IDENTIFY INFLUENTIAL VARIABLES THAT AFFECT MUNICIPAL SOLID WASTE GENERATION AND FORECAST MUNICIPAL SOLID WASTE GENERATION IN SRI LANKA

Dissanayaka D.M.S.H.*, Vasanthapriyan S.

Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

*shalikahiranthi94@gmail.com**

ABSTRACT

Municipal Solid Waste Management (MSWM) is one of the primary tasks of metropolitan local authorities in developing countries. For efficient and effective waste management schemes and scheduling, accurate forecast of Municipal Solid Waste (MSW) generation is essential, due to the uncertainties and unavailability of sufficient MSW generation information and resources in developing countries. The objectives of this paper are to identify influential variables that affect the amount of MSW generation and to predict the future MSW in Sri Lanka by consuming linear, nonlinear models and machine learning technique and propose a model for forecast future MSW generation using influential variables. Socio economic data and waste generation data are collected from Department of Census and Statistics and National Solid Waste Management Support Center. Data preparation is done with substitute missing values by average values. Pearson correlation and Principal Component Analysis is used to find correlation among influential variables. Linear model, Non-linear model and machine learning model are used to forecast municipal solid waste generation in Sri Lanka. Relatively Linear regression analysis, artificial neural network (ANN) and Random forest used as linear models, Non-linear model and machine learning model. Relatively Correlation coefficient of linear regression classification, random forest classification and ANN are R^2=0.6973, R^2=0.9608 and R^2=0.9923. Based on correlation coefficient, ANN provide higher accurate result than linear regression and random forest models. ANN is conducted by designing appropriate network architecture with one neuron demand in the hidden layer. Based on the analysed result a model was proposed to forecast future MSW generation with four influential variables that are municipal solid waste generation, total population, GDP growth rate, and Crude birth rate.

Keywords: *Forecast, Influential Variables of Solid Waste*

Generation, Solid Waste, Random Forest Model

1 INTRODUCTION

Solid waste from human activity is a serious problem, especially in many developing countries, due to changes in consumption patterns and uncontrolled urbanization (Ghinea et al., 2016). Rapid industrialization and demographic explosions forced people to move from villages to cities, producing thousands of tons of solid waste every day. The way this waste is handled, stored, and collected can pose risks to the environment and public health and increase management costs (Marandi & Ghomi, 2016). Solid waste management (MSW) is a matter of concern for the authorities of major cities in the world. Municipal Solid Waste Management (MSWM) is one of the main functions of local institutions in cities in developing countries. Successful planning of a solid waste management scheme is critically dependent on the predictability of solid waste production (Sodanil & Chatthong, 2014). It is more important to have MSW generation prediction to manage MSW facilities, manage expenditures, increase efficiency in waste management, and manage the impact on waste recycling and land-filling (Sun & Chungpaibulpatana, 2017).

There are a number of organizations involved in waste management at various phases in Sri Lanka, including the National Solid Waste Management Support Centre, Ministry of Local Government and Provincial Councils, Ministry of Megapolis and Western Province Development, Ministry of Mahaweli Development and Environment, Central Environmental Authority, Urban Development Authority, Western Province Waste Management Authority, Local Authorities (Admin, 2019). With the Western Province responsible for nearly 60 percent of solid waste production, Sri Lanka produces 7000 MT of solid waste per day but daily waste collection per day 3500MT by local authorities (LA) (Admin, 2019). An average of 1-0.4 kg of landfill per day is generated by each individual. There are 335 local authorities responsible for municipal solid waste in Sri Lanka. Only half of the waste produced is collected, according to the Waste Management Authority and the Central Environmental Authority (Admin, 2019). For optimizing landfill storage and processing technologies, a straight forward knowledge of the features and patterns of solid waste generation is crucial. Managing an efficient solid waste management plan process is a complex scenario. As a basis and prerequisite for an effective MSW program, it is very important to quantify and predict the generation of municipal solid waste (Sun & Chungpaibulpatana, 2017).

The major objectives of this research are to identify influential variables that affect the amount of Municipal Solid Waste generation in Sri Lanka, to predict the future Municipal Solid Waste generation in Sri Lanka using linear, nonlinear and machine learning models and propose a model to forecast municipal solid waste generation in Sri Lanka.

2 LITERATURE REVIEW

Several works of literature have been studied to identify the variables, tools, and techniques to forecast municipal solid waste generation and from which some of the concepts related to the forecast municipal solid waste generation are explained.

In the article (Sun & Chungpaibulpatana, 2017) they provide a model for forecasting municipal solid waste generation in Bangkok using linear and nonlinear models with Matlab tool. Total municipal solid waste), population indicators (Total number of residents, Native residents, Total native people aged 15-59 years and Total people aged 15-59 years), dwelling indicator (Number of households), economic indicator (Income per household), and external indicator (Number of tourists) are considered as variables that can be affect to MSW generation. In this research, linear regression analysis and ANN used to forecast MSW generation.

The review of models used to forecast municipal solid waste generation, (Sodanil & Chatthong, 2014) addressed the time series forecasting model for the amount of solid waste generated in Bangkok using artificial neural networks, and suggestions a model for solid waste forecasting. Rapidminer tool is used to analyze and develop a forecast model. The backpropagation algorithm is used to train the artificial neural network model.

Municipal solid waste streams analysis and forecasting of solid waste quantities generated Time series modeling is widely used in prediction studies (Marandi & Ghomi, 2016). The focus of this research is on analyzing, comparing, and choosing for Tehran city the best time series model to predict the amount of solid waste is generated. The results showed that the MAPE, MAD, and RMSE measures used to predict the amount of solid waste generated in the next years have been outperformed by the ARIMA (2, 1, 0) model. R software is used to done prediction and simulation. They also discussed stationary tests, model identification, diagnostic checking, and scenario reduction in the article.

The accurate prediction of waste production is an essential step to plan waste management schemes appropriately since different variables can influence variations in waste production (Ghinea et al., 2016). The application of predictive and prognosis models are useful tools, as reliable support for decision-making processes. The following indicators, such as resident number, population status, urban population, and municipal waste amount, were used to estimate the quantity of solid waste as reference variables in the prognostic models. Waste Prognostic Tool, regression analysis, and time series analysis used to forecast municipal solid waste generation and configuration by considering the case study of Iasi Romania. Six solid waste categories (paper, plastic, metal, glass, biodegradable, and other waste) were determined for the regression models. Accuracy Measures have been calculated and findings have shown that the S-curve progression model is the best for MSW.

3 METHODOLOGY

Data collected from Municipal councils (MC), urban councils (UC) and Pradeshiya Sabha (PS) are the local authorities (LA) responsible for MSWM. Waste related data are collected out of 335 LA from the National Solid Waste Management Support Center in the Ministry of Internal and Home Affairs and Provincial Councils and Local Government and Central Environmental Authority (CEA). Socio-economic indicators are collected from the Department of Census and Statistics, Ministry of National Policies and Economic Affairs in Sri Lanka. Data preparation is done with substitute missing values by the mean value of each variable. Pearson correlation used to search for multicollinearity between variables. Principal Component Analysis (PCA) is one of the multidimensional statistical methods used to reduce the complexity of the input variables and to prevent multicollinearity. Linear regression analysis, Artificial Neural Networks (ANN) and random forest models are used as a linear model, non-linear model, and machine learning model to forecast municipal solid waste generation in Sri Lanka.

4 DATA ANALYSIS AND RESULTS

Pearson correlation analysis and principal component analysis is conducted to reduce the complexity of the input variables and to prevent multicollinearity when the number of observations includes many variables. Principal component analysis with ranker attribute select four variables that are MSW generation per day, Crude birth rate, GDP growth rate, and Total population with the highest correlation coefficient. Table 1 shows the correlation coefficient related to the MSW generation per day with other variables.

Table 1: Correlation Ranking Filter corresponding with MSW generation per day

Correlation	Attributes
0.36187	Crude Birth Rate
0.18203	GDP Growth Rate
-0.00988	Total Population
-0.04142	GDP per capita USD
-0.07866	Mean house hold income per month
-0.10424	Unemployment Rate
-0.10673	Crude Death Rate
-0.25543	Labour Force
-0.35475	Tourists Arrivals

The correlation coefficient of linear regression analysis is R=0.6973. Figure 1 shows that the train predicts for targets form the year 2009 to 2017 using linear regression analysis. Table 2 shows the evaluation of the training data. And figure 2 shows that the predicted MSW generation for the next 8 years using linear regression analysis.

1 Step-ahead prediction for: MSW Generation per Day

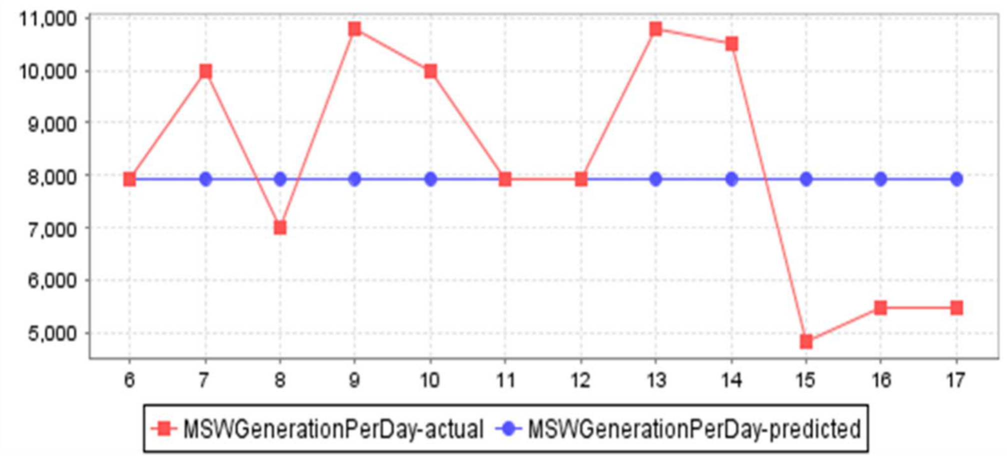

Figure 1: Linear Regression-train predict for targets data

Table 2: Linear Regression-Evaluation on training data

Target	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead	5-step-ahead	6-step-ahead	7-step-ahead	8-step-ahead
N	12	11	10	9	8	7	6	5
Mean absolute error	1780.5218	1942.369	1931.3856	2040.6732	1940.982	1925.0933	2245.9083	2695.0494
Direction accuracy	18.1818	20	22.2222	25	28.5714	33.3333	20	25
Root mean squared error	2121.5202	2215.8555	2231.5608	2330.958	2259.5608	2287.6496	2470.944	2706.7835

Future Forecast for: MSW Generation per Day

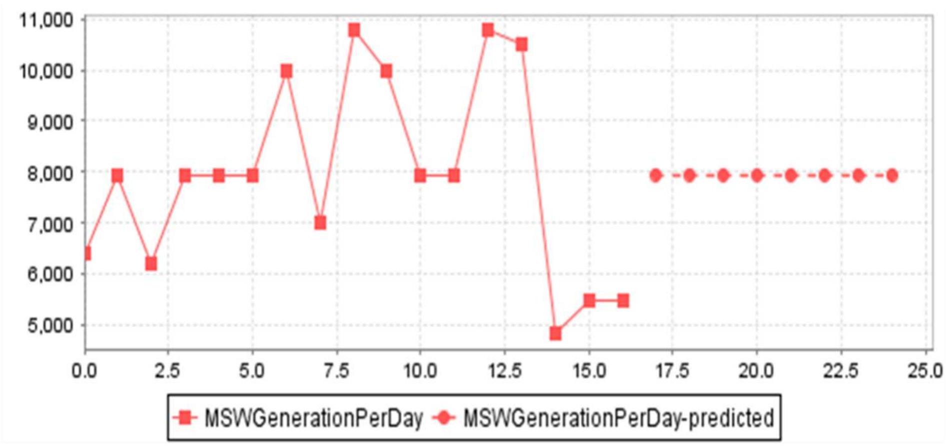

Figure 2: Linear Regression- train future predict

The correlation coefficient of random forest classification is, R=0.9608. Figure 3 shows that the train predicts for targets form the year 2009 to 2017 using random forest analysis. Table 3 shows the evaluation of the training data. Random forest analysis provides less root mean squared error and higher accuracy than linear regression analysis. And figure 4 shows that the predicted MSW generation for the next 8 years using random forest analysis.

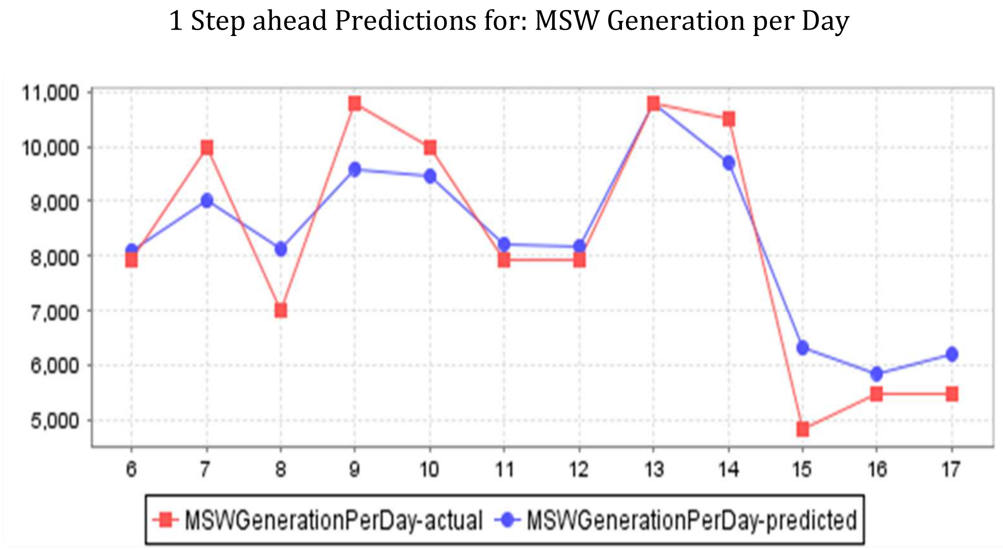

Figure 3: Random Forest- train predict for targets data

Table 3: Random Forest -Evaluation on training data

Target	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead	5-step-ahead	6-step-ahead	7-step-ahead	8-step-ahead
N	12	11	10	9	8	7	6	5
Mean Absolute Error	659.4396	776.9878	944.8459	1029.6985	1080.3582	1249.3232	1266.4317	1393.0393
Direction Accuracy	72.7273	70	66.6667	62.5	71.4286	66.6667	80	50
Root Mean Squared Error	796.6584	889.1105	1058.8659	1156.9377	1188.7841	1285.8332	1323.3955	1500.1245

Future Forecast for: MSW Generation per Day

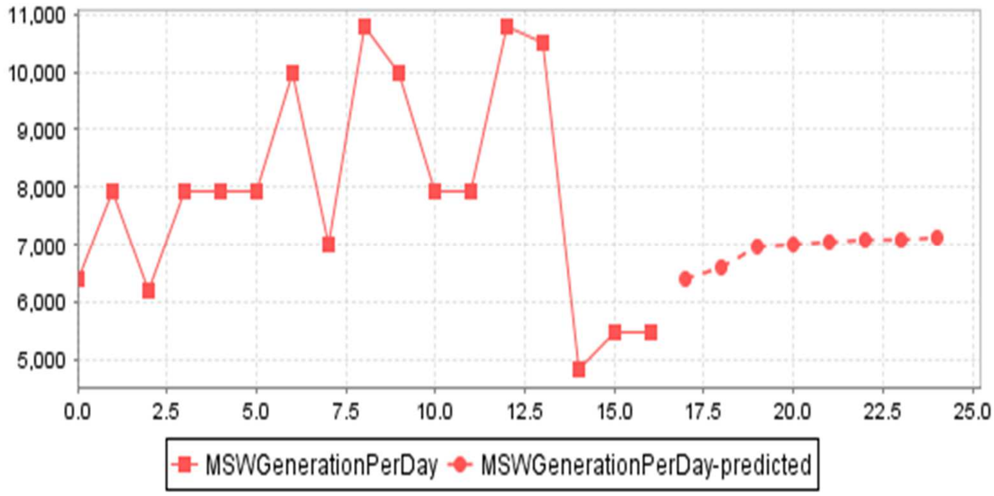

Figure 4: Random Forest- train future predict

The correlation coefficient of artificial neural network (ANN) is R=0.9923 which is better than linear regression and random forest models. ANN is conducted by designing appropriate network architecture with one neuron demand in the hidden layer. Figure 5 shows that the train predicts for targets form the year 2009 to 2017 using ANN. Table 4 shows the evaluation of training data. Figure 6 shows that the predicted MSW generation for the next 8 years using ANN.

1 Step ahead Predictions for MSW Generation per Day

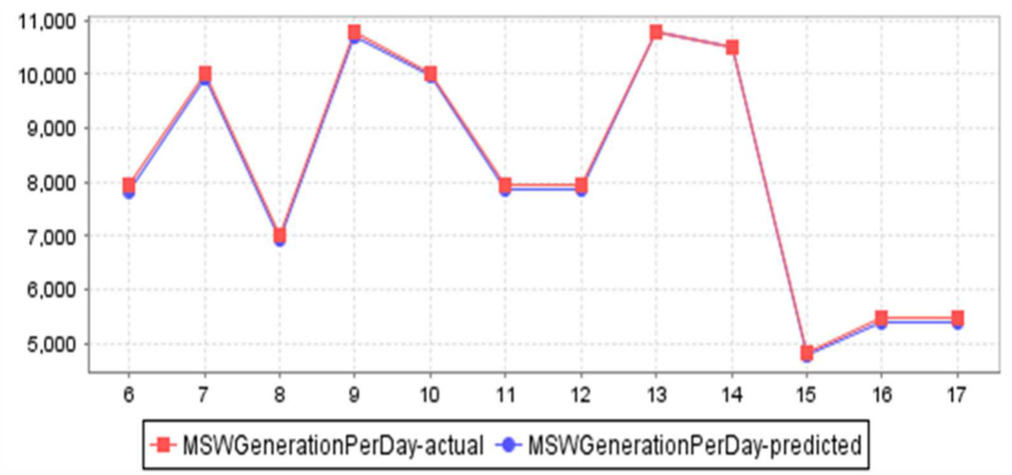

Figure5: ANN- train predict for targets data

Table 4: ANN- Evaluation on training data

Target	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead	5-step-ahead	6-step-ahead	7-step-ahead	8-step-ahead
N	12	11	10	9	8	7	6	5
Mean Absolute Error	59.6618	87.3286	177.0643	230.4398	232.6616	692.5526	926.9902	485.0177
Direction Accuracy	81.8182	80	66.6667	62.5	71.4286	50	60	75
Root Mean Squared Error	66.5912	119.491	294.3764	325.7232	277.6276	977.6214	1095.0928	622.0864

Future Forecast for: MSW Generation per Day

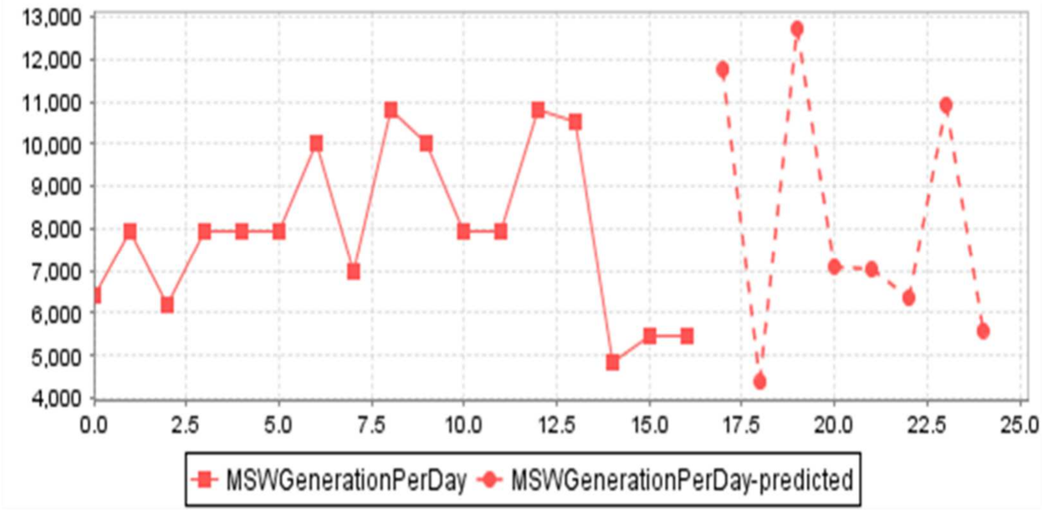

Figure 6: ANN- train future predict

5 DISCUSSION

Total population, mean household income per month, crude birth rate, crude death rate labor force, unemployment rate, GDP per capita USD, GDP growth rate, and tourist arrivals are the total variables that can be considered as affect to solid waste management generation. According to PCA and Pearson correlation analysis total population, Crude birth rate, and GDP growth rate can be considered as MSW generation influential variables. The nonlinear model, ANN provides the least root mean squared error and higher direction accuracy than random forest and linear regression analysis relatively machine learning model and linear model. To convert gathered data into meaningful information and knowledge to have insights about the future, the proposed MSW generation prediction model is shown in figure 7.

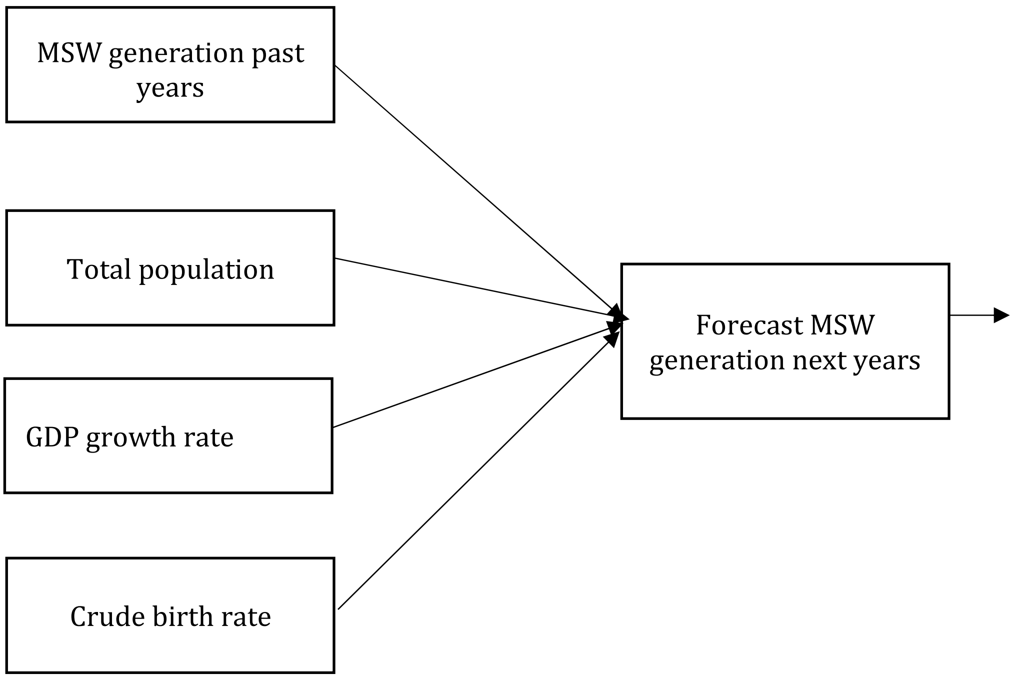

Figure 7: proposed MSW generation prediction model

6 CONCLUSION AND CONTRIBUTIONS

The appropriate forecasting model of the MSW generation is an important tool in MSW management systems and planning. The research methodology is performed using annual time series data of the amount of solid waste generation between 2001 and 2017 provided by the National Solid Waste Management Support Center and Central Environmental Authority. Total population, mean household income per month, crude birth rate, crude death rate labor force, unemployment rate, GDP per capita USD, GDP growth rate, and tourist arrivals are the total variables that can be considered as affect to solid waste management generation. Data per processing is done by replacing missing values with the mean value of the related variable. Total population, crude birth rate, and GDP growth rate are considered as the influential variables since they exist with a high coefficient. ANN, the nonlinear model, gives highly accurate results compare with regression and random analysis. The ANN multilayer perceptron is used to train and forecast municipal solid waste generation by using the Weka tool. Furthermore, one neuron in the hidden layer is enough to give a good result of the MSW prediction. This study demonstrates the influential variables affecting the amount of MSW generation and offers an appropriate model to forecast MSW in Sri Lanka. The developed model can be used for further studies on the investigation of sustainable solutions for MSW management, manage MSW facilities, manage expenditures, increase efficiency in waste management, manage the impact on waste recycling and suitable technologies for Sri Lanka MSW disposal and power generation. There was no comprehensive work dedicated to identifying influential variables that affect the amount of Municipal Solid Waste generation in Sri Lanka. The forecast Municipal Solid Waste generation in Sri Lanka using the linear regression, ANN and random forest models is a novel approach. The proposed model to forecast municipal solid waste generation in Sri Lanka is a novel approach.

REFERENCES

Admin, E. (2019). Status of Waste Management in Sri Lanka. Retrieved 19 September 2019, from https://efl.lk/status-waste-management-sri-lanka/

Ghinea, C., Drăgoi, E., Comăniţă, E., Gavrilescu, M., Câmpean, T., Curteanu, S., & Gavrilescu, M. (2016). Forecasting municipal solid waste generation using prognostic tools and regression analysis. *Journal of Environmental Management*, *182*, 80-93. doi: 10.1016/j.jenvman.2016.07.026

Marandi, F., & Ghomi, S. (2016). Time series forecasting and analysis of municipal solid waste generation in Tehran city. *2016 12Th International Conference On Industrial Engineering (ICIE)*. doi: 10.1109/induseng.2016.7519343

Sodanil, M., & Chatthong, P. (2014). Artificial neural network-based time series analysis forecasting for the amount of solid waste in Bangkok. *Ninth International Conference On Digital Information Management (ICDIM 2014)*. doi: 10.1109/icdim.2014.6991427

Sun, N., & Chungpaibulpatana, S. (2017). Development of an Appropriate Model for Forecasting Municipal Solid Waste Generation in Bangkok. *Energy Procedia*, *138*, 907-912. doi: 10.1016/j.egypro.2017.10.134