PalArch's Journal of Archaeology of Egypt / Egyptology

USE OF MACHINE LEARNING IN PREDICTING THE GENERATION OF SOLID WASTE

[1]Tushar Rathod, [2]Manoj Hudnurkar, [3]Suhas Ambekar

Symbiosis Centre for Management and Human Resources Development,

Symbiosis International (Deemed University), Pune, India

Email: suhas_ambekar@scmhrd.edu[3]

Tushar Rathod, Manoj Hudnurkar, Suhas Ambekar: Use of Machine Learning in Predicting the Generation of Solid Waste -- Palarch's Journal Of Archaeology Of Egypt/Egyptology 17(6). ISSN 1567-214x

Keywords: Machine Learning, Solid Waste Management, Ensemble Techniques, Bagging and Boosting.

ABSTRACT

This study is focussed on the betterment of municipal solid waste management and it has an intention to build machine learning models so that we municipalities can predict solid waste generation with the help of demographic and socio-economic variables. These machine learning models are built by placing domestic municipal solid waste quantities together with demographic and socio-economic variables of 200 regions around the area of Akola city, Maharashtra. We are using four machine learning algorithms viz., decision tree, multiple linear regression, random forest regression, and XGBoost regression algorithms to build predictive models. We needed the data on solid waste generated in those regions and it has been collected from local authorities and the website of Akola Municipal Corporation. From Indian Census and local authorities, we got data for demographic and socio-economic variables. The data pre-processing required in the initial stages of the model building was performed in Python and Microsoft Excel. After this, the models were built with machine learning algorithms and the results of this study showed that these algorithms can be used to estimate solid waste generated in the areas depending on the variables with a good prediction performance. The random forest regression model has the best performance, describing the minimum difference between actual and predicted values of solid waste generated. In this study, we have following the approach where we demonstrate the feasibility of building machine learning models that help in regional waste planning using data collection, data cleaning, data pre-processing, and modelling of available data.

1. Introduction

There are a few issues that are given less attention and solid waste management (SWM) is one of those municipalities' programs. In every region of India, the amount of municipal solid waste generated is increasing rapidly due to population growth and urbanization. This issue is presenting many challenges to authorities in solid waste management. Due to limited financial resources and other difficulties, this issue needed serious attention and hence the Municipal Solid Waste Rules, 2000. It is applied to every municipal authority in India and these rules were mandatory for the administrative authorities to undertake responsibilities for all activities relating to municipal solid waste management. Solid waste management activities can create new jobs and help in boosting the economy. There are many use-cases of this solid waste like it can be used to produce electricity while mitigating the environmental impacts.

A survey of MSWM practices in Indian municipality corporations suggests that the underestimation of generation rates is a major concern in India and therefore, other problems like lack of use of technology, lack of reliable information, and underestimating resource requirements. Before the corona outbreak in the year 2020, the Indian economy was growing at a good rate. India is a developing country and there is a vast history of developing countries that are likely to help in growing the rates of solid waste generation.

The present approaches for MSW management are resulting in inefficient utilization of time and resources because of the uncertainty in solid waste generation. There are a lot of modern technologies and tools like digitization, planting sensors, and solid waste optimized disposal systems that can be used to do efficient waste management mainly towards optimal utilization of time and resources. But all such waste management efforts are facing issues because of many challenges. All this starts with the incapability of local authorities handling solid waste in a better way. The main obstacle to this is that waste management is done at disaggregate levels and it involves different channels, making the data collection and compilation difficult. Many factors are contributing to this like a rapidly increasing population, building new factories, shops, and restaurants, frequent public gatherings, etc. Hence judging the amount of waste generation has become a very difficult task resulting in inefficient waste management.

The objective of this study is use the data of the solid waste generated in the region and the variables impacting the generation in order to predict the future waste generation. This paper intends to provide an overview of the methodology discussing how can the predictive models can be used for solid waste management. The research paper has been divided into eight parts, namely, Literature review, Data collection and pre-processing, Machine learning approaches, Results, Discussions, Conclusions, Managerial implications and limitations.

2. Literature Review:

This study is based on the use of data analytics techniques like descriptive analytics and predictive analytics which will streamline the processes in solid

waste management at the municipality level. There are various research papers published on similar lines and all these research papers formed the base for this study.In India, there are many things that are helping to accelerate the generation of solid waste in cities multiple times more than they did 10 years back. There is a huge difference between the quantity and composition of solid waste generated in the western countries and India (Shannigrahi et al., 1997). Some research papers have mentioned that there is a large fraction (40-60%) of solid waste in urban areas containing compostable materials (Sharholy et al. 2008). In rural households a relative percentage of solid waste is organic, but that is also decreasing because people in rural parts of India have started using materials that create more waste.

After analyzing a few factors like population growth, a number of shops, clinics, schools, etc. in the region there are patterns found in solid waste generation. Advanced topics in machine learning are helping in estimating the amount of solid waste generated. These topics include machine learning, deep learning, forecasting, neural networks, etc. Artificial intelligence methods such as artificial neural networks (ANNs) have been used to model solid waste generation data that is in time series pattern (Abbasi and El Hanandeh, 2016). A model was built which was using SARIMA i.e., seasonal autoregressive moving average modelling technique to forecast daily and monthly residential waste data for the long term (Navarro-Esbrí et al. 2002).

In this study, we are trying to judge or approximate the amount of solid waste to be generated in the region. This is very essential in solid waste management. It is very certain that the solid waste generation prediction can be done with the help of many machine learning approaches that we often use to build predictive models on different types of data. In this research, we will be trying to establish a statistical relationship between the data collected on solid waste generated in a particular region with other explanatory variables taken from the same region.

3. Research Methodology:

3.1. Description of the study area:

We will be studying the regions of Akola. It is a city in the Vidarbha region in Maharashtra state. It is the third-largest city in the Vidarbha region. It is located about 580 km east of Mumbai, the state capital, and 250 km away from Nagpur which is also known as the second capital of Maharashtra. The city is located at an altitude of 926 ft to 1035 ft above from the sea level. It is placed at longitude 77.07° East and latitude 20.7° North. According to the 2011 India census data, a population of Akola was around 426,815 and spread across 124 sq.km of an area in the region. Recently some data on the population has been released and the population of Akola has increased by 5.8% to 450,707.

Akola is a rapidly growing city in various avenues thus it generates an enormous quantity of waste on an everyday basis, thus the job of the corporation becoming challenging. The daily generation of solid waste of Akola is on average 130-135 million tons. It is from various sources like households, industrial areas, hospitals and schools, market places, and so on.

3.2. Data availability:

The data required for the prediction of solid waste generation comes from various data sources. We have defined two different data sources viz., Municipal data, and Census data. The municipal authority of Akola collects the amount of solid waste generated. This type of data collection is done for every region in Akola. The data is available on the Akola Municipal Corporation website. This amount has been collected on per day basis and also it is the dependent variable in this study. And the unit of this data is Million tons. The collected municipal solid waste observations have been collected area-wise because we want to analyse and predict the amount of solid waste accordingly. We have collected the data for 200 areas in the Akola city.

3.3. Demographic and Socio-economic variables:

The most useful variable that can help in determining the total solid waste generates in the region is the population of that region. But in this study, the population may not be the perfect variable to understand the generation of solid waste. Hence we are using Population Density in people per sq.km. After analysing the dataset we can see that population density is directly correlated with the amount of solid waste generated. The other factors that are important in measuring the amount of solid waste generated come from the behavioural nature and quality of life of people in the region. These factors have most often been contributing to an increase in the amount of waste generation. Some of these factors are the average income level, consumer expenditure, and purchasing power indices, and these variables have a positive correlation with the amount of waste generation. There are a few more factors that strongly influence the solid waste generation. These include the number of households, schools, hospitals, clinics, shops, malls. All these factors are positively correlated with the amount of waste generation. This information can be collected by doing some interviews in the local area. There are a few variables that can impact the amount of waste generation but data collection for those variables is not available at the municipal level considered in this study. Some of those variables are the type of occupation, kind of employment, tourism arrivals and expenditure, etc.

3.4. Data Collection:

All possible variables needed for the study are collected, for example, the amount of solid waste generated, population density (number of inhabitants per square km area), dwelling indicator (Number of households, Number of Shops, Hospitals, and Schools) and economic indicator (Average Income per household). Table 1 contains a description of the variables, their types, and their units. There was a study conducted by an organization The Resources and Livelihoods Group, Prayas, Pune on Municipal Solid Waste Management in Akola. We have used the data from this study as well.

All the variables required for the machine learning modelling part in this study are very significant collected from different sources at the municipal level. One of the most important variables i.e., population density collection is not that

difficult. The best data source from where this data can be taken from is the Indian census program. The use of census data guarantees the availability of many parameters for almost all regions in Akola. There is a lot of significance of recycling, yet the variables related to such policies and programs are not included in this study and that is because the data for these variables are not available on the municipal level. For the remaining variables, the data is collected from local authorities and the Akola Municipal Corporation website. Some of the data is collected through personal visits or people interviews. For example, we came to know about a number of malls and schools from local people. Thus all the required data for this study were collected and some of the discrepancies were resolved while data handling or exploratory data analysis.

Table 1: Variables used in the machine learning algorithm**.**

No	Name of the variable	Symbol	Type of Variable	Measure
1.	Solid Waste	Y	Dependent	Tonnes/day
2.	Population Density	Y_Pd	Independent	People/sq.km
3.	Avg. Income Level	Y_AIL	Independent	Rs. in '000
4.	No of Flats	Y_F	Independent	Number
5.	No of Bungalows	Y_B	Independent	Number
6.	No of Houses	Y_H	Independent	Number
7.	No of Shops	Y_SH	Independent	Number
8.	No of Malls	Y_M	Independent	Number
9.	No of Schools	Y_SL	Independent	Number
10.	No of Hospitals/clinics	Y_H	Independent	Number

3.5. Data pre-processing:

Fig 1 sketches the whole process followed in this study showcasing the overall methodology of the research. After the data collected from various sources, it needs to be cleaned and get ready for machine learning modelling. The data pre-processing step involves various techniques for data handling. After the in-depth analysis, the collected data was transformed into variables suitable for modelling. We used Excel and Python programming language for all the pre-processing steps. This process includes loading the data into the pandas data frame, combining the data collected from different resources, cleaning the data, handling outliers, handling missing values, and scaling the data. After performing all these techniques the data gets ready to feed into machine learning algorithms for building the predictive models.

In data pre-processing, some basic checks need to do done like handling outliers, missing values, etc. For checking if there are any outliers in the dataset we have used Interquartile range (IQR) filtering (S. Kanan and S. Krishnan, 2015). There can be various types of outliers present in the dataset. Some outliers can be artificial and some of them can be genuine. The reporting discrepancies from the municipality authorities or incorrect estimates can generate outliers in the dataset. The outlier in the dataset can be found with the help of upper and lower limits which are derived from the first and third quartiles of the data. An interquartile range i.e., IQR is the difference between Q1 and Q3. The upper limit is calculated by Q3+(1.5*IQR) and the lower limit

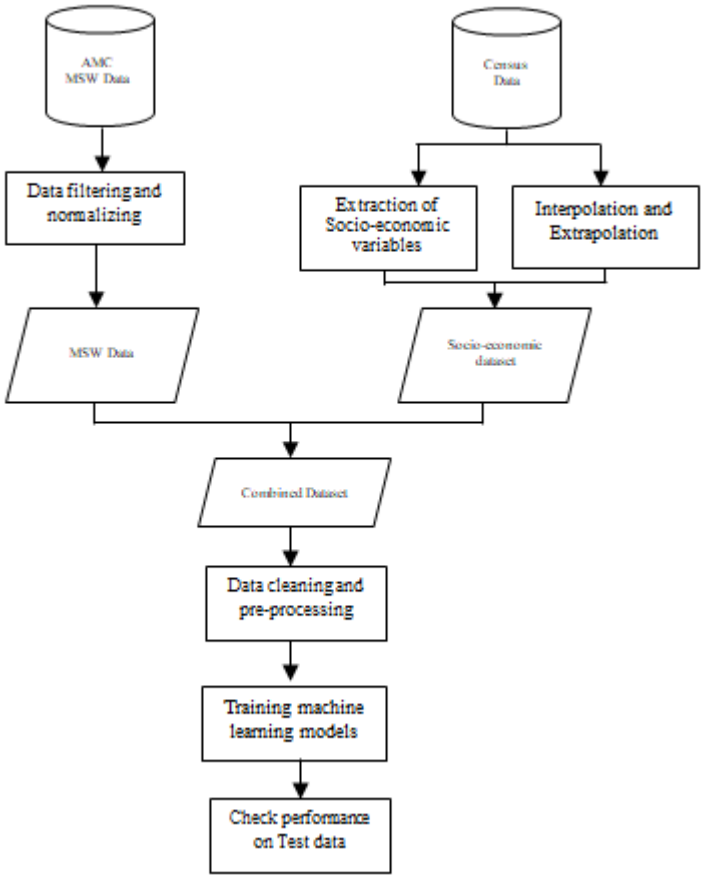
is Q1-(1.5*IQR). If the data points stand outside of these limits, those are considered as outliers. In our dataset, there were only a few outliers and laying just outside of the data range. Also, these data points were genuine as in some regions the solid waste generated was in big numbers compared to other regions. Hence we kept these outliers in the dataset.

Another check is for the missing data. The reason for missing values can be different for different columns from the same dataset. There are mainly three reasons for missing data namely MCAR, MAR, and MNAR. MCAR i.e., Missing completely at random, occurs when someone just forgets to enter the data in a particular cell, or someone is just lazy to fill it, nothing serious here. Second is MAR i.e., Missing at random, this occurs when missing-ness is not random. For example, participants accidentally omitting an answer to a particular question on a questionnaire. Lastly, the third reason for missing data is MNAR i.e., Missing not at random. The missing-ness depends on the value of that variable. To understand this let us take an example, participants with certain characteristics skipping a certain question on a questionnaire deliberately. There are various ways to handle missing data. The easiest of them is to delete the rows with missing values. But, this approach has an issue of the possibility of losing some important data. Hence the ideal way of handling missing values is imputation. This means missing values are substituted by others. The imputation can be done by various methods. Two famous imputation methods are used widely in the industry viz., KNN imputation, and Multiple Imputation by Chained Equations, which is commonly known as MICE. For KNN imputation, we have a scikit-learn library named KNNImputer. This library tries to impute the values based on the concept of the KNearestNeighbour algorithm (Ruilin Pan and Tingsheng Yang, 2015). On the other hand, the MICE algorithm runs multiple regression models and each missing value is modelled conditionally depending on the of the non-missing values. If we compare these two techniques, the MICE imputation technique performs better than the KNN imputation technique in nominal missing imputation but performs worse in the imputation of continuous and ordinal variables. In our dataset we have continuous variables hence we have used KNN imputation.

After all these data pre-processing, the data looks good to apply machine learning algorithms. But there is an issue with the magnitude of the different columns. For example, Population density columns have figures in hundreds and that of column No of Hospitals is in the range of 0 to 3. This imbalanced characteristics of the data magnitude can affect the accuracy of the model building. To resolve this issue it is very important to do the scaling of the dataset. We have two options viz., normalization and standardization. In the normalization scaling technique, the values are shifted and rescaled to draw in the range between 0 and 1. This technique is also called as Min-Max scaling. Another scaling technique is standardization. In standardization, the values are transformed so that they come around the mean with a unit standard deviation.

Hence the mean of the attribute becomes zero and then the final data distribution has a unit standard deviation.

Fig 1. Research Methodology



3.6. Machine learning approaches:

There are two main types of machine learning problems viz., regression and classification. In this study, we have solid waste generated in million tons as a dependent variable and it is a continuous variable. Hence, it is a regression problem. There are multiple machine learning approaches to tackle regression problems. Also, this is a supervised machine learning as we know the dependent variable. There are many supervised regression approaches to solve this type of problem statement. One of the basic machine learning techniques is Multiple linear regression, which is usually known as MLR. It behaves the same as a linear regression it's just this algorithm has several independent variables. This algorithm is a supervised machine learning algorithm and its main objective is to construct a model that establishes a linear relationship between dependent and independent variables (Gülden and Neşe, 2013). In other terms, multiple linear regression tries to find out how the dependent variable changes with the changes in independent variables. This model is simple to build and the interpretation of the output of this algorithm is straightforward.

Another machine learning algorithm that we have used is the decision tree (Lingjian, Songsong and Sophia, 2017). The decision tree based on a flow chart like tree structure to predict the dependent variable based on the independent. This algorithm falls under the category of supervised learning machine learning and decision tree regression is used for the continuous dependent problem. The decision tree regression first observes the features in data and trains a model in the structure of a tree trying to predict a dependent variable. The creation of the tree structure in this algorithm has a few steps. The nodes and branches of the tree are decided with the help of the values of entropy and information gain. These values are calculated for each variable and variable with the highest value for information gain is picked as a root node. And this process is performed repeatedly until the whole tree structure is completed.

There are some ensemble machine learning techniques (Thomas et al. 2000) viz., boosting, and bagging. These algorithms perform better than the above algorithms. The ensemble machine learning algorithm that reduces bias and variance is known as boosting. This boosting algorithm tries to convert weak learners into strong learners. The example of a boosting algorithm is the XGBoost machine learning algorithm. On the other hand, bagging is an ensemble technique that makes predictions by combining the results of multiple classifiers model. The example of a bagging machine learning method is a random forest. This machine-learning algorithm constructs multiple decision trees and tries to solve the regression or classification problem by involving results from those decision trees which run in parallel with each other (Leo et al. 2001). These trees do not interact with each other while building the whole structure of the random forest. The random forest machine learning algorithm combines the results of every decision tree and aggregates them with some modifications. In the decision tree algorithm, some hypermeters can be passed to it while building up the model and these values ensure fair use of all predictive features by not relying heavily on a few individual features. While generating the splits in the random forest each tree draws a random sample from the original data set and it helps in feeding randomness to the model that helps in preventing the overfitting.

In the boosting approach, we have used the gradient boosting machine learning technique for solving the regression problem in this study. XGBoost is a perfect example of gradient boosting and it is an open-source library. This name stands for eXtreme Gradient Boosting. This algorithm produces a prediction by ensembling the weak decision trees. This algorithm is very well known for its accuracy, feasibility, and efficiency (R. Santhanam and N. Uzir, 2017). It also has extra features for computing feature importance and doing cross-validations. In the boosting method, one model is built and the result of this model is passed onto another model. The first model tries to fit the new model to new residuals of the previous predictions. This way trying to minimize the loss in the latest predictions. So, in the end, you are updating the model using gradient descent. XGBoost Regressor takes different parameters as arguments. All these machine learning algorithms can be penalized through

both L1 and L2 regularization. This regularization helps in controlling the overfitting of the model which turns out to be very useful in building genuine and good performing machine learning models. In every iteration, subsampling will occur at least once. There is one important parameter that has been used quite often in this algorithm and that is learning_rate and this number ranges from 0 to 1. It helps in preventing overfitting (Ying et al. 2019).

There are libraries present for all these machine learning algorithms. We just have to load them in the python programming language and perform the modelling part. We have used the train and test splitting strategy in this study and there is a scikit-learn library available for this splitting. It takes many parameters like a dependent variable, independent variables, test size or train size, random state, etc. This splitting is random and random_state parameter is used to make sure the reproducibility of the split. In this, we have divided the pre-processed dataset into two parts in the 70:30 ratio which are named as training dataset and testing dataset respectively (Yun and Goodacre, 2018). For building the machine learning model we used the training dataset and then it has been testing with the testing dataset to check the performance of the model. This process required a few iterations until we get acceptable results. These iterations were based on hyper-parameter tunings like learning_rate in XGBoost algorithm and random forest algorithm or number of decision trees in the random forest algorithm.

We have built different supervised machine learning regression models using all these four algorithms with few iterations. All the machine learning predictive models are successful in predicting the dependent variable based on all the independent variables.

4. Results:

The performance of the supervised machine learning models on a regression problem is evaluated with the help of metrics. These metrics compare the actual quantities of solid waste with that of the predicted ones for the regions. Some of these metrics are R^2 or Coefficient of Determination, Adjusted R^2, Mean Squared Error, and Root Mean Squared Error.

We trained the above four machine learning models on the pre-processed data. To compare the performance of these machine learning algorithms, we have used Root Mean Squared Error i.e., RMSE and Adjusted R square value. In the study, we have the actual values of waste generated for the regions in the testing dataset. Using the built predictive models we have estimated the amount of solid waste for these same regions with the help of independent variables. The difference between these values is calculated and each value is squared. Then the average value of these numbers is calculated. The small error should be also penalized, hence the squaring of the differences. And finally we take the square root of this value which is known as RMSE. There is a skicit-learn library in a python programming language named metrics which has the direct function to calculate root mean squared error. The researchers are also interested in how much variance in the dependent variable can be explained by the independent variables. And this can be understood with the help of the

coefficient of determination (denoted by R^2). It is generally called as R squared. The value of the coefficient of determination lies between 0 and 1. A good machine learning model has this value close to 1. It means that the independent variables used in the model are able to explain the variation in the dependent variable very well. Sometimes coefficient of determination can be misleading because every time a new independent variable added, the value of the coefficient of determination increases even if that variable is not adding an explanation to the variance in the dependent variable. Hence there is a new metric to check the performance of the machine learning model that is adjusted R^2.

The collected data has been divided into train and test. After applying models on the training dataset we have calculated RMSE and Adjusted R square values are calculated. The summary of the Prediction Performance of the different models is given in Table 2. The model with a minimum value of the RMSE and maximum value for adjusted R^2 is picked as a more reliable machine learning algorithm. The values for all four algorithms are mentioned in Table 2. The value for the RMSE for a good performing model should be as little as possible on the testing dataset and the adjusted R square value should be as high as possible. In the machine learning models we have used in this study, we have got RMSE values within the range of 0.15 to 0.18. These values are very close to each other, hence the differentiating factor is the values of adjusted R square. The multiple regression algorithm and decision tree algorithm have adjusted R square of 59% and 50% respectively. The algorithms using ensemble machine learning technique viz., random forest regression, and XGBoost regression have performed better than multiple linear regression and decision tree regression. Based on values from Table 2, the random forest regression has given better values compared to other algorithms in both the metrics. The value of RMSE for the Random Forest regression model has been achieved equal to 0.1544. It means there is very little difference between actual and predicted values. Also, an adjusted R^2 value of 0.6758 i.e., 67.58% variation in the test dataset can be explained by independent variables.

Table 2: Results of machine learning algorithms.

Machine Learning Model	RMSE	Adjusted R2 value
Multiple Linear Regression	0.1598	0.5954
Decision Tree	0.1747	0.5044
Random Forest	0.1544	0.6758
XGBoost Regression	0.1528	0.6236

5. Discussions:

The results of this study support the findings of (Kannangara et al., 2018; Soni et al., 2019), relating to modelling and predicting the municipal solid waste generation. Some of the important factors contributing to the generation of this solid waste are population density, the average income level of the people living in the region, the number of shops, malls, schools, hospitals, etc. in the region. Several researchers have developed predictive models for the solid waste generation (Mahmood et al., 2018; Pan et al., 2019), while many others

have analyzed the factors impacting to the waste generation (Chhay et al., 2018; Liu et al., 2019; Rybová et al., 2018).

Unfortunately, due to the differences in geography, social-economic factors of the regions of the world, it is really difficult to find one fit all predictive model for this problem statement. In this study, we have tried to concentrate on the factors which are very generic and can apply to most of the regions in the world. The study focuses on the very common variables or factors that contribute to the solid waste generation. And having a good amount of data for these variables for a region, the predictive models can generate very promising numbers. Using this, the local administration can deploy its resources accordingly in solid waste management.

6. Managerial Implications:

In this study, we have seen how to predict solid waste generation so that it will become easy for its management. Various recommendations are applicable for waste management but minimizing solid waste generation would be the first preference. The first recommendation in this direction would be to try to manage the materials efficiently and sensibly so that it doesn't enter the life-cycle of waste. For doing this, the assent of reusable items would be the best alternative. The most helpful implementation of this would be to replace plastic with glass or metal in product manufacturing. This way they can be reused later preventing the generation of solid waste. Another recommendation would be to used materials in a very efficient and optimal way without throwing anything away. Hence the reduction in solid waste can make a difference in building a sustainable city. The Akola Municipality Corporation should proactively start encouraging people to implement the above suggestions. Some other policies like making people pay taxes based on the amount of trash that they create. This can encourage people to control waste generation and start using reusable items. Akola city is a hub of many industrial and agricultural businesses and hence these concepts should be applied strictly in these sectors. All the steps towards preventing people from generating more and avoidable waste generation and start using reusable and recyclable products could take some time. Hence making powerful policies and arrangements for the proper disposal of waste is very important.

There are many proposals considered for solid waste management i.e. after the waste has been generated. The first thing that comes into mind is the landfill. Akola region has a very limited area dedicated to these landfills and transferring the waste to other landfills far away from Akola could become a financial burden. There are difficulties in opening new landfills due to high land values and other local concerns in the region. Hence in the long term, AMC should focus on collaborating with some partners to explore new ways for waste disposal methods. The thinking process and actions should ensure the continued prosperity of the region and enhanced environment.

7. Conclusion:

In this research, we have successfully developed supervised machine learning regression models. These models are useful to some extent in predicting the amount of municipality solid waste in different regions in Akola city using

demographic and socio-economic parameters at the municipal level. We have found all the important variables which are influencing the waste generation and offer suitable models to predict the amount of MSW generation in different areas in Akola city depending on the values of those variables.

This study proves the successful application of machine learning algorithms at the municipality level of a city. In machine learning, there are many options as we have seen in this study, and the bagging approach gives more reliable results than any other machine learning algorithm. It produced a model which is explaining 67.58% of the variation in the amount of solid waste generated given the available data in the study. The results demonstrate that given a good amount of data and sufficient demographic and socio-economic data, these approaches can be built better models with a higher explanation for the variation in the amount of waste and much lower error for the prediction of solid waste generation.

Akola municipality corporation can make use of these developed machine learning models in further studies on building new sustainable solutions or expedite the MSW disposal system. These solid waste prediction models can be used across various municipalities in all states of India provided the input parameters used in this research are available with all the municipalities in enough quantity. The proper interpretation and understanding of the results of these predictive models can help municipalities gaining a unique opportunity to design and optimize their time and resources making waste management operations more efficient and optimal.

8. Limitations:

The study has some limitations while implementing at the ground level. The first limitation is the availability of the genuine data regarding the generation of solid waste and other parameters like demographics and socio-economic. There is a thumb rule in machine learning and that is 'Garbage In, Garbage Out'. It means the outcome of the machine learning models is totally dependent on the amount and quality of input data. If we don't provide a quality data to the machine learning algorithm in good amount, it will not build a good model and the predictions of such model are not dependable. Hence, a proper data collection or storage of the dependent and independent variables used in this study is one the biggest task and many municipality in India do not have that system in place.

References

Municipal Solid Waste Rules (Management and Handling), 2000 by Ministry of Environment and Forests.

Akola Municipality Corporation official website. https://akola.gov.in/

Indian census data for Akola city (2011). https://censusindia.gov.in/

Study by the organization named The Resources and Livelihoods Group, Prayas, Pune solid waste management.

Shannigrahi, A.S., Chatterjee, N., Olaniya, M.S., 1997. Physico-chemical characteristics municipal solid wastes in mega city. Indian Journal of Environmental Protection 17 (7), 527–529.

Sharholy, M., Ahmad, K., Mahmood, G., Trivedi, R.C., 2005. Analysis of municipal solid waste management systems in Delhi – a review. In: Book of Proceedings for the second International Congress of Chemistry and Environment, Indore, India, pp. 773–777.

Abbasi and El Hanandeh (2016). Paper on a model for Assessing Waste Generation Factors and Forecasting Waste Generation using Artificial Neural Networks., 78: 2-19(10 pages)

Navarro-Esbrí (2002). Seasonal autoregressive moving average modelling technique used for forecasting solid waste generated. Resources Conservation and Recycling 35(3):201-214.

S. Kanan, Somasundaram Krishnan, (2015). A Review of Outlier Prediction Techniques in Data Mining 10(9):1021-1028

Ruilin Pan, Tingsheng Yang, Jianhua Cao, Ke Lu (2015). Missing data imputation by K-nearest neighbours based on grey relational structure and mutual information: Applied Intelligence 43(3)

Gülden K.; Neşe G., (2013). A Study on Multiple Linear Regression Analysis., 1(3): 3-20 (8 pages)

Lingjian Y.; Songsong L.; Sophia T., (2017). A regression tree approach using mathematical programming., 50: 2-34 (9 pages).

Thomas G D., (2000). Ensemble Methods in Machine Learning (15 pages)

Leo B, (2001). Assessing how the Random forests work (33 pages)

Ramraj Santhanam, Nishant Uzir, (2017). Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets

Xue Ying (2019). An Overview of Overfitting and its Solutions.,

Yun Xu, Royston Goodacre (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning.

Kannangara, M.; Dua, R.; Ahmadi, L.; Bensebaa, F., (2018). Modelling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches.Waste Manage., 74: 3–15 (13 pages).

Pan, A.; Yu, L.; Yang, Q, (2019). Characteristics and forecasting of municipal solid waste generation in China. Sustainability, 11(5): 1-11 (11 pages).

Chhay, L.; Reyad, M.; Suy, R.; Islam, M.; Mian M., (2018). Municipal solid waste generation in China: influencing factor analysis and multi-model forecasting. J.Mater Cycles Waste Manage., 20(3): 1761–1770 (10 pages)