Comparing of ARIMA and RBFNN for short-term forecasting

Haviluddin[a,1,*], Ahmad Jawahir[b,2]

[a] *Dept. of Computer Science, Faculty of Mathematics and Natural Science, Mulawarman University - Indonesia*
[b] *Researcher at ICT of Mulawarman University - Indonesia*
[1] *haviluddin@unmul.ac.id* *; [2] *ahmadjawahirabd@gmail.com*

ARTICLE INFO

Article history:
Received March 22, 2015
Revised March 30, 2015
Accepted March 31, 2015

Keywords:
ARIMA
RBFNN
MSE
Tourist arrival

ABSTRACT

Based on a combination of an autoregressive integrated moving average (ARIMA) and a radial basis function neural network (RBFNN), a time-series forecasting model is proposed. The proposed model has examined using simulated time series data of tourist arrival to Indonesia recently published by BPS Indonesia. The results demonstrate that the proposed RBFNN is more competent in modelling and forecasting time series than an ARIMA model which is indicated by mean square error (MSE) values. Based on the results obtained, RBFNN model is recommended as an alternative to existing method because it has a simple structure and can produce reasonable forecasts.

I. Introduction

Currently, time series forecasting methods are constantly evolving where this method is a quantitative approach with past data as a basis for forecasting [1]. Therefore, various forecasting techniques based on mathematics is one of the oldest models (i.e. autoregressive-AR, moving average-MA, exponential smoothing-ES and autoregressive integrated moving average-ARIMA) in which many of researchers have been using these techniques. Some researchers have proposed ARIMA models to predict network traffic in ICT at Mulawarman University in East Kalimantan in the period of June 20-24, 2013 [2]. In the economics area, ARIMA models have been used for estimation of Malaysia Crude Oil Production (MCOP) from January 2005 to May 2010 [3]. In the hydrologic area, ARIMA models have been proposed for the forecasting of monthly inflow of Dez dam reservoir from 1960 to 2007. The statistics related to the first 42 years were used to train the models and the 5 past years were used to forecast [4]. All those researchers have confirmed that by using ARIMA, good results and accuracy can be obtained. Although mathematics models are proved to be reasonably powerful, but it still has some obstacles especially when applied to non-linear data.

For that reason, many researchers have also tried to apply artificial neural networks-ANNs (i.e. backpropagation-BPNN, radial basis function-RBFNN, and recurrent neural network-RNN) to improve the prediction accuracy by using data non-linear. An approach using ANNs has been proposed to predict network traffic by using BPNN [5] and predict the students' achievement by using RBFNN [6]. In the economics area, ANNs models have been used for stock market predictions [7, 8]. In the hydrologic area, ANNs models have been proposed by researchers to predict the weather, wind speed, and rainfall [9, 10].

However, one of the important issues on ANNs is the training or learning of the networks in which to find a set of optimal network parameters. These issues are the drawbacks of ANNs (i.e. over fitting, local minimum, and slow convergence). Then, hybrid models by using mathematics or ANNs models itself is a solution to improve of ANNs performances. Recently, numerous researchers have been trying related model combining as an alternative in prediction area including, ARIMA with RBFNN, ARIMA with BPNN, BPNN, RBFNN with genetic algorithm (GA), particle swam optimization (PSO) has been proposed to provide better prediction performance [1, 7, 8, 11, 12]. Therefore, this paper will apply two models, namely ARIMA and RBFNN that have been developed and compared in order to predict the tourist quantity to Indonesia. Section 2 describes the architectures of ARIMA and RBFNN models. Section 3 explains the time series predictor and models. Section 4 describes the analysis and discussion of the results. Finally, conclusions are summarized in Section 5.

II. Methodology

In this section, a brief information on the general tourist quantity prediction models is presented including time series models, ARIMA, and RBFNN.

A. Time Series

The time series is a dataset of observations ordered in time. A time series is an ordered sequence of observations and many ways are used to forecast the time series data. In principle, a time series model is used to predict the values of data ($y_{t+1}, y_{t+2},\ldots,y_{t+n}$) based on the data ($x_{t+1}, x_{t+2},\ldots, x_{t+n}$). In this experiment, data tourist quantity 1974-2013 (40 years of samples) was captured from BPS website http://www.bps.go.id, Table 1 and Fig. 1. Then, the data are analyzed by using MATLAB R2013b. The ARIMA and RBFNN were engaged.

Table 1. Real tourist arrival to Indonesia 1974-2013

1974	**1975**	**1976**	**1977**	**1978**	**1979**	**1980**	**1981**	**1982**	**1983**
313.452	366.293	401.237	433.393	468.614	501.430	561.178	600.151	592.046	638.855
1984	**1985**	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**
700.910	749.351	825.035	1.060.347	1.301.049	1.625.965	2.177.566	2.569.870	3.064.161	3.403.138
1994	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
4.006.312	4.324.229	5.034.472	5.185.243	4.606.416	4.727.520	5.064.217	5.153.620	5.033.400	4.467.021
2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
5.321.165	5.002.101	4.871.351	5.505.759	6.234.497	6.323.730	7.002.944	7.649.731	8.044.462	8.802.129

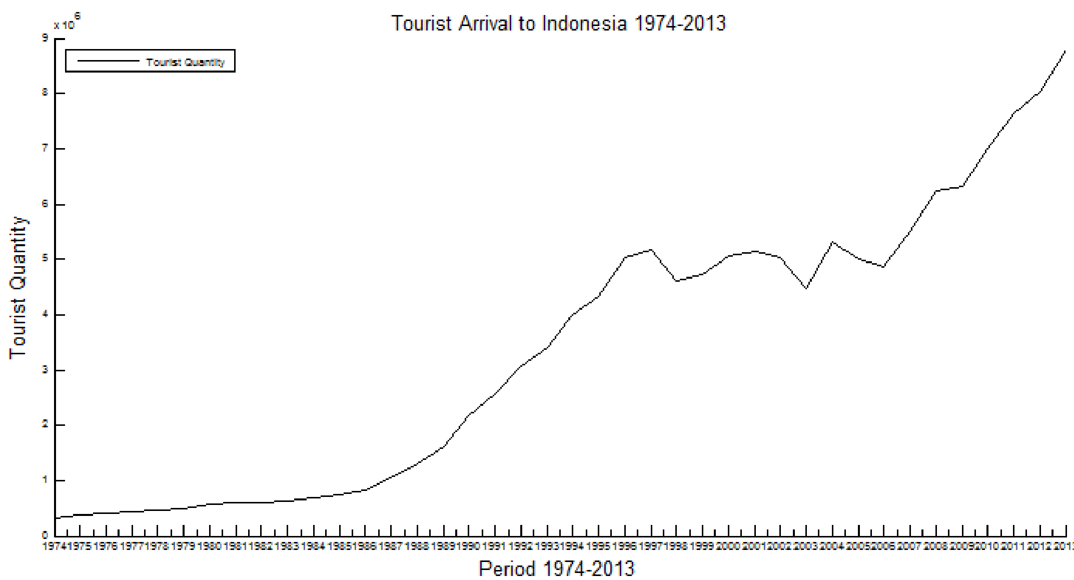

Fig. 1. Plots tourist arrival to Indonesia in period 1974-2013 (BPS, 2014)

B. ARIMA

One of the famous methods used in forecasting a time series data is ARIMA. The ARIMA method is used to analyze a time series data in which it is designed by integrating the AR (autoregressive) and MA (moving average) methods. The ARIMA (p, d, q) is a general method that is formulated with respect to the data series that are stationary only, where, p is the number of processes in AR, d is the number of differencing a time series of data to be stationary, and finally, q is the number of processes in MA. According to the Box-Jenskins methodology [13], there are four forecasting stages, that includes; (1) identification model; The data series will be carefully examined in order to determine whether the series contains a trend, seasonality, cycles or random phenomena. After that, the sample ACF and PACF of the original series are computed and examined in order to further confirm that the time series data is stationary. If the sample ACF decays very slowly, it indicates that differencing processes are needed, (2) parameter estimation; the purpose of model validation is to ensure that the right model is used. In this study, it can be done by using *t-statistic* and *p-value*, (3) model checking; the purposed model needs to be hypothesized and to have diagnostic test before it can be used for forecasting. In this test, we checked by *p-value* > α 0.05, and (4) forecasting; the forecasted values in confidence limit (upper and lower limits) provide 95% confidence interval. In this study, we used the *trial and error* method to get good model and prediction.

C. RBFNN

The RBFNN emerged as a variant of ANN in late 80's is a kind of feed-forward neural network (FFNN). The RBFNN structure has a three-layer FFNN which includes an input layer, single only of hidden layer with RBF neurons (Euclidean distance between the input signal vector and parameter vector of the network) and an output layer with linear neurons. Hence, the RBFNN has a unique training algorithm including supervised and unsupervised as well. Furthermore, RBFNN learning philosophy can be differentiated into two stages: first stage, self-organizing learning stage, solving the center and change of the hidden layer base functions; second stage, mentor learning stage, this stage is unwinding weights which is between the hidden layer and output layer [11, 12]. In this study, we used three layers and Euclidean function as an activation function (1). Furthermore, in this experiment we used the mean square error (MSE), then comparing the predicted output with the desired output between ARIMA and RBFNN. The architecture of RBFNN as shown in Fig. 2.

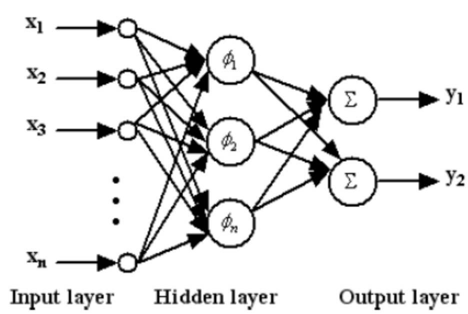

Fig. 2. The RBFNN architecture [12]

$$Y = \sum_{j=1}^{m} W_{jm}.\varphi\text{, where: Y output value, } \varphi = \text{hidden layer value, W = weights (0-1)} \quad (1)$$

The algorithm of RBFNN to analyze within time series data characteristics is:

1. Initialization of the network; randomly selecting some training and testing samples as the vectors $P_{(t-0)}=[p_{(t-5)}, p_{(t-4)}, ..., p_{(t-n)}]$, where *n* is a series data.
2. Find, D_{ij} distance between i to j *i,j=1,2,…,Q*, where Q is input-output vectors, R is input variable.

$$D_{ij} = \sqrt{\sum_{k=1}^{R}(p_{ik} - p_{jk})^2} \quad (2)$$

3. Find $a1$, where $a1$ is a result activation from distance data multiply bias, spread is constant

$$a1_{ij} = e^{-(b1*D_{ij})^2} \quad (3)$$

$$b1 = \frac{\sqrt{-\ln(0.5)}}{spread} \quad (4)$$

4. Calculation weights and biases, where w_{ij} is a new weights, w_{ij} (t) is a current weights, α is a learning rate.

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)[x_i - w_{ij}(t)] \quad (5)$$

III. Experimental

A. Analysis using ARIMA

The first analysis, tourist quantity data were tested by using ARIMA technique. Based on ARIMA Box-Jenskins rules, the data were listed in a sequence from 1974-2013 or contained 40 samples. In this experiment, we studied many models including ARIMA (1,0,0), (1,1,0), (1,1,1), (1,1,2), (2,0,0), (2,1,0), (2,1,1), (2,1,2), then decided to choose the best ARIMA (2,1,2) as a model for predicting as shown in Fig 3 and 4.

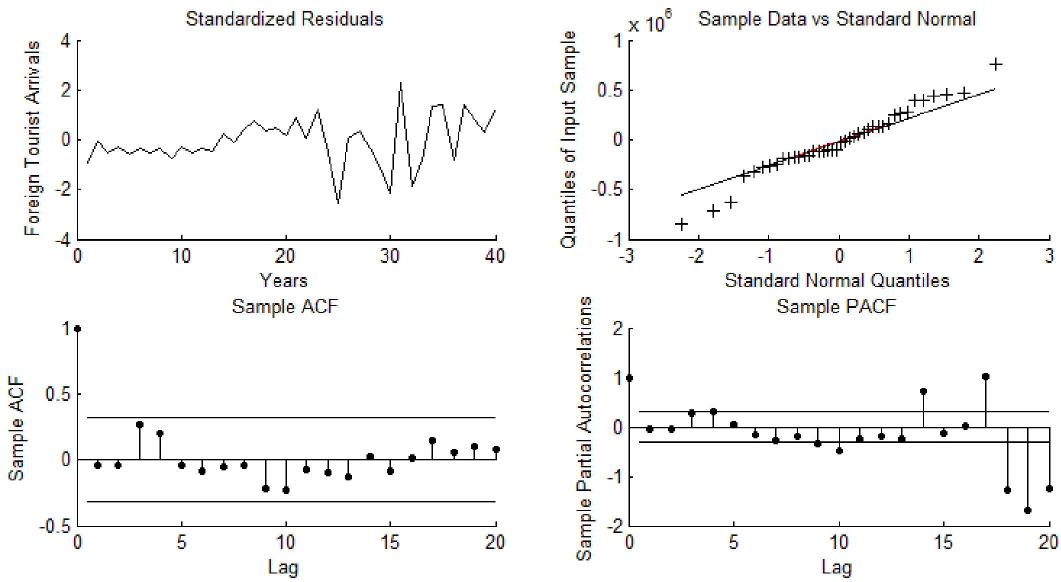

Fig. 3. Plots check goodness of fit model of ARIMA(2,1,2)

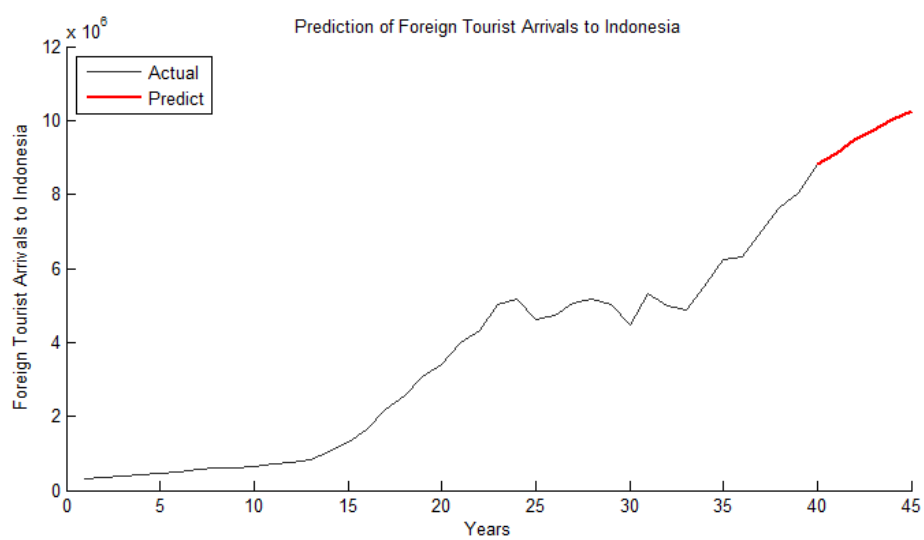

Fig. 4. Performance and plots forecast of ARIMA(2,1,2)

B. Analysis using RBFNN

In the second experiment, the tourist arrivals to Indonesia data were tested using RBFNN technique. Based on ANN's rules, the data were divided into training and testing data. The inputs and tests data would be normalized. The aim of the normalization process is to get the data with a smaller size that represents the original data without losing its own characteristics. In this experiment, the training data was 86% (30 samples series data) and testing was 14% (5 samples series data) as shown in Table 2. The normalization formula form is as follow,

Vol. 1, No. 1, March 2015, pp. 15-22

$$\bar{X} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{6}$$

where, X is the actual value of samples, X_{max} for maximum value, and X_{min} is the minimum value. In MATLAB function, the RBFNN can creating by *newrb(P,T,error_goal,spread)* function, which is this function create RBFNN structure, automatically selected the number of hidden layer and made the error to 0. In this study, we tried the sum-square error (SSE) goal values were 0.001, 0.002, and 0.003. The spread value of 200 was settled. In this experiment, we decided the RBFNN with SSE value was 0.001, spread was 200 as a good model. The RBFNN results are shown in Fig 5, 6 and 7.

Table 2. Real tourist arrival data after normalization

Group		**Input Neurons P=[p(t-5),p(t-4),p(t-3),p(t-2),p(t-1)]**					**Output Neurons T**
		p(t-5)	**p(t-4)**	**p(t-3)**	**p(t-2)**	**p(t-1)**	
Training	1	0.000	0.006	0.010	0.014	0.018	0.022
	2	0.006	0.010	0.014	0.018	0.022	0.029
	3	0.010	0.014	0.018	0.022	0.029	0.034
	4	0.014	0.018	0.022	0.029	0.034	0.033
	5	0.018	0.022	0.029	0.034	0.033	0.038
	6	0.022	0.029	0.034	0.033	0.038	0.046
	7	0.029	0.034	0.033	0.038	0.046	0.051
	8	0.034	0.033	0.038	0.046	0.051	0.060
	9	0.033	0.038	0.046	0.051	0.060	0.088
	10	0.038	0.046	0.051	0.060	0.088	0.116
	11	0.046	0.051	0.060	0.088	0.116	0.155
	12	0.051	0.060	0.088	0.116	0.155	0.220
	13	0.060	0.088	0.116	0.155	0.220	0.266
	14	0.088	0.116	0.155	0.220	0.266	0.324
	15	0.116	0.155	0.220	0.266	0.324	0.364
	16	0.155	0.220	0.266	0.324	0.364	0.435
	17	0.220	0.266	0.324	0.364	0.435	0.472
	18	0.266	0.324	0.364	0.435	0.472	0.556
	19	0.324	0.364	0.435	0.472	0.556	0.574
	20	0.364	0.435	0.472	0.556	0.574	0.506
	21	0.435	0.472	0.556	0.574	0.506	0.520
	22	0.472	0.556	0.574	0.506	0.520	0.560
	23	0.556	0.574	0.506	0.520	0.560	0.570
	24	0.574	0.506	0.520	0.560	0.570	0.556
	25	0.506	0.520	0.560	0.570	0.556	0.489
	26	0.520	0.560	0.570	0.556	0.489	0.590
	27	0.560	0.570	0.556	0.489	0.590	0.552
	28	0.570	0.556	0.489	0.590	0.552	0.537
	29	0.556	0.489	0.590	0.552	0.537	0.612
	30	0.489	0.590	0.552	0.537	0.612	0.698
Testing	31	0.590	0.552	0.537	0.612	0.698	0.708
	32	0.552	0.537	0.612	0.698	0.708	0.788
	33	0.537	0.612	0.698	0.708	0.788	0.864
	34	0.612	0.698	0.708	0.788	0.864	0.911
	35	0.698	0.708	0.788	0.864	0.911	1.000

Vol. 1, No. 1, March 2015, pp. 15-22

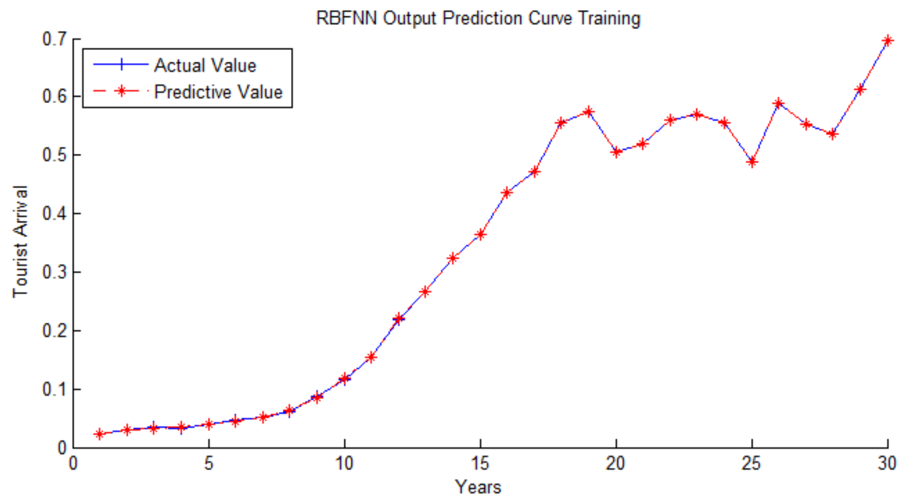

Fig. 5. Performance and plots training of RBFNN

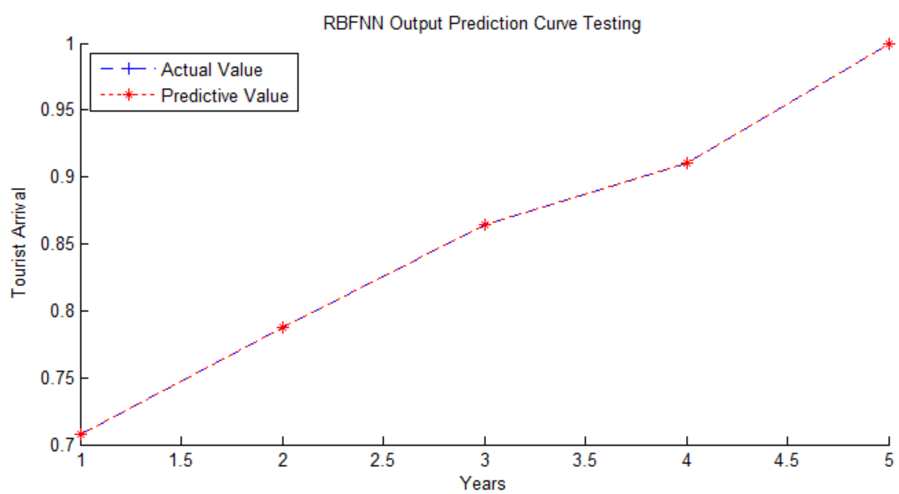

Fig. 6. Performance and plots testing of RBFNN

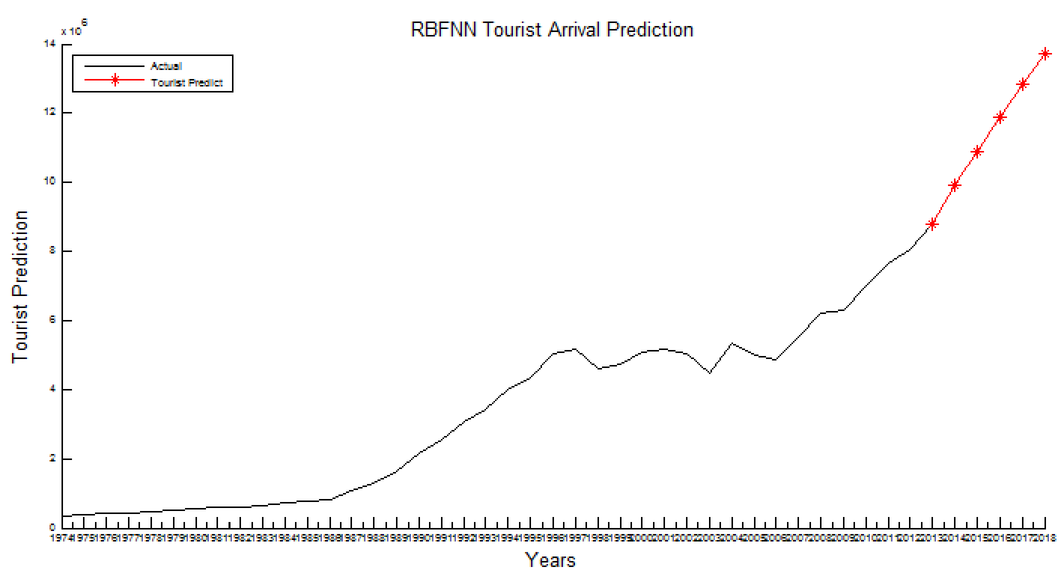

Fig. 7. Performance and plots forecast of RBFNN

IV. Results and Discussions

This section describes the test of tourist arrival data using two different models. Table 3 shows that the error prediction of ARIMA and RBFNN. We choose the MSE as an error prediction. The ARIMA error prediction was 0.00722784 and RBFNN was 0.00098188. This mean that the tourist arrival results had a good prediction accuracy by using the RBFNN technique with the setting parameters, spread was 200 and error goal was 0.001. In this study, to compare the predicted output with the desired output, MSE was predefined, as shown in Table 4. Then, the best results of MSE by using RBFNN, which that mean the RBFNN was good accuracy. The comparison prediction between ARIMA and RBFNN models of 5 years ahead, as shown in Fig. 8.

Table 3. Comparison of MSE from ARIMA and RBFNN models

Models	MSE
ARIMA (212)	0.00722784
RBFNN Error_goal = 0.001 Spread = 200	0.00098188

Table 4. Predicton results of tourist arrivals to Indonesia in 2014-2018

Years	ARIMA (211)	RBFNN
2014	9.128.791	9.908.224
2015	9.464.387	10.891.264
2016	9.724.224	11.892.736
2017	10.019.368	12.865.536
2018	10.259.238	13.727.744

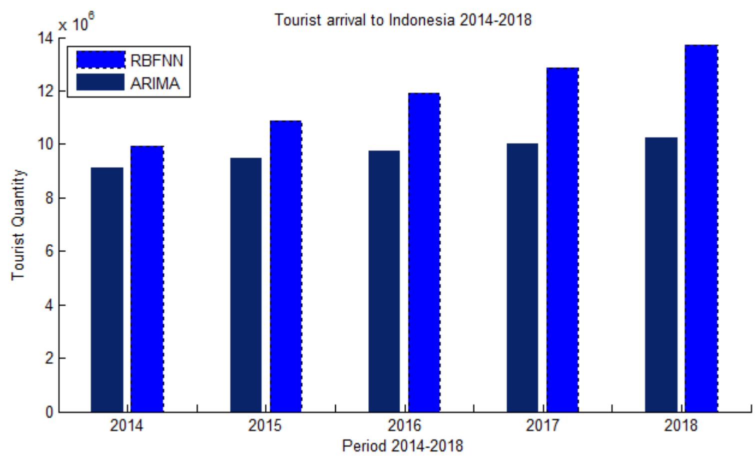

Fig. 8. Plots bar forecast of ARIMA and RBFNN

V. Conclusions

This paper has presented the performance comparison of statistical and machine learning techniques, namely ARIMA and RBFNN, in learning time series data. The mean squared errors are computed for each model and compared. Based on the results obtained, the RBFNN algorithm is found to be more efficient than ARIMA in modelling time series dataset related to tourist quantity of Indonesia. Furthermore, the future works include a comparison of a few ANN methods and the optimization process in order to obtain more accurate forecasting results.

Acknowledgment

We thank the anonymous peer reviewers for carefully revising our manuscript. Hopefully this research can be useful.

References

[1] G. Chen, K. Fu, Z. Liang, T. Sema, C. Li, P. Tontiwachwuthikul, and R. Idem, "The genetic algorithm based back propagation neural network for MMP prediction in CO_2-EOR process," *Fuel,* vol. 126, pp. 202–212, 2014.

[2] Haviluddin and R. Alfred, "Forecasting Network Activities Using ARIMA Method," *Journal of Advances in Computer Networks,* vol. 2, pp. 173-179, 2014.

[3] N. M. Yusof, R. S. A. Rashid, and Z. Mohamed, "Malaysia Crude Oil Production Estimation: an Application of ARIMA Model," in *2010 International Conference on Science and Social Research (CSSR 2010)*, Kuala Lumpur, Malaysia, 2010.

[4] M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir," *Journal of Hydrology,* vol. 476, pp. 433–441, 2013.

[5] Haviluddin and R. Alfred, "Daily Network Traffic Prediction Based on Backpropagation Neural Network," *Australian Journal of Basic and Applied Sciences,* vol. 8(24), pp. 164-169, 2014.

[6] Haviluddin, A. Sunarto, and S. Yuniarti, "A Comparison between Simple Linear Regression and Radial Basis Function Neural Network (RBFNN) Models for Predicting Students' Achievement.," in *International Conference on Education 2014 (ICEdu14) 4th - 6th June 2014.*, Universiti Malaysia Sabah - Kota Kinabalu, Malaysia, 2014, pp. 99-308.

[7] Y. Perwej and A. Perwej, "Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm," *Journal of Intelligent Learning Systems and Applications,* vol. 4, pp. 108-119, 2012.

[8] L. Yizhen, Z. Wenhua, l. Lin, j. Wu, and L. Gang, "The forecasting of Shanghai Index trend Based on Genetic Algorithm and Back Propagation Artificial Neural Network Algorithm," in *The 6th International Conference on Computer Science & Education (ICCSE 2011)*, SuperStar Virgo, Singapore, 2011.

[9] K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar, "A Rainfall Prediction Model using Artificial Neural Network," *2012 IEEE Control and System Graduate Research Colloquium (ICSGRC 2012),* 2012.

[10] M. Majumder and R. N. Barman, "Application of Artificial Neural Networks in Short-Term Rainfall Forecasting," *Application of Nature Based Algorithm in Natural Resource Management,* 2013.

[11] J. Wu, J. Long, and M. Liu, "Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm," *Neurocomputing,* vol. 148, pp. 136–142, 2015.

[12] J. W. Yu, "Rainfall time series forecasting based on Modular RBF Neural Network model coupled with SSA and PLS," *Journal of Theoretical and Applied Computer Science,* vol. 6, pp. 3-12, 2012.

[13] G. E. P. Box, G. M. Jenskins, and G. C. Reinsel. (2008). *Time Series Analysis Forecasting and Control 4th Edition*.