Signifikan: Jurnal Ilmu Ekonomi
Volume 6 (2), October 2017
P-ISSN: 2087-2046; E-ISSN: 2476-9223
Page 289– 300

Understanding Structure of Poverty Dimensions in East Java: Bicluster Approach

Budi Yuniarto[1], Robert Kurniawan[2]
[1 2] Department of Computational Statistics, Institute of Statistic, Jakarta
[1] byuniarto@stis.ac.id, [2]robertk@stis.ac.id

Abstract

Poverty is still become a main problem for Indonesia, where recently, the view point of poverty is not just from income or consumption, but it's defined multidimensionally. The understanding of the structure of multidimensional poverty is essential to government to develop policies for poverty reduction. This paper aims to describe the structure of poverty in East Java by using variables forming the dimensions of poverty and to investigate any clustering patterns in the region of East Java with considering the poverty variables using biclustering method. Biclustering is an unsupervised technique in data mining where we are grouping scalars from the two-dimensional matrix. Using bicluster analysis, we found two bicluster where each bicluster has different characteristics.
Keywords: poverty, multidimensional poverty, biclustering.

Abstrak

Kemiskinan masih menjadi permasalahan utama di Indonesia, dimana saat ini kemiskinan tidak lagi dipandang dari sisi pendapatan atau konsumsi, tapi sekarang kemiskinan didefinisikan secara multidimensional. Pemahaman akan struktur kemiskinan multidimensi sangat penting bagi pemerintah untuk mengembangkan berbagai kebijakan untuk mengentaskan kemiskinan. Tulisan ini bertujuan untuk menjelaskan struktur kemiskinan di provinsi Jawa Timur dengan menggunakan variabel-variabel pembentuk dimensi kemiskinan serta untuk menyelidiki setiap pola pengelompokkan yang terbentuk dengan menggunakan metode bicluster. Analisis bicluster adalah suatu teknik unsupervised dalam data mining yang mengelompokkan data dalam suatu matriks dua dimensi. Hasil penelitian mengungkapkan terdapat dua bicluster dengan setiap bicluster memiliki karakteristik yang berbeda.
Kata kunci: kemiskinan, kemiskinan multidimensi, biclustering.

Received: January 31, 2017; Revised: June 11, 2017; Approved: June 20, 2017

INTRODUCTION

Poverty is still become the main problem for developing countries, including for Indonesia. Previously, poverty was initially considered a person's inability to obtain sufficient level of income to meet basic living standards (Citro, 1995). According to Badan Pusat Statistik (Indonesia Statistics Agency), absolute poverty is determined based on the failure to meet minimum basic needs (food and non-food). The value of the minimum basic needs is defined as the poverty line. Word Bank (2002) states that poverty is the loss of welfare. Bapenas (2004) in BPS (2016) defines poverty as a condition that the person is not able to meet their basic rights to maintain and develop a more dignified life. Sabiti and Effendi (2017) said that poverty is one of the problems in economy. But recently, the development of the study of poverty has shifted significantly where the structure of poverty is seen not just from income or consumption, but defined multidimensionally. Poverty incidence can be defined by many variables which arranged in some dimensions. In the multidimensional concept, poverty should be viewed from various dimensions such as the dimensions of education, health, quality of life, democracy and freedom of access to economic society (Sen, Equality of What?, 1980) (Sen, Development as Freedom, 1999).

According to Adi (2005), multidimensional poverty can be viewed in three dimensions, i.e the dimension of the micro, dimension of the mezzo and dimension of the macro. The macro dimensions described the gap between rural and urban areas which are the source of poverty, the mezzo-dimensional occur because of the weakness of social trust in communication and organization, and the micro dimension occurs because of the mentality of almost instantaneous and fast paced.

Meanwhile, *Multidimensional Poverty Index* (Budyantoro, et.al, 2013), describe three poverty dimension for Indonesian households, i.e dimension of health, dimension of education, and dimension of living standard. Multidimensional Poverty Index (MPI) is an index to measure poverty by identifying cross-dimensional deprivation and show the number of poor people in multidimensional (www.undp.org). Gangga and Otok (2013) and Yuniarto and Kurniawan (2016) use these poverty dimensions on their research with some modification.

Understanding of the structure of multidimensional poverty is essential to government, where it usefull to develop policies for poverty alleviation. Poverty

http://journal.uinjkt.ac.id/index.php/signifikan
DOI: 10.15408/sjie.v6i2.4769

alleviation programs should not only done to reduce the absolute poverty dimension, which measured by Head Count Index, index of depth of poverty, and index of poverty severity.

East Java province was chosen as the area of this study because in the year 2014, East Java economic growth reached 5.86 percent, is second highest in Java after economic growth of Jakarta Capital City which was 5.95 percent. On the other hand, the number of poor people in East Java in the same year reached 4,748 million people (the province with the highest number of poor people in Java) (Badan Pusat Statistik, 2015). This shows that high economic growth in East Java was not in line with the decrease of poverty. Head Count Index in Indonesia on March 2016 was about (10.86 percent) or about 28.01 million peoples. It was decreased by 0.58 million people (11.22 percent) compared to March 2015. Jawa Timur, as one of largest province in Indonesia, had Head Count Index about 12,05 percent on March 2016, decreased compared to March 2015 which was 12,34 percent.

This paper aims to describe the structure of multi-dimensional poverty in East Java by using indicator variables of the poverty dimensions. Also this paper aims to investigate any clustering patterns in the region of East Java with considering the poverty variables using biclustering analysis. This research is also expected to contribute to local governments in which they have an understanding the identification of poverty pattern by region and poverty dimension variables simultaneously. So, poverty alleviation programs will be more effective and targeted.

METHOD

Biclustering

Biclustering is an unsupervised technique in data mining where we are grouping scalars from the two-dimensional matrix. Biclustering is different from clustering, since biclustering performs simultaneous clustering on both dimensions of a data matrix. Using of this technique will uncover submatrices with elements that manifest a similar behavior. These submatrices consist of a subset of columns which are considered to determine the assignment of rows.

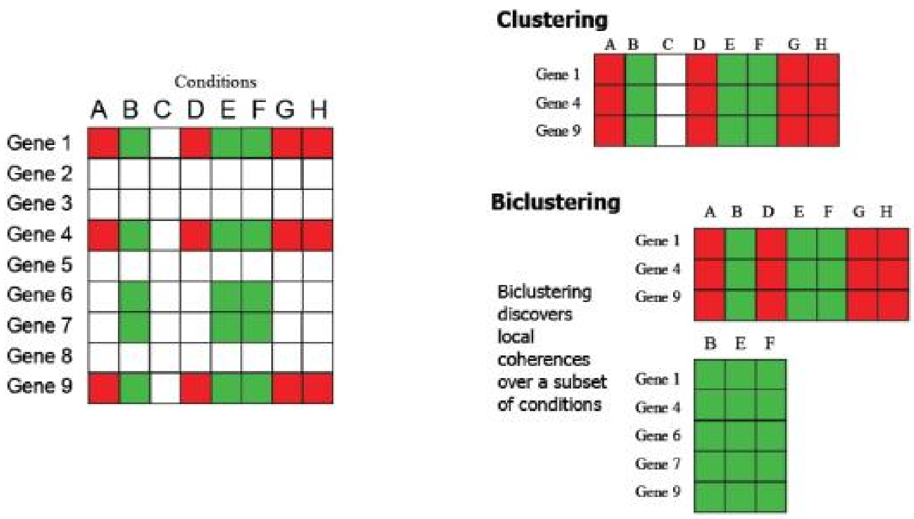
Idea of Bicluster technique first developed in 1972 by J.A. Hartigan (Mina, 2010). He proposed a model and a simultaneous technique for clustering cases and variables on voting dataset under the name Direct Clustering. In 2000, Cheng and

DOI: 10.15408/sjie.v6i2.4769

Church applied biclustering technique in the field of bioinformatics, where biclustering introduced to microarray gene expression data (Cheng & Church, 2000). Since then biclustering technique become more popular in various field.

Visualization of biclustering procedure, compared to clustering procedure, can be seen in Figure 1. On the left is the gene array matrix or data matrix used, the clustering outcome is on the right top panel, and the biclusters is on the right bottom panel. As biclustering become more popular, numerous biclustering algorithm have been proposed. Different algorithms, use different types of scoring schemes or constraints for bicluster detection. Some of the algorithms (Bozdag, Kumar, & Catalyurek, 2010) are Cheng and Church algorithm, HARP algorithm by Yip, Correlated Pattern Biclusters (CPB) algorithm by Bozdağ, Bimax algorithm by Prelic, Statistical Algorithmic Method for Bicluster Analysis (SAMBA) algorithm, and Order Preserving Submatrices (OPSM) algorithm.

Figure 1. Comparison of clustering and biclustering outcome



Source: (Mina, 2010)

To understand how biclustering works, assumed a data matrix *n* x *m* with set of rows X = {x_1, …, x_n} and set of columns Y = {y_1, …, y_m}. A bicluster defined as a submatrix k x s with set of rows I = {i_1, i_2, …, i_k} (where I ⊂ X and k ≤ n) and set of columns J = {j_1, j_2, …, j_s} (where J ⊂ Y and s ≤ m) and b_{ij} the expression value of gene i under condition j. A bicluster B = (I, J) is composed of a subset of rows I ⊂ X and a subset of columns in J ⊂ Y, where all a_{ij} , for i ∈ I and j ∈ J, are expected to fit to a

http://journal.uinjkt.ac.id/index.php/signifikan
DOI: 10.15408/sjie.v6i2.4769

predetermined target pattern possibly with small deviations. The homogeneity of each bicluster, depends on the biclustering method used in analysis. Many algorithms have been implemented using di_erent methods and scoring schemes to discover them.

Cheng and Church Algorithm

Cheng and Church (CC) algorithm (Cheng & Church, 2000) tries to find maximal biclusters with a high similarity score, so CC algorithm categorized as a greedy algorithm. This similarity score is called MRS (Mean Residue Score), and a collection of rows and columns is called a bicluster if this score is below a certain level δ defined before by user. The residue ε_{ij} is defined as following:

$$\varepsilon_{ij} = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ} \qquad (1)$$

where a_{iJ} , a_{Ij} and a_{IJ} are the mean of the entries in of row i, column j, and the entire bicluster, respectively, for $i \subset I$, and $j \subset J$. CC algorithm considered minimization of MSR as their objective. Then we define the mean residue score of a bicluster (I,J) as:

$$H(I,J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} \left(a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}\right)^2 \qquad (2)$$

Data and variables

This research use data from National Socio-economic Survey (Susenas) in 2013. The observation unit is regency/city in East Java province and the variables to be used in this research are taken from poverty structural model of Gangga and Otok (2013) as follows.

1. Head Count Index / P0 (Y_1)
2. Poverty gap index / P1 (Y_2)
3. Poverty severity index / P2 (Y_3)
4. The percentage of unemployment of poor people age 15 and older (X_1).
5. The percentage of poor people age 15 and older who work in the agricultural sector (X_2).
6. Percent of percapita expenditure on food (X_3).
7. The percentage of poor people age 15 and older who do not complete primary school (X_4).
8. Literacy Rate of poor people aged 15-55 years old (X_5).
9. Mean year school multiplied (X_6).
10. The percentage of women using contraceptives in poor households (X_7).

DOI: 10.15408/sjie.v6i2.4769

11. Percentage of toddlers in poor households whom the delivery process assisted by health personnel (X_8).
12. The percentage of poor households with per capita floor area ≤ 8 m.sq (X_9).
13. The percentage of poor households using clean water sources for drinking (X_{10}).
14. The percentage of poor households with latrines owned or jointly owned (X_{11}).
15. Life expectancy (X_{12}).

RESULT AND DISCUSSION

There are many definitions of poverty dimensions. Naraya *et al* (1999), says that poverty is composed of many interlocking dimensions. First, as a base, poverty is seen as a lack of food. Second, poverty has important psychological dimensions, such as powerlessness, voicelessness, dependency, shame, and humiliation. Third, poverty means lack of access to basic infrastructure: roads, transport, and water. Fourth is education, and the fifth is health. As well as the last, the poor rarely speak of income, they are more focused on how to manage what they have.

One measure for measuring multidimensional poverty is the Multidimensional Poverty Index (MPI). MPI was first published in the Human Development Report of 2010 prepared by UNDP. MPI constituent dimensions are the same as constituent dimension of the Human Development Index, i.e the dimension of Health, Education, and Living Standard. In this research, we used modified MPI structures, where variables Y1, Y2, and Y3 are absolute poverty measurements, X1, X2, and X3 are indicator variables of Economic dimension. Meanwhile variables X4, X5, and X6 are indicator variables of Human Resources dimension, and X7 to X12 are indicator variables of Health dimension.

Results

We use RcmdrPlugin.BiclustGUI version 1.0.6 (De Troyer and Martin, n.d) to run the CC algorithm. RcmdrPlugin.BiclustGUI is a plugin package for R-commander. The package is a Graphical User Interface (GUI) in which several biclustering methods can be executed, followed by diagnostics and plots of the results.

In data preparation, variables X6, X7, X8, X10, X11, and X12 multiplied by -1 to change the direction of their association with Head Count Index or P0.

http://journal.uinjkt.ac.id/index.php/signifikan
DOI: 10.15408/sjie.v6i2.4769

The parameter we used are $\delta = 1$ and $\alpha = 2$. Using this parameter, we find two bicluster. First bicluster has size 31 rows and 11 columns, and second bicluster has size 7 rows and 8 columns, where rows are observation units and columns are variables. The membership of each bicluster is given in Table 1.

From the heatmap of two bicluster, we can see that both of bicluster are very homogen. Membership of first bicluster, consist of 31 regencies/cities, which are Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Tuban, Lamongan, Gresik, Sampang, Pamekasan, Sidoarjo, Mojokerto, Jombang, Madiun, Ngawi, Bojonegoro, Sumenep, Kota Blitar, Kota Malang, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, and Kota Batu, which they had same characteristics in 11 variables, i.e variable X7, X8, X9, X10, and X11, which they are related to Health dimension, then variables X4, X5, and X6, which they are related to Human Resources dimension, meanwhile variables which are related to Economy dimension is only X1. The interesting thing is variable Y1 (Head Count Index) was not include in the first bicluster.

Table 1. Detail of each bicluster

Bicluster 1			Bicluster 2
Variables: X7, X8, X9, X10, X11, X4, X5, X6, X1, Y2, Y3			Variables:X7, X8, X9, X4, X6, Y2, Y3
Regency/city:			Regency/city:
1. Pacitan	13. Tuban	24. Sumenep	1. Probolinggo
2. Ponorogo	14. Lamongan	25. Kota Blitar	2. Pasuruan
3. Trenggalek	15. Gresik	26. Kota Malang	3. Nganjuk
4. Tulungagung	16. Sampang	27. Kota Pasuruan	4. Magetan
5. Blitar	17. Pamekasan	28. Kota Mojokerto	5. Bangkalan
6. Kediri	18. Sidoarjo	29. Kota Madiun	6. Kota Kediri
7. Malang	19. Mojokerto	30. Kota Surabaya	7. Kota Probolinggo
8. Lumajang	20. Jombang	31. Kota Batu	
9. Jember	21. Madiun		
10. Banyuwangi	22. Ngawi		
11. Bondowoso	23. Bojonegoro		
12. Situbondo			

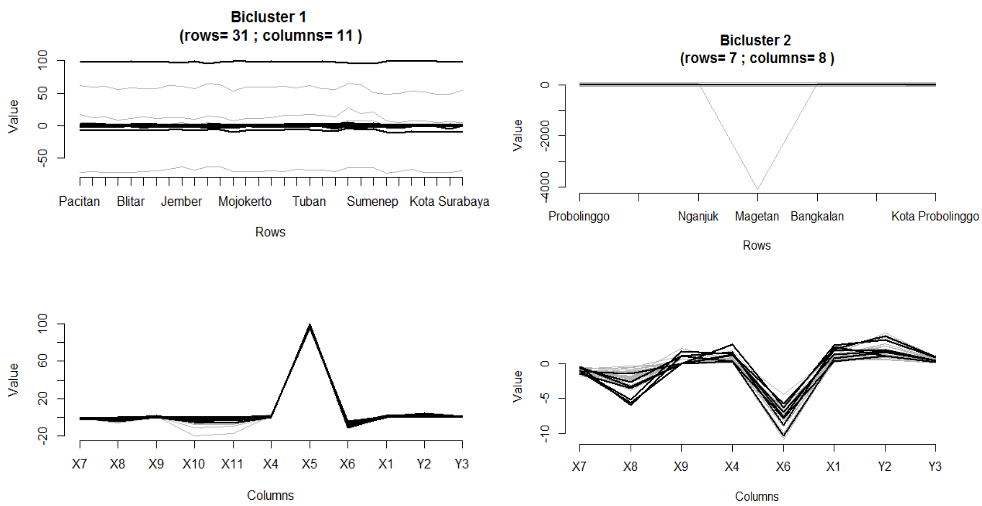
Then the second bicluster consist of 7 regencies/cities, which are Probolinggo, Pasuruan, Nganjuk, Magetan, Bangkalan, Kota Kediri, and Kota Probolinggo, and they had same characteristic in 8 variables, i.e X7, X8, and X9 (which are related to Health dimension), X4 and X6 (which are related to Human Resources dimension), and variables which are related to Economic dimension is only X1 include in the bicluster.

DOI: 10.15408/sjie.v6i2.4769

Same as first bicluster, variable Y1 (Head Count Index) was not enter the second bicluster.

Figure 2 represent parallel coordinate plots where we can check the pattern similarity for observations and variables. From Figure 2, we can see that both bicluster generally doesn't have extreme value. We can't say the bicluster categorize as poor regencies/cities, they just have same characteristics in certain variables. From two bicluster, Bicluster 1 is most interesting where bicluster 1 is a group of regencies/cities which are have same characteristic dominantly in Health dimension (except for life expectancy) and Human Resources dimension.

Figure 2. Parallel coordinates plots of bicluster 1 and bicluster 2



Discussions

The absence of variable Y1 in both bicluster implied that the pattern of Head Count Index was not similar to the pattern of poverty determinants in bicluster 1 and and bicluster 2, where most of them are categorized as Health dimension and Human Resources dimension. This was in line with Gangga and Otok (2013) who found that the health dimension and human resources dimension has no significant relationship with absolute poverty in East Java. This is slightly different from Wagle (2005) who said, in fact, the theoretically supported dimensions of poverty are all related, and able to reveal more realistically the overall poverty status of households. Theoretically, the economic well-being, one of several dimension of poverty, helps transform capability into other activities indicative of living conditions. This implied that the poverty

http://journal.uinjkt.ac.id/index.php/signifikan
DOI: 10.15408/sjie.v6i2.4769

alleviation program conducted by government should not only aimed to reduce poverty in economic dimension.

Meanwhile, to make an interpretation of each bicluster, Table 2 shows the mean and standar deviation of variables from each bicluster. As shown in the table 2 below, bicluster 2 was contrast compared to bicluster 1. For example, the mean of variable X7 of bicluster 2 lower than mean of X7 for East Java, while the mean of variable X7 of bicluster 1 is higher than the East Java.

Table 2. Mean and standard deviation of each bicluster

	Bicluster 1				**Bicluster 2**				**East Java**	
	mean		std.dev		mean		std.dev		mean	std.dev
X7	1.11		0.34	X7	0.9		0.34	X7	1.07	0.34
X8	1.96	*	0.92	X8	3.94	*	1.67	X8	2.32	1.32
X9	0.16		0.46	X9	0.56		0.72	X9	0.24	0.53
X10	1.5	*	1.6	X4	1.26		0.82	X10	2.48	3.74
X11	1.29	*	1.56	X6	7.64		1.54	X11	2.15	3.29
X4	0.59		0.43	X1	1.61		0.85	X4	0.71	0.57
X5	98.02	*	1.02	Y2	2.1		1.11	X5	97.61	1.52
X6	7.81		1.56	Y3	0.5		0.3	X6	7.78	1.54
X1	0.81	*	0.4					X1	0.96	0.59
Y2	1.73		0.89					Y2	1.8	0.93
Y3	0.39		0.23					Y3	0.41	0.24

* = significance at level $\alpha = 0.05$ compare to East Java (mean)

Member of bicluster 1 are regencies and cities which have percentage of toddlers in poor households whom the delivery process assisted by health personnels, percentage of poor households using clean water sources for drinking, percentage of poor households with latrines owned or jointly owned and percentage of unemployment of poor people age 15 and older were significantly lower than East Java (mean). These three variables can be considered as household's individual characteristic household's individual characteristic has direct impact to poverty. The individual characteristics of the household are the cause of poverty. The main focus of attention to be taken in poverty reduction efforts in East Java is to improve the characteristics of individual households. Meanwhile the Literacy Rate of poor people aged 15-55 years old is higher than East Java, means that regencies/cities in bicluster 1 is poorer in some health dimension and a economic dimension but they were better in literacy rate as indicator of Human Resources dimension.

DOI: 10.15408/sjie.v6i2.4769

Bicluster 2 is a group of regencies/cities which have similarity in less variables compared to bicluster 1, which is have percentage of toddlers in poor households whom the delivery process assisted by health personnel's higher than East Java. This results show that the pattern in poverty dimension in East Java which formed the biclusters were only occurs in several indicator variables, which are include in economic dimension, health dimension and human resources. But, bicluster analysis could not any pattern of regencies or cities which involve absolute poverty. This is slightly different compared to Manurung and Santoso (2015) who found three cluster of regencies/cities using cluster analysis based on poverty determinants variables. However, the patterns found in the bicluster will be useful for governments to formulate poverty alleviation policies at various dimensions appropriately.

CONCLUSION

The algorithm biclustering Cheng and Church (CC) applied to East Java's poverty data produced two bicluster. First bicluster consisted of 31 regencies / cities with 11 variables and second bicluster consists of 7 regencies / cities with eight variables. The biclusters indicates there are a different pattern between the Head Count Index (Y1) with the pattern of variables of other dimensions in each bicluster.

What needs to be considered is bicluster 1, where there are some variables of poverty indicators are lower than the East Java level significantly. Cheng and Church algorithm is an algorithm that produces a constant bicluster types (for row and column), so further research is needed using the other algorithm that generates the coherent bicluster type to more understand the structure of poverty dimension and other poverty dimensions may be included.

REFERENCES

Adi, I. R. (2005). Kemiskinan Multidimensi (Multidimensional Poverty). *Makara seri Sosial Humaniora, 9*(1): 91-100.

Badan Pusat Statistik. (2016). *Penghitungan dan Analisis Kemiskinan Makro Indonesia 2016 (Measurement and Analysis of Indonesian Macro Poverty 2016)*. Jakarta: Badan Pusat Statistik.

http://journal.uinjkt.ac.id/index.php/signifikan
DOI: 10.15408/sjie.v6i2.4769

Bozdag, D., A.S. Kumar., & U.V.Catalyurek. (2010). Comparative analysis of biclustering algorithms. *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology.* New York.

Budyantoro, S., V. Fanggidae., W. Saputra., A. Maftuchan., & D.R. Artha. (2013). Multidimensional Poverty Index (MPI): Konsep dan Pengukurannya di Indonesia (Multidimensional Poverty Indes: Concept and Measurement in Indonesia). *PRAKARSA Economic Working Paper*. Jakarta: PRAKARSA.

Cheng, Y., & G. M. Church. (2000). Biclustering of expression data. *ISMB-00 Proceedings* (pp. 93 - 103). *International Conference on Intelligent Systems for Molecular Biology*.

Citro, C. F. (1995). *Measuring Poverty: A New Approach.* Washington DC: National Academic Press.

De Troyer, E. & O. Martin. (n.d.). *RcmdrPlugin.BiclustGUI: 'Rcmdr' Plug-in for Biclustering. R package version 1.0.6.* Retrieved from CRAN: https://CRAN.R-project.org/package=RcmdrPlugin.BiclustGUI.

Gangga, A., & B.W. Otok. (2013). Pemodelan Kemiskinan di Jawa Timur dengan Structural Equation Model – Partial Least Square (Poverty Modelling in East Java Using Structural Equation Model-Partial Least Square). *Statistika, 1*(2): 121-130.

Junaidi, Z., D. Wisadirana., & S. Kanto. (2015). Analisis Determinan Kemiskinan di Jawa Timur (Determinant Analysis of Poverty in East Java). *Wacana, 18*(1), 41-50.

Manurung, F. A., & D.B. Santoso. (2015). Pemetaan Kemiskinan Kabupaten/Kota di Provinsi Jawa Timur (Poverty Mapping of District in East Java). *Working Paper.* Malang: Universitas Brawijaya.

Mina, E. (2010). Applying Biclustering to understand the molecular basis of phenotypic diversity. (*Unpublished Thesis*). Utrecht: Utrecht University.

Narayan, D., R. Patel., K. Schafft., A. Rademacer., & S. Koh-Schulte. (1999). Can Anyone Hear Us? Voices from 47 Countries. *Poverty Group, PREM*. World Bank.

DOI: 10.15408/sjie.v6i2.4769

Sabiti, M.B. & J. Effendi. (2017). Islamic Microfinance and Its Impact on Poverty Reduction of Two Village Models in Bogor, Indonesia. *Signifikan: Jurnal Ilmu EKonomi*. Vol. 6 (1): 87-102.

Sen, A. (1980). Equality of What? *McMurrin S Tanner Lectures on Human Values, 1*.

Sen, A. (1999). *Development as Freedom*. Oxford: Oxford University Press.

Wagle, U. (2005, November). Multidimensional Poverty Measurement with Economic Well-being, Capability, and Social Inclusion: A Case from Kathmandu, Nepal. *Journal of Human Development, 6*(3): 321-330.

World Bank Institute. (2002). *Dasar-dasar Analisis Kemiskinan (The Basics of Poverty Analysis). Edisi Terjemahan.* Jakarta: Badan Pusat Statistik.

Yuniarto, B., & Kurniawan, R. (2017). Modified multiblock partial least squares path modeling algorithm with backpropagation neural networks approach. *Statisticts and its Applications: Proceedings of the 2nd International Conference on Applied Statistics (ICAS II), 2016. 1827.* AIP Publishing.

http://journal.uinjkt.ac.id/index.php/signifikan
DOI: 10.15408/sjie.v6i2.4769