BAREKENG: Journal of Mathematics and Its Applications
June 2024 Volume 18 Issue 2 Page 1317–1332
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss2pp1317-1332

SPATIALLY INFORMED INSIGHTS: MODELING PERCENTAGE POVERTY IN EAST JAVA PROVINCE USING SEM WITH SPATIAL WEIGHT VARIATIONS

Ashabul Akbar Maulana[1], Achmad Fauzan[2*]

[1,2]Department of Statistics, Faculty of Mathematica and Natural Sciences, Universitas Islam Indonesia
Jln. Kaliurang No.Km. 14,5, Krawitan, Umbulmartani, Kec. Ngemplak, Kabupaten Sleman,
Daerah Istimewa Yogyakarta 55584, Indonesia

*Corresponding author's e-mail: * **achmadfauzan@uii.ac.id***

Article History:
Received: 7th February 2024
Revised: 3rd March 2024
Accepted: 6th May 2024
Published: 1st June 2024

Keywords:
Poverty;
Spatial Error Model;
Spatial Weight Matrix.

ABSTRACT

The East Java Province is one of Indonesia's regions grappling with a notably elevated poverty rate, accounting for 11.32% of the populace. A strategic approach to comprehending and redressing this issue involves applying spatial analysis, wherein spatial factors are intricately integrated into the modeling and cartographic representation of poverty data. The primary objective of this research is to discern the principal determinants influencing the incidence of poverty in East Java Province, employing data reflective of the population's poverty percentages within the province for the year 2021. The study incorporates six pivotal variables, namely the population poverty rate, open unemployment rate, labor force participation rate, average years of schooling, adjusted per capita expenditure, and the gross regional domestic product (GRDP), predicated on adjusted expenditure. Diverse weighting schemes are applied based on distance and contiguity. The optimal predictive model utilized is the Spatial Error Model (SEM) incorporating a Distance Band Weighing (DBW) mechanism with a designated maximum distance (d_{max}) of 75000 meters. Outcomes indicate that the variable wielding the most substantial influence on the poverty percentage in East Java Province is the average years of schooling. Specifically, an increase in the pursuit of formal education manifests as a negative correlation to the poverty percentage, implying an inverse relationship. Moreover, the SEM model adheres to the requisite assumptions, encompassing the normality of residuals, homogeneity of residuals, and non-spatial autocorrelation of residuals. Comparative analyses show that the SEM model using DBW results in lower MAE, MSE, RMSE, AIC, and MAPE values compared to linear regression. Additionally, the SEM model has higher pseudo-R^2 values. Likelihood ratio tests highlight significant differences, with SEM being more efficient and providing better explanatory power for dataset variations.

How to cite this article:

A. A. Maulana and A. Fauzan., "SPATIALLY INFORMED INSIGHTS: MODELING PERCENTAGE POVERTY IN EAST JAVA PROVINCE USING SEM WITH SPATIAL WEIGHT VARIATIONS," *BAREKENG: J. Math. & App.*, vol. 18, iss. 2, pp. 1317-1332, June, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Based on data retrieved from the Central Statistics Agency (BPS), there was an increase in the number of people living in poverty in the East Java Province in 2021, amounting to 153.63 thousand people. This number increased by 3.48% from the previous year. The impact of this increase worsened the poverty rate in East Java Province from 4.42 million people (11.02%) in 2020 to 4.57 million people (11.32%) in 2021, an increase of 0.30%, which made the province become the third poorest on the entire island of Java. Many factors can cause poverty, including unemployment, participation, level of formal education, and income **[1]**.

One of the analyses in statistics used for examining and modeling relationships between variables is regression analysis. The assumptions that must be met in regression analysis include: (1) the relationship between the dependent variable (y) and the predictor (x) is linear; (2) the error has a mean of zero; (3) the error has a constant variance; (4) the error is not correlated (autocorrelation or with response); and (5) the error is normally distributed **[2]**. But sometimes, these regression assumptions are not met. When related to data containing location, one method that can be used to analyze this data is spatial data analysis, including spatial regression, which is caused by spatial dependence, such as the spatial error model (SEM), spatial lag of X (SLX), spatial autoregressive (SAR), and so on **[3]**, **[4]**. SAR is widely regarded as the most prevalent specification and the most universally applicable approach to conceptualizing spatial dependency. Alternatively, it is feasible to incorporate spatial correlation by including the error factor in the regression equation. While SAR considers spatial dependency to be of significant importance, SEM regards it as an unwanted factor. This model only tries to guess the regression parameters for the critical explanatory variables. It does not look at the possible importance of geographical clustering or spatial autocorrelation, which could mean more than just attributional dependency. Instead of assuming that a geographical lag affects the dependent variable, SEM estimates a model that relaxes the traditional regression model assumption that errors must be independent **[5]**. Meanwhile, spatial regression models caused by spatial heterogeneity can use Geographically Weighted Regression (GWR), Geographically Weighted Poisson Regression models, or other models **[6]**, **[7]**.

Several studies that use spatial analysis, including Sihombing **[8]**, are conducting research on variables that are factors of poverty, such as income levels, consumption, health, education, and relationships in society using the SAR model. Tumanggor dan Simamora **[9]** identified factors that influence the Human Development Index using the SAR model. Safari **[10]** used the SEM model to determine the factors that influence food security in South Sulawesi province. Yulianto and Ayuwida **[11]** aimed to model the level of poverty in East Java Province using spatial regression. The data used in this research is the poverty level of East Java Province in 2015 as the dependent variable, as well as a number of independent variables including Female Head of Household, Number of out-of-school children aged 7-18 years, Number of disabled individuals, and six other independent variables. Through this research, using the Spatial Error Model (SEM) method using the Queen Contiguity weighing matrix, the most influential factors on the poverty rate in East Java Province in 2015 were determined to be the number of disabled individuals and unprotected drinking water sources.

Apart from that, the research conducted by Jelita **[12]** carried out spatial modeling on Gini ratio data for 2015-2017 as a response variable as well as population size, number of poor people, per capita expenditure, and the district/city Human Development Index in East Java Province for 2015-2017 as a predictor variable using K-Nearest Neighbor and Distance Band as a spatial weight matrix. The results of this research, using the Spatial Error Model (SEM) as a spatial regression model and KNN as a weighing matrix, found that the number of people living in poverty was the factor that had the most influence on the Gini ratio of East Java Province in 2015-2017. Other research conducted by Aziah et al. **[13]** studied the influence of education, per capita income, and the population living in poverty in East Java Province using panel data regression. The results of this research show that education and per capita income have a significant, negative effect on the. Meanwhile, population size has a positive effect on the regency/city poverty in East Java Province.

Based on these previous research, Muryani **[14]**, Azizi **[15]**, Alam **[16]**, and Widiantari **[17]**, this research aims to determine the factors that most influence the increase in the percentage of poor people in East Java Province in 2021. Several socio-economic factors are used, such as the open Unemployment Rate, labor force participation rate, average years of schooling, product per capita, and GRDP based on East Java Province expenditure in 2021. Moreover, this research considers different spatial weights when analyzing the distribution of poverty in East Java Province. Spatial weights can reflect location effects and spatial patterns in the data [18], which can help understand the relationship between socio-economic factors and poverty levels in the region. This research anticipates that the East Java Provincial government take action to reduce the percentage of poverty by paying more attention to the factors that have the most significant influence.

2. RESEARCH METHODS

2.1 Data and Source of Data

The data used is secondary data retrieved from the East Java Province Central Statistics Agency (BPS) Website in 2021 (https://jatim.bps.go.id/). This research data uses six (6) variables, which are: population poverty Rate (Y), open unemployment rate (X_1), labor force participation rate (X_2), average years of schooling (X_3), adjusted per capita expenditure (X_4), gross regional domestic product (GRDP) based on adjusted expenditure (X_5). The data used is presented in Table 1.

Table 1. Research Data

No	Regency	Y (%)	X_1 (%)	X_2(%)	X_3(Year)	X_4 (Million Rupiah)	X_5 (Trillion Rupiah)
1	Bangkalan	21.57	8.07	68.66	5.96	8673	17152779
2	Banyuwangi	8.07	5.42	72.32	7.42	12217	55471065
3	Batu	4.09	6.57	73.74	9.31	12887	11471435
4	Blitar	9.65	3.66	70.44	7.5	10757	25700019
5	Bojonegoro	13.27	4.82	71.84	7.38	10221	65839509
6	Bondowoso	14.73	4.46	73.89	5.94	10690	13921654
7	Gresik	12.42	8	69.43	9.56	13280	101318686
8	Jember	10.41	5.44	68.97	6.49	9410	54688719
9	Jombang	10	7.09	70.69	8.55	11394	28553448
10	Kediri	11.64	5.15	69.34	8.08	11127	29361672
11	Bllitar City	7.89	6.61	69.96	10.35	13816	4924572
12	Kediri City	7.75	6.37	67.35	10.15	12359	86485594
13	Madiun City	5.09	8.15	66.87	11.37	16095	10748101
14	Malang City	4.62	9.65	67.59	10.41	16663	53309702
15	Mojokerto City	6.39	6.87	67.09	10.47	13610	4976490
16	Pasuruan City	6.88	6.23	71.66	9.33	13354	5914585
17	Probolinggo City	7.44	6.55	69.71	8.95	12245	8361142
18	Lamongan	13.86	4.9	70.72	8.04	11510	27896543
19	Lumajang	10.05	3.51	66.19	6.67	9203	22623402
20	Madiun	11.91	4.99	67.77	7.82	11658	13372330
21	Magetan	10.66	3.86	73.31	8.36	11833	13417032
22	Malang	10.5	5.4	68.49	7.43	10163	68619103
23	Mojokerto	10.62	5.54	70.47	8.64	12844	60198699
24	Nganjuk	11.85	4.98	64.24	7.78	12172	18640685
25	Ngawi	15.57	4.25	72.88	7.26	11459	13823456
26	Pacitan	15.11	2.04	80.57	7.61	8887	11107402
27	Pamekasan	15.3	3.1	65.88	6.7	8804	11496236
28	Pasuruan	9.7	6.03	69.03	7.41	10297	107630268
29	Ponorogo	10.26	4.38	72.63	7.55	9851	14619969
30	Probolinggo	18.91	4.55	73.24	6.12	10969	23664388
31	Sampang	23.76	3.45	70.19	4.86	8790	13984568
32	Sidoarjo	5.93	10.87	66.47	10.72	14578	141000359
33	Situbondo	12.63	3.68	71.63	6.62	9996	13715834
34	Sumenep	20.51	2.31	75.63	5.92	9000	24161351
35	Surabaya	5.23	9.68	67.3	10.5	17862	407726799
36	Trenggalek	12.14	3.53	72.36	7.56	9743	12959018
37	Tuban	16.31	4.68	73.77	7.18	10380	43984689
38	Tulungagung	7.51	4.91	72.26	8.34	10807	27390424

2.2 Spatial Weight Matrix

The spatial weighting matrix ($W_{n\times n}$) is a crucial component in spatial analysis, providing a standardized representation of the relationships between n locations concerning a specified row or constant.

Notably, the diagonal elements of matrix $\boldsymbol{W}$, denoted as w_{ii}, are set to zero under the assumption that no spatial unit is contiguous with itself, signifying an absence of influence from a location onto itself [19]. The determination of matrix $\boldsymbol{W}$ typically involves one of two methods: (1) reliance on the distance between locations or (2) consideration of contiguity. In the context of spatial analysis, the distance measure (d_{ij}) between the centroid of location-i with coordinates (u_i, v_i) and location-j with coordinates (u_j, v_j) can be computed using various metrics, including Minkowski, Euclidean, and Manhattan distances. Here, u_i and v_i represent the latitude and longitude coordinates of location-i, while u_j and v_j denote the corresponding coordinates for location-j [20]. The general formulation of the spatial weighting matrix, accommodating n locations, is succinctly expressed through Equation (1) [21]. This matrix serves as a fundamental tool for comprehending the intricate spatial relationships inherent in the dataset, contributing to the robustness of spatial analyses within the framework of geographic information systems and statistical modeling.

$$\boldsymbol{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ W_{21} & W_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \cdots & w_{nn} \end{bmatrix} \quad (1)$$

The various weighting matrices utilized in this study are presented in Table 2 [22].

Table 2. Variation in Spatial Weighting Matrices

No	Basis of spatial weighting matrix calculation	Types of weighting matrices	Concept	Equation
1.	Based on contiguity.	Queen Contiguity (Contiguity of sides and angles)	The weight w_{ij} is designated as 1 for locations that exhibit both side and angle adjacency with the observed location, and as 0 otherwise	$w_{ij} = \begin{cases} 1, & \text{if i and j are neighbors} \\ 0, & \text{if i and j are not neighbors} \end{cases}$
2.	Based on distance.	Inverse Distance Weight (IDW)	The distance serves as a measure of spatial proximity. As the distance d_{ij} diminishes, the weight w_{ij} proportionally increases, signifying that the weight is inversely related to the distance.	$w_{ij} = d_{ij}^{-\alpha}$, where $\alpha = 1$. If $\alpha = 2$, commonly known as the Power Distance Weight
3.		k-Nearest Neighbor (k-NN)	Let d_{ij} denote the distance between the centroids of location-i and location-j, where $i \neq j$. Subsequently, these distances, denoted as $d_{ij(1)} \leq d_{ij(2)} \leq \cdots \leq d_{ij(\text{n}-1)}$, are arranged in ascending order. Subsequently, the k-NN locations from location-i are established, denoted as $N_k(i) = \{j(1), j(2), \dots, j(k)\}$, with k taking values from 1 to n-1. Subsequently, it is considered neighbors if the distance between those locations is among the-k nearest neighbors.	The k-nearest neighbor locations from location-i are assigned values according to the following criteria. $w_{ij} = \begin{cases} 1, & j \in N_k(i) \\ 0, & \text{other.} \end{cases}$
4.		Threshold Weight/ Distance Band Weight (DBW)	In the threshold weight approach, a predetermined threshold distance, denoted as d_{max}, is established. This value signifies the maximum distance for determining spatial dependence between location-i and location-j. Locations with distances smaller than the threshold are considered neighbors.	$w_{ij} = \begin{cases} 1, 0 \leq d_{ij} \leq d_{max} \\ 0, d_{ij} > d_{max} \end{cases}$

No	Basis of spatial weighting matrix calculation	Types of weighting matrices	Concept	Equation
5.		Uniform Weight	Uniform weighing allocates equal weight to all observed areas, providing each region with the same weight value, irrespective of its relationship to other areas. This approach is commonly adopted when the observed area exhibits homogeneity with other regions [23]. The uniform weighing matrix is defined as follows."	$W_{ij} = \frac{1}{n_i}$ where n_i is the area around area i.

2.3 Testing for Spatial Dependence

The determination of a suitable spatial dependence model for the research data is initiated by a thorough examination of spatial dependence. Broadly, two (2) primary tests for spatial dependence are employed: (1) Moran's Index and (2) Lagrange Multiplier Test (LM). Moran's Index is utilized to discern the Spatial Error Model (SEM), whereas the LM test is deployed to identify models such as Spatial Autoregressive (SAR), SEM, or Generalized Spatial Model (GSM). In spatial econometrics, the LM test was introduced by Anselin [24] for the Spatial Autoregressive (SAR) and Generalized Spatial Model (GSM). Subsequently, Burridge [25] extended the LM test for the Spatial Error Model (SEM). Anselin et al. [26] developed a robust LM test. Both the LM test and the robust LM test operate under the assumption of normally distributed errors. For the SEM model, Kelejian and Robinson [27] proposed a computationally straightforward test that circumvents the need for normality assumptions contingent upon a sufficiently large sample size.

The equations for the Lagrange Multiplier (LM) and robust LM tests for the SAR, SEM, and GSM are presented in **Equations (2)-(6)**. In the SAR model, the null hypothesis (H_0) assumes the absence of spatial lag dependence (ρ=0), with the alternative hypothesis (H_1) proposing the existence of spatial dependence (ρ≠0). The test statistics for LM in SAR (LM_{SAR}) and robust LM (RLM_{SAR}) are articulated in **Equations (2)** and **(3)**.

$$LM_{SAR} = \frac{\left(\frac{\boldsymbol{e}'\boldsymbol{W}\boldsymbol{y}}{\hat{\sigma}^2{}_{ML}}\right)^2}{\frac{\left(\boldsymbol{W}\boldsymbol{X}^*\widehat{\boldsymbol{\beta}^*}\right)'\boldsymbol{M}\boldsymbol{W}\boldsymbol{X}^*\widehat{\boldsymbol{\beta}^*}}{\hat{\sigma}^2{}_{ML}} + tr[\boldsymbol{W}^2 + \boldsymbol{W}'\boldsymbol{W}]} \tag{2}$$

$$RLM_{SAR} = \frac{\left(\frac{\boldsymbol{e}'\boldsymbol{W}\boldsymbol{y} - \boldsymbol{e}'\boldsymbol{W}\boldsymbol{e}}{\hat{\sigma}^2{}_{ML}}\right)^2}{\left(\frac{\left(\boldsymbol{W}\boldsymbol{X}^*\widehat{\boldsymbol{\beta}^*}\right)'\boldsymbol{M}\boldsymbol{W}\boldsymbol{X}^*\widehat{\boldsymbol{\beta}^*} + tr\left[\widehat{\boldsymbol{W}^2} + \boldsymbol{W}'\boldsymbol{W}\right]}{\hat{\sigma}^2{}_{ML}}\right) + tr[\boldsymbol{W}^2 + \boldsymbol{W}'\boldsymbol{W}]} \tag{3}$$

where $\boldsymbol{e}$ is the error vector from the regression model, $\boldsymbol{y}$ is the dependent variable vector, $\boldsymbol{W}$ is the spatial weighting matrix, $\boldsymbol{X}$ is the predictor matrix, $\widehat{\boldsymbol{\beta}}$ is the estimated regression parameter, tr [.] is the trace of a matrix, and $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}^*\left(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}'$, and $\hat{\sigma}^2{}_{ML}$ is the maximum likelihood function for the error variance of the regression model. $\boldsymbol{X}^* = \left(\boldsymbol{i}_n, x_1, \cdots, x_p\right)$ is a matrix of constants and predictors of size $n \times (p+1)$ and $\boldsymbol{\beta}^* = (\beta_0, \beta_1, \cdots, \beta_p)'$ is the coefficient vector of regression parameters of size $(p+1) \times 1$. The test criteria are that H_0 is rejected if $M_{SAR} > \chi^2_{(1),\alpha}$ or $RLM_{SAR} > \chi^2_{(1),\alpha}$. For the SEM model, the hypotheses under investigation are: $H_0: \lambda = 0$ (no spatial error dependence) and $H_1: \lambda \neq 0$ (there is spatial error dependence). The test statistics for LM and robust LM for the SEM model are presented in **Equations (4)** and **(5)**.

$$LM_{ERR} = \frac{\left(\frac{\boldsymbol{e}'\boldsymbol{W}\boldsymbol{e}}{\hat{\sigma}^2{}_{ML}}\right)^2}{tr[\boldsymbol{W}^2 + \boldsymbol{W}'\boldsymbol{W}]} \tag{4}$$

$$RLM_{ERR} = \frac{\left(\frac{e'We}{\hat{\sigma}^2_{ML}} - tr[W^2 + W'W]\left(\frac{(WX^*\widehat{\beta^*})'MWX^*\widehat{\beta^*}}{\hat{\sigma}^2_{ML}} + tr[W^2 + W'W]\right)^{-1}\left(\frac{e'Wy}{\hat{\sigma}^2_{ML}}\right)\right)^2}{tr[W^2 + W'W] - (tr[W^2 + W'W])^2\frac{(WX^*\widehat{\beta^*})'MWX^*\widehat{\beta^*}}{\hat{\sigma}^2_{ML}} + tr[W^2 + W'W]} \quad (5)$$

The testing criteria dictate the rejection of the null hypothesis (H_0) if either $LM_{SEM} > \chi^2_{(1),\alpha}$ or $RLM_{SEM} > \chi^2_{(1),\alpha}$. In the Generalized Spatial Model (GSM), the hypotheses under scrutiny are as follows: $H_0: \rho = 0$ or $\lambda = 0$ (signifying no spatial dependence), and $H_1: \rho \neq 0$ or $\lambda = 0$ (indicating spatial dependence in lag and error terms). The test statistic for LM in GSM is delineated by **Equation (6)**.

$$LM_{GSM} = \frac{\left(\frac{e'Wy}{\hat{\sigma}^2_{ML}} - \frac{e'We}{\hat{\sigma}^2_{ML}}\right)^2}{\left(\frac{(WX^*\widehat{\beta^*})'MWX^*\widehat{\beta^*}}{\hat{\sigma}^2_{ML}} + tr[W^2 + W'W]\right) - tr[W^2 + W'W]} + \frac{\left(\frac{e'We}{\hat{\sigma}^2_{ML}}\right)^2}{tr[W^2 + W'W]} \quad (6)$$

The test criteria are that H_0 is rejected if $LM_{GSM} > \chi^2_{(2),\alpha}$.

2.4 Spatial Dependency Model

In linear regression models, the assumption that must be met is that observations are independent. Regression analysis is an analytical approach that characterizes the linear relationship between two or more variables: the independent variable and the dependent variable. The primary goal of regression analysis is to estimate the variability in the dependent variable influenced by the independent variable in a given observation **[28]**. As outlined by Supanggat **[29]**, the regression model is defined by **Equation (7)**.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad (7)$$

In instances characterized by spatial dependence in observations, it becomes imperative to augment the regression model with a weight matrix that encapsulates the interdependence among locations. Such dependencies may manifest in the dependent variable, predictors, errors, or their combinations. The Generalized Spatial Nested (GNS) model can be formally expressed through **Equation (8) [30]**.

$$\boldsymbol{y} = \rho \boldsymbol{W_1 y} + \boldsymbol{X^*\beta^*} + \boldsymbol{W_2 X\gamma} + \boldsymbol{u}, \text{ where } \boldsymbol{u} = \lambda \boldsymbol{W_3 u} + \boldsymbol{\varepsilon} \quad (8)$$

Assuming $\boldsymbol{\varepsilon}$ follows a normal distribution with a mean (μ) equal to 0 and a variance (σ^2) equal to $\sigma^2 \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix of size $n \times n$, then ε is a random variable distributed as $N(0, \sigma^2\boldsymbol{I})$, **y** is a dependent variable vector of size $n \times 1$, ρ is the autoregressive coefficient of the lagged dependent variable, $\boldsymbol{W_1}$ is the spatial weighting matrix for the dependent variable of size $n \times n$, $\boldsymbol{X}^* = (\boldsymbol{i}_n, x_1, \cdots, x_p)$ is a matrix of constants and predictors of size $n \times (p+1)$, $\boldsymbol{i}_n$ is a vector with elements valued as one of size $n \times 1$, $\boldsymbol{\beta}^* = (\beta_0, \beta_1, \cdots, \beta_p)'$ is the coefficient vector of regression parameters of size $(p+1) \times 1$, $\boldsymbol{W_2}$ is the spatial weighting matrix for predictors of size $n \times n$, $\boldsymbol{X} = (x_1, \cdots, x_p)$ is the matrix of predictors of size $n \times p$, $\boldsymbol{\gamma}$ is the autoregressive coefficient vector of size $p \times 1$, $\boldsymbol{W_3}$ is the spatial weighting matrix for errors of size $n \times n$, u is the assumed autocorrelated error vector of size $n \times 1$, λ is the autoregressive coefficient of errors, and $\boldsymbol{I}$ is the identity matrix of size $n \times n$. The determination of the spatial regression model can be observed in **Figure 1**.

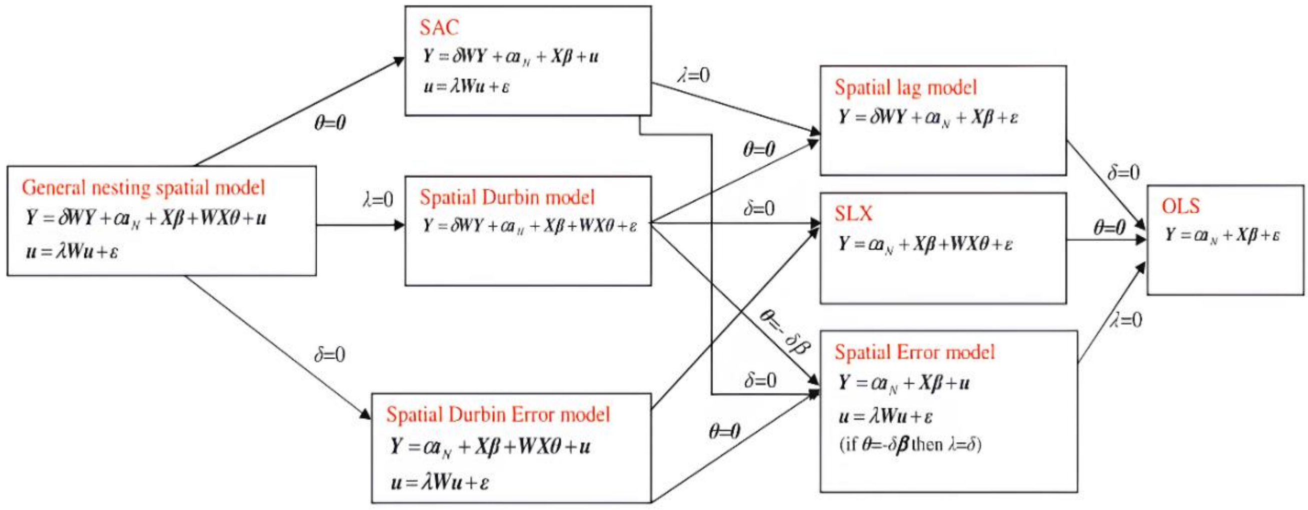


Figure 1. **Taxonomy Spatial Dependence Model [30]**

Based on **Figure 1** and **Equation (8)** several spatial models that may be employed, such as the SAR, SEM, SLX, and GSM are presented in **Table 3**.

Table 3. **Spatial Model Variations**

No	Model	Model equation	Description
1	Spatial Autoregressive Model (SAR)	$\boldsymbol{y} = \rho \boldsymbol{W_1 y} + \boldsymbol{X^* \beta^*} + \boldsymbol{\varepsilon}$	○ ε_i is assumed to follow a normal distribution, be stochastically independent, identically centered around zero with a variance of ragam σ^2 ($\varepsilon \sim N(0, \sigma^2 \boldsymbol{I})$). ○ SAR is a linear regression model wherein spatial autocorrelation is present in the dependent variable.
2	Spatial Error Model (SEM)	$\boldsymbol{y} = \boldsymbol{X^* \beta^*} + \boldsymbol{u}$ $\boldsymbol{u} = \boldsymbol{\lambda W_3 u} + \boldsymbol{\varepsilon}$	○ SEM is a linear regression model characterized by spatial autocorrelation in its error term. ○ The prediction in SEM involves three components: (1) the smoothing factor ($\boldsymbol{X\beta}$), referred to as the trend, (2) the spatial factor ($\boldsymbol{\lambda W u}$) or signal, and (3) the Fit, representing the summation of the trend and signal **[31].**
3	Spatial lag of X (SLX)	$\boldsymbol{y} = \boldsymbol{X^* \beta^*} + \boldsymbol{W_2 X \gamma} + \boldsymbol{\varepsilon}$	SLX is a regression model characterized by spatial autocorrelation among the predictor variables.
4	General Spatial Model (GSM)/ Spatial Autoregressive Confused (SAC)/ Spatial Autoregressive Moving Average (SARMA)	$\boldsymbol{y} = \rho \boldsymbol{W_1 y} + \boldsymbol{X^* \beta^*} + \boldsymbol{u}$ $\boldsymbol{u} = \boldsymbol{\lambda W_3 u} + \boldsymbol{\varepsilon}$	SAC consists of an autoregressive component in the dependent variable (ρ) and an autoregressive component in the error term (λ).

2.5 Testing Assumptions and Goodness of Fit Test

Subsequent to obtaining an appropriate model, the subsequent phase involves scrutinizing the assumptions inherent in the derived model. Assumptions for spatial models encompass (1) normally distributed residuals, (2) homogeneity of residual variances, and (3) the absence of spatial autocorrelation in residuals **[32]**, **[33]**. The normality of residuals is assessed through the Kolmogorov-Smirnov test **[34]**. The homogeneity of residual variances is tested using the Breusch-Pagan test **[35]**, whereas the absence of spatial autocorrelation in residuals is examined through Moran's index **[36]**.

A goodness-of-fit test for the regression model is essential to assess the model's effectiveness in predicting the relationship between the independent and dependent variables, both overall and for each utilized independent variable **[37]**. The goodness-of-fit test involves employing various measures, namely: (1) Mean Absolute Error (MAE), (2) Mean Square Error (MSE), (3) Root Mean Square Error (RMSE), (4) pseudo-R^2, (5) Akaike Information Criterion (AIC), and (6) Mean Absolute Percentage Error (MAPE) **[38]**. A good model is characterized by lower values of MAE, MSE, RMSE, AIC, and MAPE. This is because a smaller difference between predictions and actual data indicates greater accuracy in the generated predictions [39]. Similarly, models with lower AIC values are considered more optimal as they efficiently organize data explanations with the minimum number of parameters [40]. Meanwhile, a higher pseudo-R^2 value elucidates a better-fitting model [41]. The accuracy of the prediction percentage for forecast error (MAPE) is presented in Table 4 [42].

Table 4. MAPE Criteria for Model Evaluation

No	MAPE Value	Prediction Accuracy
1	MAPE $\leq$ 10%	Precise prediction
2	10% $< MAPE \leq$ 20%	Reliable prediction
3	20% $< MAPE \leq$ 50%	Prudent prediction
4	$MAPE >$ 50%	Poor forecasting

Next, to compare whether a spatial model significantly differs from a linear model, the likelihood ratio (LR) test can be conducted. Additionally, LR can be employed to assess the suitability of the formed model. The LR test criterion is H_0 is rejected if $\chi^2_{LR} > \chi^2_{\alpha;p}$ or p-value $< \alpha$, where H_0 defines that the alternative model is not deemed suitable [43].

2.6 Research Flow

The research flowchart is depicted in
Figure 2.

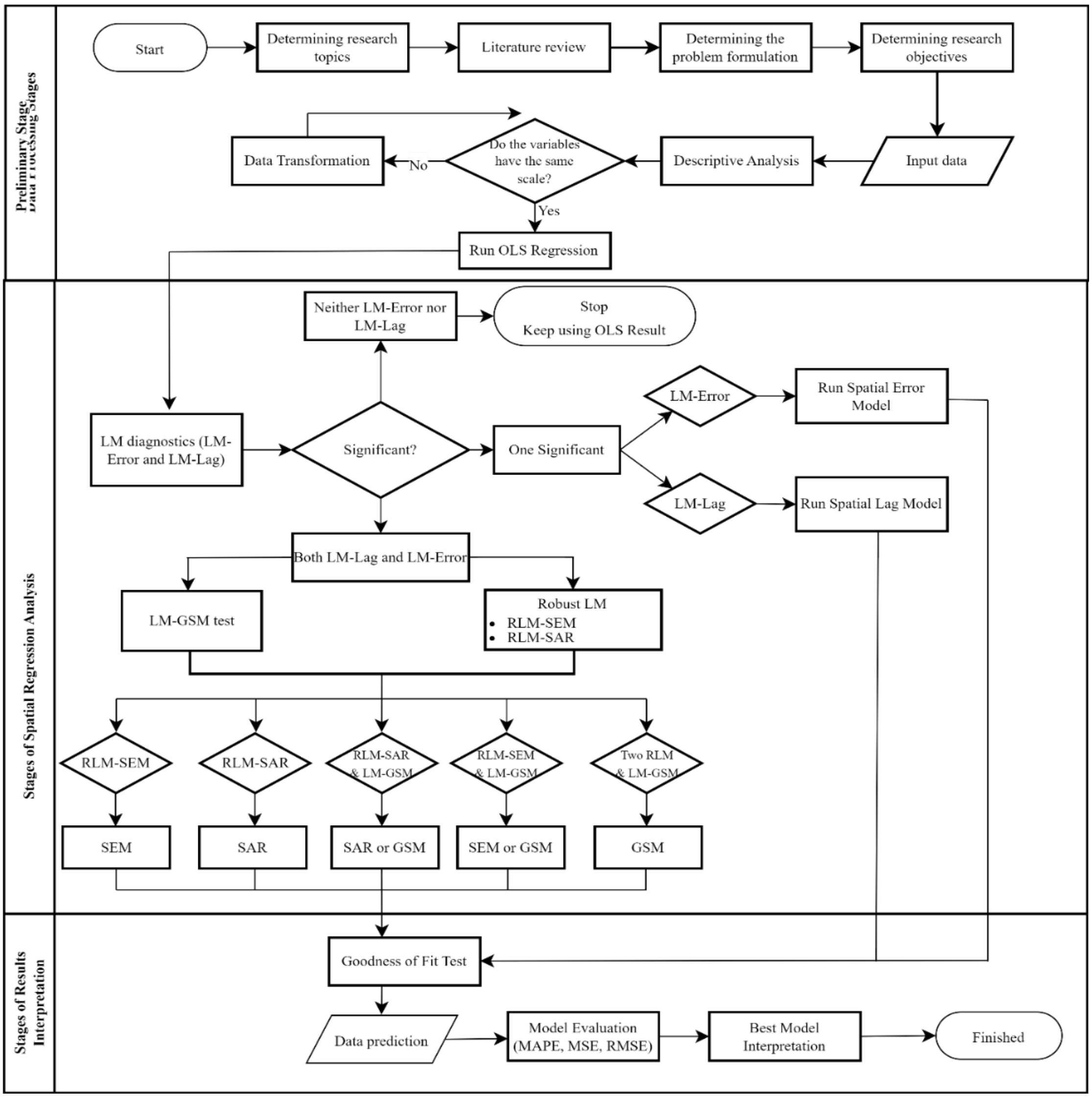


Figure 2. Research Flow Diagram [19]

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

The primary objective of conducting descriptive analysis is to articulate data in a manner that is both accessible and engaging for readers, facilitating the retrieval of pertinent information. In this study, a comprehensive descriptive analysis was undertaken on the Population Poverty Rate data for East Java Province in 2021. The data was stratified into three distinct categories: Low, Medium, and High. **Figure 3** visually illustrates the distribution of poverty levels.

Figure 3. **Percentage Poverty in the East Java Province at Regency/ City Level**

Figure 3 presents a comparison of the poverty rate by regency/city in East Java Province in 2021. In 2021, the average poverty rate in East Java Province was 11.32%, where Batu City was the area with the lowest poverty rate in East Java Province at 4.09%, as Batu City is a tourist area. Meanwhile, Sampang Regency was the area with the highest poverty in East Java Province at 23.76%. The number is associated with farmers making up the majority of Sampang Regency residents. Farmers in this area have relatively low incomes, so they are unable to improve the population welfare. The subsequent step involves the examination of multicollinearity. Multicollinearity testing is carried out through the computation of Variance Inflation Factor (VIF) values, with the results presented in **Table 5**.

Table 5. Variance Inflation Factor (VIF) values

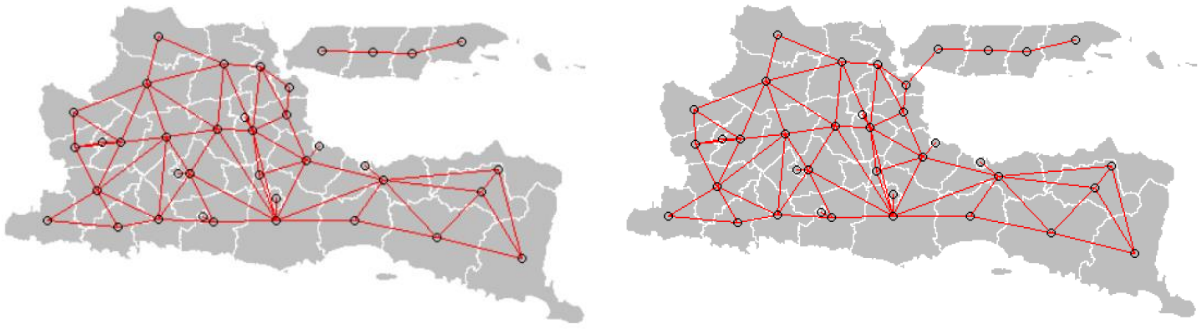
Variable	X_1	X_2	X_3	X_4	X_5
VIF Value	3.31	1.37	4.56	5.76	1.53

According to **Table 5**, the Variance Inflation Factor (VIF) values for all five predictors are sufficiently low (less than 10) **[44]**. Consequently, it can be inferred that there is an absence of multicollinearity among all employed predictors. Following this, linear regression modeling ensues, accompanied by tests for the significance of its parameters. A backward stepwise regression elimination test **[45]** is executed to ensure the retention of only significant variables. Initially encompassing five (5) variables, subsequent elimination reveals a sole significant variable, namely the Average Years of Schooling variable (X_3). In line with **Equation (7)**, the regression model equation is elucidated in **Equation (9)**.

$$\hat{y} = 31.048 - 2.447(X_3) \tag{9}$$

3.2 Estimation of Spatial Model Parameters

Preceding the computation of the spatial regression model, the formulation of a spatial weighting matrix takes precedence. In this study, the spatial weighting matrix incorporates varied weights, as delineated in **Table 2**. As illustrated in **Figure 4** (a), a visual representation of the contiguity chain plot is presented utilizing the Queen-contiguity approach. Modification of the contiguity criteria was necessary; hence, Surabaya City and Bangkalan District were deemed contiguous, facilitated by the direct connection through the Surabaya-Madura Bridge (Suramadu), as elucidated in Figure 4 (b).



(a) (b)

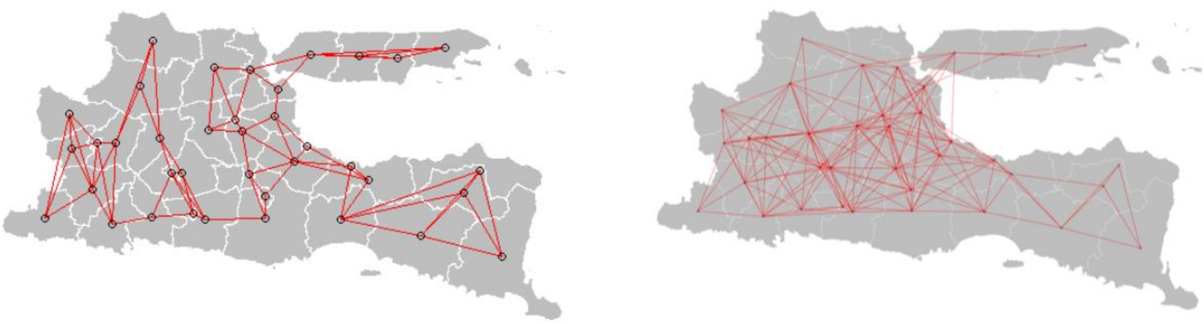
Figure 4. Chain Plot of Spatial Adjacency between Locations Based on Contiguity (a) Queen Contiguity Approach, (b) Modified Queen Contiguity Approach

In the computation of the distance-weighted matrix, it is imperative to ensure that coordinates are represented in a projective coordinate system for computational efficiency. In this study, the Universal Transverse Mercator (UTM) coordinate system for zone 49S was employed, and the conversion from geographic coordinates to projective coordinates was facilitated using QGIS software [46]. Subsequently, the Euclidean distance metric was applied. The distance (d_{ij}) between the central point of location-i with coordinates (u_i, v_i) and location-j with coordinates (u_j, v_j) is detailed in Equation (10).

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + (v_i - v_i)^2} \tag{10}$$

The variable u_i denotes the latitude coordinate of location-i, while u_j corresponds to the latitude coordinate of location-j. Similarly, v_i signifies the longitude coordinate of location-i, and v_j represents the longitude coordinate of location-j. Utilizing Equation (10) in conjunction with Table 2.

The weighting assignment in the k-NN matrix will be performed by first determining the value of k, then calculating the Euclidean distance. Subsequently, the region with the nearest distances will be assigned a weight of $\frac{1}{k}$ for the specified k value. In Figure 5 (a) using $k = 3$ as an example. Like k-NN, the weighting assignment in the radial distance matrix is initially determined by establishing a threshold (d_{max}) as a reference for weight assignment. Subsequently, Euclidean distances are calculated, and weights are assigned to areas with distances less than the specified threshold. As an example, in Figure 5 (b) a threshold value of 75000 m is used.



(a) (b)

Figure 5. Chain Plot of Spatial Adjacency between Locations Based on Distance (a) k-NN (k=3), (b) DBW (d_{max}=75000 m)

After obtaining the spatial weighting matrices as depicted in Figure 4 and Figure 5, the subsequent step involves conducting the Lagrange Multiplier (LM) test using Equations (2) through Equations (6). In the SAR model, the test criteria dictate rejecting the null hypothesis (H_0) if $M_{SAR} > \chi^2_{(1),\alpha}$ or $RLM_{SAR} > \chi^2_{(1),\alpha}$, or if the p-value is less than α. Meanwhile, for SEM, the criteria include rejecting H_0 if $LM_{SEM} > \chi^2_{(1),\alpha}$ or $RLM_{SEM} > \chi^2_{(1),\alpha}$. A significance level of 5% is applied, resulting in $\chi^2_{(1),\alpha} = 3.841$ The LM test results are detailed in **Table 6**.

Table 6. The LM Test Results

No	Spatial weighting matrix	SAR				SEM				GSM		Decisions	Conclusion
		LM_{SAR}	p-val	RLM_{SAR}	p-val	LM_{ERR}	p-val	RLM_{ERR}	p-val	GSM	p-val		
1.	Queen Contiguity	3.00	0.08	0.05	0.82	6.56	0.01	3.61	0.06	6.61	0.04	Do not reject H_0 in LM_{SAR}, but reject H_0 in LM_{ERR} and GSM	SEM
2.	Modified Queen Contiguity	0.71	0.39	2.04	0.15	5.19	0.02	6.52	0.011	7.23	0.03	Do not reject H_0 in LM_{SAR}, but reject H_0 in LM_{ERR}, RLM_{ERR}, and GSM	SEM
3.	IDW	0.35	0.55	0.22	0.64	1.53	0.21	1.40	0.24	1.75	0.42	All examinations of LM do not lead to the rejection of the null hypothesis (H_0)	Linear Regression
4.	k-NN (k=1)	0.33	0.57	0.92	0.34	0.06	0.81	0.65	0.42	0.98	0.61	All examinations of LM do not lead to the rejection of the null hypothesis (H_0)	Linear Regression
5.	k-NN (k=3)	3.06	0.08	0.15	0.70	10.63	0.00	7.72	0.00	10.78	0.00	Do not reject H_0 in LM_{SAR}, but reject H_0 in LM_{ERR}, RLM_{ERR}, and GSM	SEM
6.	k-NN (k=5)	0.58	0.45	4.22	0.04	14.01	0.00	17.65	2.6E-05	18.23	0.00	Do not reject H_0 in LM_{SAR}, but reject H_0 in LM_{ERR}, RLM_{ERR}, and GSM	SEM
7.	DBW ($d_{max} = 55000m$)	0.88	0.35	1.14	0.29	7.38	0.01	7.64	0.01	8.52	0.01	Do not reject H_0 in LM_{SAR}, but reject H_0 in LM_{ERR}, RLM_{ERR}, and GSM	SEM
8.	DBW ($d_{max} = 60000m$)	1.36	0.24	0.56	0.45	7.42	0.01	6.62	0.01	7.98	0.02	Do not reject H_0 in LM_{SAR}, but reject H_0 in LM_{ERR}, RLM_{ERR}, and GSM	SEM
9.	DBW ($d_{max} = 65000m$)	4.39	0.03	0.12	0.73	15.41	0.00	11.13	0.00	15.53	0.00	Reject H_0 in LM_{SAR}, LM_{ERR}, RLM_{ERR}, and GSM	SEM or GSM
10.	DBW ($d_{max} = 70000m$)	3.68	0.05	0.14	0.70	13.75	0.00	10.22	0.00	13.90	0.00	Do not reject H_0 in LM_{SAR}, but reject H_0 in LM_{ERR}, RLM_{ERR}, and GSM	SEM
11.	DBW ($d_{max} = 75000m$)	4.93	0.03	0.03	0.85	16.24	0.00	11.34	0.00	16.27	0.00	Reject H_0 in LM_{SAR}, LM_{ERR}, RLM_{ERR}, and GSM	SEM or GSM
12.	Uniform Weight	0.51	0.47	0.00	1.00	0.51	0.47	0.00	1.00	0.51	0.77	All examinations of LM do not lead to the rejection of the null hypothesis (H_0)	Linear regression

According to **Table 6,** among the various feasible weightings considered, the spatial regression model was established utilizing 9 specific weightings: Queen Contiguity, Modified Queen Contiguity, k-NN ($k = 3$, and $k = 5$), and DBW (d_{max}=55000 m, d_{max}=60000 m, d_{max}=65000 m, d_{max}=70000 m, and d_{max}=75000 m). In **Table 6**, the use of the number of k nearest neighbors is done by trial and error and is assumed to use The First Law of Geography from Tobler which relates that the closer the object is, the greater the influence it will have **[47]**. While the determination of the d_{max} values was performed iteratively, commencing with a distance of 55000m. This was necessitated by the minimum threshold required to generate an invertible weighting matrix. If the distance falls below this threshold, a row in the weighting matrix contains elements that are all zero or lack neighboring regions, resulting in a matrix with a determinant of 0. Consequently, the weighting matrix becomes non-invertible, rendering it incapable of solving the equation **[48].** From these 9 weightings, AIC values were computed and are presented in **Table 7.**

Table 7. **Comparison of AIC Values among Various Spatial Regression Models**

Evaluations	Models	*Queen Contiguity*	*Modified Queen Contiguity*	*k-NN* ($k = 3$)	*k-NN* ($k = 5$)	DBW (d_{Max}= 55000 m)	DBW (d_{Max}= 60000 m)	DBW (d_{Max}= 65000 m)	DBW (d_{Max}= 70000 m)	DBW (d_{Max}= 75000 m)
AIC Value	SEM	184.23	185.50	182.85	181.38	183.32	184.22	181.05	181.63	180.98
	GSM	-	-	-	-	-	-	182.60	-	182.85

Based on the findings presented in **Table 7**, the SEM model incorporating DBW weighting (d_{max}=75000 m) emerges as the model with the most favorable AIC value. Subsequent analyses will therefore focus on the application of SEM spatial regression with DBW weighting (d_{max}=75000). The formulation of the SEM model is detailed in **Equation (11)**. The components derived from the SEM are presented in **Table 8**.

$$\hat{y} = 31.27 - 2.45(\mathrm{X}_3) + \varepsilon_i, \text{with } \varepsilon_i = 0.64 \sum_{j=1}^{n} W_{ij}\varepsilon_j \tag{11}$$

Table 8. **The Components of the SEM with DBW Weighting (d_{max}=75000 m)**

No	Regency/ City	Fit	Trend	Signal
1	Bangkalan	17.53	16.69	0.85
2	Banyuwangi	11.09	13.12	-2.03
3	Batu	7.91	8.49	-0.58
4	Blitar	12.39	12.92	-0.53
5	Bojonegoro	13.78	13.21	0.56
6	Bondowoso	15.15	16.74	-1.59
7	Gresik	8.28	7.88	0.40
8	Jember	13.92	15.39	-1.47
9	Jombang	10.26	10.35	-0.09
10	Kediri	10.99	11.50	-0.51
11	Bllitar City	5.11	5.95	-0.83
12	Kediri City	5.97	6.44	-0.46
13	Madiun City	3.38	3.45	-0.07
14	Malang City	5.02	5.80	-0.78
15	Mojokerto City	5.78	5.65	0.12
16	Pasuruan City	8.21	8.44	-0.23
17	Probolinggo City	7.97	9.37	-1.40
18	Lamongan	12.44	11.60	0.84
19	Lumajang	13.54	14.95	-1.41
20	Madiun	12.01	12.14	-0.13
21	Magetan	11.08	10.82	0.27
22	Malang	12.08	13.09	-1.01
23	Mojokerto	10.10	10.13	-0.03
24	Nganjuk	12.17	12.24	-0.06
25	Ngawi	13.33	13.51	-0.17
26	Pacitan	12.38	12.65	-0.27
27	Pamekasan	17.67	14.88	2.79
28	Pasuruan	12.74	13.14	-0.40
29	Ponorogo	12.98	12.80	0.18
30	Probolinggo	14.27	16.30	-2.02
31	Sampang	20.78	19.38	1.40
32	Sidoarjo	4.90	5.04	-0.14
33	Situbondo	12.49	15.07	-2.59
34	Sumenep	18.33	16.79	1.55
35	Surabaya	6.07	5.58	0.49
36	Trenggalek	12.63	12.77	-0.14
37	Tuban	14.75	13.70	1.04
38	Tulungagung	10.72	10.87	-0.14

The test statistics for each parameter of **Equation (11)** are detailed in **Table 9**. Criteria for parameter testing (β_0 and β_1) involve rejecting the null hypothesis (H_0) if $|Z_{cal}| > Z_{tab}$ or if the p-value is less than α. With α set at 5%, the critical Z_{tab} value is 1.96. The coefficients of the autoregressive error are tested using the Wald test, with the null hypothesis rejected if the Wald value surpasses $\chi^2_{\alpha,1}$ (3.841) **[49]**.

Table 9. **Significance Testing of Each Parameter**

Parameters	Z_{cal}	p-val	Decisions	Conclusions
$\beta_0 = 31.27$	13.15	< 2.2e-16	Rejected H_0	Significant parameters
$\beta_1 = -2.45$	-8.96	< 2.2e-16	Rejected H_0	Significant parameters
$\lambda = 0.64$	Wald statistic (15.25)	9.38e-05	Rejected H_0	Significant parameters

According to **Equation (11)**, the variable (X_3) demonstrates a negative coefficient of -2.45 concerning the poverty rate in East Java Province. This implies that an increase in the level of formal education pursued by residents in the area is associated with a reduction in the poverty rate. The results of this study are in line with several other studies, including Hofmarcher **[50]**, Brown **[51]**, and Tilak **[52]**. In essence, a higher attainment of formal education by the population in East Java Province is correlated with a decreased poverty percentage in the region. This finding suggests practical implications, such as optimizing compulsory education for children, particularly up to the high school level, especially in areas identified as high-risk in **Figure 3**. Moreover, the mandate for compulsory education could gradually be extended to encompass diploma and bachelor's levels. This emphasis on education extends beyond formal schooling to include the familial environment. Undoubtedly, the family plays a pivotal role in shaping the educational trajectory of a child **[53]**. **Equation (11)** yields an autoregressive error coefficient (λ=0.64), signifying that the error in a district/city will increase by 0.64 times the average error of its neighboring areas, assuming other variables remain constant. This finding is followed by an assessment of the SEM model assumptions, as detailed in **Table 10**.

Table 10. Testing the Assumptions of the SEM Model.

No	Testing the Assumptions of the SEM Model	Test statistics value and p-value	Test criteria	Decisions	Conclusions
1.	Residual Normality using the Kolmogorov-Smirnov test.	D = 0.127 p-val= 0.123	The null hypothesis (H_0) is rejected if $D_{cal} > D_{tab}$ b or if the p-val is less than α. In this context, H_0 posits that the residuals adhere to a normal distribution **[54].**	$BP = 0.127 < 0.210 = D_{tab}$ or p-val= $0.123 > 0.05 = \alpha$	There is insufficient evidence to reject H_0; thus, it can be concluded that the residuals in the SEM model exhibit a normal distribution.
2.	Homogeneity of residuals using the Breusch-Pagan (BP) test.	BP = 0.832 p-val = 0.361	The null hypothesis (H_0) is rejected if $BP > \chi^2_{(df),\alpha}$ or p-val $< \alpha,$ H_0 posits the assumption of homogeneity of residual variances is satisfied **[55].**	$BP = 0.832 < 3.841 = \chi^2_{(1),0.05}$ or p-val= $0.361 > 0.05 = \alpha$	There is insufficient evidence to reject H_0; thus, it can be concluded that the variance of residuals in the SEM model is homogeneous
3.	Non-autocorrelation of residuals using Moran's Index.	$Z_{cal} = 0.714$ p-val=0.237	The null hypothesis (H_0) is rejected if $Z_{cal} > Z_{tab}$ or p-val < α, H_0: defines the absence of spatial autocorrelation in the residuals **[31].**	$Z_{cal} = 0.714 < 1.96 = Z_{tab}$ or p-val= $0.237 > 0.05 = \alpha$	There is insufficient evidence to reject H_0; thus, it can be concluded that there is no spatial autocorrelation in the residuals.

Next, the determination of model goodness is continued, as presented in **Table 11.**

Table 11. Determination of Model Goodness

Model	MAE	MSE	RMSE	Pseudo-R^2	AIC	MAPE (%)
SEM using DBW (d_{max} =75000 m)	1.85	5.16	2.27	76.19%	180.98	18.77
Linear Regression	2.17	7.20	2.68	66.81	188.85	21.70

Based on **Table 11**, the SEM model using DBW yields smaller values for MAE, MSE, RMSE, AIC, and MAPE compared to the linear regression model. Similarly, the pseudo-R^2 value obtained for SEM is larger than that of the linear regression model. Thus, based on these metrics, the SEM model outperforms linear regression. According to **Table 4,** the MAPE value for the SEM model falls within the category of good forecasting. Furthermore, from the analysis of the likelihood ratio (LR) test, a LR value of χ^2_{LR} =7.614 is obtained, which is greater than $\chi^2_{0.05;1} = 3.841$, or $p_{Value} = 0.005$<0.05=α. Therefore, it can be stated that the SEM model with DBW (d_{max}=75000 m) is more efficient than the linear regression model and provides a significant improvement in the model's ability to explain variation in the data.

4. CONCLUSION

Based on spatial analysis, among the twelve simulated weightings considering both distance and contiguity variations, the best model for analyzing the poverty rate in East Java Province is the Spatial Error Model (SEM) using Distance Band Weight (DBW) with a maximum distance value of 75000 m. After conducting backward stepwise regression elimination, only one predictor variable is found to be significant out of the initial five, namely, Average Years of Schooling (X_3), with the SEM model equation being $\hat{y} = 31.27 - 2.45(X_3) + \varepsilon_i$ where $\varepsilon_i = 0.64\sum_{j=1}^{n} W_{ij}\varepsilon_j$. The variable X_3 has a negative influence on the Population Poverty Rate, implying that a higher level of formal education pursued by residents in East Java

Province tends to reduce the poverty percentage in the region. The SEM model yields smaller values for MAE, MSE, RMSE, AIC, and MAPE compared to the linear regression model. Similarly, the pseudo-R^2 value obtained for SEM is larger, indicating that SEM outperforms linear regression based on these metrics. The likelihood ratio test also reveals a significant difference between the SEM and linear regression models, with the SEM model showing superior performance.

REFERENCES

[1] R. T. Ashari and M. Athoillah, "Analisis pengaruh tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, upah minimum, indeks pembangunan manusia, pertumbuhan ekonomi dan jumlah penduduk terhadap kemiskinan di kawasan tapal kuda," *Journal of Development Economic and Social Studies*, vol. 2, no. 2, pp. 313–326, 2023, doi: 10.21776/jdess.2023.02.2.08.

[2] A. Efendi, N. W. S. Wardhani, R. Fitriani, and E. Sumarminingsih, *Analisis Regresi: Teori dan Aplikasi dengan R*. Malang: UB Press, 2020.

[3] M. Firdaus, T. Irawan, F. S. Ahmad, H. Siregar, D. Siswara, and R. Jakariya, *Aplikasi Model Ekonometrika dengan R Studio*, 1st ed. IPB Press, 2021.

[4] A. Djuraidah, *Monograph Penerapan dan Pengembangan Regresi Spasial dengan Studi Kasus pada Kesehatan, Sosial, dan Ekonomi*. Bogor: IPB Press, 2020.

[5] J. Medina Ariza and R. Solymosi, *Crime Mapping and Spatial Data Analysis using R*. Boca Raton: Chapman and Hall/CRC, 2023. doi: 10.1201/9781003154914.

[6] R. E. Caraka and H. Yasin, *Geographically Weighted Regression (GWR) Sebuah Pendekatan Regresi Geografis*. MOBIUS, 2017.

[7] F. Ramdani, *Geospatial Data Science: Geospatial Data Processing and Analysis using R*. Malang: Pena Persada, 2021.

[8] A. O. Sihombing, "Spatial analysis of poverty in North Sumatera," *Journal of Analytical Research*, vol. 1, no. 1, pp. 64–77, 2022.

[9] A. Tumanggor and E. Simamora, "Pemodelan faktor-faktor yang mempengaruhi indeks pembangunan manusia di Sumatera Utara menggunakan regresi spasial," *Jurnal riset rumpun matematika dan ilmu pengetahuan alam (JURRIMIPA)*, vol. 2, no. 2, 2023.

[10] I. Y. Safitri, M. A. Tiro, and Ruliana, "Spatial regression analysis to see factors affecting food security at district level in South Sulawesi province," *ARRUS journal of mathematics and applied science*, vol. 2, no. 2, pp. 60–72, Mar. 2022, doi: 10.35877/mathscience740.

[11] S. Yulianto and A. A. Cika, "Model tingkat kemiskinan provinsi Jawa Timur dengan analisis regresi spasial," in *Proceeding of Seminar nasional matematika dan pendidikan matematika (6th SENATIK)*, Semarang: Universitas PGRI Semarang, 2021.

[12] R. A. Jelita, "Penerapan matriks pembobot jarak K-Nearest Neighbour dan Distance Band dalam regresi panel spasial SAR dan SEM pada gini rasio di provinsi Jawa Timur tahun 2015-2017," Universitas Brawijaya, Malang, 2020.

[13] E. Wahyu Azizah, Sudarti, and H. Kusuma, "Pengaruh pendidikan, pendapatan perkapita dan jumlah penduduk terhadap kemiskinan di provinsi Jawa Timur," *Jurnal Ilmu Ekonomi*, vol. 2, no. 1, pp. 167–180, 2018.

[14] Muryani and P. A. Pamungkas, "The impact of unemployment rate, labor force, capital, inflation rate, and government expenditure on economic growth in Indonesia," *American Journal of Engineering Research (AJER*, vol. 7, no. 3, pp. 109–119, 2018, [Online]. Available: www.ajer.org

[15] S. Azizi, "The impacts of workers' remittances on human capital and labor supply in developing countries," *Econ Model*, vol. 75, pp. 377–396, Nov. 2018, doi: 10.1016/j.econmod.2018.07.011.

[16] S. Alam, Y. Ansari, N. Sha, and K. Khan, "The influence of unemployment and labor force participation rates on economic development in GCC countries: A cointegration approach," *Journal of infrastructure, policy and development*, vol. 8, no. 2, Dec. 2023, doi: 10.24294/jipd.v8i2.2962.

[17] I. A. A. Widiantari, H. Sahri, and I. Suriadi, "The influence of labor force participation rate, total population and minimum wage on unemployment in West Nusa Tenggara Province in 2017-2021," *Journal of Economics, Finance and Management Studies*, vol. 06, no. 05, May 2023, doi: 10.47191/jefms/v6-i5-46.

[18] S. J. Rey, D. Arribas-Bel, and L. J. Wolf, *Geographic data science with python*. CRC Press, 2023.

[19] A. Djuraidah and R. Anisa, *Analisis statistika untuk data geospasial: dengan ilustrasi penerapan menggunakan R*. Bogor: IPB Press, 2023.

[20] M. Sarrias, "Notes on spatial econometrics," Del Maure, 2020.

[21] T. Purwaningsih, "Kajian pengaruh matriks pembobot spasial dalam model data panel spasial," Thesis, IPB University, Bogor, 2014.

[22] S. Shekar and H. Xiong, *Encyclopedia of GIS*. SpringerScince+Buisiness Media, 2007.

[23] D. Anggraeni, A. Prahutama, and S. Andari, "Aplikasi Generalized Space Time Autoregressive (GSTAR) pada pemodelan volume kendaraan masuk tol semarang," *Media Statistika*, vol. 6, no. 2, Dec. 2013, doi: 10.14710/medstat.6.2.61-70.

[24] L. Anselin, "Lagrange Multiplier test diagnostics for spatial dependence and spatial heterogeneity," *Geogr Anal*, vol. 20, no. 1, pp. 1–17, 1988, doi: 10.1111/j.1538-4632.1988.tb00159.x.

[25] P. Burridge, "On the cliff-ord test for spatial correlation," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 42, no. 1, pp. 107–108, Sep. 1980, doi: 10.1111/j.2517-6161.1980.tb01108.x.

[26] L. Anselin, A. K. Bera, R. Florax, and M. J. Yoon, "Simple diagnostic tests for spatial dependence," *Reg Sci Urban Econ*, vol. 26, no. 1, pp. 77–104, Feb. 1996, doi: 10.1016/0166-0462(95)02111-6.

[27] H. H. Kelejian and D. P. Robinson, "Spatial autocorrelation," *Reg Sci Urban Econ*, vol. 22, no. 3, pp. 317–331, Sep. 1992, doi: 10.1016/0166-0462(92)90032-V.

[28] A. Somantri, *Aplikasi statistika dalam penelitian*. Bandung: Bandung Pubisher, 2006.

[29] A. Supangat, *Statistika : dalam kajian deskriptif, inferensi, dan nonparametrik*. Jakarta: Kencana Pernada Media, 2010.

[30] J. P. Elhorst, “Spatial panel data models,” in *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, M. M. Fischer and A. Getis, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. doi: 10.1007/978-3-642-03647-7.

[31] K. Kopczewska, *Applied Spatial Statistics and Econometrics*. Routledge, 2020. doi: 10.4324/9781003033219.

[32] S. F. Higazi, D. H. Abdel-Hady, and S. A. Al-Oulfi, “Application of spatial regression models to income poverty ratios in middle delta contiguous counties in Egypt,” *Pakistan Journal of Statistics and Operation Research*, vol. 9, no. 1, p. 93, Feb. 2013, doi: 10.18187/pjsor.v9i1.272.

[33] G. Chi and J. Zhu, “Spatial regression models for demographic analysis,” *Popul Res Policy Rev*, vol. 27, no. 1, pp. 17–42, 2008, [Online]. Available: http://www.jstor.org/stable/41217935

[34] F. C. Cardoso, R. A. Berri, G. Lucca, E. N. Borges, and V. L. D. de Mattos, “Normality tests: a study of residuals obtained on time series tendency modeling,” *Exacta*, Apr. 2023, doi: 10.5585/2023.22928.

[35] A. G. Klein, C. Gerhard, R. D. Büchner, S. Diestel, and K. Schermelleh-Engel, “The detection of heteroscedasticity in regression models for psychological data,” *Psychol Test Assess Model*, vol. 58, no. 4, p. 567, 2016.

[36] Y. Chen, “An analytical process of spatial autocorrelation functions based on Moran’s index,” *PLoS One*, vol. 16, no. 4, p. e0249589, 2021.

[37] D. N. Gujarati, *Basic Econometrics*, 4th edition. Tata McGraw Hill, 2004.

[38] C. Sammut and G. I. Webb, Eds., *Encyclopedia of machine learning*. Boston, MA: Springer US, 2010. doi: 10.1007/978-0-387-30164-8.

[39] A. K. Sahoo, S. Mallik, C. Pradhan, B. S. P. Mishra, R. K. Barik, and H. Das, “Intelligence-based health recommendation system using Big Data analytics,” in *Big Data Analytics for Intelligent Healthcare Management*, Elsevier, 2019, pp. 227–246. doi: 10.1016/B978-0-12-818146-1.00009-X.

[40] A. A. Grasa, *Econometric Model Selection: A New Approach*, vol. 16. Dordrecht: Springer Netherlands, 1989. doi: 10.1007/978-94-017-1358-0.

[41] G. A. J. Hemmert, L. M. Schons, J. Wieseke, and H. Schimmelpfennig, “Log-likelihood-based Pseudo- R^2 in Logistic Regression,” *Sociol Methods Res*, vol. 47, no. 3, pp. 507–531, Aug. 2018, doi: 10.1177/0049124116638107.

[42] C. D. Lewis, *Industrial and business forecasting methods*. Butterworth-Heinemann, 1982.

[43] J. Sudrajat, S. Wira Rizki, and H. Perdana, “Perbandingan model regresi parametrik eksponensial dan Weibull pada data survival tersensor interval,” *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, vol. 07, no. 3, pp. 231–238, 2018.

[44] N. Shrestha, “Detecting multicollinearity in regression analysis,” *Am J Appl Math Stat*, vol. 8, no. 2, pp. 39–42, Jun. 2020, doi: 10.12691/ajams-8-2-1.

[45] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression Analysis*, Sixth. John Wiley and Sons, Inc, 2021.

[46] K. Chang, *An Introduction to Geographic Information Systems*, 9th ed. McGraw-Hill Education, 2018. doi: 10.1201/b12440-10.

[47] W. R. Tobler, “A Computer Movie Simulating Urban Growth in the Detroit Region,” *Econ Geogr*, vol. 46, p. 234, Jun. 1970, doi: 10.2307/143141.

[48] S. Axler, *Undergraduate texts in mathematics linear algebra done right*, 4th edition. San Frascisco: Springer, 2024.

[49] H. Yasin, B. Waryanto, and A. R. Hakim, *Regresi Spasial: Aplikasi dengan R*. Ponorogo: WADE Group, 2020.

[50] T. Hofmarcher, “The effect of education on poverty: A European perspective,” *Econ Educ Rev*, vol. 83, p. 102124, Aug. 2021, doi: 10.1016/j.econedurev.2021.102124.

[51] P. Brown and D. James, “Educational expansion, poverty reduction and social mobility: Reframing the debate,” *Int J Educ Res*, vol. 100, p. 101537, 2020, doi: 10.1016/j.ijer.2020.101537.

[52] J. B. G. Tilak, “Education and Poverty,” *Journal of Human Development*, vol. 3, no. 2, pp. 191–207, Jul. 2002, doi: 10.1080/14649880220147301.

[53] S. N. Cave, M. Wright, and S. von Stumm, “Change and stability in the association of parents’ education with children’s intelligence,” *Intelligence*, vol. 90, p. 101597, Jan. 2022, doi: 10.1016/J.INTELL.2021.101597.

[54] T. Wendling Gonçalves de Oliveira, R. Schmitz, J. Danilo Zea Camaño, A. Paula Dalla Corte, and C. Roberto Sanquetta, “Behavior of the diametric distribution of ecological groups in a mixed ombrophilous forest fragment,” *Floresta*, no. 3, pp. 413–424, 2018, doi: 10.5380/rf.v48i3.

[55] D. T. Utari, *Analisis regresi terapan dengan R*, 2nd ed. Yogyakarta: Universitas Islam Indonesia, 2023.