E-ISSN 2621–086X

Indonesian Journal of Applied Statistics

Vol. 7, No. 2, pp. 177-194, November 2024
Journal homepage: https://jurnal.uns.ac.id/ijas/index
Doi: 10.13057/ijas.v7i2.88784

Application of the Mixed Geographically Weighted Regression Model to Identify Influencing Factors for Literacy Development Index of Indonesian Society's in 2022

Zulhijrah, Ruliana, Aswi*
Statistics Study Program, Universitas Negeri Makassar, Makassar, Indonesia

***Corresponding author:** aswi@unm.ac.id

Submitted: 22-Jun-2024 **Revised:** 01-Feb-2025 **Accepted:** 03-Feb-2025

Abstract. *The mixed geographically weighted regression (MGWR) method is a combination of a linear regression model and a geographically weighted regression (GWR) model. The MGWR model can produce parameter estimates that have global parameter estimates, and other parameters that have local parameters according to the observation location. This method can be used in epidemiological studies that are influenced by spatial heterogeneity. The aim of this research is to determine and model the factors that influence the Community Literacy Development Index (CLDI) in Indonesia based on MGWR modeling. The data used in this research is CLDI data in Indonesia in 2022 along with the factors that are thought to influence it. The results of this research indicate that the MGWR model outperforms both the linear regression and GWR models, as it yields the lowest Akaike information criterion (AIC) value and an* R^2 *value of 96.54%. Based on the modeling results, several factors influencing CLDI were identified, including the percentage of libraries, the adequacy ratio of library collections, the average length of schooling, and the level of participation in organized learning.*

Keywords: *Literacy; literacy development index; mixed geographically weighted regression; spatial*

1. INTRODUCTION

Literacy comprises a collection of specialized talents and skills, including reading, writing, speaking, arithmetic, and problem-solving, which are essential for everyday life [1]. The Education Development Center (EDC), as described by the National Institute for Literacy, provides a more detailed definition of literacy, emphasizing that it encompasses an individual's capacity to utilize their potential, rather than solely their ability to read and write. The United Nations Educational, Scientific and Cultural Organization (UNESCO) further elucidates that literacy entails a repertoire of foundational abilities, particularly cognitive skills related to reading and writing, which are acquired independently of one's learning environment, as well as the sources and methods through which they are acquired. According to UNESCO, an individual's conception of literacy is influenced by academic proficiency, national context, institutional frameworks, cultural values, and personal experiences.

According to a 2019 study conducted by the Program for International Student Assessment (PISA), administered by the Organization for Economic Co-operation and Development (OECD), Indonesia is ranked 62nd out of 70 countries in terms of literacy rates. This places Indonesia among the countries with the lowest literacy rates [2]. Meanwhile, according to UNESCO, only 0.001

percent of the Indonesian population shows an interest in reading. This indicates that only 1 in 1,000 Indonesians have a preference for reading. Furthermore, in a separate study conducted by Central Connecticut State University in March 2016, titled "World's Most Literate Nations Ranked", Indonesia was ranked 60th out of 61 countries in terms of interest in reading [3].

This information underscores the need for further research into the literacy challenges facing Indonesia. An informed society is a hallmark of modernity, achievable only through the cultivation of a strong interest in reading among its populace [4]. High-quality human resources (HR) are essential for Indonesia to achieve its goal of becoming a developed nation by 2045. One strategy to enhance the quality of human resources is by fostering an informed society characterized by a strong interest in reading. Several studies have been conducted on literacy in Indonesia, employing qualitative analysis to demonstrate its impact on cognitive processes and behavior [1]. Additionally, research indicates a direct correlation between a country's quality and the literacy levels of its population. The extent to which individuals engage in reading books significantly influences their insight, cognition, and behavior [5].

Libraries represent a fundamental pillar in fostering a more cohesive literacy landscape [6]. Consistent with the objectives outlined by the National Library of the Republic of Indonesia for the period 2020–2024, the primary goal of library initiatives is to cultivate a culture of literacy, innovation, and creativity, thereby fostering a lifelong learning society. The community literacy development index (CLDI) plays a crucial role in advancing community literacy and fostering a passion for reading, with ongoing efforts to enhance its effectiveness, thereby addressing the challenges outlined in the Sustainable Development Goals (SDGs) agenda for 2030 [7].

Promoting literacy as an essential component of lifelong learning, and in alignment with the 2030 Sustainable Development Agenda, UNESCO has placed a high priority on literacy worldwide. UNESCO has developed a strategy aimed at enhancing global literacy, with a particular focus on children and adults. This strategy involves establishing a strong foundation through early childhood care and education, ensuring all children have access to quality basic education, improving functional literacy levels among adolescents and adults lacking basic literacy skills, and fostering an environment conducive to literacy development. Research on the CLDI was conducted in Probolinggo Regency [8] yielding a 2021 CLDI value of 10.67. This value suggests that a significant number of residents in the area do not engage in reading during their leisure time.

Over the past decade, there has been a considerable increase in the utilization of spatial data. Spatial data serves as a foundational element for constructing and sustaining various applications, information systems, and data based on location, making it one of the most crucial components today. Moreover, spatial data and information are assuming an increasingly significant role in education, particularly in literacy initiatives, aiding ongoing efforts to enhance literacy levels. One approach to analyzing the role of spatial data and information in literacy is the geographically weighted regression (GWR) method, which enables more accurate modeling of spatial relationships.

The GWR method is a statistical approach commonly employed to analyze data exhibiting spatial effects, allowing for the modeling of diverse relationships within a spatial visualization framework. Unlike global regression, the GWR method can model relationships while considering the spatial component, particularly distance. At the core of the GWR method lies the proximity between regions, represented by a weighting matrix. Greater proximity between regions corresponds to higher weight values [9]. Consequently, the GWR method offers greater accuracy in statistically analyzing spatial relationships among multiple variables, effectively addressing issues related to spatial heterogeneity [10].

The GWR model evolved from conventional regression models; however, in certain instances, the spatial variability of certain coefficients may be insignificant or overlooked. This occurs because the GWR model treats each location's characteristics differently, potentially ignoring or eliminating similarities across locations. Consequently, the mixed geographically weighted regression (MGWR) model was developed [11]. The MGWR model combines aspects of global regression with GWR, particularly when some predictor variables affecting the response variable are global, while others are local [12].

Research has been conducted to map the factors influencing reading literacy activities in Indonesia using the GWR application [13]. The analysis results indicate that the GWR model exhibits a higher goodness-of-fit compared to the linear regression model, thus offering superior modeling capabilities among various regression techniques. Based on the research conducted thus far, there has been no study utilizing MGWR to model the Community Literacy Development Index while considering its influencing factors in Indonesia. Therefore, this research aims to examine and identify the community literacy development index in Indonesia using MGWR. The objective is to determine the factors influencing the CLDI in Indonesia in 2022.

2. MATERIALS AND METHODS

2.1. Data Sources and Analysis Design

This study employs a quantitative research approach, which primarily focuses on numerical data analysis. The data utilized in this research is secondary data, obtained from existing sources rather than collected by the researcher. The data used in this study consist of one dependent variable and six independent variables across 34 provinces in Indonesia. The data were obtained from the 2023 publication of Indonesian Statistics, available on the official website of the Badan Pusat Statistik (BPS). The variables examined in this research are given in Table 1.

Table 1. Operational definition of variables

Variables	Definition
CLDI (Y)	The CLDI serves as a measure of the initiatives undertaken by local governments to promote and enhance libraries as a means of lifelong learning, aiming to cultivate a culture of literacy within the community
The percentage of the population (X_1)	The percentage of the population in each province of Indonesia is calculated by dividing the population of each province by the total population of Indonesia, and then multiplying the result by 100%.
The percentage of the number of libraries (X_2)	The percentage of the number of libraries for each province in Indonesia is obtained by dividing the number of libraries in each province by the total number of libraries in Indonesia, and then multiplying the result by 100%.
The library collection adequacy ratio (X_3)	The library collection adequacy ratio is the ratio between the quantity of reading materials held by a library and the quantity of reading materials deemed sufficient to meet the needs and interests of the library's served readership.

Variables	Definition
The average length of schooling (X_4)	The average length of schooling is defined as the average duration, in years, of formal education for individuals aged 15 years and older. This measurement considers a standard schooling period of 12 years, irrespective of grade repetition or interruptions in attendance.
Participation rates in organized learning (X_5)	Participation rates in organized learning serve as an indicator to assess the extent of young people's engagement in formal educational settings, particularly during the year preceding their enrollment in primary school.
The illiteracy rate (X_6)	The illiteracy rate is calculated by dividing the number of individuals unable to read in each province of Indonesia by the total number of illiterate individuals in Indonesia, and then multiplying the result by 100%.

In this study, data was processed using the MGWR method. The data analysis steps conducted were as follows: (1) analysis using linear regression; (2) classical assumption testing; (3) Testing for spatial effects; (4) GWR modeling; (5) MGWR modeling; (6) selection of the best model based on the highest R^2 value and lowest Akaike information criterion (AIC); and (7) interpretation of the results and drawing conclusions.

2.2. Linear Regression Analysis

Regression analysis is a method used to determine the relationship between variables. The general model for regression analysis is expressed as follows [14].

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i \, ; i = 1, 2, \dots, n \quad (1)$$

In this model y_i represents the response variable at the i^{th} observation, x_{ik} denotes the k^{th} regression coefficient at the i^{th} observation, β_0 represents the intercept, β_k is the regression coefficient for the k^{th} predictor variable, and ε_i is the residual at the i^{th} observation ($\varepsilon_i \sim IIDN(0, \sigma^2)$). In conducting regression analysis, several assumptions must be met, including: normality test, multicollinearity test, autocorrelation test, and heteroscedasticity test.

2.2.1 Normality Test

The Kolmogorov-Smirnov normality test can be employed to assess the normality of data [15].

Hypothesis:

H_0 : the residuals are assumed to follow a normal distribution..

H_1 : the residuals do not follow a normal distribution..

The null hypothesis H_0 is rejected if the p-value is less than the significance level $\alpha = 0.05$.

2.2.2. Multicollinearity Test

The multicollinearity test is employed to determine whether independent variables in the regression model are correlated with each other. The variance inflation factor (VIF) is a measure used to detect multicollinearity. The VIF is calculated as follows:

$$VIF = \frac{1}{1 - R_j^2} \quad (2)$$

where R_j^2 is the coefficient of determination among variables. If the VIF value is less than 10, it suggests that there is no multicollinearity among the independent variables [16].

2.2.3. Autocorrelation Test

Autocorrelation suggests that the attribute value in one area is related to the attribute value in another nearby area. Moran's I test is commonly used as a test statistic to detect spatial autocorrelation [17]. According to the hypothesis:

H_0 : the null hypothesis is that there is no autocorrelation

H_1 : the alternative hypothesis is that there is autocorrelation.

Test statistic:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j}^{n} w_{ij}(X_i-\bar{X})(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} \tag{3}$$

Test statistic:

$$Z_{value} = \frac{I-E(I)}{\sqrt{var(I)}} \tag{4}$$

Reject H_0 if $|Z_{value}| > Z_{\frac{\alpha}{2}}$.

2.2.4. Heteroscedasticity Test

The heteroscedasticity test is used to determine whether the data is homogeneous or not (homoscedasticity). The Breusch-Pagan test can be employed to detect heteroscedasticity in the data [18]. The hypotheses are as follows:

H_0 : the null hypothesis is that there is no heteroscedasticity.

H_1 : the alternative hypothesis is that there is heteroscedasticity.

Test statistic:

$$BP = \frac{1}{2}(\boldsymbol{f}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{f}) \tag{5}$$

The vector element *f* is denoted as $f_i = \frac{\varepsilon_i^2}{\sigma^2} - 1$ using the following formula.

The null hypothesis H_0 is rejected if the p-value is less than the significance level $\alpha = 0.05$.

2.3. Geographically Weighted Regression (GWR)

GWR is a spatial method that incorporates the geographical characteristics of each region as one of the factors influencing the response variable. GWR extends by incorporating geographic coordinates for each parameter at each location. The GWR model obtained is utilized to predict the response variable's magnitude using the resulting parameters, where each parameter is determined based on the object's location. At the core of the GWR method lies the proximity between regions, as indicated by a weighting matrix. Greater proximity between regions corresponds to higher weight values. The general equation for GWR is as follows [19].

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p}\beta_k(u_i, v_i)\, x_{ik} + \varepsilon_i \tag{6}$$

where (u_i, v_i) represent the coordinates of the i^{th} observation location.

The method used to estimate GWR parameters is weighted least squares, applied uniquely at each observation location [20]. The approach is as follows:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{Y} \tag{7}$$

where $\boldsymbol{W}$ represents the spatial weighting matrix for location i.

The weighting in the GWR model employs a kernel function to determine weighting values that can represent the spatial relationships between observations. The spatial weighting options in GWR include the fixed kernel function and the adaptive kernel function [21].

2.3.1 Fixed Kernel Gaussian Function

The function of the fixed kernel Gaussian is as follows.

$$w_j(u_i, v_i) = exp\left\{-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right\} \tag{8}$$

where $w_j(u_i, v_i)$ represents the weight assigned to location j based on coordinates (u_i, v_i).

2.3.2. Kernel Gaussian Adaptive Function

The function of the adaptive kernel Gaussian is as follows.

$$w_j(u_i, v_i) = exp\left\{-\frac{1}{2}\left(\frac{d_{ij}}{h_i}\right)^2\right\} \tag{9}$$

where $d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2}$ represents the formula for calculating the Euclidean distance. This distance calculation is utilized to determine the parameter values between the location (u_i, v_i) and the location (u_i, v_i). Meanwhile h is a parameter referred to as the smoothing parameter (*bandwidth*).

In the process of determining the optimal bandwidth, the cross-validation (CV) method can be utilized by adjusting the model's variance [22]. Mathematically, it can be defined as follows:

$$CV = \sum_{i=1}^{n}[y_i - \hat{y}_{\neq i}(h)]^2 \tag{10}$$

where $\hat{y}_{\neq i}(h)$ represents the estimated value of y_i when the observation at location (u_i, v_i) is excluded. The optimal h value is determined by minimizing the CV.

The model suitability test for the GWR model is based on the following hypotheses [23]:

H_0 : the GWR model is equivalent to the global regression model.

H_1 : the GWR model differs from the global regression model.

Test statistic:

$$F = \frac{\frac{SSE(H_0)}{v_1}}{\frac{SSE(H_1)}{\delta_1}} \tag{11}$$

Explanation:

$SSE(H_0) = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$ with $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$,

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

$df_1 = \frac{v_1^2}{v_2}$, $v_i = tr\left([(\boldsymbol{I} - \boldsymbol{H}) - (\boldsymbol{I} - \boldsymbol{L})^T(\boldsymbol{I} - \boldsymbol{L})]^i\right)$,

$df_2 = \frac{\delta_1^2}{\delta_2}$, $\delta_i = tr\left([(\boldsymbol{I} - \boldsymbol{L})^T(\boldsymbol{I} - \boldsymbol{L})]^i\right), i = 1, 2, \ldots$

The null hypothesis H_0 is rejected $F_1 \geq F_{\alpha, df_1, df_2}$ at the 5% significance level.

To identify predictor variables that have partial (local) influence, the following hypotheses can be formulated [21]:

H_0 : There is no influence of the independent variable on the dependent variable.

H_1 : There is an influence of the independent variable on the dependent variable.

Test statistic:

$$T_{value} = \frac{\hat{\beta}_k(u_i,v_i) - \beta_k(u_i,v_i)}{Se\hat{\beta}_k(u_i,v_i)} \tag{12}$$

where,

$\hat{\beta}_k(u_i, v_i)$ is the observed value of the k^{th} predictor variable at the i^{th} observation location.

$Se\hat{\beta}_k(u_i, v_i)$ is the standard error of the k^{th} predictor variable at the i^{th} observation location.

Reject H_0 if $|t_{value}| > t_{\frac{\alpha}{2};df}$ at the significance level α = 5%.

2.4. Mixed Geographically Weighted Regression (MGWR)

In MGWR model, some coefficients in the GWR model are assumed to be constant, while others vary according to the observed location. The equation for the MGWR model is as follows [24]:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i)\, x_{ik} + \sum_{k=p+1}^{q} \beta_k x_{ik} + \varepsilon_i \tag{13}$$

The parameter estimates in the MGWR model are the same as those estimated in the GWR model using the WLS method [11]. The parameter estimates for global variables are as follows:

$$\hat{\boldsymbol{\beta}}_g = \left[\boldsymbol{X}_g^T(\boldsymbol{I} - \boldsymbol{S_l})^T(\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{X}_g\right]^{-1}\boldsymbol{X}_g^T(\boldsymbol{I} - \boldsymbol{S}_l)^T(\boldsymbol{I} - \boldsymbol{S}_l)\boldsymbol{Y} \tag{14}$$

Meanwhile, the parameter estimates for local variables are as follows:

$$\hat{\boldsymbol{\beta}}_l = (\boldsymbol{X}_l^T\boldsymbol{W}(u_i, v_i)\boldsymbol{X_l})^{-1}\boldsymbol{X}_l^T\boldsymbol{W}(u_i, v_i)\left(\boldsymbol{Y} - \boldsymbol{X}_g\boldsymbol{\beta}_g\right) \tag{15}$$

where:

$$S_l = \begin{bmatrix} (X_{l1}^T W(u_1, v_1) X_l)^{-1} X_l^T W(u_1, v_1) \\ (X_{l2}^T W(u_2, v_2) X_l)^{-1} X_l^T W(u_2, v_2) \\ \vdots \\ (X_{ln}^T W(u_n, v_n) X_l)^{-1} X_l^T W(u_n, v_n) \end{bmatrix}$$

The model suitability test (goodness of fit) is conducted by testing the significance of geographical factors that influence local variables. The hypotheses for this test are as follows:

H_0: there is no difference between the MGWR and GWR models.,

H_1: there is a difference between the MGWR and GWR models.

Test statistic:

$$F_1 = \frac{DSS_1/v_1}{SSE(H_1))/u_1} \tag{16}$$

where:

$DSS_1 = \boldsymbol{Y}^T[(\boldsymbol{I} - \boldsymbol{H}) - (\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S})]\boldsymbol{Y}$

$SSE(\boldsymbol{H}_1) = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y}$

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

$u_i = tr\left(\left[(\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S})^i\right]\right)$

$v_i = tr\left(\left[(\boldsymbol{I} - \boldsymbol{H}) - (\boldsymbol{I} - \boldsymbol{S})^T(\boldsymbol{I} - \boldsymbol{S})^i\right]\right)$

$df_1 = \frac{v_1^2}{v_2}$ and $df_2 = \frac{u_1^2}{u_2}$.

Reject H_0 if $F_1 \geq F_{\alpha, df_1, df_2}$ at the 5% significance level.

Significance testing of parameters was conducted for each variable. The significance of parameters for global variables was tested using the following hypotheses:

H_0: global variables do not have a significant effect,

H_1: global variables have a significant effect.

Test statistic:

$$T_{gvalue} = \frac{\hat{\boldsymbol{\beta}}_k}{\hat{\sigma}\sqrt{\boldsymbol{g}_{kk}}} \text{ where } df = \frac{u_1^2}{u_1} \tag{17}$$

Reject H_0 if $\left|T_{gvalue}\right| \geq t_{\frac{\alpha}{2};df}$.

Next, the significance of parameters for local variables was tested using the following hypotheses:

H_0: local variables do not have a significant effect on the *i^th^* location,

H_1: local variables have a significant effect on the *i^th^* location.

Test statistic:

$$T_{lvalue} = \frac{\hat{\boldsymbol{\beta}}_k(u_i,v_i)}{\hat{\sigma}\sqrt{\boldsymbol{m}_{kk}}} \text{ where } df = \frac{u_1^2}{u_1} \tag{18}$$

Reject H_0 if $|T_{lvalue}| \geq t_{\frac{\alpha}{2};df}$.

where $\boldsymbol{g}_{kk}$ is the diagonal element of the $\boldsymbol{G}\boldsymbol{G}^T$ matrix and $\boldsymbol{m}_{kk}$ is the diagonal element of the matrix $\boldsymbol{M}_i{\boldsymbol{M}_i}^T$.

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

$$\boldsymbol{S}_g = \boldsymbol{X}_g\left({\boldsymbol{X}_g}^T\boldsymbol{S}_g\right)^{-1}{\boldsymbol{X}_g}^T$$

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

2.5. Selection of the Best Model

The best model is determined using the AIC and the coefficient of determination (R^2), where a lower AIC value and a higher R^2 indicate a better model. The AIC value represents the likelihood of the model minimizing information loss (error). The AIC value is determined using the following formula.

$$AIC = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n + tr(L) \tag{19}$$

The coefficient of determination is utilized to measure the proportion of variance in the research data explained by the obtained regression model [22]. The R^2 value can be calculated using the following formula.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{20}$$

3. RESULTS AND DISCUSSION

3.1. Description of the Distribution of CLDI in Indonesia

The distribution of the CLDI in Indonesia in 2022 is illustrated in Figure 1.

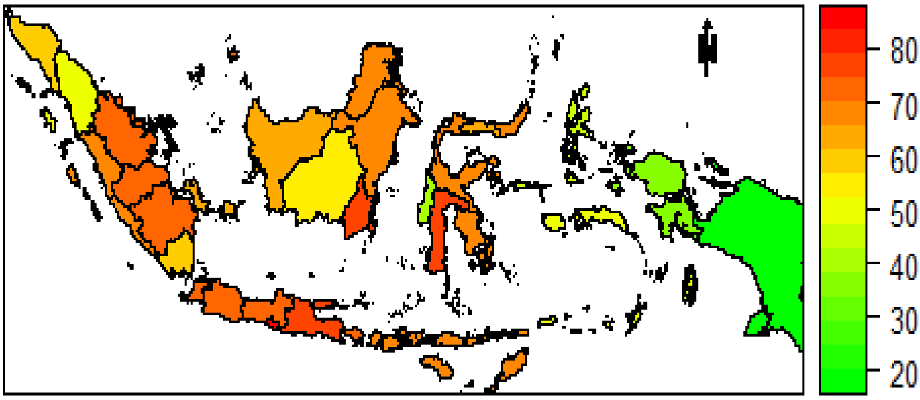


Figure 1. Map of CLDI Distribution in Indonesia

Figure 1 illustrates that the average CLDI in Indonesia is 64.48, with a median of 67.78. The CLDI ranges between the lowest and highest provinces, reaching 63.61, indicating significant differences among provinces in Indonesia. The highest CLDI (red) was recorded in DI Yogyakarta Province at 83.63, followed by DKI Jakarta Province with an CLDI of 80.63, representing a difference of 2.96. Meanwhile, the lowest CLDI (green) was observed in Papua Province, with an CLDI of 20.02.

3.2. Regression Analysis of CLDI

A linear regression model is employed to examine the relationship between CLDI in Indonesia and the factors believed to influence it. The regression model formulated based on the results above is as follows:

$$\hat{Y} = -35.5121 + 0.3158X_1 + 1.4294X_2 + 19.6995X_3 + 2.3479X_4 + 0.4529X_5 + 0.1080X_6$$

The factors that significantly influence CLDI are the percentage of the number of libraries (X_2) and the library collection adequacy ratio (X_3). This is evident from the value of Pr(>| t |) which is less than 0.05.

The regression model formed with two predictor variables, namely the percentage of the number of libraries (X_2) and the library collection adequacy ratio (X_3) is as follows:

$$\hat{Y} = 25.287 + 1.794\mathrm{X}_2 + 22.153\mathrm{X}_3$$

The formed regression model has an F_{value} (19.88) > F_{table} (2.46) with p-value 9.56×10^{-9}. Additionally, an $R^2 =$ 0.7744 was obtained, indicating that 77.44% of CLDI can be explained by the independent variables used. The remaining 22.56% is attributed to other unobserved factors.

After obtaining the regression model, the assumptions of the regression model were tested. The results are presented below.

3.2.1. Normality Test

Normality assumption testing in this study utilized the Kolmogorov-Smirnov test. The results of the Kolmogorov-Smirnov test yielded a value of D = 0.0933 with a p-value of 0.6369. This indicates that the residual data is normally distributed, as the p-value = 0.6369 > α = 0.05.

3.2.2. Multicollinearity Test

Testing for multicollinearity assumption in this study utilized the VIF test statistic. The VIF value for each predictor variable is presented in Table 2.

Table 2. VIF values for independent variables

	X_1	X_2	X_3	X_4	X_5	X_6
VIF	1.77	1.92	1.29	2.73	2.17	3.38

According to Table 2, all VIF values for the predictor variables are below 10, indicating the absence of multicollinearity among the predictor variables.

3.2.3. Autocorrelation Test

Autocorrelation assumption testing in this study employed Moran's I test. The results of Moran's I test yielded a test statistic value of 0.6016 with a p-value of 5.37×10^{-5}. These results indicate the presence of autocorrelation, as the p-value $5.37\times10^{-5} < \alpha = 0.05$.

3.2.4. Heteroscedasticity Test

Heteroscedasticity assumption testing in this study utilized the Breusch-Pagan test. The results of the Breusch-Pagan test yielded a BP test statistic value of 22.328 with a p-value of 0.00106. These results indicate the presence of heterogeneity in the data, as the p-value 0.00106 $< \alpha = 0.05$. Therefore, further analysis is required to address the data heterogeneity, namely by employing the GWR method.

3.3. Geographically Weighted Regression Modeling of CLDI

In GWR modeling, the first step is to calculate the Euclidean distance between observation locations. After calculating the Euclidean distance, weighting calculations for each observation are performed using either the Gaussian kernel function or the adaptive Gaussian kernel. The selection of the optimal bandwidth is carried out using the cross validation (CV) method for each weighting function to minimize the CV.

Table 3. Optimal bandwidth selection

Kernel Function	Bandwith Value	CV	R^2	AIC
Fixed Gaussian	13.6995	1120.286	0.9043	200.643
Adaptive Gaussian	0.3824	1104.277	0.9177	197.431

Based on the analysis results presented in Table 3, the optimal bandwidth value for the adaptive Gaussian kernel function is 0.3824, with a corresponding CV of 1104.277. This optimal bandwidth will be utilized to determine the weighting for each region in Indonesia.

Next, a GWR model suitability test was conducted using the F-test. The test results yielded an F-test statistic value of 2.2436 with a corresponding p-value of 0.0314. This indicates a significant difference between the obtained GWR model and the classical regression model, as the p-value (0.0314) is less than the significance level α = 0,05. Additionally, the R^2 value obtained in the GWR model is 0.9177. This implies that 91.77% of CLDI can be explained by the independent variables in the GWR model, while the remaining 8.23% is attributed to other unobserved factors.

Subsequently, tests were conducted to assess the geographical influence of each predictor variable in the GWR model (Equation 7). The results of these tests are presented in Table 4.

Table 4. Results of simultaneous testing for geographical influences

	F value	DF_1	DF_2	$Pr(>)$	Explanation
Intercept	1.4507	9.2696	23.924	0.2215	
X_1	0.2795	10.9756	23.924	0.9840	not significant
X_2	5.9981	7.6677	23.924	0.0003	significant
X_3	0.4764	14.6533	23.924	0.9282	not significant
X_4	6.4755	13.2539	23.924	4.26×10^{-5}	significant
X_5	0.7739	9.6448	23.924	0.6486	not significant
X_6	2.9460	17.5715	23.924	0.0075	significant

Based on Table 4, it is evident that the factors significantly influencing CLDI based on GWR modeling are the percentage of the number of libraries (X_2), the average length of schooling (X_4) and the proportion of illiterates (X_6).

In the previous test, the geographic influence of each variable as a whole was considered. However, in the GWR model, the variables influencing each region differ. This can be observed through a partial test using the t-test for each region. The t_{value} results are compared with the t_{table} value of 1.7011. Below are the parameter estimates for each region in Indonesia:

$$\hat{Y}_{Aceh} = -46.73022 + 0.9443723X_2 + 19.69200X_3 + 5.1239558X_4$$

$$\hat{Y}_{Bali} = -83.46752 + 0.9989248X_2 + 17.27572X_3 + 5.4055179X_4 + 0.7074332X_5$$

$$\hat{Y}_{Bangka\ Belitung} = -62.67978 + 0.7280965X_2 + 19.42949X_3 + 6.5774014X_4$$

$$\hat{Y}_{Banten} = -79.60929 + 0.7516402X_2 + 19.00981X_3 + 6.7469516X_4$$

$$\hat{Y}_{Bengkulu} = -53.76739 + 0.7995749X_2 + 19.83006X_3 + 6.0762435X_4$$

$$\hat{Y}_{Gorontalo} = -20.04551 + 1.5120685X_2 + 19.23413X_3 + 0.3762398X_5$$

$$\hat{Y}_{DKI\ Jakarta} = -82.15262 + 0.7500050X_2 + 18.78327X_3 + 6.7713420X_4$$

$$\hat{Y}_{Jambi} = -44.61758 + 0.7988789X_2 + 19.94659X_3 + 5.9074345X_4$$

$$\hat{Y}_{Jawa\ Barat} = -90.80865 + 0.7384239X_2 + 18.44331X_3 + 6.9621874X_4$$

$$\hat{Y}_{Jawa\ Tengah} = -103.39152 + 80.7344324X_2 + 17.64019X_3 + 7.1129114X_4 + 0.7652902X_5$$

$$\hat{Y}_{Jawa\ Timur} = -94.34372 + 0.8398481X_2 + 17.35592X_3 + 6.3982067X_4 + 0.7355606X_5$$

$$\hat{Y}_{Kalimantan\ Barat} = -61.54594 + 0.4496929X_1 + 0.8175247X_2 + 17.86655X_3 + 5.8896547X_4$$

$$\hat{Y}_{Kalimantan\ Selatan} = -78.10524 + 0.4879604X_1 + 0.8954427X_2 + 16.98042X_3 + 5.6971349X_4 + 0.6340290X_5$$

$$\hat{Y}_{Kalimantan\ Tengah} = -67.85496 + 0.4654944X_1 + 0.8607532X_2 + 17.44490X_3 + 5.7604477X_4$$

$$\hat{Y}_{Kalimantan\ Timur} = -36.51440 + 0.4456779X_1 + 1.1558313X_2 + 18.46384X_3$$

$$+3.7207522X_4$$

$$\hat{Y}_{Kalimantan\ Utara} = -36.51440 + 0.4456779X_1 + 1{,}1558313X_2 + 18{,}46384X_3$$
$$+3.7207522X_4$$

$$\hat{Y}_{Kepulauan\ Riau} = -39.76437 + 0.8624309X_2 + 18.80452X_3 + 5.3287361X_4$$

$$\hat{Y}_{Lampung} = -67.21352 + 0.8175247X_2 + 19.51773X_3 + 6.6144875X_4$$

$$\hat{Y}_{Maluku} = -15.89289 + 1.7457949X_2 + 19.59222X_3 + 0.4449400X_5$$

$$\hat{Y}_{Maluku\ Utara} = -10.35509 + 1.7078747X_2 + 19.72168X_3 + 0.3800607X_5$$

$$\hat{Y}_{NTB} = -65.28025 + 1.1869943X_2 + 17.85991X_3 + 4.1276132X_4 + 0.6211300X_5$$

$$\hat{Y}_{NTT} = -50.18575 + 1.4100192X_2 + 18.33898X_3 + 0.5875462X_5$$

$$\hat{Y}_{Papua} = -19.33500 + 1.7056640X_2 + 19.73653X_3 + 0.4449622X_5$$

$$\hat{Y}_{Papua\ Barat} = -14.46313 + 1.7428706X_2 + 19.74557X_3 + 0.4251978X_5$$

$$\hat{Y}_{Riau} = -39.29458 + 0.8144798X_2 + 19.90754X_3 + 5.7096532X_4$$

$$\hat{Y}_{Sulawesi\ Barat} = -42.22771 + 1.2511588X_2 + 17.93419X_3 + 3.2006122X_4$$
$$+ 0.4630220X_5$$

$$\hat{Y}_{Sulawesi\ Selatan} = -42.33189 + 1.3171925X_2 + 18.15075X_3 + 0.4912400X_5$$

$$\hat{Y}_{Sulawesi\ Tengah} = -25.72545 + 1.4389682X_2 + 18.70654X_3 + 0.4048710X_5$$

$$\hat{Y}_{Sulawesi\ Tenggara} = -32.96218 + 1.4579566X_2 + 18.44856X_3 + 0.4925933X_5$$

$$\hat{Y}_{Sulawesi\ Utara} = -15.32434 + 1.5941408X_2 + 19.39984X_3 + 0.3764689X_5$$

$$\hat{Y}_{Sumatera\ Barat} = -44.44364 + 0.8270860X_2 + 19.84681X_3 + 5.7380127X_4$$

$$\hat{Y}_{Sumatera\ Selatan} = -54.49185 + 0.7723848X_2 + 19.85154X_3 + 6.2425819X_4$$

$$\hat{Y}_{Sumatera\ Utara} = -41.55735 + 0.8612139X_2 + 19.74477X_3 + 5.4820823X_4$$

$$\hat{Y}_{DI\ Yogyakarta} = -101.37467 + 0.7668268X_2 + 17.56537X_3 + 6.9143237X_4$$
$$+ 0.7618670X_5$$

From this modeling, it is evident that the variables influencing each region differ, indicating variations in influencing factors across locations. This is discernible through a partial test using the t-test for each region. The t_{value} results are compared with the t_{table} value of 1.7011, with the assumption that $|t_{value}| \geq t_{table}$. The factors influencing CLDI in each province with GWR modeling can be seen in Table 5.

Table 5. Factors influencing CLDI as per GWR modeling

Variables	Provinces
X_1	West Kalimantan, South Kalimantan, Central Kalimantan and East Kalimantan
X_2	All provinces exhibit influence.
X_3	All provinces exhibit influence.

Variables	Provinces
X_4	All provinces are affected except Gorontalo, Maluku, North Maluku, East Nusa Tenggara, Papua, West Papua, South Sulawesi, Central Sulawesi, Southeast Sulawesi, and North Sulawesi.
X_5	Bali, Gorontalo, Central Java, East Java, South Kalimantan, Maluku, North Maluku, NTB, NTT, Papua, West Papua, West Sulawesi, South Sulawesi, Central Sulawesi, Southeast Sulawesi, North Sulawesi, and DI Yogyakarta.
X_6	-

3.4. Mixed Geographically Weighted Regression Modeling of CLDI

Based on the analysis of the GWR model with adaptive Gaussian Kernel weighting, two predictor variables, X_2 and X_3 exhibit global influence. Conversely, predictor variables X_1, X_4 and X_5 demonstrate local influence. Consequently, an MGWR model was constructed using these variables and adaptive Gaussian kernel weighting. The results of the MGWR model suitability test, conducted using the F test, are presented in Table 6.

Table 6. MGWR Model fit test results

	F	*p-value*
F_1	7.452	1.051×10^{-6}
F_{Global}	36.977	4.22×10^{-14}
F_{Lokal}	16.819	7.24×10^{-11}

Based on Table 6, we conclude that there is a significant difference between the MGWR model and the classical regression model because the *p-value* = $1.051\times10^{-6} < \alpha = 0.05$. Meanwhile, for both global and local parameters, the p-values obtained are smaller than α = 0.05 indicating that both global and local variables have a significant effect on the response variable simultaneously.

The parameters generated in the MGWR model consist of both global and local parameters. Consequently, each region in Indonesia will have different parameters, while some will remain the same. The parameter estimates obtained for each region through MGWR analysis are as follows:

$$\hat{Y}_{Aceh} = 73.2427631 + 0.855851X_2 + 16.938397X_3$$

$$\hat{Y}_{Bali} - 328.8291416 + 0.855851X_2 + 16.938397X_3$$

$$\hat{Y}_{Bangka\ Belitung} = 16.9151174 + 0.855851X_2 + 16.938397X_3$$

$$\hat{Y}_{Banten} = -95.0614185 + 0.855851X_2 + 16.938397X_3$$

$$\hat{Y}_{Bengkulu} = 116.1027189 + 0.855851X_2 + 16.938397X_3 - 0.9080594X_5$$

$$\hat{Y}_{Gorontalo} = -23.251670 + 0.855851X_2 + 16.938397X_3$$

$$\hat{Y}_{DKI\ Jakarta} = -100.7122594 + 0.855851X_2 + 16.938397X_3$$

$$\hat{Y}_{Jambi} = 121.6903105 + 0.855851X_2 + 16.938397X_3 - 0.9505327X_5$$

$\hat{Y}_{Jawa\ Barat} = -137.5382837 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Jawa\ Tengah} = -123.4237081 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Jawa\ Timur} = -137.7988010 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Kalimantan\ Barat} = -40.6549286 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Kalimantan\ Selatan} = -100.9604101 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Kalimantan\ Tengah} = -92.8796330 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Kalimantan\ Timur} = 58.6396157 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Kalimantan\ Utara} = 58.6396157 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Kepulauan\ Riau} = 4.82525497 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Lampung} = -13.5363652 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Maluku} = -45.979821 + 0.855851X_2 + 16.938397X_3 - 4.072096X_4$

$\hat{Y}_{Maluku\ Utara} = -47.187603 + 0.855851X_2 + 16.938397X_3 - 3.821426X_4$

$\hat{Y}_{NTB} = -58.8838168 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{NTT} = -30.8795324 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Papua} = -61.801663 + 0.855851X_2 + 16.938397X_3 - 12.565899X_4$

$\hat{Y}_{Papua\ Barat} = -42.89371 + 0.855851X_2 + 16.938397X_3 - 12.06340X_4$

$\hat{Y}_{Riau} = 122.2698552 + 0.855851X_2 + 16.938397X_3 - 0.9494914X_5$

$\hat{Y}_{Sulawesi\ Barat} = -30.0603966 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Sulawesi\ Selatan} = -30.0827738 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Sulawesi\ Tengah} = -23.2677144 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Sulawesi\ Tenggara} = -25.36458772 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Sulawesi\ Utara} = -22.3788078 + 0.855851X_2 + 16.938397X_3$

$\hat{Y}_{Sumatera\ Barat} = 122.4042404 + 0.855851X_2 + 16.938397X_3 - 0.9401086X_5$

$\hat{Y}_{Sumatera\ Selatan} = 166.8826466 + 0.855851X_2 + 16.938397X_3 - 0.9155449X_5$

$\hat{Y}_{Sumatera\ Utara} = 96.9622546 + 0.855851X_2 + 16.938397X_3$

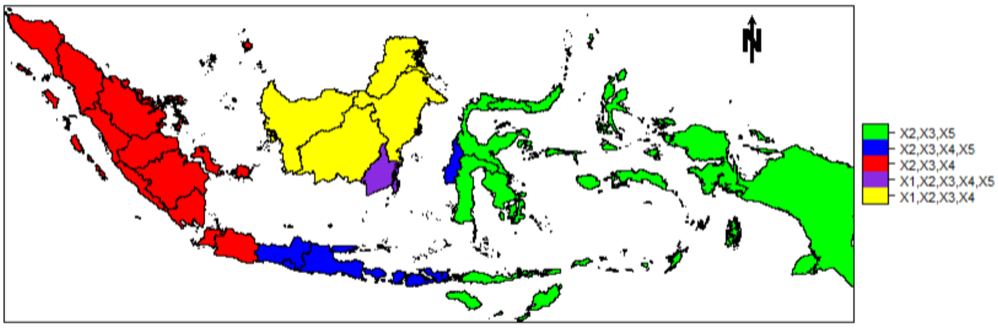
$\hat{Y}_{DI\ Yogyakarta} = -124.5652549 + 0.855851X_2 + 16.938397X_3$

The factors influencing CLDI in each province using MGWR modeling are presented in Table 7.

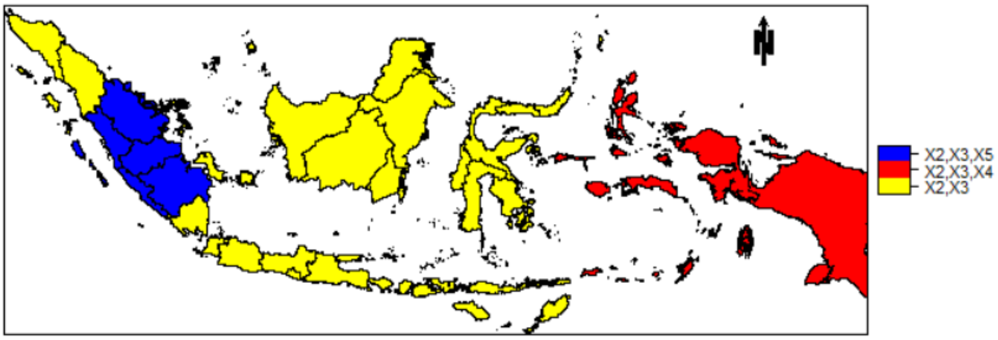
Table 7. Factors influencing CLDI Based on MGWR modeling

Variables	Provinces
X_1	-
X_2	All provinces exhibit influence.
X_3	All provinces exhibit influence.
X_4	Maluku, North Maluku, Papua and West Papua
X_5	Bengkulu, Jambi, Riau, North Sumatera and South Sumatera

Variable significance distribution map for GWR model and MGWR model is given in Figure 2 (a) and (b), respectively.



(a)



(b)

Figure 2. Variable significance distribution map (a) GWR model (b) MGWR model

3.5. Selection of the Best Model

The best model selection is based on the comparison of R^2 and AIC values. A higher R^2 value and a lower AIC value indicate a better model compared to others. This comparison is presented in Table 8.

Table 8. R^2 value and AIC model selection

Model	R^2	AIC
Linear Regression	0.7744	229.7824
GWR	0.9177	197.4314
MGWR	0.9654	172.1055

Based on Table 8, it can be observed that the linear regression model achieved an R^2 of 77.44% with an AIC of 229.7824. The GWR model showed an R^2 value of 91.77% with an AIC of 197.4214. Furthermore, the MGWR model demonstrated the highest R^2 value of 96.54% with an AIC of 172.1055. These results indicate that all three methods yielded models suitable for modeling, as their R^2 values exceeded 50%. However, the MGWR method is preferred for CLDI modeling, given its superior performance in producing a model with the highest R^2 value and the lowest AIC value.

4. CONCLUSION

From the results of MGWR modeling, it was found that the factors influencing the community literacy development index in Indonesia globally in each province are the percentage of the number of libraries and the adequacy ratio of library collections. Additionally, certain factors have a localized influence, affecting only specific regions. For instance, the average years of schooling significantly impact the Maluku, North Maluku, Papua, and West Papua regions. Meanwhile, the level of participation in organized learning significantly affects the Bengkulu, Jambi, Riau, West Sumatra, and South Sumatra regions.

REFERENCES

[1] R. Rintaningrum, "Explaining the important contribution of reading literacy to the country's generations: Indonesian's perspectives," *Int. J. Innov. Creat. Chang.*, vol. 5, no. 3, pp. 936–953, 2019, [Online]. Available: https://www.ijicc.net/images/Vol_5_Iss_3/Part_2_2020/5310_Ratna_2019_E_R.pdf

[2] S. Nugrahanto and D. Zuchdi, "Indonesia PISA result and impact on the reading learning program in Indonesia," *Adv. Soc. Sci. Educ. Humanit. Res.*, vol. 297, no. 18, pp. 373–377, 2019, doi: 10.2991/icille-18.2019.77.

[3] L. Tahmidaten and W. Krismanto, "Permasalahan budaya membaca di indonesia (studi pustaka tentang problematika & solusinya)," *Sch. J. Pendidik. dan Kebud.*, vol. 10, no. 1, pp. 22–33, 2020, doi: 10.24246/j.js.2020.v10.i1.p22-33.

[4] X. Yang, S. Huo, and X. Zhang, "Visual-spatial skills contribute to Chinese reading and arithmetic for different reasons: A three-wave longitudinal study," *J. Exp. Child Psychol.*, vol. 208, no. 42, pp. 1–16, 2021, doi: 10.1016/j.jecp.2021.105142.

[5] A. Permatasari, "Membangun kualitas bangsa dengan budaya literasi," in *Prosiding Seminar Nasional Bulan Bahasa UNIB*, 2015, pp. 146–156.

[6] N. Kurniasih, "Revitalization of library space function in digital era: comparative study of library space in TEMPO's Data and Analysis Center and Library of Universitas Indonesia," in *Proceedings of The 1st Biannual Congress of International Library, Archives and Information Science Discovery & Technology 2016*, 2016, pp. 111–121.

[7] W. Nashihuddin, "Peran perpustakaan sebagai media literasi digital masyarakat," *J. Polit. dan Sos. Kemasyarakatan*, vol. 147, no. 2, pp. 1–6, 2020, doi: 10.13140/RG.2.2.28221.82407.

[8] F. Shoimah, "Mengukur indeks pembangunan literasi masyarakat (IPLM) Kabupaten Probolinggo," *J. PANGRIPTA*, vol. 6, no. 1, pp. 1164–1175, 2023.

[9] D. Yu, Y. Zhang, X. Wu, D. Li, and G. Li, "The varying effects of accessing high-speed rail system on China's county development: A geographically weighted panel regression analysis," *Land use policy*, vol. 100, no. 5, pp. 1–11, 2021, doi: 10.1016/j.landusepol.2020.104935.

[10] D. O. Hasibuan, R. D. Bekti, E. Sutanta, and I. W. J. Pradnyana, “Application of the geographically weighted regression method to the human development index and visualization on the Tableau Dashboard,” *3rd Int. Conf. Inf. Technol. Secur. Appl.*, vol. 3, no. 1, pp. 39–51, 2022.

[11] D. R. S. Saputro, H. N. Astuti, P. Widyaningsih, S. Sutanto, and R. Setiyowati, “Characteristics of parameter mixed geographically weighted regression model: Global (a-group) and local (b-group),” *J. Phys. Conf. Ser.*, vol. 1776, no. 1, pp. 1–7, 2021, doi: 10.1088/1742-6596/1776/1/012044.

[12] F. Kusumaningati, J. Utomo, M. Zach, A. Puspitawati, and N. Chamidah, “Modeling of school participation rate for senior high school in Indonesia using mixed geographically weighted regression model,” in *Proceedings ofthe 2nd International Conference Postgraduate School (ICPS)*, 2018, no. Icps, pp. 939–943. doi: 10.5220/0007554409390943.

[13] A. Hapsery and D. Trishnanti, “Aplikasi geographically weighted regression (gwr) untuk pemetaan faktor yang mepengaruhi indeks aktivitas literasi membaca di Indonesia,” *J. Ris. dan Apl. Mat.*, vol. 05, no. 02, pp. 80–91, 2021.

[14] M. A. Tiro, *Analisis Korelasi dan Regresi*, Edisi Ketiga. Makassar: Andira Publisher, 2010.

[15] A. Quraisy, “Normalitas data menggunakan Uji Kolmogorov-Smirnov dan Saphiro-Wilk,” *J. Heal. Educ. Econ. Sci. Technol.*, vol. 3, no. 1, pp. 7–11, 2020.

[16] Yuliana and D. R. S. Saputro, “Algoritme least angle regression untuk model geographically weighted least absolute shrinkage and selection operator,” *Semin. Mat. dan Pendidik. Mat. UNY*, pp. 139–144, 2017.

[17] J. Tang, F. Gao, F. Liu, C. Han, and J. Lee, “Spatial heterogeneity analysis of macro-level crashes using geographically weighted Poisson quantile regression,” *Accid. Anal. Prev.*, vol. 148, no. 4, pp. 1–20, 2020, doi: 10.1016/j.aap.2020.105833.

[18] S. Kartika and G. Kholijah, “Penggunaan metode geograhically weighted regression (gwr) untuk mengestimasi faktor dominan yang mempengaruhi penduduk miskin di Provinsi Jambi,” *J. Mat. Theory Appl.*, vol. 2, no. 2, pp. 37–45, 2020, [Online]. Available: https://ojs.unsulbar.ac.id/index.php/Mathematics/article/view/998

[19] X. Xu, X. Luo, C. Ma, and D. Xiao, “Spatial-temporal analysis of pedestrian injury severity with geographically and temporally weighted regression model in Hong Kong,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 69, no. 2, pp. 286–300, 2020, doi: 10.1016/j.trf.2020.02.003.

[20] L. Anselin, *Spatial Econometrics: Methods and Models*, vol. 85, no. 411. 1988. doi: 10.2307/2290042.

[21] D. W. S. Yusuf, E. M. P. Hermanto, and W. Pramesti, “Pemodelan geographically weighted regression (GWR) pada persentase kriminalitas di Provinsi Jawa Timur tahun 2017,” *Indones. J. Stat. Its Appl.*, vol. 4, no. 1, pp. 156–163, 2020, doi: 10.29244/ijsa.v4i1.557.

[22] N. Lutfiani, S. Sugiman, and S. Mariani, “Pemodelan geographically weighted regression (GWR) dengan fungsi pembobot kernel Gausian dan Bi-Square,” *UNNES J. Math.*, vol. 8, no. 1, pp. 82–91, 2019. doi: 10.15294/ujm.v8i1.17103

[23] K. Amelia, L. O. Asril, and L. Febrianti, “Pemodelan incident rate demam berdarah dengue di Indonesia yang berkaitan dengan faktor lingkungan menggunakan metode geographically weighted regression (GWR),” *Ekol. J. Ilm. Ilmu Dasar dan Lingkung. Hidup*, vol. 20, no. 2, pp. 64–73, 2020. doi: 10.33751/ekologia.v20i2.2167

[24] M. Y. Darsyah, R. Wasono, and M. F. Agustina, “The model of mixed geographically weighted regression (MGWR) for poverty rate in central java the model of mixed

geographically weighted regression," *Int. J. Appl. Math. Stat.*, vol. 53, no. 6, pp. 114–121, 2015. https://www.researchgate.net/publication/303812347