Cloud based Forecast of Municipal Solid Waste Growth using AutoRegressive Integrated Moving Average Model: A Case Study for Bengaluru

Rashmi G[1]
Department of Computer Science & Engineering
Research Scholar, RNSIT
Bengaluru, India

S Sathish Kumar K[2]
Department of Information Science & Engineering
Professor, RNSIT
Bengaluru, India

***Abstract*—Forecasting the quantity of waste growth in upcoming years is very much required for assessing the existing waste management system. In this research work, time series forecast model, ARIMA (Autoregressive Integrated Moving Average), is used to predict future waste growth from 2021 to 2028 for Bengaluru, largest city in Karnataka. Eight years old historical solid waste dataset from 2012 to 2020 is used to make predictions. This dataset is preprocessed and only time bounded variables like days, month, year and waste quantity in tons are used in this research work to obtain accurate prediction. The model is implemented in python in Google Colab free cloud's Jupyter notebook. As ARIMA is time bounded, forecast made by the model is accurate and performance of the model is evaluated using metrics such as Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Coefficient of Determination (R^2). Outcomes revealed that ARIMA (0, 1, 2) model with the lowermost RMSE (753.5742), MAD (577.4601), and MAPE (11.6484) values and the maximum R^2 (0.9788) value has a greater forecast performance. The outcomes attained from the model also showed that the total volume of yearly solid waste to be produced will rise from about 50,300 tons in 2021 to 75,600 tons in 2028.**

Keywords*—*Cloud Computing; Machine Learning; Time Series Forecasting; Waste Management System; ARIMA; Predictive Modeling

I. Introduction

Today's global technologies are popularly driven by cloud computing and machine learning. Both of these are contributing to every organization's business growth. Machine learning today facilitates users to create models which can be used to make predictions by training them to automatically learn from past data. Various machine learning approaches [7] such as supervised and unsupervised require huge amount of storage which is a challenging task for machine learning professionals. Cloud computing [9] contributes in such scenarios by providing all the resources and services required to ease the tasks. Machine learning makes brainy applications where as cloud computing provides storing and refuge services to access these applications. Cloud computing [2] thus helps in enhancing and expanding machine learning applications. Recently, these two technologies together gave birth to a new technology which is known as intelligent Cloud [15].

Cloud Computing (CC) [2] is one of the easy, flexible and quickly growing and most demanded technologies meant for delivering services requested by the users on demand over the Internet. The constraints such as cost, computational processing power, storage, analysis etc. involved in traditional approach have led to the raise of cloud computing [16]. It allows us to access various applications and data remotely without letting us install any software's explicitly in personal laptops. The various services provided by CC "are generally categorized into Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS)" [2].

Cloud computing provides all the resources required to develop, run and deploy machine learning models on demand. Machine learning needs huge amount of data storage, computing power and many servers to concurrently work on models as presented in Fig. 1.

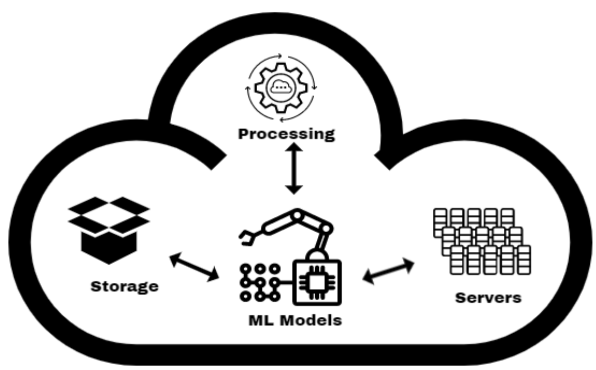


Fig. 1. Cloud Computing and Machine Learning.

Some of the key capabilities behind cloud with machine learning is that cloud's pay per use service which is good for organizations who aspire to influence machine learning competences for their business without much spending. It offers the elasticity to work with machine learning features without having advanced data science skills. It helps us in ease of testing several machine learning skills and scales up as projects go into production and demand rises. Due to these capabilities, many cloud service providers today are offering lots of machine learning services for everyone without having background expertise of Artificial Intelligence (AI) and Machine Learning (ML).

Metropolitan cities today are loaded with huge population impacting solid waste growth such as food waste, plastics, bottles, sanitary waste, construction waste etc. impacting our surrounding environment. To minimize the effect of waste growth, it is necessary to understand and analyze the speed at which solid waste is being created. Existing waste management systems do not have automated techniques incorporated for exact prediction of solid waste growth [5]. Due to lack of data, incomplete data and other challenges such as poor strategies, they are not performing efficiently. To overcome this, machine learning approach can be used [1].

II. Ease of Use

Today's municipal waste management systems [8] are inefficient to perform waste analysis and take precautionary measures due to numerous loop holes such as lack of data, lack of technical expertise, lack of efficient strategies, lack of planning etc. Inaccurate prediction may be the reason for well-known shortfalls in waste administration arrangement such as unnecessary or inadequate disposal arrangement, waste collection, landfilling and recycling divisions. Accurate forecast is very much needed in case of metropolitan cities like Bengaluru, New Delhi etc. in India as they are highly populated impacting waste growth so that an appropriate action can be taken prior. These actions are not only to develop and improve existing systems but also help to alert the public so as to encourage decrease of waste and also recycle the solid waste produced. If the waste generated is not handled well, it may affect environment and living organisms' health. Due to noteworthy influence of waste growth on the environment, waste management systems [12] resulting minimal impacts on universe and zones required to be established. "Various methods of forecasting solid waste growth [6] can be generally categorized into five key clusters: descriptive statistical approach, regression approach, material flow approach, time series approach [13] and artificial intelligence approach" [5].

In [20], authors have used ANN for forecasting waste growth in Poland. Various explanatory variables were used to reveal the impact of socio-economic and demographic variables on the amount of waste generated. Performance of the models are measured using MSE (Mean Squared Error) and R^2 metrics. The results proved that ANN is cost efficient approach in foreseeing the waste growth.

In [21], authors have developed hybrid Multilayer Perceptron (MLP) deep learning automated method to classify the waste dumped by community in the metropolitan area. Their experiments employed camera to capture images of waste and sensors to recognize the essential features. The experimental results proved that hybrid approach is capable to achieve more than 90% accuracy.

In [24], authors have carried out their research work to foresee waste growth for Mashhad city for different seasons using ANN on time series data. In [23], authors used weekly time series data to predict waste growth in Mashhad using SVM along with PCA (Principal Component Analysis). Kumar et al. used time series data which holds the yearly MSW (Municipal Solid Waste) [14] produced in New Delhi, India. Different models are used to predict waste growth and the model's performance is assessed using RMSE and the IA values.

In [17], authors presented ARIMA model to foresee solid waste growth for Arusha city, Tanzania. Monthly generated waste data for the last few years 2008 to 2013 was used to carry out the research. The result proved that ARIMA (1, 1, 1) is well suited for forecasting "in terms of MAPE, MAD and RMSE measures".

In [18], authors presented "ARIMA model to forecast solid waste growth in the Kumasi Metropolitan Assembly (KMA)". The results showed that ARIMA (1, 1, 1) is well suited for predicting solid waste growth in the KMA.

In [19], authors presented "ARIMA model to forecast healthcare waste growth for the hospitals of Garhwal region of Uttarakhand, India". The performance of the model was analyzed using R^2 value, MSE and MAE metrics and proved that ARIMA is best suited for forecast.

In [22], authors developed "ARIMA model to forecast the municipal solid waste growth of Abuja city, Nigeria. The results proved that an ARIMA (1, 1, 9) is the optimal model for forecast".

In [25], authors developed ARIMA model for forecasting amount of solid waste growth for Karur town, Tamil Nadu. Monthly based historical data was used for the year 2015 to 2017 and the results proved that ARIMA is best for prediction.

In [26], authors presented "ARIMA, Support Vector Regression (SVR) [4, 11], Grey model and Linear Regression (LR) model to forecast medical waste growth of Istanbul city, Turkey". Historical dataset used for forecast was from 1995 to 2017. Various performance metrics such as MAD, MAPE, RMSE and R^2 were used to assess the models performance. Outcome showed that ARIMA (0, 1, 2) model is well suited for waste growth forecast.

III. Machine Learning Time Series Forecast Model

Some of the key challenges in Statistics and Data Science today are time series and forecasting. A data is said to be time series data when it is bounded to time like days, months and years. When this data is used to predict future values, then it is called as Time series data.

ARIMA [3] is one of the popularly used machine learning algorithm for time series forecasting. It predicts future values using past data **(**autoregressive**,** moving average).

ARIMA Model

It is a class of linear models that uses historical data to estimate forthcoming values. ARIMA [10] enclosing three components, Auto Regressive (AR), Integrated (I) and the Moving Average (MA) contribute to the ultimate forecast. Two of the Key concepts behind these models are Stationarity and Autocorrelation.

Stationarity tells that observations/data are time independent whereas autocorrelation relates the same set of observations but across diverse timing. The different components of ARIMA are explained as follows.

Auto Regressive (AR)

This component uses autocorrelation concept, where the dependent features depend on the past values.

The general equation is:

$$Xt = \propto_1 + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_n X_{t-n} \quad (1)$$

As shown in (1), an observation X at time t, X_t, depends on $X_{t-1}, X_{t-2}, ..., X_{t-n}$, $\emptyset_1, \emptyset_2, .. \emptyset_n$ determines coefficient of lags that the model evaluates, $\propto_1$ is the intercept term, and where n is called the lag order which represents the number of previous lag samples or observations to be considered by the model.

Integrated (I)

This component of ARIMA transforms non-stationary time-series data to a stationary by accomplishing prediction on the difference between any two pair of observations instead directly on the data itself.

$$A_t = M_{t+1} - M_t \quad .....c = 1$$
$$B_t = A_{t+1} - A_t \quad .....c = 2 \quad (2)$$

As shown in (2), Differencing tasks which can be achieved many times levels (**M →A** *and* **A →B**), depends on the hyper parameter **c** that is set while training the ARIMA model.

Moving Average(MA)

This component performs some kind of aggregation on the historical time series data in terms of residual error epsilon (ε) thus reducing noise in the data.

$$X_t = \propto_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \cdots + \omega_n \varepsilon_{t-n} + \varepsilon_t \quad (3)$$

The terms ε indicate the residual errors from the aggregation operation as shown in (3) and n is another hyper parameter that specifies the time window for the moving average's residual error. X_t depends on the lagged forecast errors.

Following is the generic steps followed for ARIMA.

Step 1: Visualization of Time Series Data

Step 2: If data is non stationary, then convert it to stationary

Step 3: Make the Correlation and AutoCorrelation graphs

Step 4: Build the model using data

Step 5: Make predictions using the model

IV. Results and Discussions

The ARIMA model used here for waste growth forecast is implemented in python. Jupyter notebook from Google Colab which is a free cloud service is used for the implementation.

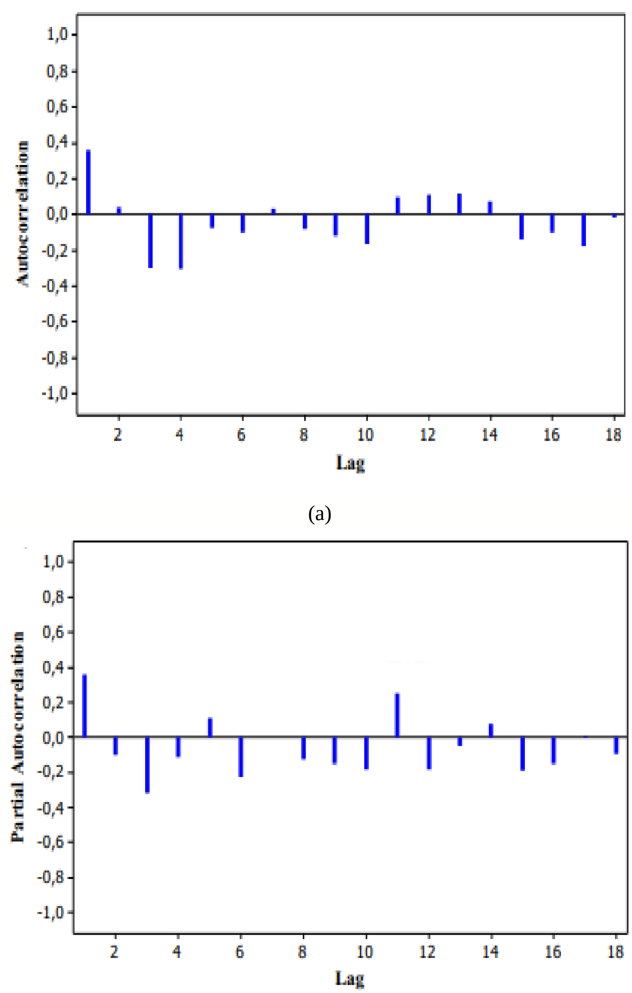
Autocorrelation (AC) and Partial Autocorrelation (PAC) graphs shown in Fig. 2 (a) and (b) respectively are used to analyse and forecast future waste growth. They basically indicate how many days of previous data need to be considered to forecast future values which is known as lags. To calculate AR, three values, p, d, q need to be chosen, where, p represents AR model lags, d represents Differencing, q represents MA lags.

ARIMA model used here is to predict future trends of waste growth. The model needs stationary data to determine AR and MA components. Since the data used in this research work is non stationary, d=1, first order differencing was done.

Autocorrelation graph in Fig. 2(a) showed that the series is stationary after first differencing. The arrangements of the AC and PAC graphs of the differenced series were examined for the initial computation of autoregressive (p) and moving average orders (q) in ARMA (p, q) model.

For an AR model, the number of nonzero partial autocorrelations gives the most extreme lag of x that is used as a predictor.

Once, AC and PAC computation and analysis was done, ARIMA model was invoked on the dataset which gave the results shown in Table I and the forecast graph obtained is shown in Fig. 3.



(a)

(b)

Fig. 2. (a) Autocorrelation and (b) Partial Autocorrelation after First Differencing .

Equations used to measure the performance of the model are RMSE shown in (4), MAD shown in (5), MAPE% shown in (6), and R^2 shown in (7).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Wreal_i - Wexpected_i)^2}{n}} \quad (4)$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|Wreal_i - Wexpected_i| \quad (5)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Wreal_i - Wexpected_i|}{|Wreal_i|} \times 100\% \quad (6)$$

$$R^2 = \left[\frac{\sum_{i=1}^{n}(Wreal_i - \overline{Wreal_i})(Wexpected_i - \overline{Wexpected_i})}{\sum_{i=1}^{n}(Wreal_i - \overline{Wreal_i})^2 \times \sum_{i=1}^{n}(Wexpected_i - \overline{Wexpected_i})^2}\right]^2 \quad (7)$$

Where, $Wreal_i$ and $Wexpected_i$ denote the real and expected value of i^{th} data point value, respectively. $\overline{Wreal_i}$ and $\overline{Wexpected_i}$ are the average of the real and expected value of i^{th} data point value. Also, n indicates the total number of data values. The performance of the model was measured by computing R^2 value. It accepts values between 0 and 1, and values very close to 1 which indicates better fitting.

TABLE I. ARIMA MODEL PERFORMANCE

Model	RMSE	MAD	MAPE	R^2
ARIMA (1,0,2)	854.2914	635.4722	11.5771	0.9767
ARIMA (0,1,0)	960.1165	713.0000	13.5741	0.9690
ARIMA (1,2,1)	1045.2257	777.5254	15.3163	0.9683
ARIMA (0,1,2)	753.5742	577.4601	11.6484	0.9788

Various ARIMA models are also made for selecting the model and their performance analysis is done using various metrics. As shown in Table I, the ARIMA (0, 1, 2) model has the highest R^2 (0.9788) and lowest RMSE (753.5742), MAD (577.4601), and MAPE (11.6484) and hence it is chosen as best model.

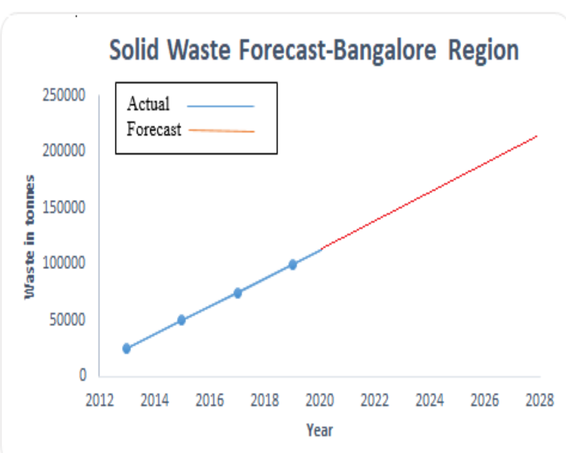


Fig. 3. Solid Waste Growth Forecast using ARIMA.

V. CONCLUSION

It is very important to contribute and enhance the existing condition and scenarios of waste management in the crowded smart city, Bengaluru which can only be attained with precise waste assessment. Hence, the goal of this research work is to deliver an appropriate model to assess the quantity of waste produced. In this context, ARIMA (0, 1, 2) was chosen as the best model and used to forecast the waste growth of Bengaluru based on eight years of historical data. The outcomes of this research work can help waste management authorities to develop a reliable waste forecast model, which can be a significant foundation of information for Bengaluru. In addition, previous data about the volume of waste produced can be used for both the planning and design of future services.

VI. LIMITATIONS AND FUTURE RESEARCH WORK

This research work targets to the development and improvement of waste management practices in smart cities through forecasting waste generation. It shows the development of a systematic process where time based factors affecting waste generation in smart cities have been determined to study and forecast waste growth. Unavailability of the continuous waste data and also socio-economic and demographic variables affecting solid waste generation makes it difficult to foresee solid waste growth for the developing countries like India.

The research work can be extended in the future by incorporating more input features, more socio-economic parameters. Other ML, AI or deep learning techniques can be used in future to handle complex scenarios and to achieve better accuracy.

REFERENCES

[1] X. Cuong Nguyen, T. Thanh Huyen Nguyen, D. Duong La, Gopalakrishnan Kumar, Eldon R. Rene, D. Duc Nguyen, S. Woong Chang, W. Jin Chung, X. Hoan Nguyen, V. Khanh Nguyen, Development of machine learning - based models to forecast solid waste generation in residential areas: A case study from Vietnam, Resources, Conservation and Recycling, Volume 167, 2021, 105381, ISSN 0921-3449, https://doi.org/10.1016/j.resconrec.2020.105381.

[2] Butt, Umer Ahmed, Muhammad Mehmood, Syed Bilal Hussain Shah, Rashid Amin, M. Waqas Shaukat, Syed Mohsan Raza, Doug Young Suh, and Md. Jalil Piran. 2020. "A Review of Machine Learning Algorithms for Cloud Computing Security" *Electronics* 9, no. 9: 1379. https://doi.org/10.3390/electronics9091379Ceylan Z, Bulkan S, Elevli S. Prediction of medical waste generation using SVR, GM (1,1) and ARIMA models: a case study for megacity Istanbul. J Environ Health Sci Eng. 2020 Jun 19;18(2):687-697.

[3] Ş. T. Özçelik and F. Boray Tek, "Forecasting and Analysis of Domestic Solid Waste Generation in Districts of Istanbul with Support Vector Regression," *2020 5th International Conference on Computer Science and Engineering (UBMK)*, Diyarbakir, Turkey, 2020, pp. 366-371. doi: 10.1109/UBMK50275.2020.9219368.

[4] Soni, U., Roy, A., Verma, A. *et al.* Forecasting municipal solid waste generation using artificial intelligence models—a case study in India. *SN Appl. Sci.* **1,** 162 (2019). https://doi.org/10.1007/s42452-018-0157-x

[5] D. M. S. H. Dissanayaka and S. Vasanthapriyan, "Forecast Municipal Solid Waste Generation in Sri Lanka," *2019 International Conference on Advancements in Computing (ICAC)*, Malabe, Sri Lanka, 2019, pp. 210-215, doi: 10.1109/ICAC49085.2019.9103421.

[6] Miyuru Kannangara, Rahul Dua, Leila Ahmadi, Farid Bensebaa, Modeling and prediction of regional municipal solid waste generation

and diversion in Canada using machine learning approaches, Waste Management, Volume 74, 2018, pp. 3-15, ISSN 0956-053X.

[7] Maya Chavan, T . R. Pattanshetti, “Survey on Municipal Waste Collection Management in Smart City”, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 01, Jan-2018, p-ISSN: 2395-0072 e-ISSN: 2395-0056.

[8] M. Talha, A. Upadhyay, R. Shamim and M. S. Beg, "A cloud integrated wireless garbage management system for smart cities," *2017 International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)*, Aligarh, 2017, pp. 175-179, doi: 10.1109/MSPCT.2017.8363999.

[9] Chauhan, Ankur & Singh, Amol. (2017). An ARIMA model for the forecasting of healthcare waste generation in the Garhwal region of Uttarakhand, India. International Journal of Services Operations and Informatics. 8. 10.1504/IJSOI.2017.086587.

[10] C. Dai, Y.P. Li, G.H. Huang, A two-stage support-vector-regression optimization model for municipal solid waste management – A case study of Beijing, China, Journal of Environmental Management, Volume 92, Issue 12, 2011, pp. 3023-3037.

[11] S. Lebersorger, P. Beigl, Municipal solid waste generation in municipalities: Quantifying impacts of household structure, commercial waste and domestic fuel, Waste Management, Volume 31, Issues 9–10, 2011, pp. 1907-1915, ISSN 0956-053X, https://doi.org/10.1016/j.wasman.2011.05.016.

[12] Mwenda A, Kuznetsov D, Mirau S, Time series forecasting of solid waste generation in Arusha city- Tanzania. Mathematical Theory and Modeling, Vol.4, No.8, 2014, pp. 29-39, ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online).

[13] Hanandeh, Ali & Abbasi, Mariam. (2016). Forecasting municipal solid waste generation using artificial intelligence modelling approaches. Waste Management. 56. 10.1016/j.wasman.2016.05.018.

[14] Aazam, Mohammad et al. “Cloud-based smart waste management for smart cities.” 2016 IEEE 21st International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD) (2016): pp. 188-193.

[15] C. Ji, Y. Li, W. Qiu, U. Awada and K. Li, "Big Data Processing in Cloud Computing Environments," *2012 12th International Symposium on Pervasive Systems, Algorithms and Networks*, San Marcos, TX, 2012, pp. 17-23.

[16] Amon Mwenda , Dmitry Kuznetsov , Silas Mirau, “Time Series Forecasting of Solid Waste Generation in Arusha City – Tanzania”, Mathematical Theory and Modeling www.iiste.org ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online) Vol.4, No.8, 2014.

[17] Owusu-Sekyere, Ebenezer. (2013). Forecasting and planning for solid waste generation in the Kumasi metropolitan area of Ghana: An ARIMA time series approach. International Journal of Sciences. 2. 69-83.

[18] Chauhan, Ankur & Singh, Amol. (2017). An ARIMA model for the forecasting of healthcare waste generation in the Garhwal region of Uttarakhand, India. International Journal of Services Operations and Informatics. September 13, 2017pp 352-366 8. 10.1504/IJSOI.2017.086587.

[19] Kulisz, M.; Kujawska, J. Prediction of municipal waste generation in Poland using neural network modeling. Sustainability 2020, 12, 10088.

[20] Chu, Yinghao & Huang, Chen & Xie, Xiaodan & Tan, Bohai & Kamal, Shyam & Xiong, Xiaogang. (2018). Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling. Computational Intelligence and Neuroscience. 2018. 1-9. 10.1155/2018/5060857.

[21] U. A. Dodo, E. C. Ashigwuike and J. N. Emechebe, "Municipal Solid Waste Generation Forecast using an ARIMA Model: A Focus on Abuja City, Nigeria," *2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)*, 2022, pp. 1-5, doi: 10.1109/NIGERCON54645.2022.9803108.

[22] Noori, R, Khakpour, A, Omidvar, B, et al. (2010b) Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. Expert Systems with Applications 37: 5856–5862.

[23] Jalili, Mojtaba & Noori, Rahimullah. (2007). Prediction of Municipal Solid Waste Generation by Use of Artificial Neural Network: A Case Study of Mashhad. International Journal of Environmental Research (ISSN: 1735-6865) Vol 2 Num 1. 2.

[24] Vignesh.C, Rameshkumar.M , Dr.S.Anand Kumar Varma, Time Series Forecasting of Solid Waste Generation in Karur City -Tamil Nadu, SSRG International Journal of Civil Engineering (SSRG - IJCE) – Volume 5 Issue 2 - February 2018 ISSN: 2348 – 8352 .

[25] Ceylan, Z., Bulkan, S. & Elevli, S. Prediction of medical waste generation using SVR, GM (1,1) and ARIMA models: a case study for megacity Istanbul. *J Environ Health Sci Engineer* **18,** 687–697 (2020). https://doi.org/10.1007/s40201-020-00495-8.