International Journal of Engineering & Technology, 7 (2.2) (2018) 65-69

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Prediction of electricity sales using neural based inverse distance weighting method

Bustani[1], Sunu Pradana[1], Mulyanto[2*], Nurjanana[3]

[1]*Department of Electrical Engineering, State Polytechnic of Samarinda, East Kalimantan, Indonesia*
[2]*Department of Information Technology, State Polytechnic of Samarinda, East Kalimantan, Indonesia*
[3]*Department of Economic Study, Mulawarman University, East Kalimantan, Indonesia*
**Corresponding author E-mail: yanto1294@gmail.com*

Abstract

Prediction of electricity sales becomes important for State Electricity Company of Indonesia (PLN) in order to estimate the Statement of Profit and Loss in next year. To obtain good predictive results may require many variables and data availability. There are many available methods that do not require so many variables to get predicted results with a good approximation. The aim of this study was to predict electricity sales by using an interpolation method called IDW (Inverse Distance Weighting). Several data samples are mapped into Cartesian coordinates. The data samples used are power connected to the household (X-axis), to industry (Y-axis), and electricity sales (Z value). Firstly, the sampled data clustered by using SOM algorithm. The Z value in each cluster is predicted by using the IDW method. The prediction results of IDW method are then optimized using ANN-BP (Artificial Neural Network Back Propagation). The trained net structure is then used to predict the electricity sale in next year.

Keywords: *connected power, electricity sales, SOM algorithm, IDW method, ANN-BP*

1. Introduction

Until 2015, electricity consumption in Indonesia is still dominated by household sector up to 40%, followed by industrial sector (38%), commercial sector (17%), and other sectors (5%). Electricity sales of 206 TWh in 2016 are predicted to increase in the next period.

In the year 2016, electricity energy sold in the island of Java is 71.81%, while the outside of the island of Java is 28.19%. The amount of electrical energy sold per capita in the island of Java is 1.057.5 kWh, while the outside of the island of Java is 543.6 kWh. The number of power plant units in the island of Java is 408 units, while the outside of the island of Java is 5,794 units. The amount of electrical energy capacity installed in the island of Java is 36,712.15 MW, while the outside of the island of Java is 17,420.13 MW. The amount of electrical energy produced in the island of Java is 1.29 GW, while the outside of the island of Java is 4.56 GW.

Profit (loss) of State Electricity Company of Indonesia in the year 2016 before the subsidy amounted to (31,627,846) rupiahs, while after subsidy amounted to 28,813,674 rupiahs. In the year 2015, profit (loss) of State Electricity Company of Indonesia amounted to 20,682,588 rupiahs. This means a decline in profit performance (loss) for State Electricity Company of Indonesia. This evidenced by the existence of government intervention to subsidize State Electricity Company of Indonesia in 2016.

To anticipate such conditions, one of them is how to be able to predict the electricity sales in the following period to estimate the income statement in next year. With reference to the estimated sales of electricity then the State Electricity Company of Indonesia can do everything necessary as a follow-up in its operation

Forecasting is the activity of predicting what will happen in the future with a relatively long time. To do so, it needs accurate data in the past, so it can be seen the prospects of future situations and conditions. Past data is usually modeled in the form of time series. Statistical methods have been widely used in many forecasting cases, where these methods are based on the assumption that time series data can be stated stationary. When a prediction is made it should be converted back to the original series.

There are many researches done that apply statistical methods for forecasting as in {Formatting Citation}. To improve the performance of forecasting results, many researchers have applied machine learning methods and their combinations.

SOM (Self Organizing Map) which is one type of ANN (Artificial Neural Network) that uses unsupervised learning method, has been combined with EML (Extreme Machine Learning) type SLFFN (Single Layer Feed-Forward Network) to accurately forecast wind energy generation [5]. ANN-based ANS (Artificial Neural System) has also been used to estimate the Optimum Duration of Road Projects in [6]. FFNN (Feed-Forward Neural Network) model was also used to predict compressive strength of the concrete after the microwave curing to ascertain the predictability of neural network models. The results indicate that the neural network models have a good scope for further study and implementations [7]. PNN (Probabilistic Neural Network) was used in Energy Management Scenario as a classifier to estimate the vulnerable point of the given network [8].

The combination of methods between statistics and machine learning is also used for various forecasting and prediction activities. Some researchers have done so with the aim of improving outcomes, but some of them also did comparisons between the two types of methods. ARIMA (Auto Regressive Integrated Moving Average), is one type of time series statistical data model, has been combined with EXP (exponential) model and ANN has been used to forecast Financial Time-Series in [9]. The K-Mean algo-

rithm is one of the well-known clustering algorithms, as well as SOM, has been combined with SLFFNN to predict heart disease in [10].
Time series data can be viewed as a spatial dataset if projected into N-dimensional coordinates. If unknown values are assumed to be in an unobserved location in N-dimensional coordinates, then the interpolation concept can be used to predict the values in the unobserved location. The interpolation concept is defined as a process for predicting a value at a point that is not a sample point, based on the sample point value around a non-sample point.
IDW (Inverse Distance Weighting) which is one of the deterministic interpolation methods, has been combined with Kriging, which is one of the stochastic interpolation methods, to predict spatial information about temperature and humidity at specific locations in [11]. IDW has also been used for interpolation of nitrate and chloride content to zone the tubular well density and groundwater quality in the urban area of the city of Lençóis, Bahia, through geo-processing techniques in the GIS platform [12]. In image processing area, an adaptive decision-based IDW interpolation algorithm has been used for the elimination of high-density salt and pepper noise in images [13].
In this study, all the data samples used will be projected in three-dimensional coordinates. The data samples used are power connected to the household (X-axis), to industry (Y-axis), and electricity sales (Z value) obtained from period 2002 – 2016 [14-20]. Firstly, the sampled data clustered by using SOM algorithm. The Z value in each cluster is predicted by using the IDW interpolation method. The prediction results of IDW method are then optimized using ANN-BP to obtain a smaller MAPE (Mean Absolute Percentage Error). The trained net structure is then used to predict the electricity sale in next year.

2. Materials and Methods

2.1. SOM (Self Organizing Map)

SOM is one type of ANN classified as unsupervised learning, in the form of intra-layer containing neurons that will arrange themselves based on the input of certain values in a group called the cluster. During the self-assembly process, clusters that have a weight vector best suited to the input (having the shortest distance) will be selected as winners. The adaptive process is performed through weight adjustment for the winning weights. The intra-layer connecting the input layer and the output layer is a $k \times n$ sizes matrix where k is the number of clusters desired and n is the number of training data. The SOM algorithm [21, 22] is shown in Figure 1.

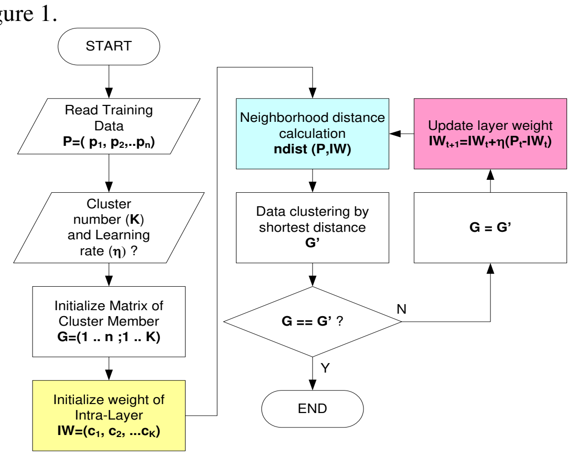


Fig. 1: SOM algorithm

2.2. IDW (Inverse Distance Weighting)

The IDW method is a deterministic interpolation method that estimates values at an unmeasured point by using a combination of linear values at a known sample point around the point that is not measurable. A value closer to an unmeasured location is considered as an estimated value. With this assumption, it can be concluded that all values closer to the observation point will have heavier weight [11]. The IDW method uses the Euclidean distance function with an interpolator expressed as:

$$Z_P = \sum_{i=1}^{N} w_i . Z_i$$

$$w_i = \frac{1/\left(\sqrt{\left(X_i - X_P\right)^2 + \left(Y_i - Y_P\right)^2}\right)}{\sum_{j=1}^{N} 1/\left(\sqrt{\left(X_i - X_P\right)^2 + \left(Y_i - Y_P\right)^2}\right)} \quad (1)$$

In Eq. (1), (X_i, Y_i) and (X_P, Y_P) are the location of the sampled point and interpolation point in Cartesian coordinate, respectively. Z_i and w_i are the measured values and their weight at sampled point (X_i, Y_i) respectively. Finally, Z_P and N are the value at interpolation point (X_P, Y_P) and the number of sampled point respectively.

2.3. ANN (Artificial Neural Network)

ANNs are composed of a number of simple elements that operate in parallel. Artificial neural networks are trained in such a way as to be able to produce outputs that approximate specific target outputs based on specific inputs. Back-Propagation (BP) is the generalization of the Widrow-Hoff learning rule while the term back-propagation refers to the way in which gradient errors are calculated for nonlinear multilayer networks [21, 23]. The ANN-BP is shown in Figure 2.

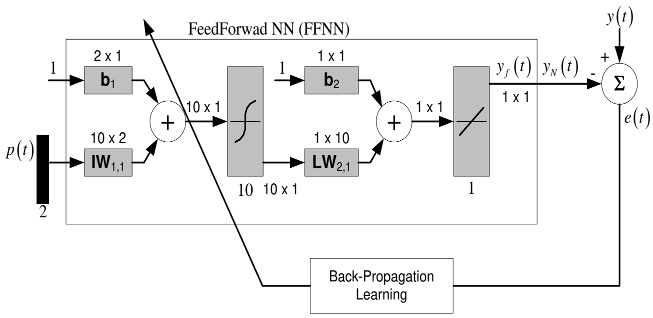


Fig. 2: ANN-BP

The created net structure uses 10 hidden neurons whereas target error $e(t) = 10^{-5}$. The performance function of ANN-BP training is using MSE (Mean Squared Error) declared:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(y\left(t_i\right) - y_{NN}\left(t_i\right) \right)^2 \quad (2)$$

In Eq. (2), N is the number of training data, $y(t_i)$ is the ith training target, $y_{NN}(t_i)$ is the ith output of ANN-BP. The implementation of ANN-BP is conducted using MATLAB programming tools.

2.4. The proposed method

The proposed method in this study is a combination of SOM, IDW, and ANN methods that are interrelated in several steps as follows:

1. Mapping the sampled data in *X-Y* coordinate.
2. Clustering the sampled data by using SOM algorithm with 3 clusters.
3. Implement the IDW method in each cluster to estimate Z value as the electricity sales prediction.
4. Improve the quality of prediction result by using NNBP.
5. Use the trained net structure to predict the next year's electricity sales.

Performance of prediction result is measured by using MAPE which is declared as:

$$APE(i) = \frac{\left|\text{actual}(i) - \text{prediction}(i)\right|}{\text{actual}(i)} \times 100\%$$
$$MAPE = \frac{1}{N}\sum_{i=1}^{N} APE(i) \tag{3}$$

The variance of the distance between data in a cluster and Z value predicted by using IDW will be used as training input and actual electricity sales as training targets. The variance of the distance is declared as:

$$Var = \left(\sum_{i=1}^{n-1}\left(x_i - \overline{x}\right)^2\right) / (n-1) \tag{4}$$

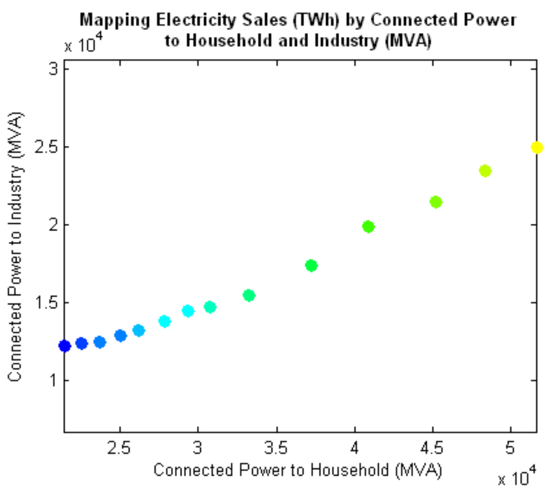
In Eq. (4), x_i is the distance between pairs of data, $\overline{x}$ is average of distance, n is the number of data. The distance between pairs of data is calculated by using Euclidean distance.

3. Results and discussions

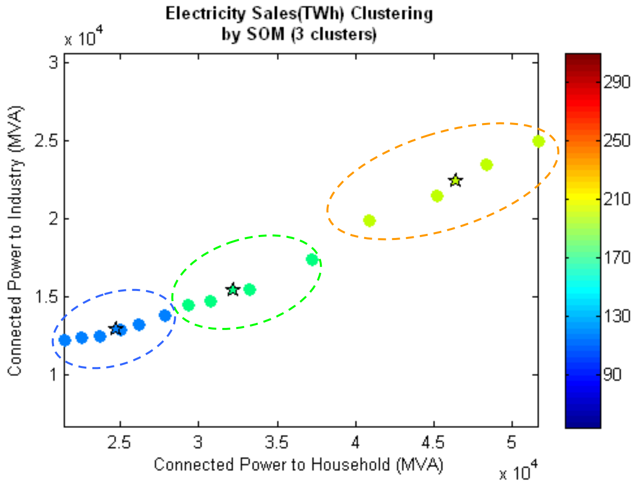
The data used in this study is shown in Table 1. The sampled data will be mapped are from the year 2002-2015, while the data in the year 2016 will be used as test data. The result of mapping and clustering of sampled data are shown in Figure 3. The result of electricity sales prediction by using IDW and optimized IDW are shown in Figure 4, Table 2 and 3.

Table 1: Raw data

Year	Connected Power (MVA)		Electricity Sales (GWh)
	Household (X)	Industry (Y)	Z
2002	21,342	12,284	87,089
2003	22,468	12,414	90,441
2004	23,658	12,533	100,097
2005	25,006	12,961	107,032
2006	26,101	13,292	112,610
2007	27,777	13,881	121,247
2008	29,334	14,531	129,019
2009	30,700	14,790	134,582
2010	33,203	15,566	147,297
2011	37,183	17,478	159,867
2012	40,869	19,981	173,991
2013	45,214	21,544	187,541
2014	48,374	23,542	198,602
2015	51,655	25,024	202,846
2016	55,285	26,570	216,004



(a)



(b)

Fig. 3: (a). Mapping electricity sales by connected power, (b). Clustering results

Table 2: The result of electricity sales prediction by using IDW

Year	Connected Power (MVA)		Electricity Sales	Cluster	Predict by IDW	APE
	House hold (X)	Industry (Y)	Z			
2002	1,342	12,284	87,089	1	99,634	14.40%
2003	22,468	12,414	90,441	1	99,463	9.98%
2004	23,658	12,533	100,097	1	100,489	0.39%
2005	25,006	12,961	107,032	1	104,354	2.50%
2006	26,101	13,292	112,610	1	105,621	6.21%
2007	27,777	13,881	121,247	1	103,981	14.24%
2008	29,334	14,531	129,019	2	140,271	8.72%
2009	30,700	14,790	134,582	2	138,164	2.66%
2010	33,203	15,566	147,297	2	139,639	5.20%
2011	37,183	17,478	159,867	2	139,121	12.98%
2012	40,869	19,981	173,991	3	193,758	11.36%
2013	45,214	21,544	187,541	3	190,948	1.82%
2014	48,374	23,542	198,602	3	191,477	3.59%
2015	51,655	25,024	202,846	3	191,423	5.63%
MAPE						**7.12%**

Table 3: The result of electricity sales prediction by using Optimized IDW

Year	Connected Power (MVA)		Electricity Sales	Cluster	Predict by Opt. IDW	APE
	House hold (X)	Industry (Y)	Z			
2002	1,342	12,284	87,089	1	87,300	0.21%
2003	22,468	12,414	90,441	1	92,959	2.53%
2004	23,658	12,533	100,097	1	95,672	4.40%
2005	25,006	12,961	107,032	1	107,187	0.15%
2006	26,101	13,292	112,610	1	112,797	0.18%
2007	27,777	13,881	121,247	1	120,726	0.50%
2008	29,334	14,531	129,019	2	130,228	0.86%
2009	30,700	14,790	134,582	2	133,705	0.63%
2010	33,203	15,566	147,297	2	146,960	0.24%
2011	37,183	17,478	159,867	2	159,607	0.19%
2012	40,869	19,981	173,991	3	174,971	0.51%
2013	45,214	21,544	187,541	3	188,277	0.39%
2014	48,374	23,542	198,602	3	193,126	2.86%
2015	51,655	25,024	202,846	3	202,061	0.41%
MAPE						**1.00%**

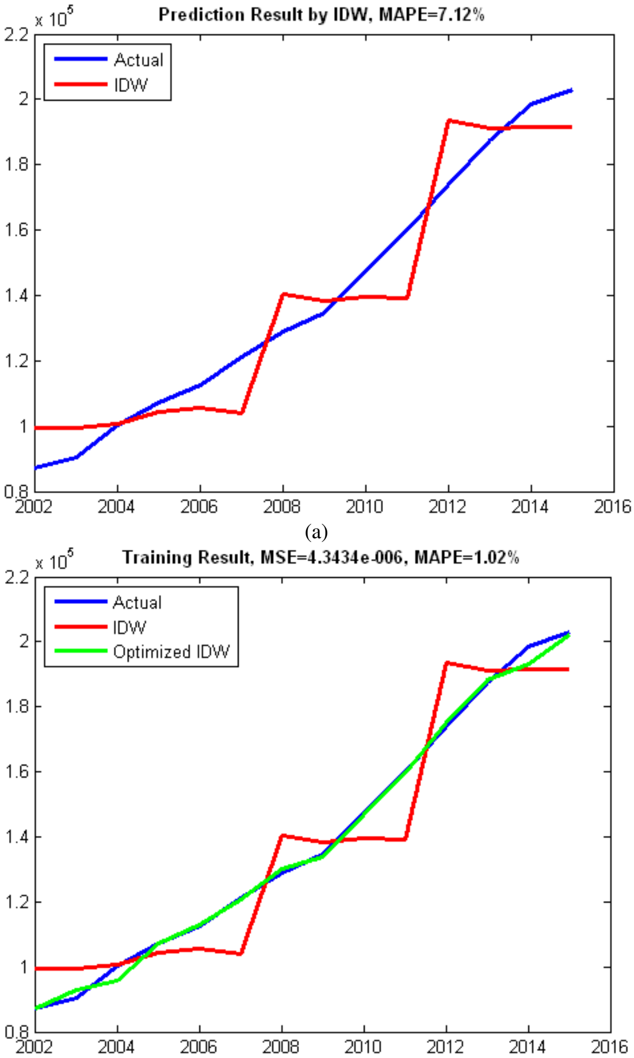


Fig. 4: (a).Prediction result by IDW, (b). Prediction result by Optimized IDW

From Figure 5, Table 2 and 3, there is an improvement of MAPE from 7.12% to 1.00% after being optimized by using ANN-BP. This indicates that the optimization is successful with performance improvement of:

$$\frac{|7.12-1.00|}{7.12}\times 100 = 85.90\%$$

Before predict the electricity sales in 2016 it is necessary to calculate the closest distance between the test data and the final centroid of the clustering results to find out where the test data is located in the clusters, as follows:

$$XY_{2016,dist} = dist\left(W_{som}, \begin{bmatrix} X \\ Y \end{bmatrix}_{2016}\right)$$

$$= dist\left(\begin{bmatrix} 24,780 & 12,992 \\ 32,946 & 15,718 \\ 46,986 & 22,744 \end{bmatrix}, \begin{bmatrix} 55,285 \\ 26,570 \end{bmatrix}\right) = \begin{bmatrix} 33,390 \\ 24,836 \\ 9,138 \end{bmatrix}$$

From the calculation results as above obtained that the test data is in cluster 3. Variance distance between test data with all cluster members 3, and electricity sales prediction by IDW for all clusters members 3 are used as input to predict by using the trained net structure. The Predicted result is $Z_{2016} = 209,042$. The performance of predicted results is calculated by using APE:

$$APE = \frac{|216,004-209,042|}{216,004}\times 100\% = 3.22\%$$

4. Conclusions

Optimization of IDW using ANN-BP resulted in a performance increase of 85.90% with a final MAPE of 1.00%. The result of electricity sale prediction in the year 2016 was obtained at 209,042 GWh with MAPE of 3.22%. From the results of this study can be stated that the proposed method is good enough to be applied in the case of electricity sales prediction with input in the form of data connected power to household and to industry. For long-term predictions, each predicted results in the next year is involved as a sample data, then the process is repeated from the beginning, and so on until it reaches the target year of prediction.
Future work is how to improve the predicted results by IDW that have much smaller prediction errors. Optimization of IDW prediction results using ANN-BP in this study still needs to minimize the number of its iterations.

Acknowledgement

The authors would like to express their heartfelt thanks to The Modern Computing Research Center, Department of Information Technology, State Polytechnic of Samarinda, for providing all their support.

References

[1] G. Anosh and E. P. Mishra, "Time series analysis model to forecast rainfall for Allahabad region," *Journal of Pharmacognosy and Phytochemistry,* vol. 6, pp. 1418-1421, (2017).

[2] J. Byun, Y. Han, I. P. Gorlov, J. A. Busam, M. F. Seldin, and C. I. Amos, "Ancestry inference using principal component analysis and spatial analysis: a distance-based analysis to account for population substructure," *BMC Genomics,* vol. 18, (2017).

[3] S. Saeed, L. Hussain1, I. A. Awan, and A. Idris, "Comparative Analysis of different Statistical Methods for Prediction of $PM_{2.5}$ and PM_{10} Concentrations in Advance for Several Hours," *IJCSNS International Journal of Computer Science and Network Security,* vol. 17, (2017).

[4] P. K. Sahu and R. Shrivastava, "Prediction of Tool Life based on Empirical Mode Decomposition and Gaussian Process Regression," *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE),* vol. 14, pp. 52-58, (2017).

[5] K. H. Tan, T. Logenthiran, and W. L. Woo, "Forecasting of wind energy generation using Self-Organizing Maps and Extreme Learning Machines," pp. 451-454, 2016.

[6] A. M. Al-saadi, S. Kh. Zamiem, L. A. A. Al-Jumaili, M. JameelJubair, and H. Abdalla Al- Hashemi, "Estimating the Optimum Duration of Road Projects Using Neural Network Model," *International Journal of Engineering and Technology,* vol. 9, pp. 3458-3469, (2017).

[7] T. R. Neelakantan, S. Ramasundaram, and R. Vinoth, "Prediction of Concrete Strength Using Microwave Based Accelerated Curing Parameters by Neural Network," *International Journal of Engineering and Technology (IJET),* vol. 5, pp. 157-164, (2013).

[8] G. Saha, K. Chakraborty, and P. Das, "Probabilistic Neural Network Based Voltage Stability Monitoring of Electrical Transmission Network in Energy Management Scenario," *International Journal of Engineering and Technology (IJET),* vol. 9, pp. 4434-4442, (2017).

[9] M. Khairalla, Xu-Ning, and N. T. AL-Jallad, "Hybrid Forecasting Scheme for Financial Time-Series Data using Neural Network and Statistical Methods," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 8, pp. 319-327, (2017).

[10] A. Malav, K. Kadam, and P. Kamat, "PREDICTION OF HEART DISEASE USING K-MEANS and ARTIFICIAL NEURAL NETWORK as HYBRID APPROACH to IMPROVE ACCURACY," *International Journal of Engineering and Technology (IJET),* vol. 9, pp. 3081-3085, (2017).

[11] E. Oktavia, Widyawan, and I. W. Mustika, "Inverse Distance Weighting and Kriging Spatial Interpolation for Data Center Thermal Monitoring," presented at the 2016 1st International Conference on Information Technology, Information Systems and

Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, (2016).
[12] J. Batista Mattos, Silva, Kaique Brito, J. M. Lima, and L. A. Oliveira, "Zoning of the use and quality of groundwater as a subsidy for the management of water resources: The case of the urban area of the municipality of Lençóis, Bahia, Northeastern Brazil," *Journal of Hyperspectral Remote Sensing,* vol. 7, pp. 40-49, (2017).
[13] V. Kishorebabu, G. Packyanathan, H. Kamatham, and V. Shankar, "An adaptive decision based interpolation scheme for the removal of high density salt and pepper noise in images," *EURASIP Journal on Image and Video Processing,* vol. 2017, (2017).
[14] PLN, "The Statistics of National Electrical Company of Indonesia Year 2010," S. E. C. o. Indonesia, Ed., ed. Jakarta: Center of State Electrical Company of Indonesia, (2010).
[15] PLN, "The Statistics of National Electrical Company of Indonesia Year 2011," S. E. C. o. Indonesia, Ed., ed: Center of State Electrical Company of Indonesia, (2011).
[16] PLN, "The Statistics of National Electrical Company of Indonesia Year 2012," S. E. C. o. Indonesia, Ed., ed: Center of State Electrical Company of Indonesia, (2012).
[17] PLN, "The Statistics of National Electrical Company of Indonesia Year 2013," S. E. C. o. Indonesia, Ed., ed: Center of State Electrical Company of Indonesia, (2013).
[18] PLN, "The Statistics of National Electrical Company of Indonesia Year 2014," S. E. C. o. Indonesia, Ed., ed: Center of State Electrical Company of Indonesia, (2014).
[19] PLN, "The Statistics of National Electrical Company of Indonesia Year 2015," S. E. C. o. Indonesia, Ed., ed: Center of State Electrical Company of Indonesia, (2015).
[20] PLN, "The Statistics of National Electrical Company of Indonesia Year 2016," S. E. C. o. Indonesia, Ed., ed: Center of State Electrical Company of Indonesia, (2016).
[21] M. H. Beale and M. T. Hagan. ((2015)). *Neural Network Toolbox™ MATLAB R2015a – User's Guide*.
[22] Haviluddin, A. Yunianta, A. H. Kridalaksana, Z. Arifin, B. Kresnapati, F. Rahman, A. F. O. Gaffar, H. Y. Irawan, M. Mulyo, and A. Pranolo, "Modelling of Network Traffic Usage Using Self-Organizing Maps Techniques," in *2016 2nd International Conference on Science in Information Technology (ICSITech)*, 2016, pp. 334-338.
[23] Haviluddin and R. Alfred, "A Genetic-Based Backpropagation Neural Network for Forecasting in Time-Series Data," in *The 2015 International Conference on Science in Information Technology (ICSITech 2015)*, Yogyakarta, Indonesia, 2015, pp. xxx-xxx.