

# Pengelompokkan Data Rekam Medis pada Penyakit Diabetes menggunakan Metode *Divisive Analysis Clustering*

## *Clustering Medical Record Data on Diabetes Disease using Divisive Analysis Clustering Method*

<sup>1</sup>Mayang Safhira Tarigan\*, <sup>2</sup>Muhammad Siddik Hasibuan

<sup>1,2</sup>Program Studi Ilmu Komputer, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Sumatera Utara, Kota Medan, Sumatera Utara, Indonesia.

\*e-mail: [mayangstrg13@gmail.com](mailto:mayangstrg13@gmail.com)

(received: 28 June 2024, revised: 21 July 2024, accepted: 23 July 2024)

### Abstrak

Diabetes merupakan penyakit kronis yang ditandai dengan tingginya kadar gula (glukosa) dalam yang melebihi batas normal, sekitar 1,5 juta kematian setiap tahun secara langsung dikaitkan dengan diabetes. Tujuan dari penelitian ini Memperoleh hasil analisis *cluster* optimal pada pengelompokkan data rekam medis pada penyakit diabetes dan memperoleh nilai *silhouette coefficient* hasil analisis *cluster* optimal pada pengelompokkan data rekam medis pada penyakit diabetes di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara dengan menggunakan algoritma *divisive analysis* serta menentukan variabel yang menjadi penyebab pasien tersebut mengalami diabetes. Hasil penelitian menunjukkan *Cluster* optimal dengan menggunakan algoritma *divisive analysis* adalah 2 *cluster* dengan nilai sebesar 0,468582 yang menyatakan bahwa struktur *cluster* yang terbentuk pada pengelompokkan ini adalah klaster yang tepat. Dan untuk Variabel yang menjadi penyebab pasien bisa mengalami diabetes yaitu Usia ( $X_2$ ) dan Tingkat gula darah ( $X_8$ ). hal itu dikarenakan dua variabel tersebut yang memiliki nilai rata-rata tertinggi diantara dua cluster dan semua variabel tersebut.

**Kata kunci:** pengelompokan data, rekam medis, diabetes, *divisive analysis clustering*

### Abstract

Diabetes is a chronic disease characterized by high levels of sugar (glucose) in excess of normal limits, about 1.5 million deaths each year are directly attributed to diabetes. The purpose of this study is to obtain the optimal cluster analysis results on grouping medical record data on diabetes and obtain the silhouette coefficient value of the optimal cluster analysis results on grouping medical record data on diabetes at TK-II Putri Hijau Kesdam I / B Hospital, Medan-North Sumatra by using the divisive analysis algorithm and determining the variables that cause the patient to develop diabetes. The results showed that the optimal cluster using the divisive analysis algorithm was 2 clusters with a value of 0.468582 which stated that the cluster structure formed in this grouping was the right cluster. And for variables that cause patients to develop diabetes, namely age ( $X_2$ ) and blood sugar level ( $X_8$ ). it is because these two variables have the highest average value among the two clusters and all these variables.

**Keywords:** data clustering, medical records, diabetes, *divisive analysis clustering*

## 1 Pendahuluan

Diabetes merupakan penyakit kronis yang ditandai dengan tingginya kadar gula (glukosa) dalam yang melebihi batas normal [1][2]. Menurut situs resmi *World Health Organization* (WHO), sekitar 422 juta orang diseluruh dunia menderita diabetes, dengan mayoritas dari mereka tinggal di negara-negara berpenghasilan rendah dan menengah. Bahkan lebih mencemaskan, sekitar 1,5 juta kematian setiap tahun secara langsung dikaitkan dengan diabetes. Indonesia juga tidak terlepas dari beban penyakit diabetes, yang jumlah kasusnya tidak kalah banyak. Data dari sumber databox

menunjukkan bahwa Indonesia berada di posisi ke-5 dalam daftar negara-negara dengan kasus diabetes tertinggi setelah Amerika Serikat [3][4].

Pada dasarnya, *clustering* atau pengklasteran adalah proses *data mining* untuk melihat pola pendistribusian data yang akan digunakan untuk melihat karakteristik dari data [5][6]. Dalam *data mining* ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu pengelompokan hierarki (*hierarchical clustering*) dan pengelompokan non hierarki (*non hierarchical clustering*) [7] [8]. Sedangkan untuk analisis klaster adalah salah satu analisis multivariat yang dapat digunakan untuk meringkas dan mempermudah interpretasi data [9].

Penelitian sebelumnya yang dilakukan oleh Kasoqi [10] menerapkan metode yang sama. Akan tetapi, perbedaan dari penelitian sebelumnya dengan penelitian yang akan saya lakukan adalah dari data yang akan saya ambil untuk bahan penelitian saya tersebut. Pada penelitian sebelumnya menjelaskan bahwasanya penelitian ini dilakukan untuk pengelompokan desa atau kelurahan di Kutai Kartanegara. Dimana, penelitian ini memperoleh *cluster optimal* yang terbentuk pada pengelompokan desa/kelurahan di Kabupaten Kutai Kartanegara berdasarkan variabel yang diduga sebagai penentu ketertinggalan desa dengan menggunakan *divisive analysis* adalah sebanyak 2 *cluster* yaitu *cluster 1* dan *cluster 2*. *Cluster 1* beranggotakan 230 desa/kelurahan dan *cluster 2* beranggotakan 2 desa/kelurahan. Dan untuk nilai *silhouette coefficient* yang bertujuan untuk validasi data hasil *clustering* desa/kelurahan di Kabupaten Kutai Kartanegara yaitu berdasarkan variabel yang diduga sebagai penentu ketertinggalan desa dengan menggunakan metode *divisive analysis* menghasilkan 2 *cluster* yang terbentuk pada pengelompokan ini adalah *strong structure*.

Pada penelitian ini, akan dipaparkan suatu metode klaster, yaitu, *Divisive Analysis Clustering* tujuannya untuk memperoleh hasil analisis *cluster optimal* pada pengelompokan data rekam medis pada penyakit diabetes dan memperoleh nilai *silhouette coefficient* hasil analisis *cluster optimal* pada pengelompokan data rekam medis pada penyakit diabetes di Rumah Sakit [11]. Metode ini sendiri memiliki arti yaitu proses pengklasteran yang didasarkan pada persamaan nilai rata-rata antar objek [12]. Jika sebuah objek memiliki persamaan nilai rata-rata terbesar, maka objek tersebut akan terpisahkan dan berubah menjadi *splinter group* [13].

## 2 Tinjauan Literatur

Konfigurasi dari kajian terdahulu untuk penelitian “Pengelompokan Data Rekam Medis Pada Penyakit Diabetes Menggunakan Metode *Divisive Analysis Clustering*” di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara dapat dibagi menjadi dua bagian, yaitu: penelitian dengan judul Pengelompokan Desa Atau Kelurahan di Kutai Kartanegara Menggunakan Algoritma *Divisive Analysis* [10], Penelitian ini memperoleh *cluster optimal* yang terbentuk pada pengelompokan desa/kelurahan di Kabupaten Kutai Kartanegara berdasarkan variabel yang diduga sebagai penentu ketertinggalan desa dengan menggunakan *divisive analysis* adalah sebanyak 2 *cluster* yaitu *cluster 1* dan *cluster 2*. *Cluster 1* beranggotakan 230 desa/kelurahan dan *cluster 2* beranggotakan 2 desa/kelurahan. Dan untuk nilai *silhouette coefficient* yang bertujuan untuk validasi data hasil *clustering* desa/kelurahan di Kabupaten Kutai Kartanegara yaitu berdasarkan variabel yang diduga sebagai penentu ketertinggalan desa dengan menggunakan metode *divisive analysis* menghasilkan 2 *cluster* yang terbentuk pada pengelompokan ini adalah *strong structure*.

Selanjutnya penelitian dengan judul “Pemetaan Zona Resiko Penularan Covid-19 di Sulawesi Selatan Menggunakan Plot Dendrogram Hierarchical Clustering” [14], Hasil akhir dari karya tulis ini adalah tersampainya peta zonasi dan plot dendrogram hierarchical clustering tingkat provinsi, sampai dengan tingkat kabupaten kota di Provinsi Sulawesi Selatan. Adanya peta tematik yang komprehensif memuat clustering kabupaten/kota, dan karakteristik kewilayahan akan dapat menjadi early warning (peringatan dini) stakeholder decision maker terkait penanganan pandemic Covid-19 khususnya di Provinsi Sulawesi Selatan. Dengan zonasi clustering berjenjang, dapat diketahui kabupaten/kota yang memiliki kemiripan karakteristik, yaitu kasus kumulatif Covid-19 sejauh apa saat ini, dan angka kepadatan penduduknya seperti apa, juga jauh atau tidaknya dari pusat penularan Covid-19 di Provinsi Sulawesi Selatan (epicentrum).

Perbandingan Antara Metode Agglomeratif, Metode Divisif dan Metode K-Means Dalam Analisis Klaster [13], Hasil pengelompokan untuk keenam metode agglomeratif, divisif dan metode k-means pada contoh data sampel: sepuluh data observasi tingkat polusi udara di Amerika Serikat

memberikan hasil 5 atau 6 jumlah kluster. Hal ini didukung dengan validasi dengan plot RMSSTD dan CD beserta plot RS dan SPR untuk masing-masing metode agglomeratif. Pengelompokan kesepuluh kota ditunjukkan pada tabel 3 dan 4, untuk jumlah kluster ( $k$ ) = 5 atau 6.

Dengan demikian, konfigurasi dari ketiga kajian terdahulu yaitu menunjukkan bahwa ada kekurangan dan menjadi dasar untuk mengembangkan penelitian selanjutnya pada bidang yang sama. Dari ketiga itu juga, dapat dilihat bahwasanya masih sedikit yang membahas tentang metode *Divisive Analysis Clustering* (DIANA). Hal itu dikarenakan metode Agglomerative lebih mudah dibandingkan metode DIANA. maka tujuan dari penelitian ini adalah: Memperoleh hasil analisis *cluster* optimal pada pengelompokan data rekam medis pada penyakit diabetes di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara dengan menggunakan algoritma *divisive analysis*. Dan memperoleh nilai *silhouette coefficient* hasil analisis *cluster* optimal pada pengelompokan data rekam medis pada penyakit diabetes di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara dengan menggunakan algoritma *divisive analysis* dan menentukan variabel yang menjadi penyebab pasien tersebut mengalami diabetes.

### 3 Metode Penelitian

Proses pengambilan data dilakukan di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara, yang beralamatkan di Jalan Putri Hijau, No. 17 Medan, Sumatera Utara. Penelitian ini dilakukan dengan mengumpulkan variabel faktor resiko kelompok pasien diabetes. Kemudian, diikuti dengan pengumpulan data sampel responden yaitu data rekam medis pada pasien yang berkunjung.

#### 3.1 Divisive Analysis Clustering (DIANA)

*Divisive Analysis Clustering* (DIANA) adalah teknik pengelompokan hierarki, dimana hal ini dapat membangun hierarkinya dalam urutan terbalik sehingga dikenal sebagai pendekatan top-down. Metode ini juga membuat kelompok dengan membagi data awal dan merupakan kebalikan dari algoritma *Agglomerative Hierarchical Clustering*, dimana data yang memakai metode ini dikelompokkan secara *agglomerative*. Pada proses metode DIANA ini awalnya seluruh objek dikelompokkan pada *cluster* yang sama, kemudian dilanjutkan dengan pemisahan *cluster* secara rekursif sampai data individu telah dipecah menjadi *cluster* tunggal [11]. Pengelompokan *divisive analysis* juga merupakan metode pengelompokan hierarki yang memecah belah dengan menggunakan strategi *top-down*, dimana data dipecah menjadi *sub-cluster* yang lebih kecil hingga setiap data memiliki kluster sendiri. Tahapan dari pengelompokan *divisive analysis* dalam penelitian ini adalah sebagai berikut [10]:

1. Melakukan standarisasi data.
2. Mendeteksi multikolinieritas dengan menghitung *Variance Inflation Factor* (VIF) dari setiap variabel. Nilai VIF dapat dihitung dengan menggunakan persamaan (1) sebagai berikut.

$$VIF_j = \frac{1}{1 - R_j^2} \quad (1)$$

Dimana  $R_j^2$  adalah koefisien determinasi dari variabel bebas  $X_j$  yang diregresikan terhadap variabel bebas lainnya. Dikatakan terjadi multikolinieritas apabila nilai VIF melebihi 10, yang berarti estimasi yang akan dilakukan kurang baik. Salah satu cara yang paling mudah dalam mengatasi multikolinieritas ialah dengan mengeluarkan salah satu variabel yang memiliki nilai VIF paling tinggi.

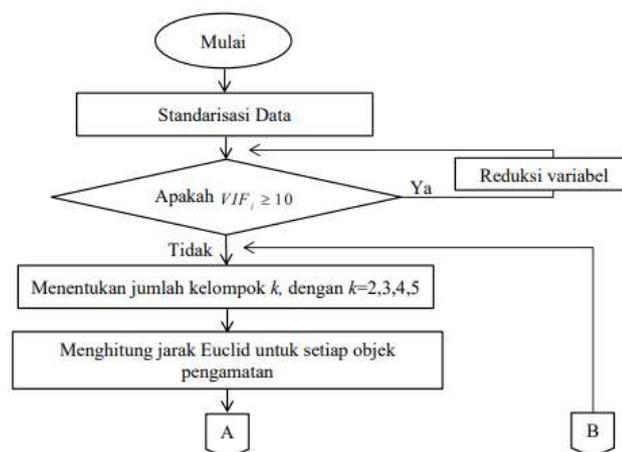
3. Melakukan pengelompokan data dengan menerapkan algoritma *divisive analysis* untuk setiap  $k = 2, 3, 4$ , dan 5 dari metode *elbow*, yaitu salah satu metode yang merupakan proses evaluasi sekaligus juga digunakan untuk menentukan nilai  $k$  optimum dari perulangan proses uji nilai  $k$ . pencarian nilai  $k$  optimum dilakukan dengan melakukan perbandingan SSE (*Sum of Square Error*) yang disajikan dalam bentuk grafik. Nilai  $k$  optimum akan ditemukan saat kondisi *elbow* (patahan) telah ditemukan pada grafik tersebut [12]. Proses tersebut dapat dilihat pada lampiran 7. Setelah itu akan dilakukan tahapan sebagai berikut:

- a. Menghitung jarak *Euclid* dengan menggunakan persamaan (2).

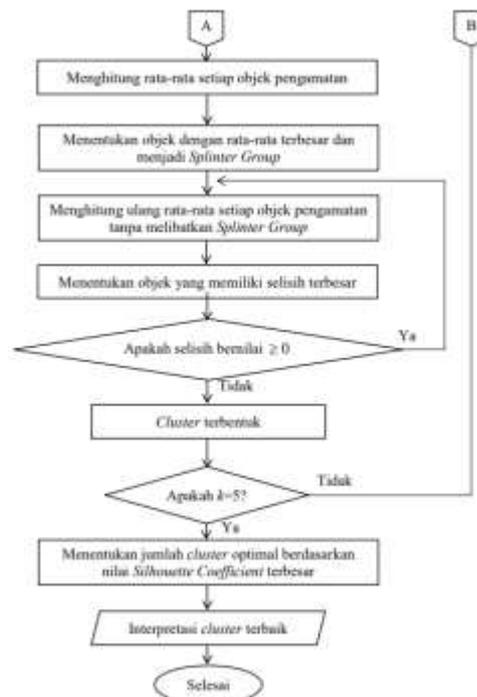
$$d(i, r) = \sqrt{(x_{i1} - x_{r1})^2 + (x_{i2} - x_{r2})^2 + \dots + (x_{iq} - x_{rq})^2} \quad (2)$$

- b. Menghitung rata-rata setiap objek berdasarkan matriks jarak *Euclid* dengan menggunakan persamaan (3).
- $$\bar{d}_i = \frac{1}{n-1} \sum_{r=1}^n d(i,r) , r \neq i \quad (3)$$
- c. Menentukan objek yang memiliki nilai rata-rata yang terbesar dan berubah menjadi *splinter group*.
  - d. Menghitung selisih nilai antara elemen matriks *splinter group* dengan nilai rata-rata setiap objek yang tersisa.
  - e. Mengulang langkah (b) sampai (d) sedemikian, sehingga semua nilai selisih antara elemen matriks *splinter group* dengan nilai rata-rata bernilai negative dan *cluster* terbagi menjadi dua cluster baru.
4. Menghitung nilai *silhouette coefficient* dan menentukan nilai *k* optimal berdasarkan nilai *silhouette coefficient* terbesar[10].

Dan berikut merupakan flowchart DIANA ditampilkan dalam Gambar 1 dan 2.



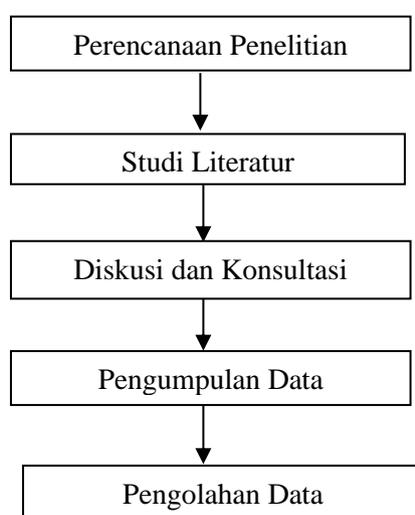
Gambar 1. Flowchart DIANA



Gambar 2. Flowchart DIANA (Lanjutan)

Hal ini dimulai dengan menempatkan semua objek dalam satu *cluster* besar, yang merupakan *root* hierarki. Kemudian membagi *root cluster* menjadi beberapa sub-*cluster* yang lebih kecil, dan secara berulang memecah *cluster* tersebut menjadi yang lebih kecil. Proses partisi berlanjut sampai setiap *cluster* di level terendah yang hanya mengandung satu objek, atau objek dalam sebuah *cluster* cukup mirip satu sama lain atau bisa disebut dengan *splinter group* [10].

Kerangka penelitian adalah merupakan langkah yang dilakukan dalam menyelesaikan masalah yang dibahas, digunakan untuk membantu peneliti dalam menyusun kerangka yang jelas tahapannya [15]. Terdapat beberapa tahapan yang harus dilalui untuk melakukan penelitian ini, ditampilkan dalam Gambar 3.



Gambar 3. Kerangka penelitian

### 3.2 Rencana Pembahasan

Perancangan yaitu tahap dari perancangan, penggambaran dan pembuatan sketsa atau pengetahuan dari beberapa elemen yang terpisah ke dalam satu kesatuan yang utuh dan berfungsi memberikan gambaran yang jelas kepada instalasi yang membutuhkan. Perancangan model ini terdiri dari metode jarak *euclidean* yang ada di algoritma *Divisive Analysis Clustering* (DIANA).

Diabetes merupakan penyakit yang paling sering ditemui dan banyak pasien yang menderita di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara. Sehingga jenis penyakit ini yang paling memungkinkan untuk dilakukannya identifikasi. Pengidentifikasian terhadap kelompok pasien diabetes dilakukan berdasarkan dua faktor yang ditemukan, yaitu: faktor yang tidak dapat diubah dan faktor resiko yang dapat diubah. Berikut merupakan variabel penelitian dari faktor-faktor resiko diatas, antara lain:

- Jenis kelamin ( $X_1$ ), dimana pasien diabetes tersebut berjenis kelamin laki-laki atau perempuan. Jika pasien tersebut laki-laki, maka akan bernilai 1 dan jika pasien tersebut perempuan, maka akan bernilai 0 pada data tersebut.
- Usia ( $X_2$ ), dimana pada variabel ini menunjukkan usia pada pasien yang mengidap penyakit diabetes.
- Hipertensi ( $X_3$ ), dimana variabel ini apakah pasien tersebut mengalami tekanan darah tinggi atau tidak. Apabila pasien tersebut mengalami tekanan darah tinggi maka akan bernilai 1. Dan apabila pasien tersebut tidak mengalami tekanan darah tinggi maka akan bernilai 0 pada data yang tersedia.
- Penyakit jantung ( $X_4$ ), dimana pada variabel ini apakah pasien tersebut mengalami penyakit jantung atau tidak. Apabila pasien tersebut mengalami penyakit jantung maka akan bernilai 1. Dan apabila pasien tersebut tidak mengalami penyakit jantung maka akan bernilai 0 pada data yang tersedia.
- Riwayat merokok ( $X_5$ ), dimana pada variabel ini apakah pasien mengalami riwayat merokok seperti saat melakukan pemeriksaan pasien tersebut perokok aktif, pernah merokok, mantan perokok, tidak pernah merokok, ataupun tidak ada informasi dari pasien tersebut. Maka, dari kelima macam diatas saya akan membagi kedua kategori. Yaitu apabila pasien tersebut

- perokok aktif, pernah merokok, ataupun mantan perokok maka akan bernilai 1. Dan apabila pasien tersebut tidak pernah merokok, ataupun tidak ada informasi dari pasien tersebut maka akan diberi nilai 0 pada data tersebut.
- BMI ( $X_6$ ), dimana pada variabel ini berisikan informasi tentang *Body Mass Index* yang mana isinya berfungsi untuk mengetahui berat badan pada pasien tersebut ideal atau tidak. Hal ini dilakukan berdasarkan berat dan tinggi badan pada pasien.
  - Tingkat HbA1c ( $X_7$ ), adalah pemeriksaan untuk mengukur rata-rata kadar gula darah selama tiga bulan terakhir pada pasien. Dan pada data ini akan ditunjukkan berapa tingkat HbA1c pada setiap pasien yang diperiksa.
  - Tingkat gula darah ( $X_8$ ), dimana pada variabel ini berisikan informasi tingkat gula darah pada setiap pasien.
  - Diabetes ( $X_9$ ), dimana penentuan apakah pasien tersebut mengalami diabetes atau tidak. Jika pasien mengalami diabetes, maka akan diberikan nilai 1. Dan apabila pasien tidak mengalami diabetes, maka akan diberikan nilai 0 pada data tersebut.

Pada variabel diatas merupakan gambaran data yang diperlukan didalam penelitian metode DIANA untuk mengelompokkan data rekam medis pasien yang beresiko diabetes berdasarkan atribut-atribut yang relevan. Metode DIANA dilakukan untuk menempatkan semua objek dalam satu *cluster* besar, yang merupakan *root* hierarki. Kemudian membagi *root cluster* menjadi beberapa sub-*cluster* yang lebih kecil, dan secara berulang memecah *cluster* tersebut menjadi yang lebih kecil. Proses partisi berlanjut sampai setiap *cluster* di level terendah yang hanya mengandung satu objek, atau objek dalam sebuah *cluster* cukup mirip satu sama lain atau bisa disebut dengan *splinter group*.

## 4 Hasil dan Pembahasan

Pada tahap ini akan dijelaskan hasil penelitian yang didapatkan, untuk lebih rincinya sebagai berikut:

### 4.1 Pembahasan

Data yang digunakan dalam penelitian ini adalah data rekam medis pada penyakit diabetes di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara. Data tersebut terdiri dari 232 pasien yaitu sebagai berikut di jelaskan dalam Tabel 1.

**Tabel 1. Data pasien diabetes awal**

Jenis Kelamin	Usia	Hipertensi	Penyakit Jantung	Riwayat Merokok	Bmi	Tingkat HbA1c	Tingkat Gula Darah	Diabetes
Perempuan	80.0	0	1	Tidak pernah	25.19	6.6	140	0
Perempuan	54.0	0	0	Tidak ada informasi	27.32	6.6	80	0
Laki-laki	28.0	0	0	Tidak pernah	27.32	5.7	158	0
...	...	...	...	...	...	...	...	...
Perempuan	76.0	0	0	Tidak ada informasi	27.32	5.0	160	0

Data akan dianalisis menggunakan metode *divisive* dalam *clustering*, dimana jumlah *cluster* yang akan digunakan adalah 2, 3, 4, dan 5 *cluster*. Evaluasi hasil *clustering* menggunakan metode *silhouette coefficient*. Variabel yang digunakan pada penelitian ini ada 9 dengan jenis data numerik, dan berikut variabel yang akan digunakan dalam penelitian ini: Jenis kelamin ( $X_1$ ), Usia ( $X_2$ ), Hipertensi ( $X_3$ ), Penyakit jantung ( $X_4$ ), Riwayat merokok ( $X_5$ ), BMI ( $X_6$ ), Tingkat HbA1c ( $X_7$ ), Tingkat gula darah ( $X_8$ ), Diabetes ( $X_9$ ).

Variabel diatas diubah dengan menggunakan sintaks *colnames* pada *rstudio*, agar nantinya dapat memudahkan penulis untuk melakukan proses perhitungan tersebut. Kemudian untuk variabel X1 dan X5 diubah dengan menggunakan sintaks *if* (Lampiran 2), agar nantinya dapat memudahkan penulis untuk melakukan proses perhitungan tersebut. Dan untuk data yang sudah diubah, di jelaskan dalam Tabel 2.

**Tabel 2. Data pasien setelah diubah**

X1	X2	X3	X4	X5	X6	X7	X8	X9
0,00	80,00	0,00	1,00	0,00	25,19	6,60	140,00	0,00

0,00	54,00	0,00	0,00	0,00	27,32	6,60	80,00	0,00
1,00	28,00	0,00	0,00	0,00	27,32	5,70	158,00	0,00
0,00	36,00	0,00	0,00	1,00	23,45	5,00	155,00	0,00
...	...	...	...	...	...	...	...	...
0,00	76,00	0,00	0,00	0,00	27,32	5,00	160,00	0,00

Setelah itu, kita akan melakukan proses standarisasi data. Dimana hal tersebut merubah nilai data asli menjadi bentuk *Z-Score* dengan tujuan untuk membuat rentang nilai yang pada semua variabel. Untuk *syntax*, dapat dilihat pada lampiran 5. Dan untuk hasil standarisasi, dalam Tabel 3.

**Tabel 3. Hasil standarisasi data**

X1	X2	X3	X4	X5	X6	X7	X8	X9
- 0,75174 5406	1,69 3274 591	- 0,233 0458	4,9669 88727	- 0,5230 11581	- 0,3034 84585	0,85223 7937	- 0,08757 6177	- 0,30648 4888
- 0,75174 5406	0,53 9867 803	- 0,233 0458	- 0,2004 61428	- 0,5230 11581	0,0084 32933	0,85223 7937	- 1,51839 8218	- 0,30648 4888
1,32450 3811	- 0,61 3538 985	- 0,233 0458	- 0,2004 61428	- 0,5230 11581	0,0084 32933	0,03634 027	0,34167 0436	- 0,30648 4888
- 0,75174 5406	- 0,25 8644 588	- 0,233 0458	- 0,2004 61428	1,9037 62156	- 0,5582 90445	- 0,59824 6805	0,27012 9334	- 0,30648 4888
...	...	...	...	...	...	...	...	...
- 0,75174 5406	1,51 5827 393	- 0,233 0458	- 0,2004 61428	- 0,5230 11581	0,0084 32933	- 0,59824 6805	0,38936 4504	- 0,30648 4888

#### 4.2 Hasil

Berikut merupakan hasil statistika deskriptif pada data rekam medis pada penyakit diabetes di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara di jelaskan dalam Tabel 4.

**Tabel 4. Statistika deskriptif**

Variabel	Rata-rata	Simpangan Baku	Nilai Minimum	Nilai Maksimum
X <sub>1</sub>	0,37	0,48	0	1
X <sub>2</sub>	41,83	22,54	0,1	80,00
X <sub>3</sub>	0,05	0,22	0	1
X <sub>4</sub>	0,04	0,19	0	1
X <sub>5</sub>	0,20	0,40	0	1
X <sub>6</sub>	27,26	6,83	13,79	63,48
X <sub>7</sub>	5,66	1,10	3,50	9,00
X <sub>8</sub>	143,67	41,93	80	300
X <sub>9</sub>	71,88	0,28	0	1

Berdasarkan pada tabel 4.1, dapat dilihat pada variabel jenis kelamin rata-rata yang dimiliki adalah 0,37. Selain itu, simpangan baku dari variabel ini adalah 0,48. Dan untuk nilai minimum di variabel ini adalah 0, sedangkan untuk nilai maksimum pada variabel ini adalah 1. Dimana pada variabel ini menjelaskan bahwasanya pasien yang berjenis kelamin laki-laki lebih banyak dibandingkan pasien yang berjenis kelamin perempuan.

Variabel usia biasanya termasuk faktor terpenting pada penentuan apakah pasien tersebut mengalami diabetes atau tidak. Rata-rata pada variabel ini adalah 41,83 dan simpang baku pada variabel ini adalah 22,54. Nilai minimum pada variabel ini adalah 0,1 dan nilai maksimum pada variabel ini adalah 80,00. Dimana, pada variabel ini dijelaskan bahwasanya bayi yang berusia 1 bulan pun bisa

mengalami yang namanya diabetes. Begitupun untuk yang berusia 80 tahun pun sangat besar kali untuk mengalami yang namanya diabetes.

Variabel selanjutnya dalam hal menentukan apakah pasien tersebut memiliki diabetes selanjutnya adalah hipertensi atau tekanan darah pasien tersebut tinggi atau tidak. Rata-rata pada variabel ini adalah 0,05 dan simpang baku pada variabel ini adalah 0,22. Nilai minimum pada variabel ini adalah 0 dan nilai maksimum pada variabel ini adalah 1.

Setelah itu, ada variabel penyakit jantung. Rata-rata pada variabel ini adalah 0,04 dan simpang baku pada variabel ini adalah 0,19. Nilai minimum pada variabel ini adalah 0 dan nilai maksimum pada variabel ini adalah 1.

Selanjutnya adalah riwayat merokok yang dapat memicu pasien mengalami diabetes. Rata-rata pada variabel ini adalah 0,20 dan simpang baku pada variabel ini adalah 0,40. Nilai minimum pada variabel ini adalah 0 dan nilai maksimum pada variabel ini adalah 1.

Sama halnya dengan variabel diatas, variabel BMI atau *Body Mass Index* juga dapat menentukan apakah pasien tersebut mengalami diabetes atau tidak. Rata-rata pada variabel ini adalah 27,26 dan simpang baku pada variabel ini adalah 6,83. Nilai minimum pada variabel ini adalah 13,79 dan nilai maksimum pada variabel ini adalah 63,48. Dan pada variabel ini, dapat kita lihat bahwasnya pasien yang memiliki nilai BMI dibawah standar BMI pun bisa mengalami diabetes, apalagi pasien obesitas.

Kemudian ada variabel tingkat HbA1c. Rata-rata pada variabel ini adalah 5,66 dan simpang baku pada variabel ini adalah 1,10. Nilai minimum pada variabel ini adalah 3,50 dan nilai maksimum pada variabel ini adalah 9,00. Dan pada variabel ini, dapat kita lihat bahwasnya pasien yang memiliki nilai BMI dibawah standar BMI pun bisa mengalami diabetes, apalagi pasien obesitas.

Variabel selanjutnya yaitu hal yang terpenting dalam menentukan apakah pasien tersebut memiliki diabetes selanjutnya adalah hipertensi atau tingkat gula darah, dikarenakan variabel ini merupakan variabel yang memiliki nilai tertinggi diantara semua variabel diatas. Rata-rata pada variabel ini adalah 143,67 dan simpang baku pada variabel ini adalah 41,93. Nilai minimum pada variabel ini adalah 80 dan nilai maksimum pada variabel ini adalah 300.

Dan variabel terakhir yaitu variabel yang menyatakan apakah pasien tersebut diabetes atau tidak. Rata-rata pada variabel ini adalah 71,88 dan simpang baku pada variabel ini adalah 0,28. Nilai minimum pada variabel ini adalah 0 dan nilai maksimum pada variabel ini adalah 1.

Sebelum memasuki tahap analisis *cluster*, terlebih dahulu dilakukan standarisasi data dan selanjutnya dilakukan pendektasian multikolinieritas. Salah satu metode untuk mendeteksi multikolinieritas adalah dengan menghitung *Variance Inflation Factor (VIF)* dari setiap variabel. Nilai VIF dapat dihitung dengan menggunakan persamaan (1).

$$VIF_1 = \frac{1}{1 - R_1^2} = \frac{1}{1 - 0,4172148} = 1,043538$$

$$VIF_2 = \frac{1}{1 - R_2^2} = \frac{1}{1 - 0,242417} = 1,319987$$

...

$$VIF_9 = \frac{1}{1 - R_1^2} = \frac{1}{1 - 0,4450387} = 1,801927$$

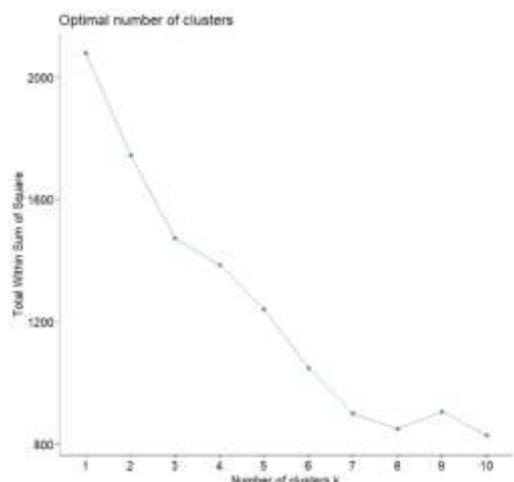
Nilai VIF dapat dilihat secara lengkap pada Tabel 5.

**Tabel 5. Nilai VIF setiap variabel**

Variabel	VIF
X <sub>1</sub>	1,043538
X <sub>2</sub>	1,319987
X <sub>3</sub>	1,118546
X <sub>4</sub>	1,155483
X <sub>5</sub>	1,106961
X <sub>6</sub>	1,101127
X <sub>7</sub>	1,362814
X <sub>8</sub>	1,365798
X <sub>9</sub>	1,801927

Dalam penelitian ini didapatkan hasil bahwa tidak terjadi multikolinieritas pada semua variabel, dikarenakan nilai VIF pada masing-masing variabel kurang dari 10 ( $VIF < 10$ ) dan dapat dilanjutkan ke proses pengelompokkan dengan metode *divisive analysis*.

Kemudian kita akan menentukan ada berapa cluster yang akan dihitung dengan menggunakan metode *divisive* dalam *clustering*. Hal ini menggunakan metode *elbow* pada Rstudio untuk dapat hasil yang akurat, dan dapat dilihat pada Gambar 4.



**Gambar 4. Cluster dengan menggunakan elbow**

Dan dapat kita lihat, bahwasanya hasil analisis *elbow* menunjukkan dataset dikelompokkan pada awal 2 *cluster* dan berakhir pada 5 *cluster*. Hal itu dikarenakan kurva menurun dari *cluster* 2 ke *cluster* 3 dan membentuk siku yang tajam, kemudian turun menuju ke *cluster* 4 dan 5 yang membentuk kurva landai. Sehingga dapat disimpulkan bahwa dataset berdasarkan *elbow* menunjukkan adanya 5 *cluster* yang akan dipakai untuk perhitungan *divisive analysis*.

Selanjutnya, data dianalisis menggunakan metode *divisive* dalam *clustering*. Pertama, menghitung jarak *Euclid* menggunakan persamaan (1). Perhitungan jarak *Euclid* untuk objek ke-2 dan objek ke-1, dan begitupun hingga objek ke-232 dan objek ke-231 adalah sebagai berikut.

$$d(2,1) = \sqrt{(x_{2,1} - x_{1,1})^2 + \dots + (x_{2,8} - x_{1,8})^2 + (x_{2,9} - x_{1,9})^2}$$

$$= \sqrt{(0 - 0)^2 + \dots + (80 - 140)^2 + (0 - 0)^2} = 5,49339900$$

Setelah itu, menghitung nilai rata-rata jarak *Euclid* setiap objek menggunakan persamaan (2). Perhitungan nilai rata-rata jarak *Euclid* setiap objek dilanjutkan hingga objek ke-232 adalah sebagai berikut:

$$\bar{d}_i = \frac{1}{232-1} (73,495+22,113+\dots+12,309) = 6,150506$$

Hasil perhitungan dapat dilihat pada Tabel 6.

**Tabel 1. Nilai rata-rata jarak euclid setiap objek**

D	1	2	3	...	232	Rata-Rata
1	0,00000000	5,49339900	6,10589670	...	5,93611010	6,150506
2	5,49339900	0,00000000	3,12517240	...	2,26083570	3,376598
3	6,10589700	3,12517240	0,00000000	...	2,44170790	3,120053
...	....	....	...	....	...	....
232	5,93611000	2,26083570	2,44170790	...	0,000000	3,168895

Selanjutnya, menentukan objek yang memiliki nilai rata-rata terbesar. Objek yang memiliki nilai rata-rata terbesar akan dipisah dan berubah menjadi *splinter group*. Berdasarkan tabel 4.2, objek yang memiliki nilai rata-rata yang terbesar adalah objek ke-5 yaitu sebesar 7,733492 sehingga objek tersebut menjadi *splinter group* 1. Kemudian menghitung ulang nilai rata-rata setiap objek yang tersisa tanpa melibatkan objek ke-5 yang telah menjadi *splinter group* 1. Setelah itu, menghitung nilai elemen matriks *splinter group* 1 dengan nilai rata-rata setiap objek yang tersisa. Hasil perhitungan dapat dilihat pada Tabel 7 berikut.

**Tabel 2. Selisih antara nilai elemen matriks splinter group 1 dengan nilai rata-rata setiap objek**

D	Rata-rata	Splinter 1 (Objek ke-5)	Selisih
1	3,883786088	5,43675700	-1,55297091
2	3,8739602	7,79816500	-3,92420480
3	3,876206185	7,96951200	-4,09330582
...	...	...	...
232	3,148441788	7,87303800	-4,72459621

Selanjutnya, menentukan objek yang memiliki nilai selisih terbesar antara elemen matriks *splinter group* dengan nilai rata-rata. Jika nilai selisih tersebut bernilai positif, maka objek yang memiliki nilai terbesar bergabung dengan *splinter group*. Berdasarkan tabel 4.4, objek ke-43 dan 120 memiliki nilai selisih terbesar dan positif yaitu 0,83131117 dan 1,30512410. Sehingga objek tersebut bergabung dengan *splinter group* 1. Kemudian, menghitung selisih nilai antara elemen matriks *splinter group* 1 dengan nilai rata-rata setiap objek yang tersisa. Hasil perhitungan dapat dilihat pada Tabel 8.

**Tabel 3. Selisih antara nilai elemen matriks splinter group 1 dengan nilai rata-rata setiap objek yang tersisa**

D	Rata-rata	Splinter 1 (Objek ke-43)	Splinter 1 (Objek ke-120)	Selisih
1	3,851563797	4,86079000	5,75292000	-6,76214620
2	3,841566648	7,36005100	7,80355900	-11,32204335
3	3,843832819	7,54423200	7,55152700	-11,25192618
...	...	...	...	...
232	3,10735255	7,63821600	8,02701400	-12,55787745

Selanjutnya, mengulangi tahapan diatas sampai nilai selisih terbesar antara elemen matriks *splinter group* 1 dengan nilai rata-rata yang tersisa bernilai negative semua. Karena ada beberapa objek yang memiliki nilai selisih terbesar dan negative, maka penulis memasukkan salah satu nilai ke dalam Tabel 5.

Proses iterasi untuk 2 *cluster* dihentikan sampai terbentuk *splinter group* 2. Untuk pembentukan 3 *cluster*, proses iterasi berhenti pada *splinter group* 3. Begitupun hingga pembentukan 5 *cluster*. Hasil pengelompokkan *divisive analysis* untuk 2, 3, 4, dan 5 *cluster* dapat dilihat pada Tabel 9,10,11,12,13,14,15 dan 16 berikut.

**Tabel 4. Hasil pengelompokkan menggunakan algoritma DIANA**

No.	Cluster
1	1
5	1
24	1
27	1
...	...
23	2

**Tabel 5. Hasil pengelompokkan menggunakan algoritma DIANA lanjutan**

No.	Cluster
25	2
26	2
28	2
30	2
...	...
63	2

**Tabel 6. Hasil pengelompokkan menggunakan algoritma DIANA lanjutan**

No.	Cluster
64	2
65	2
67	2
68	2

...	...
96	2

**Tabel 7. Hasil pengelompokan menggunakan algoritma DIANA lanjutan**

No.	Cluster
99	2
100	2
101	2
...	...
132	2

**Tabel 8. Hasil pengelompokan menggunakan algoritma DIANA Lanjutan**

No.	Cluster
133	2
134	2
135	2
138	2
...	...
167	2

**Tabel 9. Hasil pengelompokan menggunakan algoritma DIANA Lanjutan**

No.	Cluster
169	2
171	2
172	2
173	2
...	...
199	2

**Tabel 10. Hasil pengelompokan menggunakan algoritma DIANA lanjutan**

No.	Cluster
201	2
202	2
203	2
204	2
...	...
230	2

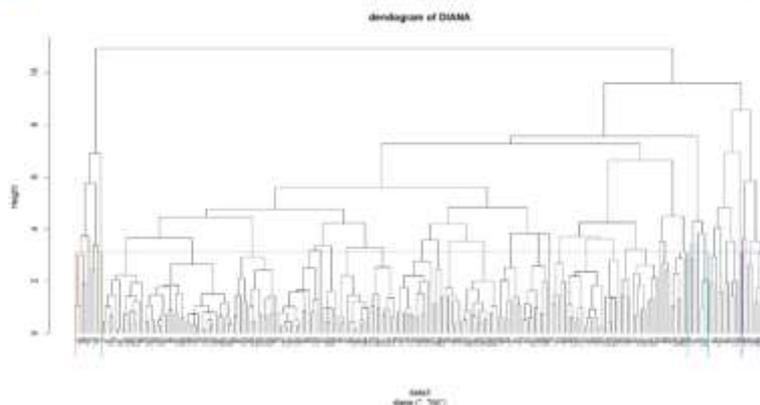
**Tabel 11. Hasil pengelompokan menggunakan algoritma DIANA lanjutan**

No.	Cluster
231	2
232	2
7	3
41	3
...	...
200	5

Berdasarkan Tabel diatas, dapat diketahui hasil algoritma *Divisive Analysis* pada pembentukan 2, 3, 4, dan 5 klaster. Pada pembentukan 2 klaster didapatkan hasil berupa ada 197 data yang termasuk kepada klaster tersebut, yaitu data ke- 2, 3, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 42, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 57, 58, 59, 62, 63, 64, 65, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 96, 98, 99, 100, 101, 102, 103, 104, 106, 107, 108, 109, 110, 111, 112, 113, 114, 116, 117, 118, 119, 121, 122, 123, 124, 125, 128, 129, 130, 131, 132, 133, 134, 135, 138, 139, 140, 142, 144, 145, 146, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 164, 165, 166, 167, 169, 171, 172, 173, 174, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 227, 228, 229, 230, 231, dan 232.

Pada pembentukan 3 kluster didapatkan hasil berupa ada 11 data yang termasuk kepada kluster tersebut, yaitu data ke- 7, 41, 54, 56, 60, 82, 97, 136, 137, 170, dan 175. Pada pembentukan 4 kluster didapatkan hasil berupa 7 data yang termasuk kepada kluster tersebut yaitu data ke- 29, 55, 61, 115, 127, 141, dan 226. Pada pembentukan 5 kluster didapatkan hasil berupa ada 8 data, yaitu data ke- 39, 88, 95, 105, 126, 43, 47, dan 200 yang termasuk kepada kluster tersebut.

Dan Gambar 5 adalah tampilan plot dendrogram dari hasil kluster diatas menggunakan algoritma *Divisive Analysis* yang telah dijelaskan oleh penulis diatas ditampilkan dalam Gambar 5.



**Gambar 5. Plot dendrogram algoritma DIANA**

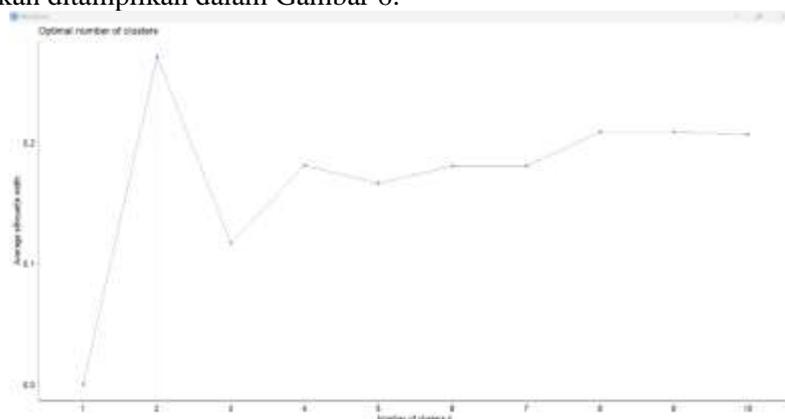
Dapat kita lihat pada gambar 4.2 tersebut, garis yang berwarna hijau pekat lah yang paling banyak kluster-nya dibandingkan kluster yang lain seperti pada tabel sebelumnya.

Setelah melewati tahapan dalam DIANA, maka hal selanjutnya yang harus dilakukan adalah evaluasi hasil *clustering* dengan menggunakan metode *Silhouette Coefficient*. Hal ini dilakukan untuk mengetahui kualitas dari hasil pengelompokkan yang sudah dilakukan, dan berikut adalah hasil *Silhouette Coefficient* nya di jelaskan dalam Tabel 17.

**Tabel 12. Nilai silhouette coefficient**

Jumlah Cluster	SC
2	<b>0,468582</b>
3	0,468582
4	0,4300823
5	0,4091697

Berdasarkan Tabel 17, dapat dilihat bahwasanya nilai SC untuk evaluasi pengelompokkan data rekam medis pada penyakit diabetes di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara dengan menggunakan algoritma *divisive analysis* memiliki nilai SC yang berbeda-beda. Nilai SC yang terbesar adalah pengelompokkan pada 2 *cluster*, yaitu 0,468582. Dan berikut untuk grafik hasil SC yang telah dilakukan ditampilkan dalam Gambar 6.



**Gambar 6. Cluster dengan menggunakan nilai silhouette coefficient**

Dan dapat dilihat juga dari Gambar 6 bahwasanya titik kluster berada pada kluster 2. Oleh karena itu dapat diputuskan bahwa pengelompokkan optimal pada pengelompokkan data rekam medis pada

penyakit diabetes di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara dengan menggunakan algoritma *divisive analysis* adalah 2 cluster.

#### 4.3 Karakteristik Hasil Analisis Cluster

Berdasarkan hasil pengelompokan untuk 2 cluster, dilakukan perhitungan nilai rata-rata variabel untuk melihat karakteristik dari masing-masing cluster di jelaskan dalam Tabel 18.

**Tabel 13. Nilai rata-rata variabel setiap cluster**

Variabel	Cluster 1	Cluster 2
X <sub>1</sub>	0,3609	0,50
X <sub>2</sub>	41,57	72
X <sub>3</sub>	0,04348	1
X <sub>4</sub>	0,03043	1
X <sub>5</sub>	0,2087	1
X <sub>6</sub>	27,29	23,73
X <sub>7</sub>	5,667	4,90
X <sub>8</sub>	143,7	140,5
X <sub>9</sub>	0,08696	0

Berdasarkan Tabel 18, dapat disimpulkan bahwa ada dua penyebab pasien bisa mengalami diabetes yaitu Usia (X<sub>2</sub>) dan Tingkat gula darah (X<sub>8</sub>). hal itu dikarenakan dua variabel tersebut yang memiliki nilai rata-rata tertinggi diantara dua cluster dan semua variabel tersebut. Dan alasan mengapa cluster 1 dimasukkan juga yaitu karena pengelompokan ada 2 cluster, dimana cluster 1 pun termasuk ke dalam cluster tersebut.

## 5 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, maka dapat diambil kesimpulan adalah *Cluster optimal* yang terbentuk pada pengelompokan data rekam medis pada penyakit diabetes di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara dengan menggunakan algoritma *divisive analysis* adalah 2 cluster. Nilai *silhouette coefficient* optimal untuk validasi data hasil *clustering* rekam medis pada penyakit diabetes di Rumah Sakit TK-II Putri Hijau Kesdam I/BB, Medan-Sumatera Utara dengan menggunakan algoritma *divisive analysis* menghasilkan 2 cluster dengan nilai sebesar 0,468582 yang menyatakan bahwa struktur *cluster* yang terbentuk pada pengelompokan ini adalah klaster yang tepat. Dan untuk Variabel yang menjadi penyebab pasien bisa mengalami diabetes yaitu Usia (X<sub>2</sub>) dan Tingkat gula darah (X<sub>8</sub>). hal itu dikarenakan dua variabel tersebut yang memiliki nilai rata-rata tertinggi diantara dua cluster dan semua variabel tersebut.

## Referensi

- [1] A. M. Argina, "Penerapan Metode *Klasifikasi K-Nearest Neighbor* pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [2] A. Simanjorang, Megawati, and Agusdiana, "Analisis Pengaruh Kualitas Pelayanan Kesehatan Terhadap Kepuasan Pasien Ruang Penyakit Dalam di RSUD Dr. Fauziah Bireuen," vol. 2, no. 2, pp. 116–130, 2019.
- [3] M. A. Wiratama and W. M. Pradnya, "Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 1, p. 1, 2022, doi: 10.23887/janapati.v11i1.45282.
- [4] Y. Yosiana Nabuasa, "Analisis Dan Perancangan Sistem Informasi Rekam Medis Sebagai Sarana Pelayanan Kesehatan Pada Puskesmas Oesapa Kota Kupang," *J. Sist. Inf.*, vol. 2, no. 1, pp. 40–46, 2021.
- [5] M. Rosadi, D. Aulia Nurhasanah, and M. Siddik Hasibuan, "Clustering Panjang Ruas Jalan di BBPJN Sumut Menggunakan *Algoritma K-Means*," *J. Comput. Sci. Informatics Eng.*, vol. 02, no. 1, pp. 29–38, 2023, doi: 10.55537/cosie.v2i1.567.
- [6] I. Nur and L. Fitriana, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Keluarga Sehat Menggunakan Metode Klaster Hirarki dan Non Hirarki," *J. Paradig. J. Multidisipliner Mhs. Pascasarj. Indones.*, vol. 2, no. 1, pp. 27–36, 2021.
- [7] N. Nirwana and M. Nor Hayati, "Categoric Data Grouping By Algorithm Quick Robust Clustering Using Links (QROCK)(Case Study: Status of Value Addrd Tax Payments at the <http://sistemasi.ftik.unisi.ac.id>

- Samarinda Ulu Primary Tax Office in 2018),” vol. 9, no. 1, 2021.
- [8] M. Z. Nasution and M. S. Hasibuan, “Pendekatan Initial *Centroid Search* untuk Meningkatkan Efisiensi Iterasi *Klustering K-Means*,” *Techno.Com*, vol. 19, no. 4, pp. 341–352, 2020, doi: 10.33633/tc.v19i4.3875.
- [9] S. N. Dwiyantri, K. Nisa, A. Sutrisno, N. Herawati, and P. Lampung, “Analisis Kluster untuk Data Kategorik Menggunakan Metode K-Modes ( Studi Kasus : Data Pasien COVID-19 di RSUD Dr . H . Abdul Moeloek,” *J. Siger Mat.*, vol. 03, no. 02, pp. 61–68, 2022.
- [10] I. A. Kasoqi, M. N. Hayati, and R. Goejantoro, “Pengelompokan Desa Atau Kelurahan Di Kutai Kartanegara Menggunakan *Algoritma Divisive Analysis*,” *J. Stat. Univ. Muhammadiyah Semarang*, vol. 9, no. 2, p. 101, 2021, doi: 10.26714/jsunimus.9.2.2021.101-108.
- [11] M. R. Gunawan, “Perbandingan Pengelompokan pada Wilayah Berdasarkan Tingkat Kerusakan Lingkungan Di Indonesia Menggunakan Analisis Kluster K- Means, Hirarki, Dan Ensemble,” vol. 33, no. 1, pp. 1–12, 2022.
- [12] A. W. Fuadah, F. N. Arifin, and O. Juwita, “Optimasi K-Klasterisasi Ketahanan Pangan Kabupaten Jember Menggunakan Metode Elbow,” *INFORMAL Informatics J.*, vol. 6, no. 3, p. 136, 2021, doi: 10.19184/isj.v6i3.28363.
- [13] D. Rachmatin and K. Sawitri, “Perbandingan antara metode agglomeratif, metode divisif dan metode k-means dalam analisis kluster,” *Semin. Nas. Mat. UNPAR*, vol. 1, pp. 9–17, 2019.
- [14] M. I. Mubarak and A. Rusyiana, “Pemetaan Zona Resiko Penularan Covid-19 Di Sulawesi Selatan Menggunakan Plot Dendrogram Hierarchical Clustering,” *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 55–64, 2021, doi: 10.34123/semnasoffstat.v2020i1.441.
- [15] M. S. Hasibuan, H. Harahap, U. Pembangunan, P. Budi, U. Islam, and N. Sumatera, “Penerapan *Metode Haar-Like Feature* dan *Algoritma Adaboost* Dalam Penentuan Klasifikasi Hama,” vol. 4307, no. 1, pp. 87–93, 2024.