Int. J. Advance Soft Compu. Appl, Vol. 8, No. 3, December 2016
ISSN 2074-8523

A Segmentation Group by Kohonen Self Organizing Maps (SOM) and K -Means Algorithms

(Case Study : Malnutrition Cases in Central Java of Indonesia)

Ayundyah Kesumawati[1] and Dewi Setianingsih[1]

[1]Department of Statistics
Faculty of Mathematics and Natural Science Islamic University of Indonesia
e-mail: ayundyah.k@gmail.com, dewisetianingsih6@gmail.com,

Abstract

Malnutrition is the condition caused by low consumption of energy and protein in daily food intake. Central Java is one of the provinces in Indonesia which has high number cases of malnutrition. Therefore, researchers classify areas in Central Java by malnutrition based on factor of facilities and health workers as well as demographic factors to assist the government in making decisions in reducing malnutrition. In this paper, K-means algorithm and the Kohonen Self Organizing Maps (SOM) are used. K-means clustering is a conventional method where grouping cannot be conducted without prior assumption test. Assumptions test is a statistical requirement that must be fulfilled in a statistical analysis in order to obtain more precise analysis results. SOM is an algorithm which can be employed to analyze high-dimensional data and grouping that does not require the assumption test. From the results of the cluster evaluation, the value of Sum Square Error (SSE) shows that clustering with SOM method results SSE value which is smaller than K-means method clustering. It means that SOM method has a higher degree of similarity than K-means clustering method has.

Keywords: *Malnutrition, K - means, Self Organizing Maps (SOM), Sum Square Error (SSE).*

1 Introduction

Malnutrition is a health problem which often arises in developed countries like Indonesia. Malnutrition is a condition caused by low consumption of energy and protein in daily food intake. In general, there are two factors causing malnutrition namely direct and indirect causes. Direct causes include infectious diseases and lack of food security. Indirect causes include lack of food availability at the household level, inadequate parenting, low health access, environmental hygiene and healthy living behavior. Poverty is a socioeconomic problem which contributes to the malnutrition. It is the most significant cause of malnutrition which is caused by inability to access health facilities. In addition, biological and environmental factors may have contributed [1].

The Central Bureau of Statistics is a Non-Government Organization that directly responsible to the President. The Central Bureau of Statistics carries out government duties in the field of statistics corresponding legislation. Central Java is one of the provinces in Indonesia which has high number cases of malnutrition. A survey conducted by the Central Bureau of Statistics show that the prevalence of malnutrition in Central Java showed 3.3% and malnutrition reached 12.4%. It means that in Central Java there are 157 malnourished children in every 1,000 children [2].

To reduce the health problems, the government holds a National Health Insurance program. National Health Insurance is a government program which aims to provide a comprehensive guarantee of health insurance for all Indonesian people to be able to live healthy , productive and prosperous. Government is responsible for the implementation of National Health Insurance for individual health. Good Nutrition role and which emphasizes promote and preventive action will be the key in succeeding National Health Insurance. People's nutrition, especially in young children, will strengthen their immunity so they do not get sick easily [3].

The purpose of this study is to find out the results of regional grouping cities/ regencies in Central Java province with K -means and Self Organizing Maps methods and reveal which method has the best clustering results based on malnutrition cases, power factor and health facilities, as well as population factors in the province Central Java 2014.

2 Clustering K-Means and Kohonen Self Organizing Maps (SOM)

Cluster analysis is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). The purpose of cluster analysis is to place objects into groups, or clusters, suggested by the data, not defined a priori, so that objects in a given cluster tend to be similar to each other in some sense, and objects in different clusters tend to be dissimilar. The result of

clustering objects must show high similarity in the cluster (homogeneity) and the inter cluster (heterogeneity) [4].

2.1 Clustering K-Means

K -Means is a method of grouping data which seeks to divide the existing data into two or more groups. This portioning method is conducted in order to differentiate data based on their characteristics. Data which has similar characteristic will be put in the same group. The purpose of the data classification is to minimize the objective function set in the process of grouping, which generally seek to minimize the variation within a group and maximize the variation between groups [5]. The assignment steps are shown below

1. Determining randomized center groups.
2. Then calculating the distance between the centers of each object with every group. To calculate the distance of the object to-i at the center of the group to-k used Euclidean distance *formula, namely*

$$d_{ik} = \sqrt{\sum_{j=1}^{m} (x_{ij} - c_{kj})^2} \quad (1)$$

With:

d_{ik} : distance object to-i in the center cluster to-k

x_{ij} : value object to-i in the variable j

c_{kj} : center cluster to-k in the variable j

m : sum of variable that used

i : object to-i (i = 1,2,3...., n)

k : cluster to-k (k = 1,2,3...., nk)

j : research variable (j = 1,2,3...., m)

3. An object will be a member of to-k a group when the distance between that object to the center of k-value group is smaller than the distance to another group.
4. Furthermore, classifying the objects that become members in each group.
5. Determining the value of the group's head can be calculated by finding the average value of the object, which is member of the group, with the following formula.

$$c_{kj} = \frac{\sum_{i=1}^{n} x_{ij}}{p}; x_{ij} \in cluster\ to\text{-}k \quad (2)$$

with:

c_{kj} : center cluster to-k in the variable j

x_{ij} : value object to-i in the variable j that include cluster to-k

p : the number of members of the group formed to-k.

6. Repeat steps 2 to 5 until there are no objects which are moved to another group.

2.2 Clustering Self Organizing Maps (SOM)

Kohonen neural network algorithm or Self Organizing Map (SOM) is an artificial neural network method introduced by Professor Teuvo Kohonen in 1982. SOM is one of the topology of Unsupervised Artificial Neural Network (ANN Unsupervised) where the training process does not require supervision [5].

Kohonen self-contained within the SOM and there are two important characteristics of this network which explains that the SOM can visualize and analyze high-dimensional data. Furthermore, the network can be used for clustering, dimensionality reduction, classification, vectors quantizazion and data mining [6].

SOM is a grouping method which provides the arrangement of classes based on the topology. SOM trains iteratively through several epochs. An epoch is defined as the process of all input patterns so that each input pattern is processed as many as the number of the epoch. SOM is a single layer neural network where neurons are set along the n - dimensional grid. SOM Kohonen network consists of two layers, they are the input layer and output layer. Each output neuron receives input through the weights connected to the input, so the weight vector has the same dimension as the input vector [7].

Let the set of m values field to record the nth into an input vector and a set of m weighting to the output node specific j be the weight vector following is a step-by-step algorithm SOM for every vector x, do [8]:

1. Initializing the weights is typically using the middle grades (middle point) or using random values [9].
2. Competition. For each output j, count value $D(W_j, X_n)$ from distance function. Determining the winner node J that minimizes $D(W_j, X_n)$ from all node output.

$$D(W_j, X_n) = \sqrt{\sum_i \left(W_{ij} - X_{ni}\right)^2} \quad (3)$$

3. Cooperation. Identifying all the output node j in a neighborhood node winner J is defined by measuring environment R. Determining renewal weight value.

$$W_{ij,new} = W_{ij,current} + \eta(X_{ni} - W_{ij,current}) \quad (4)$$

4. Updating learning rate (η) and the size of the R environment as needed. Changes in levels of learning (LR/α/ η) 0 < η <1, with the formula $\alpha(t+1) = \theta\alpha(t)$, θ is a decrease in the level of learning, decrease with the change of time t [Laurence , 1994]. Environmental sized R contains an index of all nodes within a radius R of vertices winner i * . $N_i(d) = \{j, d_{ij} \leq R\}$ [9].

5. Stop treatment when the termination criterion is achieved. Criteria for the dismissal may be restrictions on the number of iterations or when η = 0[8].

3 Experimental Result

3.1 Determination Number of Cluster

Automatically determining the number of clusters has become one of the most difficult problems in the data clustering. Most methods for automatically determining the number of clusters is thrown into a selection problem model. Typically, the clustering algorithms run with different values of K; the best value of K is then selected based on the previous determination criteria [10].

SOM can be substituted with K-Means simultaneously; the SOM training algorithm also generates the same algorithm with K-Means. In determining the number of groups, a validity of the group is needed. The validation of this group aims to determine whether the number of group is formed optimum or not. One of the validity indexes that can be used to determine whether the number of group is optimum is an index Davies Bouldin [11].

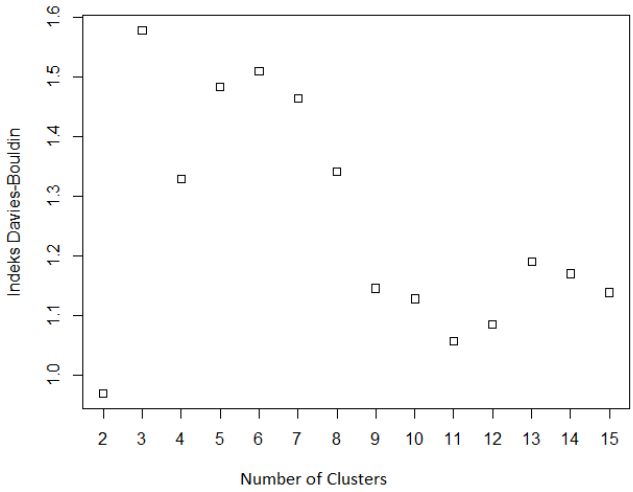

Fig. 1: Davies Bouldin Index

There is no fixed rule set in determining the number of cluster related to malnutrition cases. Good value of IDB is the smallet value. IDB smallest current value formed 2 clusters. If the number of clusters is taken as 2, then it allows the large distance in a cluster. From Figure 1, the researchers decide to create 4 clusters because forming 4 clusters will be better than forming of 3, 5, 6, 7 or 8 clusters. When formed 4 clusters would be no better than the formed 9 until 15 clusters. If the cluster is formed around 9 until 15, it is likely allowing the small distance between a cluster with another one. It is better to form not too many clusters. In addition, a good cluster should have a maximum distance to other clusters. On the other hand, the distance among the members of the cluster should minimum.

3.2 Assumption Test

3.2.1 Outlier Multivariate

Outlier is data which is deviated too far from the other data in a data set. The existence of this outlier of data will make an analysis of a series of data to be biased, or do not reflect to the actual phenomenon. The term outlier is often associated with extreme values, both large and small extremes. Mahalanobis distance for each observation can be calculated and it can show the observation distance of the average of all the variables in a multidimensional space [12].

Table 1: Mahalanobis Distance

District	MAH_1	p_MAH_1	District	MAH_1	p_MAH_1
Dist, Cilacap	25,4391	0,99748	Dist, Kudus	4,39331	0,11633
Dist, Banyumas	8,1093	0,47683	Dist, Jepara	3,78255	0,07487
Dist, Purbalingga	5,10675	0,1751	Dist, Demak	4,53256	0,12699
Dist, Banjarnegara	5,11585	0,17591	Dist, Semarang	4,65097	0,13639

Dist, Kebumen	9,01894	0,56447	Dist, Temanggung	4,53851	0,12746
Dist, Purworejo	8,57455	0,52256	Dist, Kendal	6,46278	0,30714
Dist, Wonosobo	6,40329	0,30103	Dist, Batang	5,39131	0,20105
Dist, Magelang	7,07607	0,3708	Dist, Pekalongan	15,7852	0,9285
Dist, Boyolali	5,26665	0,18953	Dist, Pemalang	10,512	0,68935
Dist, Klaten	6,13532	0,2737	Dist, Tegal	6,25026	0,28537
Dist, Sukoharjo	8,01659	0,46752	Dist, Brebes	13,3498	0,85259
Dist, Wonogiri	8,30242	0,49602	City Magelang	14,9596	0,90795
Dist, Karanganyar	6,37648	0,29827	City Surakarta	20,8079	0,98647
Dist, Sragen	4,51894	0,12593	City Salatiga	14,1116	0,88159
Dist, Grobogan	9,82864	0,63546	City Semarang	20,6869	0,98588
Dist, Blora	4,76215	0,14547	City Pekalongan	7,87	0,4527
Dist,Rembang	6,02108	0,26219	City Tegal	7,9107	0,45683
Dist, Pati	5,9319	0,25329			

If d-squared mahalanobis is bigger than the value of chi-square on the free degree, which amount is as many as the variables and p_mahalanobis is < 0.001, the data indicates multivariate outliers. Chi-square value which data is 35, the degrees of freedom (n) is 35 and alpha 5 % is 49,801. Based on Table 1, there is no mahalanobis distance which value is bigger than chi-square value. In other words, there is no indication of multivariate outliers. Mahalanobis value also shows that there is no value that is less than 0.001, then there is no indication of a multivariate outlier.

3.2.2 Multicolinearity

Multicolinearity can be seen from the value of tolerance and the opponent variance inflation factor (VIF). Tolerance measures the variability of the chosen independent variables that are not explained by other independent variable. Cutoff value, which is commonly used to indicate the presence of multicollinearity, is the tolerance value ≤ 0.1 or is equal to the VIF value ≥ 10. In searching of Tolerance and VIF value, a new variable is created; that is the sum of all variables. The new variable is used as the dependent variable and the existed variables is used as the independent variables. Here is the value of tolerance output and VIF .

Table 2: Tolerance and VIF Value

Variable	Tolerance	VIF	Variable	Tolerance	VIF
Zscore (X_1)	0,733	1,364	Zscore (X_6)	0,643	1,556
Zscore (X_2)	0,422	2,370	Zscore (X_7)	0,222	4,495
Zscore (X_3)	0,442	2,261	Zscore (X_8)	0,064	15,717
Zscore (X_4)	0,202	4,940	Zscore (X_9)	0,063	15,814
Zscore (X_5)	0,331	3,022			

Hypothesis testing

i. Hypothesis

H_0: No Multicolinearity

H_1 : Multicolinearity

ii. The level of significance (α): 5% = 0.05

iii. The critical point: Tolerance < 0.1 and VIF> 10 then refused H_0

iv. Decision
From Figure 2 it is known that there is value Tolerance < 0.1 and VIF > 10 on the X_8 and X_9 so refused H_0.

v. Conclusion
By using 95% confidence level data that is refused stating happen correlation between the independent variable or occur multicolinearity.

From testing the correlation between independent variables, it is revealed that there is multicolinearity. In this study, the data that causes multicolinearity, which is X_8 and X_9, will be disposed from clustering analysis with the K-means method.

If the data with multicolinearity is maintained, three dimensions will be formed. The first dimension will have 2.5 times more chances (five items are compared with the two other items) to affect the measure of similarity, second dimension and third dimensions.

3.3 Clustering K-Means

K-Means is included in partitioning cluster. This clustering allows data which is included in one cluster to be transferred to another cluster during the next processing. K-means clustering method classify the data based on the closest distance to the center. The results of the cluster which is formed from the K-means method is depending on the cluster's initial center value initiation [13]. Therefore, the experiments in this research are performed as many as thirty times. The result which has constant member and appearing the most will be taken for clustering process. From the data that has been standardized and has gone through the process of assumption, four groups are formed. These groups have been quantified at baseline and following the membership in each cluster:

Table 3: Cluster membership with K-means method

Cluster	Number of member	District/City
1	15	Purbalingga District, Banjarnegara District, Purworejo District, Wonosobo District, Boyolali District, Sragen District, Blora District, Rembang District, Jepara District, Demak District, Temanggung District, Batang District, Pekalongan District, Pemalang District, Tegal District

2	9	Cilacap District, Banyumas District, Kebumen District, Magelang District, Klaten District, Wonogiri District, Grobogan District, Pati District, Brebes District
3	6	Sukoharjo District, Karanganyar District, Kudus District, Semarang District, Kendal District, Semarang City
4	5	Magelang City, Surakarta City, Salatiga City, Pekalongan City, Tegal City

Table 3 above shows the membership in each cluster. Districts/cities that fit into one cluster is a region which has similar characteristics based on the existing data. Cluster 1 consists of 15 districts/cities, cluster 2 consists of nine districts/cities, cluster 3 consists of 6 districts/cities and cluster 4 consists of five districts/cities. To be able to distinguish clusters are formed and do the profiling/see the characteristics that form on each cluster.

Table 4: Profilling of cluster with K-means method

Variable	**Average**			
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Cases of Malnutrition (X_1)	89	145	40	20
The number of health center (X_2)	15	23	13	9
The number of doctors (X_3)	36	49	75	26
The number of midwifes (X_4)	282	407	185	0
The number of poor people (X_5)	133.67	211.98	87.83	24.44
Average member of family (X_6)	4	4	4	4
Inhabitant (X_7)	967	991	1886	7057
Average length of schooling (X_8)	7	7	8	9
HDI (X_9)	66.72	67.31	73.17	75.77

Cluster 1 has an average number of malnutrition cases is as many as 89 and it is still relatively high. Despite the average of facilities and health personnel are quite high, it does not make the number of cases of malnutrition low. This is likely caused by the average number of poor people is high, that is 133.67 thousand inhabitants, and average HDI value is 66.72. This value is considered low if it is compared to other clusters.

Cluster 2, with 145 cases, has the highest average value in the case of severely malnourished. According to Table 3, facilities and health workers factors in cluster 2 is considered high, it has an average number of health centers as many as 23, the average number of doctors is 49 and the average number of midwives as many as 407. The high number of facilities and health personnel factors do not make cluster 2 has a low average case of malnutrition. This is likely caused by the

average number of poor people is 211.98 thousand inhabitants and the average HDI value is not very high.

Cluster 3 has average cases of malnutrition as many as 40 cases and it is still below cluster 1 and 2. In addition, the average facilities and health workers factors are high enough; it includes the average number of health centers is 13, the average number of the doctors is 75 and the average number of midwives is 185. The density in cluster 3 is 1886 inhabitants/km2. Cluster 3 also has a higher HDI than clusters 1 and 2. The HDI in this cluster is 73.17

Cluster 4, with 20 cases, has the lowest average case of malnutrition among the other clusters. In addition, the average population density, length of schooling, and HDI is high. This is likely the reason why malnutrition cases in this cluster is low.

3.4 Clustering Self Organizing Maps (SOM)

SOM is a grouping method which provides class arrangement based on the topology. SOM trained iteratively through a number of iterations/epoch. An epoch is defined as the process of all input patterns, so each input pattern is processed as many as the epoch is [11]. The criteria for the dismissal may be restrictions on the number of iterations, or when the learning rate (η) = 0 [8].

The process of iteration in this research will be discontinued after 10000 iterations. Results iteration can be seen in the training progress figure below.

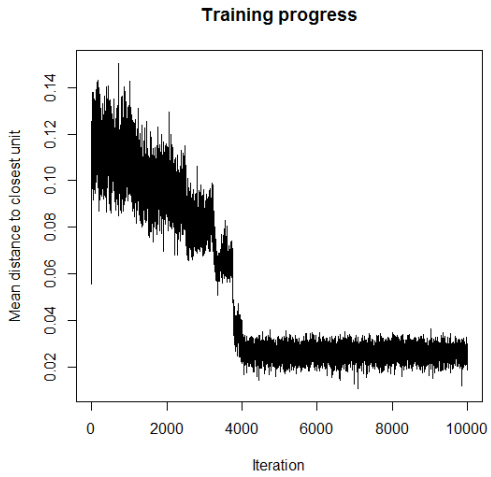

Fig. 4: Training Progress

Figure 4 above explains the amount of training progress based on the number of the iteration process. It can be seen that iteration around 4000 above shows an increasingly convergent process. The picture above also explains that the more the number of iteration is performed, the smaller mean distance is gained. As the result, the clustering results will be better. The training progress is begun by stabilizing the mean of distance cluster which value is around 0.02.

SOM Kohonen network consists of two layers, namely the input layer and output layer. Each output neuron receives input through the weights connected to the input, so the weight vector has the same dimension as the input vector [7]. In the process of putting vector input to layer input, the distance of input to the weight connected to layer input will be calculated. The shortest distance will be the winner. The process of grouping using the R software will produce a diagram fan.

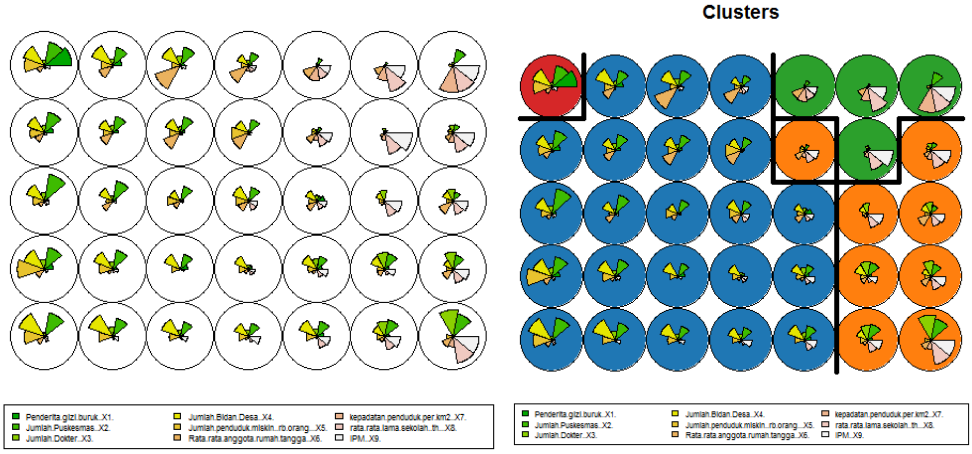

Fig. 5: Fan diagram

Fig. 5 is a fan diagram using rectangular display with 7 x 5 grid. The fan diagram shows that the distribution of the variables used in the map. Patterns can be seen by examining the dominant color.

The process of understanding the diagram in SOM algorithm is when the diagram has color and it is given boundary with the vectors which are visualized in a plot mapping. The figure above shows that the formed four clusters are represented by a different color. Each color has a different characteristic. If it is seen by naked eye, it would be difficult to determine the characteristics of each cluster because the number of variables which is used is quite a lot. The characteristics of the cluster can be seen from the average value of each variable data but before will be members of each cluster formed. The experiment takes 30 times with software R the .The result which has constant member and appearing the most will be taken.

Table 5: Cluster membership with SOM method

Cluster	Number of member	District/City
1	23	Banyumas District, Purbalingga District, Banjarnegara District, Kebumen District, Purworejo District, Wonosobo District, Magelang District, Boyolali District, Klaten District, Wonogiri District, Sragen District, Grobogan District, Blora District, Rembang District, Pati District, Jepara District, Demak District, Temanggung District, Batang District, Pekalongan

		District, Pemalang District, Tegal District, Brebes District
2	6	Sukoharjo District, Karanganyar District, Kudus District, Semarang District, Kendal District, Semarang City
3	5	Magelang City, Surakarta City, Salatiga City, City, Pekalongan, Tegal City
4	1	Cilacap District

From table 5 above, it is known that cluster 1 consists of 23 districts/cities, cluster 2 is composed of 6 districts/cities, cluster 3 is composed of five districts/cities and cluster 4 consists of one district. After that, the average of the data variables owned by each cluster will be calculated. This step is done in order to reveal the formed characteristic.

Table 6: Profilling of cluster with SOM method

Variable	**Average**			
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Cases of Malnutrition (X1)	92	40	20	533
The number of health center (X2)	18	13	9	24
The number of doctors (X3)	41	75	26	36
The number of midwifes (X4)	326	185	0	399
The number of poor people (X5)	159.7	87.83	24.44	239.8
Average member of family (X6)	4	4	4	4
Inhabitant (X7)	984	1886	7057	788
Average length of schooling (X8)	7	8	9	6
HDI (X9)	66.93	73.17	75.77	67.25

Cluster 1 is the second largest malnutrition cases with 92 cases after cluster 4. This amount is considered quite high. When it is viewed from a factor of facilities and health personnel, cluster 1 has an average number of health centers as many as 18, the average number of doctors is as many as 41 and the average number of midwives is as many as 326. Even though the number of facilities and health personnel factor is high, it does not make the average of malnutrition cases to be low. This is likely caused by the average number of poor people is quite high, that is 159.70 thousand. In addition, cluster 1 also has the lowest average of HDI value among other cluster, that is equal to 66.93.

Cluster 2 has an average case of malnutrition as many as 40 cases. Besides cluster also has an average number of health centers is 13, the average number of doctors is as many as 75 and the average number of midwives is as many as 185. The average factor of facilities and health workers is quite high but it does not create the mean average malnutrition in cluster 2 is low. If it is viewed from the average value of the number of poor, the average number of poor people in cluster 2 is quite high; that is 87.83 thousand inhabitants. This is likely to be the cause of a high average cases of malnutrition. When it is compared with cluster 1, cluster 2

has higher average value of HDI, that is 73.17, and it is likely to cause the average case of malnutrition in cluster 2 is lower than one cluster.

Cluster 3 is a cluster that has low average case of malnutrition, that is 20 cases. Although the average of facilities and health worker factors is low, it does not make cluster 3 has a high average number of malnutrition cases. The facilities and health worker factors include 9 health centers, 26 doctors and no midwives. This is likely caused by the people's average length of schooling and the average value of the HDI is higher than other clusters.

Cluster 4 is a cluster with the average case of malnutrition is very high, that is 533 cases. The average factor of facilities and health workers of cluster 4 is high; it includes 24 health centers, 36 doctors and 399 midwives. The high rate of malnutrition is likely due to the number of poor people, which is 239.80 thousand inhabitants, is higher than the other clusters. In addition, the average length of schooling is low. The people in cluster 4 mostly only take their education for 6 years. In other words, the majority of the people in cluster 4 are elementary school graduate.

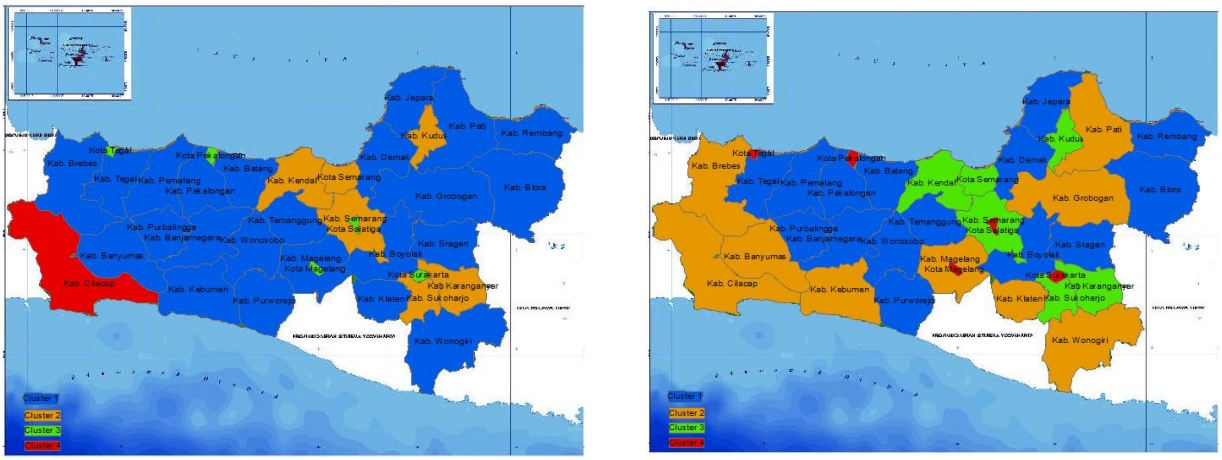

Fig. 5: Map of Central Java result SOM (left) and K-Means (right) method

3.5 Evaluation Result of Clustering

In evaluating the cluster result using k -means and SOM method, the difference between observation and the average of every cluster or Sum Square Error (SSE) is calculated by using value sum of square. If all the cases in a cluster are all identical, the value of SSE is equal to 0. The greater the amount of K is, the smaller its SSE value will be. Good cluster is a cluster with small amount of K and small value of SSE.

Table 7: Value of Sum Square Error from result clustering

Method	Cluster	SSE every cluster	SSE every method
K-Means	1	46.73	138.37
	2	44.59	
	3	27.93	
	4	19.13	
SOM	1	90.07	137.13
	2	27.93	
	3	19.13	
	4	0	

The table above describes the value of SSE on each cluster by using K-means and SOM method. If SSE value in every cluster is added, it will result a whole SEE value of all employed methods. From the table, it is revealed that the value of SSE cluster using SOM method has smaller SSE than K-means method has. This shows that cluster forming using SOM method has similarities among members in a cluster which is better than K-means method.

If it is seen from the value of SSE on each cluster with K-means method, cluster 1, with 46.73, has the highest SSE value and cluster 4, which SSE value is 19.13, has the smallest value. With the SOM method, cluster 1 also has the highest SSE value which is equal to 90.07. Meanwhile, cluster 4, which SSE value is 0, becomes the cluster with the smallest SSE value. This is indicating that the similarity in characteristics of the members in cluster 4 is better than cluster 1. 0 value in cluster 4 using SOM method shows that the members in this cluster are similar. It is showed by fact that after applying SOM method, cluster 4 only consists of one district, namely Cilacap district.

4 Conclusion

From the result of analysis, researchers concluded as follows:

The group formed by the method of K-means and SOM has particularly similar characteristics, that is a group will have high average number of malnutrition case along with high average number of poor people, low average number of the length of schooling, and low HDI. Meanwhile, a group will have low average case of malnutrition along with high average number of the length of schooling, low average number of poor people, and high HDI.

The results of cluster evaluation showed that Sum Square Error (SSE) value with SOM method has smaller value than K-Means method has. This value showed that the cluster formation by employing SOM method is better than K-Means method for this research because the results of the cluster with this method has a better degree of similarity in term of cluster formation.

For further research could be added by another variables that supporting the main cases, and compared with another algorithm.

References

[1] Ramadani, I.R., Rahmawati, R. and Hoyyi, A. 2013. Analisis Faktor-Faktor yang Mempengaruhi Gizi Buruk Balita di Jawa Tengah dengan Metode Spatial Durbin Model, *Journal Gaussian*, Vol. 2, No. 4, 333-343.

[2] Badan Pusat Statistik Republik Indonesia. 2016. Konsep Kemiskinan. http://www.bps.go.id/Subjek/view/id/23 (Las accessed: Feb 2016).

[3] Dinas Kesehatan Provinsi Jawa Timur. 2014. Permasalahan Gizi Buruk Bersifat Multidimensi yang Menyangkut Kemiskinan, Gaya Hidup dan Sosial Budaya. http://dinkes.surabaya.go.id/portal/index.php/artikelkesehatan/permasalahan-gizi-buruk-bersifatmultidimensiyangmenyangkutkemiskinan-gaya-hidup-dan-sosial-budaya/ (Last accessed: Feb 2016).

[4] JR. Joseph F.Hair, Anderson, R.E, Tatham, R.L and Black, W.C,. 1998. Multivariate Data Analysis 5th Edition, Prentice-Hall International, Inc

[5] Kohonen, Teuvo. 1990. The Self –Organizing Map. *Proceeding of The IEEE*, Vol. 78, No.9, September 1990.

[6] Bacao, Fernando and Lobo Victor. 2015. Introduction to Kohonen's Self Organizing Maps. Universidade Nova De Lisboa, Instituto Superior De Estatistica e Gestao De Informacao (Las accessed: Jan 2015).

[7] Bação, F. dan Lobo, V. 2010. Introduction to Kohonen's Self-Organizing Maps. Instituto Superior de Estatistica E Gestao de Informacao, Universidade Nova de Lisboa, Portugal.

[8] Larose DT. 2004. Discovering Knowledge in Data: An Introduction to Data mining, USA: John Wiley&Sons Inc.

[9] Demuth H, Beale M. 2003. Neural Network Toolbox For Use with MATLAB®. USA: The MathWorks, Inc.

[10]Jain, Anil.K. 2010. "Data clustering: 50 years beyond K-means". *Journal Elsevier*, 651-666.

[11]Lobo Victor J.A.S. 2009. Application of Self Organizing Maps to the Maritime Environment, Springer Verlag Berlin Heidelberg.

[12] Hair. J.R., Anderson. R,E., Tatham. R.L., and Black W.C. 2006. Multivariate Data Analysis with Readings 3th Edition, New York : Macmillan Publishing Company.

[13]Wehrens, Ron dan Buydenss, Lutgarde M.C. 2007. Self and Super-organizing Maps in R : The Kohonen Package. Journal of Statistical Software. October 2007, Volume 21. Issue 5.