

Implementasi K-Means Untuk Pengelompokan Makanan Cepat Saji Bagi Penderita Penyakit Obesitas

Yoannes Dion Pradvenanta, Rastri Prathivi*

Fakultas Teknologi Informasi Dan Komunikasi, Teknik Informatika, Universitas Semarang, Semarang, Indonesia

Email: ¹yoannesdion11@gmail.com, ²vivi@usm.ac.id

Email Penulis Korespondensi: vivi@usm.ac.id

Submitted: 04/06/2024; Accepted: 23/06/2024; Published: 23/06/2024

Abstrak—Satu dari delapan orang di dunia hidup dengan obesitas, sebuah statistik yang memprihatinkan karena menunjukkan peningkatan yang signifikan dibandingkan tahun 1990. Obesitas pada orang dewasa telah lebih dari dua kali lipat, dan obesitas pada orang remaja dan anak-anak telah meningkat empat kali lipat. Salah satu faktor obesitas kualitas makanan kurang baik. Banyak masyarakat yang kurang memperhatikan kualitas makanan mereka salah satunya adalah konsumsi makanan cepat saji dikarenakan Konsumsi makanan cepat saji dapat dikatakan baik jika frekuensi makanannya 1 kali dalam seminggu, jika lebih dari itu dan berlebihan dikatakan tidak baik. Dengan demikian, perlu adanya model pengelompokan makanan cepat saji yang membantu penderita obesitas dalam memilih makanan cepat saji. Algoritma K-means merupakan salah satu model yang ideal untuk pengelompokan makanan cepat saji. Penerapan algoritma K-Means untuk mengelompokkan makanan cepat saji berdasarkan nutrisinya. Hasil analisis menggunakan elbow method menunjukkan nilai $k=5$, kemudian mempertimbang 3 evaluasi terhadap nilai $k=5$ yaitu Sum Square Error (SSE), Silhouette Score, dan Davies Bouldin Index (DBI). Hasil tersebut data disegmentasi dengan mempertimbangkan kandungan nutrisi yang negatif dan positif bagi penderita obesitas. Hasil segmentasi data didapati *cluster* yang cukup sehat pada label_0 dengan 244 data dan *cluster* yang tidak sehat pada label_2 dengan 25 data. Dari *cluster* label_0 244 data dapat menjadi pilihan makanan cepat saji yang cukup sehat bagi penderita obesitas.

Kata Kunci: K-Means; Klasterisasi; Makanan Cepat Saji; Obesitas; Segmentasi Data

Abstract—One in eight people in the world lives with obesity, a statistic that is worrying as it shows a significant increase compared to 1990. Obesity in adults has more than doubled, and obesity among adolescents and children has quadrupled. One factor in obesity is poor food quality. A lot of people who don't pay much attention to the quality of their food one of them is eating fast food because fast food consumption can be said to be good if the meal frequency is 1 time a week, if more than that and excess is said not good. Thus, there is a need for a fast food grouping model that helps obese people choose fast foods. The K-means algorithm is one of the ideal models for grouping fast foods. The results of the analysis using the elbow method show $k=5$, then consider three evaluations against the $k=5$ value: Sum Square Error (SSE), Silhouette Score, and Davies Bouldin Index (DBI). The results were data segmented taking into account the negative and positive nutrient content for obese patients. The data segmentation results found a fairly healthy cluster on label_0 with 244 data and an unhealthy cluster in label_2 with 25 data. From the cluster label_0, 244 of the data could be a healthy fast food choice for obesity patients.

Keywords: K-Means; Clustering; Fast Food; Obesity; Data Segmentation

1. PENDAHULUAN

Obesitas adalah kondisi di mana tubuh memiliki banyak lemak baik secara keseluruhan maupun sebagian [1]. Pada tahun 2022, satu dari delapan orang di dunia hidup dengan obesitas, sebuah statistik yang memprihatinkan karena menunjukkan peningkatan yang signifikan dibandingkan tahun 1990. Obesitas pada orang dewasa telah lebih dari dua kali lipat, dan obesitas pada orang remaja telah meningkat empat kali lipat. Data menunjukkan bahwa pada orang dewasa (18 tahun ke atas) 890 juta mengalami obesitas di tahun 2022. Keadaan ini tidak hanya menimpa orang dewasa, namun juga pada anak-anak dan remaja (5-19 tahun) dengan 160 juta mengalami obesitas. Dampak dari obesitas ini pun tidak bisa diabaikan. Pada tahun 2019, hampir 5 juta kematian akibat Penyakit Tidak Menular (PTM) seperti penyakit kardiovaskular, diabetes, kanker, gangguan neurologis, penyakit pernapasan kronis, dan gangguan pencernaan disebabkan oleh *Body Mass Index (BMI)* yang lebih tinggi dari optimal [2]. Data tersebut menunjukkan bahwa obesitas telah menjadi masalah kesehatan global yang serius. Genetika, kurangnya aktivitas fisik, dan kualitas makan adalah beberapa penyebab obesitas[3]. Banyak masyarakat yang kurang memperhatikan kualitas makanan mereka Salah satunya adalah mengonsumsi terlalu banyak makanan cepat saji.

Makanan cepat saji seringkali menjadi pilihan mudah dikarenakan penyajiannya yang cepat sehingga dapat menghemat waktu dan juga harga yang terjangkau. Namun, di balik kepraktisannya, makanan cepat saji umumnya memiliki total energi, kolesterol, karbohidrat, kalori, lemak, gula, natrium yang tinggi dan rendah serat serta vitamin [4][5]. Banyaknya lemak yang terkandung pada makanan cepat saji yang sebagiannya terakumulasi dalam tubuh menyebabkan obesitas [6]. Konsumsi makan makanan cepat saji setidaknya sekali seminggu dapat dikatakan baik. Namun, jika makan lebih dari sekali dalam seminggu dapat dikatakan itu tidak baik. dan berlebihan dikatakan tidak baik yang akan berdampak buruk bagi kesehatan seperti obesitas [7][8]. Kandungan nutrisi dalam makanan cepat saji memainkan peran penting dalam perkembangan obesitas.

Oleh karena itu, perlu adanya model pengelompokan makanan yang tepat dapat membantu penderita obesitas dalam memilih makanan yang cukup sehat pada makanan cepat saji. Algoritma K-means adalah salah satu model yang ideal untuk pengelompokan makanan cepat saji. Algoritma ini dapat mengelompokkan makanan berdasarkan

kandungan nutrisi pada makanan cepat saji, sehingga memudahkan penderita obesitas dalam memilih makanan cepat saji.

Beberapa penelitian sebelumnya dilakukan di berbagai bidang terkait penerapan dan penggunaan algoritma k-means. Penelitian yang dilakukan oleh Intan Saleha Tinendung dan Ilka Zufria menggunakan algoritma k-means untuk mengelompokkan status stunting pada anak-anak, karena dampak stunting pada anak dapat menyebabkan banyak gangguan kesehatan lainnya. Hasil penelitian menunjukkan bahwa tiga cluster terdiri dari 31 anak dengan status stunting, 43 anak dengan status normal, dan 27 anak dengan status pertumbuhan cepat [9]. Pada penelitian sebelumnya yang lain oleh Ragil Satria Wicaksana, dkk melakukan penelitian dengan menggunakan algoritma k-means untuk menentukan faktor motivasional yang mendorong pengunjung untuk hadir di JHF#2. Metode Elbow membagi cluster dengan 25 variabel. Hasil penelitian menunjukkan bahwa terdapat tiga kelompok motif yang berbeda yakni edukasi, rekreasi, dan transaksional [10]. Penelitian oleh Pita Sari, dkk melakukan penelitian dengan menggunakan algoritma k-means untuk menentukan kelompok penduduk miskin yang layak menerima bantuan karena banyak kelompok penerimaan bantuan yang salah sasaran. Hasil penelitian menunjukkan *cluster* yang tepat berjumlah 5 *cluster* dan nilai *Sum of Square Error* yaitu 30.00-10.000. Analisis dari penelitian tersebut terdapat kecamatan Pagar Alam Selatan dan Dempo Selatan yang harus di prioritaskan dalam penargetan program bantuan di Kota Pagar Alam [11].

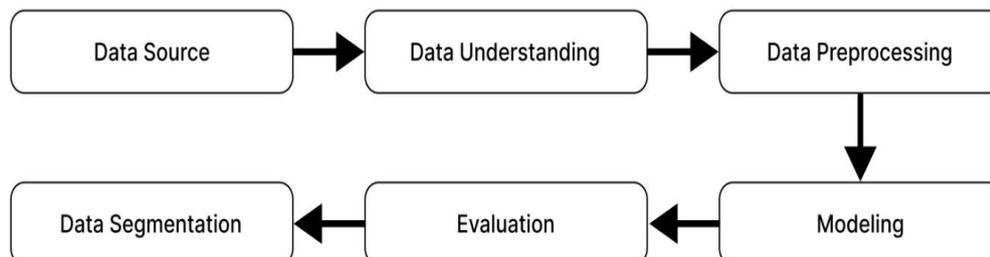
Penelitian oleh Muhammad Yamin Nurzaman dan Betha Nurina Sari yang melakukan penelitian menggunakan algoritma k-means untuk mengetahui jumlah para petani yang memadai diakarenakan provinsi Jawa Barat memiliki wilayah pertanian yang besar, sehingga memiliki potensi yang sangat besar untuk menghasilkan jumlah produksi pertanian yang besar. Hasil penelitian menunjukkan 2 *cluster* yaitu Pada *cluster* 1 memiliki 9743 data dan *cluster* 2 memiliki 881 data. Pada penelitian tersebut untuk menentukan K *cluster* terbaik menggunakan *Elbow Method* dengan melihat juga evaluasi *Silhouette Coefficient* [12]. Penelitian oleh Aditia Yudhistira dan Rio Andika menggunakan algoritma k-means untuk mengidentifikasi dan membentuk kelompok data siswa berdasarkan nilai akademik, nilai sikap, dan nilai disiplin yang bertujuan untuk meningkatkan nilai siswa selama proses pembelajaran selanjutnya. Penelitian tersebut mengevaluasi jumlah cluster menggunakan *silhouette coefficient* dan menemukan bahwa tiga cluster dengan nilai *silhouette coefficient* 0.489 adalah yang terbaik. Metode Elbow digunakan untuk menghitung nilai *Sum Square Error* (SSE) dalam penentuan *cluster*[13].

Penelitian ini menggunakan algoritma K-Means untuk mengelompokkan makanan cepat saji berdasarkan nutrisi yang ada dalam makanan cepat saji. Pada penelitian ini untuk menentukan jumlah *cluster* optimal menggunakan *Elbow Method* dengan mempertimbangkan tiga evaluasi hasil dari *cluster* yang terbaik yaitu menggunakan *Sum Square Error* (SSE), *Silhouette Score*, dan *Davies Bouldin Index* (DBI). Setelah mendapatkan jumlah *cluster* terbaik segmentasi *cluster* makanan cepat saji yang cukup sehat. Di mana setiap *cluster* mewakili kelompok makanan cepat saji dengan kandungan nutrisi yang berbeda. Penelitian ini bertujuan untuk membantu penderita obesitas dalam memilih makanan cepat saji yang cukup sehat dengan mengelompokkannya berdasarkan kandungan nutrisi pada makan cepat saji. Dengan memahami kelompok makanan cepat saji yang sesuai dengan kandungan nutrisi mereka, penderita obesitas dapat membuat pilihan yang lebih bijak saat memesan makanan cepat saji dan mengelola berat badan mereka dengan lebih baik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian adalah serangkaian langkah sistematis untuk melakukan penelitian. Tahapan penelitian dilakukan untuk mempermudah proses penelitian supaya hasil penelitian dapat tersusun secara baik. Pada gambar 1 menunjukkan tahapan penelitian.



Gambar 1. Tahapan Penelitian

Pada Gambar 1 dapat dilihat tahapan penelitian ini, dimana pada gambar tersebut tahapan penelitian ini dimulai dari *Data Source* pengambilan dataset untuk melakukan penelitian. Kedua, *Data Understanding* yang dimana memahami dan mengidentifikasi dataset untuk menentukan model. Ketiga, *Data preproccesing* mengubah dataset untuk dianalisis dan dilakukan *modeling*. Keempat, *Modeling* yang menerapkan model algoritma K-Means dari dataset yang sudah siap. Kelima, *Evaluation* melakukan evaluasi terhadap *cluster* yang sudah ditentukan. Terakhir dari tahapan penelitian *Data Segmentation* dengan menentukan *cluster* mana yang baik dari semua *cluster* yang sudah ada.

a. *Data Source*

Pada penelitian ini, dataset didapatkan dari website Kaggle. Nama dari dataset penelitian ini adalah *Fastfood Nutrition*. Dataset ini berisi informasi tentang kandungan nutrisi dari berbagai jenis makanan cepat saji yang populer di seluruh dunia. Dalam dataset ini, terdapat data sebanyak 515 baris dan 17 atribut.

b. *Data Understanding*

Data Understanding berarti mempelajari data yang akan digunakan dan menemukan subset yang menarik untuk membuat pernyataan sementara [14]. Dalam tahap *data understanding* ini sebagai berikut:

1. *Identification Dataset*

Mengidentifikasi dataset dengan memahami setiap atribut yang ada. Memeriksa jenis data yang dimiliki setiap atribut seperti numerik (bilangan) atau kategorikal (kategori).

2. *Missing Value Identification*

identifikasi keberadaan *missing value* (nilai hilang) data yang tidak tersedia dalam dataset. Mengidentifikasi *missing value* setiap baris pada dataset.

3. *Duplicated Data Identification*

Identifikasi data duplikat pada baris data dalam dataset yang memiliki nilai yang sama pada semua atribut. Mengidentifikasi data duplikat setiap baris pada dataset

c. *Data preprocessing*

Data Preprocessing adalah diprosesnya data untuk menyajikan data sehingga mesin dapat dengan mudah memahaminya [15]. Dalam proses *Data preprocessing* ada beberapa tahap seperti

1. Pemilihan atribut yang sesuai untuk modeling menggunakan k-means.

2. Mengisi data yang *missing value* pada dataset menggunakan *mean*.

3. Menghapus data yang duplikasi pada dataset

4. Mengecek atribut yang mengindikasikan *outlier* dan menormalisasikan data

d. *Modeling*

Tahap pemodelan menentukan dan menerapkan metode data mining yaitu algoritma yang akan digunakan dengan menentukan nilai parameter yang optimal [16]. Dalam tahap modeling ini, mengimplementasikan algoritma k-means. K-Means merupakan algoritma klusterisasi yang menggunakan metode partisi untuk membagi objek pengamatan ke dalam kelompok. Metode ini berusaha membagi data kedalam kelompok titik pusat (centroid) [17]. Dasar dalam menghitung algoritma k-means sebagai berikut [18]:

1. Tentukan jumlah *cluster* (k) yang diinginkan

2. Memilih centroid *cluster* secara random

3. Hitung jarak setiap data terhadap centroid menggunakan rumus Euclidean yaitu:

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

$d(x,y)$ melambangkan jarak antara dua titik data, di mana x merepresentasikan titik data pada suatu objek dan y mewakili titik pusat. Jumlah objek dalam konteks ini dilambangkan dengan i .

4. Menggabungkan semua data menjadi kelompok berdasarkan jarak minimum antara centroid dan data untuk memperoleh *cluster* baru.

5. Menghitung kembali centroid (pusat *cluster* baru) untuk setiap *cluster*. Rumusnya adalah:

$$CI = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x} \quad (2)$$

CI merujuk pada titik pusat baru dari sebuah cluster, di mana x_1 hingga x_n merupakan nilai-nilai yang mewakili cluster ke-1 hingga cluster ke-n. Total data keseluruhan dilambangkan dengan x .

6. Setelah didapatkan centroid baru, melakukan iterasi selanjutnya dengan mengulangi langkah 3 sampai 5 sampai tidak ada lagi data yang dipindahkan dari *cluster*.

Sebelum mengimplementasikan algoritma k-means perlu adanya penentuan *cluster* yang sesuai. Dalam menentukan *cluster* yang sesuai menggunakan *Elbow Method*. Metode Elbow dapat digunakan untuk menguji performa tingkat konsistensi jumlah cluster yang tepat. Ditandai dengan lekukan pada grafik dengan kriteria siku [19].

e. *Evaluation*

Evaluation adalah suatu proses untuk melakukan pengukuran hasil evaluasi dari model yang telah diimplementasikan sebelumnya di tahap modeling [20]. Pada tahap *evaluation* ini, melakukan evaluasi tingkat keberhasilan dan keakuratan pada *cluster* dari proses modeling yang sudah dilakukan. Beberapa cara untuk melakukan evaluasi yaitu *Sum Square Error (SSE)*, *Silhouette Score*, dan *Davies Bouldin Index (DBI)*.

f. *Data Segmentation*

Data segmentation pada penelitian ini adalah menentukan *cluster* makanan cepat saji mana yang cukup sehat untuk dimakan. Dalam menentukan *cluster* makanan cepat saji yang cukup sehat yaitu dengan melihat rata-rata seluruh nutrisi dari setiap *cluster*. Pada setiap *cluster* memiliki kandungan nutrisi yang positif dan negatif. Kandungan nutrisi yang positif adalah kandungan nutrisi yang baik dan sehat bagi kesehatan untuk penderita penyakit obesitas. Kandungan nutrisi yang negatif adalah kandungan nutrisi yang memperburuk dan membahayakan kesehatan bagi

penderita penyakit obesitas. *Cluster* dibandingkan satu sama lain dengan memperhatikan kandungan nutrisi yang positif dan negatif.

3. HASIL DAN PEMBAHASAN

3.1 Data Source

Data Source dalam penelitian ini membaca dataset *Fastfood Nutrition*. Pada Tabel 1 penjelasan mengenai setiap atribut pada dataset.

Tabel 1. Penjelasan Atribut

Atribut	Penjelasan
restaurant	Nama restoran tempat makanan cepat saji
item	Nama item makanan cepat saji
calories	Total kandungan kalori makanan cepat saji
cal_fat	Kalori dari lemak makanan cepat saji.
total_fat	Kandungan lemak total makanan cepat saji
sat_fat	Kandungan lemak jenuh makanan cepat saji
trans_fat	Kandungan lemak trans makanan cepat saji.
cholesterol	Kandungan kolesterol makanan cepat saji
sodium	Kandungan natrium makanan cepat saji
total_carb	Total kandungan karbohidrat makanan cepat saji
fiber	Kandungan serat makanan cepat saji
sugar	Kandungan gula makanan cepat saji
protein	Kandungan protein makanan cepat saji
vit_a	Kandungan vitamin A makanan cepat saji
vit_c	Kandungan vitamin C makanan cepat saji
calsium	Kandungan kalsium makanan cepat saji
salad	(Tidak ada informasi terkait)

Pada Tabel 1 dapat dilihat 16 terdapat informasi yang jelas mengenai atributnya dan 1 atribut tidak terdapat informasi yang jelas yaitu atribut salad. Setelah mengetahui penjelasan atribut pada dataset kemudian melanjutkan pada tahap *data understanding* untuk dianalisis lebih lanjut.

3.2 Data Understanding

Dalam *Data understanding* ini, memahami dataset makanan cepat saji untuk menganalisis temuan penting yang berguna untuk mengimplementasikan algoritma k-means yang berpotensi untuk mengelompokkan makanan cepat saji bagi penderita penyakit obesitas. Cara untuk menganalisis dataset sebagai berikut:

a. Analisis Dataset

Tabel 2. Informasi Dataset

Column	Non-Null Count	Dtype
restaurant	515 non-null	object
item	515 non-null	object
calories	515 non-null	int64
cal_fat	515 non-null	int64
total_fat	515 non-null	int64
sat_fat	515 non-null	float64
trans_fat	515 non-null	float64
cholesterol	515 non-null	int64
sodium	515 non-null	int64
total_carb	515 non-null	int64
fiber	503 non-null	float64
sugar	515 non-null	int64
protein	514 non-null	float64
vit_a	301 non-null	float64
vit_c	305 non-null	float64
calsium	305 non-null	float64
salad	515 non-null	object

Pada Tabel 2 dapat diidentifikasi bahwa terdapat 17 atribut dan 515 baris data. Dari semua atribut sudah sesuai dengan tipe data tersebut. Terdapat 3 atribut kategorikal dan 14 atribut numerik. Atribut kategorikal yaitu restaurant, item, dan salad. Atribut numerik yaitu calories, cal_fat, total_fat, sat_fat, trans_fat, kolesterol, sodium, total_carb, fiber, sugar, protein, vit_a, vit_c, dan calcium. Dataset ini memiliki beberapa atribut penting untuk mengelompokkan makanan cepat saji bagi penderita penyakit obesitas yaitu fiber, protein, vit_a, vit_c, dan calcium. Atribut yang tidak penting yaitu salad dikarenakan nilai semua data hanya 'other'.

b. Check Missing Value

Untuk memastikan kualitas data dan menghindari bias dalam analisis, penting untuk memeriksa nilai yang hilang dalam dataset. Jumlah missing value Terdapat 5 atribut yaitu fiber dengan 12, protein dengan 1, vit_a dengan 214, vit_c dengan 210, dan calcium dengan 210. Nilai yang hilang akan diatasi pada tahap preprocessing

c. Check Duplicate Data

Keberadaan data duplikat dalam dataset dapat menyebabkan redundansi dan berpotensi memengaruhi pemodelan algoritma k-means. Terdapat 2 baris data yang duplikat baris yang memiliki nilai identik di semua atribut yang akan dihapus pada tahap preprocessing.

3.3 Data preprocessing

Sebelum melakukan pengelompokan data makanan cepat saji menggunakan algoritma k-means, perlu dilakukan beberapa langkah *preprocessing* data untuk memastikan data siap untuk dianalisis. Langkah-langkah *data preprocessing* yang dilakukan dalam penelitian ini yaitu:

a. Pemilihan Atribut

Dari *data understanding* terdapat 1 atribut yang tidak penting dikarenakan tidak informasi terkait atribut tersebut yang berguna untuk penerapan algoritma k-means yaitu salad. Oleh karena itu, Atribut salad dihapus dari dataset.

b. Isi Missing Value

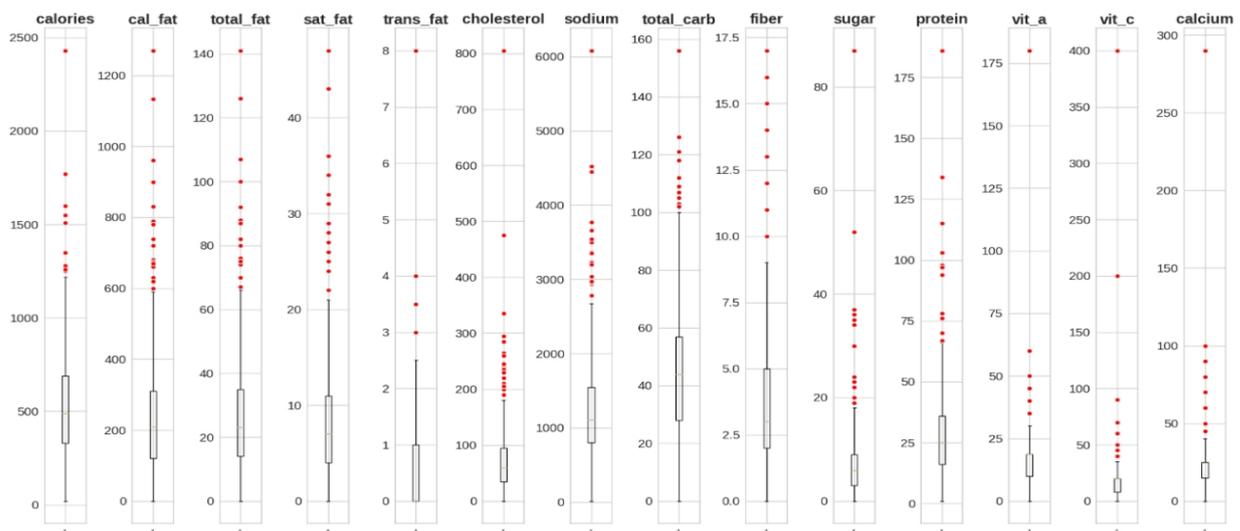
Langkah selanjutnya adalah mengisi nilai yang hilang (*missing value*) dalam data. Dalam penelitian ini, nilai yang hilang seperti fiber, protein, vit_a, vit_c, dan calcium diisi dengan nilai rata-rata dari atribut yang sama.

c. Drop Duplicate Data

Langkah selanjutnya adalah menghapus data duplikat. Dalam penelitian ini, hanya 2 data yang duplikat dihapus karena dapat mempengaruhi pemodelan.

d. Identifikasi Outlier dan menormalisasikan data

Langkah selanjutnya adalah memeriksa outlier dalam dataset dan menormalisasikan. Outlier adalah data yang jauh berbeda dari nilai-nilai lain dalam data.



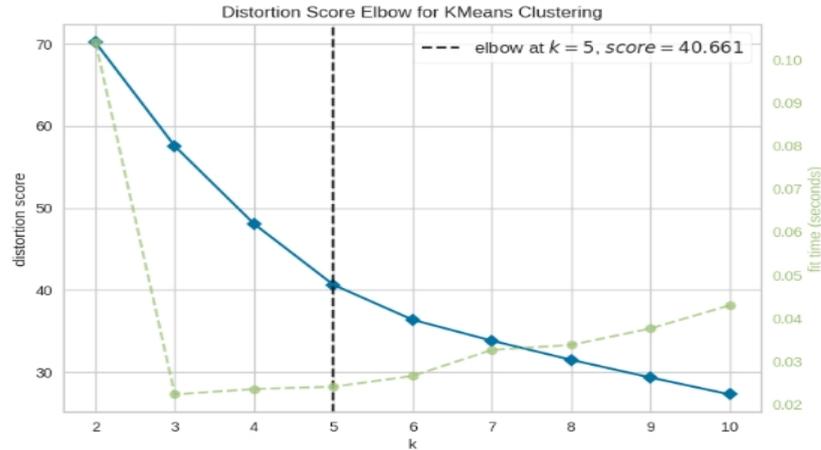
Gambar 2. Outlier Pada Setiap Atribut

Dalam penelitian ini, outlier diidentifikasi menggunakan metode *Interquartile Range (IQR)*. Pada Gambar 2 Terdapat semua atribut ada outlier dengan jumlah calories 9 data, cal_fat 21 data, total_fat 18 data, sat_fat 24 data, trans_fat 14 data, kolesterol 23 data, sodium 18 data, total_carb 14 data, fiber 46 data, sugar 14, protein 12 data, vit_a 35 data, vit_c 70 data, dan calcium 44 data. Jumlah index yang outlier pada dataset adalah 149 data. Dalam penelitian ini outlier tidak dihapus dikarenakan menghilangkan hampir sebagian besar dataset dapat mengganggu karakteristik makanan cepat saji yang sebenarnya. Hal ini dapat menyebabkan kesimpulan yang tidak akurat dan berpotensi menyesatkan penderita penyakit obesitas mengenai nutrisi pilihan makanan cepat saji. Oleh sebab itu,

penanganannya yaitu dengan menormalisasikan data. Normalisasi data adalah proses mengubah skala data sehingga semua fitur memiliki kontribusi yang seimbang terhadap algoritma k-means. Dalam penelitian ini, normalisasi data dilakukan menggunakan metode *Min-Max Scaling* sehingga semua nilai berada dalam rentang [0, 1] atau [-1, 1]

3.4 Modeling

Setelah melakukan semua langkah *data preprossesing*, dataset siap untuk pemodelan menggunakan algoritma K-Means. Dalam penentuan jumlah *cluster* yang optimal untuk pengelompokan makanan cepat saji menggunakan *elbow method*. Pada Gambar 3 merupakan grafik hasil penentuan jumlah cluster pada penelitian ini



Gambar 3. Garfik Elbow Method

Pada Gambar 3 dapat dilihat grafik dari *elbow method cluster* yang optimal menunjukkan pada k=5, Maka digunakan k=5 sebagai nilai k yang optimal untuk melakukan pengolahan data dengan implementasi algoritma K-Means.

```

0s ▶ model = KMeans(n_clusters=5,init='k-means++',random_state=42)
    model.fit(df_mod)
    
```

KMeans

KMeans(n_clusters=5, random_state=42)

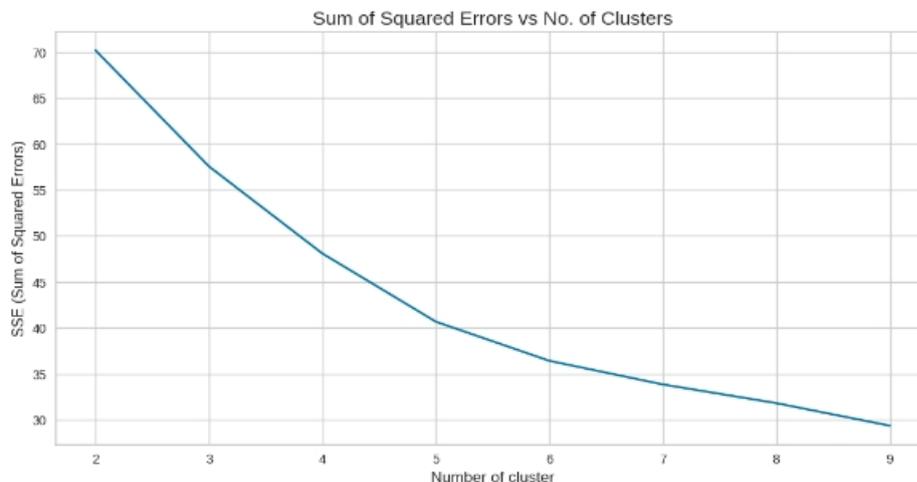
Gambar 4. Penerapan Algoritma K-Means

Pada Gambar 4 melakukan *clustering* menggunakan algoritma k-means dengan k=5. Kemudian menetapkan label *Cluster* yang dihasilkan oleh algoritma K-Means ke variabel label.

3.5 Evaluation

Pada bagian ini, mengevaluasi *cluster* yang optimal yang dihasilkan oleh *elbow method* sebagai berikut:

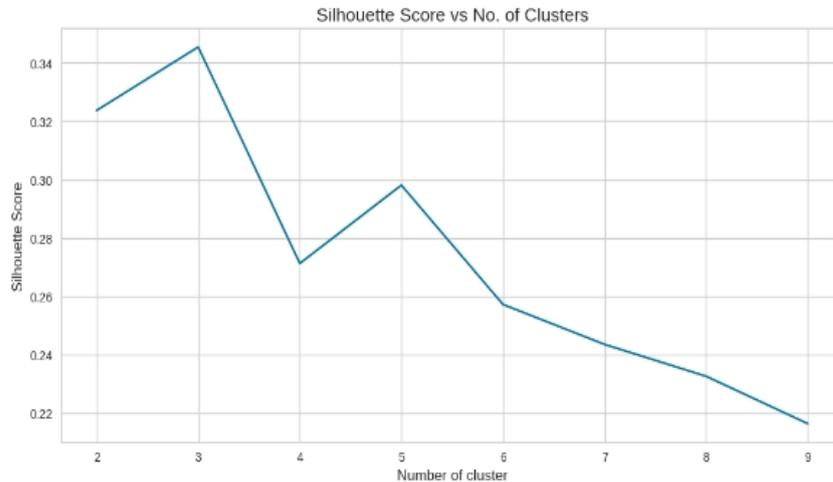
- a. *Sum of Squared Errors (SSE)*



Gambar 5. Grafik nilai *Sum of Squared Errors*

Sum of Squared Errors mengukur total jarak kuadrat antara setiap titik data dan centroid kelompoknya. Nilai *Sum of Squared Errors* yang lebih rendah menunjukkan pengelompokan yang lebih baik, yang berarti bahwa titik data dalam kelompok tertentu lebih dekat satu sama lain dan lebih jauh dari kelompok lain. Pada Gambar 5 dapat dilihat grafiknya nilai *Sum of Squared Errors* yang rendah pada cluster 5.

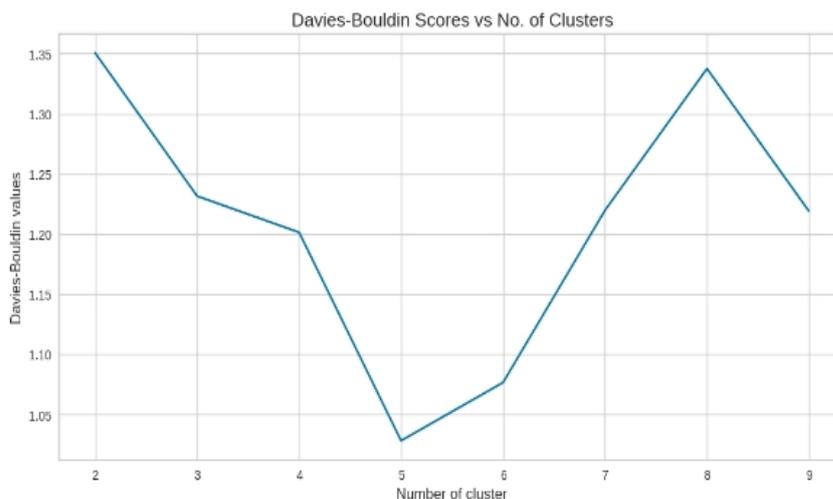
b. *Silhouette score*



Gambar 6. Grafik Nilai *Silhouette score*

Silhouette score mengukur tingkat kemiripan setiap titik data dengan kelompoknya sendiri dibandingkan dengan kelompok terdekatnya. Nilai *silhouette score* yang lebih tinggi menunjukkan pengelompokan yang lebih baik, di mana titik-titik data lebih dekat dengan centroid kelompoknya dan lebih jauh dari centroid kelompok lain. Pada Gambar 6 grafik *Silhouette score* dapat dilihat *cluster* yang nilainya cukup tinggi pada cluster 3 dan 5.

c. *Davies Bouldin Index (DBI)*



Gambar 7. Grafik Nilai Davies Bouldin Index

Davies Bouldin Index merupakan metrik yang mengukur rasio antara diameter intra-cluster (jarak maksimum antar titik data dalam kelompok) dan diameter inter-cluster (jarak minimum antar centroid kelompok). Nilai *Davies Bouldin Index* yang lebih rendah menunjukkan pengelompokan yang lebih baik. Pada Gambar 7 dapat dilihat grafik *cluster* yang nilai *Davies Bouldin Index* paling rendah pada *cluster* ke 5.

3.6 Data Segmentation

Pada tahap *data segementation*, menentukan *cluster* mana yang cukup sehat untuk pilihan makanan cepat saji bagi penderita penyakit obesitas. Diketahui hasil dari *clustering* pada algoritma k-means pada dataset dibagi menjadi 5 *cluster*, dimana setiap data memiliki cluster masing-masing. Pada dataset terdapat kolom labels dimana kolom tersebut hasil pembagian *clustering* algoritma k-means. Pada *Data segmentation* yang pertama memfilter data berdasarkan nilai pada kolom labels dan menghitung jumlah baris data yang memenuhi kriteria filter.

```

0s [54] label_0 = df[df['labels']==0]
      label_1 = df[df['labels']==1]
      label_2 = df[df['labels']==2]
      label_3 = df[df['labels']==3]
      label_4 = df[df['labels']==4]
      print("label 0",len(label_0))
      print("label 1",len(label_1))
      print("label 2",len(label_2))
      print("label 3",len(label_3))
      print("label 4",len(label_4))

label 0 244
label 1 175
label 2 25
label 3 9
label 4 60
    
```

Gambar 8. Jumlah Data Pada Setiap Label

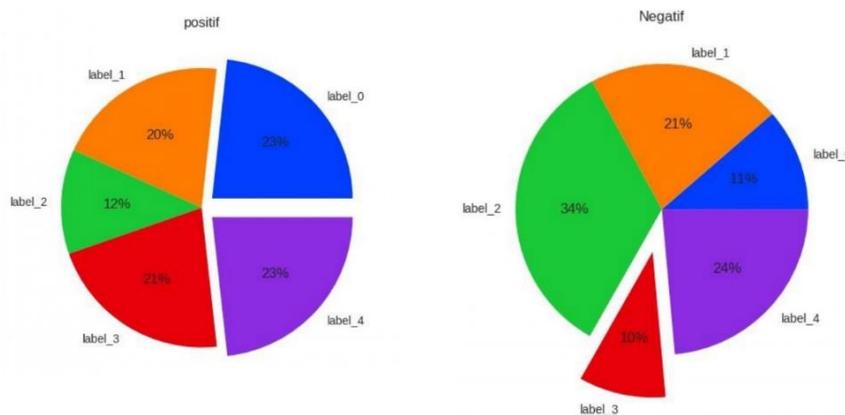
Pada Gambar 8 dapat dilihat setiap labels terdapat label_0 244 data, label_1 175 data, label_2 25 data, label_3 9 data, dan label_4 60 data. Selanjutnya Memisahkan kandungan nutrisi yang baik (positif) dan tidak baik (negatif) untuk kesehatan penderita penyakit obesitas.

```

0s [54] positive_ntr = ['protein','calcium','vit_a','vit_c','fiber']
      negative_ntr = ['sugar','calories','total_carb','total_fat','sodium',
                    'cholesterol','cal_fat','sat_fat','trans_fat']
    
```

Gambar 9. Pemisahan Kandungan Nutrisi Positif Dan Negatif

Pada Gambar 9 kandungan nutrisi yang positif yaitu protein, calcium, vit_a, vit_c, dan fiber. Kandungan nutrisi yang Negatif yaitu sugar, calories, total_carb, total_fat, sodium, kolesterol, cal_fat, sat_fat, dan trans_fat. Berikutnya pembuatan dictionary untuk setiap label dengan rata-rata nutrisi positif dan negatif. Hasil nilai rata-rata atribut nutrisi positif dan negatif untuk setiap label (0, 1, 2, 3, dan 4) divisualisasikan dengan piechart



Gambar 10. Persentase Kandungan Nutrisi Positif Dan Negatif Pada Setiap Label

Pada gambar 10 dapat dilihat kandungan nutrisi positif label_0 23%, label_1 20%, label_2 12%, label_3 21%, dan label_4 23%. Kandungan nutrisi yang positif paling tinggi pada label_0 dan labels_4, sedangkan yang paling rendah pada label_2. kandungan nutrisi negatif label_0 11%, label_1 21%, label_2 34%, label_3 10%, dan label_4 24%. Kandungan nutrisi yang negatif paling tinggi pada label_2, sedangkan yang paling rendah pada label_3. Selanjutnya menentukan labels yang cukup sehat untuk direkomendasikan sebagai pilihan makanan cepat saji bagi penderita penyakit obesitas.

```

0s [63] # label_name = (positive,negative)
      label_0_pn = (23,11)
      label_1_pn = (20,21)
      label_2_pn = (12,34)
      label_3_pn = (21,10)
      label_4_pn = (23,24)
      labels_pos_neg = [label_0_pn,label_1_pn,label_2_pn,label_3_pn,label_4_pn]

0s [64] healthy_level = [pos-neg for pos,neg in labels_pos_neg]
      healthy_level

[12, -1, -22, 11, -1]
    
```

Gambar 11. Pemilihan Label Yang Cukup Sehat Untuk Dimakan

Pada Gambar 11 diatas menunjukkan hasil pemilihan labels yang direkomendasikan adalah label_0 dikarenakan kandungan nutrisi positif lebih banyak dibandingkan label_1, label_2, label_3, dan label_4. Untuk makanan cepat saji tidak sehat terdapat pada label_2 karena kandungan nutrisi negatif lebih banyak.

4. KESIMPULAN

Dari semua tahapan penelitian yang dilakukan dapat ditarik kesimpulan, Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan makanan cepat saji berdasarkan nutrisinya yaitu dengan atribut calories, cal_fat, total_fat, sat_fat, trans_fat, kolesterol, sodium, total_carb, fiber, sugar, protein, vit_a, vit_c, dan calsium. Hasil analisis dengan menggunakan *elbow method* menunjukkan bahwa nilai $k = 5$ menghasilkan pengelompokan yang optimal, kemudian mempertimbang 3 evaluasi terhadap nilai $k=5$ yaitu *Sum Square Error (SSE)*, *Silhouette Score*, dan *Davies Bouldin Index (DBI)*. Dengan nilai $k=5$ cluster yang sudah ada diterapkan algoritma k-means. Hasil cluster algoritma k-means yaitu label_0 dengan 244 data, label_1 175 data, label_2 25 data, label_3 9 data, dan label_4 60 data. Hasil cluster tersebut dilakukan segmentasi data dengan mempertimbangkan kandungan nutrisi yang positif dan negatif bagi penderita penyakit obesitas. kandungan nutrisi positif pada atribut fiber, protein, vit_a, vit_c, dan calsium. Kandungan negatif pada atribut calories, cal_fat, total_fat, sat_fat, trans_fat, kolesterol, sodium, total_carb, dan sugar. Hasil dari segmentasi data didapati cluster yang cukup sehat pada label_0 dengan 244 data dan cluster yang tidak sehat pada label_2 dengan 25 data. Dari 244 data tersebut dapat menjadi pilihan makanan cepat saji yang cukup sehat bagi penderita penyakit obesitas. Dengan memahami cluster makanan cepat saji yang sesuai dengan kandungan nutrisinya, penderita penyakit obesitas dapat menjadi pilihan yang lebih bijak saat memesan makanan cepat saji dan mengelola berat badan mereka dengan lebih baik.

REFERENCES

- [1] M. U. Zamzami Hasibuan and P. A., "Sosialisasi Penerapan Indeks Massa Tubuh (IMT) di Suta Club," *Cerdas Sifa Pendidik.*, vol. 10, no. 2, pp. 84–89, 2021, doi: 10.22437/csp.v10i2.15585.
- [2] World Health Organization, "Obesity and overweight," *www.who.int*, 2024. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> (accessed May 15, 2024).
- [3] I. Risky, S. Program, I. K. Masyarakat, and K. Masyarakat, "Literatur Review: Pengaruh Pola Makan Dan Kurangnya Aktivitas Fisik Terhadap Terjadinya Obesitas," *JK J. Kesehatan.*, vol. 1, no. 1, pp. 170–176, 2023, [Online]. Available: <https://jurnalkesehatan.joln.org/index.php/health/article/download/18/28/41>.
- [4] S. Aghnia *et al.*, "Hubungan Tingkat Pengetahuan Gizi, Status Gizi terhadap Pola Konsumsi Fast Food dan Soft Drink pada Siswa SMKN 1 Kota Tegal The Relationship of Nutritional Knowledge Level, Nutritional Status to Consumption Patterns Fast Food and Soft Drink for Students o," vol. 1, no. 1, pp. 13–18, 2023.
- [5] E. Ranggalayuni and N. Nuraini, "Faktor yang berhubungan dengan Konsumsi Makanan cepat Saji pada Mahasiswa di Institusi Kesehatan Helvetia Medan," *JUMANTIK (Jurnal Ilm. Penelit. Kesehatan)*, vol. 6, no. 3, p. 278, 2021, doi: 10.30829/jumantik.v6i3.9977.
- [6] D. Alfora, E. Saori, and L. N. Fajriah, "Pengaruh konsumsi makanan cepat saji terhadap gizi remaja," *FLORONA J. Ilm. Kesehatan.*, vol. 2, no. 1, pp. 43–49, 2023, doi: 10.55904/florona.v2i1.688.
- [7] A. Dewi, Ilma Nuria Sulrieni, and Melvia Solfia Ningsih, "Faktor - Faktor Yang Berhubungan dengan Kejadian Obesitas Pada Siswa MAN 1 Kota Padang," *SEHATMAS J. Ilm. Kesehat. Masy.*, vol. 2, no. 1, pp. 159–171, 2023, doi: 10.55123/sehatmas.v2i1.1356.
- [8] I. Pamela, "Perilaku Konsumsi Makanan Cepat Saji Pada Remaja Dan Dampaknya Bagi Kesehatan," *Ikesma*, vol. 14, no. 2, p. 144, 2018, doi: 10.19184/ikesma.v14i2.10459.
- [9] I. S. Tinendung and I. Zufria, "Pengelompokan Status Stunting Pada Anak Menggunakan Metode K-Means Clustering," *J. Media ...*, vol. 7, pp. 2014–2023, 2023, doi: 10.30865/mib.v7i4.6908.
- [10] R. S. Wicaksana, D. Heksaputra, T. A. Syah, and F. F. Nur'aini, "Pendekatan K-Means Clustering Metode Elbow Pada Analisis Motivasi Pengunjung Festival Halal JHF#2," *J. Ilm. Ekon. Islam*, vol. 9, no. 3, p. 4162, 2023, doi: 10.29040/jiei.v9i3.10591.
- [11] P. Sari, R. Syahri, and T. Informatika, "ANALISIS CLUSTERING DATA PENDUDUK MISKIN MENGGUNAKAN ALGORITMA K-MEANS," vol. 8, no. 2, pp. 2194–2199, 2024.
- [12] M. Yamin Nurzaman and B. Nurina Sari, "Implementasi K-Means Clustering Dalam Pengelompokan Banyaknya Jumlah Petani Berdasarkan Kecamatan Di Provinsi Jawa Barat," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. September, pp. 131–144, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>.
- [13] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.
- [14] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsembung," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.
- [15] D. S. Rahayu, J. Afifah, and S. Intan, "Classification of Diabetes Mellitus Using C4.5 Algorithm, Support Vector Machine (SVM) and Linear Regression Klasifikasi Penyakit Diabetes Melitus Menggunakan Algoritma C4.5, Support Vector Machine (SVM) dan Regresi Linear," *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, vol. 1, no. 1 SE-, pp. 56–63, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas/article/view/550>.
- [16] D. K. Sitingjak, B. N. Sari, and I. Maulana, "Clustering Daerah Penyumbang Sampah Berdasarkan Provinsi di Indonesia Menggunakan Algoritma K-Means," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 16, pp. 137–146, 2022, [Online]. Available: <https://doi.org/10.5281/zenodo.7059032>.



- [17] O. Meliza, T. Susanti, T. Informatika, S. Selatan, and S. L. Review, "Implementasi k-means : sebuah studi literatur," vol. x, no. x, pp. 8–11, 2024, doi: 10.12345/juri.
- [18] A. S. Wahyuni, E. Haerani, E. Budianita, and L. Afrianti, "Pemanfaatan Algoritma K-Means Dalam Menentukan Potensi Hasil Produksi Kelapa Sawit," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, pp. 0–7, 2023, doi: 10.30865/json.v5i2.7226.
- [19] Z. F. Daldiri, M. Rafly, and I. Veritawati, "Clustering Daftar Harga Rumah di Jakarta Dengan Algoritma K-Means," vol. 3, no. 2, 2022.
- [20] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.