

Neural Network Prediction for Efficient Waste Management in Malaysia

Siti Hajar Yusoff, Ummi Nur Kamilah Abdullah Din, Hasmah Mansor, Nur Shahida Midi,
Syasya Azra Zaini

International Islamic University Malaysia Jalan Gombak, 53100 Kuala Lumpur, Malaysia,
Phone: (+603) 6196 4000, Fax: (+603) 6196 4053, Malaysia

Article Info

Article history:

Received May 13, 2018

Revised Jul 14, 2018

Accepted Jul 28, 2018

Keywords:

ANN prediction algorithm

MSWM in malaysia

Prediction of SWG

R2

Visual gene developer

ABSTRACT

Maintaining current municipal solid waste management (MSWM) for the next ten years would not be efficient anymore as it has brought many environmental issues such as air pollution. This project has proposed Artificial Neural Network (ANN) based prediction algorithm that can forecast Solid Waste Generation (SWG) based on population growth factor. This study uses Malaysian population as sample size and the data for weight is acquired via authorized Malaysia statistics' websites. All data will be normalized in the pre-processing stage before proceeding to the prediction using Visual Gene Developer. This project evaluated the performance using R2 value. Two hidden layers with ten and five nodes were used respectively. The result portrayed that there will be an increase of 29.03 percent of SWG in year 2031 compared to 2012. The limitation to this study is that the data was not based on real time as it was restricted by the government.

*Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Siti Hajar Yusoff,

International Islamic University Malaysia,

Jalan Gombak, 53100 Kuala Lumpur, Malaysia,

Phone: (+603) 6196 4000, Fax: (+603) 6196 4053, Malaysia.

Email: siti Yusoff@iium.edu.my

1. INTRODUCTION

Preventing the jeopardizing of mankind's health as well as maintaining a greener and cleaner environment is vital. One of the methods is by practicing a well-planned MSWM in a country. No matter where your country is, MSWM will be one of the important departments in a government because this department plays a key role in terms of the country cleanliness. Without this department, the country hygiene ecosystems will be affected and gives a bad impact in tourism industries. As part of the developing countries, Malaysia is not excluded from waste management problem.

Currently, Malaysia is facing big challenges when most of the available landfills have been closed because it has reached the maximum allowed disposal capacity. Another challenge is when human's population increased over time, it will increase the number of waste disposal too at an unknown amount. This unknown amount is caused by lack of real and accurate statistical information regarding the waste disposal. Meanwhile, using the conventional collection method for different household size's is another challenge that has to be faced by Malaysian authority due to non-systematic way of waste management.

Due to these challenges, one of the central state of Malaysia known as Selangor has taken its own initiatives towards a better city. One of the initiatives that Selangor has taken is a step toward smart city known as "Smart Selangor". With a moto of "future Selangor, beyond smart", Selangor has already come up with a few effective solutions that lead to smart city. One of the major concern in smart city is the waste management. In order for smart city to be manifest, there are some crucial challenges that must be faced by any organization including limited technology and funding.

National Solid Waste Management Department (NSWMD) is one of the departments under the Ministry of Urban Wellbeing, Housing and Local Government in Malaysia. One of the aims of the National Solid Waste Management Policy is to have an established management system that can be accommodated by every level of the community [1]. In order to practise this policy, it is very important for the government to have a genuine set of weight waste data.

As years passed by, the statistic of population is also increasing. According to a press released by the Department of Statistics Malaysia in July 2017, it is estimated that Malaysia is experiencing an increment of 1.3% population growth in 2017 compared to 2016 with approximately 32.0 Million people are living in Malaysia in 2017 [2]. Undeniably, waste generation will continually to increase with the growth of population over time. A statistic has outlined from year 2012 until year 2015, the increased amount of waste generated is increasing from 32,800 tonnes per day to 38,500 tonnes per day [3]. This problem is mainly resulted from the increment of population and variation in the household size [4]. Same situation also happens in Indonesia, where the increased in population is greatly triggering the MSWM issues [5].

As there are more people, more resources such as food will be consumed. Sadly, some people use this resource yet carelessly littering. As an example, some might throw the garbage away by the roadside, thus ignoring the ethical issues and the laws made by the government. This would be the most unethical practice towards the nature. Common environmental issues related to poor management of MSWM can be identified such as air pollution, water pollution as well as excessive generation of methane gas. This cycle will continuously to repeat if early prevention is taken for granted. Consequently, negative impacts that bring harm to the environment will slowly be the alarming issues to the society [6].

On the other hand, human overpopulation is one of the most unavoids causes for the environmental issue. As the population growth is escalating rapidly, there will be more people who will consume more resources. Undoubtedly, the excessive natural resource consumption for the development of the country will contribute to the same problem that this project has discussed earlier which is the increased of waste generation. Currently, MSWM in Malaysia do not have the exact statistic of how much waste is generated and how many times the waste bins get full per day. Without these statistics, it is very hard for the government to provide ample spaces of the compost sites and to plan for the garbage pick-up schedule for the future.

Therefore, it is very important to predict the amount of waste generated to ease the process in managing future MSWM. Recently, there are many researches on forecasting the SWG based on prediction models. Prediction models can give information about the future SWG based on many performance's criterion such as Mean Square Error (MSE), Mean Absolute Percentage Error (MAPE) and R^2 . Many studies suggested using ANN as the prediction tool [7], [8].

Sun & Chungpaibulpatana used MLP under ANN model and Pearson Correlation to predict SWG in Bangkok [9]. At the beginning of the research, few modelling techniques have been explored based on few influences such as population growth and household income. Also, interpolation technique has been applied during data collection stage due to some missing values. Neural fitting tool has been used to select, create and train data of the network based on MSE and regression analysis. For MLP, one neuron hidden layer has been applied that results in the acceptable fitting value R^2 of 0.96. During the evaluation stage, the performance for both techniques has been compared. The results managed to illustrate that ANN model is much more accurate compared to Principal Component Analysis-Regression (PCA-Regression) by 10% based on R^2 value. However, the values of MSE for both PCA-Regression and ANN model were very high which are 221805.2 and 63929 respectively.

In addition, many researchers use ANN as their classification model to predict many things [10]-[12]. Litta et al. stressed that the forecasting of thunderstorm is one of the toughest prediction tasks. However, the study has used ANN as their classification method to forecast the incoming thunderstorm based on the obtained meteorological parameters. Six learning algorithms were used and the performances have been compared. The results outlined that the Levenberg-Marquardt has outperformed the rest of the algorithm to predict the thunderstorm in terms of the statistical measures. The outcome of the study concluded that ANN is best used to predict any real-time data with less errors. Therefore, this project will use ANN as the classification model as it has been widely known to portrait the best results as compared to any other models while R^2 value will be used to evaluate the performance of the prediction algorithm.

The main objective of this research is to design efficient prediction algorithm for waste management to predict the generation of waste based on population growth in Malaysia. The remaining section of this paper comprises of few main parts. Section 2 will be explaining about the conducted research method while section 3 will lay out the results and the analysis of the experimentation. As a conclusion, the last section will conclude the paper.

2. RESEARCH METHOD

This section will discuss about the proposed methodology that can predict the SWG based on population growth. For this factor, the study has chosen Malaysia as the sample size. This section will be divided into three stages; data acquisition, pre-processing and evaluation. Section 2.1 will explain on the method for collection of data of waste amount generated and number of population in Malaysia. Whilst, section 2.2 describes the process of pre-processing on the collected data. Lastly, Section 2.3 explains on the steps to evaluate the data.

2.1. Data Acquisition

Earlier, this project has planned to obtain a real-time latest data with one of the MSWM contractors in Selangor. However, due to some confidential issues, the contractor could not provide any data for this project. Due to this limitation, this project acquired the number of population and the amount of waste generated via authorised websites [2], [3]. Then, these data will undergo the pre-processing stage in Section 2.2.

2.2. Pre-processing

As mentioned in the previous Section 2.1, this project will use ANN as the classification model. Firstly, the data of number of population and SWG must be pre-processed before proceeded to neural network training, due to noise reduction and the undesired ANN learning rate [4].

Saini et al. mentioned that the first step in the pre-processing stage is to obtain a trend line. This concept is called Stationary Chain Concept [4]. To meet this concept, statistical measures such as mean, need to be constant for some time and it can be achieved by observing the trend line. The reason of achieving the Stationary Chain Concept is to make sure that the trained model will be in the range of the observed data. In this study, the trend line is obtained via MATLAB. With the curve fitting tool application in MATLAB, few sets of trend line will be displayed. Then, the R^2 value of the trendlines will be compared and the trendline which has the highest value of R^2 will be chosen. R^2 is one of the widely used statistical measures. R^2 calculation is shown in Equation 1 [13]. The closer the value of R^2 to 1, the more the variability of response surrounding the mean and the accurate the result is.

$$R^2 = \frac{\text{Explained variation}}{\text{Total variation}} \quad (1)$$

The next step in the pre-processing stage is to train and to predict the data using ANN classification model. This step will be done via Visual Gene Developer, one of the softwares that can be used to train and predict any data. The default algorithm of feed forward neural network with back propagation learning which will perfectly train the network. The equation for this algorithm is shown in Equation 2 [14].

$$b_j = f(\sum_{i=1}^n (w_{ij}a_i) - T_j) \quad (2)$$

where a_i and b_j will be the input and output variable respectively. f will be the transfer function, w_{ij} will be the weight factor between two nodes and T_j will be the internal threshold. The procedure for ANN training in Visual Gene Developer is laid out as in Figure 1. After setting up the architecture, all data must undergo a process called normalization.

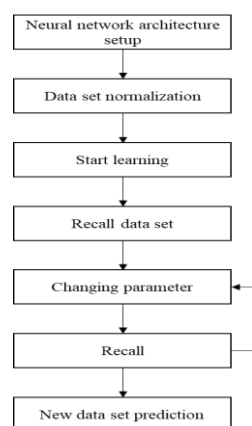


Figure 1. ANN training procedure [12]

One of the main reasons to perform the normalization is to alter the scale of the acquired data to be in the range of 0 to 1. All data must be in this range to perform ANN training. The normalization is done via the following formula in Equation 3.

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3}$$

where x is the variable, x_n will be the normalized variable, while x_{max} and x_{min} are the maximum and minimum of the input variables respectively.

The overview of the Visual Gene Developer is shown in Figure 2. The next step after normalization is to train the data and changing the parameter at the training setting so that the total cycles status will achieve the maximum input of the training cycle. After the maximum cycle is obtained, a set of prediction is generated. The pseudocode for this procedure is laid out as follow:

```

Function Main()
Call NeuralNet.OpenNN_Once("Sample SinCos - Trained network.vgn")
NeuralNet.InputData(1)=0.28
NeuralNet.InputData(2)=0.6393389
Call NeuralNet.PredictNN()
Output2=NeuralNet.OutputData(2)
Main=Output(2)
End Function
    
```

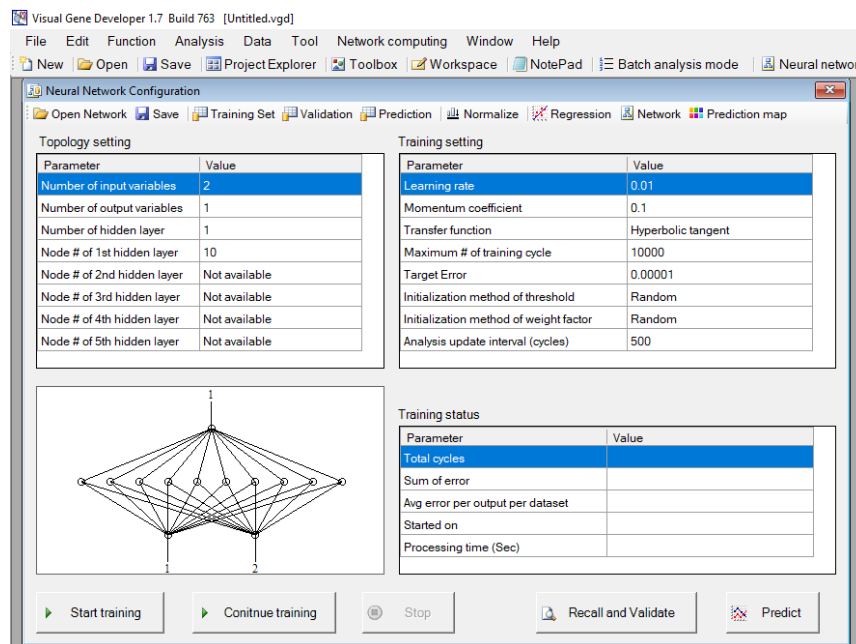


Figure 2. ANN lay out in visual gene developer software

2.3. Evaluation

In this section, the predicted data from the pre-processing stage will then undergo the evaluation process where the values of R^2 from different numbers of hidden layers and nodes will be compared. ANN is one of the statistical prediction tools that can do the job of a complex recognition from the inputs, to the network (i.e. experimentation data) and to the output [15]. ANN architecture is made from several artificial neurons. Figure 3 shows the example of ANN architecture consists of two inputs and two hidden layers with five and ten nodes each. These nodes represent the body of the ANN architecture. The higher the number of hidden layers in an architecture, the complex the fitting will be. However, it will need an extra computational power to process the input data. After testing a few combinations of the number of the hidden layers and nodes, the combination which will give the closest number of R^2 to 1 will be chosen.

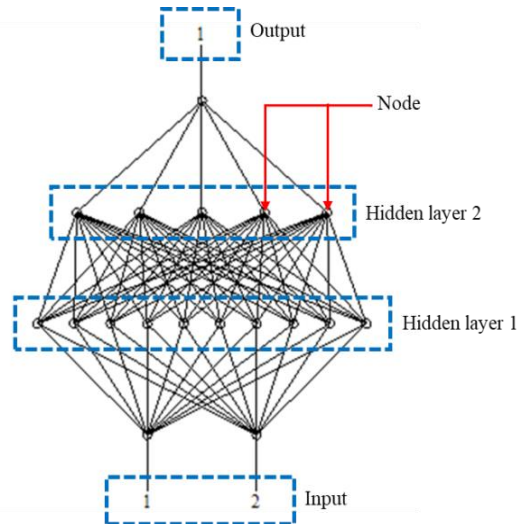


Figure 3. Example of ANN architecture

3. RESULTS AND ANALYSIS

In this section, the steps to obtain forecasted SWG based on population growth effect using ANN is portrayed. To depict the steps, this section will explain stages mentioned in Section 2 in Section 3.1, Section 3.2 and Section 3.3. Whilst, Section 3.4 proceed to the prediction of SWG in Malaysia until year 2031.

3.1. Data Acquisition

The data for the number of population in Malaysia and the amount of SWG from year 2012 until 2017 have been obtained from two websites [2], [3]. Table 1 shows the data obtained and it is plotted in Figure 4. In Table 1, it can be noted that as year increase; the weight of waste is also increased. There is an increase of 4.94 percent of weight of waste from 2016 to 2017. Whilst, for a period of six years there is a significant increase in weight of waste by 22.97 percent and population growth of 7.76 percent. Figure 4 shown as the number of population grow, the weight of waste increase too. The amount of waste produced is directly proportional to the growth of population.

Table 1. Data for Malaysia's Population and SWG from 2012 to 2017

Year	Population	Weight of waste (million ton/day)
2012	29,170,456	32.869
2013	29,706,724	34.763
2014	30,228,017	36.66
2015	30,723,155	38.563
2016	31,187,265	40.566
2017	31,624,264	42.672

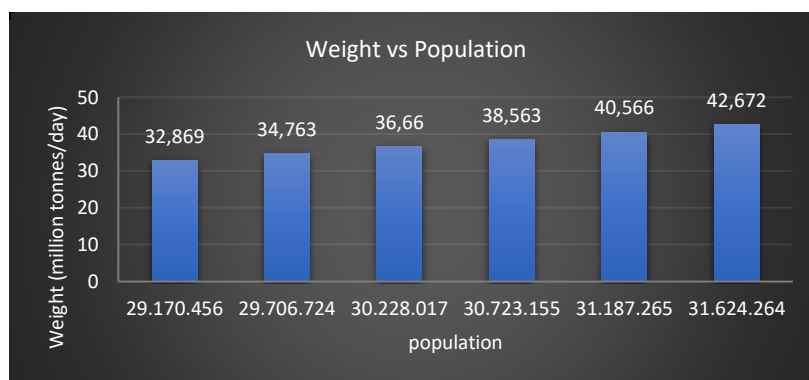


Figure 4. Data for Malaysia's population and SWG from 2012 to 2017

3.2. Pre-processing

Data for waste weight in Figure 4 has undergone the first step in pre-processing stage which is to obtain the trend line. Table 2 summarized few trend lines obtained via curve fitting tool in MATLAB while the best trendline obtained is plotted in Figure 5. There are many trend lines that have been applied during this project such as exponential, polynomial and power. As in Table 2, all R² value somehow lead to accurate fitting because all values are very close to 1. Whilst, the highlighted row in Table 2 shows the best trend line obtained. The reason why this project did not consider choosing row 1 and row 6 which has R² value of 1 is because to avoid over fitting. The equation shown in Figure 5; $y = 4e-06*x - 83$, is the trend line's equation which resulted in R² value of 0.9985, referring to row 3 in Table 2 where y is the amount of weight and x refers to the number of population.

Table 2. Trend Lines Via Curve Fitting Tool for Population Growth

Combination	Trend line	Fitting options	R ² value
1	Exponential	term: 1	0.9975
2	Exponential	term: 2	1
3	Polynomial	degree: 1 (linear)	0.9985
4	Polynomial	degree: 1 (robust: LAR)	0.9984
5	Polynomial	degree: 1 (robust: Bisquare)	0.9981
6	Polynomial	degree: 2	1
7	Power	term: 1	0.9434

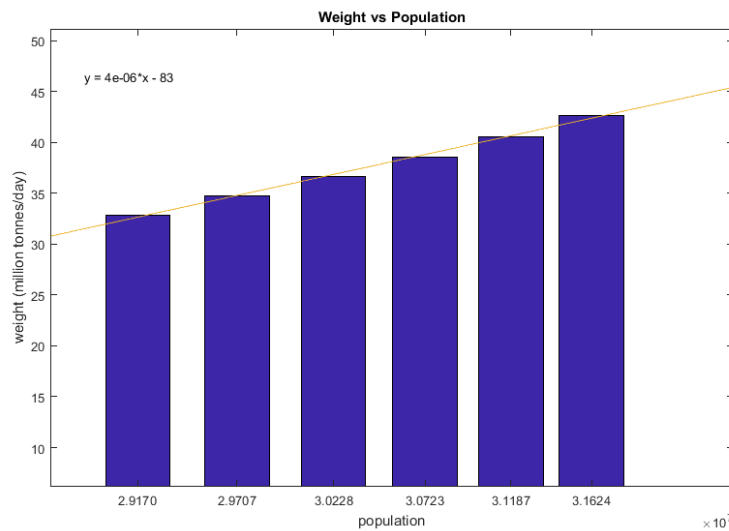


Figure 5. Trend line plotted on waste weight data for population growth

The second step in this pre-processing stage is to predict the production of waste using ANN via Visual Gene Developer. As been portrayed in Section 2, before proceeding to the prediction, all data must be normalized. Using Visual Gene Developer, the normalization in this project was easily done as shown in Figure 6. This ‘max number’ is taken as x_{max} as shown in the Equation 3 from Section 2 under Section 2.2. Thus, this normalization done in this software is as shown in Equation 3 in Section 2.2. After the normalization is done, the project has proceeded to the training and prediction.

For this step, a few sets of combination between the number of hidden layers and the number of nodes were applied. The sample of the regression line of R² value is shown in Figure 7. In this Figure 7, there are 2 lines which are grey line presenting the threshold value whilst the blue line represents the result obtained from the prediction. The closer the blue line to the grey line; the threshold value which hold the value of R² of 1, the accurate result of different combination will be tested and the resulted R² value will be depicted in Table 3 in Section 3.3.

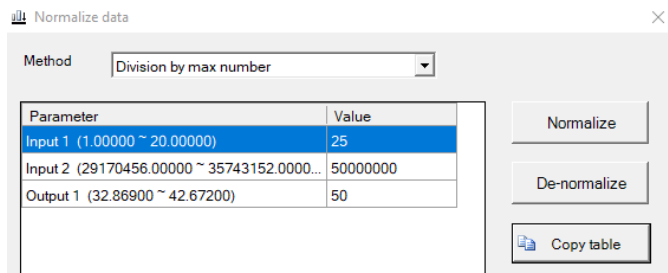


Figure 6. Data normalization for population growth

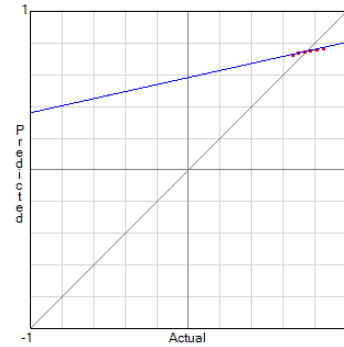


Figure 7. Sample of regression graph

3.3. Evaluation

In this section, few sets of regression lines obtained in previous Section 3.2 will be compared and evaluated. The value of R^2 obtained are summarized in Table 3. As depicted in Table 3, the highest R^2 value is 0.9886625 when the number of the applied hidden layers are two and the number of nodes for the first layer is ten while the second layer is five as been highlighted in combination number 4. The regression line is portrayed in Figure 8. As being shown in Figure 8, the blue line is very close to the threshold line (grey line). This shows the accuracy of the algorithm with reference to the obtained R^2 value.

Table 3. Summary of Performance of Prediction Model based on Different Hidden Layers and Number of Nodes for population growth factor

Combination	Hidden layer	1st node	2nd node	R^2 value
1	1	5	-	0.9871439
2	1	10	-	0.9869655
3	2	5	10	0.996362
4	2	10	5	0.9851083
5	2	2	8	0.9907039
6	2	8	2	0.9928292

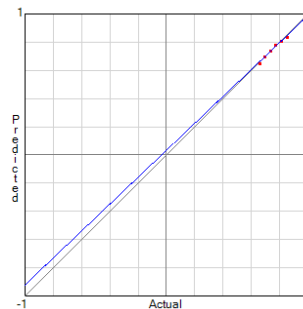


Figure 8. Regression line of the highest combination output of $R^2 = 0.9886625$

From Table 3, the project considered the combination that gives the highest R^2 value of 0.9886625 as the prediction algorithm. Next, this project will use this prediction algorithm to predict the amount of waste generated and to compare it with the observed weight for year 2012 until year 2017. The result of the comparison will be depicted in Figure 9. Figure 9 shows the accuracy of the prediction algorithm as there is not much difference in the observed and the predicted line. As shown in Figure 10, the value of the sum of error for this algorithm is only 0.001594916.

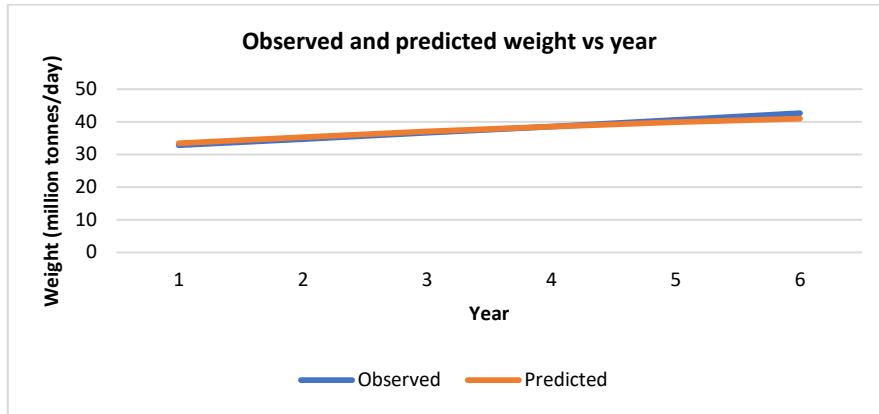


Figure 9. Comparison of the observed and predicted weight for population growth

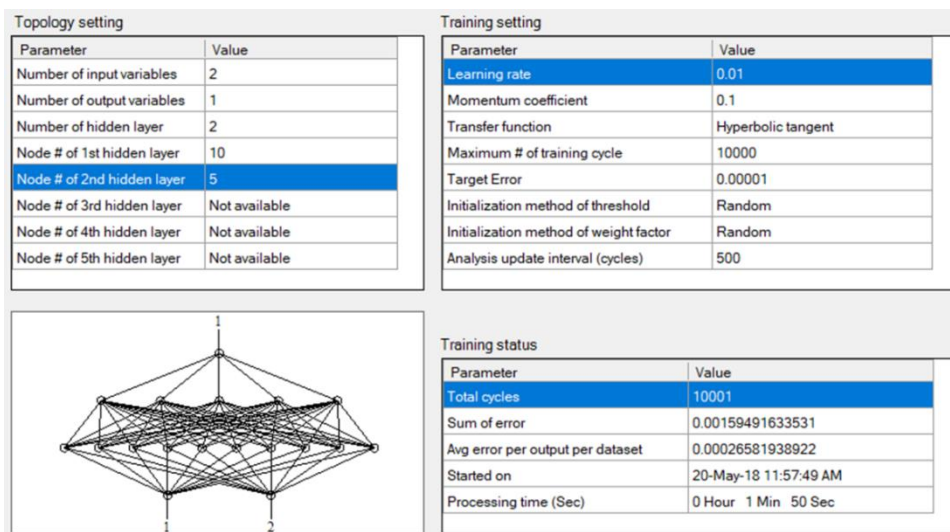


Figure 10. Visual Gene Developer experimentation layout for population growth

3.4. Results

Finally, after obtaining the best prediction algorithm in previous Section 3.3, this project will proceed to forecast the amount of waste generated until year 2031. The criterion of the prediction algorithm deduced from the previous stage are:

1. ANN
2. Two hidden layers
 - a. First hidden layer: 10 nodes
 - b. Second hidden layer: 5 nodes

The predicted SWG in Malaysia until year 2031 is shown in Figure 11. Year 1 in Figure 11 represents year 2012, the first year’s sample is up to 20 years which refers to year 2031. As shown in the Figure 11, in year 20 (refers to year 2031), the number of waste that will be generated is 47.2 million tonnes per day, compared to year 2012 where the amount of waste was only 33.5 million tonnes per day. There is an increase of 29.03 percent of weight of waste from 2012 to 2031. Figure 11 shown as the number of population grow, the weight of waste increase too. The amount of waste produced is directly proportional to the growth of population.

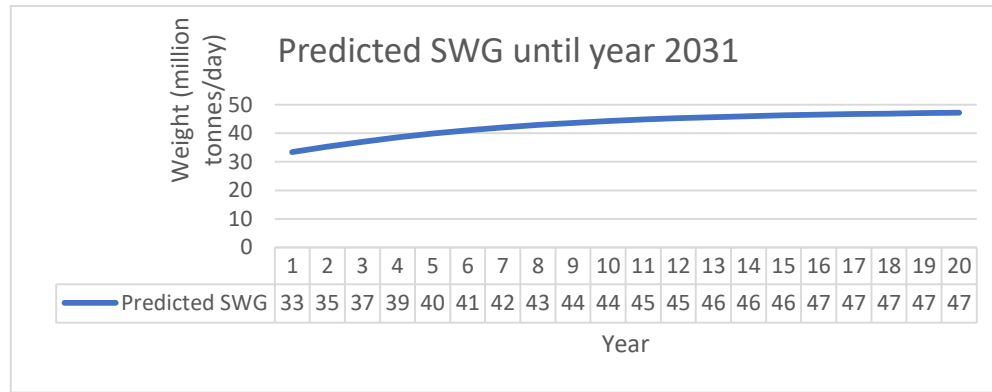


Figure 11. Prediction of SWG for twenty years

In relation to the above predicted waste in 2031, this study deduced that current system of managing SWG and the amount of waste disposal area are needed to be revised time by time. As mentioned in Section 1, this prediction of SWG will help the authorities to help preparing sufficient disposal land early with regards to the excessive amount of waste that will be generated.

4. CONCLUSION

Poor MSWM will lead to many environmental and health issues such as excessive amount of methane gas production and malaria. Therefore, in this project, prediction algorithms are proposed to provide the forecasted SWG based on population growth factor. Prediction algorithm plays a very important role not only in MSWM in Malaysia but also in handling the waste. This algorithm will provide the management personnel to have the estimation and how to handle the SWG in the future. Based on the experimentation results, it shows that the objectives of this project have been achieved. In addition, the result in Section 3 indicated that the prediction of SWG based on population growth factor is best suited when ANN is used with two hidden layers where the number of nodes for the first layer is ten and the second layer is five. The result also shows that the prediction algorithm has predicted the rate of increment of SWG is 29.03 percent for the next twenty years.

However, the limitation in this study is that data for population growth factor can only be obtained via authorized websites due to some restriction mentioned by one of the authorities handling MSWM in Malaysia. On the other hand, room for improvement can always be proposed in any project. Further suggestions can be considered to achieve better results. There are two recommendations that can be considered to improve future work for this project such as the use of other prediction algorithm such as Adaptive Neuro-Fuzzy Inference System (ANFIS) or Nonlinear Autoregressive Network with Exogenous Inputs (NARX) and to consider more SWG factors such as household size and household income.

ACKNOWLEDGEMENTS

This work was partially supported by Ministry of Higher Education Malaysia (Kementerian Pendidikan Tinggi) under Fundamental Research Grant Scheme (FRGS) number FRGS17-038-0604.

REFERENCES

- [1] Jabatan Pengurusan Sisa Pepejal Negara and Kementerian Kesejahteraan Bandar, Perumahan dan Kerajaan Tempatan. Survey on Solid Waste Composition, Characteristics & Existing Practice of Solid Waste Recycling in Malaysia [Internet]. Selangor; 2013. Available from: http://jpspn.kpkt.gov.my/resources/index/user_1/Sumber_Rujukan/kajian/Final_Report_REVz.pdf
- [2] Department of Statistics Malaysia Official Portal [Internet]. Dosm.gov.my. 2017 [cited 6 July 2018]. Available from: <https://www.dosm.gov.my/v1/>
- [3] Solid Waste Management Lab 2015 [Internet]. Kuala Lumpur: Performance Management and Delivery Unit (PEMANDU); 2015 [cited 6 July 2018]. Available from: http://www.kpkt.gov.my/resources/index/user_1/Attachments/hebahan_slider/slaid_dapatan_makmal.pdf
- [4] Solid Waste Management Lab 2015 [Internet]. Kuala Lumpur: Performance Management and Delivery Unit (PEMANDU); 2015 [cited 6 July 2018]. Available from: http://www.kpkt.gov.my/resources/index/user_1/Attachments/hebahan_slider/slaid_dapatan_makmal.pdf

- [5] Saini R, Ahuja N, Bahukhandi K. Futuristic Projection of Solid Waste Generation in Dehradun City of Uttarakhand using Supervised Artificial Neural Network-Non-Linear Autoregressive Neural Network (NARnet). International Journal of ChemTech Research [Internet]. 2017 [cited 6 July 2018];10(13):283-299. Available from: [http://www.sphinxssai.com/2017/ch_vol10_no13/3/\(283-299\)V10N13CT.pdf](http://www.sphinxssai.com/2017/ch_vol10_no13/3/(283-299)V10N13CT.pdf)
- [6] Prajati G, Padmi T, Rahardyan d. Projection of Big Cities Waste Management and Cost Based on Economic and Demographic Factors in Indonesia. IOP Conference Series: Earth and Environmental Science. 2017;97:012014.
- [7] Labs E. Overflowing garbage bins: 5 impacts on health and environment, and how to prevent | Ecube Labs [Internet]. Ecube Labs. 2016 [cited 6 July 2018]. Available from: <http://ecubelabs.com/blog/overflowing-garbage-bins-5-impacts-on-health-and-environment-and-how-to-prevent/>
- [8] Yang Y, Hu J, Zhang M. Predictions on the Development Dimensions of Provincial Tourism Discipline Based on the Artificial Neural Network BP Model. Indonesian Journal of Electrical Engineering and Computer Science. 2014;3(2).
- [9] E J, P S, S A. Prediction of Lead-Acid Battery Performance Parameter: An Neural Network Approach. Indonesian Journal of Electrical Engineering and Computer Science. 2013;2(1).
- [10] Sun N, Chungpaibulpatana S. Development of an Appropriate Model for Forecasting Municipal Solid Waste Generation in Bangkok. Energy Procedia. 2017;138:907-912.
- [11] Lee K, Chung N, Hwang S. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. Ecological Informatics. 2016;36:172-180.
- [12] Yin C, Rosendahl L, Luo Z. Methods to improve prediction performance of ANN models. Simulation Modelling Practice and Theory. 2003;11(3-4):211-222.
- [13] Litta A, Mary Idicula S, Mohanty U. Artificial Neural Network Model in Prediction of Meteorological Parameters during Premonsoon Thunderstorms. International Journal of Atmospheric Sciences. 2013;2013:1-14.
- [14] Owens M. Sec 10.3 Coefficient of Determination and Standard Error of the Estimate. Presentation presented at; 2016.
- [15] Artificial neural network [Internet]. Visualgenedeveloper.net. 2018 [cited 6 July 2018]. Available from: http://www.visualgenedeveloper.net/Func_ANN.html
- [16] Wagner M, Ansell D, Kent D, Griffith J, Naimark D, Wanner C et al. Predicting Mortality in Incident Dialysis Patients: An Analysis of the United Kingdom Renal Registry. American Journal of Kidney Diseases. 2011;57(6):894-902.