SPECTRUM
Journal of Social Sciences

Vol. 02, No. 02 (2025) 71-80, doi: 10.61552/SJSS.2025.02.002 - http://spectrum.aspur.rs

K-MEANS ALGORITHM AND SELF ORGANIZING MAPS (SOM) FOR POVERTY CLUSTERS IN MALUKU, NORTH MALUKU, PAPUA AND WEST PAPUA PROVINCES

Gabriella Haumahu
Muhammad Yahya Matdoan[1]

Received 11.07.2024.
Revised 19.08.2024.
Accepted 15.09.2024.

Keywords:

K-Means, Poverty, Self Organizing Maps (SOM).

Original research

A B S T R A C T

Based on data published by the Central Statistics Agency (BPS), it shows that provinces that are classified as poor are provinces in eastern Indonesia, especially Maluku, North Maluku, Papua and West Papua. In addition, based on the Presidential Decree, the provinces of Papua, West Papua, Maluku and North Maluku are also provinces that contribute most of the regencies/cities that are classified as the poorest regions. One of the classification or clustering methods is K-Means and Self Organizing Maps (SOM). SOM is an artificial neural network method that is used to group (clustering) data based on data characteristics/features. The technique used in the SOM method is done by creating a network that stores information in the form of node connections with the specified training set. The k-means clustering algorithm can be summarized by selecting the closest distance to the center and then calculating the new center based on the clustering results. This is done until there is no change in cluster members. The data used in this study is secondary data obtained from BPS publications with district/city sample units in the Provinces of Maluku, North Maluku, Papua and West Papua with a total of 63 districts/cities. This study obtained the results that by using the K-Means Cluster, 3 clusters were formed with details of cluster 1 consisting of 33 regencies/cities, Cluster 2 consisting of 15 regencies/cities and cluster 3 consisting of 15 regencies/cities. Furthermore, by using Self Orzanizing Maps (SOM) 4 clusters were formed with details of cluster 1 consisting of 45 districts/cities, Cluster 2 consisting of 2 districts/cities and cluster 2 consisting of 14 districts/cities.

1. INTRODUCTION

The problem of poverty is not just the number and percentage of poor people. Another dimension to consider is the depth and severity of poverty (Alkire & Santos, 2013). Policies for overcoming poverty by the government, apart from being able to reduce the number of poor people, must also be able to reduce the depth and severity of poverty (Moser, 1998).

Based on data released by the Central Statistics Agency (BPS Provinsi Maluku, 2022) for 2022, it shows that the percentage of poor people is mostly from Eastern Indonesia. The poverty percentage rate in Maluku Province in 2021 is 16.30%, Papua Province is 27.38%, and West Papua Province is 21.82% (BPS, 2022).

[1] Corresponding author: Muhammad Yahya Matdoan
Email: keepyahya@gmail.com

The data released by the Central Statistics Agency (BPS) does not provide detailed information about the poverty enclaves where the poor live. To conduct a survey on this matter is also difficult, because of various constraints such as cost and time. Efforts that make it possible to do this are grouping regions based on poverty indicators so that it can be determined which areas have a high number of poor people. This of course makes it easier to monitor the success of the poverty eradication program from year to year, by observing the ups and downs of the number of poor people in districts/cities.

The results of grouping regions based on poverty indicators will be used to determine whether these areas can be categorized as underdeveloped areas or not, and whether these areas can be categorized as poor areas or not. The areas that have been determined based on relative poverty indicators reflect the actual situation, as poverty enclaves where the poor live.

One of the classification or clustering methods is Self Organizing Maps (SOM). SOM is an artificial neural network method that is used to group (clustering) data based on data characteristics/features. According to (Harli et al., 2016; Kohonen, 1995; Kusumah et al., 2017) SOM is a topological form of Unsupervised Artificial Neural Network (Unsupervised ANN) where the training process does not require supervision (target output). This method makes it possible to describe multidimensional data into smaller dimensions, usually one or two dimensions. This simplification process is carried out by reducing the vector that connects each node. This method is also known as Vector Quantization. The technique used in the SOM method is carried out by creating a network that stores information in the form of node connections with the specified training set.

Several studies using SOM in grouping include (Vesanto & Alhoniemi, 2000) Clustering of the Self–Organizing Map, (Tasdemir & Merényi, 2009) Principal Component Analysis and Self Organizing Map for Visualizing and Classifying Fire Risk in Forest Regions, (Pandit et al., 2011) Classification of Indian power Coals Using K-means Clustering and Self Organizing Map neural network.

According to (Tasdemir & Merényi, 2009) Self Organizing Maps (SOM) have deficiencies in visualizing data cluster structures, namely SOM can only provide relative relationships between data points, this results in visualizing cluster results having difficulty in determining boundaries between clusters. To overcome the lack of visualization of SOM grouping, a grouping method is needed that uses the concept of distance and has almost the same characteristics as SOM. Therefore in this study the k-means method is used, where the k-means uses a similarity measure to group objects. This similarity can be translated into the concept of distance. Two objects are said to be similar if the distance between the two objects is close. The higher the distance value, the higher the dissimilarity value.

According to (Hendayanti et al., 2018; Kohar, 2021) the k-means clustering algorithm can be summarized by selecting the closest distance to the center followed by calculating the new center based on the grouping results. This is done until there is no change in cluster members. Several studies using the k-means method include (Niu et al., 2012) A Clustering Method Based on K-Means Algorithm, (Gera et al., 2021) Fault tolerant decentralized K-Means clustering for asynchronous large-scale networks, Speeding up k-means algorithm by GPUs.

2. LITERATURE REVIEW

2.1 K-Means Clustering

K-Means is one of the algorithms proposed by MacQueen 1996. K-means clustering is one of the non-hierarchical methods that can partition data into several clusters. Data with the same or similar characteristics are collected in the same group as well. While data with different characteristics are clustered in different groups. K-means clustering is iterative, numerical and falls into the unsupervised learning category. Here is the k-means clustering algorithm:

1. Determine the amount of data to cluster.
2. Determine the number of clusters ($\boldsymbol{K}$) to be formed.
3. Determine the initial cluster centroid randomly. Determining the initial centroid is done randomly from the available objects as many as $\boldsymbol{K}$ clusters. Then, to calculate the centroid of the next Ith cluster using the equation that will be shown at a later stage.
4. Calculate the distance of each object to each centroid of each cluster.
   The distance between objects and the centroid of each cluster can use the euclidean distance in Equation 1. until the closest distance of each data to the centroid is found.

$$\boldsymbol{d(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}} \qquad (\mathbf{1})$$

   with:
   $\boldsymbol{d(X,Y)}$ = The distance between object $\boldsymbol{X}$ and object $\boldsymbol{Y}$.
   $\boldsymbol{n}$ = Number of cluster variables.
   $\boldsymbol{x_n . y_n}$ = The nth data of object $\boldsymbol{X}$ and $\boldsymbol{Y}$.
5. Calculate the new centroid value with the formed cluster membership. The new centroid value is obtained from the average (mean) of the cluster concerned using the following Equation 2.

$$\boldsymbol{V_{ij} = \frac{1}{N_i}\sum_{k=0}^{N_i} X_{kj}} \qquad (\mathbf{2})$$

   with:
   $\boldsymbol{V_{ij}}$ = Centroid of $\boldsymbol{i}$-th cluster for $\boldsymbol{j}$-th variable.
   $\boldsymbol{N_i}$ = Number of members of the $\boldsymbol{i}$-th cluster.
   $\boldsymbol{X_{kj}}$ = The $\boldsymbol{k}$-th data value of the $\boldsymbol{j}$-th variable for the cluster.
   $\boldsymbol{i.k}$ = The index of the cluster.
   $\boldsymbol{j}$ = Variabel indexes.

2.2 Self Organizing Maps (SOM)

Self organising maps (SOM) were developed by Teuvo Kalevi Kohonen in 1982. SOM is one of the techniques in artificial neural networks that is used to separate data into several clusters (groups). each data that has similar objects to one another is grouped into one group. SOM itself is an unsupervised learning method. unsupervised learning which does not require a given output target, the purpose of this algorithm is to determine the set of centroids (neurons) to represent clusters with topological constraints. topology refers to the arrangement of centroids on the output grid, the most commonly used grid topologies are hexagonal and rectangular(Melin & Castillo, 2021). One of the advantages of SOM algorithm for visualisation and cluster analysis is that it can be used to explore clustering and relationships in high-dimensional data by projecting the data into two dimensions that clearly show regions of similarity (Ozcalici & Bumin, 2020). The following is the network structure of the SOM method can be seen in Figure 1 below **(**Kapita & Abdullah, 2020).

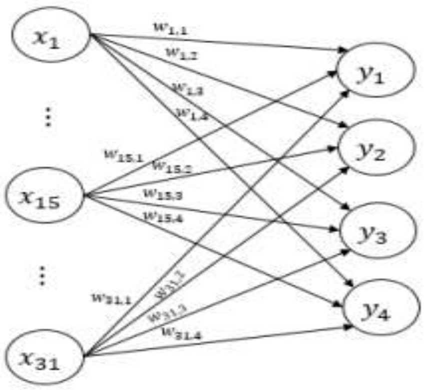

Figure 1. Arsitektur SOM

SOM is a two-layer neural network that can represent N-dimensional features into 2-dimensional or one-dimensional representations, which aims to retrain the neural network without much modification to the algorithm. The SOM algorithm has two layers, namely the input layer and the competition or output layer as shown in Figure 1 (Sun et al., 2020). Initialise by assigning random weights or coordinates to neurons that represent the cluster centroid. Each iteration involves selecting the data and figuring out the neuron closest to it using Equation 3.

$$D_j = \sqrt{\sum_{i=1}^{m} \left(W_{ji} - x_i\right)^2} \quad (3)$$

with:
W_{ji} = The connecting weight between the input neuron and the output neuron.
x_i = Neuron at the i-th input layer.
m = The number of variables.
This is a competitive approach where neurons will compete with each other to be selected the closest neuron. The weight of the selected neuron is adjusted in such a way that it is close to the data point while there is no change in other neurons using the following Equation 4.

$$W_{ji}(baru) = W_{ji}\,(lama) + \alpha\left[x_i - W_{ji}(lama)\,\right] \quad (4)$$

where the neuron weights will decrease gradually and is called the learning rate using the following Equation 5. (Kumar & Mohan, 2018).

$$\alpha(t) = \alpha_i\left(1 - \frac{t}{t_{max}}\right) \quad (5)$$

with:
α_i = The i-th initial learning rate.
t_{max} = Maximum number of iterations.
Each node from the input will be connected to all nodes in the output layer, but the nodes in the output layer are not connected to each other. Nodes in the output layer are map nodes (Andriani et al., 2018).
The SOM algorithm has a working principle by reducing neurons in its neighbours, in the end there will only be one neuron that is selected and becomes the winner (Nasution et al., 2020). The SOM method has three important components as follows (Solikin et al., 2021). The stages of the SOM network pattern clustering algorithm are as follows (Hartatik & Cahya, 2020).

1. Initialise the weights W_{ji} randomly with the weight matrix column being the number of elements in a vector and the weight matrix row being the maximum number of clusters to be formed.
2. Calculate the distance between the input value and the weight D_j using the euclidean distance as follows (Kusumah et al., 2017).
3. Determine the minimum value by looking at the results of the calculation of the distance vector D_j, then changes will be made to the weights using the following formula (Kusumah et al., 2017):
$W_{ji}(baru) = W_{ji}\,(lama) + \alpha\left[x_i - W_{ji}(lama)\,\right]$

At the stage of obtaining new weights requires a learning rate value (α) is $0 \leq \alpha \leq 1$. To make changes to the learning rate can use the following formula (Hidayatin et al., 2019).
$\alpha t{=}\alpha i 1{-}ttmax$

4. The state of the test stop can be known by calculating the weight of Wji (new) with Wji (old) when the weight of Wji is known not to change or the change is not much, it can be said that the test stops and has reached convergence.

3. METHODOLOGY

3.1 Data source

The data used in this study is secondary data that comes from the Central Bureau of Statistics and the Health Office in the Provinces of Maluku, North Maluku, Papua and West Papua in 2021. The type of data used is data that comes from data collection at a certain point in time in various regions /area (data cross-section). This data will be grouped to identify regions based on the indicators studied. The object of research is 63 regencies/cities in the Provinces of Maluku, North Maluku, Papua and West Papua which will be observed in 2021.

3.2 Operational Definition of Research Variables

To obtain a clear picture of poverty by taking into account regional conditions in Maluku Province, the variables used are the following variables: Total Population (X1), Percentage of Poor People (X2), Total Poor People (X3), Population Density per km2 (X4), Gross Regional Domestic Income (GRDP) (X5), and Expenditure per capita (X6).

3.3 Research Methods and Stages

The method and stages of research that will be carried out to answer the formulation of the problem in this study are.

1. Collecting data related to poverty indicators which are the variables in this study.
2. Grouping with k-means analysis.
    a. The number of groups used here is the same as the number of groups used in Self Organizing Maps (SOM).
    b. Determine the cluster center, the cluster center used is the average of the cluster centers formed in the Self Organizing Maps (SOM).
    c. Perform calculations of the distance between the data and the center formed using the euclidean distance.
    d. Allocate each data to the nearest centroid/average, by looking at the smallest distance to the centroid.
    e. Return to step b, if there are still people who change groups, or if there is a change in the centroid value. This process is carried out until convergence.
3. Doing regional grouping based on the data collected in accordance with the research variables using the SOM method.
    a. Initialize the weight on each node, determine the neighboring topology parameters. Determine the learning rate parameter and determine the maximum number of training iterations.
    b. Choose a vector randomly
    c. Each input vector is calculated by the distance with the weight of the node and the minimum distance is selected as the Best Matching Unit (BMU). To determine the BMU, the method used is to calculate the distance between the nodes and the input using the Euclidean distance.
    d. Determines the BMU neighbor nodes. After the BMU has been determined, the next step is to determine which nodes are BMU's neighboring nodes. What needs to be done is to determine the distance/radius from the BMU.
    e. This process will stop when the results reach convergence.
4. The clusters formed were then tested with Manova Analysis to see if there was no overlapping between each group so that they could be used to clarify the boundaries of the SOM grouping visualization.
5. After obtaining the optimum cluster, then the grouping results are interpreted based on the variables used.
6. Summarizing and Interpreting the results obtained from the groupings formed using the k-means algorithm and Self Organizing Maps.

4. DISCUSSION AND ANALYSIS

4.1 Overview of Research Variables

Before the calculation process, data collection is first carried out. Data collection was carried out indirectly or data was obtained from other parties, namely the Central Bureau of Statistics for the Provinces of Maluku, North Maluku, Papua and West Papua. The data used are the total population (X1), the percentage of poor people (X2), the number of poor people (X3), population density per km2 (X4), GRDP (X5) and per capita expenditure (X6).

Table 1. Descriptive Statistics

	N	Min	Max	Mean
Total population	63	22860	424730	137683.93
Percentage of Poor Population	63	3.55	41.66	23.30
Number of Poor Population	63	4.05	78.18	24.57
Population Density km2	63	1.08	1848.19	96.62
GRDP	63	-2.29	162.00	5.56
Per Capita Expenditures	63	6335	2709254	1020005.30

Based on Table 1, it shows that the number of sample units in this study were 63 districts/cities spread across the Provinces of Maluku, North Maluku, Papua and West Papua. The highest population of regencies/cities is 424,730 people and the least is 22,860 people. Furthermore, the highest percentage of poor people is 41.66 percent and the lowest is 3.55 percent. Furthermore, the highest number of poor people is 78.18 percent and the lowest is 1.08 percent. Furthermore, the highest population density per km2 is 1848.19 and the lowest is 1.08. Furthermore, the highest GRDP was 162 and the lowest was -2.29. Furthermore, the highest per capita expenditure was 2709254 and the lowest was 6335. For more detail on the research variables used in this study, they can be broken down as follows.

4.2 Designing/Designing Research

a. Data Standardization

Data standardization is carried out when there are significant differences in units between the variables studied. The following is the result of standardizing the variables used in the research.

Table 2. Standardization of Research Variables

Regency/City	X1	...	X6
Maluku Tenggara Barat	-0.14266	...	-1.65548
Maluku Tenggara	-0.1577	...	-1.65345
Maluku Tengah	3.00907	...	-1.6491
Buru	-0.01353	...	-1.64893
Kepulauan Aru	-0.36447	...	-1.65338
Seram Bagian Barat	0.8077	...	-1.65158
Seram Bagian Timur	0.02712	...	-1.65034
Maluku Barat Daya	-0.58177	...	-1.65453
Buru Selatan	-0.63913	...	-1.65341
Ambon	2.20099	...	-1.64272
Tual	-0.49649	...	-1.65373
Halmahera Barat	-0.03201	...	0.00232
Halmahera Tengah	-0.83732	...	0.17322
Kepulauan Sula	-0.33955	...	-0.41717

Halmahera Selatan	1.19511	...	-0.12235
Halmahera Utara	0.65258	...	0.04579
Halmahera Timur	-0.4689	...	0.01141
Pulau Morotai	-0.64556	...	0.08197
Pulau Taliabu	-0.82752	...	-0.48435
Ternate	0.71479	...	1.07552
Tidore Kepulauan	-0.22575	...	0.41152
Merauke	0.98552	...	0.92392
Jayawijaya	1.42155	...	0.66942
Jayapura	0.32279	...	0.71922
Nabire	0.34835	...	1.12555
Kepulauan Yapen	-0.24607	...	0.3715
Biak Numfor	-0.02571	...	0.15413
Paniai	0.89925	...	-0.49582
Puncak Jaya	0.94301	...	2.38163
Mimika	1.87236	...	0.83075
Boven Digoel	-0.76492	...	0.02077
Mappi	-0.29462	...	-0.57654
Asmat	-0.2731	...	0.01541
Yahukimo	2.28592	...	-0.21391
Pegunungan Bintang	-0.62379	...	-0.06796
Tolikara	1.07542	...	-0.31101
Sarmi	-1.00463	...	-0.04029
Keerom	-0.79174	...	0.06763
Waropen	-1.08257	...	0.32742
Supiori	-1.20369	...	0.09971
Mamberamo Raya	-1.05557	...	-0.3049
Nduga	-0.312	...	2.7588
Lanny Jaya	0.63948	...	1.56751
Mamberamo Tengah	-0.90702	...	-0.06088
Yalimo	-0.35953	...	1.46657
Puncak	-0.23282	...	0.54157
Dogiyai	-0.20825	...	-0.68112
Intan Jaya	-0.00805	...	-0.27394
Deiyai	-0.39015	...	-0.52089
Kota Jayapura	2.79181	...	1.13152
Fakfak	-0.54372	...	0.76293
Kaimana	-0.78335	...	0.38302
Teluk Wondama	-0.99666	...	0.35785
Teluk Bintuni	-0.50597	...	0.8476
Manokwari	0.59984	...	0.67435
Sorong Selatan	-0.88598	...	-0.15096
Sorong	-0.1648	...	0.70504
Raja Ampat	-0.75771	...	0.17519
Tambrauw	-1.11432	...	0.01237
Maybrat	-0.98875	...	0.8716
Manokwari Selatan	-1.0539	...	0.3096
Pegunungan Arfak	-1.03516	...	-0.39343
Kota Sorong	1.59427	...	1.20783

b. Outlier Detection

Based on research data that has been standardized, if there is data whose value is not between ± 2.5, it means that the data is an outlier. The results of observations on outlier data are shown in Table 2 previously. Based on the previous Table 2, it can be seen that no value exceeds ± 2.5. So it can be concluded that the data used in this study did not contain outliers.

c. Test Assumptions

There are assumptions that must be met in cluster analysis, namely:

Assumption of Sufficient Sample

To find out whether the sample used is sufficient for analysis, it can be seen from the Kaiser Meyer Olkin (KMO) value.

H0 : The sample is not sufficient for further analysis

H1 : The sample is sufficient for further analysis

Test Statistics

$$\text{KMO} = \frac{\sum_{i=1}^{\rho}\sum_{j=1}^{\rho} r_{ij}^2}{\sum_{i=1}^{\rho}\sum_{j=1}^{\rho} r_{ij}^2 + \sum_{i=1}^{\rho}\sum_{j=1}^{\rho} \rho_{ij}^2}$$

Test Criteria:

If the KMO value is > 0.5 then it fails to receive H0 or the sample is eligible for further analysis.

The results of the KMO test can be seen in Table 3 below.

Table 3. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		**.628**
Bartlett's Test of Sphericity	Approx. Chi-Square	117.944
	Df	15
	Sig.	.000

Based on Table 3, the KMO and Barlett's tests are useful for determining the feasibility of a variable, whether it can be further processed using this cluster analysis technique or not. You do this by looking at the KMO MSA value. If the KMO MSA value is greater than 0.50 then the cluster analysis technique can be continued. Based on the table 3, it can be seen that the KMO MSA value is 0.628 > 0.50 and the Barlett's Test of Spherictу (Sig.) value is 0.000 <0.05, so the cluster analysis in this study can be continued because it meets the first requirement.

4.3 Multicollinearity

The second assumption is multicollinearity. To determine whether there is multicollinearity, it can be seen from the correlation values in the correlation matrix. It is said to be multicollinearity if the correlation value is greater than 0.80. Furthermore, the results of the multicollinearity test can be seen in Table 4.

Table 4. Multicollinearity Test

	X1	X2	X3	X4	X5	X6
X1	1	-.186	.683**	.336**	-.048	.046
X2	-.186	1	.442**	-.332**	-.188	.108
X3	.683**	.442**	1	-.023	-.130	.090
X4	.336**	-.332**	-.023	1	-.041	.034
X5	-.048	-.188	-.130	-.041	1	.041
X6	.046	.108	.090	.034	.041	1

Based on Table 4, it shows that all variables have a Pearson correlation value of less than 0.80 so it can be concluded that there are no cases of multicollinearity in the research variables.

4.4 K-Means Cluster

K-means clustering is a non-hierarchical cluster analysis method that seeks to partition existing objects into one or more clusters or groups of objects based on their characteristics, so that objects with the same characteristics are grouped in the same cluster and objects with different characteristics. grouped into another cluster. The K-Means Clustering method attempts to group existing data into several groups, where data in one group have the same characteristics as each other and have different characteristics from data in other groups. Cluster Analysis Model with RapidMiner is presented on Figure 2.

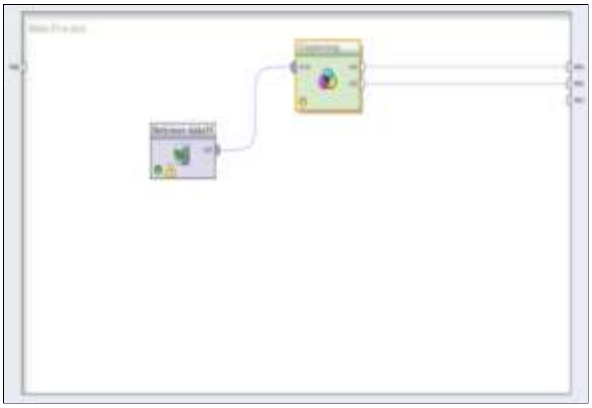

Figure 2. Cluster Analysis Model with RapidMiner

The following is the first appearance of the data clustering process before the operation is carried out.

Table 5. Initial Cluster Centers

	Cluster		
	1	2	3
Zscore(X1)	115474.0000	107921.0000	424730.0000
Zscore(X2)	36.2600	37.1800	19.8400
Zscore(X3)	40.7800	41.1700	74.5900
Zscore(X4)	14.3400	84.6400	53.0000
Zscore(X5)	.7200	2.7100	2.6500
Zscore(X6)	1351616.0000	2709254.0000	10243.0000

Based on Table 5, shows the results of the temporary process of grouping data that is carried out because this process is just the beginning, it is necessary to do the next process.

Table 6. Iteration History

Iteration	Change in Cluster Centers		
	1	2	3
1	160956.557	130066.268	264098.293
2	32251.548	322545.677	.000
3	27420.684	189007.953	.000
4	37666.144	100123.566	49701.012
5	23221.314	88031.526	.000
6	23715.814	65552.897	.000
7	22306.184	54166.788	.000
8	35941.195	62609.279	.000
9	.000	.000	.000

Based on Table 6, it shows that the iteration process in grouping clusters from the initial table results in an iterative process of 9 times. In iterations 1-8 there are many insignificant centeroids and in iteration 9 there are significant centeroids. So, all clusters have been formed and the iteration stops at iteration 9.

Table 7. Final Cluster Centers

	Cluster		
	1	2	3
Zscore(X1)	102229.9142	190200.8823	169329.6363
Zscore(X2)	23.6391	23.8358	21.4445
Zscore(X3)	18.7957	33.4611	29.2554
Zscore(X4)	28.2928	190.2470	169.3636
Zscore(X5)	7.5017	3.9576	1.8872
Zscore(X6)	992833.9708	1730279.2005	8763.5454

Based on Table 7, shows the results of the final process in clustering which forms 3 clusters for each variable. The variables in the final cluster centers table are the results for standardized values with positive numbers meaning that the data is above the total average.

Table 8. Regency/City Clusters Formed

District	Cluster	Distance
Maluku Tenggara	3	46704.66
Maluku Tengah	3	255404.7
Buru	3	32974.98
Ambon	3	178398.4
Tual	3	79019.54
Ternate	3	54065.74
Tidore Kepulauan	3	279494
Merauke	3	150383.5
Jayapura	3	270758.2
Mimika	3	237780.5
Kota Jayapura	3	214512.7
Teluk Bintuni	3	216203.2
Manokwari	3	297398.4
Sorong	3	286808.1
Kota Sorong	3	103786
Maluku Tenggara Barat	2	45319.97
Seram Bagian Timur	2	29067.84
Pulau Taliabu	2	272889.1
Jayawijaya	2	311658.6
Nabire	2	28576.95
Puncak Jaya	2	748965.2
Nduga	2	982426.4
Lanny Jaya	2	249677
Yalimo	2	206824.9
Fakfak	2	264585.3
Kaimana	2	264632.4
Sorong Selatan	2	81650.58
Raja Ampat	2	139395.7
Maybrat	2	229655.5
Manokwari Selatan	2	226304
Kepulauan Aru	1	66423.61
Seram Bagian Barat	1	45403.56
Maluku Barat Daya	1	87162.31
Buru Selatan	1	92622.07
Halmahera Barat	1	43212.5
Halmahera Tengah	1	140444.4
Kepulauan Sula	1	228284.5
Halmahera Selatan	1	156901.4
Halmahera Utara	1	112224.7
Halmahera Timur	1	35392.95
Pulau Morotai	1	81655.81
Kepulauan Yapen	1	254926.7
Biak Numfor	1	125948.4
Paniai	1	301843.9
Boven Digoel	1	54759.37
Mappi	1	325932.8
Asmat	1	37795.16
Yahukimo	1	273946.5
Pegunungan Bintang	1	28053.66
Tolikara	1	213800.7
Sarmi	1	60432.82
Keerom	1	79433.91
Waropen	1	237543.1
Supiori	1	118672.8
Mamberamo Raya	1	172350.2
Mamberamo Tengah	1	52060.81
Puncak	1	359026.4
Dogiyai	1	390198.5
Intan Jaya	1	144784.3
Deiyai	1	291782.3
Teluk Wondama	1	253398.8
Tambrauw	1	78907.28
Pegunungan Arfak	1	222907.8

Based on Table 8, it shows that there are 3 clusters formed in the classification of districts/cities based on poverty indicators in the Provinces of Maluku, North Maluku, Papua and West Papua. With the following details. Distribution of district/city classifications based on poverty indicators is presented on Figure 3.

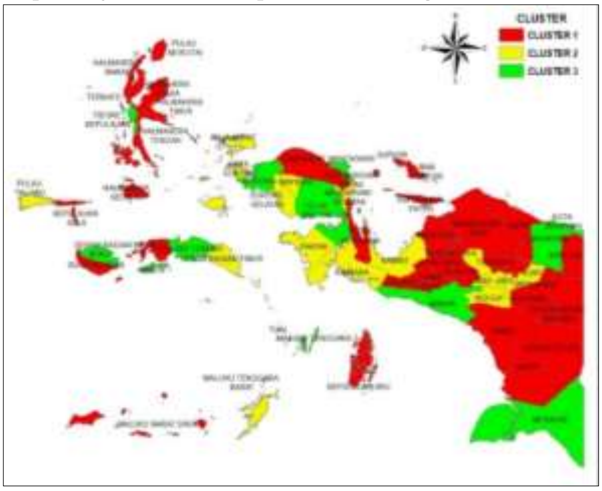

Figure 3. Distribution of district/city classifications based on poverty indicators

Cluster 1 consists of the Aru Islands, West Seram, Southwest Maluku, South Buru, West Halmahera, Central Halmahera, Sula Islands, South Halmahera, North Halmahera East Halmahera, Morotai Island, Yapen Islands, Biak Numfor, Paniai, Boven Digoel, Mappi , Asmat, Yahukimo, Mountains of Stars, Tolikara, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya, Mamberamo Tengah, Puncak, Dogiyai, Intan Jaya, Deiyai, Teluk Wondama Tambrauw, Arfak Mountains.

Cluster 2 consists of West Southeast Maluku, East Seram, Taliabu Island, Jayawijaya, Nabire, Puncak Jaya, Nduga, Lanny Jaya, Yalimo, Fakfak, Kaimana, South Sorong, Raja Ampat Maybrat, South Manokwari.

Cluster 3 consists of the Regencies of Southeast Maluku, Central Maluku, Buru, Ambon, Tual, Ternate, Tidore Islands, Merauke, Jayapura, Mimika, Jayapura City, Bintuni Bay, Manokwari, Sorong, Sorong City.

Table 9. Distances between Final Cluster Centers

Cluster	1	2	3
1		742673.808	986355.410
2	742673.808		1721642.170
3	986355.410	1721642.170	

Based on Table 9, Distances between final cluster centers show the distance between clusters, the greater the value/number, the greater/width the distance between clusters. Cluster 1 is 742673.808 from cluster 2 and 986355.410 from cluster 3. The distance between Cluster 2 and Cluster 1 is 742673.808 and with Cluster 3 is 1721642.170. The distance between Cluster 3 and Cluster 1 is 986355.410 and with cluster C is 1721642.170..

To test the significance level between clusters and find out the differences in each cluster, it is necessary to do an ANOVA test. The provisions for using the F number in cluster analysis are that the greater the calculated F number (if a hypothesis test is carried out, the calculated F will be greater than the F table) and the significance level (sig) < 0.05; the greater the difference between the three clusters formed. The following are the results of the Anova test as follows.

Table 10. ANOVA test

	Cluster		F	Sig.
	Mean Square	Df		
Zscore(X1)	50948512969.34	2	.612	.003
Zscore(X2)	23.38	2	.201	.018
Zscore(X3)	1376.18	2	.474	.015
Zscore(X4)	185315.69	2	.498	.014
Zscore(X5)	161.99	2	.370	.039
Zscore(X6)	9925430911261.27	2	.436	.000

Based on Table 10, the decision rules are as follows.

H0 = The three clusters have no significant difference
H1 = The three clusters have significant differences

Reject the null hypothesis (H0) if the p-value <0.05

Based on Table 10, it shows the p-value of the six variables used in the study, namely the Total Population (X1) variable of 0.003, the Percentage of Poor People (X2) of 0.018, the Number of Poor People (X3) of 0.015, Population Density per km2 (X4) of 0.14, GRDP (X5) of 0.039 and Expenditures Per Capita (X6) of 0.00. So it can be concluded that there is a significant difference between the six variables because the p-value obtained is <0.05.

Table 11. Number of Cases in each Cluster

Cluster	1	33.000
	2	15.000
	3	15.000
Valid		63.000
Missing		.000

Based on Table 11, Number of cases in each cluster, provides an overview of the number of Regencies/Cities included in each cluster. Cluster 1 consists of 33 Regencies/Cities, cluster 2 consists of 15 Regencies/Cities and cluster 3 consists of 15 Regencies/Cities.

4.5 Self Organizing Maps

Self-Organizing Map (SOM) or often called topology-preserving is a technique in Neural Networks that aims to visualize data by reducing data dimensions through the use of self-organizing neural networks so that humans can understand high-dimensional data that is mapped in the form low-dimensional data. In Table 12, the clustering of districts/cities in Maluku, North Maluku, Papua and West Papua Provinces is presented using the SOM method.

Table 12. Clusters formed

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Kep.Tanimbar	Halmahera Selatan	Ambon	Maluku Tengah
Maluku Tenggara	Mimika	Ternate	Jayawijaya
Buru			Nabire
Kepulauan Aru			Paniai
SBB			Puncak Jaya
SBT			Yahukimo
MBD			Tolikara
Buru Selatan			Nduga
Tual			Lanny Jaya
Halmahera Barat			Yalimo
Halmaher Tengah			Puncak
Kepulauan Sula			Kota Jayapura
Halmahera Utara			Manokwari
Halmahera Timur			Kota Sorong
Pulau Morotai			
Pulau Taliabu			
Tidore Kepulauan			
Merauke			
Jayapura			
Kepulauan Yapen			
Bika Numfor			
Boven Digoel			
Mappi			
Asmat			
Pegunungan Bintang			
Sarmi			
Keerom			
Waropen			
Supiori			
Memberamo Raya			
Memberamo Tengah			
Dogiyai			
Intan Jaya			
Fakfak			
Kaimana			
Teluk Wondama			
Teluk Bintuni			
Manokwari			
Sorog Selatan			
Sorong			
Raja Ampat			
Tambrauw			
Maybrat			
Manokwari Selatan			
Pegunungan Arfak			

Based on Table 12, it shows that there are 4 clusters formed in the classification of poverty in the Provinces of Maluku, North Maluku, Papua and West Papua. With the following details. Distribution of district/city classifications based on poverty indicators is presented on figure 4.

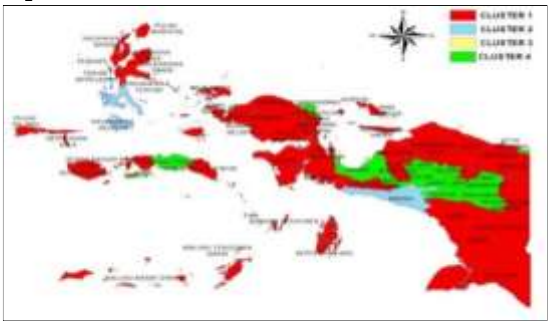

Figure 4. Distribution of district/city classifications based on poverty indicators

Cluster 1 consists of the Tanimbar Islands, Southeast Maluku, Buru, Aru Islands, SBB, SBT, MBD, South Buru, Tual, West Halmahera, Central Halmaher, Sula Islands, North Halmahera, East Halmahera, Morotai Island, Taliabu Island, Tidore Islands , Merauke, Jayapura, Yapen Islands, Bika Numfor, Boven Digoel, Mappi, Asmat, Bintang Mountains, Sarmi, Keerom, Waropen, Supiori, Memberamo Raya, Memberamo Tengah, Dogiyai, Intan Jaya, Fakfak, Kaimana, Teluk Wondama, Teluk Bintuni, Manokwari, South Sorog, Sorong, Raja Ampat, Tambrauw, Maybrat, South Manokwari, Arfak Mountains.

Cluster 2 consists of South Halmahera Regency and Mimika Regency.

Cluster 3 consists of Ambon City and Ternate City of Ternate.

Cluster 4 consists of the Regencies of Central Maluku, Jayawijaya, Nabire, Paniai, Puncak Jaya, Yahukimo, Tolikara, Nduga, Lanny Jaya, Yalimo, Puncak, City of Jayapura, Manokwari, City of Sorong. Visualization of grouping with the SOM method in two dimensions is presented in Figure 5.`

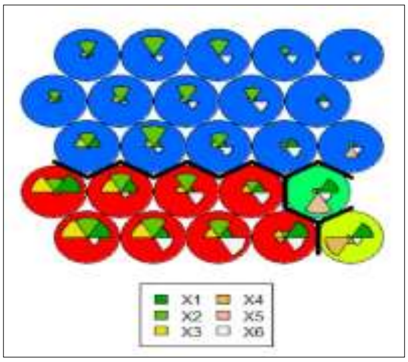

Figure 5. Codes Plot

The results of SOM grouping are in two-dimensional form using color codes and node distances as a differentiator between groups. This can be seen in Figure 5, which shows the results of SOM grouping. In the figure it can be seen that every city/district that has similar characteristics (variable value) is in the same group. Color codes and node distances for each data group are distinguished by a certain color. A very striking color difference indicates that this group is indeed very different from other groups and is stated as one group.

5. CONCLUSION

Based on the results and discussion, this study obtained the following results.

1. District/city classification based on poverty indicators in Maluku, North Maluku, Papua and West Papua Provinces using the K-Means Cluster Analysis obtained 3 clusters formed with details of cluster 1 consisting of 33 districts/cities, Cluster 2 consisting

of 15 districts/cities and Cluster 3 consists of 15 regencies/cities.

a. Cluster 1 consists of the Aru Islands, West Seram, Southwest Maluku, South Buru, West Halmahera, Central Halmahera, Sula Islands, South Halmahera, North Halmahera East Halmahera, Morotai Island, Yapen Islands, Biak Numfor, Paniai, Boven Digoel, Mappi , Asmat, Yahukimo, Mountains of Stars, Tolikara, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya, Mamberamo Tengah, Puncak, Dogiyai, Intan Jaya, Deiyai, Teluk Wondama Tambrauw, Arfak Mountains.
b. Cluster 2 consists of West Southeast Maluku, East Seram, Taliabu Island, Jayawijaya, Nabire, Puncak Jaya, Nduga, Lanny Jaya, Yalimo, Fakfak, Kaimana, South Sorong, Raja Ampat Maybrat, South Manokwari.
c. Cluster 3 consists of the Regencies of Southeast Maluku, Central Maluku, Buru, Ambon, Tual, Ternate, Tidore Islands, Merauke, Jayapura, Mimika, Jayapura City, Bintuni Bay, Manokwari, Sorong, Sorong City.

2. Classification of districts/cities based on poverty indicators in the provinces of Maluku, North Maluku, Papua and West Papua using Self-Organizing Maps (SOM) obtained 4 clusters formed with details of cluster 1 consisting of 45 districts/cities, Cluster 2 consisting of 2 districts/cities and cluster 2 consists of 14 districts/cities.
   a. Cluster 1 consists of the Tanimbar Islands, Southeast Maluku, Buru, Aru Islands, SBB, SBT, MBD, South Buru, Tual, West Halmahera, Central Halmaher, Sula Islands, North Halmahera, East Halmahera, Morotai Island, Taliabu Island, Tidore Islands , Merauke, Jayapura, Yapen Islands, Bika Numfor, Boven Digoel, Mappi, Asmat, Bintang Mountains, Sarmi, Keerom, Waropen, Supiori, Memberamo Raya, Memberamo Tengah, Dogiyai, Intan Jaya, Fakfak, Kaimana, Teluk Wondama, Teluk Bintuni, Manokwari, South Sorog, Sorong, Raja Ampat, Tambrauw, Maybrat, South Manokwari, Arfak Mountains.
   b. Cluster 2 consists of South Halmahera Regency and Mimika Regency.
   c. Cluster 3 consists of Ambon City and Ternate City of Ternate.
   d. Cluster 4 consists of the Regencies of Central Maluku, Jayawijaya, Nabire, Paniai, Puncak Jaya, Yahukimo, Tolikara, Nduga, Lanny Jaya, Yalimo, Puncak, City of Jayapura, Manokwari, City of Sorong.

References:

Alkire, S., & Santos, M. E. (2013). A multidimensional approach: Poverty measurement & beyond. *Social indicators research*, *112*(2), 239-257.

Andriani, D. P., Rahmatika, F., & Susanto, M. (2018). Upaya sustainabilitas UKM susu melalui pengendalian kualitas kandungan kadar lemak susu menggunakan statistical quality control method. *Industri Inovatif: Jurnal Teknik Industri*, *8*(1), 1-8.

BPS Provinsi Maluku (2022). Maluku Dalam Angka 2021, Badan Pusat Statistik ProvinsiMaluku, Ambon.

Gera, N., Fatta, D. D., Garg, R., & Malik, S. (2021). Which are online shopping determinants? Analysing ease and convenience to use, prior shopping experience, online benefits, social influence in India. *International Journal of Electronic Marketing and Retailing*, 12(1), 19-35.

Harli, E., Fauzi, A., & Kusmanto, T. H. (2016). Pengelompokkan Kelas Menggunakan Self Organizing Map Neural Network pada SMK N 1 Depok. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 2(2), 90-95.

Hartatik, H. and Cahya, A. S. D. (2020). Clusterisasi Kerusakan Gempa Bumi di Pulau Jawa Menggunakan SOM. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, 2(02), 25–34.

Hendayanti, N. P. N., Putri, G. A. M. A., & Nurhidayati, M. (2018). Ketepatan Klasifikasi Penerima Beasiswa STMIK STIKOM Bali dengan Hybrid Self Organizing Maps dan Algoritma K-Mean. *Jurnal Varian*, 2(1), 1-7.

Hidayatin, I., Adinugroho, S., and Dewi, C. (2019). Pengelompokan Wilayah berdasarkan Penyandang Masalah Kesejahteraan Sosial (PMKS) dengan Optimasi Algoritme K-Means menggunakan Self. *Jurnal Pengembangan Teknologi*, 3(8):7524–7531.

Kapita, S. N. & Abdullah, S. D. (2020). Pengelompokkan Data Mutu Sekolah Menggunakan Jaringan Syaraf Tiruan Kohonen-SOM. *JIKO Jurnal Informatika dan Komputer*, 3(1):56–61.

Kohar N., I. (2021). Pemetaan Pasar Tradisional Di Kabupaten Ponorogo Berbasis Webgis Menggunakan Metode K-Means Clustering (Doctoral dissertation, Universitas Muhammadiyah Ponorogo).

Kohonen, T. (1995). Learning vector quantization. In *Self-organizing maps* (pp. 245-261). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kumar, R. D. & Mohan, P. H. (2018). A New Clustering Method Using an Augmentation to the Self Organizing Maps. Proceedings of the 8th International Conference Confluence 2018 on Cloud Computing, Data Science and Engineering, Confluence 2018, pp 739–742.

Kusumah, R. D., Warsito, B., & Mukid, M. A. (2017). Perbandingan metode k–means dan self organizing map (Studi kasus: pengelompokan kabupaten/kota di jawa tengah berdasarkan indikator indeks pembangunan manusia 2015). *Jurnal Gaussian*, 6(3), 429-437.

Melin, P. & Castillo, O. (2021). Spatial and temporal spread of the COVID-19 pandemic using self organizing neural networks and a fuzzy fractal approach. Sustainability, 13(15), 8295.

Moser, C. O. (1998). The asset vulnerability framework: reassessing urban poverty reduction strategies. *World development*, *26*(1), 1-19.

Nasution, M. H. (2020). Pengendalian Kualitas Produksi Paving Block Pada CV. KSM seruni Menggunakan Peta Kendali X-Bar dan R Chart. PhD thesis, Universitas Sumatera Utara.

Niu, L. Y., Guan, Y. S., Chen, Y. Z., Wu, L. Z., Tung, C. H., & Yang, Q. Z. (2012). BODIPY-based ratiometric fluorescent sensor for highly selective detection of glutathione over cysteine and homocysteine. *Journal of the American Chemical Society*, 134(46), 18928-18931.

Ozcalici, M. & Bumin, M. (2020). An integrated multi-criteria decision making model with Self-Organizing Maps for the assessment of the performance of publicly traded banks in Borsa Istanbul. *Applied Soft Computing Journal*, 90:106166.

Pandit, Y. P., Badhe, Y. P., Sharma, B. K., Tambe, S. S., & Kulkarni, B. D. (2011). Classification of Indian power coals using K-means clustering and Self Organizing Map neural network. *Fuel*, *90*(1), 339-347.

Solikin, A. F., Kusrini, K., and Wibowo, F. W. (2021). Analisis Cluster Data Interkomparasi Anak Timbangan dengan Algoritma Self Organizing Maps. *Jurnal Teknik Informatika dan Sistem Informasi*, 7(2):435–448.

Sun, X., Wang, T., Cai, D., Hu, Z., Chen, J., Liao, H., Zhi, L., Wei, H., Zhang, Z., Qiu, Y., Wang, J., and Wang, A. (2020). Cytokine and Growth Factor Reviews Cytokine storm intervention in the early stages of COVID-19 pneumonia. *Cytokine and Growth Factor Reviews*, 53(March), 38–42.

Tasdemir, K., & Merényi, E. (2009). Exploiting data topology in visualization and clustering of self-organizing maps. *IEEE Transactions on Neural Networks*, *20*(4), 549-562.

Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on neural networks*, *11*(3), 586-600.

Gabriella Haumahu
Department of Statistics, Pattimura University
Indonesia
ghaumahu@gmail.com
ORCID: 0009-0009-7983-6970

Muhammad Yahya Matdoan
Department of Statistics, Pattimura University
Indonesia
keepyahya@gmail.com
ORCID: 0000-0001-6185-9300