ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

Volume 7 Nomor 1, Februari 2022, halaman 53-65

MODELING OF THE NUMBER OF TOURISTS WITH AUTOREGRESSIVE INTEGRATED MOVING AVERAGE AND RECURRENT ARTIFICIAL NEURAL NETWORK

PEMODELAN JUMLAH WISATAWAN DENGAN *AUTOREGRESSIVE INTEGRATED MOVING AVERAGE* DAN *RECURRENT ARTIFICIAL NEURAL NETWORK*

Rahmat Hidayat[1*], Besse Helmi Mustawinar[2]
[1*]Universitas Cokroaminoto Palopo, Kota Palopo Sulawesi Selatan, dayatmath@gmail.com
[2]Universitas Cokroaminoto Palopo, Kota Palopo Sulawesi Selatan, emmy.emm92@gmail.com

ABSTRAK

Pariwisata telah menjadi kawasan prioritas bagi pembangunan ekonomi Indonesia. Pariwisata diharapkan menjadi salah satu pendorong utama pertumbuhan ekonomi Indonesia melalui penciptaan lapangan kerja dan usaha, pendapatan mata uang asing dan pembangunan infrastruktur. Selain itu, pariwisata dapat digunakan untuk memperkenalkan identitas dan budaya masyarakat. Oleh karena itu, pembangunan pariwisata akan terus berlanjut dan meningkat melalui perluasan dan pemanfaatan sumber daya dan kemungkinan pariwisata nasional. Dalam penelitian ini akan diramalkan jumlah kedatangan wisatawan asing dengan menggunakan metode ARIMA dan model Elman-RNN, mengingat bahwa pola kedatangan wisatawan tidak selalu linier. Data yang digunakan adalah data hasil survey Badan Pusat Statistika. Data dibagi menjadi dua bagian yaitu data *in-sample* dan data *out-sample*. Dari kedua model tersebut, model dari jaringan Elman-RNN adalah model terbaik dengan nilai MAPE dan RSME paling kecil.
Kata Kunci : Pariwisata, ARIMA, RNN

ABSTRACT

Tourism has become a priority area for Indonesia's economic development. Tourism is expected to be one of the main drivers of Indonesia's economic growth through job and business creation, foreign currency earnings and infrastructure development. In addition, tourism can be used to introduce the identity and culture of the community. Therefore, tourism development will continue and increase through the expansion and utilization of national tourism resources and possibilities. In this study, the number of foreign tourist arrivals will be predicted using the ARIMA model and the Elman-RNN model, given that the pattern of tourist arrivals is not always linear. The data used is the data from the survey results of the Central Statistics Agency. The data is divided into two parts, namely in-sample data and out-sample data. Of the two models, the model of the Elman-RNN network is the best model with the smallest MAPE and RSME values.
Keywords: *Tourism, ARIMA, RNN*

How to Cite: Hidayat, R & Mustawinar, B. H. (2022). Modeling Of The Number Of Tourists With Autoregressive Integrated Moving Average And Recurrent Artificial Neural Network. *Mathline: Jurnal Matematika dan Pendidikan Matematika*, Vol. 7 No. 1, 53-65.
DOI: https://doi.org/10.31943/mathline.v7i1.262

PRELIMINARY

Tourism has become a priority area for Indonesia's economic development. Tourism is expected to be one of the main drivers of Indonesia's economic growth through job creation and business, foreign currency income and infrastructure development (Wicaksana and Fatoni, 2021). In addition, tourism can be used to introduce the identity and culture of the community. Therefore, tourism development will continue and increase through the expansion and utilization of national tourism resources and possibilities.

In the development of national tourism, attitudes towards life and environmental quality must be maintained. Tourism development follows various tourism sector development programs. Not only medium and large tourism industries The tourism industry can be profitable on a small scale (Shahraki and Keshtegar, 2019).

Autoregressive Integrated Moving Average (ARIMA) and Recurrent Artificial Neural Network (RNN) are several prediction methods that can be used to model the number of foreign tourist arrivals. ARIMA assumes that the mathematical model is a linear function of some previous observations and random error. A number of studies have shown that nonlinear models may be able to explain and predict time series better than linear models if the prediction accuracy of the linear model is relatively low and the prediction error is large (Zhang, 2003). Artificially RNN is an example of a nonlinear technique that has a flexible functional form and contains several parameters that cannot be interpreted such as a parametric model (Suhartono, 2007). RNN can perform nonlinear modeling without first knowing the relationship between input and output variables.

In his research, Cho uses three methods: exponential smoothing, ARIMA, and RNN to predict the number of tourists who come to Hong Kong from various countries. The results of the analysis show that RNN gives the best results in predicting the number of tourist arrivals. Gong and Li (2020) in their research results show that the ANN method gives the best results.

Tourist visits tend to show a moving average (MA) pattern in the ARIMA modeling results. Saayman (2010) who predicts the number of tourists coming to South Africa, and Widyawati (2013) who predicts the number of airplane passengers at Ngurarai International Airport, find that the ARIMA model includes the MA pattern. About RNN modeling which contains MA components is the Elman Recurrent Neural Network (RNN). Therefore, the network chosen for forecasting is the RNN network. RNN uses the network output as input to obtain the next output, allowing the RNN to explain the effects of the AR and MA sequences simultaneously (Endharta and Suhartono, 2009). This study models the

number of foreign tourists who come to Indonesia in two different ways. One uses the ARIMA model and the other uses the RNN model. The best model is sought from the two methods by comparing the MAPE and MSE values.

METHOD

This research was conducted using a quantitative approach. Quantitative research is research that involves theory, design, hypotheses and determining the subject. Then it is supported by data collection, data processing and data analysis before writing conclusions. The methods or theories used in this research are:

1. ARIMA

A time series is a continuous series of observations. Observation settings are generally time-based, especially evenly spaced, but time series settings can also be spatially based. Some of the goals in time series studies include understanding and explaining formation mechanisms, predicting future values, and optimizing control systems (Wei, 2006).

ARIMA models are divided into non-seasonal ARIMA models, seasonal ARIMA models and a combination of non-seasonal and seasonal ARIMA models or often referred to as multiplicative seasonal ARIMA. In general, non-seasonal ARIMA models consist of autoregressive (AR) models, moving average (MA) models, ARMA models and ARIMA models. The ARIMA model is a time series model that is not stationary with respect to the mean and requires a differencing process of d to be stationary. The general form of the ARIMA model on the order of p, q with differencing as much as d or ARIMA (p,d,q) is as follows (Guo et al, 2019).

$$\phi_p(B)(1-B)^d Z_t = \phi_q(B)a_t$$

2. Artificial Neural Network (ANN) and Recurrent Neural Network (RNN)

An Artificial Neural Network (ANN) is an information processing system with a specific performance similar to a biological neural network. ANN was developed as a general mathematical model of biological neurons based on the following assumptions (Fausett, 1994).

1. Information processing is carried out by many simple elements called neurons.
2. Signals are sent between neurons via connecting links.
3. Each link is usually assigned a weight that doubles the transmitted signal.

4. Each neuron has a corresponding network input activation function to determine its output signal.

ANN has several components such as neurons, layers, activation functions, and weights. ANN modeling can be considered as a network form of the ANN itself, including the number of neurons in the input layer, hidden layer, output layer, and transfer function (Endharta and Suhartono, 2009). The ANN architecture that is widely used is the Recurrent Neural Network (RNN). In RNN there is a relationship between units that form a loop, thus allowing the movement of information from the output unit to return to the input unit (Liboschik, 2017; Lu, 2019; Oral, 2019).

Elman Recurrent Neural Network (ERNN) is an ANN with an architecture in which there is a group of units that receive feedback signals from past data. These unit groups are called context units. The weight of the feedback connection to the context unit is fixed, and the information processing remains sequential based on time, so that in training it is not more difficult than the backpropagation network in general (Park, 2018; Ma, 2018).

3. Analysis Method

In this study, we used two prediction methods, namely ARIMA model and Elman RNN model, and compared the accuracy of each prediction result. The stages of the analysis are as follows:

1. Divide the data into two parts: in-sample data and out-of-sample data. The data in the sample used is 84 data and the data outside the sample is 12 data. The modeling uses sample data.
2. Check for data stationarity. If the distribution of the data is not constant, perform a BoxCox transform to stabilize the data. If the data is unstable the Silent average makes the difference.
3. After the mean and variance of the data are stable, the next step is to make time series plots, ACF plots, and PACFs.
4. Identify intermediate model estimates based on ACF and PACF plots.
5. Estimating the ARIMA model parameters and then testing the significance of the parameters.
6. Use the ARIMA model to predict a sample data of future tourist arrivals from a country.
7. Calculation of the Mean Absolute Percentage Error (MAPE) and Mean Square Error (RMSE) values from the out-sample prediction of the ARIMA model.

8. Implementation of Elmann Recurrent Neural Networks (RNN) training with significant input delay from the ARIMA model.
9. The use of the model from the RNN training to predict sample data from foreign tourist arrivals in the future.
10. Using the RNN training network to calculate MAPE and RMSE values from the outsampled prediction results.
11. Compare the calculated MAPE and RMSE values for each method to determine which method is more appropriate.

RESULT AND DISCUSSION

1. Descriptive Analysis of the Number of Foreign Tourists Coming to Indonesia

Tourism has become a priority sector in Indonesia's economic development. Tourism is needed to be one of the main drivers in increasing the speed of economic growth in Indonesia through job creation and business opportunities, foreign exchange earnings, and infrastructure development. In addition, tourism can also be used to introduce national identity and culture. Thus, tourism development will continue and be improved through expansion and utilization of national tourism resources and potential.

As one of the export commodities that cannot be seen in real terms, the role of tourism is increasing in the Indonesian economy. In developing international tourism, a targeted and appropriate program is needed in order to increase the number of foreign tourist arrivals (tourists). This can be done by increasing marketing activities and improving various facilities needed by foreign tourists, such as immigration services, transportation facilities, banking, accommodation, restaurants, travel agencies, and so on.

To improve marketing activities, appropriate planning is needed based on quantitative and qualitative information about foreign tourists in the past. With the Covid-19 pandemic that has hit Indonesia since early 2020, information about foreign tourists is increasingly needed as material for evaluation and development planning in the future. The unavailability of data in the past will make it difficult to make a careful and targeted plan. The statistical data presented in this publication is intended to meet one of these needs.

The number of foreign tourist visits to Indonesia until before 2020 shows positive growth. The highest number of visits was recorded in 2019 which touched 16.11 million visits. However, with the outbreak of the COVID-19 pandemic, a number of countries, including Indonesia, issued several crossing restrictions for foreigners to enter Indonesia. This has an impact on the decline in the number of foreign tourist visits in 2020.

Throughout 2020, the number of foreign tourist arrivals only reached 4.05. The average pattern of foreign tourist arrivals per month can be seen in Figure 1 below.

Table 1. Average Number of Foreign Tourists per Month

Month	Observation	Average	St. Deviation
January	8	956341.1	247162
February	8	930005.3	209854.2
March	8	931785	299349
April	8	870853.8	382950
Mei	8	876122.3	360023.7
June	8	926802.3	394331.6
July	8	991271.9	465664.6
Augustus	8	1015165	461753.6
September	8	954188.8	407733.6
October	8	922840.4	385076.9
November	8	879956.9	348231.3
December	8	994364.3	390421.3

Figure 1. Average Foreign Tourist Arrivals per Month

2. ARIMA Method

In determining the ARIMA model, the data to be used is in-sample data. The pattern of foreign tourist arrivals from January 2013 to December 2020 can be seen in Figure 2 below.

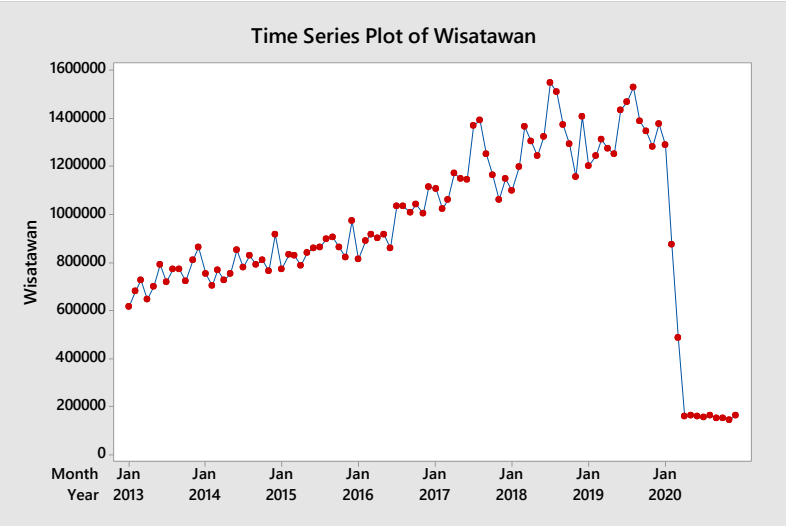

Figure 2. Time Series Plot of Foreign Tourist Arrivals Per Month

First, check whether the data is stationary or not. To find out the stationarity of the data in the variance, the Box-Cox test was performed, while to determine the stationarity of the mean, it was seen from the ACF plot of the data.

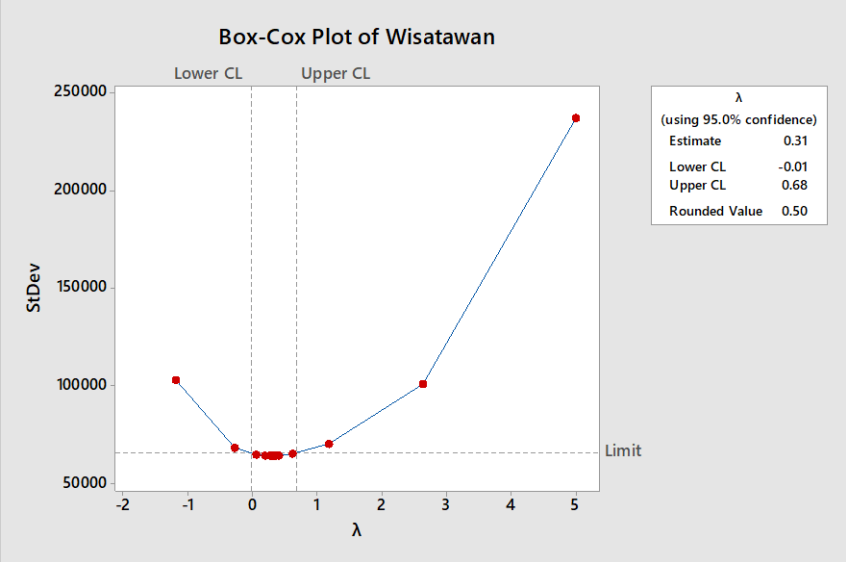

Figure 3. Box-Cox Plot of Monthly Foreign Tourist Arrival Data

A data is said to be stationary in variance if the Rounded Value is one. Based on the picture above, it can be seen that the Rounded Value is 0.5, meaning that the data is not stationary in the variance. So the transformation is carried out. The results can be seen in the following output.

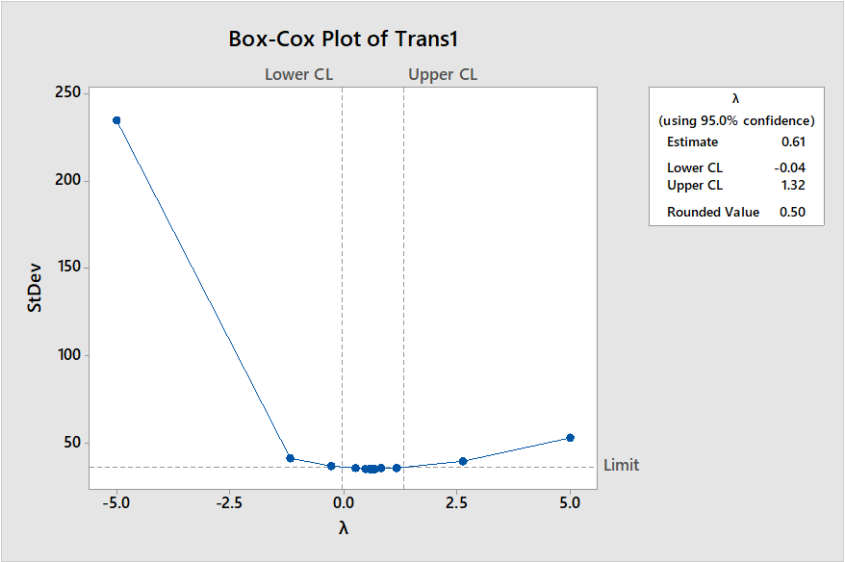

Figure 4. Plot of the first Box-Cox Transformation of Arrival Data Foreign Tourists Per Month

Based on the output above, it can also be seen that the variance is not stationary yet, so that Indonesia is now undergoing a transformation. The results of the transformation can be seen in the output below.

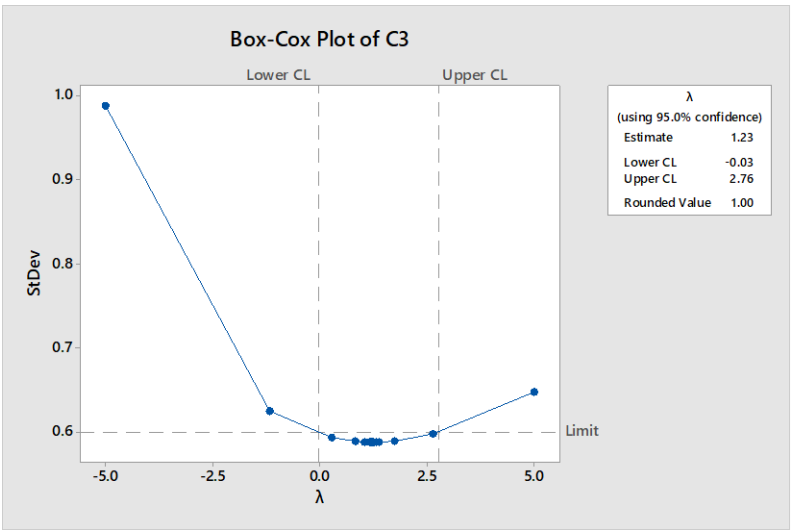

Figure 5. Plot of the second Box-Cox Transformation of Arrival Data Foreign Tourists Per Month

Based on the Box-Cox plot of the in-sample data in Figure 5, it can be seen that the value of is 1, which indicates that the data is stationary in variance.

Furthermore, it is tested whether the data is stationary in the average by looking at the ACF plot. The results can be seen in the following image:

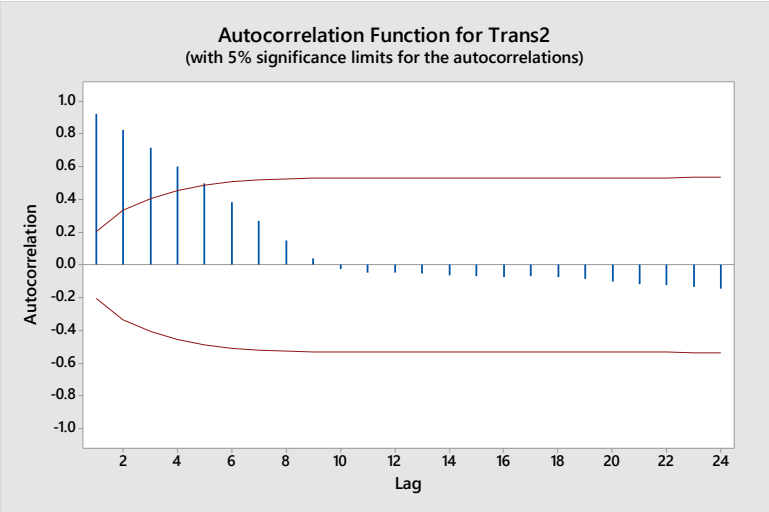

Figure 6. ACF Plot of Monthly Foreign Tourist Arrivals

The plot of the ACF data (see Figure 6) shows that there are five lags outside the confidence interval. So it can be concluded that the data is not stationary in the mean. Therefore, 1 lag differencing will be performed on the data. The results of the first differencing are as follows

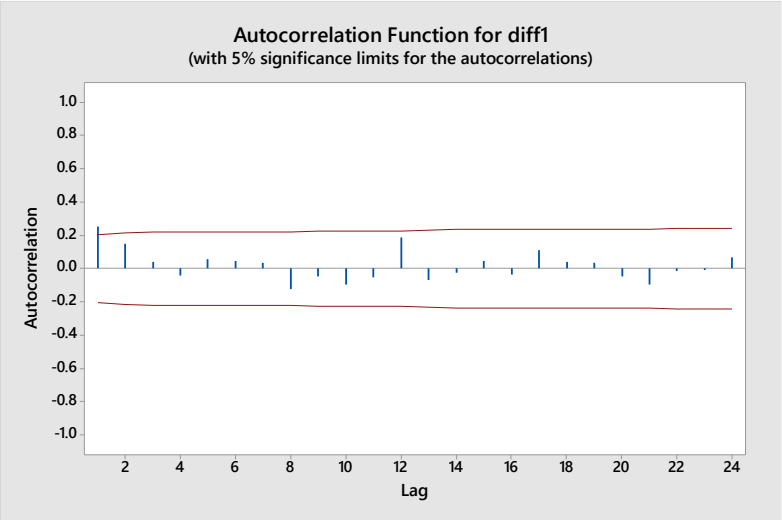

Figure 7. ACF Plot of 1 Lag Data Differencing

Based on the picture above, it can be seen that the data is stationary with respect to the average. Furthermore, the PACF plot is also seen as follows:

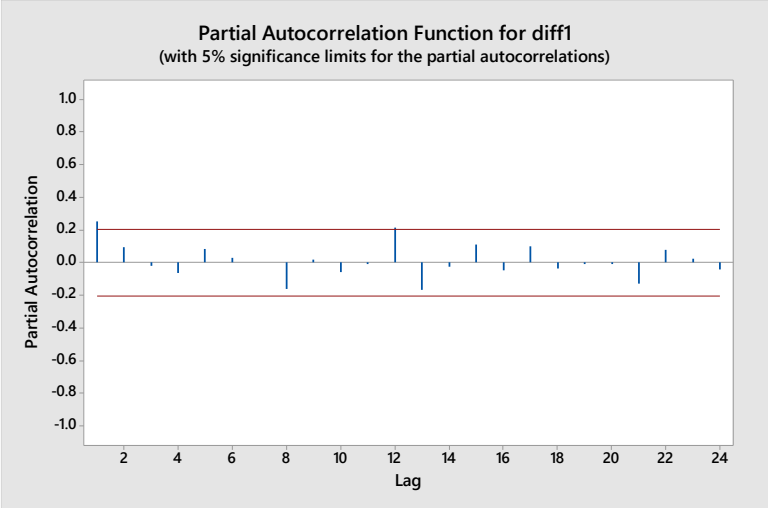

Figure 8. PACF Plot of 1 Data Differencing

Based on the results that have been obtained previously, so that the tentative model is obtained:

Table 2. Possible ARIMA Model

Possible Model
ARIMA (2,1,1)
ARIMA (2,1,0)
ARIMA (1,1,1)
ARIMA (1,1,0)
ARIMA (0,1,1)
ARIMA (0,1,0)

The model that meets the requirements of parameter significance, residual white noise and normally distributed is ARIMA (1,1,0). Furthermore, an assessment of the criteria from the out-sample of the number of tourists is carried out as follows:

Table 3. The best ARIMA model

Model	**MSE**
ARIMA (1,1,0)	27.627

Table 3 shows the criteria for evaluating the best model based on the smallest RMSE value. In the model of the number of train passengers, the best model for forecasting is ARIMA (1,1,0) which has the smallest value when compared to other models. The ARIMA model (1,1,0) is used as a model for the number of tourists, namely

$$\phi_1(B)(1-B)^1 Z_t = \theta_0(B)a_t$$

$$(1 - B - \phi_1 B + \phi_1 B^2)Z_t = a_t$$

$$Z_t - BZ_t - \phi_1 BZ_t + \phi_1 B^2 Z_t = a_t$$

$$Z_t = Z_{t-1} + \phi_1 Z_{t-1} - \phi_1 Z_{t-2} + a_t$$

$$Z_t = Z_{t-1} + 0.2586Z_{t-1} - 0.2586Z_{t-2} + a_t$$

Table 3 shows that the forecast results for the number of tourists in 2021 using the ARIMA model (1,1,0), the number of train passengers for January 2021 to December 2021 ranges from 166545 to 166899 tourists.

Table 4. Forecast of the Number of Tourists

Month	Forecast
January	166855
February	166855
March	166894
April	166899
Mei	166899
June	166899
July	166899
Augustus	166899
September	166899
October	166899
November	166899
December	166899

3. Elman-Recurrent Neural Network Model

In modeling with the neural network method, it is necessary to determine the input, activation function, and the number of units in the hidden layer. After the data is inputted, the data is divided into 2 parts, namely in sample data and out-sample data. Then, an Elman-recurrent neural network (ERNN) is formed that uses a sigmoid tangent transfer function in the hidden layer, and a linear transfer function in the output layer. The update of weights and biases on ERNN is done by using gradient descending momentum and adaptive learning rate methods. The training process is carried out using in-sample data, with a maximum number of epochs of 1000, a learning rate of 0.1 and a momentum of 0.8. The number of hidden layers used in ERNN varies from ERNN with 1 hidden layer to ERNN with 10 hidden layers. It aims to determine the number of hidden layers that provide training data results with the best out sample criteria.

Based on Table 5, it can be seen that the best recurrent network from the in-sample data is the ERNN network with 4 units in the hidden layer. This is indicated by the MAPE

and MAE values generated from the out-sample data with the smallest values, namely 5.417 and 14541. The letter k in Table 5 shows the amount of weight in the network. Therefore, the NN network that will be created is a network (5,4,1) which is a network with 5 inputs, 4 units in the hidden layer with a sigmoid tangent transfer function, and 1 unit in the output layer with a linear function.

Table 5. Criteria for the Goodness of the RNN Network

Hidden Unit	**k**	**in-Sample Criteria**		**out-Sample Criteria**	
		AIC	**MSE**	**MAPE**	**MSE**
1	8	25.33	1262	5.673	21.461
2	15	26.83	1193	4.553	22.467
3	22	24.54	2141	**3.562**	**20.456**
4	29	25.74	2612	5.674	22.449
5	36	27.11	1398	5.673	21.984
6	43	27.45	1683	5.624	23.726
7	50	24.55	2362	4.235	22.846
8	57	25.73	1482	6.421	21.947
9	64	26.52	1208	6.367	22.461
10	71	25.91	1499	6.456	23.947

4. Comparison of ARIMA Model and Elman-RNN Network

Model ARIMA yang akan dibandingkan adalah model ARIMA (1,1,0) dan jaringan Elman-RNN. Plot hasil dari pengolahan data out-sample digunakan untuk menentukan model yang terbaik. The more the out-sample forecast of a method approaches the actual out-sample data, the better the method.

Table 6. Out-Sample Goodness Criteria

Metode	Out-Sample Method	
	MAPE	MSE
ARIMA	5.248	27.627
RNN	**3.562**	**20.456**

In Table 6, it can be seen that the out-sample criteria for both methods are good. The MAPE and MSE values from the RNN network are smaller than the ARIMA model,

namely MAPE at 3.562% and MSE at 27.627. It can be concluded that the Elman-RNN network is the best method for predicting the number of foreign tourist arrivals to Indonesia.

CONCLUSION

Based on the analysis and discussion that has been carried out, it can be concluded as follows:

1. The ARIMA model used to predict the number of foreign tourist arrivals to Indonesia is the ARIMA model (1,1,0), with a MAPE value of 5.248% and an out-sample RMSE of 27,627.
2. The RNN method used produces a MAPE value of 3.562%, an out-sample RMSE of 20,456.
3. The best method based on both methods to predict the number of foreign tourist arrivals to Indonesia is the Elman-RNN method because the MAPE value is smaller than the MAPE value of the ARIMA model.

ACKNOWLEDMENT

We would like to thank the LPPM of Cokroaminoto University, Palopo for providing a Research Grant so that this research can be completed.

REFERENCES

Fausett, L., (1994). *Fundamentals of Neural Networks: Architecture, Algorithms, and Applications,* Prentice Hall.

Guo S, Li D, Li MY. 2019. Strict Stationarity Testing and GLAD Estimation of Double Autoregressive Models. *Journal of Econometrics*, 211, 319–337.

Gong, H dan Li, D. (2020). On the three-step non-gaussian quasi-maximum Likelihood estimation of heavy-tailed double Autoregressive models. *Journal of Time Series Analysis*, 43, 1-15.

Liboschik, T., Fokianos, K., dan Fried, R. (2017) tscount: An R package for Analysis of Count Time Series Following Generalized Linear Models. *Journal of Statistical Software*, 82(5), 1–51

Lu, M, AbouRizk, S. M. and Hermann, U. H. (2019). Estimating labor productivity using probability inference neural network. *Journal of Computing in Civil Engineering*, 14(4), 241–248.

Oral, E.L and Oral, M. (2019). Predicting construction crew productivity by using self-organizing maps. *Automation in Construction*, 19, 791–797.

Park, M.S., Eom, J.K., Heo, TY., dan Song, J. (2018). Intervention Analysis of the Impact of Opening a New Railway Line on Passenger Ridership in Seoul. *KSCE Journal of Civil Engineering*, 20, 2524-2534.

Ma, E., Liu, Y., Li, J., dan Chen, S. (2018). Anticipating Chinese tourists arrivals in Australia: A time series analysis. *Tourism Management Perspectives*, 17, 50-58.

Saayman, A., & Saayman, M. (2010). Forecasting tourist arrivals in South Africa. *Acta Commercii*, 10, 281-293.

Shahraki, A. S. &Keshtegar, A. (2019). Determining the Efficiency of Economic Tourism Industry in Chababar Free Zone by Using Data Envelopment Analysis (DEA) Method. *Journal of Iran Economic Rev*, 23, 1019–1039.

Suhartono. (2007). FeedForward Neural Networks untuk Pemodelan Runtun Waktu. *Disertasi*, Jurusan Matematika, Universitas Gadjah Mada.

Suhartono & Endharta, A. J. (2009). Short Term Electricity Load Demand Forecasting in Indonesia by Using Double Seasonal Recurrent Neural Networks. *International Journal of Mathematical Models and Methods in Applied Sciences*, *3* (3), 171-178.

Wei, W.W., (2006). *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.), Addison Wesley.

Wicaksana, A K., & Fatoni, E Z. (2021). Measurement and Benchmarking of Tourism Efficiency in Java-Indonesia-Nusa Tenggara Islands Using Data Envelopment Analysis. *Jurnal Kepariwisataan Indonesia*, 15(2), 86-102.

Widyawati, K.A. (2013). *Prediksi Jumlah Penumpang Pesawat di Bandar Udara Internasional Ngurah Rai dengan Model Arima, Analisis Intervensi dan Pendekatan Perubahan Struktur*, Institut Teknologi Sepuluh November.

Zhang, G. P. (2003). Time Series Forecasting Using Hybrid ARIMA and Neural Network Model. *Journal of Neurocomputing*, 50, 159-175.

Zhu Q, Zheng Y, Li G. 2018. Linear double autoregression. *Journal of Econometrics*, 207, 162–174.