p-ISSN: 2963-7856| e-ISSN: 2961-9890
Available online at jerkin.org/index.php/jerkin
Jurnal Pengabdian Masyarakat dan Riset Pendidikan
Volume 5 No. 1, Juli - September 2026, pp 915-928

Clustering Tingkat Risiko Obesitas Menggunakan Algoritma K-Means Dengan Optimasi Metode Elbow

Ahmad Dani[1]*, Muhammad Khairul Fahri[2], Harry Dwiki Pramadan[3], Fadli Anwar Sinaga[4], Aji Febrian Sani[5], Muhammad Fauzan[6], Bambang Irwansyah[7]

[1,2,3,4,5,6,7] Fakultas Teknik, Prodi Teknik Informatika, Universitas Asahan, Jl. Jend. A. Yani, Kisaran Naga, Kec. Kota Kisaran Timur, Kisaran, Sumatera Utara.
E-mail: adani8175@gmail.com
*Corresponding Author

https://doi.org/10.31004/jerkin.v5i1.7189

ARTICLE INFO

Article history:
Received: 29 Jun 2026
Revised: 05 Jul 2026
Accepted: 11 Jul 2026

Kata Kunci:
Data Mining, Clustering, K-Means, Metode Elbow, Obesitas.

Keywords:
Data Mining, Clustering, K-Means, Elbow Method, Obesity.

ABSTRACT

Obesitas merupakan salah satu permasalahan kesehatan yang terus meningkat dan berpotensi menyebabkan berbagai penyakit kronis. Pengelompokan tingkat risiko obesitas diperlukan untuk membantu mengidentifikasi karakteristik individu berdasarkan kondisi kesehatan dan gaya hidup. Penelitian ini bertujuan menerapkan algoritma K-Means dengan optimasi metode Elbow untuk mengelompokkan data obesitas. Metode Elbow digunakan untuk menentukan jumlah cluster yang optimal sebelum proses clustering dilakukan menggunakan algoritma K-Means. Dataset yang digunakan terdiri dari 200 data dengan atribut tinggi badan, berat badan, indeks massa tubuh (IMT), usia, dan jumlah konsumsi makanan utama per hari (NCP). Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah tiga cluster. Proses clustering menggunakan algoritma K-Means menghasilkan tiga kelompok data, yaitu Cluster 1 sebanyak 137 data (68,50%), Cluster 2 sebanyak 45 data (22,50%), dan Cluster 3 sebanyak 18 data (9,00%). Hasil pengelompokan menunjukkan bahwa algoritma K-Means dengan optimasi metode Elbow mampu mengelompokkan data berdasarkan kemiripan karakteristik sehingga dapat dimanfaatkan sebagai informasi pendukung dalam identifikasi tingkat risiko obesitas dan pengambilan keputusan di bidang kesehatan.

Obesity is a growing health issue with the potential to cause various chronic diseases. Categorizing obesity risk levels is essential for identifying individual characteristics based on health conditions and lifestyle. This study aims to apply the K-Means algorithm, optimized using the Elbow method, to group obesity-related data. The Elbow method is employed to determine the optimal number of clusters prior to performing the K-Means clustering process. The dataset comprises 200 records, featuring attributes such as height, weight, Body Mass Index (BMI), age, and the number of main meals consumed per day (NCP). The results indicate that the optimal number of clusters is three. The K-Means clustering process yielded three groups: Cluster 1 with 137 records (68.50%), Cluster 2 with 45 records (22.50%), and Cluster 3 with 18 records (9.00%). The findings demonstrate that the K-Means algorithm, combined with Elbow method optimization, effectively groups data based on characteristic similarities; thus, it can serve as a valuable tool for identifying obesity risk levels and supporting decision-making in the healthcare sector.

How to Cite: Ahmad Dani, et al. (2026), Clustering Tingkat Risiko Obesitas Menggunakan Algoritma K-Means Dengan Optimasi Metode Elbow, 5(1). https://doi.org/10.31004/jerkin.v5i1.7189

PENDAHULUAN

Obesitas merupakan salah satu masalah kesehatan yang menjadi perhatian dunia karena jumlah penderitanya terus mengalami peningkatan setiap tahun. Kondisi ini terjadi akibat penumpukan lemak tubuh yang berlebihan dan dapat meningkatkan risiko berbagai penyakit seperti diabetes, hipertensi, penyakit jantung, serta gangguan metabolisme lainnya. Menurut World Health Organization (2024),

obesitas telah menjadi salah satu tantangan kesehatan masyarakat yang memerlukan perhatian serius karena berdampak terhadap kualitas hidup dan tingkat kematian masyarakat. Dengan semakin banyaknya data kesehatan yang tersedia, diperlukan suatu metode yang mampu mengolah dan menganalisis data tersebut sehingga dapat memberikan informasi yang bermanfaat dalam mengidentifikasi tingkat risiko obesitas pada individu (Hasby et al., 2025).

Perkembangan teknologi informasi telah mendorong pemanfaatan data mining dalam berbagai bidang, termasuk kesehatan. Data mining digunakan untuk menemukan pola atau pengetahuan baru dari sekumpulan data yang besar sehingga dapat membantu proses pengambilan keputusan. Salah satu teknik yang sering digunakan dalam data mining adalah clustering, yaitu proses pengelompokan data berdasarkan tingkat kemiripan karakteristik tertentu. Teknik clustering memungkinkan data yang memiliki karakteristik serupa berada dalam kelompok yang sama dan terpisah dari kelompok lain yang memiliki karakteristik berbeda (Wijaya, 2020).

Algoritma K-Means merupakan salah satu metode clustering yang populer karena memiliki proses perhitungan yang sederhana dan mampu mengelompokkan data dalam jumlah besar secara efisien. Metode ini bekerja dengan menentukan pusat cluster (centroid) dan mengelompokkan data berdasarkan jarak terdekat terhadap centroid tersebut. Penelitian yang dilakukan oleh Setiawati et al. (2024) menunjukkan bahwa algoritma K-Means mampu menghasilkan pengelompokan data yang baik sehingga dapat digunakan untuk menganalisis pola dan karakteristik data kesehatan secara lebih efektif.

Meskipun demikian, salah satu tantangan dalam penggunaan algoritma K-Means adalah menentukan jumlah cluster yang optimal. Pemilihan jumlah cluster yang kurang tepat dapat menyebabkan hasil pengelompokan menjadi kurang representatif. Untuk mengatasi permasalahan tersebut, metode Elbow dapat digunakan sebagai pendekatan dalam menentukan nilai cluster terbaik. Metode ini dilakukan dengan membandingkan nilai Sum of Squared Error (SSE) pada beberapa jumlah cluster sehingga diperoleh titik perubahan yang menunjukkan jumlah cluster optimal (Wijaya, 2020). Penelitian yang dilakukan oleh Ashari et al. (2023) juga menunjukkan bahwa metode Elbow dapat digunakan sebagai salah satu teknik evaluasi yang efektif dalam menentukan jumlah cluster pada algoritma K-Means.

Beberapa penelitian sebelumnya telah menerapkan algoritma K-Means dalam pengelompokan data obesitas. Hasby et al. (2025) berhasil mengelompokkan data risiko obesitas menggunakan metode K-Means sehingga diperoleh beberapa kelompok dengan karakteristik yang berbeda. Selain itu, Santosa et al. (2024) mengombinasikan algoritma K-Means dengan metode Elbow untuk mengategorikan tingkat risiko obesitas dan memperoleh jumlah cluster yang optimal. Sementara itu, Setiawati et al. (2024) membandingkan beberapa algoritma clustering pada dataset obesitas dan menunjukkan bahwa K-Means memiliki performa yang baik dalam proses pengelompokan data.

Berdasarkan penelitian-penelitian tersebut, dapat diketahui bahwa algoritma K-Means memiliki kemampuan yang baik dalam mengelompokkan data obesitas. Namun, masih diperlukan penelitian yang berfokus pada pengelompokan tingkat risiko obesitas dengan memanfaatkan metode Elbow sebagai optimasi dalam menentukan jumlah cluster yang optimal. Oleh karena itu, penelitian ini menerapkan algoritma K-Means dengan optimasi metode Elbow pada dataset obesitas untuk menghasilkan pengelompokan tingkat risiko obesitas yang lebih efektif. Hasil penelitian diharapkan dapat memberikan informasi mengenai karakteristik setiap kelompok risiko obesitas serta menjadi pendukung dalam proses analisis dan pengambilan keputusan di bidang kesehatan.

METODE

Metode penelitian merupakan tahapan yang digunakan untuk mencapai tujuan penelitian secara sistematis. Pada penelitian ini, metode yang digunakan adalah clustering menggunakan algoritma K-Means dengan optimasi metode Elbow untuk menentukan jumlah cluster yang optimal. Dataset yang digunakan berupa data obesitas yang terdiri dari beberapa atribut yang berkaitan dengan kondisi fisik individu. Proses penelitian dilakukan melalui beberapa tahapan, yaitu pengumpulan data, preprocessing data, penentuan jumlah cluster menggunakan metode Elbow, proses clustering menggunakan algoritma K-Means, serta analisis hasil pengelompokan.

Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang meliputi pengumpulan data, preprocessing data, penentuan jumlah cluster menggunakan metode Elbow, proses clustering

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

menggunakan algoritma K-Means, serta analisis hasil clustering. Tahapan penelitian dilakukan secara sistematis untuk menghasilkan pengelompokan tingkat risiko obesitas yang optimal berdasarkan karakteristik data yang digunakan.

Tahap pertama adalah pengumpulan dataset obesitas yang diperoleh dari sumber dataset publik. Dataset tersebut berisi sejumlah atribut yang berkaitan dengan kondisi fisik dan karakteristik individu, seperti tinggi badan, berat badan, indeks massa tubuh (IMT), usia, dan jumlah konsumsi makanan utama per hari.

Tahap kedua adalah normalisasi data. Pada tahap ini dilakukan pemeriksaan data untuk menyamakan rentang nilai setiap atribut sehingga tidak terdapat atribut yang mendominasi proses perhitungan jarak.

Tahap ketiga adalah penentuan jumlah cluster menggunakan metode Elbow. Metode ini dilakukan dengan menghitung nilai Sum of Squared Error (SSE) pada beberapa nilai K. Hasil perhitungan kemudian divisualisasikan dalam bentuk grafik Elbow untuk menentukan jumlah cluster yang paling optimal.

Tahap keempat adalah proses clustering menggunakan algoritma K-Means. Data dikelompokkan berdasarkan jarak terdekat terhadap centroid menggunakan perhitungan Euclidean Distance. Proses iterasi dilakukan hingga tidak terjadi perubahan anggota cluster atau centroid telah mencapai kondisi stabil.

Tahap terakhir adalah analisis hasil clustering. Hasil pengelompokan dianalisis berdasarkan jumlah anggota setiap cluster dan karakteristik centroid yang terbentuk untuk mengidentifikasi tingkat risiko obesitas pada masing-masing kelompok.

Algoritma K-Means

Pada penelitian ini, algoritma K-Means digunakan untuk mengelompokkan tingkat risiko obesitas berdasarkan atribut usia, tinggi badan, berat badan, indeks massa tubuh, dan jumlah konsumsi makanan utama per hari. Pengukuran jarak antara data dan centroid dilakukan menggunakan metode Euclidean Distance.

Persamaan Euclidean Distance ditunjukkan pada Persamaan (1).

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (xi - yi)^2}$$

Keterangan:
$d(x,y)$ = jarak antara data dan centroid
xi = nilai data ke-i
yi = nilai centroid ke-i
n = jumlah atribut

Adapun tahapan algoritma K-Means adalah sebagai berikut:

1. Menentukan jumlah cluster (K).
2. Memilih centroid awal secara acak.
3. Menghitung jarak setiap data terhadap centroid menggunakan Euclidean Distance.
4. Mengelompokkan data ke cluster dengan jarak terdekat.
5. Menghitung centroid baru berdasarkan rata-rata anggota cluster.
6. Mengulangi proses hingga centroid tidak mengalami perubahan atau telah mencapai kondisi konvergen.

Metode Elbow

Metode Elbow merupakan salah satu metode yang digunakan untuk menentukan jumlah cluster optimal pada proses clustering. Metode ini bekerja dengan menghitung nilai Sum of Squared Error (SSE) atau Within Cluster Sum of Squares (WCSS) dari beberapa nilai cluster (K). Nilai SSE menunjukkan tingkat kedekatan data terhadap centroid pada setiap cluster. Semakin kecil nilai SSE, maka semakin baik kualitas pengelompokan data yang dihasilkan.

Penentuan jumlah cluster optimal dilakukan dengan mengamati grafik hubungan antara jumlah cluster (K) dan nilai SSE. Titik yang membentuk pola menyerupai siku (*elbow*) menunjukkan jumlah cluster yang paling optimal untuk digunakan pada proses clustering. Penggunaan metode Elbow dapat

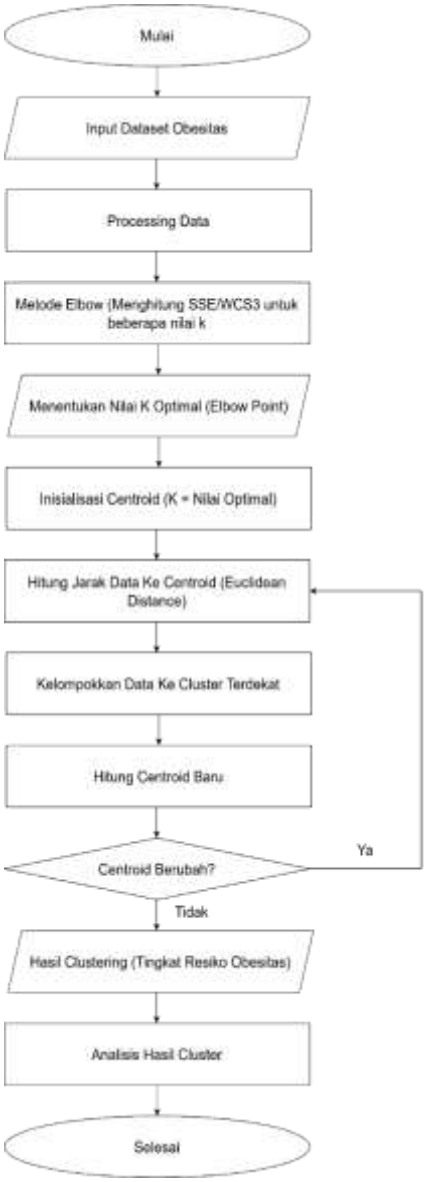
membantu mengurangi subjektivitas dalam menentukan jumlah cluster sehingga hasil pengelompokan menjadi lebih akurat (Ashari et al., 2023).
Perhitungan nilai SSE dapat dilakukan menggunakan Persamaan (2).

$$SSE = \sum_{i=1}^{k} \sum_{x \epsilon Ci} (x - ci)^2$$

Keterangan:
SSE = Sum of Squared Error
k = jumlah cluster
x = data pada cluster
Ci = cluster ke-i
ci = centroid cluster ke-i

Pada penelitian ini, metode Elbow digunakan untuk menentukan jumlah cluster optimal sebelum proses clustering menggunakan algoritma K-Means dilakukan. Nilai cluster yang menghasilkan titik siku (*elbow point*) dipilih sebagai jumlah cluster terbaik untuk proses pengelompokan tingkat risiko obesitas.

Flowchart Penelitian



Gambar 1. Flowchart Penelitian

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

Berdasarkan Gambar 1, penelitian dimulai dengan pengumpulan dataset obesitas yang kemudian melalui tahap preprocessing. Selanjutnya dilakukan optimasi jumlah cluster menggunakan metode Elbow. Nilai cluster optimal yang diperoleh digunakan pada proses clustering menggunakan algoritma K-Means. Hasil clustering kemudian dianalisis untuk mengetahui karakteristik setiap kelompok risiko obesitas yang terbentuk.

HASIL DAN PEMBAHASAN

Analisis Data

Penelitian ini menggunakan dataset obesitas yang terdiri dari 200 data individu. Dataset tersebut memuat beberapa atribut yang digunakan sebagai variabel dalam proses clustering, yaitu tinggi badan (TB), berat badan (BB), indeks massa tubuh (IMT), usia, dan jumlah konsumsi makanan utama (NCP). Seluruh atribut tersebut merupakan data numerik sehingga dapat diproses menggunakan algoritma K-Means.

Tabel 1. Data Kriteria

No	Atribut	Keterangan
1	TB	Tinggi Badan
2	BB	Berat Badan
3	IMT	Indeks Massa Tubuh
4	Usia	Umur
5	NCP	Jumlah Konsumsi Makanan Utama

Dalam menentukan keputusan dari hasil clustering peneliti memiliki 3 cluster seperti:

Tabel 2. Clustering

No	Kelas	Keterangan
1	Cluster 1	Obesitas Rendah
2	Cluster 2	Obesitas Sedang
3	Cluster 3	Obesitas Tinggi

Sebelum dilakukan proses clustering, data dianalisis untuk mengetahui kelengkapan atribut yang digunakan serta memastikan bahwa seluruh data dapat diproses pada tahap berikutnya. Selanjutnya dilakukan proses normalisasi menggunakan metode Min-Max agar setiap atribut memiliki rentang nilai yang sama sehingga tidak terdapat atribut yang mendominasi proses perhitungan jarak.

Tabel 3. Dataset

NO	NAMA	TB	BB	IMT	USIA	NCP
1	Ramlah	155	55	22.89	46	3
2	Mahmul	160	52	20.31	72	3
3	Mai Saroh Hasibuan	150	66	29.33	32	3
4	Aswat	160	60	23.4	46	3
5	Herlina	161	59	22.76	44	1
6	Yogi Manurung	165	65	23.88	35	3
7	Budi	165	70	25.71	41	3
8	Adi Syahputra	170	73	25.26	36	3
9	Saiful Amri Butar Butar	166	64	23.23	54	3
10	Mujirin	160	57	22.27	63	3
11	Zulkifli Marpaung	155	55	22.89	56	3
12	Ansi Fariwani	167	63	22.59	45	3
13	Mazla Ramadani	165	70	25.71	35	3
14	Alya Hafiza Sinaga	150	60	26.67	23	3
15	Bambang Priadi	170	65	22.49	45	1
...	...	...	...	...	...	...

200	Arifin Ilham	158	60	24.03	19	4

Normalisasi Data

Tahap normalisasi dilakukan menggunakan metode Min-Max Normalization. Tujuan normalisasi adalah mengubah seluruh nilai atribut ke dalam rentang 0 sampai 1 sehingga setiap atribut memiliki bobot yang seimbang pada saat proses perhitungan jarak menggunakan algoritma K-Means.
Rumus normalisasi Min-Max yang digunakan adalah sebagai berikut.

$$Xi = \frac{X - Xmin}{Xmax - Xmin}$$

dengan:
Xi= nilai hasil normalisasi.
X= nilai data.
Xmin = nilai minimum atribut.
Xmax= nilai maksimum atribut.

Perhitungan Normalisasi

Tabel 4. Nilai Minimum dan Maksimum Atribut

Variabel	Min	Max
TB	90	176
BB	9	85
IMT	11.11	34.05
Usia	1	72
NCP	1	4

Data 1: Ramlah
TB = 155
BB = 55
IMT = 22.89
Usia = 46
NCP = 3
Perhitungan:
TB : $\frac{155-90}{176-90} = 0.7558$
BB: $\frac{55-9}{85-9} = 0.6053$
IMT: $\frac{22.89-11.11}{34.05-11.11} = 0.5135$
Usia: $\frac{46-1}{72-1} = 0.6338$
NCP: $\frac{3-1}{4-1} = 0.6667$
Hasil normalisasi data 1:

Tabel 5. Hasil Normalisasi

TB	BB	IMT	Usia	NCP
0.7558	0.6053	0.5135	0.6338	0.6667

Hasil normalisasi seluruh data kemudian disimpan ke dalam tabel normalisasi dan digunakan sebagai masukan pada proses clustering.

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

Gambar 2. Normalisasi Data

Berdasarkan hasil normalisasi, seluruh atribut telah berada pada rentang 0 sampai 1 sehingga proses perhitungan jarak Euclidean dapat dilakukan secara lebih objektif tanpa dipengaruhi oleh perbedaan skala antaratribut.

Penentuan Jumlah Cluster Menggunakan Metode Elbow

Setelah proses normalisasi selesai dilakukan, tahap berikutnya adalah menentukan jumlah cluster yang optimal menggunakan metode Elbow. Metode ini dilakukan dengan menghitung nilai Sum of Squared Error (SSE) pada beberapa variasi jumlah cluster.

Perhitungan SSE

Data 1 (Ramlah) terhadap Cluster 1

$$(0{,}7558 - 0{,}7558)^2 + (0{,}6053 - 0{,}6053)^2 + (0{,}5135 - 0{,}5135)^2 + (0{,}6338 - 0{,}6338)^2 + (0{,}6667 - 0{,}6667)^2$$
$$= 0{,}0000 + 0{,}0000 + 0{,}0000 + 0{,}0000 + 0{,}0000$$
$$= 0{,}0000$$

Data 2 (Mahmul) terhadap Cluster 1

$$SSE_2 = (0{,}8140 - 0{,}7558)^2 + (0{,}5658 - 0{,}6053)^2 + (0{,}4010 - 0{,}5135)^2 + (1{,}0000 - 0{,}6338)^2 + (0{,}6667 - 0{,}6667)^2$$
$$= 0{,}0034 + 0{,}0016 + 0{,}0127 + 0{,}1341 + 0{,}0000$$
$$= 0{,}1518$$

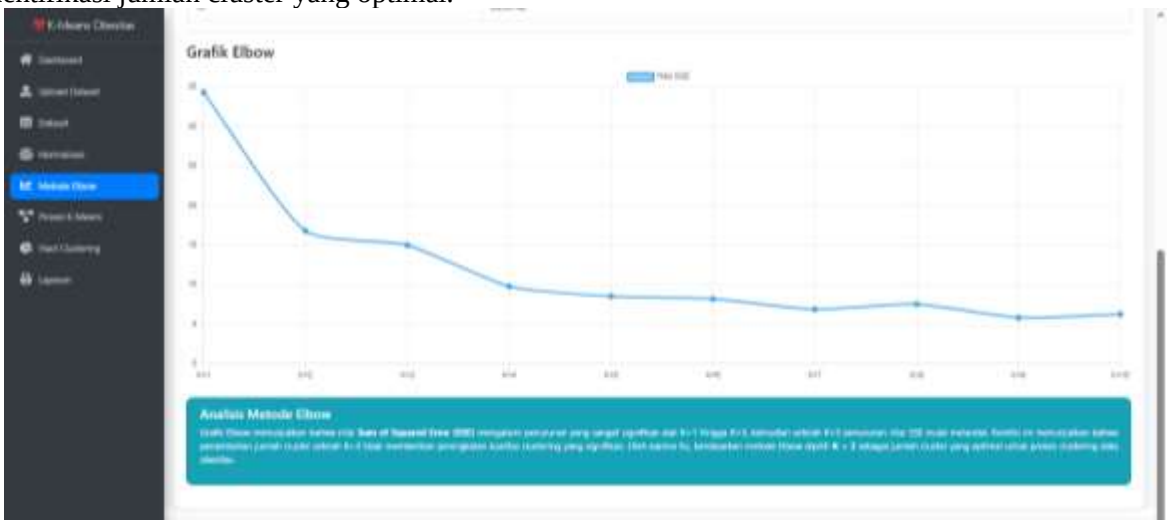
Data 3 (Mai Saroh Hasibuan) terhadap Cluster 1

$$SSE_3 = (0{,}6977 - 0{,}7558)^2 + (0{,}7500 - 0{,}6053)^2 + (0{,}7942 - 0{,}5135)^2 + (0{,}4366 - 0{,}6338)^2 + (0{,}6667 - 0{,}6667)^2$$
$$= 0{,}0034 + 0{,}0209 + 0{,}0788 + 0{,}0389 + 0{,}0000$$
$$= 0{,}1420$$

Perhitungan yang sama dilakukan terhadap seluruh data pada dataset. Selanjutnya seluruh nilai SSE dijumlahkan sehingga diperoleh nilai SSE untuk setiap jumlah cluster (K). Berdasarkan hasil perhitungan menggunakan metode Elbow diperoleh nilai SSE sebagai berikut.

Gambar 3. Nilai SSE

Nilai SSE tersebut kemudian divisualisasikan menggunakan grafik Elbow untuk mempermudah identifikasi jumlah cluster yang optimal.



Gambar 4. Grafik Elbow

Berdasarkan grafik Elbow terlihat bahwa penurunan nilai SSE terjadi secara signifikan hingga K = 3. Setelah titik tersebut penurunan SSE mulai melandai sehingga jumlah cluster yang digunakan pada penelitian ini ditetapkan sebanyak 3 cluster.

Proses Clustering Menggunakan Algoritma K-Means

Setelah jumlah cluster ditentukan, proses clustering dilakukan menggunakan algoritma K-Means dengan jumlah cluster sebanyak tiga.

Tahap pertama adalah menentukan centroid awal yang digunakan sebagai pusat cluster.

Gambar 5. Centroid Awal

Selanjutnya dihitung jarak antara setiap data terhadap seluruh centroid menggunakan metode Euclidean Distance.

Perhitungan Euclidean Distance

Data 1 : Ramlah

Perhitungan d(C1)

d(C1) = √((0.7558−0.7558)2 + (0.6053−0.6053)2 + (0.5135−0.5135)2 + (0.6338−0.6338)2 + (0.6667−0.6667)2)

= √(0.000000 + 0.000000 + 0.000000 + 0.000000 + 0.000000)

= √0.000000

= 0.0000

Perhitungan d(C2)

d(C2) = √((0.7558−0.9302)2 + (0.6053−0.7368)2 + (0.5135−0.4961)2 + (0.6338−0.7887)2 + (0.6667−0.6667)2)

= √(0.030422 + 0.017313 + 0.000304 + 0.024003 + 0.000000)

= √0.072042

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

= 0.2684

Perhitungan d(C3)
d(C3) = √((0.7558−0.7907)2 + (0.6053−0.6711)2 + (0.5135−0.5632)2 + (0.6338−0.2535)2 + (0.6667−1.0000)2)
= √(0.001217 + 0.004328 + 0.002470 + 0.144614 + 0.111111)
= √0.263740
= 0.5136

Cluster Terpilih C1

Data 2 : Mahmul
Perhitungan d(C1)
d(C1) = √((0.8140−0.7558)2 + (0.5658−0.6053)2 + (0.4010−0.5135)2 + (1.0000−0.6338)2 + (0.6667−0.6667)2)
= √(0.003380 + 0.001558 + 0.012649 + 0.134100 + 0.000000)
= √0.151688
= 0.3895

Perhitungan d(C2)
d(C2) = √((0.8140−0.9302)2 + (0.5658−0.7368)2 + (0.4010−0.4961)2 + (1.0000−0.7887)2 + (0.6667−0.6667)2)
= √(0.013521 + 0.029259 + 0.009031 + 0.044634 + 0.000000)
= √0.096445
= 0.3106

Perhitungan d(C3)
d(C3) = √((0.8140−0.7907)2 + (0.5658−0.6711)2 + (0.4010−0.5632)2 + (1.0000−0.2535)2 + (0.6667−1.0000)2)
= √(0.000541 + 0.011080 + 0.026297 + 0.557231 + 0.111111)
= √0.706260
= 0.8404

Cluster Terpilih : C2

Data 3 : Mai Saroh Hasibuan
Perhitungan d(C1)
d(C1) = √((0.6977−0.7558)2 + (0.7500−0.6053)2 + (0.7942−0.5135)2 + (0.4366−0.6338)2 + (0.6667−0.6667)2)
= √(0.003380 + 0.020949 + 0.078811 + 0.038881 + 0.000000)
= √0.142021
= 0.3769

Perhitungan d(C2)
d(C2) = √((0.6977−0.9302)2 + (0.7500−0.7368)2 + (0.7942−0.4961)2 + (0.4366−0.7887)2 + (0.6667−0.6667)2)
= √(0.054083 + 0.000173 + 0.088905 + 0.123983 + 0.000000)
= √0.267145
= 0.5169

Perhitungan d(C3)
d(C3) = √((0.6977−0.7907)2 + (0.7500−0.6711)2 + (0.7942−0.5632)2 + (0.4366−0.2535)2 + (0.6667−1.0000)2)
= √(0.008653 + 0.006233 + 0.053378 + 0.033525 + 0.111111)
= √0.212901

= 0.4614

Cluster Terpilih : C1

Data 4 : Aswat
Perhitungan d(C1)
d(C1) = √((0.8140−0.7558)2 + (0.6711−0.6053)2 + (0.5357−0.5135)2 + (0.6338−0.6338)2 + (0.6667−0.6667)2)
= √(0.003380 + 0.004328 + 0.000494 + 0.000000 + 0.000000)
= √0.008203
= 0.0906

Perhitungan d(C2)
d(C2) = √((0.8140−0.9302)2 + (0.6711−0.7368)2 + (0.5357−0.4961)2 + (0.6338−0.7887)2 + (0.6667−0.6667)2)
= √(0.013521 + 0.004328 + 0.001574 + 0.024003 + 0.000000)
= √0.043426
= 0.2084

Perhitungan d(C3)
d(C3) = √((0.8140−0.7907)2 + (0.6711−0.6711)2 + (0.5357−0.5632)2 + (0.6338−0.2535)2 + (0.6667−1.0000)2)
= √(0.000541 + 0.000000 + 0.000754 + 0.144614 + 0.111111)
= √0.257020
= 0.5070

Cluster Terpilih : C1

Data 5 : Herlina
Perhitungan d(C1)
d(C1) = √((0.8256−0.7558)2 + (0.6579−0.6053)2 + (0.5078−0.5135)2 + (0.6056−0.6338)2 + (0.0000−0.6667)2)
= √(0.004867 + 0.002770 + 0.000032 + 0.000793 + 0.444444)
= √0.452908
= 0.6730

Perhitungan d(C2)
d(C2) = √((0.8256−0.9302)2 + (0.6579−0.7368)2 + (0.5078−0.4961)2 + (0.6056−0.7887)2 + (0.0000−0.6667)2)
= √(0.010952 + 0.006233 + 0.000139 + 0.033525 + 0.444444)
= √0.495293
= 0.7038

Perhitungan d(C3)
d(C3) = √((0.8256−0.7907)2 + (0.6579−0.6711)2 + (0.5078−0.5632)2 + (0.6056−0.2535)2 + (0.0000−1.0000)2)
= √(0.001217 + 0.000173 + 0.003065 + 0.123983 + 1.000000)
= √1.128438
= 1.0623

Cluster Terpilih : C1

Data 6 : Yogi Manurung
Perhitungan d(C1)

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

d(C1) = √((0.8721−0.7558)2 + (0.7368−0.6053)2 + (0.5567−0.5135)2 + (0.4789−0.6338)2 + (0.6667−0.6667)2)
= √(0.013521 + 0.017313 + 0.001862 + 0.024003 + 0.000000)
= √0.056699
= 0.2381

Perhitungan d(C2)
d(C2) = √((0.8721−0.9302)2 + (0.7368−0.7368)2 + (0.5567−0.4961)2 + (0.4789−0.7887)2 + (0.6667−0.6667)2)
= √(0.003380 + 0.000000 + 0.003671 + 0.096013 + 0.000000)
= √0.103064
= 0.3210

Perhitungan d(C3)
d(C3) = √((0.8721−0.7907)2 + (0.7368−0.6711)2 + (0.5567−0.5632)2 + (0.4789−0.2535)2 + (0.6667−1.0000)2)
= √(0.006625 + 0.004328 + 0.000043 + 0.050784 + 0.111111)
= √0.172891
= 0.4158

Cluster Terpilih : C1

Data 7 : Budi
Perhitungan d(C1)
d(C1) = √((0.8721−0.7558)2 + (0.8026−0.6053)2 + (0.6364−0.5135)2 + (0.5634−0.6338)2 + (0.6667−0.6667)2)
= √(0.013521 + 0.038954 + 0.015112 + 0.004959 + 0.000000)
= √0.072546
= 0.2693

Perhitungan d(C2)
d(C2) = √((0.8721−0.9302)2 + (0.8026−0.7368)2 + (0.6364−0.4961)2 + (0.5634−0.7887)2 + (0.6667−0.6667)2)
= √(0.003380 + 0.004328 + 0.019703 + 0.050784 + 0.000000)
= √0.078195
= 0.2796

Perhitungan d(C3)
d(C3) = √((0.8721−0.7907)2 + (0.8026−0.6711)2 + (0.6364−0.5632)2 + (0.5634−0.2535)2 + (0.6667−1.0000)2)
= √(0.006625 + 0.017313 + 0.005363 + 0.096013 + 0.111111)
= √0.236425
= 0.4862

Cluster Terpilih : C1

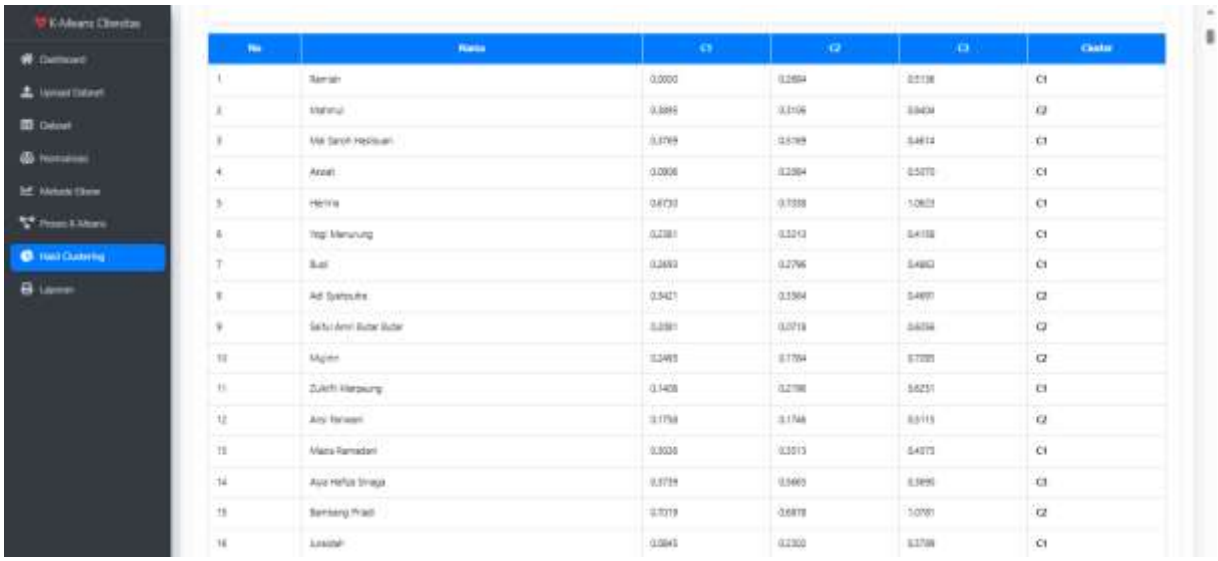
Data 8 : Adi Syahputra
Perhitungan d(C1)
d(C1) = √((0.9302−0.7558)2 + (0.8421−0.6053)2 + (0.6168−0.5135)2 + (0.4930−0.6338)2 + (0.6667−0.6667)2)
= √(0.030422 + 0.056094 + 0.010674 + 0.019837 + 0.000000)
= √0.117027
= 0.3421

Perhitungan d(C2)

d(C2) = √((0.9302−0.9302)2 + (0.8421−0.7368)2 + (0.6168−0.4961)2 + (0.4930−0.7887)2 + (0.6667−0.6667)2)
= √(0.000000 + 0.011080 + 0.014581 + 0.087483 + 0.000000)
= √0.113143
= 0.3364

Perhitungan d(C3)
d(C3) = √((0.9302−0.7907)2 + (0.8421−0.6711)2 + (0.6168−0.5632)2 + (0.4930−0.2535)2 + (0.6667−1.0000)2)
= √(0.019470 + 0.029259 + 0.002875 + 0.057330 + 0.111111)
= √0.220045
= 0.4691

Cluster Terpilih : C2



Gambar 6. Menghitung Jarak Dengan Euclidean Distance

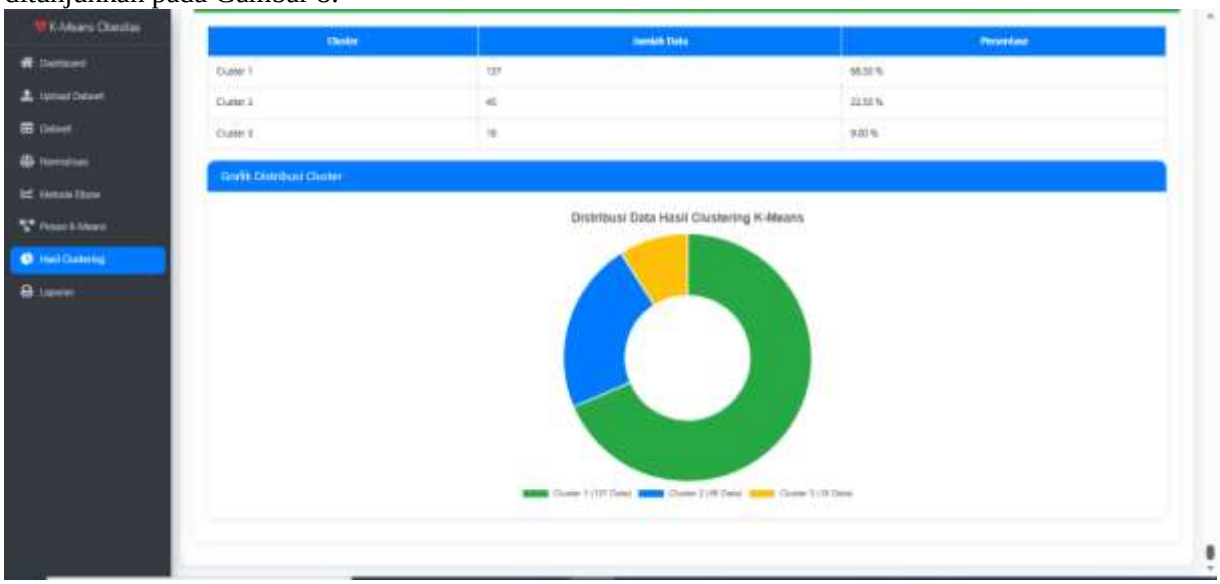
Setelah seluruh data dikelompokkan ke dalam cluster terdekat, nilai centroid baru dihitung dengan mengambil rata-rata setiap atribut pada anggota cluster. Proses pembaruan centroid dilakukan secara berulang hingga nilai centroid tidak mengalami perubahan atau anggota cluster tetap. Nilai centroid baru tersebut kemudian digunakan pada iterasi berikutnya untuk menghitung kembali jarak Euclidean dan proses clustering.

Gambar 7. Centroid Baru

ISSN 2963-7856 (print), ISSN 2961-9890 (online)

Hasil Clustering

Hasil akhir proses clustering menghasilkan tiga kelompok data berdasarkan tingkat kemiripan atribut yang digunakan pada penelitian ini. Distribusi jumlah anggota pada masing-masing cluster ditunjukkan pada Gambar 8.



Gambar 8. Grafik Clustering

Berdasarkan hasil clustering diperoleh tiga kelompok data yang memiliki karakteristik berbeda. Setiap cluster menunjukkan tingkat kemiripan yang tinggi antaranggota di dalam cluster serta memiliki perbedaan karakteristik dengan cluster lainnya. Hasil pengelompokan ini dapat digunakan sebagai dasar dalam mengidentifikasi kelompok individu berdasarkan karakteristik obesitas yang dimiliki.

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Means dengan optimasi metode Elbow berhasil diterapkan untuk mengelompokkan data tingkat risiko obesitas berdasarkan atribut tinggi badan, berat badan, indeks massa tubuh (IMT), usia, dan jumlah konsumsi makanan utama per hari (NCP). Metode Elbow berhasil menentukan jumlah cluster yang optimal, yaitu sebanyak tiga cluster. Hasil clustering menunjukkan bahwa dari 200 data yang dianalisis diperoleh Cluster 1 sebanyak 137 data (68,50%), Cluster 2 sebanyak 45 data (22,50%), dan Cluster 3 sebanyak 18 data (9,00%). Pengelompokan tersebut menunjukkan adanya perbedaan karakteristik antarcluster berdasarkan tingkat kemiripan data. Dengan demikian, penerapan algoritma K-Means yang dioptimasi menggunakan metode Elbow dapat membantu proses pengelompokan data obesitas secara efektif serta memberikan informasi yang dapat mendukung analisis dan pengambilan keputusan di bidang kesehatan.

Kalau jurnal ini akan dikirim ke dosen atau dipublikasikan, saya juga menyarankan mengganti penyebutan "Cluster 1, Cluster 2, dan Cluster 3" dengan deskripsi karakteristik masing-masing cluster (misalnya berdasarkan nilai centroid), bukan langsung menyebut "obesitas rendah/sedang/tinggi", karena K-Means secara alami hanya membentuk kelompok, bukan memberi label tingkat risiko. Label seperti "rendah", "sedang", dan "tinggi" baru boleh digunakan jika memang didasarkan pada analisis centroid masing-masing cluster.

UCAPAN TERIMA KASIH

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Ucapan terima kasih juga disampaikan kepada dosen pembimbing, seluruh dosen, serta semua pihak yang telah memberikan bimbingan, masukan, dan dukungan sehingga penelitian ini dapat diselesaikan dengan baik.

REFERENSI

Hasby, A., Bangun, B., & Masrizal, M. (2025). Data Mining Dalam Clusterisasi Risiko Tinggi Obesitas Menggunakan Metode K-Means Clustering. *Building of Informatics, Technology and Science (BITS), 7*(1), 863-872.

Santosa, R. R., Amali, A., Anindia, A., & Hamim. (2024). *K-Means clustering approach to obesity risk categorization with RapidMiner*. Jurnal SIGMA, 15(1).

Setiawati, E., Fernanda, U. D., Agesti, S., Iqbal, M., & Herjho, M. O. A. (2024). *Implementation of K-Means, K-Medoid and DBSCAN algorithms in obesity data clustering*. International Journal of Applied Technology and Innovation Science (IJATIS), 1(1), 23–29.

Wijaya, B. (2020). *Clustering menggunakan metode K-Means dengan bantuan metode Elbow*. Jurnal Komputer dan Informatika, 15(2).

Ashari, I. F., Nugroho, E. D., Baraku, R., Yanda, I. N., & Liwardana, R. (2023). *Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index evaluation on K-Means algorithm*. Journal of Applied Informatics and Computing, 7(1), 95–103.

World Health Organization. (2024). *Obesity and overweight*. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

ISSN 2963-7856 (print), ISSN 2961-9890 (online)