Day-ahead Base, Intermediate, and Peak Load Forecasting using K-Means and Artificial Neural Networks

Lemuel Clark P. Velasco, Noel R. Estoperez, Renbert Jay R. Jayson,
Caezar Johnlery T. Sabijon, Verlyn C. Sayles
Mindanao State University-Iligan Institute of Technology
Iligan City, The Philippines

Abstract—**Industries depend heavily on the capacity and availability of electric power. A typical load curve has three parts, namely, base, intermediate, and peak load. Predicting the three (3) system loads accurately in a power system will help power utilities ensure the availability of the supply and to avoid the risk for over- or under- utilization of generation, transmission, and distribution facilities. The goal of this research is to create a suitable model for day-ahead base, intermediate and peak load forecasting of the electric load data provided by a power utility company. This paper presents an approach in predicting the three (3) system loads using K-means clustering and artificial neural networks (ANN). The power utility's load data was clustered using K-means to determine the daily base, intermediate and peak loads that were then fed into an ANN model that utilized Quick Propagation training algorithm and Gaussian activation function. It was found out that the implemented ANN model generated 2.2%, 1.84%, and 1.4% as the lowest MAPE for base, intermediate, and peak loads, respectively, with highest MAPE below the accepted standard error rate of 5%. The results of this study clearly suggest that with the proper method of data preparation, clustering, and model implementation, ANN can be a viable solution in forecasting the day-ahead base, intermediate, and peak load demand of a power utility.**

Keywords—*K-means clustering; artificial neural networks; base intermediate and peak load; day-ahead load forecasting*

I. INTRODUCTION

Electricity needs to be consumed the moment it is generated. Thus, electric companies should plan how much energy is needed to be bought from suppliers in order to meet consumers' demands [1]-[3]. Failure to provide industries and consumers will result to tremendous loss of resources while obtaining surplus amount of energy will result to underutilization of electric utility resources. If the power utility cannot provide the proper amount of electricity as soon as the consumer demanded it, the power quality would lead into service interruption [1], [3]. Hence, it is a pivotal point to obtain an accurate forecast of electric power systems in order to meet the changing power consumption consumers. A typical load curve has three parts namely base, intermediate, and peak loads. Base load is the minimum level of electricity required for a period of twenty four hours and provides power that keeps running constantly [4]. Intermediate and peak load are

the next to be brought in-line whenever the demand increases above the base electric load.

Predicting the three system load: base, intermediate, and peak load accurately in a power system will help power utility companies ensure the availability of the supply as well as avoid the risk for over or under utilization of generation, transmission, and distribution facilities [4], [5]. Having accurate estimates of base load demand in a power systems will also aid electric companies to possibly meet the demand of the market as this contributes to the continuous large amount of electricity in any system. Moreover, predicting the precise peak energy usage will further help the company in determining the need of purchasing the new and existing resources, as well as the type of resources that are necessary to meet the consumers demand. Furthermore, having knowledge of the peak energy usage will help the electric utility plan for the extension of existing facilities and installation of new power plants to reliably meet consumer demand [5]. K-means clustering technique and Artificial Neural Networks (ANN) model are suggested as suitable approach in clustering and forecasting load profile that can determine base, intermediate and peak loads [2], [3], [6], [7]. K-means is a widely used unsupervised learning algorithm that solves clustering problems or data categorizing. Moreover, it follows the simple and easy way in classifying datasets into assumed k clusters [6], [7]. While ANN is a mathematical model that has powerful classification and gained a significant performance in most researches mainly load forecasting due to its flexibility and its generality. Being one of the popular machine learning technique, supervised ANN has been widely used and has been proven to predict promising forecast of electric load but there are still gaps in coming up with a serial process in its usage along with unsupervised K-means [1], [2], [6].

In the Philippines, a power utility company faces a major problem in determining the base, intermediate, and peak load in their decision making since estimating these system loads is currently done by assigning assumed values resulting to inaccurate guess. This paper attempted to present a new technique for data preparation that is based on K-means clustering in order to determine the three system loads. Furthermore, an ANN model was also introduced and implemented to predict the day-ahead base, intermediate and peak loads. With proper data analysis along with appropriate

data clustering for data to be fed into a prediction model, this study aims develop a forecasting tool which power utilities can use to augment the gap of supply and demand in electric load.

II. METHODOLOGY

A. Load Data Preparation and K-means Clustering

This study used monthly electric load data for three years from 2012-2014 that has been used from an existing system of an electric power utility company. As shown in Table I, the worksheet of the data contain three sheets corresponding to the three metering points of the power utility company which contains the metering point name, the date, time, kilowatt delivered (KW_DEL), kilowatt per hour delivered (KWH_DEL) and kilovolt amps reactive hours delivered (KVARH_DEL).

TABLE I.   FORMAT OF THE RAW LOAD DATA

SEIL	BDATE	TIME	KW_DEL	KHW_DEL	KVARH_DEL
XXX	XXX	XXX	XXX	XXX	XXX
XXX	XXX	XXX	XXX	XXX	XXX

Electric load data was grouped in terms of its metering points by creating three databases while eliminating columns that was not be used in the study. Before importing the data into the database, corrections were made through filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies in order to avoid erroneous data that could make the model uncertain [8]. The three metering points were then aggregated to come up with the total load duration of the entire locality represented by the three metering points.

Cluster centers by day of the electric load data was then determined using K-means clustering method. K-means clustering is a partitional clustering method that attempts to find the number of clusters (k), which are represented by centroids [6], [9]. Cluster parameters such as number of clusters (k) and maximum number of iterations (i) were then determined. Choosing the number of k clusters has no agreed upon solution, thus finding the number of k clusters varies on the number of clusters desired. Three basic steps in K-means clustering were then conducted [6], [7], [9]. Firstly, data items were partitioned into I initial cluster, where data consists of data points from x_1, x_2..,x_n and each point is assigned to the nearest centroid. Thereafter, it identified the group of data points and assigned each point to one group or one cluster. Secondly, assignment was done to an item to which the mean or the centroid was the nearest. Similarity of clusters was computed using mean value of the objects which were also considered as cluster centers or centroids. Lastly, an iterative process of assigning the centroids until no more reassignments of cluster centers was done. This process will stop iterating when there will be no point to navigate each cluster to another and when the centroid will remain the same [7], [10]. A graphical representation of the resulting cluster centers was generated with the load curve and load duration curve in a day in order to present the percentages of the base, intermediate, and peak load.

The clustered data was then partitioned into two data sets: training set and testing set. The training set was used for training and adjusting the weights on the neural network while testing set was used for testing the design of the network to confirm the actual predictive power. If the input data of the predictive model is not normalized, the training of the network would be slow [11]. There are various normalization methods that would produce values either from a range of 0 to 1 or -1 to 1. However, in this study only Min-max normalization method was used as suggested by authors dealing with load prediction using ANN [8], [12]. After normalizing the training set, the normalized load data, that contains the inputs and their corresponding expected outputs, will then be fed into the neural network for training. On the other hand, the normalized dataset will be inputted into the model, trained and tested, the resulting outputs of the neural network will undergo denormalization as shown in (1) to show the actual value where $x = x_1, x_2,...x_n$, y is the denormalized data and z_i is the normalized data.

$$y = z_i \, (max(x) - min(x)) + min(x) \qquad (1)$$

B. ANN Model Implementation

An ANN model with multilayer perceptron as an architecture having input layer, hidden layer, and output layer was implemented in a load prediction system. Fig. 1 shows that the ANN model has eight input neurons consisting of the day of the week, holidays or non-holidays, weekends or weekdays, week number, and month. The ANN used four hidden neurons along with Quick Propagation training algorithm and Gaussian activation function. The signals of the ANN were multiplied with bias weights and mapped into three output nodes indicating the day-ahead base, intermediate, and peak load consumption. The ANN model was implemented through desktop-based software with the use of Encog library in order to achieve the training and testing results. Encog library is a Java-based library which provides interchangeable models with efficient, internal implementations and supports machine learning models with choice of training algorithms [12]. System features and use cases that will carry out the processes needed for the clustering of the actual data up to the load prediction, starting from data loading, clustering, and ending in generating of the predicted values. The functions were grouped into classes such as the data scaling, clustering, database querying and the ANN model class for the training, testing, and forecasting functions. Other features such as export of clustered datasets and charts were also included for the purpose of easier comparison of the data.

A validation set from January 2015, outside the training and testing sets was used to test the model's accuracy. Clustering of the validation set was also conducted in order to determine their actual base, intermediate and peak loads while actual load of every past day was appended to the model to predict its day-ahead load. An important criterion in evaluating the prediction accuracy of the forecasting model is to compute the measure of error. The accuracy of the prediction model can often be defined as forecasting error, which is the difference between the actual load and the predicted load [1], [3], [5], [11]. In order to evaluate the performance of the neural

network model quantitatively, error measures was calculated in this study. Denormalization of the ANN output was conducted and used Mean Absolute Percentage Error (MAPE) which measures the error in terms of percentage and calculated as the average percentage error to compare the denormalized data to the actual data [2], [11], [12]. After denormalizing the data, the forecasted values were evaluated by comparing it to the clustered data to check if the forecasted value is close enough to the clustered data. A graphical representation of the computations was then generated for the purpose of illustrating the comparison between the clustered actual and predicted base, intermediate and peak load values.
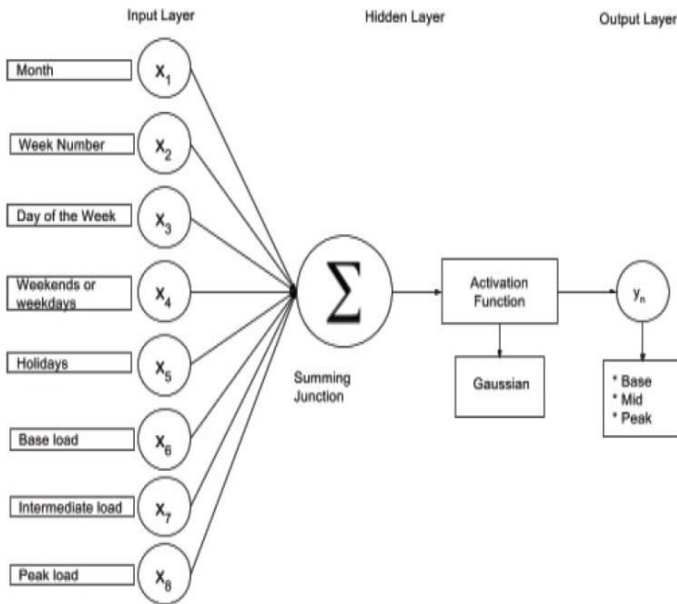

Fig. 1.   Block diagram of the implemented ANN Model.

III.   RESULTS AND DISCUSSION

A. Load Data Preparation and K-means Clustering Results

Based on the raw load dataset, the data used as input for clustering were the date, time and the kilowatt delivered (KW_DEL) which is the consumed load. The same columns were also used by researches in predicting electric load which disregarded SEIL, KWH_DEL and KVARH_DEL [2], [12]. Certain parameters were set in clustering the raw electric load data such as the number of clusters (k) and the maximum number of iterations (i). The number of clusters was set into three which represents base, intermediate, and peak load. Using K-means clustering technique, the intervals associated with the base, intermediate, and peak load were calculated. In the first step, the number of clusters was assumed to be three and initial cluster centers were determined.

Table II shows the first i=1 iteration that calculated the three cluster centers. These cluster centers has the sample values of 17,402.2929 KW for base, 20,023.5783 KW for intermediate, and 25,302.8772 KW for peak. These cluster centers will be reassigned until the groupings of data points will stay the same as exemplified by groups of authors [6], [7], [10].

TABLE II.   SAMPLE CLUSTER CENTERS ITERATION

i	Base	Intermediate	Peak
1	17, 402.2929	20,023.5783	25,302.8772
2	16,820.9705	20,538.0019	26,137.5521
3	16,820.9705	20,645.4180	26,202.7391
4	16,820.9705	20,645.4180	26,202.7391

Fig. 2 depicts the k cluster grouping points from the 96 observations of a day's 15-minute load intervals. Blue represents the base clusters, orange represents the intermediate clusters, and red represents the peak cluster points while the star sign represents the initial cluster center (k) of the data points. In the second step, K-means converges at 4th iteration, i = 4 depicting the iteration of the resulting cluster centers. K-means converges faster whenever the initial cluster centroids are selected [7], [13]. It shows that k clusters converge at iteration 4 and after 4th iteration there is no reassignment of cluster centers. The final cluster centers were determined after there was no more changes occurring in clustering the datasets and when final cluster centers have found the natural grouping of points in each cluster centers with the same values. Final cluster centers has the value of 16,820.9705 KW for base, 20,645.4180 KW for intermediate, and 26,202.7391 KW for peak, this means that these values will no longer be reassigned.
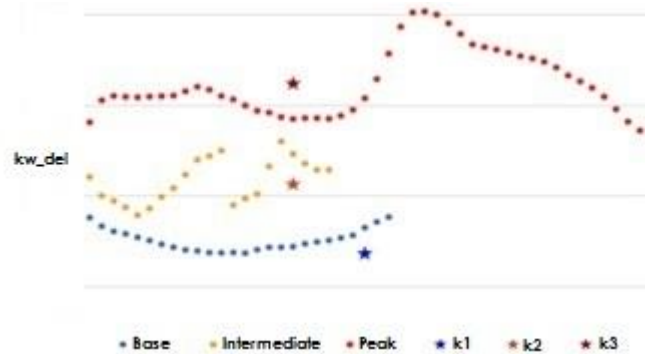

Fig. 2.   Sample k clusters.

Load duration curve is the arrangements of all load levels into descending order of magnitude. It can be beneficial for electric power system in economic dispatching, system planning, and reliability evaluation. Moreover, it helps to determine if there is a need to come up with replacement decisions due to an over or under loading condition [14]. With the three clustered data produced, it was then mapped out together with the load duration curve to effectively show how much base, intermediate, and peak load demand is required in the electric power system. Fig. 3 describes the sample load duration curve. It is show that for a particular day, the electric power system requires 59.01% for base electric load. This means that 59.01% will be used for the continuous supply of large amount of electricity. Coal-fired plants and nuclear units

are appropriate for the base load station [4], [14]. On the other hand, intermediate electric load requires 27.83% of the time period. Combine cycle units are used in the intermediate power plants. Moreover, the system requires 13.16% peak electric load. Diesel, hydro and gas turbines belongs in the peak load station categories [15].
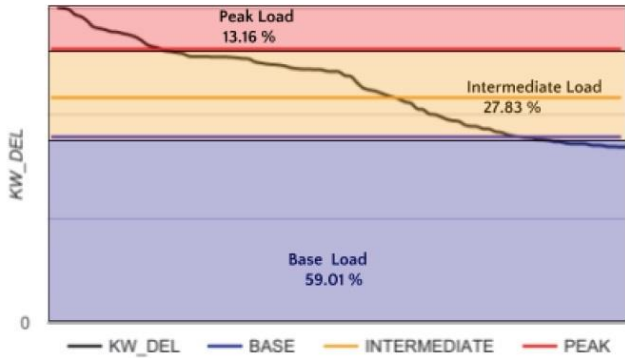

Fig. 3.   Sample load duration curve.

Out of the 101,760 observations, 3,180 emerged after clustering. The original dataset having the three metering points has 96 observations each day representing the 15-minute interval load data each day. After the clustering process, the aggregated consumption of the locality was grouped in 1,060 days of three data points representing the base, intermediate, and peak clusters. The clustered load data was then divided into training set and testing where 70% of the data given was partitioned into training data comprising the clustered load data from 2012 to 2013 and the remaining 30% of the clustered 2014 load data was set for testing data . As shown in Table III, the clustered data was normalized using min-max normalization that produces an output by the range of 0 to 1 since no negative values was observed in the clustered data. Min-Max normalization performs a linear transformation on the original data and preserves the relationship among data values [11]. This normalization technique was used in order for the data to be fed into the neural network for training dataset and to test the accuracy of the prediction.

TABLE III.   SAMPLE NORMALIZED DATA RESULTS

Date	Base	Intermediate	Peak
XXX	0.240666598	0.23465889	0.238462481
XXX	0.239323519	0.232076194	0.245447171

B. ANN Model Implementation Results

Using clustered data of a validation set, MAPE was calculated for a week's daily prediction in order to evaluate the performance of the model. As shown in Fig. 4, highest MAPE were 3.90%, 4.80%, and 4.68%, respectively for base, intermediate and peak loads. Lowest MAPE was 2.2%, 1.84%, and 1.4% respectively for base, intermediate and peak loads with peak loads having majority of the lowest MAPE should the three be compared. The MAPE of the clustered and the

predicted peak load has always been the lowest among the three except on January 5 where it also has its highest MAPE. The clustered intermediate and peaks loads seem to show a trend different from that of base load. It can also be observed that except on January 6 and 7, all of the three loads were all together increasing or decreasing in MAPE values until base load has broken the uniformity on the last two days. Base load is generally steady as validated in the figure showing intermediate load adjusting with the performance of the base and the peak load [5], [14]. According to studies, acceptable error range for load prediction is between 3%-3.5% while the corporate tolerance error of the power utility being tested is below 5% [2], [11], [12], [15]. We can then infer that the performance of the clustering and the prediction is acceptable.

The forecasted values were denormalized and compared to the clustered values of the validation set. In order to assess the forecasted result, visualization was made in each category to compare the forecasted results and the clustered data of the base, intermediate, and peak load. As shown in Fig. 5, the forecasted values were compared to clustered data of the validation set's base load which are generally bought by power utility companies on bilateral and long-term contracts [4], [14]. It can be observed that on January 2, 2015, the highest MAPE for the base load of 3.90% shows the only instance that predicted base load is below the actual clustered base load.

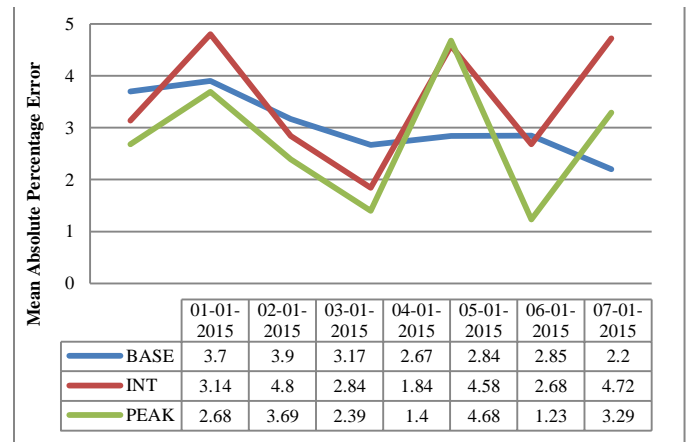|	01-01-2015	02-01-2015	03-01-2015	04-01-2015	05-01-2015	06-01-2015	07-01-2015
BASE	3.7	3.9	3.17	2.67	2.84	2.85	2.2
INT	3.14	4.8	2.84	1.84	4.58	2.68	4.72
PEAK	2.68	3.69	2.39	1.4	4.68	1.23	3.29

Fig. 4.   Performance of the implemented forecasting model.

Fig. 5.   Comparison between clustered base load vs. predicted base load.
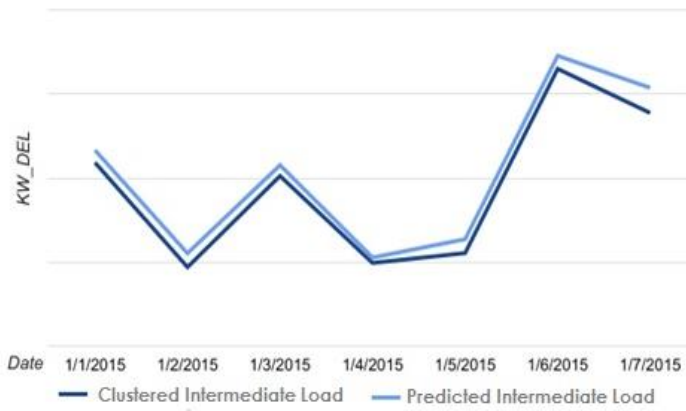
Fig. 6. Comparison between clustered intermediate load vs. predicted intermediate load.

Fig. 6 shows the predicted intermediate load being compared to clustered intermediate load which are usually bought by power utility companies on spot markets because of affordability. As shown, the forecasted intermediate values for the intermediate load were close enough to the actual intermediate values. Since MAPE values can either be above or below the predicted and the actual, it shows here that predicted load is always above the clustered intermediate load.

Fig. 7 shows the comparison between the forecasted and the clustered peak load. It can be observed that on January 1, 2, and 3 predicted peak load is higher than the clustered peak load while in the January 4, 5, and 6 shows the clustered peak load higher than the predicted peak load. January 5, 2015 has the highest MAPE of 4.68% followed by the lowest MAPE of 1.23%. Although peak load has the same trend with that of the intermediate load, this does not always mean uniformity in occurrence on that the predicted is also higher than the actual since MAPE are absolute zero values. What is notable is that regardless whether whichever is higher or lower than the other, the trend of the intermediate load adjusts with that of the peak load being evident on the MAPE [5], [14], [16].
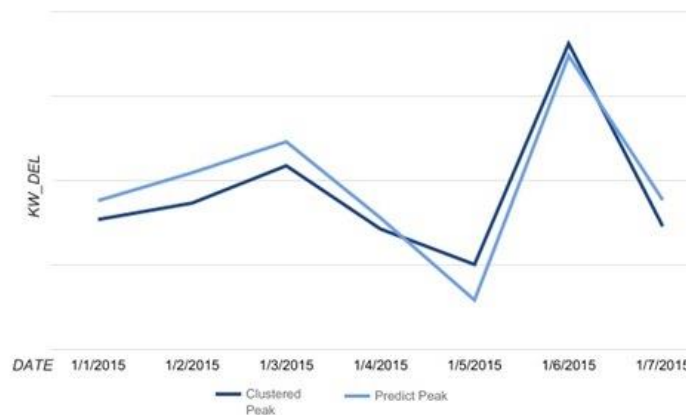

Fig. 7. Comparison between clustered peak load vs. predicted peak load.

IV. CONCLUSION AND RECOMMENDATIONS

This study attempted to develop a forecasting tool that will predict the day-ahead base, intermediate and peak loads by clustering the historical delivered daily electric loads using K-means and employing ANN to forecast the day-ahead clustered loads. The goal was achieved by data preparation, K-Means clustering, ANN implementation, and comparison of the forecasted results with the derived clustered base, intermediate, and peak loads. After successfully determining the daily base, intermediate and peak loads using K-Means, a multilayer perceptron neural network with eight input neurons, four hidden neurons along with Quick Propagation training algorithm and Gaussian activation function was implemented in Encog. Highest MAPE of the forecasting tool were 3.90%, 4.80%, and 4.68% while lowest MAPE were 2.2%, 1.84%, and 1.4% respectively for base, intermediate and peak loads. Techniques for the clustering the daily loads other than K-means clustering technique could be investigated to determine the daily base, intermediate and peak loads.

For future work, it is recommended that performance analysis on different training algorithms and activation functions be conducted to develop a more optimized model. Aside from month, day of the week, weekend/weekdays, holiday/non-holidays indicators that this study used as input neurons, other additional factors that can affect the training process of the neural network such as temperature and weather variables can be considered by future researches as essential factors of the prediction model to generate better results of the forecasting model. This study aims to help electric system decision makers by discussing concepts of developing a close to accurate forecasting technique in predicting the base, intermediate, and peak loads in order for power utilities to come up with better short, medium and long term decisions.

REFERENCES

[1] K.Y. Lee, Y.T. Cha and J.H. Park, "Short-term load forecasting using an artificial neural network", IEEE Transactions on Power Systems, Vol. 7, Issue 1, February 1992.

[2] L.C. Velasco, C. Villezas, P.N. Palahang, and J.A. Dagaang, "Next day electric load forecasting using Artificial Neural Networks", International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management, IEEE DOI: 10.1109/HNICEM.2015.7393166, December 2015.

[3] H. Quan, D. Srinivasan, and A. Khosravi, "Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals", IEEE Transactions on Neural Networks and Learning Systems, Vol. 25, Issue 2, pp: 303 – 315, February 2014.

[4] M. Ndlovu and T. Lekalakala, "Renewable energy options and base load stations for South African community", 8th Conference on the Industrial and Commercial Use of Energy, IEEE, October 2011.

[5] L. Jin, L. Ziyang, S. Jingbo, S. Xinying, "An efficient method for peak load forecasting", The 7th International Power Engineering Conference, IEEE, Print ISBN: 981-05-5702-7 , DOI: 10.1109/IPEC.2005.206877, December 2005.

[6] S. Nuchprayoon, "Electricity load classification using K-means clustering algorithm", 5th Brunei International Conference on Engineering and Technology, IEEE, Print ISBN: 978-1-84919-991-9, DOI: 10.1049/cp.2014.1061, June 2015.

[7] A, Alkilany, A. Ahmed, H. Said, and A. Abu Bakar, "Application of the k-means clustering algorithm to predict load shedding of the Southern Electrical Grid of Libya", DOI: 10.1109/INTECH.2014.6927763, October 2014.

[8] L.C. Velasco, A.R. Granados,J. M. Ortega and K.V. Pagtalunan, "Medium-Term Water Consumption Forecasting using Artificial Neural Networks" 17th Conference of the Science Council of Asia, National Research Council of the Philippines, June 2017.

[9] S.A. Azad, A. B. M. Shawkat Ali, and P. Wolfs, "Identification of typical load profiles using K-means clustering algorithm", Asia-Pacific World Congress on Computer Science and Engineering, IEEE, DOI: 10.1109/APWCCSE.2014.7053855, November 2014.

[10] M. Ghofrani, M. de Rezende, R. Azimi, and M. Ghayekhloo, "K-means clustering with a new initialization approach for wind power forecasting", Transmission and Distribution Conference and Exposition, IEEE, 10.1109/TDC.2016.7519931, July 2016.

[11] L.C. Velasco, P.N. Palahang, C. Villezas, and J. A. A. Dagaang, "Performance Analysis of Different Combination of Training Algorithms and Activation Functions in Predicting the Next Day Electric Load", 16th Philippine Computing Science Congress, Computing Society of the Philippines, March 2016.

[12] L.C. Velasco, P. Bokingkito, and J. T. Vistal, "Week-Ahead Load Forecasting using Multilayer Perceptron Neural Network for a Power Utility", 17th Conference of the Science Council of Asia, June 2017.

[13] W. L. Zhao, C.H. Deng, and C.W. Ngo, "Boost K-Means", Computer Vision and Pattern Recognition, Cornell University Library, December 2016.

[14] A. Salimi-beni, D. Farrokhzad, M. Fotuhi-Firuzabad, S.J. Alemohammad, "A New Approach to Determine Base, Intermediate and Peak-Demand in an Electric Power System", International Conference on Power System Technology, IEEE, DOI: 10.1109/ICPST.2006.321928, October 2006.

[15] A. Singh and V.K. Tripathi, "Load Forecasting Using Multilayer Perceptron Neural Network", International Journal of Engineering Science and Computing, DOI 10.4010/2016.1135, ISSN 2321 3361, pp. 4548 - 4551, May 2016.

[16] F. Ueckerdt and R.Kempener, "From Base load to peak: renewables provide a reliable solution", International Renewable Energy Agency, 2015.