

IDENTIFIKASI POLA KARAKTERISTIK KLINIS PASIEN DIABETES MELLITUS MENGGUNAKAN K-MEANS CLUSTERING DI RS JUANDA KUNINGAN

Indah Hanisa, Bambang Karmanto, Totok Subianto, Elfi

Rekam Medis dan Informasi Kesehatan, Poltekkes Kemenkes Tasikmalaya

Jl. Babakan Siliwangi No.35, Kahuripan, Kec. Tawang, Kab. Tasikmalaya, Jawa Barat 46115, Indonesia

hanisaindah655@gmail.com

ABSTRAK

Diabetes mellitus merupakan penyakit kronik dengan prevalensi yang terus meningkat dan menjadi salah satu beban utama dalam pelayanan kesehatan. Namun, pemanfaatan data rekam medis di rumah sakit masih terbatas pada fungsi administratif dan belum dimanfaatkan secara optimal untuk analisis pola pasien. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi pola karakteristik klinis pasien diabetes mellitus menggunakan metode K-Means Clustering di RS Juanda Kuningan. Metode yang digunakan adalah pendekatan data mining terhadap 200 data rekam medis pasien tahun 2025 dengan teknik purposive sampling. Variabel yang dianalisis meliputi usia, tekanan darah, kadar gula darah, dan indeks massa tubuh (IMT). Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah tiga kelompok yang merepresentasikan kondisi stabil, risiko sedang, dan risiko tinggi, dengan perbedaan utama pada kadar gula darah dan indikator klinis lainnya. Temuan ini menunjukkan bahwa metode K-Means mampu mengelompokkan pasien secara objektif serta memberikan gambaran pola kondisi pasien yang dapat mendukung pengambilan keputusan berbasis data dalam pelayanan kesehatan.

Kata kunci: Diabetes Mellitus, Rekam Medis, Data Mining, K-Means Clustering, Rumah Sakit

1. PENDAHULUAN

Diabetes mellitus merupakan penyakit kronis yang ditandai dengan gangguan metabolisme glukosa akibat ketidakcukupan produksi atau fungsi insulin. Penyakit ini terus menunjukkan peningkatan prevalensi dan menjadi salah satu permasalahan utama dalam sistem pelayanan kesehatan global. Dampaknya tidak hanya berkaitan dengan kondisi klinis pasien, tetapi juga berpengaruh terhadap beban ekonomi serta kualitas hidup, sehingga diperlukan pengelolaan yang lebih efektif dan berbasis data [1], [2], [3].

Di Indonesia, jumlah penderita diabetes juga menunjukkan tren peningkatan yang signifikan, baik pada tingkat nasional maupun regional. Di Kabupaten Kuningan, diabetes mellitus menjadi salah satu masalah kesehatan dengan jumlah kasus yang cukup tinggi, termasuk di RS Juanda Kuningan yang memiliki jumlah kunjungan pasien diabetes yang besar. Namun, pemanfaatan data rekam medis di rumah sakit masih terbatas pada fungsi administratif dan belum dimanfaatkan secara optimal untuk analisis pola pasien [4], [5], [6].

Kondisi tersebut menunjukkan adanya kesenjangan antara ketersediaan data dengan pemanfaatannya dalam mendukung analisis klinis dan pengambilan keputusan.

Rekam medis elektronik memiliki potensi besar untuk dimanfaatkan sebagai sumber data strategis dalam analisis klinis, termasuk untuk mengidentifikasi karakteristik pasien dan mendukung pengambilan keputusan berbasis data. Pemanfaatan ulang data rekam medis telah terbukti dapat mendukung penelitian, deteksi dini, serta perencanaan intervensi penyakit kronik secara lebih efektif [7], [8], [9].

Salah satu pendekatan yang dapat digunakan dalam pengolahan data rekam medis adalah data mining dengan metode K-Means Clustering. Metode ini mampu mengelompokkan data berdasarkan kemiripan karakteristik sehingga dapat mengidentifikasi pola tersembunyi dalam data pasien. Beberapa penelitian sebelumnya menunjukkan bahwa K-Means efektif dalam mengelompokkan pasien berdasarkan parameter klinis, namun sebagian besar penelitian masih menggunakan dataset publik atau belum memanfaatkan data rekam medis rumah sakit secara langsung [10], [11], [12], [13].

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengidentifikasi pola karakteristik klinis pasien diabetes mellitus menggunakan metode K-Means Clustering pada data rekam medis di RS Juanda Kuningan. Melalui pendekatan ini, diharapkan dapat dihasilkan pengelompokan pasien secara objektif berdasarkan kemiripan karakteristik klinis, sehingga dapat memberikan gambaran kondisi pasien serta mendukung pengambilan keputusan berbasis data dalam pelayanan kesehatan.

2. TINJAUAN PUSTAKA

2.1. Diabetes Mellitus

Diabetes mellitus merupakan penyakit metabolik kronis yang ditandai dengan peningkatan kadar glukosa darah akibat gangguan produksi maupun kerja insulin. Kondisi ini menyebabkan gangguan metabolisme karbohidrat, lemak, dan protein yang dapat menimbulkan komplikasi jangka panjang apabila tidak dikelola dengan baik [14], [15].

Diabetes mellitus tipe 2 merupakan jenis yang paling umum terjadi dan umumnya berkaitan dengan penurunan fungsi sel beta pankreas serta resistensi

insulin. Selain itu, berbagai faktor risiko seperti usia, pola hidup, dan kondisi metabolik turut berperan dalam peningkatan kejadian diabetes mellitus [16].

2.2. Rekam Medis

Rekam medis merupakan dokumen yang memuat Rekam medis merupakan kumpulan data yang mencatat identitas pasien serta seluruh rangkaian pelayanan kesehatan yang diterima, mulai dari pemeriksaan, tindakan medis, hingga terapi yang diberikan. Informasi tersebut berperan sebagai sumber utama dalam mendukung proses pelayanan kesehatan dan harus dikelola secara sistematis, akurat, serta berkesinambungan [17].

Seiring perkembangan teknologi, penerapan rekam medis elektronik memungkinkan proses pencatatan dan pengelolaan data pasien dilakukan secara digital. Hal ini memberikan kemudahan dalam akses, integrasi, serta pemanfaatan data secara lebih efisien. Selain mendukung pelayanan kesehatan, data rekam medis juga memiliki nilai strategis sebagai sumber data penelitian, khususnya dalam analisis berbasis data mining [18].

2.3. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan suatu proses terstruktur yang bertujuan untuk menggali informasi atau pola yang bernilai dari kumpulan data dalam jumlah besar. Proses ini dilakukan melalui serangkaian tahapan yang saling berkaitan, mulai dari pemilihan data yang relevan, pembersihan dan penyiapan data, hingga transformasi ke dalam bentuk yang sesuai untuk dianalisis. Pada tahap inti, teknik data mining digunakan untuk mengidentifikasi pola atau hubungan yang bermakna, kemudian hasil yang diperoleh dievaluasi guna memastikan validitas serta kegunaannya dalam mendukung pengambilan keputusan [19], [20].

2.4. Data Mining

Data mining merupakan suatu proses analisis yang dilakukan untuk menggali informasi bernilai dari kumpulan data berukuran besar melalui identifikasi pola, hubungan, maupun kecenderungan tertentu. Proses ini memanfaatkan berbagai teknik analitis untuk mengubah data mentah menjadi informasi yang lebih bermakna, sehingga dapat digunakan sebagai dasar dalam mendukung pengambilan keputusan secara lebih efektif [21].

2.5. Clustering

Clustering merupakan salah satu teknik dalam data mining yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik antar data. Berbeda dengan metode klasifikasi, teknik ini tidak memerlukan label awal, sehingga pola yang terbentuk berasal dari struktur data itu sendiri. Melalui proses ini, data yang memiliki karakteristik serupa akan tergabung dalam satu kelompok, sehingga memudahkan dalam

memahami pola serta distribusi data secara lebih terstruktur [22], [23].

2.6. K-Means Clustering

K-Means Clustering merupakan metode pengelompokan data yang membagi data ke dalam beberapa cluster berdasarkan kemiripan karakteristik dengan menggunakan centroid sebagai pusat cluster [10]. Proses K-Means dilakukan secara iteratif dengan menghitung jarak data terhadap centroid dan memperbarui pusat cluster hingga terbentuk kelompok data yang stabil [24].

Proses clustering K-Means dilakukan melalui langkah-langkah berikut:

- Menentukan jumlah cluster (K).
- Menentukan titik pusat awal (centroid) secara acak.
- Menghitung jarak antara setiap data dengan masing-masing centroid menggunakan Euclidean Distance:

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

$D(x, y)$ = jarak antara data dan centroid

x_i = nilai atribut ke- i pada data

y_i = nilai atribut ke- i pada centroid

- Menempatkan setiap data ke dalam cluster dengan centroid terdekat.
- Menghitung ulang posisi centroid dengan rumus:

$$C_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i \quad (2)$$

Keterangan:

C_j = centroid baru pada cluster ke- j

n_j = jumlah data pada cluster ke- j

x_i = data ke- i dalam cluster

- Proses diulang hingga posisi centroid tidak berubah (*konvergen*).

2.7. Aplikasi RapidMiner

RapidMiner merupakan perangkat lunak yang digunakan untuk melakukan analisis data dan penerapan metode data mining, termasuk K-Means Clustering, melalui antarmuka yang memudahkan proses pengolahan data [25].

2.8. Penelitian Terdahulu

Penelitian mengenai pengelompokan pasien diabetes menggunakan metode K-Means Clustering telah banyak dilakukan dengan memanfaatkan berbagai parameter klinis untuk mengidentifikasi pola dan tingkat risiko penyakit.

Penelitian oleh Maharani et al. (2025) mengembangkan sistem berbasis web untuk mengelompokkan tingkat keparahan diabetes mellitus menggunakan algoritma K-Means dengan tiga kategori, yaitu ringan, sedang, dan berat. Penelitian tersebut menggunakan data rekam medis dengan variabel klinis seperti HbA1c, kadar gula darah puasa, tekanan darah, dan indeks massa tubuh. Hasil penelitian menunjukkan bahwa metode K-Means

mampu menghasilkan pengelompokan yang akurat serta mendukung pengambilan keputusan klinis melalui sistem yang terintegrasi secara real-time [26].

Selanjutnya, penelitian oleh Setyadji et al. (2023) melakukan analisis klaster pada data pasien diabetes di Puskesmas Mojokerto menggunakan algoritma K-Means. Hasil penelitian menunjukkan terbentuknya dua kelompok utama pasien berdasarkan karakteristik seperti usia, jenis kelamin, dan diagnosis penyakit. Clustering ini membantu dalam memahami pola dan karakteristik pasien sehingga dapat digunakan sebagai dasar dalam pengembangan strategi pelayanan kesehatan [27].

Penelitian lain oleh Prasetyo dan Oktafiani (2026) menerapkan algoritma K-Means dengan jumlah cluster sebanyak tiga untuk mengelompokkan pasien diabetes berdasarkan faktor risiko. Hasilnya menunjukkan adanya tiga kelompok utama, yaitu pasien dengan kondisi sehat, pasien dengan risiko tinggi (obesitas dan kadar glukosa tinggi), serta pasien dengan risiko sedang. Penelitian ini juga menemukan adanya hubungan antara variabel usia, BMI, dan kadar glukosa terhadap tingkat risiko diabetes [28].

Selain itu, penelitian oleh Matsway dan Firmansyah (2025) menggunakan metode K-Means untuk mengelompokkan pasien berdasarkan risiko diabetes dengan memanfaatkan data klinis seperti glukosa, tekanan darah, dan BMI. Hasil clustering menunjukkan tiga kelompok utama, yaitu pasien tanpa risiko, risiko rendah, dan risiko tinggi. Penelitian ini menegaskan bahwa K-Means efektif dalam membantu proses deteksi dini dan pengelolaan diabetes [12].

Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa metode K-Means Clustering memiliki kemampuan yang baik dalam mengelompokkan data pasien diabetes berdasarkan karakteristik klinis maupun faktor risiko. Namun, sebagian besar penelitian masih berfokus pada analisis clustering tanpa mengintegrasikan hasilnya secara optimal ke dalam sistem pendukung keputusan berbasis data rekam medis secara spesifik. Oleh karena itu, penelitian ini dilakukan untuk mengoptimalkan pemanfaatan data rekam medis dalam proses clustering guna memperoleh informasi yang lebih akurat dan aplikatif.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif kuantitatif dengan desain eksploratif melalui teknik data mining. Data yang digunakan merupakan data sekunder berupa rekam medis pasien Diabetes Mellitus tahun 2025 di RS Juanda Kuningan. Teknik pengambilan sampel dilakukan menggunakan purposive sampling, yaitu pemilihan data berdasarkan kriteria tertentu yang sesuai dengan kebutuhan analisis.

Populasi dalam penelitian ini adalah seluruh data rekam medis pasien Diabetes Mellitus yang tercatat pada Sistem Informasi Manajemen Rumah Sakit (SIMRS) RS Juanda Kuningan tahun 2025. Sampel

penelitian merupakan data pasien yang memenuhi kriteria, yaitu memiliki data lengkap pada variabel yang digunakan dalam proses clustering. Setelah dilakukan seleksi dan pembersihan data, diperoleh sebanyak 200 data pasien yang digunakan dalam analisis.

Variabel yang digunakan dalam penelitian meliputi usia, tekanan darah sistolik (TDS), tekanan darah diastolik (TDD), kadar gula darah, dan indeks massa tubuh (IMT). Variabel jenis kelamin dan komplikasi tidak digunakan dalam proses clustering karena bersifat kategorikal, sedangkan algoritma K-Means bekerja berdasarkan perhitungan jarak antar data numerik. Oleh karena itu, analisis difokuskan pada variabel numerik agar hasil pengelompokan lebih sesuai dengan karakteristik metode yang digunakan.

Pengolahan data dilakukan melalui tahapan Knowledge Discovery in Database (KDD), yang meliputi seleksi data, pra-pemrosesan data, transformasi data, data mining, dan evaluasi. Pada tahap seleksi data, dilakukan pemilihan data pasien Diabetes Mellitus tahun 2025 yang memiliki variabel sesuai kebutuhan penelitian. Pada tahap pra-pemrosesan, dilakukan penghapusan data duplikat berdasarkan nomor rekam medis agar satu pasien hanya direpresentasikan oleh satu data. Selain itu, dilakukan pengecekan data kosong pada setiap variabel penelitian. Data yang tidak lengkap dikeluarkan dari dataset agar tidak memengaruhi hasil clustering.

Tahap transformasi dilakukan dengan menyesuaikan format data agar dapat diproses dalam aplikasi RapidMiner Studio. Data numerik kemudian dinormalisasi menggunakan metode Z-score. Normalisasi dilakukan karena setiap variabel memiliki satuan dan rentang nilai yang berbeda, seperti usia dalam tahun, tekanan darah dalam mmHg, kadar gula darah dalam mg/dL, dan IMT dalam kg/m². Dengan normalisasi, seluruh variabel berada pada skala yang seimbang sehingga tidak ada variabel yang mendominasi perhitungan jarak pada algoritma K-Means.

Proses data mining dilakukan menggunakan algoritma K-Means Clustering pada aplikasi RapidMiner Studio. Tahapan implementasi dimulai dari mengimpor dataset menggunakan operator *Read Excel*, kemudian melakukan normalisasi data menggunakan operator *Normalize*. Setelah itu, proses clustering dilakukan menggunakan operator *K-Means* dengan pengujian beberapa jumlah cluster, yaitu K=2, K=3, dan K=4. Algoritma K-Means bekerja dengan menentukan centroid awal, menghitung jarak setiap data terhadap centroid menggunakan Euclidean Distance, menempatkan data pada cluster dengan jarak terdekat, kemudian memperbarui centroid hingga posisi cluster stabil.

Tahap evaluasi dilakukan untuk menentukan jumlah cluster yang paling representatif. Evaluasi kualitas clustering menggunakan nilai *Average Within Centroid Distance*, yaitu nilai rata-rata jarak data

terhadap centroid dalam satu cluster. Semakin kecil nilai jarak yang diperoleh, semakin rapat data dalam cluster tersebut. Selain mempertimbangkan nilai evaluasi matematis, pemilihan jumlah cluster juga memperhatikan kemudahan interpretasi hasil agar kelompok pasien yang terbentuk dapat menggambarkan karakteristik klinis secara jelas.

Hasil akhir dari proses clustering berupa pembagian pasien ke dalam beberapa cluster berdasarkan kemiripan karakteristik klinis. Setiap cluster kemudian dianalisis berdasarkan nilai centroid untuk mengetahui kecenderungan masing-masing kelompok pasien. Interpretasi hasil dilakukan dengan membandingkan nilai centroid pada setiap variabel sehingga diperoleh gambaran kelompok pasien dengan kondisi relatif stabil, risiko sedang, dan risiko tinggi.

4. HASIL DAN PEMBAHASAN

4.1. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa rekam medis pasien Diabetes Mellitus tahun 2025 yang diperoleh dari Sistem Informasi Manajemen Rumah Sakit (SIMRS) RS Juanda Kuningan. Setelah dilakukan proses pembersihan data, diperoleh sebanyak 200 data pasien yang memenuhi kriteria untuk dianalisis.

Variabel yang digunakan dalam penelitian meliputi usia, tekanan darah sistolik (TDS), tekanan darah diastolik (TDD), kadar gula darah (GDP), serta indeks massa tubuh (IMT). Variabel tersebut dipilih karena merepresentasikan kondisi klinis utama pasien diabetes.

Tabel 1. Dataset

USIA	TDS	TDD	GDP	IMT
57	144	93	428	24.89
72	105	75	117	19.02
68	120	80	308	28.80
56	120	80	130	31.34
77	150	80	204	22.22
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

4.2. Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan data yang digunakan telah bersih dan siap untuk dianalisis. Tahapan yang dilakukan meliputi penghapusan data duplikat, penanganan data kosong, serta normalisasi data.

Penghapusan data duplikat dilakukan karena data yang digunakan merupakan data kunjungan pasien, sehingga satu pasien dapat tercatat lebih dari satu kali. Oleh karena itu, dilakukan penyaringan berdasarkan nomor rekam medis sehingga setiap pasien direpresentasikan oleh satu data.

Selanjutnya dilakukan pengecekan terhadap data yang tidak lengkap. Data yang memiliki nilai kosong pada variabel penelitian dihapus agar tidak mempengaruhi hasil clustering.

Normalisasi data dilakukan menggunakan metode Z-score untuk menyamakan skala antar

variabel. Proses ini bertujuan agar setiap variabel memiliki kontribusi yang seimbang dalam perhitungan jarak pada algoritma K-Means.

Rumus Z-score yang digunakan adalah sebagai berikut:

$$Z = \frac{X - \mu}{\sigma} \quad (3)$$

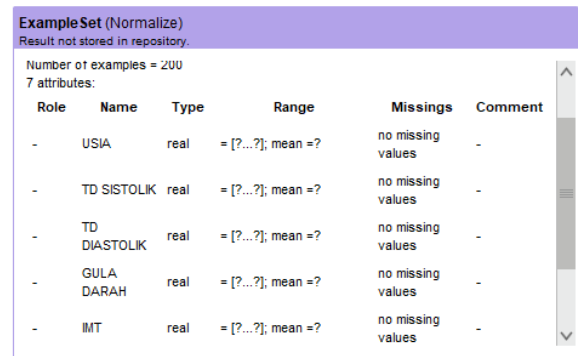
Keterangan:

Z = nilai hasil normalisasi

X = nilai asli variabel

μ = rata-rata variabel

σ = standar deviasi variabel

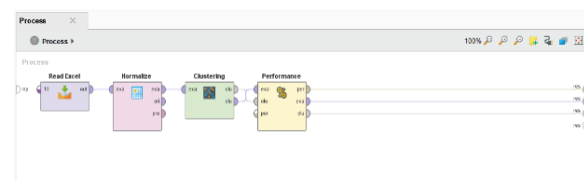


Role	Name	Type	Range	Missings	Comment
-	USIA	real	= [...?]; mean = ?	no missing values	-
-	TD SISTOLIK	real	= [...?]; mean = ?	no missing values	-
-	TD DIASTOLIK	real	= [...?]; mean = ?	no missing values	-
-	GULA DARAH	real	= [...?]; mean = ?	no missing values	-
-	IMT	real	= [...?]; mean = ?	no missing values	-

Gambar 1. Hasil normalisasi data menggunakan metode Z-score pada RapidMiner

4.3. Implementasi K-Means Clustering

Proses clustering dilakukan menggunakan aplikasi RapidMiner Studio. Tahapan implementasi disusun dalam bentuk alur proses sebagai berikut:



Gambar 2. Alur proses implementasi K-Means Clustering pada RapidMiner

a. Import Data (Read Excel)

Data yang telah dipersiapkan diimpor ke dalam RapidMiner menggunakan operator *Read Excel*. Data yang digunakan merupakan data yang telah melalui tahap pra-pemrosesan dan siap untuk dianalisis.

b. Normalisasi Data (Normalize)

Data kemudian dinormalisasi menggunakan metode Z-score sebagai bagian dari tahapan pra-pemrosesan untuk menyamakan skala antar variabel.

c. Proses Clustering (K-Means)

Proses clustering dilakukan menggunakan algoritma K-Means dengan variasi jumlah cluster yaitu $k = 2$, $k = 3$, dan $k = 4$. Pengujian ini bertujuan untuk memperoleh jumlah cluster yang paling optimal dalam merepresentasikan pola data pasien.

d. Evaluasi Hasil (Performance)

Evaluasi dilakukan untuk menilai kualitas hasil clustering dan membantu dalam menentukan jumlah cluster yang paling optimal.

4.4. Penentuan Jumlah Cluster

Penentuan jumlah cluster (k) dilakukan dengan menguji beberapa variasi nilai k , yaitu $k = 2$, $k = 3$, dan $k = 4$. Evaluasi kualitas clustering dilakukan menggunakan nilai *Average Within Centroid Distance*, di mana nilai yang lebih kecil menunjukkan jarak antar data dalam cluster yang lebih rapat.

Hasil pengujian menunjukkan bahwa nilai terbaik secara matematis diperoleh sebagai berikut.

Tabel 2. Hasil pengujian nilai k

K	Nilai
2	-4.124
3	-3.558
4	-3.050

Meskipun $k = 2$ memiliki nilai evaluasi terbaik, hasil pengelompokan yang dihasilkan masih bersifat umum dan belum mampu menggambarkan variasi karakteristik pasien secara rinci. Sementara itu, $k = 4$ menghasilkan pembagian data yang terlalu terfragmentasi sehingga sulit untuk diinterpretasikan.

Sebaliknya, $k = 3$ memberikan keseimbangan antara kualitas clustering dan kemudahan interpretasi, sehingga dipilih sebagai jumlah cluster yang paling representatif dalam penelitian ini.

Hasil evaluasi clustering untuk $K=3$ ditampilkan pada Gambar 3 berikut.

PerformanceVector	
Performance	PerformanceVector:
	Avg. within centroid distance: -3.558
	Avg. within centroid distance_cluster_0: -2.177
	Avg. within centroid distance_cluster_1: -3.160
	Avg. within centroid distance_cluster_2: -7.230
	Davies Bouldin: -1.566

Gambar 3. Hasil evaluasi nilai clustering berdasarkan jumlah $k = 3$

4.5. Hasil Clustering

Berdasarkan hasil penentuan jumlah cluster pada tahap sebelumnya, digunakan nilai $k = 3$ sebagai jumlah cluster dalam penelitian ini. Proses clustering menggunakan algoritma K-Means menghasilkan tiga kelompok pasien dengan karakteristik yang berbeda.

Cluster Model	
Description	Cluster 0: 47 items
	Cluster 1: 47 items
	Cluster 2: 80 items
	Cluster 3: 26 items
	Total number of items: 200

Gambar 4. Jumlah anggota pada masing-masing cluster

Jumlah anggota pada masing-masing cluster adalah 97 pasien pada cluster 0, 60 pasien pada cluster 1, dan 43 pasien pada cluster 2.

Distribusi tersebut menunjukkan bahwa sebagian besar data terkonsentrasi pada cluster 0, sehingga kelompok ini merepresentasikan karakteristik pasien yang paling umum. Sementara itu, cluster 1 dan cluster 2 memiliki jumlah anggota yang lebih sedikit, yang mengindikasikan adanya kelompok pasien dengan karakteristik yang lebih spesifik.

Perbedaan distribusi ini menunjukkan bahwa data pasien Diabetes Mellitus tidak bersifat homogen, melainkan terbagi ke dalam beberapa kelompok dengan pola karakteristik yang berbeda. Hasil ini menjadi dasar untuk analisis lebih lanjut terhadap karakteristik masing-masing cluster.

4.6. Karakteristik Tiap Cluster

Untuk mengetahui perbedaan karakteristik masing-masing cluster, digunakan nilai centroid yang merepresentasikan nilai rata-rata setiap variabel dalam satu cluster.

Tabel 3. Hasil pengujian nilai k

Atribut	Cluster0	Cluster1	Cluster2	Cluster3
Usia	-0.960	0.351	0.340	0.055
TDS	0.051	1.218	-0.717	-0.089
TDD	-0.028	0.106	-0.197	0.465
GD	-0.051	-0.359	-0.395	1.956
IMT	1.196	-0.357	-0.366	-0.392

Berdasarkan hasil clustering dengan $k = 3$, diperoleh tiga kelompok pasien dengan karakteristik yang berbeda, yaitu kelompok kondisi relatif stabil, risiko sedang, dan risiko tinggi.

a. Cluster 0

Cluster ini ditandai dengan nilai usia di atas rata-rata, namun diikuti dengan nilai gula darah, tekanan darah, dan IMT yang berada di bawah rata-rata. Hal ini menunjukkan bahwa meskipun usia menjadi faktor risiko, kondisi klinis pasien dalam kelompok ini relatif lebih terkendali.

b. Cluster 1

Cluster ini didominasi oleh pasien dengan usia lebih muda, namun memiliki nilai IMT dan tekanan darah yang lebih tinggi. Kondisi ini mengindikasikan adanya pengaruh faktor gaya hidup dan metabolik yang berpotensi meningkatkan risiko perkembangan penyakit.

c. Cluster 2

Cluster ini memiliki nilai gula darah paling tinggi dibandingkan cluster lainnya, yang menunjukkan kondisi hiperglikemia yang signifikan. Hal ini mengindikasikan bahwa kelompok ini berada pada tingkat risiko yang lebih tinggi.

Perbedaan karakteristik antar cluster menunjukkan bahwa kondisi pasien Diabetes Mellitus tidak bersifat homogen, melainkan dipengaruhi oleh kombinasi beberapa faktor klinis. Variabel gula darah menjadi faktor utama dalam membedakan tingkat

kondisi pasien, yang kemudian diperkuat oleh variabel lain seperti IMT dan tekanan darah.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, metode K-Means Clustering mampu mengelompokkan data pasien Diabetes Mellitus berdasarkan karakteristik klinis menjadi tiga cluster dengan tingkat risiko yang berbeda, yaitu kondisi relatif stabil, risiko sedang, dan risiko tinggi. Proses clustering menggunakan variabel numerik berupa usia, tekanan darah, gula darah, dan indeks massa tubuh menunjukkan bahwa nilai $K = 3$ merupakan jumlah cluster yang paling representatif karena memberikan keseimbangan antara kualitas pengelompokan dan kemudahan interpretasi. Hasil penelitian juga menunjukkan bahwa kondisi pasien tidak bersifat homogen, melainkan dipengaruhi oleh kombinasi beberapa variabel klinis, di mana gula darah menjadi faktor utama dalam membedakan tingkat risiko pasien. Oleh karena itu, metode K-Means dapat dimanfaatkan sebagai alat bantu dalam memahami pola kondisi pasien dan mendukung pengambilan keputusan di bidang kesehatan. Untuk penelitian selanjutnya, disarankan menambahkan variabel yang lebih beragam, menggunakan data dengan cakupan yang lebih luas, serta membandingkan metode clustering lain agar diperoleh hasil yang lebih optimal.

DAFTAR PUSTAKA

- [1] World Health Organization, "Diabetes," *World Health Organization*, Geneva, 2024.
- [2] K. L. Ong *et al.*, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021," *The Lancet*, vol. 402, no. 10397, pp. 203–234, Jul. 2023, doi: 10.1016/S0140-6736(23)01301-6.
- [3] International Diabetes Federation, "The Diabetes Atlas," *International Diabetes Federation*, Brussels, 2025.
- [4] International Diabetes Federation, "Diabetes in Indonesia," *International Diabetes Federation*, Brussels, 2024.
- [5] Survei Kesehatan Indonesia, "Penyakit tidak menular (hipertensi dan diabetes melitus)," *Survei Kesehatan Indonesia*, Jakarta, 2023.
- [6] Dinas Kesehatan Provinsi Jawa Barat, "Jumlah penderita diabetes melitus berdasarkan kabupaten/kota di Jawa Barat tahun 2019–2024," *Open Data Jabar*, Bandung, 2024.
- [7] F. Agiwahyunto, D. Ernawati, J. Prasetya, and M. T. Abiyasa, "Analisis Data Rekam Medis Elektronik di Rumah Sakit (Studi Perancangan Dan Evaluasi Sistem Informasi Analisis Kuantitatif)," *J-REMI J. Rekam Med. Dan Inf. Kesehat.*, vol. 5, no. 3, pp. 219–233, Jun. 2024.
- [8] T. Robbins, S. N. Lim Choi Keung, S. Sankar, H. Randevara, and T. N. Arvanitis, "Diabetes and the direct secondary use of electronic health records: Using routinely collected and stored data to drive research and understanding," *Digit. Health*, vol. 4, p. 2055207618804650, Jan. 2018, doi: 10.1177/2055207618804650.
- [9] J. Aswinasih, A. Susanto, and H. Hosizah, "Optimalisasi Analisis Data Rekam Medis Elektronik Menggunakan Business Intelligence di Rumah Sakit Islam Jakarta Cempaka Putih," *J. Manaj. Inf. Kesehat. Indones.*, vol. 8, no. 1, p. 78, Mar. 2020, doi: 10.33560/jmiki.v8i1.266.
- [10] L. Azzahra and Amru Yasir, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilmu Komput. Dan Sist. Inf.*, vol. 3, no. 1, pp. 1–10, Jan. 2024, doi: 10.70340/jirsi.v3i1.88.
- [11] F. Feronika, N. A. Ramdhan, and R. M. H. Bhakti, "Penggunaan Algoritma K-Means dalam Pengelompokan Pasien Diabetes Mellitus Berdasarkan Parameter Klinis di Puskesmas Brebes," *ELKOM J. Elektron. Dan Komput.*, vol. 18, no. 1, pp. 356–366, Jun. 2025.
- [12] D. A. Matsway and H. Firmansyah, "Clustering Pasien Dengan Risiko Diabetes Menggunakan Algoritma K-Means Clustering Patients with Diabetes Risk Using the K-Means Algorithm," vol. 02, no. 03, 2025.
- [13] Y. Perwira, W. Apriani, A. Zein, and S. L. Alfari, "CLUSTERIZATION OF DIABETES HEALTH INDICATORS WITH K-MEANS CLUSTER ALGORITHM," vol. 10, 2022.
- [14] R. P. Fadhillah, R. Rahma, A. Sepharni, R. Mufidah, B. N. Sari, and A. Pangestu, "Klasifikasi Penyakit Diabetes Mellitus Berdasarkan Faktor-Faktor Penyebab Diabetes menggunakan Algoritma C4.5," *JUPI J. Ilm. Penelit. Dan Pembelajaran Inform.*, vol. 7, no. 4, pp. 1265–1270, Nov. 2022, doi: 10.29100/jupi.v7i4.3248.
- [15] D. G. Saputri *et al.*, "Diabetes Melitus sebagai Gangguan Endokrin: Tinjauan Patofisiologi dan Pendekatan Diagnosis," *J. Ilm. Multidisiplin*, vol. 1, no. 6, pp. 4382–4387, 2025.
- [16] D. R. Rediningsih and I. P. Lestari, "FAKTOR RISIKO KEJADIAN DIABETES MELITUS TIPE II DI DESA KEMAMBANG," *Health J. Ilm. Kesehat.*, vol. 4, no. 2, Aug. 2022, doi: 10.35473/proheallth.v4i2.1507.
- [17] Kementerian Kesehatan Republik Indonesia, "Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 tentang Rekam Medis," *Kementerian Kesehatan Republik Indonesia*, Jakarta, 2022.
- [18] N. Mayasari, "Sosialisasi Pengetahuan Tentang Nilai Guna Rekam Medis Bagi Tenaga Rekam Medis Di Rumah Sakit Umum Selaguri Padang," *J. Community Engagem. Health*, vol. 3, no. 2, pp. 335–338, Sep. 2020.
- [19] R. M. Faska, S. K. Gusti, E. Budianita, and F. Syafria, "PENGARUH TEKNIK PENYEIMBANGAN DATA PADA

- KLASIFIKASI PENYAKIT NAFLD DENGAN ALGORITMA SVM,” vol. 7, no. 2, 2025.
- [20] R. Sinaga, “Data Mining untuk Evaluasi Kualitas Layanan Persalinan: Studi Komparatif RapidMiner dan SPSS,” vol. 2, no. 1, 2025.
- [21] M. Y. D. Deviyati, H. Oktavianto, G. Wijaya, and D. Irawan, “IMPLEMENTATION OF DATA MINING USING THE K-MEANS CLUSTERING METHOD TO DETERMINE TYPES OF EYE DISEASE AT WULUHAN HEALTH CENTER,” *J-STEM J. Sci. Technol. Eng. Math.*, vol. 1, no. 1, pp. 1–16, 2024.
- [22] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, p. 012017, Apr. 2018, doi: 10.1088/1757-899X/336/1/012017.
- [23] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *J. Ilm. Inform. Dan Ilmu Komput. JIMA-Ilk.*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [24] Y. Saputra Sy, “Klasterisasi Pasien Rawat Inap Peserta BPJS Berdasarkan Jenis Penyakit Menggunakan Algoritma K-Means,” *J. Sistim Inf. Dan Teknol.*, Sep. 2022, doi: 10.37034/jsisfotek.v5i2.162.
- [25] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, “ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER,” *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.
- [26] S. Maharani, Y. Wendra, and R. Rahim, “The Implementation of the K-Means Clustering Algorithm Based on the Severity Level of Diabetes in Patients Using a Website Platform,” vol. 4, no. 7, 2025.
- [27] Ananda Elang Satriatama Setyadji¹, Ari Prasetyo Wibowo², I Gusti Ngurah Arnold Matthew D³, Reyhan Bayu Pratama⁴, Tegar Alwinata Masyhuda⁵, Yohannes, and Alexander Agusti Sinaga⁶, Endah Purwanti^{7*}, Indah Werdiningsih^{8*}, “Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan Karakteristik Pasien,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 3, pp. 172–182, Jul. 2023.
- [28] K. B. Prasetyo and D. Oktafiani, “Clustering Data Diabetes Menggunakan Algoritma K-Means,” vol. 3, 2026.