DOI: https://doi.org/10.34123/jurnalasks.v18i1.909

Available online at https://jurnal.stis.ac.id/

Jurnal Aplikasi Statistika &
Komputasi Statistik
Vol. 18 No. 1, June 2026
e- ISSN: 2615-1367, p-ISSN:2086-4132

Analysis of Factors Influencing Waste Generation in East Java

Syefa Ilmi Beandita Putri[1*], Sri Pingit Wulandari[2]

[1,2]*Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
**Corresponding Author: E-mail address:* beandita.putri@gmail.com

ARTICLE INFO

Article history:
Received 22 Dec, 2025
Revised 12 June, 2026
Accepted 25 June, 2026
Published 30 June, 2026

Keywords:
East Java; Geographically Weighted Regression; SDGs Point 12; Spatial Heterogeneity; Waste Generation

Abstract

Introduction/Main Objectives: Waste accumulation poses a serious threat to environmental sustainability and hinders the achievement of Sustainable Development Goal (SDG) 12 regarding responsible consumption and production patterns. **Background Problems:** East Java consistently ranks second highest in waste generation among Indonesian provinces; this paper investigates the demographic, economic, and environmental determinants of waste generation, specifically addressing the research question of how these factors vary across regencies. **Novelty:** This study extends previous waste generation studies by applying Geographically Weighted Regression (GWR) to the East Java context, initially considering demographic, economic, and environmental variables, and identifying spatial variations in the significant determinants of waste generation. **Research Methods:** Secondary data from 35 regencies/cities in 2023 were analyzed using GWR with a Bisquare Fixed kernel, which was selected as the optimal weighting function compared to Fixed kernels and OLS. **Finding/Results:** Surabaya City recorded the highest waste generation, while the GWR model achieved a goodness-of-fit of 92.72%, higher than the multiple linear regression model. The results confirm that the influence of waste generation determinants is not uniform across regions, indicating significant spatial heterogeneity in East Java.

1. Introduction

Waste-related issues are critical environmental problems that require serious attention from both governments and the public. Along with changing consumption patterns, waste management has become a national issue, as higher consumption levels tend to increase waste volume and diversify its composition. This issue is also closely aligned with Sustainable Development Goal (SDG) 12, which aims to ensure sustainable consumption and production patterns. In 2023, only about 59.74% of waste in Indonesia was managed correctly, equivalent to approximately 20,441,184.59 tons out of a total waste generation of 43,375,225.12 tons [1]. East Java Province, which has the second-largest population in Indonesia at approximately 41.53 million people, is also the province with the second-highest waste generation, amounting to 5,947,865.29 tons in 2023.

These conditions indicate that waste generation in East Java is not only an environmental issue but also a regional planning issue that requires evidence-based, spatially specific policy responses. Global waste management studies emphasize that reliable waste data and local-level analysis are essential for designing effective strategies for waste prevention, collection, treatment, and reduction [2], [3]. Therefore, the benefit of this study is to provide empirical information on how demographic, economic, and environmental factors influence waste generation differently across regencies and cities in East Java. The results can help local governments identify which factors should be prioritized in each area. For example, regions where population-related factors have a stronger influence may require the strengthening of household waste reduction programs, waste sorting at source, collection services, and community-based 3R facilities. Meanwhile, regions where micro and small industrial activities have a stronger influence may require targeted assistance, technical guidance, and monitoring of business waste management. Thus, the findings of this study are expected to support more targeted waste management policies rather than applying the same policy approach to all regions.

Several previous studies have examined the determinants of municipal solid waste generation using different analytical approaches, study areas, and explanatory variables. Popli et al. [4], analyzed solid waste generation in an urban region of Laos using sociodemographic and economic parameters and showed that these factors are relevant for predicting waste generation rates. Lu et al. [5] also emphasized the importance of data-driven modeling for municipal solid waste generation across different cities, noting that waste generation patterns may vary with urban characteristics and local conditions. More recently, Fontaine et al. [6] developed a framework for predicting solid waste generation by incorporating socioeconomic and demographic factors with municipal waste collection data, confirming that population-related and socioeconomic variables remain important in explaining waste generation.

In the Indonesian context, Putri [7] applied Geographically Weighted Regression (GWR) to model waste generation in Central Java and found that GWR outperformed multiple linear regression by capturing spatial variability across regions. The significant variables in Putri's study were total population, expected years of schooling, and the open unemployment rate. However, Putri's study was limited to Central Java and focused primarily on sociodemographic and socioeconomic variables. In addition, previous studies have generally focused on different location, used global regression or prediction approaches, or failed to examine how the determinants of waste generation vary locally across regencies/cities in East Java. Therefore, the present study differs from Putri's research in terms of study locus, data period, and variable coverage. This study focuses on regencies and cities in East Java in 2023 and initially integrates demographic, economic, and environmental variables, including access to improved sanitation and food management sites. Thus, the novelty of this study lies not only in the use of GWR but also in its application to the East Java context, with broader initial variable coverage, and in identifying local variations in the determinants of waste generation across regencies and cities.

Geographically Weighted Regression (GWR) is an extension of linear regression that produces location-specific parameter estimates for each observation, allowing the model to reflect spatially varying relationships between waste generation and its influencing factors at the regency and city levels [8]. In spatial analysis, relationships can be represented through area-based or point-based approaches. In this study, spatial relationships are area-based, as the data structure and interpretation are based on the regency/city administrative area as the unit of analysis. Waste generation, demographic, economic, and environmental variables are aggregated for each regency/city, so the findings are interpreted as regional characteristics rather than individual-level behavior. However, the GWR estimation process requires spatial coordinates to measure geographical proximity between observations. Therefore, each regency/city is represented by its latitude and longitude as a representative point of the administrative area. These coordinates are used to calculate Euclidean distances and construct the spatial weighting matrix in the GWR model [9].

The use of point-based coordinates is considered appropriate in this study because the main objective is to examine spatial heterogeneity, namely, whether the influence of demographic, economic, and environmental factors on waste generation differs across locations. A point-based distance-weighting approach assigns larger weights to nearby regencies/cities than to more distant regions, consistent with the local-estimation principle of GWR [8]. In contrast, an area-based contiguity approach is better suited to models that explicitly model spatial dependence between neighboring polygon areas. Since this study aims to estimate location-specific regression parameters rather than model spatial autocorrelation using an adjacency matrix, representative point coordinates serve as the basis for the GWR weighting scheme. Nevertheless, the area-based nature of the data remains central to interpretation, as all results are reported and discussed at the regency/city level.

2. Material and Methods

The target population of this study comprises all 38 regencies/cities in East Java Province, a region identified as one of the largest contributors to waste in Indonesia. The research context focuses on environmental sustainability, specifically on modeling waste-generation patterns to support Sustainable Development Goal (SDG) 12 on responsible consumption and production. The unit of analysis is the administrative area at the regency/city level. However, the final sample used in this study consists of 35 regencies/cities because waste generation data for three regions, namely Bondowoso Regency, Probolinggo Regency, and Pasuruan Regency, were not available in the SIPSN database for 2023. Therefore, these three regions were excluded from the analysis due to data unavailability, as they did not follow the number of regencies/cities used in previous research on Central Java. Thus, the use of 35 regencies/cities in this study was determined solely by the completeness of the 2023 waste generation data.

Data collection processes were conducted correctly using secondary data sources. The data regarding waste generation was retrieved from the National Waste Management Information System (SIPSN) under the Ministry of Environment and Forestry. Meanwhile, demographic and economic data were obtained from Statistics Indonesia (BPS) official publications for East Java Province, and environmental health data were sourced from the East Java Provincial Health Office. The temporal scope of the data is the year 2023, selected to ensure the most recent and comprehensive representation of the province.

2.1 Preparations

The preparations for this research included collecting secondary data from 35 regencies/cities in East Java Province in 2023. The data were obtained from the official website of Statistics Indonesia (BPS) and the National Waste Management Information System (SIPSN). In this study, p represents the number of independent variables, n indicates the number of observations (35 regencies/cities), Y_i is the value of the dependent variable (waste generation) on i-th observation, β_0 is the constant, β_j represents the regression coefficient of independent variable X_j, X_{ij} indicates the value of j-th independent variable on i-th observation. The independent variables used are Total Population (X_{i1}) measured in persons, Expected Years of Schooling (X_{i2}) measured in years, Open Unemployment Rate (X_{i3}) measured in percent, Number of Micro and Small Industrial Enterprises (X_{i4}) measured in units/enterprises, Percentage of Households with Access to Proper Sanitation (X_{i5}) measured in percent, and Percentage of Food Management Sites (X_{i6}) measured in percent.

2.2 Performing Multicollinearity Checking

Multicollinearity is a condition in which independent variables are highly correlated. The multicollinearity test identifies multicollinearity using the Variance Inflation Factor (VIF) criterion and correlation coefficient value. If the VIF value is greater than 10, then multicollinearity occurs [10]. The VIF value can be obtained using Formula (1):

$$VIF_j = \frac{1}{1 - R_j^2} \tag{1}$$

$$R_j^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Where VIF_j denotes the Variance Inflation Factor for the j-th independent variable, and R_j^2 denotes the coefficient of determination obtained from regressing the j-th independent variable on all other independent variables in the model. In the calculation of R_j^2, SSR represents the regression sum of squares, SST represents the total sum of squares, $\hat{y}$ is the predicted value for the i-th observation, y_i is the observed value for the i-th observation, $\bar{y}$ is the mean value of the observed data, and n denotes the number of observations. A higher R_j^2 value indicates that the j-th independent variable can be strongly explained by the other independent variables, which may lead to a higher VIF_j value and indicate potential multicollinearity. Furthermore, the Pearson correlation coefficient measures the linear relationship between two variables; the higher the absolute value of the coefficient, the stronger the relationship [11]. The calculation of the correlation coefficient value is explained in Formula (2).

$$r_{x_a, x_b} = \frac{n\sum_{i=1}^{n} x_{ai} x_{bi} - \left(\sum_{i=1}^{n} x_{ai}\right)\left(\sum_{i=1}^{n} x_{bi}\right)}{\sqrt{n\sum_{i=1}^{n} x_{ai}^2 - \left(\sum_{i=1}^{n} x_{ai}\right)^2}\sqrt{n\sum_{i=1}^{n} x_{bi}^2 - \left(\sum_{i=1}^{n} x_{bi}\right)^2}} \tag{2}$$

Where r_{x_a,x_b} denotes the Pearson correlation coefficient between the a-th independent variable x_a and the b-th independent variable x_b. The notation x_{ai} represents the observed value of variable x_a in the i-th observation, while x_{bi} represents the observed value of variable x_b in the i-th observation. The symbol n denotes the total number of observations. The value of the correlation coefficient ranges from

-1 to 1. A value close to 1 indicates a strong positive linear relationship, a value close to -1 indicates a strong negative linear relationship, and a value close to 0 indicates a weak linear relationship [12].

2.3 Multiple Linear Regression Modeling

The initial stage involves modeling the relationship between waste generation and the predictor variables using Multiple Linear Regression (Ordinary Least Squares - OLS) to obtain a global model [10]. The mathematical equation for the OLS model is expressed as follows:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i \tag{3}$$

Where y_i denotes the value of the dependent variable for the i-th observation, namely waste generation; β_0 denotes the intercept or constant term β_j denotes the regression coefficient of the j-th independent variable; X_{ij} denotes the value of the j-th independent variable for the i-th observation; ε_i denotes the error term for the i-th observation; (i = 1, 2, ..., n) denotes the observation index; and (j = 1, 2, ..., k) denotes the independent variable index. Furthermore, n denotes the number of observations, and k denotes the number of independent variables in the model.

The method used to estimate the parameters is Ordinary Least Squares (OLS). Parameter estimates are stated in Formula 4.

$$\hat{\beta} = (X^T X)^{-1}(X^T y) \tag{4}$$

The significance of the global parameters is evaluated using the F-test for simultaneous influence and the t-test for partial influence. The IIDN residual test is a requirement in multiple linear regression analysis. The IIDN residual test is conducted to determine whether the residuals of the regression model meet the requirements of being identical, independent, and normally distributed [13].

2.4 Performing Spatial Heterogeneity Test

The spatial heterogeneity test was conducted using the Breusch–Pagan test. This test examines whether the variance of the residuals from the global regression model is constant across observations [14]. Spatial heterogeneity means that the relationship between waste generation and its influencing factors is not uniform across regencies/cities [15]. Therefore, if spatial heterogeneity is detected, using Geographically Weighted Regression (GWR) is appropriate, as it allows regression parameters to vary locally at each observation location. The Breusch-Pagan test formula is shown in Formula (5).

$$BP = \frac{1}{2}\boldsymbol{f}^T \boldsymbol{Z}(\boldsymbol{Z}^T \boldsymbol{Z})^{-1}\boldsymbol{Z}^T \boldsymbol{f} \tag{5}$$

Where BP denotes the Breusch Pagan test statistic, $\boldsymbol{f}$ denotes an n × 1 vector whose i-th element is f_i, $\boldsymbol{Z}$ denotes the n × p matrix of standardized independent variables, $\boldsymbol{Z}^T$ denotes the transpose of matrix $\boldsymbol{Z}$ and $(\boldsymbol{Z}^T \boldsymbol{Z})^{-1}$ denotes the inverse matrix of $\boldsymbol{Z}^T \boldsymbol{Z}$. Furthermore, f_i is shown in Formula (6).

$$f_i = \left(\frac{\varepsilon_i^2}{\sigma^2} - 1\right) \tag{6}$$

Where ε_i denotes the residual value of the i-th observation, σ^2 denotes the residual variance, n denotes the number of observations, and p denotes the number of parameters in the model. The null hypothesis of the Breusch-Pagan test states that there is no spatial heterogeneity, while the alternative hypothesis states that there is spatial heterogeneity across observation locations.

2.5 Geographically Weighted Regression (GWR) Modeling

GWR enhances the global model by allowing regression parameters to vary across locations, capturing local variations [8]. The GWR model Equation is expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{k} \beta_{ij}(u_i, v_i) X_{ij} + \varepsilon_i \tag{7}$$

Where y_i denotes the value of the dependent variable for the i-th observation, namely waste generation; $\beta_0(u_i, v_i)$ denotes the intercept parameter at the location of the i-th observation; $\beta_{ij}(u_i, v_i)$ denotes the local regression coefficient of the j-th independent variable at the location of the i-th observation; X_{ij} denotes the value of the j-th independent variable for the i-th observation; ε_i denotes the error term for the i-th observation; u_i denotes the latitude coordinate of the i-th observation; v_i denotes the longitude coordinate of the i-th observation; i = 1, 2, …, n denotes the observation index; and j = 1, 2, …, k denotes the independent variable index. Furthermore, n represents the number of observations, while k represents the number of independent variables used in the model.

The spatial weighting function in GWR plays an important role because it represents the spatial proximity among observation locations. The weighting function is determined by first calculating the Euclidean distance between locations. The Euclidean distance d_{il} denotes the distance between the i-th and l-th observation locations, which is determined using latitude u_i and longitude v_i coordinates [16]. The Euclidean distance is calculated using Formula 8.

$$d_{il} = \sqrt{\left(u_i - u_l\right)^2 + (v_i - v_l)^2} \quad (8)$$

Where d_{il} denotes the Euclidean distance between the i-th observation location and the l-th observation location. The symbol u_i denotes the latitude coordinate of the i-th observation, while u_l denotes the latitude coordinate of the l-th observation. Furthermore, v_i denotes the longitude coordinate of the i-th observation, while v_l denotes the longitude coordinate of the l-th observation. The indices i and l represent different regency/city observation locations. A smaller d_{il} value indicates that two regions are geographically closer, while a larger d_{il} value indicates that two regions are farther apart [17].

The GWR spatial weighting function can be calculated using kernel functions. Kernel functions are commonly used in GWR to assign different weights to observations based on their geographical proximity to the target location. These kernel functions can be specified using either fixed or adaptive bandwidths [8], [18]. Fixed kernel functions use the same bandwidth for all observation locations, while adaptive kernel functions allow the bandwidth to vary according to the spatial distribution of the observations. The kernel functions considered in this study include Fixed Gaussian, Fixed Bisquare, Fixed Tricube, Adaptive Gaussian, Adaptive Bisquare, and Adaptive Tricube. The weighting function chosen for this analysis is the Fixed Bisquare Kernel, which was selected based on its lowest AIC and CV values compared to other kernels. The h value is a non-negative parameter known as bandwidth. Bandwidth determines the extent of spatial influence around each target location and is used to assign weights to other observation locations in the GWR model [8], [17]. The method for determining the bandwidth value is using Cross Validation, as described in Formula 9 [9].

$$CV(h) = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(h)]^2 \quad (9)$$

To estimate the parameters, a spatial weighting matrix is required. This study utilizes the Euclidean distance metric to measure the proximity between locations i and j.

The GWR parameter model estimation used is the Weighted Least Square (WLS) method by providing different weightings for each observation location [8]. The GWR parameter estimation is explained by the Formula in 10.

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X}^T \boldsymbol{W}(u_i, v_i) \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}(u_i, v_i) \boldsymbol{y} \quad (10)$$

Finally, the model's performance is evaluated and compared against the global OLS model using the Coefficient of Determination (R^2) and Akaike Information Criterion (AIC). A higher R^2 and a lower AIC indicate a better-fitting model [19]. To provide a clearer overview of the analytical procedure used in this study, the steps of the Geographically Weighted Regression analysis are summarized in the flow chart shown in Figure 1.

3. Results and Discussion

The characteristics of waste generation data and its influencing factors based on descriptive statistics are presented in Table 1.

Table 1. Descriptive statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Waste Generation (Y_i)	169,939.01	123,696.13	27,988.20	657,016.64
Total Population (X_{i1})	1,081,209	699,795	135,414	2,893,698
Expected Years of Schooling (X_{i2})	13.57	0.90	11.97	15.77
Open Unemployment Rate (X_{i3})	4.70	1.46	1.71	8.05
Number of Micro and Small Industrial Enterprises (X_{i4})	22,251	15,155	2,009	67,609
Percentage of Households with Access to Improved Sanitation (X_{i5})	86.02	11.23	50.30	98.18

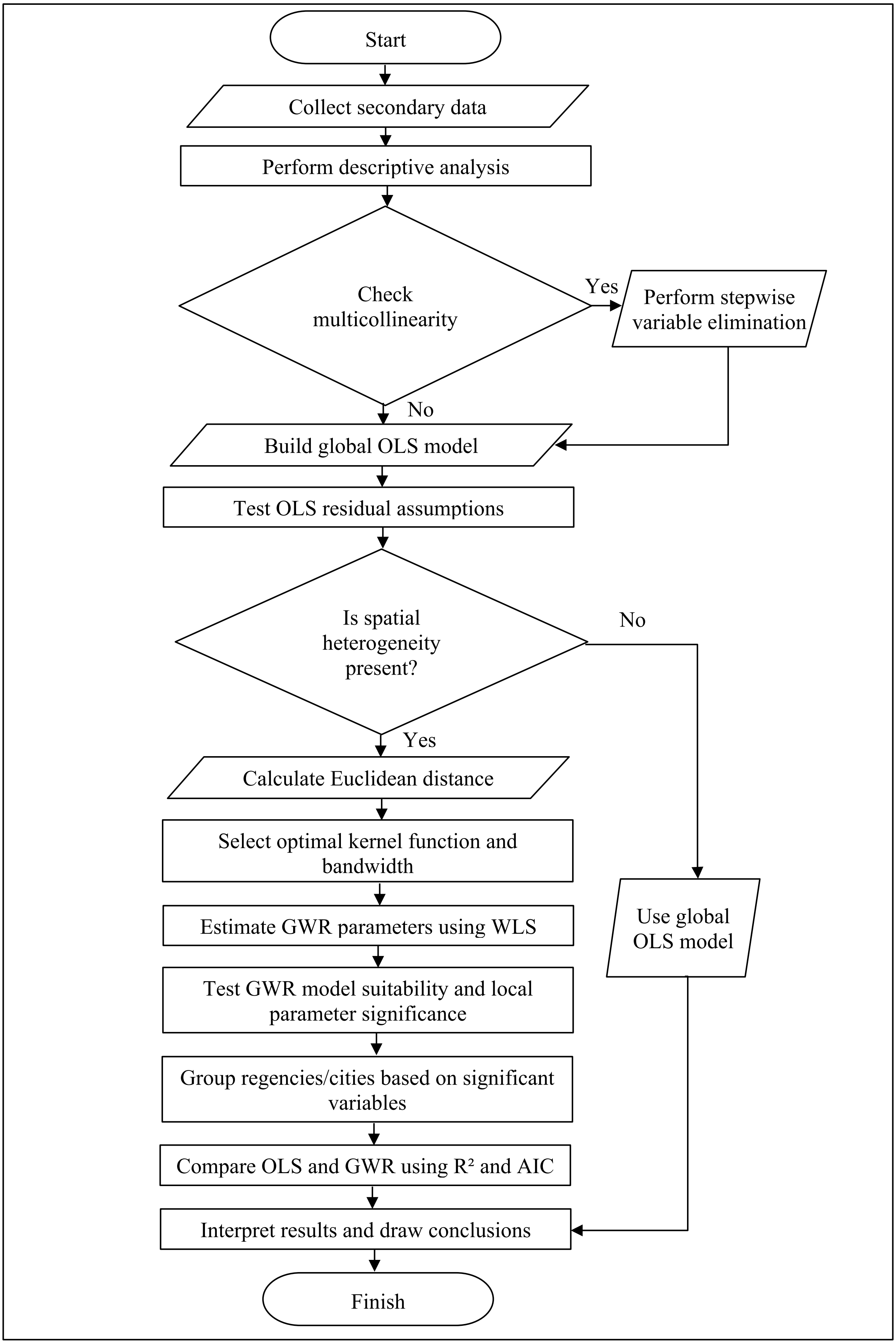

Figure 1. Flow chart

Table 1 shows that annual waste generation in East Java in 2023 averaged 169,939.01 tons per year, with a standard deviation of 123,696.13 tons per year, indicating substantial variation across regencies and cities. Blitar Regency records the lowest waste generation, while Surabaya City has the highest. The population variable also exhibits considerable regional disparity, with an average of 1,081,209 people and a standard deviation of 699,795, primarily driven by Surabaya City, the most populous area.

Other socioeconomic factors, including expected years of schooling, the open unemployment rate, the number of micro and small industrial enterprises, access to improved sanitation, and the percentage of food management facilities, display notable variation among regencies and cities. This variation reflects differences in educational attainment, labor market conditions, economic activity, and basic infrastructure across East Java. The spatial characteristics of waste generation are further illustrated through map visualizations presented in Figure 1.

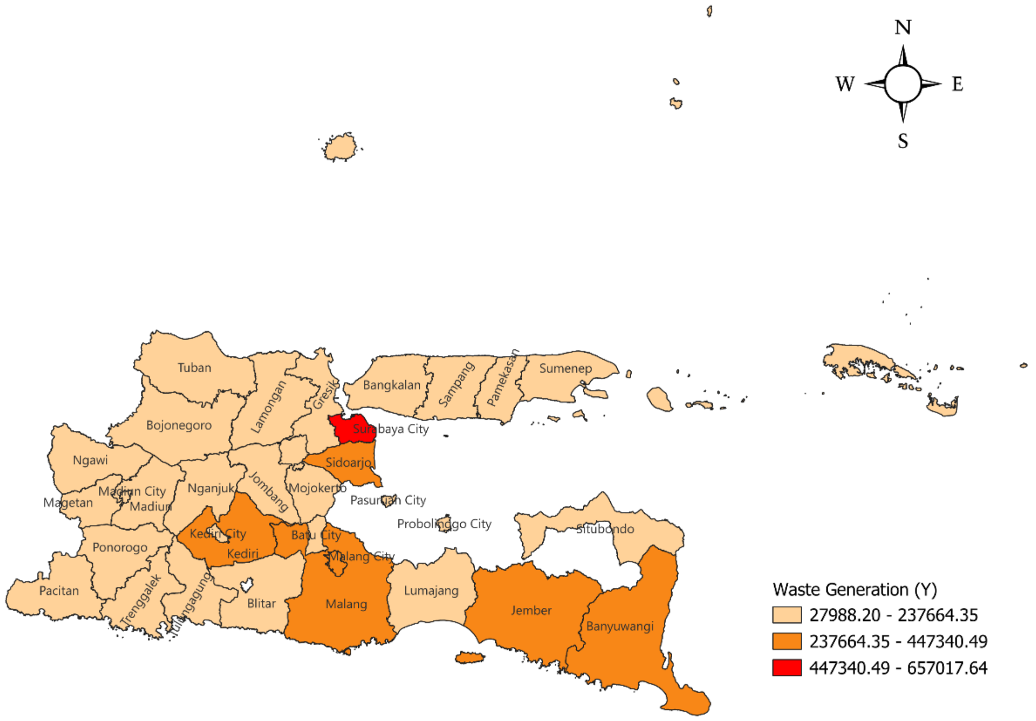

Figure 2. Mapping of waste generation in East Java Province

3.1 Multicollinearity Checking

Multicollinearity was examined using correlation coefficients and variance inflation factors (VIFs). The assessment of multicollinearity based on correlation coefficients is presented in Table 2.

Table 2. Correlation matrix

	Y	**X_{i1}**	**X_{i2}**	**X_{i3}**	**X_{i4}**	**X_{i5}**
X_{i1}	0,908					
	0,000					
X_{i2}	0,128	-0,066				
	0,462	0,708				
X_{i3}	0,375	0,297	**0,418**			
	0,026	0,083	**0,012**			
X_{i4}	0,237	**0,440**	**-0,378**	**-0,433**		
	0,170	**0,008**	**0,025**	**0,009**		
X_{i5}	-0,021	-0,133	**0,518**	**0,406**	**-0,521**	
	0,905	0,445	**0,001**	**0,015**	**0,001**	
X_{i6}	0,029	-0,041	**0,311**	**0,325**	-0,219	**0,412**
	0,867	0,813	**0,069**	**0,057**	0,207	**0,014**

Table 2 indicates that several pairs of predictor variables exhibit statistically significant correlations (p-values < 0.10). Furthermore, the results of the Variance Inflation Factor (VIF) analysis show that all predictor variables have VIF values below 10. Although the VIF values do not exceed the commonly used threshold, the presence of significant correlations among predictor variables suggests potential multicollinearity. Therefore, variable elimination was performed using the stepwise method, resulting in the final set of predictor variables consisting of X_{i1}, X_{i2}, and X_{i4}.

3.2 Multiple Linear Regression Modeling

The parameter estimates of the multiple linear regression model for factors influencing waste generation are expressed as follows:

$$\hat{Y} = -251554 + 0.17X_1 + 19157X_2 - 1.16X_4$$

The estimated multiple linear regression model indicates that, holding other variables constant, a one-person increase in total population (X_1) in East Java Province is associated with an increase in waste generation of 0.17 tons/year. Furthermore, a one-year increase in expected years of schooling (X_2) is associated with an increase in waste generation of 19,157 tons/year, assuming other variables remain constant. Meanwhile, a one-unit increase in the number of micro and small industrial enterprises (X_4) is associated with a 1.16-ton/year decrease in waste generation, ceteris paribus.

The negative coefficient of the number of micro and small industrial enterprises (X_4) is not fully consistent with the theoretical expectation that industrial activity tends to increase waste generation. This result may be related to differences in the characteristics of micro and small enterprises across regions, including the implementation of waste minimization, reuse of production materials, recycling, or practices oriented towards a circular economy. Circular economy approaches emphasize reducing waste, keeping materials in use, and reusing or recycling resources, so an increase in the number of business units does not always directly lead to higher residual waste generation [20]. Therefore, the negative sign of (X_4) should be interpreted cautiously as an indication that the relationship between micro and small industrial activity and waste generation may vary across regencies/cities. This condition supports the need for spatially local analysis using GWR.

Subsequently, a simultaneous significance test was conducted. Based on the analysis, the F-test statistic value of 71.73 exceeds the critical value $F_{(0.1;3;31)}$ is 2.27 and is supported by a p-value of 0.00, which is smaller than the significance level α is 0.1. Therefore, the null hypothesis is rejected, indicating that at least one predictor variable has a statistically significant effect on waste generation. Furthermore, the results of the partial significance tests are presented in Table 3. The t-test results indicate that variables X_{i1}, X_{i2}, and X_{i4} have statistically significant effects on waste generation.

Table 3. Partial significance test statistics

Variable	t-statistic	$t_{(0,1/2;28)}$	P-Value
X_{i1}	13.70	1.70	0.00
X_{i2}	2.01		0.05
X_{i4}	1.84		0.08

Subsequently, the results of the identical residuals test show that the test statistic value of 22.74 exceeds the critical value $F_{(0.1;1;33)}$ is 2.86 and is supported by a p-value of 0.00, which is smaller than the significance level α is 0.1. Therefore, the null hypothesis is rejected, indicating that the residuals of the regression model are not identical. This result suggests the presence of spatial heterogeneity due to differences in characteristics across regencies and cities.

The independence of residuals test yields a Durbin–Watson statistic of 2.17, which is smaller than 4 − dU is 2.35 and greater than dU is 1.65. Thus, the null hypothesis is accepted, indicating that the residuals of the regression model are independent.

The normality test of residuals produces a test statistic value of 0.13, which is smaller than the critical value $KS_{(0.1;35)}$ is 0.20 and is supported by a p-value of 0.56, which is greater than the significance level α is 0.1. Therefore, the null hypothesis is accepted, indicating that the residuals of the regression model are normally distributed.

3.3 Performing Spatial Heterogeneity Test

The spatial heterogeneity test using the Breusch–Pagan test yields a test statistic value of 16.79, which exceeds the critical value of 6.25 and is supported by a p-value of 0.00, which is smaller than the significance level α is 0.1. Therefore, the null hypothesis is rejected, indicating the presence of spatial heterogeneity. Consequently, the analysis can be extended using the Geographically Weighted Regression (GWR) approach.

3.4 Geographically Weighted Regression (GWR) Modeling

In the GWR analysis, the weighting values were determined using kernel functions, as presented in Table 4. Table 4 indicates that the Fixed Bisquare kernel function was selected for the GWR analysis, as it provides the best overall model performance based on the evaluation criteria.

Table 4. Comparison of optimal Kernel Functions

Kernel Function	CV	R^2	AIC
Fixed Gaussian	110489182025	0,88	850,28
Adaptive Gaussian	110964903827	0,88	849,67
Fixed Bisquare	120870672140	0,93	836,76
Adaptive Bisquare	121002609225	0,90	845,95
Fixed Tricube	119746559578	0,92	839,05
Adaptive Tricube	133953245074	0,90	844,33

After selecting the Fixed Bisquare kernel function, the factors influencing waste generation were modeled for the first observation location (u_1, v_1), namely Pacitan Regency, which is used as an example for interpreting the modeling results. The waste generation model for Pacitan Regency is expressed as follows:

$$\hat{Y} = -548164.14 + 0.21X_1 + 36448.47X_2 + 0.74X_4$$

This model indicates that, holding other variables constant, a one-person increase in population results in an increase of waste generation by 0.21 tons per year. Furthermore, a one-year increase in expected years of schooling, ceteris paribus, leads to an increase in waste generation by 36,448.47 tons per year. In addition, an increase of one percent in the number of micro and small industrial enterprises, while other variables remain constant, increases waste generation by 0.74 tons per year. An increase in expected years of schooling, accompanied by higher waste generation, indicates that improvements in educational attainment in Pacitan Regency have not yet been fully aligned with increased awareness and behavioral changes in waste management [21], [22]. In line with this, population growth reflects rising household consumption, which directly increases waste generation [4], [23]. In addition to demographic factors, the growth in the number of micro and small industrial enterprises suggests that business activities also contribute to increased waste generation when not supported by responsible waste management practices [24]. These conditions indicate that both demographic and economic factors are important determinants of waste generation in Pacitan Regency. Therefore, efforts in education and outreach are needed for communities and micro- and small-business actors on waste reduction, waste segregation, and sustainable waste management, particularly among groups with increasing consumption and production activities, alongside local educational and economic development.

The results of the GWR model goodness-of-fit test show that the test statistic value of 1.73 exceeds the critical value $F_{(0.1;31;23,34)}$ is 1.68 and is supported by a p-value of 0.08, which is smaller than the significance level α is 0.1. Therefore, the null hypothesis is rejected, indicating a statistically significant difference between the Ordinary Least Squares (OLS) and the GWR models. This result confirms that the GWR approach provides a better representation of the data's spatial variation.

Following the confirmation of model suitability, partial significance tests of the GWR model parameters were conducted using a significance level of α is 0.1 and degrees of freedom of 23.34, with a critical value of $t_{(0.1;23.34)}$ is 1.71 The results indicate that the total population variable (X_{i1}), is statistically significant in 35 regencies and cities, expected years of schooling (X_{i2}), and is significant in 26 regencies and cities, and the number of micro and small industrial enterprises (X_{i4}) is significant in 11 regencies and cities. These findings highlight the spatially varying influence of explanatory variables on waste generation across East Java, as presented in Table 5.

Table 5. Partial significance test of GWR model parameters

No.	Regency/City	$\lvert t_{X_{i1}} \rvert$	$\lvert t_{X_{i2}} \rvert$	$\lvert t_{X_{i4}} \rvert$	$t_{(\frac{0,1}{2};23.34)}$
1.	Pacitan Regency	**12.87**	**2.09**	0.60	1.71
2.	Ponorogo Regency	**13.91**	**2.39**	0.58	1.71
3.	Trenggalek Regency	**13.74**	**2.65**	0.59	1.71
4.	Tulunganggung Regency	**13.78**	**2.68**	0.98	1.71
5.	Blitar Regency	**13.03**	**2.56**	1.58	1.71
6.	Kediri Regency	**14.04**	**2.44**	**1.72**	1.71
7.	Malang Regency	**12.24**	**2.47**	1.71	1.71
8.	Lumajang Regency	**10.15**	**1.93**	1.10	1.71
9.	Jember Regency	**9.24**	1.29	1.53	1.71
10.	Banyuwangi Regency	**6.19**	0.78	0.99	1.71
11.	Situbondo Regency	**7.26**	0.73	1.02	1.71
12.	Sidoarjo Regency	**12.27**	**2.16**	**1.95**	1.71
13.	Mojokerto Regency	**13.72**	**2.28**	**2.36**	1.71
14.	Jombang Regency	**14.28**	**2.30**	**2.31**	1.71
15.	Nganjuk Regency	**14.30**	**2.24**	**1.82**	1.71
16.	Madiun Regency	**14.19**	**2.28**	1.06	1.71
17.	Magetan Regency	**13.88**	**2.15**	0.73	1.71
18.	Ngawi Regency	**14.21**	**1.96**	1.43	1.71
19.	Bojonegoro Regency	**14.16**	1.61	**2.31**	1.71
20.	Tuban Regency	**13.66**	1.22	**2.51**	1.71
21.	Lamongan Regency	**13.35**	**1.84**	**2.38**	1.71
22.	Gresik Regency	**12.53**	**1.94**	**2.13**	1.71
23.	Bangkalan Regency	**10.43**	1.46	1.69	1.71
24.	Sampang Regency	**9.33**	1.19	1.19	1.71
25.	Pamekasan Regency	**9.31**	0.75	1.15	1.71
26.	Sumenep Regency	**8.20**	0.59	0.83	1.71
27.	Kediri City	**14.05**	**2.46**	1.63	1.71
28.	Blitar City	**13.31**	**2.62**	1.33	1.71
29.	Malang City	**12.24**	**2.47**	1.71	1.71
30.	Probolinggo City	**10.06**	**1.79**	1.17	1.71
31.	Pasuruan City	**10.79**	**2.04**	1.42	1.71
32.	Mojokerto City	**13.71**	**2.28**	**2.36**	1.71
33.	Madiun City	**14.24**	**2.23**	1.16	1.71
34.	Surabaya City	**14.23**	**2.23**	1.15	1.71
35.	Batu City	**13.11**	**2.50**	**2.00**	1.71

Note: Values shown in bold indicate statistically significant parameters.

Based on the results of the analysis, a summary of the significant variables for each regency and city in East Java is presented in Table 6.

Table 6. Significant variables for each regency/city

No.	Regency/ City	Variables	Significant Factors	No.	Regency/ City	Variables	Significant Factors
1	Pacitan Regency	X_1, X_2	Demographic	19	Bojonegoro Regency	X_1, X_4	Demographic, Economic
2	Ponorogo Regency	X_1, X_2	Demographic	20	Tuban Regency	X_1, X_4	Demographic, Economic
3	Trenggalek Regency	X_1, X_2	Demographic	21	Lamongan Regency	X_1, X_2, X_4	Demographic, Economic
4	Tulungagung Regency	X_1, X_2	Demographic	22	Gresik Regency	X_1, X_2, X_4	Demographic, Economic
5	Blitar Regency	X_1, X_2	Demographic	23	Bangkalan Regency	X_1	Demographic
6	Kediri Regency	X_1, X_2, X_4	Demographic, Economic	24	Sampang Regency	X_1	Demographic
7	Malang Regency	X_1, X_2	Demographic	25	Pamekasan Regency	X_1	Demographic
8	Lumajang Regency	X_1, X_2	Demographic	26	Sumenep Regency	X_1	Demographic
9	Jember Regency	X_1	Demographic	27	Kediri City	X_1, X_2	Demographic
10	Banyuwangi Regency	X_1	Demographic	28	Blitar City	X_1, X_2	Demographic
11	Situbondo Regency	X_1	Demographic	29	Malang City	X_1, X_2	Demographic
12	Sidoarjo Regency	X_1, X_2, X_4	Demographic, Economic	30	Probolinggo City	X_1, X_2	Demographic
13	Mojokerto Regency	X_1, X_2, X_4	Demographic, Economic	31	Pasuruan City	X_1, X_2	Demographic
14	Jombang Regency	X_1, X_2, X_4	Demographic, Economic	32	Mojokerto City	X_1, X_2, X_4	Demographic, Economic
15	Nganjuk Regency	X_1, X_2, X_4	Demographic, Economic	33	Madiun City	X_1, X_2	Demographic
16	Madiun Regency	X_1, X_2	Demographic	34	Surabaya City	X_1, X_2	Demographic
17	Magetan Regency	X_1, X_2	Demographic	35	Batu City	X_1, X_2, X_4	Demographic, Economic
18	Ngawi Regency	X_1, X_2	Demographic				

Table 6 shows that the variables significantly influencing waste generation vary across regencies and cities. Regencies and cities were further grouped based on similarities in the significant variables affecting waste generation, as presented in Table 7.

Table 7 summarizes the grouping of regencies and cities based on the variables that significantly influence waste generation. Grouping is important because it shows that the determinants of waste generation are not uniform across East Java. Each group reflects a different local mechanism in which demographic and economic factors influence waste generation. Therefore, the same waste management policy may not be equally effective for all regencies/cities.

Group 1 consists of Jember Regency, Banyuwangi Regency, Situbondo Regency, Bangkalan Regency, Sampang Regency, Pamekasan Regency, and Sumenep Regency, in which only the total population (X_1) is significant. This indicates that waste generation in these regions is primarily driven by population size. In this group, the effects of expected years of schooling (X_2) and the number of micro and small industrial enterprises (X_4) are not statistically significant, suggesting that household and population-related waste generation is more dominant than education-related consumption patterns or small industrial activities. Therefore, waste management strategies in this group should prioritize household waste reduction, source-based waste sorting, improvements in waste collection services, and community-based waste management programs.

Table 7. Grouping of regencies/cities based on significant variables

Group	Regency/City	Variables	Number of Regencies/Cities	Significant Factors
1	Jember Regency, Banyuwangi Regency, Situbondo Regency, Bangkalan Regency, Sampang Regency, Pamekasan Regency, Sumenep Regency	X_1	7	Demographic
2	Pacitan Regency, Ponorogo Regency, Trenggalek Regency, Tulunganggung Regency, Blitar Regency, Malang Regency, Lumajang Regency, Madiun Regency, Magetan Regency, Ngawi Regency, Kediri City, Blitar City, Malang City, Probolinggo City, Pasuruan City, Madiun City, Surabaya City	X_1, X_2	17	Demographic
3	Bojonegoro Regency, Tuban Regency	X_1, X_4	2	Demographic, Economic
4	Kediri Regency, Sidoarjo Regency, Mojokerto Regency, Jombang Regency, Nganjuk Regency, Lamongan Regency, Gresik Regency, Mojokerto City, Batu City	X_1, X_2, X_4	9	Demographic, Economic

Group 2 consists of regions where total population (X_1) and expected years of schooling (X_2) are significant. This group includes several regencies and cities such as Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Malang, Lumajang, Madiun, Magetan, Ngawi, Kediri City, Blitar City, Malang City, Probolinggo City, Pasuruan City, Madiun City, and Surabaya City. The significance of X_2 indicates that waste generation in these regions is not only related to the number of residents but also to social characteristics associated with education. Higher expected years of schooling may reflect changes in lifestyle, consumption behavior, access to goods and services, and public activities, which can increase the volume and diversity of waste generated. Therefore, this group requires not only population-based waste management but also environmental education programs, school-based waste literacy, and campaigns to strengthen responsible consumption behavior.

Group 3 consists of Bojonegoro Regency and Tuban Regency, where the total population (X_1) and the number of micro and small industrial enterprises (X_4) are significant. This pattern indicates that both household activities and local economic activities influence waste generation in these regions. The significance of X_4 suggests that micro and small enterprises contribute to waste generation through production, packaging, distribution, and trade activities. Therefore, waste management policies in this group should combine household waste management with specific assistance for micro and small enterprises, such as training on waste reduction, recyclable packaging, separation of production waste, and monitoring of business-related waste disposal.

Group 4 consists of Kediri Regency, Sidoarjo Regency, Mojokerto Regency, Jombang Regency, Nganjuk Regency, Lamongan Regency, Gresik Regency, Mojokerto City, and Batu City, where the total population (X_1), expected years of schooling (X_2), and the number of micro and small industrial enterprises (X_4) are all significant. This group indicates a more complex pattern of waste generation because demographic, social, and economic factors simultaneously influence it. The regions in this group require more integrated waste management policies, including household waste services, education-based behavioral change programs, and targeted waste management assistance for micro and small enterprises. The significance of the three variables in this group indicates that waste generation is not only a population issue but is also closely related to human development and local economic activity.

The spatial distribution of regencies and cities in East Java according to these groupings is illustrated in Figure 3.

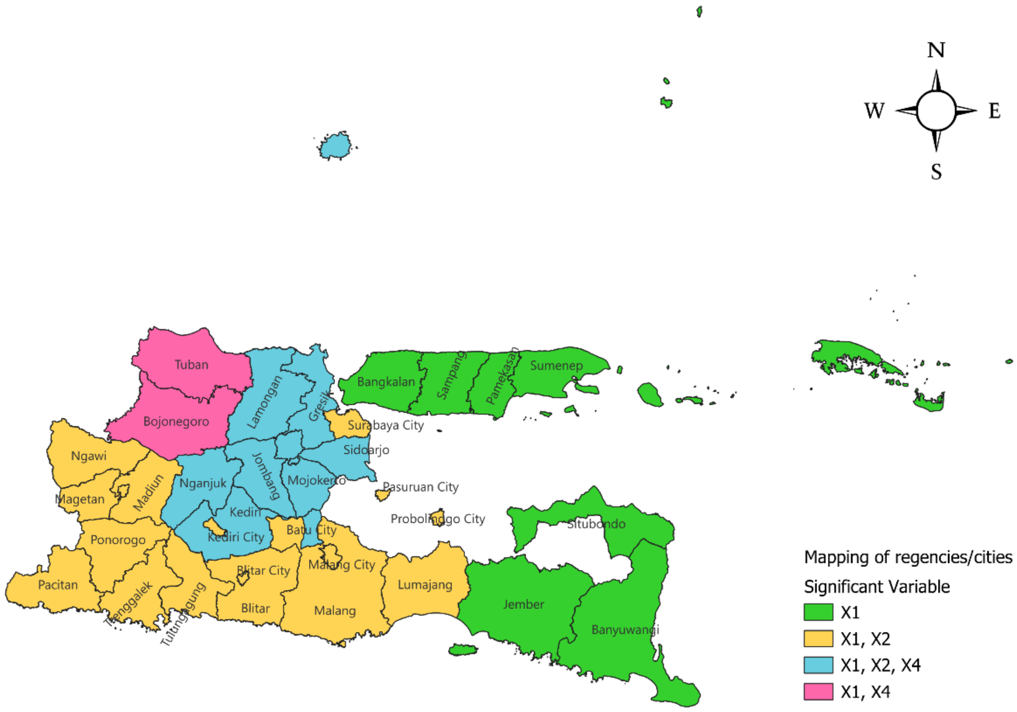

Figure 3. Mapping of regencies/cities based on significant variables

The selection of the best model was conducted by comparing multiple linear regression and Geographically Weighted Regression (GWR) using the coefficient of determination (R^2) and the Akaike Information Criterion (AIC), as presented in Table 8.

Table 8. Best model selection

Method	R2	AIC
Multiple linear regression	86.19%	856.57
GWR	92.72%	836.76

Table 8 shows that the GWR model has a higher R^2 and a lower AIC than the multiple linear regression model, indicating superior model performance. This result suggests that GWR better captures spatial heterogeneity across regions, whereas multiple linear regression applies a single global Equation to all regions.

4. Conclusion

Based on the results of the Geographically Weighted Regression model with a fixed bisquare kernel, the influence of demographic and economic factors on waste generation varies across East Java's regencies and cities, indicating spatial heterogeneity. The GWR model captures this spatial variation and achieves a goodness-of-fit of 92.72%, making it more representative than the multiple linear regression model. The results show that total population (X_1), expected years of schooling (X_2), and the number of micro and small industrial enterprises (X_4) have different levels of significance across regions. Meanwhile, environmental variables initially considered in the early stage of the analysis were not included in the final model because they were eliminated during variable selection due to multicollinearity. For local governments, these findings indicate that waste management policies should be tailored to regional characteristics. Regions where population is the dominant factor should prioritize household waste reduction, waste sorting at the source, and improvement of waste collection services, while regions where micro and small industrial enterprises also have a significant effect should strengthen outreach, technical assistance, and monitoring of business-related waste management.

This study has several limitations. First, the analysis covers only 35 regencies/cities in East Java because waste generation data for Bondowoso Regency, Probolinggo Regency, and Pasuruan Regency were unavailable in the SIPSN database for 2023. Therefore, the results do not fully represent all administrative regions in East Java. Second, this study uses cross-sectional data from a single year, so it cannot capture changes in waste generation patterns over time. Third, the analysis is limited to variables available from secondary data sources, and several potentially relevant factors, such as income level, population density, tourism activity, consumption patterns, land use, waste management facilities, and local waste management policies, were not included in the model. In addition, the use of representative coordinates at the regency/city level may not fully capture spatial variation within each administrative area. Therefore, future research is recommended to use more complete data covering all regencies and cities in East Java, apply multi-year or panel data, include additional socioeconomic, infrastructure, and waste management variables, and compare GWR with other spatial approaches such as spatial lag models, spatial error models, or spatial panel models.

Ethics approval

Not required.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

Data obtained via the Statistics Indonesia of East Java Province website https://jatim.bps.go.id/, the National Waste Management Information System (SIPSN) website https://sipsn.menlhk.go.id/, and the East Java Provincial Health Office publication (East Java Province Health Profile).

Credit Authorship

Syefa Ilmi Beandita Putri: Conceptualization, Methodology, Data Collection, Data Analysis, Software Development, Visualization, Writing–Original Draft. **Sri Pingit Wulandari:** Conceptualization, Methodology, Manuscript Review, Research Advisor, Writing–Review.

References

[1] SIPSN, “Timbulan Sampah.” Accessed: May 18, 2025. [Online]. Available: https://sipsn.kemenlh.go.id/sipsn/public/data/timbulan

[2] F. Kaza, S.; Yao, L.; Bhada-Tata, P.; Van Woerden, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. World Bank.* Washington DC: The World Bank, 2018.

[3] U. N. E. Programme, “Global Waste Management Outlook 2024: Beyond an Age of Waste,” Nairobi, 2024. [Online]. Available: https://www.unep.org/resources/global-waste-management-outlook-2024

[4] K. Popli, C. Park, S. Han, and S. Kim, “Prediction of Solid Waste Generation Rates in Urban Region of Laos Using Socio-Demographic and Economic Parameters with a Multi Linear Regression Approach,” *MDPI*, vol. 13, no. 6, 2021, doi: https://doi.org/10.3390/su13063038.

[5] W. Lu, W. Huo, H. Gulina, and C. Pan, “Development of machine learning multi-city model for municipal solid waste generation prediction,” vol. 16, no. 9, 2022.

[6] L. Fontaine, R. Legros, and J. Frayret, “Solid waste generation prediction model framework using socioeconomic and demographic factors with real-time MSW collection data,” 2025, doi: 10.1177/0734242X241231414.

[7] F. E. Putri, “Pemodelan Timbulan Sampah Berdasarkan Faktor Sosioekonomi dan Sosiodemografi di Jawa Tengah Menggunakan Geographically Weighted Regression,” Institut

Teknologi Sepuluh Nopember Surabaya, 2024.
[8] A. S. F. C. B. M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: John Wiley & Sons, 2002.
[9] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, “Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity,” vol. 28, no. 4, 1996.
[10] D. C. . E. A. P. G. G. V. Montgomery, *Introduction to Linear Regression Analysis (5th ed.)*. Wiley, 2013.
[11] N. N. Kachouie, “Association Factor for Identifying Linear and Nonlinear Correlations in Noisy Conditions,” pp. 1–19, 2020, doi: 10.3390/e22040440.
[12] X. Wang and M. Braun, “Explainable machine learning-based fatigue assessment of 316L stainless steel fabricated by laser-powder bed fusion,” *Int. J. Fatigue*, vol. 190, no. August 2024, p. 108588, 2025, doi: 10.1016/j.ijfatigue.2024.108588.
[13] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models Fifth Edition*. New York: McGraw-Hill/Irwin, 2005.
[14] S. W. Tyas and L. A. Puspitasari, “Geographically weighted generalized poisson regression model with the best kernel function in the case of the number of postpartum maternal mortality in east java,” *MethodsX*, vol. 10, no. January, p. 102002, 2023, doi: 10.1016/j.mex.2023.102002.
[15] R. Cano-guervos, J. Chica-olmo, and J. Chica-garcia, “Journal of Retailing and Consumer Services The effect of disruptive change on the spatial variation of commercial rental prices : The case of the COVID-19 pandemic,” *J. Retail. Consum. Serv.*, vol. 82, no. September 2024, p. 104111, 2025, doi: 10.1016/j.jretconser.2024.104111.
[16] S. W. Tyas, Gunardi, and L. A. Puspitasari, “Geographically weighted generalized poisson regression model with the best kernel function in the case of the number of postpartum maternal mortality in east java,” *MethodsX*, vol. 10, no. January, p. 102002, 2023, doi: 10.1016/j.mex.2023.102002.
[17] M. N. Lessani and Z. Li, “SGWR : similarity and geographically weighted regression,” *Int. J. Geogr. Inf. Sci.*, vol. 38, no. 7, pp. 1232–1255, 2024, doi: 10.1080/13658816.2024.2342319.
[18] I. Gollini, B. Lu, M. Charlton, C. Brunsdon, and P. Harris, “GWmodel : An R Package for Exploring Spatial,” vol. 63, no. 17, 2015.
[19] C. Sutherland, D. Hare, P. J. Johnson, D. W. Linden, R. A. Montgomery, and E. Droge, “Practical advice on variable selection and reporting using Akaike information criterion,” no. Dic, 2023.
[20] N. Nadja *et al.*, “Conceptualizing the Circular Economy (Revisited) An Analysis of 221 Definitions,” *Resour. Conserv. Recycl.*, vol. 194, p. 107001, 2026, doi: 10.1016/j.resconrec.2023.107001.
[21] J. K. D. D. G. V. M. A. P. Dinis, “Raising Awareness on Solid Waste Management through Formal Education for Sustainability : A Developing Countries,” vol. 6, no. 1, 2021, doi: https://doi.org/10.3390/recycling6010006.
[22] W. Fadhullah, N. Iffah, N. Imran, S. Norkhadijah, S. Ismail, and M. H. Jaafar, “Household solid waste management practices and perceptions among residents in the East Coast of Malaysia,” *BMC Public Health*, vol. 22, no. 1, pp. 1–20, 2022, doi: 10.1186/s12889-021-12274-7.
[23] Z. Zhang *et al.*, “Municipal solid waste management challenges in developing regions : A comprehensive review and future perspectives for Asia and Africa,” *Sci. Total Environ.*, vol. 930, no. April, p. 172794, 2024, doi: 10.1016/j.scitotenv.2024.172794.
[24] L. K. Ncube, A. U. Ude, E. N. Ogunmuyiwa, R. Zulkifli, and I. N. Beas, “An Overview of Plastic Waste Generation and Management in Food Packaging Industries,” *MDPI*, vol. 6, no. 1, 2021, doi: https://doi.org/10.3390/recycling6010012.