A Waste City Management System for Smart Cities Applications

Dung D. Vu
Electrical Engineering Department
École de Technologie Supérieure
Montréal, Canada
do-dung.vu.1@ens.etsmtl.ca

Georges Kaddoum
Electrical Engineering Department
École de Technologie Supérieure
Montréal, Canada
georges.kaddoum@etsmtl.ca

***Abstract*—This paper presents a new method of smart waste city management which makes the environment of the city clean with a low cost. In this approach, the sensor model detects, measures, and transmits waste volume data over the Internet. The collected data including trash bin's geolocation and the serial number is processed by using regression, classification and graph theory. Thenceforth a new method is proposed to dynamically and efficiently manage the waste collection by predicting waste status, classifying trash bin location, and monitoring the amount of waste. Then, this latter recommends the optimization of the route to manage the garbage truck efficiently. Finally, the simulation results are presented and estimated.**

***Index Terms*—Big data, machine learning, optimization, waste management, IoT**

I. Introduction

A smart city is an urban development vision to integrate both information and communication technology (ICT) and Internet of things (IoT) technology which leads to the next technological revolution [1] and manages the city's assets securely. IoT is a framework in which all *things* have a representation and a presence on the Internet. More specifically, the IoT aims at offering new applications and services bridging the physical and virtual worlds, in which Machine-to-Machine (M2M) communication represents the baseline interface that enables the interactions between things and applications on the cloud such as environment monitoring [2], [3], object tracking [4], traffic management [5], health-care and patient remote monitoring [6], and smart home technology [7] and [8]. Organizations can use IoT to drive considerable cost savings by improving asset utilization, enhancing process efficiency and boosting productivity. IoT combines the exponential growth of smart devices, the confluence of low-cost technologies (sensors, wireless networks, big data, and computing power), the pervasive connectivity and the massive volumes of big data. Hence, IoT and big data are two sides of the same coin. An IoT device generates continuous streaming information in a scalable way. Companies must handle the high volume of stream data and perform actions on that one. These actions can be event correlation, matrix calculation, statistic preparation and analytics which are applied to construct the smart city. In the smart city, the waste management system is a crucial point of living environment, and its quality is considered seriously. A possible smart waste city management system requires a way to cluster the trash bin locations, detect the status of waste in each bin, and process this collected data. The result of this work will be a valuable input data for the garbage truck management system which calculates the most optimal route to prevent the hazard of damage, pollution of waste, and resource consumptions. To manage the waste of a smart city, the system incorporates a model for sharing data between truck drivers in real time to perform garbage collection in [9]. A waste collection solution based on providing intelligence of trash cans including sensors and IoT prototype, which can read, collect, and transmit trash volume data via wireless network was proposed in [10] and [11]. An adaptive large neighborhood search algorithm of finding optimal cost routes of garbage trucks was presented in [12], which processes the data of all emptied trash bins and the driven waste to disposal sites while respecting customer time windows. An improved dynamic route planning was discussed in [13], where the authors enhanced a guided variable neighborhood threshold meta-heuristic adapted to the problem of waste collection. On the other hand, the most important part of a waste management system was the smart bin management presented in [14]–[16]. In these studies, the collected data from the sensors is sent over the internet to a dedicated for monitoring its status. While the network of ultrasonic sensors enabled smart bins to connect through the cellular network and to generate a large amount of data which was further analyzed and visualized at real time to gain insights about the status of waste around the city [14], a smart waste management with self-describing objects detected the kind of waste based on its Radio-frequency identification (RFID) information [15]. However, in their design they did not know the hazard of each bin such as explosion or flame from the bottle of perfume, batteries, and electronic wastes. In [16], a hazard detection method was proposed to detect and prevent these issues. Hence, a waste city management was conducted by an optimal garbage truck routes algorithm based on the status of the smart bin.

However, we observed that citizens in the city tend to throw their waste into the trash bins without a particular time of a day and the manager needs to add more bins during the holiday event time. As previous works do not solve this issue, the novelty of our work is to propose a new method to automatically classify the location of trash bins, predict the

978-1-5386-0585-1/17/$31.00 ©2017

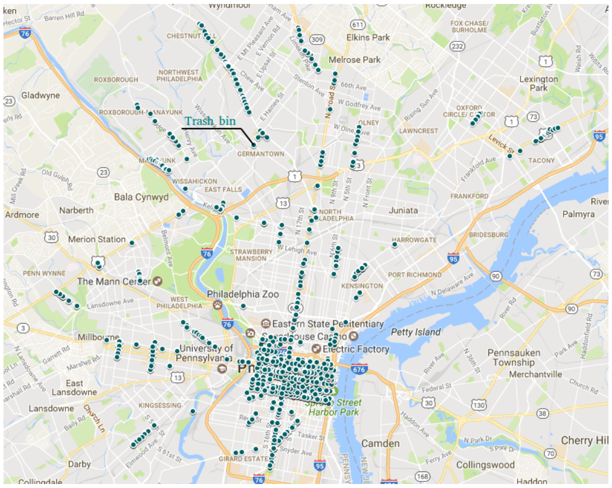


Fig. 1: Philadelphia trash bin.

status of each one and optimize the garbage truck routes. The aim of this proposition is to help the city's waste management system to operate more efficiently.

Contributions: Compared to the works mentioned above, this paper presents a new method of smart waste city management which provides a clean and hygienic environment to the city resident. In this approach, the collected data from the trash bin has been transferred over the Internet to the server which contains the status of each bin. The contributions of this paper are summarized as follows:

- The number of working clusters in each city are clustered and the location of new trash bins are classified automatically.
- A regression algorithm is applied to predict the situation of waste which can avoid the overload trash phenomenon while the garbage truck is coming to gather the trash bins.
- The priority weight of each bin is considered to conduct the optimal garbage truck routes algorithm more efficiently.

This paper is organized as follows: in section II, a new method of smart waste city management (SWCM) is explained and the structure of the algorithm is proposed. Section II-B2 presents the SWCM simulation model. Finally, the related concluding remarks are discussed in section IV.

II. IoT-Based Smart waste management model

In this section, we describe how to collect waste metadata associated with their statuses and locations. We evaluate our model with real big data in order to validate its output result.

A. Data acquisition

To obtain a set of waste data, we use an open source database[1], which has a significant amount of geo-location information and status of trash bins at the largest city in the Commonwealth of Pennsylvania, United Stated named Philadelphia. Figure 1 illustrates the distribution of trash

[1] https://www.opendataphilly.org/dataset

TABLE I: The trash bin dataset.

Field	Description
sn	The serial number of each trash bin
time_stamp	The milestone of recording data
level	The amount of waste in each trash bin (GREEN, YELLOW, RED) at the given time
lat	The latitude of each trash bin
lon	The longitude of each trash bin

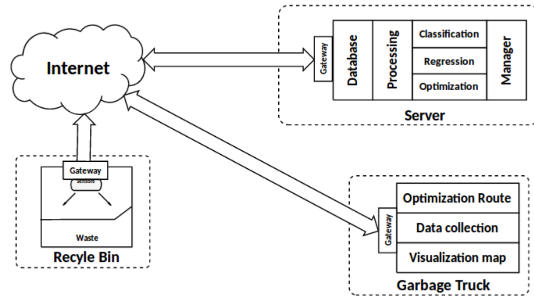


Fig. 2: The smart waste management system overview.

bin (green node) in Philadelphia city. Since there are many contents in this data, we removed the content that was not related to our scenario. For data analysis, five essential fields from this metadata are presented in the Table I.

B. System description

1) The overview of functionality: the proposed system is based on the waste status of each trash bin in the city. The collected data is sent over the Internet to the server where it is stored and processed. At here, it is used for monitoring and predicting the status of each trash bin daily. Moreover, it will be utilized for calculating the optimal garbage truck routes, accordingly. The prediction status of each bin can be analyzed based on the given training data before it occurs. Then, it will be considered to update the weight of the trash bin accordingly, which is the most important input parameter of optimal garbage truck routes algorithm. The system overview is shown in the Figure 2.

2) Data processing and classifications: the status of each bin is not homogeneous and significantly differs from each other according to the state of each location. In this section, we introduce an algorithm that can be used to dynamically and efficiently manage waste collection strategies.

K-means [17] is an unsupervised machine learning algorithm that groups a dataset into a user-specified number (k) of clusters. We use it to make the working cluster of each garbage truck. However, this algorithm is somewhat naive, it clusters the data into k clusters, even if k is not the right number of clusters to use. Therefore, when using k-means clustering, users need some way to determine whether they are using the right number of clusters.

In this paper, we use Elbow method [18] and [19] to validate the number of clusters. The collecting routes are the traveling cycles containing a set of trash bins within a given cluster. The optimization of these cycles is a combinatorial optimization

TABLE II: The definition of each variable.

Variable	Description
$S = \{s_i \mid i \in (1,n)\}$	Status of each bin i
$L = \{l_i \mid i \in (1,n)\}$	Location of each bin i
$W = \{w_i \mid i \in (1,n)\}$	Weight of each bin i
$T = \{t_{di} \mid d \in (1,m); i \in (1,n)\}$	Collected data milestone of each bin i
$\mathcal{D} = \{S, L, W, T\}$	Input data
$\mathcal{M} = \{M_j \mid j \in (1,k)\}$	Optimal route of each garbage truck within each cluster M_j
$\mathcal{DC} = \{DC_j \mid j \in (1,k);$ $DC_j = \{S_j, L_j, W_j, T_j\}\}$	Working cluster
k	Number of working clusters
η	Threshold of amount waste
n	Number of trash bins in the dataset
$m()$	Dynamic vector of number trash bins in each working cluster

problem. By considering a large number of routes, we use a Genetic Algorithm (GA) [20] which is relatively fast in providing near optimal solutions. Since the garbage truck needs time to collect every trash bin, it is very delightful if a status of a trash bin can be predicted. Hence, after predicting, the system will recommend which one should be collected to prevent the overload phenomenon. In this paper, we use the Logistic regression algorithm [2] to predict the status of each trash bin based on its historical data.

The overall procedure is summarized in Algorithm 1. In this algorithm, the parameters are defined in Table II. We suppose that the input dataset $\mathcal{D}$ is the database in Section II. At the given time stamp t, a logistic regression (LR) algorithm is applied to predict the status of each bin. If the prediction status s_{jh} of the bin is greater than the given threshold η, the weight w_{jh} is updated to 1, respectively, then the GA algorithm is utilized to minimize the driving distance for visiting the selected bins and returning to the headquarters. Therefore, we find the optimal garbage truck routes in each cluster M_j which helps us to prevent the collected trash bin more efficiently.

Algorithm 1

Input: $\mathcal{D}, \eta$
Output: $\mathcal{M}$
Initialization: $j, k, h, m() = 0, k = [1; 15]$
k^* = Elbow($\mathcal{D} \rightarrow$L, n)
$\mathcal{DC}$ = K-means($\mathcal{D} \rightarrow$L, k^*)
At a given time stamp t in the near future.
for $j = 1$ to k^* **do**
  m_j = size($\mathcal{DC}_j$)
  for $h = 1$ to m_j **do**
    s_{jh} = LR($DC_j \rightarrow S_j, DC_j \rightarrow T_j$)
    if ($s_{jh} \geq \eta$) **then**
      $w_{jh} = 1$
      M_j = GA($DC_j \rightarrow L_j, DC_j \rightarrow W_j$)
    end if
  end for
  Plot (M_j)
end for

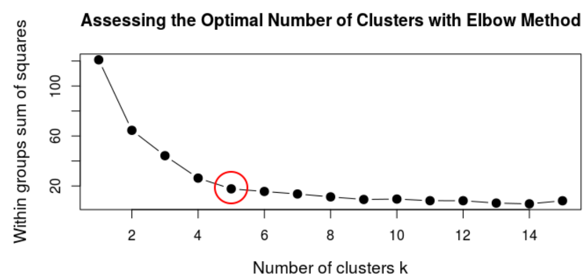


Fig. 3: The optimal value of number working clusters.

III. Analysis Result and Discussion

In this section, using the proposed Algorithm 1 in Section II-B, we first analyze the number of working cluster and then show the experimental result. We simply assume $\eta = 0.5$, which can also be set to another value to monitor the amount of waste.

A. Number of working cluster

One method of validating the number clusters for the algorithm is the Elbow method. It looks at the percentage of variance explained as a function of the number of clusters: one should choose a number of clusters so that adding another cluster does not give much better modeling of the data. In general, $k \in [1; \infty]$ but for the simple experiment, we choose $k \in [1; 15]$ in our work. To estimate the optimal value of k^*, we use the Elbow($\mathcal{D} \rightarrow$L, n) which considers the value of within-cluster variance $WS(DC_j)$ of a cluster DC_j which is defined as $\sum_{l_i \in DC_j} \|l_i - \bar{l_j}\|^2$, where $\bar{l_j}$ is the mean of cluster DC_j (also called the cluster centroid, its values are the coordinate-wise average of the data points in DC_j), and $\{l_1, l_2, ..., l_n\}$ is the set of trash bin locations. The total within-cluster scatter is $WSS = \sum_{j=1}^{k} \sum_{l_i \in DC_j} \|l_i - \bar{l_j}\|^2$ for k clusters and n observations A good choice value of k^* is considered if this value tends to change slowly and remains less changing as compared to other k's[3]. We see a very clear Elbow at $k^* = 5$ in Figure 3, which indicates that 5 is the best number of clusters. Using the value of k^* above, we apply the K-means algorithm to make the working clusers for our system. The output is represented in the Figure 4. If the manager tends to add more trash bins into our system, the system will automatically update these working clusters.

B. Predicting the status of trash bin

In fact, when the garbage trucks are running on the road, the status of trash bins can be modified. It is very delightful to predict the status of a trash bin, then update the weight of this one. After that, the system will update the optimal garbage truck route. A Logistic regression algorithm is applied

[2]https://nlp.stanford.edu/manning/courses/ling289/logistic.pdf

[3]https://www.r-bloggers.com/finding-optimal-number-of-clusters/

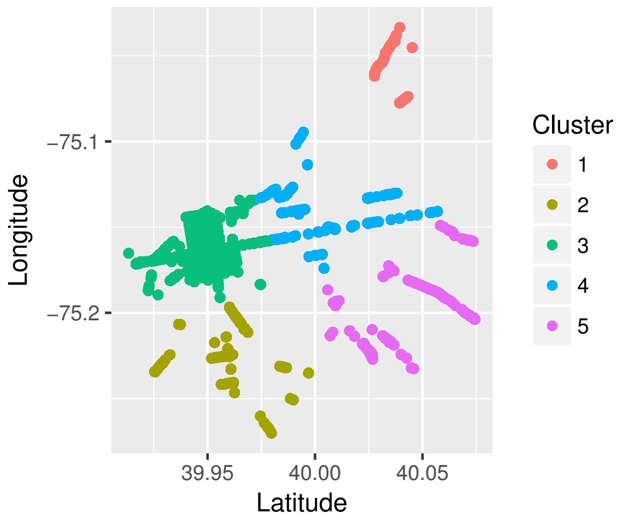


Fig. 4: The working clusters in Philadelphia.

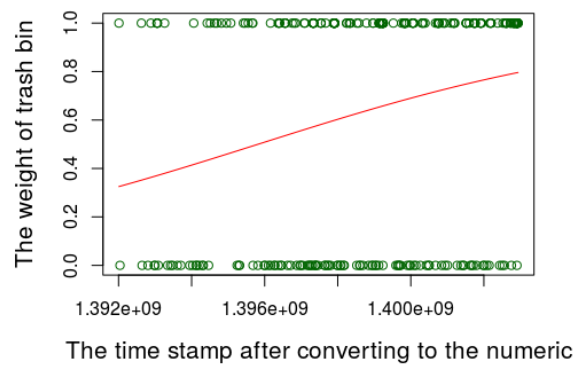
to work on this issue. We pick up randomly one trash bin in one cluster whose serial number and location are 171758 and $(39.95204^o, -75.15911^o)$, respectively, to analyze the output result for convenience. Fistly, we convert all the date to the numeric data type[4] which is utilized for the Logistic regression algorithm. Figure 5a represents the relation between weight (green circle) and the time stamp of this trash bin. In this figure, the red line represents the trend of trash bin's weight. By considering a given time, if the red line approaches the value 1, the weight of trash bin will be updated to 1, respectively. Since the limitation of the data, the area under the curve value AUC in Figure 5b is 0.66. Hence, the optimal garbage truck route will be constructed based on the new weight.

C. Optimal route of garbage truck in each cluster

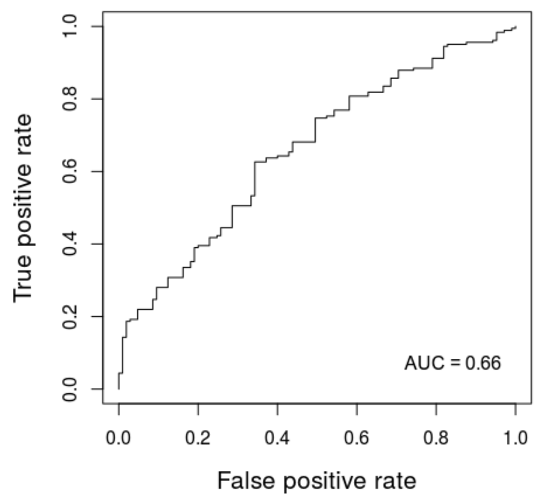
In the each cluster dataset $\mathcal{DC}_j$, where $j \in [1 : k^*]$, we observed that the original system has three status levels of bins such as RED, YELLOW, and GREEN. Let we assume that when $\eta = 0.5$, the trash bin's level is RED or YELLOW and its weight is equal to 1, otherwise it is equal to 0.

Since we need to update the optimal garbage truck routes daily for collecting the high weight trash bin, so we would like to pick up a milestone randomly which is "2014-06-16" in our dataset to make the simulation. By using the GA algorithm, the optimal garbage truck routes of each working cluster are constructed based on the weight and coordinate of trash bins. In the Figure 6, the red point is the high weight trash bin and the arrow is the route of garbage truck. It means that the garbage trucks collected all the trash bins whose weight is 1 by using the optimal routes. Hence, the system will help the smart city to reduce the traffic congestion, fuel consumption, and pollution.

[4]https://cran.r-project.org/web/packages/lubridate/lubridate.pdf



(a) The LR of the 171758^{th} trash bin.



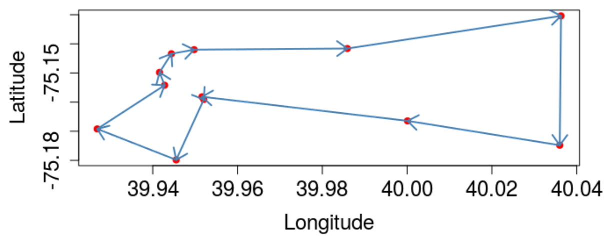
(b) The ROC curve of the output LR of the 171758^{th} trash bin.

Fig. 5: The logistic regression result for the 171758^{th} trash bin.

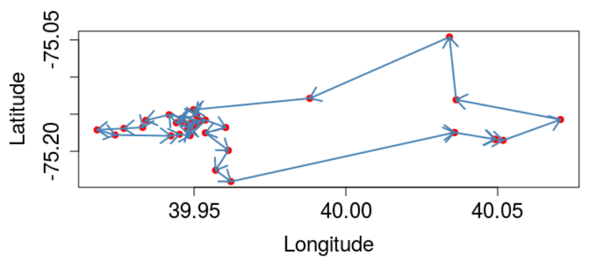
IV. Conclusion and Future Works

In this paper, we first clustered the working area by using the open source database of Philadelphia. Especially, we introduced an algorithm that automatically makes the working clusters and calculates the optimal garbage truck routes. Also, we use the Logistic regression to predict and update the weight of each trash bin. They will be used for creating the new optimal garbage route which reduces the pollution and fuel consumption more efficiently.

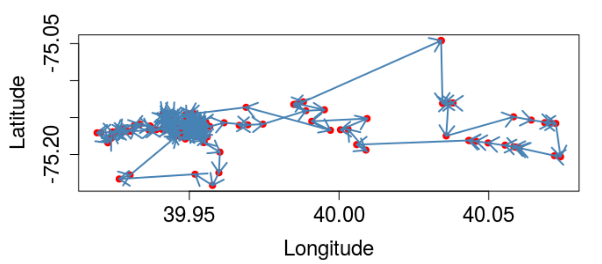
Future work has to be done to further improve the system such as: the data should be collected in a better way, in other words, not only the amount of waste must be collected but also the status of hazard (e.g., temperature, poison gas, and liquid), then the estimated priority of weight will be calculated



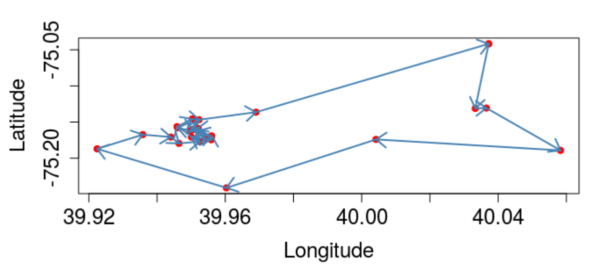
(a) Cluster 1.



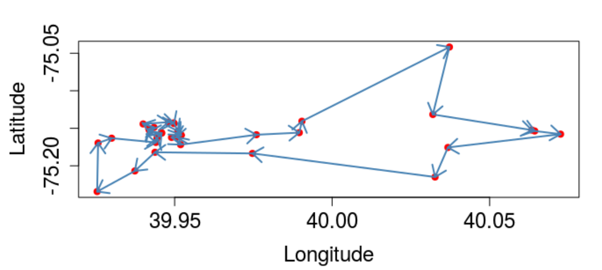
(b) Cluster 2.



(c) Cluster 3.



(d) Cluster 4.



(e) Cluster 5.

Fig. 6: The optimal route of garbage truck at 2014-06-16.

more accurately. Moreover, the input parameters of optimizing garbage truck routes should be added the traffic congestion information, fuel consumption, forecast news and the optimal garbage truck routes within each cluster need to be fitted with the street map data.

REFERENCES

[1] F. C. Delicato, P. F. Pires, T. Batista, E. Cavalcante, B. Costa, and T. Barros, "Towards an IoT eco-system," in *Proc. of the 1st ACM Int. Workshop on Software Engineering for Systems-of-Systems, SESoS'13*, pp. 25–28, 2013.

[2] M. T. Lazarescu, "Design of a WSN platform for long-term environmental monitoring for IoT application," *IEEE J. Emerging and Selected Topics in Circuits and Systems*, vol. 3, no. 1, 2013, pp. 45–54.

[3] S. D. T. Kelly, N. K. Suryadevara, and S. C. Mukhopadhyay,"Towards the implementation of IoT for environmental condition monitoring in homes", *IEEE Sensors J.*, vol 13, no 10, 2013, pp. 3846–3853.

[4] K. Gama, L. Touseau, and D. Donsez, "Combining heterogeneous service technologies for building an internet of things middle-ware", *Computer Communications*, vol 35, no 4, 2012, pp. 405–417.

[5] L. Foschini, T, Taleb, A. Corradi, and D. Bottazzi, "M2M-based metropolitan platform for IMS-enabled road traffic management in IoT", *IEEE Commun. Mag.*, vol 49, no 11, 2011, pp. 50–57.

[6] G. Matar, J. M. Lina, J. Carrier, A. Riley, and G. Kaddoum, "Internet of Things in sleep monitoring: An application for posture recognition using supervised learning", in *IEEE 18th Int. Conf. on e-Health Networking, Applicat. and Services (Healthcom)*, pp. 1–6, 2016.

[7] S. Tozlu, M. Senel, W. Mao, and A. Keshavarzian, "Wi-Fi enabled sensors for internet of things: A pratical approach", *IEEE Commun. Mag.*, vol 50, no 6, 2011, pp. 134–143.

[8] X. Li, R. Lu, X. Liang, X. Shen, J. Chen, and X. Lin, "Smart community: An internet of things appilication", *IEEE Commun. Mag.*, vol 49, no 11, 2011, pp. 68–75.

[9] A. Medvedev, P. Fedchenkov, A. Zaslavsky, T. Anagnostopoulos, and S. Khoruzhnikov, "Waste management as an IoT enabled service in smart cites", in *15th ed, St. Petersburg, Russia: Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, pp. 104–115, 2015.

[10] J. M. Gutierreza, M. Jensenb, M. Heniusa, and T. Riaze, "Smart waste collection system based on location intelligence", in *Procedia Computer Science*, vol 61, pp. 120–127, 2015.

[11] J. Hong, S. Park, B. Lee, J. Lee, D. Jeong, and S. Park, "IoT-based smart garbage system for efficient food waste management", *The Scientific World J.*, 2014.

[12] K. Buhrkala, A. Larsena, and S. Ropkea, "The waste collection vehicle routing problem with time windows in a city logistics context", in *Procedia Social and Behavioral Sciences* , vol 39, pp. 241–254, 2012.

[13] T. Nuortio, J. Kytöjoki, H. Niska, and O. Bräysy, "Improved route planning and scheduling of waste collection and transport", *Expert Syst with Applicat.*, vol 30, no 2, 2006, pp. 223–232.

[14] N. Sharma, N. Singha, and T. Dutta, "Smart bin implementation for smart cities", *Int. J. Scientific & Eng. Research* , vol 6, no 91, 2015, pp. 787–79.

[15] Y. Glouche and P. Couderc, "A smart waste management with self-describing objects", in *2nd Int. Conf. on Smart Syst., Devices, and Technologies*, pp. 63–70, 2013.

[16] A. Sinhan and P. Couderc, "Smart bin for imcompatible waste items", in *9th Int. Conf. on Autonomic and Autonomous Syst.*, pp. 40–45, 2013.

[17] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation", *IEEE Trans. on Pattern Anal. and Mach. Intell.,* vol 24, no 7, 2002, pp. 881–892.

[18] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in K-means clustering", *Int. J. of Advance Research in Computer Sci. and Manage. Stud.*, vol 1, no 6, 2013, pp. 90–95.

[19] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques", *J. of Intell. Inform. Syst.*, vol 17, no 2/3, 2001, pp. 107–145.

[20] J. M. Gutierrez, M. Imine, and O. B. Madsen, "Network planning using GA for regular topologies", in *Proc. of IEEE Int. Conf. on Commun.*, pp. 5258–5262, 2008.