European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 08, Issue 02, 2021

Minimising The Estimation Error Of Forecasting The Electricity Consumption In Malaysia

Norazliani MD LAZAM[1], Nur Izzati SHARIL[2], Suraya MOHD[3], Norsyafika Azwa MOHD SHARIFF[4], Nur Farah Haifa MD KAMAL[5]

[1]*Centre for Actuarial Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Malaysia*
[2]*CIMB Bank Berhad*
[3] *Blink Trading Sdn Bhd*

e-mail: norazliani@tmsk.uitm.edu.my

Abstract: This paper presents a study on minimising the estimation error of forecasting the electricity consumption in Malaysia. A robust and accurate forecasts of electricity consumption are deemed crucial for the supplier to arrive on fair estimations of electricity supply optimally. Thus, identifying the best model to forecast the electricity consumption accurately may hinder energy wastage. This research aims to examine which model gives the least error in estimating the future electricity consumptions in Malaysia. Two models were tested namely Artificial Neural Network (ANN) and Regression Methods. In analysing these models, this research applies the Microsoft Excel and SAS Enterprise Miner (SAS) software. The data were extracted from the Department of Statistics Malaysia (DOSM), CEIC Data Company and The Statistics Portal. Results indicate that ANN produces least error as compared to the Regression Method as the former fits the data well whilst the latter overfits the data. The ANN model uses NNTool from MATLAB is used for forecasting future electricity consumption. The forecasted values (2020-2022) proved to provide more interpretable forecasts. This study may benefit the electricity supplier, consumers and also the Government of Malaysia, in particular the Ministry of Energy and Natural Resources. It may provide insights on estimating the optimum amount of energy to be generated. This will definitely increase the savings and reduce wastage from every angle. Ultimately, the environment is saved too.

Keywords: artificial neural network; electricity consumption; forecasting; regression.

1. INTRODUCTION

Electricity consumption is deemed crucial nowadays for survivals. It has now become an essential daily needs, and one cannot think of a life without electricity. Almost all devices, tools and machineries used at homes, businesses and industries are functioning due to the power from the electricity. The demands for electricity in Malaysia are foreseen to increase between five to six per cent yearly as the urbanisation and industrial development increase rapidly in our country. In Malaysia, there is an efficacious reserve new capacity being generated consistently, with an increase of 4,780 megawatt (MW) on average [1]. Undoubtedly, electricity helps us in every way to make our daily activities easier by supplying power for home applicants, industrial manufacturing machineries, along with its capability to

supply power for medical machineries in treating patients as well as developing new medical technology that gives good impact to humanity.

Nonetheless, there are also negative impacts arise from massive electricity consumptions. The process of generating new electricity may harm the environment. Moreover, the greenhouse effect can arise from the process of generating the electricity. In addition, the production of the electricity may also contribute greatly to all kind of pollutions such as air pollution, water pollution, global warming, destroy animal population, and create more health problem towards human beings. Thus, a proper planning of generating the electricity is crucial in order to reduce wastage and at the same time, saving our environment. To address this issue, a robust and plausible model that can forecast an accurate and useful electricity consumption is essential. This may help to arrive on fair estimation of electricity consumption optimally [2].

In this study, two models namely Artificial Neural Network (ANN) and Regression were used in forecasting the electricity consumption. The model with the least error i.e ANN is chosen to forecast the electricity consumption amounts for the next 3 (three) years. Therefore, this paper fulfills the aim of achieving greater efficiency in estimating an optimum amount of energy production and consequently achieve further substantial of cost savings.

Literature Review

The methods being considered in forecasting the electricity consumption are Multiple Linear Regression and Artificial Neural Network (ANN). For Multiple Linear Regression, the analysis is conducted by using the Microsoft Excel. Meanwhile, a software tool for applying the Artificial Neural Network (ANN) method is using the SAS Enterprise Miner.

Multiple Linear Regression

Regression provides a convenient solution where prediction can be estimated by using the equation. Regression will also provide the R squared value with perception of assessing how good the model is. The R squared value ranges from 0 to 1, which 1 indicates a perfect model.

A study conducted by [3], found that the variations in a household electricity consumption can be explained by using a multiple regression model. They further commented that multiple regression model best working with extensive household surveys and large datasets of electricity consumption. Their study was to find the simplified correlations of the factors that affect the changes in electricity consumption. The factors that they studied were in building characteristics such as socio-economic, personal and socio-demographic characteristics of households. Despite being useful in predicting the future values, there are some drawbacks found in this method. Normally, due to the complexity of the data, this model tends to overfit the data. Hence, some vital components of this model like p-values, R-squared value and regression coefficients will be misleading. According to [4], the model should at least have 10 to 15 datasets for each term of the estimation. In rectifying the overfitting problem, some approaches like cross-validation, shrinkage and resampling may be carried out. Cross validation can be used to detect any overfitting sampling by generalising the data. On the other hand, shrinkage and resampling can help in determining how can the model strategically fits the data.

Artificial Neural Network (ANN)

Artificial Neural Network (ANN) was first invented in 1985 by a psychologist named Frank Rosenblatt. Initially, its first purpose was to model how human brain processed a visual data and learned to recognise objects [5]. ANN has Multi-Layer Perceptron that basically

composed of three layers of neuron (nodes) namely input layer, hidden layer and output layer. This multi-layer features in ANN enable the model to process more complex input patters. [6] used Deep Neural Network (DNN) in creating a reliable DNN architecture that better fit the complex input patterns. Another study conducted by [2], used ANN algorithms to pick nonlinear patterns in a dataset. They concluded that the normalisation process makes it easy to compare weights of the linkages among neurons in a given block of the ANN model. A study conducted by [7], found a better way that enables consumers to better understand their load curve of energy consumptions and leverage efficiently on the new energy resources with consideration on the high energy prices. ANN was used to predict the short-term prediction of total power consumption in buildings.

ANN has been widely used in many studies for forecasting and estimation. ANN helps in overcoming the restrictions of the traditional methods. [8] analysed several research that used ANN as their methods. In the study, it proves that ANN works best in predicting values. Another study by [9] to predict a short term electricity load, ANN method produced excellent results for forecasting electricity loads in buildings. The application of ANN has also been extended in other areas of research. For instance, ANN has been widely used in research relates to agricultural activities. [10] combined the capabilities of ANN with the human like fuzzy logic system to establish a system that can early detect the paddy diseases. Another study by [11] used the ANN to classify the types of soil that suitable for agricultural activities. In finance, the ANN model is also used to develop a model for pricing the Islamic Options instruments [12].

Electricity Consumption in Malaysia

In Malaysia, electricity is one of the primary industries which regulated by the Department of Electricity and Gas Supply. Other parties that have vested interests into this business include development and research institutions, company of energy supply and services, and also consumers. The Government of Malaysia controls the energy policy of Malaysia, while PETRONAS and Tenaga Nasional Berhad (TNB) play as major bodies that stand as government-linked companies, dominating this business. In Peninsular Malaysia, electricity is regulated by Tenaga Nasional Berhad, while in East Malaysia, Sarawak Energy in charge of Sarawak and Sabah Electricity in charge of both Sabah and Labuan respectively.

The capacity of electricity power that was installed in Peninsular Malaysia in 2014 was recorded at 8.636 megawatt. Ever since 1955, the generation of the electricity has been increasing tremendously until now. In 1955, the electricity generated was recorded at 948 million kWh. It went up to 1070 million kWh in two years and dropped slightly once in 1958, then kept rising each year until 2015, the last data in our record. By 2015, total consumption of electricity in Malaysia reached about 141147 million kWh. This increasing trend shows a positive landmark of our country's growth. This evolution proves our country is being at par with other developing and developed countries in terms of its development.

Impact on Inefficient Electricity Consumption

Major consumer of electricity comes from industrial sector and its share has been increased over times due to industrialisation, where the consumption of electricity is largely consumed. Theoretically, high energy consumption leads to high economic growth. It is obvious that electricity acts as an essential element for commercial, industrial and residential growth which directly leads to an economic growth of a developing country. As much as electricity helps in providing a better life and facilitating many processes in promoting country's development, inefficient of electricity consumption will jeopardise the country's

economic growth. Wasteful consumption of electricity may give negative impact on the development of a country.

Generating the electricity power requires huge energy and resources. Current resources of electricity power generation in Malaysia are coming from three major fossil fuel sources which are coal, natural gas and fuel-oil. This production has resulted to a greenhouse effect as the process of generating electricity, releases the carbon dioxide and other greenhouse gases at the same time. A lavish and over-produced of this may provide negative impact towards the mankind and the environment. The excessive produce of the unhealthy gases may create other health problems like cancer and genetic mutation. This may also destroy the animal populations. To the environment, the presence of unwanted chemical or other elements in air, may affect the quality of the ecosystem [13]. This may cause to global warming and also pollutions of air and water. Therefore, it is important for the electricity supplier to generate an optimum amount of electricity, neither over-produce nor under-produce, as this will give negative impacts to the country, society and environment. An efficient and effective planning in generating the electricity is deemed crucial as to ensure that our country continues to expanding its economy and at the same time preserving the environment and its mankind. Eventually, Malaysia will enjoy benefits from the increases of economic status, employment rates, standards of living, development of technology and many more. In achieving this, it is crucial to have a plausible projections of the electricity consumption. This will help the energy supplier and the government to efficiently estimate the productions of the electricity. This will help to reduce wastage and provide an optimum consumption of the electricity, besides preserving our environment for our future generations.

2. METHODOLOGY

Data Collection

Data on the electricity consumptions were obtained on a monthly basis from the Department of Statistics Malaysia (DOSM), CEIC Data Company and The Statistics Portal, for the period of 9 years, from January 2010 until December 2018. Two models were used in fitting the electricity consumption namely Artificial Neural Network (ANN) and Regression Methods. In fitting both models, Microsoft Excel and SAS Enterprise Miner (SAS) software were used. Whichever model that produces the least error and closeness fitting with the actual data will then be chosen for forecasting. The forecast rates from the robust model is expected to be accurate and reliable. Hence, using that model, this study estimates the electricity consumptions of Malaysia for the next three (3) years, from 2019 until 2020.

Multiple Linear Regression

Multiple linear regression is used in this study since this research used many independent variables in predicting the outcomes. In forecasting the future electricity consumption, many factors are believed contributing to the excessive usage of the electricity. [5] found that factors like urbanisation, population, employment, foreign direct investment and gross domestic product influence the usage of the energy consumption. The more rapid the economy grows, the energy consumption will increase substantially. The equation of the multiple linear regression for the factors that may contribute to the excessive usage of electricity can be described as:

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 + b_6 X_6 + b_7 X_7 + b_8 X_8 + \varepsilon \quad (1)$$

Where:

$Y = Future\ Electricity\ Consumption\ (kWh)$
$X_1 = Urbanisation\ Rate\ (\%)$
$X_2 = Population\ (number)$
$X_3 = Employment\ (number)$
$X_4 = Unemployment\ (number)$
$X_5 = Gross\ Domestic\ Product - GDP\ (\%)$
$X_6 = Foreign\ Direct\ Investment - FDI(USD)$
$X_7 = Current\ Price\ of\ Electricity\ Consumption\ (MYR)$
$X_8 = Total\ Electricity\ Consumption\ (kWh)$
$a = Intercept\ of\ Regression\ Line$
$b = Slope\ of\ Regression$
$\varepsilon = Error\ Term$

In estimating the Future Electricity Consumption, the input variables are Urbanisation Rate, Population, Employment, Unemployment, GDP, FDI and Total Electricity Consumption whist the Current Price of Electricity Consumption is the control variable. Two types of Multiple Linear Regression were used in this study namely Linear Regression and Stepwise Linear Regression. Consequently, the better type of Regressions will be tested further in comparison with the ANN method.

Artificial Neural Network (ANN)

Artificial Neural Network (ANN) was used in this study as it simplifies the process of analysing each parameter. Inspired by the nature of neurons in a brain, ANN connects the group of nodes and identify each node weights and interconnections. As shown in this study (Figure 1), ANN model with hidden nodes of 1, 2, 3 and 4 are being analysed in finding the best model that estimates the electricity consumption. The best model will be then used to further compare with the regression method.

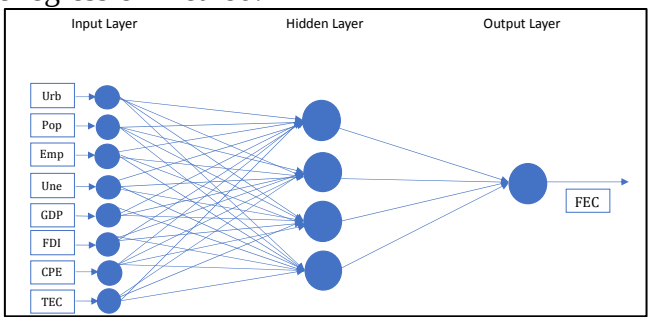

Figure 1 Artificial Neural Network (ANN) Model

NNTool by MATLAB

Upon recognising which model between regression and ANN that gives the less error in fitting the electricity consumption data, NNTool will be used to forecast the electricity consumption, based on the best model. NNTool is a software under MATLAB that builds and trains variables. A series of build and train processes were conducted to arrive at the most stable state of model. Then, the model will be used to forecast the values for future electricity consumption for Malaysia.

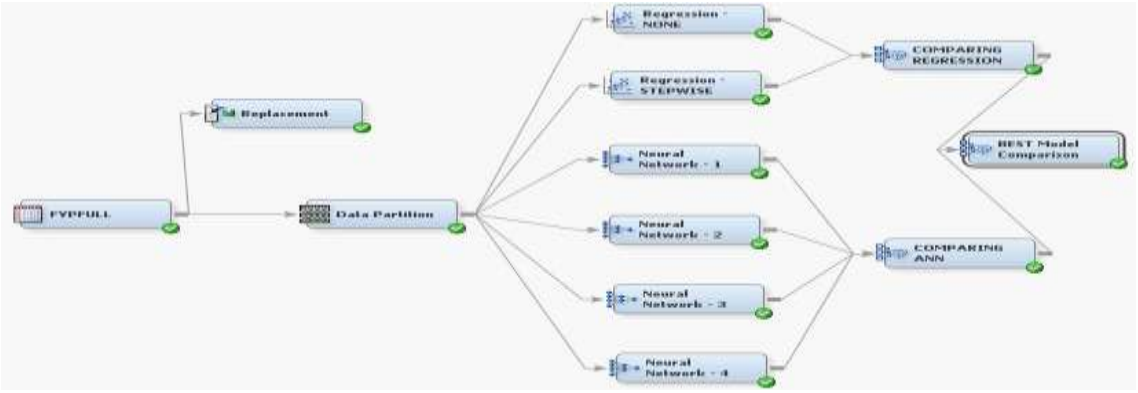

Figure 2 Methodological Framework in Determining the Best Model to Forecast the Electricity Consumption in Malaysia

Figure 2 shows the methodological framework in determining the best model to forecast the electricity consumption for the next 3 (three) years. All the inputs data have been through the partitioning process in order to better manage the data. These partitioned data will be estimating using the Multiple Regression and ANN methods. Basically, there are six types of models that comprise of two models from the regression analysis, and another four models from the ANN method to be used in this analysis. Both methods will then identify which model that best fits the data. As mentioned earlier, regression model is represented by linear regression and stepwise linear regression models, whilst ANN will be representing by four models with different hidden nodes, ANN-1, ANN-2, ANN-3 and ANN-4. Comparatively, the best model with the least error from each method between re

3. RESULTS AND DISCUSSION

Determining the Best Model of Multiple Linear Regression

Table 1 shows the analysis results of two types of regression models namely Linear Regression and Stepwise Linear Regression. The values of the gap of both types were calculated in finding which types of regression fits the data well. From Table 1, the Linear Regression function named as Regression-none produces higher gap value for both Average Squared Error (ASE) and Mean Squared Error (MSE) hence Regression-none is considered overfit. Reasons of this overfitting is mainly due to the complexity of the model that not suitable for normal regression technique. Thus, the Regression-none model is rejected. Comparatively, Regression-stepwise recorded highest valid values of ASE and MSE respectively with a similar value of 505494.5523 each. Hence, Regression-stepwise is chosen as the best model for regression.

Table 1 Error Measures and Parameter Values for Regression Type

Model Description	**Valid: Average Squared Error**	**Train: Average Squared Error**	**GAP**	**Valid: Mean Square Error**	**Train: Mean Square Error**	**GAP**	**No Underfit Model**
Regression - STEPWISE	505494.5523	459159.6702	46334.88	505494.5523	493171.4977	12323.05	Best Model
Regressio	579799.77	408945.82	170854	579799.77	465075.64	114724	Overfit

n - NONE	5	21		5	08	.1	Model

Determining the Best Model of Artificial Neural Network (ANN)

Four types of ANN model with different hidden nodes were established in determining which model fits the data well. Results show in Table 2 indicate that ANN-4 overfitted the model with the highest gap value of 119160.2 and 57227.03 respectively. Hence, left with ANN-1, ANN-2 and ANN-3 to be verified as the best model. Hypothetically, the model with the highest valid values of error measures is considered as the well fitted model. Since ANN-1 is recorded the highest valid values for both Average Squared Error (ASE) and Mean Squared Error (MSE) with 223906.265 and 223906.265 respectively, thus ANN-1 is selected as the best model for ANN method.

Table 2 Error Measures and Parameter Values for ANN Type

Model Description	**Valid: Average Squared Error**	**Train: Average Squared Error**	**GAP**	**Valid: Mean Squared Error**	**Train: Mean Squared Error**	**GAP**	**No Underfit Model**
ANN - 3	50019.15274	49062.45332	956.6994	50019.15274	86230.97857	-36211.8	
ANN - 2	73791.33717	58625.92461	15165.41	73791.33717	82934.23482	-9142.9	
ANN - 4	166079.2289	46919.05241	119160.2	166079.2289	108852.2016	57227.03	Overfit Model
ANN - 1	223906.265	196804.7031	27101.56	223906.265	232952.5058	-9046.24	Best Model

Determining the Best Method for Forecasting the Electricity Consumption

In progression from the previous exercises, further analysis was comparatively conducted between Regression-stepwise and ANN-1 models. As shown in Table 3, the Regression-stepwise model produces higher gap value for both Average Squared Error (ASE) and Mean Squared Error (MSE), so this is considered as overfit. Thus, this model is being rejected. On the other hand, ANN-1 recorded highest valid values for both ASE and MSE at the same amount of 223906.265 respectively. Therefore, ANN-1 model proved to be the best model for forecasting.

Table 3 Error Measures and Parameter Values for Regression and ANN Type

Model Description	**Valid: Average Squared Error**	**Train: Average Squared Error**	**GAP**	**Valid: Mean Square Error**	**Train: Mean Square Error**	**GAP**	**No Underfit Model**
ANN - 1	223906.265	196804.7031	27101.56	223906.265	232952.5058	-9046.24	Best Model
Regression -	505494.5523	459159.6702	46334.88	505494.5523	493171.4977	12323.05	Overfit Model

STEPWISE							

Forecasting the Electricity Consumption for 2020-2022 using the Best Method – ANN

Upon identifying ANN-1 as the best and the most stable model for prediction, this study uses NNTool technique for forecasting the future electricity consumption for Malaysia. Figure 3 shows the diagram of the ANN by NNTool under MATLAB.

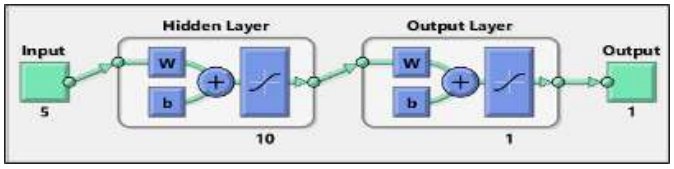

Figure 3 ANN for NNTool

A series of build and train processes were conducted repetitively to arrive at the most stable state of model of ANN model. The most stable state model will be used to forecast the values of future electricity consumption in Malaysia for the period of 2020 to 2022. Figure 4 to Figure 6 show the last three training results of the model.

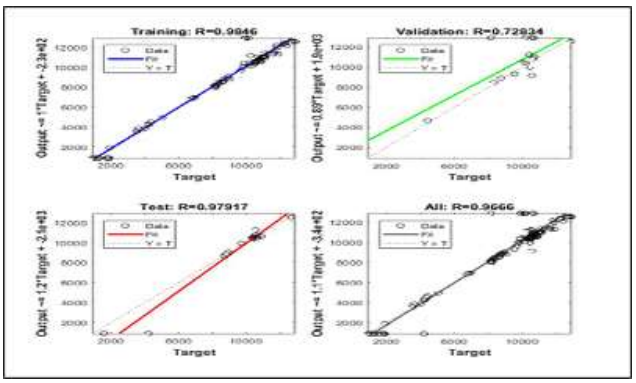

Figure 4 ANN Training for (Last 3 Train)

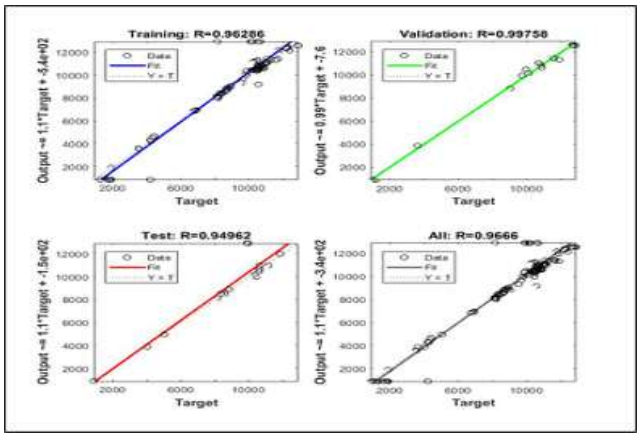

Figure 5 ANN Training (Last 2 Train)

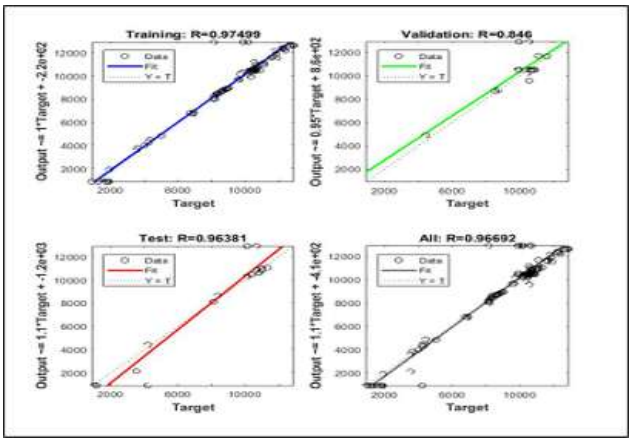

Figure 6 ANN Training (Last Train)

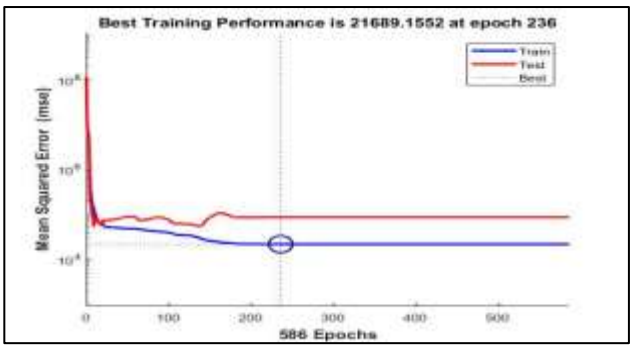

Figure 7 The Best Training Performance of the ANN Model

The movements of the train and the test dataset are then plotted and the best training performance is detected at epoch 236 with training performance recorded at 21689.1552 (Figure 7). Subsequently, the forecast values of the future electricity consumption are then projected. Figure 8 shows the projected value of the future electricity consumption in Malaysia (in kWh million) for the period of 2020 - 2022. The expected amount of electricity to be consumed by Malaysia for the next three years show a consistent and persistent increase over the time projected. This projection is believed to optimise the amount of future electricity to be generated for Malaysia. Concurrently, this will help to reduce wastage of unplanned and massive energy productions and at the same time preserving the environment for our future generations.

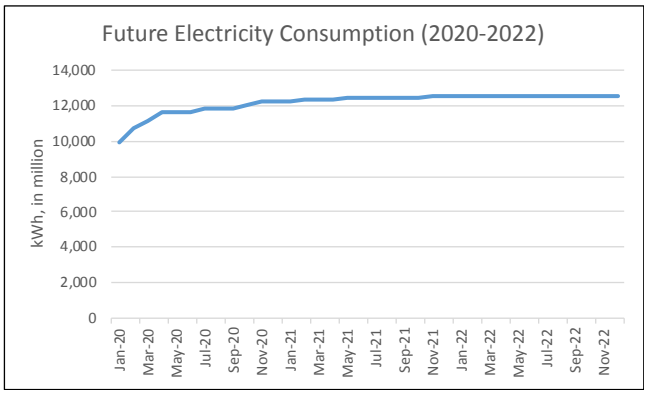

Figure 8 3-year Projection of Future Electricity Consumption in Malaysia

4. CONCLUSIONS

The electricity consumption data from 2010 to 2018 were used in identifying the best forecasting model for estimating future electricity consumptions in Malaysia. In comparing the methods of Multiple Linear Regression and Artificial Neural Network (ANN), ANN method with ANN-1 type proved to be the best and plausible model to generate the forecast values of monthly electricity consumption for the period of 2020 to 2022 for Malaysia. The forecast values generated are more interpretable and reliable. This forecast may help the Government of Malaysia, in particular the Ministry of Energy and Natural Resources to estimate the optimum amount of energy to be generated for the country. This will definitely improve our government strategies on spending the essential needs wisely and reduce wastage from excessive and undirected productions. Ultimately, the environment is saved too.

Acknowledgements

The main author would like to acknowledge the management, lecturers, staff and students of the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA. This research would not have been possible without their help and support.

REFERENCES

[1] A Khamis et.al. Energy & electricity consumption analysis of Malaysian power demand. *4th International Power Engineering and Optimization Conference (PEOCO)*, Shah Alam. 2010; 463-467.

[2] G Escrivá-Escrivá, C Álvarez-Bel, C Roldán-Blay and M Alcázar-Ortega. New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. *Energy and Buildings*. 2011; **43**, 3112-3119.

[3] I Laicane, D Blumberga, A Blumberga and M Rosa. Reducing household electricity consumption through demand side management: The role of home appliance scheduling and peak load reduction. *Energy Procedia*. 2015; **72**, 222-229.

[4] J Frost. Overfitting regression models: Problems, Detection & Avoidance, Available at: https://statisticsbyjim.com/regression/overfitting-regression-models/

[5] G O*f*ücu, OF Demirel and S Zaim. Forecasting electricity consumption with neural networks and support vector regression. *Procedia-Social and Behavioral Sciences*. 2012; **58**, 1576-1585.

[6] M Shamsuddin, S Abdul-Rahman and A Mohamed. Towards applying Deep Neural Network for complex input patterns. *Recent Advances in Mathematical and Computational Methods*. 2015.

[7] K Alexx. *Artificial Neural Networks, ComputerWorld. Computerworld-IT News, Features, Blogs, Tech Reviews*. Career Advice 12. 2001.

[8] AS Ahmad, MY Hassan, MP Abdullah, HA Rahman, F Hussin, H Abdullah and R Saidur. A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*. 2014; **33**, 102-109.

[9] PA Gonzalez and JM Zamarreno. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and buildings*. 2005; **37**, 595-601.

[10] MA Kahar, S Mutalib, Abdul-Rahman and Shuzlina. Early detection and classification of paddy diseases with Neural Networks and Fuzzy Logic. *In Proceedings of the 17th International Conference on Mathematical and Computational Methods in Science and Engineering*, MACMESE. 2015, 248-257.
[11] S Mutalib, S Jamian, S Abdul-Rahman and A Mohamed. Soil classification: An application of self-organising map and k-means. *2010 10th International Conference on Intelligent Systems Design and Applications*, Cairo. 2010; 439-444.
[12] ZAAK Adzhar and FH Tafri. Islamic options pricing model via Artificial Neural Network: Benchmarking to Black-Scholes. *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, Langkawi. 2012; 1-6.
[13] H Ahmad, NS Musa, N Nazamid and NA Zaharin. Classification of high and low level of PM10 concentrations in Klang and Shah Alam, Malaysia. *Malaysian Journal of Computing*. 2019; **4**, 325-334.