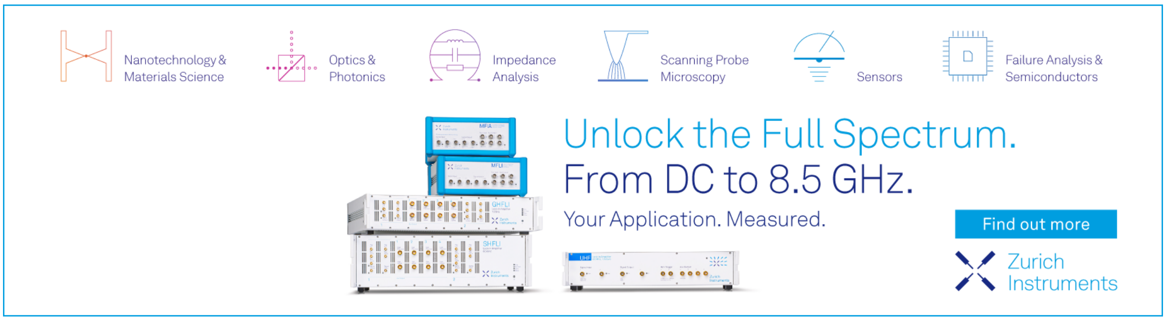
RESEARCH ARTICLE | JANUARY 30 2025

Implementation of forecasting ginger harvest using seasonal autoregressive integrated moving average method

Achmad Jauhari ✉; Devie Rosa Anamisa; Fifin Ayu Mufarroha

AIP Conf. Proc. 3250, 020012 (2025)
https://doi.org/10.1063/5.0241277



Implementation of Forecasting Ginger Harvest using Seasonal Autoregressive Integrated Moving Average Method

Achmad Jauhari[1, a)], Devie Rosa Anamisa[1, b)] and Fifin Ayu Mufarroha[1, c)]

[1]*Informatics Engineering Department, University of Trunojoyo Madura, Bangkalan, Indonesia*

[a)] Corresponding author: jauhari@trunojoyo.ac.id
[b)] devros_gress@trunojoyo.ac.id
[c)] fifin.mufarroha@trunojoyo.ac.id

Abstract. Herbal plant farming is one of the holders of an essential role in the economy of Madura. Therefore, the Madurese government is very concerned about ginger farmers developing ginger production to meet market demand so that the Madurese economy increases. In addition, 2019 data from the Central Statistics Agency show that the herbal farming sector in Madura has been achieved by ginger farming with the most significant number of commodities compared to other herbal plants. However, in recent years, ginger yields have not been able to meet the very high market demand. Therefore, to meet consumer demand for the availability of ginger, this study uses forecasting analysis by developing a system using forecasting methods. The method used in this study is the Seasonal-Autoregressive Integrated Moving Average (SARIMA) hybrid method. This method is a good method for modeling forecasting. The data used in this study are data on ginger production and harvested area from January 2015 to December 2019. And the results of this study are in the form of yield forecasting data for the following year. The test results with a data range of five produce small MAPE and RMSE, namely 43.94% and 14579.338. This shows that the SARIMA method has been able to predict future crop yields and can be used as a reference for the government in determining crop yields according to market demand.

INTRODUCTION

Agriculture, both food crops and herbal plants spread across Madura, plays a vital role in the economic and social fields because these agricultural products can generate substantial income to build the economy of the Madurese population [1]. So that this can have a positive impact on Madura farmers in cultivating plants to increase agricultural yields and be able to meet market demand. However, in recent years, agricultural production in Madura has decreased. One of the causes is the lack of accuracy of farmers in predicting agricultural yields for the future, so preparations must be prepared to increase crop yields so that market demand is fulfilled. Many commodities produced by farmers from the agricultural sub-sector include ginger. With the high demand for ginger in the world market for ginger commodities, which continues to increase from year to year, farmers are required to increase their production further to meet the demand for ginger domestically and abroad. Therefore, this research develops a ginger yield forecasting system. Forecasting is the art and science of predicting future events [2]. Forecasting for crop yields is the basis for the most important planning decisions to meet market demand. The goal of a forecasting system is to predict by producing accurate results, but this is never possible, forecasting can reduce uncertainty about the future [3].

In several previous studies, many researchers have applied forecasting systems to forecasting methods, but the search for forecasting models that produce a small error rate is still being carried out. In 2017 of research by Kafara

Proceedings of the International Conference on Informatics, Mechanical, Industrial, and Chemical Engineering (ICIMICE2023)
AIP Conf. Proc. 3250, 020012-1–020012-7; https://doi.org/10.1063/5.0241277
Published under an exclusive license by AIP Publishing. 978-0-7354-5079-0/$30.00

et al. [4] regarding forecasting rainfall with the seasonal ARIMA approach called SARIMA, which uses time series analysis, namely the Box-Jenkins method for SARIMA modelling, which has produced the smallest MAPE and RMSE with static forecasting in the forecasting model for one period (months) ahead. Whereas in 2018, it was carried out by Durrah et al. [5] concerning forecasting the number of airplane passengers using the SARIMA method, where in the process of predicting the number of passengers it has shown a seasonal pattern with the best pattern ARIMA (0,1,1)(0,0,1)12 and the MAPE result is 9.78%. Then developed in 2021 by Rizki and Taqiyyuddin [6] regarding the application of the SARIMA model to predict the inflation rate in Indonesia and produce a MAPE value of 5.19%. This shows that the SARIMA model is a method that can be used as a reference in determining steps to anticipate problems in the economic sector. However, in previous years the application of the SARIMA model in predicting ginger crop yields had not been carried out. This method is usually used for single time series analysis forecasting data.

The important thing to consider in analyzing time series data is the accuracy of the model. And a good model is a model that has the smallest MAPE and RMSE values. Therefore, this study tries to apply the SARIMA method for forecasting future ginger yields to be able to produce increased production to meet market demand. The SARIMA method is the development of the Autoregressive Integrated Moving Average (ARIMA) model on time series data that has a seasonal pattern6. SARIMA is also a time series analysis method, just like trend, moving average, or naive analysis [7]. The purpose of this study was to forecast ginger crop yields for the next 12 months. The benefits of forecasting the yield of this ginger plant are expected to be able to help ginger farmers in taking certain policies to overcome increased yields to meet market demand.

METHODS

Time series data is historical data collected in a certain period of time sequentially [8]. The data used in this forecasting are ginger yields from 2015-2019 obtained from the Pamekasan Madura Agriculture Service, with a population of 250 farmers cultivating ginger in Madura. Ginger plant is an export commodity with good quality, quality, and taste, so it has good market opportunities, domestically and abroad [9]. And the stages carried out in this study have been described in Figure 1. This stage includes literature study, data collection and pre-processing, forecasting modelling with the SARIMA method, system requirements analysis, and design and implementation of a web-based forecasting system. The literature study is the initial stage carried out in this research which aims to determine the topic and analyze the background of the problems related to fluctuating ginger commodity yields [10]. Then the data is pre-processed to prepare the data so that it can be used for forecasting models, where the form of data pre-processing is done by filling in the missing values and overcoming the outlier values [11,12]. The pre-processing results have been simplified into numbers 0 and 1, as shown in Table 1. As for the SARIMA stages, it can be seen in Figure 2, where in the early stages an initial analysis of data patterns was carried out, after which the normality and stationarity tests of historical data were carried out. After that, statistical tests were carried out with parameter estimates. The statistical test produces residual independence. From these residuals, independence and normality will be checked, so that the SARIMA model used is good or not. The best SARIMA model is the model with the analysis results showing the smallest error value. The error value in this study was obtained by calculating the MAPE AND RMSE values.

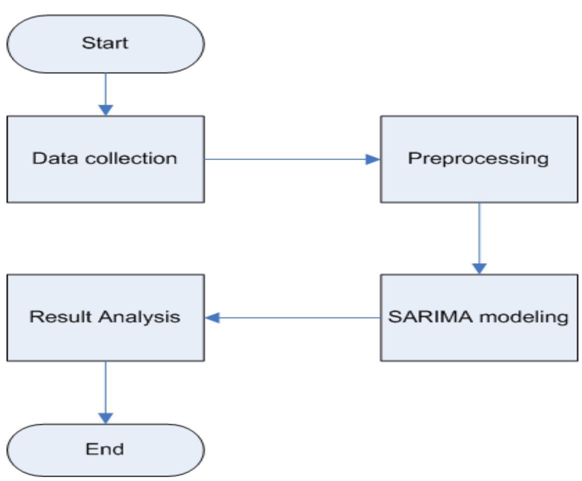


FIGURE 1. Flowchart of stages of research

For forecasting using the SARIMA method, use the auto-ARIMA source code [13]. In the source code, several parameters are set, starting from the minimum arima parameter, the maximum arima parameter, the differencing parameter¸ to the seasonal parameter. Using auto-ARIMA can automatically detect the best model parameters from SARIMA without having to manually define them and do trial and error for the best model. The SARIMA method is one of the developments of the ARIMA method, so the basic knowledge of the SARIMA method is the same as the ARIMA method [14,15]. There are Autoregressive (AR) methods, Moving Average (MA), and also Autoregressive merging with Moving Average, namely Autoregressive Moving Average (ARMA) [16]. For the equation of the autoregressive method using Eq. (1). Apart from the autoregressive equation, there are also similarities from the moving average method, as in Eq. (2). After there are autoregressive and moving average equations, then there is an equation from the combination of the autoregressive method and also the moving average, namely the autoregressive moving average (ARMA) with Eq. (3) [17].

TABLE 1. Result of preprocessing data

No	Production Amount (kg)	No	Production Amount (kg)	No	Production Amount (kg)	No	Production Amount (kg)
Min	**10500**						
Min	**15400**						
1	0.3271	81	0.9225	151	0.5994	201	0.2723
2	0.3980	82	0.7136	152	0.1103	202	0.0090
3	0.5661	83	0.5648	153	0.6022	203	0.5540
4	0.9656	84	0.0340	154	0.0513	204	0.7744
5	0.0944	85	0.8417	155	0.7686	205	0.5321
6	0.6425	86	0.9927	156	0.1290	206	0.9719
7	0.4452	87	0.4989	157	0.2324	207	0.9614
8	0.2650	88	0.9147	158	0.7615	208	0.2293
9	0.7773	89	0.1937	159	0.1739	209	0.0389
10	0.0000	90	0.1978	160	0.7496	210	0.4998
11	0.7859	91	0.8731	161	0.7312	211	0.4238
12	0.9514	92	0.5018	162	0.5679	212	0.0348
13	0.1337	93	0.7347	163	0.7255	213	0.0128
14	0.5951	94	0.8253	164	0.9700	214	0.1414
15	0.2236	95	0.8386	165	0.4957	215	0.7677
16	0.4160	96	0.5907	166	0.5521	216	0.3376
17	0.8617	97	0.6161	167	0.6994	217	0.2006
18	0.5780	98	0.6365	168	0.6541	218	0.3386
19	0.7049	99	0.3273	169	0.8888	219	0.9919
20	0.5917	100	0.1642	170	0.7617	220	0.9563
…	….	…	…	…	…	…	…
51	0.9124	130	0.6269	180	0.2359	230	0.9267
52	0.9240	131	0.7042	181	0.0427	231	0.5396
53	0.7562	132	0.5903	182	0.3422	232	0.7370
54	0.5466	133	0.9340	183	0.6394	233	0.7296
55	0.8599	134	0.7491	184	0.0698	234	0.4405
56	0.9780	135	0.8947	185	0.4865	235	0.1565
57	0.1122	136	0.5441	186	0.8137	236	0.4264
58	0.7921	137	0.0368	187	0.4802	237	0.5265
59	0.9163	138	0.1639	188	0.7167	238	0.5907
60	0.4030	139	0.9171	189	0.2363	239	0.1143
61	0.9178	140	0.3723	190	0.6690	240	0.5622
62	0.0667	141	0.9009	191	0.4320	241	0.1511
63	0.8547	142	0.2234	192	0.3216	242	0.8176
64	0.3336	143	0.2643	193	0.2149	243	0.5206
65	0.6623	144	0.3478	194	0.3294	244	0.4357
66	0.6161	145	0.3889	195	0.6661	245	0.9221
67	0.4294	146	0.0621	196	0.1951	246	0.1313
68	0.9975	147	0.8884	197	0.5761	247	0.6004
69	0.5300	148	0.8223	198	0.6708	248	0.2817
70	0.0428	149	0.1515	199	0.0108	249	0.6220
51	0.9124	150	0.6692	200	0.6473	250	0.7577
…	….	…	…	…	…	…	…

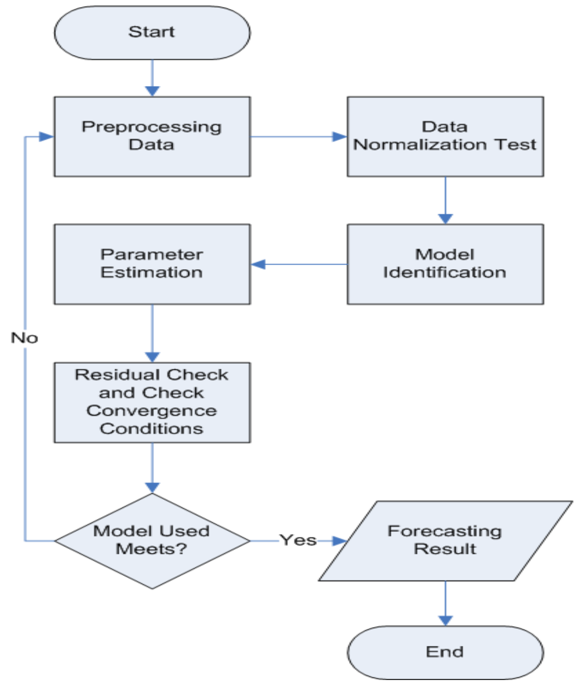


FIGURE 2. Stages of the SARIMA method

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_p Y_{t-p} - e_t \quad (1)$$

$$Y_t = \emptyset_0 + \emptyset_1 e_{t-1} + \emptyset_2 Y e_{t-2} + \cdots + \emptyset_p Y e_{t-q} \quad (2)$$

$$Y_t = \gamma_0 + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \cdots + \lambda_p e_{t-q} \quad (3)$$

After modelling using the SARIMA method, it is necessary to measure the performance of the model. Performance measurements are carried out to evaluate and determine the level of accuracy of the model made. Forecasting results will be compared with actual data using MAPE and RMSE calculations. MAPE is the percentage error resulting from the average value of the difference between the actual data and the forecasted data [18]. The smaller the MAPE and RMSE error values generated in the implemented model, the closer the predicted values are to real data, can be seen in Table 2 [19]. While the RMSE is a measure that has theoretical relevance in statistical modelling and is more sensitive to outlier data, so it is more recommended to use it in selecting the accuracy of modelling results.

TABLE 2. Result of preprocessing data

Ranges of MAPE	Forecasting Models
<10%	Very Good
10%-20%	Good
20%-50%	Quite Good
>50%	Bad

ANALYSIS AND DISCUSSION

The data used in this research is ginger yield data from January 2017 to December 2021 obtained from the Pamekasan Agriculture Service, Madura. The graph of the dataset can be seen in Figure 3. From Figure 3, it can be seen that ginger yields are seasonal data, which is characterized by a pattern of repeated high increases in specific periods and repeated every year. This is proven by several points in certain months and years where there is an erratic increase and decrease in harvest yields. In 2016, ginger yields were seen to decrease, as can be seen in Figure 4. Therefore, this study uses SARIMA modelling to forecast ginger yields next year. A good SARIMA model is a model that has the smallest MAPE and RMSE values. The results for the diagnosis of SARIMA can be seen in Table 3. Table 3 summarizes the parameter estimation results of the twelve existing models. After going through the stages of model identification, parameter estimation, and residual test, based on table 3, the best model for ginger yields is SARIMA

(0,0,1) (1,0,1)12. Before forecasting, it will first be seen how accurate the model is in predicting ginger yield data. The model's accuracy will be seen by comparing the actual value of the monthly ginger yield in the sample period used in forming the model with the forecasting value. The goodness of the model from the forecasting results to the actual value statistically can also be seen from the smallest MAPE and RMSE values, so it was selected as the best model suitable for forecasting ginger yields. The last stage carried out in this study is to forecast ginger yields for the period January to December 2019 (data testing) using the SARIMA (0,0,1) (1,0,1)12 model. The SARIMA (0,0,1) (1,0,1)12 model is the only model that satisfies the significant parameter test and white noise test where the model residuals are random and normally distributed, making it feasible to predict Madura ginger yields. The calibration results in this study use comparative forecasting data with actual data for the following year, namely 2020, because the forecast is too far from the sample period, as seen in Figure 3.

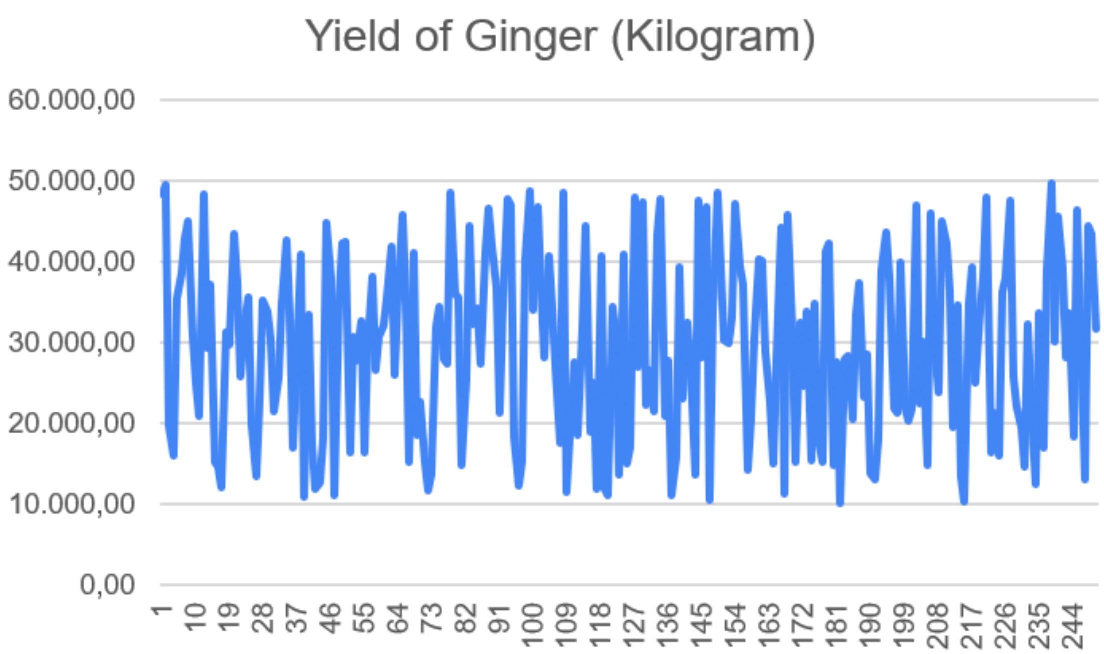


Figure 3. Ginger dataset from January 2015-December 2019

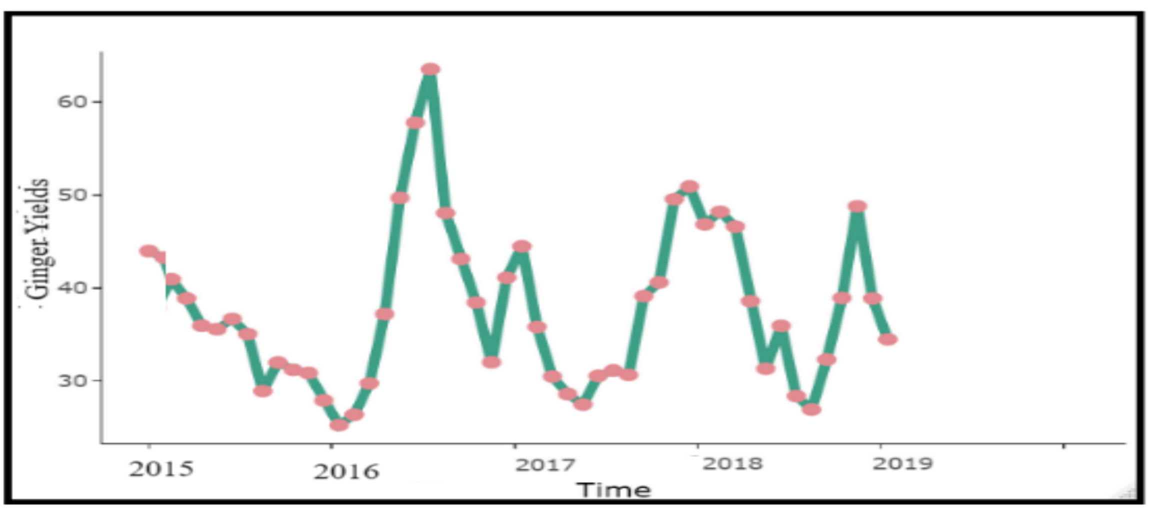


Figure 4. Ginger harvest data exploration plot

TABLE 3. Diagnosis of the SARIMA model

Model	Alpha	Time-stamp	MAPE	RMSE
SARIMA $(0,0,1)(1,0,1)_{12}$	0.001	5	35,9387	11599,219
SARIMA $(0,0,1)(1,0,1)_{12}$	0.001	10	52,6581	11276,190
SARIMA $(0,0,1)(1,0,1)_{12}$	0.001	15	43,9435	14579,338
SARIMA $(1,1,1)(1,1,0)_{12}$	0.001	5	53,4932	11676,122
SARIMA $(1,1,1)(1,1,0)_{12}$	0.001	10	65,4521	11716,113
SARIMA $(1,1,1)(1,1,0)_{12}$	0.001	15	57,4742	15586,345
SARIMA $(0,1,1)(2,1,0)_{12}$	0.001	5	56,1232	12643,112
SARIMA $(0,1,1)(2,1,0)_{12}$	0.001	10	67,3563	12713,123
SARIMA $(0,1,1)(2,1,0)_{12}$	0.001	15	58,0812	16341,371
SARIMA $(0,1,1)(2,1,1)_{12}$	0.001	5	57,1112	13643,502
SARIMA $(0,1,1)(2,1,1)_{12}$	0.001	10	63,1561	12813,148
SARIMA $(0,1,1)(2,1,1)_{12}$	0.001	15	59,1812	17341,161

CONCLUSION

Using the SARIMA method with the SARIMA (0,0,1) (1,0,1)12 model for forecasting the yield of Madura ginger produces a more accurate forecasting value with an error value of 43.94% and 14579.338. So the results of predicting the harvest of Madura ginger for the next period can be used as a reference for decision-making in determining ginger farmer policies in increasing their production. The limitations of this study include forecasting that can only be presented in the short term, forecast numbers that are too far from the sample data set will have large residuals due to changing conditions and crop yield situations. This forecasting is also purely built from the movement of the value of the harvest data itself in the previous period. Subsequent studies can create models of relevant variables that affect the needs that affect the harvest of Madura ginger.

ACKNOWLEDGMENTS

Thank you to the University of Trunojoyo Madura, for the opportunity to do research And the author also thanks the Department of Agriculture and Horticulture in the Sumenep office for their assistance in providing data in this research.

REFERENCES

1. E. Widiaswanti, R. Yunitarini, and R. Indriartiningtias, "Determination of Ginger Simplicial Marketing Strategy in Madura with EFE and IFE Matrix Approach," in MATEC Web of Conferences, (EDP Sciences, France, 2022), p. 8007.
2. R. Erlina and A. Rialdi, Journal Teknik Pertanian Lampung **9**(3), 257–263 (2020).
3. A. Muqtadir, S. Suryono, and V. Gunawan, Scientific Journal of Informatics **3**(2), 159–166 (2016).
4. Z. Kafara, F. Y. Rumlawang, and L. J. Sinay, Barekeng: Jurnal Ilmu Matematika dan Terapan **11**(1), 63–74 (2017).
5. F. I. Durrah, Y. Yulia, T. P. Parhusip, and A. Rusyana, Journal of Data Analysis **1**(1), 1–11 (2018).
6. M. I. Rizki and T. A. Taqiyyuddin, Jurnal Sains Matematika dan Statistika **7**(2), 62–72 (2021).

7. M. Giovani, I. Anggriani, and S. A. W. D. Simatupang, BAREKENG: Jurnal Ilmu Matematika dan Terapan **16**(4), 1487–1496 (2022).
8. A. M. Al'afi, W. Widiarti, D. Kurniasari, and M. Usman, Jurnal Siger Matematika **1**(1), 10–15 (2020).
9. B. R. Murthy, S. G. Rao, and S. K. N. Umar, J Pharmacogn Phytochem **9**(2S), 317–320 (2020).
10. S. Maddodi, International Journal Of Information System And Computer Science **5**(3), (2021).
11. L. Alexander and S. Sergey, "Predictive models for metrological data of engineering systems," in Journal of Physics: Conference Series, (IOP Publishing, Bristol, 2021), p. 12046.
12. S. Sridevi, S. Rajaram, C. Parthiban, S. SibiArasan, and C. Swadhikar, "Imputation for the analysis of missing values and prediction of time series data," in 2011 international conference on recent trends in information Technology (ICRTIT), (IEEE, NY, 2011), pp. 1158–1163.
13. A. Iswari, Y. Angraini, and M. Masjkur, Indonesian Journal of Statistics and Its Applications **6**(1), 132–146 (2022).
14. P. Chen, A. Niu, D. Liu, W. Jiang, and B. Ma, "Time series forecasting of temperatures using SARIMA: An example from Nanjing," in IOP conference series: materials science and engineering, (IOP Publishing, Bristol, 2018), p. 52024.
15. M. Soto-Ferrari, O. Chams-Anturi, and J. P. Escorcia-Caballero, "A time-series forecasting performance comparison for neural networks with state space and ARIMA models," in Conference: International Conference on Industrial Engineering and Operations Management, (IEOM Society International, US, 2020), pp. 155–164.
16. L. Ruohong, "Research on Applications of ARMA in Forecasting of Time Series," in 2015 International Conference on Social Science and Technology Education, (Atlantis Press, Paris, 2015), pp. 74–77.
17. A. Isiaka, A. Isiaka, and A. Isiaka, International Journal of Research in Business and Social Science (2147-4478) **10**(1), 205–234 (2021).
18. N. Abu, W. N. Syahidah, M. M. Afif, and S. Z. Nordin, "SARIMA and Exponential Smoothing model for forecasting ecotourism demand: A case study in National Park Kuala Tahan, Pahang," in Journal of Physics: Conference Series, (IOP Publishing, Bristol, 2021), p. 12118.
19. A. Assidiq, P. Hendikawati, and N. Dwidayati, Unnes Journal of Mathematics **6**(2) 129–142 (2017).