APPLICATION OF ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION (RGWR) FOR MODELING STUNTING TODDLERS IN WEST JAVA 2024

(Thesis)

By

RIZKIA SYEFIRA BANISA
2217031117

FACULTY OF MATHEMATICS AND NATURAL SCIENCES
UNIVERSITY OF LAMPUNG
BANDAR LAMPUNG
2026

ABSTRACT

APPLICATION OF ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION (RGWR) FOR MODELING STUNTING TODDLERS IN WEST JAVA 2024

By

RIZKIA SYEFIRA BANISA

Global linear regression assumes that the relationship between the response and predictor variables is constant across all regions, making it less capable of accommodating spatial heterogeneity. Geographically Weighted Regression (GWR) was developed to address this issue. However, the method is sensitive to outliers, which may lead to unstable parameter estimates. One way to overcome outliers in GWR is to use Robust techniques, one of which is the MM-Estimator. This study aims to apply Robust Geographically Weighted Regression (RGWR) with MM-estimation to model the number of stunting toddlers in 27 regencies and cities in West Java in 2024. The predictor variables include the number of health facilities, the percentage of exclusive breastfeeding, the percentage of households with proper sanitation, the adult literacy rate, and the number of malnourished toddlers. The results indicate the presence of spatial heterogeneity and nine outlier locations. The RGWR model, constructed using an adaptive Bisquare kernel weighting function, outperformed the GWR model with an AIC value of 404.0302 and an MSE value of 0.0584. Furthermore, the effects of predictor variables on the number of stunting toddlers varied across regions, indicating differences in regional characteristics among regencies and cities in West Java.

Keywords: Robust Geographically Weighted Regression, Spatial Heterogeneity, MM-estimation, Outlier, Stunting Toddlers.

ABSTRAK

PENERAPAN ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION (RGWR) UNTUK PEMODELAN JUMLAH BALITA STUNTING DI JAWA BARAT TAHUN 2024

Oleh

RIZKIA SYEFIRA BANISA

Regresi linear global mengasumsikan bahwa hubungan antara variabel respon dan variabel prediktor bersifat sama pada seluruh wilayah sehingga kurang mampu mengakomodasi heterogenitas spasial. *Geographically Weighted Regression* (GWR) dikembangkan untuk mengatasi permasalahan tersebut, namun metode ini sensitif terhadap keberadaan pencilan yang dapat menyebabkan estimasi parameter menjadi tidak stabil. Salah satu cara untuk mengatasi pencilan pada GWR adalah dengan menggunakan metode penduga Robust yaitu Estimasi-MM. Penelitian ini bertujuan menerapkan *Robust Geographically Weighted Regression* (RGWR) dengan Estimasi-MM untuk memodelkan jumlah balita stunting di 27 kabupaten/kota di Jawa Barat tahun 2024. Variabel prediktor yang digunakan meliputi jumlah fasilitas kesehatan, persentase ASI eksklusif, persentase rumah tangga dengan sanitasi layak, tingkat melek huruf, dan jumlah balita dengan gizi buruk. Hasil analisis menunjukkan adanya heterogenitas spasial dan 9 lokasi pencilan. Model RGWR yang dibangun menggunakan fungsi pembobot kernel *Bisquare* adaptif menghasilkan kinerja yang lebih baik dibandingkan model GWR, dengan nilai AIC sebesar 404,0302 dan MSE sebesar 0,0584. Selain itu, pengaruh variabel prediktor terhadap jumlah balita stunting berbeda pada setiap wilayah, yang menunjukkan adanya variasi karakteristik antar kabupaten/kota di Jawa Barat.

Kata Kunci: *Robust Geographically Weighted Regression*, Heterogenitas Spasial, Estimasi-MM, Pencilan, Balita Stunting.

APPLICATION OF ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION (RGWR) FOR MODELING STUNTING TODDLERS IN WEST JAVA 2024

RIZKIA SYEFIRA BANISA

Thesis

Submitted as a Partial Fulfillment of the Requirements for the Degree of

BACHELOR OF MATHEMATICS

In the

Department of Mathematics

Faculty of Mathematics and Natural Sciences

FACULTY OF MATHEMATICS AND NATURAL SCIENCES
UNIVERSITY OF LAMPUNG
BANDAR LAMPUNG
2026

Thesis Title	:	**APPLICATION OF ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION (RGWR) FOR MODELING STUNTING TODDLERS IN WEST JAVA 2024**
Student Name	:	**Rizkia Syefira Banisa**
Student ID Number	:	**2217031117**
Study Program	:	**Mathematics**
Faculty	:	**Mathematics and Natural Sciences**

APPROVED BY

1. Advisory Committee

Dr. Khoirin Nisa, S.Si., M.Si.	**Dr. Bernadhita H. S. U, S.Si., M. Sc.**
NIP. 197407262000032001	NIP. 199206302023212034

2. Head of the Department of Mathematics

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

VALIDATED BY

1. Examination Committee

Head	: **Dr. Khoirin Nisa, S.Si., M.Si.**	....
Secretary	: **Dr. Bernadhita H. S. U, S.Si., M.Sc**	....
Examiner Non-Advisor	: **Widiarti, S.Si., M.Si.**	....

2. Dean of the Faculty of Mathematics and Natural Sciences

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 19711001 200501 1 002

Date of Thesis Examination: **June 25, 2026**

STUDENT'S THESIS DECLARATION

The undersigned below:

Name : **Rizkia Syefira Banisa**

Student ID Number : **2217031117**

Study Program : **Mathematics**

Thesis Title : **APPLICATION OF ROBUST GEOGRAPHICALLY WEIGHTED REGRESSION (RGWR) FOR MODELING STUNTING TODDLERS IN WEST JAVA 2024**

Hereby declares that this thesis is the result of my own work. If, at any time in the future, this thesis is proven to be plagiarized or written by another person, I am willing to accept sanctions in accordance with the applicable academic regulations.

Bandar Lampung, June 25, 2026

Author

Rizkia Syefira Banisa

NPM. 2217031117

AUTHOR BIOGRAPHY

Rizkia Syefira Banisa was born in Prabumulih, South Sumatera, on July 6, 2003. She is the first child of five siblings. She began her first education at TK Aisyiyah 2 Prabumulih. Then, she completed her elementary school at SDN 1 Prabumulih from 2009–2015, her junior high school at SMPN 10 Prabumulih from 2015–2018, and her senior high school at SMAN 6 Prabumulih from 2018–2021.

In 2022, she was admitted to the Undergraduate Program in Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, through the SBMPTN admission pathway. During her studies, she actively participated in student organizations, serving as a member of the External Affairs Division of the Mathematics Student Association (HIMATIKA) in 2023, she also served as an Expert Staff member of the Media and Information Department (MEDINFO) in Badan Eksekutif Mahasiswa (BEM) 2024.

In addition to her organizational involvement, she actively participated in various committee activities, including the Mathematics Department Anniversary (DINAMIKA) and MIPA Expo. Furthermore, she completed an internship at PT PLN Nusantara Power Unit Pembangkitan Bandar Lampung in the Engineering and Quality Assurance (QA) Division at the end of 2024. She also participated in the Community Service Program (KKN) in Pematang Wangi Village, Tanjung Senang District, Bandar Lampung City.

WORDS OF INSPIRATION

“And that each person will have only what they have worked for.”

(QS. An-Najm: 39)

“If you never try, you'll never know just what you're worth..”

(Coldplay)

“Every cloud has a silver lining.”

DEDICATION

All praise and gratitude be to Allah SWT, whose endless mercy, blessings, and grace have enabled the author to complete this thesis. Every milestone achieved in this journey was made possible only through the strength and guidance provided by Him.

Dearest Father, Mother, and Siblings

Thank you for limitless love, prayers, and encouragement that kept me moving forward. You are the reason I am here today.

Supervisors and Examiner

Thank you for your mentorship, valuable insights, and dedication to guiding me through this academic endeavor.

Dearest Friends

Thank you for the laughter, support, care, ,and togetherness shared with the author throughout this journey.

Beloved Alma Mater, University of Lampung

ACKNOWLEDGEMENTS

Praise and gratitude be to the Almighty Allah SWT for His mercy, grace, and ease so that the author could complete the thesis entitled “Application of Robust Geographically Weighted Regression (RGWR) for Modeling Stunting Toddlers West Java 2024” as one of the requirements for obtaining a Bachelor’s degree in Mathematics at the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung.

During the preparation of this thesis, many parties provided assistance, support, guidance, and prayers. Therefore, the author would like to express sincere gratitude to:

1. Dr. Khoirin Nisa, S.Si., M.Si., as the first supervisor who has patiently shared her expertise, direction, motivation, and invaluable mentorship with the author.
2. Dr. Bernadhita Herindri Samodera Utami, S.Si., M.Sc., as the second supervisor who has generously provided her time, constructive ideas, and guidance.
3. Mrs. Widiarti, S.Si., M.Si., as the examiner whose insightful critiques and recommendations have significantly improved the quality of this work.
4. Dr. Notiragayu, S.Si., M.Si., as the academic advisor for her direction and guidance during the author's academic journey.
5. Dr. Aang Nuryaman, S.Si., M.Si., as the Head of the Mathematics Department, Faculty of Mathematics and Natural Sciences, University of Lampung.
6. Dr. Eng. Heri Satria, S.Si., M.Si., as the Dean of the Faculty of Mathematics and Natural Sciences, University of Lampung.
7. All Lecturers and Staff of the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung, who have

provided knowledge, assistance, and valuable experiences to the author during the study period.

8. The author’s dearest parents, Papap and Mama, author’s greatest source of strength. Thank you for your unceasing prayers, genuine love, and unwavering moral and financial support that have guided me through every step of this journey until the end.
9. The author’s dearest siblings, Mia, Diaz, Daru and Dery, whose endless attention, support, and encouragement always motivate the author to pursue a brighter future.
10. The author’s dearest friends, thank you for your companionship through every joy and sorrow during this college journey, up until the completion of this thesis.
11. All parties who cannot be mentioned one by one who have helped and provided support to the author in completing this thesis.
12. Lastly, thank you to myself for never giving up throughout this university life and the journey of living independently. Thank you for surviving this far. It was never an easy path to walk, but here we are now.

The author realizes that this thesis still has shortcomings. Therefore, the author expects constructive criticism and suggestions for future improvement. Hopefully, this thesis can provide benefits to readers and parties in need.

Bandar Lampung, June 25, 2026

Rizkia Syefira Banisa

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

1.1 Background and Problem Statement

Regression is a statistical method used to explain the relationship between one dependent variable and one or more independent variables through a mathematical model. This method is widely applied in various scientific because it can provide a quantitative picture of the influence of independent variables on the dependent variable (Montgomery et al., 2012). In practice, the accuracy of regression results depends heavily on the model's suitability to the structure of the data being analyzed. When data is heterogeneous, particularly data with locational elements, conventional regression models tend to produce parameter estimates that are less representative of the actual conditions.

In reality, real-world phenomena often demonstrate differences in characteristics between regions. Different geographic, social, economic, and demographic conditions cause a variable to have unequal effects in each location. Data that depicts these variations between regions is called spatial data, which is data that contains a location element and can demonstrate diversity between regions. According to Anselin (1988), this difference is known as spatial heterogeneity, which is a condition where the relationship between dependent and independent variables is different in each observation area. The first law of geography proposed by Tobler (1970) states that "everything is interconnected, but something that is close has a greater influence than something that is far away." This statement emphasizes that each region has unique local characteristics so that global regression models are not always able to describe the variations that occur across regions.

One of the widely used approach to analyzing spatial heterogeneity is Geographically Weighted Regression (GWR). This model allows each location to have different regression parameters, thus capturing spatial variation between regions (Fotheringham et al., 2002). However, various studies have shown that GWR has weaknesses because it is sensitive to the presence of outliers or extreme data, which can cause parameter estimates to become unstable and reduce model accuracy (Wulandari et al., 2019). Robust Geographically Weighted Regression (RGWR) method, which incorporates robust statistics concepts, was developed to overcome this issue by reducing the influence of outliers, thereby producing more stable and accurate spatial models for social and health data that often contain extreme values.

Outliers are common in social and public health data, including the phenomenon of stunted toddlers. According to the World Health Organization (WHO), stunting is a condition of growth failure in children caused by chronic malnutrition, recurrent infections, and a lack of psychosocial stimulation in early life Ministry of Health of the Republic of Indonesia (2024). The estimated number is 4,482,340 children, equivalent to 19.8% of the national toddler total, with West Java Province has the highest number of stunted toddlers, at around 638,000 cases.

Several previous studies have demonstrated the effectiveness of the Robust Geographically Weighted Regression (RGWR) method in analyzing spatial data containing outliers. Nurhayati et al. (2018) applied RGWR to model diarrhea incidence in Semarang City. The results showed that RGWR model performed better than conventional GWR with a Mean Absolute Percentage Error (MAPE) value of 16.3396%, indicating an increase in model accuracy after addressing the influence of outliers. Similarly, Warsito et al. (2018) used RGWR to model the Air Polluter Standard Index (APSI) in Surabaya City and concluded that RGWR successfully handled outliers and improved parameter estimation.

Wulandari et al. (2019) implemented RGWR on Gross Regional Domestic Product (GRDP) data across Java Island and showed that the model produced predictions closer to actual conditions compared to GWR. Then, Isnaini et al., (2019) applied RGWR with M and S estimators to GRDP data in East Java and found that the M-estimator produced the smallest Mean Absolute Deviation (MAD), proving its robustness in parameter estimation. Rahayu et al. (2023) also used RGWR to analyze factors influencing maternal mortality in Indonesia and found that the model achieved high accuracy in capturing spatial variation between regions. Continuing this line of research, Arum et al., (2024) applied RGWR to examine regional economic growth in West Java and confirmed that the model outperformed the conventional GWR in terms of accuracy and parameter stability with an R^2 value of 0.9285 and the smallest Mean Squared Error (MSE) of 0.0044.

Recent studies have further strengthened these findings. Siahaan & Husein (2024) examined stunting prevalence in North Sumatra and showed that RGWR produced stable estimates with a fixed Gaussian kernel provided more stable estimates with a MAPE accuracy level of 15%. Then, Setyowati et al. (2025) applied RGWR to stunting cases across Java Island and reported that the model effectively identified regional differences and improved prediction precision.

Based on these various research findings, it can be concluded that RGWR is a more effective method for analyzing spatial data containing outliers. Therefore, this study focuses on applying the RGWR method to analyze the factors influencing stunting in West Java. The results are expected to contribute to the development of spatial analysis methods in applied statistics and form the basis for policymaking in stunting mitigation at the regional level.

1.2 Research Purposes

This study aims to built a Robust Geographically Weighted Regression (RGWR) model for the number of stunting toddlers in West Java in 2024 and compare its performance with the Geographically Weighted Regression (GWR) model.

1.3 Benefits Of Research

The expected benefit of this research is to find out what variables significantly influence the number of stunting toddlers in regencies/cities in West Java and can be used as a reference for further research.

II. LITERATURE REVIEW

2.1 Linear Regression

Linear regression is a statistical analysis method used to study the relationship between one dependent variable and one or more independent variables. The main objective of regression analysis is to model the mathematical relationship between these variables so that the value of the dependent variable can be predicted based on the value of the independent variables (Montgomery et al., 2012). Systematically, the linear regression model can be written as follows:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \qquad , i = 1,2,3, \dots , n \tag{2.1}$$

where:

y_i : Dependent variable on the observation i-th.

β_0 : Intercept/constant.

β_k : The regression coefficient parameter of the independent variable to k-th.

x_{ik} : Independent variable i-th in observation k-th.

ε_i : Error in the observation i-th.

According to Gujarati & Porter (2009), One of the methods that can be used to obtain such estimates is Ordinary Least Squares (OLS) method. The principle of the OLS method is to minimize the sum of squared residuals produced by the model, as expressed in the following equation:

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \sum_{k=0}^{p} \hat{\beta}_k X_{ik} \right)^2 \tag{2.2}$$

Then parameter estimate β is obtained using the formula

$$\boldsymbol{\hat{\beta}} = (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T Y} \tag{2.3}$$

2.2 Multicollinearity Test

According to (Montgomery et al., 2012), a serious problem that can dramatically affect the usefulness of a regression model is multicollinearity, or near-linear dependence among the regression variables. Multicollinearity indicates the existence of a linear dependence among the independent variables. The presence of this near-linear dependence can drastically affect the ability to estimate the regression coefficients.

One method to detect multicollinearity is by using the Variance Inflation Factor (VIF), which is calculated using the following formula:

$$VIF = \frac{1}{1 - R_j^2} \tag{2.4}$$

Where R_j s the coefficient of determination obtained by regressing j-th independent variable on the remaining independent variables. The presence of multicollinearity can be assessed using the following hypotheses:

H_0 : There is no multicollinearity among the independent variables.

H_1 : There is multicollinearity among the independent variables.

The decision is based on the VIF value. H_0 is rejected if all independent variables have VIF ≥ 10. If H_0 is rejected that's indicating the presence of multicollinearity.

2.3 Spatial Heterogeneity Test

Differences from one location to another are a sign of spatial heterogeneity. Spatial heterogeneity, or variation in variance across locations, is one of the characteristics of spatial data, especially in spatial analysis using a point-based approach (Siswanto et al., 2022). Spatial heterogeneity can be determined by conducting the Breusch-Pagan test with the following hypothesis (Nurhayati et al., 2018):

H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$(No spatial heterogeneity)
H_1: there is at least one $\sigma_1^2 \neq \sigma^2$ (There is spatial heterogeneity)
Statistic test:

$$BP = \frac{1}{2}\boldsymbol{f}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{f} \tag{2.5}$$

with:

$\mathbf{f} = (f_1, f_2, \ldots, f_n)^T \quad , f = \left(\frac{e_i^2}{\sigma^2} - 1\right)$

$e_i = y_i - \hat{y}_i$

where:

e_i : Least square error of the i-th observation.

$\boldsymbol{Z}$: Matrix size of $n \times (p+1)$ containing the standardized explanatory variables.

The decision-making criteria are determined based on the BP value and $p - value$. Reject H_0 if BP $> \chi^2_{(\alpha;p)}$ where p is the number of independent variables, or $p - value < \alpha$. If H_0 is rejected, it indicates that there are differences in characteristics between one region and another.

2.4 Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) is one of the methods used to estimate data that exhibits spatial variability, using the weighted least squares estimation method (Fotheringham et al., 2002). This model is a form of local linear regression that produces model parameters which are locally specific for each location where the data are collected. The GWR model can be written as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)\, x_{ik} + \varepsilon_i \,, i = 1,2, \ldots n \tag{2.6}$$

where:

y_i : Value of the response variable observation for the i–th observation.

β_0 : Intercept

u_i, v_i : Geographic coordinates (longitude and latitude).

β_k : Value of the k–th regression coefficient.

x_{ik} : Value of the k–th predictor variable at the i–th observation location.

ε_i : Error term form the i–th observation

2.4.1 Parameter Estimation of GWR Model

The parameter estimation of the GWR model is carried out using the Weighted Least Squares (WLS) method, which assigns different weights to each location (u_i, v_i) observation (Anselin, 1988). The weighting for each location (u_i, v_i) is denoted as $w_j(u_i, v_i)$, with $j = 1,2, \ldots, n$. The GWR model can be written as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)\, x_{ik} + \varepsilon_i\,, \qquad i = 1,2, \ldots n$$

The objective is then to minimize the residual sum of squares as follows:

$$\sum_{j=1}^{n} w_j(u_i v_i)\, \varepsilon_j^2 = \sum_{j=1}^{n} w_j(u_i v_i) \left[y_i - \beta_0(u_i, v_i) - \sum_{k=1}^{p} \beta_k(u_i v_i)\, x_{ik} \right]^2 \qquad (2.7)$$

If written in matrix form, it becomes:

$$\begin{aligned}
\boldsymbol{\varepsilon^T W}(u_i, v_i)\boldsymbol{\varepsilon} &= [\boldsymbol{Y} - \boldsymbol{X\beta}(u_i, v_i)]^T \boldsymbol{W}(u_i, v_i)[\boldsymbol{Y} - \boldsymbol{X\beta}(u_i, v_i)] \\
&= \boldsymbol{Y^T W}(u_i, v_i)\boldsymbol{Y} - \boldsymbol{Y^T W}(u_i, \boldsymbol{v_i})\boldsymbol{X\beta}(u_i, v_i) \\
&\quad - \boldsymbol{\beta^T}(u_i, \boldsymbol{v_i})\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{Y} + \boldsymbol{\beta^T}(u_i, v_i)\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{X\beta}(u_i, v_i) \\
&= \boldsymbol{Y^T W}(u_i, v_i)\boldsymbol{Y} - \boldsymbol{2\beta^T}(u_i, \boldsymbol{v_i})\boldsymbol{X^T W}(u_i, v_i) \\
&\quad + \boldsymbol{\beta^T}(u_i, v_i)\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{X\beta}(u_i, v_i) \qquad (2.8)
\end{aligned}$$

Then, equation (2.8) is differentiated with respect to the parameter $\beta'(u_i, v_i)$, and the derivative is set equal to zero, since the purpose of the weighted least squares (WLS) method is to minimize this function.

$$\frac{\partial \boldsymbol{\varepsilon}^T \boldsymbol{W}(u_i, v_i)\boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta^T}(u_i, v_i)} = 0$$

$$\begin{aligned}
-2\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{Y} + 2\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{X\beta}(u_i, v_i) &= 0 \\
2\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{Y} &= 2\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{X\beta}(u_i, v_i) \\
\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{Y} &= \boldsymbol{X^T W}(u_i, v_i)\boldsymbol{X\beta}(u_i, v_i)
\end{aligned}$$

$$[\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{X}]^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{X}\boldsymbol{\beta}(u_i,v_i) = [\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{X}]^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}$$

$$\widehat{\boldsymbol{\beta}}(u_i,v_i) = [\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{X}]^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{Y}$$

Hence, the parameter estimation of the GWR model is obtained as follows:

$$\widehat{\boldsymbol{\beta}}(u_i,v_i) = [\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{X}]^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{Y} \tag{2.9}$$

where

Y : Response variable with dimension $n \times 1$.

X : Predictor variable matrix.

$\widehat{\boldsymbol{\beta}}(u_i,v_i)$: Local regression coefficient at location i.

$\boldsymbol{W}(u_i,v_i)$: Diagonal weight matrix $n \times n$ is computed for each location i.

The matrix $\boldsymbol{W_i}(u_i,v_i)$ denotes the spatial weighting structure in GWR model, expressed as a diagonal matrix size $n \times n$:

$$\boldsymbol{W_i}(u_i,v_i) = \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{in} \end{bmatrix}$$

This matrix inherently exhibits discontinuous behaviour, a kernel-based weighting function is employed to generate continuous values of w_{ij} based on the spatial distance d_{ij} between observations (Fotheringham et al., 2002).

2.4.2 Spatial Weighting Function

The role of weighting in the GWR model is very important because the weighting value represents the distance between observations. One of the most commonly used methods in the weighting process of the GWR model is the kernel function. The kernel function represents the spatial proximity between observation points. There are several types of kernel functions, and one of the most widely used is the Bi-square kernel function (Fotheringham et al., 2002). Adaptive bisquare kernel function is formulated as follows:

$$w_{ij}(u_i, v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h_i}\right)^2\right]^2, & d_{ij} \leq h_i \\ 0, & d_{ij} > h_i \end{cases} \tag{2.10}$$

where:

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{2.11}$$

d_{ij} : Euclidean distance between location i and location j

h : Optimum Bandwidth

2.4.3 Bandwidth

The choice of bandwidth h is highly influential because it controls the balance between the smoothness and the accuracy of the GWR model estimation. In other words, bandwidth determines how much neighboring observations contribute to the local parameter estimates. Selecting an appropriate bandwidth is therefore essential, as it directly affects the estimator's ability to produce optimal and reliable results. A bandwidth that is too small will yield unstable estimates with high variance, whereas an overly large bandwidth will lead to oversmoothing and the loss of local spatial variation (Tizona et al., 2017).

Several techniques are available to determine the optimal bandwidth, one of which is Cross Validation (CV).

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(h)]^2 \tag{2.12}$$

Which $\hat{y}_{\neq i}(h)$ represents the predicted value of the observation y_i, obtained by excluding the i −th observation from the prediction process and using the optimal bandwidth h.

2.5 Outlier

According to Hampel et al., (1986), An outlier is an observation that deviates from the general pattern formed by the majority of the data. The presence of outliers can affect the results of statistical analysis, including bias, which can lead to errors in decision-making. In the context of spatial data, a spatial outlier is defined as a local instability or spatial object that has relatively extreme non – spatial attributes compared to other objects around it in the neighborhood (Nurhayati et al., 2018). However, sometimes outliers provide necessary information and should be retained in the data (Cohen et al., 2003).

DFFITS (Difference in Fit Standardized) is a statistic that measures the magnitude of the change in a model's predicted value when a particular observation is removed from the dataset. Mathematically, the formula for calculating the DFFITS value for the i-th observation is as follows:

$$(DFFITS)_i = t_i\sqrt{\frac{h_{ii}}{1-h_{ii}}} \tag{2.13}$$

In the formula above, the component t_i represents the studentized residual, and h_{ii} denotes the leverage or hat value, which measures the contribution of a data point to the model's prediction. The criterion for classifying a data point as an outlier is when the absolute value of DFFITS exceeds a certain cut-off threshold, namely $|DFFITS| > 2\sqrt{\frac{p}{n}}$ where p the total number of parameters (including the intercept) in OLS and n is the total number of observations.

2.6 Robust Geographically Weighted Regression (RGWR)

As previously discussed, the GWR model is sensitive to outliers because its parameter estimation is based on the Weighted Least Squares (WLS) method, which minimizes the sum of squared residuals. Therefore, a robust technique is required to address the presence of outliers. There are several estimation methods in robust regression, three of which are M-estimation (Maximum likelihood type estimator), S-estimation (Scale estimator), and MM-estimation (Modified M-estimation) (Prahutama & Rusgiyono, 2021).

Building upon this idea, the Robust Geographically Weighted Regression (RGWR) model is developed from the Geographically Weighted Regression (GWR) to improve robustness against the influence of outliers. The model maintains the fundamental principle of GWR, which assigns each observation location its own set of regression parameters, but incorporates a robust weighting system derived from MM – estimation so that parameter estimates are less affected by extreme observations (Wheeler, 2014). In this approach, each observation receives an adaptive weight based on its standardized residual through a robust Tukey Bi – square function

2.6.1 M–Estimation

M–estimation is a robust regression technique derived from the maximum likelihood estimation (MLE) principle, designed to produce parameter estimates that are less sensitive to the influence of outliers. This technique is designed to reduce the influence of extreme or outlying observations on parameter estimates. In some cases, certain data points may be removed when their presence unduly affects the model; however, this is not always appropriate, especially when those data are important or representative (Susanti et al., 2014).

M–estimation principle is to minimize the residual function ρ:

$$\hat{\beta}_M = \min_{\beta} \sum_{i=1}^{n} \rho(e) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{\hat{\sigma}_M}\right)$$

$$\hat{\beta}_M = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=1}^{p} x_{ij}\beta_j}{\hat{\sigma}_M}\right) \quad (2.17)$$

Where $\rho(e)$ is a symmetric function of the residuals that assigns a contribution from each residual to the overall objective function. The scales estimator $\hat{\sigma}_M$ can be obtained using the following formula:

$$\hat{\sigma}_M = \frac{Med|e_i - Med(e_i)|}{0.675} \tag{2.18}$$

The constant 0.675 is used to make the estimator $\hat{\sigma}_M$ unbiased for normally distributed errors with sample size n (Montgomery et al., 2012).

2.6.2 S–Estimation

The S-estimation (scale statistic) belongs to the group of high-breakdown point estimators proposed by Rousseeuw and Yohai (1984). Unlike the M-estimation method, which has the drawback of disregarding the overall data distribution and relies only on the median as a weighting measure, the S-estimator is derived from minimizing a robust scale M-estimator (Abonazel & Rabie, 2019). The S–estimation is defined as:

$$\hat{\beta}_S = min_\beta\, \hat{\sigma}_S\,(e_1, e_2, \dots, e_n) \tag{2.19}$$

With determining minimum robust scale estimator $\hat{\sigma}_S$ and satisfying

$$\min \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=1}^{p} x_{ij}\beta_j}{\hat{\sigma}_S}\right) \tag{2.20}$$

With robust scale formula:

$$\hat{\sigma}_S = \sqrt{\frac{1}{nK}\sum_{i=1}^{n} w_i e_i^2} \tag{2.21}$$

where

$$w_i = w_i^S(u_i) = \frac{\rho(u_i)}{u_i^2} \tag{2.22}$$

K is turning constant and $u_i = \frac{\varepsilon_i}{\hat{\sigma}_S}$. The initial estimate $(\hat{\sigma}_S^0)$ can be written:

$$\hat{\sigma}_S^0 = \frac{Med|e_i - Med(e_i)|}{0.675} \tag{2.23}$$

S-estimator are more robust than the M-estimator, because S-estimators have smaller asymptotic bias and smaller asymptotic variance in the case of contaminated data

2.6.3 Parameter Estimation of RGWR using MM-Estimation

MM-estimation procedure is used to estimate regression parameters using S-estimation, which is used to minimize the residual scale of residuals and obtain a robust scale, followed by the M-estimation step to produce efficient parameter estimates. The purpose of MM-estimation is to obtain estimators with a high breakdown value and higher efficiency. The breakdown value itself is a general measure of the proportion of outliers that can still be handled before the observations begin to affect the model (Yimnak & Piampholphan, 2021). The MM-estimation procedure combines two steps.

S-estimation aims to obtain initial estimate of regression parameters and robust scale. The robust scale ($\hat{\sigma}_S$) is defined by minimizing the residual scale that satisfies (Rousseeuw & Yohai, 1984):

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0\left(\frac{\varepsilon_i}{\hat{\sigma}_S}\right) = K \tag{2.24}$$

where $\rho_0(\cdot)$ is a Tukey bi-square objective function. The estimation of $\hat{\sigma}_S$ is performed iteratively until convergence is reached using equation (2.21).

After obtaining the fixed robust scale ($\hat{\sigma}_S$) from S-estimation, M-estimation is performed to estimate the regression coefficients β efficiently by minimizing the weighted loss function.

$$\min_{\beta}\sum_{i=1}^{n} \rho\left(\frac{\varepsilon_i}{\hat{\sigma}_S}\right) \tag{2.25}$$

To obtain the solution of the above function, the first derivative is taken with respect to each parameter β_j, resulting in the following equation:

$$\sum_{i=1}^{n} x_{ij}\psi(v_i) = 0, \tag{2.26}$$

Where $v_i = \frac{y_i - \sum_{j=1}^{p} x_{ij}\beta_j}{\widehat{\sigma_S}}$ and $\psi(v_i) = \rho'(v_i)$. The influence function $\psi(v_i)$ can be expressed as a robust weight function ω_i through this equation:

$$\omega_i = \omega(v_i) = \frac{\psi(v_i)}{v_i} \tag{2.27}$$

Subtituting (2.27) into the left side of equation (2.26):

$$\sum_{i=1}^{n} x_{ij}\psi(v_i) = \sum_{i=1}^{n} x_{ij}\psi(v_i)\left(\frac{v_i}{v_i}\right)$$

$$= \sum_{i=1}^{n} x_{ij}\left(\frac{\psi(v_i)}{(v_i)}\right)v_i$$

$$= \sum_{i=1}^{n} x_{ij}\,\omega_i v_i$$

$$= \sum_{i=1}^{n} x_{ij}\,\omega_i \frac{y_i - \sum_{j=1}^{p} x_{ij}\beta_j}{\widehat{\sigma_S}} \tag{2.28}$$

Because the condition $\sum x_{ij}\omega_i v_i = 0$ must be satisfied, and $v_i = \frac{y_i - \sum_{j=1}^{p} x_{ij}\beta_j}{\widehat{\sigma_S}}$then dividing by the constant $\widehat{\sigma_S}$, does not change the solution point β. so the equation (2.28) become:

$$\sum_{i=1}^{n} x_{ij}\,\omega_i y_i - \sum_{i=1}^{n} x_{ij}\,\omega_i \sum_{j=1}^{p} x_{ij}\beta_j = 0$$

$$\sum_{i=1}^{n} x_{ij}\,\omega_i y_i = \sum_{i=1}^{n} x_{ij}\,\omega_i \sum_{j=1}^{p} x_{ij}\beta_j \tag{2.29}$$

ω_i is a robust weight function (Tukey bi-square) with the following formula:

$$\omega_i = \begin{cases} \left[1 - \left(\frac{v_i}{4.685}\right)^2\right]^2, & -4.685 \le v_i \le 4.685 \\ 0, & v_i < -4.685 \text{ or } v_i > 4.685 \end{cases} \tag{2.30}$$

In matrix form, equation (2.29) can be written as:

$$\boldsymbol{X^T\omega_i Y} = \boldsymbol{X^T\omega_i X\beta}$$

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{MM}} = (\boldsymbol{X^T\omega_i X})^{-1}\boldsymbol{X^T\omega_i Y} \tag{2.31}$$

where $\boldsymbol{\omega_i}$ is an $n \times n$ diagonal matrix containing the robust weights $(\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \dots, \boldsymbol{\omega_i})$. Thus, the parameter estimator for MM-estimation is obtained in equation (2.31) for global regression.

In GWR model, parameter β varies for each location (u_i, v_i), capturing spatial heterogeneity among observations. Therefore, each local model is calibrated using a spatial kernel weighting matrix $\boldsymbol{W_i}$ of size $n \times n$. Since two sets of weights (spatial weights and robust weights) both are diagonal matrices, the equation at location (u_i, v_i), becomes:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{MM}}(u_i, v_i) = (\boldsymbol{X^T \omega_i W_i X})^{-1} \boldsymbol{X^T \omega_i W_i Y}$$

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{MM}}(u_i, v_i) = (\boldsymbol{X^T W_i^* X})^{-1} \boldsymbol{X^T W_i^* Y} \tag{2.32}$$

where $\boldsymbol{W_i^*}$ is an $n \times n$ combined weighting matrix defined as $\boldsymbol{W_i^*} = \boldsymbol{\omega_i W_i}$, The spatial weighting matrix $\boldsymbol{W_i}$ represents the geographic proximity among observations, whereas the robust weighting matrix $\boldsymbol{\omega_i}$ controls the influence of outliers in the data. Therefore, to obtain the value of $\boldsymbol{\beta_{MM}}(u_i, v_i)$, an iterative procedure known as Iteratively Reweighted Least Squares (IRLS) is employed (Fox & Weisberg, 2019).

2.6.4 Parameter Testing of RGWR Model

Parameter testing on the RGWR model is conducted to identify the factors influencing the number of stunted toddlers in every regencies/city of West Java in 2024.

1. Simultaneous Parameter Significance Test

The simultaneous test of RGWR model is conducted to determine whether the RGWR model obtained provides a better fit compared to the classical linear regression model (Fotheringham et al., 2002). The hypotheses this test are as follows:

H_0: $\beta_k(u_i, v_i) = \beta_k;\ k = 1,2, \dots, p$ and $i = 1,2, \dots, n$

(There is no significant difference between the RGWR model and OLS model)

H_1: at least one $\beta_k(u_i, v_i) \neq \beta_k$

(There is a significant difference between the OLS model and RGWR model)

Statistic test:

$$F_{cal} = \frac{(RSS_{OLS} - RSS_{RGWR})/d_1}{RSS_{OLS}/(n-p-1)} \sim F_{\left(\frac{d_1^2}{d_2}, n-p-1\right)}$$

where:

RSS_{RGWR} : $\boldsymbol{Y^T}(\boldsymbol{I} - \boldsymbol{S})^{\boldsymbol{T}}(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y}$

RSS_{OLS} : $\boldsymbol{Y^T}(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$

$\boldsymbol{H}$: $\boldsymbol{X}(\boldsymbol{X^T X})^{-\boldsymbol{1}}\boldsymbol{X^T}$

d_1 : $n - p - 1 - \delta_1$

d_2 : $n - p - 1 - 2\delta_1 + \delta_2$

δ_i : $\boldsymbol{tr}[(\boldsymbol{I} - \boldsymbol{S})^{\boldsymbol{T}}(\boldsymbol{I} - \boldsymbol{S})]$

n : number of observations

p : number of independent variables

with

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{X}_1^T(\boldsymbol{X}^T\boldsymbol{W}(u_1, v_1)\boldsymbol{\omega}_1^S\boldsymbol{X})^{-1}\boldsymbol{X}^T(u_1, v_1)\boldsymbol{\omega}_1^S \\ \boldsymbol{X}_2^T(\boldsymbol{X}^T\boldsymbol{W}(u_2, v_2)\boldsymbol{\omega}_2^S\boldsymbol{X})^{-1}\boldsymbol{X}^T(u_2, v_2)\boldsymbol{\omega}_1^S \\ \vdots \\ \boldsymbol{X}_n^T(\boldsymbol{X}^T\boldsymbol{W}(u_n, v_n)\boldsymbol{\omega}_n^S\boldsymbol{X})^{-1}\boldsymbol{X}^T(u_n, v_n)\boldsymbol{\omega}_n^S \end{bmatrix}$$

$\boldsymbol{S}$ is the projection matrix of the RGWR model, namely the matrix that projects the values of y and $\hat{y}$ in location (u_i, v_i). Reject H_0 if the statistic test $> F_{\alpha(\frac{d_1^2}{d_2}, n-p-1)}$, which indicates that the regression parameters vary across locations, thus the RGWR model is appropriate (Leung et al., 2000).

2. Partial Parameter Significance Test

Partial parameter testing is conducted to determine which explanatory variables significantly influence the number of stunting toddlers at each observation location. The hypotheses underlying this test are:
H_0: $\beta_k(u_i, v_i) = 0$ (The independent variable has no significant effect)
H_1: $\beta_k(u_i, v_i) \neq 0$ (The independent variable has significant effect)

Statistic test

$$t_{cal} = \frac{\hat{\beta}_k(u_i, v_i)}{SE(\hat{\beta}_k(u_i, v_i))}$$

with
$SE(\hat{\beta}_k(u_i, v_i))$: Standard error of the coefficient $\hat{\beta}_k(u_i, v_i)$.
Reject H_0 if $|t_{cal}| > t_{\left(\frac{\alpha}{2}, df\right)}$, which means there is a significant influence of the independent variable on the dependent variable.

2.7 Residual Diagnostic

Residual diagnostics are series of testing procedures performed on the residuals of the fitted Robust Geographically Weighted Regression (RGWR) model to verify whether the fundamental assumptions of the spatial regression model have been adequately satisfied (Comber et al., 2023). In this study, residual diagnostics were conducted through two tests, namely the residual normality test using Kolmogorov–Smirnov test and the residual spatial autocorrelation test using Moran's I index.

2.7.1 Residual Normality Test

The residual normality test was performed as a post-model diagnostic to evaluate whether the residuals produced by the RGWR model approximately follow a normal distribution. The method used for the normality test in this study is the Kolmogorov–Smirnov test. This test compares the cumulative distribution of the sample with a specified cumulative distribution, which in this case is the normal cumulative distribution (Massey Jr., 1951). The hypotheses used in this test are:

$H_0: F(x) = F_0(x)$ the residuals are normally distributed

$H_1: F(x) \neq F_0(x)$ the residuals are not normally distributed

with the test statistic defined as follows:

$$D_n = \max\ |F_n(x) - F_0(x)|$$

where

D_n : the maximum vertical distance between $F_n(x)$ and $F_0(x)$

$F_n(x)$: the empirical cumulative distribution function of the residuals

$F_0(x)$: the normal cumulative distribution function

H_0 is rejected if the $p - value < \alpha\ (0.05)$. If H_0 fails to be rejected, the residuals of RGWR model are considered normally distributed and the normally assumption is satisfied.

2.7.2 Residual Spatial Autocorrelation Test

Residual spatial autocorrelation occurs when the residual values at one location are correlated with residual values at geographically neighboring locations. The presence of spatial autocorrelation in the residuals indicates that there are spatial patterns not captured by the model, there by threatening the validity of statistical inference (Leung et al., 2000). The hypothesis used for residual spatial autocorrelation test as follows:

H_0: There is no spatial autocorrelation in the model residual

H_1: There is spatial autocorrelation in the model residual

Moran’s I statistic is defined as:

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}$$

with

$$I = \frac{n}{S_0} \; \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \, e_i e_j}{\sum_{i=1}^{n} e_i^2}$$

$$E(I) = -\frac{1}{(n-1)}$$

$$Var(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{(n^2 - 1) S_0^2} - \big(E(I)\big)^2$$

where

I : Moran's I value.

$E(I)$: Expected value (mean) of I

$Var(I)$: Variance of I

S_0 : $\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$

S_1 : $\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij} + w_{ji})^2}{2}$

S_2 : $\sum_{i=1}^{n} \sum_{j=1}^{n} \big(w_{ij} + w_{ji}\big)^2$

The decision rule for spatial autocorrelation test is H_0 is rejected if $|Z(I)| > Z_{\alpha/2}$ or $p - value < \alpha$. For an RGWR model that has successfully captured the spatial structure of the data, the residuals are expected to no longer exhibit significant spatial autocorrelation, indicating that the model has adequately extracted spatial information so that the remaining errors are spatially random

2.8 Goodness of Fit Measure

The goodness of fit measure is used to evaluate how well the model can explain the data. The Akaike Information Criterion (AIC) is one of the criteria that can be used to determine the goodness of fit of a regression model. A smaller AIC value indicates a better – fitting model, as it reflects a model that achieves a balance between goodness of fit and model complexity (Grasa, 1989). The formula for calculating AIC is as follows:

$$AIC = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n + tr(\boldsymbol{S}) \qquad (2.33)$$

where:

n : number of observations.

$\hat{\sigma}$: estimated standard deviation of the residuals.

$tr(\boldsymbol{S})$: trace of the hat matrix $\boldsymbol{S}$, which represents the effective number of parameters in the model.

Mean Square Error (MSE) provides an indication of how much error or deviation exists between the values predicted by a regression model and the actual values observed in the data. Models with lower MSE are considered better at estimating the true values, as a smaller MSE indicates that the model's predictions are closer to the actual data points Arum et al., (2024). Here is the formula for calculating MSE:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{2.34}$$

A value of MSE closer to zero indicates that the model produces more accurate predictions and fits the observed data well.

2.9 Stunting

Stunting is one of the most serious obstacles to human development, affecting approximately 162 million children under five years of age worldwide. According to the World Health Organization (WHO), stunting is defined as a condition in which a child's height-for-age is below minus two standard deviations (-2 SD) from the median of the WHO Child Growth Standards. This condition most frequently occurs among children aged 12–36 months and is often unrecognized, as physical differences between stunted and non-stunted children may not be immediately apparent (Siahaan & Husein, 2024). Several factors can contribute to the incidence of stunting, including the lack of parental knowledge regarding nutritional preparation during pregnancy and the first 1,000 days of life, inappropriate parenting practices, low birth weight conditions, and most commonly the family’s economic status (Fauziah et al., 2023).

III. RESEARCH METHODOLOGY

3.1 Time and Place of Research

This research was conducted in the even semester of 2025/2026, at the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung.

3.2 Research Data

This study uses secondary data, namely the number of stunting cases in West Java in 2024 obtained from “Open Data Jabar” (https://opendata.jabarprov.go.id/) along with its factors with observation units totaling 27 regencies/cities.

Table 1. Spatial Data Research Variables

Variables	Information	Unit
Y	Number of Stunting Toddlers	Person
X_1	Number of Health Facilities	Unit
X_2	Percentage of Exclusive Breastfeeding	Percent
X_3	Percentage of Households with Proper Sanitation	Percent
X_4	Adult Literacy Rate (≥15 years old)	Percent
X_5	Malnourished Toddlers	Person

3.3 Research Methodology

The steps taken in modeling the factors that influence the number of stunting cases in West Java using RGWR are:

1. Conduct descriptive analysis.
2. Perfom data standardization on independent variables using z – score method as follows (Cho et al., 2018):

$$Z = \frac{X - \bar{x}}{S}$$

3. Evaluate multicollinearity assumptions by observing the VIF values in equation 2.4.
4. Test spatial heterogeneity assumptions using Breusch – Pagan test in equation 2.5.
5. Estimate the GWR model.
   a. Compute the Euclidean distance.
   b. Determine optimal bandwidth using Cross Validation (CV).
   c. Calculate the spatial kernel weights using Bi-square function.
   d. Estimate the model parameters.
6. Identify outliers by computing DFFITS.
7. Estimate RGWR model using MM-estimation approach.
   a. Estimate the initial local regression coefficients $\beta^0(u_i, v_i)$ using S-estimation.
   b. Calculate residual obtained from the S-estimation process.
   c. Calculate value of robust scale $\hat{\sigma}_i = \hat{\sigma}_S$ based on the residuals from S-estimation.
   d. Compute standardized residuals $v_i = \frac{\varepsilon_i}{\hat{\sigma}_S}$.
   e. Perfom M-estimation stage by calculating robust weights $\boldsymbol{\omega_i}$ using the Tukey bisquare weighting function in Equation (2.30).
   f. Combine the spatial kernel weight $\boldsymbol{W_i}$ and the robust weight $\boldsymbol{\omega_i}$ into the robust spatial weight matrix $\boldsymbol{W_i^*} = \boldsymbol{W_i} \times \boldsymbol{\omega_i}$.

g. Estimate the RGWR parameters $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MM}}$ using WLS method with combined weighted $\boldsymbol{W}_{\boldsymbol{i}}^{*}$.

h. Repeat steps c – g iteratively using Iteratively Reweighted Least Squares (IRLS) to obtain a convergent value of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MM}}$.

8. Test the significance of RGWR model parameters using F-test for simultaneous significance and t-test for partial of each local coefficient $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MM}}$.
9. Perform residual diagnostic test.
10. Evaluate the model performance using AIC and MSE.

The research flow is presented in Figure 1.

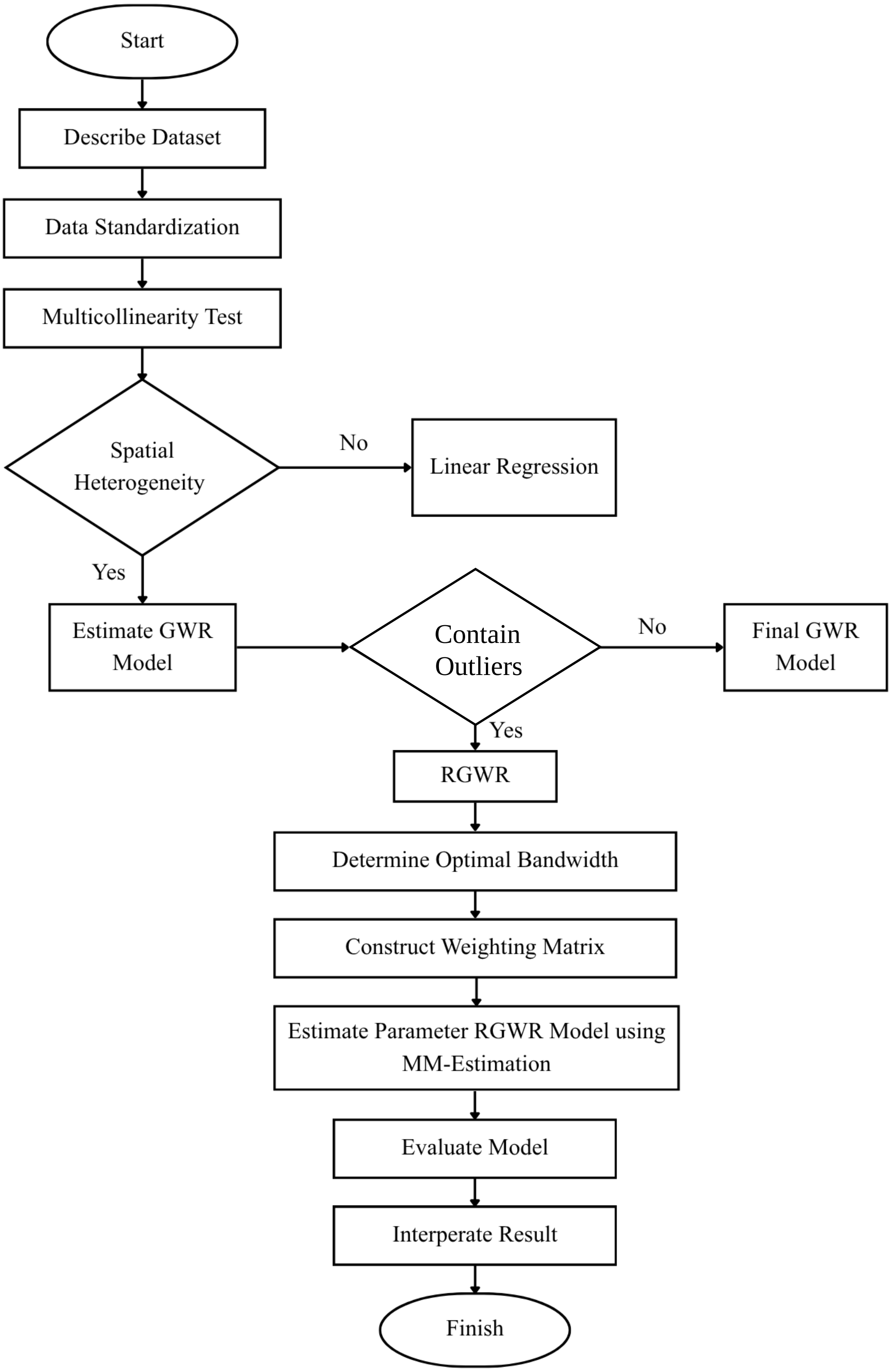

Figure 1. Flowchart RGWR Model

V. CONCLUSION

5.1 Conclusion

Based on the results of the study on Robust Geographically Weighted Regression (RGWR) modeling of the number of stunted toddlers in West Java Province in 2024, the RGWR model using MM-estimation was found to perform better than the Geographically Weighted Regression (GWR) model. This is evidenced by the lower AIC value (404.0302) and MSE value (0.0584) obtained by the RGWR model compared to the GWR model, which produced an AIC value of 486.2096 and an MSE value of 0.0698. The results suggest that considering spatial heterogeneity and reducing the influence of outliers through MM-estimation lead to improved model performance. Therefore, the RGWR model is more appropriate for modeling the number of stunted toddlers in West Java Province in 2024.

5.2 Recommendations

Future research is recommended to consider incorporating additional relevant variables in order to obtain more comprehensive results in explaining the factors influencing stunting. Furthermore, expanding the study area and increasing the number of observation units are suggested to improve the accuracy and stability of spatial parameter estimates in the RGWR model. From a methodological perspective, future studies may also consider developing the model by applying or combining it with other spatial methods to better capture more complex data characteristics

REFERENCES

Abonazel, M. R., & Rabie, A. R. 2019. The Impact of using Robust Estimations in Regression Models: An Application on the Egyptian Economy. *Journal of Advanced Research in Applied Mathematics and Statistics*. **4**(2): 8–16.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models* (Vol. 4). Springer Netherlands.

Arum, P. R., Ridwan, M., Alfidayanti, I., & Wasono, R. 2024. Robust Geographically Weighted Regression With Least Absolute Deviation (LAD) Estimation and M-Estimation on GDRP Of West Java Province. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, **18**(3): 1573–1584.

Cho, K., Yoon, T., Park, S., & Park, D. 2018. Using Standardization for Fair Data Evaluation. *International Conference on Advanced Communications Technology (ICACT 2018)*. 585–589.

Cohen, J., Aiken, L. S., Cohen, P., & West, S. G. 2015. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd Edition. Lawrence Erlbaum Associates, Inc. New Jersey.

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., & Harris, P. 2023. A Route Map for Successful Applications of Geographically Weighted Regression. *Geographical Analysis*, **55**(1): 155–178.

Fauziah, J., Trisnawati, K. D., Rini, K. P. S., & Putri, S. U. 2023. Stunting: Penyebab, Gejala, dan Pencegahan. *Jurnal Parenting dan Anak*, **1**(2): 11.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.

Fox, J., & Weisberg, S. 2019. *An R Companion to Applied Regression*. 3^{rd} Edition. Sage Publications.

Grasa, A. A. 1989. *Econometric Model Selection: A New Approach*. Springer-Verlag.

Gujarati, D. N., & Porter, D. C. 2009. *Basic econometrics.* 5^{th} Edition. McGraw-Hill Irwin.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. 1986. *Robust statistics: The approach based on influence functions* (Online-Ausg). John Wiley & Sons, Inc.

Isnaini, B., Syafitri, U. D., & Aidi, M. N. 2019. Estimating the Parameters of a Robust Geographically Weighted Regression Model in Gross Regional Domestics Product in East Java. *International Journal of Sciences*. **48**(3), 150–160.

Kementerian Kesehatan RI. 2024. *Survei Status Gizi Indonesia (SSGI) dalam Angka 2024* (352). Kementerian Kesehatan Republik Indonesia.

Leung, Y., Mei, C.-L., & Zhang, W.-X. 2000. Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model. *Environment and Planning A: Economy and Space*, **32**(1): 9–32.

Massey Jr., F. J. 1951. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*. **46**: 68–78.

Montgomery, D. C., Peck, E. A., & Vining, G. G. 2012. *Introduction to Linear Regression Analysis*. 5^{th} Edition. John Wiley & Sons, Inc.

Nurhayati, I. C., Rusgiyono, A., & Yasin, H. 2018. Robust Geographically Weighted Regression dengan Metode Mutlak Simpangan Terkecil Pada Pemodelan Kejadian Diare di Kota Semarang. *Jurnal Gaussian*. **7**(2): 143–152.

Prahutama, A., & Rusgiyono, A. 2021. Robust regression with MM-estimator for modelling the number maternal mortality of pregnancy in Central Java, Indonesia. *Journal of Physics: Conference Series*. **1943**(1): 1–7.

Rahayu, L., Ulfa, E. M., Sasmita, N. R., Sofyan, H., Kruba, R., Mardalena, S., & Saputra, A. 2023. Unraveling Geospatial Determinants: Robust Geographically Weighted Regression Analysis of Maternal Mortality in Indonesia. *Infolitika Journal of Data Science*. **1**(2): 73–81.

Rousseeuw, P. J., & Yohai, V. 1984. Robust Regression by Means of S Estimators. In J. Franke, W. H"ardle, & R. D. Martin, *Robust and Nonlinear Time Series Analysis*. Springer-Verlag.

Schober, P., Mascha, E. J., & Vetter, T. R. 2021. Statistics From A (Agreement) to Z (z Score): A Guide to Interpreting Common Measures of Association, Agreement, Diagnostic Accuracy, Effect Size, Heterogeneity, and Reliability in Medical Research. *Anesthesia & Analgesia*. **133**(6): 1633–1641.

Setyowati, S. L., Aidi, M. N., Syafitri, U. D., & Ernawati, F. 2025. Evaluating Robust Estimators in Geographically Weighted Regression for Stunting Analysis at the District-Level Across Java: A Focus on Outlier Handling. *ZERO: Jurnal Sains, Matematika Dan Terapan*. **9**(1): 56–65.

Siahaan, M. P. A., & Husein, I. 2024. Robust Geographically Weighted Regression Modeling In Cases Of Stunting Toddlers In North Sumatera Utara. *Journal of Computer Science, Information Technology and Telecommunication Engineering*. **5**(2): 686–693.

Siswanto, S., Mirna, M., Yusran, M., Syam, U. A., & Miolo, A. S. I. 2022. Identification of Factors that Influence Stunting Cases in South Sulawesi using Geographically Weighted Regression Modeling. *Jurnal Matematika, Statistika Dan Komputasi*. **19**(1): 100–108.

Susanti, Y., Pratiwi, H., Sulistijowati H., S., & Liana, T. 2014. M Estimation, S Estimation, nd MM Estimation In Robust Regression. *International Journal of Pure and Apllied Mathematics*. **91**(3): 349–360.

Tizona, A. R., Goejantoro, R., & Wasono, W. (2017). Pemodelan Geographically Weighted Regression (Gwr) Dengan Fungsi Pembobot Adaptive Kernel Bisquare Untuk Angka Kesakitan Demam Berdarah di Kalimantan Timur Tahun 2015. *EKSPONENSIAL*. **8**(1): 87–94.

Tobler, W. R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*. **46**: 234–240.

Warsito, B., Yasin, H., Ispriyanti, D., & Hoyyi, A. 2018. Robust geographically weighted regression of modeling the Air Polluter Standard Index (APSI). *Journal of Physics: Conference Series*. **1025**: 1–6.

Wheeler, D. 2014. Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression. *Environment and Planning A*. **46**(10): 2590–2607.

Wulandari, P. P., Djuraidah, A., & Wigena, A. H. 2019. Robust Geographically Weighted Regression Modeling using Least Absolute Deviation and M-Estimator. *Repositories - Dept. of Statistics, IPB University*. 238–245.

Yimnak, K., & Piampholphan, K. 2021. A Comparison of MM-estimation and Fuzzy Robust Regression for Multiple Regression Model with Outliers. *Thailand Statistician*. **19**(2): 411–419.