Jurnal Varian
https://journal.universitasbumigora.ac.id/index.php/Varian
e-ISSN 2581-2017
Vol.4, No.2, April 2021, Hal. 109-116

Cluster Analysis with Complete Linkage and Ward's Method for Health Service Data in Makassar City

Isma Muthahharah[1], Agusalim Juhari [2]
[1]STKIP Pembangunan Indonesia, Indonesia, e-mail : muthahharahisma@gmail.com
[2]STKIP Pembangunan Indonesia, Indonesia, e-mail : agusalimjuhari@gmail.com

Article history:
Received : 07-09-2020
Reviced : 20-01-2021
Accepted : 28-01-2021

Keyword:
Cluster;
Complete Linkage;
Ward's Method;
Index Davies Bouldin (IDB)

ABSTRACT

Health care facilities are a place used to organize health efforts. Health service data in Makassar City has not shown which sub-districts have excellent service criteria, good enough, and not good. Therefore, it is necessary to group sub-districts with cluster analysis using hierarchy method. The hierarchy method used in this study is only 2, namely complete linkage and ward's method. Complete linkage method is the opposite of the approach to the minimum distance principle that is the furthest distance between objects while Ward's Method is a method that aims to minimize variance between objects in one cluster. There are four health services used, namely Hospitals, Health Centers, Home Care and Telemedicine with 15 sub-districts. This study also used a validity test namely Index Davies Bouldin (IDB) to determine the criteria of health services. The results of the analysis on complete linkage formed 3 clusters, namely cluster 1 with good health services, cluster 2 with excellent health services, and cluster 3 with poor health services. In addition, ward's Method also formed 3 clusters, namely cluster 1 with good health services, clusters 2 with poor service, and cluster 3 with excellent health services.

DOI: https://doi.org/10.30812/varian.v4i2.883

A. INTRODUCTION

Health care facilities are a place used to organize health efforts. The condition of health services in most sub-districts in the city of Makassar is increasing, along with that the level of public satisfaction is also increasing. The Health Office noted, the people served in 2018 reached 7,366 or 81.77% of people in comparison with 2017 only 4,685 or 81.71% of people served directly in their homes. (Dinas Kesehatan Provinsi Sulawesi Selatan, 2015). The data released has not yet issued specific data about which sub-districts are very good health services, quite good, and not good. Based on previous research on the quality picture of ISO certified inpatient services in Makassar City, the results of the study is expected that ISO-certified Health Centers maintain aspects that patients consider satisfied with the services provided and can make improvements that patients are not satisfied with the services provided in order to improve the quality of services. (Rusmin et al., 2016). Furthermore, about the analysis of the effect of the dimension of service quality on the satisfaction of inpatients in Bhayangkara Hospital Makassar, with the results of the study is the quality of services in the form of reliability, certainty or assurance, care and concrete evidence together have a positive and significant effect on the satisfaction of inpatients at Bhayangkara Hospital Makassar. (Nasyrah et al., 2017).

The results of the study have shown an overview of community satisfaction in Hospitals, Health Centers, Homecare, and Telemedicine in Makassar city. Therefore, a method is needed that can group the sub-district

according to the quality of service. This can make it easier for the Department of Health to be able to measure the level of health services. One of the grouping methods used in this research is cluster analysis, because cluster analysis is a multivariate technique in statistical analysis that aims to group objects that have the same characteristics into a smaller group. (Rahmawati et al., 2016). There are several cluster analysis methods developed to group, namely hierarchy and non-hierarchy methods. In hierarchical method have four agglomerative methods in cluster namely Single Linkage, Complete Linkage, Average Linkage, and Ward's Method. But in this study only used two methods namely Complete Linkage and Ward's Method, where Complete linkage is clustering method at the furthest distance between objects (maximum distance), while ward's method is a variant method that aims to obtain clusters that have the smallest internal variance cluster. (Sela Oktavia, 2013). This study also conducted validity test namely Index Davies Bouldin (IDB) is one of the internal validation tests on partition-based grouping method based on the amount of data proximity to centroid of the cluster it follows and between two clusters measured by the proximity of two centroid clusters. (Starczewski, 2017).

B. LITERATURE REVIEW

1. Cluster Analysis

Cluster analysis is a multivariate technique which has the main objective of classifying objects based on their characteristics. (Ediyanto et al., 2013). Thus, the purpose of grouping is so that objects that join in a group are objects that are similar (or related) to each other and different (not related) to objects in another group. Thus, the similarity between groups between individual classes (intra-class) must be small and high between different groups (between classes), similarity is considered as a measure of distance. (Rehioui et al., 2016). The most commonly used distance size is Euclidian distance size. Euclidean distance is the distance between objects, for example two objects to-i and to-j that are in p dimensions. The equation can be written as equation (1):

$$d_{ij} = \sqrt{\sum_{k=1}^{p}(X_{ik} - X_{jk})^2} \quad (1)$$

Where:

d_{ij} : Distance between the object to-i and the object to-j

p : Number of cluster variables

X_{ik} : Data from the subject of the$i$ on the variable to-ks

X_{jk} : Data from subject to-j on variable to-k

Hierarchical grouping methods can be agglomerative and divisive. Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy cluster. (Yogita Rani, 2015). There are four agglomerative methods in cluster formation, namely Single Linkage, Complete Linkage, Average Linkage, and Ward's Method. (Aprilia A.P et al., 2016). But in this study only use 2 methods, namely:

a. Complete Linkage

The complete linkage method is the opposite of the approach to the minimum distance principle. The distance used is the furthest distance between objects. The distance between one cluster and another is measured based on the object that has the furthest distance. (Ramadhani et al., 2018). Suppose X and have Y the distance (d_{XY}) farthest, it Z must be searched with the XZ maximum YZ distance and written as follows:

$$d_{(XY)Z} = Max(d_{(XZ)}, d_{(YZ)}) \quad (2)$$

where:

$d_{(XZ)}$: distance between cluster X and Z

$d_{(YZ)}$: distance between clusters Y and Z

$d_{(XY)Z}$: distance between cluster XY and Z

b. Ward's Method

Ward's Method strives to minimize variance between objects in a single cluster. The distance between the two clusters formed in ward's method is sum of square error between the two clusters.

SSE can only be calculated if the cluster has more than one element object. (Fathia et al., 2016). The SSE equation is as follows :

$$SSE = \sum_{j=1}^{P}\left(\sum_{i=1}^{n} X_{ij}^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_{ij}\right)^2\right) \quad (3)$$

where:

X_{ij} : value for the object to-i in the cluster to-j

p : number of variables measured

n : the number of objects in the cluster formed

The distance between the two clusters is the total number of squares of the two clusters for each variable. This method is different from other methods because it uses a variance analysis approach to calculate the distance between clusters or a method of minimizing the number of squares. (Suhaeni et al., 2018).

c. IDB Validity Test

Davies Bouldin Index (IDB) aims to maximize the distance between clusters (inter clusters) with each other (separation value) and estimate the distance between points (intra cluster) in a cluster (compactness value). (Starczewski, 2017). IDB is defined as the following:

$$IDB = \frac{1}{K}\sum_{i=1}^{K} R_i \quad (4)$$

where K is the number of clusters formed $R_i = \max_{j\neq i} \frac{S_i+S_j}{d_{ij}}$ and ($S_i$andS_j point to the spread of within-clusters for i_{th} and j_{th}).

2. Health Services in Makassar

Health services are efforts organized alone or jointly in an organization to maintain and improve health, prevent and cure diseases and restore the health of individuals, families, groups and or communities. (Daud M.Liando, 2016). Types of health services used in this research are:

a. Health Centers

Health Centers is a health care facility that facilitates public health efforts and first-rate individual health efforts, with the priority of promotive and preventive efforts to achieve the highest degree of public health in the working area. (Kemenkes RI, 2016). The existence of the puskesmas is very beneficial for poor families. Through the Health Centers, at least can answer service needs adequate health, namely service health within easy reach. (Irmawati, 2017).

b. Hospital

Hospital is a health service institution that organizes individual health services in a plenary that provides inpatient, outpatient, and emergency service. (Kementerian Kesehatan RI, 2018). Besides that, the hospital also has various types health services that can be seeded for maintain patient loyalty. (Supartiningsih, 2017).

c. Home Care

Home Care is a part or continuation of continuous and comprehensive health services provided to individuals and families in their homes that aim to improve, maintain or restore health or maximize the level of independence and minimize the impact of disease. (Kemenkes RI, 2014).

d. Telemedicine

Telemedicine is a health service that is carried out remotely through the use of communication and information technology in order to provide guidance or diagnostic consultation and treatment management between the health service facilities of the care takers and those who are managed. (Kementerian Kesehatan RI, 2016).

C. RESEARCH METHODS

1. Types of Research

This type of research is applied statistical research that will provide solutions in practical inequality.

2. Study Object

This research uses secondary data, namely data obtained directly from the Makassar City Health Office. The study object of this research is 15 sub-districts in Makassar city, with details as table 1:

Table 1. Sub-districts In Makassar City

No	Sub-districts	No.	Sub-districts
1	Biring Kanaya	9	Tallo
2	Bontoala	10	Tamalanrea
3	Makassar	11	Tamalate
4	Mamajang	12	Ujung Pandang
5	Manggala	13	Ujung Tanah
6	Mariso	14	Wajo
7	Panakkukang	15	Kepulauan Sangkarang
8	Rappocini		

3. Research Variables

The research variables useare:

X_1 : Health service in Hospital (%)
X_2 : Health service in Health Centers (%)
X_3 : Health service in Home Care (%)
X_4 : Health service in Telemedicine (%)

4. Research Procedure

The stages of the research carried out were:

1. Collecting data on the quality of health and human services in 15 sub-districts in Makassar City.
2. Grouping the sub-district with complete linkage.
   a) Calculate the matrix of distance between data by using Euclidian distance calculation.
   b) Determine the distance small from the matrix distance.
   c) Calculates the distance of the cluster joined by the other clusters.
   d) Create a recent distance matrix based on the previous calculation.
   e) Repeating the step (b) to step (d) so that it forms 3 cluster health services with good health services, poor service, and excellent health services.
3. Grouping sub-districts with ward's method
   a) Pay attention to N clusters that have one respondent cluster (all respondents are considered as clusters) at this stage SSE is zero.
   b) The first cluster in the shape by selecting two of the N that have the smallest SSE value.
   c) N-1 clusters of clusters are then considered again to determine the clusters, these clusters can minimize hydrogenation. Thus, N clusters are systematically reduced by N-1.
   d) Repeat steps (c) and (d) until one cluster is obtained or all respondents merge into one cluster.
4. From the cluster formed then tested validity by using Device Bouldin Index (IDB) to determine the criteria of service quality based on centroid value.

D. RESULTS AND DISCUSSION

The results of the research from the data used are the result of the overall variables of health services in the city of Makassar, namely Health Centers, Hospitals, Home Care and Telemedicine. Furthermore, the statistical analysis will be described from each variable of health care data presented in the following table 2:

Table 2. Descriptive Statistics of health care data in Makassar City

	N	Minimum	Maximum	Mean	Std. Deviation
Health Centers	15	52	100	78.67	12.821
Hospital	15	66	100	83.07	9.098
Home Care	15	68	100	84.40	11.469
Telemedicine	15	50	100	73.73	13.025

Based on table 2 it can be seen that the number of data (N) of all variables is equal to 15. Health care data in Makassar city has a maximum value (highest value) of all variables that is 100 while the minimum value (lowest data) is Telemedicine 50, Health Centers 52, Hospital 66 and Home Care 68. Health care data also has the highest mean in Home Care variable which is 84.40 with a deviation standard of 11.469 and the lowest is in Telemedicine variables with a deviation standard of 13.025. The standard deviation value indicates the level of diversity of health care percentage data.

Cluster analysis with complete linkage method is a complete linkage method with the distance between one cluster and another cluster measured based on objects that have the furthest distance. A new distance matrix in accordance with the numbering in the study object 15 Sub-districts in Makassar city, so that for the overall results can be seen in the following table 3:

Table 3. New Distance Matrix Complete Linkage

Stage	Cluster Combined		Coefficients
	cluster 1	cluster 2	
1	2	5	17.205
2	11	14	22.361
3	4	12	20,000
4	7	8	24.331
5	1	3	32.373
6	10	13	29.326
7	1	14	34.986
8	7	15	28.775
9	1	13	30.265
10	2	9	23.664
11	1	5	26.907
12	6	9	29.732
13	4	6	36.056
14	1	4	37.789

Based on table 3 the smallest coefficient is stage 1 with a value of 17,205 where the selected clusters are number 2 and 5 while the highest coefficient is stage 14 with a value of 37,789 where the selected clusters are number 1 and 4. This stage is carried out until the desired cluster is formed. In the study, 3 clusters were formed with details of the number of clusters and the members formed can be seen in figure 1:

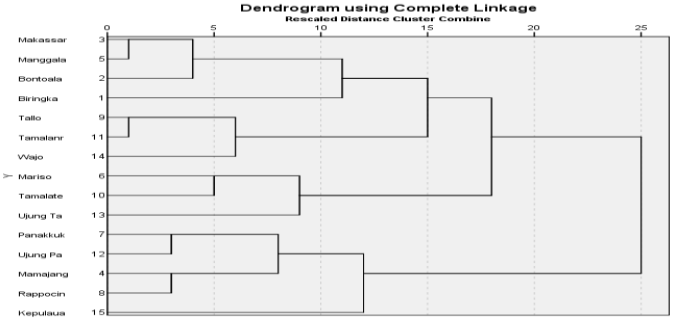

Figure 1. Dendogram Complete Linkage

Dendograms are useful for pointing out existing cluster members if multiple clusters should be formed. Figure 1 shows 3 clusters formed, then cluster 1 consisting of Makassar, Manggala, Bontoala, Biring Kanaya, Tallo, Tamalanrea and Wajo, Cluster2 consisting of Mariso, Tamalate, and Ujung Tanah. Cluster 3 consists of Panakukkang, Ujung Pandang, Mamajang, Rappocini and Kepulauan Sangkarang.

Ward's method is a hierarchy clustering method that is agglomerative or agglomeration to obtain clusters that have the smallest internal variants. SSE can be seen from the coefficient of agglomeration process. The agglomeration process starts from the two closest objects, then the distance is the closest of the many combinations of distances in accordance with the numbering in the study object 15 sub-districts in the city of Makassar. Give specific characteristics to describe the contents of the cluster drawn on the figure 2:

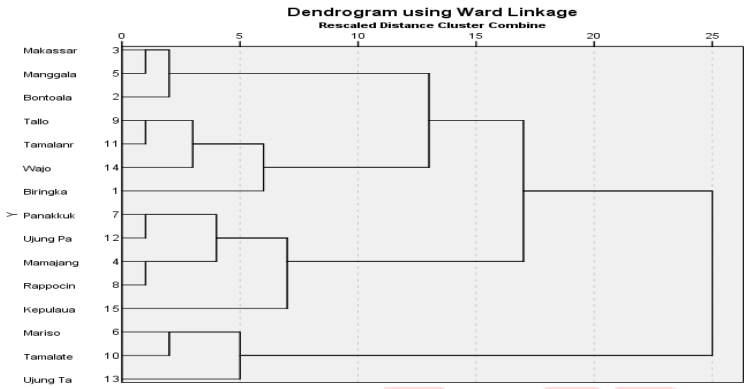

Figure 2. Dendogram Ward's Method

Based on Figure 1 shows 2 clusters formed with cluster1 consisting of Makassar, Manggala, Bontoala, Tallo, Tamalanrea, Wajo, and Biring Kanaya. Cluster 2 consists of Panakkukang, Ujung Pandang, Mamajang, Rappocini, and Kepulauan Sangkarang. Cluster 3 consists of Mariso, Tamalate, and Ujung Tanah.

From the process clustering with complete linkage and ward's method shows that cluster 1 has the same sub-district members, while cluster 2 and cluster 3 have different members. In the complete environment, the members of cluster 2 are 3 districts and cluster 3 are 5 districts, while the ward's method is cluster 2 members, namely 5 districts and cluster 3, namely 3 districts. Sub-district members who are included in the cluster with complete Linkage and ward's method are presented in the following table 3:

Table 3. Member cluster with Complete Linkage and Ward's Method

	Method	
Cluster	**Complete Linkage**	**Ward's Method**
1	Makassar	Makassar
	Manggala	Manggala
	Bontoala	Bontoala
	Biring Kanaya	Tallo
	Tallo	Tamalanrea
	Tamalanrea	Wajo
	Wajo	Biring Kanaya
2	Mariso	Mamajang
	Tamalate	Panakkukang
	Ujung Tanah	Rappocini
		Ujung Pandang
		Kepulauan Sangkarang
3	Mamajang	Mariso
	Panakkukang	Tamalate
	Rappocini	Ujung Tanah
	Ujung Pandang	
	Kepulauan Sangkarang	

After the clustering results were obtained, it was continued with the IDB validity test. IDB is one of the internal validation tests on the partition-based grouping method based on the number of data closeness to the centroid of the clusters that it is participating in and between the two clusters it is measured by the proximity of the two centroid clusters. IDB validity test on ward's method with a centroid cluster 1 value for each variable in table 5:

Table 5. Centroid Value Cluster 1

	X1	**X2**	**X3**	**X4**
Centroid Value	82	81	79	69

Based on table 5 it is known that the cluster of 1 have medium centroid value or all objects of waitern health is quite good in this case in Health Centers, Hospitals, Home Care, and Telemedicine.

Table 6. Centroid Value Cluster 2

	X1	**X2**	**X3**	**X4**
Centroid Value	86	78	93	86

Based on table 6 it is known that cluster 2 has a high centroid value or all the objects of excellent health services in this case in Health Centers, Hospitals, Home Care, and Telemedicine.

Table 7. Centroid Value Cluster 3

	X1	**X2**	**X3**	**X4**
Centroid Value	86	78	93	86

Based on table 7 it is known that cluster 3 has less centroid value or all objects of poor health in this case in Health Centers, Hospital, Home Care, and Telemedicine. For IDB test on complete linkage is done only by swapping table 6 to table 7, and table 7 to table 6 comes with explanation of each.

E. CONCLUSSION AND SUGGESTION

Based on the results of analysis conducted on health service data in the city of Makassar, the conclusions obtained that complete linkage formed 3 clusters namely cluster 1 with good health services, cluster 2 with excellent health services, and cluster 3 with poor health services. Then ward's method formed 3 clusters namely cluster 1 with good health services, clusters with poor service, and cluster 3 with excellent health services.

Apart from the two hierarchical methods used in this study, there are still average linkage and single linkage methods, as well as K-Mean for non-hierarchical methods that can form clusters so that they can be compared with the previous method with the aim of getting the best clustering method.

ACKNOWLEDGEMENT

Thank you to KEMENDIKBUD DIKTI who have given research grants and not forgetting also to the STKIP Pembangunan Indonesia who have provided support for this research to be realized properly.

REFERENCES

Aprilia A.P, N. W. A., Srinadi, I. G. A. M., & Sari, K. (2016). Pengelompokan Desa/Kelurahan Di Kota Denpasar Menurut Indikator Pendidikan. *E-Jurnal Matematika*, *5*(2), 38. https://doi.org/10.24843/mtk.2016.v05.i02.p119

Daud M.Liando, J. R. H. W. (2016). *Pelayanan Kesehatan di Pusat Kesehatan Masyarakat (Puskesmas) Amurang Timur Kapubaten Minahasa Selatan*. 6.

Dinas Kesehatan Provinsi Sulawesi Selatan. (2015). *Profil Kesehatan Provinsi Sulawesi Selatan*.

Ediyanto, Mara, N., & Intisari, N. S. (2013). Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, *02*(2), 133–136.

Fathia, A. N., Rahmawati, R., & Tarno. (2016). Analisis Klaster Kecamatan Di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode Ward Dan Single Linkage. *Jurnal Gaussian*, *5*(4), 801–810.

Irmawati, S. (2017). Kualitas Pelayanan Kesehatan di Puskemas Sangurara Kecamatan Tatanga Kota Palu. *Jurnal Katalogis*, *5*(1), 188–197.

Kemenkes RI. (2014). *Draft Konsep pengembangan pelayanan home care (kementerian kesehatan Republik Indonesia)*. 1–17.

Kemenkes RI. (2016). Buku Data Dasar Puskesmas Kondisi Desember 2015. *Pusat Data Dan Informasi Kementerian Kesehatan RI*, *53*(9), 1–169.

Kementerian Kesehatan RI. (2016). *Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.02.02/MENKES/409/2016 tentang Rumah Sakit Uji Coba Program Pelayanan Telemedicine Berbasis Video-Conference dan Teleradiologi*.

Kementerian Kesehatan RI. (2018). Peraturan Menteri Kesahatan. *New England Journal of Medicine*, *372*(2), 2499–2508.

Nasyrah, Darwis, & Hasmin. (2017). Analisis Pengaruh Dimensi Kualitas Pelayanan Terhadap. *Jurnal Mirai Management*, *2*, 133–148.

Rahmawati, L., Widya Sihwi, S., & Suryani, E. (2016). Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret). *Jurnal Teknologi & Informasi ITSmart*, *3*(2), 66. https://doi.org/10.20961/its.v3i2.654

Ramadhani, L., Purnamasari, I., & Amijaya, F. D. T. (2018). Penerapan Metode Complete Linkage dan Metode Hierarchical Clustering Multiscale Bootstrap (Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016). *Eksponensial*, *9*(2016), 1–10.

Rehioui, H., Idrissi, A., Abourezq, M., & Zegrari, F. (2016). DENCLUE-IM: A New Approach for Big Data Clustering. *Procedia Computer Science*, *83*(Ant 2016), 560–567. https://doi.org/10.1016/j.procs.2016.04.265

Rusmin, M., Aeni, S., & Yuliarma, A. F. (2016). Gambaran Mutu Pelayanan Rawat Inap Puskesmas Bersertifikasi ISO di Kota Makassar Tahun 2016. *Al-Sihah Public Health Science Journal*, *8*, 22–28.

Sela Oktavia, M. N. & N. S. (2013). Pengelompokan Kinerja Dosen Jurusan Matematika Fmipa Untan Menggunakan Metode Ward. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, *02*(2), 93–100.

Starczewski, A. (2017). A new validity index for crisp clusters. *Pattern Analysis and Applications*, *20*(3), 687–700. https://doi.org/10.1007/s10044-015-0525-8

Suhaeni, C., Kurnia, A., & Ristiyanti. (2018). Perbandingan Hasil Pengelompokan Menggunakan Analisis Cluster Berhirarki , K-Means Cluster , dan Cluster Ensemble. *Jurnal Media Infotama*, *14*(1), 31–38.

Supartiningsih, S. (2017). Kualitas Pelayananan Kepuasan Pasien Rumah Sakit: Kasus Pada Pasien Rawat Jalan. *Jurnal Medicoeticolegal Dan Manajemen Rumah Sakit 10.18196/Jmmr.2016*, *6*(1), 9–15. https://doi.org/10.18196/jmmr.6122

Yogita Rani, D. H. R. (2015). A study of hierarchical clustering algorithms. *2015 International Conference on Computing for Sustainable Global Development, INDIACom 2015*, *3*(11), 537–541.