



Grouping Diabetes Diagnosis Based on Age Range with K-means Algorithm

Allwin M.Simarmata¹, Maria Albina Pratiwi Sianipar², Sukbinder Singh³,
Intan Is Mutiara Gulo⁴, Jhon Boy Ricardo Purba⁵

^{1,2,3,4,5}Major Informatic, Faculty Teknologi and Computer Science,
^{1,2,3,4,5}Universitas Prima Indonesia Medan, South Sumatra

E-mail: ¹allwinsimarmata@unprimdn.ac.id, ²mp2754788@gmail.com, ³sukbinder999@gmail.com,
⁴intan99gulo@gmail.com, ⁵boyjhon641@gmail.com

ARTICLE INFO

ABSTRACT

Article history:

Received: July 12, 2021
Revised: August 11, 2021
Accepted: August 30, 2021

Keywords:

Group,
Range,
Age,
K-means,
Diabetes

Diabetes Mellitus has two types of disease, namely Diabetes Type I and Diabetes Type II. Research on the grouping of diabetes based on age at RSU Royal Prima Medan has not been carried out. For this reason, the researchers made a research analysis in grouping diabetes based on age. The aim of this research is to find the age group that is susceptible to being diagnosed with Type I or Type II diabetes. Researchers took sample data from patients diagnosed with diabetes of 2019 as many as 200 data. The method used in grouping the data is Kmeans. The Kmeans method has been widely used in various studies for grouping in the medical field. The results of the calculation analysis show that a sample of 200 data is Type I Diabetes at the age of 43 - 47, 56, 62, 67, 68 while Type II diabetes is at the age of 49, 55, 63. It is possible from the results of the study that patients of that age are not based on There are not many types though. The conclusion in this study is that the grouping of Type I and Type II diabetes based on age is expected to assist the hospital in identifying patients quickly so that patients can receive health services quickly as well.

Copyright © 2021 Jurnal Mantik.
All rights reserved.

1. Introduction

Diabetes mellitus is a disease that has high sugar levels in the blood. People with diabetes mellitus are usually obtained from an unhealthy lifestyle or it can also be inherited from their parents. Patients with diabetes mellitus when exposed to other diseases or exposed to injuries will make it worse. Diabetes mellitus has 2 (two) types, namely dry type diabetes and wet type diabetes. Both are categories that must be treated seriously so that healing can be achieved.

Diabetes has 2 (two) types, namely type I and type II, both of which are types that must be analyzed further for serious treatment so that it does not spread to other diseases. These types require an analysis of how spread out over how old the patient is, whether those under the age of 20 who have diabetes always fall into the Type II category or vice versa. For this reason, research is needed on the age-susceptibility of grouping diabetes by age into Type I or Type II categories. The expected benefit is that the hospital can quickly and accurately categorize Type I or Type II in patients with Diabetes Mellitus so that the supply of drugs can be properly coordinated.

According to the definition of the National Institute of Diabetes and Digestive and Kidney Diseases Health Information Center, diabetes is a disease that occurs when the sugar content in the blood is too high. Over time, high blood sugar levels can cause health problems. Among them are heart disease, nerve damage, vision, and kidney disease ^[1]. K-Means is one of the methods in the clustering category. Kmeans utilizes a number of variables in grouping, using different algorithms for each feature and concept ^[2]. One algorithm will reflect other features in different algorithms ^[3]. K-Means is an algorithm that requires k input parameters and divides a set of n objects into clusters so that the level of similarity between members in one cluster is high



while the level of similarity with members in other clusters is very low [4]. The resemblance of members to the cluster is measured by the proximity of the object to the mean value in the cluster or referred to as the cluster centroid [5].

2. Method

In this study, the research method uses the waterfall diagram technique.

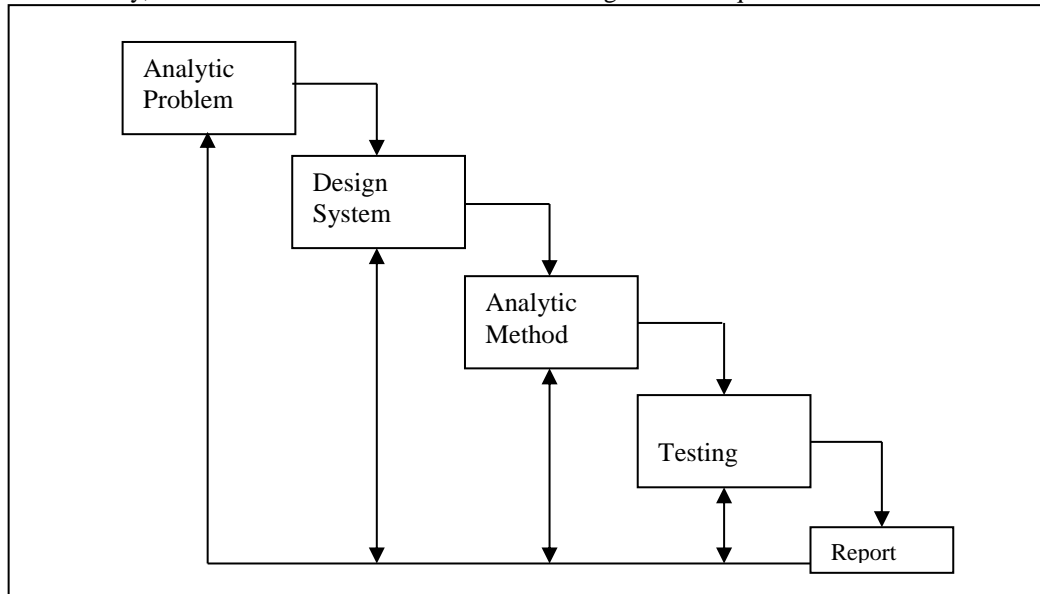


Fig 1. Research Method

Each stage will be explained as follows:

a. Problem Analysis Phase

The problem analysis stage is the process of identifying the problems that occur so that an idea can be concluded to be solved. In this study, the analysis of the problem is the grouping of Type I or Type II diabetes based on age. The data obtained will be adapted in the form of technology and interview techniques to obtain data on diabetic patients at the Royal Prima Hospital in Medan. Then it is determined that the point of the problem of diabetes which has two types will be grouped based on age using the Kmeans method.

b. Design Stage

In System Design here, namely the design of the solution to the problem. For this reason, the researchers used the Kmeans data mining method in classifying Type I and Type II diabetes based on age. Kmeans is quite often used in research for grouping in the medical field.

c. Method Analysis Phase

Kmeans method manages data in the form of age, patient name, rm number, date of admission, type of disease, gender which will be processed to generate momentum for grouping diabetes. Diabetes which has two types, namely type I diabetes and type II diabetes, is grouped based on age. Then it will show what age is the dominant type I or type II diabetes sufferer.

d. Testing Stage

In the first test stage, the researcher carried out without the participation of other parties. Testing is done by comparing to other studies with almost the same case. Where its function is to ensure the use of calculations is correct.

e. Report Stage

The final stage of this research is report generation. The report is prepared in accordance with the provisions contained in the writing guidelines.

3. Result and Discussion

In this study, the results obtained were in the category of diabetic patients in type I or type II. The data obtained is medical record data for patients at Royal Prima Hospital in 2019. The following is medical record data for patients with diabetes taken from RSU Royal Prima Dari in 2019. From 200 patient data, researchers have summarized by age. Number of patients of the same age and diagnosed with Diabetes Mellitus.

Table 1.
Data on RM Patients that have been summarized

Data To-	Month of RM Patient Year 2019	M to-	Number Of DM I [X]	Number Of DM II [Y]
1	47 years old	M ₁	10	8
2	67 years old	M ₂	12	11
3	63 years old	M ₃	10	5
4	35 years old	M ₄	5	0
5	43 years old	M ₅	15	13
6	56 years old	M ₆	18	15
7	49 years old	M ₇	13	19
8	68 years old	M ₈	10	15
9	62 years old	M ₉	2	8
10	55 years old	M ₁₀	4	7

The next process is to find the distance of each object to each centroid using the correlation formula between two objects, namely Euclidean Distance. Mn= (x,y)

M₁ = (10,8)

M₂ = (12,11)

M₃ = (10,5)

M₄ = (5,0)

M₅ = (15,13)

M₆ = (18,15)

M₇ = (13,19)

M₈ = (10,15)

M₉ = (2,8)

M₁₀=(4,7)

3.1 Creating clusters

The first step in determining the cluster is to take the centroid point at random. In this case the centroid point taken is as follows:

	X	Y
Centroid 1	10	8
Centroid 2	18	15

According to the journal by Johan Oscar Ong [11], the formula for calculating the distance of the data to the centroid point is calculated using the Euclidean Distance formula, namely:

$$D_{(i,j)} = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (Xki - Xkj)^2} \dots \dots \dots (1)$$

After going through iteration calculations, then grouping each data based on the shortest distance between the data (object) and its centroid (central point of the cluster).

Table 2.
Initialization Table The results of the calculation of the distance of each data for each - each cluster.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
C1	1	1	0	0	1	1	0	1	1	0
C2	0	0	1	1	0	0	1	0	0	1

Based on table 2 above, it can be seen that the results of the cluster in iteration 2 and the cluster in iteration 2 are the same. So the process stops here. The process will continue if the cluster members C1 and C2 in



iteration 1 and iteration 2 are different. The results of the cluster in the second iteration can be seen in the following table:

Table 3. Result

C1	47 Years Old
	67 Years Old
	43 Years Old
	56 Years Old
	62 Years Old
	68 Years Old
C2	63 Years Old
	35 Years Old
	49 Years Old
	55 Years Old

3.2 Discussions

There have been several studies conducted related to the k-Means method and diabetes, including: Abdi Raja's research uses 2 methods, namely K-Means and Fuzzy C-Means, in his research only grouping people/patients who suffer from diabetes and non-diabetes, so that the final result found is only information about diabetics and those who do not. suffer from diabetes [12]. Meanwhile, Andika Prasatya in his research also uses 2 methods, namely the combination of the K-Means method with C4.5. K-Means and was used only to predict HbA1c in terms of HbA1c examination [13]. Therefore, from previous research, there was no discussion about the type/type of diabetes and there was no discussion about the Age/Age classification of diabetics based on the type or type of diabetes, allowing the author to conduct this research so that it can be useful information. for the world of health and education.

This study uses data derived from medical record reports at RSU Royal Prima in 2019. The parameters used are the age of patients in the adult and elderly categories every month from January to December who suffer from diabetes. The cluster process is carried out to determine the type of diabetes, which has 2 (two) types, namely type I and type II. From the previous results, the next discussion is the results found in the form of clustering as follows:

- a. The hierarchical cluster process using the K-means method is able to show the relationship between each patient.
- b. The results of the cluster_K-means analysis of 10 types of patients based on age/age classification which can be grouped into two cluster_k-means, namely:
 - 1) Cluster_k-means1, is a group of patients with Type I DM, which is more than half the number of patients.
 - 2) Cluster_k-means2, is a group of DM Type II . patients

From the lack of a running system, this study applied data mining analysis using Clusstring using the K-Means algorithm. So that it is easy to determine and classify diabetic patients with the provisions of Type I DM and Type II DM.

4. Conclusion

Based on the description of the explanation and discussion of the overall material in the previous chapters, the following conclusions can be drawn:

- a. The system for classifying Type I and Type II diabetes based on age range can help Royal Prima Hospital in analyzing patients quickly and accurately in identifying diabetes.
- b. The results of the cluster_K-means analysis of 200 patients based on age/age classification which can be grouped into two cluster_k-means, namely Cluster_k-means1, is a group of Type I DM patients, which is more than half the number of patients. Cluster_k-means2, is a group of DM Type II. Patients.

The author provides several suggestions that are expected to improve the performance of the proposed system as consideration for Royal Prima Hospital, including:

- a. It is hoped that further research in Diabetes Mellitus or other diseases with the Kmeans method will add to the analysis of scientific data in the field of medicine for hospitals.
- b. According to the author, the system that the author proposes does not yet have complete features. Therefore, it is necessary to add additional features such as features for transactions in the calculation method.

5. References

- [1] Agusta, Y. (2007). K-Means–Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, 3(1), 47-60.
- [2] Agustina, S., Yhudo, D., Santoso, H., Marnasusanto, N., Tirtana, A., & Khusnu, F. (2012). Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means. Universitas Brawijaya Malang, Malang.
- [3] Asroni, A., & Adrian, R. (2015). Penerapan metode K-means untuk clustering mahasiswa berdasarkan nilai akademik dengan Weka Interface studi kasus pada jurusan Teknik Informatika UMM Magelang. *Semesta Teknika*, 18(1), 76-82.
- [4] Ediyanto, M. N. M., & Satyahadewi, N. (2013). Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. *Bimaster*, 2(02).
- [5] Handoko, K. (2016). Penerapan Data Mining dalam Meningkatkan Mutu Pembelajaran Menggunakan Metode K-MEANS Clustering. *Jurnal Nasional Teknologi dan Sistem Informasi*, 2(3), 31-40.
- [6] METISEN, Benri Melpa; SARI, Herlina Latipa. Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila. *Jurnal media infotama*, 2015, 11.2
- [7] Sani, A. (2018). Penerapan metode k-means clustering pada perusahaan. *Jurnal Ilmiah Teknologi Informasi*, 353, 1-7.
- [8] Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 4(1), 85-92.
- [9] Suprawoto, T. (2016). Klasifikasi data mahasiswa menggunakan metode k-means untuk menunjang pemilihan strategi pemasaran. *JIKO (Jurnal Informatika dan Komputer)*, 1(1).
- [10] Waworuntu, M. N. V., & Amin, M. F. (2018). Penerapan Metode K-Means Untuk Pemetaan Calon Penerima Jamkesda. *KLIK-Kumpulan Jurnal Ilmu Komputer*, 5(2), 190-200.
- [11] BASTIAN, A. (2018). Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). *Jurnal Sistem Informasi*, 14(1), 28-34. <https://doi.org/10.21609/jsi.v14i1.566>.
- [12] Meisida, Novita; soesanto, Oni; candra, Heru Kartika. K-Means untuk Klasifikasi Penyakit Karies Gigi. *klik-kumpulan Jurnal Ilmu Komputer*, 2017, 1.1: 12-22.
- [13] Sugianto, Castaka Agus; rahayu, Ayu Hendrati; gusman, Aditia. Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Cigugur Tengah. *Journal of Information Technology*, 2020, 2.2: 39-44.

