Jurnal Minfo Polgan (JMP)
Volume 15, Nomor 1, Maret 2026
DOI : https://doi.org/10.33395/jmp.v15i1.16056
e-ISSN : 2797-3298
p-ISSN : 2089-9424

Sistem Klasifikasi Kandungan Gula pada Produk Makanan dan Minuman Kemasan Berbasis Algoritma K-Nearest Neighbor (KNN)

[1]Fikri Firmansyah, [2*] Indah Purnama Sari
[1]Sistem Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
[2*]Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
*Korespondensi: indahpurnama@umsu.ac.id

Submit : 22 Mar 2026 | **Diterima** : 09 April 2026 | **Terbit** : 11 April 2026

ABSTRACT

Excessive sugar consumption from packaged food and beverage products contributes to the increasing prevalence of obesity and type 2 diabetes mellitus in Indonesia. Low public nutritional literacy often makes nutrition label information difficult to understand, making it difficult for consumers to assess the sugar content level of a product. This study aims to develop a sugar content classification system for packaged food and beverage products using the K-Nearest Neighbor (KNN) algorithm. The dataset was obtained from Open Food Facts and includes nutritional attributes such as energy, sugar, carbohydrates, fat, and protein. The research process includes data preprocessing, feature selection, and dataset splitting using train-test split or cross-validation methods. The classification results categorize products into three sugar levels: low, medium, and high. The developed system was implemented as a web-based prototype that can automatically provide sugar category information.

Keywords: *Classification, Sugar Content, K-Nearest Neighbor, Data Mining, Open Food Facts, Nutritional Information System.*

ABSTRAK

Konsumsi gula berlebih dari produk makanan dan minuman kemasan menjadi salah satu faktor meningkatnya kasus obesitas dan diabetes melitus tipe 2 di Indonesia. Rendahnya literasi gizi masyarakat menyebabkan informasi nilai gizi pada label produk sulit dipahami sehingga konsumen kesulitan menilai tingkat kandungan gula suatu produk. Penelitian ini bertujuan merancang sistem klasifikasi kandungan gula pada produk makanan dan minuman kemasan menggunakan algoritma K-Nearest Neighbor (KNN). Dataset yang digunakan berasal dari Open Food Facts dengan atribut nilai gizi seperti energi, gula, karbohidrat, lemak, dan protein. Proses penelitian meliputi data preprocessing, pemilihan fitur, serta pembagian data menggunakan metode train-test split atau cross-validation. Hasil klasifikasi membagi produk ke dalam tiga kategori kandungan gula yaitu rendah, sedang, dan tinggi. Hasil evaluasi menunjukkan bahwa algoritma KNN mampu mengklasifikasikan tingkat kandungan gula dengan performa yang baik dan sistem diimplementasikan dalam bentuk prototype berbasis web.

Kata Kunci: Klasifikasi, Kandungan Gula, K-Nearest Neighbor, Data Mining, Open Food Facts, Sistem Informasi Gizi

PENDAHULUAN

Dampak dari tingginya kadar gula darah sangat serius. IDF menyebutkan bahwa diabetes dapat memicu berbagai komplikasi seperti penyakit kardiovaskular, kerusakan ginjal, gangguan penglihatan, kerusakan saraf, dan masalah pada kulit maupun kaki. Untuk mengurangi risiko tersebut, Organisasi Kesehatan Dunia (WHO) merekomendasikan konsumsi gula bebas kurang dari 10% total energi harian, atau idealnya 5%, yang setara dengan sekitar 50 gram per hari untuk orang dewasa. Pemerintah Indonesia melalui Peraturan Menteri Kesehatan No. 30 Tahun 2013 juga menetapkan anjuran konsumsi gula sebesar 50 gram per hari (BPOM RI, 2021). Meskipun standar tersebut telah ditetapkan, pemenuhan batas konsumsi gula masih menjadi tantangan besar bagi masyarakat, terutama ketika informasi komposisi pada label produk kemasan tidak secara eksplisit menunjukkan jumlah gula, sehingga diperlukan metode klasifikasi yang lebih sistematis dan terukur untuk menentukan tingkat kandungan gula

berdasarkan atribut seperti kadar gula dalam bentuk gram atau mililiter, dengan memanfaatkan pendekatan berbasis data untuk mengatasi ketidakjelasan tersebut.

Salah satu kesenjangan utama terletak pada rendahnya pemahaman konsumen terhadap informasi nilai gizi pada label produk kemasan. Banyak konsumen yang tidak memahami arti angka-angka pada label gizi, termasuk kandungan gula per porsi maupun per 100gram. Hal ini menyebabkan masyarakat sulit melakukan kontrol konsumsi gula secara mandiri. Oleh karena itu, dibutuhkan suatu sistem yang dapat membantu masyarakat menginterpretasikan kandungan gula pada produk secara lebih mudah dan informatif, di mana sistem tersebut dirancang untuk mengklasifikasikan tingkat kandungan gula secara otomatis berdasarkan atribut nilai gizi yang tersedia pada setiap produk, seperti energi, lemak, karbohidrat, dan kategori produk, melalui proses pengolahan data yang terintegrasi untuk menghasilkan kategori rendah, sedang, atau tinggi kandungan gula.

Dalam konteks pengolahan data nutrisi, penerapan kecerdasan buatan (Artificial Intelligence/AI), khususnya algoritma K-Nearest Neighbor (KNN), menjadi salah satu pendekatan yang menjanjikan. KNN merupakan algoritma sederhana namun efektif dalam melakukan klasifikasi berdasarkan kemiripan data. Penelitian oleh Gina dkk. menunjukkan bahwa algoritma KNN dengan variasi nilai k (3, 5, dan 7) mampu mencapai akurasi hingga 99% dalam menentukan status gizi balita berdasarkan indeks antropometri (Gina Purnama Insany et al., 2023). Keberhasilan ini menunjukkan bahwa KNN mampu mengenali pola hubungan antaratribut nutrisi dan memiliki potensi yang tinggi untuk diterapkan dalam klasifikasi kandungan gula pada produk makanan dan minuman kemasan, di mana penerapan KNN melibatkan pemilihan parameter terbaik seperti nilai k optimal melalui teknik validasi silang (cross-validation) untuk meningkatkan akurasi klasifikasi, serta integrasi dengan dataset untuk membangun model yang adaptif terhadap variasi data nutrisi.

Untuk kebutuhan penelitian ini, digunakan dataset dari Open Food Facts, yaitu basis data nutrisi terbuka yang memuat informasi komposisi gizi ribuan produk makanan dan minuman dari berbagai negara. Dataset ini mencakup atribut seperti kadar gula, energi, lemak, karbohidrat, serta kategori produk. Dengan memanfaatkan dataset tersebut, model klasifikasi dapat dibangun untuk mengelompokkan produk ke dalam kategori rendah, sedang, atau tinggi kandungan gula berdasarkan kadar gula per gram atau mililiter, dan kinerja sistem ini dievaluasi melalui metrik seperti akurasi, presisi, recall, dan F1-score untuk memastikan keandalan dalam pengelompokan, termasuk penggunaan confusion matrix guna mengidentifikasi kesalahan klasifikasi dan potensi perbaikan model.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk merancang dan menganalisis model klasifikasi kandungan gula pada produk makanan dan minuman kemasan menggunakan algoritma K-Nearest Neighbor (KNN). Hasil penelitian diharapkan dapat menjadi dasar pengembangan sistem informasi gizi yang lebih mudah dipahami masyarakat, sehingga mendukung upaya pengendalian konsumsi gula dan penerapan gaya hidup sehat di Indonesia.

METODE PENELITIAN

Analisis Permasalahan

Konsumsi gula berlebih dari makanan dan minuman kemasan merupakan salah satu penyebab meningkatnya kasus obesitas dan diabetes melitus tipe 2 di Indonesia. Meskipun label nilai gizi telah diwajibkan, tingkat literasi gizi masyarakat masih rendah sehingga banyak konsumen tidak mampu memahami kandungan gula yang tercantum pada kemasan. Akibatnya, masyarakat kesulitan menilai apakah suatu produk memiliki kadar gula tinggi atau rendah.
Permasalahan ini diperparah oleh penyajian informasi gizi yang kurang mudah dipahami, sehingga konsumen lebih fokus pada rasa atau harga tanpa mempertimbangkan kandungan gula aktual. Oleh karena itu, diperlukan alat bantu yang mampu menyederhanakan informasi gizi menjadi kategori yang mudah dimengerti.

Teknologi klasifikasi berbasis algoritma K-Nearest Neighbor (KNN) dapat digunakan untuk mengelompokkan produk ke dalam kategori rendah, sedang, atau tinggi kandungan gula secara otomatis. Dengan demikian, sistem klasifikasi dapat membantu konsumen membuat keputusan yang lebih sehat tanpa harus menafsirkan label gizi secara manual.

Analisis Data

Analisis data bertujuan untuk memahami struktur, karakteristik, dan kualitas data yang digunakan dalam proses pembangunan model klasifikasi kandungan gula. Mengingat dataset Open Food Facts memiliki jumlah data yang sangat besar dan bervariasi, penelitian ini

menggunakan subset data yang telah dipilih dan diproses agar sesuai untuk penerapan algoritma K-Nearest Neighbor (KNN).

Alur Pemikiran Penelitian

Konsumsi gula tambahan yang berlebihan pada produk kemasan menjadi salah satu faktor utama meningkatnya prevalensi obesitas dan diabetes melitus tipe 2 di Indonesia. Masyarakat sering kali kesulitan mengenali kadar gula aktual karena informasi pada label kurang mudah dipahami atau terkadang tidak lengkap. Open Food Facts menyediakan basis data nutrisi terbuka yang memuat informasi gula dan komponen gizi lainnya secara terstandarisasi per 100 g, sehingga dapat dimanfaatkan sebagai sumber data untuk membangun model klasifikasi otomatis. Dengan menerapkan teknik data mining, khususnya algoritma K-Nearest Neighbors (KNN), data nutrisi tersebut dapat diolah untuk mengelompokkan produk pangan ke dalam tiga kategori kadar gula: rendah, sedang, dan tinggi. Model ini diharapkan dapat menjadi dasar pengembangan sistem atau aplikasi yang membantu konsumen maupun tenaga kesehatan dalam mengidentifikasi produk berisiko tinggi gula secara cepat dan akurat, sehingga mendukung upaya pencegahan penyakit tidak menular melalui pengendalian asupan gula harian.

Gambar 1 Diagram Kerangka Pemikiran

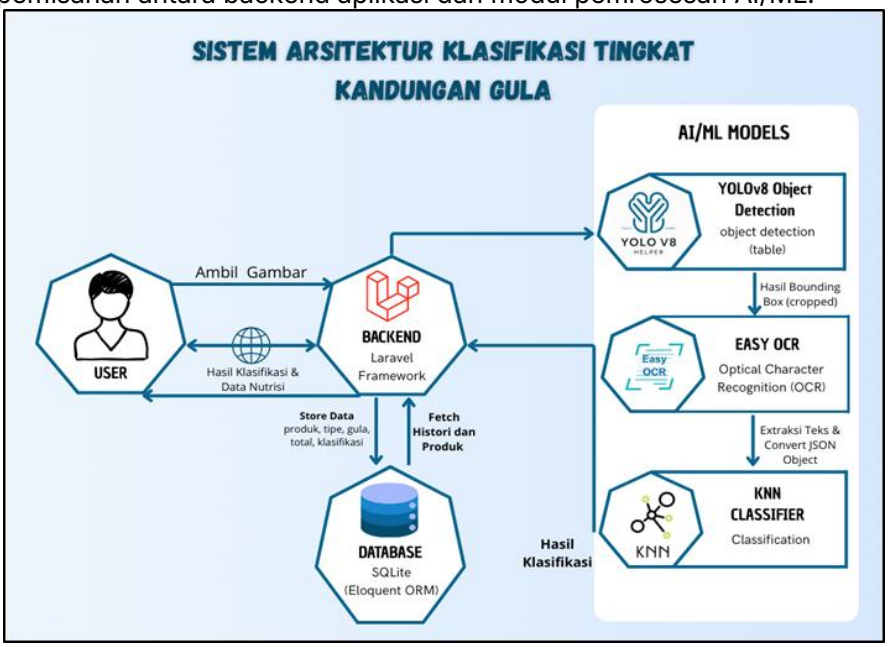
Kerangka pemikiran penelitian ini dimulai dari studi literatur untuk memahami konsep kandungan gizi, risiko konsumsi gula berlebih, serta metode klasifikasi yang relevan. Setelah itu, diidentifikasi permasalahan tingginya konsumsi gula sehingga diperlukan sistem yang mampu mengelompokkan produk berdasarkan kadar gulanya. Dataset dari Open Food Facts kemudian

dikumpulkan dan dipilih untuk mendukung kebutuhan analisis. Berdasarkan karakteristik data, algoritma K-Nearest Neighbors (K-NN) ditetapkan sebagai metode klasifikasi. Implementasi dilakukan dengan menghitung jarak antar data dan menentukan kategori gula melalui proses voting tetangga terdekat. Hasil akhirnya berupa klasifikasi tingkat kandungan gula pada produk makanan dan minuman kemasan.

HASIL DAN PEMBAHASAN

Arsitektur Sistem

Arsitektur sistem pada tahap implementasi menggambarkan integrasi aktual antara komponen aplikasi web, modul kecerdasan buatan, serta basis data dalam mendukung proses klasifikasi tingkat kandungan gula produk. Sistem dirancang menggunakan pendekatan client-server dengan pemisahan antara backend aplikasi dan modul pemrosesan AI/ML.



Gambar 2 Diagram Arsitektur Sistem

Training Model K-Nearest Neighbor

Tahap pelatihan model dilakukan menggunakan algoritma K-Nearest Neighbor (KNN) yang tersedia pada pustaka scikit-learn. Algoritma ini merupakan metode klasifikasi berbasis jarak (distance-based classifier) yang menentukan kelas suatu data berdasarkan kedekatannya terhadap sejumlah tetangga terdekat pada data latih. Penggunaan Pipeline memastikan bahwa proses scaling hanya dipelajari dari data training pada setiap lipatan validasi silang (cross validation), sehingga tidak terjadi kebocoran informasi dari data testing.

Contoh Normalisasi Data (5 Data Pertama)

Sebelum Normalisasi:

	energy-kcal_100g	carbohydrates_100g	sugar_per_carbs_ratio	sugar_per_kcal_ratio	sugar_dominance_score
4756	643.0	28.57	0.124956	0.022208	0.083857
4148	334.0	30.70	0.869707	0.319760	0.649728
4332	300.0	73.30	0.954980	0.933333	0.946321
3113	828.0	0.00	0.000000	0.000000	0.000000
679	343.0	14.29	1.000000	0.166647	0.666659

Sesudah Normalisasi (Rentang 0-1):

	energy-kcal_100g	carbohydrates_100g	sugar_per_carbs_ratio	sugar_per_kcal_ratio	sugar_dominance_score
0	0.714444	0.2857	0.124956	2.313375e-10	2.183779e-09
1	0.371111	0.3070	0.869707	3.330838e-09	1.692001e-08
2	0.333333	0.7330	0.954980	9.722222e-09	2.464378e-08
3	0.920000	0.0000	0.000000	0.000000e+00	0.000000e+00
4	0.381111	0.1429	1.000000	1.735909e-09	1.736091e-08

Gambar 3 Normalisasi Data

Sebelum proses pelatihan, dilakukan normalisasi fitur menggunakan metode Min-Max Scaling. Normalisasi ini mentransformasikan nilai fitur ke dalam rentang 0 hingga 1. Normalisasi diperlukan karena KNN menghitung jarak antar data. Perbedaan skala antar fitur dapat menyebabkan fitur dengan nilai yang lebih besar mendominasi perhitungan jarak. Oleh karena

itu, seluruh fitur harus berada pada skala yang sama agar setiap fitur memiliki kontribusi yang seimbang terhadap perhitungan jarak. Proses normalisasi dan pelatihan model diintegrasikan dalam sebuah Pipeline menggunakan pustaka scikit-learn. Penggunaan Pipeline bertujuan untuk mencegah terjadinya data leakage selama proses cross validation, karena proses normalisasi hanya dilakukan pada data training di setiap lipatan validasi.

Optimasi Hyperparameter

```
🔎 Grid Search Cross Validation...
Fitting 5 folds for each of 56 candidates, totalling 280 fits

📋 10 Kombinasi Parameter Terbaik:
    param_knn__n_neighbors param_knn__weights param_knn__metric  mean_test_score  std_test_score
33                       7          distance         manhattan          0.96575        0.007730
37                      11          distance         manhattan          0.96525        0.009334
39                      13          distance         manhattan          0.96525        0.009792
35                       9          distance         manhattan          0.96475        0.008782
43                      17          distance         manhattan          0.96475        0.008078
31                       5          distance         manhattan          0.96450        0.005788
29                       3          distance         manhattan          0.96450        0.007185
47                      21          distance         manhattan          0.96425        0.008537
45                      19          distance         manhattan          0.96425        0.008314
41                      15          distance         manhattan          0.96375        0.007706
```

Gambar 4 Grid Search Validation

Untuk memperoleh performa optimal, dilakukan pencarian kombinasi parameter terbaik menggunakan metode Grid Search Cross Validation. Parameter yang diuji meliputi: Jumlah tetangga (K) dengan rentang 3 hingga 51 (nilai ganjil); Metode pembobotan (uniform dan distance); Metode perhitungan jarak (euclidean dan manhattan); Nilai leaf size (20, 30, 40) yang mempengaruhi kecepatan komputasi pada struktur data tree.

```
🎯 Best Parameter:
{'knn__metric': 'manhattan', 'knn__n_neighbors': 7, 'knn__weights': 'distance'}
Best CV Accuracy: 0.9657

📊 Evaluasi pada data TEST...

Test Accuracy: 0.9640
```

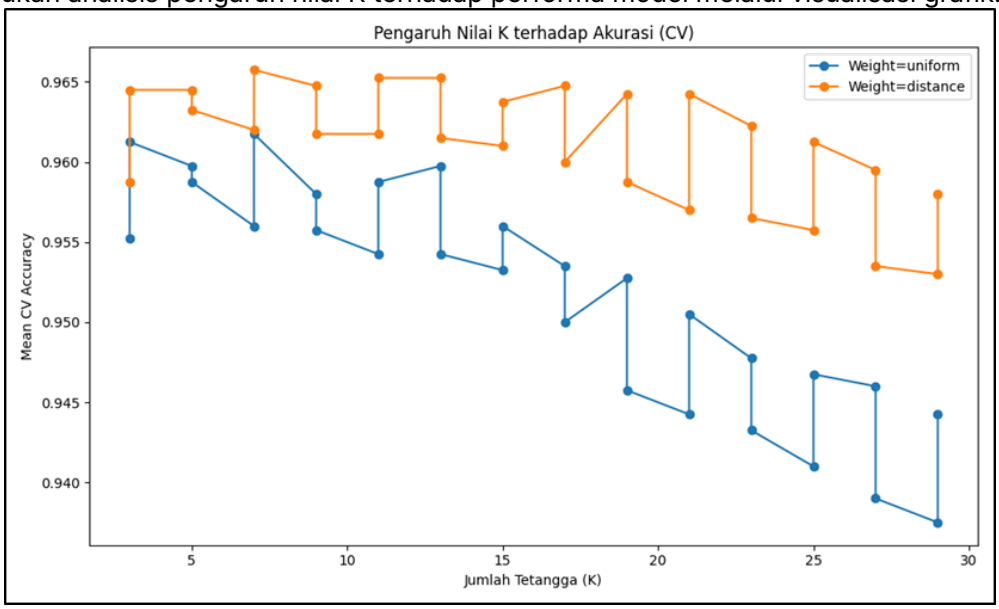
Gambar 5 Hasil Optimasi

Pada penelitian ini, dua metode perhitungan jarak digunakan, yaitu Euclidean Distance dan Manhattan Distance. Euclidean Distance menghitung jarak garis lurus antar dua titik dalam ruang multidimensi, sedangkan Manhattan Distance menghitung jarak berdasarkan jumlah selisih absolut tiap dimensi. Pengujian kedua metode ini dilakukan untuk menentukan fungsi jarak yang paling sesuai dengan karakteristik dataset, sehingga pemilihan metode tidak didasarkan pada asumsi teoritis semata, melainkan pada hasil evaluasi eksperimental. Penggunaan nilai K ganjil bertujuan untuk meminimalkan kemungkinan terjadinya tie pada proses voting mayoritas. Validasi dilakukan menggunakan metode Stratified K-Fold Cross Validation dengan 5 lipatan (n_splits=5). Teknik stratifikasi memastikan distribusi kelas tetap seimbang pada setiap lipatan data sehingga evaluasi model menjadi lebih stabil dan tidak bergantung pada satu pembagian data tertentu. Validasi dilakukan menggunakan metode Stratified K-Fold Cross Validation dengan 5 lipatan. Teknik ini memastikan distribusi kelas tetap seimbang pada setiap lipatan data sehingga evaluasi model lebih stabil dan tidak bergantung pada satu pembagian data tertentu. Proses evaluasi pada tahap Grid Search menggunakan metrik accuracy sebagai dasar pemilihan parameter terbaik. Model dengan nilai rata-rata akurasi validasi silang (mean cross-validation accuracy) tertinggi dipilih sebagai model terbaik.

Evaluasi dan Visualisasi Hasil Optimasi

1. Akurasi
   Evaluasi model dilakukan menggunakan metrik Accuracy untuk mengukur sejauh mana model mampu mengklasifikasikan data dengan benar pada dataset yang belum pernah dilihat sebelumnya (Test Set). Nilai akurasi sebesar 96.40% menunjukkan bahwa dari 100

data uji, model mampu memberikan prediksi yang benar pada sekitar 96 data. Hasil ini mengindikasikan bahwa model memiliki kemampuan generalisasi yang sangat baik dan tidak mengalami overfitting yang signifikan, karena mampu mempertahankan performa tinggi pada data yang tidak digunakan selama proses training. Setelah proses grid search, dilakukan analisis pengaruh nilai K terhadap performa model melalui visualisasi grafik.



Gambar 6 Pengaruh Nilai K Terhadap Akurasi

Grafik ini menampilkan hubungan antara jumlah tetangga (K) dengan rata-rata akurasi validasi silang untuk masing-masing metode pembobotan (uniform dan distance). Visualisasi ini membantu dalam memahami perilaku model dan memvalidasi pemilihan parameter optimal.

2. Classification Report
   Classification report digunakan untuk memberikan gambaran yang lebih mendalam mengenai performa model pada setiap kelas (0, 1, dan 2). Metrik yang digunakan meliputi Precision, Recall, dan F1-Score.

Classification Report:

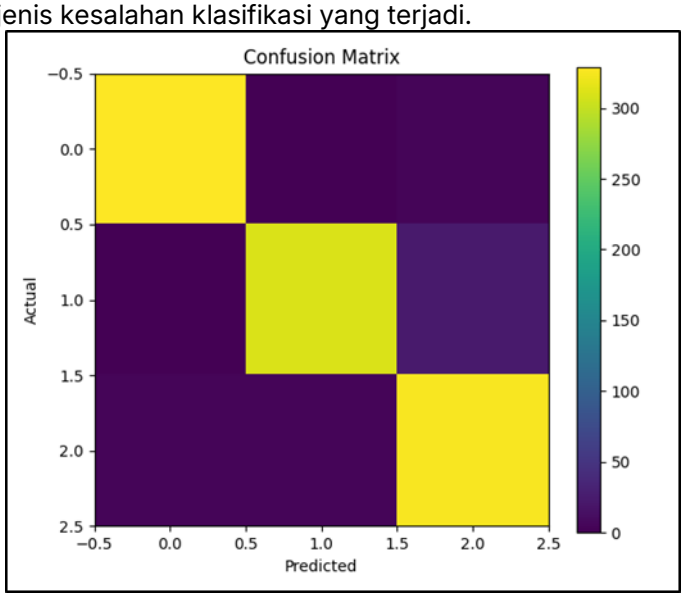
	precision	recall	f1-score	support
0	0.99	0.99	0.99	333
1	0.99	0.93	0.96	334
2	0.92	0.98	0.95	334
accuracy			0.96	1001

Gambar 7 Evaluasi Model

Pada Kelas 0, model menunjukkan performa yang hampir sempurna dengan nilai F1-score mencapai 0.99, yang menandakan keseimbangan ideal antara presisi dan sensitivitas deteksi. Sementara itu, Kelas 1 memiliki tingkat presisi yang sangat tinggi sebesar 0.99, namun dengan nilai recall sebesar 0.93. Hal ini mengindikasikan bahwa meskipun model sangat akurat saat memprediksi Kelas 1, masih terdapat sekitar 7% data asli Kelas 1 yang terklasifikasi ke dalam kelas lain. Sebaliknya, pada Kelas 2, model menunjukkan nilai recall yang sangat tinggi yaitu 0.98, namun dengan nilai presisi terendah di angka 0.92. Fenomena ini menunjukkan kecenderungan model untuk lebih inklusif dalam mengidentifikasi Kelas 2, sehingga beberapa sampel dari kelas lain berisiko terdeteksi sebagai Kelas 2 (false positive). Meskipun terdapat variasi kecil antar kelas, nilai macro average yang stabil di angka 0.96 menegaskan bahwa model ini sangat andal untuk diimplementasikan secara luas.

3. Confussion Matrix
   Visualisasi Confusion Matrix memberikan pemahaman yang lebih eksplisit mengenai

performa model dalam mengklasifikasikan setiap sampel ke dalam label yang tepat serta mengidentifikasi jenis kesalahan klasifikasi yang terjadi.



Gambar 8 Confussion Matrix

Berdasarkan matriks yang dihasilkan, terlihat dominasi warna terang pada garis diagonal utama dari kiri atas ke kanan bawah, yang merepresentasikan jumlah prediksi benar (True Positive) yang sangat tinggi untuk ketiga kelas. Kelas 0 menunjukkan tingkat keberhasilan klasifikasi yang paling konsisten dengan hampir seluruh sampel berada pada koordinat aktual-prediksi yang tepat. Analisis lebih mendalam pada area di luar diagonal utama menunjukkan adanya sedikit hamburan data, khususnya pada interaksi antara Kelas 1 dan Kelas 2. Terdapat sejumlah kecil sampel dari Kelas 1 yang terprediksi secara keliru sebagai Kelas 2, yang menjelaskan mengapa nilai recall pada Kelas 1 (0.93) sedikit lebih rendah dibandingkan kelas lainnya. Meskipun demikian, tingkat kesalahan ini sangat minim jika dibandingkan dengan total sampel uji sebanyak 1001 data. Secara keseluruhan, Confusion Matrix ini mengonfirmasi bahwa model memiliki tingkat diskriminasi yang kuat antar kelas dan mampu membedakan fitur-fitur unik dari setiap kategori dengan margin kesalahan yang sangat rendah.

KESIMPULAN

Berdasarkan hasil penelitian mengenai Sistem Klasifikasi Kandungan Gula pada Produk Makanan dan Minuman Kemasan Berbasis Algoritma K-Nearest Neighbor (KNN), maka dapat diambil beberapa kesimpulan sebagai berikut: Penelitian ini berhasil membangun sistem klasifikasi kandungan gula pada produk makanan dan minuman kemasan dengan memanfaatkan dataset dari Open Food Facts yang telah melalui proses seleksi dan preprocessing. Dataset difokuskan pada atribut nutrisi utama seperti energy_100g, carbohydrates_100g dan fitur tambahan hasil rekayasa fitur. Proses pra-pemrosesan data yang meliputi data cleaning, normalisasi (Min-Max Scaling), feature selection, dan pembagian data latih serta data uji terbukti penting dalam meningkatkan kualitas data sebelum diterapkan pada algoritma KNN. Normalisasi sangat berpengaruh karena KNN berbasis perhitungan jarak Euclidean yang sensitif terhadap perbedaan skala fitur. Algoritma K-Nearest Neighbor (KNN) dapat diterapkan secara efektif untuk mengklasifikasikan tingkat kandungan gula ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi. Berdasarkan hasil pengujian dengan metode 10-fold cross validation, nilai K optimal yang diperoleh adalah K = 5, yang menghasilkan performa klasifikasi terbaik dibandingkan nilai K lainnya. Sistem yang dikembangkan mampu memberikan hasil klasifikasi secara otomatis berdasarkan input komposisi nutrisi dari pengguna. Hasil evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score menunjukkan bahwa model memiliki tingkat kinerja yang sangat baik dalam mengelompokkan produk sesuai kategori kadar gula. Berdasarkan pengujian menggunakan metode 10-fold cross validation, model K-Nearest Neighbor (KNN) dengan nilai K optimal menghasilkan tingkat akurasi sebesar 96,40%, yang menunjukkan bahwa sistem memiliki kemampuan klasifikasi yang tinggi dan konsisten dalam

membedakan kategori rendah, sedang, dan tinggi kandungan gula. Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan data mining menggunakan algoritma KNN dapat dimanfaatkan sebagai solusi sistematis untuk membantu masyarakat memahami tingkat kandungan gula pada produk makanan dan minuman kemasan secara lebih sederhana dan informatif. Hasil penelitian melalui statistik Uji t (student t-test) dibuktikan bahwa Kualitas Pelayanan (X) berpengaruh positif dan signifikan terhadap Kepuasan Pelanggan (X) pada Seoul Korean Restaurant Palembang dengan taraf kepercayaan 95%. Hal ini menunjukkan bahwa pelayanan yang berkualitas berperan penting dalam membentuk kepuasan konsumen, selain itu juga erat kaitannya dalam menciptakan keuntungan bagi perusahaan. Semakin berkualitas pelayanan yang diberikan oleh perusahaan maka kepuasan yang dirasakan oleh pelanggan akan semakin tinggi.

UCAPAN TERIMA KASIH

Terima kasih kepada Universitas Muhammadiyah Sumatera Utara yang sudah memberikan bantuan biaya publikasi dan motivasi terhadap kami dan terima kasih kepada keluarga kami yang paling kami sayangi.

REFERENSI

Agyapong, K. B., Hayfron-Acquah, J. B., & Asante, M. (2016). An Overview of Data Mining Models (Descriptive and Predictive) (Vol. 4). www.ijournals.in

Alghifari, F., & Juardi, D. (2021). Fauzan Alghifari Penerapan Data Mining Pada Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes.

Amelisa Edwina, D., Manaf, A., & Efrida. (2015). Pola Komplikasi Kronis Penderita Diabetes Melitus Tipe 2 Rawat Inap di Bagian Penyakit Dalam RS. Dr. M. Djamil Padang Januari 2011-Desember 2012. In Andalas (Vol. 4, Issue 1). http://jurnal.

Archyde. (2025, July 9). Indonesia Masuk 5 Negara dengan Jumlah Penderita Diabetes Tertinggi. https://www.archyde.com/indonesia-ranks-among-top-5-nations-with-diabetes-a-critical-health-alert/

Arhami, M., & Nasir, M. (2020). Data mining : algoritma dan implementasi. https://librarian.stitek.ac.id/opac/detail-opac?id=453

Asri Mulyani, Sarah Khoerunisa, & Dede Kurniadi. (2025). Perbandingan Kinerja Algoritma KNN dan SVM Menggunakan SMOTE untuk Klasifikasi Penyakit Diabetes. Jurnal Nasional Teknik Elektro Dan Teknologi Informasi, 14(1), 25–34. https://doi.org/10.22146/jnteti.v14i1.15198

Sari, I.P., Ramadhani, F., Satria, A., & Apdilah, D. (2023). Implementasi Pengolahan Citra Digital dalam Pengenalan Wajah menggunakan Algoritma PCA dan Viola Jones. Hello World Jurnal Ilmu Komputer 2 (3), 146-157

Davies, T., Louie, J. C. Y., Ndanuko, R., Barbieri, S., Perez-Concha, O., & Wu, J. H. Y. (2022). A Machine Learning Approach to Predict the Added-Sugar Content of Packaged Foods. Journal of Nutrition, 152(1), 343–349. https://doi.org/10.1093/jn/nxab341

Dwi Fasnuari, H. A., Yuana, H., & Chulkamdi, M. T. (2022). PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI PENYAKIT DIABETES MELITUS. Antivirus : Jurnal Ilmiah Teknik Informatika, 16(2), 133–142. https://doi.org/10.35457/antivirus.v16i2.2445

Gina Purnama Insany, Indra Yustiana, & Sri Rahmawati. (2023). Penerapan KNN dan ANN pada klasifikasi status gizi balita berdasarkan indeks antropometri. Jurnal CoSciTech (Computer Science and Information Technology), 4(2), 385–393. https://doi.org/10.37859/coscitech.v4i2.5079

Han, J., Kamber, M., & Pei, J. (2011). Data Transformation by Normalization. Data Mining: Concepts and Techniques, 113–115. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

Hardy, O. T., Czech, M. P., & Corvera, S. (2012). What causes the insulin resistance underlying obesity? Current Opinion in Endocrinology, Diabetes, and Obesity, 19(2), 81. https://doi.org/10.1097/MED.0B013E3283514E13

Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. IJCCS (Indonesian

Journal of Computing and Cybernetics Systems), 15(2), 121. https://doi.org/10.22146/ijccs.65176

Iwan Sudipa, I. G., Azdy, R. A., Arfiani, I., Setiohardjo, N. M., & Sumiyatun. (2024). Leveraging K-Nearest Neighbors for Enhanced Fruit Classification and Quality Assessment. Indonesian Journal of Data and Science, 5(1), 30–36. https://doi.org/10.56705/ijodas.v5i1.125

Poznyak, A. V., Litvinova, L., Poggio, P., Sukhorukov, V. N., & Orekhov, A. N. (2022). Effect of Glucose Levels on Cardiovascular Risk. Cells, 11(19), 3034. https://doi.org/10.3390/CELLS11193034

Ramadhani, F., Satria, A., & Sari, I. P. (2023). Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. Hello World Jurnal Ilmu Komputer, 2(2), 58–62. https://doi.org/10.56211/helloworld.v2i2.253

Shelin Sahira, M., Dessela Putri, F., Ristyawan, A., & Daniati, E. (2024). Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 502 Penggunaan Algoritma K-Nearest Neighbor Untuk Klasifikasi Data Diabetes Pada Wanita. In Agustus (Vol. 8). Online.

Sinaga, J., Sinambela, J. L., Purba, B. C., & Pelawi, S. (2024). Gula dan Kesehatan Kajian Terhadap Dampak Kesehatan Akibat Konsumsi Gula Berlebih. Mutiara : Jurnal Ilmiah Multidisiplin Indonesia, 2, 54–68. https://jurnal.tiga-mutiara.com/index.php/jimi/index

Teguh, H. T. S. (2025). Implementasi Data Minning Clustering Dalam Mengelompokan Kasus Perceraian Yang Terjadi Di Provinsi Jawa Timur Menggunakan Algoritma K-Means. JAMI: Jurnal Ahli Muda Indonesia, 6(1), 68–83. https://doi.org/10.46510/jami.v6i1.324

Utami, E., Iskandar, A. F., Hidayat, W., Prasetyo, A. B., & Hartanto, A. D. (2021). Covid-19 Hoax Detection Using KNN in Jaccard Space. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 15(3), 255. https://doi.org/10.22146/ijccs.67392

Sari, I.P., & Batubara, I.H. (2021). Perancangan Sistem Informasi Laporan Keuangan Pada Apotek Menggunakan Algoritma K-NN. Seminar Nasional Teknologi Edukasi dan Humaniora (SiNTESa) (1), 2021.

Wijaya, R. P., Rouf, A., & Supardi, T. W. (2019). Klasifikasi Tingkat Kemurnian Bahan Bakar Minyak Berdasarkan Cepat Rambat Gelombang Menggunakan Algoritma K-Nearest Neighbor. IJEIS (Indonesian Journal of Electronics and Instrumentation Systems), 9(2), 161. https://doi.org/10.22146/ijeis.49660

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniquesWitten, Ian H, Eibe Frank, Mark A Hall, and Christopher J Pal. 2016. Data Mining: Practical Machine Learning Tools and Techniques. . 654. https://books.google.com/books/about/Data_Mining.html?hl=id&id=1SylCgAAQBAJ

Yunita, F. (2016). SISTEM KLASIFIKASI PENYAKIT DIABETES MELLITUS MENGGUNAKAN METODE K-NEAREST NEIGHBOR (K-NN). Sistem Klasifikasipenyakit Diabetes....(Fitri).

allagui.(2025, 7 3). Roboflow Universe. Retrieved from https://universe.roboflow.com/allagui/nutrition-table-detection/dataset/3