INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS

ISSN(print): 2643-9840, ISSN(online): 2643-9875

Volume 08 Issue 06 June 2025

DOI: 10.47191/ijmra/v8-i06-49, Impact Factor: 8.266

Page No. 3483-3489

Estimating Municipal Solid Waste Using Linear Regression Modelling at MRF South Jakarta

Adi Puspitarini Ayu[1], Susi A.Wilujeng[2]

[1]Magister of Environmental Management Program, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 60111

[2]Environmental Engineering Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia,60111

ABSTRACT: Accurate estimation of municipal solid waste (MSW) generation is critical for effective urban waste management planning and infrastructure optimization. However, many community-based waste management facilities, such as Siaga Municipal Recovery Facility (MRF) in South Jakarta, face challenges in direct waste quantification due to the common absence of weighbridge systems, leading to a gap in precise data for operational and strategic decision-making.

This study addresses this gap by developing and applying a simple linear regression modelling approach to estimate MSW generation. The model utilizes readily available volume data collected from various types of solid waste transport units operating at Siaga MRF. A dataset comprising 46 observational data points, including actual solid waste weight and corresponding volume measurements, was collected over several operational days. The regression model was developed and validated using a data mining approach with Python, employing the Scikit-Learn and Pandas libraries to identify the statistical relationship between conveyance volume (input) and waste weight (output).

The linear regression model yielded an R-squared (R^2) value of 0.562, indicating a moderate but practically useful level of predictive accuracy in explaining the variance in waste weight based on volume. The Mean Squared Error (MSE) was 0.0125, further supporting the model's reliability for estimation purposes in this context. These findings demonstrate that simple linear regression can serve as a feasible, low-cost, and valuable tool for estimating waste generation where direct weighing infrastructure is unavailable. The results are intended to assist policymakers, facility operators, and urban planners in enhancing capacity planning, improving daily operational forecasting, and guiding future investments in waste management technologies. Furthermore, this study reinforces the importance of applying quantitative, data-driven approaches to promote evidence-based environmental management and operational efficiency at the community scale.

KEYWORDS: Material Recovery Facility; solid waste modelling; waste estimation; waste generation

INTRODUCTION

Estimating the weight of waste entering community-based waste processing facilities, such as Siaga MRF in South Jakarta, plays a fundamental role in the design and operational planning of adequate waste treatment infrastructure. Reliable waste quantification informs not only the sizing of mechanical treatment units and storage areas, but also supports forecasting of treatment capacity, budgeting, and environmental impact assessments (World Bank, 2018). In practice, however, many small- to medium-scale waste facilities in urban Indonesia are not equipped with weighbridges or load cells that allow for direct measurement of daily incoming waste loads (Kementerian Lingkungan Hidup dan Kehutanan, 2021). This limitation results in incomplete records, which may hamper long-term planning and performance evaluations

To address this challenge, this research introduces a data-driven approach by applying linear regression modelling to predict waste weight using volume measurements of transport units. The model is developed using empirical field data collected over multiple days of observation, involving three main types of waste carriers—manual carts (A), motorized carts (B), and municipal service vehicles (C). Linear regression remains a robust and interpretable method for identifying and quantifying the relationship between variables in solid waste estimation, especially when data availability is constrained.

By transforming volume data into estimated weight through statistical modelling, this study offers a practical alternative for waste estimation in facilities without advanced weighing infrastructure. Moreover, this approach contributes to strengthening evidence-based management practices within the 3R (Reduce, Reuse, Recycle) framework, particularly at the neighbourhood and municipal scale (United Nation Environment Programme,2019). In doing so, it supports not only operational efficiency but also aligns with the broader agenda of sustainable urban waste governance.

METHOD

This study employed a structured field-based approach to estimate the weight of municipal solid waste delivered to the Siaga MRF facility by quantifying the relationship between the physical volume of waste transport units and their corresponding mass. The primary step involved measuring the dimensions (length, width, and height) of various transport vehicles—including manual carts (A), motorized carts (B), and municipal sanitation carts (C)—to compute the total volume of waste each unit carried. Concurrently, a representative subset of these transport units was manually weighed using a calibrated scale to determine the actual waste mass, enabling the derivation of an average density value for each category of carrier.

To transform these field measurements into a predictive model, a simple linear regression analysis was conducted using Python, utilizing the Scikit-Learn and Pandas libraries. In this model, volume served as the independent variable, while waste weight was defined as the dependent variable. The application of Python offered significant advantages in terms of computational flexibility, transparency of code, and reproducibility of analysis (Oliveira, Sousa, and Dias-Ferreira, 2019). The model was trained using 46 observations and subsequently evaluated using two widely accepted performance metrics in regression modeling: the Mean Squared Error (MSE) and the R-squared (R^2) coefficient. The results yielded an R^2 value of 0.562, which indicates a moderate yet meaningful ability of the model to explain variance in the target variable. An MSE of 0.0125 further supports the model's reliability, suggesting that the predicted values closely approximate the actual recorded data. This modeling approach not only addresses the lack of weighbridge infrastructure but also provides a scalable and evidence-based method for estimating waste inputs in similar urban waste collection systems (Chicco, Warrens, and Jurman, 2021).

RESULT AND DISCUSSION

This study adopted a linear regression modeling approach to address the absence of weighbridge systems in small- to medium-scale urban waste processing facilities such as Siaga MRF. By modeling the relationship between the volume of incoming waste transport units and their estimated weight, the study contributes to more accurate and data-informed waste management operations.

The research began with volume measurements of different types of waste transport units including manual carts (A), motorized carts (B), and municipal sanitation carts (C). These measurements were paired with a representative sample of manually weighed waste loads to calculate average density per vehicle type. This foundational dataset was used to develop a predictive linear regression model with volume as the independent variable and waste weight as the dependent variable.

A total of 46 observations were used, satisfying the minimum sample size suggested by Jenkins and Quintana-Ascencio (2020), who recommended a minimum of 25 samples for inference stability under high variance. The regression model was constructed and evaluated using Python, leveraging libraries such as Scikit-Learn for modeling and Pandas for data processing. The model's performance was assessed using the Mean Squared Error (MSE) and the R-squared (R^2) value. The model achieved an R^2 of 0.562, indicating a moderate explanatory power and suggesting that the model accounts for approximately 56.2% of the variance in waste weight based on transport volume.

Further, data preprocessing steps included categorizing transport units, sampling their contents over three operational days, and calculating volume via standard geometric formulas. Damanhuri and Padmi (2016) supports this technique as an effective alternative in the absence of automated measurement tools. The study also employed data integration and transformation steps to format the dataset into a structure suitable for linear modeling. These stages were critical for building a reliable model that mirrors operational realities.

The process of modeling involved defining a dictionary of volume and weight data, converting this into a DataFrame, and running a regression algorithm using Python. Oliveira, Sousa, and Dias-Ferreira (2019) underscores the suitability of Python in environmental data analysis, particularly for waste density estimation due to its strong libraries and visualization capabilities.

The approach to waste generation through density calculations begins with categorizing the type of conveyance that will be sampled by weight. The selection is based on the type of conveyance that will be analyzed for 3 consecutive days (Damanhuri and Padmi, 2016). The conveyances studied consisted of carts, LH carts, and motorized carts for sampling or manual calculation of waste weight. Waste density refers to the mass of waste per unit volume. The basic formula of density:

$$\text{specific weight}\left(\frac{\text{kg}}{\text{m}^3}\right)=\frac{\text{solid waste weight (kg)}}{\text{solid waste volume (m}^3\text{)}}$$

Tabulation of data sampling density of conveyance consisting of carts, LH carts, and motorized carts can be seen in the following table.

Table 1. Results of Waste Measurement on Transport Equipment

Cart Number	Type of Conveyance	Length (m)	Width (m)	Cart Height (m)	Garbage Height (m)	Weight (kg)
023	A	1.20	0.78	0.98	1.33	0.92
031	B	1.67	1.23	1.19	1.34	2.44
002	B	2.30	1.27	0.85	1.73	2.48
783	C	2.00	1.14	0.92	1.04	2.10
023	A	1.20	0.78	0.98	0.80	0.92
031	B	1.67	1.23	1.19	1.05	2.44
002	B	2.30	1.27	0.85	1.38	2.48
783	C	2.00	1.14	0.92	1.65	2.10
023	A	1.20	0.78	0.98	1.27	0.92
031	B	1.67	1.23	1.19	0.80	2.44
002	B	2.30	1.27	0.85	1.11	2.48
783	C	2.00	1.14	0.92	1.18	2.10

Based on the sampling data, the average density calculation for each category of conveyance can be seen in Table 2.

Table 2. Average Density of Waste for Each Transport Equipment

Type of Conveyance	Volume (m3)	Weight (kg)	Average density (kg/m3)
A	0.92	166.00	180.97
B	2.46	309.78	125.74
C	2.10	170.60	81.33

Data Selection

This stage aims to select some data that will be used in the prediction process, namely the dataset, the size of the measured.

Data Integration

The average density of the three types of conveyance is then used to calculate the weight of each conveyance that enters the MRF. Based on the volume measurement of each conveyance carried out in this research, it can be calculated the estimated waste generation from each conveyance using the average density of the conveyance category. The following is the estimated waste weight of each conveyance that enters MRF.

Table 3. Estimated Waste Generation from Each Conveyance

No	**Type of Conveyance**	**Length (m)**	**Width (m)**	**Height (m)**	**Volume(m3)**
1	A	1.20	0.78	0.98	0.92
2	A	1.10	0.60	0.80	0.53
3	A	1.25	0.79	1.05	1.04
4	A	1.56	1.20	1.50	2.81
5	A	1.43	1.10	1.20	1.89
6	B	1.67	1.23	1.19	2.44
7	B	2.00	1.25	1.40	3.50
8	C	1.00	1.10	1.00	1.10
9	A	1.54	0.80	1.00	1.23
10	B	1.70	1.30	1.20	2.65
11	B	1.57	1.13	1.20	2.13
12	B	1.60	1.30	1.20	2.50
13	A	2.23	1.54	1.33	4.57
14	C	1.90	1.00	1.10	1.10
15	B	1.60	1.20	1.20	2.30
16	A	0.00	0.00	0.00	2.50
17	B	1.60	1.20	1.10	2.11

No	Type of Conveyance	Length (m)	Width (m)	Height (m)	Volume(m3)
18	B	2.04	1.22	1.28	3.19
19	B	2.30	1.27	0.85	2.48
20	B	1.67	1.23	1.19	2.44
21	B	1.60	1.20	1.10	2.11
22	B	1.67	1.25	1.25	2.61
23	B	1.60	1.20	1.10	2.11
24	A	2.77	1.71	1.38	6.54
25	A	1.10	0.60	0.80	0.53
26	A	1.20	0.80	0.96	0.92
27	C	1.10	1.00	1.00	1.10
28	C	1.00	1.00	1.10	1.10
29	B	1.50	1.20	1.14	2.05
30	B	2.04	1.27	1.29	3.34
31	B	1.50	1.20	1.27	2.29
32	B	1.75	1.27	1.20	2.67
33	B	1.82	1.28	1.28	2.98
34	A	1.10	0.60	0.80	0.53
35	A	0.00	0.00	0.00	2.50
36	B	1.67	1.25	1.25	2.61
37	B	1.60	1.20	1.10	2.11
38	B	1.60	1.20	1.10	2.11
39	A	1.20	0.60	1.10	0.79
40	B	1.60	1.20	1.10	2.11
41	B	1.77	1.26	1.29	2.88
42	A	1.65	1.00	0.82	1.35
43	A	1.58	0.80	1.05	1.33
44	A	1.10	0.90	0.80	0.79
45	A	1.00	0.50	1.00	0.50
46	B	1.91	1.26	1.20	2.89

Data Transformation

In the data transformation stage, the collected raw data on waste conveyance volume and weight was processed and converted into a structured format suitable for data mining algorithms. This crucial step aimed to ensure the data was clean, consistent, and appropriately organized to facilitate the subsequent analysis and modelling processes. The primary goal was to prepare the dataset for the development of a linear regression model to predict waste density.

The determination of waste density values at MRF involved a statistical approach using modelling techniques. The collected survey data, which initially existed in a less structured format, was systematically organized. This involved transforming the data into a tabular structure, where each record corresponded to an observation of a waste conveyance, with distinct attributes for 'volume' and 'weight'. This structured format is essential for computational analysis.

The transformed data was then used to explore correlations between variables, specifically how the volume of conveyance relates to the weight of the waste. This involved calculating descriptive statistics and preparing the data for fitting a linear regression model, where 'volume' was designated as the independent variable and 'weight' as the dependent variable. As supported by Oliveira, Sousa, and Dias-Ferreira (2019), computational tools are effective for calculating density based on weight and volume data. The data transformation ensured that the dataset was ready for these analytical procedures, which are fundamental for calculating waste density and building a predictive model.

Data Mining

The researcher continued the data mining process with a linear regression approach. The data mining process applied in this study was to use linear regression to predict the weight of waste based on the size of the conveyance. Evaluation of prediction results was done by measuring Mean Squared Error (MSE) and the R-squared Value (R^2). At this stage, the focus was on identifying correlations between variables to determine the factors that significantly affect the weight of waste.

To facilitate this, a dataset, referred to as accumulation1, was prepared. This dataset contained two primary variables for the regression analysis: the 'volume of conveyance' (measured in cubic meters, m^3), serving as the independent variable (X), and the

'weight of waste' (measured in kilograms, kg), serving as the dependent variable (Y). A linear regression model was then developed using these variables to predict waste weight.
The results of the modeling process carried out in this study can be seen in Figure 2 below.

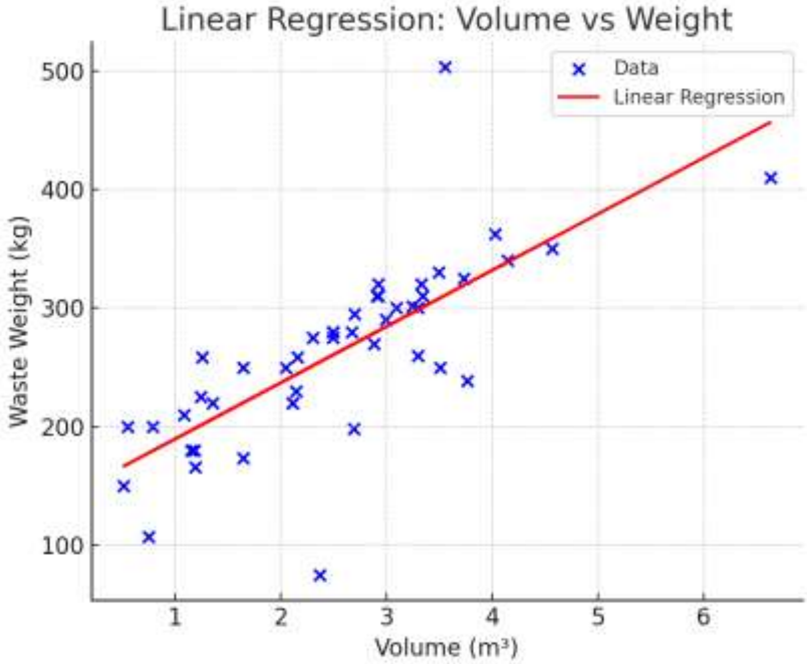

Figure 2. Density Model Interpretation

The linear regression modelling graph, as depicted in Figure 9, illustrates the relationship between the volume of the conveyance and the weight of the waste. The modelling was done using a linear regression approach, where volume acts as the independent variable and weight as the dependent variable. The purpose of this analysis is to determine how much volume contributes to the variation of waste weight, as well as to evaluate whether volume can be used as a reliable predictor in the operational context of MRF facilities. In the graph, the horizontal axis (X) represents the volume of the waste conveyance in cubic meters (m^3), which is obtained from measuring the length, width, and height of the conveyance. Meanwhile, the vertical axis (Y) displays the actual weight of the waste in kilograms (kg). The blue dots in the graph represent the actual data points, while the red line is the linear regression line generated from the model, showing the predicted trend of weight based on volume.
From the modelling results, the regression equation obtained is:
$y = 47.41x + 142.13$
In this equation, the coefficient (slope) of 47.41 indicates that for each additional cubic meter in conveyance volume, the weight of the waste is predicted to increase by approximately 47.41 kg. The intercept value of 142.13 represents the predicted weight of waste when the conveyance volume is zero, though in a practical context, this value primarily serves to position the regression line.

Evaluation and Interpretation of Model Results

To determine the level of prediction accuracy, the model was evaluated using Mean Squared Error (MSE) and the R-squared (R^2) value. The calculated metrics are:

- Mean Squared Error (MSE): 0.0125
- R-squared (R^2): 0.562

The MSE calculation, which measures the average squared difference between the actual and predicted values, yielded a value of 0.0125. This relatively small MSE suggests that the model has good prediction precision (Chicco, Warrens, and Jurman, 2021). Furthermore, an R-squared (R^2) value of 0.562 was obtained. This indicates that approximately 56.2% of the variability in the actual waste weight data can be explained by the conveyance volume through the regression model. An R^2 value above 0.5 is generally considered an indication that the model has moderate to good explanatory power in a predictive context (Yin et al., 2020).

Waste Generation Analysis

The waste generation analysis in this study was carried out by calculating the volume and estimating the weight of waste based on the type and number of conveyances operating at Siaga MRF. Data revealed a total of 46 conveyance units entering the facility, categorized into four main types: (A), (B), (C). The recapituated data indicated that (B) units were the most prevalent, with 24 units (approximately 57% of the total), and they also contributed the largest share of waste volume, transporting 60.62 m^3 (around 59.2% of the total). This highlights the dominance of (B) unit in the waste collection operations feeding into Siaga MRF. (A) type, typically used for neighborhood-scale collection or in areas with limited access, numbered 18 units and contributed 31.26 m^3 (14.2%) to the total volume. (C) units, provided as government assistance, were fewer, with 4 units transporting 4.4 m^3. Cumulatively, the total transport volume from all conveyance types recorded was 96.27m^3. This observation aligns with findings by Sari & Wibowo (2021), who noted that the dominance of motorized conveyance in waste collection systems tends to enhance route efficiency and collection frequency, particularly for area-scale facilities like Siaga MRF. To estimate the total weight of the incoming waste from this recorded volume, the previously established simple linear regression model was applied:

$y = 47.41x + 142.13$

Where:

- y = estimated weight of waste (in kilograms)
- x = volume of conveyance (in cubic meters)

Applying this model to the total recorded volume:

- Total Volume of Conveyance (x) = 96.27 m^3
- Estimated Incoming Waste Weight (y) ≈ 4.70 tons

The Siaga MRF currently receives about 4.70 tons of waste daily. This is a critical figure because the facility's operational capacity is 50 tons per day, meaning its current waste inflow is far below its maximum potential. This data clearly shows the facility isn't being used to its full potential. There's plenty of room to take in more waste without causing problems like too much unprocessed trash, environmental issues, or less efficient operations. Because the daily waste amount is low, it directly affects how resources are planned. This means we can budget fuel, schedule staff for sorting, and plan equipment maintenance efficiently, avoiding overspending. On the other hand, it also means we can easily increase resources as more waste comes in, moving towards full capacity. Accurate waste data, especially when it shows such a big gap, is crucial for local governments and waste management.

CONCLUSIONS

This study developed a linear regression model to estimate the weight of municipal solid waste entering the Siaga MRF facility in South Jakarta, using the volume of waste transport units as the primary predictor. The approach addresses the absence of weighing infrastructure at small- to medium-scale waste processing facilities by providing an alternative estimation method grounded in empirical observation.

Based on 46 field samples involving manual carts, motorized carts, and municipal sanitation vehicles, the model yielded an R^2 value of 0.562 and a Mean Squared Error (MSE) of 0.0125. These results indicate that the model explains approximately 56.2% of the variation in waste weight, with a relatively low degree of error.

The estimated total transport volume of 96.27 m^3 corresponds to approximately 4.70 tons of incoming waste per day, as calculated from the regression equation $y= 47.41\ x +142.1$. This estimation supports operational planning and capacity assessment efforts at MRF facilities by enabling waste quantification in the absence of direct measurement tools.

The approach is consistent with the principles of sustainable waste management under the 3R (Reduce, Reuse, Recycle) framework, and may be applied in similar contexts where conventional weighing equipment is unavailable. By enabling more accurate waste flow estimates, the method contributes to data-informed decision-making in community-scale waste governance.

REFERENCES

1) Chicco, Davide, Matthijs J. Warrens, and Giuseppe Jurman. 2021. “The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation.” PeerJ Computer Science 7: 1–24.
2) Damanhuri, Enri, and Tri Padmi. 2016. Pengelolaan Sampah Terpadu. Pengolaan Sampah.
3) Jenkins, David G., and Pedro F. Quintana-Ascencio. 2020. “A Solution to Minimum Sample Size for Regressions.” PLoS ONE 15 (2): 1–15.
4) Kementerian Lingkungan Hidup dan Kehutanan. 2021. “Sistem Informasi Pengelolaan Sampah Nasional (SIPSN).” 2021. https://sipsn.menlhk.go.id/sipsn/.
5) Luangpaiboon, Pongchanun, and Lakkana Ruekkasaem. 2023. “Developing A Hybrid Regression-Metaheuristic Forecasting Model for University Solid Waste Generation.” Journal of Environmental Modeling 14 (4): 245–52.

6) Oliveira, Verónica, Vitor Sousa, and Celia Dias-Ferreira. 2019. “Artificial Neural Network Modelling of The Amount of Separately-Collected Household Packaging Waste.” Journal of Cleaner Production 210: 401–9.
7) Tchobanoglous, George, Hilary Theisen, and Samuel Vigil. 1993. “Integrated Solid Waste Management. New York: McGraw-Hill.”
8) United Nation Environment Programme. 2019. “Waste Management Outlook for Asia and the Pacific.” Mumbai.
9) World Bank. 2018. What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. Washington DC: World Bank Group.