

SENASTIKA Universitas Malikussaleh

SEGMENTASI RESIKO OBESITAS DAN PENYAKIT JANTUNG DENGAN ALGORITMA K-MEANS

Muhammad Dzikrillah¹, Nuzul Althaf², Sylva Putri Utari³, Suhaiba Nasyira Hariono*⁴, Munirul Ula*⁵

1,2,3,4 Teknik Informatika, Universitas Malikussaleh, lhokseumawe. Aceh

Email: ¹muhammad.210170249@mhs.unimal.ac.id, ²nuzul.210170258@mhs.unimal.ac.id, ³sylva.210170247@mhs.unimal.ac.id, ⁴suhaiba.210170253@mhs.unimal.ac.id, ⁵munirulula@unimal.ac.id

Abstrak

Penelitian ini bertujuan untuk melakukan segmentasi risiko obesitas dan penyakit jantung dengan menerapkan algoritma K-Means dan mengidentifikasi faktor-faktor yang memengaruhi tingkat obesitas pada pasien pria dan wanita. Faktor-faktor yang dianalisis mencakup usia, tinggi badan, berat badan, riwayat keluarga dengan berat badan berlebih, frekuensi konsumsi makanan tinggi kalori, frekuensi konsumsi sayuran, jumlah makan utama dalam sehari, kebiasaan ngemil, kebiasaan merokok, asupan air, pemantauan kalori, frekuensi aktivitas fisik, durasi penggunaan perangkat teknologi, konsumsi alkohol, dan jenis transportasi yang digunakan. Hasil penelitian menunjukkan bahwa analisis terbaik dari setiap metode dan optimasi K-Means menghasilkan nilai Davies-Bouldin Index optimal pada $n_clusters = 3$ dengan nilai DBI 1.6545, serta nilai terbaik untuk Silhouette Coefficient pada $n_clusters = 4$ dengan rata-rata 0.2445. Temuan ini menunjukkan bahwa K-Means merupakan metode yang paling efektif untuk segmentasi risiko obesitas dan penyakit jantung. Meskipun nilai evaluasi clustering mungkin tidak optimal, hal ini tidak selalu mencerminkan kualitas pemodelan; bisa jadi ini menunjukkan bahwa data tidak terkelompok dengan jelas atau perlu penyesuaian parameter model. Oleh karena itu, evaluasi sebaiknya digabungkan dengan pemahaman tentang domain dan analisis eksploratif untuk memastikan model clustering sesuai dengan kebutuhan analisis.

Keywords: *Segmentasi, Risiko Obesitas, Penyakit Jantung, Algoritma K-Means, Silhouette Coefficient dan Davies-Bouldin Index*

1. PENDAHULUAN

Obesitas telah menjadi masalah kesehatan yang signifikan dengan prevalensi yang meningkat. Diperkirakan 1,6 miliar orang dewasa di dunia mengalami kelebihan berat badan, dan 400 juta di antaranya obesitas. Pada tahun 2015, diperkirakan 2,3 miliar orang akan overweight dan 700 juta diantaranya obesitas. Penyakit jantung terjadi saat aliran darah ke otot jantung tersumbat, menyebabkan jantung tidak berfungsi dengan baik dan berpotensi mengalami kerusakan serius atau penyempitan arteri koroner[1]. Faktor risiko terbagi menjadi yang tidak dapat diubah, seperti usia, jenis kelamin, dan keturunan, serta yang dapat diubah, seperti hipertensi, kolesterol tinggi, obesitas, diabetes, kurang aktivitas fisik, dan konsumsi alkohol berlebih. Menurut peneliti Haryadi dkk, [2] Serangan jantung merupakan penyakit mematikan dengan angka kematian mencapai 12,90% dari semua penyakit jantung.

Kurangnya akses informasi tentang penyakit jantung menyebabkan peningkatan angka kematian setiap tahun, sehingga dibutuhkan sistem klasifikasi untuk memberikan informasi dan melakukan pengecekan dini[3]. Obesitas umumnya disebabkan oleh ketidakseimbangan antara asupan makanan dan aktivitas fisik, dengan individu obesitas cenderung makan berlebihan dan kurang berolahraga. Namun, obesitas dipengaruhi oleh berbagai faktor, termasuk genetik, fisiologis, metabolik, psikologis, sosioekonomik, gaya hidup, dan budaya[4]. Obesitas berisiko meningkatkan penyakit kronis seperti hipertensi, diabetes, dan kanker, serta dapat menimbulkan stigma sosial bagi penderitanya.

Penyakit jantung dan obesitas menjadi dua isu kesehatan yang saling terkait dan semakin mendesak untuk diatasi dalam masyarakat modern[5]. Dengan meningkatnya angka kematian akibat penyakit jantung, terutama yang disebabkan oleh kurangnya akses informasi dan pemahaman yang tepat mengenai kondisi ini, muncul kebutuhan mendesak akan sistem klasifikasi yang dapat membantu dalam deteksi dini dan penanganan penyakit. Di sisi lain, obesitas yang sering kali berakar dari pola hidup yang tidak seimbang, menjadi pemicu berbagai penyakit kronis yang berkontribusi pada meningkatnya risiko kematian[6].

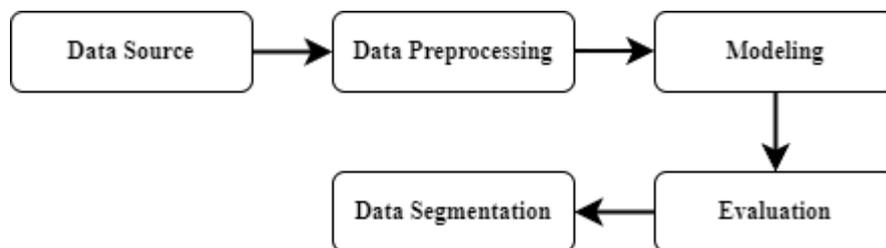
Penyakit jantung dan obesitas menjadi dua isu kesehatan yang saling terkait dan semakin mendesak untuk diatasi dalam masyarakat modern. Dengan meningkatnya angka kematian akibat penyakit jantung, terutama yang disebabkan oleh kurangnya akses informasi dan pemahaman yang tepat mengenai kondisi ini, muncul kebutuhan mendesak akan sistem klasifikasi yang dapat membantu dalam deteksi dini dan penanganan penyakit. Di sisi lain, obesitas yang sering kali berakar dari pola hidup yang tidak seimbang menjadi pemicu berbagai penyakit kronis yang berkontribusi pada meningkatnya risiko kematian. Pemahaman yang lebih mendalam tentang pola obesitas

dan faktor-faktor risikonya diperlukan untuk mendukung upaya pencegahan dan pengobatan yang lebih efisien. Salah satu pendekatan yang bisa digunakan dalam analisis data kesehatan adalah metode pengelompokan (clustering)[7].

Tujuan dari penelitian ini adalah untuk mengidentifikasi faktor-faktor penyebab obesitas dan penyakit jantung, serta mengeksplorasi hubungan antara keduanya[8]. Penelitian ini menggunakan algoritma K-Means dengan evaluasi hasil clustering melalui Silhouette Score dan Davies-Bouldin Index[9]. Kedua metode evaluasi internal ini membantu mengukur kualitas pengelompokan data berdasarkan karakteristik risiko dan dampak sosial yang dihadapi individu dengan kondisi tersebut. Silhouette Score mengukur seberapa baik data dalam sebuah cluster dibandingkan dengan cluster lainnya, sementara Davies-Bouldin Index mengevaluasi jarak antar cluster, memberikan gambaran tentang homogenitas dan kualitas cluster yang dihasilkan[10]. Beberapa penelitian sebelumnya mendukung penggunaan metrik ini dalam mengevaluasi hasil clustering. Penelitian ini diharapkan memberikan kontribusi signifikan dengan mengusulkan penggunaan Silhouette Score dan Davies-Bouldin Index sebagai metrik evaluasi untuk menilai efektivitas clustering K-Means, yang dapat bermanfaat bagi masyarakat dalam upaya pencegahan obesitas dan penyakit jantung.

2. METODE PENELITIAN

Tahapan penelitian merupakan rangkaian langkah terstruktur yang diterapkan agar proses penelitian menjadi lebih mudah dan hasilnya tersusun dengan baik. Pada gambar 1 menggambarkan tahapan-tahapan penelitian.



Gambar 1. Tahapan Penelitian

Gambar 1 memperlihatkan tahapan penelitian, dimulai dari pengambilan dataset pada tahap Data Source untuk memulai penelitian. Selanjutnya, pada tahap Data Preprocessing, dataset diolah agar siap untuk dianalisis dan dimodelkan. Pada tahap ketiga, dilakukan Modeling dengan menerapkan algoritma K-Means pada dataset yang telah diproses. Kemudian, pada tahap Evaluation, evaluasi dilakukan terhadap cluster yang telah terbentuk. Terakhir, pada tahap Data Segmentation, cluster terbaik dipilih dari semua cluster yang telah dihasilkan.

a. Data Source

Data yang digunakan dalam penelitian ini berasal dari dataset yang dapat diakses di situs web data publik dengan mencari menggunakan kata kunci “Obesity/CVD Risk Clustering with K-Means” di platform Kaggle.com.

Tabel 1. Dataset Penyakit Obesitas Dan Jantung

No	Gender	Age	V1	V2	...	V11	V12	V13	V14	V15	V16	V17
1	female	21	1,62	64	...	2	no	0	1	no	public_trnsprt	normal_weight
2	female	21	1,52	56	...	3	yes	3	0	sometimes	public_trnsprt	normal_weight
3	male	23	1,8	77	...	2	no	2	1	frequently	public_trnsprt	normal_weight
4	male	27	1,8	87	...	2	no	2	0	frequently	walking	overweight_level_i
...
2106	female	21,28	1,72	131,3		1,79	no	1,72	0,89	sometimes	public_trnsprt	obesity_type_iii
2107	female	20,97	1,71	131,4		1,72	no	1,67	0,90	sometimes	public_trnsprt	obesity_type_iii
2108	female	21,98	1,74	133,7		2,00	no	1,34	0,59	sometimes	public_trnsprt	obesity_type_iii
2119	female	22,52	1,75	133,6		2,05	no	1,41	0,64	sometimes	public_trnsprt	obesity_type_iii
2110	female	24,36	1,73	133,3		2,85	no	1,13	0,58	sometimes	public_trnsprt	obesity_type_iii
2111	female	23,66	1,73	133,4		2,86	no	1,02	0,71	sometimes	public_trnsprt	obesity_type_iii

b. Data Preprocessing

Data Preprocessing adalah proses pengolahan data agar dapat dipahami oleh mesin dengan lebih mudah [11]. Proses ini mencakup beberapa langkah, seperti:

1. Memilih atribut yang relevan untuk digunakan dalam modeling dengan algoritma k-means.
2. Mengisi nilai yang hilang dalam dataset menggunakan rata-rata.
3. Menghapus data yang duplikat dari dataset.
4. Memeriksa atribut yang mengindikasikan adanya outlier dan melakukan normalisasi data [12].

c. Modeling

Dengan menentukan nilai parameter yang optimal, tahap modeling ini mengimplementasikan algoritma k-means. K-Means adalah algoritma klasterisasi yang membagi objek pengamatan ke dalam kelompok menggunakan metode partisi [13]. Algoritma ini membagi data ke dalam kelompok berdasarkan titik pusat (centroid). Langkah-langkah dalam perhitungan algoritma k-means adalah sebagai berikut:

1. Tentukan jumlah cluster (k) yang diinginkan.
2. Pilih centroid cluster secara acak.
3. Hitung jarak setiap data terhadap centroid menggunakan rumus Euclidean:

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2} \quad (1)$$

Di mana $d(x, y)$ melambangkan jarak antara dua titik data; x adalah titik data pada objek, dan y adalah titik pusat. Jumlah objek dilambangkan dengan i .

4. Kelompokkan data berdasarkan jarak minimum antara centroid dan data untuk membentuk cluster baru.
5. Hitung ulang centroid (pusat cluster baru) untuk setiap cluster menggunakan rumus:

$$CI = \frac{x1 + x2 + x3 + \dots + xn}{\sum x} \quad (2)$$

CI adalah titik pusat baru dari sebuah cluster, dengan $x1$ hingga xn mewakili nilai dalam cluster ke- n .

6. Ulangi langkah 3 hingga 5 hingga tidak ada data yang berpindah cluster.

Sebelum mengimplementasikan algoritma k-means [14], perlu dilakukan penentuan jumlah cluster yang sesuai. Metode Elbow digunakan untuk mengukur konsistensi jumlah cluster yang optimal, yang ditandai dengan adanya lekukan pada grafik yang menyerupai siku.

d. Evaluation

Evaluation adalah proses pengukuran hasil dari model yang telah diimplementasikan pada tahap modeling sebelumnya. Pada tahap ini, dilakukan penilaian terhadap tingkat keberhasilan dan akurasi cluster yang dihasilkan dari proses modeling. Beberapa metode evaluasi yang digunakan antara lain Silhouette Score, dan Davies Bouldin Index (DBI).

e. Data Segmentation

Data segmentation digunakan untuk mengelompokkan individu berdasarkan faktor-faktor yang memengaruhi risiko obesitas dan penyakit jantung, seperti usia, berat badan, riwayat keluarga, pola makan, kebiasaan merokok, dan aktivitas fisik. Dengan algoritma K-Means, individu dikelompokkan ke dalam segmen dengan karakteristik serupa, sehingga dapat memudahkan identifikasi kelompok berisiko tinggi untuk intervensi yang lebih tepat dalam pencegahan dan pengobatan.

3. HASIL DAN PEMBAHASAN

3.1 Data Source

Data Source dalam penelitian ini menggunakan dataset Obesity/CVD Risk. Tabel 2 menjelaskan setiap atribut yang terdapat dalam dataset tersebut.

Tabel 2. Deskripsi *Dataset Obesity-CVD Risk*

Atribut	Keterangan	Penjelasan
Gender	Jenis kelamin	Male= 51% dan female = 49%
Age	Umur(tahun)	Umur pasien “min = 14, dan maks = 61”.
Height	Tinggi Badan	Tinggi badan pasien “min = 1,45 cm dan maks = 1,98 cm”.
Weight	Berat Badan	Berat badan pasien “min = 39 kg, dan maks = 173 kg”.
FHWO	family_history_with_overweight	Memiliki Keluarga yang berat badannya tinggi dengan kategori “yes = 82% dan no 18%”.
FAVC	Frequent consumption of high caloric food	Frekuensi mengonsumsi makanan berkalori tinggi dengan kategori “yes = 88% dan no = 12%”.
FCVC	Frequency of consumption of vegetables	Frekuensi Konsumsi Sayur perhari dengan kategori 1(sedikit), 2(sedang) dan 3(banyak).
NCP	Number of main meals	Jumlah makan dalam sehari dengan kategori 1(sedikit), 2(sedang) dan 3(banyak).
CAEC	Consumption of food between meals	Konsumsi makanan atau cemilan di antara waktu makan dengan kategori “no = 5%, sometimes = 84%, frequently = 11% dan always = 5%”
SMOKE	Merokok	pasien dibedakan dengan kategori “no dan yes”
CH20	Consumption of water daily	Jumlah Air yang diminum dalam sehari dengan kategori 1(sedikit), 2(sedang) dan 3(banyak).
SCC	Calories consumption monitoring	Pemantauan konsumsi jumlah kalori dengan kategori “yes dan no”
FAF	Physical activity frequency	Frekuensi aktivitas fisik dengan kategori 0(tidak), 1(jarang), 2(sering) dan 3(selalu).
TUE	Time using technology devices	Menghabiskan waktu menggunakan perangkat teknologi dengan kategori 0(tidak), 1(jarang) dan 2(selalu).
CALC	Consumption of alcohol	Jumlah alkohol yang dikonsumsi dengan kategori no(tidak), sometimes(kadang), frequently(sering), dan always(selalu)
MTRANS	Transportation used	Transportasi yang digunakan sehari hari pasien dengan kategori “Public transportation, automobile, motorbike, bike, dan walking”.
Nobeyesdad	Tingkat obesitas dikurangi	Mengacu pada kategori atau tingkat obesitas yang ditentukan seperti: Insufficient_Weight, Normal_Weight, Overweight_Level_I, Overweight_Level_II, Obesity_Type_I, Obesity_Type_II, Obesity_Type_III

Dataset ini mencakup data untuk estimasi tingkat obesitas pada individu berdasarkan kebiasaan makan dan kondisi fisik mereka. Data berisi 17 atribut dan 500 catatan, catatan diberi label dengan variabel kelas NObesity (Tingkat Obesitas), yang memungkinkan klasifikasi / pengelompokan / Regresi data menggunakan nilai Berat Badan Kurang, Berat Badan Normal, Berat Badan Berlebih I, Berat Badan Berlebih II, Jenis Obesitas I, Jenis Obesitas II, dan Jenis Obesitas III. Setelah menggunakan algoritma K-Means untuk klusterisasi dataset, kami menemukan bahwa tiga klaster memberikan segmentasi terbaik dalam menggambarkan risiko obesitas individu. Klaster-klaster tersebut dibagi berdasarkan faktor gaya hidup seperti frekuensi aktivitas fisik, kebiasaan mengonsumsi makanan berkalori tinggi, serta faktor genetik seperti riwayat keluarga dengan obesitas, dll. Setelah

mengetahui penjelasan atribut pada dataset kemudian melanjutkan pada tahap *data processing* untuk dianalisis lebih lanjut.

3.2 Data Preprocessing

Sebelum penerapan algoritma K-Means, dilakukan praproses data untuk memastikan kualitas data yang baik [11]. tahapan praproses meliputi:

1. Konversi data kategorikal menjadi numerik: Karena K-means menggunakan jarak Euclidean, data kategorikal perlu diubah ke format numerik[15]. Ini dapat dilakukan dengan metode seperti Label Encoding atau One-Hot Encoding. Data ini terdapat pada tabel 3 dibawah ini.

Tabel 3. Konfersi Data Kategorikal-Numerik

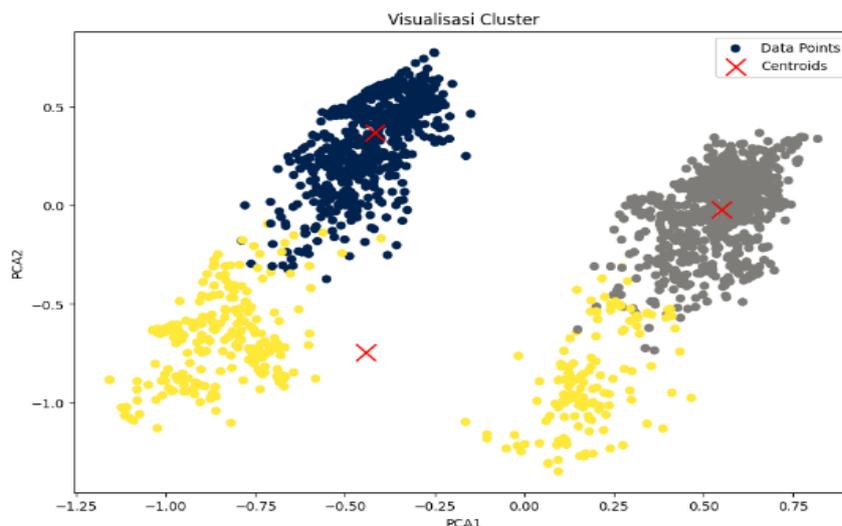
Index	Gender	Age	V1	V2	...	V13	V14	V15	V16	V17
0	0	21	1,62	64	...	0	1	3	3	1
1	0	21	1,52	56	...	3	0	2	3	1
2	1	23	1,8	77	...	2	1	1	3	1
3	1	27	1,8	87	...	2	0	1	4	5
4	1	22	1,78	89,8	...	0	0	2	3	6

2. Normalisasi data: Data numerik sering kali perlu dinormalisasi untuk memastikan skala yang konsisten. Hal ini penting karena K-means peka terhadap skala variabel[12]. Salah satu metode untuk melakukan ini adalah MinMaxScaler, yang mengubah data sehingga fitur-fitur berada dalam rentang [0, 1]. Normalisasi data merupakan proses penyesuaian skala data agar semua fitur memberikan kontribusi yang seimbang terhadap algoritma k-means dan berikut ini hasil setelah di normalisasi.

Tabel 4. Hasil Normalisasi Data

Index	Gender	Age	V1	V2	...	V13	V14	V15	V16	V17
0	0	0,14	0,32	0,18	...	0	0,5	1	0,75	0,16
1	0	0,14	0,13	0,12	...	1	0	0,66	0,75	0,16
2	1	0,19	0,66	0,28	...	0,66	0,5	0,33	0,75	0,16
3	1	0,27	0,66	0,35	...	0,66	0	0,33	1	0,83
4	1	0,17	0,62	0,37	...	0	0	0,66	0,75	1

3. Visualisasi cluster: Setelah model di-fit, hasil klusterisasi dapat divisualisasikan menggunakan plot dua dimensi jika data telah direduksi, atau dengan teknik visualisasi multidimensi lainnya.



Gambar 2. Hasil visualisasi cluster

Centroid dari setiap kluster, yang merupakan titik tengah dari kluster, dapat ditampilkan pada visualisasi untuk memberikan pemahaman tentang pusat setiap kluster.

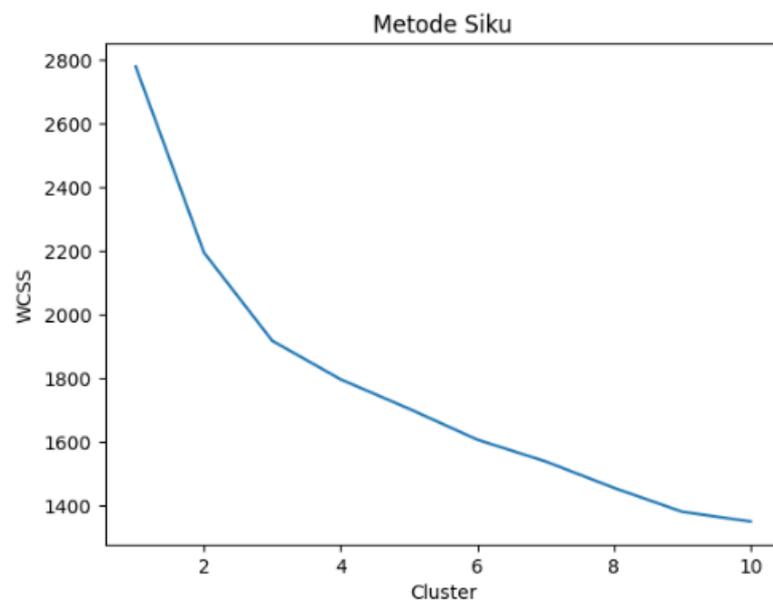
3.3 Modelling

Setelah menyelesaikan semua langkah dalam data preprocessing, dataset siap untuk menentukan model tingkat risiko terkena obesitas dan penyakit jantung yang dikategorikan sebagai "0 = Rendah, 1 = Sedang, dan 2 = Tinggi". pada tabel dibawah ini

Tabel 5. Hasil Klasterisasi

Gender	Age	V1	V2	...	V13	V14	V15	V16	V17	Cluster	Resiko Terkena(Cluster Label)
Female	21	1,62	64	...	0	1	no	Public_Transport	Nrml_Weight	0	Rendah
Female	21	1,52	56	...	3	0	Sometimes	Public_Transport	Nrml_Weight	0	Rendah
Male	23	1,8	77	...	2	1	Frequently	Public_Transport	Nrml_Weight	1	Sedang
Male	27	1,8	87	...	2	0	Frequently	Walking	Overweight_Lvl_I	2	Tinggi
...
Female	21,28	1,72	131,33	...	1,72	0,89	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	20,97	1,71	131,40	...	1,67	0,90	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	21,98	1,74	133,74	...	1,34	0,59	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	22,52	1,75	133,68	...	1,41	0,64	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	24,36	1,73	133,34	...	1,13	0,58	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah
Female	23,66	1,73	133,47	...	1,02	0,71	Sometimes	Public_Transport	Obesity_Type_III	0	Rendah

Kemudian untuk menentukan jumlah kluster yang optimal, metode elbow digunakan. Ini melibatkan plot dari jumlah kluster terhadap inersia (total jarak dalam kluster) dan mencari titik "siku" di mana peningkatan jumlah kluster tidak lagi memberikan penurunan inersia yang signifikan.

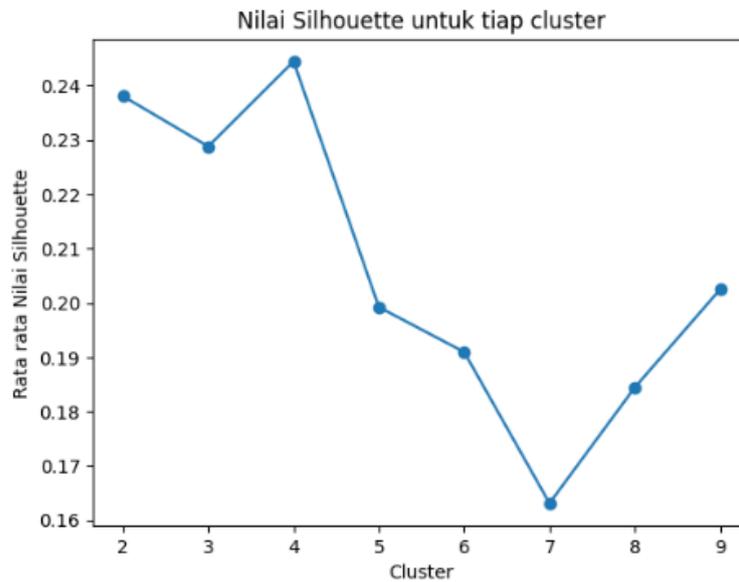


Gambar 3. Grafik Elbow Method

Gambar 3 memperlihatkan grafik elbow method yang menunjukkan penurunan signifikan pada WCSS hingga mencapai sekitar 3 cluster, setelah itu penurunan menjadi lebih lambat. Oleh karena itu, maka nilai $k = 3$ ditetapkan sebagai nilai optimal untuk pengolahan data menggunakan algoritma K-Means.

3.4 Evaluation

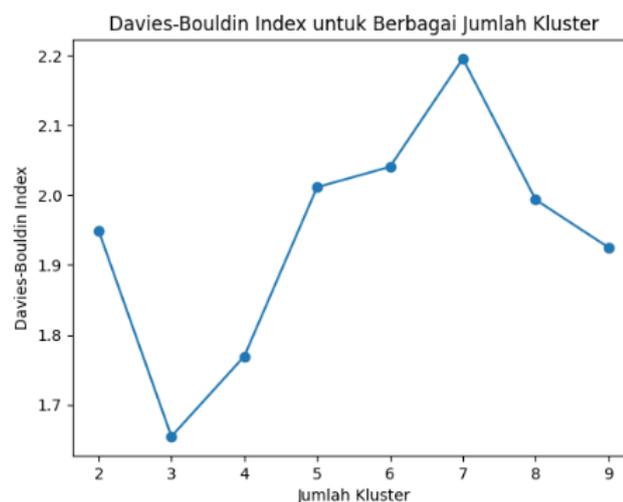
a. Silhouette Score



Gambar 4. Grafik Nilai Silhouette Score

Silhouette score mengukur kemiripan setiap titik data dengan kelompoknya dibandingkan dengan kelompok terdekat[16]. Nilai yang lebih tinggi menunjukkan pengelompokan yang lebih baik, di mana titik data lebih dekat ke centroid kelompoknya dan lebih jauh dari centroid kelompok lain[17]. Gambar 6 menunjukkan bahwa cluster 2 dan 6 memiliki nilai silhouette score yang cukup tinggi.

b. Davies Bouldin Index (DBI)



Gambar 5. Grafik Nilai Davies Bouldin Index

Davies Bouldin Index mengukur rasio antara jarak rata-rata intra-cluster (jarak titik data dalam kelompok) dan jarak minimum antar centroid kelompok. Nilai yang lebih rendah menunjukkan pengelompokan yang lebih baik[18]. Gambar 7 menunjukkan bahwa cluster ke-3 memiliki nilai Davies Bouldin Index terendah.

3.5 Data Segmentation

Pada tahap data segmentation, yaitu menentukan cluster mana berdasarkan dari dataset diatas, peneliti mengidentifikasi dan mendapatkan sejumlah faktor yang memengaruhi tingkat obesitas pada pasien pria dan wanita dengan penyakit jantung. Faktor-faktor tersebut mencakup usia, tinggi badan, berat badan, riwayat keluarga dengan berat badan berlebih (*family_history_with_overweight*), frekuensi konsumsi makanan tinggi kalori (FAVC), frekuensi konsumsi sayuran (FCVC), jumlah makan utama dalam sehari (NCP), kebiasaan ngemil (CAEC), kebiasaan merokok (SMOKE), jumlah air yang diminum setiap hari (CH2O), pemantauan konsumsi kalori (SCC), frekuensi aktivitas fisik (FAF), durasi penggunaan perangkat teknologi (TUE), konsumsi alkohol (CALC), transportasi yang digunakan sehari-hari (MTRANS), serta tingkat obesitas yang dikurangi (NObesidad).

3. DISKUSI

Hasil penelitian kami menunjukkan bahwa algoritma K-Means efektif dalam melakukan segmentasi risiko obesitas dan penyakit jantung dengan mempertimbangkan berbagai faktor seperti pola konsumsi makanan, aktivitas fisik, dan kebiasaan hidup sehari-hari. Temuan ini sejalan dengan jurnal terdahulu yang juga menggunakan algoritma K-Means untuk mengelompokkan makanan cepat saji berdasarkan kandungan nutrisinya guna membantu penderita obesitas memilih makanan yang lebih sehat.

Namun, terdapat beberapa perbedaan dalam pendekatan dan hasil penelitian. Jurnal terdahulu yang berjudul **"Implementasi K-Means Untuk Pengelompokan Makanan Cepat Saji Bagi Penderita Penyakit Obesitas"** yang ditulis oleh Yoannes Dion Pradvenanta dan Rastri Prathivi yang dipublikasikan pada bulan Juni Tahun 2024 berfokus pada segmentasi makanan cepat saji bagi penderita obesitas, di mana data diambil dari 515 sampel makanan, dan hasil segmentasi menunjukkan bahwa dari 5 cluster yang dihasilkan, 244 sampel makanan dapat dianggap sebagai pilihan yang lebih sehat. Di sisi lain, penelitian kami lebih menitikberatkan pada segmentasi risiko obesitas dan penyakit jantung pada individu dengan dataset yang mencakup 2111 sampel pasien. Dengan menggunakan 3 hingga 4 cluster, hasil penelitian kami memberikan pemahaman yang lebih komprehensif mengenai faktor risiko yang terkait dengan obesitas dan penyakit jantung, termasuk riwayat keluarga, pola makan tinggi kalori, dan frekuensi aktivitas fisik.

Dalam hal prevalensi obesitas, perbandingan data dari kedua jurnal menunjukkan tren yang mengkhawatirkan. Berdasarkan jurnal kami, terdapat 1,6 miliar orang dewasa di dunia yang mengalami kelebihan berat badan, dengan 400 juta di antaranya menderita obesitas parah. Angka ini diperkirakan terus meningkat menjadi 2,3 miliar orang yang mengalami kelebihan berat badan dan 700 juta dengan obesitas pada tahun 2015. Sementara itu, jurnal terdahulu melaporkan bahwa pada tahun 2022, satu dari delapan orang di dunia menderita obesitas, termasuk 890 juta orang dewasa dan 160 juta anak-anak dan remaja. Hal ini mempertegas bahwa obesitas menjadi salah satu masalah kesehatan global yang semakin serius, dengan dampak signifikan terhadap risiko penyakit tidak menular, termasuk penyakit jantung, diabetes, dan gangguan metabolik lainnya.

Evaluasi model yang digunakan pada kedua penelitian juga memperlihatkan hasil yang berbeda. Penelitian kami menunjukkan bahwa nilai terbaik untuk Davies-Bouldin Index (DBI) adalah 1.6545 pada 3 cluster, dan Silhouette Coefficient sebesar 0.2445 pada 4 cluster, yang menunjukkan pemisahan cluster yang cukup baik. Di sisi lain, jurnal terdahulu menggunakan Sum Square Error (SSE), Silhouette Score, dan DBI untuk menilai clustering makanan cepat saji, dengan hasil optimal pada 5 cluster.

Perbandingan ini menunjukkan bahwa meskipun kedua studi menggunakan algoritma yang sama, fokus analisis dan populasi yang diteliti berbeda. Penelitian kami lebih berfokus pada identifikasi faktor risiko dan segmentasi pasien, sedangkan jurnal terdahulu lebih fokus pada segmentasi makanan berdasarkan kandungan nutrisinya untuk membantu penderita obesitas membuat pilihan yang lebih sehat. Kedua pendekatan ini saling melengkapi dalam upaya mengurangi risiko obesitas dan penyakit terkait.

4. KESIMPULAN

Dari penelitian ini dapat disimpulkan bahwa algoritma K-Means efektif untuk segmentasi risiko obesitas dan penyakit jantung berdasarkan berbagai faktor. Tiga cluster teridentifikasi: Cluster 0 (risiko rendah), Cluster 1 (risiko sedang), dan Cluster 2 (risiko tinggi). Usia juga berpengaruh, di mana individu yang lebih tua cenderung memiliki risiko lebih tinggi. Temuan ini menunjukkan pentingnya pemantauan faktor risiko dan perlunya pendekatan terfokus dalam pencegahan. Nilai evaluasi yang tidak optimal tidak selalu mencerminkan buruknya pemodelan, tetapi bisa jadi karena kurangnya pengelompokan jelas dalam data atau perlunya penyesuaian parameter. Oleh karena itu, evaluasi harus dikombinasikan dengan pemahaman domain dan analisis eksploratif untuk memastikan kesesuaian model clustering dengan kebutuhan analisis.

DAFTAR PUSTAKA

- [1] F. Rahmadayanti, S. Muntari, dan R. Putriani, “PENGELOMPOKAN TINGKAT RESIKO PENYAKIT JANTUNG BERDASARKAN USIA MENGGUNAKAN ALGORITMA K-MEANS,” 2023.
- [2] D. Haryadi dan D. Marini Umi Atmaja, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Risiko Penyakit Jantung,” *Journal of Informatics and Communications Technology*, vol. 3, no. 2, hlm. 051–066, 2021.
- [3] J. Wala dan R. Umar, “Implementasi K-Means Clustering pada Pengelompokan Pasien Penyakit Jantung,” 2024.
- [4] H. H. Satoto*, “TINJAUAN PUSTAKA Patofisiologi Penyakit Jantung Koroner Coronary Heart Disease Pathophysiology,” 2014.
- [5] Sonia Indhira dan Billy Hendrik, “Implementation Of The K-Means Clustering Algorithm For Grouping Heart Disease Risk Levels,” *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, vol. 3, no. 3, hlm. 154–170, Okt 2023, doi: 10.55606/juisik.v3i3.677.
- [6] Y. Dion Pradvenanta dan R. Prathivi, “Implementasi K-Means Untuk Pengelompokan Makanan Cepat Saji Bagi Penderita Penyakit Obesitas,” *Technology and Science (BITS)*, vol. 6, no. 1, 2024, doi: 10.47065/bits.v6i1.5279.
- [7] R. Wahyusari, S. Tinggi, dan T. Ronggolawe, “Penerapan Algoritma K-Medoids Untuk Mengelompokkan Status Obesitas,” 2024.
- [8] I. A. T. H. A. B. S. Ice J. Johanis1, “Media Kesehatan Masyarakat FAKTOR RISIKO HIPERTENSI, MEROKOK DAN USIA TERHADAP,” *Media Kesehatan Masyarakat*, vol. 2, no. 1, 2020, [Daring]. Tersedia pada: <https://ejurnal.undana.ac.id/MKM>
- [9] F. A. B. N. S. Abyan Rasyid Aryadi, “Analisis Klaster K-Means pada Data Rata-Rata Konsumsi Kalori dan Protein Menurut Provinsi dengan Metode Davies Bouldin Index,” 2022.
- [10] D. Hartama dan I. Sudahri Damanik, “Pengelompokan Algoritma K-Means dan K-Medoid Berdasarkan Lokasi Daerah Rawan Bencana di Indonesia dengan Optimasi Elbow, DBI, dan Silhouette,” *Technology and Science (BITS)*, vol. 6, no. 2, 2024, doi: 10.47065/bits.v6i2.5851.
- [11] G. Fouriza Ibanez, W. Giri, dan R. Wiriasto, “Kombinasi Principal Component Analysis dengan Algoritma K-Means untuk Klasterisasi Data Stunting,” *Media Online*, vol. 5, no. 1, hlm. 131–141, 2024, doi: 10.30865/klik.v5i1.1977.
- [12] L. Mutawalli, S. Fadli, dan S. Supardianto, “Komparasi Metode Perhitungan Jarak K-Means Paling Baik Terhadap Pembentukan Pola Kunjungan Wisatawan Mancanegara,” *Journal of Information System Research (JOSH)*, vol. 5, no. 1, hlm. 159–166, Okt 2023, doi: 10.47065/josh.v5i1.4377.
- [13] Y. R. Nasution dan M. Eka, “PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA APLIKASI MENENTUKAN BERAT BADAN IDEAL,” 2018.
- [14] A. Satriawan, R. Andreswari, dan O. N. Pratiwi, “SEGMENTASI PELANGGAN TELKOMSEL MENGGUNAKAN METODE CLUSTERING DENGAN RFM MODEL DAN ALGORITMA K-MEANS TELKOMSEL CUSTOMER SEGMENTATION USING CLUSTERING METHOD WITH RFM MODEL AND K-MEANS ALGORITHM,” 2021.
- [15] A. Saputra dan R. Yusuf, “Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, hlm. 1346–1361, Jul 2024, doi: 10.57152/malcom.v4i4.1516.
- [16] K. Amrulloh, T. Hendro Pudjiantoro, P. Nurul Sabrina, dan A. Id Hadiana, “Comparison Between Davies-Bouldin Index and Silhouette Coefficient Evaluation Methods in Retail Store Sales Transaction Data Clusterization Using K-Medoids Algorithm,” 2022. [Daring]. Tersedia pada: www.kaggle.com,
- [17] Y. Hasan, “Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan Dbscan,” 2024.

- [18] I. Firman Ashari, E. Dwi Nugroho, R. Baraku, I. N. Yanda, dan R. Liwardana, “Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta,” 2023. [Daring]. Tersedia pada: <http://jurnal.polibatam.ac.id/index.php/JAIC>