

Characteristics and segmentation of social problems with kohonen self-organizing maps

Reza Aditya Pratama ^{1,*}, Tiara Shafira ², Faisal Ardiansyah ³, RB Fajriyah Hakim ⁴

Statistics Program, Faculty of Mathematics and Natural Science, Islamic University of Indonesia, Indonesia

¹ rezaadityauui@gmail.com *; ² tiarashafira60@gmail.com; ³ faisalardiansyah000@gmail.com; ⁴ fajriyahakim@yahoo.com

* corresponding author

ARTICLE INFO

Article history

Received November 3, 2016

Revised December 22, 2016

Accepted January 10, 2017

Keywords

Social problem

Clustering

Kohonen

ABSTRACT

Indonesia is a country with a low Human Development Index, it shows the number of quality and healthy standard in Indonesia is still poor. Indonesia also have various social problems such as overcrowding, poverty, unemployment, bad education level. This problem can bring negative impact for our society like increasing of crime rate. For identification phase of social problems and crime, Indonesian government does not integrate social problems which is identified can affect the crime and use descriptive statistics only. Further diagnosis required for cases of social issues. The purpose and benefits of this research is to determine the characteristics of the social problems in Indonesia, introduce and make segmentation using Kohonen Self Organizing Map's algorithm. Hopefully the results of this analysis can helps government for make public policy in general, specifically future policy about social problems in Indonesia. Using Kohonen algorithm effective for visualizing of high-dimensional data by reducing the dimensions of ann-dimensional input into lower dimension while maintaining its original topological relations. Based of clustering result of provinces in Indonesia, it divided into 5 group and each group has similar characteristics.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

Social problems are phenomenon and symptom of reality in social life. Identification of social problems in society is different between one figures with others. Social problems are mismatch between element of culture or society which endanger social life [1]. Social problems is contradict situation of society values, where people agreed to take some action for change those situation [2].

Certainly there are social problem in our community or society like juvenile delinquency, population problems, environmental pollution, as well as other social problems such as poverty, social inequality, etc. The existence of those social problems in community have positive or negative impact for the community itself. One of various impact because social problems is increasing of crime rate [3].

Specially discuss about one of an impact of social problems is the increasing of crime rate. Although it was not main factor, but it affect the psychology of perpetrators. There are various theoretical reviews regarding the cause of crime related to economic issues change crime behavior directly. This theory includes decreasing of economy, comparative deterioration in socio-economic, reduced of opportunities in formal sectors, urbanization and economic growth that potentially lead the integration with poorer. Poverty have a correlation with crime [4], result of psychologists and sociologist's research prove that someone which is life under poverty, vulnerable commit a crime. In poor condition, people will experience anxiety, become very emotional, easily frustrated, even doing suicidal. Poverty also become the causes of crime as alternative in order of living his/her life [4].

Hundred million of the world's population today, and one of the reason incidence of insecurity of their life are not surely terrorism only, but extreme poverty. Therefore, factor of density will make crime, it become worse if socio economic factors does not support community needs.

Population density will become worse if it was not accompanied by new job chance, it causing unemployment. It mean more unemployment will affect more crime [4], [5]. It is evident that crime rate is higher in urban areas than in rural areas. Then, effective method except keeping public security and safety, government should open more jobs for their people, if unemployment decrease so do the crime rate will decrease.

Except density, poverty and unemployment factor, there is another social problem that is educational inequality. Low level of education have significant influence with crime rate. Nowadays, the increasing of human resources quality can be seen from the average of education level in several region in Indonesia. This increase is a result of the growing demand for education to get a job with a better income, due to gain employment in the modern sector is dependent on their education [6]. There is a negative correlation between level of education and the level of crime. First, higher education can bring or obtain legal employment. Secondly, highly educated person would tend to think of a criminal act, because the benefits are too small. So education will indirectly affect crime through increased wages. Someone who is graduates from lower level of education may have a poorer skill than someone which is graduated from high school or university. The analysis conducted by Ehrlich mention that education is important for population, because education helps to determine the benefits from legal or illegal activities [7].

Human Development Index (HDI) generally can show the society quality, it able to measure the achievement of human development life, HDI is built through basic three dimensional approach. The dimensions include long and healthy life; knowledge, and a decent life. The third dimension has a wide sense as related to many factors. To measure the dimension of health, use life expectancy. Furthermore, about the knowledge indicator using literacy rate and average length of school. As for measuring the dimension of decent life used indicator of the ability of purchase number of basic needs as seen from the average amount of spending per capita as the income approach which represents the achievement of development for a decent life.

In 2013 HDI of Indonesia those year do not change from 108th from 187 [8]. Except Singapura (9), Brunei (30), Malaysia (62) and Thailand (89), other ASEAN country such as Myanmar (150), Laos (139), Cambodia (136), Vietnam (121) and Filipina (117) placed in low rank. Indonesia's HDI is 0,684 in 2013, there are upward movement which is not so far from previous year, ie 0,681. This report also highlight the lack of decent jobs especially for young people. That is major challenge in Asia and Pacific. 22% of the unemployment rate in Indonesia is relatively high. Education must be accessible in eastern Indonesia like Papua. These areas often do not experiencing the improvement of living standard due to limited access to basic social services.

In general, crime itself is a behavior that violates the law and social norms, so people will against it. In many cases the crime occurred due to several factors. Factors that causing crime like biological factors, sociologically composed of economic factors (economic system, population, changes of market prices, financial crisis, lack of employment and unemployment), mental factors (religion, literature, newspapers, movies), physical factors such as climatic conditions and others, and personal factors (age, race and nationality, alcohol, war) that whole thing is part of a social problem and it needs to be fix by analyzing the social problems.

In identification stage, the government of Indonesia in Indonesian Crime Statistics 2013 through BPS is focused on the use of the Susenas Model and Statistik Potensi Desa, not integrating with social problems which is identified has an effect on crime. BPS only use descriptive statistics to determine the incidence and prevalence crime.

Developing of computer, software, and research measurement instrument, made research in various scientific fields capable to collect and analyze data set which is size and its dimensions continues to increase.

Author considers that need some method that can solve that problem, especially about social problem in Indonesia. Using Kohonen because it is an effective algorithm for visualizing of high-dimensional data by reducing the dimensions of ann-dimensional input into lower dimension while maintaining its original topological relations. In addition, Kohonen SOMs is a nonparametric

approach requires no assumptions about the distribution of the population. Therefore, researchers will identify and perform clustering for social problems which is particularly affect the rate of criminality.

Based on those background, the problem is how the characteristics and segmentation of social problems in Indonesia using Kohonen algorithm to clustering provinces in Indonesia. The purpose of this paper is to know the characteristics of the social problems in Indonesia, introduce and make segmentation algorithm using Kohonen's Self Organizing Map for provinces based on similar characteristics in terms of the social issues that affect crime rate.

Boundary of this paper, author using data obtained from BPS Indonesia for most recent year in 2013. Author using variables that have a significant effect on crime as the negative impact of social problems in Indonesia that based source of news and previous research. These variables such as population density, the school enrollment rate of children aged more than 15 years, unemployment, poverty, and Human Development Index in 2013. The method of analysis is using descriptive and analysis of algorithms cluster with Kohonen's Self Organizing Map.

2. Literature Review

Variable which is used by author is Human Development Index (HDI), HDI intends to shift the focus of development towards three factors. These three factors are difficult to measure precisely and so 'proxies' deemed to be the best indicators of the level of these targets are chosen to form the indices instead. Furthermore, the proxies chosen are relevant and available indicators in very potential region of study, enabling international comparison. Human Development Report the HDI combines three dimensions, there are:

- A long and healthy life: Life expectancy at birth
- Educational Index Mean years of schooling and Expected years of schooling
- A decent standard of living: GNI per capita (PPP US\$)

The formula to count HDI is shown at equation 1. The index $X_{(i,j)}$ is HDI component index i for region to j . The variable i is 1, 2, 3 and the variable J is 1, 2, ..., k region.

$$HDI = \frac{1}{3} \sum Index X_{(i,j)} \quad (1)$$

Poverty Rate is a percentage of the population under the poverty line. Poverty line is the sum of the food poverty line (FPL) and Non-Food Poverty Line (NFPL) [9]. Residents who have an average monthly per capita expenditure below the poverty line are categorized as poor. Food Poverty Line (FPL) is the minimum food expenditure, which is equivalent to 2100 kilocalories per capita per day. Non-Food Poverty Line (NFPL) is the minimum requirement for housing, clothing, education and health. Package of non-food commodities basic needs represented by 51 types of commodities in urban and 47 rural commodity.

$$Poverty\ Line = FPL + NFPL \quad (2)$$

FPL = Food Poverty Line

NFPL = Non Food Poverty Line

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[\frac{z-y_i}{z} \right]^{\alpha} \quad (3)$$

Based on the equation 3, the variable α is 0. The variable z is the poverty line. The variable y_i is the average expenditure per capita of population in a month under poverty line ($i=1,2,3,\dots,q$), $y_i < z$. the variable q is count of population that live under poverty line and variable n is the population.

Population Density is the number of inhabitants per unit area [9]. Ceude population density shows number of people for every kilometer square area. The total area is the total land area in an administration. School Enrollment is the proportion of school children at age level of education in the age group in accordance to the education level.

Unemployment figures are the percentage of the population who is unemployed compared with the productive population of an area. Unemployment is a person belonging to productive and for a period does not work, and willing to accept the job, and looking for work.

Segmentation is an attempt to classify objects of observation into several groups. Objects of observation that are in one group are generally more homogeneous than objects of observation in other group.

Descriptive statistics are methods relating to the collection and presentation data to become useful information. Descriptive statistics describe an overview of the object using sample or population it can also be presented as Pareto diagrams and tables [10].

Clustering used to analyze different group of data, similar like classification, but the grouping has not been defined before executable with data mining tool. Usually using a neural network or statistical methods. Clustering divide the items into several groups based on the data mining tool. The principle of clustering is to maximize the similarity between members of the class and minimizing inter-cluster similarity. Clustering can be performed on the data that has some attributes. Many clustering algorithms require the function of distance to measure the similarity between data. Also required a method for normalizing various attributes of the data. Clustering algorithms used to determine segment of data sets into sub groups which is equal [11].

Self-Organizing Map (SOM) is a form of unsupervised neural network that produces a low (typically two) dimensional representation of input space of one set of training samples. Self-Organizing Map (SOM) network is a neural network based method for dimension reduction. SOM can learn from complex, multi-dimensional data and transform them into a map of fewer dimensions, such as a two-dimensional plot. The two-dimensional plot provides an easy-to-use graphic aluser interface to help decision-maker visualize the similarities between consumer preference patterns. For example in AT & T data set, there are 68 variables. It would be difficult to visually classify consumers based on all these attributes because grouping must be done in a 68-dimensional space. By mapping the information contained in 68-variable set into a two-dimensional plot, one can visually group customers with similar preferences into clusters. These relationships can be translated into an appropriate type of structure that genuinely represents the underlying relationships between market segments. Hence, SOM networks can be used to build a decision support system for marketing management. It is designed to capture topologies and hierarchical structures of higher dimensional input spaces. Unlike most neural network applications, SOM performs unsupervised training, i.e., during the learning (training) stage, SOM processing input units in network and adjusts their weights primarily based on the lateral feedback connections. The nodes in the network converge to form clusters to represent groups of nodes with similar properties. A two-dimensional map of the input data is created in such a way that the orders of the interrelationships among objects are preserved. The number and composition of clusters can be visually determined based on the output distribution generated by the training process [12].

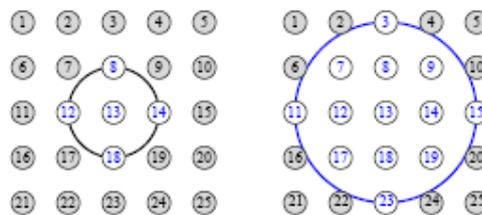


Fig. 1.Environment Illustration [12]

Fig. 1 illustrates environment concept. Left one figure show environment with $R=1$ it surrounding point 13. Right one show environment from $R=2$. Environment topology generally using 3 type of grid there is grid, hexagonal and random topology. Distance function with Euclidean.

$$D(w_j, x_n) = \sqrt{\sum_i (w_{ij} - x_{ni})^2} \quad (4)$$

This is several Relative research that associated with this paper, previous research becomes very important in order to know relationship between previous and this research. There are research that related to social issues nor analysis methods that used previously. One research result is clustering based on crime location as impact of social problems and notice socio-economic factors [13]. Another research show count of population and unemployment has an effect with crime, while the number of industry and poverty indirectly effect on crime [4], [5].

Pinheiro separate company into several segmentation and take a policy for monitoring certain consumer groups using Kohonen's Self Organizing Maps [14]. Kiang give conclusion that show segmentation of SOM algorithm and result demographic profile is better than K Means clustering [15].

3. Method

This research used secondary data obtained from website of Badan Pusat Statistika that accessible through <http://bps.go.id>. The data that used are the factors that affect the crime rate in Indonesia like population density, school enrolment rate of children aged more than 15 year, unemployment rate, poverty and HDI [16]. This research conducted on January 2015. Latest data show 2013 data only and covering a whole Indonesia territory. This research using descriptive statistics and Kohonen Self Organizing Map's.

4. Result and Discussion

This research using descriptive analysis that able to describe characteristics of provinces in Indonesia based on several variables that affect crime rates based on previous studies and using of Kohonen algorithm make segmentation.

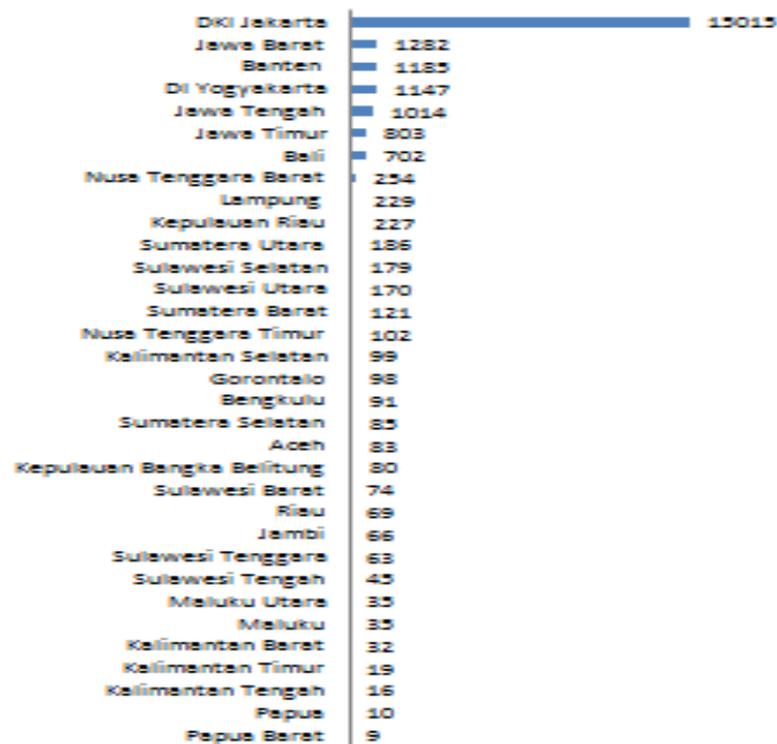


Fig. 2. Characteristics of Population Density in Indonesia

Based on the Fig. 2, Jakarta is in first place with 15.015 inhabitants/km² and then followed by West Java, Banten, DIY, and Central Java. All provinces in Java have high population density but lowest one is West Papua with the density 9 inhabitants/ km² only. There is inequality of population density between Java and another islands. It is able to make another social, economic, or cultural problems if government made no effort for population distribution.

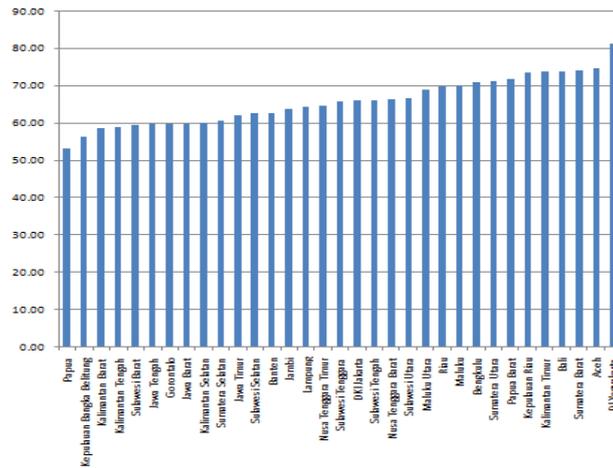


Fig. 3. School enrolment rate of children aged more than 15 year (APS)

The lowest school enrollment rates (APS) is Papua, followed by Bangka Belitung, West Kalimantan, Central Kalimantan and West Sulawesi. Fig. 3 shows a big proportion of children who is more than 15 year cannot feel the education facility. Special Region of Yogyakarta has a numeric percentage of children aged above 15 years of schooling highest followed Aceh, West Sumatra, Bali and East Kalimantan. Java Island, Jakarta occupies the first position with APS is 66.09 only.

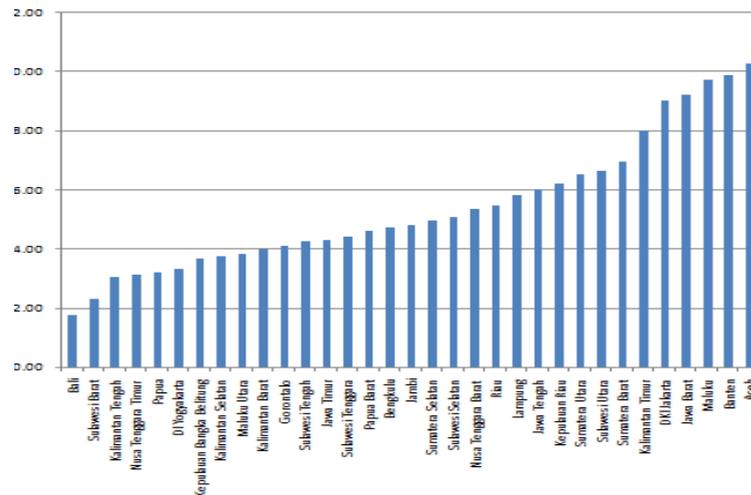


Fig. 4. Unemployment Rate in Indonesia

Fig. 4 illustrates that lowest unemployment rate is Bali, while highest unemployment rate is Aceh. Big five which is have a high unemployment rate include 3 province in Java Island like Banten, West Java and Jakarta when those province have a high population density too. Government should preparing a new jobs to decrease unemployment rate. If decrease unemployment program succeed, it can increase the economy in society.

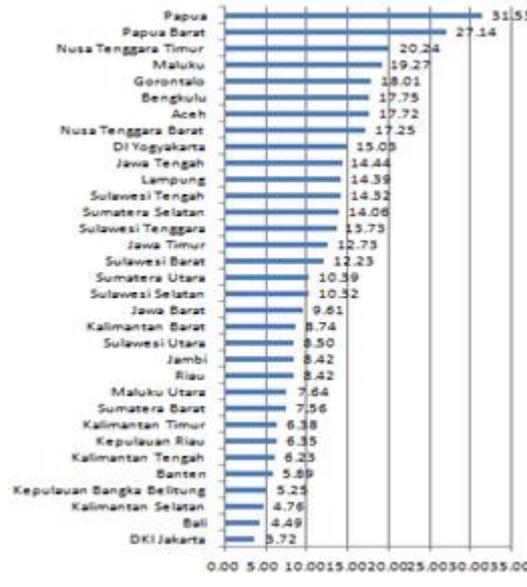


Fig. 5. Poverty in Indonesia

Based on Fig. 5, highest poverty rate, and top five province with highest poverty rates are mostly in East Indonesia such as Papua, West Papua, Maluku and Gorontalo. NTT locates in central region and capital city of Indonesia has lowest poverty rate followed by Bali, South Kalimantan, Bangka Belitung and Banten.

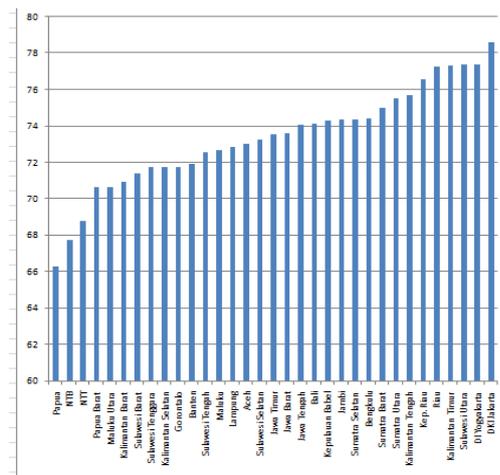


Fig. 6. Human Development Index

Papua, West Nusa Tenggara, East Nusa Tenggara, West Papua and North Maluku province with a low HDI, while Jakarta, Yogyakarta, North Sulawesi, East Kalimantan and Riau is a province with a high HDI. Papua province occupies the lowest position compared to other provinces.

Variables of this research related with social, education, population and economy and this is the result of clustering using Kohonen algorithm.

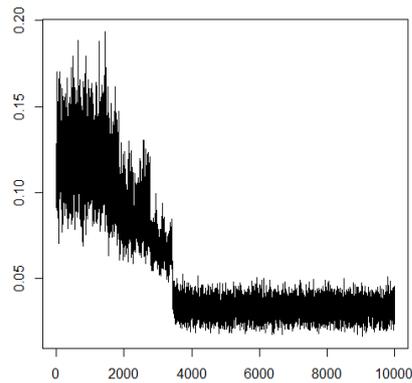


Fig. 7. Training Progress

Based on Fig. 7, author using 10.000 iteration but in 4000th iteration show the data begin to constant. Based on output with R, average distance of this clustering is 0,1285484.

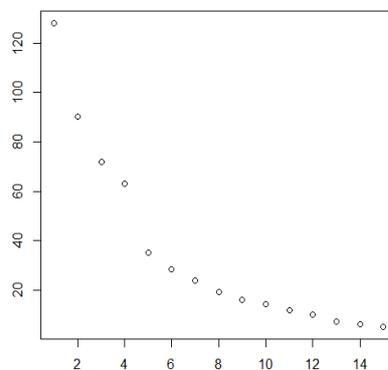


Fig. 8. Within Cluster Sum of Squares

Fig. 8 shows the within cluster sum of squares (WCSS), if researcher pick one cluster only then the curve will show sharp decline, so when researcher take 2, 3, or 4 cluster then the sharpness still visible. More clusters formed so the members will also less.

Fig. 9 show the clustering using Kohonen that divide population density, APS, HDI, unemployment and poverty line. To clarify the cluster result, the researcher needs to use boundaries. Researchers have determined will make 5 clusters and here are the results:

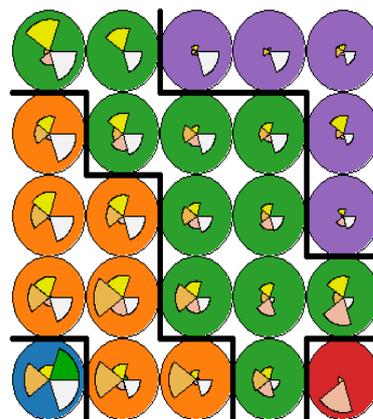


Fig. 9. Kohonen SOM Algorithm Result

After obtaining these results, researcher will explain results of clustering using Kohonen Self Organizing Maps. Blue circle shows high density, unemployment, HDI and APS, and lower poverty. Orange circle show high APS levels, unemployment, poverty, medium HDI and low density. Green

circle has a low density, APS, unemployment, poverty and medium HDI. The red circle has a high poverty rate, but the density, APS, unemployment and HDI are low. Purple circle show medium HDI, but density, APS, unemployment and poverty are low. After obtaining these results, the researchers made a clustering between provinces in Indonesia and some of the variables that have been described previously.

Table 1. Cluster Result

| Cluster | Cluster Result |
|---------|--|
| I | DKI Jakarta |
| II | Aceh, Sumatra Utara, Sumatra Barat, Riau, Kepulauan Riau, Jawa Barat, Banten, Kalimantan Timur, Sulawesi Utara, Maluku |
| III | Jambi, Sumatra Selatan, Bengkulu, Lampung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, NTB, NTT, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Gorontalo, Papua Barat. |
| IV | Papua |
| V | Kepulauan Bangka Belitung, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Barat dan Maluku Utara. |

Based on Table 1, DKI Jakarta have high density, unemployment, HDI and APS, and low poverty. Cluster II, Aceh, Sumatra Utara, Sumatra Barat, Riau, Kepulauan Riau, Jawa Barat, Banten, Kalimantan Timur, Sulawesi Utara dan Maluku have same characteristics that high APS levels, unemployment, poverty, medium HDI and low density. Cluster III, there are Jambi, Sumatra Selatan, Bengkulu, Lampung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo dan Papua Barat low density, APS, unemployment, poverty and medium HDI. Papua itself the only member of cluster IV that have high poverty rate, but the density, APS, unemployment and HDI are low. Kepulauan Bangka Belitung, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Barat dan Maluku Utara are become cluster V where HDI, density, APS, unemployment and poverty are low.

5. Conclusion

Summary of this research that various problem about population, education, social etc. can be fixed with implementation of Kohonen Self Organizing Map's algorithm as one of alternative method for clustering which is every member of cluster have similar characteristics. This research, author make 5 cluster which is containing several provinces that have similarity about social problem than can influent crime rate.

For next research, it will be better if adding other variables especially social problem in society that influent crime rate, adding spatial variables can make a better cluster. Government should be doing their job for national development, it will help decreasing social problem and minimizing crime rate.

References

- [1] D. Wolke, W. E. Copeland, A. Angold, and E. J. Costello, "Impact of bullying in childhood on adult health, wealth, crime, and social outcomes," *Psychol. Sci.*, vol. 24, no. 10, pp. 1958–1970, 2013.
- [2] D. Weinberg, "On the social construction of social problems and social problems theory: A contribution to the legacy of John Kitsuse," *Am. Sociol.*, vol. 40, no. 1–2, pp. 61–78, 2009.
- [3] R. Martinez Jr, R. Rosenfeld, and D. Mares, "Social disorganization, drug market activity, and neighborhood violent crime," *Urban Aff. Rev.*, vol. 43, no. 6, pp. 846–874, 2008.
- [4] A. Šileika and J. Bekerytė, "Theoretical issues of relationship between unemployment, poverty and crime in sustainable development," *J. Secur. Sustain. Issues*, vol. 2, no. 3, 2013.
- [5] M. J. Lin, "Does unemployment increase crime? Evidence from US data 1974–2000," *J. Hum. Resour.*, vol. 413–436, no. 2, pp. 413–436, 2008.

-
- [6] D. Suryadarma, A. Suryahadi, and S. Sumarto, "The Measurement and Trends of Unemployment in Indonesia: The Issue of Discouraged Workers," Jakarta, 2005.
- [7] P. R. Ehrlich, A. N. Ehrlich, and J. P. Holdren, *Human Ecology*. San Francisco: Freeman Company, 1973.
- [8] E. Sofilda, P. Hermiyanti, and M. Hamzah, "Determinant Variable Analysis of Human Development Index in Indonesia (Case for High and Low Index at Period 2004-2013)," *OIDA Int. J. Sustain. Dev.*, vol. 8, no. 9, pp. 11–27, 2015.
- [9] Badan Pusat Statistik, "Statistik Kriminal 2013," *Statistik Kriminal 2013*, 2013. [Online]. Available: <https://www.bps.go.id/publication/2013/12/23/0ca75fd275be1372ea0ac3ec/statistik-kriminal-2013.html>. [Accessed: 15-Jan-2015].
- [10] P. J. Bickel and E. L. Lehmann, *Descriptive statistics for nonparametric models*. Boston, MA: Springer, 2012.
- [11] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2005.
- [12] H. Demuth and M. Beale, *Neural Network Toolbox*, User's Gui. Natick, MA: The MathWorks, Inc, 2000.
- [13] R. Sujatha and D. D. Ezhilmaran, "An Adaptive Method for Analyzing and Predicting the Crime Locations by Means of AMABC and ARM," *J. Theor. Appl. Inf. Technol.*, vol. 59, no. 1, pp. 45–56, 2014.
- [14] C. A. R. Pinheiro, A. G. Evsukoff, and N. F. F. Ebecken, "Neural Network to identify and prevent bad debt in Telephone Companies," in *IEEE 2005 International Conference on Data Mining: Workshop on Data Mining Case Studies*, 2005, pp. 125–139.
- [15] M. Y. Kiang, M. Y. Hu, and D. M. Fisher, "An extended self-organizing map network for market segmentation—a telecommunication example," *Decis. Support Syst.*, vol. 42, no. 1, pp. 36–47, 2006.
- [16] Badan Pusat Statistik, "BPS-Statistics Indonesia." 2015.