purpose of this stage is to detect all residual correlations and lags in the estimation data that have been obtained and appropriately used in forecasting. The testing is conducted to determine whether the model is suitable or not. If it is not, then it is necessary to re-identify using a new model to achieve better results. If the residual turns out to be white noise, it indicates that the selected model is good. However, if the residual is not white noise, then the selected model is incorrect or inappropriate. Thus, it is necessary to search for better specifications.

2.2.5 Forecasting

Once the suitable model has been obtained from the diagnostic checking process, the next step is to initiate forecasting using the minitab 16.0 software. To evaluate the accuracy and reliability of the ARIMA model's forecasting outcomes, one can determine the MAD (Mean Absolute Deviation), which is the absolute average error value, MSE (Mean Squared Error), which is the average value of ranked error, and MAPE (Mean Absolute Percent Error), which is the average value of

absolute error percentage. The precision and quality of the forecasting results can be determined by the magnitude of the error value; the smaller the error value, the more precise and optimal the results will be.

3 RESULTS AND DISCUSSION

3.1 Data Identification

At this stage, data on voucher sales for three months, from November 2022 to January 18, 2023, were obtained. The sample data to be used for forecasting consists of 27 observations, covering the sales data for the three-month period, as illustrated in Figure 1.

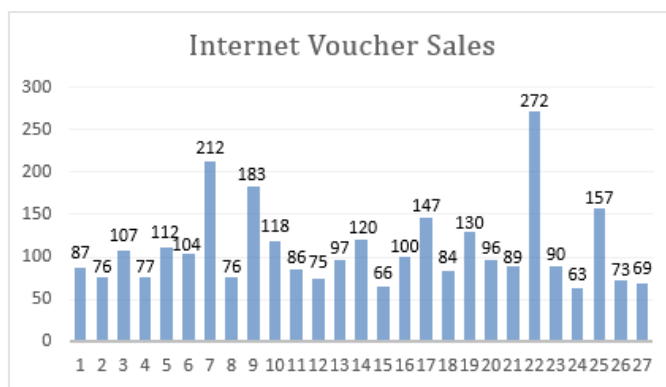


Figure 1. Internet Voucher Sales Data

To further analyze the training data, Minitab 16.0 software was used, and the results are presented in Figure 2. The time series plot in Figure 2 shows that there are seasonal fluctuations in voucher sales, with significant increases and decreases occurring in different months.

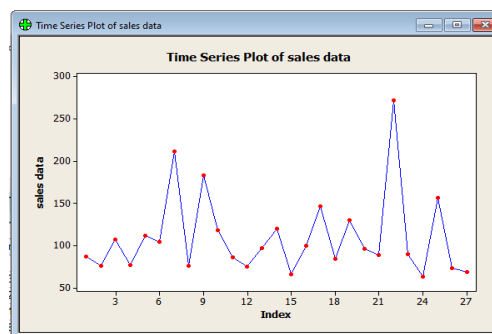


Figure 2. Time Series voucher sales data

3.2 Stationary Testing in Variants

The variables were tested using the Box-Cox transformation method in the Minitab software. Data can be considered stationary in variants if the lambda value (λ) is equal to 1. However, if (λ) is not equal to 1, then it needs to be transformed to the value of 1.

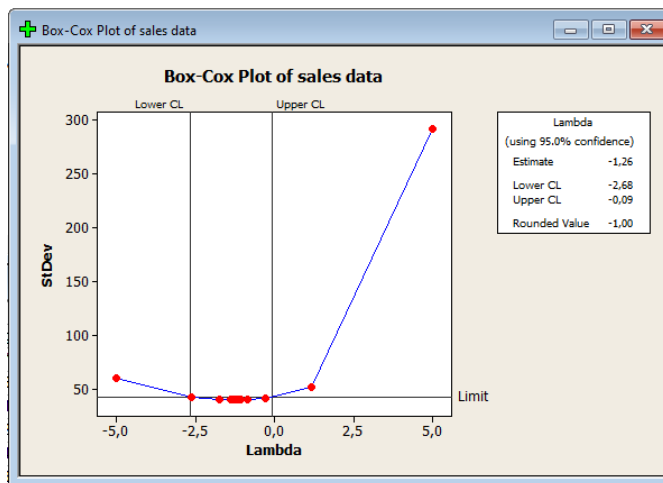
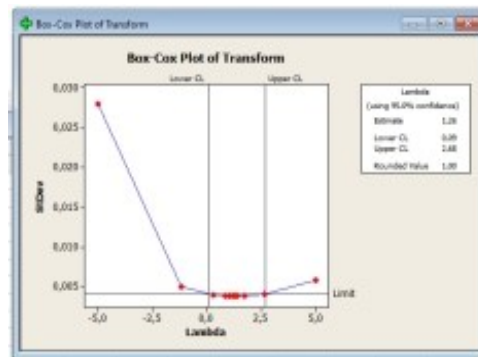


Figure 3. Check λ Values with Box Cox

As shown in Figure 3, after testing the voucher sales data using the Box-Cox transformation, it yielded a rounded value of -1.00, indicating that the data was not stationary and needed to be transformed. The process of calculating the data transformation can also be performed manually using theoretical guidelines such as Table 2.

Table 2. Manual calculation of transformation values

Data Training	Number of Vouchers Sold	Value Transform -1,00
1	87	$\frac{1}{87} = 0,0114943$
2	76	$\frac{1}{76} = 0,0131579$
3	107	$\frac{1}{107} = 0,0093458$
...		...
27	69	$\frac{1}{69} = 0,0144928$



After applying the Box-Cox transformation to the data in the trial using Minitab 16.0 software, the rounded value results showed a value of 1.00, indicating that the data can be considered stationary in terms of variability.

3.3 Stationary In Average (Mean)

A stationary in average can be seen by looking at the graph of the autocorrelation function, if the ACF graph shows a rapidly falling pattern, it means that the data has met the assumption of stationariness in the mean.

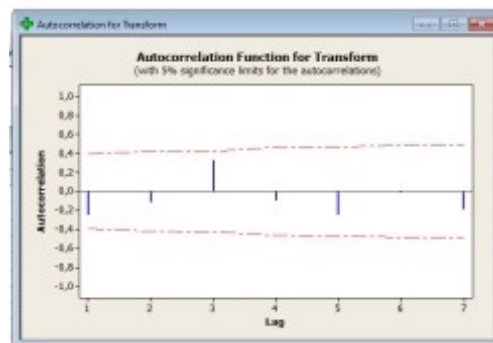


Figure 6 Non-stationary data autocorelation

To determine stationarity in the mean, the graph of the autocorrelation function can be examined. If the ACF graph exhibits a rapid decline, it indicates that the data meets the assumption of stationarity in the mean. Figure 6 depicts the autocorrelation function. Stationarity in the mean will be achieved if the correlogram graph on the autocorrelation function does not exceed the significance limit or tolerance value of α 5%, or 0.05. If the pattern is not stationary, it is necessary to carry out a differencing process to transform and manage non-stationary values, as presented in Table 3.

Table 3 Non-Stationary Values

Lag	Non-Stationary values
1	0,254314
2	-0,118043
3	0,326615
4	-0,096435
5	0,251647
6	-0,014350
7	-0,189784

This can be observed in Table 4, which comprises 7 lags. The significance limit of the autocorrelation value from lag 1 to lag 7 is no more than $\alpha=5\%$, indicating that the voucher sales data in the observation data satisfy the stationary assumption against the average. After all the data is deemed stationary, the subsequent step involves analyzing the model using ACF and PACF.

Table 4. Stationary ACF Values

Lag	Stationary Values
1	-0,568351
2	-0,072646
3	0,288432
4	-0,072872
5	-0,113937
6	0,128526
7	-0,165093

3.4 ARIMA Model Identification

The identification of the ARIMA model is basically done by analyzing the already Stationary ACF and PACF charts.

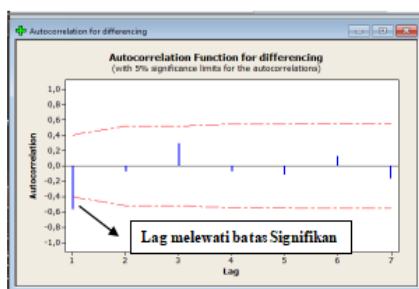


Figure 7 Voucher Sales Data
Autocorrelation

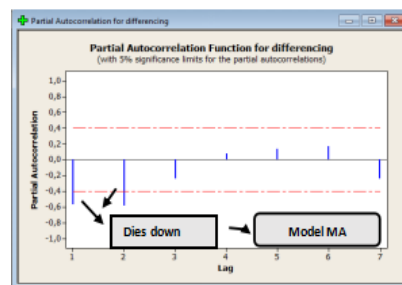


Figure 8 Partial Autocorrelation

As depicted in Figures 7 and 8, the ACF and PACF chorograms are used to analyze the significance of autocorrelation and data stability. The seasonal ARIMA model with order (p, d, q) is determined as $\text{arima}(2,1,1)$ from the partial image, where there are two lags beyond the significant limits and differencing of 1. The overfitting process is then carried out to determine the combination of models to be used, with p being 2, d being a maximum of 1, and q being a maximum of 1.

To verify the accuracy of the $\text{arima}(2,1,1)$ model, minitab 16.0 software can be used with the condition that $p < 2$ on the PACF chart, as shown in figure 8. The obtained model should not have $p < 2$, hence the models obtained are $(2,1,1)$, $(2,1,0)$, and $(1,1,1)$. All three models must be tested for their feasibility.

1) ARIMA model test results $(2,1,1)$

Table 5 Model Testing Results $(2,1,1)$

Type	Coef	SE Coef	T	P
AR 1	-0,671	0,323	-2,08	0,049
AR 2	-0,432	0,267	-1,62	0,119
MA 1	0,344	0,341	1,01	0,324
Constant	0,454	7,124	0,06	0,950

In testing the ARIMA model $(2,1,1)$ on MA1 and AR 2 still has a paramater greater than 0.05 so this model cannot be used.

2) ARIMA model test results $(1,1,1)$

Table 6 Model Test Results $(1,1,1)$

Type	Coef	SE Coef	T	P
AR 1	-0,231	0,221	-1,05	0,306
MA 1	0,939	0,1733	5,42	0,000
Constant	0,228	1,315	0,17	0,864

Testing the $\text{arima}(1,1,1)$ still has parameters greater than 0.05 so it can be used in AR1.

3) ARIMA model test results $(2,1,0)$

Table 7 Model Test Results $(2,1,0)$

Type	Coef	SE Coef	T	P
AR 1	-0,91	0,176	-5,16	0,000
AR 2	-0,577	0,181	-3,19	0,004
Constant	0,33	10,72	0,03	0,975

To test the ARIMA model (2,1,0), parameters were examined and no values exceeded 0.05, indicating that this model is suitable for forecasting. Subsequently, ACF and PACF charts were tested using the model as shown in Figure 10 and 11. It is evident from the charts that none of the lags cross the significance limit, suggesting that the model is appropriate for forecasting purposes.

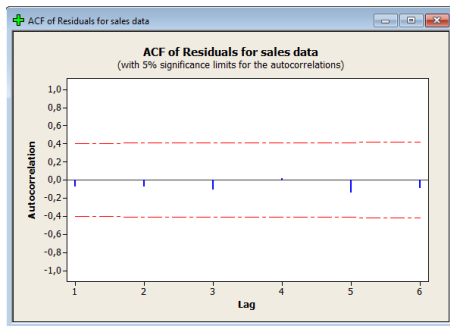


Figure 10 Graph ACF Model
(2,1,0)

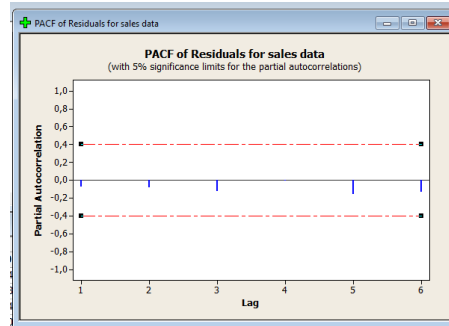


Figure 11 Graph PACF Model
(2,1,0)

3.5 Forecasting

Forecasting is a crucial aspect of any business that helps it plan for the future and make informed decisions. In the case of voucher sales, forecasting is done to predict the sales trend for the upcoming periods, i.e., period 28 to period 40. Accurate forecasting of voucher sales data is essential for diamond cell counters to make informed decisions regarding the amount of voucher stock to be maintained to avoid any potential shortages. By analyzing the sales data, they can project the expected sales volume and plan the stock accordingly to ensure that they have enough vouchers in stock to meet the customer demand. This, in turn, can help diamond cell counters streamline their business operations, optimize their resources, and maximize their profits.

Table 8 Forecasting results

<i>Period</i>	<i>Forecast</i>	<i>Lower</i>	<i>Upper</i>	<i>Actual</i>
20	100,25	-93,08	293,59	96
21	100,08	-97,92	298,09	89
22	100,54	-101,99	303,07	272
23	100,56	-106,63	307,74	90
24	100,61	-110,94	312,16	63
25	100,88	-114,98	316,75	157
26	100,94	-119,23	321,11	73
27	101,06	-123,24	325,37	69

28	101,25	-127,15	329,65
29	101,34	-131,1	333,78
30	101,48	-134,89	337,86
31	101,63	-138,64	341,91
32	101,75	-142,35	345,85
33	101,89	-145,97	349,75
34	102,03	-149,55	353,61
35	102,15	-153,08	357,39
36	102,29	-156,54	361,13
37	102,43	-159,97	364,82
38	102,56	-163,34	368,46
39	102,69	-166,67	372,06
40	102,83	-169,96	375,61

Table 8 presents the results of the forecasting process using the ARIMA model (2,1,0) with the help of minitab 16.0 software. This model was selected as it had parameters that were statistically significant, and further tests confirmed its suitability for forecasting voucher sales. The aim of this forecasting exercise was to predict the sale of vouchers for period 28 to period 40, which would help the diamond cell counters in their stock calculations.

To determine the accuracy of the forecasting results, the Mean Absolute Deviation (MAD), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE) were calculated. The smaller the value of these metrics, the more accurate and reliable the forecasting results would be. By examining the results in Table 8, it can be seen that the ARIMA model produced accurate predictions for voucher sales during the forecast period. These results can be used by diamond cell counters to help them manage their stock levels more effectively and avoid any potential shortages.

Table 9 Error Calculation Formula

No	Actual Data	Forecasting	Error	Error Absolut	Square of Error	Absolute Values of error
1.	87	*	*	*	*	*
2.	76	*	*	*	*	*
3.	104	97,780	6,220	6,220	38,693	5,98
4.	212	90,872	121,128	121,128	14672,000	57,14
...	...	...	...	...	...	...
26	73	100,937	-27,937	27,937	780,470	38,27
27	69	101,065	-32,065	32,065	1028,151	46,47

No	Actual Data	Forecasting	Error	Error Absolut	Square of Error	Absolute Values of error
		TOTAL	317,430	800,590	65068,676	628,77

Error Calculation Formula:

- 1) MAD (Mean Absolute Deviation)

$$MAD = \frac{\sum |Actual - forecast|}{n}$$

$$= \frac{800,590}{27} = 29,65$$

- 2) MSE (Mean Squared Error)

$$MSE = \frac{\sum |Actual - forecast|^2}{n}$$

$$= \frac{65068,676}{27} = 2409,95$$

- 3) MAPE (Mean Absolute Percent Error)

$$MAPE = \frac{\sum (|Actual - forecast| / Actual * 100)}{n}$$

$$= \frac{628,77}{27} = 23,3 \%$$

After conducting the forecasting process using the ARIMA model (2,1,0) in the period 28 to period 40, the accuracy of the results was evaluated by calculating the error rate in terms of Mean Absolute Deviation (MAD), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). The accuracy of the model is determined by how small the values of these error rates are, as the smaller the value, the more accurate and reliable the forecasting results.

As shown in Table 8, the ARIMA model (2,1,0) has a MAD value of 29.65, which indicates the average difference between the actual values and the predicted values. The MSE value obtained is 2409.95, which represents the average squared difference between the actual and predicted values. Lastly, the MAPE value obtained is 23.3%, which represents the percentage of the average absolute difference between the actual and predicted values to the actual values.

Overall, the values of the error rates in the ARIMA model (2,1,0) indicate that the forecasting results are sufficient and feasible for this study. Therefore, the model can be used by diamond cell counters to estimate the amount of voucher stock needed to prevent shortages.

4 CONCLUSION

Based on the results of the forecasting using the Arima method, it can be concluded that the most suitable Arima model for predicting internet voucher sales is the Arima model (2,1,0) with a MAD error value of 29.65, a MSE value of 2409.95, and a MAPE value of 23.3%. This model can be considered sufficient and feasible for the purposes of this study. Furthermore, the stability of internet voucher sales during the period 28 to period 40 indicates that there is no need to excessively increase the stock of internet vouchers. Diamond cell counters can maintain the same amount of stock as in the previous period or order stock vouchers in the same amount.

REFERENCES

- [1] S. Wardah, "Kemasan Bungkus (Studi Kasus : Home Industry Arwana Food Tembilahan)," 2016.
- [2] S. E. Rumagit and A. SN, "Prediksi Pemakaian Listrik Kelompok Tarif Menggunakan Jaringan Syaraf Tiruan dan ARIMA," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 7, no. 2, p. 189, 2013, doi: 10.22146/ijccs.3359.
- [3] M. H. Hamirsa and R. Rumita, "Usulan Perencanaan Peramalan (Forecasting) Dan Safety Stock Persediaan Spare Part Busi Champion Type Ra7yc - 2 (Ev - 01 / Ew - 01 / 2) Menggunakan Metode Time Series Pada Pt Triangle Motorindo Semarang," vol. 2, 1927.
- [4] W. S. Rahayu, P. T. Juwono, and W. Soetopo, "Analisis Prediksi Debit Sungai Amprong Dengan Model Arima (Autoregressive Integrated Moving Average) Sebagai Dasar Penyusunan Pola Tata Tanam," *J. Tek. Pengair.*, vol. 10, no. 2, pp. 110–119, 2019, doi: 10.21776/ub.pengairan.2019.010.02.04.
- [5] S. P. Elvani, A. R. Utary, and R. Yudaruddin, "Peramalan jumlah produksi tanaman kelapa sawit dengan menggunakan metode arima (autoregressive integrated moving average)," *Jurnal Manajemen*, vol. 8, no. 1, pp. 95–112, 2016.
- [6] M. Personal and R. Archive, "Munich Personal RePEc Archive Forecasting irish inflation using ARIMA models," no. 11359, 2008.
- [7] A. Lusiani, "Pemodelan Autoregressive Integrated Moving Average (Arima) Curah Hujan Di Kota Bandung Modelling Of Autoregressive Integrated Moving Average (Arima) Rainfall In Bandung," pp. 9–25, 2010.
- [8] A. F. Kurniawan, S. F. Pane, and R. M. Awangga, "Prediksi Jumlah Penjualan Rumah di Bojongsoang ditengah Pandemi Covid-19 dengan Metode ARIMA," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1479, 2021, doi: 10.30865/mib.v5i4.3121.
- [9] E. Munarsih *et al.*, "Peramalan Jumlah Pengangguran di Provinsi Sumatera Selatan dengan Metode Autoregressive Integrated Moving Average (ARIMA)," *J. Penelit. Sains*, vol. 19, pp. 1–5, 2017.

- [10] M. S. Pradana, D. Rahmalia, and E. D. A. Prahastini, "Peramalan Nilai Tukar Petani Kabupaten Lamongan dengan Arima," *J. Mat.*, vol. 10, no. 2, p. 91, 2020, doi: 10.24843/jmat.2020.v10.i02.p126.
- [11] L. Farosanti, H. Mubarok, and Indrianto, "Analisa Peramalan Penjualan Alat Kesehatan dan Laboratorium di PT. Tristania Global Indonesia Menggunakan Metode ARIMA," *JIMP J. Inform. Merdeka Pasuruan*, vol. 7, no. 2, pp. 14–18, 2022, [Online]. Available: <http://dx.doi.org/10.37438/jimp.v7i1.428>.
- [12] I. B. B. Mahayana, I. Mulyadi, and S. Soraya, "Peramalan Penjualan Helm dengan Metode ARIMA (Studi Kasus Bagus Store)," *Inferensi*, vol. 5, no. 1, p. 45, 2022, doi: 10.12962/j27213862.v5i1.12469.