

Implementasi Algoritma K-Means dan NLP dengan Metode Semantic Embedding dan Cosine Similarity pada Aplikasi Rekomendasi Nutrisi Makanan

Dimas Dwi Saputra, Astri Salwa Anisa, Liah Nopiansah, Yayan Hendrian*

^{1,2,3,4}Universitas Bina Sarana Informatika

e-mail: 17230718@bsi.ac.id, 17230802@bsi.ac.id, 17230731@bsi.ac.id

email korespondensi: yayan.yhn@bsi.ac.id (*)

Abstrak - Asupan nutrisi dan kalori adalah fokus utama untuk sebagian masyarakat yang ingin atau sedang melakukan program untuk menaikkan berat badan, menurunkan berat badan, serta menjaga berat badan. Kebanyakan orang lupa untuk memperhatikan kebutuhan asupan nutrisi dan kalori mereka, sehingga program yang mereka lakukan gagal atau tidak sesuai keinginan. Oleh karena itu, dibutuhkan waktu lebih untuk mencatat serta menghitung asupan nutrisi serta kalori yang dikonsumsi setiap harinya. Penelitian ini membuat program yang mampu memberikan rekomendasi makanan dengan menghitung kalori yang dibutuhkan berdasarkan perhitungan BMR. Aplikasi ini mengambil data makanan Indonesia yang sudah tersedia pada *website* milik *Kementerian Kesehatan Republik Indonesia* untuk menghitung kalori yang dibutuhkan. Dengan menggunakan Algoritma K-Means untuk mengelompokkan jenis makanan berdasarkan makanan tinggi protein, rendah kalori, dan makanan seimbang. NLP (*Natural Language Processing*) dengan model berbasis SentenceTransformer untuk memahami serta merespon masukan/input dari pengguna. Implementasi ini diharapkan mampu memberikan dampak bagi masyarakat dalam memilih makanan sesuai kebutuhan nutrisi dan kalori mereka.

Kata kunci: K-Means, NLP, Virtual Assistant

Abstract - Nutrient and calorie intake is the main focus for some people who want or are undergoing a program to gain weight, lose weight, or maintain their weight. Most people forget to pay attention to their nutrient and calorie intake needs, causing the programs they follow to fail or not meet their expectations. Therefore, more time is needed to record and calculate the nutrients and calories consumed daily. This study developed a program capable of providing food recommendations by calculating the required calories based on BMR calculations. This application uses Indonesian food data available on the website of the Ministry of Health of the Republic of Indonesia to calculate the required calories. It uses the K-Means algorithm to classify food types based on high-protein foods, low-calorie foods, and balanced foods. NLP (Natural Language Processing) with a model based on Sentence Transformer to understand and respond to user input. This implementation is expected to be able to have an impact on society in choosing foods according to their nutritional and calorie needs.

Keywords: K-Means, NLP, Virtual Assistant, Food Recommendations

PENDAHULUAN

Teknologi merupakan sebuah perkembangan pada perangkat lunak maupun perangkat keras yang didasari berdasarkan ilmu pengetahuan disertai kebutuhan manusia yang terus berkembang membantu pekerjaan manusia menjadi lebih mudah (Taufik et al., n.d.). Perkembangan teknologi membawa dampak perubahan yang sangat pesat bagi kehidupan manusia, salah satunya dengan adanya perkembangan teknologi berbasis *Artificial Intelligence* (AI). *Machine Learning* (ML) adalah cabang AI yang dapat bekerja secara mandiri tanpa harus di program ulang oleh manusia (Alfarizi M. Riziq Sirfatullah et al., 2023). Algoritma *machine learning* membutuhkan data untuk bekerja, yaitu data training untuk melatih algoritma dan data *testing* untuk menguji performannya pada data baru. Dengan memanfaatkan teknologi AI ini, komputer dapat melakukan tugas tertentu yang hampir sama dengan manusia. Salah satu tugas yang dapat memanfaatkan *machine learning* dan AI contohnya mengklasifikasikan makanan sehat berdasarkan kandungan nutrisinya sekaligus rekomendasi bagi pengguna agar dapat menyesuaikan konsumsi makanan sesuai kebutuhan. Dengan munculnya *platform* digital dan ketersediaan data dalam jumlah besar, sistem rekomendasi makanan telah menjadi alat yang kuat untuk membantu orang menemukan makanan baru dan lezat.



Pola makan sehat berarti mengonsumsi zat gizi makro dan mikro sesuai dengan kebutuhan tubuh, tidak kekurangan atau kelebihan (Cena & Calder, n.d.). Keseimbangan antara kalori yang masuk dan keluar dapat mencegah penyakit tidak menular. Kebutuhan kalori tiap orang berbeda tergantung tinggi badan, berat badan, usia, jenis kelamin, dan tingkat aktifitas. Obesitas merupakan masalah kelainan ditandai dengan adanya penimbunan yang berlebihan terhadap jaringan lemak pada tubuh, sehingga menjadi pemicu terhadap keseimbangan antara energi yang masuk dan keluar (Kurnia Saraswati et al., 2020). Untuk menentukan rekomendasi makanan bagi pengguna, maka dilakukan *clustering* sebagai metode yang mampu mengelompokkan makanan berdasarkan makanan tinggi protein, rendah kalori, dan makanan seimbang. Salah satu metode yang mampu mengelompokkan makanan tersebut yaitu K-Means.

Metode K-Means adalah metode *clustering* yang membagi data menjadi kluster berdasarkan karakteristik yang sama. Kelebihan K-Means adalah mudah diimplementasikan, efektif untuk data besar, dan dapat menangani data dengan banyak dimensi. Kekurangan K-Means adalah sensitif terhadap inisialisasi *centroid* dan perlu menentukan jumlah kluster (K) sebelumnya. *Davies Bouldin Index* (DBI) digunakan untuk mengevaluasi kualitas kluster (Sulistiyawati & Supriyanto, 2021).

Penelitian lain juga menggunakan penelitian yang sama, yaitu menggunakan K-Means (Syahri et al., 2023), penelitian tersebut mengelompokkan data dari tahun 2018 hingga 2013 dari *database* seperti *Google Scholar*. Peneliti bertujuan agar data-data yang dihasilkan mampu dikelompokkan berdasarkan pihak-pihak yang berkaitan. Penelitian yang dilakukan oleh (Dan et al., 2021) melakukan *clustering* untuk mengelompokkan penelitian – penelitian yang saling berkaitan dengan menggunakan metode K-Means. Penelitian lain juga melakukan *clustering* atau pengelompokkan data berdasarkan kesamaan karakteristik (Metode et al., 2024). Sedangkan di penelitian lainnya, peneliti menganalisis penerapan data *mining* dalam mengelompokkan jumlah data kunjungan wisatawan. Penyelesaian ini menggunakan Metode k-means yang meningkatkan perekonomian masyarakat di daerah wisata (Afiyanti et al., 2023).

Dalam penelitian ini, penulis mengembangkan *chatbot* dengan menggunakan *Natural Language Processing* (NLP), yaitu sebuah kombinasi antara ilmu komputer dan bidang kecerdasan buatan yang berkaitan dengan linguistik. Dengan adanya NLP komputer mampu memahami bahasa manusia. NLP menganalisis teks dengan menggunakan *keyword* pada teks tersebut, metode ini digunakan agar dapat membantu dalam meringkas teks dan mengidentifikasi subjek teks (Prasetyo et al., 2021).

METODE PENELITIAN

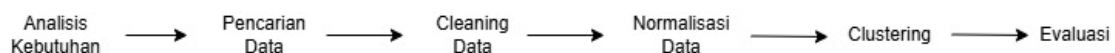
Dalam penelitian ini, dilakukan beberapa tahapan yang mencakup:

1) Tahap Pengumpulan Data

Pada tahap awal penelitian, penulis mencari studi literatur yang sesuai dengan tujuan utama penelitian ini yaitu mengembangkan sebuah aplikasi rekomendasi makanan dengan *chatbot* yang mampu merespon pertanyaan *user* dan mampu melakukan log atau catatan konsumsi user yang akan langsung dikalkulasikan dan disimpan kedalam laporan harian. Kemudian membuat aplikasi ini berkembang menjadi *Nutrition Assistant*, yaitu sistem cerdas yang mampu memberikan rekomendasi makanan, menghitung kebutuhan kalori, dan melakukan pemantauan konsumsi asupan kalori perharinya.

2) Tahap *Clustering* menggunakan algoritma K-Means

Berikut ini adalah alur pengerjaan yang dilakukan peneliti:



Sumber: penelitian penulis (2026)

Gambar 1. Alur pengerjaan menggunakan K-Means

a) Analisis Kebutuhan

Menentukan indikator yang mempengaruhi perhitungan kalori seseorang berdasarkan hasil studi literatur dengan membaca beberapa jurnal dan artikel ditemukanlah bahwa jenis kelamin, berat badan, tinggi badan, umur, dan level aktivitas menjadi *indicator*. Kemudian untuk menghitung berapa kebutuhan kalori, disini penulis menggunakan metode BMR dengan rumus yang dikemukakan oleh Mifflin-St Jeor dan juga tingkat aktivitas fisik (Science From Indonesia, 2021).

Rumus Mifflin-St Jeor:

Pria: $BMR = (10 \times \text{berat badan dalam kg}) + (6.25 \times \text{tinggi badan dalam cm}) - (5 \times \text{usia dalam tahun}) + 5$

Wanita: $BMR = (10 \times \text{berat badan dalam kg}) + (6.25 \times \text{tinggi badan dalam cm}) - (5 \times \text{usia dalam tahun}) - 161$

Table 1 Aktivitas Fisik

Jenis Aktivitas	Keterangan	Faktor Aktivitas
Tidak Aktif	Jarang berolahraga bahkan tidak sama sekali	1,2
Ringan	Berolahraga 1-3 kali/minggu	1,375
Sedang	Berolahraga 3-5 kali/minggu	1,55
Sangat Aktif	Berolahraga 6-7 kali/minggu	1,725
Ekstrim	Ekstra aktif seperti atlet	1,9

Sumber: https://123dok.com/document/7q05251y-hubungan-makan-aktivitas-fisik-dismenore-siswi-kelas-model.html#google_vignette

Setelah hasil dari perhitungan BMR telah didapatkan, maka langkah selanjutnya adalah mengalikan dengan faktor aktivitas yang ada pada tabel aktivitas fisik berdasarkan intensitas aktivitasnya.

$$\text{Target Kalori} = \text{BMR} \times \text{Faktor aktivitas.}$$

b) Pencarian Data

Aplikasi ini berfokus kepada rekomendasi makanan berdasarkan kebutuhan kalori yang dihasilkan lewat perhitungan target kalori sebelumnya dan beberapa informasi mengenai kandungan nutrisi yang ada pada setiap jenis makanan seperti Protein, Karbohidrat, serta Lemak. Informasi tersebut sangat diperlukan untuk tujuan *Bulking* dan *Cutting*. Penulis mencari informasi tersebut dan mendapatkan dataset yang diberikan oleh Kementerian Kesehatan Republik Indonesia pada website [Andrafarm TKPI\(2019\)](#), data tersebut sudah cukup untuk mewakili jenis makanan olahan dan tunggal ada di Indonesia. Data tersebut juga sudah berisikan informasi kalori, protein, karbohidrat, lemak, serat mikronutrien yang ada pada makanan-makanan tersebut.

c) Pembersihan Data (*Data Cleaning*)

Hal yang sangat penting untuk dilakukan setelah mendapatkan data informasi yang diinginkan adalah melakukan *data cleaning* atau biasa disebut juga *Pre-processing* data, yaitu melakukan serangkaian tahap yang memastikan bahwa data benar-benar bersih dari data yang berisi kosong dengan mengubah nilai [-] menjadi nilai [0].

d) Normalisasi Data

Setelah membersihkan data dari nilai yang “kotor” kemudian lakukan normalisasi data dengan menggunakan *Min-Max Normalization* (0-1 Scaling). Normalisasi untuk merubah keseluruhan nilai menjadi rentang 0 sampai dengan 1, hal ini di fungsikan untuk mengubah nilai kandungan nutrisi agar memiliki skala yang sama sehingga menghindari dominasi antara nilai yang lebih besar terhadap nilai yang lebih kecil. Normalisasi ini dilakukan terhadap atribut energi, protein, karbohidrat, serta lemak. Dengan adanya normalisasi, maka rekomendasi makanan akan lebih efektif.

Contoh: Nilai pada energi yang lebih besar dapat mempengaruhi terhadap nilai protein yang lebih kecil.

Rumus:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Keterangan:

X' = Data hasil normalisasi

X = Data yang ingin dinormalisasikan

Xmin = Data terkecil pada suatu kolom baris data

Xmax = Data terbesar pada suatu kolom baris data

e) Clustering

Algoritma yang digunakan dalam penelitian ini adalah K-Means, alasan penulis menggunakan algoritma ini adalah karena K-Means cocok untuk data yang bernilai numerik seperti pada nutrisi makanan. K-Means juga mudah untuk diinterpretasikan karena hasil yang ditampilkan berupa *centroid* yang mewakili *cluster*-nya masing-masing. Berikut langkah-langkah yang dilakukan penulis untuk *clustering* menggunakan algoritma K-Means:

1) Menentukan jumlah cluster (k)

Menggunakan metode *Elbow* untuk menentukan berapa jumlah *cluster* (k) yang tepat dengan menggunakan rumus WCSS (*Within-Cluster Sum of Squares*):

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

Keterangan:

k = jumlah *cluster*
 C_i = *cluster* ke - i
 x = titik data pada *cluster* C_i
 μ_i = *centroid*/pusat dari *cluster* C_i

2) Memberikan label *cluster*

Tahap berikutnya adalah memberikan label nama *cluster* menjadi *Bulking*, *Cutting*, dan *Maintain*. Hal ini diperlukan untuk menginterpretasikan hasil. Mengelompokkan hasil rata-rata nilai yang ada pada atribut energi, protein, karbohidrat, dan lemak kedalam label yang sudah dibuat berdasarkan makanan tinggi protein, rendah kalori, dan makanan seimbang.

3) Evaluasi

Untuk mengevaluasi apakah data sudah mirip dengan *cluster*-nya sendiri dibandingkan dengan *cluster* lain, maka penulis melanjutkan dengan menghitung nilai kualitas *clustering* yang sudah ditentukan dengan menggunakan DBI (*Davies–Bouldin Index*) semakin kecil nilai yang dihasilkan (non-negatif ≥ 0) maka dianggap semakin bagus atau valid hasilnya menggunakan rumus:

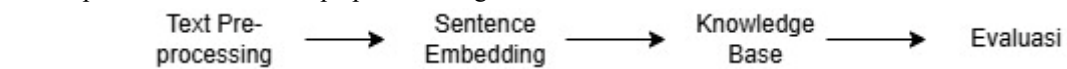
$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right)$$

Keterangan:

k = jumlah *cluster*
 S_i = rata-rata jarak antar titik data dengan *centroid* pada *cluster* i
 S_j = rata-rata jarak antar titik data dengan *centroid* pada *cluster* j

3. Tahap Analisa Intent Menggunakan NLP

Dalam tahap ini dilakukan beberapa proses sebagai berikut:



Sumber: <https://dqlab.id/nlp-machine-learning-untuk-sentiment-analysis>

Gambar 2. Tahapan analisa intent menggunakan NLP

1) *Text Pre-processing*

Input yang di terima oleh sistem akan langsung diubah menjadi *lowercase* yang artinya semua huruf yang ada akan diubah menjadi huruf kecil yang difungsikan menyelaraskan dengan yang lain.

2) *Sentence Embedding*

Pada penelitian ini penulis menggunakan metode *sentence embedding* dengan jenis *semantic* yang digunakan untuk mengubah kata menjadi bentuk *vector*. Penulis menggunakan model (*sentence-transformers/all-MiniLM-L6-v2*) yang mampu mendukung Bahasa Indonesia dan sangat cocok untuk digunakan sebagai *intent detection*.

3) *Knowledge Base*

Pada tahap ini data makanan beserta hasil vektor hasil embeddingnya disimpan kedalam database, tahap ini disebut dengan *knowledge base*. Informasi mengenai makanan mulai dari nama makanan dan nutrisi akan disimpan di database yang nantinya ketika *user* bertanya, maka sistem akan memberikan informasi yang relevan.

4) *Evaluasi*

Untuk membandingkan kemiripan sebuah kata antara *embedding* yang di *input user* dengan *embedding knowledge base* yang sudah dibuat maka dilakukan *Cosine Similarity* dengan mengukur nilai kedekatan antar kata untuk menentukan intent yang paling relevan atau bisa disebut *cosine distance* (jarak konsius).

Rumus:

$$\text{Cosine Similarity } (A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

Keterangan:

$A \cdot B$ = dot produk
 $\|A\|$ = Panjang vektor A
 $\|B\|$ = Panjang vektor B

Tabel 2. Rentang Nilai

Nilai	Keterangan
1.0	Sangat mirip
0.7 – 0.9	Mirip
0.4 – 0.6	Cukup mirip
< 0.4	Tidak mirip

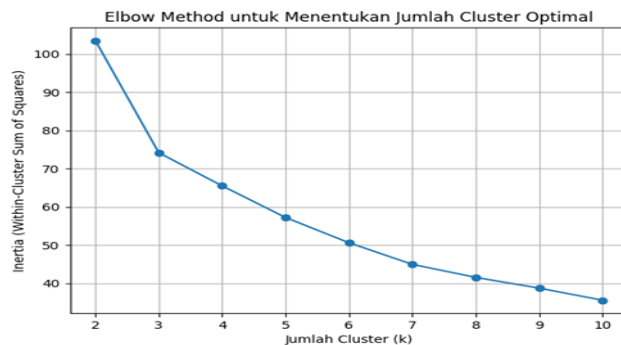
Sumber: <https://builtin.com/machine-learning/cosine-similarity>

HASIL DAN PEMBAHASAN

1. Proses Pengolahan Data

Data yang ada pada dataset penelitian ini berjumlah 1148 baris dengan 25 kolom data yang sebelumnya sudah melewati langkah *cleaning* dan juga penambahan atribut. Namun yang hanya kami gunakan hanya beberapa atribut saja seperti Id_Makanan, Nama_Makanan, Air, Energi, Protein, Lemak, Karbohidrat, Serat, Abu, Kalsium, Fosfor, Besi, Natrium, Kalium, Tembaga, Seng, Retinol, β -karoten, Karoten, Thiamin, Riboflavin, Niasin, Vitamin C, BDD, Kelompok.

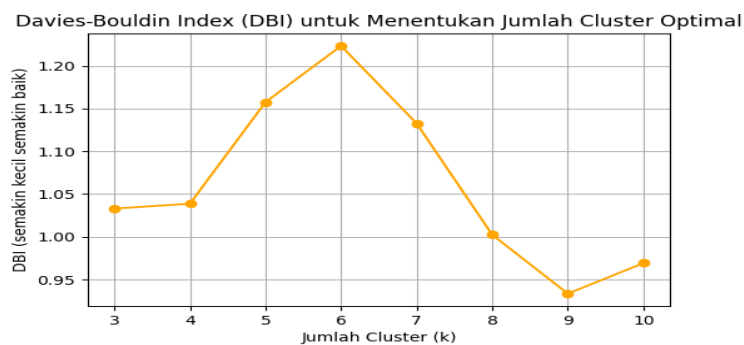
Pada tahap awal, penulis melakukan *cleaning* data dengan merubah nilai data yang kotor seperti mengubah nilai [-] menjadi nilai [0] supaya algoritma K-Means menghasilkan output yang optimal karena semua di *clustering* berdasarkan nilai numerik. Setelah dinormalisasi, selanjutnya mencari jumlah *cluster* yang tepat dengan menggunakan metode Elbow. Disini penulis melakukan normalisasi berdasarkan atribut Energi, Protein, Lemak, Karbohidrat, Serat, Vitamin C, Niasin, Riboflavin. Normalisasi ini difungsikan agar tidak ada kesenjangan yang terlampaui jauh dari atribut-atribut tersebut. (Prianto & Bunyamin, 2020)



Sumber: Hasil Penelitian (2026)

Gambar 3 Grafik Metode *Elbow*

Terlihat ada pembelokan garis pada titik 3 yang signifikan yang menandakan bahwa titik tersebut merupakan jumlah cluster yang cocok untuk penulis gunakan. Untuk memastikan apakah memang $k = 3$ yang cocok, maka dilakukan evaluasi dengan menentukan kedekatan antar data terhadap *centroid*/titik pusat *cluster* dengan menggunakan metode DBI setelah *code* K-Means. (Kadang et al., 2025)



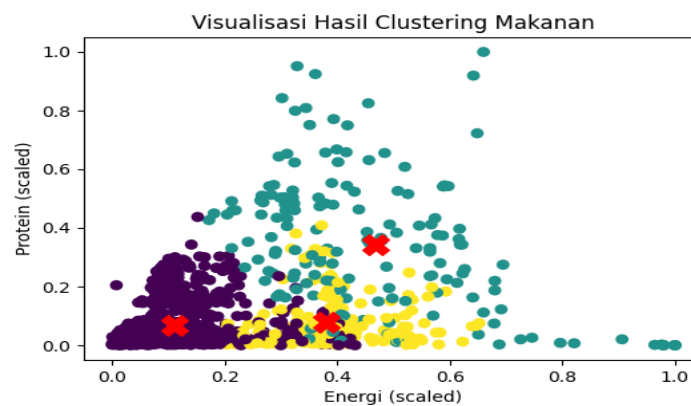
Sumber: Hasil Penelitian (2026)

Gambar 4 Grafik DBI

k = 3,	DBI = 1.0331
k = 4,	DBI = 1.0387
k = 5,	DBI = 1.1576
k = 6,	DBI = 1.2235
k = 7,	DBI = 1.1327
k = 8,	DBI = 1.0024
k = 9,	DBI = 0.9335
k = 10,	DBI = 0.9692

Sumber: Hasil Penelitian (2026)
Gambar 5 Hasil DBI

Gambar grafik yang didapatkan dengan menggunakan metode DBI terlihat bahwa titik 9 dan 10 adalah titik yang nilainya terkecil, namun pada kasus ini pada rekomendasi makanan yang paling logis adalah dengan menggunakan 3 cluster karena jika menggunakan 9 atau 10 maka akan sulit untuk diinterpretasikan dan tidak relevan terhadap program yang dibuat dan sebelumnya memang sudah ditentukan jumlah cluster yang paling cocok dengan menggunakan metode Elbow. Setelah jumlah cluster ditentukan maka langkah selanjutnya adalah menentukan atribut yang ingin dilakukan clustering dan memberikan label terhadap centroid sesuai dengan jumlah k = 3. (Gatto & Awangga, 2023)



Sumber: Hasil Penelitian (2026)
Gambar 6 Grafik K-Means hasil clustering makanan

Cluster	Energi	Protein	Lemak	Karbohidrat
0	107.902597	49.384416	20.796104	101.031169
1	426.633540	254.788820	203.130435	117.118012
2	348.778802	58.847926	70.216590	649.221198


Sumber: Hasil Penelitian (2026)
Gambar 7 Hasil Clustering K-Means

Pada gambar 6 terdapat tanda X merah yang diinterpretasikan sebagai centroid dan mengapa hanya Protein dan Energi saja yang ditampilkan pada grafik, karena agar hasil dapat ditampilkan dalam dua dimensi digunakanlah dua fitur Energi dan Protein untuk interpretasikan sumbu X dan sumbu Y. Pada centroid yang berada di kiri bawah dengan warna ungu gelap, bermakna bahwa makanan yang dekat dengan centroid tersebut adalah makanan yang rendah kalori dan protein kecil diinterpretasikan sebagai makanan untuk *Cutting*, Untuk centroid yang berada di tengah bawah dengan warna kuning bermakna bahwa makanan dengan energi sedang dan protein sedang diinterpretasikan sebagai makanan untuk *Maintain*, dan untuk centroid yang berada di tengah atas dengan warna hijau bermakna bahwa makanan dengan energi tinggi dan protein tinggi diinterpretasikan sebagai makanan untuk *Bulking*.

Sedangkan pada gambar 7 terlihat bahwa pada cluster 0 Energi bernilai paling rendah dengan protein, lemak dan karbohidrat yang rendah ini dapat diinterpretasikan bahwa cluster tersebut merupakan cluster untuk *Cutting*, untuk cluster 1 Energi bernilai paling tinggi dengan protein dan lemak sangat tinggi dan juga karbohidrat sedang ini dapat diinterpretasikan bahwa cluster tersebut merupakan cluster untuk *Bulking*, dan untuk cluster 2 Energi bernilai tinggi dengan protein dan lemak sedang dan juga karbohidrat yang sangat tinggi ini dapat diinterpretasikan bahwa cluster tersebut merupakan cluster untuk *Bulking*.

2. Implementasi Sistem Perhitungan Target Kalori

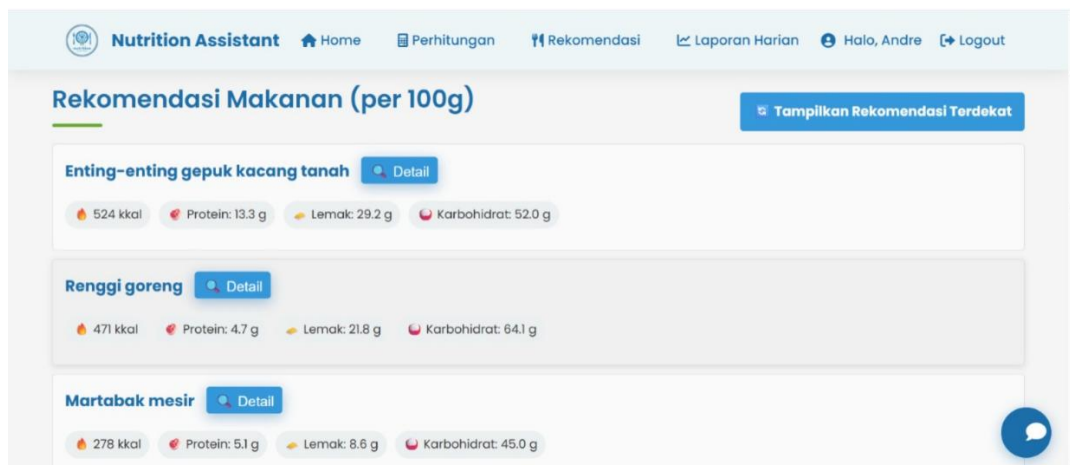
Aplikasi ini diimplementasikan dalam bentuk *website* dengan menggunakan *framework Bootstrap* dan menggunakan bahasa utamanya yaitu *JavaScript*.



Sumber: Hasil Penelitian (2026)

Gambar 8 Menghitung Target Kalori

Gambar 8 adalah bagian untuk user dapat mengetahui berapa jumlah kalori yang dibutuhkan sehari-harinya, dengan memasukkan data jenis kelamin, berat badan, tinggi badan, umur, level aktivitas dan tujuan yang ingin dicapai. Setelah pengguna memasukkan semua data yang diperlukan maka sistem akan otomatis menampilkan hasil perhitungan dari target kalori. Setelah target kalori sudah diketahui maka sistem akan mengelompokkan makanan berdasarkan *goal/tujuan* dari *user*, dapat dilihat dari gambar dibawah ini.



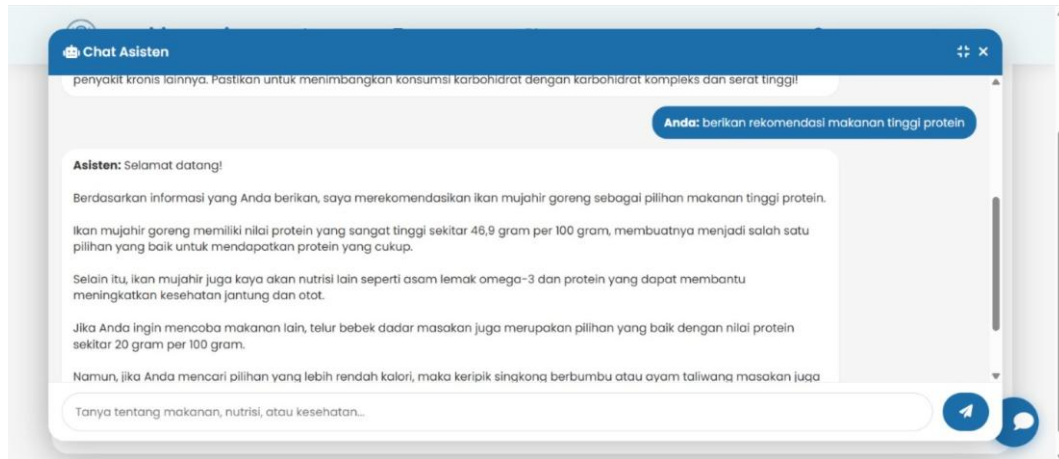
Sumber: Hasil Penelitian (2026)

Gambar 9 Rekomendasi Makanan Sesuai Kriteria

Gambar 9 menampilkan hasil rekomendasi makanan berdasarkan hasil target kalori yang telah dihitung. Menampilkan keseluruhan informasi mengenai Nama Makanan, Energi, Protein, Karbohidrat, Lemak dan beberapa informasi yang ditampilkan pada detail keseluruhan.

3. Implementasi Penerapan NLP

Dalam metode NLP ini penulis menggunakan *Word Embedding* yang menggabungkan antara *Semantic Embedding* dan *Cosine Similarity*.(Mannepalli et al., 2024)



Sumber: Hasil Penelitian (2026)

Gambar 10 Halaman Chatbot

Pada Gambar 10 adalah hasil simulasi pada saat user melakukan interaksi terhadap *chatbot*. Terlihat bahwa *chatbot* mampu merespon input yang diberikan oleh *user* dengan *Knowledge Base* yang sebelumnya sudah penulis buat, hal ini menandakan bahwa *Sentence Embedding* dan *Cosine Similarity* dapat berjalan dengan baik.

KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means dan *Natural Language Processing* (NLP) pada pembuatan aplikasi rekomendasi untuk nutrisi makanan yang mampu disesuaikan dengan kebutuhan kalori pengguna. Kebutuhan kalori dilakukan dengan menggunakan perhitungan metode BMR yang digabungkan dengan aktivitas fisik, sehingga dapat menghasilkan target harian kalori yang lebih spesifik. Algoritma *K-means* digunakan untuk mengakses kumpulan data makanan yang kemudian mengelompokkan data menjadi tiga *cluster* utama, yaitu *Cutting*, *Maintain*, dan *Bulking* berdasarkan kekuatan energi dan nutrisi, dengan menggunakan metode Elbow dan dinilai menggunakan *Davies-Bouldin Index* (DBI) agar jumlah kluster lebih optimal. Pada penerapan NLP berbasis *Semantic Embedding* dan *Cosine Similarity* mampu merespons dan memahami input pengguna dengan efektif melalui *chatbot*. Aplikasi website ini berjalan dengan baik dalam melakukan perhitungan terhadap kebutuhan kalori dan mampu memberikan rekomendasi makanan. Dengan demikian, aplikasi ini berkapasitas menjadi *Nutrition Assistant* yang mampu membantu dalam mengelola makanan sesuai dengan kebutuhan dan tujuan untuk kesehatan.

REFERENSI

- Afiasari, N., Suarna, N., & Rahaningsih, N. (2023). Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means E-commerce K-Means melakukan analisis penerapan Data Mining dalam mengelompokkan jumlah. *Jurnal SAINTEKOM (Sains, Teknologi, Komputer, Dan Manajemen)*, 13(1), 100–110.
- Alfarizi M. Riziq Sirfatullah, Al-farish Muhammad Zidan, Taufiqurrahman Muhammad, Ardiansah Ginan, & Elgar Muhammad. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)*, 2(1), 1–6.
- Cena, H., & Calder, P. C. (n.d.). *Defining a Healthy Diet : Evidence for the Role of Contemporary Dietary Patterns in Health and Disease. Nutrients [revista en Internet] 2020 [acceso 19 de julio de 2021]; 12(334): 1-15.* 1–15.
- Chow, Y. Y., Haw, S. C., Naveen, P., Anaam, E. A., & Mahdin, H. Bin. (2023). Food Recommender System: A Review on Techniques, Datasets and Evaluation Metrics. *Journal of System and Management Sciences*, 13(5), 153–168. <https://doi.org/10.33168/JSMS.2023.0510>
- Dan, I., Intech, T., Kamila, C., Adiyatma, M. A. S., Namang, G. R., Ramadhan, R., & Syah, F. (2021). *Systematic Literature Review : Penggunaan Algoritma K-Means Untuk Clustering di Indonesia dalam Bidang Pendidikan.* 2(1), 19–24.
- Gatto, P. A., & Awangga, R. M. (2023). *Pengelompokan Kedisiplinan Pegawai Berdasarkan Absensi Menggunakan Algoritma K-Means.* Penerbit Buku Pedia.
- Kadang, M. O., Indahsari, A. N., Sumah, J., Firdaus, F., Nasution, V. M., & Prayoga, J. (2025). *Machine Learning Tanpa Label: Panduan Lengkap Unsupervised Learning.* Serasi Media Teknologi.
- Kurnia Saraswati, S., Dhista Rahmaningrum, F., Naufal Zidane Pahsya, M., Paramitha, N., Wulansari, A., Rossa Ristantya, A., Magdalena Sinabutar, B., Estetika Pakpahan, V., & Nandini, N. (2020). MEDIA

- KESEHATAN MASYARAKAT INDONESIA Literature Review : Faktor Risiko Penyebab Obesitas. *Media Kesehatan Masyarakat Indonesia*, 70–74.
- Mannepalli, P. K., Khan, A., Khatri, P., & Chourasia, H. (2024). *NATURAL LANGUAGE PROCESSING*. Xoffencerpublication.
- Metode, R., Berbasis, K., Learning, M., Adawiyah, Q., & Defit, S. (2024). *Jurnal KomtekInfo Penerapan Algoritma K-Means Clustering untuk Mengelompokkan*. 11(4), 300–305. <https://doi.org/10.35134/komtekinfo.v11i4.563>
- Prasetyo, V. R., Benarkah, N., & Chrisintha, V. J. (2021). Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya. *Teknika*, 10(2), 114–121. <https://doi.org/10.34148/teknika.v10i2.370>
- Prianto, C., & Bunyamin, S. (2020). *Pembuatan aplikasi clustering gangguan jaringan menggunakan metode K-Means clustering*. Kreatif.
- Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25. <https://doi.org/10.33365/jtk.v15i2.1162>
- Syahri, R., Informatika, T., & Selatan, S. (2023). *ALGORITMA K-MEANS CLUSTERING : SEBUAH STUDI LITERATUR K-MEANS CLUSTERING ALGORITHM : A LITERATUR STUDY*. x(x), 1–7. <https://doi.org/10.12345/juri>
- Taufik, A., Kom, S., Bernadus Gunawan Sudarsono, M., & Kom, M. (n.d.). *Taufik, A., Sudarsono, G., Sudaryana, I. K., & Muryono, T. T. (2022). Pengantar teknologi informasi. Yayasan Drestanta Pelita Indonesia, 1-113.*