



Application of the Spatial Autoregressive (SAR) Method in Analyzing Poverty in Indonesia and the Self Organizing Map (SOM) Method in Grouping Provinces Based on Factors Affecting Poverty

Ulimazzada Islamy^{a,1,*}, Afdelia Novianti^{a,2}, Freditasari Purwa Hidayat^{a,3}, Muhammad Hasan Sidiq Kurniawan^{a,4}

^aDepartement of Statistics, Universitas Islam Indonesia, Jl.Kaliurang km 14,5, Yogyakarta 55584, Indonesia

¹18611128@students.uui.ac.id*; ²18611082@students.uui.ac.id; ³18611027@students.uui.ac.id ⁴hasan.sidiq@uui.ac.id

* corresponding author

ARTICLE INFO

ABSTRACT

Article history

Received September 3, 2021

Revised September 21, 2021

Accepted November 4, 2021

Keywords

Poverty

Self-Organizing Map (SOM)

Spatial Autoregressive (SAR)

The economy is a benchmark to determine the extent of the development of a country. Indonesia, which is now a developing country, is ranked 5th as the poorest country in Southeast Asia. Certainly, the government must pay attention because poverty until now has become one of Indonesia's main problems. Ending poverty everywhere in all its forms is goal 01 of the Sustainable Development Goals (SDGs) program. One of the efforts that can be done is by planning as part of the implementation of the target, namely eliminating poverty and providing appropriate social protection for all levels of society to achieve the SDGs. Therefore, it is deemed important to do a spatial analysis by making a model of poverty estimation in Indonesia and grouping to identify areas in Indonesia that have the highest poverty mission. The clustering method used in this grouping is Self-Organizing Map (SOM). In this study, Spatial Autoregressive (SAR) analysis was used to create a predictive model. This is because poverty is very likely to have a spatial influence or be influenced by location to other areas in the vicinity. Furthermore, the region with the highest mission was grouped using the Self Organizing Map (SOM) clustering based on variables that significantly affect the rate of poverty in Indonesia. From the results of the analysis, four clusters were obtained, each of which had its characteristics to classify 34 provinces in Indonesia. The clusters formed included cluster 1 consisting of 17 provinces, cluster 2 consisting of 9 provinces, cluster 3 consisting of 1 province, and cluster 4 consisting of 7 provinces.

1. Introduction

Poverty can be caused by the scarcity of basic needs or the difficulty of accessing education and employment. It is defined as the deprivation of welfare. Poverty is a complex problem and has always been a scourge for developing countries, like Indonesia, where this country is ranked as the 5th poorest country in Southeast Asia.

Poverty is one of the fundamental problems that are complex and multidimensional so that it becomes the concern for the government in any country. It affects various aspects such as social aspects, economic aspects, cultural aspects, and other aspects. In consideration to high economic growth as one of the indicators of development success [1], efforts are needed on how to improve the economy as a way to reduce the poverty rate in Indonesia.

The main problem in efforts to reduce poverty in Indonesia today is related to the uneven distribution of economic growth throughout Indonesia; this is evidenced by the high-income disparity between regions. In addition, poverty is also a causal relationship (circular causality) meaning that the high poverty rate occurs due to low per capita income, and low per capita income occurs due to low per capita investment [2].

The first of the 17 SDGs goals is "End Poverty in All Forms Everywhere". The main objective must be the theme of development, the main and sustainable agenda underlying various other development goals such as infrastructure, tourism, food and energy, and others [3]. Research conducted by Tri Santoso stated that poverty alleviation has links with other goals proclaimed by the SDGs. Further goals such as a world without hunger, health, prosperity, quality education, gender equality, clean water and sanitation, clean and affordable energy until poverty can at least be reduced. Talking about poverty is a complex and global problem Indonesia is still facing and the Government of Indonesia is still trying to deal with [4].

The research that became the reference entitled "Analysis of Poverty in Central Java Province Using the Spatial Regression Method" by [5]. The results showed that population working in the agricultural sector (X2) was the only variable that was significant or affected in the multiple linear regression model. For the SAR and SEM models, the significant variables on poverty are the community participation level (X1) and population working in the agricultural sector (X2).

The other research was conducted by [6] with the title "Grouping Regencies Based on Poverty Characteristics in East Nusa Tenggara Province Using the Self Organizing Maps (SOM) Algorithm" which aimed to determine the grouping of mikes owned by each Regency in East Nusa Tenggara Province to see the characteristics of poverty in each group formed in each cluster. Using Self Organizing Maps (SOM) method, the results of this study showed 6 clusters, each of which had its own characteristics. The clusters formed included cluster 1 consisting of 1 regency, cluster 2 consisting of 5 regencies, cluster 3 consisting of 2 regencies, cluster 4 consisting of 6 regencies, cluster 5 consisting of 5 regencies and cluster 6 consisting of 2 districts.

2. Method

2.1. Data and Data Sources

This study used secondary data obtained from the website of the Central National Statistics Agency (bps.go.id) with the variables being the Total Poverty in Indonesia (Y), Human Development Index (X1), Life Expectancy (X2), Open Unemployment Rate (X3), Population Working in the Agricultural Sector (X4), Literacy Rate (X5), and Average Length of Schooling (X6).

2.2. Sustainable Development Goals (SDGs)

Until the end of the 20th century, there are still many problems that burden the world started from poverty, healthy living, depleted energy, gender equality, to environmental pollution. The world responded by agreeing on a declaration known as the Millennium Development Goals (MDGs). With the end of the MDGs in 2015, a concept was developed in the context of the post-2015 development agenda called the Sustainable Development Goals (SDGs).

The SDGs, which were agreed upon by world leaders at the United Nations session, contain 17 aspects of sustainable development goals, which contain 169 formulated targets. The concept of SDGs is needed as an agenda in the new development accommodating all the changes occurred in post-2015. The main goal is to obtain a universal common goal that is able to maintain a balance of the three dimensions of sustainable development, environmental, social and economic.

2.4. Descriptive Statistics

Descriptive statistics is a method related to the collection and presentation of data so as to provide useful information [7]. It serves to describe or provide an overview of the object under study without analyzing and making conclusions that apply to the public [8]. The data presented in descriptive statistics are usually in the form of a data concentration measure [9]. Apart from measuring the concentration of data, it can also be presented in the form of Pareto diagrams and tables.

2.5. Spatial Autoregressive (SAR)

Spatial Autoregressive Model is a model that combines a simple regression model with spatial lag on the dependent variable using cross section data. Cross section data refer to data collected at certain times and places, generally reflecting a phenomenon in one period of time only [10]. Spatial lag occurs when the observation value of the dependent variable at a location is correlated with the observed value of the dependent variable in the surrounding location. According to [11] the SAR model can be stated as follows:

$$Y_i = \rho \sum_{j=1}^n w_{ij} Y_j + \sum_{k=1}^n \beta_{ik} X_{ik} + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, \sigma^2) \quad (1)$$

with:

- Y_i = response variable at location i
- X_{ik} = predictor variable at location i
- w_{ij} = element of the spatial weight matrix W in the i to j
- ε_i = error at location i which is normally distributed and independent with zero mean and variance 2
- ρ = parameter of spatial lag coefficient of response variable
- β = regression coefficient parameter

2.6. Self-Organizing Map (SOM)

Self-Organizing Map (SOM), as its name implies, is a network that does not require any special supervision. The word map means that this method uses maps in the weighting of input data [12]. SOM is used as a visualization, and analysis tool for high-dimensional data, clustering, classification, and data mining. It is considered as a spatial form of K-Means group analysis.

2.7. Research Stage

The stages of research carried out in this study are as follows:

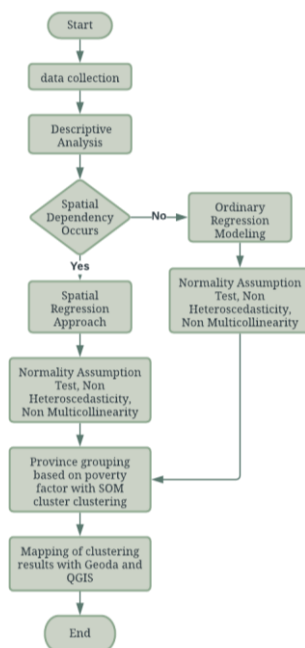


Fig. 1. Research stage.

Having collected the data, the descriptive statistics was analyzed. Then, a spatial dependency test was carried out, if there was no spatial dependencies, to find out the variables that had a significant effect. Ordinary regression analysis test was used followed by the assumption test, but if there was a spatial dependencies the spatial regression analysis would be carried out followed by an assumption test. After seeing what variables had an effect, the variables were grouped using the SOM method and then a map was made as a visualization.

3. Results and Discussion

3.1. Descriptive Analysis

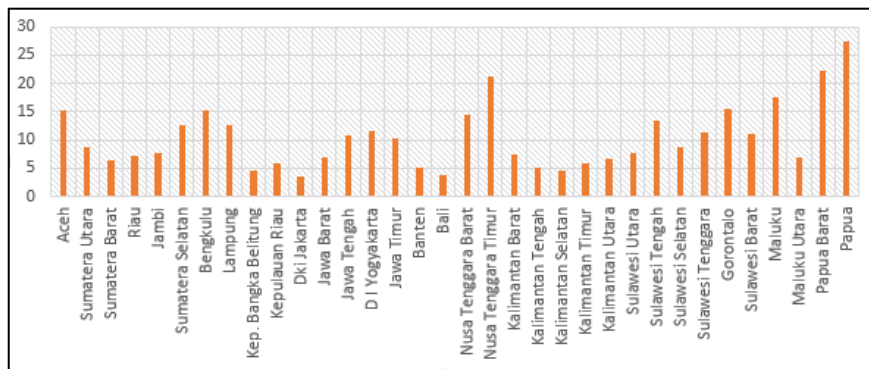


Fig. 2. Poverty percentage in 2019.

As shown in Fig. 2, Papua Province became the province with the highest poverty at 27.53%. It was then followed by West Papua Province at 22.17%. Meanwhile, DKI Jakarta became the province with the lowest poverty at 3.47%.

3.2. Spatial Weighting

Table 1. Spatial Weighting

Group	Number of Neighbor	Province	Total	Wight
1	0	Bali, Bangka-Belitung, Riau Islands, Maluku, North Maluku, West Nusa Tenggara, East Nusa Tenggara	7	0
2	1	Aceh, Irian Jaya Barat, East Java, Papua, North Kalimantan, Papua, North Sulawesi, Yogyakarta	7	1
3	2	Banten, Gorontalo, Jakarta Raya, West Kalimantan, South Kalimantan, Lampung, West Sulawesi, Southeast Sulawesi	8	$\frac{1}{2}$
4	3	West Java, Central Java, Central Kalimantan, Riau, South Sulawesi, South Sumatra, North Sumatra	7	$\frac{1}{3}$
5	4	Bengkulu, Jambi, East Kalimantan, Central Sulawesi, West Sumatra	5	$\frac{1}{4}$

The formation of the SAR model requires weighting so that it is necessary to analyze the neighbors of each province. In this study, the queen contiguity method was used, namely the adjacency seen based on the contact of the angles and sides for the weighting of the matrix. Then the matrix was converted into a standardized spatial weighting matrix.

3.3. Lagrange Multiplier Test

Lagrange multiplier test is used to see whether there is a spatial dependencies. Then, it will be known also as the type of spatial dependencies that can be used.

Table 2. Lagrange Multiplier Test

Method	p-value	Sign	Decision	Conclusion
LMerr	0.62645	>	Failed to reject H ₀	No spatial dependencies (SEM)
LMLag	0.01451	<	Reject H ₀	There is a spatial dependencies (SAR)
RLMerr	0.34605	>	Failed to reject H ₀	No spatial dependencies (Robust SEM)
RLMLag	0.01005	<	Reject H ₀	There is a spatial dependencies (Robust SAR)
SARMA	0.03234	<	Reject H ₀	There is a spatial dependencies (SARMA)

Because H₀ was rejected, it can be concluded that there was a spatial dependency so that it would proceed to make a spatial autoregressive (SAR) model.

3.4. Identification Factors and SAR Model

Table 3. Identification Factors

No	Coefficient	P-value	Sign	Decision
1	W_Poverty	0.00188	<	Reject H ₀
2	β_0 (Intercept)	0.00013	<	Reject H ₀
3	β_1 (X5)	0.02429	<	Reject H ₀
4	β_2 (X2)	0.00171	<	Reject H ₀
5	β_3 (X4)	0.02247	<	Reject H ₀

Based on the results of the comparison of test statistics with a value of $\alpha = 0.05$, it can be concluded that the variables W_Poverty, Intercept, X5, X2 and X4 had a decision to reject H₀. In other words, all of these variables were the significant variables in the SAR model.

Based on the results of the test of factors that affected poverty in the SAR model above, the following model can be formed:

$$\hat{y}_i = 89.9421 + 0.272319 \sum_{j=1, i \neq j}^n W_{ij} Y_j - 0.378856 X5 - 0.792074 X2 + 0.12717 X4$$

3.5. Assumption in SAR Model

In the SAR model, the assumption test that must be met is that the residuals must be normally distributed, no heteroscedasticity, and no multicollinearity.

Table 4. Assumption Test

No	Assumption	P-value	Sign	Decision
1	Normality test	0.4191	>	Failed to reject H ₀
2	Heteroscedasticity	0.1279	>	Failed to reject H ₀
3	Multicollinearity	1.2636 (X5) 1.2686 (X2) 1.2627 (X4)	<	Failed to reject H ₀

Based on this value, because the p-value for normality and heteroscedasticity test is above 0.05, the assumption is fulfilled, while for the multicollinearity test the assumption is fulfilled because the three variables have a value below 10.

3.6. Determination of the Number of Cluster

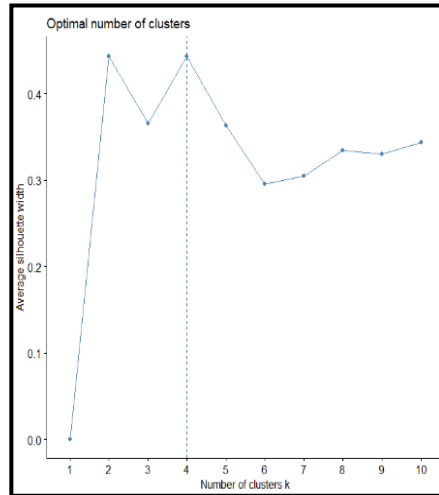


Fig. 3. Silhouette plot.

The plot of the silhouette method shows that the optimal number of clusters is 4, so that 4 clusters would be formed in the Papua province poverty grouping.

3.7. Result and Cluster's Profiling

Table 5. Cluster Member

Cluster	Total	Members
1	17	North Sumatra, West Sumatra, South Sumatra, Lampung, Riau Islands, West Java, Central Java, Yogyakarta, East Java, Banten, Bali, West Kalimantan, South Kalimantan, North Sulawesi, South Sulawesi, Southeast Sulawesi, North Maluku.
2	9	Aceh, Bengkulu, West Nusa Tenggara, East Nusa Tenggara, Central Sulawesi, Gorontalo, West Sulawesi, Maluku, West Papua.
3	1	Papua
4	7	Riau, Jambi, Bangka Belitung Islands, Jakarta, Central Kalimantan, East Kalimantan, North Kalimantan

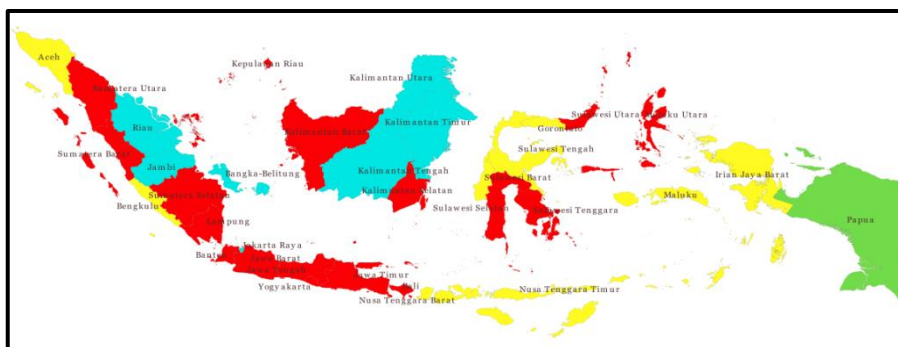


Fig. 4. Grouping visualization.

Table 4. Assumption Test

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Poverty (Y)	8.32588	16.2311	27.53	5.76
X5	70.8385	67.2394	65.69	71.7686
X2	87.7212	90.8933	98.81	67.5743
X4	98.6318	97.8256	78.1	99.5171
Category	Moderate	High	Very High	Low

As seen in **Fig. 4** and **Table 4**, the provinces included in cluster 1 were the areas with moderate poverty rate. Cluster 2 consisted of 9 provinces as the areas with high poverty rate. Cluster 3 consisted of 1 province as an area with a very high poverty rate. Meanwhile, Cluster 4 consisted of 7 provinces as the areas with low poverty rate.

4. Conclusion

Based on the results of the analysis carried out, the following conclusions are obtained:

1. An overview of poverty in Indonesia in 2019 showed that the province with the highest poverty was Papua Province at 27.53%. While, DKI Jakarta was the province with the lowest poverty at 3.47%.
2. There are 3 variables significantly affect poverty in Indonesia, i.e. AMH, AHH, and PSP variables with the SAR model as follows:

$$\hat{y}_i = 89.9421 + 0.272319 \sum_{j=1, i \neq j}^n W_{ij} Y_j - 0.378856 AMH - 0.792074 AHH + 0.12717 PSP$$

3. The number of clusters formed from the results of the analysis of the SOM method obtained 4 clusters. Cluster 1 (red) consisted of 17 provinces, cluster 2 (yellow) consisted of 9 provinces, cluster 3 (green) consisted of 1 province, and cluster 4 (blue) consisted of 7 provinces.
4. The characteristics contained in each cluster showed that in clusters 1 and 2 there were no prominent characteristics and the percentage of poverty was almost the same as members of one cluster and was also relatively low from other clusters. Cluster 3 showed the highest percentage of poverty with the lowest life expectancy and literacy rate but with the highest value of the population working in the agricultural sector. Meanwhile, in cluster 4, the variable population working in the agricultural sector showed the lowest percentage of poverty and it had the highest life expectancy and literacy rate from other clusters.

References

- [1] Rustam, "Perencanaan Pertumbuhan Ekonomi Jawa Timur Dalam Rangka Mengurangi Angka Pengangguran Dan Kemiskinan," p. 6(1), 2010.
- [2] N. d. Baiq, "Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Kemiskinan di Indonesia," *media.neliti*, p. 34, 2012.
- [3] bappenas, "Tujuan Pembangunan Berkelanjutan SDGs," [Online]. Available: <http://sdgs.bappenas.go.id/tujuan-1/>.
- [4] I. d. S. T. Raharjo, "Sustainable Development Goals (SDGs) dan Pengentasan Kemiskinan," <http://jurnal.unpad.ac.id/share/article/view/13198/6032>, p. 159.
- [5] N. I. Septiana, "Analisis Kemiskinan di Provinsi Jawa Tengah Menggunakan Metode Regresi Spasial," *Dspace*, 17 04 2015.

- [6] F. Lazumi, "Pengelompokan Kabupaten/Kota Berdasarkan Karakteristik Kemiskinan di Provinsi Nusa Tenggara Timur Menggunakan Algoritma Self Organizing Maps (SOM)," 25 01 2018. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/5621>.
- [7] R. E. Walpole, Pengantar Statistika Edisi Ke-3, Jakarta: PT. Gramedia Pustaka Utama, 1995.
- [8] Sugiyono, Metode Penelitian Kuantitatif Kualitatif dan R&D, Bandung: Alfabeta, 2007.
- [9] D. Kuswanto, Statistik Untuk Pemula dan Orang Awam, Jakarta: Laskar Aksara, 2012.
- [10] E. Sugiarto and Kusmayadi, Metodologi Penelitian dalam Bidang Kepariwisata, Jakarta: Gramedia Pustaka, 2000.
- [11] L. Anselin, Spatial Econometrics: Methods and Models, Dordrecht: Academic Publishers, 1999.
- [12] Guthikonda, "ohonen Self Organizing Maps," 2005. [Online]. Available: <http://www.shy.am/wp-content/uploads/2009/01/kohonen-self-organizing-maps-shyam-guthikonda.pdf>.