Welfare Mapping of Region in Indonesia based on Health and Nutrition Indicators using Self Organizing Mapping (SOM) Method

D U Wutsqa¹, Kismiantini², R Subekti³

¹,²,³ Department of Mathematics, Yogyakarta State University, Yogyakarta, Indonesia

dhoriva_uw@uny.ac.id, kismi@uny.ac.id, retnosubekti@uny.ac.id

Abstract. Health and nutrition are important indicators of public welfare. Availability map of the region on those indicators is necessary for constructing an effective plan of welfare equal distribution in health and nutrition. This study aims to create a map of the welfare index on health and nutrition indicators over Indonesian provinces. This research employs a soft computing approach specifically a self-organizing mapping (SOM). SOM is a method based on the neural network concept. The data are drawn from the Statistics Indonesia. Sixteen variables are considered as factors to create the map. They could be grouped into six factors namely illness, mortality, life expectancy, nutritional conditions, breastfeeding and immunization, clean water facilities and latrines. A Davies Bouldin Index (DBI) is used to determine the optimal number of clusters. The best mapping is obtained by five clusters as corresponding to the least DBI value. Thirty-four provinces in Indonesia are mapped into five clusters. The characteristics of each cluster are described by six factors of health and nutrition indicators.

1. Introduction

Equal distribution of people's welfare is still an important issue for Indonesia. One indicator of people's welfare is health and nutrition. The quality of human life can be seen from the health quality and nutritional adequacy. High quality health and adequate nutrition are investment on productive resource development. Indonesia consists 34 provinces spread out from Sabang to Merauke, each province has different characteristics with different levels of welfare. Equitable programs to improve health and nutrition will be right on target if the government knows the conditions of each region. So, the development plans can be made in accordance with regional needs.

Health and nutrition indicators can be viewed from various variables, including morbidity rate, infant mortality rate, life expectancy at birth, and percentage of households with clean drinking water sources, etc. Each province has different characteristics. Some provinces are dominated by urban, rural, inland, or water areas. This is thought to affect different levels of health and nutrition for different regions. The mapping of Indonesia's territory based on these variables will be able to provide information about similarities and differences of the regions. From the results of the mapping, it can also be further explored the characteristics of each region in the same group.

Regional mapping corresponds to the clustering method. Recently, the clustering methods are developed by using a soft computing approach. Neural network is a very popular soft computing approach. Self-Organizing Mapping (SOM) is an effective clustering method that works based on the principles of the neural network [1] [2]. Classification of breast cancer [3], map of Indonesian territory
based on pollution levels [4], student preference [5], monitor of large industrial processes [6], situation in the world lifestyle and health nutrition [7], polyethylene production [8], maritime [9] are the examples of the application of SOM. In this study, we use SOM to generate a people's welfare map of region in Indonesia based on health and nutrition indicators.

Research on people's welfare mapping can be approached by supervised and unsupervised learning. The supervised learning involves dependent and independent variables and relates to the classification analysis. Welfare classifying studies have been reported by using bagging logistic regression [10] and using Support Vector Machine [11]. The unsupervised learning does not need dependent variables and relates to clustering analysis, which specifically intends for welfare mapping. It has been carried out the welfare mapping, but the indicators used are more general and not specific to nutritional and health indicators [12]. The method used is K-Nearest Neighbor (KNN). Mapping on nutritional, health, together with lifestyle has been constructed using SOM. The variables of nutrition are concentrated to the nutritional contents of foods such as fat, protein, etc., and the variables of health are tuberculosis, obesity and mortality. In this research, we map people's welfare using SOM with the different variables.

We entail sixteen variables published by Statistics Indonesia as the variables of health and nutrition indicator. They can be grouped as illness, mortality, life expectancy, nutritional conditions, breastfeeding and immunization, clean water facilities and latrines. The results of this study are expected to provide information about clusters of regions in Indonesia based on nutritional and health indicators and characteristics of the regions in the same cluster. This information can be used by government in development planning to improve people's welfare on health and nutrition indicator in accordance with regional needs.

2. Self Organizing Map

Self Organizing Map (SOM) is a class of the neural network introduced by Professor Teuvo Kohonen. Self Organizing Map (SOM) is a class of the neural network introduced by Professor Teuvo Kohonen in 1982. It has a topology of unsupervised neural network whose training process does not require target. The architecture of SOM consists of two layers, namely the input layer and the output layer. Each neuron in the input layer is connected to each neuron in the output layer. Each neuron in the output layer represents a class (cluster) from the input that has been given. The network architecture of SOM is depicted in Figure 1.

![Figure 1. Architecture of Self Organizing Map.](image)

SOM is an algorithm that proceeds the data in high-dimensional vector space into two-dimensional vector spaces located in nearby locations. Input data are in the form of n-tuples vector which will be
grouped into maximum $m$ groups (called sample vectors). Network output is the group that is closest or similar to the given input. There are several measures of similarity. The minimum Euclidean distance is very popular measurement [13]. A Kohonen method is a competitive learning algorithm in the SOM. The algorithm does not require an activation of the accumulated signal, because there are no effect on the selection of winner neurons. Therefore, the training process error is not calculated at each training iteration. The training process will stop after a certain number of iterations. The SOM steps are as follows [14]

**Step 0** Initiate the weights randomly.

Specify the learning rate and topological neighborhood parameters.

**Step 1** While stopping condition is false. do Steps 2-8.

**Step 2.** For each input vector $x$, do Steps 3-5.

**Step 3.** For each $j$, compute the weights using Euclidean Distance.

$$D(j) = \sqrt{\sum_i (w_{ij} - x_i)^2}$$

**Step 4.** Find index $J$ such that $D(J)$ is a minimum.

**Step 5.** For all units $j$ within a specified neighborhood of $J$, and for all $i$:

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha [x_i - w_{ij}(\text{new})]$$

**Step 6.** Update learning rate.

In the next step, the learning rate value used is $\text{learning\_rate\_new} = \alpha \times b$ where the value of $b$ is between 0 and 1. At the end of the iteration, the value of $\alpha$ will lead to a minimum learning rate value.

**Step 7.** Reduce radius of topological neighborhood at specified times.

**Step 8.** Test stopping condition.

The iteration will stop if the threshold is fulfilled. The threshold value is said to be fulfilled if the parameter value has been fulfilled.

3. **Method**

This study uses secondary data available on the page of the Statistics Indonesia (known in Indonesia as BPS or Badan Pusat Statistik). There are many indicators of people’s welfare, but we focus on health and nutrition indicators. The sample units are the 34 provinces in Indonesia. The steps of analysis are as follows:

**Step 1.** Define the input variables and the number of clusters.

**Step 2.** Normalize the input variables.

**Step 3.** Set the normalized data of input variables as a matrix $n \times p$, where $n$ is the number of provinces in Indonesia and $p$ is the number of variables.

**Step 4.** Perform the SOM analysis from step 0 to step 8 to the input matrix. The result of this step is the cluster of the provinces in Indonesia.

**Step 5.** Calculate the Davies Bouldin Index (DBI) of the clusters obtained in step 4 using the formula [15]

$$DBI = \frac{1}{K} \sum_{i=1}^{K} \max_{i \neq j} (R_{ij})$$

where

$$R_{ij} = \frac{\text{var} (c_i) + \text{var} (c_j)}{\|c_i - c_j\|}.$$ $c_i, c_j$ are the $i$th or $j$th cluster centroid.

In this study, we perform the clustering process from the number of cluster 2 until 6. The optimal number of clusters is determined by considering the least DBI value. The result of the analysis is the provinces clusters. In the same cluster, the provinces have similar characteristics based on the variables used. The provinces clusters are presented in the form of a geographical map that processes
using Quantum Geographic Information System (GIS) software. The characteristics of each cluster are described by observing the bar chart of the average of each variable in each cluster.

4. Result and Discussion

4.1 Data Description

Data are taken from annual book published by Statistics Indonesia [15]. Sixteen variables in Table 1 are used in this study for representing public health and nutrition in Indonesia. The sixteen variables are then grouped based on the condition of public health, public facilities, and infrastructure and services provided by the government. Also, the variables restricted to which related to public health conditions and community owned infrastructure. The research variables are illness, mortality, life expectancy, nutritional conditions, breastfeeding and immunization, clean water facilities and latrines.

Table 1. Variables of public health and nutrition indicator

| Variable | Explanation |
|----------|-------------|
| $x_1$    | Percentage of population that has health complaints |
| $x_2$    | Morbidity Rates in Indonesia |
| $x_3$    | Average duration of illness suffered by residents |
| $x_4$    | Percentage of population aged 0-59 months who received complete immunization |
| $x_5$    | Percentage of infants aged 0-5 months who are given exclusive breastfeeding |
| $x_6$    | Infant mortality rate |
| $x_7$    | Neonatal mortality rate |
| $x_8$    | Life expectancy at birth |
| $x_9$    | The percentage of mothers giving birth to children born alive in the last two years in health services |
| $x_{10}$ | The percentage of toddlers aged 0-59 months according to nutritional status with index BB / U. |
| $x_{11}$ | The percentage of infants aged 0-23 months according to nutritional status with index TB / U |
| $x_{12}$ | The percentage of infants aged 0-59 months according to nutritional status with index TB / U |
| $x_{13}$ | The percentage of infants aged 0-23 months according to nutritional status with index BB / TB |
| $x_{14}$ | Nutritional status of adults> 18 years old based on BMI |
| $x_{15}$ | Percentage of households with clean drinking water sources |
| $x_{16}$ | Percentage of households with privately owned latrines |

4.2 SOM clustering on welfare data

The clustering is proceeded by following all the steps in Section 3. The first step of SOM method is data normalization. The SOM model is obtained by estimating the weights in the model architecture. The number clusters in the SOM model are the same as the number of neurons. The lowest IBD deals with the optimal number of clusters. The IBD values for 2-6 clusters are given in Figure 2.
Figure 2. Plot of DBI values.

The minimum IBD value was obtained at the number of clusters 5. So, the map of provinces consists five clusters. The object (province) will be in one cluster if its Euclid inter cluster distance is minimum at that cluster. Table 2 provides the Euclidean inter cluster of each province in each cluster (neuron) and the location of the provinces in the clusters.

| Neuron                      | 1    | 2    | 3    | 4    | 5    | cluster |
|-----------------------------|------|------|------|------|------|---------|
| Aceh                        | 1.49 | 0.71 | 1.50 | 0.95 | 0.75 | 2       |
| North Sumatera              | 1.38 | 0.42 | 1.97 | 1.39 | 1.39 | 2       |
| West Sumatera               | 0.74 | 0.33 | 1.11 | 1.31 | 1.42 | 2       |
| Riau                        | 0.73 | 0.28 | 1.55 | 0.94 | 0.95 | 2       |
| Jambi                       | 0.97 | 0.19 | 1.68 | 1.09 | 0.93 | 2       |
| South Sumatera              | 0.74 | 0.26 | 2.04 | 1.89 | 1.44 | 2       |
| Bengkulu                    | 1.43 | 0.37 | 1.84 | 1.79 | 1.38 | 2       |
| Lampung                     | 0.47 | 0.09 | 1.11 | 1.02 | 1.15 | 2       |
| Bangka Belitung Island      | 0.21 | 0.64 | 1.96 | 1.97 | 1.94 | 1       |
| Riau Island                 | 0.51 | 1.17 | 2.34 | 1.92 | 2.30 | 1       |
| DKI Jakarta                 | 0.32 | 1.20 | 2.81 | 2.97 | 3.15 | 1       |
| West Jawa                   | 0.33 | 0.28 | 1.59 | 1.42 | 1.70 | 2       |
| Central Jawa                | 0.28 | 0.71 | 1.51 | 1.94 | 2.25 | 1       |
| DI Yogyakarta               | 0.66 | 1.49 | 2.24 | 2.95 | 3.44 | 1       |
| East Jawa                   | 0.23 | 0.46 | 1.28 | 1.23 | 1.63 | 1       |
| Banten                      | 0.60 | 0.19 | 1.00 | 0.90 | 0.98 | 2       |
| Bali                        | 0.50 | 1.05 | 2.52 | 3.15 | 3.09 | 1       |
| West Nusa Tenggara          | 1.76 | 1.45 | 0.43 | 2.00 | 1.96 | 3       |
| East Nusa Tenggara          | 3.70 | 2.89 | 0.75 | 1.33 | 2.24 | 3       |
| West Kalimantan             | 2.72 | 1.85 | 1.34 | 0.28 | 1.38 | 4       |
| Central Kalimantan          | 1.90 | 1.01 | 0.71 | 0.21 | 0.66 | 4       |
| South Kalimantan            | 1.38 | 1.05 | 0.29 | 0.73 | 0.94 | 3       |
| East Kalimantan             | 0.40 | 0.95 | 2.27 | 1.37 | 2.14 | 1       |
| North Kalimantan            | 0.67 | 1.03 | 2.58 | 1.38 | 2.20 | 1       |
| North Sulawesi              | 0.73 | 0.38 | 1.72 | 1.85 | 1.47 | 2       |
| Central Sulawesi            | 1.91 | 0.94 | 1.03 | 0.82 | 0.29 | 5       |
| South Sulawesi              | 1.30 | 1.08 | 1.23 | 0.39 | 1.27 | 4       |
| South east Sulawesi         | 1.25 | 0.59 | 0.80 | 0.97 | 1.03 | 2       |
The results of SOM analysis with five clusters are presented in the map as geographic information systems, so that information is easier to be read and interpreted. The geographical map of the level of people's welfare on health and nutrition indicators are presented in Figure 3. Provinces on the same cluster are displayed in the same color.

![Map of Indonesia with welfare data](image)

**Figure 3. Welfare Map of Provinces in Indonesia.**

Figure 3 shows that the adjacent provinces tend to fall in the same cluster. Cluster 1 is dominated by provinces in Java and Bali. Two nearby provinces in Kalimantan and two islands closed to Sumatra are also included in cluster 1. Cluster 2 is dominated by provinces in Sumatra, but it also includes West Java and Banten. This information demonstrates that the health and nutrition conditions in West Java and Banten are more similar to the provinces in Sumatra island than the provinces in Java. Cluster 3 is dominated by the provinces in Nusa Tenggara. Meanwhile, Cluster 4 is dominated by provinces in Kalimantan and Cluster 5 is dominated by provinces in Papua. The closest provinces to Papua are Maluku and Central Sulawesi, which are also included in Cluster 5. The descriptions of each cluster for each variable are presented in Figure 4-7.
Figure 4. Diagram of variables grouped in (a) illness (b) breastfeeding and immunization.

Figure 5. Diagram of variables grouped in (a) mortality and (b) life expectancy.

Figure 6. Diagram of variables grouped in nutritional conditions.
From Figure 4-7, we can observe the different level of variables in each cluster. The characteristics of each cluster are set up based on the dominant variables in the cluster compared to other clusters. Cluster 1 on the clustering map is red and has 9 provinces. It demonstrates the highest level on variables morbidity rate, average duration of illness suffered by residents, the percentage of mothers giving birth to children born alive in the last two years in health service, percentage of households with privately owned latrines, and percentage of households with clean drinking water sources. It has characteristic in term of lowest neonatal mortality rate and the percentage of infants aged 0-59 months according to nutritional status with index Height/Age. The low percentage of neonatal mortality possibly relates to the high percentage of mother giving birth to children born alive in health service.

Cluster 2 on the clustering map is blue and has 12 provinces. There are no prominent variables in cluster 2, most variables have moderate level. The variables having less than 20% are morbidity rates, average duration of illness suffered by residents, infants aged 0-23 months according to nutritional status with index Height/Age, the percentage of infants aged 0-23 months according to nutritional status with index Weight/Height, and nutritional status of adults > 18 years old based on BMI. However those percentages are almost the same to other clusters.

Cluster 3 on the clustering map is green and has 5 provinces. It shows the highest percentage of population that has health complaints, infants aged 0-59 months according to nutritional status with index Height/Age, completed immunization and exclusive breastfeeding, and relatively high infant mortality rate. As other clusters, the morbidity rate and the average duration of illness suffered are relatively low.

Cluster 4 on the clustering map is yellow and has 3 provinces. It delivers the highest percentage of infants aged 0-23 and 0-59 months according to nutritional status with index Height/Age. The illness, exclusive breastfeeding and complete immunization, mortality and life expectancy factors are similar to those of Cluster 2.

Cluster 5 on the clustering map is pink and has 5 provinces. The population that has health complaints, aged 0-59 who received complete immunization, and mother giving birth to children born alive in health service have the lowest percentage in comparison with the other four clusters. It also demonstrates the low nutritional status of toddler population with respect to Weight/Height, Height/Age, and Weight/Height, and adult with respect to BMI. Cluster 5 has the highest infant mortality rate.

If we compare among clusters, it can be stated that Cluster 1 has the highest illness problem compared to other clusters, but it has the least nutritional problems, the lowest mortality rate, high life expectancy, and the most adequate water and latrines facilities. In general, Cluster 1 has better health and nutrition indicators than others. The results are reasonable since the provinces included in Cluster 1 are provinces in Java, Bali and a small part of Sumatra. Nutritional problems mainly occur in provinces in Clusters 3 and 4, i.e. the provinces of Nusa Tenggara, parts of Kalimantan and Sulawesi. Nusa Tenggara as Papua, also has a serious problem of high infant mortality.
5. Conclusion

In this study, the SOM method has been used to construct a map of provinces in Indonesia based on the health and nutrition indicators with 16 variables. The DBI criterion can be applied to determine the map with suitable number of clusters. The proposed method yields a map with five clusters. The Quantum GIS has displayed the geographical map of provinces in Indonesia based on the health and nutrition indicators. The bar charts of the average percentages of the variables in each cluster are described to explain the characteristics of each cluster. The map shows that the cluster tends to contain the nearby provinces. This evidence indicates that the nearby provinces have similar conditions with respect to the variables of health and nutrition indicators.

Acknowledgments

We acknowledge the financial support to our research provided by DIPA Faculty UNY 2018.

References

[1] Du K L and Swamy MNS 2006 Neural Networks in Soft Computing Frame Work (London: Springer Verlag)
[2] Bacao F, Lobo V, and Painho M 2005 Self-organizing Maps as Substitutes for K Means Clustering, in Lecture Notes in Computer Science Sunderam V S et al.. Editors (Berlin Heidelberg: Springer-Verlag) pp 476–483
[3] Anand S, Vinod V, and Rampure A 2015 International Journal of Applied Engineering Research. 10 64 p 274-281
[4] Hernawati K, Insani N, Bambang S H M, and Waryanto N H, Sahid 2017 Proc. of The 4th Int. Conf. on Research Implementation and Education of Mathematics and Science (AIP Publishing)
[5] Hilmi M N, Wilandari Y, and Yasin H 2015 Jurnal Gaussian 4 1 pp 53 – 60
[6] Dua I, Dominguez M, and Cuadrado AA, Fuertes JJ 2008 Expert. Syst. Appl 34 4 2953–65
[7] Yasir M, Mudassar A, Xi C, and Timo H 2011 Self-Organizing Maps of Nutrition. Lifestyle and Health Situation in the World. WSOM in LNCS 6731. Laaksonen J and Honkela T (Eds.) (Berlin Heidelberg: Springer-Verlag) pp 160–167
[8] Abonyi J, Nemeth S, Vincze C, and Arva P 2003 Computers in Industry 52(3) 221–234
[9] Victor JAS and Lobo 2009 Application of Self-Organizing Maps to the Maritime Environment. V.V. Popovich et al. (eds.). Information Fusion and Geographic Information Systems Lecture Notes in Geoinformation and Cartography. DOI: 10.1007/978-3-642-00304-2_2 (Berlin Heidelberg: Springer-Verlag)
[10] Ningrum E S 2012 Classification of Household Welfare in Malang City with a Bagging Logistic Regression Approach ( Indonesian : Klasifikasi Kesejahteraan Rumah Tangga di Kota Malang dengan Pendekatan Bagging Regresi Logistik (Surabaya :ITS)
[11] Sahitayakti R P dan Fithriasari K 2015 Jurnal Sains dan Seni ITS 4 2 2337-3520 (Print 2301-928X) p D-229–D-235
[12] Wina P S, Elin C, Dana I S, and Handrie N 2016 IEEE Student Conf. on Research and Development (SCOReD)
[13] Siang J 2009 Artificial Neural Networks and Programming using Matlab (Indonesian: Jaringan Syaraf Tiruan dan Pemogramannya Menggunakan Matlab (2nd Ed.) (Yogyakarta: Andi Offset)
[14] Faussset L 1994 Fundamentals of Neural Networks: Architectures, Algorithms, and Applications (New Jersey: Prentice-Hall)
[15] Davies D and Bouldin D 1979 IEEE Trans. Pattern Anal. Mach. Intell. 1 2 pp. 224-227