



Hospital Performance Assessment Clustering in North Sumatra by using K-Means Algorithm

Adi Putra Lumban Gaol¹, Agustina Simangunsong²

Informatics Engineering Study Program, STMIK Pelita Nusantara, Medan, Indonesia.

Article Info

Keywords:

Hospital,
Data Mining,
K-Means Clustering.

ABSTRACT

The hospital is an integral part of the entire health care system that serves patients with various types of services. The hospital assessment system is an effort to provide a tool that encourages the hospital to continuously improve the quality and safety of services. Thus the hospital must apply hospital accreditation standards, including other standards that apply to hospitals in accordance with the description in the Hospital Accreditation Standards. Data mining is a process of finding meaningful relationships, patterns, and trends by examining a large set of data stored in storage using pattern recognition techniques such as statistical and mathematical techniques. In this research the data mining algorithm used is the K-Means Clustering Algorithm. The simplest and most common method is because K-Means has the ability to classify large amounts of data in a relatively fast and efficient computation time.

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



Corresponding Author:

Adi Putra Lumban Gaol,
Informatics Engineering Study Program, STMIK Pelita Nusantara, Medan, Indonesia,
Jl. Iskandar Muda No.1, Merdeka, Kec. Medan Baru, Kota Medan, Sumatera Utara 20154.
Email: adiputra.bggaolz@gmail.com

1. INTRODUCTION

The hospital as a health service institution as well as an entity that has responsibilities to stakeholders, and stakeholders must be observant in seeing changes caused by technological advances, the influence of globalization, higher levels of competition, and increasingly critical patient behavior in choosing health services. In the end, this is a trigger for hospitals to always provide optimal service to the community.

The hospital is an integral part of the overall health care system that serves patients with various types of services. Health is a big role in every activity to maintain and improve the health or standard of living of the community, aiming to achieve optimal health in hospitals, the need for quality resources for the community and the development of the world of health services business in the era of globalization lately continues to increase both physically quality and quantity. In improving the quality and quantity of hospital services, it also requires a management system that can mobilize all existing human resources so that it will have an impact on performance achievement.

According to Hasibuan, performance can be defined as a result of work that has been achieved by a person or organization in carrying out the tasks assigned to him based on skills, experience and sincerity and timeliness. The hospital assessment process is designed to improve the safety culture and quality culture in the hospital, so that it always strives to improve its quality and services. Through the assessment process, one of the benefits of hospitals can increase public confidence that hospitals focus on patient safety and service quality[1].

The hospital assessment system is an effort to provide a tool that encourages hospitals to continuously improve the quality and safety of services. Thus, hospitals must apply hospital

accreditation standards, including other standards that apply to hospitals in accordance with the elaboration in Hospital Accreditation Standards.

Data mining is a process of finding meaningful relationships, patterns, and trends by examining large sets of data stored in storage using pattern recognition techniques such as statistical and mathematical techniques[2]–[5]. Data mining techniques that are able to extract information from multidimensional data sets that have been obtained, in addition to extracting data that are connected to other data can also be done by data mining techniques using clustering. Clustering is a method for finding and grouping data that have similar characteristics (similarity) between one data and another[6]–[8].

The K-Means algorithm is the simplest and most common method[9]. This is because K-Means has the ability to group large amounts of data in a relatively fast and efficient computation time[10]. This algorithm has the advantages of being easy to implement and run, relatively fast, easy to adapt and the most widely practiced in data mining tasks[11], [12].

2. RESEARCH METHOD

The research framework is the stages carried out during research. The research framework is made to facilitate the achievement of research results, to be able to complete research on time and research can run as expected. The research framework used can be seen in Figure 1 below:

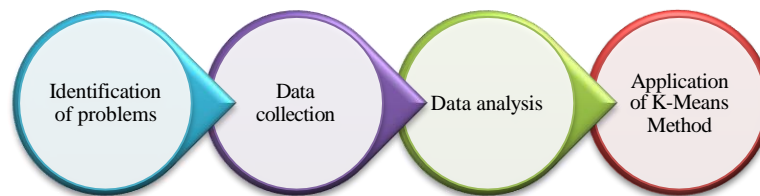


Figure 1. Research Framework

A. Identification of problems

At this stage, the authors identify and formulate problems in the study, this aims to find out the problems experienced. so that the author can determine a work plan and determine what data will be needed in this study.

B. Data collection

There are 2 sources of data obtained in writing the thesis, namely primary data and secondary data.

1. Primary data is data obtained directly from the first source, namely individuals or individuals who require further management such as the results of interviews or the results of filling out questionnaires. The primary data used by the author is in the form of a questionnaire distributed to all hospitals in North Sumatra.
2. Secondary Data is secondary data that has been further processed and presented properly by primary data collectors or other parties. Secondary data that the authors collect from the Hospital Management, Hospital Staff and Patients and Families.

C. Application of K-Means Method

K-Means is a simple clustering algorithm that is unsupervised. Suppose D is a dataset of n objects, and k is the number of clusters to be formed, the partitioning algorithm organizes these objects into k ($k \leq n$) partitions, where each partition represents a cluster. Each cluster is formed to optimize partition criteria, such as a difference function based on distance, so that objects in a cluster are similar, while objects in different clusters are not similar in terms of dataset attributes. The equation for calculating the distance between data on K-Means uses the Euclidian Distance (D) formula shown in Eq.

3. RESULTS AND DISCUSSION

In this study, hospital clustering will be carried out using the K-Means Clustering algorithm based on hospital data to look for similarities. From the clusters formed, the best clusters are searched for then cluster labeling is carried out. Not all hospital data will be included in the calculation of the

system, only some attributes that are useful and whose distribution is not too random. Because data that is too random will make the mining process take a long time and the level of accuracy is low. Hospital data that will be used as an attribute of the calculation include the number of patients, drug availability and number of facilities.

Data analysis is a process or effort to process data into new information so that the characteristics of the data become easier to understand and useful for solving problems, especially those related to research. Analysis of the data displayed in this study includes data analysis of Hospitals in North Sumatra. The data obtained from the research results can be seen in table 1 as follows.

Table 1. List of Hospitals to be Tested

No	Hospital Name	Address
1	RS Colombia Asia Hopital Medan	2A, Jl. Listrik, Petisah Tengah, Kec. Medan Petisah, Kota Medan, Sumatera Utara
2	RS Islam Malahayati	Jl. Pangeran Diponegoro No 2-4, Petisah Tengah, Medan Petisah, Kota Medan, Sumatera Utara 20112.
3	RS Sari Mutiara Medan	Jl. Kapten Muslim 79 Medan.
4	RSUP Haji Adam Malik Medan	No.17, Jl. Bunga Lau, Kemenangan Tani, Kec. Medan Tuntungan, Kota Medan, Sumatera Utara
5	RS Hermina Medan	RS Hermina Medan, Sei Sikambing C. II, Kec. Medan Helvetia, Kota Medan, Sumatera Utara
6	RS Imelda Medan	Pulo Brayon Darat I, Kec. Medan Tim., Kota Medan, Sumatera Utara
7	RS Siti Hajar Medan	jl. Jamin Ginting No.2, Merdeka, Kec. Medan Baru, Kota Medan, Sumatera Utara 20222
8	RSU Vina Estetika	Jl. Iskandar Muda No.119, Sei Sikambing D, Kec. Medan Petisah, Kota Medan, Sumatera Utara 20111
9	RS Bunda Thamrin Medan	28, Jl. Sei Batang Hari, Babura Sunggal, Kec. Medan Sunggal, Kota Medan, Sumatera Utara
10	RS Advent Medan	4, No.KM, Jl. Gatot Subroto, Sei Sikambing D, Kec. Medan Petisah, Kota Medan, Sumatera Utara
11	RS Estomihi	235, Jl. Sisingamangaraja, Sudirejo II, Kec. Medan Kota, Kota Medan, Sumatera Utara

The data is hospital data in North Sumatra, only a few hospitals in North Sumatra. The data consists of the name of the hospital and the address. The assessment process will be carried out using the K-Means method.

Analysis Method

K-Means can also be interpreted as a Clustering method which includes: in the partitioning approach. The K-Means algorithm is a centroid model. Centroid mode is a model that uses centroids to create clusters. The centroid is the midpoint of a cluster. The centroid is a value. Centroid is used to calculate the distance of a data object to the centroid. A data object is included in the cluster if it has the shortest distance to the centroid of the cluster.

The software system developed in this thesis is to implement the segmentation concept for hospital assessment based on hospital data which is used as a calculation parameter, namely, the number of patients, the availability of drugs and the number of facilities.

Attributes used in hospital assessment include:

Table 2. Attribute

No	Attribute Name	Information
1.	Number of Patients	Represents the number of patients registered in the hospital
2.	Drug Availability	Represents the amount of availability of drugs needed in the hospital
3.	Number of Facilities	It is a facility or service provided by a hospital

The following are 10 samples of hospital data used for the implementation of the K-Means calculation in the hospital assessment application.

Table 3. Hospital Data

No	Hospital Name	Number of Patients	Drug Availability	Number of Facilities
1	RS Colombia Asia Hopital Medan	98	90	80
2	RS Islam Malahayati	20	65	60
3	RS Sari Mutiara Medan	50	20	45
4	RSUP Haji Adam Malik Medan	40	40	80
5	RS Hermina Medan	85	86	80
6	RS Imelda Medan	40	50	73
7	RS Siti Hajar Medan	50	40	73
8	RSU Vina Estetika	30	60	45
9	RS Bunda Thamrin Medan	40	50	77
10	RS Advent Medan	40	75	80
11	RS Estomihi	10	60	80

For grouping the data using the K-Means algorithm. This process begins with normalizing data from the database. After the data is normalized, the data is ready for clustering. Existing data will be grouped into 3 groups. The steps for grouping data are as follows:

1. Determine the center of the cluster at random, suppose we determine
 - a. Cluster Center [1] = (81, 65, 65)
 - b. Cluster Center [2] = (65, 81, 65)
 - c. Cluster Center [3] = (65, 65, 81)

Calculate the distance of each existing data to each cluster center using the Euclidean distance formula. The distance between two points can be calculated by:

$$d = |x - y| = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Where:

- d : distance
- p : data dimension
- x : first data point,
- y : second data point,

For example, to calculate the distance of parameter data from the first hospital to the first cluster center is:

$$C_{11} = \sqrt{(98 - 81)^2 + (90 - 65)^2 + (80 - 65)^2} = 33,7490$$

$$C_{12} = \sqrt{(20 - 81)^2 + (65 - 65)^2 + (60 - 65)^2} = 61,2045$$

$$C_{13} = \sqrt{(50 - 81)^2 + (20 - 65)^2 + (45 - 65)^2} = 58,1893$$

Table 4. Cluster Position in the First Iteration

No	Hospital Name	Centroid 1			Centroid 2			Centroid 3		
		81	65	65	65	81	65	65	65	65
1	RS Colombia Asia Hopital Medan	33,7490			37,3496			41,4125		
2	RS Islam Malahayati	61,2045			48,0208			49,6588		
3	RS Sari Mutiara Medan	58,1893			65,9241			59,5482		
4	RSUP Haji Adam Malik Medan	50,3090			50,3090			35,3694		
5	RS Hermina Medan	26,1151			25,4950			29,0172		
6	RS Imelda Medan	44,3846			40,6201			30,2324		

7	RS Siti Hajar Medan	40,6201	44,3846	30,2324
8	RSU Vina Estetika	55,0090	45,4532	50,4579
9	RS Bunda Thamrin Medan	45,2769	41,5932	29,4278
10	RS Advent Medan	44,7883	29,7657	26,9443
11	RS Estomihi	72,7392	60,7536	55,2558

- A data will be a member of a cluster that has a distance smaller than the center of the cluster. Suppose for the first data, the smallest distance is obtained in the fourth cluster, so that the first data will be a member of the fourth cluster.
- Calculate the new cluster center. For the first cluster, there are 2 data, namely data 1 and 3 so that:

$$C11 = \frac{98 + 50}{2} = 74$$

$$C12 = \frac{90 + 20}{2} = 55$$

$$C11 = \frac{80 + 45}{2} = 62,2$$

So:

$$C1 = (74, 55, 62.2)$$

For the second cluster, there are 3 data, namely data 2, 5 and 8.

$$C11 = \frac{20 + 85 + 30}{3} = 45$$

$$C12 = \frac{65 + 86 + 60}{3} = 77,33$$

$$C11 = \frac{80 + 80 + 45}{3} = 61,66$$

So:

$$C2 = (45, 77.33, 61.66)$$

For the third cluster, there are 6 data, namely data namely 4, 6, 7, 9, 10 and 11

$$C11 = \frac{40 + 40 + 50 + 40 + 40 + 10}{6} = 36,66$$

$$C12 = \frac{40 + 50 + 40 + 50 + 75 + 60}{6} = 52,5$$

$$C11 = \frac{80 + 73 + 73 + 77 + 80 + 80}{6} = 77,16$$

So:

$$C3 = (36.66, 52.5, 77.16)$$

Cluster Center of the First Iteration:

$$C1 = (74, 55, 62.2)$$

$$C2 = (45, 77.33, 61.66)$$

$$C3 = (36.66, 52.5, 77.16)$$

Repeat step 2 until the data position has not changed.

Table 5. Cluster Position in Second Iteration

No	Hospital Name	Centroid 1		Centroid 2			Centroid 3			
		74	55	62.5	45	70.3	61.2	36.6	52.5	77.16
1	RS Colombia Asia Hopital Medan	45.9047			59.4296			71.9448		
2	RS Islam Malahayati	54.9749			25.6168			26.9948		
3	RS Sari Mutiara Medan	45.9047			53.2561			47.6311		
4	RSUP Haji Adam Malik Medan	41.0761			35.7941			13.2434		
5	RS Hermina Medan	37.2922			46.7071			58.8760		
6	RS Imelda Medan	35.9339			23.8094			5.8925		
7	RS Siti Hajar Medan	30.1869			32.7651			18.7453		

8	RSU Vina Estetika	47.6156	24.6891	33.6955
9	RS Bunda Thamrin Medan	37.2994	25.9529	4.1699
10	RS Advent Medan	43.1537	19.5675	22.9213
11	RS Estomihi	66.5375	40.8397	27.8458

Table 6. Cluster Position in the Third Iteration

No	Hospital Name	Centroid 1			Centroid 2			Centroid 3		
		77.6	65.3	68.3	30	66.6	61.6	36	48	76.6
1	RS Colombia Asia Hopital Medan	34.0293			74.1926			74.9637		
2	RS Islam Malahayati	58.2666			10.2740			28.6454		
3	RS Sari Mutiara Medan	58.0086			53.4373			44.4810		
4	RSUP Haji Adam Malik Medan	46.8681			33.8706			9.5686		
5	RS Hermina Medan	24.8394			61.1137			62.1012		
6	RS Imelda Medan	40.9349			22.4993			5.7410		
7	RS Siti Hajar Medan	37.8012			35.2073			16.5215		
8	RSU Vina Estetika	53.3385			17.9505			34.3301		
9	RS Bunda Thamrin Medan	41.5812			24.7565			4.4899		
10	RS Advent Medan	40.5996			22.4845			27.5056		
11	RS Estomihi	68.8718			27.9384			28.8367		

Table 7. Cluster Position in Fourth Iteration

No	Hospital Name	Centroid 1			Centroid 2			Centroid 3		
		91.5	88	80	25	65	66.25	44	40	69.6
1	RS Colombia Asia Hopital Medan	6.8007			78.3776			74.3246		
2	RS Islam Malahayati	77.7254			8.0039			35.9605		
3	RS Sari Mutiara Medan	87.0129			55.6916			32.2670		
4	RSUP Haji Adam Malik Medan	70.4006			32.2344			11.1427		
5	RS Hermina Medan	6.8007			65.0389			62.4912		
6	RS Imelda Medan	64.3836			22.2612			11.2942		
7	RS Siti Hajar Medan	63.8376			35.9939			6.8963		
8	RSU Vina Estetika	76.1002			22.3955			34.6577		
9	RS Bunda Thamrin Medan	64.0722			23.7815			13.0675		
10	RS Advent Medan	53.1154			22.6729			36.7309		
11	RS Estomihi	86.1756			20.9538			40.7941		

Because in the 3rd and 4th iterations the position of the cluster does not change, the iteration is stopped. So that the last cluster center is obtained from the K-Means calculation:

Cluster Center [0]=(91.5, 88, 80)

Cluster Center [1]=(25, 65, 66.25)

Cluster Center [2]=(44, 40, 49.6)

From the results of the cluster that has been formed, the results of the K-Means calculation are as follows:

Table 8. Calculation Results of K-Means Clustering

No	Hospital Name	Predicate
1	RS Colombia Asia Hopital Medan	Very good
2	RS Islam Malahayati	Good
3	RS Sari Mutiara Medan	Enough
4	RSUP Haji Adam Malik Medan	Less
5	RS Hermina Medan	Less
6	RS Imelda Medan	Less
7	RS Siti Hajar Medan	Less
8	RSU Vina Estetika	Less
9	RS Bunda Thamrin Medan	Less
10	RS Advent Medan	Less
11	RS Estomihi	Less

4. CONCLUSION

Hospital Assessment System in North Sumatra Using K-Means Algorithm Clustering Method, several conclusions can be drawn, including Based on the results of the tests that have been carried out, the results of the obtained from the results of the calculation is one hospital with a very good predicate is Columbia Asia Hospital Hospital, for the good predicate, Hermina Hospital, the rest got enough predicate and less predicate. The process of determining the scale of parameter values has a big influence on the K-algorithm. Means Clustering in determining the hospital predicate group. The tests that have been carried out can be concluded that the data mining classification using the Clustering K-Means method can be used for the hospital assessment process.

REFERENCES

- [1] D. W. Sitohang and A. Rikki, "Implementasi Algoritma K- Means Clustering untuk Mengelompokkan Data Gizi Balita pada Kecamatan Garoga Tapanuli Utara," vol. 02, pp. 80–92, 2019.
- [2] J. P. Gultom and A. Rikki, "Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin," *Kumpul. Artik. Karya Ilm. Fak. Ilmu Komput.*, vol. 02, no. 01, pp. 11–19, 2020.
- [3] P. Marpaung, "DATA MINING FOR DETERMINING BOOK LOAN PATTERNS IN-LIBRARY USING APRIORI ALGORITHM," *INFOKUM*, 2019. <http://infor.seaninstitute.org/index.php/infokum/article/view/8> (accessed Oct. 30, 2020).
- [4] T. Ramdhany, "Implementasi Data Mining Menggunakan Algoritma Apriori," *Pros. SINTAK 2017*, 2017.
- [5] P. S. Hasugian and S. Panjaitan, "The Data Mining of Cell Phone Most Interested Using," *J. INFOKUM, Vol. 7, No.1, Desember 2018*, vol. 7, no. 1, pp. 7–15, 2018.
- [6] F. Rahmawati and N. Merlina, "Metode Data Mining Terhadap Data Penjualan Sparepart Mesin Fotocopy Menggunakan Algoritma Apriori," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, 2018, doi: 10.33558/piksel.v6i1.1390.
- [7] A. Amrin, "Data Mining Dengan Regresi Linier Berganda Untuk Peramalan Tingkat Inflasi," *J. Techno Nusa Mandiri*, 2016.
- [8] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," *J. Tek. Komput. AMIK BSI*, 2018.
- [9] A. Rohmatullah, D. Rahmalia, and M. S. Pradana, "KLAUSTERISASI DATA PERTANIAN DI KABUPATEN LAMONGAN MENGGUNAKAN ALGORITMA K-MEANS DAN FUZZY C MEANS," *J. Ilm. Teknosains*, vol. 5, no. 2, 2020, doi: 10.26877/jitek.v5i2.4254.
- [10] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [11] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Cigugur Tengah," *J. Inf. Technol.*, vol. 2, no. 2, 2020, doi: 10.47292/joint.v2i2.30.
- [12] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.