JURNAL STATISTIKA DAN APLIKASINYA

e-ISSN 2620 – 8369
VOLUME X (ISSUE X) MONTH YEAR: XX-XX
https://doi.org/10.21009/JSA.081XX

ANALYSIS OF THE PERCENTAGE OF POOR POPULATION BY DISTRICT/CITY IN EAST JAVA PROVINCE IN 2022 USING GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

Ammar Hanafi [1*], Bryan Jonathan[2], Renata Shaula Alfino Ritonga[3]

[1,2,3]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia
Kampus Baru UI, Kota Depok, 16424, Indonesia

Corresponding author's e-mail: * **ammar.hanaafi@ui.ac.id**

Article History:

Received: date, month year
Revised: date, month year
Accepted: date, month year
Published: date, month year
Available online.

Keywords:

geographically weighted regression; poverty; spatial analysis; East Java; socio-economic factors

ABSTRACT

Poverty remains a major socio-economic challenge in East Java Province, Indonesia, with notable disparities across its districts and cities. This study aims to analyze the spatial relationship between the percentage of poor population and four independent variables: open unemployment rate, access to proper drinking water, expected years of schooling, and cigarette consumption among individuals aged 30–39. Using data from BPS Jawa Timur for 2022, a Geographically Weighted Regression (GWR) approach was applied to account for spatial heterogeneity. The analysis revealed that all four variables significantly influence poverty rates across the region, with variations in their spatial impacts. GWR with a Bisquare kernel provided the best model fit, outperforming global linear regression models based on lower AIC values and higher R-squared values. The study highlights the importance of localized poverty alleviation strategies, emphasizing improvements in education access, basic infrastructure, and public health interventions. Despite providing valuable spatial insights, the study is limited to the selected variables and a single year's data, suggesting that future research could explore additional socio-economic factors and temporal dynamics. This research contributes original empirical evidence on the spatial determinants of poverty in East Java, supporting more targeted and effective policy interventions.

1. Introduction

Poverty remains a persistent and complex socio-economic issue that affects the quality of life and economic stability of communities worldwide. In Indonesia, East Java Province represents a significant case, as it is one of the regions with notable disparities in poverty levels across its regencies and cities. The persistent poverty in East Java not only hinders regional development but also reflects broader national challenges in achieving equitable growth and improving human welfare. Given the multi-dimensional nature of poverty, identifying and analyzing the socio-economic factors that influence its prevalence is crucial for policymakers and development practitioners.

Previous studies have highlighted several key factors that are often correlated with poverty, including unemployment rates, access to basic infrastructure, education levels, and health-related behaviors. In this study, we focus on four independent variables: the Open Unemployment Rate (Tingkat Pengangguran Terbuka, TPT) [1], Access to Proper Drinking Water [2], Expected Years of Schooling [3], and the Average Number of Cigarettes Consumed by individuals aged 30–39 [4]. Each of these variables is selected based on its significant impact on the economic and social dynamics of a

Journal homepage: https://journal.unj.ac.id/unj/index.php/statistika
Journal e-mail: jsa@unj.ac.id
Research Article · Open Access

population. Understanding how these variables relate spatially to poverty can provide a comprehensive view of the underlying issues contributing to poverty in East Java.

The primary objective of this research is to analyze the spatial relationship between poverty rates and the selected independent variables across the regencies and cities of East Java. By employing spatial statistical methods, we aim to capture not only the magnitude of these relationships but also their spatial patterns and variations. This approach enables the identification of regional clusters and anomalies, offering insights that conventional non-spatial analyses might overlook. Through this analysis, we seek to contribute valuable empirical evidence that can inform targeted poverty alleviation programs, with strategies tailored to the specific characteristics of each region.

To achieve this, data were collected from official publications of Badan Pusat Statistik (BPS) Jawa Timur for the year 2022. The analysis involves exploratory spatial data analysis (ESDA) techniques, including mapping, correlation analysis, and spatial regression models. By integrating spatial considerations into the analysis, this study addresses both the social and geographical dimensions of poverty, providing a more nuanced understanding of how socio-economic variables interact across space to influence poverty outcomes in East Java.

2. Methods

Material and Data

This study utilizes data from the Central Bureau of Statistics (BPS) covering 38 districts/cities in East Java Province for the year 2022. The dependent variable is the percentage of the poor population, while the independent variables are open unemployment rate, percentage of households receiving PKH, mean years of schooling, and population density. Spatial data is obtained in the form of district/city boundary shapefiles, which include coordinate information used in spatial analysis.

Research Method

Multiple Linear Regression

Multiple Linear Regression (MLR) models the linear relationship between a dependent variable and two or more independent variables. The general form of the model is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \tag{1}$$

where y_i is the dependent variable, x_{ij} are the independent variables, β_j are the regression coefficients, and ε_i is the error term.

The parameters are estimated using the Ordinary Least Squares (OLS) method, which minimizes the Sum of Squared Errors (SSE). Significance of parameters is tested using partial t-tests and F-tests for joint significance, while diagnostic tests are used to evaluate assumptions including multicollinearity, using VIF; normality, using Kolmogorov-Smirnov test; autocorrelation using Durbin-Watson test; and homoscedasticity using scatterplot inspection.

Spatial Heterogeneity Detection

To determine whether spatial heterogeneity is present, the Breusch-Pagan test is used with the following hypotheses:

H_0: $No\ spatial\ heterogeneity$

H_1: $Spatial\ heterogeneity\ exists$

A significant result with p-value < 0.05 supports the presence of spatial heterogeneity, reinforcing the appropriateness of a geographically weighted approach.

Geographically Weighted Regression (GWR)

Geographically Weighted Regression (GWR) is a local form of linear regression used to model spatially varying relationships. Unlike traditional regression models that assume the relationship between dependent and independent variables is constant across space, GWR allows regression coefficients to vary by location, which is suitable for data with spatial heterogeneity. The GWR model for observation *i* located at coordinates $\left(u_i, v_i\right)$ is defined as:

$$Y_i = \beta_0\left(u_i, v_i\right) + \sum_{k=1}^{p} \beta_k\left(u_i, v_i\right)X_{ik} + \varepsilon_i,\ i = 1,2, \ldots, n \tag{2}$$

Where:

Y_i: value of the dependent variable at location *i*

X_{ik}: value of the *k*-th independent variable at location *i*

$\left(u_i, v_i\right)$: spatial coordinates of location *i*

$\beta_k\left(u_i, v_i\right)$: local regression coefficients

ε_i: random error term at location *i*

Each observation has its own local regression equation. Estimation is performed using Weighted Least Squares (WLS), which assigns higher weights to nearby observations and lower weights to distant ones.

Estimation of GWR Parameters

The estimated coefficient vector at location *i*, $\hat{\beta}\left(u_i, v_i\right)$, is obtained by minimizing the Weighted Sum of Squared Errors (JKGT):

$$JKGT = \sum_{i=1}^{n} w_i\left(u_i, v_i\right)\left[y_i - \beta_0\left(u_i, v_i\right) - \sum_{k=1}^{p} \beta_k\left(u_i, v_i\right)x_{ik}\right]^2 \tag{3}$$

In matrix form, the solution is:

$$\hat{\beta}\left(u_i, v_i\right) = \left(X^T W\left(u_i, v_i\right)X\right)^{-1} X^T W\left(u_i, v_i\right)Y \tag{4}$$

Where:

X: matrix of independent variables (including intercept)

Y: vector of observed dependent variable values

$W\left(u_i, v_i\right)$: spatial weight matrix for location i, typically a diagonal matrix

Spatial Weighting Functions

Weights are calculated using kernel functions, which define how proximity affects the influence of nearby data points [5]. Common kernels include:

a. Gaussian Kernel

$$w_{ij} = exp\left(-\frac{d_{ij}^2}{2h^2}\right) \tag{5}$$

b. Bisquare Kernel (if $d_{ij} < h$, else 0)

$$w_{ij} = \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2 \tag{6}$$

c. Tricube Kernel (if $d_{ij} < h$, else 0):

$$w_{ij} = \left(1 - \left(\frac{d_{ij}}{h}\right)^3\right)^3 \tag{7}$$

Where:

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

h: bandwidth parameter controlling the influence range

Bandwidth Selection

Bandwidth h is crucial to model performance and is typically selected by minimizing Cross-Validation (CV) or Akaike Information Criterion (AIC). The formula for Cross-Validation score is expressed as:

$$CV = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_{\neq i}\right)^2 \tag{8}$$

Where $\widehat{Y}_{\neq i}$ is the fitted value for Y_i obtained by excluding the i-th observation during model fitting.

Local Parameter Significance Testing

To test whether a local coefficient $\beta_k\left(u_i, v_i\right)$ is significantly different from zero at location i, a t-statistic is computed as:

$$t_k\left(u_i, v_i\right) = \frac{\widehat{\beta}_k\left(u_i, v_i\right)}{SE\left(\widehat{\beta}_k\left(u_i, v_i\right)\right)} \tag{9}$$

Where the standard error is given by:

$$SE\left(\widehat{\beta}_K\left(u_i, v_i\right)\right) = \sqrt{C_{kk} \cdot \widehat{\sigma}^2} \tag{10}$$

With:

$$C = \left(X^T W\left(u_i, v_i\right) X\right)^{-1} \tag{11}$$

$\widehat{\sigma}^2$: estimated local variance of residuals

Model Evaluation

Two key criteria are used to assess model performance:

a. Akaike Information Criterion (AIC)

AIC is a measure of the relative quality of a statistical model to another model. Lower values of AIC indicate a better model, adjusting for model complexity. The AIC score of a model is calculated with the following formula.

$$AIC = e^{\frac{2k}{n}} \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n} \quad (12)$$

Where:

k: number of parameters

n: number of observations

$\hat{\varepsilon}_i$: error at observation *i*

b. Coefficient of Determination (R^2)

The coefficient of determination is a measure of the variation in the dependent variable that can be predicted by the independent variable. Higher values of R^2 indicate better explanatory power. R^2 is calculated with the following formula:

$$R^2 = 1 - \frac{JKG}{JKT} \quad (13)$$

Where:

JKG : sum of squares error ($JKG = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$) (14)

JKT: total sum of squares ($JKT = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2$) (15)

y_i: *i*-th observation data

$\bar{y}$: mean of observation data

The GWR model with the best combination of low AIC and high R^2 is selected as the final model for interpretation.

3. Results

3.1 Data

Table 1. Data

Loc	Lat	Long	Y	X1	X2	X3	X4
Pacitan	-8,204614	111,51370	4,76	6,39	98,54	14,43	68,28
Ponorogo	-7,867827	112,7344	4,72	7,62	95,66	14,83	70,53
Trenggalek	-8,05	112,5239	3,79	8,43	99,12	14,4	70,97
…	…	…	…	…	…	…	…
Madiun	-7,629714	111,51370	4,76	6,39	98,54	14,43	68,28
Surabaya	-7,289166	112,7344	4,72	7,62	95,66	14,83	70,53

Batu	-7,8671	112,5239	3,79	8,43	99,12	14,4	70,97

Source: BPS Jawa Timur

The data used as seen in **table 1** consists of the locations, latitude, longitude, percentage of poor population as Y, open unemployment rate as X1, access of decent drinking water as X2, school length expectation as X3, and average amount of cigarettes used by people aged 30-39 as X4.

3.2 Analysis of Global Regression

An estimate of the parameters of global regression is done using the Ordinary Least Square method which minimizes the sum square error. The results are achieved using R Studio and are shown in **table 2**.

Table 2. Result of Ordinary Least Square

Coefficients	Estimate	Standard Error	t-value	Pr(> | t |)
Intercept	48.20624	8.01505	6.014	9.25e-07
X1	-0.86716	0.19862	-4.366	0.000118
X2	-0.20367	0.06923	-2.942	0.005925
X3	-1.86688	0.40005	-4.667	4.91e-05
X4	0.14256	0.02385	5.976	1.03e-06

Source: RStudio

According to RStudio, this model can explain 80.83% variance of the data and all four of the variables are significant towards the percentage of poor population. As such we achieve the best global regression model:

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

$$\hat{y} = 48.20624 - 0.86716x_1 - 0.20367x_2 - 1.86688x_3 + 0.14256x_4$$

As explained before, this model can explain 80.83% variance of the data and has a SSE of 1.98. For F-Statistic Global Regression, the achieved p-value is 2.086e-11 so H0 is rejected as such we conclude that at least one independent variable is significant towards the dependent variable. Next, we check the assumptions.

3.3 Assumption Test

Table 3. Result of Assumption Test

Assumption	Statistical Test	Result	Interpretation
Residual Normality	Shapiro-Wilk Test	W = 0.99191 p-value = 0.9934	Residuals are normally distributed (p > 0.05)
Multicollinearity	Variance Inflation Factor (VIF)	VIF X1 = 1.174 VIF X2 = 1.053 VIF X3 = 1.221 VIF X4 = 1.041	No multicollinearity detected (VIF < 5)

Residual Autocorrelation	Durbin-Watson Test	DW = 2.384 p-value = 0.8331	No autocorrelation detected ($p > 0.05$)
Homoscedasticity	Breusch-Pagan Test	BP = 12.169 df = 4 p-value = 0.01614	Indication of heteroscedasticity ($p < 0.05$)

Source: RStudio

As seen in **table 3**, all assumptions are as we hoped and we can proceed with Geographically Weighted Regression.

3.4. Geographically Weighted Regression

3.4.1. Selecting Independent Variable

Pseudo-stepwise method is done to select the independent variables to be used for Geographically Weighted Regression. The selection models are displayed on **Figure 1 to 6**, Gaussian kernel, Bisquare kernel, and Tricube kernel are used.

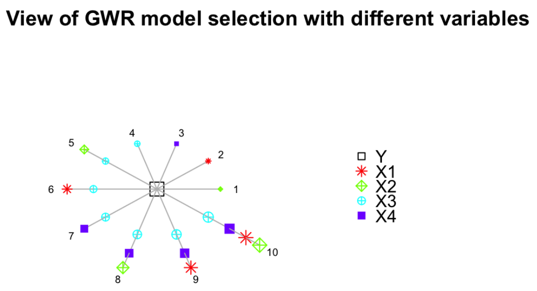

Figure 1

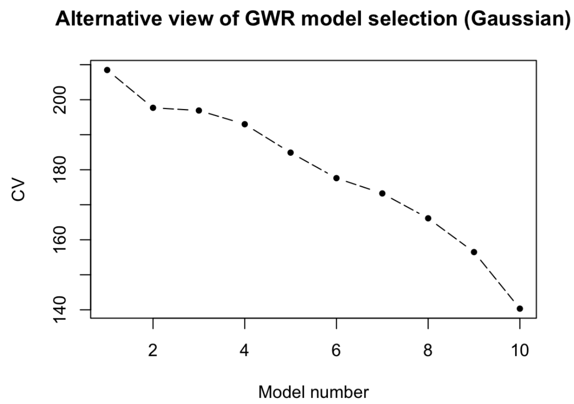

Figure 2

According to **Figure 1**, pseudo-stepwise selection is conducted through 10 steps, starting from utilizing one independent variable and progressively adding more variables up to using all four independent variables in the model. According to **Figure 2**, the CV value consistently decreases as more variables are included, indicating that adding additional independent variables improves the predictive performance of the Geographically Weighted Regression (GWR) model. As a result, all four independent variables are utilized in the final GWR model with the Gaussian kernel as the weight function.

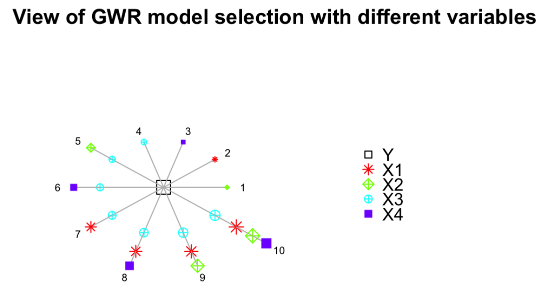

Figure 3

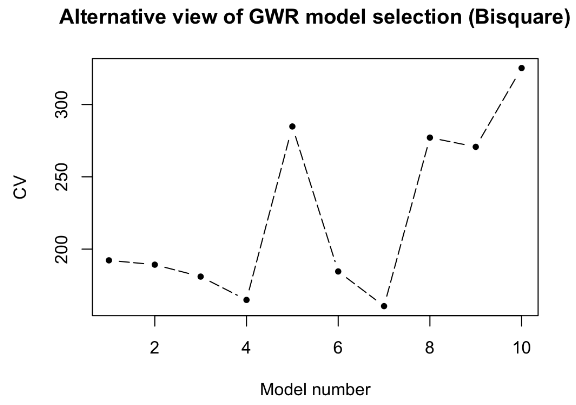

Figure 4

According to **Figure 3**, pseudo-stepwise selection is conducted by 10 steps, starting from utilizing one independent variable and progressively adding variables up to the full set of four independent variables. According to **Figure 4**, the CV value fluctuates across the models, indicating that the model's predictive accuracy changes depending on the combination of variables. The model with the lowest CV value will be selected, which in this case is Model 7, meaning that the optimal GWR model under the Bisquare kernel uses a specific subset of the available variables rather than including all independent variables.

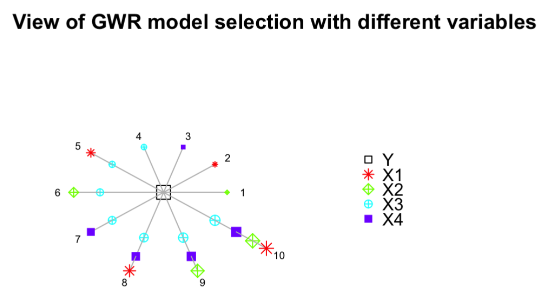

Figure 5

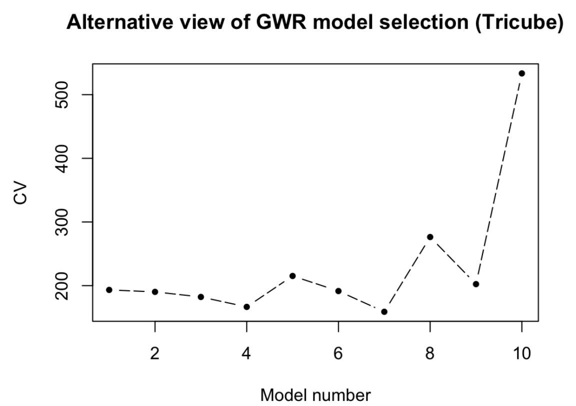

Figure 6

According to **Figure 5**, pseudo-stepwise selection is conducted through 10 steps, starting from utilizing one independent variable and progressively adding more variables until all four independent variables are included in the model. According to **Figure 6**, the CV value fluctuates across the models, and a sharp increase is observed when all variables are used. This indicates that including all independent variables in the Geographically Weighted Regression (GWR) model with Tricube kernel as the weight function results in poorer predictive performance. Therefore, not all independent variables should be utilized in the final GWR model under the Tricube kernel setting.

3.4.2. Optimal Bandwidth Selection

Optimal Bandwidth Selection will be done to select the bandwidth as the parameter of the weight function:

1. Gaussian Kernel Weight Function

   Optimal Bandwidth Selection will be done to select the bandwidth as the parameter of the Gaussian Kernel weight function. RStudio is used and the optimal bandwidth for Gaussian Kernel function is achieved with 6.643368, hence the Gaussian Kernel function is as such:

$$K_G(d_{ij}) = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{6.643368}\right)^2\right]$$

2. Bisquare Kernel Weight Function

   Optimal Bandwidth Selection will be done to select the bandwidth as the parameter of the Bisquare Kernel weight function. RStudio is used and the optimal bandwidth for Bisquare Kernel function is achieved with 6.644454, hence the Bisquare Kernel function is as such:

$$K_G(d_{ij}) = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{6.644454}\right)^2\right]^2$$

3. Tricube Kernel Weight Function

   Optimal Bandwidth Selection will be done to select the bandwidth as the parameter of the Tricube Kernel weight function. RStudio is used and the optimal bandwidth for Tricube Kernel function is achieved with 6.644414, hence the Tricube Kernel function is as such:

$$K_G(d_{ij}) = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{6.644414}\right)^3\right]^3$$

3.4.3. Best Model Selection

The Akaike Information Criterion (AIC) scores and R-Squared values obtained from the global linear regression model, the Geographically Weighted Regression model with Gaussian kernel weighting function, the Geographically Weighted Regression model with Bisquare kernel weighting function, and the Geographically Weighted Regression model with Tricube kernel weighting function are shown in **Table 4**.

Table 4. Best Model Selection according to R-Squared and AIC

Model	AIC	R-Squared
Global Linear Regression	166.3807	0.8083126
GWR with Gaussian Kernel	158.9220	0.8114638
GWR with Bisquare Kernel	157.2626	0.8225263

GWR with Tricube Kernel	157.9161	0.8181444

Source: RStudio

Based on **Table 4**, it can be observed that the AIC score for the global linear regression model is higher and shows a noticeable difference compared to the AIC scores of the GWR models. The lower R-Squared value of the global linear regression model also indicates that it is less effective in modeling this case. Thus, it can be concluded that the Geographically Weighted Regression (GWR) models provide a better fit for this case compared to the global linear regression model.

Among the GWR models, the model with the Bisquare kernel function achieves the lowest AIC score and the highest R-Squared value compared to the Gaussian and Tricube kernel models. Therefore, it can be concluded that the Geographically Weighted Regression (GWR) model with the Bisquare kernel is the most suitable model for this case. Accordingly, the GWR model with the Bisquare kernel will be used to analyze the data in this study.

3.5. Results

3.5.1. Distribution of Significant Variables

The following figure presents the spatial distribution of significant variables, illustrating which variables have a statistically significant influence on population density.

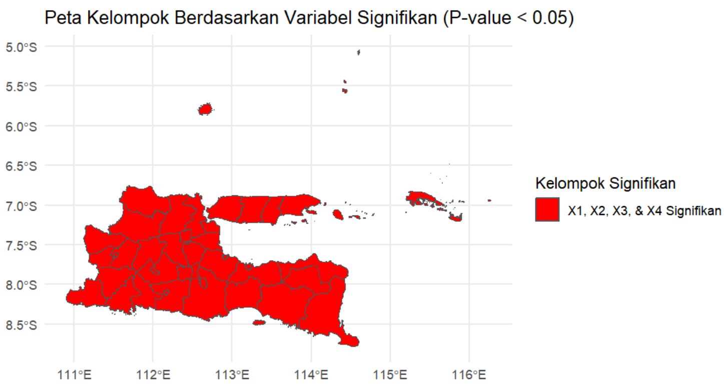

Figure 7

Figure 7 shows the spatial distribution of significant variables based on a significance level of 5% (p-value < 0.05). It can be observed that across the entire East Java (Jawa Timur) region, all four independent variables (X1, X2, X3, and X4) are statistically significant in influencing the dependent variable. This is indicated by the uniform red color across the map, which represents areas where X1, X2, X3, and X4 simultaneously have significant effects. The result suggests that the relationships between the predictors and the dependent variable are consistently strong throughout the region, with no areas showing non-significance.

3.5.2. Spatial Variation of Parameter Estimates β_i

The following section explains the spatial variation of the parameter estimates β for each explanatory variable based on the results of the Geographically Weighted Regression (GWR) model using the Bisquare kernel weighting function.

- **Spatial Variation of Parameter Estimates** β_1

The following figure presents the spatial distribution of the parameter estimate X_1, illustrating the influence of the open unemployment rate on the percentage of the poor population.

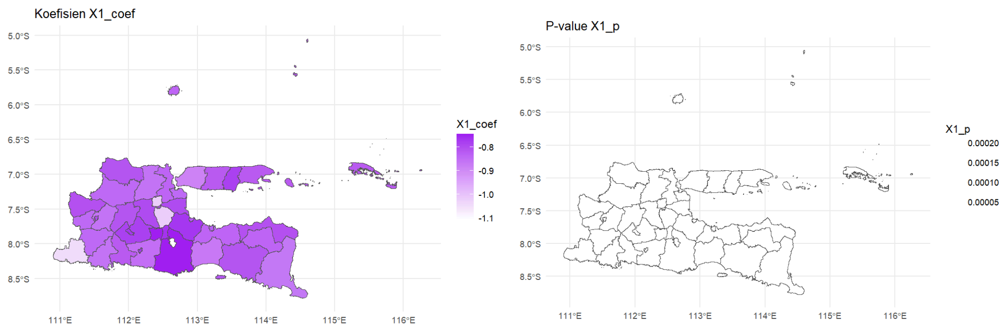

Figure 8. Coefficient Map and P Value Map of X1

The purple color on the coefficient map of X_1 (open unemployment rate) indicates a negative relationship between the open unemployment rate and the percentage of the poor population. This suggests that an increase in the open unemployment rate is associated with a decrease in the percentage of the poor population. Regions with darker purple shades, such as Kabupaten Malang, Kabupaten Jember, and Kabupaten Kediri, demonstrate a stronger negative influence of open unemployment on the poor population percentage. Based on the p-value map, all areas are shown in white, indicating that the relationship between the open unemployment rate and the percentage of poor population is statistically significant across the entire East Java region.

- **Spatial Variation of Parameter Estimates** β_2

The following figure presents the spatial distribution of the parameter estimate X_2 , illustrating the influence of the access of decent drinking water on the percentage of the poor population.

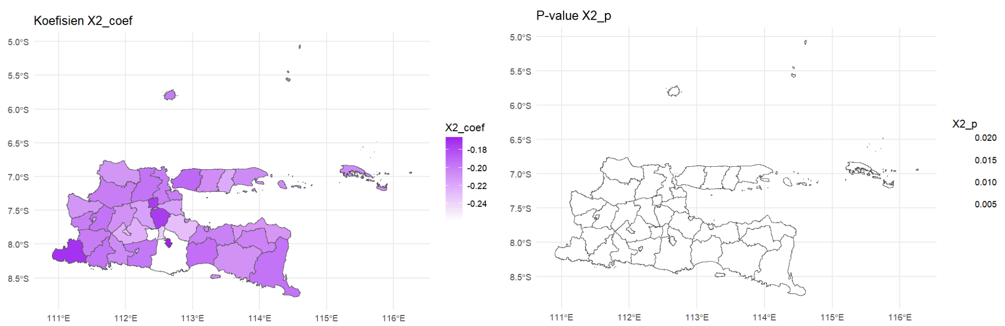

Figure 9. Coefficient Map and P Value Map of X2

The purple color on the coefficient map of X_2 (access to decent drinking water) indicates a negative relationship between access to decent drinking water and the percentage of the poor population. This suggests that an increase in access to decent drinking water is associated with a decrease in the percentage of poor population. Regions with darker purple shades, such as Kabupaten Pacitan, Kabupaten Lumajang, and Kabupaten Mojokerto, demonstrate a stronger negative influence of access to drinking water on the poor population percentage. Based on the p-value map, all areas are shown in white, indicating that the relationship between access to decent drinking water and the percentage of the poor population is statistically significant across the entire East Java region.

- **Spatial Variation of Parameter Estimates** β_3

The following figure presents the spatial distribution of the parameter estimate X_3 , illustrating the influence of the school length expectation on the percentage of the poor population.

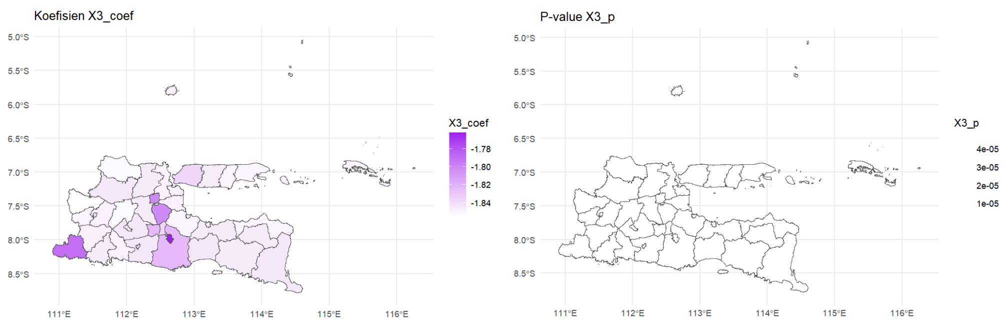

Figure 10. Coefficient Map and P Value Map of X3

The purple color on the coefficient map of X_3 (school length expectation) indicates a negative relationship between school length expectation and the percentage of the poor population. This suggests that an increase in the expected length of schooling is associated with a decrease in the percentage of the poor population. Regions with darker purple shades, such as Kabupaten Pacitan, Kota Malang, and Kabupaten Mojokerto, demonstrate a stronger negative influence of school length expectation on the poor population percentage. Based on the p-value map, all areas are shown in white, indicating that the relationship between school length expectation and the percentage of the poor population is statistically significant across the entire East Java region.

- **Spatial Variation of Parameter Estimates** β_4

The following figure presents the spatial distribution of the parameter estimate X_4 , illustrating the influence of average amount of cigarettes used by people aged 30-39 on the percentage of the poor population.

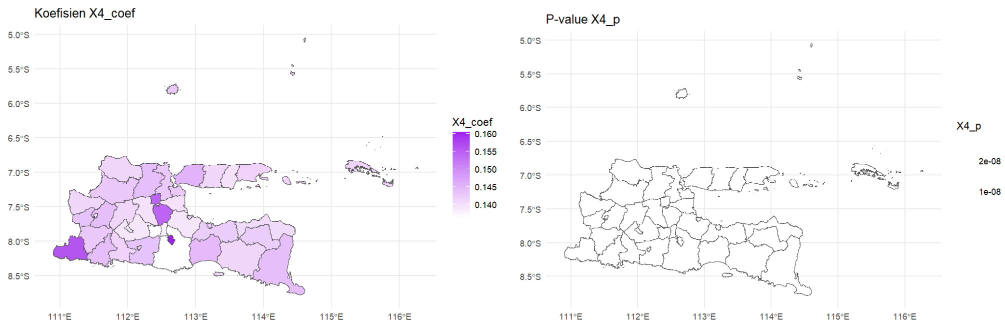

Figure 11. Coefficient Map and P Value Map of X4

The purple color on the coefficient map of X_4 (average amount of cigarettes used by people aged 30–39) indicates a positive relationship between cigarette consumption and the percentage of the poor population. This suggests that an increase in the average amount of cigarette use among individuals aged 30–39 is associated with an increase in the percentage of poor population. Regions with darker purple shades, such as Kabupaten Pacitan, Kabupaten Mojokerto, and Kabupaten Jombang, demonstrate a stronger positive influence of cigarette consumption on the poor population percentage. Based on the p-value map, all areas are shown in white, indicating that the relationship between cigarette consumption and the percentage of the poor population is statistically significant across the entire East Java region.

3.5.3. Spatial Variation of Residuals

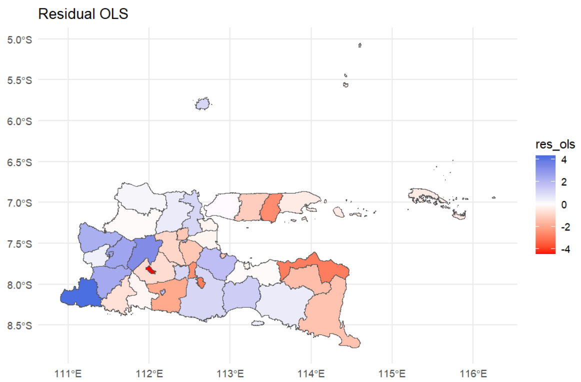

Figure 12. Map of Residual OLS

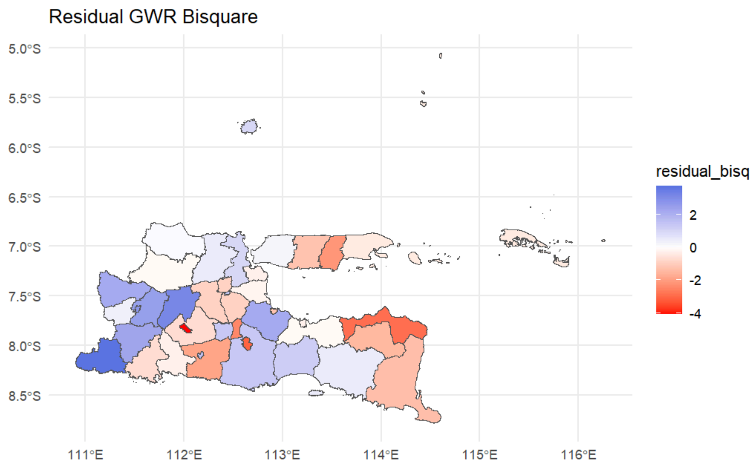

Figure 13. Map of Residual GWR Bisquare

Table 5.

Model	Min	Max
Model Global	-4.479115	4.326362
Model GWR Gaussian	-4.063486	3.712704

Source: RStudio

Figures 12 and 13 show the residual maps for the global linear regression model (OLS) and the Geographically Weighted Regression (GWR) model with the Bisquare kernel, respectively. In both maps, the blue color represents positive residuals, while the red color represents negative residuals. Based on the minimum and maximum residual values, the residuals in the global linear regression model range from -4.479115 to 4.326362, while the residuals in the GWR Bisquare model have a narrower range, from -4.063486 to 3.712704.

This indicates that the GWR Bisquare model produces residuals with smaller prediction errors compared to the global linear regression model. Therefore, spatial analysis using the GWR Bisquare model is better suited to capture local variations in the data than the global regression model.

4. Discussions

The results of the analysis reveal notable spatial patterns in the relationship between poverty and the selected socio-economic factors in East Java. The relationship between the Open Unemployment Rate (TPT) and poverty, for instance, shows a clear spatial clustering effect in several regencies. Areas with high unemployment rates also tend to exhibit higher poverty levels, particularly in urban centers where job opportunities may be more abundant but also highly competitive. This finding aligns with existing literature that suggests unemployment is one of the most direct drivers of poverty, particularly in regions experiencing economic transition [1].

Access to proper drinking water, another key variable, demonstrates a less straightforward, but still significant, spatial relationship with poverty. In regions where access to clean water is limited, poverty levels tend to be higher, especially in rural areas. These findings are consistent with studies indicating that inadequate access to basic services, such as clean water, exacerbates the effects of poverty and reduces opportunities for economic mobility. Rural communities, in particular, face compounded challenges in securing basic needs, which reinforces cycles of poverty and limits long-term economic development [2]. This spatial analysis suggests that improving access to basic services could be a critical intervention in poverty alleviation programs.

The Expected Years of Schooling (HLS) also show a strong association with poverty rates, particularly in more developed regencies where education levels tend to be higher. These regions exhibit lower poverty rates, suggesting that education plays a critical role in reducing economic disparities. This result is in line with global studies which assert that education is one of the most effective tools for breaking the cycle of poverty, as it equips individuals with the skills necessary to participate in the labor market and increases their earning potential [3]. The spatial pattern observed in East Java further underscores the importance of investing in education infrastructure to reduce regional disparities in poverty levels.

Lastly, the number of cigarettes consumed by individuals aged 30-39 appears to have a less direct but still notable correlation with poverty. Areas with higher rates of smoking tend to have higher poverty levels, possibly reflecting broader social determinants of health and well-being. Smoking has been linked to lower productivity and higher healthcare costs, both of which contribute to the perpetuation of poverty. While this finding is consistent with some public health literature [4], it suggests the need for integrated poverty reduction policies that also address health behaviors and their long-term economic consequences.

Overall, the results of this study provide valuable insights into the complex and multifaceted nature of poverty in East Java. The spatial analysis not only confirms the significant roles of unemployment, education, and access to basic services but also highlights the need for region-specific interventions that consider local socio-economic conditions. Further research could explore additional variables, such as healthcare access and economic diversification, to further refine the understanding of poverty dynamics in this region.

5. Conclusion

This study aimed to analyze the spatial relationship between the percentage of poor population and four independent variables which are open unemployment rate, access to decent drinking water, school length expectation, and cigarette consumption among individuals aged 30–39 across districts and cities in East Java Province using the Geographically Weighted Regression (GWR) method. Based on the results, it was found that all four variables significantly influenced the percentage of poor population throughout the region, as indicated by p-values less than 0.05 across all locations.

The model comparison showed that GWR models outperformed the global linear regression model, with GWR models achieving lower AIC values and higher R-Squared values. Among the GWR models, the Bisquare kernel produced the best performance, achieving an AIC of 157.2626 and an R-Squared value of 0.8225263, compared to the global model which had an AIC of 166.3807 and an R-Squared value of 0.8083126. Therefore, GWR with the Bisquare kernel was selected as the final model for interpreting spatial patterns.

The analysis of the spatial variation of coefficients revealed that the open unemployment rate, access to decent drinking water, and school length expectation each showed a negative relationship with the percentage of poor population, meaning improvements in these variables are associated with lower poverty levels. Meanwhile, cigarette consumption exhibited a positive relationship, suggesting higher smoking rates among adults aged 30–39 are linked to higher poverty rates. Furthermore, the residual analysis showed that the GWR Bisquare model produced a narrower residual range (from -4.0635 to 3.7127) compared to the global model (from -4.4791 to 4.3264), indicating better predictive accuracy and capturing local variation more effectively.

In conclusion, the research objectives were successfully achieved. The findings highlight the importance of addressing unemployment, improving access to basic needs such as clean water, promoting education, and tackling public health issues such as smoking in efforts to alleviate poverty in East Java. Future research is encouraged to incorporate additional socio-economic and demographic factors, as well as temporal analysis, to provide a more comprehensive understanding of poverty dynamics in the region.

6. References

[1] Badan Pusat Statistik Jawa Timur, "Tingkat Pengangguran Terbuka (TPT) menurut Kabupaten/Kota," BPS Jawa Timur, 2022. [Online]. Available: https://jatim.bps.go.id/id/statistics-table/2/NTQjMg==/tingkat-pengangguran-terbuka-tpt-menurut-kabupaten-kota.html [Accessed: 27 April 2025].

[2] Badan Pusat Statistik Jawa Timur, "Persentase Rumah Tangga Menurut Kabupaten/Kota Sumber Air Minum Bersih dan Akses Air Minum Layak," BPS Jawa Timur, 2022. [Online]. Available: https://jatim.bps.go.id/id/statistics-table/1/MjkwMCMx/persentase-rumah-tangga-menurut-kabupaten-kota-sumber-air-minum-bersih-dan-akses-air-minum-layak-2022.html [Accessed: 27 April 2025].

[3] Badan Pusat Statistik Jawa Timur, "Harapan Lama Sekolah," BPS Jawa Timur, 2022. [Online]. Available: https://jatim.bps.go.id/id/statistics-table/2/NDY4IzI=/harapan-lama-sekolah.html [Accessed: 27 April 2025].

[4] Badan Pusat Statistik Jawa Timur, "Rata-rata Jumlah Batang Rokok per Minggu yang Dihisap Penduduk Usia 5 Tahun ke Atas yang Merokok Tembakau Sebulan Terakhir di Jawa Timur Menurut Kabupaten/Kota Kelompok Umur," BPS Jawa Timur, 2022. [Online]. Available: https://jatim.bps.go.id/id/statistics-table/1/Mjc4MiMx/rata-rata-jumlah-batang-rokok-per-minggu-yang-dihisap-penduduk-usia-5-tahun-ke-atas-yang-merokok-tembakau-sebulan-terakhir-di-jawa-timur-menurut-kabupaten-kota-kelompok-umur-2022.html [Accessed: 27 April 2025].

[5] Lutfiani, N., Sugiman, S., & Mariani, S. "Pemodelan Geographically Weighted Regression (GWR) dengan Fungsi Pembobot Kernel Gaussian dan Bi-Square." *UNNES Journal of Mathematics*, 8(1), pp. 82-91, 2019.