

Prediction of Solid Garbage Waste Generation in Smart Cities using Naive Bayes Algorithm

Rashmi G, S Sathish Kumar

Abstract: Smart cities which are becoming overcrowded today are making human beings life miserable and prone to more challenges on daily basis. Overcrowded is leading to vast generation of wastes contributing to air pollution and in turn is affecting health causing various diseases. Even though various measures are taken to recycle wastes, the rate at which it is being produced is becoming higher and higher. This paper deals with prediction of waste generation using Naive Bayes machine learning algorithm(Classifier) based on the statistics of previous waste datasets. The datasets used for the future prediction are obtained from reliable sources. The implementation of the algorithm is done in Pyspark using Anaconda Jupyter. The performance of the classifier on the datasets is analyzed with confusion matrix and accuracy metric is used to rate the efficiency of the classifier. The accuracy obtained indicates that algorithm can be effectively used for real time prediction and it gives more accurate results for huge input datasets based on independence assumption.

Keywords: Machine Learning, Big Data, Naive Bayes Classifier, PySpark, Solid Garbage Waste

I. INTRODUCTION

The increase in solid garbage waste such as plastic waste, building and devastation waste, biomedical waste, hazardous waste, e-waste, municipal solid waste) in smart cities is due to increasing population and urbanization which is resulting various challenges for living beings.

A. Introduction to Waste Generation in Smart Cities

As per the survey, Bengaluru city is producing 6233 tonnes of solid waste every day and is expected to reach 13911 tonnes in next ten years. Hence, it is very important to predict the rate at which waste is being produced which helps waste management system to analyse the consequences and prevent the same by planning and taking some measures. Waste management includes several networks and stake holders and gathering and compilation of data is very hard. Waste management is taken care by many people including local experts and private freelancers. With accurate data, solid waste could be achievable another home of energy to substitute presently utilized diesel power generation. Collection of solid waste information is thought-provoking because of unconditional influences and rules comprised.

Revised Manuscript Received on December 12, 2019.

RASHMI G, Computer Science and Engineering, RNSIT Research Centre, Bengaluru, India. rashmi4mahendrakar@gmail.com

S SATHISH KUMAR, Department of Information Science and Engineering, RNSIT, Bengaluru, India. sathish_tri@yahoo.com

B. Introduction to Machine Learning

Today machine learning is the most demanded technology in the market for predictive analysis. Machine learning algorithms make your coding smoother by automatic learning from the input data given and they deal with two stages. Namely, training stage and testing stage.

Training Stage

This stage deals with random selection of data known as training data. Prepares a table involving different features of the training data along with the expected outcomes. Then, the data is fed into machine learning algorithm which learns correlation between the features and outcomes.

Testing Stage

This stage makes use of the model developed in training phase to test the data and performs predictive analysis.

Classification of Machine learning algorithms is shown in Fig 1.

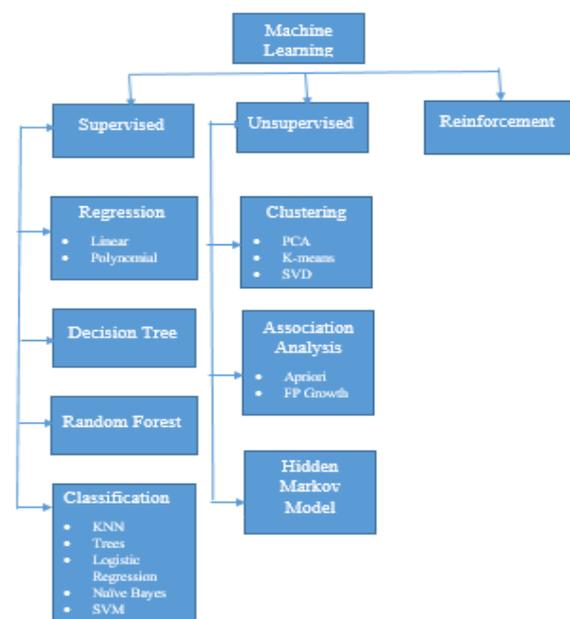


Fig. 1. Machine learning algorithms classification

II. LITERATURE REVIEW

Predictive analysis is one of the fastest growing technology available for estimating or predicting the amount of waste in smart cities.

Following are some of the contributions by the various researchers towards forecasting the solid garbage waste generation using different models.

Miyuru Kannagara, Rahul Dua, Leila Ahmadi, Farid Bensebaa[2] proposed models for forecasting of solid waste generation at Municipality and change based on demographic and socio economic variables, with strategic application of creating Canada wise solid waste records. Models were created by mapping uptown municipal solid waste masses with socio-economic and demographic factors of 220 municipalities in the state of Ontario, Canada. Decision tree and neural network machine algorithms were used to construct the models. Socio economic variables were achieved from census data of Canada at local and municipal levels. A data preprocessing and amalgamation framework was built in Matlab software to generate datasets with acceptable information capacity and quality for modeling. Outcomes revealed that waste models can be created effectively using machine learning algorithms with good prediction performance. Neural network models performed well explaining 72% of difference in the data. The tactic recommended in this research work proves the possibility of making tools that benefits in state waste scheduling by means of sourcing, recycling, integrating and exhibiting of widely accessible information from several bases.

Maya Chavan, T. R. Pattanshetti[3] proposed a review of literature study which says management of waste is one of the major issues in the crowded cities because of increasing population. Proper care needs to be taken care by the corresponding authorities to overcome this challenge. The large unmanaged gathering of solid waste garbage is contributing to air pollution affecting living beings' health. To overcome this problem, an effective garbage waste management system needs to be developed. The proposed survey paper contains various thoughts of solving this problem based on few algorithms which can be used.

Kontokosta, C., Hong, B., Johnson, N. E., & Starobin, D[12] proposed a new approach which uses amalgamation of machine learning and minor area assessment techniques to forecast daily and weekly based generation of waste at the construction scale. They collected ten years' datasets from the Department of Sanitation sub divisions, New York on distinct construction features, locality socioeconomic properties, climate and chosen route level gathering data. They applied neural network model and gradient boosting regression trees to assess weekly and daily rejected and reprocessing capacities for each of the more than 700,000 housing belongings of the city. Outcome of the work indicate that the proposed approach holds good for predicting waste generation of the buildings with a high accuracy.

Hanandeh, Ali & Abbasi[8] developed a model for exact forecasting of solid waste generation at municipal level which helps waste management centers to plan enhanced waste administration systems. They used four algorithms which includes Support Vector Machine, Adaptive Neuro Fuzzy Inference System, Artificial Neural Network and K-Nearest Neighbors to forecast monthly generation of waste in the City of Logan Council zone in Queensland, Australia. Outcomes indicated that artificial intelligence models have improved

forecast ability and can be used to develop forecast models for solid waste generation at municipal level. Machine learning algorithms can consistently forecast waste generation monthly basis. Adaptive Neuro Fuzzy formed the most precise predictions of the peeks whereas K-Nearest Neighbor was efficient enough to predict averages of waste capacity on monthly basis. Depending on the outcome, the overall yearly waste produced in the City of Logan is expected to touch 9.4×10^7 kg by 2021 whereas the highest monthly waste is expected to touch 9.37×10^6 kg.

III. NAÏVE BAYES CLASSIFIER FOR FORECASTING SOLD GARAGE WASTE GENERATION

Naïve Bayes is the accurate, fastest and straight forward classification algorithm. which performs well with huge amount of data. It is used in various applications such as spam filtering, text classification, sentiment analysis etc.

It has two stages. Namely, training stage and testing stage shown in Fig. 2. Training stage is also called learning stage which trains its model on a given data set. Testing stage is also known as evaluation stage which evaluates the model performance based on various parameters such as accuracy, error, precision and recall.

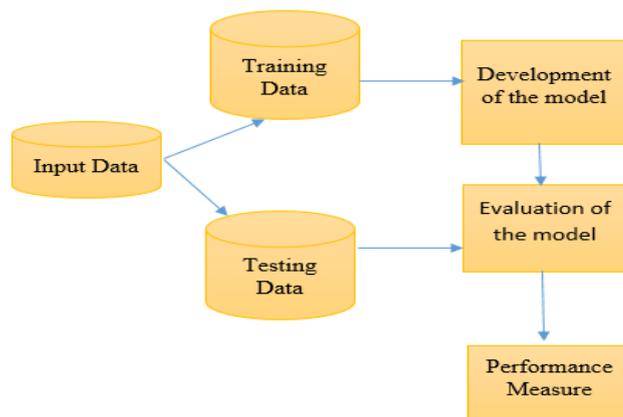


Fig. 2. Stages of Naïve Bayes Approach

This approach use Bayes theorem based on the assumption that the occurrence of specific attribute (feature or variable) in a class is independent of some other attributes and hence it is termed as Naïve Bayes. This approach is very beneficial for huge data sets.

Bayes theorem computes subsequent or later probability $P(c|x)$ using $P(c)$, $P(x)$ and $P(x|c)$ as shown in (1).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability
Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c) \tag{1}$$

Where, $P(c|x)$ is the later likelihood of aimed class given a feature(predictor).

P(c) is the past likelihood of a class.
P(x|c) is the likelihood of predictor given class.
P(x) is the previous likelihood of predictor.

Algorithm

Input: A CSV file containing input datasets

Output: Accuracy of prediction

Step1: Load and parse the input data file.

Step2: Divide the given input data sets roughly into training data (90%) and test data (10%).

Step 3: Train the Naïve Bayes Model.

- Compute the previous likelihood for given class labels.
- Calculate possible probability with each feature for each class.
- Place the obtained values in Bayes formula.
- Obtain the class having greater probability, given the input fits to the greater probability class.

Step 4: Make prediction and check accuracy.

Step 5: Save and load the model.

IV. EXPERIMENTS AND RESULTS

The prediction of waste generation is done using Anacondada3 Jupyter. The Naïve Bayes algorithm has been implemented in Pyspark (python for Spark). The input given is CSV file contains 5000 data sets collected from Central Pollution Control Board. To measure accuracy of the proposed method, the dataset is partitioned into training and testing data where 90% of the random samples are taken for building training model and 10% for testing model. The results show that the average accuracy of prediction achieved using proposed approach is 84%. The approach works well with multiple features and huge datasets.

Various parameters also known as variables such as population, economy, education, geography, climate and quantity of waste can be considered for predicting solid waste generation.

Population deals with percentage of Population, Age groups, gender, age groups.

Economy deals with poverty, income, vehicles, construction rate, unemployment.

Education deals with illiteracy rate, libraries, cultural activities.

Dwelling deals with houses, number of people in the house.

Geography(Area), Climate and Quantity of Waste.

The proposed work considers only population, Geography and Waste quantity as parameters (features or variables) for computing accuracy of prediction of waste generation as shown in the sample input data (see Table I).

In developing relationships between Waste Generation Prediction(WGP) and variables, we have used Naïve Bayes classifier as follows.

P (WGP | Population, Geography, Quantity of Waste) =

$$P(\text{Population}) \times P(\text{WGP} | \text{Population}) \times P(\text{Geography}) \times P(\text{WGP} | \text{Geography}) \times P(\text{Quantity of waste}) \times P(\text{WGP} | \text{Quantity of waste})$$

Table-I: Sample Input Datasets

Sl No	City	Population	Area(sq.km)	Waste Quantity(Tonnes per Day)
1	Kavaratti	10119	4	3
2	Gangtok	29354	15	13
3	Itanagar	35022	22	12
4	Daman	35770	7	15
5	Silvassa	50463	17	16
6	Panjim	59066	69	32
7	Kohima	77030	30	13
8	Port Blair	99984	18	76
9	Shillong	132867	10	45
10	Simla	142555	20	39
11	Agartala	189998	63	77
12	Gandhinagar	195985	57	44
13	Dhanbad	199258	24	77
14	Pondicherry	220865	19	130
15	Imphal	221492	34	43
16	Aizwal	228280	117	57
17	Jammu	369959	102	215
18	Dehradun	426674	67	131
19	Asansol	475439	127	207
20	Kochi	595575	98	400

Performance or Efficiency of the Naïve Bayes classifier is examined with confusion matrix (see Table II) and is measured using accuracy metric.

Table-II. Confusion Matrix

	Prediction	
	Relevant	Irrelevant
Relevant	TP	FN
Irrelevant	FP	TN

Where,

TN (True Negative) indicates the number of right predictions that an occurrence is inappropriate.

FP (False Positive) indicates the number of wrong predictions that an occurrence is appropriate.

FN (False Negative) indicates the number of wrong predictions that an occurrence is inappropriate.

TP (True Positive) indicates the number of right predictions that an occurrence is appropriate.

Accuracy is well-defined as the percentage of the overall predictions that were right and is given by

$$\text{Accuracy}(\%) = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FN} + \text{FP} + \text{TP}}$$

The accuracy obtained by Naïve Bayes for different size of datasets (see Table III).

Table-III: Accuracy of prediction for different size of data sets

Dataset Size	Accuracy (%)
1000	79.8
2000	80.0
3000	80.4
4000	80.9
5000	81.2



V. CONCLUSION

The paper contributes to the growth of waste management practices based on the data about the waste being generated and the waste generating parameters. Naïve Bayes algorithm can be effectively used for forecasting waste generation and it achieved an average of 80% of accuracy. Due to the limitations of data availability, limited attributes such as city, population, area and waste quantity were considered for prediction. Performance of Naïve Bayes is analysed with confusion matrix predictive bounds such as True Positive, True Negative, False Negative and False Positive and evaluation is done using accuracy matrix. Naïve Bayes results in best accuracy only when the independence assumption holds good. The current research work can be enhanced by considering more attributes and other performance metrics such as precision, F-Measure and recall to forecast waste generation in smart cities.

REFERENCES

1. Zhang, Zhenying & Zhang, Yuxiang & Wu, Dazhi. (2019). Hybrid model for the prediction of municipal solid waste generation in Hangzhou, China. *Waste Management & Research*. 37. 0734242X1985543. 10.1177/0734242X19855434.
2. 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, Pages 3-15, ISSN 0956-053X.
3. 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.
4. Yadav, Saneh & Sohal, Asha. (2017). Review Paper on Big Data Analytics in Cloud Computing.
5. D. D. Vu and G. Kaddoum, "A waste city management system for smart cities applications," *2017 Advances in Wireless and Optical Communications (RTUWO)*, Riga, 2017, pp. 225-229.
6. 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.
7. Sharmin, Sadia & Al-Amin, Sikder Tahsin. (2016). A Cloud-based Dynamic Waste Management System for Smart Cities. 1-4. 10.1145/3001913.3006629.
8. 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.
9. 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): 188-193.
10. Chandio, Aftab & Tziritas, Nikos & Xu, Cheng-Zhong. (2015). Big-Data Processing Techniques and Their Challenges in Transport Domain. *ZTE Communications*. 13. 50-59. 10.3969/j.issn.1673-5188.2015.01.007.
11. 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.
12. Kontokosta, C., Hong, B., Johnson, N. E., & Starobin, D. (Accepted/In press). Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities. *Computers, Environment and Urban Systems*.
13. Grazhdani, Dorina. (2015). Assessing the variables affecting on the rate of solid waste generation and recycling: An empirical analysis in Prespa Park. *Waste management (New York, N.Y.)*. 48. 10.1016/j.wasman.2015.09.028.

AUTHORS PROFILE



Rashmi G has done her Bachelor of Engineering in CSE from Acharya Institute of Technology, Bengaluru and M.Tech in CSE from RV College of Engineering, Bengaluru. She is currently a research scholar at RNSIT, Bengaluru and pursuing Ph.D in Big Data domain, Visvesvaraya Technological University. She has around 10 publications in reputed journals. Her area of interest includes Big Data and Cloud Computing.



Dr. S Sathish Kumar has done B.E degree from Madurai Kamaraj University, Tamilnadu, India, M.E degree from Regional Engineering College, Tiruchirappalli, Bharathidasan University, Tamilnadu, India. and Doctor of Philosophy Degree from Dr. M. G. R University, Chennai, Tamilnadu, India. Currently he is working as a Professor in R N S Institute of Technology, Bengaluru, Karnataka, India. He is having more than 25 years of experience. He has served as a BOE/BOS member of various Universities and colleges. He has published research papers in various journals and conferences. His field of Interest includes Data Mining, Big Data, Bio Medical Engineering and Cloud Computing.