IOP Conf. Series: Materials Science and Engineering **807** (2020) 012017 doi:10.1088/1757-899X/807/1/012017

Application of SKATER and Ward's methods in grouping Indonesian provinces based on monthly expenditure per capita of food commodity groups

Yudi Setyawan[1], Rokhana Dwi Bekti[1] and Firman Isarlin[1]

[1]Department of Statistics, Faculty of Applied Sciences, IST AKPRIND Yogyakarta, Jl. Kalisahak 28 Komplek Balapan, Yogyakarta 55222, Indonesia

Corresponding e-mail: setyawan@akprind.ac.id

Abstract. Clustering is a fundamental data mining instrument that intends to find inherent cluster structure in data. Spatial clustering methods are usually used to assess the demographic data characterization. This study aims to classify provinces in Indonesia based on monthly expenditure per capita according to food commodity groups by using Ward's and Spatial 'K'luster analysis by tree edge removal (SKATER) methods and to identify a better classification between the two methods. The variables of this research constitute percentages of expenditure per capita for 14 groups of food commodities of 34 provinces in Indonesia during March 2018. The results of the first analysis (excluding outliers) revealed that SKATER method produced standard deviation rasio of 0.236, better than Ward's method that produced standard deviation rasio of 0.370. However, from the second analysis (including outliers), the outcomes showed that the Ward's method generated standard deviation rasio of 0.170, better than SKATER method that delivered standard deviation rasio of 0.199. Moreover, it can be concluded that the second analysis is better than the first analysis because it produced smaller standard deviation ratios based on the Ward's and SKATER methods contrasted with the first one.

1. Introduction

People's welfare is always an interesting topic to discuss. As in every country, the main goal in development is to improve people's welfare. In Indonesia, people's welfare is one of the state goals stated in the Preamble of the 1945 Indonesian Constitution [1]. One important thing to note in order to identify the society welfare is the expenditure per capita. The expenditure per capita is the cost incurred for the consumption of all household members for a month divided by the number of household members [2]. The proportion of expenditure for food and non-food can reveal the pattern of household consumption. The composition of household expenditure can be used as a measure to assess the level of economic well-being of the population [2-4]. The lower the percentage of expenditure on food on total expenditure, the better the level of welfare. Moreover, household expenditures are differentiated according to food and non-food groups. Changes in one's income will affect the shift in spending patterns. The higher the income, the higher the expenditure on non-food. Thus, expenditure patterns can be used as a tool to measure the level of welfare of the population, where changes in composition are used as an indication of changes in welfare levels [2].

A method is needed to describe the percentage of monthly expenditure per capita according to food commodity groups in each province in Indonesia so that it can be classified in several groups. The statistical method that can be used is cluster analysis. Johnson and Wichern in Suhaeni [5], states that, cluster analysis is a multivariate analysis method to group the n objects into m clusters ($m \leq n$) based on their characteristics. Grouping is based on the nature of similarities or the nature of dissimilarities between objects. Cluster analysis is important for investigating the number of clusters of natural data in several areas. Clustering data is a complex assignment including the choice between a wide range of methods and strategies, parameters and performance measurements, with implications in some true issues [6]. Different clustering methods have been proposed [7-9]. However, it is very hard to choose the method best suited to the type of data [10-14].

Clustering techniques are divided in two classes namely hierarchical and non-hierarchical methods. Hierarchical clustering methods have two different sub-classes, agglomerative and divisive approaches. Non-hierarchical clustering methods are divided in four sub-classes namely partitioning, density-based, grid-based and other approaches [15].

Agglomerative clustering treats every data point as a singleton cluster and then progressively, step by step unions two most similar clusters, called linkages, until all data points have been converged into a single final cluster [16]. Several methods used to choose pairs of clusters to merge such as single linkage, complete linkage, average linkage, centroid, and Ward's methods.

In this research, Ward's method was used because it is the only agglomerative clustering procedures that relies upon the sum of squares criterion, resulting groups minimize within-group dispersion at each binary combination [17]. In Ward's method, the similarity of objects within a cluster is not a single measure of similarity, but the sum of squares across all variables in a cluster [18].

In many cases, spatial clusterings had been applied to classify spatial data such as socio-demographic and geographic data [19-21]. The researchers used a spatial clustering namely the spatial 'K'luster analysis by tree edge removal (SKATER) method. The SKATER method was introduced by Assunção et al. [22] and Bekti [23]. This method uses an algorithm that transforms regional data into partition charts. This method partitions locations that are not neighboring and do not have similar characteristics. The results of Assunçãu works [22] showed that SKATER is an efficient method that produced good quality results and is significantly faster. In this research both methods would be applied to classify 34 provinces in Indonesia based on monthly expenditure per capita according to food commodity groups and the results would be compared.

2. Methodology

2.1. Data Collection Method

The data in this study are percentages of monthly expenditure per capita according to food commodity groups on the total monthly expenditure per capita in 34 provinces in Indonesia during March 2018. Data used in this study are data sourced from the official website of the Indonesian Central Statistics Agency [2].

2.2. Research Variables

The variables in this study constitute 14 groups of food commodities on monthly expenditure per capita of 34 provinces in Indonesia during March 2018. The data units are in percent (%). The variables are shown in Table 1.

Table 1. Research Variables

Variables	
Grains (X_1)	Fruits (X_8)
Tubers (X_2)	Oil and Coconuts (X_9)
Fish/Shrimp/Squid/Shellfish (X_3)	Beverage Ingredients (X_{10})
Meat (X_4)	Spices (X_{11})

IOP Conf. Series: Materials Science and Engineering **807** (2020) 012017 doi:10.1088/1757-899X/807/1/012017

Eggs and Milk (X_5)	Other Food Ingredients (X_{12})
Vegetables (X_6)	Processed Food and Beverage (X_{13})
Beans (X_7)	Cigarettes and Tobacco (X_{14})

2.3. Stages of Research

The first stage in this research is to identify spatial patterns and descriptions which is carried out to see the characteristics of the research data used and to look at the description of research data based on 34 provinces in Indonesia. Detecting outliers data is done by looking at the z-score. If there is no outlier or outlier problem can be overcome, then continue to the stages in point c. If there is an outlier in the data that cannot be overcome, then do two types of analyzes namely first and second analyses, then compare them. The first analysis is as in the stages c, but eliminate the outlier data. The second analysis is as in point c using all data, included outliers.

Multi-collinearity test was conducted next. If there is multi-collinearity, then do the main component analysis. Otherwise, do not perform it. After that, cluster analysis was performed with Ward's method by using squared Euclidean distance and Ward's method. The cluster number was determined by using RMSSTD (Root Mean Squared Standard Deviation) criterion. Then, the each formed cluster was interpreted.

The cluster analysis was conducted by using the spatial 'K'luster analysis by tree edge removal (SKATER) method with several stages which are determining weighting factor,where the weight used in this study is binary or queen contiguity, and identifying spatial autocorrelation. The identification of spatial autocorrelation was using Moran's I statistics. Then, the results of clustering obtained from the Ward's and SKATER methods were compared by comparing their standard deviation ratios. Lastly, conclusion was made based on the results.

3. Results and discussions

3.1. Identifying Spatial Patterns and Descriptions

3.1.1. Descriptive analysis

The desription of the data used in this research consisted of the mean, standard deviation, minimum, and maximum values (in percents) of the 14 variables and depicted as in Table 2.

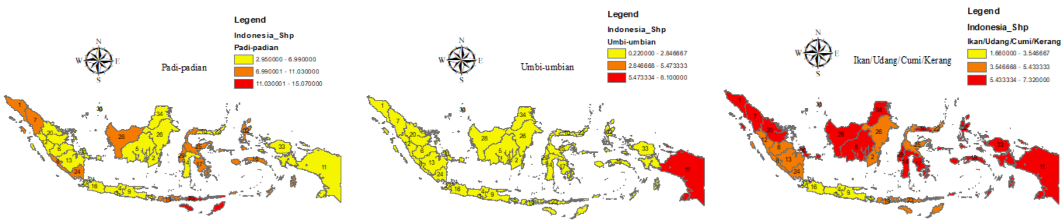
Table 2. Descriptive Analysis of the Data

Variable	Mean	StDev	Minimum	Maximum
X_1	6.600	2.091	2.949	15.072
X_2	0.729	1.333	0.225	8.099
X_3	4.996	1.473	1.664	7.320
X_4	1.919	0.661	0.560	3.409
X_5	2.809	0.348	2.015	3.389
X_6	3.950	0.984	2.438	6.703
X_7	0.836	0.241	0.437	1.548
X_8	2.440	0.527	1.366	4.157
X_9	1.300	0.298	0.688	1.861
X_{10}	1.629	0.355	0.901	2.551
X_{11}	1.021	0.218	0.567	1.430
X_{12}	0.901	0.202	0.549	1.298
X_{13}	15.008	2.674	8.470	19.803
X_{14}	5.967	1.196	2.941	7.774

IOP Conf. Series: Materials Science and Engineering **807** (2020) 012017 doi:10.1088/1757-899X/807/1/012017

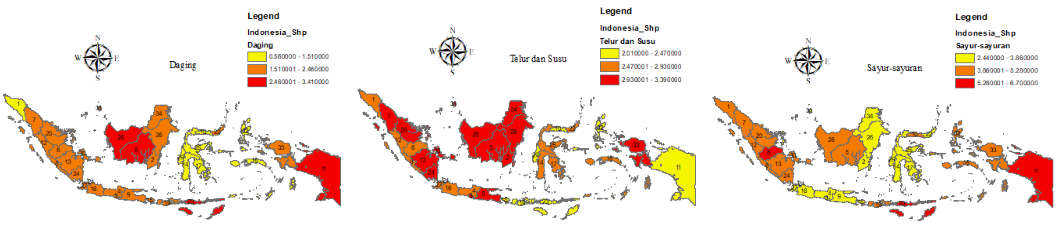
3.1.2. Spatial patterns

Spatial patterns based on 14 variables are as shown as follow in Figure 1, Figure 2, Figure 3, Figure 4 and Figure 5.



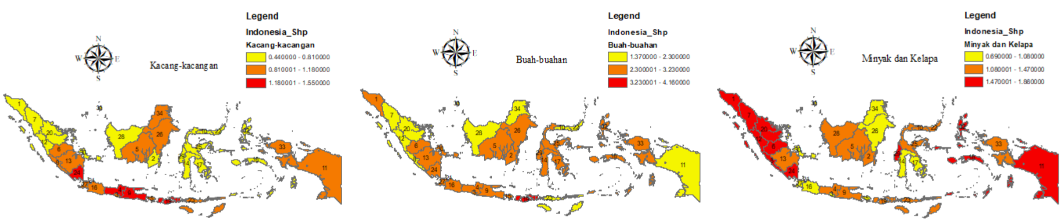
Grains Tubers Fish/shrimp/squid/shellfish

Figure 1. Spatial patterns of grains, tubers, and fish/shrimp/squid/shellfish variables



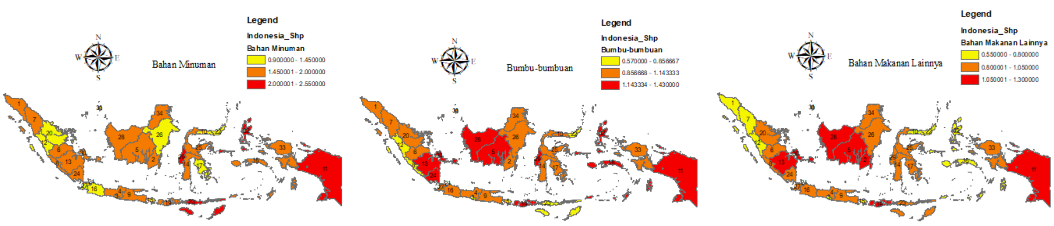
Meat/egg Milk Vegetables

Figure 2. Spatial patterns of meat/egg, milk, and vegetable variables



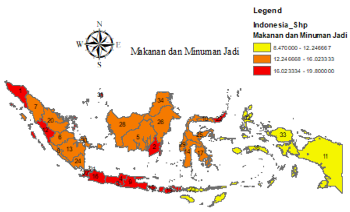
Beans Fruits Oil/coconut

Figure 3. Spatial patterns of bean, fruit, and oil/coconut variables

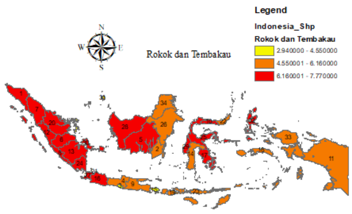


Beverage Spice Other food ingredient

Figure 4. Spatial patterns of beverage, spice, and other food ingredient variables



Processed food and beverage



Cigarette and tobacco

Figure 5. Spatial patterns of processed food and beverage, cigarette and tobacco variables

3.2. Outliers Detection

In this study, outliers were detected by standardizing or changing each research variable into the z-score. Research data that have z-scores $\leq$ -3.00 or z-scores $\geq$ 3.00 are categorized as outliers. Based on the analysis carried out, it was found that the data categorized as outliers are as in Table 3.

Table 3. Outlier Observation

Variable	Observation number
X_1 (grains)	23 (East Nusa Tenggara)
X_2 (tubers)	11 (Papua)
X_8 (fruits)	19 (West Nusa Tenggara)

To overcome the outlier problem, several data transformations were carried out. The results of the seven data transformations are shown in Table 4.

Table 4. Data Transformation Results

No.	Transform	Result	Variable-Observation number
1	Natural logarithm (ln)	*Outlier*	(X_3-11)
2	Squared root	*Outlier*	(X_1-23) and (X_2-11)
3	Logarithm	*Outlier*	(X_1-23) and (X_2-11)
4	Arcsin	*Outlier*	(X_1-23), and(X_2-11)
5	Inverse	*Outlier*	($X_1 - 21$)
6	Inverse squared	*Outlier*	(X_1-21), (X_2-2), (X_3-3), (X_4-32), (X_8-28), (X_9-21), (X_{10}-21), (X_{13}-11), and (X_{14}-3)
7	Inverse squared root	*Outlier*	(X_1-21), (X_2-11), (X_3-3), (X_4-32), and (X_{14}-3)

Because there are still outliers in the research data even though seven types of transformation have been carried out, the next step to deal with outliers is to eliminate provinces that contain outliers. But in this study all variables were included in the analysis. Therefore, the researchers decided to carry out two types of analyses namely first and second analyses. The first analysis excluded provinces containing outlier values while the second analysis included provinces with outlier values. Then, provinces containing outliers would be identified. Finally, the results would be compared to get the better method for both analyses.

3.3. First and Second Analyses

3.3.1 Multicollinearity test

In this study, multicollinearity test was identified from the Pearson correlation coefficients. It is said that there is multicollinearity if there is a Pearson correlation coefficient greater than 0.6. In the first analysis, it can be concluded that there is multicollinearity which is characterized by the Pearson correlation coefficients greater than 0.6, namely the correlation coefficients between X_1 and X_{19}, X_1 and X_{10}, and X_9 and X_{14}. In the second analysis, there is multicollinearity in the research data, this is indicated by the Pearson correlation coefficients greater than 0.6, namely the correlation coefficients between X_1 and X_{10}, X_6 and X_9, X_6 and X_{10}, X_9 and X_{10}, and X_{11} and X_{12}. The next step was performing main component analysis in order to eliminate the correlation through the transformation of the origin variable to the new variable that is not correlated.

For the main component analysis, the new main component is formed by searching for the cummulative proportion greater than 80%. In the first analysis, based on the cummulative proportions obtained, the cummulative proportion of W_5 has a value of 0.867 or 86.7% which is greater than 80%, so that components W_1, W_2, W_3, W_4, and W_5 can explain 86.7% of the total proportion variance. Furthermore, those components would be used in the cluster analysis. In the second analysis, based on the cummulative proportion obtained, the cummulative proportion of W_5 has a value of 0.827 or 82.7%, therefore components W_1, W_2, W_3, W_4, and W_5 would be used in the cluster analysis.

For the cluster analysis using Ward's method, the first analysis where the cluster amount was determined, the number of clusters used was determined by finding the smallest RMSSTD value. Based on the calculation, the RMSSTD value of the 2-cluster grouping is 0.233, while the RMSSTD of the 3-cluster grouping is 0.235. Therefore the 2-cluster grouping is a better choice. Based on the dendrogram in Figure 6, the cluster membership for 2-cluster grouping is shown in Table 5.

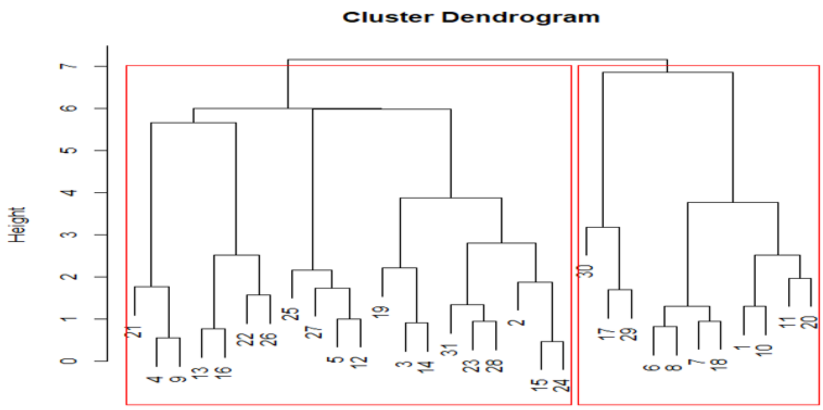


Figure 6. Dendrogram of 2-cluster grouping using Ward's method

Table 5. Cluster membership for first analysis using Ward's method

Cluster	Number of members	Members
1	11	Aceh, Jambi, North Sumatera, Bengkulu, Gorontalo, West Sumatera, Maluku, Riau, North Sulawesi, North Maluku, and West Papua
2	20	South Kalimantan, Special Region of Yogyakarta, Central Java, Central Kalimantan, East Java, South Sumatera, South Sulawesi, Bali, West Java, Southeast Sulawesi, Special Capital Region of Jakarta, Lampung, Central Sulawesi, East Kalimantan, Banten, West Kalimantan, West Sulawesi, Bangka-Belitung Islands, Riau Islands, and North Kalimantan

Profiling or characteristics was then performed. The first cluster consists of 11 provinces. In this cluster, among 14 variables there are 7 commodities having high levels and other 7 commodities having low levels of percentage of expenditure for food commodities. This result is similar to the second cluster consisting of 20 provinces, there are 7 commodities having high levels and other 7 commodities having low levels of percentages. Most provinces in cluster 1 have low levels of welfare and most provinces in cluster 2 have high levels of welfare. In second analysis where cluster amount was determined, based on the calculation of RMSSTD, the results showed that the RMSSTD value of 2 clusters is 0.274, while the RMSSTD 3 cluster is 0.263. Therefore the 3-cluster grouping is more preferable. Based on the dendrogram in Figure 7, the cluster membership for 3-cluster grouping is shown in Table 6.

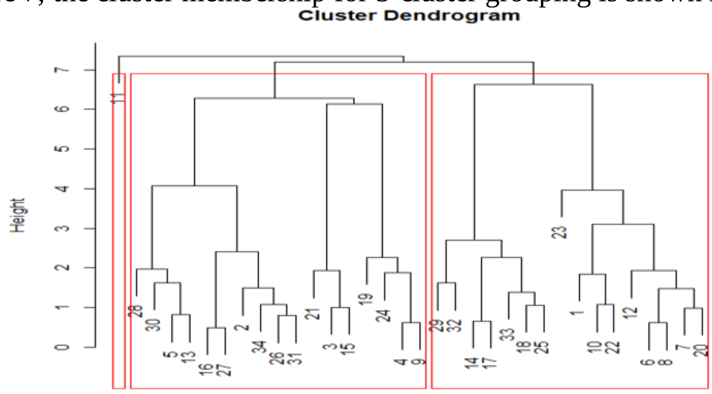


Figure 7. Dendrogram of 3-cluster grouping for second analysis using Ward's method

Table 6. Cluster membership for second analysis using Ward's method

Cluster	Number of members	Members
1	16	Aceh, Jambi, North Sumatera, Bengkulu, Gorontalo, West Sumatera, South Sulawesi, Southeast Sulawesi, Maluku, Riau, North Sulawesi, East Nusa Tenggara, Central Sulawesi, West Sulawesi, North Maluku, and West Papua
2	17	South Kalimantan, Special Region of Yogyakarta, Central Java, Central Kalimantan, East Java, South Sumatera, Bali, West Java, West Nusa Tenggara, Special Capital Region of Jakarta, Lampung, East Kalimantan, Banten, West Kalimantan, Bangka Belitung Islands, Riau Islands, and North Kalimantan
3	1	Papua

The first cluster consists of 16 provinces with medium percentages of expenditure for food commodities. The second cluster consists of 17 provinces with low percentages of expenditure for food commodities. The third cluster consists of one province with a high percentage of expenditure on food commodities. Most provinces (56.25%) in cluster one have low levels of welfare. In contrary, most provinces (58.81%) in cluster two have high levels of welfare.

3.3.2. Spatial autocorrelation test (by Moran's I test) and SKATER method
For the first analysis in Moran's test, based on the results of Moran's I test, it can be concluded that variables W_1, W_2, W_3, W_4 and W_5 have spatial autocorrelations, which is indicated by each of these variables having a p-value of less than 0.05. The results showed that the RMSSTD value for 2-cluster grouping was 0.213, while the RMSSTD value for 3-cluster grouping was 0.211. Therefore the 3-cluster grouping is more preferable. By using SKATER method, the minimum spanning tree partition and the members of each cluster for the first analysis are shown in Figure 8 and Table 7.

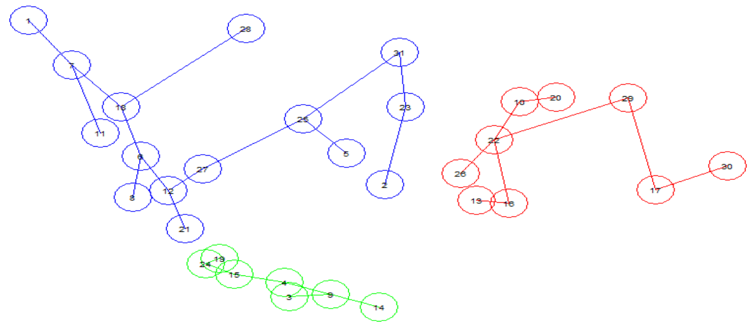


Figure 8. 3-cluster grouping using MST partition for first analysis

Table 7. Cluster membership for first analysis using SKATER method

Cluster	Number of members	Members
1	9	Gorontalo, South Sulawesi, Southeast Sulawesi, Maluku, North Sulawesi, Central Sulawesi, West Sulawesi, North Maluku, and West Papua
2	7	Special Region of Yogyakarta, Central Java, East Java, Bali, West Java, Special Capital Region of Jakarta, and Banten
3	15	Aceh, South Kalimantan, Central Kalimantan, Jambi, North Sumatera, Bengkulu, West Sumatera, South Sumatera, Riau, Lampung, East Kalimantan, West Kalimantan, Bangka Belitung Islands, Riau Islands, and North Kalimantan

The first cluster consists of 9 provinces with low percentages of expenditure for food commodities. The second cluster consists of 7 provinces with high percentages of expenditure for food commodities. The third cluster consists of 15 provinces with moderate percentages of expenditure on food commodities. Most provinces (66.67%) in cluster one have moderate levels of welfare, all provinces in cluster two have high levels of welfare, some provinces (26.67%) in cluster 3 have low levels of welfare.

In the second analysis in the Moran's test, the results are obtained that the variables W_1, W_2, W_4, and W_5 are said to have spatial autocorrelation, which is indicated by each variable having a p-value of less than 0.05. The results obtained are that the RMSSTD value for the 2-cluster grouping is 0.262, while the RMSSTD value for the 3-cluster grouping is 0.259. Therefore the 3-cluster grouping is more preferable.

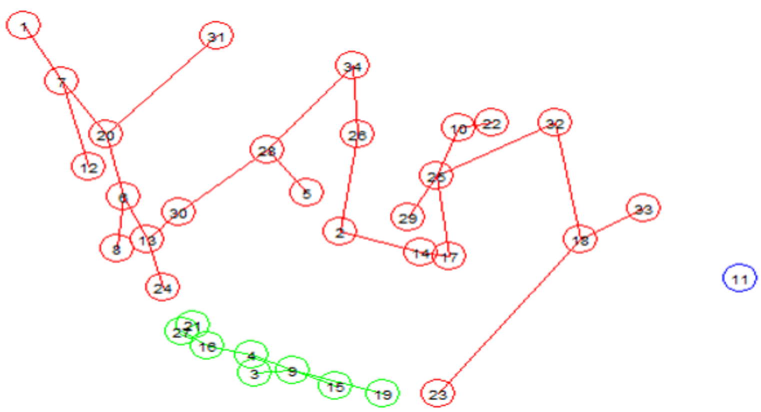


Figure 9. 3-cluster membership using MST partition for second analysis

Table 8. Cluster membership for second analysis using SKATER method

Cluster	Number of members	Members
1	25	Aceh, South Kalimantan, Central Kalimantan, Jambi, North Sumatera, Bengkulu, Gorontalo, West Sumatera, South Sumatera, South Sulawesi, Southeast Sulawesi, Maluku, Riau, North Sulawesi, East Nusa Tenggara, Lampung, Central Sulawesi, East Kalimantan, West Kalimantan, West Sulawesi, Bangka Belitung Islands, Riau Islands, North Maluku, West Papua, and North Kalimantan
2	8	Special Region of Yogyakarta, Central Java, East Java, Bali, West Java, West Nusa Tenggara, Special Capital Region of Jakarta, and Banten
3	1	Papua

The first cluster consisted of 25 provinces with moderate percentages of expenditure for food commodities. The second cluster consists of 8 provinces with high percentages of expenditure for food commodities. The third cluster consists of 1 province with low percentage of expenditure on food commodities. Some provinces (40%) in cluster one have high levels of welfare, most provinces (87.5%) in cluster two have high levels of welfare, one province (100%) in cluster 3 has a low level of welfare (see Figure 9 and Table 8).

3.4. Comparison Between Methods and Analysis

The standard deviation ratio values between the different methods and analysis is shown in Table 9 below.

Table 9. Standard Deviation Ratio Values between Analysis and Method

Method	Standard deviation	
	First Analysis	Second Analysis
Ward	0.37	0.18
SKATER	0.23	0.20

For the first analysis, the SKATER method is much better than Ward's method because it has a smaller standard deviation ratio than Ward's method. However, for the second analysis Ward's method gave a slightly better result than the SKATER method. Furthermore, from the results of the study, it can be concluded that the second analysis is better than the first analysis because it has a smaller standard deviation ratio value using both the Ward's and SKATER methods.

From this research it turned out that the SKATER and the Ward's methods do not always give satisfied result and this case supports the conclusion in [10-14]. To improve the result, modified SKATER method can be applied [24]. Modified SKATER technique is performed by modifying the computation of weights and the partition procedure to oblige the homogeneity within and the heterogeneity between clusters.

4. Conclusion

Based on the results discussed in previous section, it can be concluded that:

1. From the first analysis, the Ward's method produced 2 clusters where the first cluster consisted of 11 provinces with high welfare levels while the second cluster consisted of 20 provinces with low welfare levels. While the SKATER method obtained 3 clusters where the first cluster numbered 9 provinces with moderate welfare levels, the second cluster amounted to 7 provinces with high welfare

levels, and the third cluster amounted to 15 provinces with low welfare levels. In this case the SKATER method is better because it has less standard deviation ratio than that of the Ward's method.

2. In the second analysis the Ward's method produced 3 clusters where the first cluster numbered 16 provinces with moderate welfare levels, the second cluster numbered 17 provinces with high welfare levels, and the third cluster numbered 1 province with a low welfare level. Whereas the SKATER method obtained 3 clusters where the first cluster numbered 25 provinces with moderate welfare levels, the second cluster numbered 8 provinces with high welfare levels, and the third cluster amounted to 1 province with a low welfare level. In this case the Ward's method is slightly better than the SKATER method.
3. Based on the two analyses above, it can be concluded that the second analysis is better than the first analysis because it generates smaller standard deviation ratios for both the SKATER and Ward's methods.

References

[1] Alwi W and Hasrul M 2018 Analisis klaster untuk pengelompokan kabupaten/kota di propinsi Sulawesi Selatan berdasarkan indikator kesejahteraan rakyat *Jurnal MSA* **6**(1) pp 35-42.
[2] Anonimous 2018 *Consumption expenditure of population of Indonesia by province, March 2018.*
[3] Zezzaa A, Carlettoa C, Gennarib P and Jolliffea D 2017 Food counts measuring food consumption and expenditures in household consumption and expenditure surveys (HCES) Introduction to the special issue *Food Policy* **72** pp 1-6.
[4] Sununtnasuk C and Fiedler JL 2017 Can household-based food consumption surveys be used to make inferences about nutrient intakes and inadequacies? a Bangladesh case study *Food Policy* **72** pp 121-31.
[5] Suhaeni 2018 Perbandingan hasil pengelompokan menggunakan analisis cluster berhirarki, K means cluster, dan cluster ensemble (studi kasus data indikator pelayanan kesehatan ibu hamil). *Jurnal Media Infotama* **14**(1) pp 31-8.
[6] Rodriguez M Z, Comin C H, Casanova D, Bruno O M, Amancio D R and Costa LD 2019 Clustering algorithms: a comparative approach. *PLoS ONE* **14**(1).
[7] Jing L, Ng M K and Huang JZ 2007 An entropy weighting K-means algorithm for subspace clustering of high-dimensional sparse data *IEEE Transactions on Knowledge and Data Engineering* **19**(8) pp 1026–41.
[8] Suzuki R and Shimodaira H 2006 Pvclust: an R package for assessing the uncertainty in hierarchical clustering *Bioinformatics* **22**(12) pp 1540–2.
[9] Camastra F and Verri A 2005 A novel kernel method for clustering *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(5) 801–5.
[10] Rujasiri P and Chomtee B 2009 Comparison of clustering techniques for cluster analysis *Kasetsart J (Nat Sci)* **43** pp 378-88.
[11] Guha S, Rastogi R and Shim K 2001 Cure: an efficient clustering algorithm for large databases *Information Systems* **26**(1) pp 35-58.
[12] Aggarwal CC and Reddy CK 2013 *Data clustering: algorithms and applications* 1st ed: Chapman & Hall/CRC.
[13] Karypis G, Han E H and Kumar V 1999 Chameleon: Hierarchical clustering using dynamic modeling *Computer* **32**(8) pp 68-75.
[14] Huang J, Sun H, Kang J, Qi J, Deng H and Song Q 2013 ESC: An efficient synchronization-based clustering algorithm *Knowledge-Based Systems* **40** pp 111-22.
[15] Gülağız FK and Şahin S 2017 Comparison of hierarchical and non-hierarchical clustering algorithms *International Journal of Computer Engineering and Information Technology* **9**(1) pp 6-14.
[16] Rafsanjani M K, Varzaneh Z A, and Chukanlo N E 2012 A survey of hierarchical clustering algorithms *The Journal of Mathematics and Computer Science* **5**(3) pp 229-40.

[17] Murtagh F and Legendre P 2014 Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification* **31**.
[18] Hair F H J, Black W C, Babin B C, and Anderson R E 2014 *Multivariate data analysis* 7th ed.
[19] Gelawa Y A, Magalhães RJS, Assefaa Y, and Williams G 2019 Spatial clustering and socio-demographic determinants of HIV infection in Ethiopia, 2015–2017 *International Journal of Infectious Diseases* **82** pp 33-9.
[20] Hughey S M, Kaczynski A T, Porter D E, Hibbert J, and Liu J 2018 Spatial clustering patterns of child weight status in a southeastern US county *Applied Geography* **99** pp 12-21.
[21] Liew H P, Brooks T 2017 Spatial clusters of AIDS in Indonesia *Health Policy and Technology* **6**(2) pp 208-13.
[22] Assunção R M, Corrêa M, Câmara N G, and Freitas C D C 2006 Efficient regionalisation techniques for socio-economic geographical units using minimum spanning trees *International Journal of Geographical Information Science* **20**(7) pp 797-811.
[23] Bekti RD 2015 Metode spasial skater untuk pengelompokan lokasi berdasarkan fasilitas air bersih dan sanitasi *Jurnal Teknologi* **8**(1) pp 53-8.
[24] Rinaldi I K Y, Theresia M B, Efendi A, and Pramoedyo H 2017 Application of modified spatial 'k'luster analysis by tree edge removal (SKATER) method on the level of crime in Way Kanan district Lampung *Cauchy-Jurnal Matematika Murni dan Aplikasi* **4**(4) pp 155-160.