KnE Life Sciences

ISMoPHS 2020
The 2nd International Scientific Meeting on Public Health and Sports
(ISMoPHS 2020)
Volume 2021

Conference Paper

Prediction of Dengue Fever Cases in Malang City using a Neural Network Model

Nanta Sigit[1] and Ida Ayu P K[2]

[1]Management Health Information, STIKes Panti Waluya, Malang, Indonesia
[2]Pharmacy, STIKes Panti Waluya, Malang, Indonesia

ORCID:
Nanta Sigit: https://orcid.org/0000-0002-5538-4991

Corresponding Author:
Nanta Sigit
nantasigit1991@gmail.com

Published: 25 March 2021

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the ISMoPHS 2020 Conference Committee.

Abstract

Dengue fever has been declared endemic in many cities of Indonesia, one of them being the Malang City. In 2015, the incidence of dengue fever in the region was recorded at 1,629 with 13 deaths. There are many factors that contribute to the disease. The factors associated with dengue-fever transmission include population density, population mobility, quality of housing and attitude of life. However, the factors that can trigger dengue fever are environmental in nature, and include changes in temperature, humidity and rainfall, which cause mosquitoes to lay eggs more often and facilitates a rapid reproduction of the dengue virus. Parasites and disease carriers (mosquitoes) are very sensitive to climatic factors, especially temperature, rainfall, humidity, water levels and wind. Therefore, this study aimed to develop a suitable model for forecasting dengue fever in Malang City based on the Transfer Function and Artificial Neural Network (ANN). Data used were dengue fever data from 2004 to 2019. The results showed that the smallest RMSE, MAPE and SMAPE values of the two models were ANN models.

Keywords: Artificial Neural Network (ANN), Transfer Function, dengue fever

1. Introduction

Malang City is one of the cities declared as an endemic area of dengue fever. In 2015, the number of dengue fever sufferers was 1629 with 13 deaths. There are many factors that contribute to the disease, as well as dengue fever. These factors come from the individual himself or from the environment. Several factors associated with dengue fever transmission include population density, population mobility, quality of housing and attitude of life. While the factors that can trigger dengue fever are environmental factors, including changes in temperature, humidity, and rainfall, which cause mosquitoes to lay eggs more often and the dengue virus reproduces rapidly. Parasites and disease carriers (mosquitoes) are very sensitive to climatic factors, particularly temperature, rainfall, humidity, water levels and wind.

In order to anticipate the increase in many cases of dengue fever, the Minister of Health, the Directorate General of PPM & PLP, and the Head of the East Java Health Office have issued various regulations and policies. One of the regulations is the implementation of Mosquito Nest Eradication (PSN) through community empowerment known as 3M (Burying, Covering, and Draining) eradication. However, various efforts that have been made have not resulted in optimal and unsatisfactory results so that the handling of cases is still late.

Dengue data is periodic data, namely data that is presented within a certain time. Periodic data is closely related to forecasting or prediction. One of the efforts to handle dengue fever cases is to predict the rate of disease progression in the future. This prediction or forecasting can help optimize prevention efforts from an early age so that delays in handling no longer occur.

The forecasting process can be done using many methods, including smoothing, decomposition, regression, ARIMA Box Jenskins, and so on. In subsequent developments, a new method was developed using an Artificial Neural Network (ANN) or Neural Network (NN) system.

Neural Network (NN) is an information processing system that has similar characteristics to biological neural networks. The NN model is effectively used for forecasting the nonlinear model. The NN forecasting method continued to be developed from the 1940s to around 1988 the artificial neural network function was developed. One of the forecasting methods classified in the NN model is the artificial neural network (ANN) model. In this study, we will predict the many cases of dengue fever in the city of Malang using a comparison of the Transfer Function model and the Artificial Neural Network to predict dengue fever in Malang.

2. Material and Method

To predict dengue fever sufferers for a long period of time, from 2014 to 2019. Data is sourced from the Malang City Health Office, which is then analysed using Transfer Function modelling and Artificial Neural Network modelling. The forecasting process can be done using many methods, including smoothing, decomposition, regression, ARIMA Box Jenskins, and so on. In subsequent developments, a new method was developed using an Artificial Neural Network (ANN) or Neural Network (NN) system. Neural Network (NN) is an information processing system that has similar characteristics to biological

DOI 10.18502/kls.v0i0.8870

neural networks. The NN model is effectively used for forecasting the nonlinear model. The NN forecasting method continued to be developed from the 1940s to around 1988 the artificial neural network function was developed. One of the forecasting methods classified in the NN model is the artificial neural network (ANN) model.

3. Results

There are two indicators that will be used in this problem, namely the many cases of dengue fever in Malang city, the response variable (Y), the population density in the city of Malang with the population density unit / km^2 (X). The data will be divided into two, namely in-sample data and out-of-sample data. In sample data used for modelling is data from January 2014 to December 2018. Meanwhile, out of sample data is data from January to December 2019. The data will be modelled according to ARIMA and then modelled with the Transfer Function model. Here is the Time Series Plot between response variables and predictor variables:

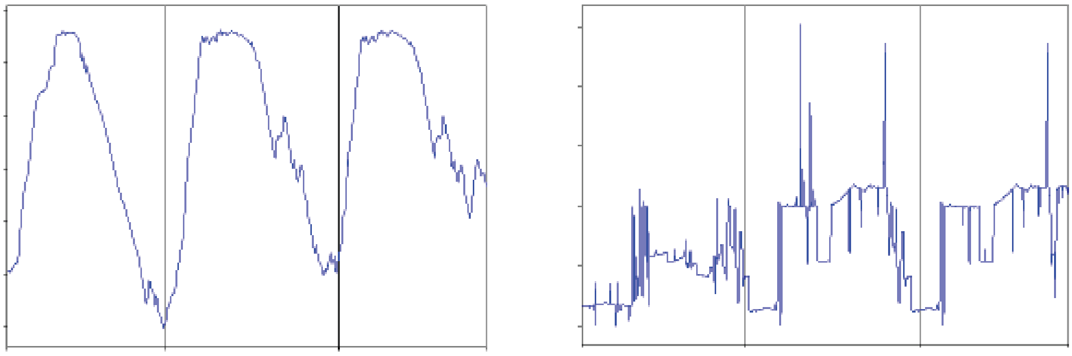

Figure 1: Population density in Malang city. (Source: Author’s own work)

Figure 1 shows that the series of population density variables in Malang city are not stationary well in the mean. Because the Time Series Plot of population density data in Malang City shows a seasonal pattern of 365 days (1 year), differencing is done at lag-365 but the data is still not stationary. Therefore, the differencing data at lag-365 are peered again at lag-1. In the Time Series, the population density data plot in Malang city does not show any seasonal patterns. So the differencing process is only carried out at lag-1.

The ACF plot on population density data in Malang city shows that there is a cut off in some lags. The lag can be used as a reference to get the ARIMA model order. So that we get the possibility of the ARIMA (3,0,0) $(0,1,0)^{365}$ and ARIMA (0,0,3) $(0,1,0)^{365}$ models. ACF and PACF plots show that the ACF and PACF cut off at lag 1, 2, 3, 4, 13, 15, 23, and 24. So that the estimated order of the ARIMA model (0,1, [1,2,3,4,13, 15,23, 24]) and ARIMA ([13, 15], 1, [1,2,3,4,23, 24]).

Parameters in the ARIMA (3,0,0) (0,1,0) and ARIMA (0,0,3) $(0,1,0)^{365}$ model are significant because they have a p-value that is smaller than a = 0,05. The residual diagnostic check is used to determine the feasibility of the ARIMA model that is formed. all parameters in the ARIMA model (0,1, [1,2,3,4,13,15,23, 24]) and ARIMA ([13, 15],1,[1,2,3,4,23, 24]) and the result is that all parameters are significant because they have a p-value smaller than a = 0.05. The residual diagnostic check is used to determine the feasibility of the ARIMA model that is formed. The input series used in transfer function modelling must meet the white noise assumptions.

The results of the residual white noise test for the ARIMA (3,0,0) $(0,1,0)^{365}$ and ARIMA (0,0,3)(0,1,0) models with a significance level of 0,05 have fulfilled the white noise assumption. because the p-value of the residuals on the tested lags has a value greater than 0.05. The results of residual white noise testing for ARIMA (0,1, [1,2,3,4,13,15,23, 24]) and ARIMA ([13, 15], 1, [1,2,3, 4], 23,24]) with a significance level of 0.05, it fulfils the white noise assumption because the p-value of the residuals on the tested lags has a value greater than 0.05. The model selected as the best model is seen based on the smallest AIC value. Table 2 is an in-sample AIC of each model.

The AIC value in Table 2 shows that for population density data in the city of Malang that has the smallest AIC value is the ARIMA model $(3.0.0)(0.1.0)^{365}$. Whereas for population density data in Malang city that has the smallest AIC value is the ARIMA model ([13, 15],1,[1,2,3,4,23, 24]). The next stage in transfer function modelling is pre-writing the input and output series.

3.1. Model Transfer Function

Transfer Function Model after obtaining the model in the ARIMA order, the pre-whitening process is carried out. The prewhitening process of the input and output series is carried out after obtaining the model in the ARIMA order. In the Multiinput Transfer Function model, the output series pre-whitening process is carried out for each input series.

Following is the prewhitening for the output series on population density in Malang city

$$\beta_t = Y_t - Y_{t-365} + 0,16843Y_{t-1} + 0,16843Y_{t-366} + 0,18664Y_{t-2} + 0,18664Y_{t-367} + 0,14640Y_{t-3} + 0,1464Y_{t-368}$$

TABLE 1: Estimation and significance test of ARIMA model parameters.

ARIMA Model	Parameter	Estimate	*P-value*
(3,0,0)(0,1,0)^{365}	$\varnothing_1$	0.16843	<0.0001
	$\varnothing_2$	0.18664	<0.0001
	$\varnothing_3$	0.1464	0.0001
(0,0,3)(0,1,0)^{365}	θ_1	−0.16921	<0.0001
	θ_2	−0.18749	<0.0001
	θ_3	−0.17841	<0.0001
(0.1,[1,2,3,4, 13,15,23,24])	θ_1	0.07687	0.0113
	θ_2	0.20928	<0.0001
	θ_3	0.18237	<0.0001
	θ_4	0.08637	0.0045
	$\varnothing_{13}$	−0.07043	0.0185
	$\varnothing_{15}$	−0.06133	0.0420
	θ_{23}	0.08849	0.0036
	θ_{24}	−0.15214	<0.0001
([13, 15],1,[1,2, 3,4,23,24])	θ_1	0.09013	0.0031
	θ_2	0.21458	<0.0001
	θ_3	0.19349	<0.0001
	θ_4	0.08544	0.0050
	$\varnothing_{13}$	0.08120	0.0070
	$\varnothing_{15}$	−0.15304	<0.0001
	θ_{23}	0.09156	0.0032
	θ_{24}	0.09880	0.0015

TABLE 2: Criteria for the best input series model.

Model ARIMA	AIC
(3,0,0)(0,1,0)^{365}	1.188.184
(0,0,3)(0,1,0)^{365}	2.357.657
(0,1,[1,2,3,4,13,15,23, 24])	3662.248
([13, 15],1,[1,2,3,4,23, 24])	3654.556

Following is the prewhitening for the output series for residents in Malang city

$$Z = W$$

Where:

$$Z = Y_t + 0.08120Y_{t-13} + 0.08120Y_{t-14} - 0,15304Y_{t-15} - 0,15304Y_{t-16}$$

$$W = \beta_t + 0.09013\beta_{t-1} + 0.21458 + 0,19349\beta_{t-3} + 0,08544\beta_{t-4} + 0,09156\beta_{t-23} - 0,09880\beta_{t-24}$$

The formation of the initial transfer function model is by obtaining a Cross-Correlation Function (CCF) between the input and output series which has undergone a pre-accounting process. CCF shows the strength of the relationship between the input series and the output series.

From the CCF plot, the order b, r, and s are obtained, for the models (0.0, [1, 15]) and (0.0 [1]). Estimation and Significance of Order Parameters (b, r, s) as follows.

TABLE 3: Estimation and significance test of parameters.

Models b, r and s	**Parameter**	**Estimate**	***P-value***
0,0,[1, 15]	ω_0	0.16843	<0.0001
	ω_1	0.18664	<0.0001
	ω_{15}	0.1464	00.0001
(0,0,[1])	ω_0	–0.16921	<0.0001
	ω_1	–0.17841	<0.0001

Order b, r, and s obtained b = 0; r = 0; s = [1, 15] (Table 3). The estimation of the first model is an indication of the CCF between the input and output series, it can be seen that the lag-0, 1, and 15 are not significant. This can be seen from the p-value of the parameter which is greater than a = 0.05 so that the parameter is not significant in the model. So the input series has no effect on the model. The orders b, r, and s obtained are b = 0; r = 0; s = [1, 15] (Table 4). The estimation of the first model is an indication of the CCF between the input and output series, it appears that the lag-0 and 1 are significant. This can be seen from the p-value parameter which is smaller than a = 0.05 so that the parameter is significant in the model. So the input series has an effect on the model.

To determine the feasibility of the transfer function model based on the initial prediction model, residual diagnostics was carried out. The identification of the noise series and the residual diagnostic check shows that the residuals from the transfer function have fulfilled the white noise assumption, it can be seen from the p-value which is greater than a = 0.05. Then the noise component can be modelled with the ARMA model in the dengue fever transfer function model.

ACF plot and PACF plot noise series model b = 0; r = 0; s = [1] indicates that cuts off occurs at lag2,3,5 and 7. This indicates that the predicted ARMA models are ARMA (0, [1,2,3,5, 8]) and ARMA ([1, 2,3,5, 8], 0).

Based on the detected model, it is necessary to estimate and the significance of the parameters of the ARMA noise series model to find the best model to explain the relationship between dengue fever data patterns. The estimation results and parameter significance are shown in Table 4.

Table 4: Estimation and test of the significance of parameters.

Model ARMA	Parameter	Estimate	*P-value*
(0,[2,3,5, 7])	θ_1	0.24290	<0.0001
	θ_2	0.29159	<0.0001
	θ_3	0.17084	<0.0001
	θ_5	0.10593	0.0007
	θ_8	–0.12962	<0.0001
	ω_0	4.89952	<0.0001
	ω_1	–2.47027	<0.0001
([2,3,5, 7],0)	$\varnothing_1$	–0.15241	<0.0001
	$\varnothing_2$	–0.25320	<0.0001
	$\varnothing_3$	–0.21864	<0.0001
	$\varnothing_5$	–0.20983	<0.0001
	$\varnothing_8$	0.01684	0.5693
	ω_0	4.65066	<0.0001
	ω_1	–2.14965	<0.0001

Table 4 shows that the predicted ARMA model for the noise series model b = 0; r = 0; s = [1] namely the ARMA (0, [1,2,3,5, 8]) and ARMA ([1,2,3,5, 8], 0) models. Both estimate models show significant parameter estimates. This can be seen from the p-value which is smaller than a = 0.05. Residual testing of the models that have been formed.

Table 5: White noise test results.

Model ARMA	Lag	*P-value*
(0, [1,2,3,5, 8])	6	0.1929
	12	0.2971
	18	0.3843
	24	0.4310
([1,2,3,5, 8],0)	6	<0.0001
	12	<0.0001
	18	<0.0001
	24	<0.0001

Table 5 shows that the results of the residual white noise test for the ARMA model (0, [1,2,3,5, 8]) with a significance level of 0.05 have fulfilled the white noise assumption because the p-value of the residuals on the lags tested has a value greater than = 0.05. Meanwhile, ARMA ([1,2,3,5, 8], 0) has not fulfilled the white noise assumption.

Table 6: Normality test results.

Model ARMA	D_{hitung}	*P-value*
(0,[1,2,3,5, 8])	0.248902	<0.0100
([1,2,3,5, 8],0)	0.260961	<0.0100

Graphical Summary For the residual data on dengue fever, it shows that there is one data that is the highest and out of bounds, with a kurtosis value of -0.177365 and a P-value <0.05, it can be said that the residual value in dengue fever cannot be normal, so it was decided to ignore assumptions, because in this forecasting case, the most important point is in the forecast results.

The next assumption that must be fulfilled is the residual cross-correlation (α_t) with the input series (α_t) fulfils the white noise assumption.

TABLE 7: Cross-correlation residuals with input series.

Model ARMA	**Lag**	***P-value***
(0, [1,2,3,5, 8])	5	0.6658
	11	0.1894
	17	0.1415
	23	0.1924
([1,2,3,5, 8],0)	5	0.2930
	11	0.1105
	17	0.1838
	23	0.2124

Table 7 explains that the p-value from the results of the cross-correlation residual (α_t) and the input series (α_t) has a p-value greater than a = 0.05, which means that the residual has white noise.

The model chosen as the best model is the ARMA model (0, [1,2,3,5, 8]) because it fulfils all white noise assumptions. So that mathematically the ARMA noise series model (0, [1,2,3,5, 8]) with the relationship between the input series and the output series occurs at lag-1 can be stated as follows

$$y_t = (3.63828x_t) + (-3.05979x_{t-1}) + a_t(1 - 0.24290B^1 - 0.29159B^2 - 0.17084B^3 - 0.10593B^5 + 0.1296B^8)$$

Where:

$y_t = Y_t - Y_{t-1}$

$x_t = X_t - X_{t-1}$

$x_{t-1} = X_{t-1} - X_{t-2}$

so that the final model obtained is

$$y_t = 3.63828X_t - 6.69807X_{t-1} + 3.05979X_{t-2} - a_t - 0.24290a_{t-1} - 0.29159a_{t-2} - 0.17084a_{t-3} - 0.10593a_{t-5} + 0.12962a_{t-8} + Y_{t-1}$$

The final transfer function model obtained explains that the pattern of dengue fever data today has a dependency with dengue fever yesterday and has a linear correlation with population density now, day 1 and 2 days before and forecast errors now day 1,2, 3, 5, and 8 days earlier in both regions.

3.2. Single Artificial Neural Network Model

Dengue data that will be modelled in a single ANN include variables X and Y. Previously, the data needs to be pre-processed so that the data is in the range 0 to 1. Then determine the input and iterations to obtain ANN parameters through the backpropagation optimization process.

The input determination is obtained from significant lags in the Transfer Function model. Based on Table 10, it can be seen that the lags selected as input from single ANN for electrical energy are lag 1 for dengue fever variables and lag 0.1 and 2 for population density variables.

Table 8: Criteria for selecting the Single ANN model with RMSE, MAPE and SMAPE.

Input(Yt)	**Input(Xt)**	**Model**	**RMSE**	**MAPE**	**SMAPE**
Lag 1	Lag 0, 1, dan 2	NN(1)**	56.86657	56.679	45.103
		NN(2)*	55.99797	56.491	45.168
		NN(3)	65.66727	62.354	56.012
		NN(4)	64.71017	60.769	57.175
		NN(5)	75.95881	70.038	70.391

*The model with the smallest error

**The model with the second smallest error

The ANN model chosen is NN (2) where RMSE, SMAPE and MAPE have the smallest value among all models.

Table 9: Bias and weighted values for NN (2).

Bias		**Pembobot**			
		β_{11}	0.27361		
		β_{12}	–4.07711		
a_{01}	2.11371	β_{21}	–1.11948	B_{01}	–0.23714
		β_{22}	0.76123		
a_{11}	–3.20892	β_{31}	–0.02409		
		β_{32}	0.5043	B_{02}	–0.23714
a_{12}	–1.21087	β_{41}	–0.27185		
		β_{42}	2.06035		

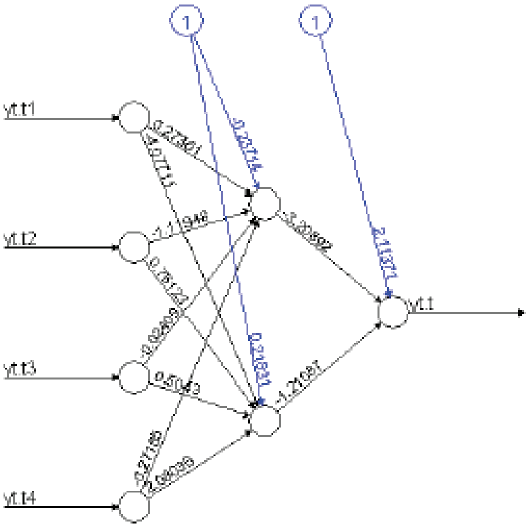

Figure 2: NN Architecture (2). (Source: Author's own work.)

Mathematically, the NN (2) model with the input of lag 1 for Y_t and lag 0, 1, 2 can be written as follows.

$$Z_t = a_0 + a_{11} f^{(h_1)} + a_{12} f^{(h_1)}$$

Where, $f^{(h_1)} = \frac{1}{1+e^{-(B_{01}+B_{11}y_{t,t1}+B_{21}y_{t,t2}+B_{31}y_{t,t3}+B_{41}y_{t,t4})}}$

$f^{(h_2)} = \frac{1}{1+e^{-(B_{02}+B_{12}y_{t,t1}+B_{22}y_{t,t2}+B_{32}y_{t,t3}+B_{42}y_{t,t4})}}$

$y_{t,t1} = Y_{t-1}$

$y_{t,t2} = X_t$

$y_{t,t3} = X_{t-1}$

$y_{t,t4} = X_{t-2}$

with weighted and bias values as in table 9.

4. Discussion

Comparison of forecasting can be seen from the time series plot of the forecast results between the transfer function method, single ANN and out sample data. The following is the time series plot of the two models against the out sample data.

Meanwhile, for selecting the best model from the two models, it can be done by looking at the ratio of the smallest RMSE, MAPE, and SMAPE values. The comparison of these errors can be seen in table 10.

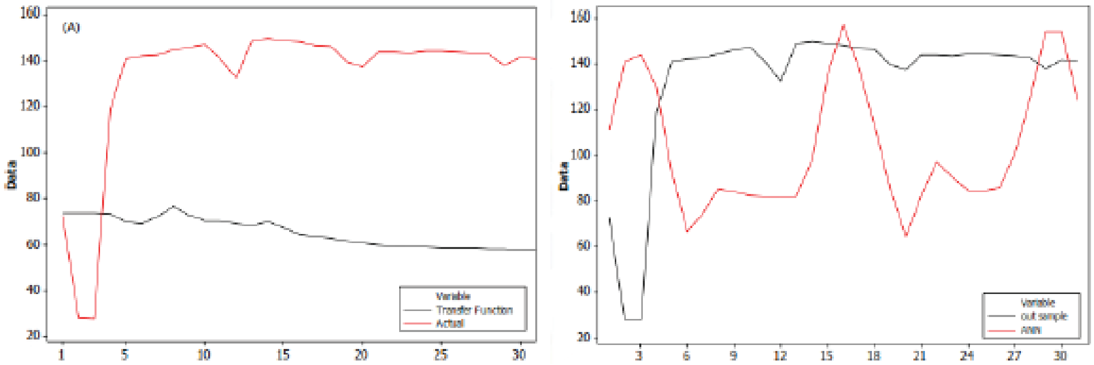

Figure 3: (A) Time series plot of Transfer Function model with data out sample. (B) Time series plot of Single ANN model without sample data. (**Source:** Author's own work.)

Table 10: The RMSE, MAPE and SMAPE values for the Transfer Function and Single ANN methods

Method	RMSE	MAPE	SMAPE
Transfer Function	75.53158	61.64293	74.20904
Single ANN	55.99797	56.491	45.168

Based on the RMSE, MAPE, and SMAPE values, it can be seen that the error value in NN is smaller than the error value in the ARIMA method. This suggests that the ANN method forecast is more appropriate for dengue fever forecasting data.

5. Conclusion

From the results and analysis, it was found that the single ANN method produced a smaller RMSE compared to other models in modelling dengue fever

Acknowledgement

The authors would like to express their gratitude to Hibah Eksternal PDP 2020 funded by Kementerian Riset dan Teknologi / Badan Riset dan Inovasi Nasional Republik Indonesia (Kemenristek/BRIN) for funding this work, Malang for the use of data.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Gao, Y., *et al*. (2020). Deep Learning and Transfer Learning Models of Energy Consumption Forecasting for a Building with Poor Information Data. *Energy & Buildings*, vol. 223, issue 1, pp. 110-6.

[2] Güven, İ. and Simsir, F. (2020). Demand Forecasting with Color Parameter in Retail Apparel Industry Using Artificial Neural Networks (ANN) and Support Vector Machines (SVM) Methods. *Computers & Industrial Engineering*, vol. 147, issue 1, pp. 1-10.

[3] Hardiana, V. M. D. (2013). Susilaningrum, and Suhartono, Peramalan Jumlah Tamu di Hotel "X" dengan Pendekatan Arima, Fungsi Transfer, dan Anfis. *Jurnal Sains Dan Seni Pomits*, vol. 2, issue 2, pp. 183-8.

[4] Jia, B., Dong, R. and Du, J. (2020). Ozone Concentrations Prediction in Lanzhou, China, using Chaotic Artificial Neural Network. *Chemometrics and Intelligent Laboratory Systems*, vol. 204, issue 1, pp. 1-8.

[5] Jiang, H., *et al*. (2020). Prediction of Output Power with Artificial Neural Network using Extended Datasets for Stirling Engines. *Applied Energy*, issue 271, pp. 115-23.

[6] Kashani, H. R. K., *et al*. (2020). Prediction Value of Preoperative Findings on Meningioma Grading Using Artificial Neural Network. *Clinical Neurology and Neurosurgery*, vol. 196, issue 20, pp. 1-7.

[7] Ma'rufah, N., Rahayu, S.P. and Suhartono. (2013). Peramalan Pendapatan Operasional Bank Menggunakan Metode Fungsi Transfer dan Neural Network. *Jurnal Sains Dan Seni Pomits*, vol. 2, issue 2, pp. 219-24.

[8] Nurina, D. L. and Irhamah. (2013). Peramalan Volume Pemakaian Air Sektor Rumah Tangga di Kabupaten Gresik dengan Menggunakan Fungsi Transfer. *Jurnal Sains Dan Seni Pomits*, vol. 2, issue 2, pp. 260-4.

[9] Nurvitasari, Y. and Irhamah. (2012*)*. Pendekatan Fungsi Transfer Sebagai Input Adaptive Neuro-Fuzzy Inference System (ANFIS) dalam Peramalan Kecepatan Angin Rata-Rata Harian di Sumenep. *Jurnal Sains Dan Seni ITS*, vol. 1, issue 1, pp. 62-8.

[10] Ye, R. and Dai, Q. (2020). Implementing Transfer Learning across Different Datasets for Time Series Forecasting Pattern Recognition 2021. *Pattern Recognition,* vol. 109, issue 1, pp. 1-6.