JURNAL STATISTIKA DAN APLIKASINYA

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

DETERMINANTS OF POVERTY IN EAST JAVA: A COMPARATIVE APPROACH USING SPATIAL AND MACHINE LEARNING MODELS

Ammar Hanafi [1*], Bryan Jonathan[2], Renata Shaula Alfino Ritonga[3***], Yekti Widyaningsih[4****]**

[1,2,3,4]Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Indonesia
UI Depok Campus, Depok, 16424, Indonesia
Corresponding author's e-mail: *ammar.hanafi@ui.ac.id **bryanjonathan201@gmail.com
***renata.shaula@ui.ac.id ****yekti@sci.ui.ac.id

ABSTRACT

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

Keywords:
East Java; machine learning; poverty mapping; spatial econometrics; spatial heterogeneity

Poverty in East Java presents a spatially-varied challenge. Traditional econometric models often fail to capture complex non-linear relationships and underlying spatial patterns, leading to suboptimal policy recommendations. This study analyzes the key socio-economic factors of poverty across 38 regions in East Java by comparing spatial regression models with a machine learning approach, aiming to evaluate the trade-off between model interpretability and predictive accuracy. Using 2024 cross-sectional data, OLS, SLM, SEM, and Spatial Random Forest (SRF) models are estimated. After diagnostic tests for spatial autocorrelation, models are evaluated using AIC, BIC, R-squared, RMSE, and a spatial analysis of the models' residuals. Diagnostic tests found global spatial parameters (ρ and λ) to be insignificant, rendering the parsimonious OLS model sufficient for interpretation. However, the SRF model proved unequivocally superior for prediction, achieving a Pseudo R-squared of 0.976 versus the OLS R-squared of 0.75. Mean Years of Schooling was identified as the most important predictor. Notably, the OLS residuals exhibited significant spatial heterogeneity, a pattern not present in the SRF model, indicating unobserved local patterns. The study's limitations include its cross-sectional nature, which precludes temporal analysis, and its focus on a predefined set of variables. The novelty lies in its direct comparative framework, demonstrating the trade-off between the inferential clarity of econometrics and the predictive power of machine learning. It highlights that significant spatial heterogeneity can persist even when global spatial autocorrelation is statistically absent.

How to cite this article:
Author(s), "ARTICLE TITLE", Journal Statistika dan Aplikasinya, vol. x, iss. x, pp. xx – xx, Month, Year

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

1. Introduction

Poverty is a complex socio-economic issue and is still a major challenge in Indonesia, including in East Java Province. Based on data from Badan Pusat Statistik (BPS), the percentage of poor people in East Jawa in March 2024 was recorded at 9.79%, with the number of poor people reaching 3893 million people. Although there was a decrease compared to the previous year, poverty disparities between regions are still significant. For example, Sampang Regency recorded a poverty rate of 20.83% in 2024.

The phenomenon of poverty in East Java can be seen in the gap that exists between the more developed and less developed regions. In more developed areas, poverty rates are generally lower and there is more access to basic services such as education, health, and infrastructure. In contrast, in more isolated areas, poverty rates tend to be higher, which is associated with a lack of access to these services. Poverty is not only influenced by economic factors, but also by social and geographical factors. Inequality in income distribution, education levels, and limited access to health services are some of the important aspects that need to be considered in an effort to reduce poverty. In addition, lower economic mobility in certain regions can exacerbate existing inequalities.

This phenomenon encourages the need for a more in-depth analysis of the causes and distribution of poverty at the district/city level in East Java Province. To analyze the influence of factors that affect the poverty rate, a spatial data analysis approach is needed that considers the interconnectedness between regions. Spatial regression and Random Forest models are two methods that can be used for this purpose. Spatial regression allows the identification of relationships between regions, while Random Forest can handle data complexity well. An analysis that takes into account inter-regional linkages and considers factors that influence poverty will provide a clearer picture of the socio-economic conditions in the province, as well as enable more targeted policy making to reduce disparities between regions and improve people’s welfare.

Research by Ayudia and Putri (2024) utilized the Spatial Autoregressive Model (SAR) to analyze factors influencing poverty in East Java. They found that variables such as income inequality and education significantly affect the poverty rate. Similarly, a study by Maulana and Fauzan (2024) employed the Spatial Error Model (SEM) to identify determinants of poverty in East Java, highlighting the importance of education and economic indicators. These studies underscore the complexity of poverty, influenced by various interconnected factors. Understanding these factors is crucial for developing targeted policies to alleviate poverty. Therefore, this study aims to analyze the determinants of poverty in East Java using spatial regression models, considering factor such as income inequality, education, healthcare access, and infrastructure development.

2. Methods

This section outlines the materials, data, and analytical procedures employed in this study. It begins by describing the dataset and variables, followed by a detailed explanation of the research methodology, which includes the construction of the spatial weights matrix, diagnostic tests for spatial effects, and the specification of the models used for comparative analysis.

2.1. Material and Data

This research utilizes secondary, cross-sectional data for the year 2024, sourced from the official publications of Statistics Indonesia (BPS) for the East Java Province. The geographical scope of the study covers all 38 districts and cities within East Java. The spatial data, comprising the administrative boundary polygons for each district/city, was obtained from a shapefile corresponding to the official BPS administrative divisions.

The variables selected for this study are defined as follows:

- Dependent Variable (Y): Poverty Rate. This is the percentage of the population living below the official poverty line, as defined by BPS. It serves as the primary indicator of economic welfare in each region.
- Independent Variables (X):
  1. Gini Ratio (X1) : An index measuring income or consumption inequality within a population, ranging from 0 (perfect equality) to 1 (perfect inequality). It is included to examine the relationship between inequality and poverty.
  2. Mean Years of Schooling (X2) : The average number of years of formal education completed by the population aged 15 and over. This variable serves as a key proxy for human capital.
  3. Medical Personnel (X3): The total number of registered medical professionals (doctors, specialists, etc.) in a region. It is used as an indicator of access to healthcare services.
  4. Motor Vehicles (X4): The total number of registered motorcycles in a region. This variable is included as a proxy for economic mobility and household-level economic activity, particularly relevant in both urban and rural contexts in Indonesia.

2.2. Analytical Method

This study employs a comparative analytical approach to investigate the factors influencing poverty. The methodology follows a sequential process: (1) preliminary diagnostic testing for spatial effects, (2) estimation of multiple regression and machine learning models, and (3) a comparative assessment to determine the most suitable model for both interpretation and prediction. The primary software used for the analysis includes Python (with the pysal library) and R (with the spdep and spatialRF packages).

2.3. Spatial Weights Matrix Construction

A fundamental step in any spatial analysis is the formal definition of "neighborhood" relationships, which is accomplished by constructing a spatial weights matrix (). For this study, the matrix was constructed using the K-Nearest Neighbors (KNN) method with k=5. In this specification, each region is defined as having exactly five neighbors based on the shortest distances between regional centroids. The matrix is row-standardized, where the sum of the weights for each row equals one.

2.4. Diagnostic Testing for Spatial Effects

To formally test for the presence of spatial patterns, two primary diagnostic tests were conducted.

- Global Spatial Autocorrelation (Moran's I): Moran's I was used to test for the presence of global spatial autocorrelation in both the dependent and independent variables. The Moran's I statistic is calculated as:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

  where n is the number of regions, xi is the value of the variable at location i, x¯ is the mean of the variable, and wij is the spatial weight between locations i and j. A significant positive Moran's I indicates spatial clustering, while a significant negative value indicates spatial dispersion.
- Lagrange Multiplier (LM) Tests: After estimating a baseline OLS model, LM tests were applied to the residuals to diagnose the specific form of spatial dependence. The LM-Lag test assesses whether a Spatial Lag Model (SLM) is appropriate, while the LM-Error test assesses the suitability of a Spatial Error Model (SEM).

2.5. Model Specification

Four distinct models were estimated to compare their explanatory and predictive power.

1. Ordinary Least Squares (OLS): The standard non-spatial linear regression model serves as a benchmark. The model is specified as : $\mathbf{Y}=\mathbf{X}\beta+\epsilon$, where Y is the vector of the dependent variable, X is the matrix of independent variables, β is the vector of coefficients, and ϵ is the vector of error terms.
2. Spatial Lag Model (SLM): This model accounts for spatial dependence by incorporating a spatially lagged dependent variable, suggesting that poverty in one region is influenced by poverty in

neighboring regions (spillover effects). The model is : $\mathbf{Y}=\rho\mathbf{WY}+\mathbf{X}\beta+\epsilon$, where ρ (rho) is the spatial autoregressive coefficient.

3. Spatial Error Model (SEM): This model assumes that the spatial dependence exists in the error term, suggesting that unobserved factors in one region are correlated with those in neighboring regions. The model is specified as : $\mathbf{Y}=\mathbf{X}\beta+\mathbf{u}$, where $\mathbf{u}=\lambda\mathbf{Wu}+\epsilon$ and λ (lambda) is the spatial error coefficient.
4. Spatial Random Forest (SRF): A non-parametric machine learning model, the Random Forest is an ensemble of decision trees capable of capturing complex, non-linear relationships and interactions between variables. The spatialRF package in R was used, which evaluates the model for residual spatial autocorrelation and provides variable importance metrics. This model is included primarily to assess the maximum predictive accuracy achievable and to identify the most influential predictors without assuming a linear functional form.

2.6. Model Comparison and Selection

The models were compared using several criteria. For the regression models (OLS, SLM, SEM), the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used, with lower values indicating a better model fit relative to model complexity. For assessing predictive power, particularly when comparing against the SRF, R-squared (and Pseudo R-squared) and Root Mean Squared Error (RMSE) were utilized. Finally, a visual analysis of the spatial distribution of the models' residuals was performed to check for any remaining spatial patterns.

3. Results

This chapter presents the empirical findings of the analysis conducted to examine the factors influencing the poverty rate across the districts and cities of East Java. The chapter begins with a presentation of descriptive statistics, followed by an essential preliminary analysis of spatial autocorrelation using Moran's I. Subsequently, it details the estimation and comparative analysis of three spatial regression models (OLS, SLM, SEM) and a machine learning model (Spatial Random Forest). The chapter concludes with a detailed interpretation of the selected model and a diagnostic analysis of its residuals to provide a comprehensive discussion of the findings.

3.1. Descriptive Statistics

The study utilizes cross-sectional data from 38 districts and cities within the province of East Java for the year 2024. The dataset was pre-processed, and no missing or invalid values were found, resulting in a final sample size of 38 observations for the analysis. The variables include the poverty rate (Y) as the dependent variable, alongside four independent variables: Gini Ratio (X1), Mean Years of Schooling (X2), number of Medical Personnel (X3), and number of Motor Vehicles (X4). A summary of the descriptive statistics is presented in Table 1.

Table 1. Descriptive Statistics.

Variable	Description	Mean	Std. Dev.	Min.	Max.
Y	Poverty Rate (%)	9.782	4.213	3.06	20.83
X1	Gini Ratio	0.332	0.037	0.233	0.435
X2	Mean Years of Schooling	8.464	1.703	5.08	12.11
X3	Medical Personnel	48.158	34.175	11	228
X4	Motor Vehicles	516191.7	498888	140363	303475.4

Source: BPS East Java (2024), processed by the author.

The data reveals significant variation across the regions. The poverty rate ranges from a low of 3.06% to a high of 20.83%, indicating substantial regional disparity. Similarly, the variables for Medical Personnel (X3) and Motor Vehicles (X4) display extremely high standard deviations, suggesting significant inequality in healthcare access and economic activity proxies across East Java.

3.2. Classical Assumption Diagnostics

To validate the robustness of the Ordinary Least Squares (OLS) model as a baseline for comparison, a series of diagnostic tests were conducted. These tests assess whether the model adheres to the core assumptions of classical linear regression, ensuring the reliability of its coefficient estimates.

3.2.1. Normality of Residuals

The assumption that the model's error terms are normally distributed was evaluated using the Jarque-Bera test. The test yielded a p-value of 0.8954. As this value is substantially greater than the standard alpha level of 0.05, the null hypothesis of normality cannot be rejected. This result provides strong evidence that the residuals are normally distributed, satisfying this critical assumption for valid inference.

3.2.2. Homoscedasticity

The assumption of homoscedasticity, which posits that the variance of the residuals is constant across all observations, was examined using the Breusch-Pagan test. The test produced a p-value of 0.7825. Since this p-value is well above the 0.05 threshold, the null hypothesis of constant variance (homoscedasticity) is not rejected. This indicates that the model does not suffer from issues of heteroskedasticity.

3.2.3. Multicollinearity

Table 2. Variance Inflation Factor (VIF) Analysis.

Variable	VIF
X1	1.64423
X2	1.65923
X3	8.61686
X4	8.75864

Source: Author's calculation using R.

The potential for multicollinearity—high correlation among independent variables—was assessed using the Variance Inflation Factor (VIF). The analysis revealed that the variables for Gini Ratio (X1) and Mean Years of Schooling (X2) have VIF values well below 5, indicating no multicollinearity concerns. While the VIF values for Medical Personnel (X3) and Motor Vehicles (X4) were found to be greater than 5 (approaching 9), they remained below the more critical threshold of 10. This suggests a moderate but tolerable level of multicollinearity that does not critically undermine the stability or interpretation of the model. Therefore, all predictor variables were retained in the analysis.

3.3. Spatial Autocorrelation Analysis

To investigate the presence of spatial dependency wherein the value of a variable in one location is related to the values in neighboring locations, Moran's I test was performed on each variable. The null hypothesis (H_0) posits a random spatial distribution (no spatial autocorrelation), while the alternative hypothesis (H_1) suggests the presence of a spatial pattern. The results are summarized in Table 2.

Table 3. Moran's I Test for Spatial Autocorrelation.

Variable	Description	Moran's I	p-value	Interpretation
Y	Poverty Rate (%)	0.1998	0.0385	Significant
X1	Gini Ratio	0.1817	0.052	Almost Significant
X2	Mean Years of Schooling	0.2857	0.0047	Significant
X3	Medical Personnel	0.1337	0.0428	Significant
X4	Motor Vehicles	0.1984	0.0066	Significant

Source: Author's calculation using R.

The analysis reveals that the dependent variable, Poverty Rate (Y), and three of the four independent variables (X2, X3, X4) exhibit statistically significant spatial autocorrelation ($p < 0.05$). The positive Moran's I statistics indicate a pattern of positive spatial clustering, where districts with high poverty rates tend to be geographically adjacent to other high-poverty districts, and vice-versa. This finding violates a key assumption of classical linear regression (independence of observations) and strongly motivates the use of spatial regression models.

3.4. Model Estimation and Comparison

Based on the evidence of spatial effects, four models were estimated: Ordinary Least Squares (OLS), Spatial Lag Model (SLM), Spatial Error Model (SEM), and Spatial Random Forest (SRF).

3.4.1. Regression Model Results

The initial OLS model was estimated as a benchmark, followed by the SLM and SEM to account for spatial dependencies. Table 3 presents the coefficient estimates for the OLS model, which will be used for final interpretation, while Table 4 summarizes the key parameters and fit statistics for all three regression models.

Table 4. Final Interpreted Model: OLS Estimation Results.

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	32.6307	3.316	9.8404	0.0000
X1	-31.6307	12.6519	-2.5001	0.0176
X2	-1.5083	0.2770	-5.4447	0.0000
X3	0.0693	0.0310	2.2377	0.0321
X4	0.00001	0.0000	-2.6070	0.0136

Source: Author's calculation using Python.

Diagnostic tests on the OLS residuals, including the Lagrange Multiplier (LM) tests for spatial lag and spatial error, yielded insignificant results. This provided an early indication that the spatial effects might be fully accounted for by the independent variables, making OLS a potentially sufficient model.

Table 5. Comparative Statistics for Regression Models.

Model	R-squared	Log-Likelihood	AIC	BIC	Spatial Par.	p-value
OLS	32.6307	-81.6840	173.368	181.556	-	-
SLM	-31.6307	-81.0053	175.06	184.886	$\rho = -0.0780$	0.5448
SEM	-1.5083	-81.6571	173.314	181.502	$\lambda = -0.1030$	0.6463

Source: Author's calculation using Python.

As shown in Table 4, the spatial parameters for both SLM (Rho, ρ) and SEM (Lambda, λ) are statistically insignificant. While the SEM has the lowest AIC and BIC values, the difference compared to OLS is negligible. Given the insignificance of the spatial parameters and the principle of parsimony, the OLS model is selected as the most appropriate model for interpretation.

3.4.2. Spatial Random Forest (SRF) Model Results

For predictive purposes, an SRF model was estimated. The model demonstrated superior predictive performance, as detailed in Table 6.

Table 6. Spatial Random Forest Performance Metrics.

Metric	Value	Interpretation
R-squared (OOB)	0.6534	65.3% of variance explained by out-of-bag samples.
Pseudo R-squared	0.976	Superior model fit compared to regression models.
RMSE	1.116	Low prediction error relative to the mean poverty rate.

Source: Author's calculation using R.

The variable importance analysis from the SRF model, shown in Table 7, identifies Mean Years of Schooling (X2) as the most critical predictor of poverty, followed by the Gini Ratio (X1).

Table 7. Variable Importance

Variable	Importance
X2	3.762
X1	1.448

X4	1.370
X3	0.794

Source: Author's calculation using R.

Critically, diagnostic tests on the SRF residuals found them to be normally distributed (Shapiro-Wilk p = 0.510) and, most importantly, free of any significant spatial autocorrelation. This indicates the SRF model successfully captured the complex relationships in the data, leaving behind only random, non-spatial noise. These results reinforce the conclusion that while the regression models offer clear interpretability, the SRF is superior for accurate prediction of poverty rates across East Java.

3.5. Residual Analysis of the OLS Model

Although the OLS model was selected for interpretation, a visual analysis of its residuals is essential to understand its limitations. Figure 2 presents the spatial distribution of the OLS model's residuals.

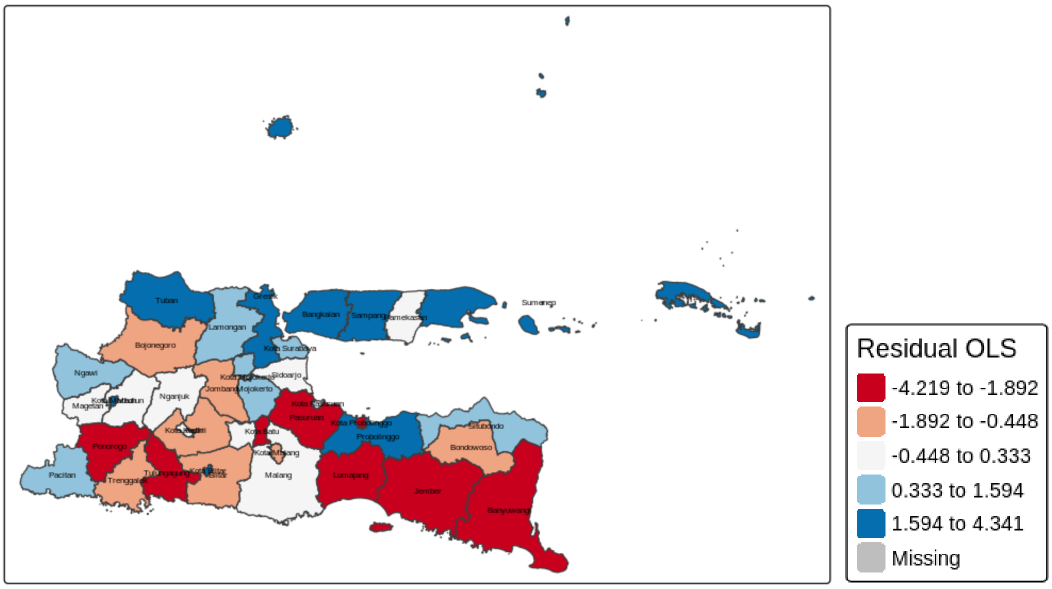

Figure 2. Spatial Distribution of OLS Model Residuals

The residual map reveals a clear spatial pattern that was not captured by the global statistics. A distinct cluster of negative residuals (blue areas), where the model over-predicted poverty, is evident in the southern and eastern regions of the province. Conversely, clusters of positive residuals (red areas), where the model under-predicted poverty, are found in the central and northern areas.

This visual evidence of spatial heterogeneity is a key finding. It suggests that while a global spatial effect was not significant, there are localized, unobserved factors influencing poverty that vary systematically between the northern and southern regions of East Java. This could be related to cultural differences, distinct economic structures, or specific regional policies, presenting a clear direction for future research.

3.6. Residual Analysis of the Spatial Random Forest Model

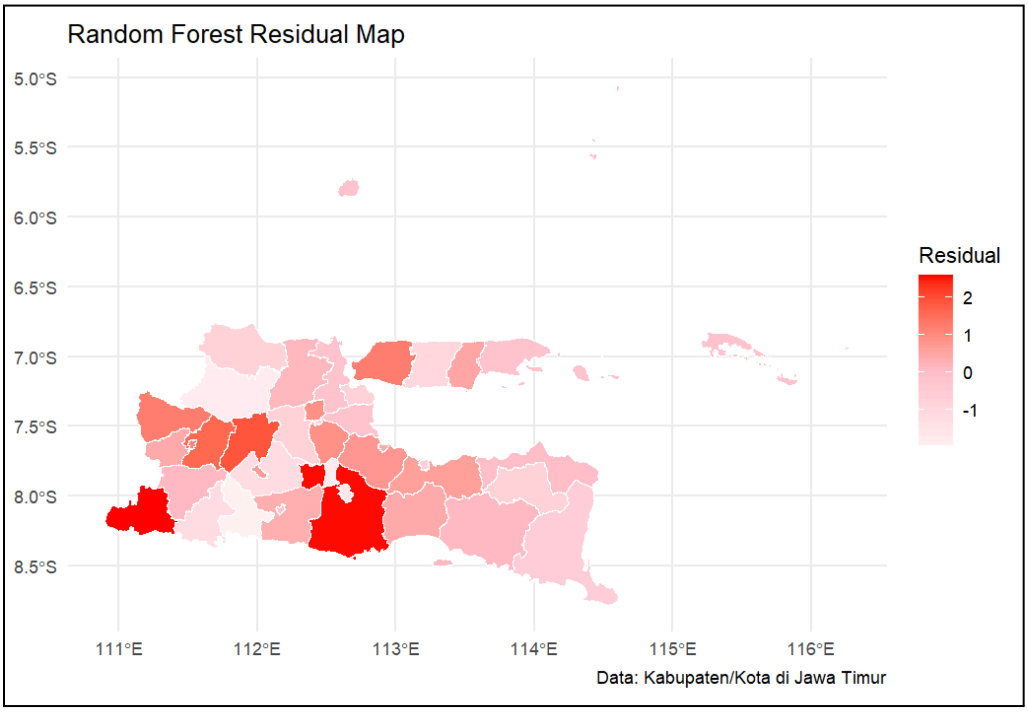

Figure 3. Spatial Distribution of Random Forest Model Residuals

Figure 3 illustrates the spatial distribution of the residuals generated by the Spatial Random Forest (SRF) model. In this map, orange and red colors represent areas with positive residuals where the model under-predicts the poverty rate compared to the observed data. In contrast, white and pink colors represent negative residuals, indicating areas where the model over-predicts the poverty rate. The residuals appear to be randomly distributed across the province without any visible spatial clustering. This visual evidence supports the results of the Moran's I test for the residuals, which produced small negative values and p-values well above 0.26. This indicates that there is no significant spatial autocorrelation in the residuals of the SRF model.

This random pattern of residuals suggests that the Spatial Random Forest model has successfully captured the complex spatial variability of poverty rates across East Java, leaving behind only random noise. Consequently, the SRF model stands out as the most robust approach for predictive modeling in this study, outperforming the traditional spatial regression models (OLS, SAR, and SEM). While the SRF does not provide explicit coefficients for interpretability, its ability to deliver high predictive accuracy without residual spatial bias makes it a valuable tool for policy-relevant forecasts of poverty distribution.

4. Discussions

The analysis leads to a nuanced conclusion. For explaining the directional relationships between socio-economic factors and poverty, the OLS model is the most suitable. The interpretation is as follows:

1. Mean Years of Schooling (RLS) is the most impactful factor, showing a significant negative relationship with poverty (coef. -1.51). This underscores the critical role of education in poverty alleviation in East Java.
2. Gini Ratio also has a significant negative effect (coef. -31.63). This counter-intuitive result may reflect that in areas with higher inequality, the economic growth that drives this inequality may still have poverty-reducing effects, though this relationship requires further investigation.
3. Medical Personnel shows a significant but small positive effect (coef. 0.07). This is likely not causal but rather indicates that healthcare resources are often targeted and allocated to regions with pre-existing higher poverty levels.
4. Motor Vehicles, as a proxy for economic mobility, shows a significant but very small negative effect on poverty.

For the purpose of prediction, the Spatial Random Forest model is unequivocally superior, explaining over 97% of the variance in the poverty rate. Its ability to model non-linear relationships and interactions without leaving residual spatial patterns makes it a powerful predictive tool.

5. Conclusion

This study was conducted to analyze the primary determinants of poverty in East Java and to compare the performance of parametric spatial regression models against a non-parametric machine learning approach. The analysis confirms that socio-economic factors, particularly education, significantly influence poverty rates, and it demonstrates a clear trade-off between model interpretability and predictive power. The findings directly address the research objectives, concluding that while a standard OLS regression model is sufficient for interpreting directional relationships, the Spatial Random Forest (SRF) model offers vastly superior predictive accuracy.

Answering the first research objective regarding the influence of independent variables, the final OLS model revealed that **Mean Years of Schooling** is a critical determinant, showing a strong, statistically significant negative relationship with the poverty rate. This was further corroborated by the SRF model, which identified it as the single most important predictor. Other significant factors included the **Gini Ratio** and the **Number of Motor Vehicles**, both of which were associated with lower poverty rates. Conversely, the number of **Medical Personnel** showed a slight positive correlation, which is interpreted not as a causal factor but likely as an indicator of targeted healthcare resource allocation to poorer regions.

In addressing the second objective, the comparative analysis demonstrated that the inclusion of spatial parameters in the SLM and SEM did not yield statistically significant results ($p - value >$ 0.05 for both ρ and λ). Given the negligible difference in fit statistics (e.g., AIC/BIC) between the OLS and SEM models, the more parsimonious **OLS model was deemed most appropriate for explanatory purposes**. However, for the objective of prediction, the **Spatial Random Forest model proved to be unequivocally superior**. The SRF model achieved a pseudo R-squared of 0.976, far surpassing the OLS model's R-squared of 0.75, and its residuals showed no remaining spatial autocorrelation, unlike the OLS residuals which displayed clear spatial heterogeneity.

Based on these findings, this study offers two primary implications. For policymakers, the results strongly suggest that **investments in education should be a cornerstone of poverty alleviation strategies** in East Java, as improving the Mean Years of Schooling has the most substantial impact. For future research, this study highlights the value of a dual-modeling approach. While OLS or SEM can provide valuable, interpretable insights into policy levers, machine learning models like SRF are essential for achieving high-accuracy predictions. Future work should aim to explore the observed spatial heterogeneity in the OLS residuals, potentially by incorporating additional variables related to local economic structure or cultural factors to build an even more comprehensive understanding of the spatial dynamics of poverty.

6. Acknowledgments

The Directorate of Research and Development of Universitas Indonesia (DRPM UI) funded this study as an additional paper through a grant of Hibah Publikasi Terindeks Internasional (PUTI) Q2 2022-2023 No.: NKB-668/UN2.RST/HKP.05.00/2022..

7. References

[1] A.A. Maulana, A. Fauzan. "Spatially informed insights: Modeling percentage poverty in East Java Province using SEM with spatial weight variations," *BAREKENG: Journal of Mathematics and Its Applications*, vol. 18, no. 6, pp. 1317-1332, June 2024.

[2] Badan Pusat Statistik (BPS), [Online]. Available: https://www.bps.go.id/id [Accessed: 26 May 2025].

[3] Badan Pusat Statistik (BPS), "Persentase Penduduk Miskin di Jawa Timur September 2024 turun menjadi 9,56 persen", [Online]. Available: https://jatim.bps.go.id/id/pressrelease/2025/01/15/1474/persentase-penduduk-miskin-di-jawa-timur-september-2024-turun-menjadi-9-56-persen.html [Accessed: 5 June 2025].

[4] E.W. Fox, J.M.V. Hoef. "Comparing spatial regression to random forests for large environmental data sets," *PLoS ONE*, vol. 15, no. 3, March 2020.

[5] IBM, "Apa itu random forest?", [Online]. Available: https://www.ibm.com/id-id/think/topics/random-forest [Accessed: 7 June 2025].

[6] L. Anselin, *Spatial econometrics: methods and models*. New York: McGraw Hill, 1953.

[7] N.U. Ayudia, D.E. Putri. "Pemodelan Spatial Autoregressive Model (SAR) pada kasus kemiskinan di Jawa Timur," *Jurnal Gaussian*, vol. 13, no. 2, pp. 308-1318, November 2024.

[8] R. Hasanah, Syaparuddin, Rosmeli. "Pengaruh angka harapan hidup, rata-rata lama sekolah dan pengeluaran perkapita terhadap tingkat kemiskinan pada Kabupaten /Kota di Provinsi Jambi," *E-Jurnal Perspektif Ekonomi Dan Pembangunan Daerah*, vol. 10, no. 3, pp. 223-232, September 2021.