



Spatial Modeling of Highly Dispersed Poverty Data in East Java Using the Geographically Weighted Negative Binomial Regression Method

Nova Putri Dwi Sugiarti^{1*}, Trimono², Shindi Shella May Wara³,

^{1,2,3}Program Studi Sains Data, Fakultas Ilmu Komputer,
Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia

E-Mail: ¹22083010088@student.upnjatim.ac.id,

²trimono.stat@upnjatim.ac.id, ³shindi.shella.fasilkom@upnjatim.ac.id

Received Mar 12th 2026; Revised Apr 04th 2026; Accepted Apr 24th 2026; Available Online Apr 28th 2026

Corresponding Author: Nova Putri Dwi Sugiarti

Copyright ©2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Poverty in East Java Province is a complex issue characterized by diverse regional characteristics. This study aims to model the number of poor inhabitants in East Java while accounting for overdispersion and spatial heterogeneity using the Geographically Weighted Negative Binomial Regression (GWNBR) method with four kernel weighting functions. The results indicate that the Negative Binomial regression model performs well at the global level, with an AIC of 941.01. However, to capture local variation in poverty drivers, the GWNBR model with an Adaptive Bisquare kernel proved optimal, yielding the lowest AIC of 309.32. These findings confirm the diversity of predictor variable influences across regencies and cities, as evidenced by the significance of variables in Blitar Regency. The GWNBR approach provides more accurate local parameter estimates than the global model, serving as a strategic tool for the government to design more targeted and effective poverty alleviation policies for each region.

Keywords: GWNBR, Overdispersion, Poverty Level, Spatial Effects

1. INTRODUCTION

Poverty is a multifaceted condition and is a crucial indicator in assessing a country's success in development. This situation is not only related to the financial inability to meet basic needs but is also influenced by a complex interaction between various social, economic, and geographic factors. According to data from the Badan Pusat Statistik (BPS), the national poverty rate in March 2024 was recorded at 9.03%, indicating a downward trend, but still requiring greater efforts to achieve medium-term development targets. The complexity of this problem is exacerbated by the cycle of poverty, where low quality human resources, limited job opportunities, and a lack of basic infrastructure mutually exacerbate the public's welfare [1].

East Java Province is a key region contributing significantly to the economy, but it also has the highest number of poor people in Indonesia. Although East Java recorded a single-digit poverty rate decline of 9.79% in 2024, this figure still reflects 3.98 million people living below the poverty line. Historically, East Java has experienced variations influenced by economic shocks, both global and domestic [2]. The poverty patterns observed in this province show distinct regional clustering, with areas like Surabaya City boasting a very low poverty rate of around 3.96%, while areas on Madura Island, such as Sampang Regency, still recorded a rate of 21.76%, followed by Bangkalan at 18.66% and Sumenep at 17.08%. Furthermore, several areas in the “Horseshoe” zone, such as Probolinggo, still recorded a rate of 16.31%. This suggests that poverty does not occur randomly, but is concentrated in areas with specific social and economic characteristics [3].

The dimensions of poverty in East Java have significant social impacts, going beyond mere numerical data. Disparities in access to basic needs remain evident in the quality of housing, the availability of sanitation and clean water, and the varying levels of public health across regions. Furthermore, changes in the employment sector following the pandemic and fluctuations in regional economic growth have also impacted people's vulnerability to poverty, particularly for informal workers and low-income families, whose economic resilience is more vulnerable to shocks [4]. Low levels of education in some regions also limit people's access to better and more decent jobs, thus reinforcing the cycle of structural poverty. This suggests that policies to address poverty must be developed with a differentiated approach, as each district or city has

unique root causes, from dependence on traditional agriculture in remote areas to structural unemployment and unequal job opportunities in industrial areas. Therefore, more specific strategies are needed that take into account the characteristics of each region [5].

Various previous studies have analyzed the causes of poverty using global regression models, such as multiple linear regression, to identify the factors contributing to poverty in Indonesia. For example, in the study by Gultom et al. [6] using a global approach to determine the relationship between poverty levels and explanatory variables such as unemployment, education, or health indicators, assumed to be constant across Java, demonstrates that the global approach is less able to explain variations in the influence of poverty factors during the 2020 Covid-19 pandemic. The pandemic situation has resulted in uneven impacts across regions, for example, areas with a high dependence on the service and industrial sectors were affected differently than agricultural-based areas. However, the general regression approach assumes the same relationship across regions, making it less effective in capturing spatial diversity that can lead to differences in the influence of variables in each district/city. To address this limitation, spatial-based methods such as Geographically Weighted Regression (GWR) and its development, Geographically Weighted Negative Binomial Regression (GWNBR), have begun to be used in poverty analysis. However, several studies indicate that the use of GWNBR is not fully optimal, either due to the limited number of regional samples, inappropriate bandwidth selection, or local estimation results that are not always significant in all regions. Therefore, further development and testing are still needed for this model to provide more accurate and useful results.

The variables selected in this study were determined based on their importance, both theoretically and practically, in understanding aspects of poverty in East Java. The employment sector, represented by the Open Unemployment Rate (TPT) and the Labor Force Participation Rate (TPAK), indicates potential family income, while the Population Density Rate (PDR) reflects the demographic burden in an area. Meanwhile, the Average Years of Schooling (RLS) and Life Expectancy (UHH) represent the quality of human resources. Furthermore, the Economic Growth Rate (EGR) is added to evaluate the extent to which the distributional effects of economic growth are felt at the district or city level. By combining these variables, it is hoped that a comprehensive coverage of economic, social, and demographic aspects can be achieved [7].

Based on the gaps identified in this research, the application of the GWNBR method is crucial because it addresses two key issues simultaneously. These issues include the enumerated nature of poverty data, its tendency toward overdispersion, and the spatial differences between regions. GWNBR leverages the Negative Binomial distribution to provide more consistent estimates compared to the Poisson model, while also generating local parameters that capture the differences in the influence of each variable across districts or cities. Furthermore, to address the limitations of previous research, which generally employed only one type of spatial weighting, this study compared several kernel functions, such as Gaussian, Bisquare, Tricube, and Exponential, to identify the model with the best performance based on specific criteria, such as the AIC value. This comparison is crucial because the choice of kernel and bandwidth significantly impacts the model's sensitivity to geographic proximity patterns. This approach is expected to enhance robust and comprehensive analysis, providing more precise policy recommendations for poverty alleviation that are tailored to the region's spatial characteristics.

This study introduces a novel approach by integrating the GWNBR model with a systematic comparison of several adaptive kernel weighting functions to address overdispersion and spatial heterogeneity in poverty data. Unlike previous studies that tend to rely on a single spatial weighting scheme or global modeling approach, this study evaluates four different kernel functions (Gaussian, Bisquare, Tricube, and Exponential) to determine optimal model performance. Furthermore, the use of the latest district/city-level data in East Java allows for more granular and location-specific analysis, highlighting spatial disparities often overlooked in global models. This combined approach provides a more comprehensive and robust framework for modeling poverty, offering deeper insights into local determinants and strengthening the methodological contribution of GWNBR to spatial data analysis.

2. MATERIALS AND METHOD

To address the research questions, a series of orderly and planned steps are required to ensure a clear and precise analysis. Each step is implemented sequentially to ensure consistent and accurate results. Details of the sequence of these research steps are shown schematically in Figure 1.

In addition to a structured workflow, selecting the appropriate analysis method is a crucial step in this research. The use of the Geographically Weighted Negative Binomial Regression (GWNBR) model is based on the characteristics of the data and the research objectives. The response variable, the number of poor people, is classified as count data, which is more appropriately analyzed using discrete statistical models than classical linear regression.

Furthermore, preliminary analysis indicates the presence of overdispersion, where the variance exceeds the mean, violating the assumptions of the Poisson regression model. Therefore, the Negative Binomial model is employed as a more flexible alternative to handle this condition. In addition, the poverty

data exhibit spatial heterogeneity, reflecting variations in socio-economic conditions across districts/cities in East Java. This indicates that the relationship between predictor variables and the response variable is not constant across locations. Based on these conditions, a local modeling approach is required to capture spatial variations more accurately. The GWNBR model extends the Negative Binomial regression by incorporating geographic weighting, allowing parameter estimates to vary across locations. Therefore, this method is considered the most appropriate approach as it simultaneously addresses the issues of count data, overdispersion, and spatial heterogeneity.

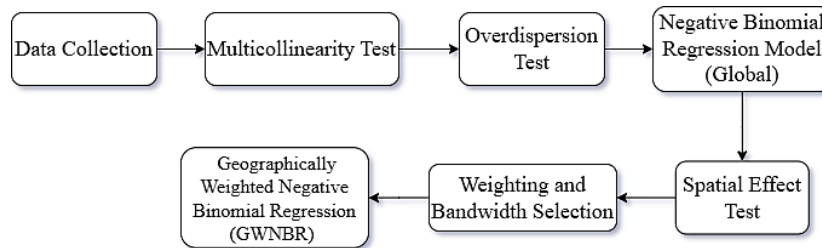


Figure 1. Research Methodology

In summary, the technical procedures of this research are summarized in the following algorithm:

Input: The response variable is the number of poor people, while there are six predictor variables related to employment, demographics, and the economy. The geographic coordinates of 38 districts/cities in East Java are also included.

Procedure:

1. Multicollinearity test: by calculating the value $VIF = \frac{1}{1-R_k^2}$. If the VIF value is < 10 , then the variable is retained.
2. Overdispersion test: tests whether the value of $\text{Var}(Y) > E(Y)$ to determine the use of the negative binomial model.
3. Spatial weighting: calculating the weighting matrix $W(u_i, v_i)$ using adaptive kernel functions from Gaussian, Bisquare, Tricube, and Exponential.
4. Bandwidth determination: finding the optimum bandwidth value (h) by minimizing the Cross-Validation (CV) value.
5. Local parameter estimation: calculating values $\hat{\beta}_k(u_i, v_i)$ through an iterative process for each location i .
6. Regional classification: by mapping the distribution of poor people based on significant variables in each district/city.

Output: estimated values of local significance parameters for each district/city and mapping of poverty distribution

2.1 Data Collection

This study uses secondary data from 38 districts/cities in East Java in 2024 sourced from the Central Statistics Agency (BPS). The response variable (Y) is the number of poor people (count data), with six predictor variables (X) covering aspects of employment (TPT, TPAK), demographics (Population Density), human resource quality (RLS, UHH), and economic growth. This dataset integrates the geographic coordinates of each region to compile a spatial weighting matrix in the GWNBR model. All data have gone through a verification and format standardization stage to ensure consistency of local parameter estimates in each observation area.

2.2 Multicollinearity Test

Multicollinearity is a condition of strong correlation between independent variables that can lead to unstable coefficient estimates. In the GWNBR model, identifying multicollinearity is crucial because socioeconomic variables tend to be spatially interrelated. R_k^2 is the coefficient of determination obtained from auxiliary regression, namely by regressing each k th predictor variable with other predictor variables. If the VIF value is known to be ≥ 10 , then the predictor variable has experienced multicollinearity [8].

2.3 Overdispersion Test

Overdispersion occurs when the variation in the data exceeds that predicted by a statistical model, particularly one that assumes the mean equals the variance, such as Poisson regression. If the actual variance is higher than predicted, the model is no longer accurate. If overdispersion is not addressed, the analysis can

be misleading because the standard error will be smaller, the p-value will appear lower, and the relationship between variables will appear significant when it is not [9].

2.4 Negative Binomial Regression Model (Global)

Negative Binomial regression is used to model overdispersed count data, a condition where the variance of the data exceeds its mean. This distribution can be considered an extension of the Poisson distribution with the addition of a dispersion parameter, making it more capable of handling high data variation. In Negative Binomial regression, the response variable is assumed to follow a distribution that is a combination of Poisson and Gamma. This model is used to find the relationship between independent and dependent variables that follow the Negative Binomial distribution [10]. The probability function for the negative binomial can be written as Equation 1.

$$f(y; \mu, \theta) = \frac{\Gamma(y + \frac{1}{\theta})}{y! \Gamma(\frac{1}{\theta})} \left(\frac{1}{1 + \theta\mu}\right)^{\frac{1}{\theta}} \left(\frac{\theta\mu}{1 + \theta\mu}\right)^y; y = 0, 1, 2, \dots \quad (1)$$

This regression model is an alternative to Poisson regression when the assumptions of equality of means and variances are not met. In general, the form of Negative Binomial regression can be expressed as Equation 2.

$$y_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}) \quad (2)$$

Where is $\beta_0, \beta_1, \dots, \beta_p$ as parameters and ε_i as the error for the i-th observation.

Parameter estimation in negative binomial regression is performed using the maximum likelihood method with the Newton-Raphson procedure. Furthermore, simultaneous model fit testing can be performed with the following hypotheses [11]:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{There must be at least one } \neq 0; k = 1, 2, \dots, p$$

With the test statistics used as follows in Equation 3.

$$D(\hat{\beta}) = -2 \ln \left[\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right] = 2 [\ln (L(\hat{\Omega})) - (L(\hat{\omega}))] \quad (3)$$

The assessment criteria are seen when rejecting H_0 if the value of the value of $D(\hat{\beta})$ is greater than $\chi^2_{(1-\alpha; df)}$, which indicates that at least one parameter in the Negative Binomial regression model has a significant effect on the response variable [12].

After conducting simultaneous testing, spatial testing was then carried out to determine which parameters would have a significant influence on the response variable with the following hypothesis:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

The test statistics used are as follows [13], as Equation 4.

$$Z_{hit} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \quad (4)$$

By deciding if you reject H_0 if the value is known $|Z_{hit}| > Z_{\alpha/2}$ yang which means that the parameter has a significant influence on the response variable in the negative binomial regression.

2.5 Spatial Effect Test

The existence of spatial influence was analyzed through spatial heterogeneity testing with the Breusch-Pagan test, to ensure that the model parameters are local and not fixed across various regions in East Java. Spatial heterogeneity is a condition where the characteristics or relationships between independent and dependent variables differ across locations [14]. The following are the hypotheses used in the Breusch-Pagan test:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \text{ (There is no spatial heterogeneity (constant residual variance))}$$

H_1 : Minimal ada satu $\sigma_n^2 \neq \sigma^2$ (There is spatial heterogeneity (residual variance is not constant))

With the test statistics used as follows [15], as Equation 5.

$$BP = \left(\frac{1}{2}\right) \mathbf{f}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{f} \sim X_{(p)}^2 \quad (5)$$

The assessment criteria are stated when H_0 is rejected if the BP test statistic value is greater than the critical value $X_{(a;p)}^2$. This condition indicates that the residual variance is not constant across locations, thus concluding that there is spatial heterogeneity in the data. In other words, at least one location has a different variance than the others.

2.6 Weighted and Bandwidth Selection

Bandwidth is a measure of distance that determines how much surrounding observations play a role in parameter estimation at a location [16]. In the GWNBR model, observation points within this range will be weighted according to the kernel function if the closer the distance to the center of the location, the greater the weight received. Determining the right bandwidth is very important to maintain a balance between model accuracy and smooth variation in estimation results. Meanwhile, this weighting function plays a role in assigning weights to each observation based on the geographic distance from the analysis location. Some kernel functions that are often used as weights include Gaussian, Bisquare, Tricube, and Exponential.

1. Gaussian Kernel (see in Figure 2)

The Gaussian kernel is a commonly used kernel function to capture non-linear patterns by indirectly projecting the data into a higher-dimensional feature space, making complex relationships easier to recognize.

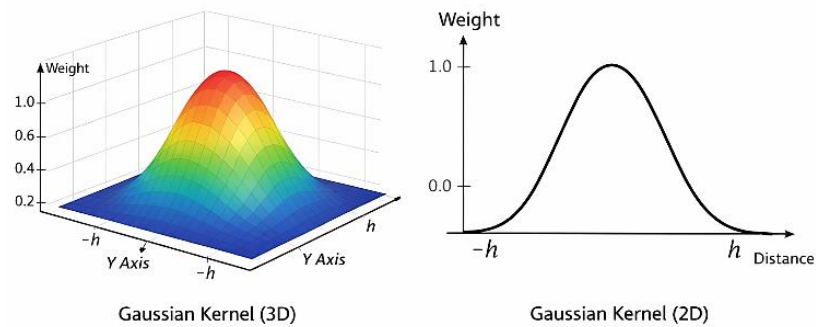


Figure 2. Gaussian Kernel Function Graph [17]

The formula for Gaussian Kernel weighting can be explained as follows [18]:

$$w_{ij} = \exp \left[-\frac{1}{2} \left(\frac{d_{ij}}{b} \right)^2 \right] \quad (6)$$

2. Bisquare Kernel (see in Figure 3)

The Bisquare (Biweight) kernel is a non-parametric kernel that weights only the data within a certain bandwidth and is commonly used in non-parametric regression and density estimation.

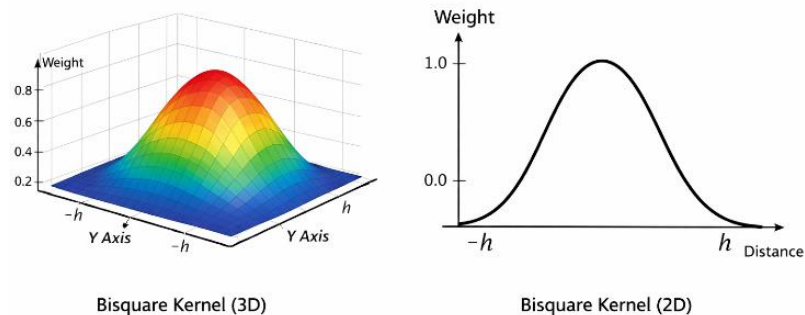


Figure 3. Bisquare Kernel Function Graph [19]

The formula for the weight can be explained as follows [20], in Equation 7.

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2, & \text{jika } d_{ij} < b \\ 0, & \text{jika } d_{ij} \geq b \end{cases} \quad (7)$$

3. Tricube Kernel (see in Figure 4)

Kernel The Tricube kernel is a non-parametric kernel with limited support that gives the largest weight to the nearest points and ignores data outside the bandwidth, making it effective for smoothing and stable estimation.

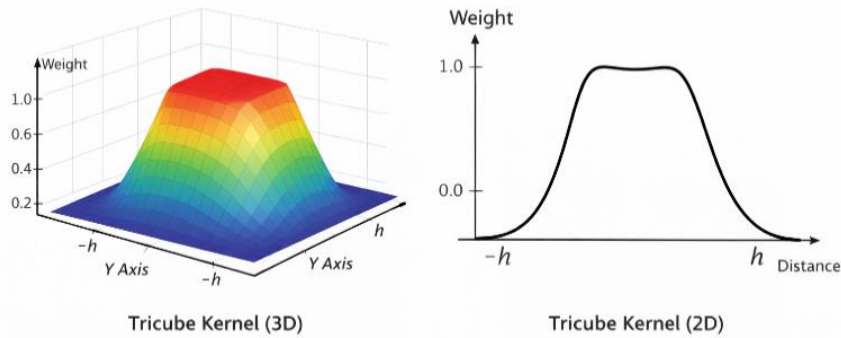


Figure 4. Tricube Kernel Function Graph [21]

The formula for the weight can be explained as follows [22], in Equation 8

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^3\right]^3, & \text{jika } d_{ij} < b \\ 0, & \text{jika } d_{ij} \geq b \end{cases} \quad (8)$$

4. Exponential Kernel (see in Figure 5)

The Exponential Kernel assigns weights based on the negative exponential of the distance, which is suitable for phenomena with rapid decline in influence and heavier tails than the Gaussian.

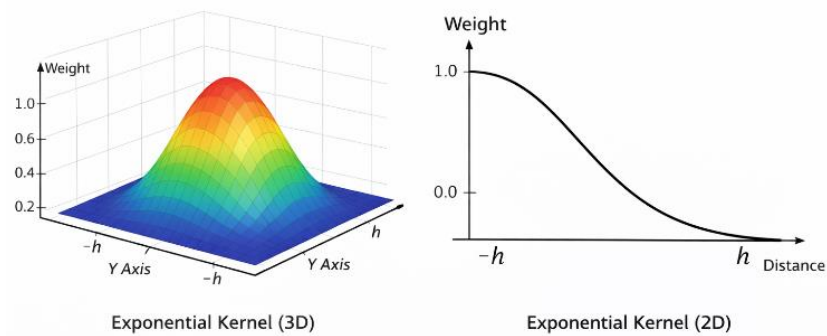


Figure 5. Exponential Kernel Function Graph [23]

The formula for the weight can be explained as follows [24]:

$$w_{ij} = \exp\left(-\frac{d_{ij}}{b}\right) \quad (9)$$

2.7 Geographically Weighted Negative Binomial Regression (GWNBR)

Geographically Weighted Negative Binomial Regression (GWNBR) is a localized extension of negative binomial regression that yields distinct parameter estimates at each location. This model is used with the assumption that the response variable follows a negative binomial distribution, which can be expressed as follows [25], as Equation 10.

$$y_i \sim BN(\mu_i, \theta_i) \quad (10)$$

If the data indicates a spatial effect, ordinary negative binomial regression is less appropriate because it assumes the same relationship across the entire area, even though each location may have different characteristics and environmental influences. Therefore, GWNBR is used because it can model relationships that vary locally. The GWNBR model can be formulated as follows [26]:

$$y_i \sim BN[\exp(\sum_k^p \beta_k(u_i, v_i)x_{ik}), \theta(u_i, v_i)] \quad (11)$$

Then the value of μ_i is connected to the explanatory variable through the link function algorithm, which can be written as follows, Equation 12.

$$\ln(\mu_i) = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)X_{i1} + \beta_2(u_i, v_i)X_{i2} + \dots + \beta_k(u_i, v_i)X_{ik} \quad (12)$$

The GWNBR model parameter estimation is performed using Maximum Likelihood Estimation. The likelihood function can be written as follows Equation 8.

$$L(\beta(u_i, v_i), \alpha) = \prod_{i=1}^n \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \quad (13)$$

GWNBR model testing is carried out in various ways, namely:

1. Testing the similarity of the GWNBR model with negative binomial regression
This test aims to see whether there is a significant difference between the GWNBR model and the negative binomial regression model [27]. The following hypothesis is used:

$$H_0 : \beta_1(u_i, v_j) = \beta_j, j = 1, 2, \dots, p, i = 1, 2, \dots, n$$

$$H_1 : \text{there is at least one } \beta_j(u_i, v_j) \neq \beta_j$$

Test statistics:

$$F_{hit} = \frac{\text{deviance model A/df}_A}{\text{deviance model B/df}_B} \quad (14)$$

2. Simultaneous and partial testing

The simultaneous test in the GWNBR model aims to evaluate the influence of all predictor variables on the response variable simultaneously. The test is conducted using the Likelihood Ratio Test (LRT) by comparing the full model to the model without predictors. Rejection of the null hypothesis indicates that at least one parameter in the model has a significant influence on the response variable. The hypotheses used are as follows:

$$H_0 : \beta_1(u_i, v_j) = \beta_2(u_i, v_j) = \dots = 0$$

$$H_1 : \text{there is at least one } \beta_k(u_i, v_j) \neq 0 ; k = 1, 2, \dots, p$$

Test statistics [28]:

$$D(\hat{\beta}) = -2 \ln \left[\frac{L(\hat{\omega})}{L(\hat{\Omega})} \right] = 2 [\ln(L(\hat{\Omega})) - \ln(L(\hat{\omega}))] \quad (15)$$

Where:

$L(\hat{\omega})$: a likelihood function for a simple model without involving predictor variables

$L(\hat{\Omega})$: a likelihood function for a complete model involving predictor variables

Next, partial parameter testing of the GWNBR model was conducted to evaluate the significance of each predictor variable's influence on the response variable at each study location. Because the GWNBR parameters are local, this testing is crucial for identifying differences in variable contributions at each observation point. The hypotheses used in this testing are as follows:

$$H_0 : \beta_k(u_i, v_j) = 0$$

$$H_1 : \beta_k(u_i, v_j) \neq 0$$

Test statistics:

$$z = \frac{\hat{\beta}_k(u_i, v_i)}{SE\hat{\beta}_k(u_i, v_i)} \quad (16)$$

Where:

$\hat{\beta}_k(u_i, v_i)$: parameter estimation for the kth variable at location (u_i, v_i)

SE : standard error of the parameter estimate

With the test criteria set by rejecting the null hypothesis if the absolute value of the test statistic $|Z|$ exceeds the critical value $Z_{\alpha/2}$. This condition shows that the parameters tested have a significant influence on the response variables at each location in the GWNBR model.

3. RESULTS AND DISCUSSION

This study uses the Geographically Weighted Negative Binomial Regression (GWNBR) model to analyze socio-economic factors on the number of poor people in East Java based on 2024 data. This model overcomes overdispersion and captures spatial heterogeneity, so it is able to show the influence of specific variables in each district/city.

3.1. Sample Data Description

This study uses data on the number of poor people in East Java as the response variable, with districts/cities as the units of analysis. Predictor variables include six socioeconomic indicators sourced from the Central Statistics Agency (BPS). A summary of descriptive statistics for all study variables, both response and predictor, is presented in detail in Table 1.

Table 1. Descriptive Statistics of Data on the Number of Poor People and Factors Suspected of Influencing Them

Variable	Mean	Median	Min	Max
Y	99733	103115	6590	240140
X_1	4.041	4.075	1.560	6.490
X_2	73.67	73.36	67.52	86.62
X_3	1937.9	855.5	410.0	8667.0
X_4	8.464	8.060	5.080	12.110
X_5	74.80	74.93	73.31	76.02
X_6	4.728	4.945	1.210	6.190

Based on the summary of descriptive statistics presented in Table 1, the response variable, the number of poor people in East Java, shows a mean of 99.733. This figure provides a general overview of the poverty level in the regencies/cities within the study area. The lowest recorded number of poor people was 6.590, while the highest number reached 240.140. This significant difference between the lowest and highest values indicates significant inequality or variation in the number of poor people across various regions in East Java.

3.2. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) on poverty levels in East Java was conducted through spatial mapping using choropleth maps. This mapping aims to identify distribution patterns, disparities between regions, and early indications of spatial heterogeneity in the number of poor people. Through this visual representation, variations in the number of people living below the poverty line are displayed through different colors, with lighter colors representing higher numbers and darker colors indicating lower numbers.

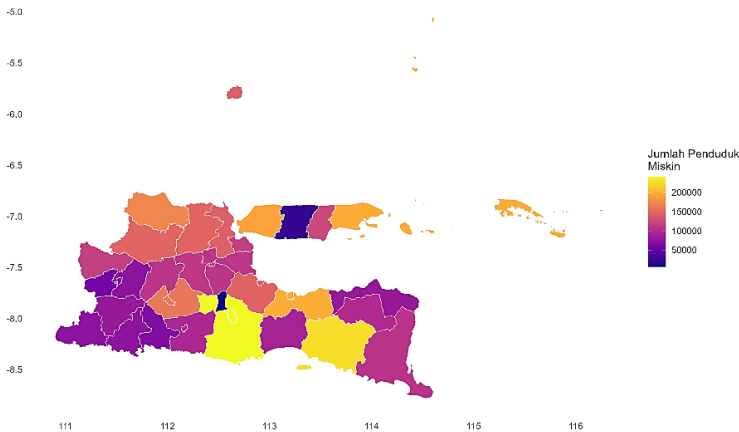


Figure 6. Visualization of the Distribution Map of the Number of Poor Residents in East Java Province

The distribution map in Figure 6 shows that the number of poor people in East Java is unevenly distributed. The southern and eastern regions tend to have higher poverty rates, while the western regions and some urban areas have lower rates. This pattern can be assumed to be due to spatial variation and clustering influenced by geographic, social, and economic factors within each district/city in East Java.

3.3. Multicollinearity Test

Multicollinearity detection was performed by evaluating the Variance Inflation Factor (VIF) values to identify a strong linear relationship between the predictor variables. The test results showed that all variables had low VIF values, ranging from 2 to 5. Since all these values were below the threshold ($VIF < 10$), it can be concluded that there was no significant multicollinearity problem, so these variables were suitable for use in further modeling.

3.4. Overdispersion Test

The poisson regression model requires equidispersion, which means equality between the expected value and variance. However, testing on the poverty data indicates a violation of this assumption in the form of overdispersion, where the variance value is found to be much larger than the expected value ($Var(Y) > E(Y)$). With a dispersion ratio of **29509.45**, it can be concluded that the standard Poisson regression model is unable to accurately capture data diversity, necessitating a negative binomial approach.

3.5. Negative Binomial Regression (Global Model)

The parameter estimation of the negative binomial regression model was performed using the Maximum Likelihood Estimation (MLE) method with the Newton-Raphson optimization algorithm. The iteration process was carried out until the convergence criterion was reached, resulting in stable and accurate parameter estimates. The results of the negative binomial regression model parameter estimation are detailed in the Table 2.

Table 2. Parameter Estimation of Negative Binomial Regression Model

	Estimate	Std. Error	Z	P-value
Intercept	-8.1440	15.7300	-0.518	0.60459
X_1	0.3065	0.1300	2.358	0.01837
X_2	0.01787	0.04189	0.426	0.66975
X_3	0.00001	0.00008	0.064	0.94896
X_4	-0.50740	0.13410	-3.785	0.00015
X_5	0.28910	0.22730	1.272	0.20344
X_6	-0.0824	0.10110	-0.815	0.41504
Deviance = 40.457			DF = 31	
AIC =				
941.01				

Based on the results of negative binomial regression modeling using a significance level of 5%, the deviance value was obtained ($D(\hat{\beta})$) by 40.457. When compared with the value $X_{31,0.05}^2$ by 44.985, then the model deviance value is much larger ($40.457 < 44.985$), so H_0 is rejected. This indicates that there is at least one predictor variable that has a significant influence on the response variable simultaneously. For partial testing with a significance level of 5%, a variable is said to be significant if the value $|Z_{hitung}|$ exceeds critical value $Z_{0.05/2}$ by 1.96 this means that only two predictor variables in the model individually have a significant influence on the number of poor people in East Java. The following table shows the negative binomial regression model:

$$\ln(\hat{\mu}) = -8.1440 + 0.3065X_1 + 0.01787X_2 + 0.00001X_3 - 0.5074X_4 + 0.2891X_5 - 0.0824X_6$$

From the estimation results, it can be observed that only two predictor variables, namely X_1 and X_4 have a significant influence on the response variable. Variable X_1 has a positive coefficient of 0.3065 with a p-value of 0.01837, indicating a statistically significant positive effect. This implies that for every one-unit increase in X_1 , the expected number of poor people increases by a factor of $\exp(0.3065) \approx 1.36$, assuming other variables remain constant. In practical terms, this suggests that higher values of X_1 contribute to an increase in poverty levels. On the other hand, variable X_4 shows a negative coefficient of -0.5074 with a p-value of 0.00015, indicating a statistically significant negative effect. This means that for every one-unit increase in X_4 , the expected number of poor people decreases by a factor of $\exp(-0.5074) \approx 0.60$. This finding suggests that improvements in X_4 are associated with a reduction in poverty levels.

Meanwhile, the other variables, namely X_2 , X_3 , X_5 , and X_6 , are not statistically significant, as their p-values exceed 0.05. This indicates that, at the global level, these variables do not have a strong enough influence to explain variations in the number of poor people in East Java. However, this does not necessarily imply that these variables are unimportant, as their effects may vary across locations and can be better captured using a spatial approach such as GWNBR. Overall, the global negative binomial model provides an initial understanding of the relationship between predictors and poverty. However, the limited number of significant variables suggests that a global model may not be sufficient to capture spatial variations, thereby justifying the use of the GWNBR model for more detailed local analysis.

3.6. Spatial Effect Testing

Spatial impact testing was conducted to identify spatial influences in the data through spatial heterogeneity testing. This process utilized the Breusch-Pagan method to determine differences in residual variation across observation locations. The Table 3 shows the results of the spatial effect testing.

Table 3. Results of Spatial Effect Test with Breusch-Pagan Values

Spatial Test	Test Statistics	Value	P-value
Spatial Heterogeneity	Breusch-Pagan (BP)	13.894	0.0308

Based on the results of this spatial effect test, the spatial heterogeneity test indicated significant heterogeneity. This is indicated by the Breusch-Pagan (BP) statistic value of 13.894 with a p-value of 0.0308. Because this p-value is smaller than the significance level of $0.0308 > 0.05$, it was decided to reject H_0 . This indicates significant spatial heterogeneity in the research data. This spatial diversity indicates that the influence of predictor variables on the number of poor people is local, making the use of the GWNBR model more appropriate than a global regression model to capture the specific characteristics of each district/city.

3.7. GWNBR Modeling

Analysis using the GWNBR model was conducted to capture spatial variation in factors influencing the number of poor people in East Java Province. Unlike global regression models, which produce a single parameter estimate for the entire region, GWNBR modeling produces local parameters for each district/city. This modeling process begins by integrating the geographic location of the region into the model, followed by determining appropriate spatial weighting for each observation location.

Based on the analysis results, the districts and cities can be classified into four distinct groups. This division is based on the similarity of predictor variable combinations that have been shown to have a significant influence in each region. Table 4 is a breakdown of the regional groupings formed according to the characteristics of their variable significance.

Table 4. Grouping of Provinces Based on Significant Variables

Group	Regency/City	Significant Variables
1	Bangkalan Regency Trenggalek Regency, Tulungagung Regency, Kediri Regency, Malang Regency, Lumajang Regency, Bondowoso Regency, Situbondo Regency, Sidoarjo Regency,	$(X_1), (X_2), (X_3), (X_4), (X_5)$
2	Mojokerto Regency, Jombang Regency, Madiun Regency, Magetan Regency, Ngawi Regency, Tuban Regency, Lamongan Regency, Gresik Regency, Sampang Regency, Blitar City, Malang City, Madiun City, Batu City	
3	Jember Regency, Nganjuk Regency, Sumenep Regency, Kediri City, Probolinggo City, Mojokerto City, Surabaya City Pacitan Regency, Ponorogo Regency, Blitar Regency, Banyuwangi	$(X_3), (X_4), (X_5), (X_6)$
4	Regency, Probolinggo Regency, Pasuruan Regency, Bojonegoro Regency, Pamekasan Regency, Pasuruan City	

The partial analysis in this study yielded parameter estimates that varied across districts/cities in East Java. Decision-making criteria were based on a comparison between $|Z_{hitung}|$ with a threshold of $Z_{0.05/2}$ of 1.96. with a threshold of $Z_{0.05/2}$ of 1.96. Rejection is stated as H_0 occurring if the value $|Z_{hitung}|$ of 1.96, which indicates that the variable is statistically significant. The test results show that the GWNBR model used involves variables that almost all have a significant influence in several districts/cities, including predictor variables such as TPT (X_1), TPAK (X_2), Population Density Rate (X_3), RLS (X_4), UHH (X_5), except for the Economic Growth Rate variable (X_6). For example, parameter testing will be conducted at a location in Bangkalan Regency. GWNBR model parameter testing in Bangkalan Regency can be seen in Table 5.

Table 5. GWNBR Model Parameter Testing in Bangkalan Regency

	Estimate	Z	P-value
Intercept	-8.144	-3499.358	0.000
X_1	0.6948	2.08865	0.000
X_2	0.0820	-36.7677	0.000
X_3	0.0001	-31.7318	0.000
X_4	-0.8243	2.68794	0.000
X_5	0.2344	1.87615	0.000
X_6	-0.0422	-0.17361	0.862

For partial testing with a significance level of 5%, a variable is said to be significant if the value of $|Z_{hitung}|$ exceeds the critical value of $Z_{0.05/2}$ of 1.96. Based on the test results table in Bangkalan Regency, it is known that almost all predictor variables have a p-value of 0.000 except for X_6 , which has a p-value > 0.05 . This means that almost all predictor variables individually have a significant influence on the number of poor people in the region. The GWNBR model formed for Bangkalan Regency is as follows:

$$\ln(\hat{\mu}) = -8.144 + 0.6948X_1 + 0.0820X_2 + 0.0001X_3 - 0.8243X_4 + 0.2344X_5 - 0.0422X_6$$

Based on the significant variables in the GWNBR model for Bangkalan Regency, it can be observed that several socio economic factors have a meaningful influence on the number of poor people. The Open Unemployment Rate (X_1) shows a positive and significant effect, where an increase of 1 percent is associated with an increase in the number of poor people by a factor of $\exp(0.6948) \approx 2.00$, assuming other variables remain constant. This indicates that unemployment is a major driver of poverty in Bangkalan Regency, where limited job opportunities may directly reduce household income and increase economic vulnerability. Similarly, the Labor Force Participation Rate (X_2) also shows a positive effect, suggesting that an increase in labor force participation does not necessarily reduce poverty. This condition may indicate that a large proportion of the workforce is engaged in low-productivity or informal sectors, which do not provide sufficient income to escape poverty. Population Density (X_3) has a positive but relatively small effect, indicating that higher population pressure may contribute to increased poverty, although its impact is not as strong as other variables. In contrast, Average Years of Schooling (X_4) has a significant negative effect, where an increase in education level reduces the number of poor people. This highlights the important role of education in improving human capital and access to better employment opportunities. Life Expectancy (X_5) shows a positive relationship with poverty, which may reflect underlying structural issues, such as unequal access to health services or demographic imbalances. Meanwhile, the Economic Growth Rate (X_6) has a negative but insignificant effect, indicating that economic growth in the region may not be evenly distributed and therefore does not directly reduce poverty levels. Overall, these findings indicate that poverty in Bangkalan Regency is influenced by a combination of labor market conditions, education level, and demographic factors. This reinforces the importance of region-specific policies, particularly in improving employment opportunities and access to education, as key strategies for poverty reduction.

3.8. Best Model Selection

The best model was selected by comparing the performance of the global regression model (Negative Binomial) with a local regression model that takes spatial elements into account, namely the Geographically Weighted Negative Binomial Regression (GWNBR) with four types of kernel weights. The criterion used to determine the most optimal model was the Akaike Information Criterion (AIC) value. A summary of the comparison of AIC values for all tested models is presented in the Table 6.

Table 6. Comparison of Model Evaluation Results with AIC Values

Regression Model	AIC
Negative Binomial Regression	941.01
GWNBR Adaptive Bisquare	309.32044
GWNBR Adaptive Gaussian	309.618422

Regression Model	AIC
GWNBR Adaptive Tricube	309.529982
GWNBR Adaptive Exponential	319.752320

Based on the comparison of Akaike Information Criterion (AIC) values in Table 6, the GWNBR Adaptive Bisquare model was selected as the best model because it had the lowest AIC value of 309.32044. This value is much smaller than the global regression model (Negative Binomial Regression), which had an AIC of 941.01. This highly significant AIC value indicates that the local GWNBR model is much more efficient at explaining the number of poor people in East Java than the global model. This result also aligns with the spatial heterogeneity test, which shows variations in characteristics across regions.

4. CONCLUSION

This study aims to model the number of poor people in East Java by addressing overdispersion and spatial heterogeneity using the Geographically Weighted Negative Binomial Regression (GWNBR) method. Based on the results, the GWNBR model captures spatial variation in poverty more accurately than the global Negative Binomial Regression model. This is evidenced by the significantly lower AIC value of the GWNBR model with the Adaptive Bisquare kernel (309.32044) compared to the global model (941.01). The findings indicate that the influence of predictor variables on poverty is not uniform across districts/cities, confirming the presence of spatial heterogeneity. The GWNBR model successfully provides local parameter estimates that reflect region-specific characteristics, allowing for a more detailed understanding of the dominant factors influencing poverty in each area, such as in Bangkalan Regency. Therefore, the results of this study answer the research objective by demonstrating that a spatial-based approach is more effective in modeling poverty data with complex regional variations.

In terms of contribution, this study provides a more comprehensive modeling framework by integrating the handling of count data, overdispersion, and spatial effects simultaneously through the GWNBR approach with multiple kernel comparisons. The results also offer practical implications for policymakers, emphasizing that poverty alleviation strategies should be tailored to local conditions rather than applied uniformly across regions. For future research, it is recommended to incorporate additional relevant variables, such as infrastructure, social assistance, or environmental factors, to improve model accuracy. Furthermore, the use of alternative spatial models or the expansion of the study area and time period may provide deeper insights into the dynamics of poverty and enhance the robustness of the analysis.

REFERENCES

[1] Oxford Poverty and Human Development Initiative, *Global Multidimensional Poverty Index 2025 – Overlapping Hardships: Poverty and Climate Hazards*, 2025, doi: 10.18356/9789211546316.

[2] J. R. Harismawan, H. R. A. H., and M. T. Wanda, “The role of financial inclusion in achieving the percentage of poor population targets in the 2024 RPJMN case study: Indonesia,” *Parahyangan Economic Development Review*, vol. 3, no. 1, pp. 234–246, 2025, doi: 10.26593/pedr.v3i1.7504..

[3] A. Salam, D. S. Pratomo, and P. M. A. Saputra, “Analisis kemiskinan pada rumah tangga di Jawa Timur melalui pendekatan multidimensi dan moneter,” *Jurnal Kependudukan Indonesia*, vol. 16, no. 2, pp. 127–138, 2021, doi: 10.14203/jki.v16i2.480.

[4] M. B. A. Fahreza and D. P. Pratama, “Analisis faktor sosial ekonomi terhadap kesejahteraan masyarakat di Provinsi Jawa Timur tahun 2023 menggunakan analisis faktor,” *Jurnal Global Ilmiah*, vol. 2, no. 2, pp. 907–916, 2024, doi: 10.55324/jgi.v2i2.136.

[5] R. Y. Y. Nugroho and S. J. I. J. Isnaini, “Analisis determinan kemiskinan di Jawa Timur tahun 2018,” *Jurnal GeoEkonomi*, vol. 11, no. 2, pp. 176–187, 2020, doi: 10.36277/geoekonomi.v11i2.120.

[6] M. R. F. Fuady, M. Fuady, and F. Aulia, “Kemiskinan multi dimensi dan indeks pembangunan manusia di Indonesia,” *TATALOKA*, vol. 23, no. 4, pp. 575–582, 2021, doi: 10.14710/tataloka.23.4.575-582.

[7] N. A. Kusuma, N. A. A. Zakiyah, and M. N. Cholida, “Faktor-faktor yang memengaruhi kemiskinan di 38 kabupaten/kota Jawa Timur,” in *Seminar Nasional Potensi dan Kemandirian Daerah*, vol. 2, 2024.

[8] N. K. Lestari, D. W. Hikmawan, I. Safitrah, and E. Widodo, “Analisis faktor lingkungan penyebab kasus diare di Jawa Barat tahun 2016 menggunakan regresi binomial negatif,” in *Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya*, 2019.

[9] D. Ruswanti, “Performance measurement of support vector machine and neural network in predicting poverty levels,” *Jurnal Gaung Informatika*, vol. 13, no. 1, 2020, doi: 10.47942/gi.v13i1.455.

[10] A. J. Syafiqoh, R. Mahardika, S. Amaria, E. Winaryati, and M. Al Haris, “Pemodelan regresi binomial negatif untuk mengevaluasi faktor-faktor yang mempengaruhi kasus tuberkulosis di Provinsi Jawa Barat,” *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, vol. 12, no. 1, pp. 15–23, 2024, doi: 10.24252/msa.v12i1.39450.

- [11] M. Fathurahman, "Regresi binomial negatif untuk memodelkan kematian bayi di Kalimantan Timur," *Eksponensial*, vol. 13, no. 1, pp. 79–86, 2022, doi: 10.30872/eksponensial.v13i1.888.
- [12] H. M. Winata, "Mengatasi overdispersi dengan regresi binomial negatif pada angka kematian ibu di Kota Bandung," *Jurnal Gaussian*, vol. 11, no. 4, pp. 616–622, 2023, doi: 10.14710/j.gauss.11.4.616-622.
- [13] A. Tamimi and I. Husein, "Application of negative binomial regression in analyzing factors that influence stunting in toddlers," *Desimal: Jurnal Matematika*, vol. 7, no. 1, pp. 103–112, 2024, doi: 10.24042/djm.v7i1.22070.
- [14] A. Prayoga, A. R. Maulana, and J. Falson, "Analisis heterogenitas spasial tingkat pengangguran terbuka di Pulau Jawa tahun 2024: Pendekatan geographically weighted regression," in *Seminar Nasional Official Statistics*, vol. 2025, no. 1, pp. 59–70, 2025, doi: 10.34123/semnasoffstat.v2025i1.2418.
- [15] D. F. Fedriansya and B. Rahman, "Analisis tren perubahan spasial fungsi lahan di kawasan pesisir," *Jurnal Plano Buana*, vol. 6, no. 1, pp. 19–27, 2025, doi: 10.36456/vbfcscd78.
- [16] P. G. O. W. Putra and I. K. N. A. Jaya, "Implementasi bandwidth management menggunakan Mikrotik RouterOS (studi kasus di PT. Rejeki Maha Bumi Lestari)," *RESI: Jurnal Riset Sistem Informasi*, vol. 1, no. 1, pp. 1–9, 2022, doi: 10.32795/resi.v1i1.2936.
- [17] W. Wang and B. Y. Jing, "Gaussian process regression: Optimality, robustness, and relationship with kernel ridge regression," *Journal of Machine Learning Research*, vol. 23, no. 193, pp. 1–67, 2022.
- [18] Z. Liu, C. Lyu, J. Huo, S. Wang, and J. Chen, "Gaussian process regression for transportation system estimation and prediction problems: The deformation and a hat kernel," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22331–22342, 2022, doi: 10.1109/TITS.2022.3155527.
- [19] W. Reza, F. C. H. Buan, and P. Angreni, "Comparison of geographically weighted regression with adaptive Gaussian and bisquare kernel on open unemployment rate in Riau Islands," *Leibniz: Jurnal Matematika*, vol. 6, no. 1, pp. 15–28, 2026, doi: 10.59632/leibniz.v6i01.705.
- [20] A. Asnita, S. Sifriyani, and M. Fauziyah, "Estimation of geographically weighted panel regression model with bisquare kernel weighting function on percentage of stunting toddlers in Indonesia," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 1, pp. 383–394, 2024, doi: 10.30598/barekengvol18iss1pp0383-0394.
- [21] Y. Pan et al., "A mask guided oriented object detector based on rotated size-adaptive tricube kernel," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024, doi: 10.1109/TGRS.2024.3377704.
- [22] R. Putra, S. W. Tyas, and M. G. Fadhlurrahman, "Geographically weighted regression with the best kernel function on open unemployment rate data in East Java Province," *Enthusiastic: International Journal of Applied Statistics and Data Science*, pp. 26–36, 2022.
- [23] T. K. J. Kwan, F. Chen, and W. Dunsmuir, "Likelihood inference of the non-stationary Hawkes process with non-exponential kernel," *arXiv preprint arXiv:2408.09710*, 2024.
- [24] N. Batra, "Modelling high-frequency data with bivariate Hawkes processes: Power-law vs. exponential kernels," *arXiv preprint arXiv:2503.14814*, 2025.
- [25] Z. Mar'ah, Z. Rais, and A. S. Haris, "Geographically weighted negative binomial regression (GWNBR) in modeling the risk factors of pneumonia disease among toddlers in Central Sulawesi Province," *VARIANSI*, vol. 5, no. 3, pp. 118–131, 2023, doi: 10.35580/variensiunm151.
- [26] A. Prahutama et al., "Geographically weighted negative binomial regression for modelling the number of dengue diseases and their mortality in East Nusa Tenggara, Indonesia," *Journal of Mathematics and Computer Science*, vol. 11, no. 6, pp. 7971–7990, 2021, doi: 10.28919/jmcs/6740.
- [27] D. A. Mustika and R. Nooraeni, "Kajian efek spasial kasus difteri dengan geographically weighted negative binomial regression (GWNBR)," *Indonesian Journal of Statistics and Its Applications*, vol. 3, no. 1, pp. 91–104, 2019, doi: 10.29244/ijsa.v3i1.185.
- [28] P. Handayani, "Pemetaan spasial dan deteksi wilayah kantong tuberkulosis menggunakan metode geographically weighted negative binomial regression dan flexibly shaped spatial scan statistic (studi kasus: jumlah kasus tuberkulosis di Provinsi Jawa Timur tahun 2018)," Bachelor's thesis, Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah Jakarta, 2021.