

Modeling Factors That Influence Poverty Level In Sumatera Using Spatial Regression Models

Deandra Sakirah Putri¹, Yekti Widyaningsih^{1*}

¹Universitas Indonesia, Jakarta, Indonesia

*Corresponding author: yekti@sci.ui.ac.id

Abstract

Poverty is a condition of a person's inability to fulfill basic needs such as food, clothing, shelter, school and health. Poverty is still a major unresolved problem in Indonesia, including the island of Sumatra. Based on BPS data in 2022, the number of poor people in Sumatera is the second after Java. This research was conducted with the aim of analyzing the factors that explain the level of poverty in Sumatera. The method that will be used is spatial regression modeling, including the General Nested Spatial Model (GNSM), Spatial Autoregressive Model (SAR), Spatial Error Model (SEM) and Spatial Autoregressive and Moving Average Model (SARMA). The dependent variable is Poverty Level (Miskin) and the independent variables include Population Density (KP), Average Years of Schooling (RLS), Human Development Index (IPM), Open Unemployment Rate (TPT) and Regency Minimum Wage (UMK). Based on the result of this study, the Spatial Autoregressive and Moving Average Model is the best model, with the smallest AIC value, namely 305.32. The only factor that has a significant influence on the poverty level is the Human Development Index (IPM).

Keywords: Spatial Regression Models, Poverty Level, Sumatera, and SARMA

Article History:

Received : 31 January 2026

Revised : 16 April 2026

Accepted : 21 July 2026

© 2026 Putri and Widyaningsih. Published

by Program Studi Statistika Fakultas MIPA

Universitas Lambung Mangkurat.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



How to Cite:

D. S. Putri and Y. Widyaningsih, *Modeling Factors That Influence Poverty Level In Sumatera Using Spatial Regression Models*, RAGAM: Journal of Statistics and Its Application, vol. 5, no. 2, pp. 111-129, 2026, [doi: 10.20527/ragam.v5i2.18241](https://doi.org/10.20527/ragam.v5i2.18241)

1. INTRODUCTION

Poverty remains a major unresolved problem in several developing countries, including Indonesia. The term poverty arises when a person or group of people are unable to meet the level of economic prosperity that is considered the minimum requirement for a certain standard of living [1]. Poverty can occur due to a large population and a lack of employment opportunities. This is one of the factors causing the high poverty level in Indonesia. In addition to employment, poverty is also related to inadequate levels of education and health. Unequal income distribution is also a contributing factor, as it affects social inequality between upper- and lower-income communities. Poverty is a fundamental issue that receives serious attention from the government, because poverty alleviation is an important aspect of achieving public welfare.

The number of poor people in Indonesia continues to increase every year. Based on Statistics Indonesia (BPS) data in 2022, the number of poor people on Sumatra Island ranks second after Java Island. This attracts the attention of researchers to examine the influencing factors. This study aims to analyze the factors that affect poverty levels on Sumatra Island. The method used is spatial regression modeling.

The selection of the spatial regression model to be used is adjusted to the conditions of the data to be analyzed. Several spatial models include Spatial Autoregressive Models (SAR), Spatial Error Models (SEM), Spatial Autoregressive and Moving Average (SARMA), General Nested Spatial Models (GNSM). Before modeling, a spatial effect test was performed using Moran's I. This was done to ensure the presence of a spatial effect in the data. If there is no spatial influence, a global model can be used, namely a linear regression model using Ordinary Least Squares (OLS) [2].

Many previous studies have discussed the factors influencing poverty, such as the study conducted by Laga and Wiwin (2022), which analyzed the factors influencing poverty. The study used variables such as provincial minimum wage, human development index, economic growth, and unemployment as factors influencing poverty. The method used was multiple linear regression analysis with the Ordinary Least Squares (OLS) model. The results showed that provincial minimum wage had a negative and significant effect on poverty, human development index had a positive but insignificant effect on poverty, economic growth had a negative but insignificant effect on poverty, unemployment had a positive and significant effect on poverty, and all independent variables simultaneously influenced poverty in East Java [3].

Furthermore, a study conducted by Adi and Eny (2022) examined the effect of gross regional domestic product (GRDP) and average years of schooling on poverty in urban areas. The variables used were gross regional domestic product (GRDP) and average years of schooling, using multiple linear regression and descriptive statistics. The results of this study indicate that gross regional domestic product has a negative and significant effect on poverty, while average years of schooling have a positive but insignificant effect on poverty [4].

Research on factors influencing poverty and unemployment in Jambi City, conducted by Harlik, Amri, and Hardiani (2013), used the variables GRDP, unemployment rate, poverty level, life expectancy, Jambi provincial minimum wage (UMP), population density, and literacy rate. The method used was multiple linear regression analysis, with the following findings: Population density, education level, and unemployment rate significantly influence poverty levels [5].

In contrast to previous studies that focused on poverty factors in East Java and Jambi, this research conducts an analysis across Sumatra, covering 145 regencies and cities. While prior research utilized multiple linear regression, this study employs spatial regression models, including the Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), Spatial Autoregressive and Moving Average (SARMA), and the General Nested Spatial Model (GNSM). The author retrieves secondary data from the BPS-Statistics Indonesia website, supplemented by spatial data to visualize the distribution of poverty levels across Sumatra. The variables include poverty levels, population density, average years of schooling, the open unemployment rate, minimum wage, and the Human Development Index (HDI) for each regency and city in Sumatra for the year 2022.

2. LITERATURE REVIEW

2.1 Spatial Regression

Spatial regression is a development of classical linear regression methods. This development is based on Tobler's First Law of Geography, which states that everything is related to everything else, but things that are closer have a greater influence than things that are farther away. This indicates that there is a spatial effect on the analyzed data [6].

Spatial Dependence

1. Moran's I Test

The Moran's I test is conducted to test whether there is autocorrelation in spatial data. The value is obtained by multiplying the deviation of a variable value in a region by the mean value of the variable.

2. Lagrange Multiplier Test

The Lagrange Multiplier test is conducted to test whether there is autocorrelation in the lag, error, or both.

Spatial Weight Matrix

The spatial weighting matrix (W) is an $n \times n$ matrix containing elements w_{ij} that describe the spatial relationship between the i -th location and the j -th location. The elements w_{ij} are defined as:

$$w_{ij} = \begin{cases} 0, & \text{if } i \text{ and } j \text{ do not have common edge} \\ a, a \neq 0, & \text{if } i \text{ and } j \text{ have common edge} \end{cases}$$

with common edge meaning there is contact or interaction between regions.

2.2 Spatial Regression Models

Spatial regression models address spatial dependence between observations in regression analysis. These models include the General Nested Spatial Model (GNSM), Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), and Spatial Autoregressive and Moving Average Model (SARMA).

General Nested Spatial Model (GNSM)

The GNSM model is a spatial regression model that includes spatial dependencies in the dependent variable, independent variables, and error ($\rho \neq 0, \gamma \neq 0$ dan $\lambda \neq 0$) [7].

$$Y = \rho WY + X\beta + WX\gamma + u \quad (1)$$

Spatial Autoregressive Model (SAR)

The SAR model is a spatial regression model in which there is spatial dependency on the dependent variable ($\rho \neq 0$) [8].

Spatial Error Model (SEM)

The SEM model is a spatial regression model in which there is spatial dependency on its errors [8].

Spatial Autoregressive and Moving Average (SARMA)

The SARMA model is a general model of spatial regression that contains spatial dependencies on the dependent variable and the error ($\rho \neq 0$ dan $\lambda \neq 0$) [9].

Selection of the Best Model

There are two criteria for selecting the best model:

1. Coefficient of determination (R^2)

The greater the R^2 value, the better the model obtained in explaining the dependent variable.

2. Akaike Info Criterion (AIC)

The smaller the AIC value, the better the model obtained [10].

3. RESEARCH METHOD

The data used in this study are secondary data obtained from the Statistics Indonesia (BPS) website in 2022 [11].

Table 1. Explanation of the variables used

No	Variable	Explanation	Unit	Data Source
1.	Poverty Level (Miskin)	Number of poor people per thousand population by district/city	Souls/1000 population	Central Statistics Agency (BPS) website 2022
2.	Population Density (KP)	Population density by district/city	Souls/ km^2	Central Statistics Agency (BPS) website 2022
3.	Average Years of Schooling (RLS)	Average length of schooling by district/city	Year	Central Statistics Agency (BPS) website 2022
4.	Open Unemployment Rate (TPT)	Open unemployment rate by district/city	Percent	Central Statistics Agency (BPS) website 2022
5.	Human Development Index (IPM)	Human development index by district/city	skor	Central Statistics Agency (BPS) website 2022
6.	Regency Minimum Wage (UMK)	District minimum wage by district/city	x Rp 1000	Central Statistics Agency (BPS) website 2022

This study utilizes data from 145 regencies and cities across Sumatra. The dependent variable is the Poverty Level (Miskin), while the independent variables include Population Density (KP), Average Years of Schooling (RLS), Open Unemployment Rate (TPT), Human Development Index (IPM) and Regency Minimum Wage (UMK). This research employs spatial regression as its main analytical tool. Prior to model estimation, the author performs a descriptive analysis and map visualization to explore the data. Subsequently, the study conducts Moran's I test to identify spatial dependence and defines the spatial weight matrix. Finally, the Lagrange Multiplier (LM) test is applied to select the most appropriate spatial model for the dataset.

4. RESULT

Descriptive statistics are shown using measures of central tendency and data dispersion presented in Table 2.

Table 2. Descriptive Statistics

Variable	Minimum	Maximum	Mean	Median	Std. Deviation	Variance
Miskin	22.134221	223.54564	100.6675	93.483344	43.969501	1933.317
KP	18.611404	9740.38	653.61447	113.31083	1500.7668	2252301
RLS	5.88	13.03	9.0094483	8.7	1.2964207	1.6807066
TPT	0.26	11.69	4.67267	4.55	2.07139	4.29066
IPM	62.93	86.28	71.66566	70.55	4.360727	19.01594
UMK	2238.094	3648.714	2834.9969	2770.794	334.09908	111622.19

Based on Table 2, the average poverty level on Sumatra Island in 2022 was 100.6675 (thousand) people, with the lowest poverty level in Sawahlunto City and the highest poverty level in West Nias Regency. The variance of poverty level data on Sumatra Island was 1933.317. The following is a mapping of each variable used in this study.

4.1 Poverty Level

The poverty level on Sumatra Island increased in 2022. Figure 1 shows a map of the poverty level on Sumatra Island, with green indicating low poverty levels and red indicating high poverty levels. The lowest poverty level is found in Sawahlunto City (blue arrow), while the highest poverty level is found in West Nias Regency (red arrow).

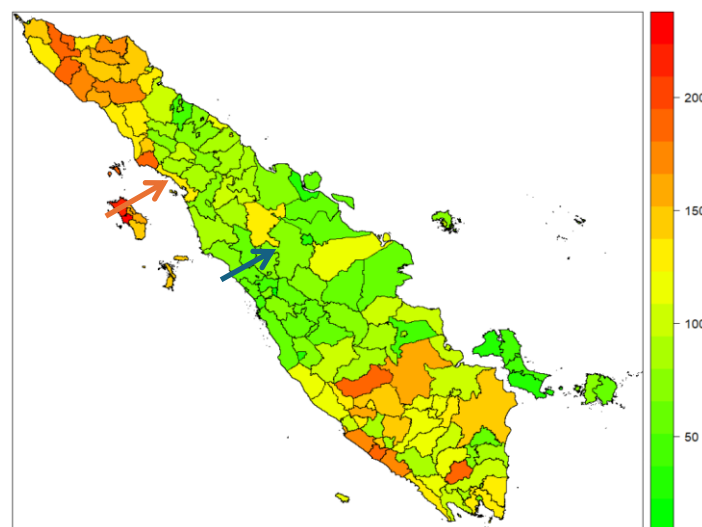


Figure 1. Poverty Level Map

4.2 Population Density (KP)

In 2022, the population density on Sumatra Island was 653,614,47 people/km². Figure 2 shows a map of the distribution of population density on Sumatra Island, with green indicating low population density and red indicating high population density. The dominant color on the map is green, indicating that the population density in each district/city is relatively low. The lowest population density is in Gayo Luwes Regency (blue arrow), while the highest population density is in Medan City (red arrow).



Figure 2. Population Density Map

4.3 Average Years of Schooling (RLS)

Figure 3 shows a map of the distribution of average years of schooling on the island of Sumatra in 2022, with green indicating low average years of schooling and red indicating high average years of schooling. The lowest average years of schooling were in Nias Regency (blue arrow), while the highest average years of schooling were in Banda Aceh City (red arrow).

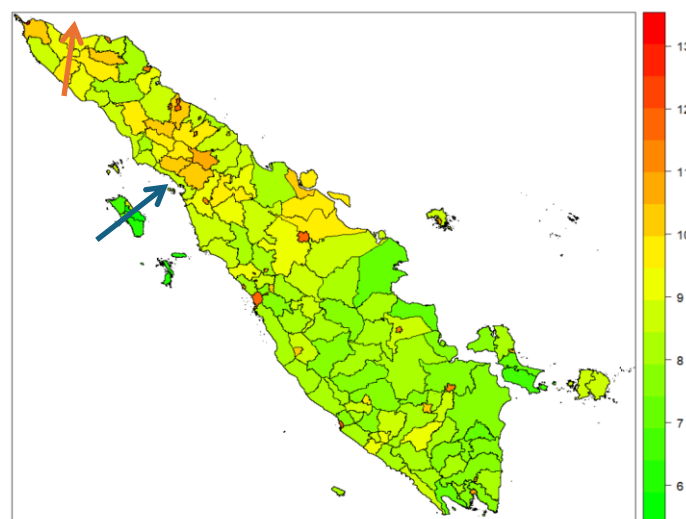


Figure 3. Average Years of Schooling Map

4.4 Open Unemployment Rate (TPT)

Figure 4 shows a map of the distribution of open unemployment rates on the island of Sumatra in 2022, with green indicating low unemployment and red indicating high unemployment. The lowest open unemployment rate was in Pakpak Bharat Regency (blue arrow), while the highest was in Padang City (red arrow).

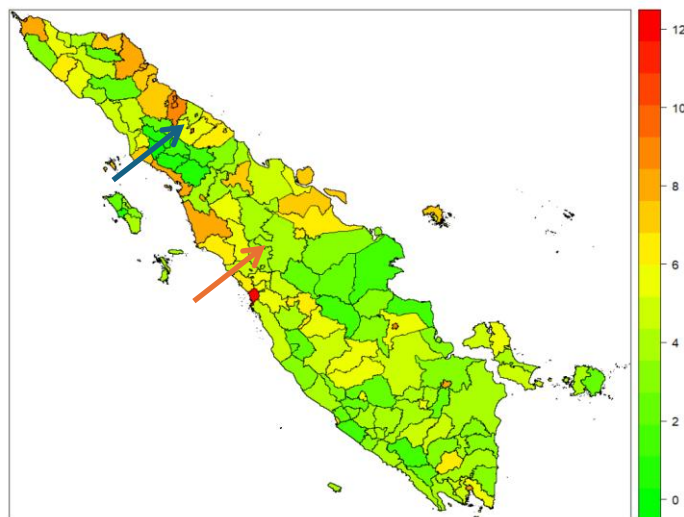


Figure 4. Open Unemployment Rate Map

4.5 Human Development Index (IPM)

Figure 5 shows a map of the distribution of the human development index on the island of Sumatra in 2022, with green indicating a low human development index and red indicating a high human development index. The lowest human development index is in West Nias Regency (blue arrow), while the highest is in Banda Aceh City (red arrow).

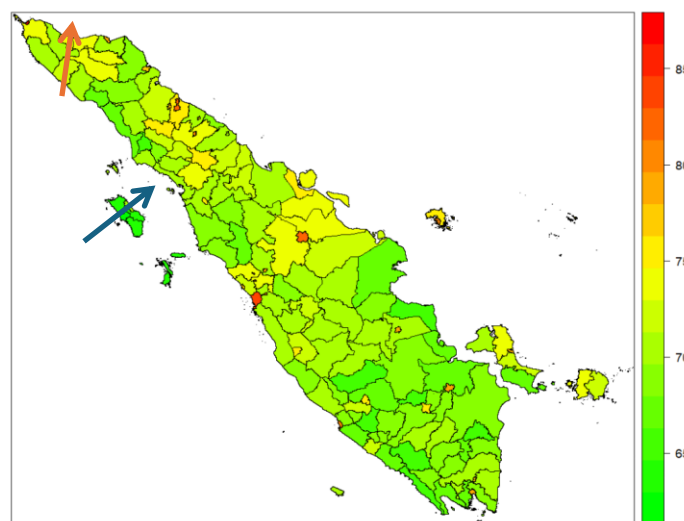


Figure 5. Human Development Index Map

4.6 Regency Minimum Wage (UMK)

Figure 6 shows a map of the distribution of district minimum wages on the island of Sumatra in 2022, with green indicating low district minimum wages and red indicating high district minimum wages. The lowest district minimum wage is in Rejang Lebong Regency (blue arrow), while the highest is in Bintan Regency (red arrow).

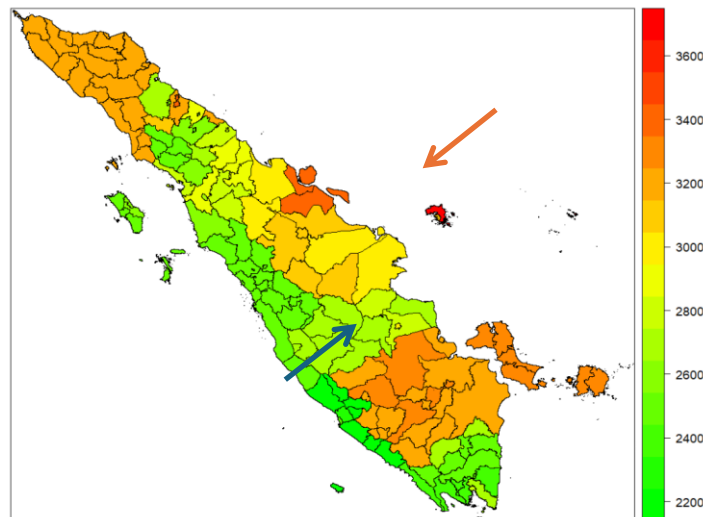


Figure 6. Regency Minimum Wage Map

Data Analysis with Linear Regression Model

Before modeling using spatial regression, a multiple linear regression model will be carried out using the Ordinary Least Square (OLS) method.

Table 3. Multiple linear regression test results

	Estimation Parameter	Std.error	p-value
Intercept	615.6	88.64	1.34×10^{-10}
KP	0.001346	0.002	0.6212
RLS	19.21	5.799	0.0011
TPT	1.4408	1.886	0.4464
IPM	-10.1465	1.772	6.17×10^{-08}
UMK	0.011119	0.009	0.2654

Based on Table 3, the multiple linear regression model is formulated as follows:

$$\widehat{Miskin}_i = 615.6 + 0.001346KP_i + 19.21RLS_i + 1.4408TPT_i - 10.1465IPM_i + 0.011119UMK_i$$

for $i = 1, 2, \dots, n$

with $p\text{-value} = 7.261 \times 10^{-08}$, $R^2 = 25.61\%$, dan $\text{adj}(R^2) = 22.93\%$. Consequently, it can be concluded that 22.93% of the variation in poverty levels can be explained by the variations in population density, average years of schooling, open unemployment rate, human development index, and regency minimum wage through the multiple linear regression model.

Linear Regression Model Assumption Test

1. Normality Test

The residual normality test was performed using the Normal Probability Plot and the Anderson-Darling test.

Normal Probability Plot

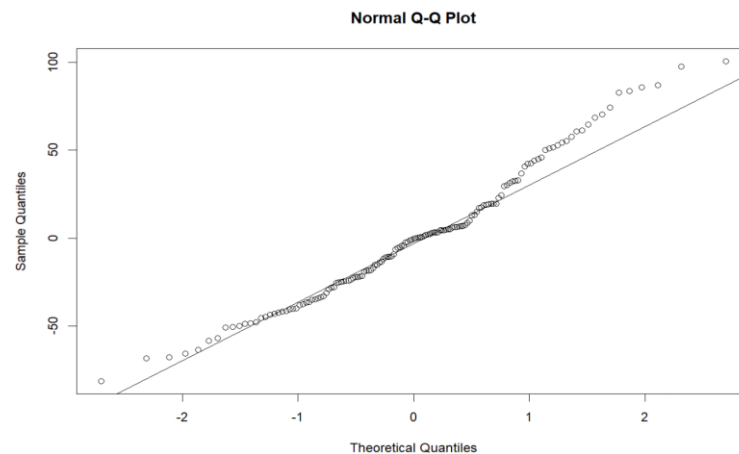


Figure 7. Normal Probability Plot

Based on the Normal Probability Plot in Figure 7, the data points are spread around the diagonal line as a representation of a normal distribution pattern, so it can be concluded that the residuals (errors) are normally distributed even though there are slight deviations.

Anderson-Darling

The null hypothesis for the normality test is that the residuals are normally distributed. The value of $A = 1.0537$ and $p\text{-value} = 0.008802$ are obtained. H_0 is rejected if the $p\text{-value} < 0.05$. The $p\text{-value} = 0.008802 < 0.05$, so H_0 is rejected. Therefore, it can be concluded that the residuals of the multiple linear regression model are not normally distributed.

2. Autocorrelation Test

The autocorrelation test uses the Durbin-Watson test. The null hypothesis for the autocorrelation test is that there is no autocorrelation. The obtained DW value 1.6575 and $p\text{-value} = 0.0151$. H_0 is rejected if the $p\text{-value} < 0.05$. It can be seen that the $p\text{-value} = 0.0151 < 0.05$, so H_0 is rejected. Therefore, it can be concluded that autocorrelation occurs in the residuals.

3. Homoscedasticity Test

The homoscedasticity test uses the Breusch-Pagan test. The null hypothesis for the homoscedasticity test is that there is homogeneity in the residuals. The obtained BP value = 7.5032 and $p\text{-value} = 0.1858$. H_0 is rejected if the $p\text{-value} < 0.05$. It can be seen that the $p\text{-value} = 0.1858 > 0.05$, then H_0 is not rejected. So it can be concluded that there is homogeneity in the residuals.

4. Multicollinearity Detection

Multicollinearity detection is done by looking at the VIF value.

Table 4. Table VIF value

Miskin	RLS	TPT	IPM	UMK
1.607076	5.463540	1.476562	5.776163	1.066683

Multicollinearity in a variable occurs if the VIF value is > 10 . From the results above, it can be seen that the VIF value of all variables is less than 10 so it can be concluded that there is no multicollinearity in the variables used in this study.

5. Residual Freedom Test

The residual independence test was performed using the Runs test. The null hypothesis for the residual independence test is that the residuals of the model are independent. The Runs value is 61 and the $p - value = 0.04555$. H_0 is rejected if the $p - value < 0.05$. The $p - value = 0.04555 < 0.05$, so H_0 is rejected. Therefore, it can be concluded that the residuals of the model are dependent.

Data Analysis with Spatial Regression Model

Following the multiple linear regression analysis, spatial regression analysis will be conducted. Moran's I test is carried out to determine whether certain variables have spatial dependencies so that autocorrelation occurs between locations.

Hypothesis

$H_0: I = 0$ (There is no spatial dependency/no autocorrelation between locations)

$H_1: I \neq 0$ (There is spatial dependency/autocorrelation between locations)

The statistical values of Moran's I test and p-value are presented in Table 5.

Tabel 4. Moran's I Test

	Moran's I	p-value
Miskin	0.632744751	2.2×10^{-16}
KP	0.017508167	0.339
RLS	0.258205759	1.046×10^{-05}
TPT	0.329132842	3.393×10^{-08}
IPM	0.302803167	3.203×10^{-07}
UMK	0.651154450	2.2×10^{-16}
Residual	0.457620007	4.633×10^{-14}

Decision rules, H_0 rejected if $p\text{-value} < 0.05$. Result:

- Variable Miskin: $p - value = 2.2 \times 10^{-16} < 0.05$, H_0 rejected.
- Variable KP: $p - value = 0.339 > 0.05$, H_0 not rejected.
- Variable RLS: $p - value = 1.046 \times 10^{-5} < 0.05$, H_0 rejected.
- Variable TPT: $p - value = 3.393 \times 10^{-8} < 0.05$, H_0 rejected.
- Variable IPM: $p - value = 3.203 \times 10^{-7} < 0.05$, H_0 rejected.
- Variable UMK: $p - value = 2.2 \times 10^{-16} < 0.05$, H_0 rejected.
- Residual: $p - value = 4.633 \times 10^{-14} < 0.05$, H_0 rejected.

With a significance level of 0.05, the results show that there is no spatial dependency on the KP variable. However, in variable Miskin, RLS, TPT, IPM, UMK and error variables, spatial dependency occurs which indicates the presence of autocorrelation between locations, thus allowing the General Nested Spatial Model (GNSM) to be suitable for use because it is in accordance with the data conditions, namely there is a significant spatial dependency on the dependent variable, independent variables, and errors (residuals). Figure 8, Figure 9, Figure 5, Figure 10, Figure 11 and Figure 12, show the Moran I plots for the variables of Miskin, KP, TPT, IPM, and UMK. The Moran's I plot is divided into 4 quadrants, which means:

1. Quadrant I, HH (High-High) shows areas that have high observation values surrounded by areas that have high observation values.
2. Quadrant II, LH (Low-High) shows areas that have low observation values but are surrounded by areas that have high observation values.
3. Quadrant III, LL (Low-Low) shows areas that have low observation values surrounded by areas that also have low observation values.
4. Quadrant IV, HL (High-Low) shows areas that have high observation values but are surrounded by areas that have low observation values.

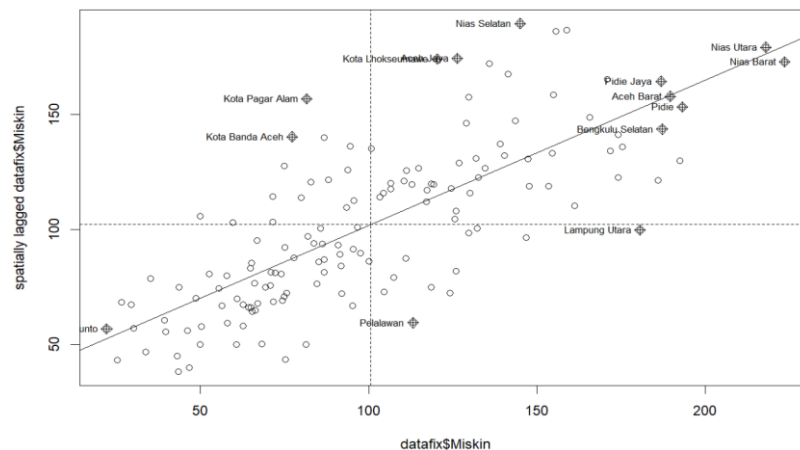


Figure 13. Moran's I plot for the poverty level variable

5. The districts/cities in quadrant I are Aceh Jaya, Pidie Jaya, West Aceh, Pidie, South Bengkulu, South Nias, North Nias, West Nias, and Lhokseumawe City.
6. The districts/cities in quadrant II are Pagar Alam City and Banda Aceh City.
7. The districts/cities in quadrant III are Sawahlunto City.
8. The districts/cities in quadrant IV are Pelalawan and North Lampung.

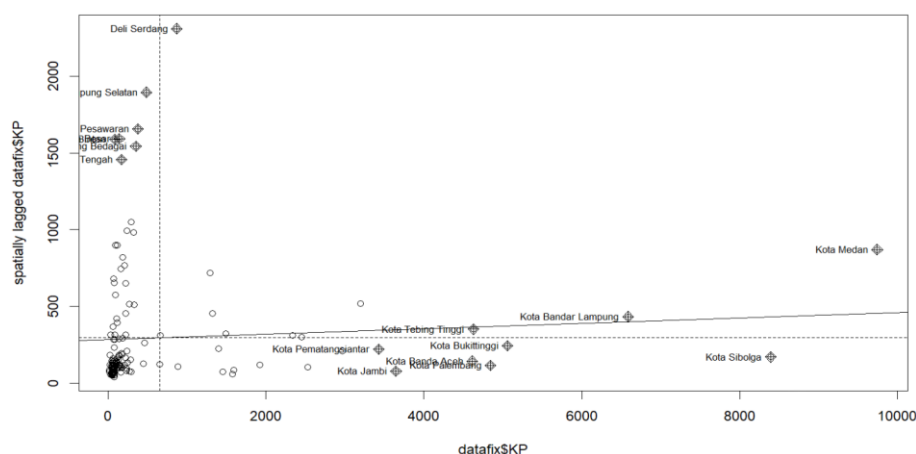
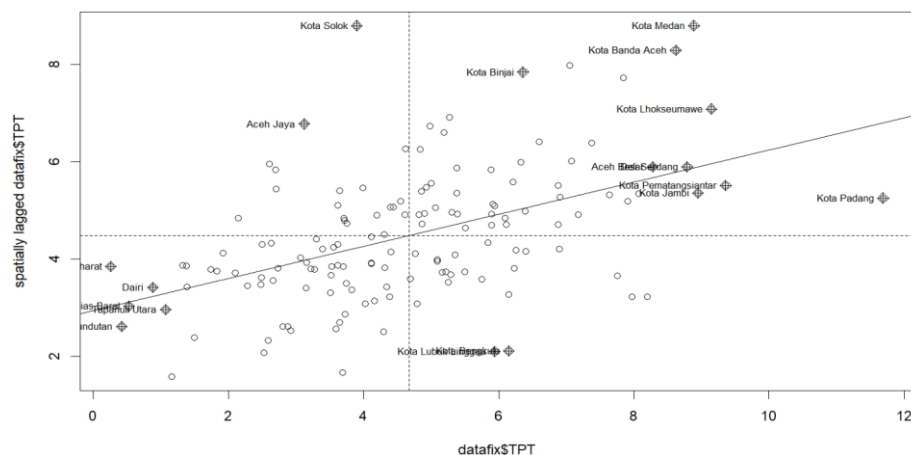


Figure 14. Moran's I plot for population density variables (KP)

1. The districts/cities in quadrant I are Deli Serdang, Medan City, Bandar Lampung City and Tebing Tinggi City.
2. The districts/cities in quadrant II are Serdang Begadai, South Lampung, Pesawaran, and Aceh Besar.

-
- A scatter plot showing the relationship between the spatially lagged variable `spatially_lagged_datafix$RRLS` (y-axis) and the original variable `datafix$RRLS` (x-axis). The x-axis ranges from 6 to 13, and the y-axis ranges from 7 to 10. A solid regression line is plotted, showing a positive correlation. A horizontal dashed line is drawn at approximately y = 8.7. A vertical dashed line is drawn at x = 9. Data points for 17 Indonesian provinces are plotted as diamond markers and labeled. Most points are clustered between x = 8 and x = 12, and y = 7.5 and y = 10.5. The labels include Nias, Nias Selatan, Nias Barat, Kota Gunungsitoli, Padang Pariaman, Bintan, Kota Banda Aceh, Kota Pekanbaru, Kota Padang Panjang, Kota Bengkulu, and Kota Pangkalpinang.

1. The districts/cities in quadrant I are Banda Aceh City, Padang City, Pekanbaru City and Bukittinggi City.
2. The districts/cities in quadrant II are Padang Pariaman and Bintan.
3. The regencies/cities in quadrant III are Nias, West Nias, South Nias, and Gunungsitoli City.
4. The districts/cities in quadrant IV are Padang Panjang City and Bengkulu City.



1. The districts/cities in quadrant I are Medan City, Binjai City, Padang City, Aceh Besar, Deli Serdang, Pematangsiantar City, Lhokseumawe City and Jambi City.
2. The districts/cities in quadrant II are Aceh Jaya and Solok City.
3. The districts/cities in quadrant III are West Nias, North Tapanuli, Pakpak Bharat, and Dairi.
4. The districts/cities in quadrant IV are Lubuk Linggau City and Bengkulu City.

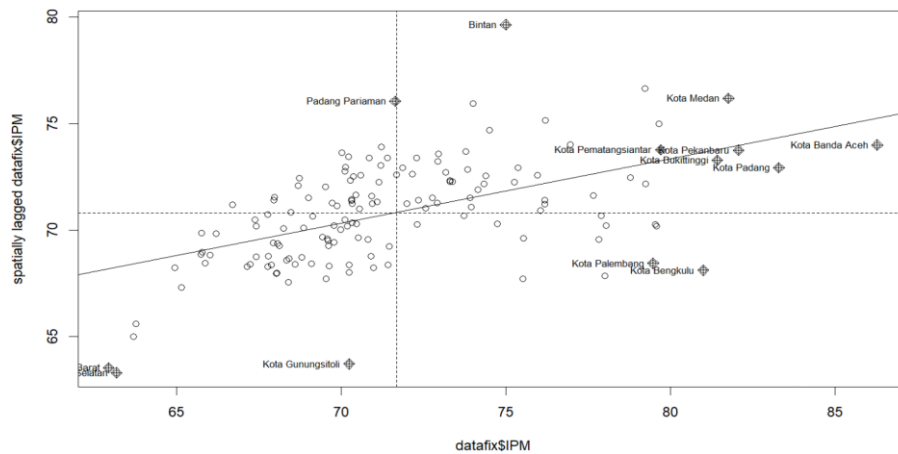


Figure 17. Moran's I plot for the human development index variable (IPM)

1. The districts/cities in quadrant I are Bintan, Medan City, Padang City, Pematangsiantar City, Bukittinggi City, Pekanbaru City and Banda Aceh City.
2. The district/city in quadrant II is Padang Pariaman.
3. The district/city in quadrant III is Gunungsitoli City.
4. The districts/cities in quadrant IV are Palembang City and Bengkulu City.

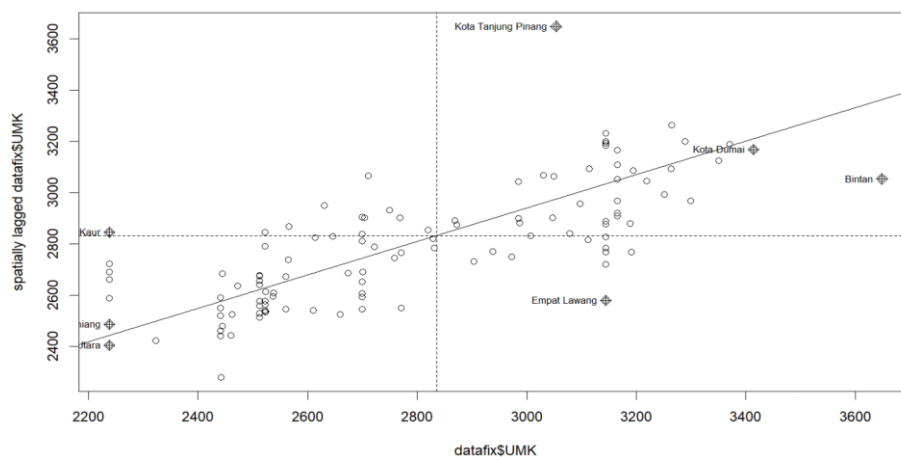


Figure 18. Moran's I plot for district minimum wage variables (UMK)

1. The districts/cities in quadrant I are Bintan, Dumai City and Tanjung Pinang City.
2. The district/city in quadrant IV is Empat Lawang.

Lagrange Multiplier Test (LM)

The Lagrange Multiplier is used to detect spatial dependencies more specifically. Table 6 displays the results of the Lagrange Multiplier test for error (LMerr) and lag (LMlag).

Table 5. Lagrange Multiplier test result

	LMerr	LMlag
Statistic parameters	52.4190	76.0873
p-value	4.483×10^{-13}	2.2×10^{-16}

The null hypothesis for the LM test is that there is no autocorrelation. Table 6 shows that Lmerr and LMlag yield significant results with a p-value < 0.05. This indicates that the SAR, SEM, and SARMA models are suitable for the data.

From the results of the Moran's I test and the Lagrange Multiplier test, the spatial models used in this study are the GNSM model, SAR model, SEM model, and SARMA model. The study employs a spatial weight matrix based on the queen contiguity criterion.

General Nested Spatial Model (GNSM)

The GNSM model is a spatial regression model for data that indicates the existence of spatial dependency on the dependent variable, independent variable and error.

Table 6. General Nested Spatial Model (GNSM) result

	Estimation	Std.error	z-value	p-value
intercept	-0.0069354	0.0498300	-0.1392	0.88931
KP	0.0364789	0.0689591	0.5290	0.59681
RLS	-0.0429723	0.1587411	-0.2707	0.78662
TPT	0.0674646	0.0670199	1.0066	0.31411
IPM	-0.1830765	0.1679491	-1.0901	0.27568
UMK	-0.1087676	0.0892410	-1.2188	0.22292
lag.KP	0.0775716	0.2041474	0.3800	0.70396
lag.RLS	0.3274129	0.1914667	1.7100	0.08726
lag.TPT	-0.0976889	0.1058087	-0.9233	0.35587
lag.IPM	-0.3333260	0.2417428	-1.3788	0.16794
lag.UMK	0.1884448	0.1087990	1.7320	0.08327

Based on Table 7, there are no significant variables because the p-value is > 0.05 for each variable. General Nested Spatial Model (GNSM) Assumption Test. The following is an assumption test on the General Nested Spatial model:

1. Residual Normality Test (*Anderson-Darling*)

The null hypothesis for the normality test is that the residuals are normally distributed. The value of $A = 1.5634$ with a p-value = 0.0004844. H_0 is rejected if the p-value < 0.05 . The p-value = 0.0004844 < 0.05 , so H_0 is rejected. So, it can be concluded that the residuals of the GNSM model are not normally distributed.

2. Residual Homogeneity Test (Breusch-Pagan)

The null hypothesis for the homogeneity test is that there is homogeneity in the residuals. The value of BP = 15.999 with a p-value = 0.09965. H_0 is rejected if the p-value is < 0.05 . The p-value = 0.09965 > 0.05 , so H_0 is not rejected. Therefore, it can be concluded that there is homogeneity in the residuals.

3. Residual Freedom Test (Runs)

The null hypothesis for the residual independence test is that the residuals of the model are independent. The value of Runs = 72 and the p-value = 0.868. H_0 is rejected if the p-value is < 0.05 . The p-value = 0.868 > 0.05 , so H_0 is not rejected. Therefore, it can be concluded that the residuals of the GNSM model are independent.

Spatial Autoregressive Model (SAR)

The SAR model is a spatial regression model in which there is spatial dependency on its dependent variable.

Table 7. Spatial Autoregressive Model (SAR) result

	Estimation	Std.error	z-value	p-value
intercept	307.3283079	68.5685581	4.4821	7.393×10^{-06}
KP	0.0010482	0.0019513	0.5372	0.5911493
RLS	6.8339498	4.2166562	1.6207	0.1050812
TPT	1.3963928	1.3518671	1.0329	0.3016336
IPM	-4.8133208	1.3185320	-3.6505	0.0002617
UMK	0.0013236	0.0071306	0.1856	0.8527353

Based on Table 8, it can be seen that only the IPM variable is significant because the p-value is < 0.05 , while the other variables have p-values > 0.05 .

Spatial Autoregressive Model (SAR) Assumption Test. The following is an assumption test on the Spatial Autoregressive model.

1. Residual Normality Test (*Anderson-Darling*)

The null hypothesis for the normality test is that the residuals are normally distributed. The value of $A = 1.7851$ with a $p - value = 0.0001376$. H_0 is rejected if the p-value < 0.05 . The $p - value = 0.0001376 < 0.05$, so H_0 is rejected. So, it can be concluded that the residuals of the SAR model are not normally distributed.

2. Residual Homogeneity Test (Breusch-Pagan)

The null hypothesis for the homogeneity test is that there is homogeneity in the residuals. The value of $BP = 8.5896$ with a $p - value = 0.1266$. H_0 is rejected if the p-value is < 0.05 . The $p - value = 0.1266 > 0.05$, so H_0 is not rejected. Therefore, it can be concluded that there is homogeneity in the residuals.

3. Residual Freedom Test (Runs)

The null hypothesis for the residual independence test is that the residuals of the model are independent. The value of $Runs = 78$ and the $p - value = 0.5046$. H_0 is rejected if the p-value is < 0.05 . The $p - value = 0.5046 > 0.05$, so H_0 is not rejected. Therefore, it can be concluded that the residuals of the SAR model are independent.

Spatial Error Model (SEM)

The SEM model is a spatial regression model in which there is spatial dependency on the error.

Table 8. Spatial Error Model (SEM) result

	Estimation	Std.error	z-value	p-value
intercept	360.5255009	87.5173036	4.1195	3.797×10^{-05}
KP	0.0014344	0.0019762	0.7258	0.4679
RLS	2.4290769	5.2928126	0.4589	0.6463
TPT	2.1348280	1.4314865	1.4913	0.1359
IPM	-3.6176618	1.6958248	-2.1333	0.0329
UMK	-0.0119527	0.0112421	-1.0632	0.2877

Based on Table 9, it can be seen that only the IPM variable is significant because the p-value is < 0.05 , while the other variables have p-values > 0.05 .

Spatial Error Model (SEM) Assumption Test. The following is an assumption test on the Spatial Error model.

1. Residual Normality Test (*Anderson-Darling*)

The null hypothesis for the normality test is that the residuals are normally distributed. The value of $A = 1.7478$ with a $p - value = 0.00017$. H_0 is rejected if the p -value < 0.05 . The $p - value = 0.00017 < 0.05$, so H_0 is rejected. So, it can be concluded that the residuals of the SEM model are not normally distributed.

2. Residual Homogeneity Test (Breusch-Pagan)

The null hypothesis for the homogeneity test is that there is homogeneity in the residuals. The value of $BP = 6.8353$ with a $p - value = 0.2332$. H_0 is rejected if the p -value is < 0.05 . The $p - value = 0.2332 > 0.05$, so H_0 is not rejected. Therefore, it can be concluded that there is homogeneity in the residuals.

3. Residual Freedom Test (Runs)

The null hypothesis for the residual independence test is that the residuals of the model are independent. The value of $Runs = 77$ and the $p - value = 0.6166$. H_0 is rejected if the p -value is < 0.05 . The $p - value = p - value = 0.6166 > 0.05$, so H_0 is not rejected. Therefore, it can be concluded that the residuals of the SEM model are independent.

Spatial Autoregressive and Moving Average Model (SARMA)

The SARMA model is a spatial regression model that contains spatial dependencies on the independent and error variables.

Table 9. Spatial Autoregressive and Moving Average Model (SARMA)

	Estimation	Std.error	z-value	p-value
intercept	-0.026984	0.030810	-0.8758	0.3811240
KP	0.016263	0.058971	0.2758	0.7827206
RLS	0.160007	0.089380	1.7902	0.0734240
TPT	0.036607	0.050789	0.7208	0.4710526
IPM	-0.377610	0.099590	-3.7916	0.0001497
UMK	0.030145	0.035839	0.8411	0.4002768

Based on Table 10, it can be seen that only the IPM variable is significant because the p -value is < 0.05 , while the other variables have p -values > 0.05 .

Spatial Autoregressive and Moving Average Model (SARMA) Assumption Test. The following is an assumption test on the Spatial Autoregressive and Moving Average model.

1. Residual Normality Test (*Anderson-Darling*)

The null hypothesis for the normality test is that the residuals are normally distributed. The value of $A = 1.5757$ with a $p - value = 0.0004516$. H_0 is rejected if the p -value < 0.05 . The $p - value = 0.0004516 < 0.05$, so H_0 is rejected. So, it can be concluded that the residuals of the SARMA model are not normally distributed.

2. Residual Homogeneity Test (Breusch-Pagan)

The null hypothesis for the homogeneity test is that there is homogeneity in the residuals. The value of $BP = 10.711$ with a $p - value = 0.05743$. H_0 is rejected if the p -value is < 0.05 . The $p - value = 0.05743 > 0.05$, so H_0 is not rejected. Therefore, it can be concluded that there is homogeneity in the residuals.

3. Residual Freedom Test (Runs)

The null hypothesis for the residual independence test is that the residuals of the model are independent. The value of $Runs = 72$ and the $p - value = 0.868$. H_0 is rejected if the p-value is < 0.05 . The $p - value = 0.868 > 0.05$, so H_0 is not rejected. Therefore, it can be concluded that the residuals of the SARMA model are independent.

Selecting The Best Model

Next, a comparison of the models will be conducted to determine the best model.

Table 10. Model comparison (*Goodness of Fit*)

	Linear Regression	GNSM	SAR	SEM	SARMA
AIC	1478.808	308.2	1408.8	1415.767	305.32
ρ	-	0.76518 (2.22×10^{-16})	0.64302 (2.22×10^{-16})	-	0.80374 (2.22×10^{-16})
λ	-	-0.47479 (0.15155)	-	0.69612 (7.771×10^{-16})	-0.5269 (6.276×10^{-05})
Residual Normality	Not valid	Not valid	Not valid	Not valid	Not valid
Residual Homogeneity	Not valid	Valid	Valid	Valid	Valid
Residual Freedom	Not valid	Valid	Valid	Valid	Valid

Based on Table 11, the SARMA model has the smallest AIC value, at 305.32. Furthermore, the p-values of ρ and λ in the SARMA model are both less than 0.05, making them significant. Therefore, the SARMA model is the best model for this data.

Best Model

Based on the analysis conducted, the SARMA model was selected to explain the variables in the data. The parameter estimation results of the SARMA model are shown in Table 11.

Table 11. SARMA model parameter estimation

Variable	Estimate	Standard Error	z value	Pr(> z)
(Intercept)	-0.026984	0.030810	-0.8758	0.3811240
KP	0.016263	0.058971	0.2758	0.7827206
RLS	0.160007	0.089380	1.7902	0.0734240
TPT	0.036607	0.050789	0.7208	0.4710526
IPM	-0.377610	0.099590	-3.7916	0.0001497
UMK	0.030145	0.035839	0.8411	0.4002768

Based on Table 12, it can be seen that only the IPM variable has a significant influence on the model so the model will be regressed with this variable.

Table 12. Latest SARMA model parameter estimation

Variable	Estimate	Standard Error	z value	Pr(> z)
(Intercept)	155.49894	34.72083	4.4785	7.515×10^{-06}
IPM	-1.94639	0.45342	-4.2927	1.765×10^{-05}
Rho (ρ)	0.82723	0.044664	18.521	$< 2.22 \times 10^{-16}$
Lambda (λ)	-0.51409	0.13068	-3.934	8.3535×10^{-05}

The resulting SARMA equation is:

$$\widehat{Miskin}_i = 155.49894 + 0.82723 \sum_{j=1, i \neq j}^n w_{ij} Miskin_j - 1.94639 IPM_i - 0.51409 \sum_{j=1, i \neq j}^n w_{ij} u_j$$

5. CONCLUSION

Based on multiple linear regression, an analysis of population density factors, average length of schooling, open unemployment rate, human development index, and district minimum wage, on the Poverty Level was conducted. From the model, value of $R^2 = 22.93\%$ was obtained, which stated that the model explained 22.93% of the total variation in the dependent variable. When testing the classical regression assumptions, with a significance level of 0.05, it was found that the data met the assumption of not multicollinearity, but did not meet the assumption of normality, there was autocorrelation, the residuals were not independent, and homoscedastic. Further testing was carried out through the Moran 1 and Lagrange Tests. It was found that spatial dependencies were found in the variables of Poverty Level (Miskin) and the independent variables include Population Density (KP), Average Years of Schooling (RLS), Human Development Index (IPM), Open Unemployment Rate (TPT) and Regency Minimum Wage (UMK). And error and there was spatial autocorrelation through the Lmlag and Lmerror models. Based on the Spatial Regression Analysis, SAR, SEM, SARMA and GNSM models were formed and further assumption tests were carried out. All models failed to meet the normality criteria, and only the multiple linear regression model failed all assumption tests except for multicollinearity. In the Goodness of Fit test, comparing the AIC values, the SARMA model was found to have the smallest AIC value, thus the model is considered good for spatial modelling. Despite its contributions, this research possesses several limitations that should be addressed in future studies. This study utilizes cross-sectional data for the year 2022, which limits the ability to observe long-term poverty trends or temporal changes in spatial clusters. Future research should consider utilizing spatial panel data to capture both spatial and temporal dynamics simultaneously. The analysis focuses primarily on socioeconomic indicators and does not account for geographical or environmental factors, such as topography and infrastructure access, which may play a role in rural poverty.

REFERENCES

- [1] Annur, R. A. (2013). Faktor-Faktor Yang Mempengaruhi Kemiskinan Di Kecamatan Jekulo Dan Mejobo Kabupaten Kudus Tahun 2013. *Economics Development Analysis Journal*, 2(4), 409–426.
- [2] Permai, S. D., Jauri, R., & Chowanda, A. (2019). Spatial autoregressive (SAR) model for average expenditure of Papua Province. *Procedia Computer Science*, 157, 537–542. <https://doi.org/10.1016/j.procs.2019.09.011>
- [3] Priseptian, L., & Primandhana, W. P. (2022). Analisis faktor-faktor yang mempengaruhi kemiskinan. *Forum Ekonomi*, 24(1), 45–53. <https://doi.org/10.30872/jfor.v24i1.10362>
- [4] Nugroho, A., & Rochaida, E. (2022). Pengaruh PDRB dan rata-rata lama sekolah terhadap kemiskinan di wilayah kota. *Jurnal Ilmu Ekonomi Mulawarman (JIEM)*, 7(4). <https://doi.org/10.29264/jiem.v7i4.9736>

- [5] Harlik, H., Amir, A., & Hardiani, H. (2013). Faktor-faktor yang mempengaruhi kemiskinan dan pengangguran di Kota Jambi. *Jurnal Perspektif Pembiayaan dan Pembangunan Daerah*, 1(2), 109–120. <https://doi.org/10.22437/ppd.v1i2.1500>
- [6] Rati, Musfika. 2013. Skripsi. Model Regresi Spasial untuk Anak tidak Bersekolah Usia Kurang 15 Tahun di Kota Medan. Medan: Universitas Sumatera Utara.
- [7] Amaliya, R. N., (2018). Pemodelan General Nesting Spatial (GNS) Pada Data Angka Harapan Hidup Kabupaten/Kota Di Jawa Timur Tahun 2016. 1–71. <https://repository.ub.ac.id/id/eprint/168540/>
- [8] Anselin, L. 1988. *Spatial econometrics: methods and models*. Springer Science & Business Media.
- [9] Widyastuti, M. N., Srinadi, I. G., & Susilawati, M. (2019). Pemodelan Jumlah Kasus Pneumonia Balita Di Jawa Timur Menggunakan Regresi Spatial Autoregressive Moving Average. *E-Jurnal Matematika*, 8(3), 236-245. <https://doi.org/10.24843/mtk.2019.v08.i03.p259>
- [10] Wei, W.W. 1990. Time Series Analysis. Addison-Wesley Publishing Company.
- [11] BPS-Statistics Indonesia. (2023). *Percentage of Poor Population by Regency/City in Sumatra 2020-2022*. <https://www.bps.go.id>