Yelli Ramalisa, Azzikra Febriyanti, Cut Multahadah

Department of Mathematics, Fakultas Sains dan Teknologi, Universitas Jambi, Indonesia

Email : yellirmalisa@gmail.com; Fazzikra.af@gmail.com ; cutmultahadah@gmail.com

Abstract. This research purpose to investigate the community health degree in Jambi and Muaro Jambi City using a non-hierarchical cluster. This method based on community health building indicators with 22 sub-districts, eleven in Jambi City and 11 in Muaro Jambi district. The results was basically 22 sub-districts in Jambi City and District. Muaro Jambi already has a fairly good IPKM value, such as South Jambi sub-district, Red Paal, Jambi Market and Teluk Danau based on analysis of hierarchical bomb.
Keywords: Hierarchical, IPKM, K-Rataan, K-Medoid

1. Introduction

The development achievements of a country are one of them based on how the country provides health insurance to the people in that country. Health status is one indicator of the success of development and is a global issue contained in the Human Development Index (HDI)/Human Development Index (HDI). IPKM is a collection of health indicators that can be easily and directly measured to describe health problems. This series of health indicators can directly or indirectly play a role in increasing long and healthy life expectancy. To find out the degree of health of the sub-district community in Jambi City and Muaro Jambi, grouping based on the 2013 GPA was limited to 14 indicators from 30 indicators in 7 groups, namely malnourished and under-fives, complete immunization, prenatal care, delivery by health workers in health facilities , proportion of sub-districts with sufficient number of doctors per population, ownership of health insurance, smoking, hypertension, diabetes mellitus, pneumonia, diarrhea, ARI, access to sanitation and access to clean water.

The degree of public health is very important in improving human development in a region / city. According to (3) data on health indicators that have been applied based on basic statistical techniques. The calculations that have been carried out are based on the average results of all indicators or are based on data distribution, for example the data in figure.1 concerning the 2012 Infant Mortality Rate (IMR). In addition, the data obtained has problems with data consistency and lack of each indicator. One solution to this problem is data mining.

Data Mining is a data processing technique that provides a number of algorithms that can be used to extract hidden information from multidimensional data sets . Data mining techniques that can explore data connectivity are non-hierarchical clusters, namely K-rataan and K-medoid. C-medoid and K-average. work by dividing data into a number of

groups analyzed similarity factors and inequality in the data and see the connectedness of a number of data. Average is a method of one method cluster analysis that is not often used. This method selects centroids based on the average value. Centroid retrieval based on mean causes grouping to be insensitive to outliers.

The K-Medoid method is a method of clustering on a median basis, so it is expected to be robust to outliers. The purpose of this study is based on the formulation of the problem is to group sub-districts in Jambi City and Muaro Jambi using a non-hierarchical cluster method based on community health building indicators. Furthermore, the study aims to obtain an overview of the level of public health based on the grouping that has been produced based on grouping indicators in the development of public health.

2. Experimental Section

The data used in this study are secondary data obtained by researchers from the Health Department of Jambi Province and Puskesmas in Jambi City. As for this research, it was conducted in Jambi City and Muaro Jambi Districts, so that the observation units were sub-districts in Jambi City. The city of Jambi consists of 11 sub-districts and 11 sub-districts in Muaro Jambi Regency, so the observation units are 22 sub-districts.

Procedure

The stages of the process carried out in the application of cluster analysis are as follows:

1. Exploring the data and ensuring there is no outlier in the sample data by calculating the distance of the mehalanobis square (D_i^2) for each observation in each cluster. Value D_i^2 then raised with $\chi_{p,1-\alpha}^2$ if $D_i^2 > \chi_{p,1-\alpha}^2$ then x or the first observation (I) is considered as outlier.
$$D_i = \sqrt{(x - \mu)' \Sigma^{-1}(x - \mu)},$$
With μ is a vector of middle values X, and Σ is a covariance matrix of X

2. Examination of freedom between variables is done by calculating the correlation value. In this study, the correlation between variables is not very strong (between independent variables) if the correlation value is between -0.80 and 0.80, on the contrary it is said that there are indications that variables are strongly correlated (Sugiyono2007). Perform Main Component Analysis (ME) for sample data.

3. Detecting outlier data.

4. Determine the amount of k (ie the number of clusters, and centroids in each cluster)

5. Calculate the distance between each object with each centroid in the K-average and k-medoid method.

6. Count the average (centroid) for the new cluster. Repeat step 4 until there is no more transfer of objects between buttons.

7. Compare the results of K-rataan and k-medoid clustering.
8. Calculate the level of misclassification on the K-average and k-medoid method and choose the best cluster.

3. Results and Discussion

Grouping with K-Medoid

Grouping by finding a number of objects (referred to as a representative or medoid object whose position is in the data center. One approach is to look for a number of objects that minimize the closest average between repetitive objects and other objects. Descriptions of groupings with medoid K are presented in Table 1 and Table 2.

The groups formed will be seen their characteristics so that they can be categorized into groups with the status of achievement of public health. In this study the bekum can say or conclude the quality of good / bad health from the community. Because there are no special assessments that have been carried out before by the health department. Here the researcher groups 3 categories and from that category will be seen a small size of each indicator of community health development.

Based on Table 1 and Table 2 Group 1 formed by K-medoid characterizes the districts with the best GPA compared to the other two groups, the third group is the sub-group that has the lowest GPA. This indicates that groups formed with K-medoid are in line with the average score of the GPA.

Table 1. Group Centers with K-Medoid

Variable	Group		
	1	2	3
X1	1	1	0.999184
X2	0.46118	0.370099	0
X3	0.314864	0.380048	0
X4	0.25412	0.138511	0.042402
X5	0.096081	0.02121	0.052234
X6	0.211351	0.226322	0.119982
X8	0.264689	0.286831	0.218922
X9	0.383917	0.383574	0.383223
X10	0.30226	0.295257	0.273825
X11	0.471748	0.488127	0.46648
X13	0.491502	0.49692	0.488415

Table 2 Number of Subdistricts in each Group

	Group					
	1		2		3	
Value	Average	Median	Average	Median	Average	Median
X1	0.999955	1	0.999805	1	0.999181	0.999181
X2	0.435779	0.451475	0.356452	0.355994	0.083438	0.083438
X3	0.347033	0.343452	0.377581	0.381136	0.121338	0.121338
X4	0.274656	0.280906	0.157328	0.140065	0.021201	0.021201
X5	0.113388	0.087653	0.005792	-0.00721	0.027257	0.027257
X6	0.232245	0.242701	0.230232	0.235045	0.126154	0.126154
X8	0.258535	0.266685	0.274558	0.27893	0.214244	0.214244
X9	0.383753	0.383984	0.383896	0.384002	0.382426	0.382426
X10	0.301662	0.302817	0.298195	0.299018	0.272785	0.272785
X11	0.484673	0.491679	0.490352	0.49092	0.465729	0.465729
X13	0.493155	0.497432	0.497232	0.497336	0.49048	0.49048
IPKM	0.864967	0.851629	0.814285	0.810066	0.640847	0.382426
Number of group member	10		10		2	

Grouping using K-Rata

Grouping using the K-Rata method is based on a comparison of the distance between the data and the center in each group. Data are grouped appropriately in groups that have the closest distance to the data. The description of the results of grouping using K-Average is shown in Table 3 and Table 4.

Table 3 Group Center with K-Rata

Variable	Group		
	1	2	3
X1	0.999977	0.999783	0.999181
X2	0.443711	0.34852	0.083438
X3	0.364689	0.359925	0.121338
X4	0.275192	0.156793	0.021201
X5	0.113966	0.005214	0.027257
X6	0.234298	0.22818	0.126154
X8	0.260048	0.273045	0.214244
X9	0.383747	0.383903	0.382426
X10	0.301996	0.297861	0.272785
X11	0.484675	0.49035	0.465729
X13	0.493119	0.497268	0.49048

Based on Table 3 and Table 4 the Group 1 formed by the average characterizes the district with the best GPA compared to the other two groups, the second group (two) is the sub-group that has the lowest GPA. This refers to the group formed with the average in line with the average score of the GPA.

Table 4 Number of Subdistricts in each Group

Variables	Group					
	1		2		3	
	Average	Median	Average	Median	Average	Median
X1	0.999978	1	0.999682	1	0.999977	0.999919
X2	0.42236	0.427213	0.30434	0.377281	0.457945	0.436918
X3	0.47259	0.469923	0.320161	0.329647	0.292755	0.356488
X4	0.26796	0.267513	0.134194	0.183458	0.280013	0.274433
X5	0.165119	0.097597	0.008888	0.007111	0.079865	0.099459
X6	0.231628	0.246354	0.211175	0.225388	0.236078	0.23546
X8	0.24746	0.252358	0.263245	0.274101	0.268439	0.272793
X9	0.383531	0.383666	0.383656	0.383776	0.38389	0.383594
X10	0.300227	0.302776	0.293682	0.301117	0.303175	0.302368
X11	0.479948	0.493506	0.486247	0.485746	0.487827	0.490383
X13	0.489034	0.498363	0.496137	0.495915	0.495842	0.497201
IPKM	0.891967	0.87823	0.780281	0.808594	0.857161	0.851629
Number of Group Member	4		12		6	

K-Medoid and K-rataan Test Goodness Test

Test the moetdoe goodness between groupings assessed by the average distance of objects in the group to the center of the group itself. In Table 5, and Table 6 shows the average object in the group to the group center and the average distance outside the group to the group center. Distance in the Group to the center The group itself can be seen in the diagonal members of the table

Table 5 The average distance of objects to the center of the K-Medoid method group

Object	Centre of Group		
	1	2	3
Group1	0.02667	0.04873	0.11841
Group 2	0.03995	0.01685	0.10231
Group 3	0.09532	0.08519	0.02422

The average value of the distance of objects in the group to the center The group looks the smallest for each group as illustrated in Table 5. This can be seen in the diagonal

members of the table having the smallest value compared to the other values on the same row.

Table 6 Distance of the average object to the center of the K-Rata method group

Object	Centre of Group		
	1	2	3
Group 1	0.02471	0.05054	0.117554
Group 2	0.04852	0.027993	0.071674
Group 3	0.01949	0.04096	0.098557

Based on Table 6, it can be seen that the average distance of objects in Group 1 is 0.02471. This value is the smallest value compared to the average distance outside the group to the center of the group. However, in Group 3 it can be seen that the smallest value occurs when the objects in the group use the group 1 center.

Table 5 and Table 6 show that Group 2 results of grouping using K-Medoin have the lowest average distance of objects in the group to the Group center. This value indicates that Group 2 is the most tight group compared to the other groups. From the two tebel groupings of IPKM values in sub-districts in Jambi City and Muaro Jambi District, the Medoid K-method method was better than the K-Rataan method.

Furthermore the goodness of the model can be seen with the value of fragility in and between groups. Diversity values and between groups of two methods are shown in the following table:

Table 7 Description of diversity within and between groups for the K-Medoid method

Peubah	Keragam dalam Kelompok (a)				Keragaman antar Kelompok (b)	Rasio Keragaman	
	1	2	3	Rata-rata		a/b	b/a
X1	6.28E-09	1.87E-07	1.8E-11	6.43E-08	9.79E-08	0.66	1.52
X2	0.00377	0.00040	0.01392	6.03E-03	0.002727	2.21	0.45
X3	0.00409	0.00082	0.02945	1.15E-02	0.004819	2.38	0.42
X4	0.00029	0.00239	0.00089	1.19E-03	0.001465	0.81	1.23
X5	0.00753	0.00101	0.00125	3.26E-03	0.003896	0.84	1.19
X6	0.00028	0.00079	7.62E-05	3.82E-04	0.000553	0.69	1.45
X8	0.00036	0.00016	4.38E-05	1.87E-04	0.000251	0.75	1.34
X9	1.72E-07	6.33E-08	1.27E-06	5.02E-07	1.71E-07	2.93	0.34
X10	1.06E-05	1.16E-05	2.16E-06	8.12E-06	1.23E-05	0.66	1.51
X11	0.00010	5.55E-05	1.13E-06	5.24E-05	0.000116	0.45	2.22
X13	7.69E-05	4.87E-06	8.52E-06	3.01E-05	4.48E-05	0.67	1.49

Table 8 Description of diversity within and between groups for the K-rataan method

Peubah	Keragam dallam Kelompok (a)				Keragaman antar Kelompok (b)	Rasio Keragaman	
	1	2	3	Rata-rata		a/b	b/a
X1	6.28E-09	1.87E-07	1.8E-11	6.43E-08	1.16E-07	0.26	3.92
X2	0.00377	0.00040	0.01392	6.03E-03	0.02741	0.22	4.60
X3	0.00409	0.00082	0.02945	1.15E-02	0.01064	0.86	1.17
X4	0.00029	0.00239	0.00089	1.19E-03	0.00969	0.18	5.64
X5	0.00753	0.00101	0.00125	3.26E-03	0.00322	1.60	0.62
X6	0.00028	0.00079	7.62E-05	3.82E-04	0.00245	0.37	2.68
X8	0.00036	0.00016	4.38E-05	1.87E-04	0.00061	0.58	1.72
X9	1.72E-07	6.33E-08	1.27E-06	5.02E-07	4.35E-07	0.53	1.90
X10	1.06E-05	1.16E-05	2.16E-06	8.12E-06	0.00016	0.15	6.77
X11	0.00010	5.55E-05	1.13E-06	5.24E-05	9.70E-05	0.56	1.77
X13	7.69E-05	4.87E-06	8.52E-06	3.01E-05	8.46E-06	0.76	1.31

Based on table 7 and Table 8 the diversity ratios in groups and between groups (a / b) have no different values. It can be seen that the average method gives a smaller diversity ratio value. In general, both methods can be said to be a good method of grouping because they produce a small diversity ratio. In general it can be said that both methods are good in grouping, this is shown by the diversity ratio between Groups (a / b) is greater than growling in Group (a).

Table 9. Medoid and K-mean Meode K-Average Value

Metode	Rata-rata jarak dalam objek	Rata-rata jarak objek di luar Kelompok	Rasio (b) dan (a)	Rata-rata jarak antar Kelompok	Nilai Hotelling	F-Hitung
K-Medoid	0.02258	0.07149	3.166076	0.061961	4.45	35.54
K-Rataan	0.05042	0.058123	1.152777	0.055555	2.38	37.66

Based on the good value of the pasa method Table 9 shows that the two methods are quite good in classifying sub-districts in Jambi City and Muaro Jambi District. But there is no method which is more favored than the two methods. On the comparison criteria for the average distance of objects outside the group with the average distance of the object to the center of the group itself, the K-Medoid method is more complicated because it has the smallest average distance compared to the K-Rata method. Therefore, it can be said that the K-Medoid method is able to produce groups that have a high density in the group. Furthermore for the diversity criteria, the K-Rata method is

superior because it is able to produce a more homogeneous group and higher diversity among groups.

Conclusions and recommendations
1. The following shows the results of grouping the values of IPKM in sub-districts in Jambi City and Kab. Muaro Jambi uses the Non Hierarchy method:
a. The results of grouping using K-Medoid provide groupings of regions as follows:

Table 10. Grouping results using the K-Medoid method

No	Kota/Kab	Kecamatan	Kelompok
1	Kota Jambi	Danau Sipin	1
2	Kota Jambi	Telanai Pura	
3	Kota Jambi	Jambi Selatan	
4	Kota Jambi	Paal Merah	
5	Kota Jambi	Pasar Jambi	
6	Kota Jambi	Danau Teluk	
7	Kota Jambi	Pelayangan	
8	Kota Jambi	Kota Baru	
9	Kota Jambi	Alam Barajo	
10	Kota Jambi	Jelutung	
11	Kota Jambi	Jambi Timur	2
12	Kab. Muaro Jambi	Bahar Selatan	
13	Kab. Muaro Jambi	Bahar Tengah	
14	Kab. Muaro Jambi	Bahar Utara	
15	Kab. Muaro Jambi	Jambi Luar Kota	
16	Kab. Muaro Jambi	Kumpeh Ulu	
17	Kab. Muaro Jambi	Maro Sebo	
18	Kab. Muaro Jambi	Mestong	
19	Kab. Muaro Jambi	Sekernan	
20	Kab. Muaro Jambi	Sungai Gelam	
21	Kab. Muaro Jambi	Kumpeh	3
22	Kab. Muaro Jambi	Taman Rajo	

Table 10 states that the IPKM of 10 sub-districts in Group 1 are in a very good category, 10 sub-districts in Group 2 have good IPKM values and 2 sub-districts in Group 3 have good GPA values.

b. The results of grouping using the Average provide groupings of regions as follows:

Table 11. Results of grouping the K-Rata method

No	Kota/Kab	Kecamatan	Kelompok
1	Kota Jambi	Jambi Selatan	1

2	Kota Jambi	Paal Merah	
3	Kota Jambi	Pasar Jambi	
4	Kota Jambi	Danau Teluk	
5	Kota Jambi	Jambi Timur	
6	Kab. Muaro Jambi	Bahar Selatan	
7	Kab. Muaro Jambi	Bahar Tengah	
8	Kab. Muaro Jambi	Bahar Utara	
9	Kab. Muaro Jambi	Jambi Luar Kota	
10	Kab. Muaro Jambi	Kumpeh	2
11	Kab. Muaro Jambi	Kumpeh Ulu	
12	Kab. Muaro Jambi	Maro Sebo	
13	Kab. Muaro Jambi	Mestong	
14	Kab. Muaro Jambi	Sekernan	
15	Kab. Muaro Jambi	Sungai Gelam	
16	Kab. Muaro Jambi	Taman Rajo	
17	Kota Jambi	Danau Sipin	
18	Kota Jambi	Telanai Pura	
19	Kota Jambi	Pelayangan	3
20	Kota Jambi	Kota Baru	
21	Kota Jambi	Alam Barajo	
22	Kota Jambi	Jelutung	

Table 10 states that the IPKM 4 sub-districts that are in Group 1 are in a very good category, 6 sub-districts in Group 3 have good IPKM values and 12 sub-districts in Group 2 have good GPA values.

c. Based on data obtained basically 22 sub-districts in Jambi City and District. Muaro Jambi already has a fairly good IPKM value even very good for several sub-districts such as South Jambi sub-district, Red Paal, Jambi Market and Teluk Danau. However, if you look at the acquisition of the IPKM value as seen in Appendix 1 for the Kec. Kumpeh and Kec. Taman Rajo has a low IPKM value even though it is already in a fairly good category.

Suggestion
Based on the results of research conducted by researchers it is expected that the city / regency government will pay more attention to the sub-districts that have low GPA values. Especially giving attention to the adequacy of doctors at health centers in each sub-district. For further research, researchers are expected to use more concrete methods to avoid repetitive grouping that occurs in one data.

References

[1] Analisis Pembangunan Wilayah. (2015). *Seri Analisis Pembangunan Wilayah Provinsi Jambi 2015.* Jambi: Pemprov Jambi.

[2] Artho, L. N., Wuisan, J., & Najoan, J. A. 2015.*Efek Serbuk Kopi Robusta (Coffea canephora) Terhadap Penyembuhan Luka Insisi Pada Kelinci (Oryctolagus cuniculus). Jurnal e-Biomedik (ebm)*, Volume 3, Nomor 3.

[3] Atthina, N &Lizda Iswari. (2014). *Klasterisasi Data Kesehatan Penduduk Untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means.* Snati. Issn 1907-5022. Yogyakarta.

[4] Badan Penelitian dan Pengembangan Kesehatan (BPPK). (2014). *Indeks Pembangunan Kesehatan Masyarakat*. Jakarta: Kementrian Kesehatan Republik Indonesia.

[5] Departemen Kesehatan.2008.*Buku Kesehatan Ibu dan Anak.*Departemen Kesehatan:Jakarta

[6] FlowrensiaY. 2010. Perbandingan penggerombolan*k-means*dan*k-medoid*pada datayangmengandung pencilan[Skripsi].Bogor:DepartemenStatistika FMIPAIPB.

[7] Johnson RA, Winchern D W . 1998 .*Applied Multivariate Statistical Analisy s4th Edition* .Prentice-Hall :London.

[8] KaufmanL,RousseeuwJP.1990.*FindingGroupsinData:AnIntroductiontoCluster Analysis*. John Wiley&Sons :New Jersey.

[9] Kementerian Kesehatan Republik Indonesia (Kemenkes RI).(2010). *24 Indikator Kesehatan dalam IPKM*. Dipublikasikasikan pada: Rabu, 01 Desember 2010. www.depkes.go.id

[10] Medriosa, Hamdani.2014.*Metode cluster analisys.*Jurnal Momentum, Vol 16 No.2: ISSN 1693752x

[11] Mega, Winda P.D. (2015). *Clustering menggunakan K-Means untuk Menentukan Status Gizi Balita.* Jurnal Informatika.Volume 15. Nomor 2.