JAIIT (Journal of Advances in Information and Industrial Technology)
Vol. 8, No. 1 Mei 2026, ISSN 2716-1935, E-ISSN 2716-1927

Implementation of the Semiparametric Geographically Weighted Logistic Regression (GWLRS) Model for Predicting Poverty Depth Index in Central Java Province

Ikmal Thariq Kadafi[1)], Trimono[*2)], Shindi Shella May Wara[3)]
[1,2,3)]Department of Data Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia
[1]22083010052@student.upnjatim.ac.id, [2*]trimono.stat@upnjatim.ac.id, [3]shindi.shella.fasilkom@upnjatim.ac.id

ARTICLE INFO

Article history:
Received 31 January 2026
Revised 12 May 2026
Accepted 30 May 2026
Available online 31 May 2026

Keywords:
Central Java
GWLRS
Poverty
Spatial Analysis
Spatial Regression

ABSTRACT

Poverty in Central Java Province remains a significant multidimensional issue, characterized by socio-economic disparities across regencies and municipalities and high rates of school dropouts among children. This study aims to evaluate the influence of socio-economic variables on poverty depth (P1 Index) at both local and global levels. The approach employed is the Geographically Weighted Logistic Regression Semiparametric (GWLRS), which integrates local and global effects. The model uses two types of spatial weights: Adaptive Gaussian Kernel and Queen Contiguity, with predictor variables including Dependency Ratio, Minimum Regional Wage (UMK), Number of Industries, Open Unemployment Rate (TPT), Adequate Housing, and Sanitation. Parameter estimation was conducted using Maximum Likelihood. The results indicate that the Dependency Ratio, Adequate Housing, and Sanitation are locally significant in 2–5 regions. Local coefficients for the Dependency Ratio range from 0.42 to 11.57 (mean 3.33), Adequate Housing from −15.808 to −2.371 (mean −5.95), and Sanitation from 1.86 to 17.27 (mean 5.71). The model correctly predicts 31 out of 35 cases, yielding an accuracy of 91.4%. The Number of Industries, UMK, and TPT are not globally significant, indicating that their effects are more stable across regions. In conclusion, the GWLRS model effectively captures the spatial heterogeneity of poverty determinants and provides quantitative insights that can support more targeted, location-based poverty alleviation policies in Central Java Province.

1. Introduction

Poverty is a multidimensional issue that remains a major concern in Indonesia. This condition is not only related to the inability to meet basic needs but is also influenced by social, cultural, economic, and spatial factors [1]. Despite the implementation of various government policies, poverty continues to be a complex problem that is difficult to address comprehensively [2]. According to the September 2024 National Socio-Economic Survey (Susenas), the national poverty rate was recorded at 8.57%, equivalent to 24.06 million people. Although this figure represents a decline compared to the previous year, it remains below the National Medium-Term Development Plan (RPJMN) 2020–2024 target of 6.5–7.5% [3]. The complexity of this issue is further reinforced by the phenomenon of the poverty trap, in which unemployment, inequality, and low economic growth continuously interact and exacerbate societal conditions [4].

Central Java Province is one of the regions that contributes significantly to the total number of poor people in Indonesia [5]. Although the poverty trend has shown a decline from 10.47% in March 2024 to 10.32% in September 2024, this rate remains higher than the national average [3]. Historically, Central Java has also experienced substantial fluctuations in poverty levels, such as an increase to 11.79% in 2021 due to the COVID-19 pandemic, before declining again in 2023 [6]. The spatial pattern of poverty in this province reveals clusters of high-poverty areas, including Kebumen, Purbalingga, Banjarnegara, Banyumas, and Cilacap, indicating that poverty is not randomly distributed but rather concentrated in regions with similar socio-economic characteristics [7].

The temporal dimension further indicates that poverty in Central Java is highly vulnerable to external shocks, including pandemics, inflation, and global economic dynamics. In 2024, the poverty gap index (P1) was recorded at 1.41, declining from 1.61 in March 2024, while the poverty severity index (P2) also showed a downward trend [3]. Nevertheless, poverty remains concentrated in rural areas with a strong dependence on traditional agricultural sectors [6]. Moreover, the impacts of poverty are reflected in the high number of out-of-school children, including more than 23,000 children in Pemalang who are not attending school, the majority of whom come from extremely poor households. This condition demonstrates that poverty has far-reaching social implications, extending beyond purely economic dimensions.

Numerous studies have attempted to analyze the spatial patterns of poverty using models such as Geographically Weighted Regression (GWR) and Geographically Weighted Logistic Regression (GWLR). These methods allow for the estimation of local parameters that vary according to regional characteristics. For instance, Saifudin et al. [8] demonstrated that GWLR is capable of modeling spatial variation in the Human Development Index in East Java; however, their study employed only a single type of spatial weighting scheme, limiting its ability to capture more complex spatial dependence. Meanwhile, other studies, such as Fitriatusakiah et al. [9], have applied semiparametric approaches but were restricted to cross-sectional data and did not examine variations in spatial weighting structures.

The variables used in this study were selected based on their theoretical and empirical relevance in explaining the depth of poverty. The Dependency Ratio reflects the household dependency burden, which can affect a family's economic capacity; the Regency/City Minimum Wage (UMK) and the number of industries represent income opportunities and regional economic potential; the Open Unemployment Rate (TPT) serves as an indicator of poverty risk through reduced household income; while access to adequate housing and proper sanitation reflects the fulfillment of basic needs and the quality of the living environment, both of which influence welfare and individuals' capabilities to escape poverty. Through this combination of variables, the study is able to capture the economic, social, and infrastructural dimensions that directly contribute to poverty levels.

Based on this research gap, the Geographically Weighted Logistic Regression Semiparametric (GWLRS) approach is considered relevant because it is capable of combining the strengths of global and local models while capturing nonlinear relationships among variables. In addition, this model allows the use of spatial weighting through an Adaptive Gaussian Kernel to assess the sensitivity of the model to different neighborhood structures. By incorporating six poverty-determining variables, namely the Dependency Ratio, UMK, TPT, number of industries, adequate housing, and proper sanitation, the GWLRS model has the potential to produce a more comprehensive analysis. Therefore, this study is important in exploring a robust, interpretable, and applicable modeling framework to support location-based poverty alleviation policies in Central Java.

2. Methods

2.1. Data and Data Source

This study employs panel data from 35 regencies/municipalities in Central Java Province, with the response variable defined as the Poverty Gap Index (P1), which is categorized into two classes based on the provincial average value. The predictor variables consist of socio-economic and demographic indicators, namely (X1) dependency ratio, (X2) regency/municipality minimum wage (UMK), (X3) number of industries, (X4) open unemployment rate (TPT), (X5) percentage of adequate housing, and (X6) access to improved sanitation. All variables are obtained from official publications of Statistics Indonesia (BPS) and sectoral data from local government agencies. In addition, the dataset includes geographic coordinates in the form of longitude and latitude for each regency/municipality, enabling

 DOI: https://doi.org/10.52435/jaiit.v8i1.767

spatial analysis within the GWLRS framework. All data are presented in panel format and have undergone data cleaning, verification, and harmonization before analysis.

2.2. Tools and Software

The analysis in this study was conducted using the Python programming language as the primary platform. Several libraries were employed to support data processing and spatial modeling, including Pandas and GeoPandas for tabular and spatial data management, and Matplotlib for visualization. To examine spatial heterogeneity and spatial autocorrelation, the study utilized the spreg module from the PySAL library, which provides statistical tests such as the Breusch–Pagan test and Moran's I. The baseline geographically weighted (GW) model was constructed using the mgwr module, which allows flexible estimation of local coefficients. In addition, the spglm module was used to support the binomial family, enabling the extension of the GWR framework into Geographically Weighted Logistic Regression (GWLR) as required by the analysis. Python was selected due to its flexibility, open-source nature, and extensive library support for spatial analysis and location-based modeling.

2.3. Geographically Weighted Logistic Regression (GWLR)

Geographically Weighted Logistic Regression (GWLR) is a statistical model used to analyze data exhibiting spatial heterogeneity. Spatial heterogeneity occurs when the same independent variables generate different responses across locations within the study area. GWLR is an extension of global logistic regression, in which model parameters are estimated locally for each observation location, resulting in location-specific parameter values [10]. Mathematically, the GWLR model can be expressed as follows.

$$\pi_{(x_i)} = \frac{\exp\left(\sum_{j=0}^{p} \beta_j(u_i,v_j)\right)}{1+p\left(\sum_{j=0}^{p} \beta_j(u_i,v_j)\right)} \tag{1}$$

The logit transformation is applied to enable a linear relationship between the independent variables and the response variable. Accordingly, the GWLR model employs a logit transformation. The functional form of the GWLR logit transformation is expressed as follows.

$$logit\big(\pi(x_i)\big) = \sum_{j=i}^{n} \beta_j(u_i, v_j)x_{ij} \tag{2}$$

In the GWLR model, the coefficient βj u_i, v_j represents a local coefficient whose value varies with the spatial location u_i, v_j allowing each observation point to have distinct parameter estimates.

2.4. Geographically Weighted Logistic Regression Semiparametric (GWLRS)

Geographically Weighted Logistic Regression Semiparametric (GWLRS) is an extension of the Geographically Weighted Logistic Regression (GWLR) model that accommodates a combination of local and global coefficients within a single framework [11]. This model is appropriate when some predictor variables exhibit spatially varying (non-stationary) effects across regions, while other variables remain constant (stationary). In general, the GWLRS model can be formulated as follows.

$$\pi(x_i) = \frac{\exp(\beta_0(u_i,v_i)+\sum_{j=1}^{k*} \beta_j\big(u_i,v_i\big)x_{ij}+\sum_k^k m=k*+1)}{+\exp(\beta_0(u_i,v_i)+\sum_{j=1}^{k*} \beta_j\big(u_i,v_i\big)x_{ij}+\sum m=k*+1\gamma_{\mathrm{m}}\, x_{im})} \tag{3}$$

The logit form of the GWLRS model is expressed as follows:

$$logit\big(\pi(x_i)\big) = \sum_{j=i}^{k*} \beta_j\big(u_i, v_i\big)x_{ij} + \sum_{m=k*+1}^{k} \gamma_m x_{im} \tag{4}$$

In the GWLRS model, the coefficient βj u_i, v_j represents a local coefficient whose value varies with the spatial location u_i, v_j allowing each observation point to have distinct parameter estimates. Meanwhile, γm denotes a global coefficient that is assumed to be constant across the entire study area. The number of variables modeled with local coefficients is denoted by k∗, while k represents the total number of predictor variables included in the model. The variables xij and xim correspond to the predictor values associated with local and global coefficients, respectively, for the i-th observation.

Parameter estimation in the GWLRS model follows the same approach as in the GWLR model, employing the Maximum Likelihood Estimation (MLE) method. Accordingly, the likelihood function is constructed as follows.

$$L(\beta(u_i, v_i)\gamma) = \prod_{i=1}^{n} P(Y_i = y_i) = \prod_{i=1}^{n} \pi(x_i)^{yi}[1 - \pi(x_i)]^{1 - yi} \tag{5}$$

After obtaining the likelihood function, a logarithmic transformation is applied, followed by the incorporation of spatial weights for each region. Parameter estimates are then derived by differentiating the resulting weighted log-likelihood function with respect to $\beta(ui, vi)$ and γ and setting the derivatives equal to zero. Consequently, the estimating equations can be expressed as follows

$$\frac{ln\,L\ \ \beta(u_i,v_i)\gamma}{\partial\beta\,(u_i,v_i)} = \sum_{i=1}^{n} y_i w_i(u_i, v_i)x_{G,i} - w_i(u_i, v_i)\pi(x_i) \tag{6}$$

2.5. Adaptive Gaussian Kernel

This function is an extension of the standard Gaussian kernel, but it is adaptive in nature because the bandwidth (h_i) is not fixed; instead, it adjusts according to the density of data around each observation point [12]. Mathematically, the Adaptive Gaussian Kernel weighting function can be expressed as follows.

$$w_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right] \tag{7}$$

In this weighting function, dij denotes the Euclidean distance between locations i and j, while hi represents the adaptive bandwidth at location i. When data density around point i is high, the value of hi becomes smaller, thereby increasing the influence of nearby observations and smoothly attenuating the influence of more distant ones.

2.6. Determination of Distance and Optimal Bandwidth

In the GWR model, spatial coordinates in the form of latitude (ui) and longitude (vi) are used to compute the distance between observation locations (dij). This distance is commonly calculated using the Euclidean distance, as expressed by the following equation:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \tag{8}$$

The determination of weights for each observation point depends on the bandwidth (h), which can be interpreted as the radius of influence around a given location. The bandwidth determines the extent to which neighboring observations contribute to the estimation process; therefore, its selection is crucial for obtaining a stable and accurate GWR model.

The optimal bandwidth can be selected using the cross-validation (CV) method, with the following objective function:

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

Where y≠i (h) denotes the predicted value at the i-th location when the corresponding observation is excluded from the estimation process (leave-one-out). The optimal bandwidth is obtained at the value of h that minimizes the CV criterion, as it indicates the lowest prediction error and the best model performance in capturing spatial patterns.

2.7. Multicolinearity Test (VIF)

A multicollinearity test is conducted to ensure that the predictor variables in the model do not exhibit excessively high linear correlations with one another, as such conditions may lead to unstable coefficient estimates and inflated parameter variances [13]. The test is performed using the Variance Inflation Factor (VIF), which is calculated based on the linear relationship between each predictor variable and the remaining predictors. The VIF value for the j-th variable is given by:

$$VIF = \frac{1}{1 - R_1^2} \tag{10}$$

 DOI: https://doi.org/10.52435/jaiit.v8i1.767

Where R_1^2 denotes the coefficient of determination obtained from regressing the j-th predictor on all other predictor variables. A variable is considered to exhibit multicollinearity when its VIF value exceeds the commonly accepted threshold of 10 [14]. Variables exceeding this threshold are considered for removal, transformation, or further evaluation to maintain the stability of the model's parameter estimates.

2.8. Spatial Heterogeneity (Breusch-Pagan)

A spatial heterogeneity test is conducted to examine whether the relationship between the response variable and the predictor variables is spatially non-stationary, that is, whether the effect of a given variable varies across regions [15]. This test is essential prior to applying the GWLRS model, as the model is specifically designed to capture local variations in regression parameters. The Breusch–Pagan test is employed for this purpose.

$$BP = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T f + \left(\frac{1}{T}\right) \left[\frac{e^T W e}{\sigma^2}\right]^2 \sim x^2_{(p)} \quad (11)$$

In the spatial heterogeneity test, the vector f_i consists of the squared residuals e^2 for each observation, with their variance denoted accordingly. The number of independent variables is represented by p, while W denotes the spatial weighting matrix [16]. The decision rule for the Breusch–Pagan (BP) test is to reject the null hypothesis H_0 if $BP > x^2$ or if the p-value is less than the significance level α. This condition indicates the presence of spatial heterogeneity, implying that geographically weighted modeling is appropriate to proceed.

3. Results and Discussions

3.1. Exploratory Spatial Data Analysis

As an initial step of the analysis, an exploration of the P1 variable was conducted to understand its distribution pattern and variation across regencies/municipalities in Central Java Province. This analysis aims to provide a preliminary overview of the spatial characteristics of the data and to identify differences in values across regions. To facilitate interpretation, P1 values are visualized using a thematic map, allowing spatial patterns to be observed more clearly.

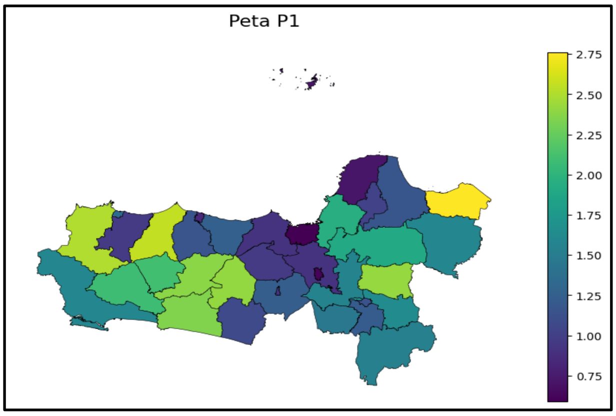

Figure 1. Thematic Map of the Spatial Distribution of P1 in Central Java

Based on the thematic map in Figure 1, the P1 values exhibit considerable spatial variation, ranging from 0.590 to 2.760. Areas with relatively high P1 values tend to be concentrated in several regencies, such as Rembang, Pemalang, Brebes, and Wonosobo, whereas lower values are generally observed in urban areas, including Semarang City and Salatiga City. This pattern indicates that the spatial characteristics of P1 are not homogeneous across regions, suggesting the presence of potential spatial non-stationarity that should be taken into account in subsequent stages of analysis.

Following the analysis of the dependent variable's spatial distribution, the next stage involves exploring the independent variables, namely the dependency ratio (DepRatio), regency/municipality minimum wage (UMK), number of industries, open unemployment rate (TPT), percentage of adequate housing, and access to improved sanitation. This analysis aims to identify the spatial patterns of each explanatory variable and to examine potential regional differences that may influence the variation in the dependent variable.

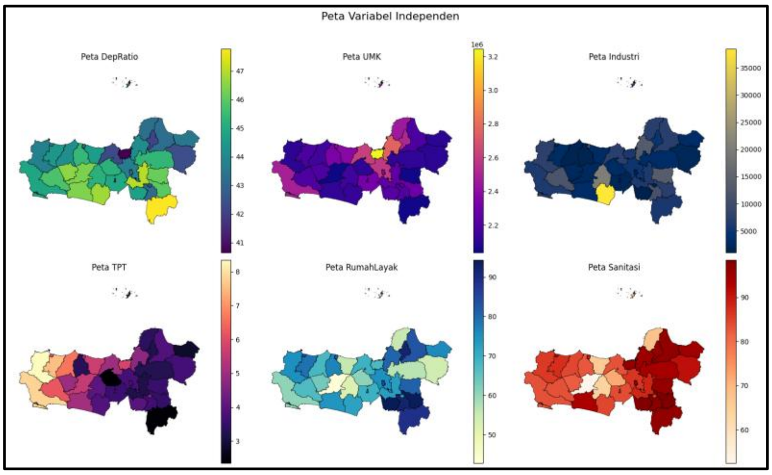

Figure 2. Thematic Maps of the Spatial Distribution of Independent Variables

Based on the thematic maps in Figure 2, each independent variable exhibits a distinct spatial pattern. The minimum wage (UMK) and the number of industries show high-value concentrations in urban areas and industrial zones, such as Semarang City, Demak, and Sragen, while regions with predominantly agrarian characteristics tend to display lower values. In contrast, the dependency ratio (DepRatio) and the open unemployment rate (TPT) demonstrate more dispersed spatial variations, with several rural areas exhibiting relatively higher values. Meanwhile, the percentage of adequate housing and access to improved sanitation tends to be higher in regions with more advanced infrastructure development. These differences in spatial patterns indicate that the effects of the independent variables on the dependent variable are likely to be location-specific, thereby supporting the use of a spatial analytical approach in the subsequent modeling stage.

3.2. Multicolinearity Test

Before further modeling, it is essential to ensure that the independent variables do not suffer from multicollinearity. Multicollinearity occurs when there are strong linear relationships among independent variables, which may lead to unstable parameter estimates and reduce the reliability of model interpretation. Therefore, a multicollinearity test is conducted using the Variance Inflation Factor (VIF) to assess the degree of correlation among the explanatory variables.

Table 1. Results of Multicolinearity Test

Variable Name	Variable	VIF
Dependency ratio	X1	2.103011
UMK	X2	1.766088
Number of industries	X3	1.167679
TPT	X4	2.374499
Adequate Housing	X5	4.025702
Sanitation	X6	3.738808

 DOI: https://doi.org/10.52435/jaiit.v8i1.767

Based on the VIF test results in Table 1, all independent variables have VIF values below 5, ranging from 1.17 to 4.03. The adequate housing and sanitation variables exhibit relatively higher VIF values compared to the other variables, indicating the presence of moderate correlation; however, these values remain within acceptable limits. Meanwhile, the dependency ratio, minimum wage (UMK), number of industries, and open unemployment rate (TPT) display low VIF values, suggesting no indication of serious multicollinearity. Therefore, it can be concluded that all independent variables are suitable for inclusion in the subsequent modeling process without the need for variable removal or transformation.

3.3. Global Logistic Regression

After confirming that there are no significant multicollinearity issues among the independent variables, the next step involves estimating a global logistic regression model. This model serves as a baseline to identify the relationship between the independent variables and the dependent variable (encoded P1), as well as to evaluate the global significance of each predictor before considering more complex spatial approaches.

Table 2. Results of the Global Logistic Regression Model

Variable	Coefficients	*p-value*	Pseudo R^2	LLR *p-value*
const	-79,7949	0.039		
X1	1,6412	0.027		
X2	0,00000181	0.493		
X3	0,00006985	0.351	**0,5718**	**0,00011**
X4	0,565	0.119		
X5	0,338	0.024		
X6	0,389	0.022		

Overall, the global logistic regression model demonstrates in Table2 shows good performance and statistical significance, as indicated by a Pseudo R-squared value of 0.5718 and a likelihood ratio test (LLR) p-value of 0.0001. These results suggest that the model is able to explain the global variation in the probability of P1 occurrence with an adequate level of goodness of fit. In general, the direction of the coefficients indicates that the probability of P1 occurrence tends to increase with higher dependency ratios, open unemployment rates, and sanitation levels, while it decreases with an increase in the percentage of adequate housing.

At the partial level, the dependency ratio has a positive and statistically significant effect on the probability of P1 occurrence, with a coefficient of 1.641 ($p = 0.027$) and an odds ratio of 5.16, indicating that the likelihood of P1 occurrence increases as the population dependency burden rises. The sanitation variable also shows a positive and significant effect, with a coefficient of 0.389 ($p = 0.022$) and an odds ratio of 1.48. In contrast, adequate housing has a negative and statistically significant effect, with a coefficient of −0.338 ($p = 0.024$) and an odds ratio of 0.71, suggesting a reduced probability of P1 occurrence as housing conditions improve. Conversely, the minimum wage (UMK), number of industries, and open unemployment rate do not exhibit statistically significant effects at the 5% significance level, indicating that their global contributions to changes in P1 probability are relatively limited. Therefore, this global logistic regression model is employed as a baseline for assessing whether spatial approaches can enhance model performance and capture local variations in the effects of explanatory variables across regions.

3.4. Spatial Heterogeneity Test

The results of the Breusch–Pagan test yield a p-value of **0.0471**, indicating the presence of heteroskedasticity, whereby the error variance is not constant across observations. This finding suggests that the relationship between the independent and dependent variables may vary across regions, implying spatial heterogeneity. Consequently, a more advanced modeling approach capable of capturing local variations is required. In this study, the Geographically Weighted Logistic Regression (GWLR) model is therefore employed to address this issue.

3.5. Geographically Weighted Logistic Regression (GWLR) Model

After estimating the global logistic regression model as a benchmark, the next step is to apply the Geographically Weighted Logistic Regression (GWLR) model to capture potential local variations in the relationships among variables. This approach allows each region to have distinct regression coefficients, thereby providing a better model fit that may not be adequately explained by the global model.

The spatial parameters in the GWLR model are determined by searching for the optimal bandwidth using the cross-validation (CV) criterion. The evaluation results indicate that the optimal bandwidth is obtained at a value of 8, which yields the minimum CV score and is subsequently used in the estimation of local parameters. Based on the GWLR estimation with the optimal bandwidth of 8, the results of the model goodness-of-fit test are presented in Table 3.

Table 3. Results of the GWLR Model Goodness-of-Fit Test

Statistic	Significant Variables
Log-Likelihood	-3,080
AIC	28,570
AICc	40,569
ENP	11,205

The results of the GWLR model goodness-of-fit test across regencies/municipalities indicate that the model achieves a log-likelihood of −3.080, an AIC value of 28.570, an AICc value of 40.569, and an effective number of parameters (ENP) of 11.205. These results suggest that the model is capable of adequately capturing local variations across regions and provides insight into the contributions of significant variables at the regency/municipality level. Furthermore, the results of the partial significance tests for local parameters are presented for five selected regions in Central Java Province.

Table 4. Results of the Partial Significance Tests of the GWLR Model by Regency

Regency	Significant Variables	Number of Variables
Semarang City	Sanitasi	1

The GWLR estimation results from Table 4 reveal the presence of spatial heterogeneity in the significance of variables across regions. Blora Regency and Semarang City exhibit three significant variables, namely the dependency ratio, adequate housing, and sanitation, whereas Karanganyar and Pati Regencies are significantly influenced only by the dependency ratio and sanitation. In Salatiga City, the dependency ratio emerges as the sole significant variable. Table 5 presents the estimated coefficients of the GWLR model by regency.

Table 5. GWLR Model Results by Regency

Regency	Significant Variables
Semarang City	$\pi_5 - \frac{\exp(-0.125679 + 3.639628x_{61}}{1 + \exp(-0.125679 + 3.639628x_{6\,1}}$

 DOI: https://doi.org/10.52435/jaiit.v8i1.767

3.6. Spatial Variability Test

Based on the results of the Geographically Weighted Logistic Regression (GWLR) modeling, several independent variables exhibit differing levels of significance across regencies. This finding indicates that the relationship between the independent variables and the dependent variable is not spatially homogeneous. Therefore, further testing is required to determine whether the variation in coefficients produced by the GWLR model is statistically significant or merely a result of random fluctuations.

Table 6. Results of the Spatial Variability Test

Variable	F3 Statistic	F Critical	Decision
DepRatio (X1)	3.296619		Local
UMK (X2)	2.873925		Local
Number of Industries (X3)	0.174645	1.965605	Global
TPT (X4)	0.466907		Global
Adequate Housing (X5)	24.121231		Local
Sanitation (X6)	11.854621		Local

The results of the spatial variability test from Table 6 indicate that not all independent variables exert a constant effect across regions. Several variables, namely X1, X2, X5, and X6, exhibit statistically significant spatial variation in their coefficients, suggesting that the influence of these variables on the probability of $P1$ differs across regencies and municipalities. In contrast, variables X3 and X4 show more stable effects and do not demonstrate strong spatial variability. These findings confirm the presence of spatial non-stationarity in the data and highlight the importance of considering further modeling approaches that combine both global and local effects.

3.7. Geographically Weighted Logistic Regression Semiparametric (GWLRS) Model

Based on the indications from the spatial variability test, the analysis was extended using the GWLRS model to capture both global and local effects of the independent variables. The estimation results of the GWLRS model are presented in Table 7 for five selected regions, distinguishing between global (*) and local parameters.

Table 7. Estimation Results of the GWLRS Model

Regency	β_0*	β_1	β_2*	β_3*	β_4*	β_5	β_6
Blora		0,127404	-0.000004			-0,093939	0,111484
Karanganyar		0,097568	-0,000003			-0,116030	0,128213
Tegal City	-1,0978	0,255614	-0,000007	0,000014	0,183556	-0,345305	0,333495
Salatiga City		0,131712	-0,000003			-0,126966	0,123837
Semarang City		0,111225	-0,000003			-0,121272	0,119677

Model adequacy was evaluated by comparing the global logistic regression, GWLR, and GWLRS models using several goodness-of-fit criteria. This comparison aims to assess the extent to which spatial approaches improve model performance relative to the global model, while accounting for differences in model complexity.

Table 8. Results of Model Goodness-of-Fit Evaluation

Model	Log-Likelihood	AIC	AICc	Number of Parameters/ENP
Global Logistic Regression	-10,332	34,664	38,812	6
GWLR	-3,080	28,5699	40,569	11,205
GWLRS	-6,991	38,597	53,697	12,307

Comparison of goodness of fit between the GWLR and GWLRS models in Table 8 shows that the GWLR yields a Log-Likelihood of −3.08 and an AIC of 28.57, whereas the GWLRS produces a Log-Likelihood of −11.21, an AIC of 48.10, and an AICc of 64.91, with an effective number of parameters (ENP) of 12.84, reflecting increased model complexity due to the separation of global and local effects. The higher AIC and ENP values in the GWLRS indicate greater model complexity; therefore, further testing using the test was conducted to evaluate the roles of global and local variables within the model. Subsequently, the simultaneous significance of global parameters in the GWLRS model was assessed using the test (Likelihood Ratio Test), as presented in Table 9.

Table 9. Results of the Simultaneous Significance Test of Global Parameters in the GWLRS Model

Model	Log-Likelihood	df	LR	*p-value*
Null	-24,131	1	-	-
GWLRS	-6,991	12,307	34,28	0,0007

Table 9 presents the results of the partial test for the GWLRS model, where the Log-Likelihood value increases from −24.131 (null model) to −6.991 in the GWLRS model. The Likelihood Ratio (LR) value of 34,28 with a p-value of 0.712 indicates that, partially, the inclusion of variables in the global GWLRS model does not provide a statistically significant improvement in model fit compared to the null model. The subsequent analysis focuses on testing the significance of global parameters, as presented in Table 10.

Table 10. Results of the Global Parameter Significance Test for the GWLRS Model

Variable	Coefficients	SE	Z	*p-value*
Intercept	-0,294118	0,393964	-0,746560	0,455329
UMK	-1,151076	0,592502	-1,942740	0,052048
Number of Industries	0,067602	0,358283	0,188683	0,850341
TPT	0,409590	0,384423	1,065466	0,286665

The partial test results indicate that none of the global variables has a statistically significant effect on the occurrence of P1 at the 5% significance level. This condition motivates further testing of the local parameter significance to evaluate the influence of variables at each district/city level using a 5% significance level. In addition, descriptive statistics of the local parameters are presented in Table 11.

Table 11. Results of the Local Parameter Significance Test for the GWLRS Model

Parameter	Min	Max	Mean	Std	Regencies
Dependency Ratio	0,420891	11,570217	3,336715	2,677463	2
Adequate Housing	-15,808047	-2,371538	-5,951888	2,896447	5
Sanitation	1,867798	17,27255	5,712441	3,553405	3,553405

The table above presents a summary of the local coefficients in the GWLRS model. The Dependency Ratio shows substantial variation, ranging from 0.4209 to 11.5702 with a mean of 3.3367, and is significant in 2 regencies. Adequate Housing has coefficients ranging from −15.8080

 DOI: https://doi.org/10.52435/jaiit.v8i1.767

to −2.3715 (mean −5.9519) and is significant in 5 regencies. Meanwhile, Improved Sanitation ranges from 1.8678 to 17.2723 with a mean of 5.7124, and is significant in 0 regencies. These results indicate the presence of spatial heterogeneity in how each variable influences the dependent outcome, reflecting variations that cannot be fully captured by a global model. Table 12 presents the parameters that are significant in the five selected regions.

Table 12. Local Parameter Significance of the GWLRS Model by Regency

Regency	Significant Local Parameters
Cilacap	(X5) Adequate Housing
Kendal	(X1) DepRatio
Pekalongan City	(X5) Adequate Housing
Salatiga City	(X1) DepRatio
Pekalongan	(X5) Adequate Housing

Table 12 shows that Adequate Housing (X5) emerges as a significant local parameter in Cilacap, Pekalongan, and Pekalongan City, indicating that housing conditions play a notable role in influencing poverty depth (P1) within these areas. Meanwhile, the Dependency Ratio (X1) is significant only in Kendal and Salatiga City, suggesting that its effect on poverty is highly localized and does not extend uniformly across the region. No other variables appear as significant local parameters in the remaining regencies/cities.

3.8. Model Evaluation

Model performance was evaluated by calculating various classification metrics, including accuracy, precision, recall, and F1-score, for each model: Global Logistic Regression, GWLR, and GWLRS. Comparing these metric values provides a quantitative overview of each model's ability to distinguish between the two classes. The results of the evaluation metrics comparison are presented in Table 13.

Table 13. Models Evaluation

Metrics	Accuracy	Precision	Recall	F1-Score
Global Logistic Regression	0,8286	0,8125	0,8125	0,8125
GWLR	1	1	1	1
GWLRS	0,9143	0,8824	0,9375	0,9091

Table 13 presents a comparison of the classification evaluation metrics for the three models: Global Logistic Regression, GWLR, and GWLRS. The Global Logistic Regression model achieves an accuracy of 0.8286, with precision, recall, and F1-score all equal to 0.8125, indicating good—though not yet optimal—predictive performance. The GWLR model attains perfect scores (1.0) across all metrics, demonstrating its strong ability to capture local spatial variation; however, because these results are based on in-sample evaluation, they may be overly optimistic and suggest potential overfitting. The GWLRS model provides a more balanced and realistic classification performance, achieving an accuracy of 0.9143, a precision of 0.8824, a recall of 0.9375, and an F1-score of 0.9091, indicating stronger predictive capability than the global model and better generalization potential than GWLR.

Compared to the reference study, which primarily highlights spatial pattern identification without a detailed assessment of model performance, this study provides clearer empirical evidence of improvement through the GWLRS framework. The global logistic regression model yields an accuracy of 0.8286 with a Pseudo R^2 of 0.5718, while the GWLRS model enhances predictive performance with an accuracy of 0.9143 and balanced precision–recall outcomes. Furthermore, unlike the reference model that treats all covariates uniformly, the GWLRS results reveal substantial spatial variability in local parameters, reflecting the importance of capturing localized determinants of poverty depth (P1).

4. Conclusion

Based on the empirical analysis, several key findings can be concluded. First, the global logistic regression model indicates that the Dependency Ratio, Adequate Housing, and Sanitation variables have a significant effect on the probability of P1 occurrence, whereas UMK, Number of Industries, and Open Unemployment Rate (TPT) are not globally significant. This finding suggests the presence of spatial heterogeneity in the effects of poverty determinants, which cannot be fully captured by a global model.

Second, modeling using GWLR and GWLRS confirms the existence of significant spatial variation in several independent variables, particularly the Dependency Ratio, Adequate Housing, and Sanitation. The differing coefficients across regencies and municipalities indicate that the effects of poverty determinants are local in nature, and the separation of global and local effects in the GWLRS provides a more accurate representation of the spatial patterns of variable influence.

Third, the performance evaluation of the GWLRS model using the confusion matrix and classification metrics demonstrates strong predictive capability, with an accuracy of 91.543%, precision of 88.24%, recall of 93.75%, and F1-score of 90.91%, reflecting the model's ability to consistently capture P1 occurrences across regions.

Overall, these findings underscore the relevance of spatial approaches in analyzing poverty determinants. The use of GWLRS enables simultaneous identification of local and global variable effects, providing a more comprehensive understanding of the spatial variation in the influence of poverty determinants. Consequently, this model offers a robust empirical basis for formulating location-based poverty alleviation policies in Central Java Province.

References

[1] Siti Hartina Daulay and Elmanani Simamora, "PEMODELAN FAKTOR-FAKTOR PENYEBAB KEMISKINAN DI PROVINSI SUMATERA UTARA MENGGUNAKAN METODE GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)," *J. Ris. RUMPUN Mat. DAN ILMU Pengetah. ALAM*, vol. 2, no. 1, pp. 47–60, Jan. 2023, doi: 10.55606/jurrimipa.v2i1.646.
[2] Y. Mansur, "Analisis Perkembangan Penduduk Miskin, Karakteristik Kemiskinan dan Kedalaman Kemiskinan di Indonesia," *J. EMT KITA*, vol. 8, no. 1, pp. 18–31, Jan. 2024, doi: 10.35870/emt.v8i1.1930.
[3] "Profil Kemiskinan di Indonesia September 2024 Provinsi Jawa Tengah," Badan Pusat Statistik Jawa Tengah. [Online]. Available: https://jateng.bps.go.id/id/pressrelease/2025/01/15/1547/profil-kemiskinan-di-indonesia-september-2024-provinsi-jawa-tengah.html
[4] F. A. Guampe, A. S. Walenta, and F. B. Kawani, "Pengaruh Pertumbuhan Ekonomi Dan Pengangguran Terbuka Terhadap Kemiskinan Di Indonesia Tahun 2001-2021," *JPEK J. Pendidik. Ekon. Dan Kewirausahaan*, vol. 6, no. 1, pp. 92–102, Jun. 2022, doi: 10.29408/jpek.v6i1.5536.
[5] A. I. Nurfarizki, B. Setiawan, F. R. Nugroho, W. Rizky, and M. R. W. Azka, "Pengaruh Indeks Pembangunan Manusia, Pertumbuhan Ekonomi dan Lama Sekolah terhadap Kemiskinan di Provinsi Jawa Tengah," 2024.
[6] K. Sari'ah, "Pengaruh Kemiskinan dan Pertumbuhan Ekonomi Terhadap Indeks Pembangunan Manusia di Provinsi Jawa Tengah dalam Prespektif Ekonomi Islam Ibnu Khaldun Tahun 2019-2023," *J. Manag.*, vol. 3, 2024.
[7] A. Anwar, "SPATIAL ANALYSIS OF REGIONAL POVERTY IN CENTRAL JAVA INDONESIA," *J. Din. Ekon. Pembang.*, vol. 5, no. 1, pp. 36–55, Jul. 2022, doi: 10.14710/jdep.5.1.36-55.
[8] T. Saifudin, L. S. Panjaitan, S. Falasifah, and Yan Dwi, "Application of Geographically Weighted Logistic Regression in Modeling the Human Development Index in East Java," *J. Apl. Stat. Komputasi Stat.*, vol. 16, no. 1, pp. 43–57, Jun. 2024, doi: 10.34123/jurnalasks.v16i1.587.

[9] F. Fitriatusakiah, A. K. Jaya, and L. P. Talangko, "Pemodelan Semiparametrik Geographical Weighted Logistic Regression pada Data Kemiskinan di Provinsi Sulawesi Selatan Tahun 2017," *ESTIMASI J. Stat. Its Appl.*, pp. 105–114, Jul. 2021, doi: 10.20956/ejsa.v2i2.11309.
[10] W. Kang and T. M. Oshan, "Scale and correlation in multiscale geographically weighted regression (MGWR)," *J. Geogr. Syst.*, vol. 27, no. 3, pp. 399–424, Jul. 2025, doi: 10.1007/s10109-025-00468-1.
[11] A. H. Azizah, N. Nurjannah, A. A. R. Fernandes, and R. Hamdan, "GEOGRAPHICALLY WEIGHTED PANEL LOGISTIC REGRESSION SEMIPARAMETRIC MODELING ON POVERTY PROBLEM," *MEDIA Stat.*, vol. 16, no. 1, pp. 47–58, Nov. 2023, doi: 10.14710/medstat.16.1.47-58.
[12] N. Nurhasanah, W. Widiarti, D. E. Nurvazly, and M. Usman, "Penerapan Model Geographically Weighted Logistic Regression dengan Fungsi Pembobot Adaptive Gaussian Kernel pada Data Kemiskinan," *Jambura J. Math.*, vol. 6, no. 2, pp. 204–211, Aug. 2024, doi: 10.37905/jjom.v6i2.26504.
[13] M. Irwansyah, R. Ruliana, and M. K. Aidid, "Analisis Regresi Balanced Panel dengan Komponen Galat Dua Arah pada Kasus Melek Huruf Masyarakat di Provinsi NTB," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 3, no. 1, p. 10, Mar. 2021, doi: 10.35580/variansiunm14644.
[14] A. Kalnins and K. Praitis Hill, "The VIF Score. What is it Good For? Absolutely Nothing," *Organ. Res. Methods*, vol. 28, no. 1, pp. 58–75, Jan. 2025, doi: 10.1177/10944281231216381.
[15] H. Yasin, B. Warsito, and A. R. Hakim, *REGRESI SPASIAL (Aplikasi dengan R)*. in 1. WADE Group, 2020.
[16] S.-L. Shen, J.-L. Cui, and X.-Q. Wu, "A simple test for spatial heteroscedasticity in spatially varying coefficient models," *J. Stat. Comput. Simul.*, vol. 91, no. 8, pp. 1580–1592, May 2021, doi: 10.1080/00949655.2020.1862112.

This page intentionally left blank

DOI: https://doi.org/10.52435/jaiit.v8i1.767