

**PENERAPAN METODE *K-MEDOIDS* UNTUK PENGELOMPOKAN
JENIS *DIABETES MELLITUS* PADA RUMAH SAKIT
MARDI WALUYO METRO**

Skripsi

Oleh

FILIPUS ARDI PRADANA



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2026

ABSTRACT

APPLICATION OF THE K-MEDOIDS METHOD FOR CLUSTERING DIABETES MELLITUS TYPES IN HOSPITAL MARDI WALUYO METRO

By

Filipus Ardi Pradana

Diabetes mellitus is one of the non-communicable diseases whose prevalence continues to increase every year, both globally and nationally. The variation in patient characteristics, such as gender and types of complications, necessitates an analytical method capable of grouping patients into homogeneous clusters. This study aims to classify types of diabetes mellitus using a non-hierarchical cluster analysis method, namely K-Medoids. This method is chosen because it is more robust to the influence of outliers compared to the K-Means method.

Cluster analysis is a method used to group objects based on the similarity of their characteristics. In clustering diabetes mellitus patient data, the presence of outliers often becomes a challenge as it can affect the clustering results.

Keywords: K-Medoids, cluster analysis, outliers, diabetes mellitus.

ABSTRAK

PENERAPAN METODE *K-MEDOIDS* UNTUK PENGELOMPOKAN JENIS *DIABETES MELLITUS* PADA RUMAH SAKIT MARDI WALUYO METRO

Oleh

Filipus Ardi Pradana

Penyakit diabetes mellitus merupakan salah satu penyakit tidak menular yang prevalensinya terus meningkat setiap tahun, baik secara global maupun nasional. Variasi karakteristik pasien, seperti jenis kelamin dan jenis komplikasi, menyebabkan perlunya metode analisis yang mampu mengelompokkan pasien ke dalam kelompok yang homogen. Penelitian ini bertujuan untuk mengelompokkan jenis *diabetes mellitus* menggunakan metode analisis kluster *non-hierarki*, yaitu *K-Medoids*. Metode ini dipilih karena lebih tahan terhadap pengaruh nilai pencilan dibandingkan metode K-Means.

Analisis kluster merupakan metode yang digunakan untuk mengelompokkan objek berdasarkan tingkat kemiripan karakteristik. Dalam pengelompokan data pasien diabetes mellitus, keberadaan nilai pencilan sering menjadi kendala karena dapat mempengaruhi hasil klusterisasi.

Kata-kata kunci: *K-Medoids*, analisis kluster, outliers, *diabetes mellitus*.

**PENERAPAN METODE *K-MEDOIDS* UNTUK PENGELOMPOKAN
JENIS *DIABETES MELLITUS* PADA RUMAH SAKIT
MARDI WALUYO METRO**

FILIPUS ARDI PRADANA

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2026

Judul Skripsi : **PENERAPAN METODE *K-MEDOIDS* UNTUK PENGELOMPOKAN JENIS *DIABETES MELLITUS* PADA RUMAH SAKIT MARDI WALUYO METRO**

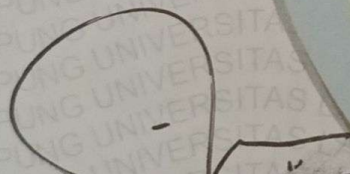
Nama Mahasiswa : **Filipus Ardi Pradana**

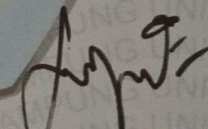
Nomor Pokok Mahasiswa : **2217031118**

Program Studi : **Matematika**

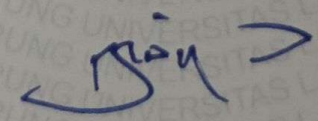
Fakultas : **Matematika dan Ilmu Pengetahuan Alam**




Dr. Subian Saidi, S.Si., M.Si.
NIP 198008212008121001


Misgiyati, S.Pd., M.Si.
NIP 198509282023212032

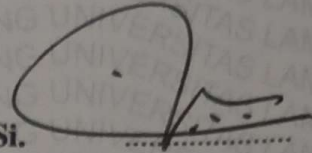
2. Wakil Dekan Bidang Akademik dan Kerjasama,
FMIPA Universitas Lampung


Mulyono, S.Si., M.Si., Ph.D.
NIP. 197406112000031002

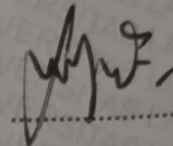
MENGESAHKAN

1. Tim Penguji

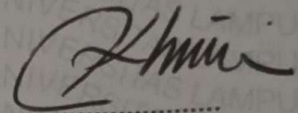
Ketua : Dr. Subian Saidi, S.Si., M.Si.



Sekretaris : Misgiyati, S.Pd., M.Si.



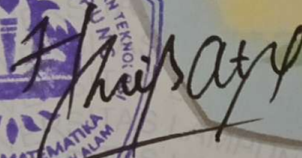
**Penguji
Bukan Pembimbing : Dr. Khoirin Nisa, S.Si., M.Si.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002



Tanggal Lulus Ujian Skripsi: 06 Mei 2026

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : **Filipus Ardi Pradana**
Nomor Pokok Mahasiswa : **2217031118**
Jurusan : **Matematika**
Judul Skripsi : **Penerapan Metode *K-Medoids* untuk Pengelompokan Jenis *Diabetes Mellitus* Pada Rumah Sakit Mardi Waluyo Metro**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 06 Mei 2026

Penulis,



Filipus Ardi Pradana

RIWAYAT HIDUP

Penulis memiliki nama lengkap Filipus Ardi Pradana yang lahir di Kota Metro pada tanggal 20 Juli 2004. Penulis merupakan anak semata wayang dari kedua orang tua penulis yang bernama Ayah Dwi Aris Tuhono dan Ibu Dian Ika Mayasari.

Penulis mengawali pendidikan di Taman Kanak-Kanak (TK) Kristen 1 Metro pada tahun 2008. Kemudian penulis melanjutkan ke jenjang Sekolah Dasar (SD) di SD Kristen 1 Metro pada tahun 2010. Penulis memasuki jenjang Sekolah Menengah Pertama (SMP) pada tahun 2016, penulis menempuh pendidikan di SMP Negeri 4 Metro. Penulis kemudian melanjutkan pada jenjang selanjutnya di SMA Negeri 1 Metro. Penulis menempuh pendidikan S1 di Universitas Lampung.

Penulis juga melakukan magang pada tahun 2024 di Direktorat Jendral Pajak yang dilakukan selama 3 bulan kemudian melanjutkan Kuliah Kerja Praktik di Kantor Pelayanan Pajak Metro selama 2 bulan.

KATA INSPIRASI

“Janganlah hendaknya kamu kuatir tentang apa pun juga, tetapi nyatakanlah dalam segala hal keinginanmu kepada Allah dalam doa dan permohonan dengan ucapan syukur”

–Filipi 4:6–

”Serahkanlah segala kekuatiranmu kepada-Nya, sebab Ia yang memelihara kamu”

–1 Petrus 5:7–

”Karena itu Aku berkata kepadamu: Janganlah kuatir akan hidupmu, akan apa yang hendak kamu makan atau minum, dan janganlah kuatir pula akan tubuhmu, akan apa yang hendak kamu pakai. Bukankah hidup itu lebih penting dari pada makanan dan tubuh itu lebih penting dari pada pakaian?”

–Matius 6:25–

”Sebab Allah memberikan kepada kita bukan roh ketakutan, melainkan roh yang membangkitkan kekuatan, kasih dan ketertiban

–2 Timotius 1:7–

PERSEMBAHAN

Dengan mengucap Puji Tuhan atas Berkah dan Karunia Tuhan Yesus Kristus sehingga skripsi ini dapat terselesaikan dengan baik dan tepat pada waktunya. Dengan rasa syukur dan bahagia, saya persembahkan rasa terimakasih saya kepada:

Ayah dan Ibuku Tercinta

Terimakasih kepada orang tuaku atas segala pengorbanan, motivasi, doa dan rela hati serta dukungannya selama ini. Terimakasih telah memberikan pelajaran berharga kepada anakmu ini tentang makna perjalanan hidup yang sebenarnya sehingga kelak bisa menjadi orang yang bermanfaat bagi banyak orang.

Dosen Pembimbing dan Pembahas

Terimakasih kepada dosen pembimbing dan pembahas yang sudah sangat membantu, memberikan motivasi, memberikan arahan serta ilmu yang berharga.

Sahabat-sahabatku

Terimakasih kepada semua orang-orang baik yang telah memberikan pengalaman, semangat, motivasinya, serta doa-doanya dan senantiasa memberikan dukungan dalam hal apapun.

Almamater Tercinta

Universitas Lampung

SANWACANA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Penerapan Metode *K-Medoids* untuk Pengelompokan Jenis *Diabetes Mellitus* Pada Rumah Sakit Mardi Waluyo Metro" dengan baik dan lancar serta tepat pada waktu yang telah ditentukan.

Dalam proses penyusunan skripsi ini, banyak pihak yang telah membantu memberikan bimbingan, dukungan, arahan, motivasi serta saran sehingga skripsi ini dapat terselesaikan. Oleh karena itu, dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Bapak Dr. Subian Saidi, S.Si., M.Si. selaku Pembimbing 1 yang telah banyak meluangkan waktunya untuk memberikan arahan, bimbingan, motivasi, saran serta dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Ibu Misgiyati, S.Pd., M.Si. selaku Pembimbing II yang telah memberikan arahan, bimbingan dan dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si. selaku Penguji yang telah bersedia memberikan kritik dan saran serta evaluasi kepada penulis sehingga dapat menjadi lebih baik lagi.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Ibu Dr. Notiragayu, S.Si., M.Si. selaku dosen pembimbing akademik.
6. Seluruh dosen, staff dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Kedua orang tua penulis, Bapak Dwi Aris Tuhono dan Ibu Dian Ika Mayasari yang senantiasa memberikan dukungan dan arahan serta kasih sayang yang

tiada hentinya, khususnya pada proses penelitian ini ayah dan ibu selalu memberikan motivasi hingga penelitian ini selesai. Terima kasih atas pengorbanan dan perjuangan untuk penulis yang diberikan dengan sepenuh hati.

8. Kepada sahabat-sahabat penulis, Ruben dan Fuad yang sudah berteman sejak lama, terima kasih karena selalu hadir di setiap waktu yang membuat penulis merasa semangat dalam menempuh pendidikan di perguruan tinggi hingga selesai.

Semoga skripsi ini dapat bermanfaat bagi kita semua. Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, sehingga penulis mengharapkan kritik dan saran yang membangun untuk menjadikan skripsi ini lebih baik lagi.

Bandar Lampung, 06 Mei 2026

Filipus Ardi Pradana

DAFTAR ISI

	Halaman
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	3
1.3 Manfaat Penelitian	3
II. TINJAUAN PUSTAKA	4
2.1 Matriks	4
2.1.1 Definisi Matriks	4
2.1.2 Determinan Matriks	4
2.1.3 <i>Invers Matriks</i>	5
2.1.4 Nilai Eigen dan Vektor Eigen	5
2.1.5 Matriks Korelasi	6
2.2 Analisis Klaster	7
2.2.1 Nilai Z	7
2.2.2 Ukuran Kemiripan	7
2.2.3 Metode <i>K-Medoids</i>	8
2.3 Lebar <i>Silhouette</i>	10
2.4 Validasi Klaster	11
2.5 Analisis Komponen Utama	12
2.6 <i>Diabetes Mellitus</i>	13
III. METODOLOGI PENELITIAN	15
3.1 Waktu dan Tempat Penelitian	15
3.2 Data Penelitian	15
3.3 Metode Penelitian	17

IV. HASIL DAN PEMBAHASAN	18
4.1 Analisis Deskriptif	18
4.2 Pra Pemrosesan Data	26
4.2.1 Pemeriksaan Pencilan	26
4.2.2 Standarisasi Data	28
4.3 Analisis Korelasi Antar Variabel	28
4.4 Analisis Kluster dengan Metode <i>K-Medoids</i>	30
4.4.1 Menentukan Jumlah Kluster	30
4.4.2 Jarak Manhattan Setiap Pasangan Objek	32
4.4.3 Menghitung v_j untuk Objek j	36
4.4.4 Melakukan Pengklasteran Berdasarkan v_j	39
4.4.5 Perbarui <i>Medoid</i>	42
4.4.6 Pengklasteran dan Menghitung Jumlah Jarak Baru	43
4.5 Evaluasi Kluster	45
4.6 Interpretasi Hasil Kluster	49
V. KESIMPULAN DAN SARAN	58
5.1 Kesimpulan	58
5.2 Saran	58
DAFTAR PUSTAKA	59

DAFTAR TABEL

Tabel	Halaman
1. Jenis-jenis <i>diabetes mellitus</i> tahun 2023	18
2. Jenis-jenis <i>diabetes mellitus</i> tahun 2024	19
3. Data pasien <i>diabetes mellitus</i> tahun 2023 terstandarisasi	28
4. Data pasien <i>diabetes mellitus</i> tahun 2024 terstandarisasi	28
5. Matriks korelasi antar variabel pasien <i>diabetes mellitus</i> 2023	29
6. Matriks korelasi antar variabel pasien <i>diabetes mellitus</i> 2024	29
7. Nilai <i>R-Square</i> dan <i>silhouette</i> pada tiap banyaknya klaster	30
8. Nilai <i>R-Square</i> dan <i>silhouette</i> pada tiap banyaknya klaster	31
9. Jarak Manhattan antar objek	34
10. Jarak Manhattan antar objek	35
11. Nilai v_j untuk masing-masing objek	37
12. Nilai v_j untuk masing-masing objek	39
13. Jarak Manhattan antara <i>medoid</i> awal dan objek	40
14. Vektor klaster awal	41
15. Jarak Manhattan antara <i>medoid</i> awal dan objek	41
16. Vektor klaster awal	42
17. Jarak Manhattan antara <i>medoid</i> awal dan objek	43
18. Vektor klaster baru	44
19. Jarak Manhattan antara <i>medoid</i> dan objek	44
20. Vektor klaster baru	45
21. Jarak minimum pada <i>intercluster</i> data 2023	46
22. Jarak maksimum pada <i>intracluster</i> data 2023	46
23. Jarak minimum pada <i>intercluster</i> data 2024	47
24. Jarak maksimum pada <i>intracluster</i> data 2024	48
25. Tabel Klaster <i>diabetes mellitus</i> 2023	50
26. Tabel Klaster <i>diabetes mellitus</i> 2023	50

27. Mean variabel pada setiap klaster 2023	51
28. Karakteristik klaster 2023	51
29. Tabel Klaster <i>diabetes mellitus</i> 2024	54
30. Tabel Klaster <i>diabetes mellitus</i> 2024	55
31. Mean variabel pada setiap klaster 2024	55
32. Karakteristik klaster 2024	56

DAFTAR GAMBAR

Gambar	Halaman
1. Metode K-Medoids	9
2. Analisis komponen utama	13
3. Diagram Batang Jenis <i>diabetes mellitus</i> Tahun 2023	20
4. Diagram Batang Jenis Kelamin yang Mendominasi <i>diabetes mellitus</i> Tahun 2023	21
5. Diagram Batang Jenis <i>diabetes mellitus</i> Tahun 2024	22
6. Diagram Batang Jenis Kelamin yang Mendominasi <i>diabetes mellitus</i> Tahun 2024	23
7. Diagram Batang GDS dan tekanan darah <i>diabetes mellitus</i> Tahun 2023	24
8. Diagram Batang GDS dan tekanan darah <i>diabetes mellitus</i> Tahun 2024	25
9. Boxplot <i>diabetes mellitus</i> 2023	26
10. Boxplot <i>diabetes mellitus</i> 2024	27
11. Grafik nilai R-Square dan <i>silhouette diabetes mellitus</i> Tahun 2023	31
12. Grafik nilai R-Square dan <i>silhouette diabetes mellitus</i> Tahun 2024	32
13. Plot Klaster <i>diabetes mellitus</i> Tahun 2023	49
14. Plot Klaster <i>diabetes mellitus</i> Tahun 2024	53

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis kluster merupakan metode yang digunakan untuk mengelompokkan objek ke dalam kelompok yang memiliki kesamaan karakteristik, sehingga setiap kelompok dinamakan kluster. Objek-objek yang terdapat dalam satu kluster memiliki tingkat kemiripan yang tinggi, sementara dengan objek pada kluster lain perbedaannya cukup jelas. Dalam pengklasteran, objek hanya boleh masuk ke dalam satu kluster saja sehingga tidak terjadi tumpang tindih (Nugroho, 2008). Analisis kluster dapat dikatakan sebagai metode yang digunakan untuk mengelompokkan data ke dalam berbagai kelas. Sasaran dari analisis kluster adalah untuk memisahkan data sehingga objek yang berada dalam satu kelompok memiliki ciri yang sama, sedangkan objek-objek dalam kelompok lainnya memiliki perbedaan (Arbelaitz, *et al.*, 2013). Proses pengklasteran dilakukan dengan dua metode, yaitu metode hierarki dan metode non hierarki (Widyadhana, dkk., 2021).

Metode analisis kluster non hierarki yang umum digunakan adalah algoritma *K-Means*. *K-means* merupakan metode pengklasteran berbasis jarak yang membagi data ke dalam sejumlah kluster. Metode *K-Means* digunakan untuk menentukan kelompok data dengan menghitung *centroid*, yang merupakan titik tengah berdasarkan rata-rata dari data dalam kelompok tersebut (Yedla, *et al.*, 2010). Metode ini sering digunakan, terutama ketika data cenderung berdistribusi secara normal. Namun, *K-Means* memiliki kelemahan saat menangani data yang memiliki nilai-nilai pencilan (*outliers*), karena *centroid* dapat terpengaruh oleh nilai-nilai ekstrem ini. Sehingga, dalam kasus ini *K-Means* mungkin menghasilkan kelompok-kelompok yang tidak optimal. Metode *K-Medoids* atau *Partitioning Around Medoids* (PAM) merupakan metode kluster non-hierarki yang lebih tahan terhadap nilai pencilan (*outliers*) dibandingkan *K-Means*, karena menggunakan *medoid* sebagai pusat kluster. Metode ini pertama kali diperkenalkan oleh Kaufman

& Rousseeuw (1989). Metode *K-Medoids* merupakan varian umum dari metode *K-Means*. Perbedaan metode *K-Medoids* dari metode *K-Means*, yaitu pada pemilihan *medoid* atau nilai tengah sebagai pusatnya. *Medoid* adalah objek yang letaknya terpusat dalam suatu kelompok bertujuan untuk mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai ekstrim yang ada pada data set ((Sindi, dkk., 2020). *K-Medoids* merupakan metode yang tidak menghitung *centroid*, tetapi mencari titik-titik data yang paling representatif dalam kelompok (*medoid*). Dengan demikian, PAM atau *K-Medoids* lebih stabil terhadap nilai-nilai ekstrem dalam data sehingga cocok untuk situasi di mana *outliers* dapat mempengaruhi hasil analisis secara signifikan.

Penelitian mengenai *K-Medoids* dengan indikator penyakit *diabetes mellitus* sebelumnya pernah dilakukan Solikhun dkk.(2021), membahas peningkatan klusterisasi metode *K-Medoids* berbasis Jarak Manhattan. Hasilnya menunjukkan bahwa 238 data diklasifikasikan ke dalam kluster non diabetes dan 10 data diklasifikasikan ke dalam kluster diabetes. Dwi dkk.(2021), juga melakukan analisis kluster *K-Medoids* atau PAM berdasarkan data pasien *diabetes mellitus*. Hasilnya terdapat 3 kluster, kluster 1 adalah *diabetes mellitus* tidak bergantung pada insulin (tipe II) merupakan penyakit yang paling banyak diderita oleh pasien dengan total kasus sebanyak 435 kasus. Pada kluster 2, *Hipertensi Esensial* merupakan penyakit yang paling banyak diderita oleh pasien dengan total kasus sebanyak 2785 kasus. Pada kluster 3, *Vunus Laseratum, Puctum* merupakan penyakit yang paling banyak diderita oleh pasien dengan total kasus sebanyak 328 kasus.

Penyakit *diabetes mellitus* merupakan salah satu penyakit tidak menular yang jumlah penderitanya terus mengalami peningkatan setiap tahunnya, baik di tingkat global maupun nasional. Berdasarkan laporan *International Diabetes Federation (IDF)*, pada tahun 2021 tercatat sebanyak 537 juta orang di seluruh dunia mengidap diabetes, dan angka ini diproyeksikan akan meningkat menjadi 643 juta pada tahun 2030 (Hossain, *et al.*, 2024). Hal tersebut menunjukkan bahwa diabetes termasuk permasalahan utama di bidang kesehatan. Karakteristik penderita diabetes sangat bervariasi, mulai dari jenis kelamin hingga jenis komplikasi yang dialami. Variasi karakteristik ini yang mendorong perlunya pendekatan analitis untuk mengelompokkan pasien agar diperoleh pola dan informasi yang jelas. Berdasarkan penelitian-penelitian di atas, maka dalam penelitian ini peneliti ingin melakukan analisis kluster dengan metode *K-Medoids* atau PAM pada data pasien diabetes Rumah Sakit Mardi Waluyo.

1.2 Tujuan Penelitian

Tujuan penelitian ini adalah menerapkan metode analisis kluster PAM atau *K-Medoids* dalam mengelompokkan jenis penyakit *diabetes mellitus* di Rumah Sakit Mardi Waluyo Metro.

1.3 Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi terhadap perkembangan pengetahuan matematika dalam analisis kluster.
2. Menerapkan metode analisis kluster PAM pada proses pengelompokan jenis penyakit *diabetes mellitus* untuk mendukung perencanaan layanan kesehatan.
3. Dapat dijadikan referensi bagi penelitian lanjutan tentang klusterisasi data.

II. TINJAUAN PUSTAKA

2.1 Matriks

2.1.1 Definisi Matriks

Menurut Anton & Rorres, matriks merupakan kumpulan bilangan, simbol, atau ekspresi yang diatur bentuk baris dan kolom. Setiap bilangan dalam matriks disebut entri atau elemen matriks. Baris merupakan susunan elemen secara horizontal, sedangkan kolom adalah susunan elemen secara vertikal. Sebuah matriks A dengan m kolom dan n baris disebut matriks $m \times n$ yang dapat dituliskan dalam bentuk:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

2.1.2 Determinan Matriks

Determinan suatu matriks didefinisikan sebagai hasil selisih antara perkalian elemen-elemen pada diagonal utama dengan perkalian elemen-elemen pada diagonal sekunder. Operasi determinan matriks hanya berlaku untuk matriks berbentuk persegi. Determinan dari matriks A dapat dinyatakan dengan $\det(A)$ atau $|A|$ (Anton & Rorres, 1994). Bentuk umum determinan matriks A dapat dituliskan sebagai berikut:

$$\det(A) = |A| = \sum_{j=1}^n a_{ij} A_{ij}$$

dengan:

n : ukuran matriks

a_{ij} : elemen matriks pada baris ke- i dan kolom ke- j

A_{ij} : kofaktor dari elemen a_{ij}

2.1.3 Invers Matriks

Menurut Anton & Rorres (1994), invers matriks adalah matriks baru yang merupakan kebalikan dari matriks asal. Jika A adalah matriks persegi, dan jika terdapat matriks B yang ukurannya sama sedemikian rupa sehingga $AB = BA = I$, dengan I melambangkan matriks identitas $n \times n$, dan perkalian yang dilakukan merupakan perkalian matriks yang umum. Jika hubungan tersebut berlaku, maka matriks B disebut sebagai invers dari matriks A , dan diberi lambang A^{-1} . Adapun invers matriks A sebagai berikut:

$$A^{-1} = \frac{1}{\det(A)} \text{Adj}A$$

2.1.4 Nilai Eigen dan Vektor Eigen

Greenbaum *et al.* (2019), menyatakan bahwa nilai eigen (λ) dan vektor eigen (v) adalah konsep utama dalam aljabar linier yang menjelaskan bagaimana suatu matriks dapat bertindak terhadap sebuah vektor. Vektor eigen merupakan vektor tak nol yang arahnya tidak berubah ketika dikenai transformasi linear oleh suatu matriks persegi, melainkan hanya mengalami peregangan atau pengecilan ukuran. Faktor skalar yang menyatakan besarnya peregangan atau pengecilan disebut nilai eigen. Proses memperoleh nilai eigen dan vektor eigen dimulai dengan menyusun persamaan karakteristik dari suatu matriks A . Persamaan ini diperoleh dengan menyatakan hubungan dasar $Av = \lambda v$, yang dapat ditulis ulang menjadi:

$$(A - \lambda I)v = 0 \tag{2.1}$$

agar persamaan tersebut memiliki solusi non-trivial ($\mathbf{v} \neq 0$), maka determinan dari matriks $(\mathbf{A} - \lambda\mathbf{I})$ harus sama dengan nol, yaitu:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \quad (2.2)$$

Persamaan tersebut dapat dikatakan sebagai persamaan karakteristik. Penyelesaian dari persamaan tersebut akan menghasilkan nilai-nilai eigen (λ). Setelah nilai eigen diperoleh, langkah berikutnya adalah menentukan vektor eigen yang bersesuaian dengan cara menyelesaikan sistem persamaan linear. Dengan demikian, vektor eigen ditemukan sebagai himpunan solusi dari sistem persamaan homogen tersebut.

2.1.5 Matriks Korelasi

Matriks korelasi adalah matriks yang menunjukkan koefisien korelasi berpasangan antar variabel. Koefisien adalah angka yang mengukur kekuatan dan arah hubungan linear antara dua variabel. Nilainya berkisar antara -1 hingga 1, dimana -1 menunjukkan korelasi negatif sempurna dan 1 menunjukkan korelasi positif sempurna, 0 menunjukkan tidak ada korelasi. Graffelman dan de Leeuw (2023), menekankan bahwa matriks korelasi tidak hanya berfungsi sebagai kumpulan angka, tetapi juga sebagai alat visualisasi. Melalui teknik seperti biplot atau korelogram. Forrester dan Zhang (2020), mengatakan bahwa sebuah matriks korelasi harus memenuhi beberapa syarat penting, seperti diagonal utama bernilai satu, bersifat simetris dan positif semidefinit. Misal ada sebuah vektor acak y berukuran $n \times 1$, dan matriks kovarian yang dinotasikan dengan V . Maka, korelasi matriks dari y dinotasikan dengan $\mathbf{R} = [\rho_{ij}]$ dimana ρ_{ij} dapat ditulis sebagai berikut:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

Matriks korelasi dari \mathbf{A} berukuran $m \times n$ dapat ditulis sebagai berikut:

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & 1 \end{bmatrix}$$

2.2 Analisis Kluster

Analisis kluster adalah kumpulan teknik eksploratori untuk mengelompokkan objek ke dalam kelompok yang bersifat seragam didalamnya dan berbeda dengan kelompok lainnya. Tujuannya adalah untuk mengungkap struktur yang tidak terlihat dalam data tanpa adanya label sebelumnya, sehingga teknik ini sering diutilisasi dalam studi kesehatan untuk mengenali fenotipe penyakit, pengelompokan risiko, serta dukungan bagi pasien dengan kebutuhan klinis yang berbeda. Tinjauan naratif dan sistematis yang terkini menunjukkan penerapan metode kluster yang luas dalam konteks klinis dan epidemiologi (Lotfus, T., *et al.*, 2022). Proses pengelompokan objek dari setiap partisi memiliki kemiripan berdasarkan matriks tertentu, dalam proses inti *clustering* atau pengelompokan data terdapat dua metode yang bisa digunakan, yaitu metode hierarki dan metode non hierarki.

2.2.1 Nilai Z

Nilai Z merupakan metode statistik yang digunakan untuk mengukur hubungan antara suatu skor dengan rata-rata dalam suatu kelompok skor (Kumar & Rao, 2015). Nilai Z dapat didefinisikan sebagai berikut:

$$Z = \frac{X - \mu}{\sigma} \quad (2.3)$$

dengan:

Z : nilai standar

X : nilai yang akan distandarisasi

μ : rata-rata distribusi

σ : simpangan baku (standar deviasi) distribusi

2.2.2 Ukuran Kemiripan

Dalam analisis kluster ukuran kemiripan atau jarak memiliki peran yang sangat penting karena menjadi dasar utama dalam menentukan seberapa dekat atau jauhnya satu objek dengan objek lainnya. Secara konseptual, ukuran jarak menggambarkan

tingkat ketidaksamaan (*dissimilarity*) antara dua individu dalam satu ruang data. Semakin kecil nilai jarak yang diperoleh, semakin tinggi tingkat kemiripan dua objek tersebut. Menurut Shirktorshidi *et al.* (2015), pemilihan ukuran jarak yang tepat memiliki dampak signifikan terhadap hasil pengelompokan. Dalam penelitiannya ditunjukkan bahwa penggunaan jarak yang berbeda seperti Euclidean, Manhattan atau Mahalanobis dapat menghasilkan struktur kluster yang bervariasi, terutama pada data berdimensi tinggi. Hal ini menunjukkan bahwa pemilihan ukuran kemiripan tidak bisa dianggap sepele karena dapat mempengaruhi validitas hasil analisis. Penelitian ini menggunakan jarak Manhattan sebagai alat ukur kemiripan yang didefinisikan sebagai berikut:

$$d(i, j) = d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (2.4)$$

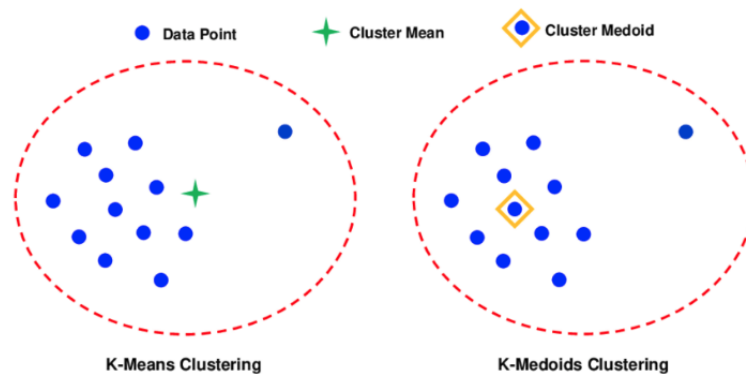
keterangan:

- d_{ij} : jarak antara objek ke-i pada objek ke-j
 x_{ik} : data dari objek ke-i pada variabel ke-k
 x_{jk} : data dari objek ke-j pada variabel ke-k
 (Suwanda, *et al.*, 2019)

2.2.3 Metode *K-Medoids*

Metode *K-Medoids* dikenal juga sebagai PAM (*Partitioning Around Medoids*) merupakan salah satu metode yang digunakan untuk proses klustering. Dalam metode ini data yang terdiri dari n objek di partisi menjadi k kluster di mana jumlah $k \leq n$. Oleh karena itu tujuannya adalah untuk menemukan k objek tersebut (Kaufman, L & Rousseeuw, P. J., 1989). Perbedaan metode *K-Medoids* dengan metode *K-Means* terletak pada penentuan pusat kluster, di mana metode *K-Medoids* nilai tengah dari data sebagai *medoids* atau perwakilan untuk pusat kluster. Sedangkan algoritma *K-Means* menggunakan nilai rata-rata dari setiap kluster sebagai pusat kluster (Kaur, *et al.*, 2014). Secara umum, *K-Medoids* bekerja dengan memilih sejumlah objek sebagai *medoid* awal, kemudian mengelompokkan objek lain berdasarkan kedekatannya terhadap *medoid* tersebut. Proses ini dilakukan secara iteratif dengan cara menukar *medoid* dengan objek lain yang bukan *medoid*, selama pertukaran tersebut dapat menurunkan total biaya atau total jarak dalam kluster. Konsep utama dari metode *K-Medoids* terletak pada pemilihan *medoid*

sebagai pusat kluster. *Medoid* didefinisikan sebagai objek data aktual yang memiliki total jarak minimum terhadap seluruh objek lain dalam kluster yang sama. Dengan demikian, *medoid* merupakan objek yang paling mampu merepresentasikan karakteristik kluster dibandingkan objek lainnya. Secara matematis, pemilihan *medoid* dilakukan dengan meminimalkan total jarak antara *medoid* dan seluruh anggota kluster. Secara umum, algoritma *K-Medoids* bekerja melalui beberapa tahapan, yaitu pemilihan *medoid* awal secara acak, pengalokasian setiap objek ke *medoid* terdekat berdasarkan ukuran jarak tertentu (misalnya Jarak Euclidean), kemudian dilakukan proses pertukaran (*swapping*) antara *medoid* dan *non-medoid* untuk memperoleh nilai total biaya yang lebih kecil. Berikut ini merupakan contoh gambar metode *K-Medoids* yang bersumber dari kaggle.



Gambar 1. Metode *K-Medoids*

Langkah-langkah analisis kluster *K-Medoids* (Gultom, *et. al.*, 2018) sebagai berikut:

1. Lakukan perhitungan jarak antara setiap pasangan objek berdasarkan ukuran jarak yang dipilih.
2. Menghitung nilai v_j untuk objek j dengan menggunakan persamaan:

$$v_j = \sum_{i=1}^n \frac{d_{ij}}{\sum_{l=1}^n d_{il}}; \quad j = 1, 2, 3, \dots, n \quad (2.5)$$

3. Urutkan nilai v_j dari yang terkecil hingga terbesar dan pilih k objek dengan nilai v_j sebagai *medoid* awal.
4. Hasilkan hasil awal dari pengklasteran dengan menentukan objek yang paling dekat dengan *medoid*.

5. Menghitung jumlah jarak dari setiap objek ke medoidnya.
6. Temukan *medoid* baru dari setiap kluster dengan cara meminimalkan jumlah jarak di dalam kluster tersebut dan perbaharui *medoid* dari setiap kluster dengan *medoid* yang baru.
7. Tentukan setiap objek ke *medoid* terdekat dan peroleh hasil pengklasteran.
8. Lakukan perhitungan jumlah jarak dari semua objek ke medoidnya.

2.3 Lebar *Silhouette*

Dalam analisis kluster, untuk menentukan seberapa baik suatu pembagian data ke dalam kelompok lebar *Silhouette* atau *Silhouette coefficient* sering digunakan sebagai ukuran validitas internal kluster. Konsep ini pertama kali diuraikan oleh Rousseeuw (1987) yang menjelaskan bahwa nilai *Silhouette* dapat menggambarkan seberapa dekat suatu objek dengan anggota dalam klasternya sendiri. Definisi lebar *Silhouette* dapat diuraikan sebagai berikut:

Jika sebuah objek yang menjadi fokus dan termasuk dalam kluster A. Kelompok lain yang tidak termasuk objek ini sebagai kluster C. Nilai $a(i)$ didefinisikan sebagai rata-rata perbedaan antara objek i dengan semua objek lain yang ada dalam kelompok A. Sedangkan nilai $c(i, C)$ adalah rata-rata perbedaan antara objek i dan semua objek yang terdapat dalam kelompok C.

$$b(i) = \min_{C \neq A} c(i, C)$$

Lebar *Silhouette* $s(i)$ didefinisikan sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.6)$$

Lebar *Silhouette* $s(i)$ bervariasi antara -1 hingga 1. Sebuah nilai yang mendekati 1 menunjukkan bahwa suatu objek memiliki kesamaan yang lebih tinggi dengan objek lain dalam kelompoknya dibandingkan dengan objek-objek dari kelompok yang berbeda, dan ini menandakan kualitas klasifikasi yang sangat baik. Sebaliknya, jika $s(i)$ hampir 0, ini menunjukkan bahwa klasifikasi objek tersebut bisa dipertanyakan karena posisinya yang ambigu antara dua kelompok. Nilai $s(i)$ yang mendekati -1

menandakan adanya kesalahan dalam pengelompokan. Oleh karena itu, rata-rata lebar *Silhouette* di setiap kelompok memberikan gambaran mengenai sejauh mana kualitas kelompok tersebut.

2.4 Validasi Klaster

Untuk mengetahui seberapa baik hasil yang diperoleh dari proses *clustering* diperlukan klaster validasi. Jika pada klasifikasi dapat diketahui dengan jelas mengenai pengukuran keberhasilan suatu proses klasifikasi dengan melihat nilai presisi, *recall*, dan akurasi yang bersifat eksak atau pasti. Pengukuran hasil proses *clustering* memiliki banyak metode seperti pengukuran berdasarkan nilai korelasinya, nilai *cohesion* dan *separation*, koefisien *silhouette*, *Dunn Index*, dan *Davis-Bouldin Index* (Saikhu & Gita, 2013).

Dunn Index adalah salah satu pengukur validitas klaster yang diajukan oleh Joseph C. Dunn. Pada jurnal penelitian Luthfi & Wijayanto (2021) yang bersumber dari penelitian Joseph C. Dunn pada tahun 1973 menjelaskan bahwa *index validitas Dunn* menghitung nilai minimum dari perbandingan antara nilai fungsi *dissimilaritas* antara dua klaster sebagai *separation* dan nilai maksimum dari diameter klaster sebagai *compactness*. Jumlah klaster terbaik ditunjukkan dengan semakin besarnya nilai *Dunn Index*. Rumus *Dunn Index* dapat ditulis sebagai berikut:

$$D = \frac{\min_{\substack{1 \leq i \leq k \\ i+1 \leq j \leq l}} (d(C_i, C_j))}{\max_{1 \leq l \leq q} \text{diam}(C_k)} \quad (2.7)$$

Keterangan:

- D : *Dunn Index*
- q : jumlah klaster
- $d(c_i, c_j)$: jarak antar klaster (*intercluster distance*)
- $\text{diam}(c_k)$: diameter klaster ke-k (*intracluster distance*)

2.5 Analisis Komponen Utama

Analisis Komponen Utama (*Principal Component Analysis*) merupakan salah satu metode statistik multivariat yang umum digunakan dalam penelitian. PCA pertama kali diperkenalkan oleh Karl Pearson pada tahun 1901 dan dikembangkan lebih lanjut oleh Harold Hotelling pada tahun 1933. Metode ini digunakan untuk mereduksi dimensi data, yakni mengubah sekumpulan variabel yang mungkin saling berkorelasi menjadi sekumpulan variabel baru yang tidak berkorelasi tanpa kehilangan banyak informasi dari data awal.

Menurut Jolliffe dan Cadima (2016), PCA berfungsi untuk mengekstraksi informasi utama dari data dengan cara mencari kombinasi linear dari variabel-variabel asli sehingga komponen pertama menjelaskan variasi terbesar dari keseluruhan data. Adapun beberapa tahapan dalam melakukan analisis komponen utama diantaranya yaitu: menghitung matriks korelasi, menghitung nilai eigen, menghitung vektor eigen, dan menghitung komponen utama. Persamaan analisis komponen utama adalah sebagai berikut:

$$\begin{aligned} KU_1 &= \alpha_{11}x_1 + \alpha_{21}x_2 + \alpha_{31}x_3 + \cdots + \alpha_{p1}x_p \\ KU_2 &= \alpha_{12}x_1 + \alpha_{22}x_2 + \alpha_{32}x_3 + \cdots + \alpha_{p2}x_p \\ &\vdots \\ KU_j &= \alpha_{1j}x_1 + \alpha_{2j}x_2 + \alpha_{3j}x_3 + \cdots + \alpha_{pj}x_p \end{aligned}$$

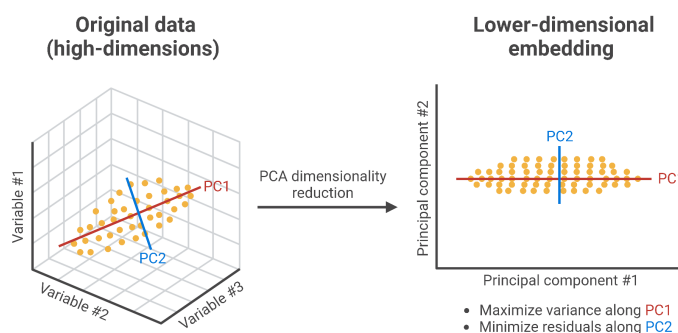
Keterangan:

KU_j : komponen utama ke-j

α_{pj} : koefisien untuk variabel ke-p dalam komponen utama ke-j

x_p : nilai variabel ke-p dari data asli

Principal Component Analysis (PCA) Transformation



Gambar 2. Analisis komponen utama

Gambar di atas merupakan contoh grafik analisis komponen utama. PC1 merupakan arah penyebaran data paling besar sedangkan PC2 arah kedua terbesar, yang dimana makna dari PC (*Principal Component*) adalah mencari arah terbaik untuk merangkum informasi data. PC1 berperan sebagai sumbu x dan PC2 berperan sebagai sumbu y.

2.6 *Diabetes Mellitus*

Diabetes mellitus merupakan tantangan kesehatan masyarakat yang signifikan di seluruh dunia. *Diabetes mellitus* dan komplikasinya merupakan penyebab utama morbiditas dan mortalitas di banyak negara berkembang. Selain predisposisi genetik, kebiasaan gaya hidup yang tidak sehat, pola makan yang tidak sehat, stres, dan kurangnya aktivitas fisik yang tepat dapat menyebabkan peningkatan dan kadar glukosa darah yang tidak terkontrol, resistensi insulin, dan obesitas, yang pada akhirnya menyebabkan diabetes melitus. Ada dua jenis utama diabetes. Diabetes tipe 1 disebabkan oleh predisposisi genetik dan kerusakan autoimun sel beta pankreas. Diabetes tipe 2 disebabkan oleh peningkatan stres, resistensi insulin, obesitas, dan gaya hidup yang tidak banyak bergerak. Meskipun hiperglikemia dapat dikontrol dengan obat-obatan farmasi, penyesuaian gaya hidup, peningkatan aktivitas fisik, dan pengurangan stres, penyebab molekuler yang mendasari diabetes dan komplikasinya masih perlu dieksplorasi. Kemajuan terbaru dalam terapi molekuler menargetkan

mekanisme molekuler yang terlibat dalam metabolisme glukosa, pensinyalan insulin, dan fungsi sel- β pankreas (Ramana , 2025).

Hiperglikemia kronis adalah gangguan metabolisme yang disebabkan oleh kekurangan sekresi insulin, gangguan kerja insulin, atau keduanya. Insulin berperan penting sebagai hormon anabolik, yang memengaruhi metabolisme karbohidrat, lipid, dan protein. Kelainan metabolisme yang terkait dengan diabetes terutama memengaruhi jaringan seperti jaringan adiposa, otot rangka, dan hati akibat resistensi insulin. Tingkat keparahan gejala dapat bervariasi tergantung pada durasi dan jenis diabetes. Individu dengan kadar gula darah tinggi, terutama mereka yang kekurangan insulin total, seperti anak-anak, dapat mengalami gejala seperti peningkatan nafsu makan, polidipsia, disuria, penurunan berat badan, peningkatan nafsu makan, dan masalah penglihatan. Beberapa penderita diabetes mungkin tidak mengalami gejala apa pun, terutama pasien diabetes tipe 2 pada tahap awal. Tanpa pengobatan yang tepat, diabetes yang tidak terkontrol dapat menyebabkan berbagai komplikasi seperti koma, kebingungan, dan dalam kasus yang jarang terjadi, kematian akibat ketoasidosis atau sindrom hiperosmolar nonketotik yang tidak diobati (Samar, *et al.*, 2023).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun ajaran 2025/2026 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder mengenai jenis *diabetes mellitus* dengan variabel jenis diabetes, golongan umur, jenis kelamin, tekanan darah serta kadar gula darah yang diambil dari Rumah Sakit Mardi Waluyo Metro. Data jenis diabetes pada tahun 2023 terdapat 17 jenis diabetes dan data jenis diabetes pada tahun 2024 terdapat 20 jenis diabetes. Adapun variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Jenis Diabetes (X_1)

Variabel ini merupakan variabel kategorik yang menunjukkan klasifikasi penyakit diabetes mellitus yang diderita pasien. Jenis diabetes dalam penelitian ini terdiri dari beberapa kategori, antara lain:

- a. IDDM dengan koma
- b. IDDM tanpa komplikasi
- c. IDDM dengan komplikasi kidney
- d. IDDM dengan komplikasi neurological
- e. IDDM dengan komplikasi unspecified
- f. NIDDM dengan koma

- g. NIDDM dengan ketoasidosis
- h. NIDDM dengan komplikasi renal
- i. NIDDM dengan komplikasi oftalmik
- j. NIDDM dengan komplikasi neurologis
- k. NIDDM dengan komplikasi sirkulasi perifer
- l. NIDDM dengan komplikasi lainnya
- m. NIDDM dengan komplikasi multipel
- n. NIDDM dengan komplikasi unspecified
- o. NIDDM tanpa komplikasi
- p. DM dengan koma
- q. DM dengan komplikasi renal
- r. DM dengan komplikasi sirkulasi perifer
- s. DM dengan komplikasi multipel
- t. DM dengan komplikasi neurological
- u. DM dengan komplikasi unspecified
- v. DM tidak spesifik multipel
- w. DM tidak spesifik

2. Golongan Umur (X_2)

Variabel ini merupakan variabel numerik yang dikelompokkan ke dalam beberapa kategori umur pasien, yaitu:

- a. 1–4 tahun
- b. 5–14 tahun
- c. 15–24 tahun
- d. 25–44 tahun
- e. 45–64 tahun
- f. ≥ 65 tahun

3. Jenis Kelamin (X_3)

Variabel ini merupakan variabel kategorik yang menunjukkan jenis kelamin pasien, terdiri dari:

- a. Laki-laki

b. Perempuan

4. Kadar Gula Darah Sewaktu (GDS) (X_4)

Variabel ini merupakan variabel numerik yang menunjukkan kadar gula darah pasien (mg/dL) yang diukur secara sewaktu.

5. Tekanan Darah Sistolik (X_5)

Variabel ini merupakan variabel numerik yang menunjukkan tekanan darah sistolik pasien (mmHg).

6. Tekanan Darah Diastolik (X_6)

Variabel ini merupakan variabel numerik yang menunjukkan tekanan darah diastolik pasien (mmHg).

3.3 Metode Penelitian

Adapun langkah-langkah penelitian ini adalah sebagai berikut:

1. Melakukan analisis deskriptif
2. Melakukan pra-pemrosesan data
 - a. Melihat pencilan data atau *outlier*
 - b. Melakukan standarisasi data
3. Melakukan analisis kluster untuk data penelitian dengan metode *K-Medoids* dengan langkah-langkah sebagai berikut:
 - a. Menentukan nilai k yang akan digunakan.
 - b. Pilih secara acak *medoid* awal sebanyak k .
 - c. Menghitung jarak Manhattan setiap pasangan objek.
 - d. Menentukan anggota kluster pada *medoids* awal.
 - e. Hitung jumlah jarak dari semua objek terhadap *medoid*nya.
 - f. Mengelompokkan setiap objek ke *medoid* terdekat.
 - g. Hitung jumlah jarak baru dari semua objek terhadap *medoid*.
4. Mengevaluasi hasil kluster dengan menggunakan koefisien *silhouette* untuk mengukur seberapa baik objek ditempatkan dalam kluster yang sesuai.
5. Menjelaskan hasil analisis kluster

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian, diperoleh kesimpulan sebagai berikut :

1. Penerapan metode *K-Medoids* terhadap indikator penyakit *diabetes mellitus* di Rumah Sakit Mardi Waluyo Metro menghasilkan jumlah kluster optimal sebanyak 2 kluster untuk data tahun 2023 dan 2024.
2. Kluster tahun 2023 terdiri dari 12 jenis *diabetes mellitus* dengan komplikasi berat dan 5 jenis *diabetes mellitus* dengan komplikasi ringan .
3. Kluster tahun 2024 terdiri dari 10 jenis *diabetes mellitus* dengan komplikasi berat dan 10 jenis *diabetes mellitus* dengan komplikasi ringan .

5.2 Saran

Berdasarkan hasil pengelompokan menggunakan metode K-Medoids pada data penyakit *diabetes mellitus*, beberapa hal yang dapat disarankan untuk pengembangan penelitian selanjutnya, yaitu memperluas variabel yang digunakan dalam analisis kluster. Selain variabel demografis dan jenis komplikasi, variabel klinis seperti HbA1c, indeks massa tubuh (IMT), serta lama menderita diabetes dapat dipertimbangkan. Penambahan variabel tersebut diharapkan mampu memberikan gambaran kluster yang lebih representatif terhadap kondisi klinis penderita *diabetes mellitus*.

DAFTAR PUSTAKA

- Antar, S. A., Ashour, N. A., Sharaky, M., Khattabm M., Ashour, N. A., Zaid, R. T., Roh. E. J., Elkamhawy, A., Al-Karmalawy, A. A. 2023. Diabetes mellitus: Classification, mediators, and complications; A gate to identify potential targets for the development of new effective treatments. *Biomedicine & Pharmacotherapy*. **168**, 115734
- Anton, H., & Rorres, C. 1994. *Elementary linear algebra: applications version*. Jhon Wiley & Sons.
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. 2013. An extensive comparative study of cluster validity indices. *Pattern Recognition*. **46**(1): 243-256
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. 2011. *Cluster Analysis*. **5** . Wiley.
- Forrester, P. J., & Zhang, J. 2020. Parametrising correlation matrices. *Journal of Multivariate Analysis*. **178**(1), 104619
- Graffelman, J., & de Leeuw, J. 2023. Improved approximation and visualization of the correlation matrix. *The American Statistician*. **77**(4): 432-442
- Greenbaum, A., Li, R. C., & Overton, M. L. 2019. First order pertubation theory for eigenvalues and eigenvectors. *Siam Publication Library*. **62**(2): 463-482
- Gultom, S., Sriadhi, S., Martiano, M., & Simarmata, J. 2018. Comparasion analysis of K-means and K-medoid with Ecludience distance algorithm, Chanberra

- distancem and Chebyshev distance for big data clustering. *IOP Conference Series: Materials Science and Engineering*. **420**(1): 120-129
- Hossain, M. J., Al-Mamun, M., & Islam, M.R. 2024. Diabetes Mellitus, the fastest growing global public health concern: Early detection should be focused. *Health Science Reports*. **7**(3)
- Jolliffe, I. T., & Cadima, J. (2016). Principal Component Analysis: A Review and Recent Developments. *The Royal Society A: Mathematical, Physical and Engineering Sciences*. **374**(2065): 1–16.
- Kaufman, L. & Rousseuw, P. J. 1989. *Finding Groups in Data an Introduction to Cluster Analysis*. Wiley, J. & Sons. New Jersey.
- Kaur, N. K, Kaur, U., & Singh, D. 2014. K-medoid Clustering Algorithm- A Review. *International Journal of Computer Application and Technology*. **01**(01): 42-45
- Kumar, M. N., & Rao, V. S. 2015. A New Methodology for Estimasting Internal Credit Risk and Bankruptcy Prediction under Basel II Regime. *Computational Economics*. **46**(1): 83-102
- Loftus, T. J., Shickel, B., Balch, A. J., Tighe, P. J., Aboot, K. L., Fazzone, B., Anderson, E. M., Rozowsky, J., Baslanti, T. O., Ren, Y., Berceci, S. A., Hogan, W. R., Efron, P. A., Moorman, J. R., Rashidi, P., Upchruch Jr, G. R. & Bihorac, A. 2022. Phenotype clustering in health care: A narrative reivew for clinicians. *Frontiers in Artificial Intelligence*. (5), 842306
- Nugroho, S. 2008. *Statistika Multivariat Terapan*. Ed ke-1. UNIB Press. Bengkulu
- Prahara, A., Ismi, D. P., & Azhari, A. 2020. Parallelization of Partitioning Around Medoids (PAM) in K-Medoids Clustering on GPU. *Knowledge Engineering and Data Science*. **3**(1): 40-49

- Ramana, K. V. 2025. Special Issue “Molecular Therapeutics for Diabetes and Related Complications”. *International Journal of Molecular Sciences*. **26**(12), 5585
- Saikhu, A., & Gita, Y.B. 2013. Implementasi Deteksi Outlier Pada Algoritma Hierarchical Clustering. Prosiding Seminar Nasional Teknologi Informasi dan Multimedia. Yogyakarta.
- Shirkhorshidi, A. S., Aghabozorgi, S., & Teh, Y. W. 2015. A comparison study on similarity and dissimilarity measures in clustering continuous data. *PLOS ONE*. **10**(12), e0144059.
- Sindi, S., Ningse, W. R. O., Sihombing, I. A., Zer, F. I. R. H., & Hartama, D. 2020. Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 di Indonesia. *Jurnal Teknologi Informasi*. **4**(01): 166-173
- Suwanda, R., Syahputra, Z., & Zamzami, M. 2019. Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K. *Journal of Physics*. **1566**(1). IOP Publishing
- Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. 2021. Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah. Prosiding Seminar Nasional Matematika. Semarang
- Yedla, M., Pathakota, S.R., & Srinivasa, T. M. 2010. Enhancing K-means clustering algorithm with improved initial center. *International Journal of computer science and information technologies*. **1**(2): 121-125