VARIANSI: Journal of Statistics and Its Application on Teaching and Research ISSN 2684-7590 (Online)
Vol. 8 No. 1 (2026), 49-62

DOI: 10.35580/variansiunm503

Comparison of Geographically Weighted Regression (GWR) and Mixed Geographically Weighted Regression (MGWR) Models (Case Study: Crime in South Sulawesi)

Indi Nur Ridwan[1], Sudarmin[2], Zakiyah Mar'ah[3*]

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Indonesia

Keywords: Criminality, Geographically Weighted Regression (GWR), Mixed Geographically Weighted Regression (MGWR).

Abstract:

The Geographically Weighted Regression (GWR) model operates by taking into account how the relationships between different factors change across geographic space. Meanwhile, the Mixed Geographically Weighted Regression (MGWR) model permits certain variables to exhibit spatially varying (local) effects, while other variables are assumed to have constant effects across all locations. Both models are relevant to be applied in crime studies influenced by variations in regional conditions. The objective of this study is to evaluate the GWR and MGWR approaches in selecting the best model to explain factors associated with crime cases in South Sulawesi. The data used include the number of crime cases in South Sulawesi in 2024 along with factors presumed to influence them. The investigation's outcomes suggest the GWR model demonstrates higher appropriateness compared to the MGWR model, evidenced by its reduced Akaike Information Criterion (AIC) score and a 98.44% coefficient of determination (R^2). Based on the best-fitting model, population density and the number of poor residents were identified as the main factors influencing criminality in South Sulawesi in 2024.

1. Introduction

Regression analysis is a statistical method used to examine the relationship between a response (dependent) variable and predictor (independent) variables, and it is generally applied to data with a continuous response variable (Lutfiani et al., 2019). Regression analysis aims to explain the relationships among variables that are the focus of the study. In a geographical context, spatial effects commonly exist, meaning that a region can be influenced by its surrounding areas (Nidyashofa & Darsyah, 2020). Differences in environmental and geographical characteristics among observation areas give rise to spatial effects characterized by spatial heterogeneity. This heterogeneity causes the influence of predictor variables on the response variable to vary across locations, resulting in non-uniform regression parameter estimates among regions. Therefore, a local spatial modeling approach is required, with parameter estimation conducted at each observation location. One regression model that estimates parameters at every observation point is Geographically Weighted Regression (GWR) (Pratiwi et al., 2020).

The GWR model is an advanced regression model derived from global regression by incorporating spatial variation. In the GWR model, all predictor variables are assumed to have effects that vary locally (Mar'ah et al., 2017). However, in practice, not all variables exhibit significant spatial variation. Therefore, the Mixed Geographically Weighted Regression (MGWR) model was developed, allowing some model parameters to remain constant across all regions while others vary by location. The MGWR model combines local and global regression approaches within a single modeling framework, resulting in parameter estimates that better reflect the spatial characteristics of each variable (Utami et al., 2021).

The number of criminal cases in South Sulawesi Province has shown a downward trend over the past three years, decreasing from 26,684 cases in 2022 to 25,533 cases in 2023, and further declining to 23,359 cases in 2024 (BPS Provinsi Sulawesi Selatan, 2025). Nevertheless, the relatively high number of cases indicates that crime remains an issue requiring serious attention in order to maintain security and social stability in the region.

Various previous studies indicate that crime rates in Indonesia are influenced by socioeconomic factors such as poverty, unemployment, GRDP per capita, the Human Development Index, education, population density, and

* Corresponding author.
E-mail address: zakiyahm@unm.ac.id.

income inequality, and that they exhibit varying spatial patterns. ((Febrianti et al., 2023). In line with these findings, Nurhuda & Jaya, (2018) in their study, applied the GWR model and found that population density, economic growth, the Gini ratio, and poverty significantly influence crime in East Java. The MGWR model has been applied in research conducted by (Widayaka et al., 2016), indicating that the MGWR model is able to identify global and local parameters more flexibly than GWR and produces better model performance based on the Akaike Information Criterion (AIC).

To date, no studies have been found that examine crime cases using the MGWR model. Based on this, the researcher is interested in implementing the GWR and MGWR models to obtain the best model for analyzing the factors that influence crime rates in South Sulawesi.

2. Literature Review

2.1 Regression analysis

Regression analysis is used to predict the relationship between more than two predictor variables and a dependent variable. Parameter estimation in the model is carried out using the Ordinary Least Squares (OLS) method, which works by minimizing the sum of squared errors (Marizal & Mulyani, 2024). The general form of a linear regression equation using the OLS approach can be written as follows (Maulani et al., 2016).

$$Y_i = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ik} + \varepsilon_i \,, \qquad i = 1,2,\dots,n \tag{1}$$

Equation (1) in the multiple linear regression model can be rewritten in matrix notation as follows.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_3 \end{bmatrix} = \begin{bmatrix} 1 & X_1 & Y_1 & \cdots & X_{1p} \\ 1 & X_2 & X_2 & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2}$$

Where:

Y: an (n × 1) column vector of observations

X: an $n \times$ (p+1) matrix with the first column equal to 1 representing the intercept

$\boldsymbol{\beta}$: a (p+1) ×1 column vector of regression coefficients

$\boldsymbol{\varepsilon}$: an (n×1) column vector of errors

The simplified form of Equation (2) can be expressed as follows:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{3}$$

The estimation of the parameter β using the Ordinary Least Squares (OLS) approach can be expressed as follows:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{4}$$

2.2 Classical Assumption Tests

2.2.1 Normality Test

The normality test aims to determine whether the data are normally distributed, using the Shapiro–Wilk test (Monica et al., 2020):

$$T_3 = \frac{1}{D}\left[\sum_{i=1}^{n} a_i (X_{n-i+1} - X_i\right]^2, \text{dengan } D = \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{5}$$

Where:

T_3 : *Shapiro-Wilk* test statistic

a_i : *Shapiro-Wilk* test coefficients

X_{n-i+1} : the $(n - i + 1)$-th ordered observation

x_i : the *i*-th observation

$\bar{x}$: the sample mean

The decision is made by accepting H_0 if T_3 > *p-value* and rejecting H_0 if $T_3 \leq$ *p-value* at the 5% significance level.

2.2.2 Multicollinearity Test

Multicollinearity testing is used to identify the degree of association among independent variables. One commonly used indicator is the Variance Inflation Factor (VIF). If the VIF value is greater than 10, it can be concluded that multicollinearity exists among the independent variables (Bakri et al., 2024). In general, the VIF value can be expressed as follows:

$$VIF_k(u_i, v_i) = \frac{1}{1-R_j^2} \tag{6}$$

Where R_j^2 is the coefficient of determination obtained from the relationship between the independent variable X_j and the other independent variables.

2.2.3 Spatial heterogeneity

Spatial heterogeneity indicates differences in relationships across observation areas. The spatial heterogeneity test aims to determine whether each location has distinct characteristics. This condition results in non-uniform regression parameters across regions and can be tested using the Breusch–Pagan method (Daulay & Simamora, 2023).

The hypotheses used are:

$H_0: \sigma_1^2 = \sigma_2^2$ (no spatial heterogeneity is identified)

H_1: at least one $\sigma_1^2 \neq \sigma^2$ (spatial heterogeneity is identified)

Test statistic:

$$BP = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T \tag{7}$$

Where the elements of the factor vector f are defined as $f_i = \frac{e_i^2}{\sigma^2} - 1$, where $e_i = y_i - \hat{y}_i$ is obtained from the OLS method. The matrix Z is an $n \times (p + 1)$ matrix containing normalized and standardized vectors for each observation.

2.3 *Geographically Weighted Regression* (GWR) Model

In the GWR model, regression parameters are estimated locally at each observation point, so the regression coefficients used to estimate the response variable y vary across locations. The GWR model can be written as follows.(Caraka & Yasin, 2017).

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)\, X_{ik} + \varepsilon_i\,, \qquad i = 1,2,\dots,n \tag{8}$$

Keterangan:

y_i : the observed value of the response variable at the i-th location

X_{ik} : the observed value of the k-th predictor variable at the i-th observation location

$\beta_0(u_i, v_i)$: the constant (intercept) parameter at the i-th observation location

(u_i, v_i) : the geographical coordinates (longitude and latitude) of the i-th observation location.

$\beta_k(u_i, v_i)$: the regression coefficient of the k-th predictor variable at the i-th observation location.

ε_i : the error term at the i-th observation location, assumed to be identical, independent, and normally distributed with zero mean and constant variance σ^2

Parameter estimation in the GWR model is performed using the Weighted Least Squares (WLS) method, by assigning different weights to each observation location based on the spatial position of the data points. The weight at each location is denoted as $w_j(u_i, v_i)$ for j = 1, 2, ..., n locations. The GWR parameter estimate $\beta_k(u_i, v_i)$ for the k-th variable at the i-th observation location is given as follows:

$$\hat{\beta}(u_i, v_i) = [\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{X}]^{-1}\boldsymbol{X^T W}(u_i, v_i)\boldsymbol{Y} \tag{9}$$

Keterangan:

$\hat{\beta}(u_i, v_i)$: the estimated parameter vector at the i-th location

$\boldsymbol{X}$: the predictor variable matrix $(n \times (p + 1))$

$\boldsymbol{W}(u_i, v_i)$: the weight matrix at the i-th location (n × n)

$\boldsymbol{Y}$: the response variable vector (n × 1)

2.4 Spatial Weighting in Geographically Weighted Regression (GWR)

In the GWR model, the selection of the weighting function is a crucial factor. In this context, the weights are represented in the form of a diagonal matrix, where each element on the main diagonal is the result of the weighting function applied to each location point. The primary role of this weight matrix is to produce parameter estimates that vary at each location, thereby allowing differences in parameters across observation locations (Lutfiani et al., 2019). Kernel functions are used to determine the weight values at each location in the GWR model. In this study, the kernel function applied is the fixed bisquare kernel.

$$\textit{Fixed Bisquare} \quad = W_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2 & \text{, for } d_{ij} \le \text{h} \\ 0 & \text{, for } d_{ij} > \text{h} \end{cases} \tag{10}$$

Where:

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{11}$$

The value his a known non-negative parameter (also referred to as the bandwidth or smoothing parameter), while d_{ij} represents the Euclidean distance between locations i and j.

2.5 Bandwidth Selection

According to Caraka & Yasin, (2017) the Cross Validation (CV) method is one approach that can be used to determine the optimal bandwidth, which is systematically defined as follows.

$$CV = \sum_{i=1}^{n} [y_i - \hat{y}_{\neq 1}(h)]^2 \tag{12}$$

Where $\hat{y}_{\neq i}(h)$ $\hat{y}_{\neq 1}(h)$ represents the estimate of y_i obtained by excluding the observation at location (u_i, v_i). The optimal bandwidth h is the value of h that minimizes the CV value.

2.6 Partial Parameter Test of the GWR Model

The significance testing of parameters is conducted partially at each location to evaluate the effect of each parameter. The purpose is to assess whether $\beta(u_i, v_i)$ has a partial effect on the response variable in the GWR model. The hypotheses are formulated as follows (Simamora & Ratnasari, 2014).

$H_0: \beta(u_i, v_i) = 0$

$H_1: \beta(u_i, v_i) \neq 0; \qquad i = 1,2, \dots, n; \;\; k = 1,2, \dots, p$

The parameter $\hat{\beta}_k(u_i, v_i)$ is estimated under the assumption that it follows a multivariate normal distribution.

Test statistic:

$$T = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma}\sqrt{g_{kk}}} \tag{13}$$

Reject H_0 jika nilai $|T| \ge t_{\left(\frac{\alpha}{2};db\right)}$ indicating that the parameter $\beta_k(u_i, v_i)$ is significant in the model.

2.7 Spatial Variability Test

According to (Leung et al., 2000), the spatial variability test aims to distinguish between global and local coefficients in the GWR model. This analysis reveals which coefficients are influenced by location and which do not affect the dependent variable. The test is conducted to evaluate whether there are significant differences in the effects of the independent variable X_k across different locations.

The test is conducted using the following hypotheses:

H_0: All local coefficients are equal, that is $\beta_1(u_1, v_1) = \beta_2(u_2, v_2) = \cdots = \beta_j(u_n, v_n)$, for $j = 0,1, \dots, k$

H_1: At least one local coefficient $\beta_j(u_i, v_i)$, $(i = 1,2, \dots, n) \neq \beta_j(u_n, v_n)$

$$F = \frac{V_j^2 / tr\left(\frac{1}{k} B_k^T \left[I - \frac{1}{k} J\right] B_j\right)}{SSE(H_1)/\delta_1} \tag{14}$$

Where:

$$V_j^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_j(u_i,v_i) - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_j(u_i,v_i)\right)^2$$

$$= \frac{1}{n}\boldsymbol{\beta}_j^T\left[\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right]\boldsymbol{\beta}_j$$

$$SSE\ (H_1) = \boldsymbol{Y^T}(\boldsymbol{I}-\boldsymbol{L})^{\boldsymbol{T}} - (\boldsymbol{I}-\boldsymbol{L})\boldsymbol{Y}$$

$$\boldsymbol{\beta}_j(u_i,v_i) = \begin{bmatrix}\beta_0(u_i,v_i)\\ \beta_1(u_i,v_i)\\ \vdots \\ \beta_k(u_i,v_i)\end{bmatrix}, \boldsymbol{\beta}_j(u_i,v_i) = \begin{bmatrix}\boldsymbol{e}_1^T[\boldsymbol{W^T W}(u_i,v_i)\boldsymbol{X}]^{-1}\boldsymbol{X^T W}(u_i,v_i)\\ \boldsymbol{e}_2^T[\boldsymbol{W^T W}(u_i,v_i)\boldsymbol{X}]^{-1}\boldsymbol{X^T W}(u_i,v_i)\\ \vdots \\ \boldsymbol{e}_j^T\left[\boldsymbol{W^T W}(u_i,v_i)\boldsymbol{X}\right]^{-1}\boldsymbol{X^T W}(u_i,v_i)\end{bmatrix}$$

Dengan

J : an $(n \times n)$ matrix with all elements equal to 1

$\boldsymbol{e}_j$: a $(k+1)$-dimensional column vector, with the j-th element equal to 1 and the other elements equal to 0.

$\boldsymbol{e}_j^T$: the transpose of vector e_j, used to extract the j-th local regression coefficient.

H_0 is rejected if the *p-value* $\leq \alpha$ or if $F \geq F_{\alpha;df_1;df_2}$ where $df_1 = \left(\frac{\gamma_1^2}{\gamma_2}\right)$ with $\gamma_i = tr\left(\frac{1}{k}B_k^T\left[I - \frac{1}{k}J\right]B_j\right)^i$; $i = 1,2$ and $df_2 = \left(\frac{\delta_1^2}{\delta_2}\right)$ with $\delta_1 = tr([(I-L)^T(I-L)]^i)$.

2.8 *Mixed Geographically Weighted Regression* (MGWR) Model

The MGWR model is a modeling approach that combines global regression with geographically weighted regression (Hakim et al., 2014). This model is used when, in GWR, some independent variables do not have a significant local effect, making certain variables global. This is what defines the MGWR model. In MGWR, some coefficients from GWR are assumed to remain constant across all observation points, while other coefficients continue to vary according to the data location (Fotheringham et al., 2002) The MGWR model can be expressed as follows (Hakim et al., 2014).

$$y_i = \beta_0(u_i,v_i) + \sum_{k=1}^{q}\beta_k(u_i,v_i)\,x_{ik} + \sum_{k=q+1}^{p}\beta_k\,x_{ik} + \varepsilon_i\,, i = 1,2,\dots,n \tag{15}$$

Where:

y_i : the observed value of the response variable at the i-th location

X_{ik} : the observed value of the k-th predictor variable at the i-th observation location

$\beta_0(u_i,v_i)$: the constant (intercept) parameter at the i-th observation location

(u_i,v_i) : the geographical coordinates (longitude and latitude) of the i-th observation location.

$\beta_k(u_i,v_i)$: the regression coefficient of the k-th predictor variable at the i-th observation location.

β_k : the regression coefficient of the k-th predictor variable observation

ε_i : the error term at the i-th observation location, assumed to be identical, independent, and normally distributed with zero mean and constant variance σ^2

Parameter estimation in the MGWR model is performed using WLS, as applied in the GWR model. The initial step involves creating a weight matrix for each observation (Fotheringham et al., 2002) The parameter estimates for the global variables can be expressed as follows.

$$\widehat{\boldsymbol{\beta}}_g = \left[\boldsymbol{X}_g^T(\boldsymbol{I}-\boldsymbol{S}_l)^{\boldsymbol{T}}(\boldsymbol{I}-\boldsymbol{S}_l)\boldsymbol{X}_g\right]^{-1}\boldsymbol{X}_g^T(\boldsymbol{I}-\boldsymbol{S}_l)^{\boldsymbol{T}}(\boldsymbol{I}-\boldsymbol{S}_l)\boldsymbol{y} \tag{16}$$

For the local variables, the parameter estimates are given as follows.

$$\widehat{\boldsymbol{\beta}}_l(u_i,v_i) = [\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i)\boldsymbol{X}_l]^{-1}\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i)\left(\boldsymbol{y} - \boldsymbol{X}_g\widehat{\boldsymbol{\beta}}_g\right) \tag{17}$$

Where:

$$\boldsymbol{S_l} = \begin{pmatrix} \boldsymbol{x}_{l1}^T[\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i)\boldsymbol{X}_l]^{-\mathbf{1}}\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i) \\ \boldsymbol{x}_{l2}^T[\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i)\boldsymbol{X}_l]^{-\mathbf{1}}\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i) \\ \vdots \\ \boldsymbol{x}_{ln}^T[\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i)\boldsymbol{X}_l]^{-\mathbf{1}}\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i) \end{pmatrix}, \qquad \boldsymbol{S}_g = \boldsymbol{X}_g\left[\boldsymbol{X}_g^T\boldsymbol{X}_g\right]^{-1}\boldsymbol{X}_g^T$$

2.9 Partial Parameter Test of the MGWR Model

According to Purhadi & Yasin (2012), the purpose of this test is to identify variables that have a significant effect on the response in the MGWR model, both globally and locally. The testing is conducted in two stages: first for the global variables and then for the local variables.

For the significance of global variables, the hypotheses used are as follows:

H_0: $\beta_k = 0$ (the global variable X_k is not significant)

H_1: $\beta_k \neq 0$ (the global variable X_k is significant)

for $k = 1,2,\dots,p$

Test statistic:

$$T_{g\ hit} = \frac{\hat{\beta}_k}{\hat{\sigma}\sqrt{g_{kk}}} \tag{18}$$

Where:

g_{kk}: the k-th diagonal element of the matrix $\boldsymbol{GG}^T$

$$\boldsymbol{G} = \left[\boldsymbol{X}_g^T(\boldsymbol{I}-\boldsymbol{S}_l)^T(\boldsymbol{I}-\boldsymbol{S}_l)\boldsymbol{X}_g\right]^{-1}\boldsymbol{X}_g^T(\boldsymbol{I}-\boldsymbol{S}_l)^T(\boldsymbol{I}-\boldsymbol{S}_l)$$

$$\hat{\sigma}^2 = \frac{\boldsymbol{y}^T(\boldsymbol{I}-\boldsymbol{S})^T(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{y}}{tr(\boldsymbol{I}-\boldsymbol{S})^T(\boldsymbol{I}-\boldsymbol{S})}$$

H_0 is rejected if $\left|T_{g\ hit}\right| \geq t_{\frac{\alpha}{2};df}$ where $df = \left(\frac{u_1^2}{u_2}\right)$.

For the significance testing of local variables, the hypotheses are as follows:

H_0: $\beta_k(u_i,v_i) = 0$ (the local variable X_k at location i is not significant)

H_1: $\beta_k(u_i,v_i) \neq 0$ (the local variable X_k at location i is significant)

for $k = 1,2,\dots,q$ and $q = 1,2,\dots,n$

Test statistic:

$$T_{l\ hit} = \frac{\hat{\beta}_k(u_i,v_i)}{\hat{\sigma}\sqrt{m_{kk}}} \tag{19}$$

Where:

$\boldsymbol{m}_{kk}$: the k-th diagonal element of the matrix $\boldsymbol{MM}^T$

$$\boldsymbol{M} = \left[\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i)\boldsymbol{X}_l\right]^{-1}\boldsymbol{X}_l^T\boldsymbol{W}(u_i,v_i)\left(\boldsymbol{I}-\boldsymbol{X}_g\boldsymbol{G}\right)$$

H_0 is rejected if $|T_{l\ hit}| \geq t_{\frac{\alpha}{2};df}$ where $df = \left(\frac{u_1^2}{u_2}\right)$.

2.10 Model Selection

Model selection is the process of evaluating how well a model fits the data being analyzed. Methods commonly used to determine model goodness-of-fit include the Akaike Information Criterion (AIC) and the coefficient of determination (R^2). These methods are described as follows:

2.10.1. *Akaike Information Criterion* (AIC)

The AIC value is not only used to determine the optimal bandwidth but also to select the best model among several models. The AIC represents a model's likelihood rank that tends to minimize information loss (error). Therefore, a lower AIC value indicates a better model compared to others (Lutfiani et al., 2019). The AIC value is calculated using the following formula:

$$AIC = 2n\, In(\hat{\sigma}) + n\, In\,(2\pi) + n + tr(S) \quad (20)$$

2.10.2. Coefficient of Determination (R^2)

The coefficient of determination (R^2) is used to measure the extent to which the independent variables contribute to the dependent variable. It can be calculated using the following formula: (Bakri et al., 2024).

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \times 100\% \quad (21)$$

3. Research Methods

This study uses secondary data obtained from the South Sulawesi Provincial Statistics Agency for the year 2024, along with astronomical location data in the form of latitude and longitude for each regency/city. The study sample includes 24 regencies/cities in South Sulawesi Province. The data consist of one response variable and several predictor variables. The response variable (Y) is the number of crime cases per regency/city, while the predictor variables include population density (X_1), number of poor residents (X_2), average years of schooling (RLS) (X_3), labor force participation rate (TPAK) (X_4), open unemployment rate (TPT) (X_5), per capita expenditure (X_6), sex ratio (X_7), and life expectancy (UHH) (X_8).

The data analysis process in this study begins with data collection, followed by linear regression analysis, then GWR, and finally MGWR. Subsequently, the results of the GWR and MGWR modeling are compared to determine the best model based on the Akaike Information Criterion (AIC) and the coefficient of determination (R^2), and the study concludes with drawing conclusions.

4. Results and Discussion

4.1. Descriptive Analysis

Before proceeding to a more in-depth analysis, a descriptive analysis was conducted to provide an overall overview of each variable in every regency/city in South Sulawesi. The descriptive statistics of the variables used are presented in **Tabel 4.1**.

Table 4. 1 Descriptive Statistics of Research Data

Variable	Min.	Max.	Mean	Standar Deviation
Number of Crimes	117	7.671	973,2	1.493,639
Population Density	45	8.282	673,2	1.659,44
Poor Population	7,94	79,53	30,69	18,27133
Average Years of Schooling (RLS)	7,01	11,570	8,626	1,136642
Labor Force Participation Rate (TPAK)	59,32	78,39	68,22	5,15989
Open Unemployment Rate (TPT)	1,51	9,71	3,709	1,960417
Per Capita Expenditure	8.319	18.386	12.109	2.111,451
Sex Ratio	93,78	106,82	99	3,393081
Life Expectancy (UHH)	67,16	74,23	70,35	1,92408

In addition to the table, the descriptive analysis is also visualized in a thematic map to show the data distribution. The map displays the number of crime cases in the 24 regencies/cities of South Sulawesi Province in 2024, as shown in the following figure.

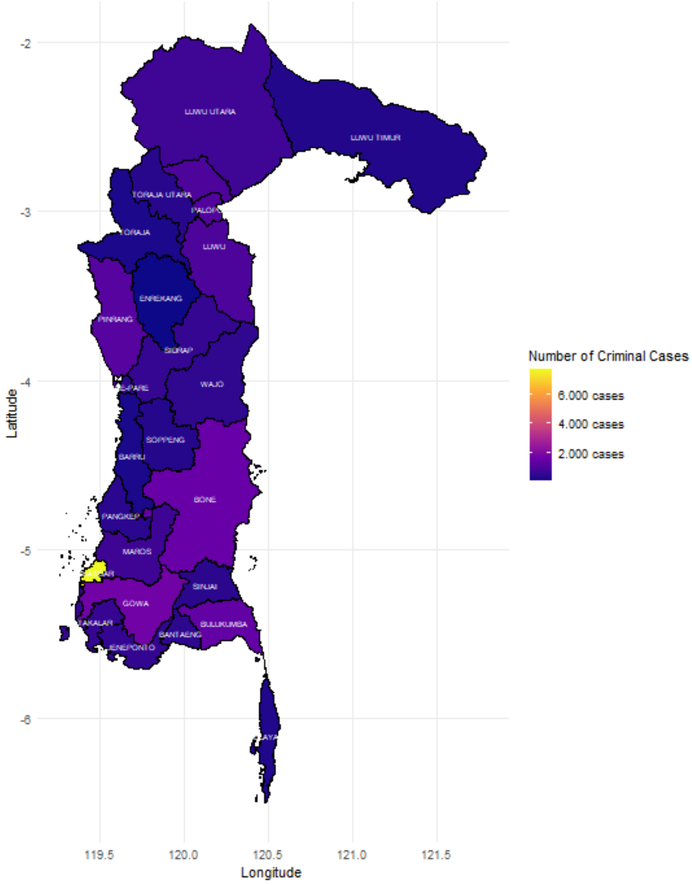

Figure 4.1 Distribution Map of Crime Cases

4.2. Assumption Testing

Assumption testing was conducted on the residuals of the fitted regression model. The assumptions tested are as follows.

4.2.1. Normality Test

The normality test was conducted using the Shapiro–Wilk test. The test results yielded a W value of 0.96322 with a p-value of 0.506. This indicates that the residuals are normally distributed, as the p-value (0.506) is greater than $\alpha = 0.05$.

4.2.2. Multicollinearity Test

The multicollinearity assumption in this study was tested using the Variance Inflation Factor (VIF). The VIF values for each predictor variable are presented in **Table 4.2**.

Table 4. 2 VIF Values of Predictor Variables

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
VIF	4,571	2,044	4,166	1,799	4,255	2,822	2,424	3,077

Table 4.2 shows that all predictor variables have VIF values less than 10, indicating that multicollinearity does not occur.

4.2.3. Spatial Heterogeneity Test

The spatial heterogeneity assumption in this study was tested using the Breusch–Pagan test. The test results yielded a BP statistic of 22.583 with a p-value of 0.003943. This indicates that spatial heterogeneity exists in the data, as the p-value (0.003943) is less than $\alpha = 0.05$.

4.3. *Geographically Weighted Regression* (GWR) Model

The GWR model utilizes geographical information for each observation area. In R-Studio, building this model requires the longitude and latitude coordinates. The weight for each observation is calculated using the Fixed Bisquare Kernel based on the optimal bandwidth, which is selected through Cross Validation (CV). The analysis results show an optimal bandwidth of 1.83122 with a CV value of 9.886226, which is used to determine the weights for each region in South Sulawesi.

Once the weights are obtained, parameter estimates for each region can be calculated. **Table 4.3** presents the GWR model parameter estimates for Kepulauan Selayar Regency.

Table 4. 3 GWR Model Parameter Estimation Results

		Coefficient	Standard Error	t_{value}	*p-value*
Kepulauan Selayar Regency	β_0	0,008	0,063	0,131	0,897
	β_1	0,604	0,135	4,461	0,000
	β_2	0,383	0,089	4,284	0,000
	β_3	0,033	0,128	0,260	0,797
	β_4	-0,046	0,086	-0,537	0,597
	β_5	-0,074	0,138	-0,534	0,599
	β_6	0,076	0,105	0,725	0,477
	β_7	-0,121	0,109	-1,115	0,279
	β_8	0,174	0,123	1,417	0,172

The following is the GWR model obtained for Kepulauan Selayar Regency.

$$\hat{Y}_{Selayar} = 0{,}008 + 0{,}604X_1{}^* + 0{,}383X_2{}^* + 0{,}033X_3 - 0{,}046X_4 - 0{,}074X_5 + 0{,}076X_6 - 0{,}121X_7 + 0{,}174X_8$$

Based on the model above, population density (X_1) and the poor population (X_2), marked with an asterisk (*), are the variables that have a significant effect. Population density (X_1) has an effect of 0.604, meaning that an increase of one unit in population density will increase crime by 0.604. The poor population (X_2) has an effect of 0.383, meaning that an increase of one unit in the poor population will increase crime by 0.383. Meanwhile, variables X_3through X_8do not have an asterisk, indicating they are not significant. This means their influence on the model is weak or there is insufficient evidence that these variables have an effect.

4.4. Partial Parameter Test of the GWR Model

In the GWR model, the partial test is conducted to evaluate the magnitude of the effect of each independent variable on the dependent variable at each observation location. A variable is considered significant if the p-value < 0.05, and not significant if the p-value ≥ 0.05.

Table 4.4 presents the group of variables that are significant in the GWR model.

Table 4. 4 GWR Group of Variables Significant in the GWR Model

Significant Variables	Regency/City
X_1 and X_2	Selayar, Bulukumba, Jeneponto, Pangkep, Soppeng, Sidrap, Pinrang, Luwu, Luwu Utara, Toraja Utara, Makassar, and Parepare.
X_1	Maros
X_2	Bantaeng, Takalar, Gowa, Barru, Bone, Wajo, Enrekang, Tana Toraja, Luwu Timur, and Palopo.

Sinjai Regency does not have any significant variables.

The following map shows the spatial variation of the combination of significant variables affecting crime in each regency/city based on the GWR model.

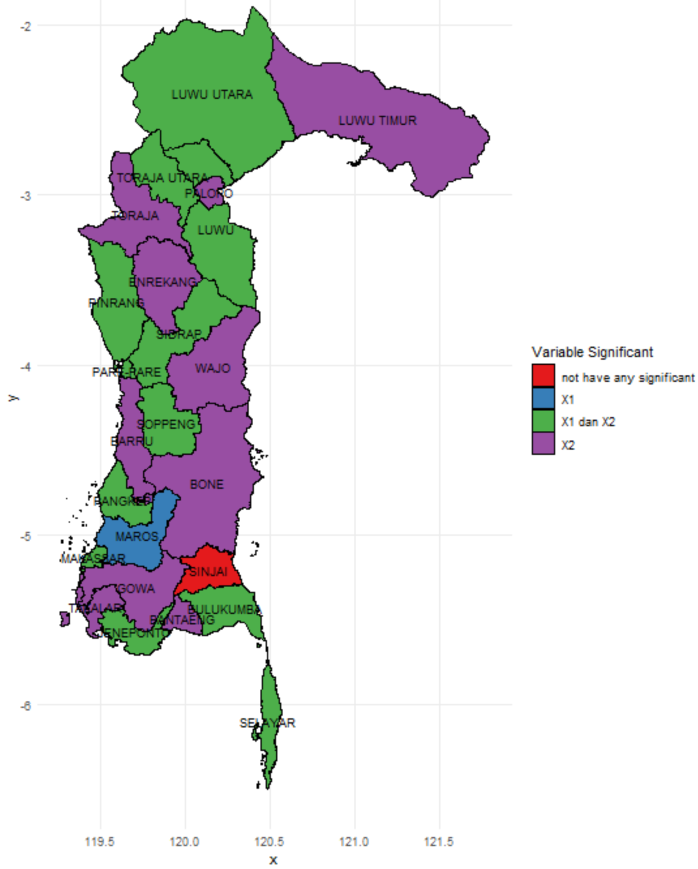

Figure 4.2 Map of Significant Variable Groups in the GWR Model

4.5. Spatial Variability Test

In the significance testing of GWR model parameters, it was found that not all variables are significant at every observation location. This suggests that some variables may not be influenced by location. If a parameter does not show varying effects across locations, it is assumed to be a global coefficient, meaning it has the same estimated value for all observation points. The spatial variability test is conducted to identify parameters that are not affected by location and to determine which coefficients will be treated as global or local in the MGWR modeling. The results of the test, calculated using R-Studio, are presented in Table 4.5. The null hypothesis is rejected if the *p-value* $\leq \alpha =$ 10% or $F \geq F_{\alpha;df_1;df_2}$

Table 4. 5 Spatial Variability Test

Variable	F_{value}	F_{table}	*p-value*	Desicion
X_1	1,156	2,618	0,426	Fail to reject H_0
X_2	2,886	2,823	0,095	Reject H_0
X_3	1,005	2,754	0,477	Fail to reject H_0
X_4	3,652	2,604	0,043	Reject H_0
X_5	3,183	2,852	0,079	Reject H_0
X_6	2,235	2,904	0,161	Fail to reject H_0
X_7	0,552	2,555	0,826	Fail to reject H_0
X_8	1,754	2,580	0,223	Fail to reject H_0

Table 4.5 shows that the variables Poor Population (X_2), Labor Force Participation Rate (TPAK) (X_4), and Open Unemployment Rate (TPT) (X_5) have a significant effect across locations, meaning their coefficients vary spatially. Meanwhile, Population Density (X_1), Average Years of Schooling (RLS) (X_3), Per Capita Expenditure (X_6), Sex Ratio (X_7), and Life Expectancy (UHH) (X_8) do not show significant differences across locations. Based on these results, the independent variables can be grouped into local and global variables. The local variables include Poor Population (X_2), TPAK (X_4), and TPT (X_5), while the global variables consist of Population Density (X_1), RLS (X_3), Per Capita Expenditure (X_6), Sex Ratio (X_7), and UHH (X_8).

4.6. *Mixed Geographically Weighted Regression* (MGWR) Model

Based on the results of the spatial variability test, the predictor variables that have a local effect are X_2, X_4, and X_5, while the predictor variables with a global effect include X_1, X_3, X_6, X_7, and X_8. These results form the basis for constructing the MGWR model for crime cases in South Sulawesi using the Fixed Bisquare Kernel weighting function.

Using the MGWR model built with the Fixed Bisquare Kernel weights, an analysis was then conducted to identify the influence of each predictor variable on the crime rate in South Sulawesi. This analysis includes the evaluation of coefficient values, standard errors, t-values, and the significance of each variable, both local and global. **Table 4.6** presents the MGWR model parameter estimation results for Makassar City.

Table 4. 6 MGWR Model Parameter Estimation Results

		Coefficient	**Standard Error**	t_{hitung}	***p-value***
Makassar City	β_0	-0,170	0,062	-2,730	0,012
	β_1	0,616	0,056	10,999	$1{,}40 \times 10^{-8}$
	β_2	0,610	0,049	12,278	$9{,}05 \times 10^{-11}$
	β_3	0,110	0,064	1,709	0,107
	β_4	-0,237	0,035	-6,618	$1{,}91 \times 10^{-6}$
	β_5	-0,316	0,048	-6,532	$2{,}29 \times 10^{-6}$
	β_6	0,094	0,059	1,592	0,132
	β_7	-0,159	0,052	-3,034	0,008
	β_8	0,238	0,061	3,898	0,001

The following is the MGWR model obtained for Makassar City.

$$\hat{Y}_{Makassar} = -0{,}170 + 0{,}616X_1{}^* + 0{,}610X_2{}^* + 0{,}110X_3 - 0{,}237X_4{}^* - 0{,}316X_5{}^* + 0{,}094X_6$$
$$-0{,}159X_7{}^* + 0{,}238X_8{}^*$$

Based on the MGWR model obtained for Makassar City, variables X_1, X_2, and X_8have a positive effect on the dependent variable, as indicated by the asterisk (*). This means that if any of these variables increase, the number of crime cases will increase according to the magnitude of their coefficients, assuming other variables remain constant. Conversely, variables X_4, X_5, and X_7have negative coefficients, indicating that an increase in these variables would actually reduce crime cases. Thus, the MGWR model for Makassar City suggests that factors X_1, X_2, and X_8contribute to increasing crime, while factors X_4, X_5, and X_7help decrease crime cases. Meanwhile, variables $X_3$and X_6do not have an asterisk (*), meaning they are not statistically significant in influencing crime cases. Although their coefficients are positive, the effects of $X_3$and $X_6$are not strong enough to explain variations in crime in the MGWR model for Makassar City.

4.7. Partial Parameter Test of the MGWR Model

The partial test in the MGWR model is conducted using the p-value of each parameter. Both global and local parameters are considered to have a significant effect on crime cases if their p-value is below the predetermined significance level ($\alpha = 0.05$). **Table 4.7** presents the group of variables that are proven to be significant in the MGWR model.

Table 4. 7 Group of Variables Significant in the MGWR Model

Significant Variables	**Regency/City**
X_1, X_2, X_4, X_5, X_7, and X_8	Jeneponto, Sinjai, Maros, Pangkep, Soppeng, Sidrap, Pinrang, Luwu, Toraja Utara, Makassar, and Pare-pare.
X_1, X_2, X_5, X_7, and X_8	Selayar, Bulukumba, Barru, Bone, Wajo, Tana Toraja, and Luwu Utara.
X_1, X_2, X_7, and X_8	Bantaeng, Takalar, Gowa, Enrekang, Luwu Timur, and Palopo.

The following map shows the spatial variation of the combination of significant variables affecting crime in each regency/city based on the MGWR model.

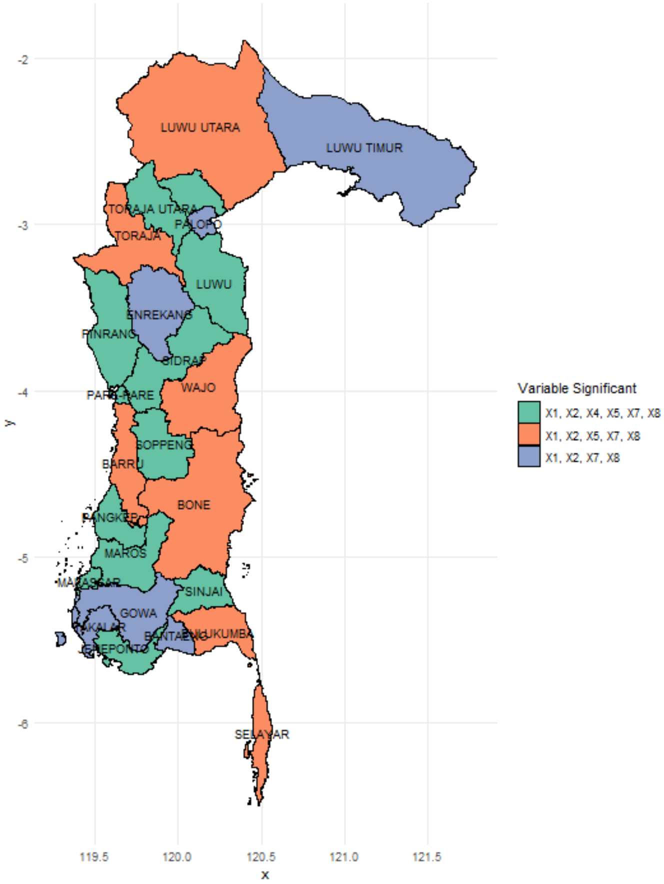

Figure 4.3 Map of Significant Variable Groups in the MGWR Model

4.8. Partial Parameter Test of the MGWR Model

To determine the best model, the R^2 and AIC values of each model are compared. A higher R^2 and a lower AIC indicate better model performance, as shown in **Table 4.8.**

Table 4. 8 R^2 and AIC Values of the Models

Model	**R^2**	**AIC**
Geographically Weighted Regression (GWR)	98,44%	-15,6256
Mixed Geographically Weighted Regression (MGWR)	97,89%	61,6401

Table 4. 9, It can be seen that the GWR model is better suited for modeling crime cases, as it has the highest $R^2$and the lowest AIC value.

5. Conclusion

In 2024, the average number of crime cases in South Sulawesi reached 1,064 cases per region, with Makassar City recording the highest number and Enrekang Regency the lowest. To analyze the factors influencing the distribution of crime, this study applied the Geographically Weighted Regression (GWR) model. The resulting GWR model can be expressed as follows.

$$\hat{Y}_{Selayar} = 0{,}008 + 0{,}604X_1{}^* + 0{,}383X_2{}^* + 0{,}033X_3 - 0{,}046X_4 - 0{,}074X_5 + 0{,}076X_6 - 0{,}121X_7 + 0{,}174X_8$$

$$\hat{Y}_{Bulukumba} = -0{,}028 + 0{,}368X_1{}^* + 0{,}371X_2{}^* + 0{,}089X_3 + 0{,}006X_4 - 0{,}026X_5 + 0{,}067X_6 - 0{,}153X_7 + 0{,}122X_8$$

$$\vdots$$

$$\hat{Y}_{Palopo} = -0{,}155 + 0{,}010X_1 + 0{,}331X_2{}^* + 0{,}189X_3 + 0{,}248X_4 + 0{,}199X_5 - 0{,}007X_6 - 0{,}154X_7 - 0{,}181X_8$$

Meanwhile, the modeling of crime cases for each regency/city in South Sulawesi using the Mixed Geographically Weighted Regression (MGWR) model is as follows.

$$\hat{Y}_{Selayar} = 0{,}009 + 0{,}616X_1^* + 0{,}406X_2^* + 0{,}110X_3 - 0{,}066X_4 - 0{,}181X_5^* + 0{,}094X_6 - 0{,}159X_7^* + 0{,}238X_8^*$$

$$\hat{Y}_{Bulukumba} = 0{,}021 + 0{,}616X_1^* + 0{,}393X_2^* + 0{,}110X_3 - 0{,}066X_4 - 0{,}174X_5^* + 0{,}094X_6 - 0{,}159X_7^* + 0{,}238X_8^*$$

$$\vdots$$

$$\hat{Y}_{Palopo} = 0{,}052 + 0{,}616X_1^* + 0{,}371X_2^* + 0{,}110X_3 - 0{,}068X_4 - 0{,}076X_5 + 0{,}094X_6 - 0{,}159X_7^* + 0{,}238X_8^*$$

Based on the analysis results, the comparison of (R^2) and AIC values between the GWR and MGWR models shows that the GWR model performs better. The higher (R^2) and lower AIC of the GWR indicate its superior ability to explain variations in crime cases in each regency/city of South Sulawesi Province in 2024 more accurately. Therefore, the GWR model is considered more appropriate for understanding the spatial distribution patterns of crime in South Sulawesi Province in 2024 compared to the MGWR model.

References

Bakri, N. A., Annas, S., & Aidid, M. K. (2024). Pendekatan Geographically Weighted Regression (GWR) untuk Menganalisis Hubungan PDRB Sektor Pertanian, Kehutanan, dan Perikanan dengan Faktor Pencemaran Lingkungan di Jawa Timur. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *6*(1), 11–20. https://doi.org/10.35580/variansiunm194

BPS Provinsi Sulawesi Selatan. (2025). *Provinsi Sulawesi Selatan dalam Angka 2025* (Vol. 12). Badan Pusat Statistik Provinsi Selatan Selatan.

Caraka, R. E., & Yasin, H. (2017). *Geographically Weighted Regression (GWR): Sebuah Pendekatan Regresi Geografis* (H. Caraka, Rezzy Eko & Yasin, Ed.; pertama). Mobius.

Daulay, S. H., & Simamora, E. (2023). Pemodelan Faktor-Faktor Penyebab Kemiskinan di Provinsi Sumatera Utara Menggunakan Metode Geographically Weighted Regression (GWR). *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam (JURRIMIPA)*, *2*(1), 47–60.

Febrianti, E., Susetyo, B., & Silvianti, P. (2023). Pemodelan Tingkat Kriminalitas di Indonesia Menggunakan Analisis Geographically Weighted Panel Regression. *Journal of Statictics*, *12*(1), 91–109.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression the analysis of spatially varying relationships*.

Hakim, A. R., Yasin, H., & Suparti. (2014). Pemodelan Persentase Penduduk Miskin di Kabupaten dan Kota Di Jawa Tengah dengan Pendekatan Mixed Geographically Weighted Regression. *Jurnal Gaussian*, *3*(4), 575–584. http://www.ejournal-s1.undip.ac.id/index.php/gaussian/article/view/8068%5Cnhttp://www.ejournal-s1.undip.ac.id/index.php/gaussian/article/download/8068/7848

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*, *32*(1), 9–32. https://doi.org/10.1068/a3162

Lutfiani, N., Sugiman, & Scolastika, M. (2019). Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-square. *UNNES Journal of Mathematics*, *8*(1), 82–91. http://journal.unnes.ac.id/sju/index.php/ujmUJM8

Mar'ah, Z., Djuraidah, A., & Hamim Wigena, A. (2017). *Semi-parametric Geographically Weighted Regression Modelling using Linear Model of Coregionalization*. 178–186.

Marizal, M., & Mulyani, S. (2024). *Analisis Jumlah Kriminal Di Indonesia dengan Geographically Weighted Regression*. 10–19.

Maulani, A., Herrhyanto, N., & Suherman, M. (2016). *Aplikasi Model Geographically Weighted Regression (GWR) Untuk Menentukan Faktor-Faktor Yang Anak Balita di Jawa Barat*. *4*(1), 46–63.

Monica, S., Susanta, A., & Yensy, N. A. (2020). Pengaruh Pendekatan Realistic Mathematics Education Terhadap Kemampuan Pemecahan Masalah Siswa. *Jurnal Penelitian Pembelajaran Matematika Sekolah*, *4*(2), 220–228.

Nidyashofa, N., & Darsyah, M. Y. (2020). Pemilihan Model Regresi Spasial Pada Tingkat Pengangguran Terbuka di Provinsi Jawa Tengah. *Jurnal Statistika*, *8*(1), 88–96.

Nurhuda, I., & Jaya, I. G. N. M. (2018). Pemodelan Kriminal di Jawa Timur dengan Metode Gographically Weighted Regression (GWR). *Jurnal Matematika "MANTIK,"* *4*(2), 150–158.

Pratiwi, N., Suyitno, S., & Siringoringo, M. (2020). Penerapan Model Geographically Weighted Logistic Regression Pada Data Status Kesejahteraan Masyarakat di Kalimantan Tahun 2017. *Jurnal Eksponensial*, *11*(1), 83–92. https://doi.org/10.30872/eksponensial.v11i1.648

Purhadi, & Yasin, H. (2012). Mixed geographically weighted regression model (case study: The percentage of poor households in Mojokerto 2008. *European Journal of Scientific Research*, *69*(2), 188–196.

Simamora, P. A., & Ratnasari, V. (2014). Pemodelan Persentase Kriminalitas Dan Faktor- Faktor Yang Mempengaruhi Di Jawa Timur Dengan Pendekatan Geographically Weighted Regression (GWR). *Jurnal Sains Dan Seni ITS*, *3*(1), 18–23.

Utami, P. F., Rusgiyono, A., & Ispriyanti, D. (2021). Pemodelan Semiparametric Geographically Weighted Regression Pada Kasus Pneumonia Balita Provinsi Jawa Tengah. *Jurnal Gaussian*, *10*(2), 250–258. https://doi.org/10.14710/j.gauss.v10i2.30945

Widayaka, P. G., Mustafid, & Rahmawati, R. (2016). Pendekatan Mixed Geographically Weighted Regression Untuk Pemodelan Pertumbuhan Ekonomi Menurut Kabupaten/Kota di Jawa Tengah. *Jurnal Gausian*, *5*(4), 727–736.