JURNAL TEKNIK KOMPUTER NETWORK
Vol. 1, No. 1, Juni 2021,
E-ISSN: XXXX-XXXX
DOI:

IMPLEMENTATION OF K-MEDOIDS ALGORITHM IN GROUPING DISEASES IN THE COMMUNITY

Risnawati Saragih[1], Solikhun[2], Irawan[3], Eka Irawan[4], Ilham Syahputra[5]
[1,2,3,4,5]Program Studi Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar, Indonesia

e-mail :risnawatisaragih83@gmail.com[1],solikhun@amiktunasbangsa.ac.id[2]
,irawaniwan56@gmail.com[3] ,eka.irawan@amiktunasbangsa.ac.id[4], Ilham@amiktunasbangsa.ac.id[5]

Keywords:

Data Mining
K-Medoids Algorithm
Disease in society

ABSTRACT

Disease is a type of disease that attacks children and adults. Disease transmission can be through direct contact, coughing, sneezing and so on. In this case, some of the complications that often make the Bukit Maraja Community Health Center community include diarrhea, allergies, influenza, HIV, gout and so on. Based on the Ministry of Health, Indonesia is still one of the countries that often suffer from diseases. Puskesmas Bukit Maraja noted that the number of cases of the latest disease has increased in the past four years. To see the results of grouping diseases that often occur or are most visited by the people of the Bukit Maraja Community Health Center, the data mining method is used with the K-Medoids Algorithm. Sources of data obtained by making observations at the Bukit Maraja Community Health Center so that it is hoped that this research can help the government and community health centers in overcoming diseases that often occur or are most visited by the Bukit Maraja Community Health Center. The results obtained from these tests are the same calculations between manuals. And rapidminer, and obtained high and low clusters, high cluster there are 5 types of diseases such as Dbd, Types, Asthma, Rubella, tuberculosis and low clusters there are 19 types of diseases, namely diarrhea, diarrhea, gout, diabetes, HIV, tetanus, scabies, ulcers , Influenza, Worms, Ringworm And Ulcer

Corresponding Author:

Solikhun,
Department of System Information,
STIKOM Tunas Bangsa Pematangsiantar,
Jend. Sudirman street bock A No 1,2,& 3 Pematangsiantar, North Sumatera, Indonesia.
Email: Solikhun@Amiktunasbangsa.ac.id

1. INTRODUCTION

Disease is a type of disease that attacks children and adults. Disease transmission can be through direct contact, coughing, sneezing and so on. In this case, some of the complications that are often experienced by the Bukit Maraja Community Health Center include diarrhea, allergies, influenza, HIV, gout and so on.

Health development must be considered because it is an investment forimprove the quality of human resources. In measuring the Human Development Index (HDI), health is one of the main components besides education and income. In Law Number 23 of 1992 concerning Health, it is stipulated that health is a state of well-being of body, soul and society that can enable everyone to live productively socially and economically[1].

Puskesmas Bukit Maraja noted that the number of disease cases has tended to increase in the past four years. The impact of environmental rejection and social problems that befall on the community encourages various parties to take early prevention. Due to the lack of environmental hygiene, complaints of disease are increasing every year. To overcome this, the government needs accurate data to find out the number of diseases each year at Bukit Marajaagar Community Health Center can provide solutions to the Bukit Maraja Community Health Center through data on the number of diseases in the community from 2015 to 2018 and it requires public concern for the environment. The grouping process can be implemented through One of the clustering methods is the K-medoid method

Clastering in data mining can be done using the K-medoid algorithm. With the k-medoid algorithm, it will produce a cluster presentation that is formed using the medoids which is built by calculating the proximity between the medoid and the non-medoid object using a distance measure. This method can minimize the amount of dissimilarity between each object and the appropriate medoid[2]

Based on the background of the problem above, the author provides a solution to classify cases of disease complaints in the BukirMaraja Community Health Center using data mining with the K-medoid clustering method.

2. RESEARCH METHOD

2.1. Research design

The research was conducted using literature studies with disease grouping in the community, the data were taken from the Bukit Maraja Community Health Center. Some of the ways this research works, namely:

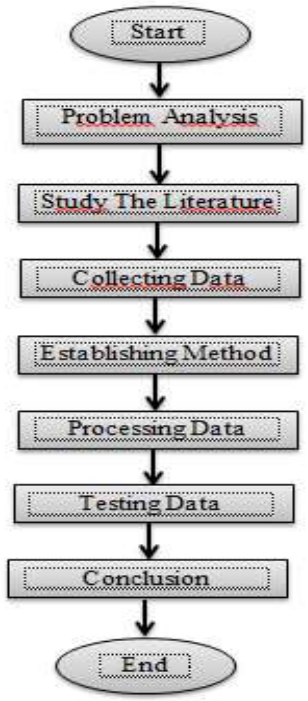

Image 1. Research Workflow

Description of the framework:

1. Problem Analysis
   Analyze the problems associated with the data on the number of diseases in the Bukit Maraja Community Health Center and determine the parameters used. In this research. The parameter used is the data on the number of diseases in the community at Puskesmas Bukit Maraja from 2015-2018.
2. Studying Literature
   This research must be based on several references that can be used to obtain information in research.
3. Collecting Data
   The data were obtained from the Bukit Maraja Community Health Center regarding disease data in the Bukit Maraja Community Health Center in 2015-2018

4. Establish Methods
   Establish a method for solving the problem. This study uses the K-Medoids Clustering method.
5. Managing Data
   Perform data processing using data mining, especially the k-medoids clustering method.
6. Test data
   Data testing was carried out using the Rapidminer 5.3 tool so that a cluster that is in accordance with the k-medoids clustering method is obtained.
7. Conclusion
   The conclusion obtained in classifying disease data in the Bukit Maraja Community Health Center is in the form of 2 clusters, namely the low cluster and the high cluster.

2.2. Data Source

The data that is processed in this research is community disease data provided by the Bukit Maraja Community Health Center. The data used are data from 2015 to 2018.

Table 1. Raw Data

	Number of Diseases of the Bukit Maraja Community Health Center			
Disease	**2015**	**2016**	**2017**	**2018**
DIARRHEA	20	16	16	16
DENGUE FEVER	90	81	93	93
TYPHUS	69	70	71	70
ASTHMA	59	53	57	58
ISPA	24	24	25	25
RUBELA	68	68	69	71
URIC ACID	20	20	20	20
DIABETES	22	23	24	22
HIV	30	34	32	28
TETANUS	19	20	20	20
SCABIES	27	27	26	25
BOILS	16	18	16	17
HYPERTENSION	31	33	32	32
RHEUMATISM	26	25	24	24
TBC	88	80	83	86
MEASLES	18	18	17	19
BREATHLESS	31	29	28	28
DENTAL CARIES	35	32	34	30
SKIN ALLERGIES	30	30	30	27
BOILS	35	37	35	37
INFLUENZA	25	27	28	23
WORMS	33	35	33	37
RINGWORM	16	19	16	17
ULCER	19	16	17	19

2.3. Data Mining

Data Mining is an activity to form large amounts of data and find knowledge in order to find information.Data mining: the process of looking for interesting patterns or information in selected data using certain techniques or methods[3].The purpose of data mining is to find, or increase knowledge of information[4].

Data mining is divided into several groups based on the tasks that can be done :[5]

1. Description (Description)
   Description is a technique of forming a pattern that is contained in data and describing the tendency of a data.
2. Estimation (Estimation)

Estimation is almost the same as classification, except that the estimation target variable is more numerical.

3. Prediction (Prediction)
   Prediction is almost the same as classification and estimation, except that in the prediction the value of the results will be in the future.
4. Classification
   In the classification there are target categorical variables. Distinguish data classes.
5. Clustering
   Clustering is a grouping or paying attention to and forming a class of objects that have similarities.
6. Association
   Association in data mining is finding attributes that appear at one time.

2.4. Clustering

Clustering is a method of analyzing data grouping. Cluster analysis is grouping data in such a way that has relatively different characteristics[6]. The purpose of the cluster is not to classify, estimate, predict the value of the target variable.[7]

2.5. Rapidminer

Rapidminer is a data science tool developed by the company of the same name that provides an integrated environment for machine learning data preparation, learning in text development and predictive analysis[8], Rapid Miner is a tool used in engineering in the environment of machine learning, data mining, text mining and predictive analytics[9].

2.3.1. Unifield Modelling Langue (UML)

Unified Modeling Language is a standard specification language used for modeling object-oriented program designs[10].

2.3.2. Use Case Diagram

Use case diagrams are modeling or techniques used in developing a software or information system[11]

2.3.6. Flowchart

Flowchart is a symbol of an algorithm to solve a problem to make it easier to check the forgotten part in analyzing the problem. Flowcharts are useful for communicating for programming a project[12], Flowchart is a systematic presentation of the process and logic of

the act of handling information or the graphical depiction of the steps and sequence of procedures of a program[13].

2.6. Algoritma K-Medoids

The K-Medoids method is a classic partition clustering technique that groups object data. The K-Medoids method is quite efficient in small datasets. The initial step for K-Medoids is to find the most representative points (medoids) in the dataset by calculating the distance from groups in all combinations of medoids so that the distance between points in the cluster is small and the distance between the points between the cluster is large.[14]. Basically, the use of this algorithm in the clustering process depends on the data obtained and the conclusions to be reached at the end of the process[15]. The first step for K-Medoids is to find the most representative points (medoids) in all combinations.

The steps for calculating the k-medoids clustering algorithm include:

1. look for as many cluster points (number of clusters)
2. looking for the closest calculation data where i = 1,… .., n; j = 1,… .., n and p is a variable.
3. Accurately search for the cluster object as the new medoid.
4. Count every object on the new medoid.
5.

Draws the result of the deviation (S) by calculating the new total distance value - the old total distance.

3. RESULTS AND ANALYSIS

3.1. Early Medoids

Table 2. Early Medoids

Name	Disease	2015	2016	2017	2018
C1	TBC	90	81	93	93
C2	BISUL	16	18	16	17

The results of calculations using the k-medoids algorithm in the Bukit Maraja Community Health Center are as follows:

Determine the number of clusters (k) of n objects is 2 clusters

1. Assuming the initial centroid that has been determined as in table 2
2. placed the non-medoids object into the cluster closest to the medoids based on the Euclidean distance.

$$D_{diare,c1} = \sqrt{\begin{matrix}((20-90)^2+(16-81)^2+(16-93)^2 \\ +(16-93)^2\end{matrix}} = 144.8551$$

$$D_{diare,c2} = \sqrt{\begin{matrix}((20-16)^2+(16-18)^2+(16-16)^2 \\ +(16-17)^2\end{matrix}} = 4.5825757$$

3.2 Discussion

3.2.1. K-medoids Algorithm Process

Some of the steps for the manual calculation process of the k-medoids algorithm in the Bukit Maraja Community Health Center are as follows:

1. Assume a predefined starting centroid
2. The following is the calculation of the distance to the number of diseases of the Pukesdes Bukit Maraja Community.

$$D_{diare,c1} = \sqrt{\begin{matrix}((20-90)^2+(16-81)^2+(16-93)^2 \\ +(16-93)^2\end{matrix}} = 144.8551$$

$$D_{diare,c2} = \sqrt{\begin{matrix}((20-16)^2+(16-18)^2+(16-16)^2 \\ +(16-17)^2\end{matrix}} = 4.5825757$$

The calculation results can be seen in table 3 below:

Table 3.Early Medoids

Name	Desease	2015	2016	2017	2018
C1	TBC	90	81	93	93
C2	BISUL	16	18	16	17

Table 4. The Result of Calculation of the 1st Iteration of the K-Medoids Algorithm

DESEASE	DISTANCE K-MEDOIDS		NEAREST	
	C1	C2		CLUSTER
DIARRHEA	144.8551	4.5825757	4.582576	C2
DENGUE FEVER	0	145.4304	0	C1
TYPHUS	39.68627	106.5223	39.68627	C1
ASTHMA	65.31462	80.224684	65.31462	C1
ISPA	129.8191	15.652476	15.65248	C2

RUBELA	41.3884	104.54186	41.3884	C1
URIC ACID	138.8488	6.7082039	6.708204	C2
DIABETES	133.3792	12.247449	12.24745	C2
HIV	117.2817	28.79236	28.79236	C2
TETANUS	139.3557	6.164414	6.164414	C2
SCABIES	126.4832	19.131126	19.13113	C2
BOILS	145.4304	0	0	C2
HYPERTENSION	115.0087	30.512293	30.51229	C2
RHEUMATISM	129.4372	16.186414	16.18641	C2
TBC	12.40967	135.19615	12.40967	C1
MEASLES	142.8461	3	3	C2
BREATHLESS	120.9752	24.718414	24.71841	C2
DENTAL CARIES	113.4725	32.403703	32.4037	C2
SKIN ALLERGIES	120.5239	25.21904	25.21904	C2
BOILS	107.0561	38.509739	38.50974	C2
INFLUENZA	127.5382	18.493242	18.49324	C2
WORMS	110.0045	35.594943	35.59494	C2
RINGWORM	145	1	1	C2
DIARRHEA	143.2411	4.2426407	4.242641	C2
JUMLAH	2609.356	895.07442		
Total Cost	3504.429991			

It was found that the results of the distance between the 1st and 2nd iterations can be seen in Table 5 below:

Table 5. Clustering Results Using Clustering

DESEASE	DISTANCE K-MEDOIDS		*NEAREST*	CLUSTER
	C1	C2		
DIARRHEA	134.5697	3.316625	3.316625	C2
DENGUE FEVER	12.40967	143.2411	12.40967	C1
TYPHUS	29.3428	104.561	29.3428	C1
ASTHMA	55.04544	78.03845	55.04544	C1
ISPA	119.6537	13.74773	13.74773	C2
RUBELA	31.06445	102.5329	31.06445	C1
URIC ACID	128.6429	5.196152	5.196152	C2
DIABETES	123.2153	10.77033	10.77033	C2
HIV	106.9813	27.40438	27.40438	C2
TETANUS	129.1743	5.09902	5.09902	C2
SCABIES	116.1895	17.37815	17.37815	C2
BOILS	135.1962	4.242641	4.242641	C2
HYPERTENSION	104.7616	28.75761	28.75761	C2
RHEUMATISM	119.1386	14.28286	14.28286	C2
TBC	0	133.0489	0	C1
MEASLES	132.6235	2.236068	2.236068	C2
BREATHLESS	110.63	22.69361	22.69361	C2
DENTAL CARIES	103.1988	30.36445	30.36445	C2

SKIN ALLERGIES	110.2452	23.45208	23.45208	**C2**
BOILS	96.7626	36.67424	36.67424	**C2**
INFLUENZA	117.3542	17.14643	17.14643	**C2**
WORMS	99.7547	33.71943	33.71943	**C2**
RINGWORM	134.7405	4.795832	4.795832	**C2**
DIARRHEA	133.0489	0	0	**C2**
JUMLAH	2383.744	862.6999		
Total Cost	3246.444			

The results found are processed into the k-medoids algorithm using Rapidminer to determine the most diseases in Puskesmas Bukit Maraja, while the final results obtained are as follows:

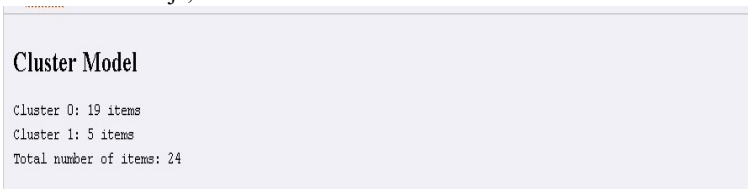

image 2. Rapidminer End Result

3.2.2. Test result

In research that is supported by the results of data processing using Sostware Rapidminer, the results of 2 clustering can be seen in Figure 2 below:

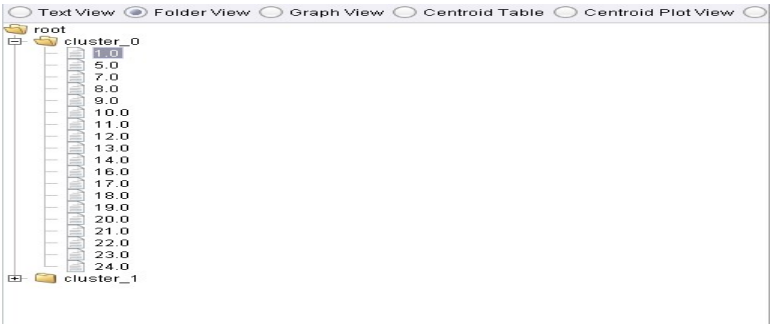

Image 3. Results of the K-Medoids Algorithm

4. CONCLUSION

Based on the research that has been done, several conclusions have been obtained, namely, The causes of various diseases have been found in Puskesmas Bukit Maraja. And data processing was carried out using Rapidminer software to obtain 2 clusters, namely the high cluster there were 5 types of disease and the low cluster there were 19 types of disease.

ACKNOWLEDGEMENTS

Thank you, hope my research may be useful for further researches as a refrence

REFERENCES(10 PT)

[1] S. S. Sundari and N. Ariani, “Penerapan Data Mining Untuk Pengelompokan Penyakit Dengan Algoritma Fuzzy C-Means (Studi Kasus : UPT Puskesmas Salawu),” *J. VOI (Voice Informatics)*, vol. 8, no. 2, pp. 63–76, 2019.

[2] L. Purba, S. Saifullah, and R. Dewi, “Pengelompokan Kasus Penyakit Aids Berdasarkan Provinsi Dengan Data

Mining K-Medoids Clustering," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 687–694, 2019, doi: 10.30865/komik.v3i1.1679.

[3] L. R. Angga Ginanjar Mabrur, "Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit," *J. Komput. dan Inform.*, vol. 1, no. 1, pp. 53–57, 2012.

[4] N. Erlangga, S. Solikhun, and I. Irawan, "Penerapan Data Mining Dalam Mengelompokan Produksi Jagung Menurut Provinsi Menggunakan Algoritma K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 702–709, 2019, doi: 10.30865/komik.v3i1.1681.

[5] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.

[6] A. P. Windarto, "Penerapan Datamining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering Method," *Techno.Com*, vol. 16, no. 4, pp. 348–357, 2017, doi: 10.33633/tc.v16i4.1447.

[7] E. Nanda, Solikun, and Irawan, "PENERAPAN DATA MINING DALAM MENGELOMPOKAN PRODUKSI JAGUNG MENURUT PROVINSI MENGGUNAKAN ALGORITMA K-MEANS," vol. 3, pp. 702–709, 2019, doi: 10.30865/komik.v3i1.1681.

[8] S. M. Dewi, A. P. Windarto, and D. Hartama, "Penerapan Datamining Dengan Metode Klasifikasi Untuk Strategi Penjualan Produk Di Ud.Selamat Selular," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 617–621, 2019, doi: 10.30865/komik.v3i1.1669.

[9] D. Novianti, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, 2019, doi: 10.31294/p.v21i1.4979.

[10] Suendri, "Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus: UIN Sumatera Utara Medan)," *J. Ilmu Komput. dan Inform.*, vol. 3, no. 1, pp. 1–9, 2018.

[11] K. Kawano, Y. Umemura, and Y. Kano, " Field Assessment and Inheritance of Cassava Resistance to Superelongation Disease 1 ," *Crop Sci.*, vol. 23, no. 2, pp. 201–205, 1983, doi: 10.2135/cropsci1983.0011183x002300020002x.

[12] S. Santoso and R. Nurmalina, "Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas (Studi Kasus Politeknik Negeri Tanah Laut)," *J. Integr.*, vol. 9, no. 1, pp. 84–91, 2017.

[13] P. Soepomo, "Membangun Aplikasi Autogenerate Script ke Flowchart untuk Mendukung Business Process Reengineering," *J. Sarj. Tek. Inform.*, vol. 1, no. 2, pp. 448–456, 2013, doi: 10.12928/jstie.v1i2.2555.

[14] I. Indriani, P. Damanik, I. S. Saragih, and I. Parlina, "Algoritma K-Medoids untuk Mengelompokkan Desa yang Memiliki Fasilitas Sekolah di Indonesia," no. September, pp. 520–527, 2019.

[15] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018, doi: 10.30865/komik.v2i1.930.