



Penggunaan Algoritma K-Means dalam Pengelompokan Pasien Diabetes Mellitus Berdasarkan Parameter Klinis di Puskesmas Brebes

Fadia Feronika¹, Nur Ariesanto Ramdhan², Raden Muhammad Herdian Bhakti³

^{1,2,3}Teknik Informatika, Universitas Muhadi Setiabudi Brebes, Indonesia

Email author: fadiaferonika310@gmail.com¹, ariesantoramdhan@gmail.com², bhakti@umc.ac.id³

Article Info

Article history:

Received Januari 3, 2025

Revised Februari 17, 2025

Accepted June 28, 2025

Keywords:

Diabetes Mellitus

Data Mining

K-Means

ABSTRAK

Diabetes Mellitus merupakan salah satu penyakit kronis yang jumlah penderitanya terus bertambah setiap tahunnya, termasuk di wilayah Puskesmas Brebes. Banyaknya pasien dengan kondisi klinis yang beragam mendorong perlunya suatu metode untuk mengelompokkan pasien berdasarkan tingkat keparahannya. Penelitian ini bertujuan untuk menerapkan algoritma *K-Means* dalam proses pengelompokan pasien Diabetes Mellitus dengan menggunakan beberapa parameter klinis, yaitu Gula Darah Puasa (GDP), kadar HbA1c, Kolesterol Total (CHOL), serta tekanan darah sistolik dan diastolik. Pendekatan yang digunakan dalam penelitian ini adalah deskriptif kuantitatif dengan metode data mining berbasis algoritma K-Means. Data yang digunakan diperoleh dari rekam medis Puskesmas Brebes. Proses klusterisasi menghasilkan tiga kelompok, yaitu kategori risiko rendah, sedang, dan tinggi. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu melakukan pengelompokan data pasien secara akurat sesuai tingkat keparahan. Hasil tersebut kemudian divisualisasikan melalui sistem berbasis web yang bertujuan untuk mempermudah pihak puskesmas dalam menganalisis kondisi pasien serta mendukung pengambilan keputusan medis yang lebih efektif.

Corresponding Author:

Fadia Feronika,

Universitas Muhadi Setiabudi Brebes

Pangeran Diponegoro No.KM2, Rw. 11, Pesantunan, Kec. Wanasari, Kabupaten Brebes, Jawa Tengah

Email: fadiaferonika310@gmail.com



1. PENDAHULUAN

Seiring dengan pertumbuhan jumlah penduduk dan perkembangan zaman, jumlah pasien yang menderita berbagai penyakit terus mengalami peningkatan. Salah satu penyakit kronis yang cukup umum adalah Diabetes Mellitus, yang sering dikenal dengan sebutan diabetes [1]. Penyakit ini termasuk dalam kategori penyakit tidak menular, namun dapat diturunkan dari orang tua kepada anak. Selain faktor keturunan, Diabetes Mellitus juga dapat dipengaruhi oleh berbagai faktor lainnya seperti usia, jenis kelamin, riwayat diabetes dalam keluarga, kelebihan berat badan (obesitas), serta pola hidup yang tidak sehat [2].

Berdasarkan perkiraan dari International Diabetes Federation (IDF), pada tahun 2019 terdapat sekitar 463 juta orang dewasa berusia 20–79 tahun di seluruh dunia yang menderita diabetes, dengan prevalensi sebesar 9,3% dari total populasi dalam rentang usia tersebut. Jumlah ini diperkirakan akan meningkat sebesar 51% pada tahun 2045, sehingga mencapai sekitar 700 juta orang. Sementara itu, angka prevalensi diabetes mellitus pada anak-anak dan remaja di bawah usia 20 tahun mencapai 1,1%. Masih berdasarkan data IDF, secara global, negara dengan jumlah kasus diabetes mellitus tertinggi pada kelompok usia 20–79 tahun adalah Tiongkok dengan 116,4 juta kasus, disusul oleh India dengan 77 juta kasus, dan Amerika Serikat sebanyak 31 juta kasus. Indonesia sendiri menjadi satu-satunya negara di kawasan Asia Tenggara yang masuk dalam daftar 10 besar negara dengan jumlah penderita diabetes tertinggi di dunia [3].

Puskesmas sebagai fasilitas layanan kesehatan tingkat pertama sering menghadapi banyak pasien dengan beragam kondisi kesehatan. Untuk memberikan penanganan yang lebih efektif dan efisien, diperlukan suatu pendekatan yang dapat mengelompokkan pasien berdasarkan kesamaan kondisi klinis mereka. Pengelompokan ini dapat membantu dalam menetapkan prioritas penanganan, mencegah komplikasi, serta mengevaluasi program pengobatan secara lebih optimal. Salah satu pendekatan yang dapat digunakan adalah penerapan metode clustering dalam data mining.

K-Means merupakan metode pembelajaran tanpa pengawasan (unsupervised learning) yang sering digunakan untuk mengelompokkan data ke dalam beberapa klaster. Metode ini memiliki sejumlah keunggulan, seperti mudah diterapkan, efisien dan efektif dalam proses pengelompokan data, hasil yang mudah dipahami, serta tingkat akurasi yang cukup baik [4]. Dalam konteks pasien Diabetes Mellitus, beberapa indikator klinis seperti gula darah puasa (GDP), kadar Hemoglobin A1C (HbA1C), kolesterol total (CHOL), serta tekanan darah (sistolik dan diastolik) dapat digunakan sebagai variabel utama dalam proses pengelompokan data. Dengan menerapkan algoritma K-Means, pasien dapat dikelompokkan ke dalam beberapa klaster yang mencerminkan tingkat risiko atau kondisi kesehatan tertentu.

Penelitian ini dilakukan di Puskesmas Brebes, yang merupakan salah satu pusat layanan kesehatan dengan jumlah pasien yang cukup banyak, termasuk penderita Diabetes Mellitus. Dengan memanfaatkan data rekam medis yang tersedia, penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam mengelompokkan pasien Diabetes Mellitus berdasarkan empat parameter klinis utama, yaitu GDP, HbA1C, CHOL, serta tekanan darah (sistolik dan diastolik). Hasil dari proses klasterisasi ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai profil pasien, sekaligus menjadi dasar bagi pihak Puskesmas dalam merancang strategi penanganan yang lebih tepat sasaran. Selain itu, penelitian ini juga menjadi langkah awal dalam pemanfaatan teknologi data mining di bidang kesehatan, khususnya pada tingkat pelayanan primer.

Dalam metode pengumpulan data kepustakaan, langkah-langkah yang dilakukan meliputi membaca, mencatat, dan mengolah bahan-bahan penelitian setelah topik penelitian ditentukan. Setelah topik ditetapkan, peneliti mencari landasan teori dengan mengumpulkan informasi yang relevan dari berbagai sumber pustaka, seperti jurnal, hasil penelitian yang tersedia di internet, dan sumber lainnya.

Penelitian berjudul "Pengelompokan Data Rekam Medis pada Pasien Diabetes Menggunakan Metode Divisive Analysis Clustering" oleh MS Tarigan yang diterbitkan pada tahun 2024 membahas efisiensi dan akurasi dalam mengelompokkan data rekam medis pasien diabetes. Studi ini menyoroti kelemahan metode klasterisasi manual yang cenderung tidak efisien dan rentan terhadap kesalahan. Sebagai solusi, penelitian ini mengusulkan penggunaan metode Divisive Analysis Clustering (DIANA) untuk mengelompokkan data pasien, yang diimplementasikan menggunakan perangkat lunak RapidMiner. Hasilnya menunjukkan bahwa metode DIANA mampu mengelompokkan data secara efektif, meningkatkan efisiensi, serta mengurangi tingkat kesalahan dibandingkan dengan cara manual. Penerapan metode ini juga diharapkan dapat membantu dalam mengidentifikasi pola pada data pasien diabetes [5].

Penelitian berjudul "Pengelompokan Tingkat Risiko Penyakit Diabetes Mellitus Menggunakan Algoritma K-Means Clustering" oleh R Gestavito dan AI Hadiana yang diterbitkan pada tahun 2024 berfokus pada analisis Diabetes Mellitus (DM), yaitu kondisi metabolik kronis yang ditandai dengan tingginya kadar gula darah akibat kekurangan insulin atau ketidakmampuan tubuh merespons insulin secara efektif. Mengingat meningkatnya prevalensi DM, terutama di negara berkembang, penelitian ini bertujuan untuk mengkategorikan tingkat risiko penyakit DM. Metode yang digunakan adalah algoritma K-Means Clustering, yaitu teknik pengelompokan data berdasarkan kesamaan karakteristik. Dalam

penelitian ini, evaluasi dilakukan dengan jumlah kluster $k = 2$, yang berarti data risiko DM dikelompokkan ke dalam dua kategori [6].

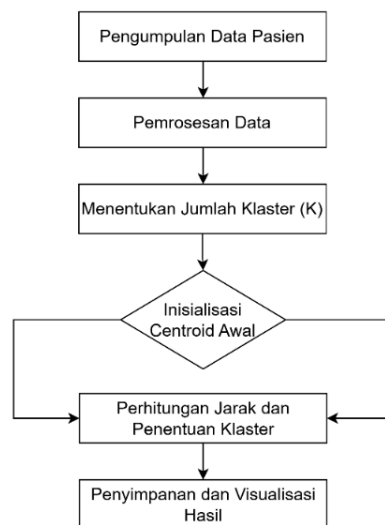
2. METODE PENELITIAN

2.1 Metodologi Penelitian

Penelitian ini menggunakan metode deskriptif kuantitatif dengan pendekatan data mining melalui penerapan algoritma K-Means untuk mengelompokkan pasien Diabetes Mellitus berdasarkan tingkat keparahan kondisi klinisnya. Data yang digunakan berasal dari 107 rekam medis pasien di Puskesmas Brebes, dengan lima atribut utama yang dianalisis, yaitu Gula Darah Puasa (GDP), kadar HbA1c, Kolesterol Total (CHOL), serta tekanan darah sistolik dan diastolik. Sebelum proses klasterisasi dilakukan, data terlebih dahulu diproses melalui tahapan pra-pemrosesan, meliputi pembersihan data (data cleaning) untuk menghilangkan nilai kosong atau tidak valid, serta normalisasi agar setiap atribut berada pada skala yang seimbang.

Klasterisasi dilakukan menggunakan algoritma K-Means dengan jumlah kluster yang ditetapkan sebanyak tiga kategori, yaitu risiko rendah, sedang, dan tinggi. Proses algoritma diawali dengan inialisasi centroid secara acak, dilanjutkan dengan perhitungan jarak antara masing-masing data ke centroid menggunakan rumus Euclidean. Data kemudian dikelompokkan ke kluster terdekat, dan posisi centroid diperbarui berdasarkan rata-rata nilai dalam kluster. Langkah ini diulang hingga nilai centroid tidak berubah secara signifikan, menandakan proses telah mencapai konvergensi.

Hasil klasterisasi dianalisis untuk melihat karakteristik dominan pada masing-masing kelompok yang merepresentasikan tingkat risiko pasien. Selanjutnya, hasil tersebut diterapkan dalam sistem berbasis web yang dirancang untuk memudahkan pihak puskesmas dalam menampilkan hasil pengelompokan dalam bentuk tabel dan grafik. Sistem ini diharapkan dapat menjadi alat bantu dalam memahami kondisi pasien secara lebih terstruktur serta mendukung pengambilan keputusan medis yang lebih tepat dan efisien.



Gambar 1. Alur Metode K-Means

2.2 Diabetes Mellitus

Diabetes Mellitus merupakan penyakit kronis yang terjadi ketika pankreas tidak mampu memproduksi insulin dalam jumlah cukup atau tubuh tidak dapat menggunakan insulin secara efektif. Metode C4.5 melibatkan variabel seperti Jenis Kelamin, Usia, Jenis Masuk Rawat, Status Pulang dari Rumah Sakit, Sumber Rujukan, Diagnosis 1–3, Hasil Tes HbA1c, Perubahan Obat, Penggunaan Obat Diabetes, dan Riwayat Rawat Inap Ulang. Pendekatan ini menjadi alternatif bagi rumah sakit atau klinik yang belum memiliki fasilitas lengkap untuk pemeriksaan langsung. Kadar gula darah tinggi atau

hiperglikemia merupakan kondisi umum pada diabetes yang tidak terkontrol dan dapat menyebabkan kerusakan serius pada berbagai sistem tubuh, terutama saraf dan pembuluh darah [7].

2.3 Data Mining

Data mining merupakan teknik yang digunakan untuk mengolah data dalam jumlah besar dengan cara mengidentifikasi pola-pola tertentu guna menghasilkan informasi yang bernilai dan bermanfaat. Data yang dianalisis dapat berasal dari berbagai sumber, seperti data warehouse, web, basis data, dan lainnya. Penerapan data mining mencakup berbagai bidang, seperti peningkatan penjualan, pengembangan pengetahuan, serta sektor lain yang membutuhkan analisis data untuk mendukung pengambilan keputusan yang lebih efektif [8].

2.4 K-Means Clustering

K-means clustering merupakan salah satu algoritma yang dapat membagi data ke dalam kluster sesuai jumlah yang diinginkan dengan terlebih dahulu menentukan centroid. Metode ini digunakan karena pengelompokan data perlu dilakukan dengan membentuk kelompok atau kelas berdasarkan taksonomi atau kategori tertentu [9].

$$d(x_i, y_i) = \sqrt{\sum (x_i - y_i)^2}$$

Dimana :

$d(x_i, y_i)$: Jarak antara objek ke-x dengan objek ke-y

x_i : Nilai objek ke-x pada variabel k

y_i : Nilai objek ke-y pada variabel k

3. HASIL DAN PEMBAHASAN

3.1. Analisis Kebutuhan

Pada tahap ini, informasi yang dibutuhkan untuk pengembangan aplikasi diperoleh melalui pengumpulan data dengan cara observasi langsung serta studi literatur yang relevan. Beberapa data yang digunakan mencakup poin-poin penting yang diperlukan untuk menentukan tingkat risiko pada pasien Diabetes Mellitus. Data yang dikumpulkan meliputi Nama Pasien, Gula Darah Puasa (GDP), kadar HbA1c, kolesterol (CHOL), serta tekanan darah sistolik dan diastolik.

3.2. Jenis Data

3.2.1 Data Yang Digunakan

Seluruh data yang digunakan dalam penelitian ini berjumlah 107 data, yang memuat informasi medis dari pasien yang telah terdiagnosis Diabetes Mellitus, meliputi:

Tabel 3. Dataset Pasien Diabetes Mellitus

Pasien	Jenis Kelamin	Usia	GDP	Hba1c	CHOL	Sistolik	Diastolik
1	L	56	162	10	221	140	80
2	L	57	179	13.3	218	160	90
3	L	57	333	12.5	261	140	90
4	L	43	158	7.4	158	120	80
5	L	45	200	6.3	222	120	80
6	P	58	285	8.7	170	130	80
7	P	64	190	9.3	244	100	70

8	L	62	129	5.9	221	120	80
9	P	69	133	5.9	274	150	80
10	P	66	288	6.9	265	140	80
11	L	60	115	6	171	100	70
12	P	73	88	6.4	270	140	80
13	P	58	117	7	264	140	90
14	L	54	114	6.9	159	150	90
15	P	66	101	8.6	216	130	80
16	L	63	160	10.9	143	130	80
17	L	64	98	6.9	187	150	80
18	L	49	110	6.3	213	110	80
19	P	44	202	10.8	195	110	80
20	P	59	234	10.3	223	150	90
21	P	40	149	7.8	234	120	70
22	P	61	109	7.9	289	160	80
23	P	51	296	13.2	234	140	80
24	L	70	180	8	192	160	80
25	P	57	230	9.4	242	100	70
26	L	63	113	7	193	100	80
27	L	51	195	9.2	176	130	80
28	P	74	89	5.1	194	100	70
29	P	62	270	10	239	170	80
30	P	73	535	10.4	204	120	80
31	P	61	255	7	160	160	90
32	P	61	151	7.1	198	130	80
33	P	49	211	11.8	241	160	80
34	P	56	140	9.2	196	150	90
35	P	50	311	9.5	261	100	80
36	P	58	218	9.3	213	110	80
37	P	46	92	8	171	130	80
38	L	65	394	6.2	180	100	80
39	P	66	124	9.1	199	150	80
40	P	31	397	10	213	90	80
41	L	79	114	5.9	206	130	80

42	P	59	191	10.4	206	130	80
43	P	68	130	10.3	198	100	80
44	P	68	217	6.2	240	120	80
45	L	62	99	6.7	211	160	90
46	P	73	90	7.1	220	120	80
47	L	59	206	6	216	110	80
48	L	75	115	5.3	281	150	90
49	P	68	220	10	150	110	70
50	P	61	100	13.3	144	130	80

3.2.2 Menentukan Jumlah Cluster

Setelah data tersebut ditetapkan, langkah selanjutnya adalah menentukan jumlah kluster yang akan dibuat. Adapun kluster yang dirancang meliputi:

Tabel 4. Keterangan Cluster

Cluster	Keterangan
1	Rendah
2	Sedang
3	Tinggi

3.2.3 Menentukan Pusat Cluster

Setelah jumlah cluster ditentukan, langkah selanjutnya adalah memilih pusat cluster awal atau centroid. Centroid awal dipilih secara acak dari data pasien yang tersedia.

Tabel 5. Penentuan Centroid Awal Secara Acak

Centroid	K1	K2	K3	K4	K5
C1	88	6.4	270	140	80
C2	179	13.3	218	160	90
C3	190	9.3	244	100	70

3.1.3 Perhitungan Jarak Data dengan Centroid

Langkah berikutnya adalah melakukan perhitungan jarak menggunakan rumus Euclidean Distance untuk mengukur kedekatan setiap data pasien dengan masing-masing centroid yang telah ditetapkan sebelumnya. Tujuan dari perhitungan ini adalah untuk menentukan seberapa dekat data pasien terhadap tiap pusat kluster. Hasil perhitungan berupa nilai jarak yang kemudian dibandingkan, dan data pasien akan dimasukkan ke dalam kluster dengan jarak terpendek. Proses pengukuran jarak ini dilakukan menggunakan rumus Euclidean Distance.

Centroid awal merupakan titik pusat dari nilai masing-masing kriteria, dengan hasil pasien Centroid 1 K1 = 88, K2 = 6.4, K3 = 270, K4 = 140, K5 = 80, lalu Centroid 2 K1 = 179, K2 = 13.3, K3 = 218, K4 = 160, K5 = 90, Centroid 3 K1 = 190, K2 = 9.3, K3 = 244, K4 = 100, K5 = 70.

Tabel 6. Hasil Perhitungan Centroid

Pasien	Jarak ke C1	Jarak ke C2	Jarak ke C3
1	48.5298	43.7514	178.367
2	71.396	41.9486	164.727
3	223.34	145.364	36.2706
4	73.8325	62.7606	195.291
5	86.3849	20.8822	140.099
6	176.722	90.8969	80.6676
7	86.903	51.9369	153.229

8	153.229	74.2046	211.008
9	63.1346	99.9955	212.778
10	180.292	107.692	64.2723
...
50	127.038	60.6984	144.952

Dalam penentuan *cluster* dilakukan berdasarkan nilai dari *Euclidean Distance* yang telah di hitung sebelumnya pada tabel 4. Untuk penentuannya, dari tiap baris tersebut yang memiliki nilai *Euclidean Distance* terendah maka akan menjadi *cluster* dari kelompok *Euclidean Distance* tersebut.

Tabel 7. Penentuan Hasil *Cluster*

Jarak ke C1	Jarak ke C2	Jarak ke C3	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
48.5298	43.7514	178.367	-	1	-
71.396	41.9486	164.727	-	1	-
223.34	145.364	36.2706	-	-	1
73.8325	62.7606	195.291	-	1	-
86.3849	20.8822	140.099	-	1	-
176.722	90.8969	80.6676	-	-	1
86.903	51.9369	153.229	-	1	-
153.229	74.2046	211.008	1	-	-
63.1346	99.9955	212.778	1	-	-
180.292	107.692	64.2723	-	-	1
...	-	-	-
127.038	60.6984	144.952	-	1	-

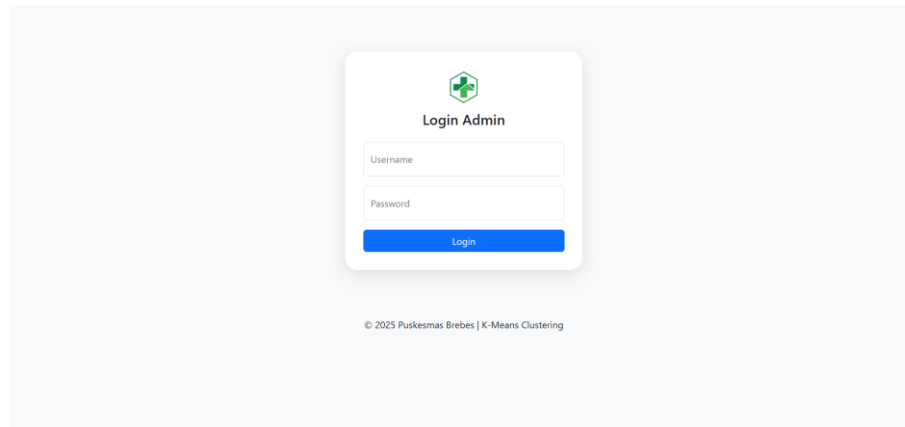
Tabel 7 menunjukkan penentuan nilai *centroid* akhir untuk ke 5 atribut di tabel 5, di mana nilai tersebut dihitung berdasarkan hasil pengelompokan pada Tabel 6 dengan menggunakan rumus *Euclidean Distance*. Kolom nilai *centroid* merupakan hasil dari perhitungan yang tercantum pada kolom perhitungan. Apabila nilai setiap *centroid* baru sama dengan nilai *centroid* sebelumnya, maka proses *K-Means* dianggap selesai. Namun, jika nilainya berbeda, maka proses dilanjutkan kembali ke langkah kedua, yaitu menghitung jarak *Euclidean* dari setiap *centroid* menggunakan nilai *centroid* yang baru.

3.1.4 Tampilan Sistem

Tampilan sistem merupakan bagian dari tampilan pengguna (*user interface*) yang berfungsi sebagai media interaksi antara user dengan aplikasi clustering K-Means. Sistem ini dibangun dalam bentuk web menggunakan PHP native, dengan Bootstrap untuk mendukung tampilan yang responsif, serta MySQL sebagai basis penyimpanan datanya.

1. Tampilan Login

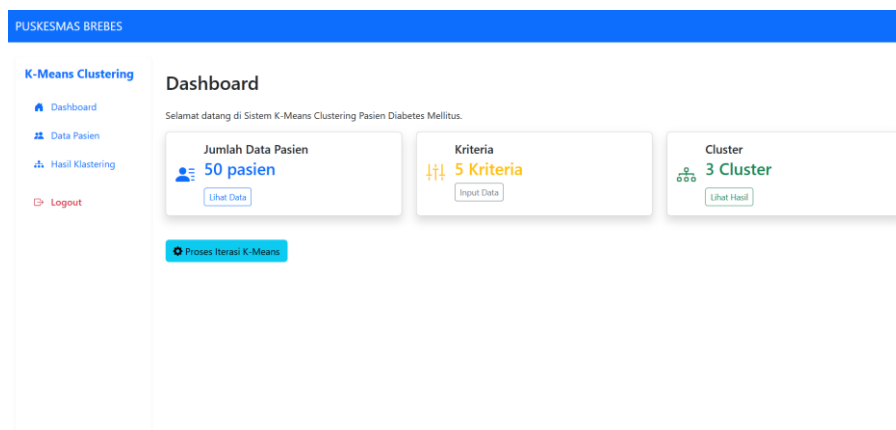
Tampilan ini berfungsi sebagai proses autentikasi pengguna sebelum masuk ke dalam sistem. Pengguna diwajibkan memasukkan username dan password untuk dapat melanjutkan akses ke fitur sistem selanjutnya.



Gambar 2. Tampilan Login

2. Tampilan Dashboard

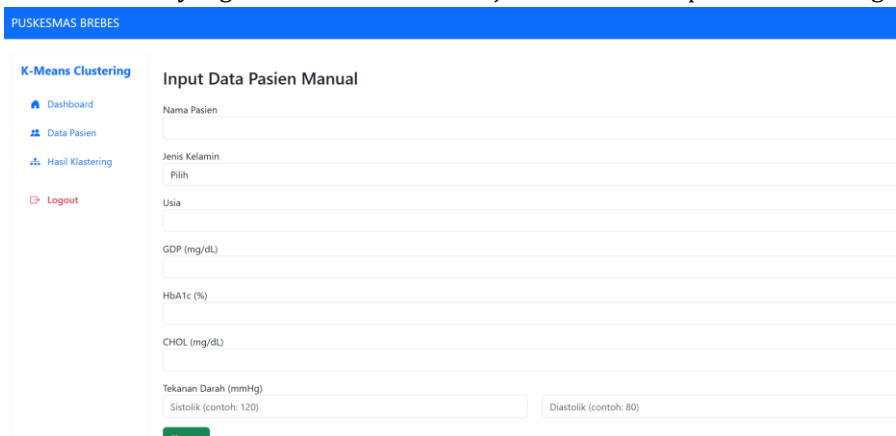
Tampilan ini merupakan tampilan utama yang muncul setelah pengguna berhasil melakukan login. Di dalamnya ditampilkan informasi seperti jumlah data pasien, jumlah kriteria yang digunakan, hasil clustering berdasarkan masing-masing kluster, serta proses iterasi yang dilakukan oleh algoritma K-Means.



Gambar 3. Tampilan Login

3. Tampilan Input Data

Halaman ini berfungsi untuk menambahkan data pasien, baik melalui input manual maupun dengan mengunggah file Excel. Data yang dimasukkan akan menjadi dasar dalam proses clustering.



Gambar 4. Tampilan Input Data

4. Tampilan Hasil Perhitungan

Tampilan ini menampilkan hasil perhitungan pada iterasi ke-5, yang mencakup data pasien, hasil perhitungan setelah diproses, serta kluster yang diperoleh dari proses tersebut.

No	Nama Pasien	Jarak ke C1	Jarak ke C2	Jarak ke C3	Cluster	Keterangan (Stadium)
1	ABDUL SYUKUR	48.5298	43.7514	178.367	1	Sedang
2	ADI WINOTO	71.396	41.9486	164.727	1	Sedang
3	ALI NURDIN	223.34	145.364	36.2706	2	Tinggi
4	AMONG BUDIONO	73.8325	62.7606	195.291	1	Sedang
5	ANDI SETYANTO	86.3849	20.8822	140.099	1	Sedang
6	ANI RIYANTI	176.722	90.8969	80.6676	2	Tinggi
7	ANITA HANDAYANI	86.903	51.9369	153.229	1	Sedang
8	ANSORI HASAN	18.9683	74.2046	211.008	0	Rendah
9	ASLIHAH	63.1346	99.9955	212.778	0	Rendah
10	CASIPAH	180.292	107.692	64.2723	2	Tinggi
11	CASMUAN	180.292	107.692	64.2723	2	Tinggi
12	CHURIYAH	56.415	96.3899	233.964	0	Rendah
13	DAIRAH	60.2753	131.223	255.336	0	Rendah
14	DALIM	49.1008	104.268	225.97	0	Rendah

Gambar 4. Tampilan Hasil Perhitungan

4. KESIMPULAN

Berdasarkan hasil penelitian dan implementasi sistem, algoritma K-Means terbukti mampu mengelompokkan data pasien Diabetes Mellitus di Puskesmas Brebes secara efektif ke dalam tiga kluster risiko: rendah, sedang, dan tinggi. Proses ini menggunakan atribut GDP, HbA1c, kolesterol, serta tekanan darah sistolik dan diastolik. Sistem yang dikembangkan telah mengotomatisasi perhitungan, menampilkan hasil dalam bentuk visualisasi, serta mendukung input data manual dan melalui file Excel. Hasil kluster disimpan secara permanen untuk mendukung analisis dan pengambilan keputusan medis secara lebih efisien.

SARAN

Untuk pengembangan selanjutnya, sistem sebaiknya mendukung pembaruan centroid otomatis saat data baru ditambahkan, serta mampu mengolah data historis pasien untuk analisis jangka panjang. Dari segi keamanan, perlu diterapkan enkripsi data dan pembatasan akses berdasarkan level pengguna. Selain itu, algoritma lain seperti DBSCAN atau Hierarchical Clustering dapat diuji sebagai pembandingan. Sistem juga disarankan diuji dengan data real-time dan skala lebih besar untuk menilai performa dan keakuratan secara menyeluruh.

REFERENCES (10 PT)

- [1] N. Ariani, R. Alfian, and E. Prihandiwati, "Tingkat Perilaku Pengobatan, Kepatuhan Minum Obat, Dan Kadar Gula Darah Pasien Diabetes Mellitus Rawat Jalan Di Rsud Briggend. H. Hasan Basry Kandangan," *J. Ilm. Manuntung*, vol. 8, no. 1, pp. 156–162, 2022, doi: 10.51352/jim.v8i1.523.
- [2] A. Nugrahaeni and E. Widianawati, "Persebaran Kasus Diabetes Melitus Pasien Rumah Sakit Telogorejo Berbasis Wilayah Kota Semarang Tahun 2020," *J. Bina Cipta Husada J. Kesehat. Dan Sci.*, vol. 18, no. 2, pp. 89–98, 2022.
- [3] Bunga Farchati, K. D. Pertiwi, and Ita Puji Lestari, "Faktor Risiko Diabetes Mellitus di Wilayah Kerja Puskesmas Gunungpati Kota Semarang," *Pro Heal. J. Ilm. Kesehat.*, vol. 5, no. 1, pp. 333–339, 2023, doi: 10.35473/proheallth.v5i1.2143.
- [4] Yovan Febriawan Nurpratama, Dhian Satria Yudha Kartika, and Reisa Permatasari, "Implementasi Folium pada Hasil Kluster Diabetes Mellitus di Puskesmas Modopuro," *J. Ilm. Tek. Inform. dan Komun.*, vol. 3, no. 3, pp. 74–85, 2023, doi: 10.55606/juitik.v3i3.623.

- [5] A. C. Method, “Pengelompokan Data Rekam Medis pada Penyakit Diabetes menggunakan Metode Divisive Analysis Clustering Clustering Medical Record Data on Diabetes Disease using Divisive,” vol. 13, pp. 1718–1731, 2024.
- [6] R. Gestavito, A. Id Hadiana, F. Rakhmat Umbara, and U. Jenderal Achmad Yani Jl Terusan Jenderal Sudirman, “Pengelompokan Tingkat Risiko Penyakit Diabetes Melitus Menggunakan Algoritma K-Means Clustering,” *J. Masy. Inform. Unjani*, vol. 8, no. 1, pp. 16–35, 2024.
- [7] L. Pebrianti, F. Aulia, H. Nisa, and K. Saputra, “Informasi Implementasi Metode Adaboost untuk Mengoptimasi Klasifikasi Penyakit Diabetes dengan Algoritma Naïve Bayes,” *J. Sist. dan Teknol.*, vol. 7, no. 2, pp. 122–127, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/view/8627%0Ahttp://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/download/8627/4296>
- [8] Lestari, Zulkarnain, Sijid, and S. Aisyah, “Diabetes Melitus: Review Etiologi, Patofisiologi, Gejala, Penyebab, Cara Pemeriksaan, Cara Pengobatan dan Cara Pencegahan,” *UIN Alauddin Makassar*, vol. 1, no. 2, pp. 237–241, 2021, [Online]. Available: <http://journal.uin-alauddin.ac.id/index.php/psb>
- [9] A. P. Wijaya, A. Premana, and N. A. Ramdhan, “Penerapan Algoritma K-Means pada Klasterisasi Data kawalcovid19.id,” *Pros. Sains Nas. dan Teknol.*, vol. 12, no. 1, p. 479, 2022, doi: 10.36499/psnst.v12i1.7294.

NOMENKLATUR

Nomenklatur disertai arti dari semua persamaan matematika ataupun nomenklatur lain di alam artikel, dituliskan pada bagian ini.

K-Means	arti dari Algoritma clustering non-hierarki yang digunakan untuk mengelompokkan data berdasarkan kedekatan terhadap pusat kluster (centroid).
Clustering	arti dari Proses pengelompokan data ke dalam beberapa kelompok berdasarkan kemiripan karakteristik.
Centroid	arti dari Titik tengah dari sebuah kluster yang merepresentasikan rata-rata nilai dari seluruh data dalam kluster tersebut.
Euclidean Distance	arti dari Metode perhitungan jarak antar dua titik dalam ruang multidimensi yang digunakan dalam proses pengelompokan.
GDP (Gula Darah Puasa)	arti dari Nilai kadar glukosa darah seseorang yang diukur setelah berpuasa minimal 8 jam.
HbA1c	arti dari Parameter yang menunjukkan kadar rata-rata gula darah dalam 2-3 bulan terakhir.
CHOL (Cholesterol)	arti dari Kadar kolesterol total dalam darah.
Sistolik	arti dari Tekanan darah saat jantung memompa darah keluar.
Diastolik	arti dari Tekanan darah saat jantung berelaksasi di antara dua detak.
Cluster 1 (Rendah)	arti dari Kelompok pasien dengan risiko rendah terhadap komplikasi Diabetes Mellitus berdasarkan parameter klinis.
Cluster 2 (Sedang)	arti dari Kelompok pasien dengan risiko sedang.
Cluster 3 (Tinggi)	arti dari Kelompok pasien dengan risiko tinggi.
Data Cleaning	arti dari Tahap prapemrosesan data untuk menghilangkan nilai kosong atau tidak valid.
Normalisasi	arti dari Proses penskalaan ulang data agar semua atribut memiliki bobot yang seimbang.
Iterasi	arti dari Pengulangan proses dalam algoritma K-Means sampai hasil clustering stabil (konvergen).
MySQL	arti dari Sistem manajemen basis data yang digunakan untuk menyimpan data pasien.
PHP Native	arti dari Bahasa pemrograman sisi server yang digunakan untuk membangun sistem berbasis web tanpa framework.
Bootstrap	arti dari Framework CSS untuk membangun tampilan web yang responsif.
Data Mining	arti dari Teknik analisis data untuk menemukan pola atau informasi bermanfaat dari sejumlah besar data.
Deskriptif Kuantitatif	arti dari Metode penelitian yang menggunakan angka/statistik untuk mendeskripsikan fenomena atau gejala.

