Environment and Ecology Research 10(1): 87-94, 2022
DOI: 10.13189/eer.2022.100108 http://www.hrpub.org

Vector Autoregressive Model with Seasonal Indicator and Feed Forward Neural Network for Modeling Rainfall in Malang and Karangkates

Eni Sumarminingsih*, Solimun, Djihan Wahyuni, Triardy Satria Wibawa

Department of Statistics, Universitas Brawijaya, Indonesia

Received November 7, 2021; Revised December 22, 2021; Accepted January 17, 2022

Cite This Paper in the following Citation Styles

(a): *[1] Eni Sumarminingsih, Solimun, Djihan Wahyuni, Triardy Satria Wibawa , "Vector Autoregressive Model with Seasonal Indicator and Feed Forward Neural Network for Modeling Rainfall in Malang and Karangkates," Environment and Ecology Research, Vol. 10, No. 1, pp. 87 - 94, 2022. DOI: 10.13189/eer.2022.100108.*

(b): *Eni Sumarminingsih, Solimun, Djihan Wahyuni, Triardy Satria Wibawa (2022). Vector Autoregressive Model with Seasonal Indicator and Feed Forward Neural Network for Modeling Rainfall in Malang and Karangkates. Environment and Ecology Research, 10(1), 87 - 94. DOI: 10.13189/eer.2022.100108.*

Abstract Water is a basic need for the survival of every living thing. Rainfall is the main source of water availability. Lack of water supply can have a tremendous negative impact. On the other hand, heavy rainfall can cause flooding which has a bad impact. Accuracy in rainfall prediction is useful in crop planning strategies and flood and drought prevention. Research on rainfall used several models including Autoregressive (ARIMA), Seasonal ARIMA (SARIMA), Vector Autoregressive (VAR) and Feed Forward Neural – Network (FFNN). The purpose of this study is to establish a VAR with Seasonal Indicator and FFNN model for rainfall in Malang and Karangkates and compare the performance of the models. The novelty of this research is that we added the seasonal indicator variable to the VAR model as an exogenous variable and as an input to the feed forward neural network model. The best VAR model for rainfall in Malang and Karangkates is the first order VAR model (VAR(1)) with seasonal indicator variables. While the FFNN model for rainfall in Malang and Karangkates is the FFNN model with the tangent hyperbolic activation function and the number of units in the hidden layer is 15 and the inputs are seasonal indicator variables, rainfall in Malang the previous day and rainfall in Karangkates the previous day. The result of this study is VAR (1) with indicator variables model which is better than the VAR - NN with indicator variables model based on RMSE, especially on testing data.

Keywords Rainfall, Vector Autoregressive, Feed Forward Neural Network

1. Introduction

The economic development of an area depends on many factors, in which water is one of the factors that plays a very important role. Water is a basic need for the survival of every living being. Rainfall is the main source of water availability and plays an important role in making crop planning strategies. Knowledge and prediction of precise rainfall is very important for crop planning in an area, especially those that use rainwater as a source of irrigation. Lack of water supply can have a tremendous negative impact on agricultural and industrial production and in turn have an impact on the country's economy. In other hand, heavy rainfall can cause flooding which has a bad impact. Floods caused by high rainfall are experienced by many regions in Indonesia, including in Malang. Precise rainfall prediction can prevent flooding with the necessary anticipatory steps.

Research on rainfall has been widely carried out by researchers from various regions and countries, such as Ali [1], Susanto and Ulama [8], Mahmud, et al. [3] and Pasaribu et al. [5]. Ali [1] modeled monthly rainfall in Baghdad using the Seasonal Autoregressive Integrated

Moving Average (SARIMA). Susanto and Ulama [8] compared the SARIMA, Feed Forward Neural Network (FFNN) and SARIMA-FFNN models to model monthly rainfall in Banyuwangi, Indonesia. The results of this study conclude that the FFNN model is the best model. Mahmud, et al. [3] modeled monthly rainfall in Bangladesh using SARIMA. Research by [5] modeled monthly rainfall in Merauke, Indonesia using the ARIMA model. The four studies that have been mentioned are univariate time series studies.

In addition to univariate time series models, multivariate time series models such as Vector Autoregressive are also used for rainfall modeling. Research on rainfall using Vector Autoregressive includes Nugroho et al. [4], Rosita et al. [6] and Rusman et al. [7]. Nugroho et al. [4] and [6] used rainfall, humidity and temperature variables, while [7] used rainfall variables in three locations, namely Bandung City, Cimahi City and West Bandung Regency.

In this study, the rainfall in Malang and Karangkates is modelled. The appropriate model for modeling two interrelated time series variables is VAR. According to Wutsqa et al. [9], Vector Autoregressive – Neural Network (VAR – NN) can improve prediction accuracy. Therefore, in this study the VAR – NN model was used to model daily rainfall in Malang and Karangkates. In this study, we also added an indicator variable that shows the season as exogenous variable in VAR model or as input in VAR-NN model.

2. Methodology

Vector Autoregressive (VAR) is a statistical model that can be used to analyze the relationship between several variables that influence each other. The Vector Autoregressive K variable model with order p according to Lutkepohl [2] can be written as:

$$\mathbf{y_t} = \mathbf{v} + \mathbf{\Phi_1} + \mathbf{y_{t\text{-}1}} + \cdots + \mathbf{\Phi_p} \mathbf{y_{t\text{-}p}} + \mathbf{u}_t \qquad (1)$$

where $\mathbf{y}_t = \left(y_{1t}, \ldots, y_{Kt}\right)'$, $\mathbf{v} = \left(v_1, \ldots, v_K\right)'$, $\mathbf{\Phi}_i$ is K by K coefficient matrix and $\mathbf{u}_t$ is white noise with a nonsingular covariance matrix $\mathbf{\Sigma}_u$. y_{it} is stationary time series. VAR modeling begins with a data stationarity test, order determination and parameter estimation. The data stationarity test used the Augmented Dickey Fuller (ADF) test, determination of order using Akaike Information Criteria Corrected (AICC). The selected order is the order with the smallest AICC value. The estimation of the parameters of the VAR model can be done using the Ordinary Least Squares (OLS) or Maximum Likelihood Estimation (MLE).

VAR-NN model used in this study is VAR – NN model with Feed Forward Neural Network type with one hidden layer. The VAR – NN architecture is built with an output layer containing variables, namely $\left[y_{1t}, y_{2t}, \ldots, y_{Kt}\right]$. The input layer contains p variable lag from the variables in the output layer so that the input layer consists $K \times p$ variables, namely $\left[y_{1t-1}, \ldots, y_{Kt-1}, \ldots, y_{1t-p}, \ldots, y_{Kt-p}\right]$. If there are h hidden units, then the weight matrix (network parameter) for the hidden layer has dimensions $(K \times p) \times h$ where

$$\mathbf{w} = \begin{bmatrix} w_{1,t-1,1} & w_{1,t-1,2} & \cdots & w_{1,t-1,h} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1,t-p,1} & w_{1,t-p,2} & \cdots & w_{1,t-p,h} \\ \vdots & \vdots & \ddots & \vdots \\ w_{K,t-1,1} & w_{K,t-1,2} & \cdots & w_{K,t-1,h} \\ \vdots & \vdots & \ddots & \vdots \\ w_{K,t-p,1} & w_{1,t-p,2} & \cdots & w_{1,t-p,h} \end{bmatrix}$$

A constant input unit is involved in the architecture, and is connected to each neuron in the hidden layer and output layer. This results in a bias vector $\boldsymbol{\alpha} = \left[\alpha_1, \alpha_2, \ldots, \alpha_h\right]'$ in *hidden layer* and $\boldsymbol{\beta} = \left[\beta_1, \beta_2, \ldots, \beta_h\right]'$ in *output layer*. Since there are K variables in the *output layer*, the weight matrix for the output layer becomes

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & \cdots & \lambda_{1,h} \\ \lambda_{2,1} & \lambda_{2,2} & \cdots & \lambda_{2,h} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{K,1} & \lambda_{K,2} & \cdots & \lambda_{K,h} \end{bmatrix}$$

The output of the VAR – NN model can be defined as

$$\mathbf{y}_t = \boldsymbol{\lambda} F\left(\left(\mathbf{yw}\right)' + \boldsymbol{\alpha}\right) + \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t$$

Where $F\left(\left(\mathbf{yw}\right)' + \boldsymbol{\alpha}\right) = \dfrac{1}{1 + \exp(-\left(\left(\mathbf{yw}\right)' + \boldsymbol{\alpha}\right)}$ for sigmoid activation function and $F\left(\left(\mathbf{yw}\right)' + \boldsymbol{\alpha}\right) = \dfrac{\exp\left(2\left(\left(\mathbf{yw}\right)' + \boldsymbol{\alpha}\right)\right) - 1}{\exp\left(2\left(\left(\mathbf{yw}\right)' + \boldsymbol{\alpha}\right)\right) + 1}$ for tanh activation function

The data used in this study is daily rainfall at the Malang Climatology Station and at the Karangkates Geophysics Station from January 1, 2019 to December 31, 2020. The data was obtained from the dataonline.bmkg.go.id website.

3. Result and Discussion

Data exploration is carried out using data plots. Figure 1 and Figure 2 are plots of rainfall variables in Malang and Karangkates respectively.

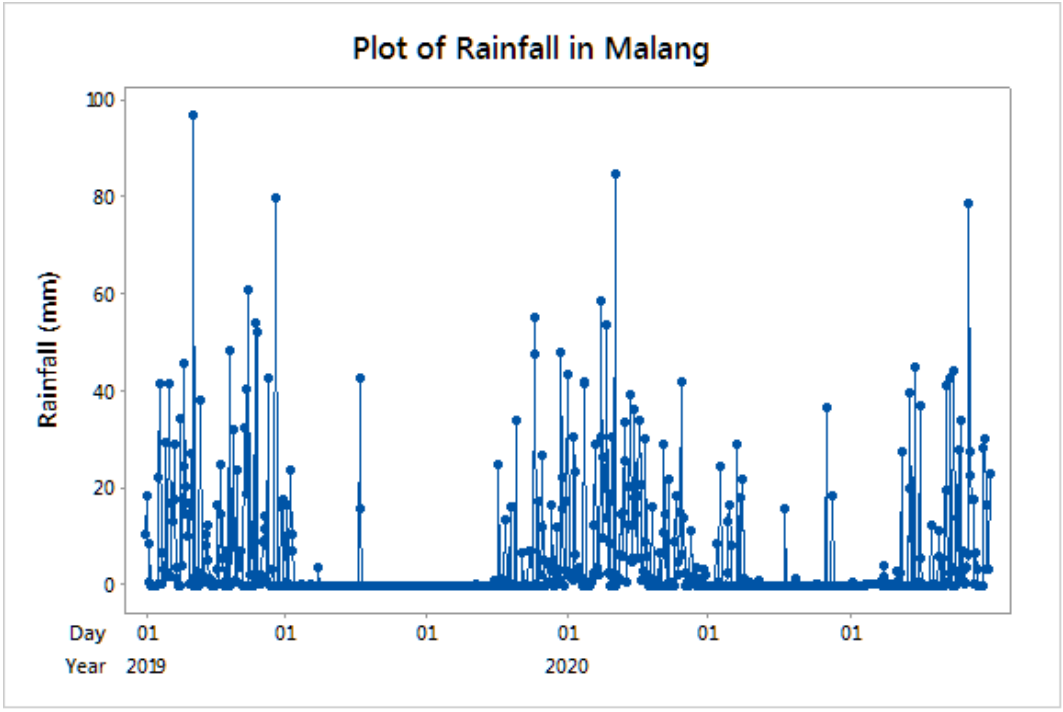

Figure 1. Plot of Rainfall in Malang

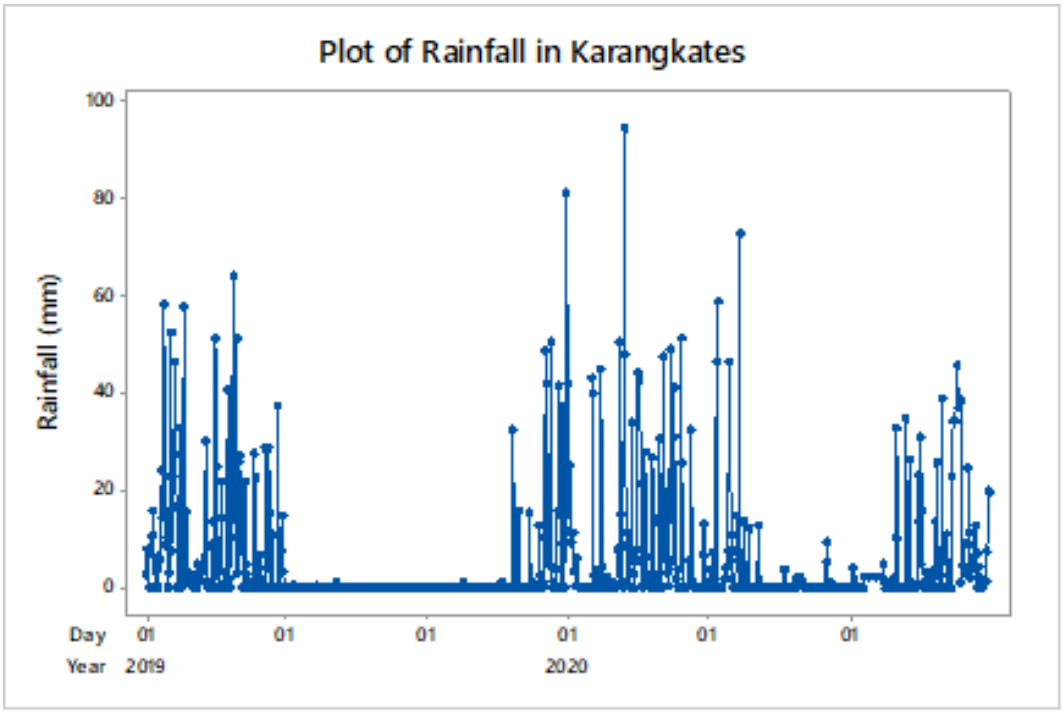

Figure 2. Plot of Rainfall in Karangkates

In general, the rainy season in Indonesia occurs from October to April. This is because the west monsoon blows in these months. However, Figure 1 shows that the rainfall in Malang started in November to mid May. Figure 2 also shows the same pattern for rainfall in Karangkates. However, it can also be seen in Figure 1 that July 2019 observed high rainfall, even though it was not received in the rainy season. Meanwhile in Karangkates this did not happen. Figure 1 also shows there is some very high rainfall. Figure 1 and Figure 2 also show quite extreme rainfall with rainfall of 80 mm – 100 mm in February and April.

The strategy in VAR-NN modeling is to do VAR modeling first. In this VAR modeling, it begins with stationarity test, order determination and parameter estimation. The order of this VAR determines the input of the VAR-NN. The stationarity test was carried out on the rainfall variable in Malang and in Karangkates using the ADF test. The ADF test gave the result that the p-value for both rainfall in Malang and Karangkates was 0.01. This shows that the rainfall variables in Malang and in Karangkates are stationary. We then employ normalization on the observed data. Before modeling, the data is divided into two, namely training data as much as 80% or as much as 585 and testing data as much as 20% or as much as 146. The training data is used to build the model and the testing data is used to assess the performance of the model. In this study, we used indicator variable that shows the season. The value of indicator variable is 1 if the day falls in the rainy season and is zero otherwise. In Indonesia, the rainy season generally occurs from October to April. However, from the plots in Figure 1 and Figure 2 it can be seen that there is a shift in the rainy season. The rainy season starts in November until around mid-May. Therefore, the value of the indicator variable is 1 for days from November to mid-May. After doing the stationarity test, the next step is order determination. In this process, Akaike Information Criterion Corrected (AICC) is used. The order chosen is the order with the smallest AICC. The order tested is from order 1 to 15 and the order with the smallest AICC is order 1. The obtained AICC is -8.480927. The complete AICC value can be seen in Table 1.

Table 1. AICC Value of VAR with Indicator Variable Model

Lag	AICC
0	-8.438282
1	-8.480927
2	-8.468818
3	-8.468072
4	-8.457424
5	-8.452159
6	-8.449138
7	-8.452823
8	-8.443054
9	-8.47801
10	-8.468038
11	-8.468503
12	-8.468038
13	-8.471513
14	-8.464857
15	-8.45248

Furthermore, the parameter estimation of the VARX model is carried out. The results of the estimation of these parameters are presented in Table 2.

Table 2. Parameter Estimation of VAR(1) with Indicator Variable Model

Equation	Variable	Parameter Estimates	p-value
y_{1t}	x_t	0.08962	0.0001
	y_{1t-1}	0.03056	0.4125
	y_{2t-1}	0.06708	0.0637
y_{2t}	x_t	0.07817	0.0001
	y_{1t-1}	-0.01196	0.7508
	y_{2t-1}	0.22850	0.0001

Note:

y_{1t} is the rainfall in Malang at time – t

y_{2t} is the rainfall in Karangkates at time – t

x_t is a seasonal indicator variable

Table 2 shows that the indicator variables are significant in forecasting rainfall, both in Malang and in Karangkates with p-value 0.0001. Meanwhile, the previous day's rainfall in both Malang (p-value=0.4125) and Karangkates (p-value=0.0637) was not significant in forecasting rainfall in Malang. The previous day's rainfall in Malang is also insignificant (p-value=0.7508) to the rainfall forecast in Karangkates, but the previous day's rainfall in Karangkates is significant (p-value=0.0001).

Next, based on the VAR modelling, NN modeling is carried out with the following conditions:

- The inputs used are seasonal indicator variables (x_t), rainfall in Malang the previous day (y_{1t-1}) and rainfall in Karangkates the previous day (y_{2t-1}).
- The output used is rainfall in Malang (y_{1t}) and rainfall in Karangkates (y_{2t}).
- The activation functions tested are Logistic Sigmoid and Tangent Hyperbolic.
- The number of neurons tested is 1, 2, 3, 4, 5, 10, 15.

The best model is chosen based on the smallest Root Mean Square Error (RMSE). Table 3 shows the RMSE of VAR-NN model.

Table 3 shows that the smallest RMSE for the VAR – NN model with the logistic sigmoid activation function is in the number of units in the hidden layer is 4 in the training data (RMSE=0.122767739) and 10 in the testing data (RMSE=0.107684216). The table also shows that the smallest RMSE for the VAR – NN model with the tangent hyperbolic activation function is found in the number of units in the hidden layer is 10 in the training data (RMSE=0.122781686) and 15 in the testing data (RMSE=0.107078258). Based on the tables, we can also see that the logistic sigmoid activation function produces a smaller RMSE than the tangent hyperbolic activation function in training but larger in testing data. Therefore, the VAR – NN model was chosen with the tangent hyperbolic activation function and the number of units in the hidden layer is 15. The architecture of the VAR – NN model is presented in Figure 3. The figure shows that there are an input layer, a hidden layer and an output layer. The input layer consists of seasonal indicator variables (x_t), rainfall in Malang the previous day (y_{1t-1}) and rainfall in Karangkates the previous day (y_{2t-1}). The hidden layer consists of 15 unit and the output layer consists of rainfall in Malang (y_{1t}) and rainfall in Karangkates (y_{2t}). The results of the estimation of model parameters, namely the weights from the input to the hidden layer are presented in Table 4 and the weights from the hidden layer to the output layer are presented in Table 5.

Table 3. RMSE of VAR – NN Model

Activation Function	The number of unit in hidden layer	RMSE of Training Data	RMSE of Testing Data
Logistic Sigmoid	1	0.123664094	0.108697141
	2	0.122830488	0.107747803
	3	0.12303246	0.108318404
	4	0.122767739	0.108160206
	5	0.122921068	0.107747803
	10	0.122941962	0.107684216
	15	0.122969815	0.107779582
Tangent Hyperbolic	1	0.123386851	0.109638259
	2	0.123025501	0.108413213
	3	0.123143752	0.108507938
	4	0.122907137	0.108571042
	5	0.123366033	0.107365717
	10	0.122781686	0.108128538
	15	0.12349089	0.107078258

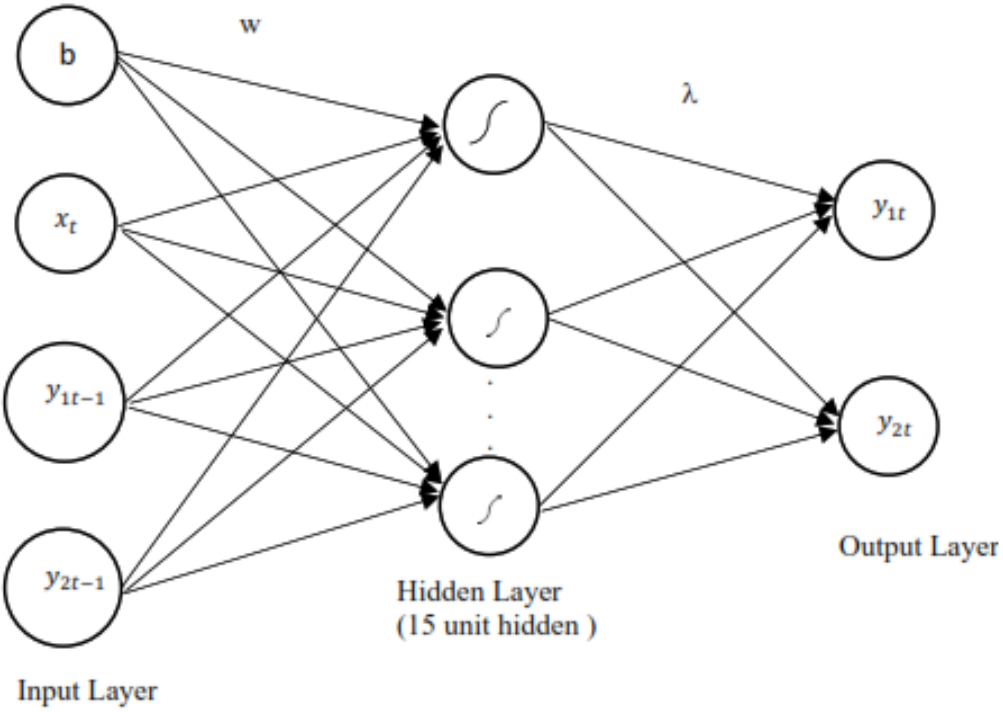

Figure 3. The Architecture of the VAR – NN model

Table 4. Weights from the Input to the Hidden Layer

Predictor (Input Layer)	*Predicted* (*Hidden Layer*)							
	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7	Unit 8
Bias	-.160	-.170	.383	-.134	-.283	-.454	.159	.436
x_t	-.179	.344	-.395	-.435	.084	.216	.360	.230
y_{1t-1}	-.375	.227	-.425	-.177	.009	.130	-.038	-.425
y_{2t-1}	-.380	.095	-.216	.030	.421	.292	.110	-.344

Table 4. Weights from the Input to the Hidden Layer (continued)

Predictor (Input Layer)	*Predicted* (*Hidden Layer*)						
	Unit 9	Unit 10	Unit 11	Unit 12	Unit 13	Unit 14	Unit 15
Bias	-.319	-.022	-.456	.169	-.065	.310	.326
x_t	.202	-.189	-.454	-.248	.336	-.060	.301
y_{1t-1}	.022	.338	-.195	.256	-.338	.029	.258
y_{2t-1}	-.218	.260	.486	.035	.237	-.057	-.169

Table 5. Weights from the Hidden Layer to the Output Layer

Predictor (Hidden Layer)	Predicted (Output Layer)	
	y_{1t}	y_{2t}
Bias	.411	.415
Unit 1	.139	.163
Unit 2	-.183	.023
Unit 3	-.276	-.398
Unit 4	-.160	-.114
Unit 5	-.412	-.118
Unit 6	.416	.007
Unit 7	-.324	.262
Unit 8	-.269	-.372
Unit 9	.093	-.380
Unit 10	-.362	-.249
Unit 11	.314	.355
Unit 12	.219	-.440
Unit 13	.295	.017
Unit 14	-.198	.149
Unit 15	.271	-.347

Based on Table 4 and Table 5, we can predict using equation

$$\hat{y}_{1t} = 0.411 + 0.139F(\theta_1) + ... + 0.271F(\theta_{15})$$

$$\hat{y}_{2t} = 0.415 + 0.163F(\theta_1) + ... - 0.347F(\theta_{15})$$

Where

F is tanh activation function,

$$\theta_1 = -0.160 - 0.179x_t - 0.375y_{1t-1} - 0.380y_{2t-1}$$

$$\vdots$$

$$\theta_{15} = 0.326 + 0.301x_t + 0.258y_{1t-1} - 0.169y_{2t-1}$$

To select proper model for rainfall modelling in Malang and Karangkates, we compare RMSE of VAR (1) with indicator variable model and VAR – NN with indicator variable model. The comparison is presented in Table 6.

Table 6. Comparison of RMSE

Model	RMSE of Training Data	RMSE of Testing Data
VAR (1) with Indicator Variable	0.128399	0.040627
VAR – NN with Indicator Variable	0.123491	0.107078

Table 6 shows that RMSE of VAR (1) with indicator variable is higher 0.004908 (=0.123491 – 0.128399) than RMSE of VAR – NN with indicator variable for training data but lower 0.066451 (=0.107078 – 0.040627) for testing data. So we conclude that VAR (1) with indicator variable model is better than VAR – NN with indicator variable model for modelling rainfall in Malang and Karangkates.

Neural Networks have been widely used in various areas, including time series data, both univariate and multivariate. Many studies have shown that neural networks have succeeded in making predictions very well. One of the weaknesses of the neural network is the difficulty in determining the input. In this study, the input is determined based on the VAR model. This study compares the VAR and neural network models with input based on the VAR (VAR - NN) model. The results of this study indicate that the VAR model is better than the VAR - NN model based on the resulting RMSE value. These results are not in line with the results obtained from [9].

4. Conclusions

This study aims to model rainfall in Malang and Karangkates simultaneously. The model used is the VAR with indicator variable model and the VAR - NN with indicator variable model. Based on the results of the study, it is concluded that the VAR (1) with indicator variables model is better than the VAR - NN with indicator variables model based on RMSE, especially on testing data.

Acknowledgments

A big thank you to the Faculty of Mathematics and Natural Sciences for the funds provided through the DPP/SPP program.

REFERENCES

[1] Ali, S. M. 2013. Time Series Analysis of Baghdad Rainfall Using ARIMA Method. *Iraqi Journal of Science* (54) Supplement No. 4 pp. 1136 – 1142.

[2] Lutkepohl, H. 2005. New Introduction to Multiple Time Series Analysis. Springer, Berlin.

[3] Mahmud, I, Bari, S. H, dan Rahman, M. T. 2017. Monthly Rainfall Forecast of Bangladesh using Autoregressive Integrated Moving Average Method. *Environmental Engineering Research.* (22) 2 pp. 162-168.

[4] Nugroho, A, Subanar, Hartati, S, Mustofa, K. 2014. Vector Autoregressive (VAR) Model for Rainfall Forecast and Isohyet Mapping in Semarang – Central Java – Indonesia. *International Journal of Advance Computer Science and Applications* (5) 11 pp. 44 – 49.

[5] Pasaribu, Y. P, Fitrianti, H, and Suryani, D. R. 2018. Rainfall Forecast of Merauke Using Autoregressive Integrated Moving Average Model. *E3S Web of Conferences* (73) 12010 pp. 1-5.

[6] Rosita, T., Zaekhan and Estuningsih, R. D. 2018. Vector Autoregressive (VAR) for Rainfall Prediction. *International Journal of Engineering and Management Research* (8) 2 pp. 96-102.

[7] Rusman, I. R, Ruchjana, B. N, and Sukono. 2019. Vector Autoregressive (VAR) Model for Rainfall Forecasting in West Java Indonesia at the Peak of the Rainy Season. *International of Recent Technology and Engineering* (8) 2S7 pp. 216 – 223.

[8] Susanto, Y and Ulama, B. S. S. 2016. Pemodelan Curah Hujan dengan Pendekatan Model ARIMA, Feed Forward Neural Network dan Hybrid (ARIMA – NN) di Banyuwangi. *Jurnal Sains dan Seni ITS* (5) 2 pp. 145 – 150.

[9] Wutsqa, D. U, Subanar, Guritno, S and Soejoeti, Z. 2006. Forecasting Performance of VAR– NN and VARMA Models. Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications. pp 194 – 200.