PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA
e-ISSN. 2829-6303
Vol. 04, No. 3, December 2025
Page 575-590

Check for updates

Modeling Illiteracy Rate in Indonesia with Spatial Regression

Alvizar Syamsul Balda[1] , **Alfira Mulya Astuti**[2*], **Parhaini Andriani**[3]

[1,2,3] Mathematics Education Study Program, Faculty of Tarbiyah and Teacher Training, Universitas Islam Negeri Mataram
Gajah Mada Street No.100, Jempong Baru, Mataram, 83116, West Nusa Tenggara, Indonesia

***E-mail Correspondence Author:** *alfiramulyastuti@uinmataram.ac.id*

Abstract
The illiteracy rate (ILR) serves as an indication of educational attainment, representing the percentage of individuals aged 15 and older who lack reading and writing skills. Despite a reduction in Indonesia's ILR to 3.33% in 2024, the objective of eliminating it entirely remains a priority to fulfill the fourth aim of The Sustainable Development Goals (SDGs) by 2030. This research seeks to estimate Indonesia's rate of illiteracy by examining relevant elements, including the number of individuals living in poverty, mean years of schooling, and gross enrollment ratio. The data is obtained from the BPS-Statistics Indonesia. This study employs spatial regression, utilizing an area-based methodology to capture spatial impacts among regions and analyze them with R software. The analysis results indicate that a) the chosen weighted matrix is k-Nearest Neighbor, b) the selected spatial model for the illiteracy rate in Indonesia is the Spatial Durbin Model (SDM), and c) mean years of schooling and gross enrollment ratio within a province significantly affect the illiteracy rate in that province, which may indirectly elevate the illiteracy rate in neighboring regions.

Keywords: *Iliteracy Rate, Spatial Regression.*

: https://doi.org/10.30598/parameterv4i1pp575-590

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright © 2025 The Author(s).
Research Article ***Open Acces***
Email: jurnalparameter@gmail.com
Homepage: https://ojs3.unpatti.ac.id/index.php/parameter
Submitted: June 2025
Accepted: December 2025

1. INTRODUCTION

Education is a significant facet of life since it substantially enhances its quality. A society's educational attainment correlates positively with the quality of its human resources. The enhancement of human resource quality serves as a more robust assurance for attaining a more sophisticated and wealthy existence. The illiteracy rate (ILR) serves as an indicator of educational attainment [1]. The illiteracy rate is the percentage of individuals aged 15 and older who are unable to read or write in Latin or other scripts within a specific region [2]. BPS-Statistics Indonesia projects that by 2024, 3.33% of the adult population aged 15 years and older will remain illiterate [3]. In comparison to the preceding year, the illiteracy rate in 2023 was around 3.47%. This figure signifies a reduction of 0.14%, reflecting advancements in enhancing literacy in Indonesia. Notwithstanding the decline, this proportion indicates that Indonesia must persist in its efforts to eliminate illiteracy entirely, hence facilitating the attainment of the fourth objective of the Sustainable Development Goals (SDGs), which is "quality education for all" by 2030 [4].

Mitigating illiteracy aligns with the directive of Allah SWT as articulated in the Qur'an, Surah Al-Isra', verse 14, which states:

اِقْرَأْ كِتٰبَكَۗ كَفٰى بِنَفْسِكَ الْيَوْمَ عَلَيْكَ حَسِيْبًاۗ

Meaning: "*Go through your book and use today as a reckoning of yourselves*" (QS. Al-Isra' verse 14).

The verse emphasizes the importance of reading and seeking knowledge as essential tools to avoid ignorance, with one specific aspect being the avoidance of illiteracy.

A reduced illiteracy rate is anticipated to enhance the welfare and quality of life for individuals in Indonesia. Therefore, investigating the determinants of illiteracy rates is of critical importance. Numerous prior researchers have conducted studies on the factors influencing illiteracy [5].

Kurniati [6] conducted an examination of the illiteracy rate in West Java Province in 2012, employing spatial regression via the Geographically Weighted Logistic Regression (GWLR) methodology. The findings indicated that the illiteracy rate in West Java Province was affected by various factors, including population size, the ratio of teachers to students, the ratio of students to schools, the percentage of the impoverished population, and the rate of primary school enrollment. In 2020, Lake et al. [1] conducted an analysis of the illiteracy rate in Papua Province, employing spatial panel data methodologies. The findings of Lake et al.'s research suggest that the determinants affecting the illiteracy rate in Papua Province include the gross regional domestic product (GRDP) per capita, the number of people living in poverty, and ratio of educators to school-age children. Satvika et al. [7] conducted a modeling analysis of the illiteracy rate in West Nusa Tenggara Province for the years 2016 to 2020, employing the panel regression method. The findings suggest that the prevalence of poverty, GRDP, the ratio of elementary school students to teachers, and the proportion of households employing computers or laptops considerably affects the illiteracy rate in West Nusa Tenggara.

Maharani [2] conducted a modeling analysis of the illiteracy rate in West Sumatra Province for the year 2014 using Geographically Weighted Regression (GWR). The analysis indicates that the factors that had a significant impact on the illiteracy rate in West Sumatra in 2014 include the net primary enrollment rate, the net enrollment rate for junior high schools, the proportion of junior high school education buildings, and the proportion of junior high school teaching staff. Furthermore, Irwansyah et al. [8] employed a spatial regression approach to map and model the illiteracy rate in East Java Province. The findings indicate that the illiteracy rate is affected by the ratio of the impoverished population, the proportion of primary school teaching staff, the proportion of junior high school teaching staff, and the net enrollment rate for individuals aged 13 to 15 years.

Previous studies demonstrate that certain variables from the economic sector are frequently utilized by researchers and have been shown to significantly impact illiteracy rates, particularly the number of individuals living in poverty. The prevalence of individuals experiencing poverty indicates the community's challenges in meeting fundamental requirements, including sustenance, apparel, and housing. Consequently, we expect that a decrease in the poverty rate may have a beneficial impact on the illiteracy rate. This study incorporates the number of individuals living in poverty as an independent variable from the economic sector to analyze its relationship with the illiteracy rate. Furthermore, the variables employed pertain to the education sector, specifically the mean years of schooling and the gross enrollment ratio. The impact of these two variables on the illiteracy rate is substantial [9].

The mean years of schooling (MYS) was utilized as an independent variable, as an extended duration of education suggests a higher level of intellectual capability within the population (literacy and numeracy) [10]. MYS represents the average duration of formal schooling attained by individuals aged 15 and older. The computation of MYS adheres to the criteria established by the United Nations Development Programme (UNDP), spanning from 0 to 15 years, where 0 signifies no educational attainment and 15 indicates a comparatively elevated degree of formal education [11]. The gross enrollment ratio is considered a factor that affects the illiteracy rate.

The Gross Enrollment Ratio (GER) quantifies the overall participation of the populace in education. GER is the percentage of pupils of all ages enrolled in junior high school education relative to the total school-age population eligible for that level. Furthermore, GER plays an important role in Indonesia by providing insight on the educational attainment of the population, particularly in a context where the community has numerous challenges in accessing education, including dropout rates and discontinuation of schooling [12].

Another rationale for utilizing the mean years of schooling and gross enrollment ratio in this analysis is to accommodate the data accessibility provided by BPS-Statistics Indonesia. The data released by BPS-Statistics Indonesia possesses significant validity and dependability, establishing it as the primary reference for comprehensive analysis. Thus, this study employs the population living in poverty, mean years of schooling, and gross enrollment ratio to model the illiteracy rate.

Prior scholars, including Maharani [2] and Kurniati [6], have integrated spatial effects into their methodologies. The two researchers concentrated their investigation on spatial effects, particularly addressing the point approach. Besides the point approach, an alternative analytical method that incorporates spatial effects is the area approach. The area approach is an investigation of spatial effects that examines intersections between neighboring locations [13]. The spatial effect of a model is defined by the inclusion of a weighted matrix (**W**). The spatial model's value is in its ability to convey information regarding the direct, indirect, and overall effects of the independent variables [14], [15], [16]. The author of this work aims to investigate the illiteracy rate using spatial regression methodology. This research differs from prior studies in its years of observation and analysis method. This research intends to model the illiteracy rate in Indonesia for 2024 using spatial regression methodology.

2. RESEARCH METHODS

2.1. Data Source

This research makes use of information from the BPS-Statistics Indonesia website, it is available at https://www.bps.go.id. The observation year for this study is 2024. The data set includes a variety of demographic and economic metrics that are critical for evaluating national trends. By examining this data, we want to provide insight into Indonesia's current socioeconomic situation.

2.2. Research Variables

The research variables used to model the illiteracy rate in Indonesia with spatial regression are presented in **Table 1.**

Table 1. Research Variables

Variable	Unit	Code
Illiteracy Rate	Percent	Y
Number of People Living in Poverty	Soul	X_1
Mean Years of Schooling	Year	X_2
Gross Enrollment Ratio	Percent	X_3

Figure 1 illustrates the relationship between the variables.

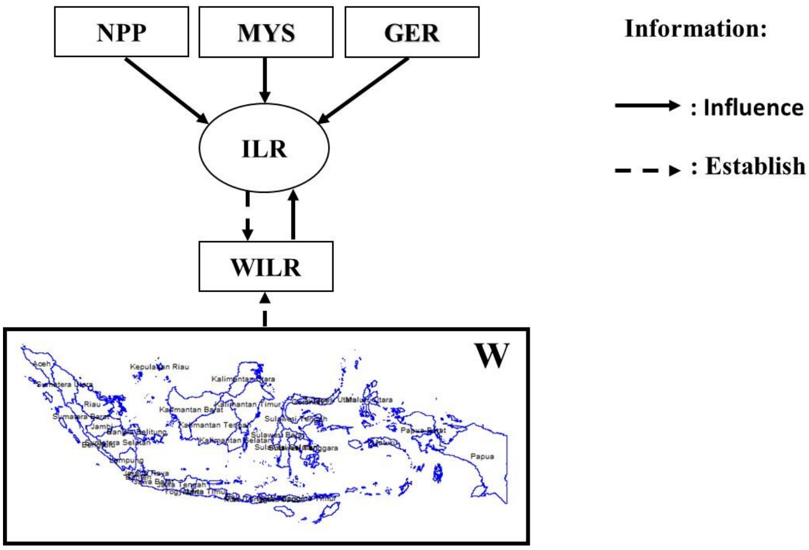


Figure 1. Relationship Between Variables

2.3. Sampling Techniques

This research employed a purposive sampling technique. In purposive sampling, a non-probability sampling method, participants are intentionally chosen by the researcher based on characteristics that align with the specific objectives of the research **[17]**.

2.4. Data Collection Technique

This research use the documentation technique for data collecting. This documentation method seeks to gather data and information from diverse sources, including documents, archives, journals, or websites pertinent to the topic [18].

2.5. Data Analysis Method

An analytical technique for evaluating the relationship between a dependent variable and one or more independent variables with respect to location or geographical factors is called spatial regression. Spatial effects are examined using the point method and the area method. The point technique employs distance data as its metric, whereas the area approach utilizes the overlap between neighboring sites. Spatial modeling with an area approach encompasses the Spatial Autoregressive Model (SAR), Spatial Durbin Model (SDM), and Spatial Error Model (SEM) [19], [20].

a. Spatial Autoregressive Models (SAR)

SAR is one of the spatial models with an area approach by taking into account spatial lag's impact just on the dependent variable. The SAR model's shape is visible in **Equation (1)**.

$$\begin{gathered} \boldsymbol{y} = \boldsymbol{\rho W y} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}_{\varepsilon}^{2}\mathbf{I}_{n}) \end{gathered} \tag{1}$$

where y is a response variable vector, ρ is the response variable's spatial lag parameter coefficient, W is a matrix of spatial weighting, X is a matrix of variables that predict, β is a vector of the coefficients for the regression parameters, ε is vector of error in the SAR model, I_n is a vector with the number 1 in it.

b. Spatial Durbin Model (SDM)

SDM is a spatial regression model with a spatial lag on the response variable (y) that resembles SAR. However, a spatial lag in the predictor variable (x) is a characteristic of SDM. The SDM model's shape is visible in **Equation (2)**.

$$\begin{gathered} \boldsymbol{y} = \boldsymbol{\rho W y} + \boldsymbol{a}\mathbf{1}_{n} + \boldsymbol{X\beta} + \boldsymbol{WX\theta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^{2}\mathbf{I}_{n}) \end{gathered} \tag{2}$$

where a is a fixed parameter, θ is a vector of the predictor variables' spatial lag parameters, ε is a vector of errors.

c. Spatial Error Model (SEM)

SEM can be applied when there is a spatial correlation between error, or when the error value at one site is connected with the error value at nearby locations. The form of the error at site i in the SEM model is a function of the error at position j, where j is a location that is close to location i.. The SEM model's shape is visible in **Equation (3)**.

$$\begin{gathered} \boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} \\ \boldsymbol{u} = \boldsymbol{\lambda W u} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}_{\varepsilon}^{2}\mathbf{I}_{n}) \end{gathered} \tag{3}$$

where u is an error vector that affects spatial, λ is spatial error parameter coefficient.

2.6. Steps of Analysis

The stages of spatial regression analysis in detail include (1) conducting a descriptive exploration of each variable, (2) examining the relationships among variables, (3) estimating parameters using the multiple linear regression method, and (4) establishing the spatial weighted matrix (**W**), which encompasses Rook contiguity, Inverse, and *k*-nearest neighbor weighted matrices for *k*=1, (5) performing spatial dependency tests utilizing the Moran Index and Lagrange Multiplier tests to identify spatial regression models, (6) estimating parameters through spatial regression techniques, (7) calculating the Akaike Information Criterion (AIC) value for each model, (8) selecting the ideal model based on the smallest AIC value, and (9) interpreting the results of the model.

3. RESULTS AND DISCUSSION

3.1. Data Exploration

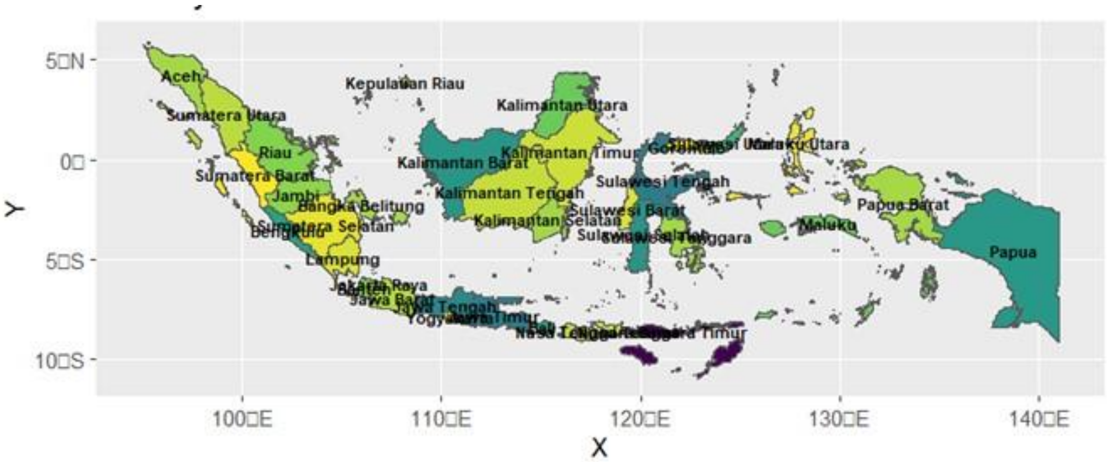
The first step in modeling the illiteracy rate in Indonesia is to explore the data used in this study. The outcomes of data exploration are displayed in **Table 2.**

Table 2. Characteristics of Research Variables

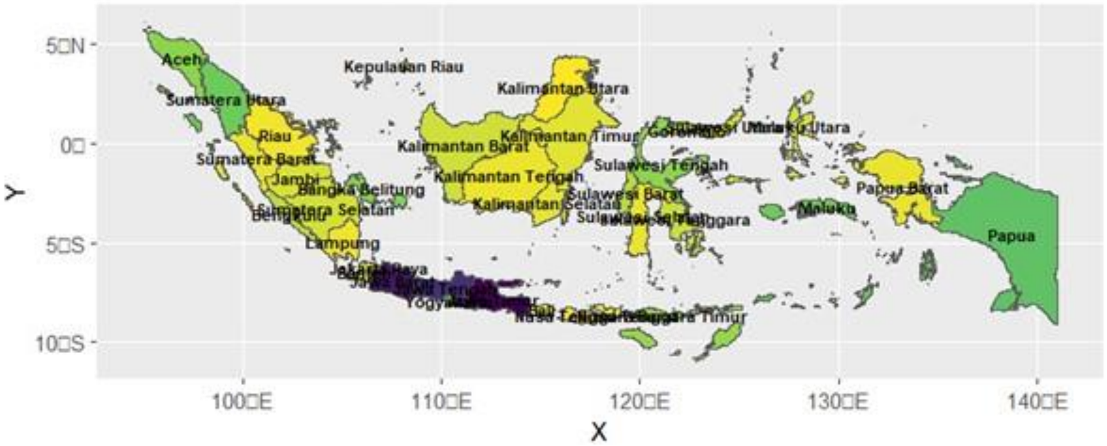
Variable	Statistic			
	Max.	Min.	Average	Standard Deviation
ILR	10.17	0.25	2.58	2.17
NPP	3893.82	41.11	683.42	986.78
MYS	11.49	8.28	9.48	0.75
GER	100.21	82.07	91.31	4.39

The analysis of illiteracy rate (ILR) in Indonesia in 2024 has an average value of 2.58%. The highest illiteracy rate reached 10.17% while the minimum value reached 0.25%. For the number of people living in poverty (NPP) in Indonesia, the average in the same period is 683.42 people, with the maximum value reaching 383.82 people and the minimum value reaching 41.11 people. Meanwhile, the mean years of schooling (MYS) in Indonesia in 2024 shows an average value of 9.48 years with the highest value reaching 11.49 years while the minimum value reaches 8.28 years. the gross enrollment ratio (GER) in Indonesia in 2024 shows an average of 91.3%, with the highest possible value of of 100.21% and a minimum amount of 82.07%.

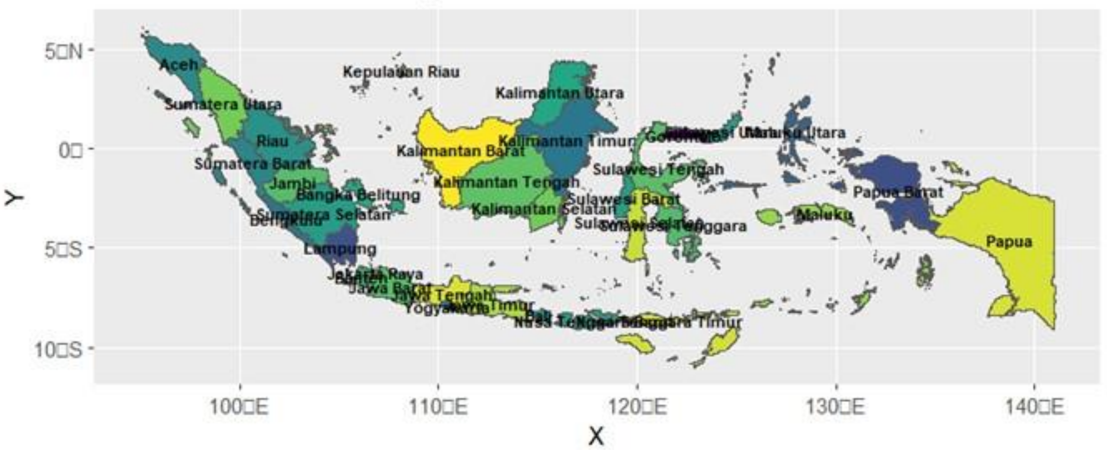
Figure 2 displays the spatial distribution of each variable: (a) ILR, (b) NPP, (c) MYS, and (d) GER. **Figure 2(a)** indicates that West Nusa Tenggara has the highest illiteracy rate, while North Sulawesi has the lowest. **Figure 2(b)** indicates that East Java has the highest population of impoverished individuals, while North Kalimantan has the lowest. **Figure 2(c)** indicates that DKI Jakarta has the highest average level of schooling, whereas West Kalimantan has the lowest. **Figure 2(d)** indicates that Papua has the highest gross enrollment rate, whereas West Sulawesi has the lowest. Evidence indicates the presence of spatial dependencies among provinces in Indonesia across all observed variables. The dependence is evidenced by the comparable coloration of adjacent provinces on the variable distribution map. Provinces exhibiting comparable values are situated in proximity to one another.


Aceh
Sumatera Utara
Riau
Sumatera Barat
Jambi
Bangka Belitung
Sumatera Selatan
Lampung
Kepulauan Riau
Kalimantan Utara
Kalimantan Barat
Kalimantan Timur
Kalimantan Tengah
Kalimantan Selatan
Sulawesi Tengah
Sulawesi Barat
Jawa Barat
Jawa Tengah
Papua Barat
Maluku
Papua
5□N
0□
5□S
10□S
100□E
110□E
120□E
130□E
140□E
X
Y

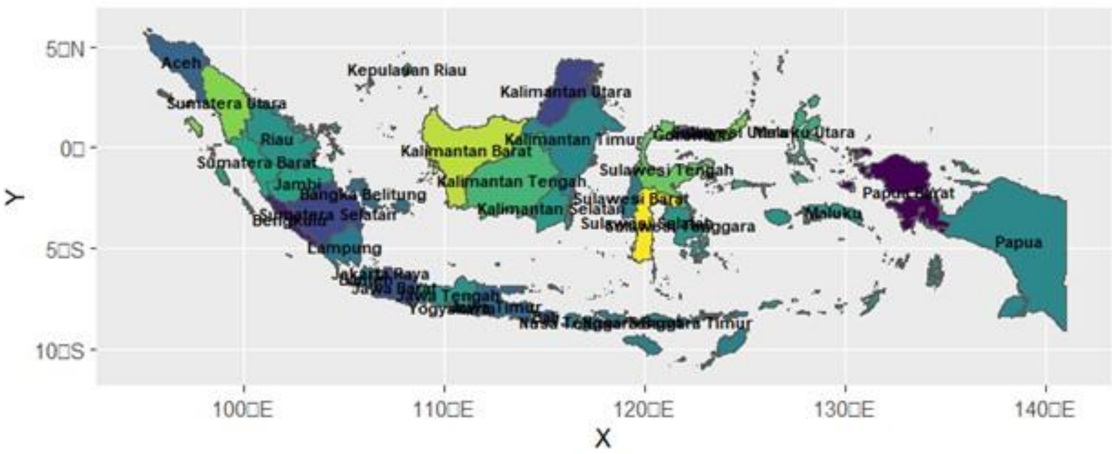
(a)


Aceh
Sumatera Utara
Riau
Sumatera Barat
Jambi
Bangka Belitung
Sumatera Selatan
Lampung
Kepulauan Riau
Kalimantan Utara
Kalimantan Barat
Kalimantan Timur
Kalimantan Tengah
Kalimantan Selatan
Sulawesi Tengah
Sulawesi Barat
Papua Barat
Maluku
Papua
5□N
0□
5□S
10□S
100□E
110□E
120□E
130□E
140□E
X
Y

(b)


Aceh
Sumatera Utara
Riau
Sumatera Barat
Jambi
Bangka Belitung
Sumatera Selatan
Lampung
Kepulauan Riau
Kalimantan Utara
Kalimantan Barat
Kalimantan Timur
Kalimantan Tengah
Kalimantan Selatan
Sulawesi Tengah
Sulawesi Barat
Jawa Barat
Jawa Tengah
Papua Barat
Maluku
Papua
5□N
0□
5□S
10□S
100□E
110□E
120□E
130□E
140□E
X
Y

(c)



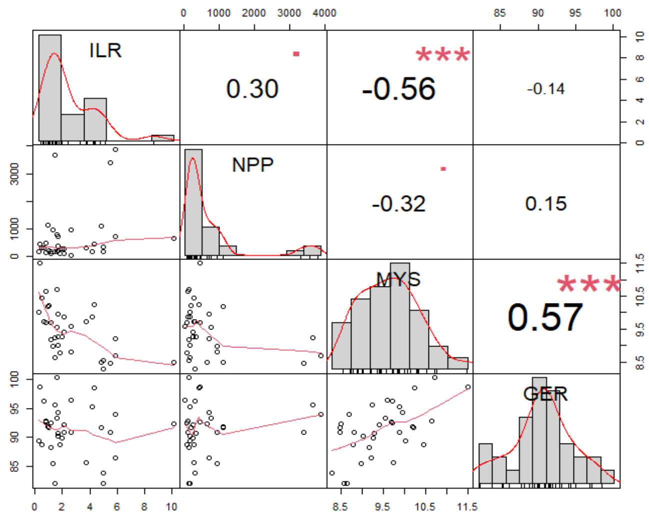
(d)

Figure 2. Distribusi Maps: (a) ILR, (b) NPP, (c) MYS, (d) GER

Source: The image is processed utilizing R Software

3.2. Correlation Between Variables

Figure 3 presents a matrix scatterplot that depicts the correlation among the variables in the study: ILR, NPP, MYS, and GER. The graphic illustrates the Pearson correlation coefficient between variables, along with the significance level indicated by an asterisk (*).



*** significant at α = 0%,

Figure 3. Scatterplot Relationship Between Variables

Source: The image is processed utilizing R Software

Figure 3 depicts the outcomes of the correlation study among the ILR, NPP, MYS, and GER variables, shown as a scatterplot. The diagonal and bottom triangles of the scatterplot illustrate the distribution of each variable. The upper triangle presents the correlation coefficient of Pearson and its level of relevance. The correlation coefficient of Pearson for the NPP variable concerning ILR is 0.30. This figure indicates that NPP exerts a favorable and substantial influence on ILR. A greater NPP corresponds to a greater ILR. The Pearson correlation coefficient for the MYS variable with respect to ILR is -0.56, which is significant at α = 0%. MYS exerts a detrimental and considerable influence on ILR. A higher MYS correlates with a lower ILR. The Pearson correlation coefficient for the GER variable with respect to ILR is -0.14. GER adversely affects ILR. An increase in the GER typically correlates with a drop in the ILR.

3.3. Multiple Linear Regression Parameter Estimation

Parameters in models of multiple regression are estimated utilizing the OLS (ordinary least square) estimator approach. This OLS method aims to minimize the sum of squared errors to obtain a regression model.. Based on **Table 3**, the regression model built is quite good in the multicollinearity test, shown in the Variance Inflation Factor (VIF) value acquired shows that each variable has a value below 10. Thus, it can be assumed that the regression model formed does not appear multicollinearity in each independent variable, so that these variables are suitable for use in further regression analysis.

Table 3. Parameter Estimation Results with Multicollinearity Test

Variable	Estimate
NPP	1.210
MYS	1.769
GER	1.651

3.4. Weighted Matrix

The spatial weight matrix is employed to ascertain the weight amid observed places determined by the neighboring relationship between places and is symbolized by (**W**) [21], [22], [23]. There are 3 weighted matrices used, namely: Rook Contiguity, Inverse Distance, and *K*-Nearest Neighbor for $k = 1$.

a. Rook Contiguity

The observation region is established using the intersecting edges along with the angle is neglected to be considered. The rook contiguity results is seen in **Figure 4.**

Figure 4. Rook Contiguity
Source: The image is processed utilizing R Software

b. Invers Distance

The inverse-distance weighting method is based on the assumption that the influence of one location on another depends on the separation between the two places.

The influence decreases with increasing distance between the two locations. In this case, the weighting matrix used is a matrix that takes into account the distance between one location and another. The inverse-distance results is seen in **Figure 5**.

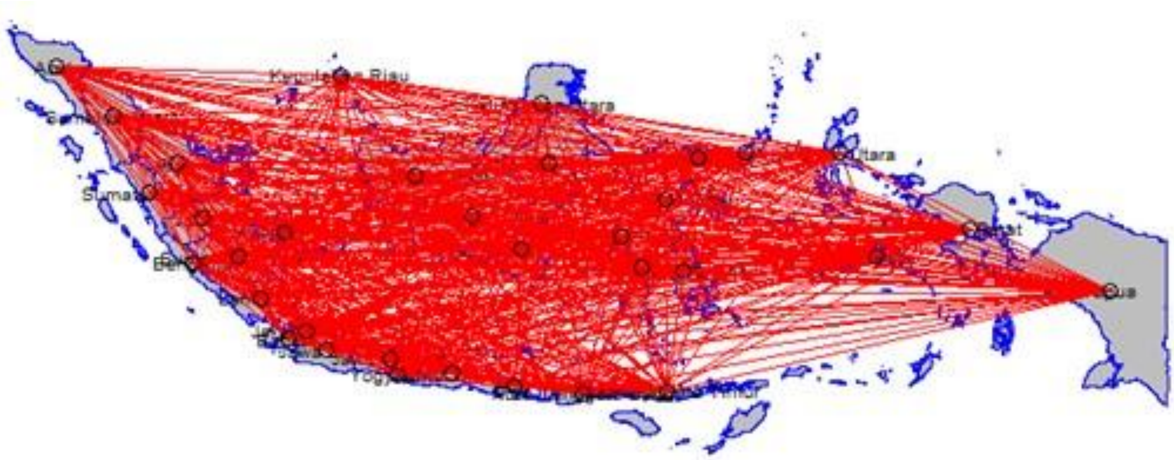


Figure 5. Invers Distance
Source: The image is processed utilizing R Software

c. K-Nearest Neighbor

Each row i in the k-nearest neighbor spatial weighting matrix has k columns j with element 1 and the other columns are 0. The k-nearest neighbor results is seen in **Figure 6**.

Figure 6. K-Nearest Neighbor
Source: The image is processed utilizing R Software

3.5. Spatial Dependency Analysis

Spatial dependency testing is conducted to see whether observations in a location affect observations in other locations that are located nearby. A statistical test is Moran's I that aims to measure the the relationship between locations on one variable or spatial dependence. To check for spatial dependence or autocorrelation between observations or places, utilize the Moran's I coefficient. Lagrange Multiplier (LM) a test to ascertain if the model has spatial effects on the dependent variable or error. The outcomes of the spatial dependency test is displayed in **Table 4**.

Table 4. Spatial Dependence Test Results with Moran I

Weighting Matrix	Moran I	p-value
Rook Contiguity	0.248	0.234
Invers Distance	-3.030	1
k-Nearest Neighbor	-0.521	0.019

Table 4 shows that in the Rook Contigiuty weighting matrix, the Moran's I value is 0.248 using a p-value of 0.234. This suggests a positive spatial correlation trend, but it is not noteworthy in terms of statistics because there is a p-value over 0.05.. In contrast, the Inverse Distance matrix yields a Moran's I value of -0.030 with a p-value 1.0, indicating the absence of data's spatial autocorrelation and the result is not significant. Meanwhile, the *K*-Nearest Neighbor matrix shows a Moran's I p-value -0.521 with a p-value of 0.019. This value is negative and statistically significant (as the p-value < 0.05), demonstrating the existence of significant negative spatial autocorrelation, where neighboring units tend to have significantly different values. Overall, of the three weighted methods, only the *k*-Nearest Neighbor matrix showed significant results in detecting spatial patterns in the data. Therefore, the *k*-Nearest Neighbor weighted matrix is the matrix used in the model of spatial regression.

Tabel 5. Results of The Spatial Dependence Test

Statistic	*k*-Nearest Neighbor	
	Estimate	p-value
Moran's I	-0.521	0.019
LMerr (SEM)	5.815	0.015
LMlag (SAR/SDM)	7.568	0.005

Table 5 shows a p-value of 0.019 and a Moran's I value of -0.521 indicating a significant negative spatial autocorrelation, which means there is a pattern of irregularity where the value at a location tends to be different from the value at the surrounding location. Furthermore, the LMerr (SEM) test results have a statistical value of 5.815 with a 0.015 p-value, and the LMlag (SAR/SDM) the test's p-value is 0.005 and its value is 7.568. The two p-values are smaller than 0.005, indicating that all three models show the existence of significant spatial effects in the data. However, based on p-value, the most suitable model is SAR/SDM.

3.6. Spatial Regression Parameter Estimation

Parameter estimation for SAR, SEM, and SDM models usually done applying the method of Maximum Likelihood Estimation (MLE). Modeling with spatial regression can be done using several models generated through the Lagrange Multiplier (LM) test which is illustrated in **Table 6**.

Table 6. SAR and SDM Estimation Results

Variable	SAR		SDM	
	Estimate	p-value	Estimate	p-value
Intercept	-3.244	0.408	-7.619	0.093
WILR (ρ)	-0.338	0.002	-0.367	0.001
NPP	-0.032	0.727	-0.099	0.278
MYS	-8.413	0.000	-8.852	0.000
GER	6.060	0.011	6.300	0.003
WNPP	-	-	0.229	0.013

Variable	SAR		SDM	
	Estimate	**p-value**	**Estimate**	**p-value**
WMYS	-	-	-0.374	0.846
WGER	-	-	2.197	0.327
AIC	**9.648**		**6.618**	

Table 6 informs that the smallest AIC value is found in the SDM. Therefore, SDM is the model utilized to simulate Indonesia's illiteracy rate, which is written in **Equation (4).**

$$\widehat{ILR} = -7{,}619 - 0{,}367WILR - 0{,}099NPP - 8{,}852MYS + 6{,}300GER + 0{,}229WNPP - 0{,}374WMYS + 2{,}197WGER \quad \textbf{(4)}$$

3.7. Assumption Test

Table 7 shows the outcomes of the normality test using two statistical methods, namely Shapiro-Wilk and Jarque-Bera Test. One crucial premise in many statistical analyses is that data is regularly distributed, which is ascertained using the normality test. In the Shapiro-Wilk test, the statistical value is 0.974 with a p-value of 0.602. Given that the p-value exceeds 0.05, there is insufficient proof to reject (H_0), which states that the data is normally distributed. In the Jarque-Bera test, the statistical value is 0.323 with a p-value of 0.850. Just like the previous test, the high p-value shows that there isn't strong proof to disprove the (H_0). Thus, the data is considered to adhere to a normal distribution.

Table 7. Parameter Estimation Results with Normality Test

Shapiro-Wilk Test		Jarque Bera Test	
Statistic	p-value	Statistic	p-value
0.974	0.602	0.323	0.850

Figure 7 shows a normal Q-Q (Quantile-Quantile) plot which utilized to determine whether the residuals of a regression model. The dots on this plot represent the residual values. If the residual data is normally distributed, then the points will follow the straight line shown by the dotted line on the plot. From **Figure 7**, it can be seen that most of the points follow a straight line pattern, which indicates that the assumption of residual normality is relatively fulfilled. Therefore, the regression model used can be considered valid in terms of the normality assumption.

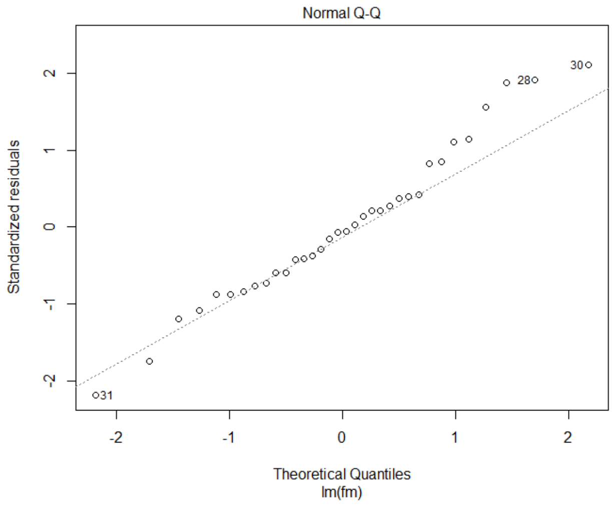


Figure 7. **Normal Test Chart**
Source: The image is processed utilizing R Software

Table 8 shows the results of estimating the model parameters by conducting an autocorrelation test using two methods, namely the Breusch-Godfrey Test and the Durbin-Watson Test. The Breusch-Godfrey (BG) test produces a BG test statistical value of 0.029 with a p-value of 0.862. Since there is insufficient data to rule out the null hypothesis because the p-value is higher than 0.05, which means there is no autocorrelation in the residuals of the model. Meanwhile, the Durbin-Watson (DW) test results show a DW test statistical value of 2.055 with a p-value of 0.524. Additionally, the absence of autocorrelation in the residuals is indicated by the Durbin-Watson value near 2. These two test results consistently show that the regression model does not contain autocorrelation problems. Hence, the model is considered to fulfill the classical assumptions related to error independence.

Table 8. **Parameter Estimation Results with Autocorrelation Test**

Breusch Godfrey Test		**Durbin Watson Test**	
Statistic	p-value	Statistic	p-value
0.029	0.862	2.055	0.524

3.8. Marginal Effect

The effect of a significant variations in the independent and dependent variables is measured based on the marginal direct effects and indirect effects. The marginal effects of MYS and GER are presented in **Table 9**.

Table 9. **Marginal Effects of Independent Variables**

Variable	**Direct**	**Indirect**	**Total**
MYS	-9.049	2.759	-6.290
GER	6.519	-1.988	4.531

Table 9 states that MYS in a province can directly reduce ILR in that province by -9.039. If MYS in a province increases, then indirectly ILR in the neighboring province also increases by 2.759. This means that an increase in MYS in a neighboring province actually has a positive influence (increase) on ILR in the province being observed, so there is an

inter-regional effect. This study dies carried out by Kevin et.al [24]. The mean years of schooling reflects the level of education attained by the population in a region. An increase in mean years of schooling indicates that people have better access to formal education, which has a direct impact on illiteracy rates. More educated people tend to have a better understanding of healthy lifestyles and utilize health services effectively, which ultimately determines quality of life and longevity.

An increase in GER in a province can directly increase ILR by 6.519 in that province. If the GER in a province increases, the illiteracy rate in the province adjacent to the observed province will indirectly decrease by -1.988. The findings of this investigation are consistent with the study carried out by Ritonga et.al [25]. Increasing gross enrollment ratio is one of the most effective strategies in reducing illiteracy. As more children receive education, especially primary education, the overall literacy of the community will improve significantly.

4. CONCLUSION

The analysis results indicate that a) the chosen weighting matrix is *k*-Nearest Neighbor, b) the spatial model employed to assess the illiteracy rate in Indonesia is the spatial Durbin model (SDM), and c) MYS and GER within a province significantly influence the illiteracy rate in that province, which may indirectly elevate the illiteracy rate in neighboring regions. Spatial modeling indicates that the phenomena of illiteracy is impacted not just by the internal characteristics of an area, but moreover by the conditions of adjacent regions. This connection underscores the necessity of considering geographical interdependence when tackling educational inequality. Policymakers must employ a comprehensive approach that considers both local and regional elements to effectively diminish illiteracy rates across provinces. Consequently, it is imperative to formulate education enhancement plans in a cohesive and cross-regional approach, including the augmentation of average years of schooling and gross enrollment rates. Regional and cooperation strategies among provinces are essential for effectively mitigating illiteracy in Indonesia.

Acknowledgments

The writers would like to express their gratitude to everyone who helped with the development of this article. Special thanks go to the supervisor for his guidance and direction, as well as to the BPS-Statistics Indonesia that have provided access to the information needed in this study.

Funding Information

This research received no external funding.

Author Contributions Statement

Table 10. The Author's Contribution

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Alvizar Syamsul Balda	✓	✓	✓		✓	✓		✓	✓		✓		✓	✓
Alfira Mulya Astuti		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		
Parhaini Andriani				✓			✓			✓		✓		

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

Conflict of Interest Statement

Authors state no conflict of interest.

Informed Consent

Not available.

Ethical Approval

Not available.

Data Availability

The data that support the findings of this study are available on request from the corresponding author, AMA.

REFERENCES

[1] B. S. Lake and E. D. Utami, "Variabel-Variabel yang Memengaruhi Angka Buta Huruf (ABH) di Provinsi Papua Tahun 2020," in *Seminar Nasional Official Statistics*, 2022, pp. 113–121.

[2] R. Maharani and W. S. Winahju, "Pemodelan Angka Buta Huruf di Provinsi Sumatera Barat Tahun 2014 dengan Geographically Weighted Regression," *J. SAINS DAN SENI ITS*, vol. 5, no. 2, pp. 361–365, 2016.

[3] BPS, "Angka Buta Huruf Menurut Provinsi dan Kelompok Umur (Persen), 2024," Badan Pusat Statistik Republik Indonesia. Accessed: Nov. 03, 2024. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/MTAyIzI=/angka-buta-aksara-menurut-provinsi-dan-kelompok-umur.html

[4] V. D. Y. Alvira Oktavia Safitri, "Upaya Peningkatan Pendidikan Berkualitas di Indonesia: Analisis Pencapaian Sustainable Development Goals (SDGs)," *J. Basicedu*, vol. 6, no. 4, pp. 7096–7106, 2022.

[5] Y. Aprianti and A. M. Astuti, "Systematic Literature Reviews : Analisis Angka Buta Huruf," *J. Math Tadris*, vol. 4, no. 2, pp. 182–201, 2024.

[6] D. F. Kurniati, "Regresi Spasial dengan Pendekatan Geographically Weighted Logistic Regression (GWLR) Studi Kasus Angka Buta Huruf Tahun 2012 di Kota dan Kabupaten Provinsi Jawa Barat," Universitas Pendidikan Indonesia, 2014.

[7] A. P. Satvika, N. L. P. Suciptawati, and M. Susilawati, "Memodelkan Angka Buta Huruf di Provinsi Nusa Tenggara Barat," *J. Educ.*, vol. 06, no. 01, pp. 1499–1507, 2023.

[8] A. Andiyono, R. D. Bekti, and E. Irwansyah, "Analisis Faktor yang Mempengaruhi Angka Buta Huruf Melalui Geographically Weighted Regression: Studi Kasus Propinsi Jawa Timur," *ComTech Comput. Math. Eng. Appl.*, vol. 4, no. 1, pp. 443–449, 2013.

[9] A. Yakin, "Pemodelan angka buta huruf di indonesia," Universitas Islam Negeri Mataram, 2024.

[10] S. Sirilius, "Hubungan Antara Pendidikan, Pengangguran, dan Pertumbuhan Ekonomi dengan Kemiskinan," *J. Ekon. KUANTITATIF Terap.*, vol. 10, no. 2, pp. 59–71, 2017.

[11] Intan Choirunnisa, "Pengaruh PMA, pertumbuhan ekonomi, belanja pemerintah, tingkat

kemiskinan dan pendidikan terhadap indeks pembangunan manusia (IPM) Kabupaten Kota di Provinsi Jawa Barat," Universitas Islam Negeri Syarif Hidayatullah, 2020.

[12] E. K. A. R. Amalia, "Kondisi Pemerataan Pendidikan," *J. Muhammadiyah, Malang*, vol. 01, no. 02, pp. 1–34, 2007.

[13] A. M. Astuti, "Model Panel Dinamis Durbin Spasial Pendekatan Modifikasi Generalized Estimating Equation," Institut Teknologi Sepuluh Nopember Suarabaya, Surabaya, 2023.

[14] I. G. N. M. Jaya and Y. Andriyana, *Analisis Data Spasial: Perspektif Bayesian*. Sumedang: Algaprint Jatinangor, 2020.

[15] J. P. LeSage and R. . Pace, *Introduction to Spatial Econometrics*. Boca Raton: CRC Press Taylor & Francis Group, 2009.

[16] A. M. Astuti, H. Ratu, and P. Negara, "Analysis of Human Development Index in West Nusa Tenggara Province with Spatial Panel Model," *Int. J. Comput. Sci. Appl. Math.*, vol. 10, no. 2, pp. 113–118, 2024.

[17] P. G. Subhaktiyasa, "Menentukan Populasi dan Sampel : Pendekatan Metodologi Penelitian Kuantitatif dan Kualitatif," *J. Ilm. Profesi Pendidik.*, vol. 9, no. 4, pp. 2721–2731, 2024.

[18] Ardiansyah, Risnita, and M. S. Jailani, "Teknik Pengumpulan Data Dan Instrumen Penelitian Ilmiah Pendidikan Pada Pendekatan Kualitatif dan Kuantitatif," *J. IHSAN J. Pendidik. Islam*, vol. 1, no. 2, pp. 1–9, 2023, doi: 10.61104/ihsan.v1i2.57.

[19] L. Anselin, "Spatial Models in Econometric Research," *Oxford Res. Encycl. Econ. Financ.*, no. March, pp. 1–54, 2021, doi: 10.1093/acrefore/9780190625979.013.643.

[20] N. Debarsy and C. Ertur, "Interaction matrix selection in spatial autoregressive models with an application to growth theory," *Reg. Sci. Urban Econ.*, vol. 75, no. November, pp. 49–69, 2019, doi: 10.1016/j.regsciurbeco.2019.01.002.

[21] A. M. Astuti, Setiawan, I. Zain, and J. D. T. Purnomo, "A Review of Panel Data on Spatial Econometrics Models," *J. Phys. Conf. Ser.*, vol. 1490, no. 1, pp. 1–13, 2020, doi: 10.1088/1742-6596/1490/1/012032.

[22] A. M. Astuti, Setiawan, I. Zain, and J. D. T. Purnomo, "The extended algorithm for quasi maximum likelihood parameter estimation," *AIP Conf. Proc.*, vol. 2360, no. September, pp. 1–5, 2021, doi: 10.1063/5.0059509.

[23] A. M. Astuti, Setiawan, I. Zain, and J. D. T. Purnomo, "A modified generalized estimating equation approach for simultaneous spatial durbin panel model: Case study of economic growth in ASEAN countries," *Decis. Sci. Lett.*, vol. 12, no. 2, pp. 369–388, 2023, doi: 10.5267/j.dsl.2023.1.001.

[24] A. V. Kevin, A. Bhinadi, and A. Syari'udin, "Pengaruh PDRB, Angka Harapan Hidup, Dan Rata Rata Lama Sekolah Terhadap Kemiskinan Di Kabupaten/Kota Provinsi Jawa Tengah Tahun 2013-2021," *SIBATIK J. J. Ilm. Bid. Sos. Ekon. Budaya, Teknol. dan Pendidik.*, vol. 1, no. 12, pp. 2959–2968, 2022, doi: 10.54443/sibatik.v1i12.482.

[25] J. R. Ritonga, S. Ginting, E. Naibaho, A. Situmorang, and K. Kunci, "Pengaruh Angka Melek Huruf Dan Angka Partisipasi Sekolah Terhadap Jumlah Kemiskinan Di Provinsi Sumatera Utara Tahun 2017-2023," *JSEH (Jurnal Sos. Ekon. dan Humaniora)*, vol. 10, no. 4, pp. 676–681, 2024.