IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 29, 2020, accepted March 2, 2020, date of publication March 6, 2020, date of current version March 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2979015

Quantum GIS Based Descriptive and Predictive Data Analysis for Effective Planning of Waste Management

IMRAN, SHABIR AHMAD, AND DO HYEUN KIM
Department of Computer Engineering, Jeju National University, Jeju 63243, South Korea

Corresponding author: Do Hyeun Kim (kimdh@jejunu.ac.kr)

This work was supported in part by the Energy Cloud R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, under Grant 2019M3F2A1073387, and in part by the Institute for Information & Communications Technology Planning & Evaluation (IITP) funded by the Korea Government (MSIT) under Grant 2019-0-01456, AutoMaTa: Autonomous Management framework based on artificial intelligent Technology for adaptive and disposable IoT.

ABSTRACT Waste has a direct impact on human health and the surrounding environment. Apart from the health aspect, many industries' growth is effected by waste material such as the food industry. Waste management authorities are interested in reducing the cost of waste management operations and searching for sustainable waste management solutions. For effective planning of waste management, reliable data analysis is required to produce results that can facilitate the planning process. Data mining and machine learning-based data analysis over the waste data can produce a more detailed, and in-time waste information generation, which can lead to effectively manage the waste amount of specific area. In this paper, a descriptive data analysis approach, along with predictive analysis, is used to produce in-time waste information. The performance of the proposed approach is evaluated using a real waste dataset of Jeju Island, South Korea. Waste bins are virtualized on its actual location on the Jeju map in Quantum Geographic Information Systems(QGIS) software. The performance results of the predictive analysis models are evaluated in terms of Mean Absolute Error(MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error(MAPE). Performance results indicate that predictive analysis models are reliable for the effective planning and optimization of waste management operations.

INDEX TERMS Waste monitoring, QGIS, descriptive analytics, predictive analytics, waste management, data analysis.

I. INTRODUCTION

Waste management is a collective name for all the processes carried out in the waste collection such as waste collecting, waste monitoring, and waste disposal, and recycling of waste. Waste management authorities should ecologically perform these waste management operations. One aspect of effectively planning waste management operations is to reduce the cost. Besides cost, authorities are interested in sustainable solutions for waste management to reduce the harmful effects of wastes on the health of humans, animals, and the environment. Another aspect of effectively planning waste management is to reduce the generation of the waste amount and use fewer resources in the transportation, disposing of, and recycling of wastes. Various waste management techniques have been developed over the years for waste management, [1], [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Navanietha Krishnaraj Rathinam.

Due to the emergence of fields such as data science, almost every problem is addressed using intelligent mechanisms integrated into data analysis. Data mining and machine learning-based hybrid mechanisms, [3] if applied to waste data, can produce a more detailed, and in-time waste generation information, enabling waste management authorities to manage the waste amount of specific area effectively. The generated waste information is based on historical waste data and future forecasts of the patterns in waste data. One of the challenges in Effective planning for waste management is Spatially and temporally accurate waste data.

Monitoring of reliable spatial and timely accurate waste data along intelligent data-driven insights will be helpful in the planning of waste management operations. Waste management principles more specifically rely on the proper allocations of waste bins, generated waste amount monitoring and analysis, waste storage capacity of the waste management sites [4]. Some more populated areas will require more waste bins than the other areas. Similarly, more resources will be required if bins are not properly allocated, such as time to travel to the waste bins, and fuel cost. Studies suggest that efficient waste collection requires finding the optimal locations, an optimal number of waste bins, the number of garbage trucks, and cleaners required in the waste collection operations [4].

Geographic Information System (GIS) application has the capability of multi factors analysis on spatial data, providing visualization and insights on the data in real-time. Therefore GIS applications in waste management systems are used since the very beginning of the GIS Technology [5]. GIS applications in waste management are used to analyze the collection of waste, transportation, [5], [6]; landfill sites selection and monitoring [4], [7]–[9]; and waste management [4], [5]. GIS tools can enable the waste authorities to reduce time and financial resources in waste management operations [4], [7]. Therefore, waste management authorities can utilize GIS tools to model components of waste management for the optimization of waste management operations in terms of costs and quality.

The raw dataset used for this study contains waste data from 38 locations in Jeju Island,South Korea. We are utilizing the mapping capability of QGIS along with predictive analysis methods, to create useful in-time information for waste management authorities. We virtualized the waste bins in QGIS as boxes in the respective locations of the bins. Each waste bin visualizes information such as waste amount, the population of the area, demographic information of the area, and the correlation between population and waste amount. The contributions of this research study as follows:

- Jeju based waste dataset enhancement for further waste analysis studies.
- Descriptive analysis of location-based waste amount.
- Descriptive analysis of seasonal waste amount.
- Descriptive analysis of the number of cleaners and the number of trucks.
- Time-series predictive analysis of location-based waste amount.
- Time-series predictive analysis of seasonal waste amount.
- Time-series predictive analysis of the number of cleaners and the number of trucks.
- QGIS based waste data visualization.

The rest of the paper is structured as follows. Section II is related work highlighting relevant contributions to the field of waste management. Section III explains the proposed approach of the study. Section IV discusses descriptive analysis, whereas section V discusses predictive analysis. The conclusion of the study and recommendations of future work is presented in section VI.

II. RELATED WORK

Distributed data can provide sustainable solutions if data integration and channelization techniques are applied. Data fusion must be done through advanced data fusion techniques to make it interpretable for both humans and machines [10]. World population statistics show that the population dwells in urban areas. The urban areas' infrastructure needs to be feasible for a healthy life [6]. Waste management directly affects industries, healthcare, and other aspects of life. One of the challenges in sustainable solutions for urban life is waste management for both developing and developed countries. Solid waste generation is increasing day by day due to the migrations from rural areas to urban areas [11]. Inefficient management of waste can lead to serious environmental problems, health issues, and increase the cost factor.

Waste management authorities need to consider the number of users who generate waste data, waste amounts, and stakeholders. It is essential to manage waste smartly, and notify the status of waste collection, provide analysis on the waste data to all the stakeholders on time. Keeping all these considerations in mind, Aazam M *et al.* proposed a cloud-based mechanism for smart waste management. Waste bins are embedded with sensors to notify the waste level to the stack holders through a cloud-based platform. This platform helps the waste management administration of urban areas to calculate the optimal route for waste collection, reducing the cost of waste collection operations.

Ravi, Swarnalakshmi *et al.* proposed the concept of a reliable waste management portal that collaborates with other waste collectors and satellite units, for efficient and fast processing of the waste in Smart cites [12]. Waste management authorities use the classification information to on time, notify waste amounts from waste bins. The classification algorithm classifies the wastes into wet, dry, normal, and hazardous categories. Due to the advancement of the IoT technologies, everything can be remotely monitored, integration of machine learning and big data analysis techniques can provide intelligent insights on data that can be considered to develop sustainable solutions.

In literature, some of the studies considered developing policy for waste recycling part of waste management. Moh and Manaf [13] discuss the waste recycling policy in Malaysia. The discussion concludes that most countries with an emerging economy use landfilling. Landfilling results in limitation of space, causing harmful effects on the environmental and savior health issues. Al-Jarallah and Aleisa [14] study characterize solid waste in Kuwait., discussing the average daily waste amount is 1.01 kg per person. Organic waste comprises 44.4 percent of the total waste amount; among the other noteworthy types of waste are 11.2 percent is film, and 8.6 percent is corrugated fibers. A smart way for notification of the waste amount can involve the stack holders effectively and hence efficiently manage the overall waste management mechanism. Bing *et al.* [15] summarize their studies with the same conclusion who investigated waste management mechanisms for European countries.

Zhang *et al.* [16] mention the applicability and importance of IoT in smart cities for industries such as the food industry. It is vital to analyze food waste, monitor which food is con-

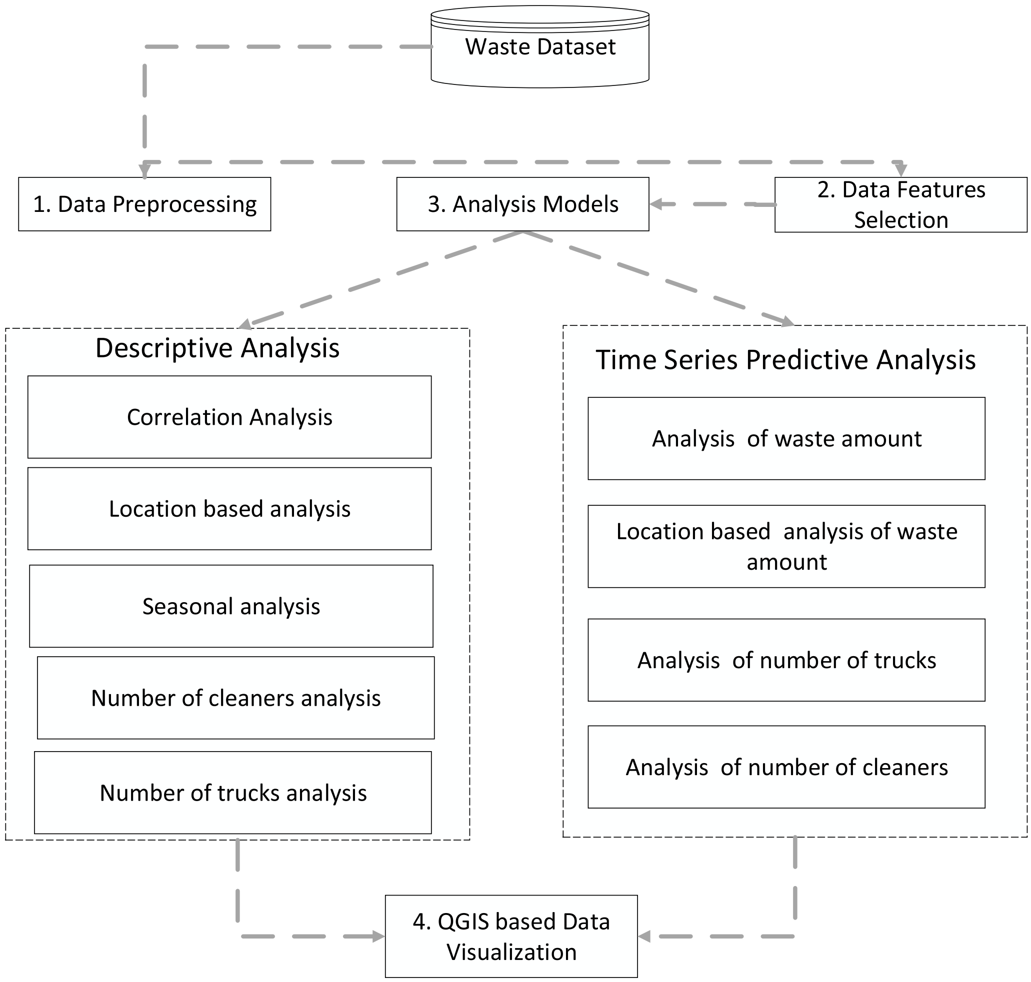

FIGURE 1. **Proposed methodology.**

sumed more by customers for the management and operation of the food industry. Mujki *et al.* proposed a system for the detection of the states of the waste-container. [17]. The status information describes the fullness of the container, status of the battery, and the covers' position. The data transmission mechanism to the broker is based on ESP8266-12E, a Wi-Fi module, where a waste management application provides an analysis of the transmitted data. Message Queue Telemetry Transport (MQTT) is used for a communication protocol that operates autonomously, with a source of battery or solar power.

An electronic system was designed by Reshmi *et al.* [18] is a solution to dispose of irregular waste. The system is embedded with sensors such as biosensors, weight, and height sensors to detect the fullness and overflow of the waste bins. The sensing data is utilized by the controller to calculate the status of the waste bins, and send notification of the waste bins status to the stack holders with the help of the GSM module. The system can achieve effective waste management with low energy costs by utilizing the solar energy system for power up the electronic system.

Guerrero *et al.* [19] provide a detailed review study of waste management carried in twenty-two countries of three continents. The study concludes that all the factors of waste management systems and stakeholders are affected by the mechanisms of waste collection, waste classification, transportation of waste, recycling, and disposal of waste. The authors' emphasis on smart and sustainable solutions is essential for the efficient handling of all the processes of waste management, including notification of the waste amount and for the recycling mechanisms, intelligent timely generated insights on waste data.

III. PROPOSED METHODOLOGY

The proposed methodology for waste analysis is consists of a series of steps given in Figure 1. Dataset preprocessing removes missing data, manages outliers, and standardized the dataset. Data features selection select useful data features for the analysis models. The descriptive analysis includes further steps such as correlation analysis of the dataset features, location-based waste amount analysis, seasonal waste amount analysis, calculation of numbers of cleaners, and the number of trucks. Predictive analysis uses time series prediction models for the analysis of waste amount, location-based analysis of waste amount, analysis of the numbers of trucks,

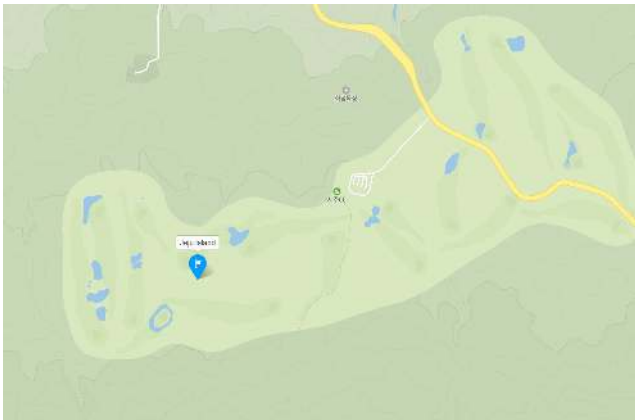

FIGURE 2. Study area-Jeju Island South Korea.

and analysis of numbers of cleaners. Lastly, we use QGIS open-source GIS software for the visualization of descriptive and predictive analysis; now we discuss the waste dataset and study area.

A. STUDY AREA AND DATA

The study area selected is 38 locations in Jeju, a self-governing province of South Korea. Figure 2 displays a Naver based map of the study locations of waste. Datasets features can be categorized as follows: features that are useful for grid visualization such as Grid ID, Grid top, Grid bottom, Grid center, Grid left, and Grid right. Population data are categorized into data by gender; population data is also classified into age classes, each class representing an age threshold. Population and its sub-features are representing the inhabitants living in the area of a waste bin by different dimensions.

Data wise distribution is the month-wise distribution of the waste data of 24 months of the years 2017 and 2018. RFID is the use of electronic tags to identify and track waste bins. The dataset also contains day-wise waste data of a week; more features are listed in Table 1. Waste data is also arranged by months, starting from January 2017 and ending December 2018. These 24 months features can be used to derive a few other features such as seasons wise waste data, waste data by year such as 2017 yearly waste, and 2018 yearly waste data. From the day-wise distribution of the waste amount, we extract features such as weekdays waste amount and weekends waste amount. From latitude and longitude, we are extracting features such as location name and postal code. Grand total feature is a cumulative feature for the sum of the total waste amount of 2017 and 2018 waste data amount.

B. DATA VISUALIZATION ON MAP

QGIS is a GIS application, and GIS stands for 'Geographical Information System.' We can use a GIS application to deal with spatial information on a computer. With a GIS application, users can open digital maps on their computers, create new spatial information to add to a map, create printed maps customized to their needs, and perform spatial analysis. In this study, we use QGIS as a visualization environment.

In Figure 3 boxes based on the grid information such as: left, top, right, bottom xcoord, ycoord are drawn, representing waste bins. Waste bins are virtualized using boxes on Jeju map in the respective locations of waste bins. Waste bins' box represents the Population and Waste data amount and its correlation. Each waste bin is clickable and displays the waste information in the feature attribute window.

TABLE 1. Jeju Island, waste data analysis dataset.

Feature	Description
Grid ID	Unique ID of the grid box
Population In Grid	Population count in a grid
Single Households	House holds Members count
Household	Single member house hold
Population by Person	Persons count In population
Male Population	Male population count
Female Population	Female popolartion count
Date wise distribution	Population by date
Population by age	Population by age groups
Sunday	Waste amount of Sundays
Monday	Waste amount of Mondays
Tuesday	Waste amount of Tuesday
Wednesday	Waste amount of Wednesday
Thursday	Waste amount of Thursdays
Saturday	waste amount of Saturdays
Friday	Waste amount of Fridays
Latitude	Latitude of the location of waste bin
Longitude	Longitude of the location of waste bin
Waste amount	Waste amount of the year 2017
Waste amount	Waste amount of the year 2018
RFID	RFID of the Waste bin
Grid left	Left dimension
Grid top	top dimension of grid
Grid right	right dimension
Grid bottom	Bottom dimension of grid
Grid center	X axis from origin
Grid center	Y axis from origin

C. DATA PRE-PROCESSING

In the process of the data preparation first, we analyzed the dataset for the missing values; One approach for handling missing data can be that we remove the whole record of the data, but we cannot take the risk of losing valuable information. For the numeric records such as waste amount by days and months, we filled the missing value with the mean value. For missing values in the latitude and longitude field, we filled the missing values by most commonly occurring values.

Dataset's categorical features, such as location, contain 38 locations in textual form. Machines are good at processing numbers. Therefore, we convert the text data into numbers, so these 38 locations' names are encoded into numbers for building machine learning models. The training process of predictive models can be improved by standardizing the dataset. In the process of standardizing our dataset, we only remove that information, which is not needed. There were duplicates data records in the dataset is collected from various data sources, so we remove duplicates records.

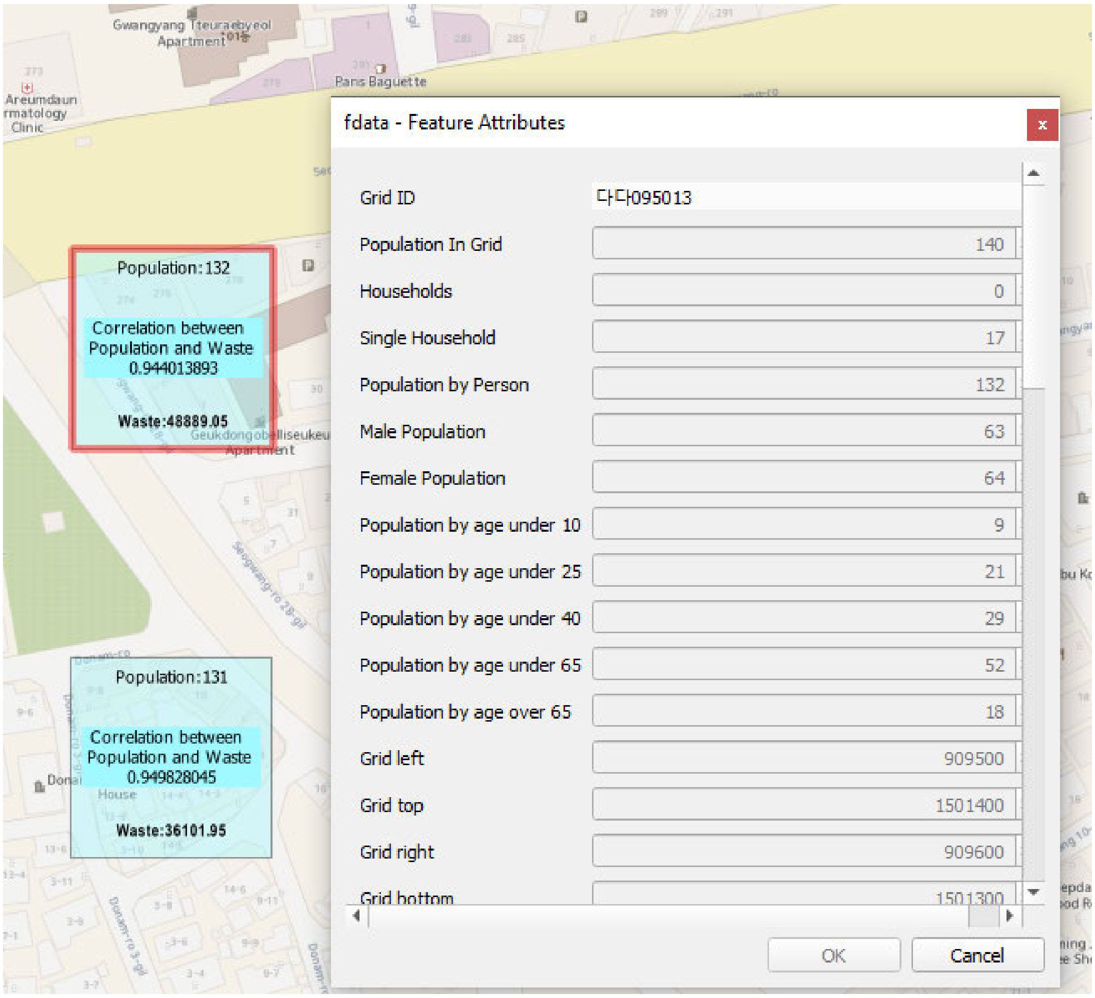

FIGURE 3. QGIS Naver map based waste bins in Jeju Island.

D. DATA FEATURES SELECTION

Features extraction is the process of calculating more features from the existing one. In our case, we calculated, year-wise, location-wise, weekdays, and weekend waste data. Location is extracted from the latitude and longitude information. Weekdays and weekend waste data are calculated from the seven days of the week. Year-wise features are calculated from the waste amount by months. Seasons wise waste amount is calculated from the month-wise waste amount. Data feature selection is the process of selecting useful features for the analysis models. Table 2 represents the extracted features from the dataset.

IV. DESCRIPTIVE ANALYTICS

Descriptive analysis steps are carried out to summarize or turn data into meaningful information. In other words, it summarized what has occurred. Waste based descriptive analysis will have some meaningful impact but will not be much helpful in waste amount forecasting. In this section, we will use descriptive analysis approaches for processing and examining our waste data sets. The analysis will obtain the correlation between dataset features, yearly location-based waste analysis, yearly seasonal analysis, numbers of cleaners, and garbage truck information. This waste analysis information is visualized in QGIS, enabling waste management authorities to make more informed, real-time, and pragmatic waste management decisions.

A. CORRELATION ANALYSIS

Correlation between two or more features explains the relationship between the features and their statistical significance. For example, if two features have the same correlation with the feature we are predicting, then we can drop one of the features. If the correlation between two or more features is positive, it means that an increase or decrease of the value in one feature will lead to an increase or decrease in the other feature. A negative correlation is the opposite of the positive correlation, so an increase in the value of one feature will lead to a decrease in the value of the other feature and vice versa. We calculated the correlation between features in three phases 1) correlation between population, genders, and waste amount. 2) correlation between days and

TABLE 2. Extracted features from waste dataset.

Feature	Description
Population by age under 25	Population by age group under 25
Population by age under 40	Population by age group under 40
Population by age under 65	Population by age group under 65
Population by age above 65	Population by age group above 65
Average waste per day	Average waste per day
city	location where the waste bin is located
Waste amount –2017	Waste amount of the year 2017
Waste amount-2018	Waste amount of the year 2018
cleaners	Represents the number of cleaners
Trucks	Represents the number of trucks
Month wise distribution	Waste amount by Month
winter 2017	Waste amount of winter season of 2017
spring 2017	Waste amount of spring season of 2017
summer 2017	Waste amount of summer season of 2017
autumn 2017	Waste amount of autumn season of 2017
winter 2018	Waste amount of winter season of 2018
spring 2018	Waste amount of spring season of 2018
summer 2018	Waste amount of summer season of 2018
autumn 2018	Waste amount of autumn season of 2018

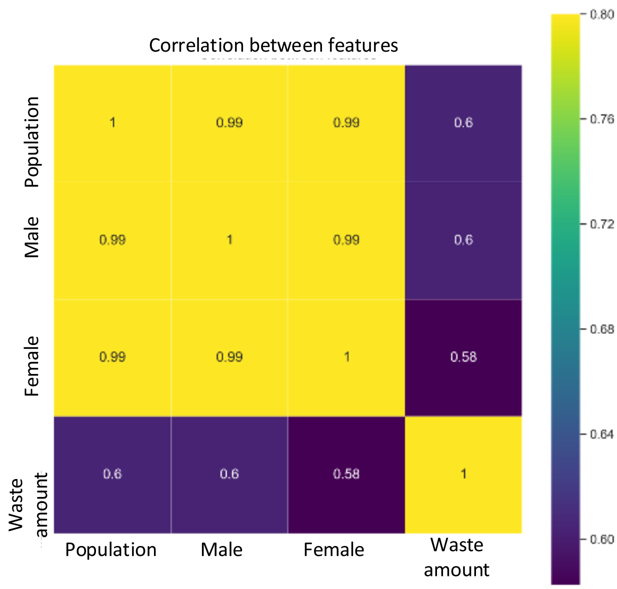

FIGURE 4. Correlation between population, gender and waste amount.

waste amount. 3) correlation between population, age groups, and waste amount.

The first phase of correlation analysis considers population, male, female, and waste amount. Pearson correlation algorithm is used to generate a correlation matrix. Figure 4 is the heatmap representation of the correlation matrix.

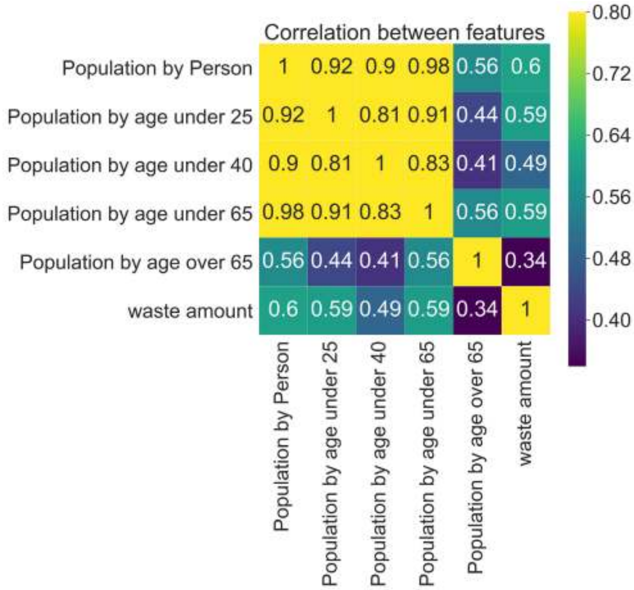

FIGURE 5. Correlation between age groups, waste amount and population.

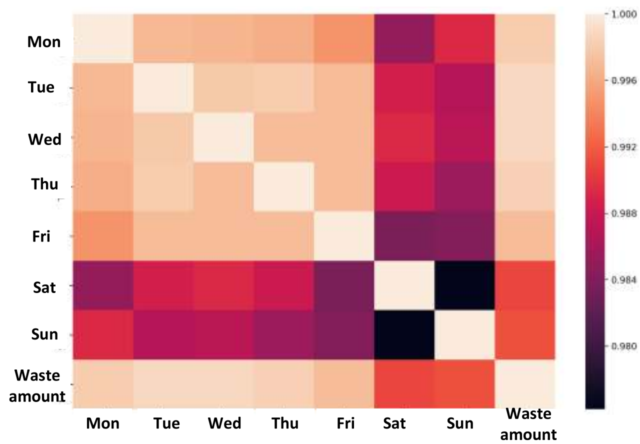

FIGURE 6. Correlation between days and Waste amount.

A heatmap is a graphical representation of data in which data values are represented as colors.

The second phase of the correlation considers population, waste amount, and populations by age groups such as 25,40,65. Figure 5 represents the correlation between age groups, waste amount, and population.

The third phase of correlation considers population and the waste amount by days of the week. The Correlation between days and Waste amount as given in figure 6. After correlation analysis, we performed a location-wise analysis of the waste amount, which forecasts the waste amount based on 38 locations. A seasons-wise analysis is carried out to forecasts the waste amount based on the four seasons of Korea: summer, winter, spring, and autumn. In the following sections, we present location and season wise analysis in detail.

TABLE 3. Jeju Island locations wise waste data amount.

Location	Waste Amount-2017	Waste Amount-2018
Noh Hyung Dong	2633713	2679536.5
Il-do-dong	825061.65	788867.8
Ildo 2-dong	780036.75	801078.55
Ido-dong	224813.1	212358.95
Ido 1-dong	246903.7	234193.85
Ido 2-dong	2232494.05	2340605.5
Donam-dong	1018883.8	1125879.7
Samdo 1-dong	416690.9	320762.15
Samdo 2-dong	276334.6	287840.65
Yongdam 1-dong	1009461.15	1031418.95
Yongdam 2-dong	268525.65	285305.95
Keonip-dong	425062.6	421195.5
Hwabuk-dong	1186413.55	1182788.25
Hwabuk 2-dong	95914.9	98052.45
Samyang-dong	1064449.05	1096983.2
Bonggae-dong	28797.45	30767.3
Hoechon-dong	88534.2	95456.25
Aradong	134658.7	137043.25
Wolpyeong-dong	9184.95	9677.05
Yeongpyeong-dong	70512.3	152880.95
Ora-dong	201995.8	233184.6
Ora 1-dong	423428	388429.45
Ora 2-dong	29078.25	34480.45
Ora 3-dong	218067.5	220830.45
Oedo-dong	752519.5	782640.75
naedo	151146.55	168507.45
Dopyeong-dong	410898.8	416925.35
Iho-dong	206015.5	241469.15
Dodu-dong	402139.55	411317.95
Samdoy-dong	284745.5	266727.95
Samdoil-dong	411517.6	479484
Yeon-dong	1768608.25	1822128.365
Ido Il-dong	520336.15	520000.25
ildo-e-dong	699651.908	699569
Samdo 2-dong	183166.45	16544
7.95 Aewol-eup	69953.4	68318.05

B. WASTE AMOUNT ANALYSIS BASED ON LOCATIONS

Locations are extracted from the latitude and longitude information through reverse geocoding. The API we developed is based on a python library called reverse-geocode. The API Returns the city location of the Jeju province based on the latitude and longitude value. Table 3 displays the location-wise waste distribution of Jeju South Korea.

Figure 7 represents the location-based analysis of Jeju Island waste. The X-axis represents locations, whereas the Y-axis represents the waste amount. Among the given locations, Noh Hyung Dong is producing maximum waste, whereas Wolpyeong-dong location is the minimum waste producer.

C. WASTE AMOUNT ANALYSIS BASED ON SEASONS

Korea has four seasons, summer, winter, spring, and autumn. Winter in Korea starts in December and ends in February. Winters are generally cold and dry, very dry. Although there used to be a lot of snow, that's not the case these days. Spring starts in March and ends in May. Spring usually announces its arrival with bright yellow forsythia flowers (kanari) along streets and highways. The summer starts in June and finishes in August. Summer is one of the proofs that Korea is a land of extremes in that it is as hot and humid, and the Korean winter is cold and dry.Autumn starts in September and finishes in November.

Figure 8 represents Season Wise Total waste Analysis. The X-axis of the graphs represents seasons such as winter, spring, summer, and autumn of Korea for the year 2017 and 2018. Y-axis represents the total waste amount(in Tons) per season. The figure shows that for the year 2017 spring season waste amount is less than the other seasons, whereas for the year 2018 autumn season waste amount (in Tons) is less than the other seasons. The maximum amount of waste is generated in the Summer season.

D. CALCULATION OF NUMBER OF TRUCKS

We can formulate an equation for calculating numbers of trucks from a simple assumption, that average garbage trucks can handle 14 tons waste. Let us assume W represents the waste amount in tons, and T_n represents the number of trucks. Equation 1 can be used to calculate the number of trucks from the waste amount. If cleaners are four, we assign them one truck.

$$T_n = round(w/4) \tag{1}$$

Figure 9 represents the calculated number of trucks using Equation 1 and number of trucks are represented at the secondary Y-axis for the Waste Amount given in the primary Y-axis.

E. CALCULATION OF NUMBER OF CLEANERS

In this section, we present how we calculated the number of cleaners from the dataset. For the number of cleaners, we formulate an equation for calculating the numbers of cleaners. Let's assume C_n represents the number of cleaners. T_n represents the number of trucks. Jeju policy for waste management' assigns 1 truck to 4 cleaners. Now we can calculate the number of cleaners C_n from T_n using equation 2.

$$C_n = round(T_n/4) \tag{2}$$

Figure 10 represents the waste amount per day and the number of cleaners needed for the handling of that waste amount. The primary Y-axis represents the waste amount in tons, whereas the secondary Y-axis represents the number of cleaners.

V. PREDICTIVE ANALYSIS

Products are turned into waste due to poor policies of the manufacturer of the product and the use of products by consumers. In the case of the food industry, the relationship between customers, food products, and the environment should be understood to reduce waste generation. This relationship is complicated, and hence statistically approximating the relationship is a challenging task. Waste monitoring data collected by the waste management authorities enables them to analyze the relationship by factors such as in which days waste is generated more. Other aspects considered are the

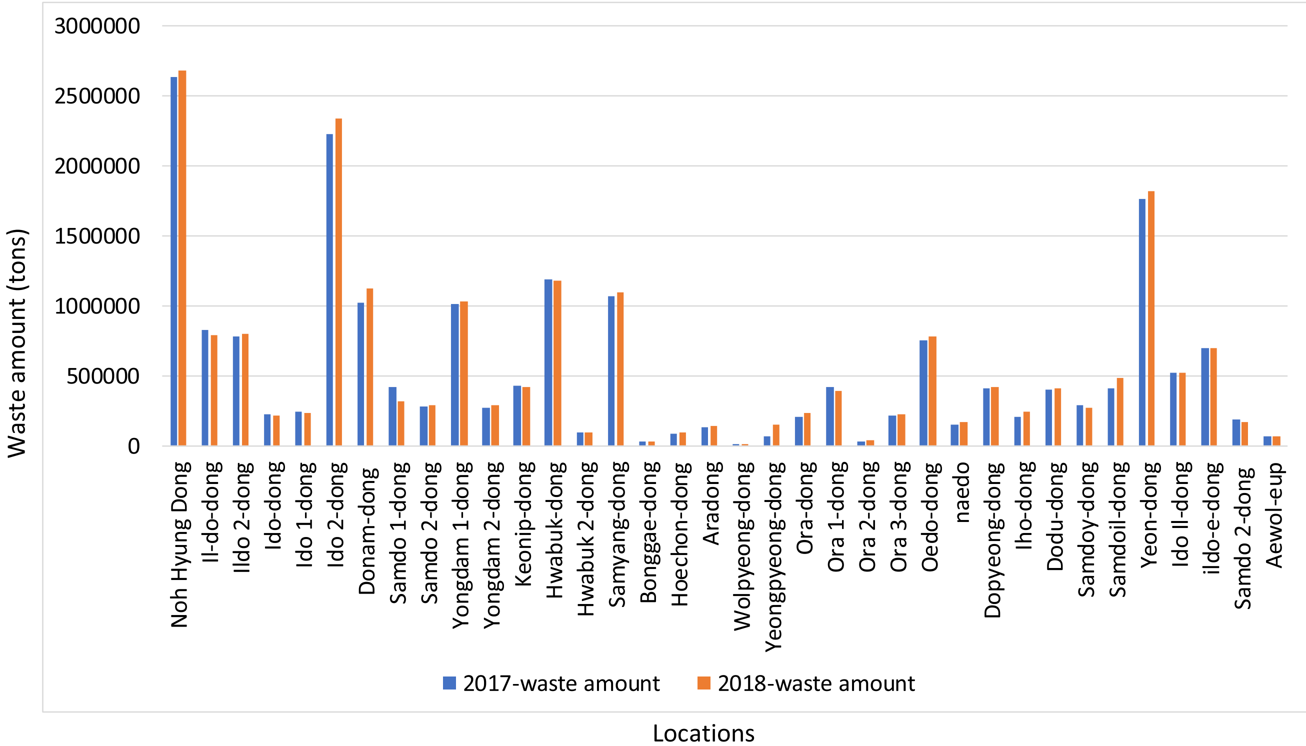

FIGURE 7. Location wise waste analysis.

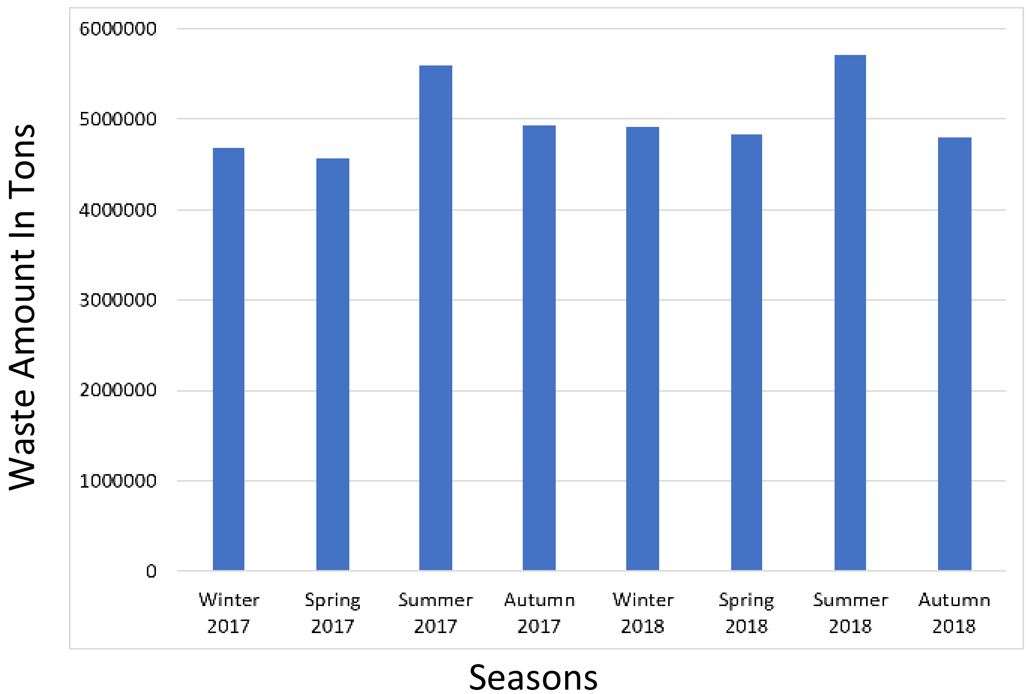

FIGURE 8. Season wise total waste analysis.

waste amount by population, season, location of the waste bins. This analysis can lead to advanced predictive analytics and building models of the spatial and temporal patterns in the waste data. Building a temporal model will require on the day wise, month wise, and year wise collection of waste amount. From the data, we can also understand patterns such as seasonal variations of the waste amount. In this study, we implement predictive time series based machine learning regression models on our dataset to forecasts all these aspects.

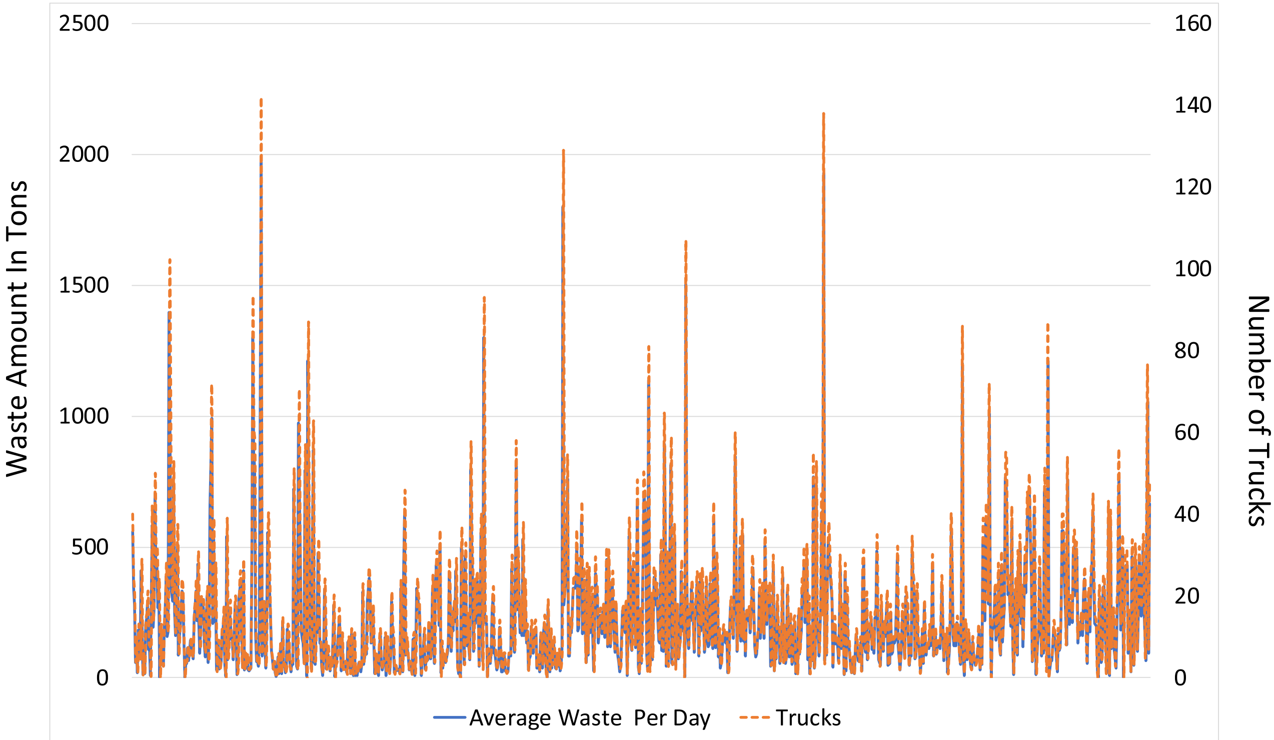

FIGURE 9. Number of trucks vs waste amount.

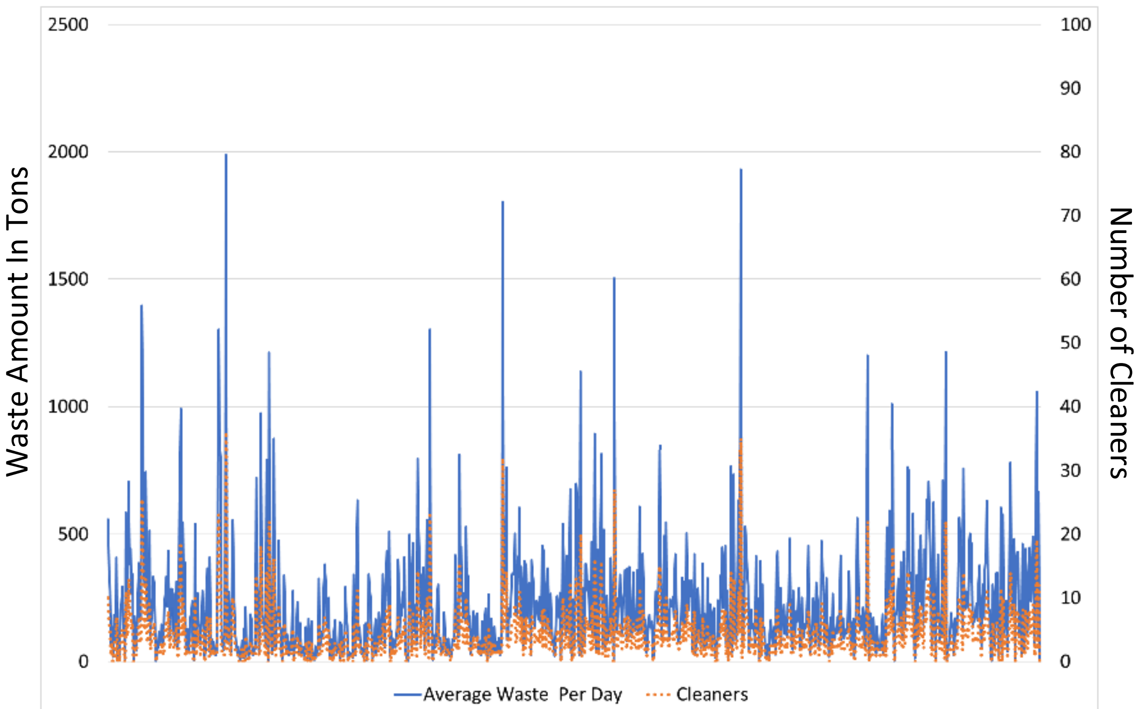

FIGURE 10. Waste amount in Tons and number of cleaner.

A. TIME SERIES PREDICTION OF WASTE AMOUNT

A time-series waste amount prediction is a series of waste amount features instances ordered by time such as dates. In a time series waste amount prediction, time is the independent variable 'X', and the prediction model goal is to forecast the waste amount 'Y' for a specific time t in the future.

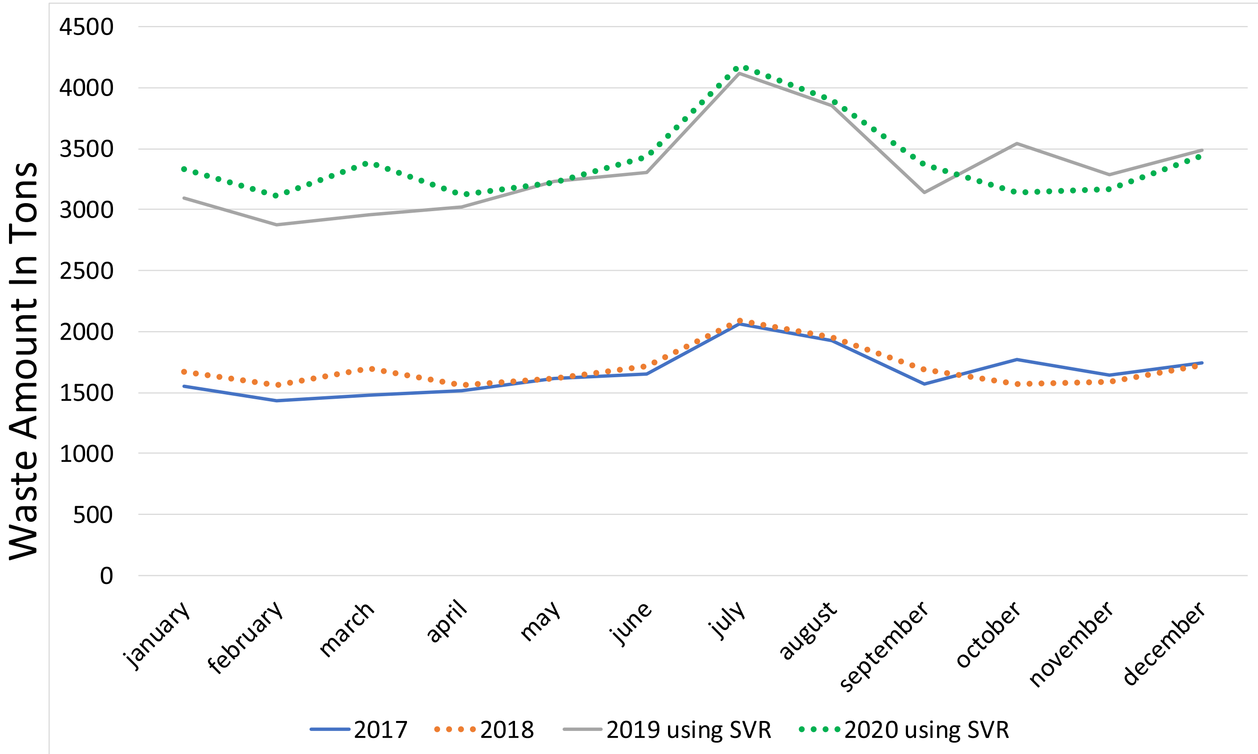

FIGURE 11. Time series prediction of waste amount.

We build Support Vector Machine Regression(SVR) based time series model on the dataset, which contains month-wise data of 2017 and 2018; based on the waste data of 2017 and 2018, we predicted month-wise waste amounts of 2019 and 2020. The reason for selecting SVR model is its small MAPE value. Figure 11 represents the predicted waste amount for the year 2019 and 2020 for the original waste data of 2017 and 2018.

B. TIME SERIES PREDICTION OF NUMBER OF TRUCKS AND CLEANERS

Once we have calculated the number of trucks and cleaners for the existing time series data of 2017 and 2018, next, we trained SVR based time series model for prediction of number of trucks and cleaners in near future. The model predicts the number of trucks and cleaners for the year 2019 and 2020. Figure 12 shows the time series prediction of the number of cleaners and trucks using SVR. Y-axis represents the number of cleaners and trucks, whereas X-axis represents months of the years. Cleaners are visualized using a bar chart, whereas the line chart is used to visualize the number of trucks. Cleaners-2017 and cleaners-2018 are the number of cleaners in the existing dataset, whereas cleaners-2019 and cleaners-2020 are the predicted number of cleaners.

C. TIME SERIES PREDICTION OF WASTE AMOUNT BASED ON LOCATIONS

In this section, we present location-based time-series predictive analysis on the dataset for prediction of the future waste amount for a specific location. The location-based waste analysis will enable the waste management authorities to develop policies for specific locations by deploying the needed waste bins in the grid area based on the waste forecasts. This will also help in route optimization for the garbage truck and cleaners management. For forecasting location-based waste analysis, we trained SVR time series regression model with the existing data of 2017 and 2018. SVR forecasts the waste amount based on specific locations. Figure 13 shows the time-series prediction of the waste amount based on locations; location ido-2-dong is the maximum waste producer, whereas wolpyeong-dong is the least waste-producing location. The X-axis of the figure displays location, whereas Y-axis represents the waste amount in tons.

D. PERFORMANCE ANALYSIS

In this section of the report, we compare the performance of the regression methods; Table 4 represents Comparison of regression methods in terms of waste amount prediction. Among the best regression methods for waste, prediction top are SVR, Random forest and LassoLars Regression. We test the models using three performance matrices for regression, MAE, RMSE, and MAPE for each model. Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Mean Absolute Error (MAE) is the average vertical distance between each point and the identity line. The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of prediction accuracy of a forecasting method in statistics.

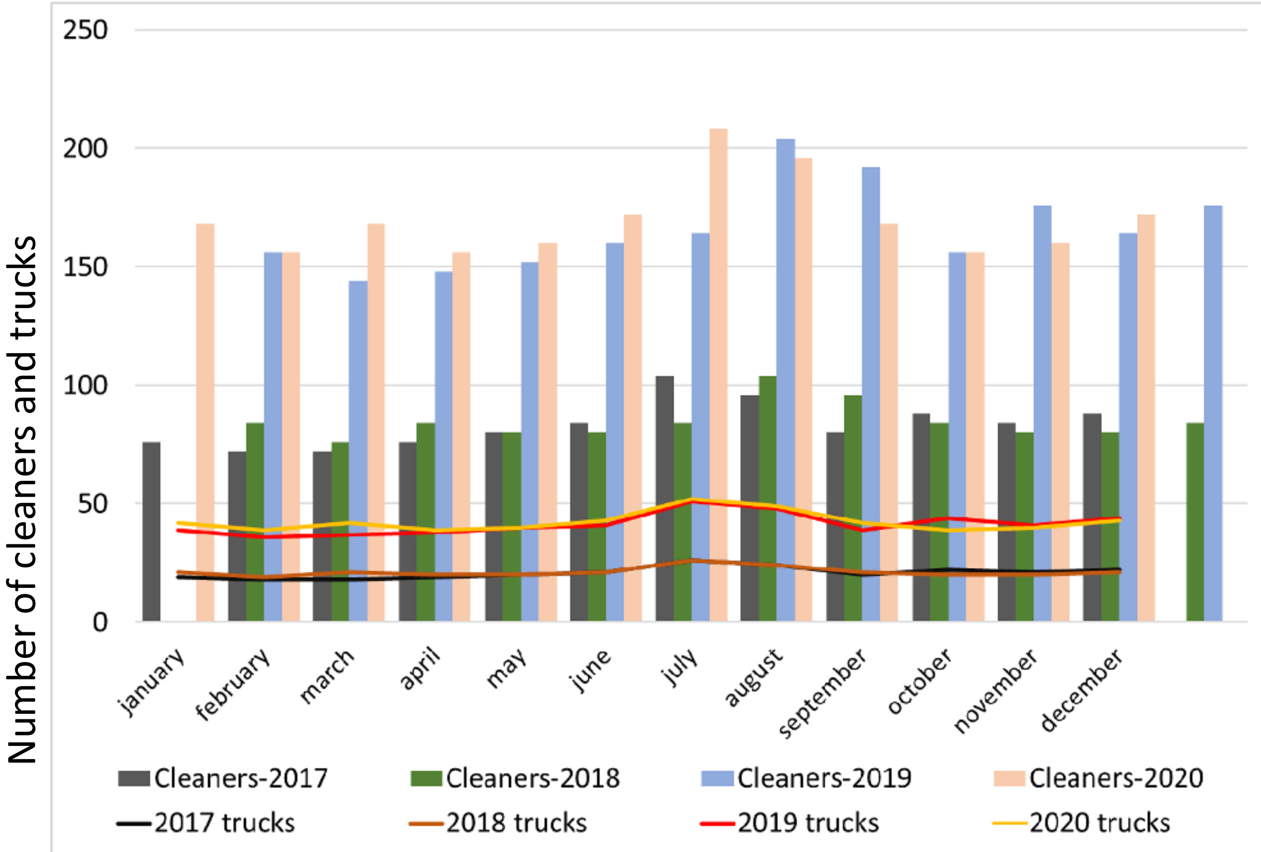

FIGURE 12. Time series prediction of number of trucks and cleaners.

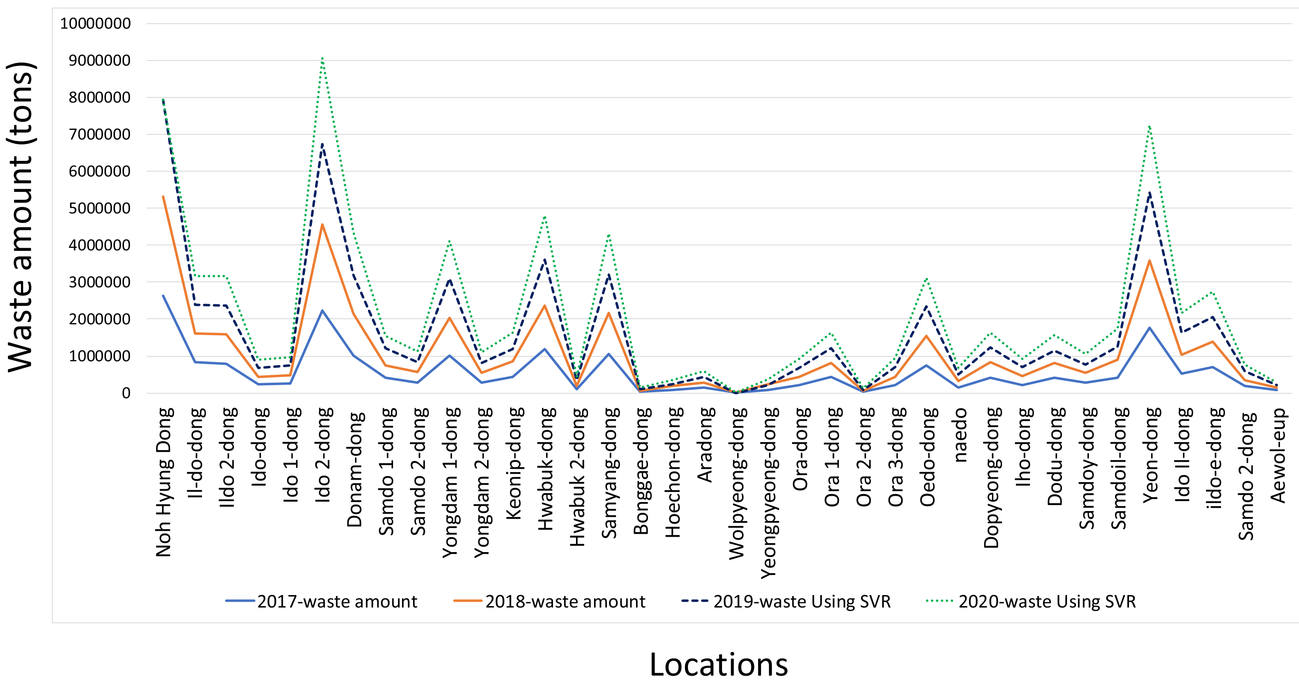

FIGURE 13. Time series prediction of the waste amount based on locations.

TABLE 4. Performance analysis using RMSE, MAE and MAPE.

RegressionModel	MAE	RMSE	MAPE
LassoLars	5315.061	6725.638	29.845
Linear	6006.247	7274.950	41.749
RandomForest	6423.536	8026.389	45.606
SGDR	7332.597	8796.588	60.140
SVR	976.412	1112.150	7.385
TheilSen	6160.402	6986.908	42.904
PassiveAggressive	5753.969	6860.840	40.332

1) MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

MAPE is used to compute an average deviation found in waste amount value from actual waste amount value. MAPE is calculated by dividing the sum of absolute differences between the actual waste amount and predicted waste amount by the machine learning algorithm we applied in this study with the total number of waste data records such as n.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{e_t}{y_t} \right| \tag{3}$$

2) ROOT MEAN SQUARED ERROR(RMSE)

RMSE is one of the standard method to measure the error of a prediction model in quantitative data such as waste amount.RMSE can be understood as a type of distance between the vector of predicted waste amount and the vector of original waste amount. RMSE can be obtained by simply taking the square root of MSE.

3) MEAN ABSOLUTE ERROR(MAE)

MAE is one of the most common way to measure error of prediction models. It measures performance of our regression models, which are applied on waste amount-continuous variable. It is the sum of absolute differences between our target waste amount and predicted waste amount. We use MAE to measure the average magnitude of errors in the waste amount prediction.

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t| \tag{4}$$

VI. CONCLUSION

Effective planning for the waste management needs reliable data analysis over the waste data to produce a more detailed, and in-time waste information generation. In this research, a descriptive data analysis approach, along with predictive analysis, is used to produce in-time waste information utilized by waste management authorities for the effective planning of waste management. The performance evaluation of the approach is evaluated using a real waste dataset of Jeju Island, South Korea. The QGIS-based visual analysis helps in accurately allocating waste bins and predicting an optimal route for the garbage truck. The future work of this study can be the use of the descriptive and predictive analysis results for generating an optimal route for the garbage truck. Furthermore, the analysis can allow policymakers to devise a policy that is cost-efficient and optimal in terms of financial and human resources. Some advanced time series prediction can also be considered on this dataset, such as LSTM, Prophet, and SARIMA based models and other intelligent approaches to identify the non-linear relationship between the features of the waste dataset.

ACKNOWLEDGMENT

Any correspondence related to this paper should be addressed to Dohyeun Kim.

REFERENCES

[1] E. Pongrácz, P. S. Phillips, and R. L. Keiski, "Evolving the theory of waste management-implications to waste minimization," in *Proc. Waste minimization Resour. Use Optim. Conf.*, 2004, pp. 7–61.

[2] E. Pongrácz and V. J. Pohjola, "Re-defining waste, the concept of ownership and the role of waste management," *Resour., Conservation Recycling*, vol. 40, no. 2, pp. 141–153, Jan. 2004.

[3] I. Bose and R. K. Mahapatra, "Business data mining—A machine learning perspective," *Inf. Manage.*, vol. 39, no. 3, pp. 211–225, Dec. 2001.

[4] J. Malczewski, "GIS-based land-use suitability analysis: A critical overview," *Prog. Planning*, vol. 62, no. 1, pp. 3–65, Jul. 2004.

[5] S. Leao, I. Bishop, and D. Evans, "Assessing the demand of solid waste disposal in urban region by urban dynamics modelling in a GIS environment," *Resour., Conservation Recycling*, vol. 33, no. 4, pp. 289–313, Nov. 2001.

[6] S. Madakam and R. Ramaswamy, "The state of art: Smart cities in India: A literature review report," *Int. J. Innov. Res. Develop.*, vol. 2, no. 12, pp. 115–119, 2013.

[7] V. R. Sumathi, U. Natesan, and C. Sarkar, "GIS-based approach for optimized siting of municipal solid waste landfill," *Waste Manage.*, vol. 28, no. 11, pp. 2146–2160, Nov. 2008.

[8] A. F. Lukasheh, R. L. Droste, and M. A. Warith, "Review of expert system (ES), geographic information system (GIS), decision support system (DSS), and their applications in landfill design and management," *Waste Manage. Res.*, vol. 19, no. 2, pp. 177–185, Apr. 2001.

[9] R. Vijay, A. Gautam, A. Kalamdhad, A. Gupta, and S. Devotta, "GIS-based locational analysis of collection bins in municipal solid waste management systems," *J. Environ. Eng. Sci.*, vol. 7, no. 1, pp. 39–43, Jan. 2008.

[10] T. J. Kalidoss and S. Ravi, "Disaster management system leveraging the power of emerging technologies," presented at the Oral Conf. IEEE ICCIC, Dec. 2016.

[11] M. Aazam, M. St-Hilaire, C.-H. Lung, and I. Lambadaris, "Cloud-based smart waste management for smart cities," in *Proc. IEEE 21st Int. Workshop Comput. Aided Modelling Des. Commun. Links Netw. (CAMAD)*, Oct. 2016, pp. 188–193.

[12] S. Ravi and T. Jawahar, "Smart city solid waste management leveraging semantic based collaboration," in *Proc. Int. Conf. Comput. Intell. Data Sci. (ICCIDS)*, Jun. 2017, pp. 1–4.

[13] Y. C. Moh and L. Abd Manaf, "Overview of household solid waste recycling policy status and challenges in Malaysia," *Resour., Conservation Recycling*, vol. 82, pp. 50–61, Jan. 2014.

[14] R. Al-Jarallah and E. Aleisa, "A baseline study characterizing the municipal solid waste in the state of kuwait," *Waste Manage.*, vol. 34, no. 5, pp. 952–960, May 2014.

[15] X. Bing, J. M. Bloemhof, T. R. P. Ramos, A. P. Barbosa-Povoa, C. Y. Wong, and J. G. A. J. van der Vorst, "Research challenges in municipal solid waste logistics management," *Waste Manage.*, vol. 48, pp. 584–592, Feb. 2016.

[16] Q. Zhang, T. Huang, Y. Zhu, and M. Qiu, "A case study of sensor data collection and analysis in smart city: Provenance in smart food supply chain," *Int. J. Distrib. Sensor Netw.*, vol. 9, no. 11, Nov. 2013, Art. no. 382132.

[17] F. Mujkic and A. Aksamovic, "Implementation of an electronic platform for aiding the waste management process," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2018, pp. 1517–1521.

[18] W. Reshmi, R. K. Sundaram, and M. R. Kumar, "Sensor unit for waste management: A better method for frequent data updating system," in *Proc. Int. Conf. Sci. Eng. Manage. Res. (ICSEMR)*, Nov. 2014, pp. 1–5.

[19] L. A. Guerrero, G. Maas, and W. Hogland, "Solid waste management challenges for cities in developing countries," *Waste Manage.*, vol. 33, no. 1, pp. 220–232, Jan. 2013.

IMRAN received the B.S. degree (Hons.) in information technology from the University of Malakand, KPK, Pakistan, and the M.S. degree in computer science from Bahria University, Islamabad, Pakistan, in 2018. He is currently pursuing the Ph.D. degree with the Department of Computer Engineering, Jeju National University, South Korea. His work experiences include full stack software development, IT trainings, and entrepreneurship. His research work mainly focused on the Internet of Things applications, machine learning, data science, and blockchain applications.

SHABIR AHMAD received the B.S. degree in computer system engineering from the University of Engineering and Technology, Peshawar, Pakistan, and the M.S. degree in computer software engineering from the National University of Science and Technology, Islamabad, Pakistan, in 2013. He is currently pursuing the Ph.D. degree with the Department of Computer Engineering, Jeju National University, South Korea. He is also a faculty member with the Software Engineering Department, University of Engineering and Technology. His research work mainly focused on the Internet of Things application, cyber-physical systems, and intelligent systems.

DOHYEUN KIM received the B.S. degree in electronics engineering and the M.S. and Ph.D. degrees in information telecommunication from Kyungpook National University, South Korea, in 1988, 1990, and 2000, respectively. He was with the Agency of Defense Development (ADD), from Match 1990 to April 1995. Since 2004, he has been with Jeju National University, South Korea, where he is currently a Professor with the Department of Computer Engineering. From 2008 to 2009, he was with the Queensland University of Technology, Australia, as a Visiting Researcher. His research interests include sensor networks, M2M/IOT, energy optimization and prediction, intelligent service, and mobile computing.