Jurnal SAINTEKOM (Sains dan Teknologi Komputasi) Vol. 01, No. 02, September, 2025
Lembaga Penelitian, Pengembangan, dan Pengabdian Kepada Masyarakat UNBIN
e-ISSN: 3109 - 1350

Received: 31 Maret 2025; Accepted: 28 April 2025; Published: Septeber 2025
DOI: 10.36350/jskom.v1i2.42

Penerapan *Metode K-Means* pada Pemetaan Persebaran Penyakit Diabetes untuk Rekomendasi Prioritas Pemberian Penyuluhan

Priscila[1*], Hardi Jamhur[2]
[1,2]Program Studi Sistem Informasi, Fakultas Informatika dan Komputer, Universitas Binaniaga Indonesia email: cilawinata44@gmail.com
*Corresponding Author

ABSTRACT

The process of mapping the distribution of diabetes is a grouping of the distribution of diabetes based on various criteria which will later be grouped (clustered) based on the distribution of diabetes sufferers, whether high, medium, low, to help health workers in providing data and information references in order to determine strategies for providing counseling about diabetes in each sub-district that can be implemented in the next period. In this study, an application was developed that can determine the distribution of diabetes sufferers using the K-Means Clustering Algorithm approach, namely by analyzing the initial data group, transforming the initial data and grouping it. In it, the variables applied are smoking, lack of physical activity, excessive sugar, excessive salt, excessive fat, lack of eating vegetables and fruits, alcohol consumption, and sugar checks. This is done to see the distribution of diabetes, in order to help health workers in providing references. In the application that was built, a feasibility test has been carried out and a feasibility percentage of 100% was obtained which can be categorized into the "Very Feasible" interpretation. User testing has been conducted using the PSSUQ questionnaire according to the PSSUQ category including Overall of 73.2%, Sysuse of 69.01%, Infoqual of 78.6%, Interqual of 71.5%, which means the application is worthy of use. And the validity test of the cluster using the Silhouette Coefficient against the K-Means algorithm applied with a value of 0.503 which means the cluster created is included in the "weak structure" category.

***Keywords**: Diabetes, Clustering, Distribution, Providing Counseling, K-Means Algorithm*

ABSTRAK

Proses pemetaan persebaran penyakit diabetes merupakan pengelompokan persebaran penyakit diabetes berdasarkan dari berbagai kriteria yang nantinya dikelompokkan (klasterisasi) berdasarkan persebaran penderita diabetes baik itu tinggi, sedang, rendah guna membantu pihak tenaga kesehatan dalam menyediakan acuan data dan informasi agar dapat menentukan strategi pemberian penyuluhan mengenai penyakit diabetes di setiap kelurahan yang dapat dijalankan di periode selanjutnya. Pada penelitian ini dikembangkan sebuah aplikasi yang dapat mengetahui persebaran penderita penyakit diabetes dengan pendekatan Algoritma K-Means Clustering yaitu dengan menganalisis kelompok data awal, mentransformasi data awal dan melakukan pengelompokkan. Didalamnya diterapkan variabel-variabel yaitu merokok, kurang aktifitas fisik, gula berlebihan, garam berlebihan, lemak berlebihan kurang makan sayur dan buah, konsumsi alkohol, dan pemeriksaan gula. Hal ini dilakukan untuk melihat persebaran penyakit diabetes, supaya dapat membantu pihak tenaga kesehatan dalam menyediakan acuan. Pada aplikasi yang dibangun telah dilakukan uji kelayakan dan diperoleh presentase kelayakan 100% yang dapat dikaterogikan kedalam interpretasi yang “Sangat Layak”. Telah dilakukan uji pengguna dengan menggunakan kuesioner PSSUQ sesuai dengan kategori PSSUQ diantaranya yaitu Overall sebesar 73,2% , Sysuse sebesar 69,01% , Infoqual sebesar 78,6% , Interqual sebesar 71,5% , yang artinya aplikasi layak digunakan. Serta telah uji validitas cluster menggunakan Silhouette Coefficient terhadap algoritma K-Means yang diterapkan dengan nilai yang di dapat sebesar 0,503 yang berarti klaster yang dibuat termasuk dalam kategori “weak structure”.

Keywords: Penyakit Diabetes, Klasterisasi, Persebaran, Pemberian Penyuluhan, Algoritma K-Means

A. PENDAHULUAN

1. Latar Belakang Masalah

Angka kematian yang disebabkan oleh penyakit diabetes ini semakin banyak dan setiap tahunnya diperkirakan akan terus meningkat angka kasus kematiannya. Namun belum dapat dilihat persebaran penyakit diabetes (tinggi, sedang, rendah) di masing-masing daerah. Pemberian penyuluhan dan edukasi kesehatan yang bertujuan untuk meningkatkan kesadaran dan pencegahan terhadap diabetes belum dapat dilakukan secara optimal, karena tidak sesuai dengan tingkat prioritas persebaran penyakit di setiap daerah. Maka pemberian penyuluhan masih belum sesuai dengan tingkat prioritas persebaran penyakit diabetes. Dengan menggunakan teknik komputasi, proses klasterisasi akan menjadi lebih efektif dan efisien.

2. Permasalahan

Dalam melakukan upaya kegiatan penyuluhan maupun sosialisasi pada masyarakat dalam pencegahan dan penanggulangannya terhadap penyakit diabetes ini, terkadang pemilihan lokasi kurang sesuai dengan tingkat kerawanan kasus diabetes. Menjadikan pada saat ini masih belum dapat diketahui tingkat persebaran mengenai pemetaan pemberian penyuluhan terhadap penyakit diabetes. Dalam hal ini dapat mempengaruhi langkah pihak terkait untuk melakukan upaya pencegahan dan penanganan pemberian penyuluhan terhadap penyakit diabetes.

Tabel 1 Penderita Diabetes Puskesmas Tanah Sareal Tahun 2023

PASIEN	FAKTOR RISIKO							PEMERIKSAAN GULA
	MEROKOK	KURANG AKTIFITAS FISIK	POLA MAKAN				KONSUMSI ALKOHOL	
			GULA BERLEBIHAN	GARAM BERLEBIHAN	LEMAK BERLEBIHAN	KURANG MAKAN BUAH DAN SAYUR		
P1	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	234
P2	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	376
P3	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	211
P4	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	231
P5	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	200
P6	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	210
P7	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	250
…	…	…	…	…	…	…	…	…

Berdasarkan dari permasalahan yang diuraikan tersebut, maka dapat di identifikasi masalah yaitu:

a. Belum dapat diketahui pemetaan persebaran penyakit diabetes di masing-masing daerah untuk rekomendasi prioritas pemberian penyuluhan.
b. Belum dapat diketahui efektifitas penerapan metode K-Means pada pemetaan persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan**.**

3. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

a. Mendapatkan hasil persebaran penyakit diabetes di masing-masing daerah.
b. Mendapatkan efektifitas penerapan metode K-Means pada pemetaan persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan.
c. Mengembangkan prototype aplikasi K-Means pada pemetaan persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan.
d. Mengukur persebaran penyakit diabetes di masing-masing daerah dan ke efektifitas penerapan metode K-Means pada pemetaan persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan.

4. Tinjauan Pustaka

a. Pengertian Data Mining

Dalam proses menemukan informasi yang menarik terdapat sebuah data yang akan diolah dengan menggunakan metode tertentu. Menurut (Arhami and Nasir, 2020, p. 2-5) data mining merupakan proses untuk menganalisa sebuah pola pada data menjadi sebuah informasi yang berguna sehingga dapat membantu dalam membuat keputusan yang lebih baik. Memiliki tujuan untuk menemukan sebuah pola yang memiliki arti dan maksud tertentu yang dapat digunakan untuk menggali dan menganalisis dari sebagian besar data yang ada. Proses yang terjadi seperti pengambilan informasi, analisa pola dan ringkasan pengetahuan yang memiliki pemahaman mengenai data serta menggiring pengukuran secara konstruktif dari area yang terlibat. Data mining memiliki beberapa metode diantaranya klasifikasi, clustering, asosiasi, estimasi, prediksi dan regresi. Pada penelitian ini menggunakan metode clustering. (Arhami and Nasir, 2020, p. 149) mengutarakan bahwa clustering merupakan data atau nilai yang belum memiliki label pada kelasnya perlu diprediksi ke dalam sebuah kelas untuk mengetahui objek tersebut akan dimasukan pada kelas yang sesuai berdasarkan kesamaan pola atau karakteristik pada kelompoknya. Pada penelitian ini dalam memetakan penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan digunakan model clustering dengan pendekatan partisi clustering.

b. Algoritma K-Means

Menurut (Santosa, 2007, p. 42) K-Means merupakan sebuah teknik klastering yang umum dikenal dan paling sederhana, dengan K-Means dapat mengelompokkan suatu objek kedalam sebuah klaster. (Prasetyo, 2014, p. 189) mengutarakan bahwa algoritma K-Means sederhana untuk diimplementasikan, mudah dipahami, umum dalam penggunaannya dan relatif cepat.

B. METODE

1. Algoritma K-Means

Dalam penelitian ini, pemetaan penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan diperlukannya sebuah algoritma. Dengan algoritma K-Means dapat dilakukan pembagian data ke dalam sebuah kelompok yang memiliki karakteristik yang serupa dikelompokkan ke dalam satu set yang serupa, sebaliknya jika memiliki karakteristik berbeda, maka akan dikelompokkan ke dalam set yang berbeda pula, hal ini dilakukan untuk meminimalkan nilai variatif pada suatu kelompok (Putra et al., 2023, p. 87-88).

Menurut (Arhami and Nasir, 2020, p. 148-149) berikut ini langkah-langkah dasar pada algoritma K-means yaitu:

1. tentukan nilai k klaster sesuai dengan yang diinginkan;
2. pilih titik-titik atau sampel yang menjadi anggota klaster secara acak;
3. tentukan nilai centroid atau titik tengah dari klaster tersebut dengan rumus;

$$M_k = \left(\frac{1}{n_k}\right)\sum_{i=i}^{nk} X_{ik}$$

4. hitung *square error* untuk setiap klaster $C_k$yang merupakan jumlah kuadrat dari jarak Euclidean antara tiap sampel dalam $C_k$dan titik tengahnya (*centroid)*, error dikenal juga dengan nama *within cluster variation (WCV)*, yaitu;

$$e_k^2 = \sum_{i=1}^{nk} (X_{ik} - M_k)^2$$

5. selajutnya jumlah dari keseluruhan error dari k-*cluster* juga dihitung dengan rumus;

$$E_k^2 = \sum_{k=1}^{k} e_k^2$$

6. kelompokkan kembali semua sampel berdasarkan jarak minimum dari masing-masing pusat $M_1, M_2, \ldots, M_k$ sehingga diperoleh distribusi baru dari sampel sesuai kelasternya, untuk memperoleh distribusi baru dari sampel baru tersebut dapat dilakukan dengan menghitung jarak masing-masing titik pusat dengan keseluruhan sampel d(M_1, x_1) … d(M_k, x_k); perhitungan jarak dari masing-masing titik tersebut dapat menggunakan beberapa metode, contohnya *Euclidean Distance,* dimana jarak antara dua titik dalam satu, dua, tiga atau sampai dimensi n dapat dihitung sebagai berikut;

$$d(p,q) = \sqrt{((p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2)}$$

7. tuliskan hasil anggota klaster baru sesuai dengan hasil yang diperoleh pada langkah ke-5 dan ulangi langkah ke-3 sampai beberapa iterasi yang nantinya akan ditemukan nilai total *square error* turun secara signifikan.

2. Teknik Analisis Data

Pengujian model yang dilakukan untuk menekankan kedekatan relasi antar objek dan mengetahui seberapa jauh klaster terpisah dengan klaster lainnya dan kombinasi dari proses agregasi dan pemisahan disebut sebagai *Silhouette Coefficient.* Berikut tahapan perhitungan *Silhouette Coefficient* menurut (Handoyo, M, & Michrandi, 2014) sebagai berikut;

(1) menghitung rata-rata jarak dari suatu data dengan permisalan i dengan semua data lain yang berada dalam satu klaster;

$$a(i) = \frac{1}{|A|-1}\sum j \in_{A, j \neq i} d(i,j)$$

dengan j adalah data lain dalam suatu klaster A dan d(i, j) merupakan jarak antara data i dengan j;

(2) menghitung rata-rata jarak dari data i tersebut dengan semua data pada klaster lain dan diambil nilai terkecilnya;

$$d(i,C) = \frac{1}{|A|}\sum j \in C\ d(i,j)$$

dengan d(i, j) merupakan jarak rata-rata data i dengan semua objek pada klaster lain C dimana A ≠ C;

(3) dengan rumus Silhouette Coefficient sebagai berikut;

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

dimana s(i) merupakan semua rata-rata pada semua data.
Untuk menilai *Silhouette Coefficient* dapat dilihat pada tabel 2 sebagai berikut:

Tabel 2 Tabel Nilai Silhouette Coefficient

No	**Nilai *Silhouette Coefficient***	**Struktur**
1	0,71 - 1,00	*Strong Structure*
2	0,51 - 0,70	*Medium Structure*
3	0,26 - 0,50	*Weak Structure*
4	≤ 25	*No Structure*

Sumber : (Rousseeuw, 1987)

C. HASIL DAN PEMBAHASAN

1. HASIL

a. Penentuan Variabel Penelitian

Variabel yang digunakan dalam pemetaan persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan adalah Merokok, Kurang Aktifitas Fisik, Gula Berlebihan, Garam Berlebihan, Lemak Berlebihan, Kurang Makan Buah dan Sayur, Konsumsi Alkohol, Pemeriksaan Gula dan Kelurahan. Tabel 3 menjelaskan definisi tiap variabel yang digunakan dalam pemetaan persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan.

Tabel 3 Variabel Penelitian

No	Variabel	Definisi
1	Merokok	Menyatakan faktor risiko diabetes dari faktor merokok
2	Kurang Aktifitas Fisik	Menyatakan faktor risiko diabetes dari faktor kurang aktifitas fisik
3	Gula Berlebihan	Menyatakan faktor risiko diabetes dari faktor gula berlebihan
4	Garam Berlebihan	Menyatakan faktor risiko diabetes dari faktor garam berlebihan
5	Lemak Berlebihan	Menyatakan faktor risiko diabetes dari faktor lemak berlebihan
6	Kurang Makan Buah Dan Sayur	Menyatakan faktor risiko diabetes dari faktor kurang makan buah dan sayur
7	Konsumsi Alkohol	Menyatakan faktor risiko diabetes dari faktor konsumsi alkohol
8	Pemeriksaan Gula	Menyatakan faktor risiko diabetes dari hasil pemeriksaan gula
9	Kelurahan	Menyatakan kelurahan dari penderita diabetes

b. Data Transformasi dan Normalisasi

Diperlukan transformasi data pada beberapa atribut, yaitu merokok, kurang aktivitas fisik, gula berlebihan, garam berlebihan, lemak berlebihan, kurang makan buah dan sayur, dan konsumsi alkohol dimana proses transformasi dilakukan dengan melihat tingkat prioritas dari masing-masing atribut dan diberi inisial berupa angka jika 1 = ya dan 2 = tidak setelah dilakukannya transformasi data maka langkah selanjutnya yaitu normalisasi dengan menggunakan *scaling*. Adapun proses perhitungan *scaling* untuk normalisasi data menurut (Han, Kamber, & Pei, 2011) adalah sebagai berikut:

$$new\ data = \frac{data}{i}$$

Keterangan:
$new\ data$ = data hasil normalisasi
$data$ = data yang digunakan
i = nilai maksimal dari kriteria

Contoh perhitungan *scaling* untuk normalisasi data pada beberapa atribut, yaitu merokok, kurang aktivitas fisik, gula berlebihan, garam berlebihan, lemak berlebihan, kurang makan buah dan sayur, dan konsumsi alkohol yaitu:

$$new\ data = \frac{1}{2} = 0{,}5$$

Maka tabel data setelah dilakukan proses transformasi dan normalisasi adalah sebagai berikut:

Tabel 4 Transformasi Data Atribut Merokok

Inisial	Merokok	Keterangan
1	0,5	Ya
2	1	Tidak

Tabel 5 Transformasi Data Atribut Kurang Aktivitas Fisik

Inisial	Kurang Aktivitas Fisik	Keterangan
1	0,5	Ya
2	1	Tidak

Tabel 6 Transformasi Data Atribut Gula Berlebihan

Inisial	Gula Berlebihan	Keterangan
1	0,5	Ya
2	1	Tidak

Tabel 7 Transformasi Data Atribut Garam Berlebihan

Inisial	Garam Berlebihan	Keterangan
1	0,5	Ya
2	1	Tidak

Tabel 8 Transformasi Data Atribut Lemak Berlebihan

Inisial	Lemak Berlebihan	Keterangan
1	0,5	Ya

Inisial	Lemak Berlebihan	Keterangan
2	1	Tidak

Tabel 9 Transformasi Data Atribut Kurang Makan Buah dan Sayur

Inisial	Kurang Makan Buah dan Sayur	Keterangan
1	0,5	Ya
2	1	Tidak

Tabel 10 Transformasi Data Atribut Konsumsi Alkohol

Inisial	Konsumsi Alkohol	Keterangan
1	0,5	Ya
2	1	Tidak

Selanjutnya terdapat atribut pemeriksaan gula yang dimana proses normalisasi dilakukan dengan menggunakan z-score. Z-score merupakan nilai standar yang berupa selisih nilai tersebut dengan nilai rata-rata yang dibagi dengan nilai standar baku sebagai pembanding posisi nilai tersebut dengan keseluruhan data. Adapun rumus Z-Score menurut (Hastie, T., Tibshirani, R., & Friedman, 2009) adalah sebagai berikut:

$$Z = \frac{(x - \mu)}{\sigma}$$

Keterangan:
x = nilai yang diamati (skor mentah)
μ = rata-rata populasi
σ = standar deviasi populasi
Z = Z-Score (Nilai Baku)

Untuk mencari nilai μ atau rata-rata dari populasi maka dapat menggunakan rumus menurut (Bluman, 2018) berikut ini:

$$\mu = \frac{\Sigma X}{N}$$

dimana:
μ = rata-rata populasi
ΣX = jumlah semua nilai dalam populasi
N = jumlah keseluruhan elemen pada populasi

Contoh perhitungan μ atau rata-rata dari populasi untuk digunakan dalam menghitung Z-Score sebagai berikut:

$$\mu = \frac{(234 + 376 + 211 + 231 + 200 + \cdots + 221)}{186}$$
$$\mu = 210{,}8656$$

Untuk mencari nilai σ atau standar deviasi populasi maka dapat menggunakan rumus menurut (DeGroot & Schervish, 2012) berikut ini:

$$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$$

e-ISSN: 3109 - 1350

dimana:
σ = standar deviasi populasi
X = nilai individu dalam populasi
μ = rata-rata populasi
N = jumlah keseluruhan elemen pada populasi
Contoh perhitungan σ atau standar deviasi populasi untuk digunakan dalam menghitung Z-Score sebagai berikut:

$$\sigma = \sqrt{\frac{((234 - 210{,}8656)^2 + (376 - 210{,}8656)^2 + (211 - 210{,}8656)^2 + (231 - 210{,}8656)^2 + (200 - 210{,}8656)^2 + \cdots + (221 - 210{,}8656)^2)}{186}}$$

$$\sigma = \sqrt{\frac{404039{,}6398}{186}}$$

$$\sigma = \sqrt{2172{,}256128}$$

$$\sigma = 46{,}607$$

Jika sudah diperoleh nilai μ atau rata-rata dari populasi dan σ atau standar deviasi populasi, maka selanjutnya dapat menghitung nilai Z-Score. Contoh perhitungan Z-Score untuk normalisasi data pada atribut pemeriksaan gula adalah sebagai berikut:

Data ke-1
$$Z = \frac{(x - \mu)}{\sigma} = \frac{(234 - 210{,}866)}{46{,}607} = 0{,}496$$

Data ke-2
$$Z = \frac{(x - \mu)}{\sigma} = \frac{(376 - 210{,}866)}{46{,}607} = 3{,}543$$

Data ke-3
$$Z = \frac{(x - \mu)}{\sigma} = \frac{(211 - 210{,}866)}{46{,}607} = 0{,}003$$

Data ke-4
$$Z = \frac{(x - \mu)}{\sigma} = \frac{(231 - 210{,}866)}{46{,}607} = 0{,}432$$

Data ke-5
$$Z = \frac{(x - \mu)}{\sigma} = \frac{(200 - 210{,}866)}{46{,}607} = -0{,}233$$
…

Data ke-186
$$Z = \frac{(x - \mu)}{\sigma} = \frac{(221 - 210{,}866)}{46{,}607} = 0{,}217$$

Tabel 11 merupakan hasil dari perhitungan Z-Score untuk normalisasi data sebagai berikut:

Tabel 11 Transformasi Data Atribut Pemeriksaan Gula

Data Ke-	Pemeriksaan Gula
1	0,496
2	3,543

Data Ke-	Pemeriksaan Gula
3	0,003
4	0,432
5	-0,233
…	…
186	0,217

Lalu terdapat atribut kelurahan dimana proses transformasi dilakukan berdasarkan dengan frekuensi terbesar yang diberi inisial mulai dari angka 1, 2, 3 dan seterusnya hingga frekuensi data terkecil agar data dapat diolah menggunakan Algoritma K-Means seperti pada tabel 12.

Tabel 12 Transformasi Data Atribut Kelurahan

Inisial	Keterangan	Frekuensi
1	Tanah Sareal	93
2	Kebon Pedes	33
3	Kedung Badak	27
4	Cibadak	8
5	Sukaresmi	6
6	Sukadamai	6
7	Mekarwangi	4
8	Kedung Waringin	4
9	Kedung Jaya	3
10	Kencana	1
11	Kayumanis	1

Tabel 13 Data Penderita Diabetes Puskesmas Tanah Sareal Tahun 2023 Hasil Transformasi

No	**Pasien**	**Merokok**	**Kurang Aktifitas Fisik**	**Gula Berlebihan**	**Garam Berlebihan**	**Lemak Berlebihan**	**Kurang Makan Buah Dan Sayur**	**Konsumsi Alkohol**	**Pemeriksaan Gula**	**Kelurahan**
1	P1	1	1	1	1	1	1	1	0,496	1
2	P2	1	0,5	1	1	1	1	1	3,543	1
3	P3	1	1	1	1	1	1	1	0,003	1
4	P4	1	1	1	1	1	1	1	0,432	1
5	P5	1	1	1	1	1	1	1	-0,233	7
6	P6	1	1	1	1	1	1	1	-0,019	1
7	P7	1	1	1	1	1	1	1	0,840	1
8	P8	1	1	1	1	1	1	1	-0,019	1
9	P9	1	0,5	1	1	1	0,5	1	-0,019	6
10	P10	1	1	1	1	1	1	1	0,217	1

No	Pasien	Merokok	Kurang Aktifitas Fisik	Gula Berlebihan	Garam Berlebihan	Lemak Berlebihan	Kurang Makan Buah Dan Sayur	Konsumsi Alkohol	Pemeriksaan Gula	Kelurahan
…	…	…	…	…	…	…	…	…	…	…
186	P186	1	0,5	1	1	1	0,5	1	0,217	5

c. Proses Algoritma K-Means

Dari data yang tertera pada tabel 13 akan dibuat klaster pemetaan persebaran penyakit diabetes untuk rekomendasi prioritas pemberian penyuluhan berdasarkan variabel yang telah ditentukan. Secara umum Algoritma K-Means bertujuan untuk membagi data menjadi beberapa kelompok dengan tahapan sebagai berikut:

1) Menyiapkan Dataset
   Dataset yang digunakan pada perhitungan ini menggunakan variabel data Merokok, Kurang Aktifitas Fisik, Gula Berlebihan, Garam Berlebihan, Lemak Berlebihan, Kurang Makan Buah dan Sayur, Konsumsi Alkohol, Pemeriksaan Gula dan Kelurahan dari data Penderita Diabetes Puskesmas Tanah Sareal Tahun 2023 yang telah ditransformasi. Dataset tersebut terdiri dari 186 baris. Dataset terdapat pada tabel 13.
2) Menentukan Jumlah Klaster
   Dari dataset penderita diabetes puskesmas tanah sareal tahun 2023 pada tabel 13 akan dibentuk menjadi 3 klaster sehingga nilai K pada perhitungan ini yaitu 3, dengan keterangan klaster 0 merupakan klaster dengan tinggi penderita diabetes, klaster 1 merupakan klaster dengan sedang penderita diabetes, dan klaster 2 merupakan klaster dengan rendah penderita diabetes.
3) Menentukan Titik Centroid
   Dataset penderita diabetes puskesmas tanah sareal tahun 2023 akan dibagi ke dalam 3 klaster, sehingga diperlukan 3 titik awal centroid untuk klaster 0, klaster 1, dan klaster 2. Titik awal centroid yang digunakan yaitu pada tabel 14.

Tabel 14 Nilai Centroid Awal

Klaster 0	Klaster 1	Klaster 2
1	1	1
0,5	0,5	1
0,5	1	1
0,5	1	1
0,5	1	1
0,5	0,5	1
1	1	1
0,196	0,260	-0,255
1	5	10

4) Hitung Jarak Dengan Centroid
   Perhitungan jarak data dengan centroid berfungsi untuk menentukan jarak terpendek pada pengelompokkan klaster, menghitung jarak data dengan centroid dengan menggunakan rumus Euclidean Distance. Euclidean Distance merupakan perhitungan jarak dari 2 buah titik dan untuk mempelajari hubungan antara sudut dan jarak dengan rumus sebagai berikut:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Keterangan:
D_e = Euclidean Distance
i = Banyaknya Objek

e-ISSN: 3109 - 1350

(x, y) = Koordinat Objek
(s, t) = Koordinat Centroid

Iterasi Pertama

Perhitungan Jarak dengan Klaster 0:

Data ke 1:

$$D(X_1, C_0)$$

$$= \sqrt{\begin{matrix}(1-1)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + \\ (1-0{,}5)^2 + (1-0{,}5)^2 + (1-1)^2 + (0{,}496-0{,}196)^2 + \\ (1-1)^2\end{matrix}}$$

$$= 1{,}158$$

Data ke 2:

$$D(X_2, C_0)$$

$$= \sqrt{\begin{matrix}(1-1)^2 + (0{,}5-0{,}5)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + \\ (1-0{,}5)^2 + (1-0{,}5)^2 + (1-1)^2 + (3{,}543-0{,}196)^2 + \\ (1-1)^2\end{matrix}}$$

$$= 3{,}493$$

Data ke 3:

$$D(X_3, C_0)$$

$$= \sqrt{\begin{matrix}(1-1)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + \\ (1-0{,}5)^2 + (1-0{,}5)^2 + (1-1)^2 + (0{,}003-0{,}196)^2 + \\ (1-1)^2\end{matrix}}$$

$$= 1{,}135$$

Data ke 4:

$$D(X_4, C_0)$$

$$= \sqrt{\begin{matrix}(1-1)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + \\ (1-0{,}5)^2 + (1-0{,}5)^2 + (1-1)^2 + (0{,}432-0{,}196)^2 + \\ (1-1)^2\end{matrix}}$$

$$= 1{,}143$$

Data ke 5:

$$D(X_5, C_0)$$

$$= \sqrt{\begin{matrix}(1-1)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + \\ (1-0{,}5)^2 + (1-0{,}5)^2 + (1-1)^2 + ((-0{,}233)-0{,}196)^2 + \\ (7-1)^2\end{matrix}}$$

$$= 6{,}118$$

…..

Data ke 186:

$$D(X_{186}, C_0)$$

$$= \sqrt{\begin{matrix}(1-1)^2 + (0{,}5-0{,}5)^2 + (1-0{,}5)^2 + (1-0{,}5)^2 + \\ (1-0{,}5)^2 + (0{,}5-0{,}5)^2 + (1-1)^2 + (0{,}217-0{,}196)^2 + \\ (5-1)^2\end{matrix}}$$

$$= 4{,}093$$

Tabel 15 Hasil Pengelompokkan Iterasi Pertama

No	Pasien	C0	C1	C2	Jarak Terdekat	Cluster
1	P1	1,158	4,069	9,031	1,158	0
2	P2	3,493	5,199	9,781	3,493	0
3	P3	1,135	4,070	9,004	1,135	0
4	P4	1,143	4,066	9,026	1,143	0
5	P5	6,118	2,178	3,000	2,178	1
…	…	…	…	..	…	..
186	P186	4,093	0,043	5,072	0,043	1

Tabel 15 merupakan hasil perhitungan dengan rumus Euclidean Distance antara tiap variabel dengan centroid Cluster 0, Cluster 1, dan Cluster 2 sehingga didapat jarak terdekat dari masing-masing atribut, kemudian untuk ditentukan masuk ke cluster masing-masing. Setelah proses pengelompokkan data lalu dilakukan penentuan titik centroid baru. Nilai centroid baru didapatkan dengan menggunakan rumus:

$$\mu_k = \left(\frac{1}{n_k}\right) \sum_{i=1}^{nk} X_{ik}$$

Keterangan:
μ_k = Titik centroid dari klaster ke-k
n_k = Banyaknya data pada klaster ke-k
X_{ik} =Data ke-I klaster ke-k

Mencari nilai centroid cluster 0 baru:

$$\mu_0 = \frac{122{,}5}{126} = 0{,}972$$
$$\mu_1 = \frac{116{,}5}{126} = 0{,}925$$
$$\mu_2 = \frac{120{,}5}{126} = 0{,}956$$
$$\mu_3 = \frac{120{,}5}{126} = 0{,}956$$
$$\mu_4 = \frac{120{,}5}{126} = 0{,}956$$
$$\mu_5 = \frac{120{,}5}{126} = 0{,}956$$
$$\mu_6 = \frac{125{,}5}{126} = 0{,}996$$
$$\mu_7 = \frac{22{,}913}{126} = 0{,}182$$
$$\mu_8 = \frac{159}{126} = 1{,}262$$

Mencari nilai centroid cluster 1 baru:

$$\mu_0 = \frac{51}{51} = 1$$
$$\mu_1 = \frac{46}{51} = 0{,}902$$
$$\mu_2 = \frac{50{,}5}{51} = 0{,}990$$
$$\mu_3 = \frac{50{,}5}{51} = 0{,}990$$
$$\mu_4 = \frac{50{,}5}{51} = 0{,}990$$
$$\mu_5 = \frac{46}{51} = 0{,}902$$
$$\mu_6 = \frac{51}{51} = 1$$
$$\mu_7 = \frac{-16{,}653}{51} = -0{,}327$$
$$\mu_8 = \frac{207}{51} = 4{,}059$$

Mencari nilai centroid cluster 2 baru:

$$\mu_0 = \frac{9}{9} = 1$$
$$\mu_1 = \frac{9}{9} = 1$$
$$\mu_2 = \frac{9}{9} = 1$$
$$\mu_3 = \frac{9}{9} = 1$$
$$\mu_4 = \frac{9}{9} = 1$$
$$\mu_5 = \frac{9}{9} = 1$$
$$\mu_6 = \frac{9}{9} = 1$$
$$\mu_7 = \frac{-6{,}261}{9} = -0{,}696$$
$$\mu_8 = \frac{80}{9} = 8{,}889$$

Lalu didapatkan centroid baru untuk Cluster 0, Cluster 1 dan Cluster 2 pada tabel 16.

Tabel 16 Centroid Hasil Iterasi Pertama

No	Cluster 0	Cluster 1	Cluster 2
1	0,972	1	1
2	0,925	0,902	1
3	0,956	0,990	1
4	0,956	0,990	1
5	0,956	0,990	1
6	0,956	0,902	1
7	0,996	1	1
8	0,182	-0,327	-0,696
9	1,262	4,059	8,889

Ulangi perhitungan jarak data dengan centroid hingga nilai titik centroid tidak berubah. Iterasi kedua dapat dilihat pada perhitungan berikut. Dari perhitungan yang dilakukan pada penelitian ini terjadi sebanyak 3 kali, titik pusat cluster tidak ada lagi perubahan dan data pertama hingga data yang terakhir tidak terdapat lagi yang bergeser dari satu cluster ke cluster yang lainnya. Dari perhitungan yang sudah dilakukan maka dapat disimpulkan bahwa hasil *clustering* dari dataset perhitungan K-Means menghasilkan pembagian objek ke masing-masing klaster, yaitu:

i) Klaster 0
Tabel 17 merupakan anggota klaster 0 dengan penderita penyakit diabetes tinggi terdiri dari 126 data penderita penyakit diabetes di puskesmas Tanah Sareal tahun 2023.

Tabel 17 Hasil Klaster 0

No	Pasien	Merokok	Kurang Aktifitas Fisik	Gula Berlebihan	Garam Berlebihan		Lemak Berlebihan	Kurang Makan Buah Dan Sayur	Konsumsi Alkohol	Pemeriksaan Gula	Kelurahan
1	P1	T	T	T	T		T	T	T	234	Tanah Sareal
2	P2	T	Y	T	T		T	T	T	376	Tanah Sareal
3	P3	T	T	T	T		T	T	T	211	Tanah Sareal
4	P4	T	T	T	T		T	T	T	231	Tanah Sareal
5	P6	T	T	T	T		T	T	T	210	Tanah Sareal
6	P7	T	T	T	T		T	T	T	250	Tanah Sareal
7	P8	T	T	T	T		T	T	T	210	Tanah Sareal
8	P10	T	T	T	T		T	T	T	221	Tanah Sareal
9	P11	T	T	T	T		T	T	T	102	Kebon Pedes
10	P12	T	T	T	T		T	T	T	206	Tanah Sareal
…	…	…	…	…	…		…	…	…	…	…
126	P185	T	T	T	T		T		T	230	Tanah Sareal

ii) Klaster 1

Tabel 18 merupakan anggota klaster 1 dengan penderita penyakit diabetes sedang terdiri dari 47 data penderita penyakit diabetes di puskesmas Tanah Sareal tahun 2023.

Tabel 18 Hasil Klaster 1

No	**Pasien**	**Merokok**	**Kurang Aktifitas Fisik**	**Gula Berlebihan**	**Garam Berlebihan**	**Lemak Berlebihan**	**Kurang Makan Buah Dan Sayur**	**Konsumsi Alkohol**		**Pemeriksaan Gula**	**Kelurahan**
1	P9	T	Y	T	T	T	Y	T		210	Sukadamai
2	P14	T	T	T	T	T	T	T		220	Cibadak
3	P17	T	T	T	T	T	T	T		243	Kedung Badak
4	P22	T	T	T	T	T	T	T		200	Cibadak
5	P24	T	T	T	T	T	T	T		200	Kedung Badak
6	P28	T	T	T	T	T	T	T		211	Kedung Badak
7	P34	T	T	T	T	T	T	T		267	Kedung Badak
8	P36	T	T	T	T	T	T	T		234	Sukadamai
9	P38	T	Y	T	T	T	Y	T		102	Sukaresmi
10	P46	T	T	T	T	T	T	T		178	Cibadak
…	…	…	…	…	…	…	…	…		…	…
47	P186	T	Y	T	T	T	Y	T		221	Sukaresmi

iii) Klaster 2

Tabel 19 merupakan anggota klaster 2 dengan penderita penyakit diabetes rendah terdiri dari 13 data penderita penyakit diabetes di puskesmas Tanah Sareal tahun 2023.

Tabel 19 Hasil Klaster 2

No	Pasien	Merokok	Kurang Aktifitas Fisik	Gula Berlebihan	Garam Berlebihan	Lemak Berlebihan	Kurang Makan Buah Dan Sayur	Konsumsi Alkohol	Pemeriksaan Gula	Kelurahan
1	P5	T	T	T	T	T	T	T	200	Mekarwangi
2	P18	T	T	T	T	T	T	T	165	Kedung Waringin
3	P47	T	T	T	T	T	T	T	199	Kencana
4	P66	T	T	T	T	T	T	T	156	Kedung Jaya
5	P67	T	T	T	T	T	T	T	102	Kedung Waringin
6	P72	T	T	T	T	T	T	T	202	Kedung Waringin
7	P99	T	Y	Y	Y	Y	Y	T	208	Mekarwangi
8	P100	T	T	T	T	T	T	T	234	Kayumanis
9	P113	T	T	T	T	T	T	T	210	Mekarwangi
10	P143	T	T	T	T	T	T	T	235	Kedung Jaya
…	…	…	…	…	…	…	…	…	…	…
13	P178	T	T	T	T	T	T	T	220	Mekarwangi

2. PEMBAHASAN

Berdasarkan hasil pengembangan terkait dengan metode yang diterapkan pada sistem informasi. Hasil diuji dengan menggunakan metode *Silhouette Coefficient*. Metode ini digunakan untuk menguji kualitas klaster, dengan hasil akhir perhitungan dari K-Means sebagai hasil klaster. Adapun proses perhitungan *Silhouette Coefficient* dengan langkah langkah seperti berikut :

a. Menghitung rata-rata jarak dari suatu data, misal data ke-1 dengan semua data lain yang berada dalam satu klaster:
Klaster 1:

$$a(i)_1$$

$$= \sqrt{\begin{array}{c}((1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(1-1)^2+(0{,}496-0{,}496)^2+(1-1)^2)+((1-1)^2+\\(1-0{,}5)^2+(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(1-1)^2+(0{,}496-3{,}543)^2+(1-1)^2)+((1-1)^2+\\(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(1-1)^2+(0{,}496-0{,}003)^2+(1-1)^2)+\cdots+((1-1)^2+\\(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(1-1)^2+(0{,}496-0{,}041)^2+(1-1)^2)\end{array}}$$

$$= \frac{115{,}336}{126} = 0{,}915$$

Klaster 2:

$$a(i)_2$$

$$= \sqrt{\begin{array}{c}((1-1)^2+(0{,}5-0{,}5)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(0{,}5-0{,}5)^2+(1-1)^2+((-0{,}019)-(-0{,}019))^2+(6-6)^2)+\\((1-1)^2+(0{,}5-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(0{,}5-1)^2+(1-1)^2+((-0{,}019)-0{,}196)^2+(6-4)^2)+\\((1-1)^2+(0{,}5-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(0{,}5-1)^2+(1-1)^2+(-0{,}019)-0{,}689)^2+(6-3)^2)+\cdots+\\((1-1)^2+(0{,}5-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(0{,}5-1)^2+(1-1)^2+((-0{,}019)-0{,}217)^2+(6-5)^2)\end{array}}$$

$$= \frac{120{,}603}{47} = 2{,}566$$

Klaster 3:

$$a(i)_3$$

$$= \sqrt{\begin{array}{c}((1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(1-1)^2+((-0{,}233)-(-0{,}233))^2+(7-7)^2)+((1-1)^2+\\(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(1-1)^2+((-0{,}233)-(-0{,}984))^2+(7-8)^2)+((1-1)^2+\\(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(1-1)^2+((-0{,}233)-(-0{,}255))^2+(7-10)^2)+\cdots+\\((1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+(1-1)^2+\\(1-1)^2+(1-1)^2+((-0{,}233)-0{,}196)^2+(7-7)^2)\end{array}}$$

$$= \frac{21{,}698}{13} = 1{,}669$$

Tabel 20 Rata-Rata Jarak Dalam Satu Klaster

Data	a(i)
Ke-1	0,915

Data	a(i)
Ke-2	3,453
Ke-3	0,860
Ke-4	0,894
Ke-5	0,867
…	…
Ke-186	1,759

Tabel 20 merupakan hasil perhitungan jarak rata-rata dari setiap data yang terletak dalam satu klister.

b. Menghitung rata-rata jarak, misal data ke-1 dengan semua data pada klaster lain dan diambil nilai terkecilnya:

Klaster 1 ke Klaster 2

$$d(1,2)$$

$$= \sqrt{\begin{matrix} ((1-1)^2 + (0{,}5-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (0{,}5-1)^2 + (1-1)^2 + ((-0{,}019) - 0{,}496)^2 + \\ (6-1)^2) + ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + (1-1)^2 + (0{,}196 - 0{,}496)^2 \\ +(4-1)^2) + ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + (0{,}689 - 0{,}496)^2 + (3-1)^2) + \cdots + \\ ((1-1)^2 + (0{,}5-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (0{,}5-1)^2 + (1-1)^2 + (0{,}217 - 0{,}496)^2 + (5-1)^2) \end{matrix}}$$

$$= \frac{145{,}566}{47} = 3{,}097$$

Klaster 1 ke Klaster 3

$$d(1,3)$$

$$= \sqrt{\begin{matrix} ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + ((-0{,}233) - 0{,}496)^2 + (7-1)^2) + ((1-1)^2 + \\ (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + ((-0{,}984) - 0{,}496)^2 + (8-1)^2) + ((1-1)^2 + \\ (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + ((-0{,}255) - 0{,}496)^2 + (10-1)^2) + \cdots + \\ ((1-1)^2 + (0{,}5-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 \\ (0{,}5-1)^2 + (1-1)^2 + (0{,}196 - 0{,}496)^2 + (7-1)^2) \end{matrix}}$$

$$= \frac{96{,}651}{13} = 7{,}435$$

Klaster 2 ke Klaster 1

$$d(2,1)$$

$$= \sqrt{\begin{matrix} ((1-1)^2 + (1-0{,}5)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-0{,}5)^2 + (1-1)^2 + (0{,}496 - (-0{,}019))^2 + (1-6)^2) + \\ ((1-1)^2 + (0{,}5-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + (3{,}543 - (-0{,}019))^2 + (1-6)^2) + \\ ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + (0{,}003 - (-0{,}019))^2 + (1-6)^2) + \cdots + \\ ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + (0{,}411 - (-0{,}019))^2 + (1-6)^2) \end{matrix}}$$

e-ISSN: 3109 - 1350

$$= \frac{616{,}492}{126} = 4{,}893$$

Klaster 2 ke Klaster 3

$$d(2,3)$$

$$= \sqrt{\begin{matrix} ((1-1)^2 + (1-0{,}5)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-0{,}5)^2 + (1-1)^2 + \left((-0{,}233) - (-0{,}233)\right)^2 + (7-6)^2) + \\ ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + \left((-0{,}984) - (-0{,}019)\right)^2 + (8-6)^2) + \\ ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + \left((-0{,}255) - (-0{,}019)\right)^2 + (10-6)^2) + \\ \ldots + ((1-1)^2 + (0{,}5-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (0{,}5-1)^2 + (1-1)^2 + \left(0{,}196 - (-0{,}019)\right)^2 + (7-6)^2) \end{matrix}}$$

$$= \frac{34{,}179}{13} = 2{,}629$$

Klaster 3 ke Klaster 1

$$d(3,1)$$

$$= \sqrt{\begin{matrix} ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-0{,}5)^2 + (1-1)^2 + \left(0{,}496 - (-0{,}233)\right)^2 + (1-7)^2) + \\ ((1-1)^2 + (0{,}5-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + \left(3{,}543 - (-0{,}233)\right)^2 + (1-7)^2) + \\ ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + \left(0{,}003 - (-0{,}233)\right)^2 + (1-7)^2) + \cdots + \\ ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + \left(0{,}411 - (-0{,}233)\right)^2 + (1-7)^2) \end{matrix}}$$

$$= \frac{736{,}399}{126} = 5{,}844$$

Klaster 3 ke Klaster 2

$$d(3,2)$$

$$= \sqrt{\begin{matrix} ((1-1)^2 + (0{,}5-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (0{,}5-0{,}5)^2 + (1-1)^2 + \left((-0{,}019) - (-0{,}233)\right)^2 + (6-7)^2) + \\ ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + \left(0{,}196 - (-0{,}233)\right)^2 + (4-7)^2) + \\ ((1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (1-1)^2 + (1-1)^2 + \left(0{,}689 - (-0{,}233)\right)^2 + (3-7)^2) + \cdots + \\ ((1-1)^2 + (0{,}5-1)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + \\ (0{,}5-1)^2 + (1-1)^2 + \left(0{,}217 - (-0{,}233)\right)^2 + (5-7)^2) \end{matrix}}$$

$$= \frac{157{,}977}{47} = 3{,}361$$

Tabel 21 Rata-Rata Jarak Pada Klaster Lain

Data	d(I, c)		
	d(I,1)	d(I,2)	d(I,3)
Ke-1	-	3,097	7,435
Ke-2	-	4,937	8,418
Ke-3	-	3,004	7,386
Ke-4	-	3,080	7,426
Ke-5	-	3,002	7,385
…	…	…	…
Ke-186	5,832	3,391	-

Tabel 21 merupakan hasil perhitungan jarak rata-rata dari data i dengan semua data pada klaster lain. Jika jarak rata-rata dari data i dengan semua data pada klaster lain sudah dilakukan, maka diambilah nilai terkecilnya. Maka diperoleh nilai terkecil pada jarak rata-rata dari data i dengan semua data pada klaster lain dapat dilihat pada tabel 22.

Tabel 22 Nilai Terkecil Pada Jarak Rata-Rata Klaster Lain

Data	b(i)
Ke-1	3,097
Ke-2	4,937
Ke-3	3,004
Ke-4	3,080
Ke-5	3,002
…	…
Ke-186	3,391

Setelah mengetahui nilai α(i) dan d(ij), hitung nilai *Sillhouette Coefficient* sebagai berikut:

Data Ke-1

$$S(i) = 1 - \frac{(3{,}097 - 0{,}915)}{\max(0{,}915, 3{,}097)} = 0{,}704$$

Data Ke-2

$$S(i) = 1 - \frac{(4{,}937 - 3{,}453)}{\max(3{,}453, 4{,}937)} = 0{,}300$$

Data Ke-3

$$S(i) = 1 - \frac{(3{,}004 - 0{,}860)}{\max(0{,}860, 3{,}004)} = 0{,}714$$

Data Ke-4

$$S(i) = 1 - \frac{(3{,}080 - 0{,}894)}{\max(0{,}894, 3{,}080)} = 0{,}710$$

Data Ke-5

$$S(i) = 1 - \frac{(0{,}867 - 3{,}002)}{\max(3{,}002, 0{,}867)} = 0{,}711$$

…

Data Ke-186

$$S(i) = 1 - \frac{(3{,}391 - 1{,}759)}{\max(1{,}759, 3{,}391)} = 0{,}481$$

Tabel 23 Silhouette Coefficient Semua Data

Data	S(i)
Ke-1	0,704
Ke-2	0,300
Ke-3	0,714
Ke-4	0,710
Ke-5	0,711
…	…
Ke-186	0,481

Tabel 23 merupakan hasil perhitungan dari *Silhouette Coefficient* semua data. Maka, diperoleh untuk *Silhouette Coefficient* secara keseluruhan yaitu dengan menghitung rata-rata nilai *Silhouette Coefficient* pada semua data sebagai berikut:

$$S(i) = \frac{0{,}704 + 0{,}300 + 0{,}714 + 0{,}710 + 0{,}711 + 0{,}662 + \cdots + 0{,}481}{186}$$

$$S(i) = 0{,}503$$

Semakin *Silhouette Score* mendekati nilai 1 artinya semakin kuat klaster. Sebaliknya jika nilai *silhouette* score semakin mendekati 0, maka semakin kurang baik pengelompokkan data kedalam suatu klaster. Maka perhitungan uji hasil dengan menerapkan *Silhouette Coefficient* dengan tiga klaster diperoleh nilai rata-rata 0,503 yang berarti masuk kedalam kategori *weak structure*.

C. KESIMPULAN

Berdasarkan dari hasil penelitian yang dilakukan, kesimpulan yang bisa diuraikan antara lain:

1. Menerapkan metode K-Means dapat memberikan persebaran penderita diabetes dengan optimal karena telah dilakukan uji akurasi dengan menggunakan *Silhouette Coefficient.*
2. Menerapkan metode K-Means dalam pemetaan persebaran penyakit diabetes di masa yang akan datang menjadi lebih efektif dari proses yang dilakukan sebelumnya.
3. Hasil pengembangan prototype untuk pemetaan persebaran penyakit diabetes adalah menampilkan hasil klaste dan ploting data.
4. Hasil uji akurasi dengan *Silhouette Coefficient* sebesar 0,503 yang berarti termasuk kedalam kategori *weak structure*, kemudian hasil kuesioner kepada pengguna sebesar 73,2% serta hasil kuesioner kepada ahli dengan pengujian blackbox sebesar 100% dan pengujian whitebox sebesar 100%.

D. DAFTAR PUSTAKA

[1] Bilal, Abdilah, dan Harbani, Arif. (2025). *Penerapan Metode Test Driven Development untuk Menguji Rest Api pada Script di Visual Basic*. Jurnal Ilmiah Saintekom, Volume 01 Nomor 01, Juni 2025; 84 – 94.

[2] Bluman, A. G. (2018). *Elementary Statistics: A Step by Step Approach* (10th ed.). McGraw-Hill Education.

[3] DeGroot, M. H., & Schervish, M. J. (2012). *Probability and Statistics* (4th ed.). Addison-Wesley.

[4] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd Editio.). Morgan Kaufmann.

[5] Handoyo, R., M, R. R., & Michrandi, S. N. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means Pada Pengelompokan Dokumen. *JSM STMIK Mikroskil*, *15*.

[6] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). Springer.

[7] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). Springer.

[8] Prasetyo, E. (2014). *Data Mining - Mengolah Data menjadi Informasi Menggunakan Matlab*. (A. Sahala, Ed.). Yogyakarta: ANDI OFFSET.

[9] Rousseeuw, P. J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 53–65.

[10] Santosa, B. (2007). *Data Mining: Teknik Pemanfataan Data Untuk Keperluan Bisnis* (Pertama.). Yogyakarta: Graha Ilmu.