Towards an Optimization Model for Household Waste Bins Location Management

Moulay Lakbir Tahiri Alaoui, Meryam Belhiah, Soumia Ziti
Intelligent Processing and Security of Systems-Faculty of Sciences, Mohammed V University in Rabat, Morocco

***Abstract*—Smart cities require effective, adaptive household waste management systems due to rapid urbanization. Traditional bin placement strategies based on placing bins equidistant among residents fail to account for actual human behavior, leading to overflowing or underused bins. This paper addresses optimizing bin location and capacity through Internet of things (IoT) technologies and data-driven decision-making by deploying LoRaWAN sensors in Tangier City as a case study; real-time usage information was then collected and analyzed. Through statistical analysis and outlier detection techniques, the proposed approach identifies bin placements that are non-optimized by using statistical analysis. It also evaluates data quality and classes bins by their usage level; results show several bins were constantly overused or underused indicating that dynamic placement and capacity adjustment would improve waste collection efficiency, reduce operational costs and enhance citizen satisfaction within a Smart City framework.**

***Keywords*—Smart City; IoT; household waste; LoRaWan; bin location; outlier detection**

I. Introduction

Cities in developing nations are rapidly expanding, increasing the challenges associated with household waste management. Data collected from networks of IoT sensors placed in waste bins provide valuable data about filling levels and enable dynamic waste collection planning. Leveraging IoT networks to monitor fill rates in real time allows waste collection to be optimized without degrading the quality of service for citizens or wasting resources by emptying half-full bins.

Despite recent advances, several shortcomings remain in existing waste management systems. Many solutions rely on static routing or periodic collection schedules, ignoring real-time variations in bin usage. Prior works often lack robustness against real-world factors such as communication failures, and temporary urban events. To communicate with the servers, some studies use both IOT and GSM [1], which is expensive.

Furthermore, most studies address filling rate without integrating additional factors such as bin moisture and temperature, which are crucial for assessing waste degradation and health risks.

These limitations explain why the problem of dynamic and efficient bin management remains partially unsolved. Existing approaches either oversimplify the complexity of urban environments or fail to incorporate reliable outlier detection, resulting in inefficient resource allocation and increased operational costs [2].

This study proposes a method to optimize bin locations and dimensions based on continuous real-time data collected from sensors embedded in waste bins. The approach also incorporates an optimized routing algorithm for waste collection vehicles, aiming to minimize fuel consumption, travel time, and human resource utilization, which together represent a significant portion of municipal operating budgets [3].

The key contributions of this work are:

- A real-time data analysis framework that integrates fill rate, moisture, and temperature measurements for dynamic waste bin management.
- An outlier detection method designed to distinguish between temporary and permanent bin overflow conditions.
- A system design that considers technical constraints such as energy limitations, frequency interference, and network security in heterogeneous IoT environments.
- An evaluation showing how optimizing bin dimensions and locations can significantly reduce waste management costs.

However, limitations remain. The current system depends on the stability of wireless communication networks and the accuracy of low-cost sensors, which may introduce measurement errors under specific urban conditions.

The remainder of this paper is organized as follows. Section II reviews related works, including the IoT paradigm, IoT network architecture, LoRaWAN technology, and the challenges faced by IoT devices. It also discusses key data quality dimensions and methods to enhance the household waste collection process. Section III presents the proposed optimization model for managing household waste bin locations, including the outlier detection method, comparison metrics, and hyperparameter tuning process. It also describes the case study conducted in Tangier City, detailing data pre-analysis, quality verification, and outlier detection results. Section IV presents and discusses the results, including data pre-analysis, quality verification, outlier detection, and identification of slow- and fast-filling bins. Section V concludes the paper and suggests future research directions.

II. Related Works

A. IoT Paradigm

IoT and internet of everything (IoE) refer to a connected world where all objects are interconnected via ubiquitous sensors [4] and devices from different manufacturer's need to exchange data. The IoT economic impact could grow to $3,352.97 billion by 2030 [5] ; the number of IoT devices may reach 75 billion [6].

Globally dispersed, diverse, and heterogeneous IoT devices provide data that influence interoperability and data quality [6].

Heterogeneous networks and sensors' challenges are the origin of communication problems between multiple nodes and layers. The following sub-chapter delves into the fundamental structure, components, and layers of IoT, exploring their applications and challenges.

B. IoT Network Architecture

The IoT architecture is divided into several layers [7], each one responsible for different functions within the ecosystem as depicted in Fig. 1:

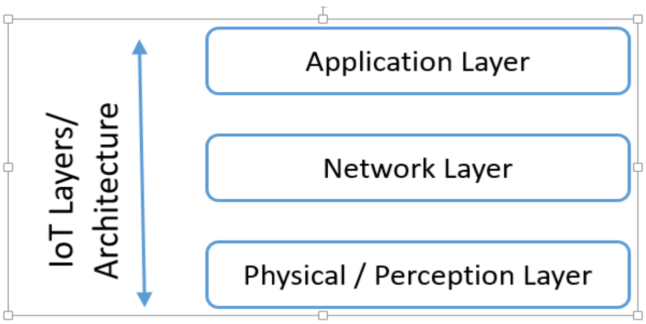


Fig. 1. IoT Architectre [7].

The physical layer, sometimes referred to as the perception layer, is in charge of actuating the environment in real time, measurement, and communication to the next layer, like temperature, bin level filling, moisture, geolocation, etc.

Maintenance of these devices poses challenges in terms of replacement and repair due to potential sensor placement in inconvenient locations [8]. This could lead to operational difficulties and even delays in data collection. Moreover, environmental conditions like high temperatures or humidity can affect sensor performance, necessitating extra care during hardware maintenance procedures. Furthermore, issues with the power supply or connectivity may make it difficult for the physical layer sensors to function. It may be necessary to regularly monitor and troubleshoot these issues in order to ensure accurate and uninterrupted data transmission. Defective equipment can lead to a malfunctioning sensor that produces inaccurate data, affecting service delivery and overall business insights.

Sensors often face limitations: being cost-effective means they aren't of the highest quality and have limited capacities. This includes issues with connectivity and short battery life for various functions, lack of precision, loss of calibration, and keeping up reporting once the device becomes faulty [9].

Noise is a major problem for sensors. The signals they rely on can be seriously disrupted by interference or physical impediments, which results in inaccurate data collection[10].

Network layer: interconnects IoT devices with the next layer [11] using universal protocols.

Application layer: controls sensors, receives data, analyzes them, and takes decisions[12].

In the next subsection, we describe the LoRaWan solution and present its benefits and drawbacks.

C. LoRaWan

LoRaWan is a member of the Low Power Wide Area Network (LPWAN) family. These devices use the medium access control protocol (MAC) mechanism to communicate with the gateway.

1) LoRaWAN Dataframe: Dataframes have the same time duration. To overcome noise and interference, LoRa uses forward error correction (FEC) codes ranging between 4/8 and 4/5 and diagonal interleaving. The symbol rate Sr depends on the bandwith Bw and the spreading factor according to the Formula (1) [13]:

$$S_r = \frac{Sp \ * \ B_w}{2^{SF}} \tag{1}$$

2) LoRaWAN Topology: LoRaWan has a star topology as per Fig. 2 [14]. Sensors can only communicate with the gateways but not with each other; gateways communicate with the server; they encapsulate raw data received from sensors in UDP/IP packets and send them to the server. The server sends downlink packets and commands. Devices are divided into three classes [15]:

Class A: has basic options needed to join a LoRaWan network. Bidirectional communication can be enabled. Class A devices are most of the time asleep, thus they consume the least of power.

Class B: more receive windows can be scheduled to get synchronized and to inform when devices are ready for downlink traffic; power consumption is higher than the first ones.

Class C: Receive windows are open continuously except during transmission, power consumption is higher for this class.

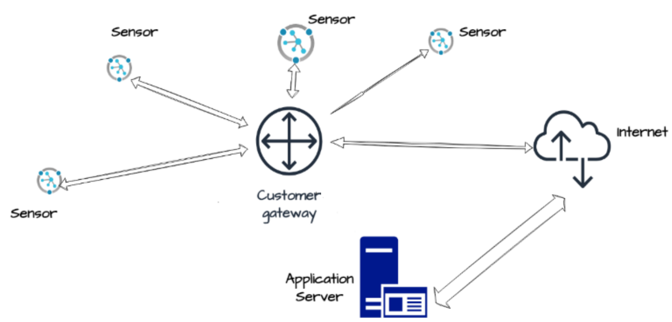


Fig. 2. LoRaWan network topology.

3) LoRaWAN transmission: The LoRa physical layer can use 125 KHz channels from the bands 433 MHz, 868 MHz, and 915 MHz to transmit, sensors communicate within 1% of the time only and transmit a small amount of data [16].

To guarantee secure and reliable communication, LoRaWan sets a number of mechanisms and joining procedures for sensors.

Once the sensor joins the network and is activated using over-the-air procedure (OTAA) [17] or by personalization (ABP) [18], the device sends a join or re-join message with needed keys and identifiers and gets a join-accept message from the server. The use of open source protocols helps to reduce the solution fees; the communication protocol MAC [19] is used between the sensors and the gateway; furthermore, to avoid financing frequency license fees, unlicensed bands may be used; the 868 MHz sub-band is unlicensed in Europe [20] and Morocco, according to the frequency regulation center [21].

Numerous technologies, including SigFox, IEEE 802.15.4g, LoRaWAN, and Z-Wave, use the same frequency, which may have an effect on signal quality and interference. In order to overcome interference, European regulations share time resources; a radio transmitting for one second cannot transmit for the next 99 seconds [22].

D. *LoRaWan Solution to IoT Devices Challenges*

With the cited information above, the LoRaWan solution overcomes most of the IoT constraints while maintaining the same quality of service:

- *Joining and re-joining procedures:* restrict the sensors allowed to send data to MGWs.
- *Power consumption*: LoRaWan is a low-power area network [15]; sensors are asleep most of the time, especially for class A devices. Sensors contain a solar panel that extends the lifespan of sensor batteries and delivers sufficient voltage to sensor units (low power transmitter). Fig. 3 shows LoraWan Sensor modules. The embedded GPS module in the device helps to inform trucks about exact geolocation and to identify the location of bins. Table I summarizes solutions to most sensor challenges and the reason behind the popularity of this technology.
- Frequency transmission fees: as per [16] free of charge frequency usage to reduce the solution cost.
- Interference using processes like listen before sending or sending 1% of the time reduces considerably the interference. Data control and preprocessing can be done on the application server.
- Distance between the sensors and the gateway can reach 6 km with a high reception rate (more than 90%) [23].
- Synchronization: Is a drawback for LoRaWan [24]
- Access to transmission channels from multiple and heterogeneous sensors is unpredictable which causes collision and loss of frames [25].

TABLE I. LoRaWan Solution to IoT Devices Challenges

Challenge	Solution	Details
Power consumption	Low power consumption	Sensors are asleep most of the time
Frequency transmission fees	Free of charge frequency usage	Used to reduce charges
Interference	Sending 1% of the time reduces considerably the interference	A limited number of devices can transmit in the same time

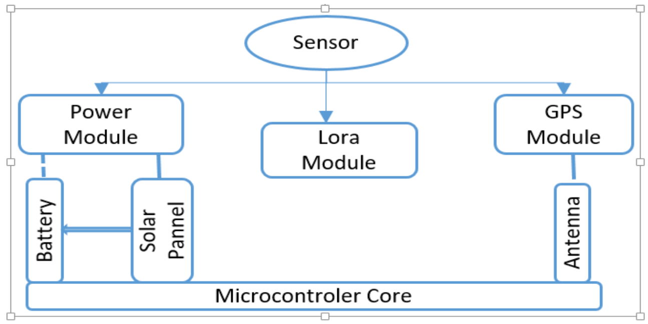


Fig. 3. LoRaWan sensor modules.

In the following section, we will present the most important dimensions regarding data quality.

E. *Data Quality Dimensions*

Data dimensions are characteristics of data quality that can reveal the data's overall quality level once they are measured correctly [26]. Data quality is a crucial parameter for services based on IoT; a control and validation of the quality are mandatory before model deployment. Herein we will define the most influencing data quality dimensions:

1) Accuracy: Evaluates the reliability, dependability, and certification of data [27]. It represents the degree to which observations are correct, trustworthy, and guaranteed error-free. It can also be defined as how a value 'v' is close to the correct value of the real world.

2) Completeness: A «NULL» value may indicate a missing value, which is an existing value in the real world but the observation is lost for a specified reason; those values may exist but are unknown, or they do not exist, or the system does not know if they exist or not [28].

3) Timeliness, volatility and currency: The temporal dimensions are sensitive since late data may be unuseful. Timeliness describes how current the data are; it can alternatively be described as the offset between the server time and the received sensor's timestamp. Volatility is a measure of how frequently data changes over time; where currency specifies how fast data are updated, it can be defined as in Formula (2) [29]:

$$\text{Currency} = \text{Age} + (\text{Delivery\_Time} - \text{Input\_Time}) \quad (2)$$

where, "age" indicates the data's original age upon receipt and "Input_time" is the server time when data are observed.

F. *Household Waste Collection Process Enhancement*

While the majority of research concentrated on management and trash classification, the current work looks at bin locations and identifies inadequate ones; it also shows outliers, including the most and least used bins; as well as inaccurate data produced by malfunctioning devices. Non-synchronized bins are sorted out to permit a full operational waste collection process. The data control part in our program can be deployed by other IoT services.

III. Towards an Optimization Model for Household Waste Bins Location Management

A. *Outlier Detection Model Presentation*

Outliers represent a rare event in a dataset; this unexpected value may be due to a measurement error or a faulty sensor, but it can also be a valuable insight [27]. In this subsection, we will present the main steps and procedures of our model that detects outliers. Data may be numbers, dates, geo-location values, etc.

Our model is built using Jupyter Notebook with required libraries and dependencies installed, like Pandas, Numpy, Matplotlib, Sickit-Learn, Pyod, Pycaret etc. The model is designed to operate in a variety of fields; it can be adapted according to each domain specification, and the following steps are performed:

- Data collection: supplied data, in text, csv, or other formats, contain multiple columns; it is filtered to keep necessary columns for our model.
- Cleanup: NaN and empty values represent a non-complete observation; it should be detected and cleaned/corrected.
- Preprocessing: data columns are interpreted as object types and need to be converted to appropriate types such as date time, integer, etc.
- Data normalization and feature selection: features should have similar scales with a low correlation level to provide better results. The data show below variables:
  - The server time (servertime) and time of the IoT device related to the observation.
  - Bin_number, Bin_index1 and Bin_index2 represent an identyer of the waste bin used by different layers.
  - Filling_level is a variable measuring the fill level of a Bin.
  - Bin_longitude and latitude represents GPS coordinates.

Time identifiers and Bin indexes are correlated as per Fig. 4. To detect outliers we use the variables bin identifier, time and fill level etc. A new variable will be introduced to classify data per day of the year.

- Data from Bin_index columns are converted to a dictionary to simplify data analysis and allow exploration of data through outlier detection methods requiring numerical values.
- Parameter tuning: we can change the ratio of data to train the model according to the data size, and other parameters can be changed according to our need.
- Pandas library helps to sort out IoT devices that overflow the server and bins rarely transmitting their filling level. We use Pycaret to detect outliers, tune, and compare the models.
- Model comparison: different outlier detection models can be compared using different metrics, as we will present in the next subsection.

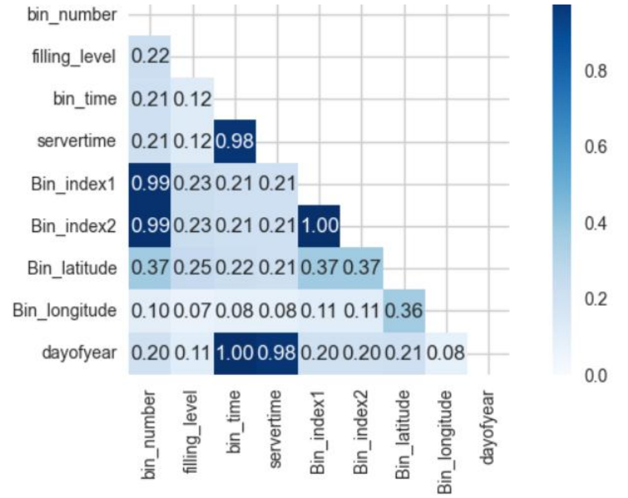


Fig. 4. Correlation matrix.

The Fig. 5 below summarizes different data treatment steps of our program:

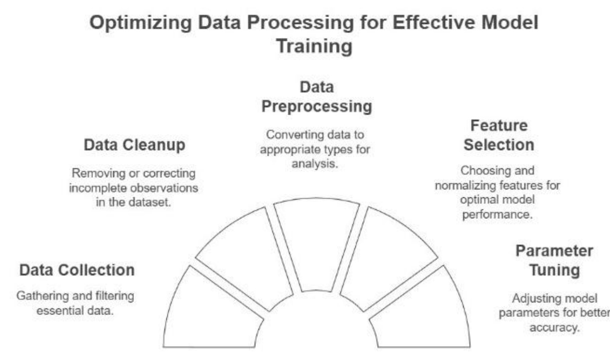


Fig. 5. Data processing.

1) Comparison metrics: The model sorts out the best method to detect outliers based on the best metrics. The main metrics used in the case of classification models are [30]:

- Accuracy: it presents the ratio of the correct prediction numbers to all the prediction numbers.

$$\text{Accuracy} = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{3}$$

$$\text{Auc}= \frac{1+\ Correct\ positif\ predicts - Incorrect\ positif\ predict}{2} \tag{4}$$

$$\text{Recall:}\quad \text{Recall} = \frac{Correct\ positif}{Correct\ positif + False\ negatif} \tag{5}$$

$$\text{Precision:}\quad \text{Precision} = \frac{Correct\ positif\ results}{Correct + False\ positif\ results} \tag{6}$$

$$\text{F1 Score:}\quad \text{F1} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad [31] \tag{7}$$

2) Outlier detection using KNN method: The number of neighbors to be given an integer K, this method calculates the distance to the k nearest neighbors. In a m-dimentional space, let A and B be two points; they can be expressed as the tulpes: (A[1], A[2],…A[m]) and (B[1], B[2]…B[m]). The distance AB can be : $\sqrt{(\sum(B[i]-A[i])2}$. An object O is an outlier if the number of neighbors within a distance r defined as a threshold is less than K.

To detect outliers, we use Pycaret 3.3.1. Setup function trains the environment, multiple parameters can be specified in this step such us features, thresholds, outlier method, the number of neighbors to be used for KNN method, etc. possible models can be listed, compared, and the best model is identified, the model is tuned and saved for future use.

3) Hyperparameter tuning: Finding outliers is not an easy task; it depends on many factors. The model should be well trained, which requires a large dataset with low-correlated features. It also depends on how rare the outliers we are interested in are. Using default parameter values for a model helps to sort out data that are very different; however, to detect outliers that are not completely different, parameter tuning is mandatory [32]. The number of neighbors can be increased and the threshold reduced. The number of dimensions may be large, which increases computational resources and time. Principal Component Analysis (PCA) is used to overcome this issue and convert correlated variables to non-correlated ones.

B. Case Study: Household Collection in Tangier City

This study concerns a part of the Tangier city, one of the biggest cities in northern Morocco; it has undergone an economic and demographic surge empowered by the establishment of the Tangier-Med port, cars and aircraft manufacturers, among other development factors. Like many growing cities, the population has risen significantly; therefore, public services have to follow the pace.

In most cities, household waste collection is planned once a day in low traffic periods. Although this approach appears to be effective, it has a number of drawbacks, and there is more to be done to optimize the collection process. Indeed, time, fuel consumption, and human resources will be overused if we include the total number of bins in everyday travel in order to empty and clean them up.

IV. Results

This paper aims to sort out the rate at which bins are filling up. Datas are gathered from different LoRaWan sensors placed in all household bins in the studied region.

Data is collected over a 10-day period, sensors calculate bins filling level and send it to a central server, where, GPS location (longitude and latitude), server time, and measurement time are included. It is important to highlight data quality efforts provided in different layers in the network.

A. Data Pre-analyzis and Data Quality Check

1) Accuracy: To monitor bin filling level, gathered data from different sensors should be reliable. A sensor hardware or software failure may cause a reduced number of observations or may flood the system with observation signals. Table II shows such abnormal behavior from sensors that need to be checked. The column “Bin Number” is an identifier of bins; “Num of measurements during 10 days” represents the sum of observations during the supervision period, while the other columns show the number of observations for each day of the year (194 refers to 13th July, 195 to 14th July..).

2) Completeness: The server discards incomplete frames and missing values; incomplete ones will lead to a « Null » value that will be discarded in our data pre-treatment program.

3) Timeliness: Data age is an essential parameter; old data do not reflect reality and cannot be used to make a decision. The time difference between the server and sensors data measurement can indicate rows to exclude from data analysis; this offset will be considered in our program.

- Bins 495 and 529 sent 60982 messages during the ten days.
- Bin 529 sent 33959 during the ten days (33506 during 3 days).
- Bin 495 sent 27007 during 3 days.
- Bins 266, 271, 20, 274, and 274 sent 2 messages each during the whole ten days. Other indicators may help to find faulty devices: 391 reported only 8 measurement values during the 10 days and passed from 10% to 100% within 12 seconds between 2024-07-20 08:58:53 and 2024-07-20 08:59:05.

A normal bin’s data are also included; « bin 28 » sent 1102 observations well spread over the monitored period, with different filling level values; those observations were continually sent to the server, with a minimum of 58 and a maximum of 143 observations per day.

B. Outlier Detection

In this subsection, we will present the result of the computational program that aims to sort out outliers that need to be analyzed and take actions accordingly. Python with machine learning libraries such us Sickit Learn, Pandas, Matplotlib, and PyOD are used to develop our computation software. After data cleanup, training, testing, and evaluation, our model detects outliers. KNN is used since it has the best performing metrics: accuracy, recall, F1-score, and precision, as it is highlighted in Table III.

Found outliers contain a few bins that are 100% filled up and that are mentioned in the above subsections (bin numbers: 121, 516, 304, 503, 200, and 352). Other bins need to be highlighted and studied (bin numbers: 312, 2, 494, 452, 511, 201, 465, 376, 208, 536, 539, 545, 547, 518, 13) as per Fig. 6. More investigations need to be done to check why each of those locations represents an outlier and different stakeholders have to be involved to take a decision according to the analysis results. Other outlier methods with parameter tuning need to be used for more accurate data analysis.

TABLE II. NUMBER OF MEASUREMENTS PER SENSOR DURING THE SUPERVISION PERIOD

Bin Number	Number of measurements during 10 days	Day of the year	195	196	197	198	199	200	201	202	203	204
		194										
Bin_Number_352	8402	6	2	4	2	1099	2658	0	0	1653	0	2978
Bin_Number_495	27023	6	2	4	2	4015	9831	13161	0	0	0	2
Bin_Number_529	33959	6	2	4	2	439	0	0	0	9978	11937	11591
Bin_Number_266	2	0	0	0	0	0	0	0	0	0	0	2
Bin_Number_271	2	0	0	0	0	0	0	0	0	0	0	2
Bin_Number_20	2	0	0	0	0	0	0	0	0	0	0	2
Bin_Number_274	2	0	0	0	0	0	0	0	0	0	0	2
Bin_Number_273	2	0	0	0	0	0	0	0	0	0	0	2
Bin_Number_391	8	0	0	0	0	0	0	0	0	6	0	2
Bin_Number_307	16	6	2	4	2	0	0	0	0	0	0	2
Bin_Number_42	16	6	2	4	2	0	0	0	0	0	0	2
Bin_Number_198	16	6	2	4	2	0	0	0	0	0	0	2
Bin_Number_219	16	6	2	4	2	0	0	0	0	0	0	2
Bin_Number_435	16	6	2	4	2	0	0	0	0	0	0	2
Bin_Number_414	20	6	2	4	2	0	0	0	0	4	0	2
Bin_Number_28	1102	72	75	114	104	116	143	124	108	92	96	58

TABLE III. BEST PERFORMING OUTLIER DETECTION ALGORITHMS

Model		Accuracy(3)	AUC(4)	Recall(5)	Prec.(6)	F1(7)	Kappa	MCC	TT (Sec)
knn	K Neighbors Classifier	0.3516	0.7917	0.3516	0.3441	0.342	0.3263	0.327	8.58
nb	Naive Bayes	0.2339	0.7486	0.2339	0.1494	0.174	0.1905	0.196	4.027
dummy	Dummy Classifier	0.108	0.5	0.108	0.0117	0.021	0	0	2.499
svm	SVM - Linear Kernel	0.0826	0	0.0826	0.1335	0.077	0.0697	0.102	787.618
qda	Quadratic Discriminant Analysis	0.0087	0	0.0087	0.001	0.0210	0	0	4.073

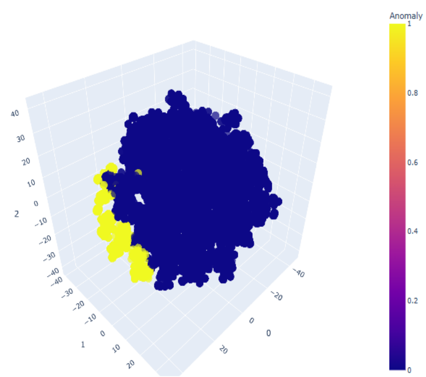


Fig. 6. Outlier bins pertaining to fill level.

The outlier detection can be used as a first step; more analysis follow to check the reasons behind those bins to be outliers. Checking outliers reduces computational resources and filters bins and sensors to monitor.

C. Slowest Filling Up Bins

Bins that fill up slowly can be deprioritized during waste collection. Unnecessary travels to those bins can be avoided.

Table IV shows multiple bins that did not exceed 30% of their capacity during 2024-07-12. These bins do not require immediate emptying and can be excluded from daily routes. During this day Bin_451 and Bin_225 recorded a 0% fill level, Bin_456, Bin_151, and Bin_44 stayed below 10% and most bins listed remained under 25%.

Table V highlights bins that remained underused over a 10-day period. Specifically, it shows the number of days during which each bin did not reach a 50% fill level.

Bin_Number_44 never reached half capacity during the entire 10-day period, Bin_Number_533, 532, 310, and 120 exceeded 50% capacity only once. Several other bins stayed below the threshold for 6 or 7 days.

Concerned bin sizes and places should be reviewed.

Table VI presents the maximum daily fill levels recorded for Bin_Number_514 over an 11-day supervision period. During 6 of these 11 days, the bin did not exceed the 50% fill threshold. Only between Days 201 and 203 did the fill level rise above 60%, with a peak of 74%. On Day 204, the fill level dropped sharply to 22%, reflecting a possible irregular usage pattern or external intervention such as manual emptying.

TABLE IV. BINS NOT REACHING 30% FILL-UP LEVEL ON 2024-07-12

bin_num	max_fill up level	bin_num	max_fill up level	bin_num	maxfill_up level	bin_num	max_fill up level
Bin_451	0	Bin_412	17	Bin_395	22	Bin_382	24
Bin_225	0	Bin_532	20	Bin_533	22	Bin_365	24
Bin_456	2	Bin_495	20	Bin_493	22	Bin_524	25
Bin_151	3	Bin_224	21	Bin_217	23	Bin_479	25
Bin_44	9	Bin_386	21	Bin_508	23	Bin_355	25
Bin_204	10	Bin_363	21	Bin_481	23	Bin_360	25
Bin_467	11	Bin_380	21	Bin_34	23	Bin_417	25
Bin_215	11	Bin_188	21	Bin_414	24	Bin_369	25
Bin_405	14	Bin_316	21	Bin_529	24	Bin_361	26
Bin_69	16	Bin_498	21	Bin_388	24	Bin_120	27
Bin_439	17	Bin_368	22	Bin_497	24	Bin_375	29

TABLE V. NUMBER OF DAYS /10 THE MAXIMUM FILL LEVEL DID NOT REACH 50%

Bin Number	N of occurrences	Bin Number	N of occurrences
44	10	343	7
533	9	515	7
532	9	350	7
310	9	412	7
120	9	40	7
488	8	417	6
506	8	182	6
535	7	456	6
467	7	514	6

Fig. 7 illustrates the fill level evolution for Bin N44, N532, and N533 over the supervision period. These bins display consistently low filling patterns. Although short spikes are observed, the majority of values remain below the 50% threshold. Bin N44, for instance, shows extended periods near zero. Bin N532 briefly exceeds 60%, then stabilizes below 30%. Bin N533 presents a single peak but quickly returns to lower levels. These trends confirm the underuse highlighted in Table V. They suggest that these bins may not require daily collection. However, the accuracy of the recorded values must be verified before operational adjustments are made.

The map below (Fig. 8) shows the geospatial distribution of bins that consistently reported low fill levels during the 10-day monitoring period in the city of Tangier. These bins, highlighted on the map, rarely exceeded 50% capacity.

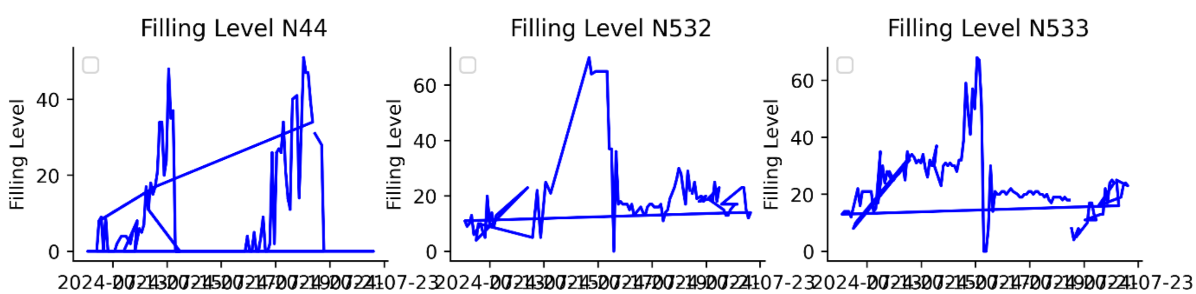


Fig. 7. Bins with a low filling level during the supervision period.

TABLE VI. MAXIMUM FILL LEVEL DURING SUPERVISION PERIOD FOR N514

day	Bin Num	level	newla	newlo	date	time
194	514	34	35.767506	-5.800178	12/07/2024	21:06:47
195	514	37	35.767506	-5.800178	13/07/2024	23:30:21
196	514	37	35.767506	-5.800178	14/07/2024	21:04:33
197	514	54	35.767506	-5.800178	15/07/2024	19:19:16
198	514	52	35.767506	-5.800178	16/07/2024	21:10:16
199	514	48	35.767506	-5.800178	17/07/2024	21:02:31
200	514	40	35.767506	-5.800178	18/07/2024	18:56:37
201	514	64	35.767506	-5.800178	19/07/2024	18:49:52
202	514	68	35.767506	-5.800178	20/07/2024	18:42:25
203	514	74	35.767506	-5.800178	21/07/2024	12:36:42
204	514	22	35.767506	-5.800178	22/07/2024	13:50:17

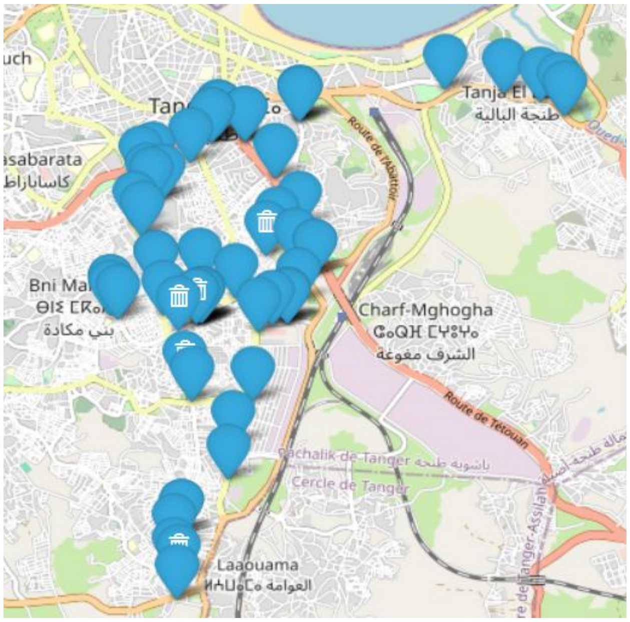
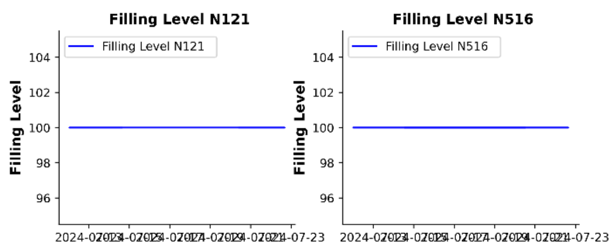


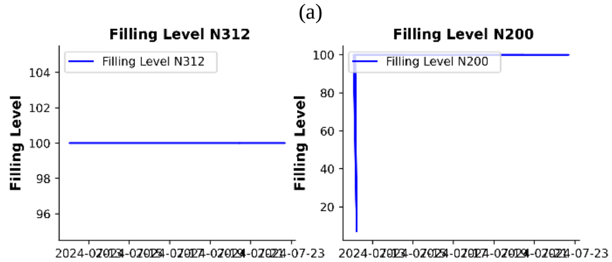
Fig. 8. Bins with a low filling level during the supervision period.

D. Fastest Filling up Bins

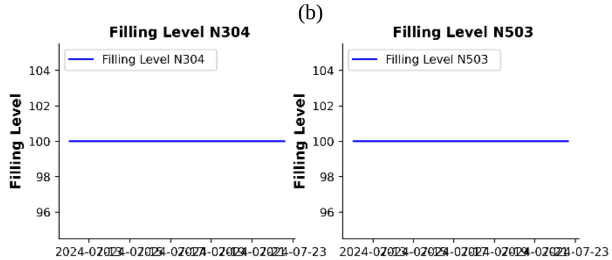
A filled-up bin can emit an unpleasant odor, which impacts the life quality of citizens. In this subsection, we will sort out the fastest-filling bins.



(a)



(b)



(c)

Fig. 9. (a) Bins with abnormal filling level for N121 and N516, (b) Bins with abnormal filling level for N121 and N516, (c) Bins with abnormal filling level for N304 and N503.

Data analysis indicates the following results: Bins 121, 516, 304, 503, 200, and 352, exhibited a constant filling level of 100% throughout the supervision period, following a small number of initial readings below that threshold. For example, Bin 352 recorded only 14 initial values below 100%, followed by 8,388 consecutive values at 100% as per Fig. 9(a), 9(b), and 9(c) which indicates a device failure that needs to be fixed.

The list below presents the quickest filled up bins; Fig. 10 shows their geolocations; those bins should be replaced with bins having a bigger capacity to answer citizens demand: [bin numbers: '121', '516', '312', '200', '304', '2', '503', '494', '452', '511', '201', '465', '376', '208', '536', '539', '545', '547', '518', '13', '541', '542'].

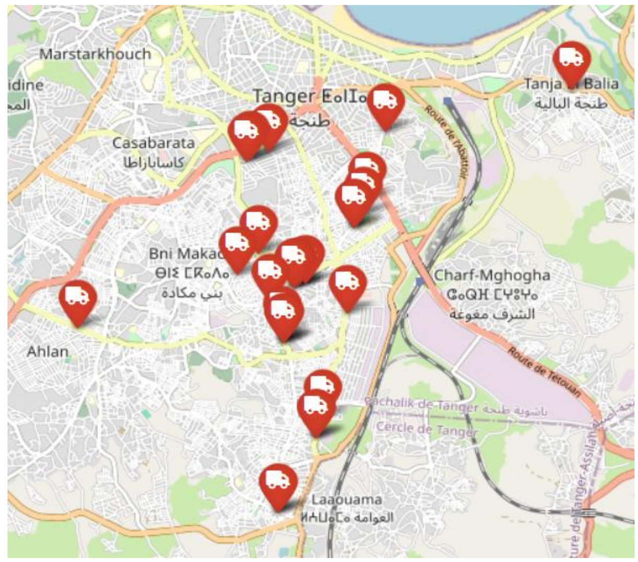


Fig. 10. Bins with a high filling level during the supervision period.

The map below combines both types of bins Fig. 11.

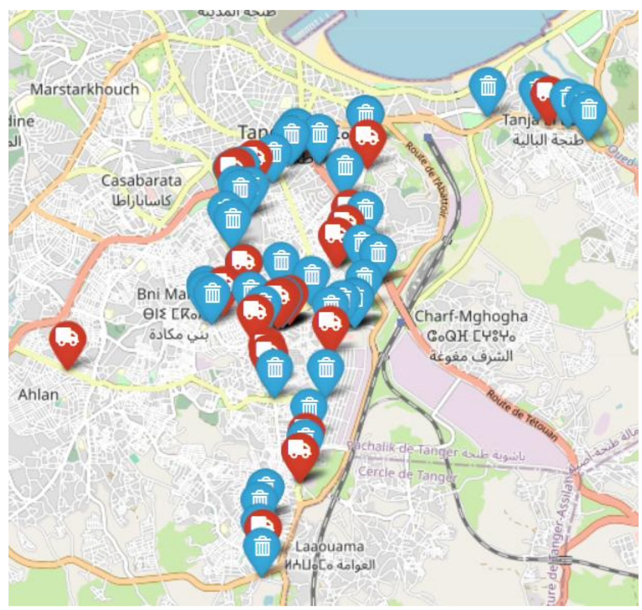


Fig. 11. Bins with a high filling level and the ones with a low filling level.

Fig. 11 presents a combined geospatial representation of waste bins with contrasting usage patterns during the supervision period. Red markers indicate bins with consistently high fill levels, while blue markers represent underused bins with persistently low fill levels.

This dual-layered view enables rapid identification of mismatches between bin capacity and local waste generation dynamics. High-fill bins highlight priority zones for:

- Capacity increase
- Additional bin deployment
- More frequent collection schedules

Low-fill bins suggest potential for:

- Relocation
- Downsizing
- Reduced collection frequency

Such spatial insights are essential for optimizing operational efficiency, minimizing collection costs, and maintaining service quality across the city. Nevertheless, sensor reliability must be verified before implementing adjustments to avoid decisions based on inaccurate data.

E. Discussion

Provided data offers valuable insights about the functional state of LoRaWan sensors, indeed:

- Several IoT-based technologies communicate filling levels via GSM texting. The expense of communication is therefore unsustainable. In just 10 days, Bin_Number_529 sent 33,959 messages. The GSM [1]method is very costly because, at 0.05 USD each SMS, for instance, that amounts to 1,697.95 USD for a single bin. Over the course of 10 days, bins 529, 495, and 352 sent 69384 messages. Scaled to a citywide network, the cost becomes huge, the proposed solution is performing better than [1] and [33] using GSM especially that LoRaWan uses free transmission band.
- Synchronization state of sensors: the dataset shows a big time offset between the server and sensors clock which may lead to incorrect observation.
- Unreliable information as demonstrate by bins 121, 516, and 352 in Fig. 9(a), Fig. 9(b) which were 100% filled up throughout the duration or a brutal filling level from 10 to 100% within 12 seconds. Fig. 9(b) indicating a faulty measure or faulty device: such behavior indicates a faulty device which facilitate the maintenance process.
- An erroneous GPS location of a sensor indicates a displacement of the bin or a faulty device measurement, which allows tracking bins in real time.
- Data control helps to maintain sensors in a healthy state and keeps transmitted data frames accurate. Maintaining IOT devices is easier by measuring data quality and avoiding traffic outage. Suspected faulty devices should be checked which allows a continuous bin fill level check.
- Truck trips can be significantly reduced by avoiding travels to bins not reaching a predefined filling threshold level; this impacts also travel time and man hours; fuel consumption and its impact on the environment can be reduced; more than that, transportation trucks can be reused and their number reduced.
- Hot seasons and special events are another critical context where waste collection efficiency directly impacts service quality and citizen satisfaction. During these periods, waste generation increases rapidly, and delayed responses can degrade urban hygiene and public perception. Monitoring the fill level of all bins in real time, while ensuring data quality, enables timely emptying of full bins. This optimizes collection routes, avoids emptying bins that are not yet full, and reduces unnecessary trips, fuel consumption, and human resource usage.

Supervising the filling level of bins for long periods indicates less used ones; a rarely used bin does not have to be emptied on a daily basis, which minimizes resource usage. The above-listed bin geolocations mentioned in the previous subsection should be reviewed; indeed, it can be displaced to more demanded locations.

On the other hand, dimensions of bins with a high level of fill-up should be resized; bigger bin sizes are needed to answer citizens' demand. Above parameters, among others, are keys of bin geolocation optimization.

There is a limitation to this study in receiving data; indeed, the analysis is impossible without valid and accurate data.

V. Conclusion

This study has enabled us to implement a program that controls IoT data quality, especially the most important dimensions such as timeliness, completeness, and accuracy.

By processing observations collected by the various devices, a full operational IoT network is maintained by early detecting faulty devices. It is possible to reduce the distance covered by bin collection trucks and reduce collection time, as well as fuel consumption and its impact on the environment.

As the location of bins is a key element in the service provided to citizens, this program allows to detect locations that are underused and that need to be displaced to a more demanded area; it also detects bins that are frequently overloaded and for which the size needs to be increased or enhances the emptying frequency, especially in hot seasons or during special events where the demand increases substantially and the service quality KPIs should be higher.

To optimize bin geolocation positions, long-term supervision needs to be put in place. Outlier method results can sort out locations that need to be highlighted. Deeper investigations regarding each outlier needed to be performed to take actions, by either increasing the bin's size, displacing the bin, or keeping the bin under surveillance.

This study focused on optimizing household waste collections based on real-time sensor data and bin usage patterns. Future work will focus on automating recycling integration at waste deposit points.

REFERENCES

[1] V. Muthukrishnan, P. Nannavare, P. Chavhan, N. Nimade, P. Kanawade, and D. Dhansade, "IOT Based Household Appliances Automation System," in 2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET), Indore, India: IEEE, Sep. 2024, pp. 1–6. doi: 10.1109/ACROSET62108.2024.10743473.

[2] H. Manoharan, Y. Teekaraman, R. Kuppusamy, and A. Radhakrishnan, "An Intellectual Energy Device for Household Appliances Using Artificial Neural Network," Mathematical Problems in Engineering, vol. 2021, pp. 1–9, Nov. 2021, doi: 10.1155/2021/7929672.

[3] M. Belhiah, M. El Aboudi, and S. Ziti, "Optimising unplanned waste collection: An IoT - enabled system for smart cities, a case study in Tangier, Morocco," IET Smart Cities, vol. 6, no. 1, pp. 27‐40, Mar. 2024, doi: 10.1049/smc2.12069.

[4] M. A. Albreem, A. M. Sheikh, M. H. Alsharif, M. Jusoh, and M. N. Mohd Yasin, "Green Internet of Things (GIoT): Applications, Practices, Awareness, and Challenges," IEEE Access, vol. 9, pp. 38833–38858, 2021, doi: 10.1109/ACCESS.2021.3061697.

[5] E. A. M. Belhiah, "An IoT-Based Sensor Mesh Network Architecture for Waste Management in Smart Cities. J," vol. 20, 2025, doi: 10.12720/jcm.20.2.153-165.

[6] A. S. Syed, D. Sierra-Sosa, A. Kumar, and A. Elmaghraby, "IoT in Smart Cities: A Survey of Technologies, Practices and Challenges," Smart Cities, vol. 4, no. 2, pp. 429–475, Mar. 2021, doi: 10.3390/smartcities4020024.

[7] M. N. Bhuiyan, M. M. Rahman, M. M. Billah, and D. Saha, "Internet of Things (IoT): A Review of Its Enabling Technologies in Healthcare Applications, Standards Protocols, Security, and Market Opportunities," IEEE Internet Things J., vol. 8, no. 13, pp. 10474–10498, Jul. 2021, doi: 10.1109/JIOT.2021.3062630.

[8] C. Wang, J. Qin, C. Qu, X. Ran, C. Liu, and B. Chen, "A smart municipal waste management system based on deep-learning and Internet of Things," Waste Management, vol. 135, pp. 20–29, Nov. 2021, doi: 10.1016/j.wasman.2021.08.028.

[9] Tahiri Alaoui, M. L., Belhiah, M., Ziti,S, "IoT-enabled Waste Management in Smart cities : A Systematic Literature Review," IJACSA.

[10] O. M. Gul, M. Kulhandjian, B. Kantarci, A. Touazi, C. Ellement, and C. D'amours, "Secure Industrial IoT Systems via RF Fingerprinting Under Impaired Channels With Interference and Noise," IEEE Access, vol. 11, pp. 26289–26307, 2023, doi: 10.1109/ACCESS.2023.3257266.

[11] A. Jahangeer, S. U. Bazai, S. Aslam, S. Marjan, M. Anas, and S. H. Hashemi, "A Review on the Security of IoT Networks: From Network Layer's Perspective," IEEE Access, vol. 11, pp. 71073–71087, 2023, doi: 10.1109/ACCESS.2023.3246180.

[12] V. Quincozes, S. Quincozes, J. Kazienko, S. Gama, O. Cheikhrouhou, and A. Koubaa, "A Survey on IoT Application Layer Protocols, Security Challenges, and the Role of Explainable AI in IoT (XAIoT)," Nov. 17, 2023, In Review. doi: 10.21203/rs.3.rs-3606636/v1.

[13] M. González-Palacio, D. Tobón-Vallejo, L. M. Sepúlveda-Cano, S. Rúa, G. Pau, and L. B. Le, "LoRaWAN Path Loss Measurements in an Urban Scenario including Environmental Effects," Data, vol. 8, no. 1, p. 4, Dec. 2022, doi: 10.3390/data8010004.

[14] M. Alenezi, K. K. Chai, Y. Chen, and S. Jimaa, "Ultra - dense LoRaWAN: Reviews and challenges," IET Communications, vol. 14, no. 9, pp. 1361‐1371, Jun. 2020, doi: 10.1049/iet-com.2018.6128.

[15] A. Proto, C. C. Miers, and T. C. M. B. Carvalho, "Classification and Characterization of LoRaWAN Energy Depletion Attacks: A Review," IEEE Sensors J., vol. 25, no. 2, pp. 2141–2156, Jan. 2025, doi: 10.1109/JSEN.2024.3504259.

[16] M. Alenezi, K. K. Chai, Y. Chen, and S. Jimaa, "Ultra - dense LoRaWAN: Reviews and challenges," IET Communications, vol. 14, no. 9, pp. 1361‐1371, Jun. 2020, doi: 10.1049/iet-com.2018.6128.

[17] K. Mikhaylov, "On the Uplink Traffic Distribution in Time for Duty-cycle Constrained LoRaWAN Networks," in 2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Brno, Czech Republic: IEEE, Oct. 2021, pp. 16–21. doi: 10.1109/ICUMT54235.2021.9631708.

[18] A. M. D. Rocha, M. A. D. Oliveira, P. José F. M., and G. G. H. Cavalheiro, "ABP vs. OTAA activation of LoRa devices: an Experimental Study in a Rural Context," in 2023 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, USA: IEEE, Feb. 2023, pp. 630–634. doi: 10.1109/ICNC57223.2023.10074553.

[19] M. Jouhari, N. Saeed, M.-S. Alouini, and E. M. Amhoud, "A Survey on Scalable LoRaWAN for Massive IoT: Recent Advances, Potentials, and Challenges," IEEE Commun. Surv. Tutorials, vol. 25, no. 3, pp. 1841–1876, 2023, doi: 10.1109/COMST.2023.3274934.

[20] M. Jouhari, N. Saeed, M.-S. Alouini, and E. M. Amhoud, "A Survey on Scalable LoRaWAN for Massive IoT: Recent Advances, Potentials, and Challenges," IEEE Commun. Surv. Tutorials, vol. 25, no. 3, pp. 1841–1876, 2023, doi: 10.1109/COMST.2023.3274934.

[21] "https://www.anrt.ma/sites/default/files/document/pnf-2021.pdf."

[22] J. R. Cotrim and C. B. Margi, "Make or Break? How LoRaWAN Duty Cycle Impacts Performance in Multihop Networks," IEEE Access, vol. 12, pp. 168925–168937, 2024, doi: 10.1109/ACCESS.2024.3494038.

[23] V. Bonilla, B. Campoverde, and S. G. Yoo, "A Systematic Literature Review of LoRaWAN: Sensors and Applications," Sensors, vol. 23, no. 20, p. 8440, Oct. 2023, doi: 10.3390/s23208440.

[24] A. Triantafyllou, P. Sarigiannidis, T. Lagkas, I. D. Moscholios, and A. Sarigiannidis, "Leveraging fairness in LoRaWAN: A novel scheduling scheme for collision avoidance," Computer Networks, vol. 186, p. 107735, Feb. 2021, doi: 10.1016/j.comnet.2020.107735.

[25] F. Loh, N. Mehling, and T. Hoßfeld, "Towards LoRaWAN without Data Loss: Studying the Performance of Different Channel Access Approaches," Sensors, vol. 22, no. 2, p. 691, Jan. 2022, doi: 10.3390/s22020691.

[26] R. Miller, H. Whelan, M. Chrubasik, D. Whittaker, P. Duncan, and J. Gregório, "A Framework for Current and New Data Quality Dimensions: An Overview," Data, vol. 9, no. 12, p. 151, Dec. 2024, doi: 10.3390/data9120151.

[27] M. L. Tahiri Alaoui, M. Belhiah, and S. Ziti, "Towards an Optimization Model for Outlier Detection in IoT-Enabled Smart Cities," in International Conference on Advanced Intelligent Systems for Sustainable Development, vol. 712, J. Kacprzyk, M. Ezziyyani, and V. E. Balas, Eds., in Lecture Notes in Networks and Systems, vol. 712. , Cham: Springer Nature Switzerland, 2023, pp. 328–338. doi: 10.1007/978-3-031-35251-5_32.

[28] C. Daraio, S. Di Leo, and M. Scannapieco, "Accounting for quality in data integration systems: a completeness-aware integration approach," Scientometrics, vol. 127, no. 3, pp. 1465–1490, Mar. 2022, doi: 10.1007/s11192-022-04266-0.

[29] W. Elouataoui, I. El Alaoui, S. El Mendili, and Y. Gahi, "An Advanced Big Data Quality Framework Based on Weighted Metrics," BDCC, vol. 6, no. 4, p. 153, Dec. 2022, doi: 10.3390/bdcc6040153.

[30] G. Naidu, T. Zuva, and E. M. Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," in Artificial Intelligence Application in Networks and Systems, vol. 724, R. Silhavy and P. Silhavy, Eds., in Lecture Notes in Networks and Systems, vol. 724. , Cham: Springer International Publishing, 2023, pp. 15–25. doi: 10.1007/978-3-031-35314-7_2.

[31] A. Tharwat, "Classification assessment methods," ACI, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.

[32] "Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure," IJIES, vol. 14, no. 6, pp. 198–207, Dec. 2021, doi: 10.22266/ijies2021.1231.19.

[33] M. U. Sohag and A. K. Podder, "Smart garbage management system for a sustainable urban life: An IoT based application," Internet of Things, vol. 11, p. 100255, Sep. 2020, doi: 10.1016/j.iot.2020.100255.