JSDS: JOURNAL OF STATISTICS AND DATA SCIENCE
VOLUME 3, No 1, March 2024
e-ISSN: 2828-9986
https://ejournal.unib.ac.id/index.php/jsds/index

Forecasting Export Value of Bengkulu Province Through Pulau Baai Harbour with ARIMA, ANN, and Hybrid ARIMA-ANN Approach

Wina Ayu Lestari[1*], Sigit Nugroho[1], Fanani Haryo Widodo[1]

[1]Department of Mathematics, Bengkulu University, Indonesia
* Corresponding Author: wyna.austry@gmail.com

Article Info

Article History:
Received: mm dd yyyy
Revised: mm dd yyyy
Accepted: mm dd yyyy
Available Online: mm dd yyyy

Key Words:
Export Value,
Forecasting,
ARIMA,
ANN,
Hybrid ARIMA-ANN.

Abstract

Forecasting is a process of predicting future events based on past event data. One of the time series models that can be used for forecasting is the Autoregressive Integrated Moving Average (ARIMA). The advantages of ARIMA are in the accuracy and flexibility of its forecasting in representing several different types of time series, but the main limitation is the linear form of the model which causes ARIMA to be unable to capture non-linear patterns in the data. An alternative model for time series modeling is Artificial Neuron Network (ANN). ANN can overcome the weaknesses of ARIMA, but cannot handle linear and nonlinear patterns of the data simultaneously. As an effort to improve forecasting accuracy, Hybrid ARIMA-ANN is carried out by taking advantage of the supremacy of ARIMA and ANN. This study aims to obtain the best model for forecasting the export value of Bengkulu Province, a model generated by the time series data of export values issued by Pulau Baai Harbour from January 2014 to June 2022. The result shows that the best model for predicting the export value of Bengkulu Province is the ARIMA-ANN hybrid model with MAAPE of 0.5289 and MASE of 0.7664.

1. INTRODUCTION

Bengkulu is a province located on the west littoral of the island of Sumatra. Bengkulu Province is rich in natural resources, so it has considerable potential for foreign trade. Bengkulu's export commodities include coal, rubber, palm shells, crude palm oil (CPO), coffee, wood, and others. Until now, the largest proportion of the total export value of Bengkulu Province is channeled through the Pulau Baai Harbour, and the main export commodity of Bengkulu is coal [1].

Based on Figure 1, it could see that the value of exports in the period from January 2014 to June 2022 fluctuated. The lowest fluctuations occurred in January and the highest in April 2022. The export value in April 2022 reached 36.13 million US$, but the export value decreased again in May and June 2022. The value of export in May and June 2022 reached 26.33 million US$ and 28.96 million US$. That value tends to be higher than the export value from 2014 to 2021.

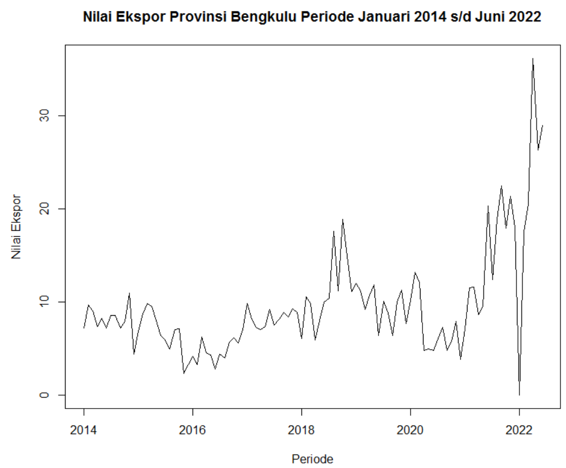

Figure 1. Bengkulu Province export value data plot for the period January 2014-June 2022

The fluctuating plot of export values indicates that the export value data is not stationary. One model that could use to make predictions with non-stationary data is the Autoregressive Integrated Moving Average (ARIMA), where the data would convert to be stationary first through differencing before being analyzed [2]. An alternative model for time series modeling is the Artificial Neuron Network (ANN) [3]. ANN is a time series forecasting model that imitates how biological neural networks work. ANN excels in nonlinear modeling, and ARIMA excels in liner modeling, but ARIMA nor ANN is not a universal model suitable for all circumstances [4]. Zhang [4] proposes the Hybrid ARIMA-ANN approach used for time series forecasting by leveraging the advantages of ARIMA and ANN.

Some previous studies, namely Guha and Bandyopadhyay [5], forecast gold prices in India using the ARIMA model and found that the ARIMA (1,1,1) model is the best model. Katoch and Sidhu [6] applied the ARIMA model to predict COVID-19 cases in India. The results obtained show that the best model, namely ARIMA (4,2,7), can predict COVID-19 cases in India. Jiang, Sharafisafa, and Shen [7] used ANN to predict the effect of heterogeneity on rock strength at different strain levels. The results showed that the ANN (7-5-1) model obtained an accuracy of 99.51%. Toga, Atalay, and Toksari [8] applied ARIMA and ANN in predicting the prevalence of COVID-19 in Turkey and compared the performance of the two models based on mean squared error (MSE) and sum square error (SSE). The results showed that ARIMA and ANN have almost the same forecasting performance. Hybrid ARIMA-ANN was conducted by Wang and Meng [9] for forecasting energy consumption in Hebei Province, China, and found that hybrid models can be an effective way to improve forecasting accuracy.

Research by applying ARIMA, ANN, and hybrid ARIMA-ANN models has been carried out by several researchers in Indonesia, including Lailiyah and Manuharawati [10] forecasting the value of exports in Indonesia by applying the ARIMA model and obtained the result that the ARIMA (1,1,0) model is the best ARIMA model. Nugraha, Suparman, and Pravitasari [11] applied the ANN model in forecasting export values in Indonesia and the results showed that the network model (12-6-1) was the optimum network. Research that applies the combination of ARIMA and ANN models has been conducted by Kamadewi and Achmad [12] with the research title "Hybrid ARIMA-ANN Modeling on Indonesian Inflation data for 2009-2020. The results obtained show that Hybrid ARIMA-ANN produces a fairly small forecasting error and has accurate forecasting accuracy.

This study compares the performance of the ARIMA, ANN, and hybrid ARIMA-ANN models in forecasting the export value of Bengkulu Province. The best model is the one that has the smallest forecasting error rate measured by Mean Arctangent Absolute Percentage Error (MAAPE) and Mean Absolute Scaled Error (MASE).

2. THEORETICAL BASIC

2.1 ARIMA Models

The ARIMA model is a forecasting model that belongs to the linear group. The ARIMA model is obtained by combining the AR(p) and MA(q) models and adding a difference component. The general form of the ARIMA model is as follows:

$$\phi(B)\nabla^d Z_t = \theta(B)e_t \tag{1}$$

where Z_t is the actual value of the observation at the t-th time, B is a backward shift operator which is defined as $B^m Z_t = Z_{t-m}$, $\phi(B)$ is the *autoregressive operator* (AR) which is defined as $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$, $\theta(B)$ is the *moving average* operator (MA) which is defined as $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)$, ∇^d is the d-th difference dan e_t is a random error at the t-th time, which is assumed to be identically and independently distributed with a mean of zero and a constant variance [2].

2.2 ANN Models

ANN is an information processing system had performance characteristics on general biological neural networks. The ANN had developed as the mathematical generalization model from human neural networks. Three characteristics of ANN are network architecture, algorithms, and activation functions [13]. The general form of the ANN model is as follows

$$Z_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \, g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} \; Z_{t-i}\right) + \varepsilon_t \tag{2}$$

where α_j $(j = 0,1,2, \dots, q)$ and β_{ij} $(i = 0,1,2, \dots, p\,;\; j = 1,2, \dots, q)$ are the connection weights, g is the activation function used, p is the number of input nodes dan q is the number of hidden nodes [4].

Backpropagation is one of the algorithms used in multilayer networks with supervised training methods. The stages of network training with the backpropagation algorithm are as follows [13]:

1) Feedforward
2) Backpropagation of Error
3) Update weights and bias

2.3 ARIMA-ANN Hybrid Models

Hybrid ARIMA-ANN is a combination model between ARIMA and ANN used for linear and nonlinear modeling, so that complex autocorrelation structures in data can be modeled more accurately [4]. The general form of the ARIMA-ANN hybrid model is as follows:

$$Z_t = L_t + N_t \tag{3}$$

where L_t denotes the linear component dan N_t denotes the nonlinear component. The Method of hybrid ARIMA-ANN consists of two steps as follows [4]:

Step 1. ARIMA is used to analyze the linear part of the problem so that a linear model has obtained. The ARIMA wasn't capable of capturing nonlinear patterns from data, then the residual contains nonlinear patterns.

$$e_t = Z_t - \hat{L}_t \tag{4}$$

Step 2. Analyze residues with ANN to obtain nonlinear patterns in the data. ANN model for residual with p input as follows:

$$e_t = f\left(e_{t-1}, e_{t-2}, \dots, e_{t-p}\right) + \varepsilon_t \tag{5}$$

The combination of forecasting becomes as follows:

$$\hat{Z}_t = \hat{L}_t \quad + \hat{N}_t \tag{6}$$

where $\hat{L}_t$ denotes the t-th predicted value of the ARIMA model dan $\hat{N}_t$ denotes the t-th predicted value of the ANN model.

3. METHOD

3.1 Data

The data used in this study is data on the export value of Bengkulu Province through the Pulau Baai Harbour from January 2014 to June 2022 obtained from the official website of BPS Bengkulu Province (https://bengkulu.bps.go.id/). Data were analyzed using ARIMA, ANN, and hybrid ARIMA-ANN models with the help of the R-studio program. Before conducting the analysis, the data will be divided into training data and testing data on a percentage of 94% and 6%. Especially for ANN data, time series data must be formed into preprocessed data first with the following formula:

$$Z_t = \left(Z_{t-1}, Z_{t-2}, \dots, Z_{t-12}\right) \tag{7}$$

where $t = 1$ (January 2015) s/d 90 (June 2022).

3.2 Stages of Data Analysis

The stages of data analysis in this study are:

(1) ARIMA analysis with the following stages:

a) Check if the data is stationary in variance and mean. If the data is not stationary in variance then perform a Box-Cox transformation, if the data is not stationary in mean then do differencing.

b) Identify the model by looking at significant lags on the ACF plot and the PACF plot to determine the q and p order, while the d order is the number of differencing.

c) Parameter estimation

Maximum Likelihood Estimation is one of the methods that can be applied to obtain parameter estimates from the ARIMA model.

d) Model diagnostic checking

A model is said to be feasible if the resulting residual is a process of white noise.

e) Calculate MAAPE and MASE of each model

f) Selection of the best ARIMA model based on the smallest MAAPE and MASE values

g) Performance testing of the best ARIMA model using testing data.

(2) ANN analysis with the following stages:

a) Normalization of preprocessed data using the min-max scaling method.

b) Designing the network architecture to be used in network training

Table 1. Network architecture design

Component	Description
Input layer	12 *neurons*
Hidden layer	1 layer : trial and error (2 s/d 11) neurons 2 layers : trial and error (2,2) s/d (11,11) neurons
Output layer	1 neuron
Activation function	Sigmoid biner
Algoritma	Backpropagation
Learning rate	0.01
Threshold	0.05

c) Network training with Backpropagation algorithm

d) Data denormalization

e) Calculate MAAPE and MASE of each model

f) Selection of the best ANN model based on the smallest MAAPE and MASE values

g) Performance testing of the best ANN model using testing data.

(3) ARIMA-ANN Hybrid Analysis

a) The data are analyzed using ARIMA, so the best and residual ARIMA models are available

b) The residuals of the ARIMA model are analyzed with ANN so that the best ANN model is available.

c) Combine ARIMA and ANN model, so the ARIMA-ANN hybrid model is available.

d) Calculate MAAPE and MASE from the ARIMA-ANN hybrid model.

e) Performance testing of the best ARIMA-ANN hybrid model using testing data.

(4) Selection of the best model for forecasting export value

(5) Forecasting export value for the three periods next.

4. RESULTS AND DISCUSSION

4.1 ARIMA Analysis Results

The results showed that the data were not stationary in variance and mean. The data was stationer after transformation and after one-time differencing. Subsequently, the ARIMA model was identified by looking at significant lags in the ACF plot and the PACF plot. The conjecture models obtained are the ARIMA model (0,1,1) and the ARIMA model (0,1,0). Estimation of ARIMA model parameters (0,1,1) using Maximum Likelihood Estimation (MLE) and obtained the following results:

Table 2. Estimated ARIMA parameters (0,1,1)

Type	coefficient	Koefiisien error standard
MA(1)	-0.5002	0.0792

The results of model diagnostic checking show that only the ARIMA model (0,1,1) meets the assumption residual of white noise, so the ARIMA model (0,1,1) is the best model for forecasting export values.

4.2 ANN Analysis Results

Network training has conducting using the backpropagation algorithm and network architecture design in Table 1. The results showed that the best ANN model is ANN (12-8-1) because it produces minimum MAAPE and MASE values. The network architecture of ANN (12-8-1) has shown in Figure 2.

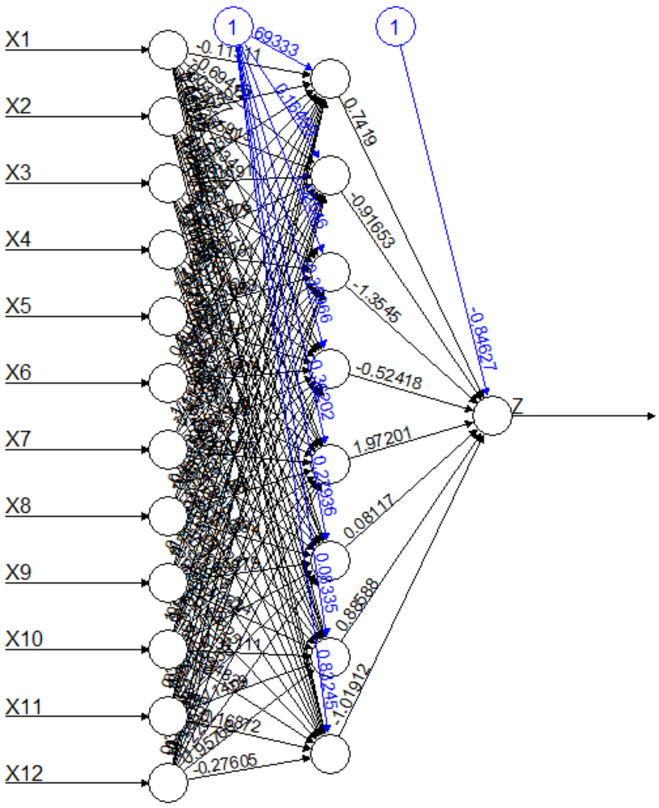

Figure 2. The Architecture of ANN (12-8-1)

4.3 ARIMA-ANN Hybrid Analysis Results

The Hybrid ARIMA-ANN model is a combined model of ARIMA and ANN, where the initial stage of analysis is to perform ARIMA analysis of export value data. ARIMA analysis has been discussed earlier, where ARIMA (0,1,1) is the best model for forecasting export value. The ARIMA model residue (0,1,1) was tested for linearity by looking at its cumulative periodogram. Figure 3 shows that the residue forms a line that does not follow the diagonal line, meaning that the ARIMA model residue (0,1,1) is not linear, which means ARIMA is not capable of capturing nonlinear patterns in the data, so implementing a hybrid ARIMA-ANN is the right solution to overcome the ARIMA deficiency.

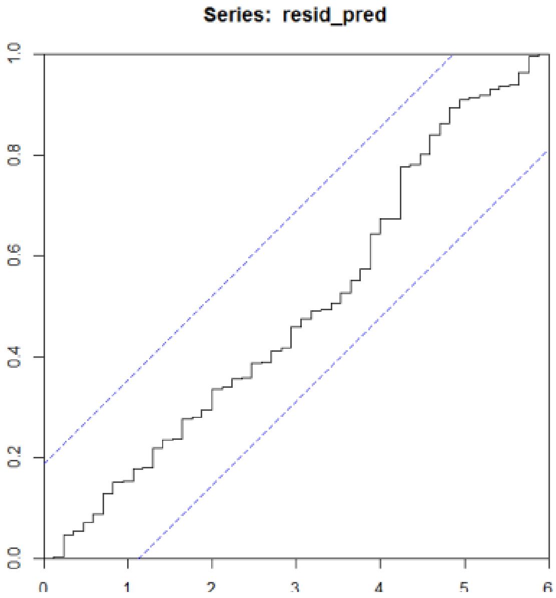

Figure 3. Cumulative periodogram of ARIMA residue (0,1,1)

Residuals of the ARIMA model (0,1,1) were then analyzed with ANN using the backpropagation algorithm and network architecture design prepared in Table 1. The results showed that the ANN model (12-8-8-1) produced minimum MAAPE and MASE values. Here's the architecture of the ANN model (12-8-8-1):

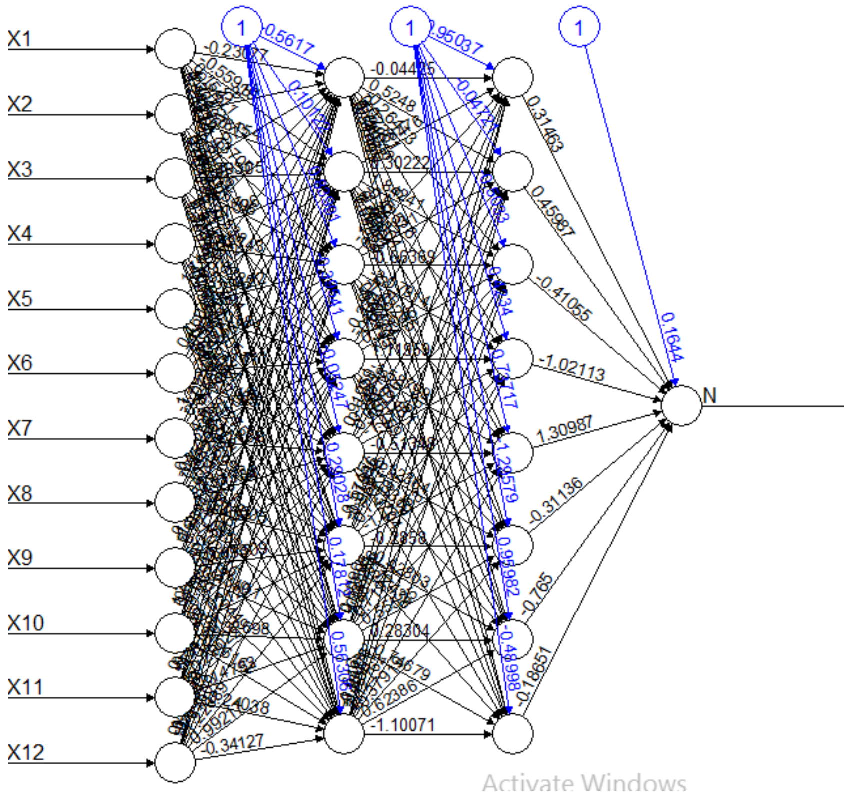

Figure 4. The Architecture of ANN (12-8-8-1)

The Hybrid ARIMA-ANN model is a combination of the ARIMA (0,1,1) and the model of ANN (12-8-8-1). The model will be used to forecast the value of exports. The mathematical model of hybrid ARIMA-ANN is as follows:

$$Z_t = L_t + N_t$$

$$= (L_{t-1}^{0.061} + 0.061e_t + 0.0305e_{t-1})^{\frac{1}{0.061}} + f\Big((0.1644) + \Big((YY_1 \times 0.3146) + (YY_2 \times 0.4599) + \big(YY_3 \times (-0.4106)\big) + \big(YY_4 \times (-1.02116)\big) + (YY_5 \times 1.3099) + \big(YY_6 \times (-0.3114)\big) + \big(YY_7 \times (-0.7650)\big) + \big(YY_8 \times (-0.1865)\big)\Big)\Big) + \varepsilon_t$$

4.4 Selection of the Best Model

After going through several stages of analysis, the best models of each approach are ARIMA (0,1,1), ANN (12-8-1), and hybrid ARIMA (0,1,1) – ANN (12-8-8-1). The performance of models had tested using testing data. Figure 5 shows that the predictions generated by the three models are not good. The plot of prediction data with the ANN models was unlike the actual data plots, but the prediction data plots with ARIMA and hybrid models were like actual data plots.

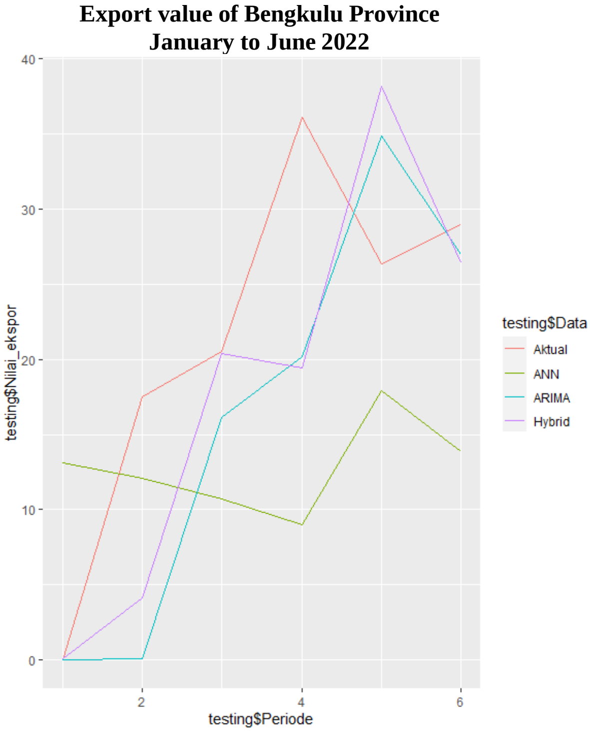

Figure 5. Comparison chart of actual values and predicted exportvalues with ARIMA, ANN, and hybrid models

The following is a comparison of MAAPE and MASE values for the ARIMA, ANN, and hybrid ARIMA-ANN models at the training and testing stage:

Table 3. Comparison of MAAPE and MASE values of ARIMA, ANN, and hybrid ARIMA-ANN models

Model	*Training*		*Testing*	
	MAAPE	MASE	MAAPE	MASE
ARIMA (0,1,1)	0.2198	0.8428	0.5606	0.8289
ANN (12-8-1)	0.2946	1.0395	0.6251	1.3557
ARIMA (0,1,1)- ANN (12-8-8-1)	0.2207	0.8364	0.5289	0.7664

Table 3 shows that the ANN model has a higher error rate than ARIMA and the ARIMA-ANN hybrid both at the training and testing stages. At the training stage, the performance of the ARIMA model is equal to ARIMA-ANN hybrid model. The ARIMA-ANN hybrid model outperforms the ARIMA and ANN models in the testing stage in terms of both MAAPE and MASE. The ARIMA-ANN hybrid model provides an increase in performance for ARIMA by 5.99% in terms of MAAPE, while in MASE it is 8.15%. The results showed that combining the two models can significantly reduce forecasting errors. Therefore, the ARIMA-ANN hybrid model is the best model for predicting the export value of Bengkulu Province for data from January 2014 to June 2022. Furthermore, the ARIMA-ANN hybrid model will be used to predict the export value of Bengkulu Province in the next three periods. The following are the results of the prediction of the export value of Bengkulu Province for the period of July to September 2022:

Table 4. Export value prediction for the period of July to September 2022

Period	Export Value (Million US$)
July 2022	27.94
August 2022	28.05
September 2022	28.03

5. CONCLUSION

Based on the discussion in this study, the ARIMA-ANN hybrid model is the best in predicting the export value of Bengkulu Province for export value data through the Baai Island Port from January 2014 to June 2022 with a MAAPE of 0.5289 and a MASE of 0.7664. The mathematical model of hybrid ARIMA-ANN is as follows:

$$Z_t = L_t + N_t$$

$$= (L_{t-1}^{0.061} + 0.061e_t + 0.0305e_{t-1})^{\frac{1}{0.061}} +$$

$$f\Big((0.1644) + \big((YY_1 \times 0.3146) + (YY_2 \times 0.4599) + (YY_3 \times (-0.4106)) + (YY_4 \times (-1.02116)) + (YY_5 \times 1.3099) + (YY_6 \times (-0.3114)) + (YY_7 \times (-0.7650)) + (YY_8 \times (-0.1865))\big)\Big) + \varepsilon_t$$

Prediction of the export value of Bengkulu Province for the next 3 periods using the ARIMA (0,1,1) - ANN (12-8-8-1) model, namely July 2022 amounting to 27.94 million US $, August 2022 amounting to 28.05 million US $ and 28.03 million US $ in September 2022.

REFERENCES

[1] BPS Bengkulu Province, Bengkulu Province in Figures Year 2022, Bengkulu: BPS Bengkulu Province, 2022.

[2] Box, G. E. P., G. M Jenkins and G. C. Reinsel, Time Series Analysis Forecasting and Control, New Jersey: Prentice-Hall, Inc., 1994.

[3] Zhang, G., B. E. Patuwo and M. Y. Hu, “Forecasting with Artificial Neural Networks: The State of The Art”, International Journal of Forecasting, 14. pp.35-62, 1998.

[4] Zhang, G. P., “Time Series Forecasting Using A Hybrid Arima and Neural Network Model, Journal of Neurocomputing, 50. pp. 159-175, 2003.

[5] Guha, B. and G. Bandyopadhyay, “Gold Price Forecasting Using ARIMA Model”, Journal of Advanced Management Science, Vol. 4, No. 2, pp. 1-5, 2016.

[6] Katoch, R. and A. Sidhu, “An Application of ARIMA Model to Forecast the Dynamics of COVID-19 Epidemic in India”, Sage Journals, 2021. https://doi.org/10.1177/0972150920988653.

[7] Jiang, S., M. Sharafisafa and L. Shen, “Using Artificial Neural Networks to Predicted Influences of Heterogeneity on Rock Strength at Different Strain Rates”, Materials Journal, Vol. 14, No. 3042, pp. 1-14, 2006.

[8] Toga, G., B. Atalay and M. D. Toksari, “COVID-19 Prevalence Forecasting Using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey”, Journal of Infection and Public Health 14 (2021) 811–816, 2021.

[9] Wang, X. and M. Meng,”A Hybrid Neural Network and ARIMA Model for Energy Consumption Forecasting”, Journal of Computers, Vol. 7, No. 5, pp. 1184-1190, 2012.

[10] Lailiyah, W. H and Manuharawati, “Application of Autoregressive Integrated Moving Average (ARIMA) Method on Forecasting Export Value in Indonesia”, Scientific Journal of Mathematics, Vol. 6, No. 3, pp. 45-52, 2018.

[11] Nugraha, A., Y. Suparman and A. A. Pravitasari, “Application of Artificial Neural Network Backpropagation to Forecast Indonesia's Export Value”, National Seminar on Statistics X (2021), Bandung, 2021.

[12] Kamadewi, R and A. I. Achmad, “Hybrid ARIMA (Autoregressive Integrated Moving Average)-ANN (Artificial Neural Network) Modeling of Indonesian Inflation Data for 2009-2020”, Proceedings of Statistics, Bandung, 2021.
[13] Fausett, L. V., 1994, *Fundamentals of Neural Networks: Architectures, Algorithms, and Aplications*, Prentice-Hall, New Jersey.