

**SISTEM KLASIFIKASI KANDUNGAN GULA PADA PRODUK
MAKANAN DAN MINUMAN KEMASAN BERBASIS
ALGORITMA K-NEAREST NEIGHBOR(KNN)**

SKRIPSI

DISUSUN OLEH

FIKRI FIRMANSYAH

NPM. 2209010131



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN**

2026

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

(KNN)

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

FIKRI FIRMANSYAH

NPM. 2209010131

PROGRAM STUDI FIKRI FIRMANSYAH

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2026

LEMBAR PENGESAHAN

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

Nama Mahasiswa : Fikri Firmansyah

NPM : 2209010131

Program Studi : Sistem Informasi

Menyetujui
Komisi Pembimbing



(Indah Purnama Sari, S.T., M.Kom.)
NIDN. 0116049001

Ketua Program Studi



(Mahardika Abdi Prawira Tanjung,
S.Kom., M.Kom)
NIDN. 0117088902

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

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

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 06 Mei 2026

Yang membuat pernyataan



Fikri Firmansyah

NPM. 2209010131

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini :

Nama : Fikri Firmansyah
NPM : 2209010131
Program Studi : Sistem Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

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

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 06 Mei 2026

Yang membuat pernyataan



Fikri Firmansyah
NPM. 2209010131

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Fikri Firmansyah
Tempat dan Tanggal Lahir : Amuntai, 06 Februari 2004
Alamat Rumah : Jl. Terusan Dusun V Bandar Setia
Telepon/Faks/HP : 081370577338
E-mail : Frmnsyah2406@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD	: SD NEGERI 2 MEULABOH	TAMAT	: 2016
SMP	: MTSN SERDANG BERDAGAI	TAMAT	: 2019
SMA	: SMKN 1 PERCUT SEI TUAN	TAMAT	: 2022

KATA PENGANTAR



Alhamdulillahirabbil'alamin, puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan kemudahan yang diberikan sehingga skripsi ini dapat diselesaikan dengan baik. Dalam proses penyusunan skripsi ini, penulis menyadari bahwa banyak dukungan, doa, bimbingan, serta bantuan dari berbagai pihak yang sangat berarti.

Dengan penuh rasa hormat dan terima kasih, penulis menyampaikan apresiasi kepada:

1. Allah SWT, atas segala rencana terbaik-Nya. Di saat penulis merasa lelah, ragu, dan hampir menyerah, pertolongan-Nya selalu datang dengan cara yang tidak pernah disangka.
2. Bapak Prof. Dr. Akrim, M.Pd., selaku Rektor Universitas Muhammadiyah Sumatera Utara (UMSU), yang telah memberikan kesempatan dan fasilitas untuk menempuh pendidikan di lingkungan akademik yang berkualitas.
3. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU, atas dukungan terhadap proses pendidikan yang berjalan dengan baik.
4. Ibu Dr. Firahmi Rizky, S.Kom., M.Kom., selaku Wakil Dekan I Sistem Informasi, atas arahan dan dukungan selama proses akademik penulis.

5. Bapak Mhd. Basri, S.Si., M.Kom, selaku Wakil Dekan III Sistem Informasi, atas arahan dan dukungan selama proses akademik penulis.
6. Bapak Mahardika Abdi Prawira, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi sekaligus Dosen Pembahas yang telah memberikan masukan, kritik, dan saran yang membangun demi penyempurnaan skripsi.
7. Bapak Mulkan Azhari, S.Kom., M.Kom., selaku Sekretaris Program Studi Sistem Informasi sekaligus Dosen Pembimbing Akademik yang telah memberikan arahan, bimbingan serta dukungan selama masa perkuliahan.
8. Ibu Indah Purnama Sari, S.T., M.Kom., selaku Dosen Pembimbing yang dengan penuh kesabaran telah meluangkan waktu, tenaga, dan pikiran dalam memberikan bimbingan, arahan, koreksi, serta motivasi. Setiap masukan yang diberikan menjadi pembelajaran berharga bagi penulis.
9. Ibu Yohani Syahra, S.Si., M.Kom., selaku Dosen Penguji yang telah memberikan masukan, kritik, serta saran yang membangun demi penyempurnaan skripsi ini.
10. Ayah dan Ibu tercinta serta Adik saya, atas doa yang tiada henti, kasih sayang, pengorbanan, serta dukungan moral dan materi yang menjadi kekuatan terbesar penulis.
11. Teman-teman seperjuangan, khususnya Dina Aulia, Aulia Shiddiq Asy-Syifa, Nabila Yusra, Dwi Shintya dan Rika serta rekan-rekan lainnya, atas kebersamaan, dukungan, dan motivasi selama masa perkuliahan hingga penyusunan skripsi ini.
12. Diri sendiri, yang telah berjuang, bertahan, dan tidak menyerah dalam menyelesaikan setiap prosesnya.

ABSTRAK

Konsumsi gula berlebih dari makanan dan minuman kemasan berkontribusi terhadap meningkatnya kasus obesitas dan diabetes melitus tipe 2 di Indonesia. Rendahnya literasi gizi menyebabkan informasi pada label produk sulit dipahami. Penelitian ini bertujuan merancang sistem klasifikasi kandungan gula untuk membantu konsumen memahami informasi gizi secara lebih mudah. Metode yang digunakan adalah algoritma K-Nearest Neighbor (KNN) dengan dataset dari Open Food Facts, menggunakan atribut energi, gula, karbohidrat, lemak, dan protein. Tahapan penelitian meliputi preprocessing data, pemilihan fitur, serta pembagian data menggunakan train-test split dan cross-validation. Hasil penelitian menunjukkan bahwa sistem mampu mengklasifikasikan produk ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi. Evaluasi model menunjukkan akurasi sebesar 96%, yang menandakan performa model sangat baik. Sistem kemudian diimplementasikan dalam bentuk prototype berbasis web. Sistem ini diharapkan dapat membantu meningkatkan pemahaman masyarakat terhadap informasi gizi dan mendukung pemilihan produk yang lebih sehat.

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

ABSTRACT

Excessive sugar consumption from packaged food and beverages contributes to the rising cases of obesity and type 2 diabetes mellitus in Indonesia. Low nutritional literacy makes label information difficult to understand. This study aims to develop a sugar content classification system to assist consumers in understanding nutritional information. The K-Nearest Neighbor (KNN) algorithm was applied using a dataset from Open Food Facts, with attributes including energy, sugar, carbohydrates, fat, and protein. The process involves data preprocessing, feature selection, and data splitting using train-test split and cross-validation. The results show that the system classifies products into three categories: low, medium, and high. The model achieved an accuracy of 96%, indicating excellent performance. The system was implemented as a web-based prototype. This system is expected to improve public understanding of nutritional information and support healthier product choices.

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

DAFTAR ISI

LEMBAR PENGESAHAN	iii
PERNYATAAN ORISINALITAS.....	Error! Bookmark not defined.
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	Error! Bookmark not defined.
RIWAYAT HIDUP	vi
KATA PENGANTAR.....	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI.....	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	6
BAB II LANDASAN TEORI	8
2.1 Konsumsi Gula dan Dampaknya Terhadap Kesehatan	8
2.2 Data Mining.....	10
2.2.1 Jenis Tugas Data Mining	11
2.2.2 Tahapan Proses data mining	12
2.3 Klasifikasi.....	13
2.4 Dataset Penelitian	15
2.5 Pra-Pemrosesan Data.....	16
2.6 Algoritma K-Nearest Neighbors (KNN)	17
2.6.1 Langkah Kerja Algoritma K-NN	18
2.6.2 Rumus jarak Euclidean Distance	19
2.6.3 Kelebihan dan kelemahan Algoritma KNN.....	21
2.7 Metodologi Penelitian	22
2.8 Penelitian Terdahulu.....	22

BAB III ANALISA DAN PERANCANGAN SISTEM.....	25
3.1 Analisis Permasalahan.....	25
3.2 Analisis Kebutuhan Sistem	26
3.2.1 Kebutuhan Fungsional	26
3.2.2 Kebutuhan Non-Fungsional.....	29
3.3 Analisis Data	30
3.3.1 Sumber Data	31
3.3.2 Deskripsi Variabel	31
3.3.3 Label Kategori	32
3.4 Pre Processing Data.....	33
3.4.1 Data Cleaning	34
3.4.2 Data Transformation.....	34
3.4.3 Feature Selection.....	35
3.4.4 Splitting Data	36
3.5 Proses Klasifikasi Menggunakan Algoritma K-Nearest Neighbor	37
3.6 Kerangka Pemikiran	39
3.6.1 Alur Pemikiran Penelitian.....	39
3.6.2 Diagram Kerangka Pemikiran (Flow Diagram).....	41
3.7 Perancangan Sistem.....	42
3.7.1 Flowchart Sistem	43
3.7.2 Diagram Use Case	45
3.7.3 Entity Relationship Diagram	47
3.8 Implementasi Prototype Web	48
3.8.1 Desain Antarmuka	48
BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM	51
4.1 Kebutuhan Sistem.....	51
4.1.1 Perangkat Keras	51
4.1.2 Perangkat Lunak Sistem	51
4.2 Implementasi Sistem	54
4.2.1 Arsitektur Sistem	54
4.2.2 Implementasi Modul Deteksi Objek (YOLOv8).....	57
4.2.3 Implementasi Modul OCR.....	58
4.2.4 Implementasi Model Klasifikasi KNN	61
4.2.5 Integrasi OCR dan YOLO ke Model	83
4.2.6 Implementasi Antarmuka Sistem.....	88

4.3 Pengujian Sistem	96
4.3.1 Blackbox Testing Halaman Register	97
4.3.2 Blackbox Testing Halaman Login	97
4.3.3 Blackbox Testing Halaman Dashboard	98
4.3.4 Blackbox Testing Halaman Klasifikasi	100
4.3.5 Blackbox Testing Halaman Histori.....	101
4.3.6 Blackbox Testing Halaman Profil.....	102
4.3.7 Blackbox Testing Navigasi Sidebar.....	103
BAB V PENUTUP.....	104
5.1 Kesimpulan.....	104
5.2 Saran.....	105
DAFTAR PUSTAKA.....	107

DAFTAR TABEL

Tabel 2.1 Penjelasan Rumus Persamaan Euclidean Distance	20
Tabel 2.2 Penelitian Terdahulu	23
Tabel 3.1 Variabel Utama Dataset	31
Tabel 4.1 Ringkasan Komponen Sistem	57
Tabel 4.2 Data Latih.....	79
Tabel 4.3 Data Uji.....	79
Tabel 4.4 Pengurutan Jarak	80
Tabel 4.5 Blackbox Halaman Register.....	97
Tabel 4.6 Blackbox Testing Halaman Login	97
Tabel 4.7 Blackbox Testing Halaman Dashboard.....	98
Tabel 4.8 Blackbox Testing Halaman Klasifikasi.....	100
Tabel 4.9 Blackbox Testing Halaman Histori.....	101
Tabel 4.10 Blackbox Testing Halaman Profil.....	102
Tabel 4.11 Blackbox Testing Navigasi Sidebar	103

DAFTAR GAMBAR

Gambar 2.1 Tahapan Proses Data Mining	13
Gambar 2.2 Cara Kerja Algoritma K-NN	19
Gambar 3.1 Diagram Kerangka Pemikiran	41
Gambar 3.2 Flowchart Sistem.....	43
Gambar 3.3 Diagram Use Case	45
Gambar 3.4 Entity Relationship Diagram (ERD)	47
Gambar 3.5 Dashboard Page.....	49
Gambar 3.6 Sugar Classification Page.....	49
Gambar 3.7 History Page	50
Gambar 4.1 Diagram Arsitektur Sistem.....	55
Gambar 4.2 Feature Engineering	63
Gambar 4.3 Balanced Sampling.....	65
Gambar 4.4 Deskripsi Dataset.....	65
Gambar 4.5 Fitur Dataset	66
Gambar 4.6 Distribusi Kelas	68
Gambar 4.7 Distribusi Fitur	68
Gambar 4.8 Hubungan Fitur Terhadap Gula.....	68
Gambar 4.9 Korelasi Antarfitur	69
Gambar 4.10 Label Encoder	70
Gambar 4.11 Train-Test Split Data.....	71
Gambar 4.12 Normalisasi Data	73
Gambar 4.13 Grid Search Validation.....	74
Gambar 4.14 Hasil Optimasi.....	74
Gambar 4.15 Accuracy Testing.....	76
Gambar 4.16 Pengaruh Nilai K Terhadap Akurasi	76
Gambar 4.17 Evaluasi Model.....	77
Gambar 4.18 Confussion Matrix.....	78
Gambar 4. 19 Probabilitas Produk 1	81
Gambar 4. 20 Probabilitas Produk 2	81
Gambar 4. 21 Probabilitas Produk 3	82

Gambar 4.22 Halaman Home.....	88
Gambar 4.23 Halaman Form Login	89
Gambar 4.24 Halaman Form Register	90
Gambar 4.25 Halaman Dashboard	91
Gambar 4.26 Halaman Klasifikasi	92
Gambar 4.27 Halaman History	93
Gambar 4. 28 Halaman Profil	94
Gambar 4.29 Halaman Change Password.....	95

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Konsumsi gula berlebih dari produk makanan dan minuman kemasan telah menjadi isu kesehatan masyarakat yang semakin mengkhawatirkan, terutama di negara berkembang seperti Indonesia. Meningkatnya ketersediaan dan keterjangkauan produk kemasan membuat masyarakat semakin bergantung pada makanan dan minuman instan, yang umumnya mengandung gula tambahan dalam jumlah tinggi. Tren ini berkontribusi pada perubahan pola makan dan meningkatnya risiko penyakit tidak menular. Pada tingkat global, kawasan Asia Tenggara mencatat 106,9 juta penderita diabetes pada tahun 2024, dan angka ini diprediksi melonjak hingga 184,5 juta pada tahun 2050 atau meningkat sekitar 73%. Indonesia menjadi salah satu negara dengan jumlah penderita diabetes tertinggi di kawasan tersebut. Data terbaru menempatkan Indonesia di posisi lima besar dunia untuk prevalensi penyakit kronis ini (Archyde, 2025). Laporan International Diabetes Federation (IDF) tahun 2024 mencatat bahwa jumlah penyandang diabetes di Indonesia mencapai 20,426.4 ribu orang (sekitar 20,4 juta jiwa), dengan prevalensi diabetes sebesar 11,3%, lebih tinggi dibandingkan rata-rata kawasan Asia Tenggara yaitu 10,8% (Diabetes Atlas, 2024). Artinya, lebih dari satu dari sembilan orang dewasa di Indonesia hidup dengan diabetes, dan jumlah ini diprediksi terus meningkat seiring perubahan gaya hidup, pola makan tinggi gula, serta minimnya aktivitas fisik.

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.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana menentukan tingkat kandungan gula pada produk makanan dan minuman kemasan ketika informasi komposisi tidak secara eksplisit menunjukkan jumlah gula, sehingga diperlukan metode klasifikasi yang lebih sistematis dan terukur?
2. Bagaimana merancang sistem yang mampu mengklasifikasikan tingkat kandungan gula secara otomatis berdasarkan atribut nilai gizi yang tersedia pada setiap produk?
3. Bagaimana menerapkan algoritma K-Nearest Neighbor (KNN) sebagai metode klasifikasi pada sistem tersebut, serta bagaimana menentukan parameter terbaik agar hasil klasifikasi lebih akurat?
4. Bagaimana mengevaluasi kinerja sistem klasifikasi berbasis KNN tersebut untuk memastikan tingkat akurasi, presisi, dan keandalannya dalam mengelompokkan kategori kandungan gula?

1.3 Batasan Masalah

Untuk menghindari pembahasan yang terlalu luas, terdapat beberapa batasan permasalahan, yaitu :

1. Data yang digunakan berasal dari dataset Open Food Facts, yang berisi informasi gizi produk makanan dan minuman kemasan, khususnya (atribut seperti energy (kcal), sugar (g), carbohydrates (g), fat (g), dan proteins (g)).
2. Klasifikasi yang dilakukan hanya berfokus pada kategori kandungan gula, yang dibagi menjadi tiga kelas, yaitu rendah, sedang, dan tinggi, berdasarkan nilai kadar gula per 100 gram atau 100 mililiter produk.
3. Algoritma yang digunakan dalam penelitian ini adalah K-Nearest Neighbor (KNN) tanpa perbandingan langsung dengan algoritma lain.
4. Pengujian kinerja sistem difokuskan pada pengukuran metrik akurasi, presisi, recall, dan f1-score, menggunakan metode train-test split atau k-fold cross validation.
5. Implementasi sistem dilakukan dalam bentuk prototype berbasis web sederhana, dengan fokus pada fungsionalitas klasifikasi dan tampilan hasil kategori kandungan gula, bukan pada tampilan antarmuka komersial atau integrasi API produk nyata.
6. Penelitian ini tidak membahas dampak medis atau klinis dari konsumsi gula terhadap kesehatan, melainkan terbatas pada proses klasifikasi data kandungan gula berdasarkan nilai gizi produk kemasan.

1.4 Tujuan Penelitian

Adapun beberapa hal yang menjadi tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk menganalisis data nutrisi dari Open Food Facts agar dapat digunakan sebagai dataset klasifikasi kandungan gula pada produk makanan dan minuman kemasan.
2. Untuk merancang sistem klasifikasi kandungan gula pada produk makanan dan minuman kemasan berbasis algoritma K-Nearest Neighbor (KNN).
3. Untuk mengimplementasikan algoritma K-Nearest Neighbor (KNN) dalam proses klasifikasi kandungan gula berdasarkan atribut nilai gizi produk.
4. Untuk menguji kinerja sistem klasifikasi yang dibangun dengan menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan f1-score, guna mengetahui tingkat efektivitas algoritma KNN dalam mengelompokkan kategori kadar gula.

1.5 Manfaat Penelitian

Adapun manfaat penelitian yang bisa didapat dari penulisan penelitian ini adalah:

a. Bagi Kampus

- 1) Sebagai sumber referensi ilmiah dalam bidang sistem informasi, khususnya terkait penerapan algoritma K-Nearest Neighbor (KNN) dalam klasifikasi data nutrisi.
- 2) Menambah koleksi penelitian di lingkungan kampus yang dapat dijadikan acuan bagi mahasiswa lain dalam mengembangkan penelitian serupa di bidang data mining dan kesehatan digital.
- 3) Mendukung peningkatan kualitas penelitian terapan di kampus dengan menyajikan studi kasus implementasi kecerdasan buatan

(AI) pada isu kesehatan masyarakat, sehingga relevan untuk integrasi dalam materi perkuliahan.

b. Bagi Penulis

- 1) Sebagai penerapan ilmu yang telah diperoleh selama perkuliahan dalam bentuk penelitian nyata di bidang kecerdasan buatan dan sistem informasi.
- 2) Sebagai pengembangan sistem klasifikasi berbasis data gizi dari komposisi makanan dan minuman kemasan.
- 3) Sebagai bukti kompetensi dalam mengintegrasikan analisis data dengan penerapan teknologi untuk mendukung solusi di bidang kesehatan dan nutrisi.

c. Bagi Masyarakat

- 1) Memberikan informasi yang mudah dipahami mengenai tingkat kandungan gula pada produk makanan dan minuman kemasan.
- 2) Meningkatkan kesadaran masyarakat terhadap konsumsi gula berlebih yang dapat berisiko terhadap penyakit seperti diabetes melitus.
- 3) Menjadi bahan edukasi gizi yang dapat membantu masyarakat memilih produk makanan dan minuman secara lebih bijak sesuai kebutuhan tubuh.

BAB II

LANDASAN TEORI

2.1 Konsumsi Gula dan Dampaknya Terhadap Kesehatan

Gula merupakan salah satu komponen utama dalam pola makan modern dan berfungsi sebagai sumber energi cepat bagi tubuh. Dalam konteks ilmu gizi, gula dibedakan menjadi gula alami yang terdapat pada buah atau susu, dan gula tambahan (added sugar) yang ditambahkan ke produk makanan serta minuman kemasan. Seiring meningkatnya konsumsi makanan dan minuman olahan, penggunaan gula tambahan menjadi salah satu penyumbang terbesar asupan gula harian masyarakat. Konsumsi gula berlebih dalam jangka panjang dapat meningkatkan peluang terjadinya Diabetes Melitus tipe 2. Asupan gula yang tinggi menyebabkan penumpukan lemak visceral, terutama di area perut. Peningkatan lemak visceral membuat sel-sel tubuh menjadi kurang sensitif terhadap insulin sehingga glukosa sulit masuk ke dalam sel. Kondisi ini memicu resistensi insulin. Ketika resistensi insulin terjadi, pankreas dipaksa bekerja lebih keras untuk memproduksi insulin dalam jumlah besar agar kadar glukosa tetap terkontrol. Namun, jika kondisi ini berlangsung terus-menerus, sel β pankreas akan mengalami kelelahan sehingga produksi insulin menurun. Saat tubuh tidak lagi mampu menghasilkan insulin yang cukup, terjadilah Diabetes Melitus tipe 2 (Hardy et al., 2012). Setelah diabetes tipe 2 berkembang, kadar gula darah akan terus meningkat apabila tidak dikendalikan. Gula darah tinggi yang berlangsung kronis menyebabkan kerusakan pada dinding pembuluh darah. Kerusakan ini mengganggu aliran darah dan berdampak pada berbagai organ tubuh. Salah satu komplikasi serius yang dapat terjadi adalah penyakit jantung, karena pembuluh darah yang

menuju jantung mengalami penyempitan atau penyumbatan sehingga menghambat proses pemompaan darah dan dapat menyebabkan gagal jantung ataupun penyakit jantung koroner(Poznyak et al., 2022).

Konsumsi gula melebihi batas yang dianjurkan, akan berdampak buruk bagi kesehatan manusia. Menurut penelitian yang dilakukan, telah menunjukkan hubungan yang kuat antara konsumsi gula yang berlebihan dan sejumlah masalah kesehatan yang serius. Efek samping dari konsumsi gula berlebihan termasuk peningkatan risiko obesitas, diabetes, penyakit jantung, dan kanker. Gula dapat menyebabkan resistensi insulin dan memengaruhi metabolisme glukosa, yang merupakan faktor risiko penting untuk berkembangnya Diabetes melitus. Selain itu, “konsumsi gula yang berlebihan dikaitkan dengan peningkatan risiko obesitas karena gula memiliki nilai energi yang tinggi dan memberikan sedikit atau tidak ada nutrisi penting bagi tubuh. Konsumsi gula yang berlebihan juga berkontribusi pada perkembangan diabetes. (Sinaga et al., 2024).

DM (Diabetes Melitus) digolongkan atas DM tergantung insulin (DM tipe 1) dan DM tidak tergantung insulin (DM tipe 2). DM tipe 2 yang tidak ditangani dengan baik akan menimbulkan berbagai komplikasi yaitu komplikasi akut dan komplikasi kronik. Komplikasi kronis DM tipe 2 dapat berupa komplikasi mikrovaskular dan makrovaskular yang dapat menurunkan kualitas hidup penderita. Penyebab utama kematian penyandang DM tipe 2 adalah komplikasi makrovaskular. Komplikasi makrovaskular melibatkan pembuluh darah besar yaitu pembuluh darah koroner, pembuluh darah otak dan pembuluh darah perifer. Spesifik diabetes yang menyerang kapiler dan arteriola retina (retinopati diabetik), glomerulus ginjal (nefropati diabetik) dan saraf-saraf perifer (neuropati diabetik)

(Amelisa Edwina et al., 2015). Melihat tingginya resiko kesehatan akibat Pola konsumsi gula berlebih dan tingginya angka diabetes, diperlukan sistem yang mampu membantu masyarakat mengenali kadar gula pada produk kemasan yang mereka konsumsi.

2.2 Data Mining

Data mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. Data mining mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. Data mining juga dikenal sebagai *Knowledge discovery*, *Knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lain-lain (Arhami & Nasir, 2020). Dalam Data Mining, proses pencarian pola atau informasi yang berguna dari data yang telah dipilih atau diolah dinamakan sebagai Knowledge Data Discovery (KDD).

2.2.1 Jenis Tugas Data Mining

Tugas dalam data mining secara umum dibedakan menjadi dua kategori utama, yaitu tugas prediktif dan tugas deskriptif. Tugas prediktif berfokus pada pembangunan model yang mampu memperkirakan nilai atau label tertentu berdasarkan data historis, sehingga sering digunakan dalam proses klasifikasi, regresi, dan prediksi. Sebaliknya, tugas deskriptif bertujuan untuk mengungkap pola, struktur, atau hubungan tersembunyi dalam data tanpa melakukan prediksi nilai tertentu, mencakup aktivitas seperti clustering, asosiasi, summarization, dan deteksi anomali. Kedua kategori ini saling melengkapi dalam memberikan pemahaman yang komprehensif terhadap data (Agyapong et al., 2016).

1. Tugas Prediktif (Predictive Tasks)

Tugas prediktif berfokus pada pembangunan model yang mampu memperkirakan nilai atau label berdasarkan pola data historis. Pendekatan ini meliputi teknik seperti klasifikasi untuk mengelompokkan data ke dalam kelas tertentu, regresi untuk memprediksi nilai numerik kontinu, serta *time series prediction* untuk memodelkan data berurutan..

2. Tugas Deskriptif (Descriptive Tasks)

Tugas deskriptif bertujuan mengungkap pola, struktur, dan hubungan dalam data tanpa melakukan prediksi nilai tertentu. Teknik yang termasuk di dalamnya meliputi clustering untuk mengelompokkan objek berdasarkan kemiripan, association rule mining untuk menemukan keterkaitan antar item, summarization untuk merangkum karakteristik data, serta anomaly detection untuk mengidentifikasi data yang menyimpang. Pendekatan ini digunakan untuk memahami karakteristik dan struktur dasar dari suatu dataset.

2.2.2 Tahapan Proses data mining

Proses data mining umumnya mengacu pada kerangka kerja *Knowledge Discovery in Databases* (KDD) yang terdiri dari beberapa tahapan yang saling terintegrasi, yaitu (Teguh, 2025):

1. Data Selection

Tahap ini mencakup pemilihan data yang relevan dari berbagai sumber untuk memastikan bahwa hanya atribut dan variabel yang sesuai dengan tujuan analisis yang digunakan. Pemilihan data relevan dari berbagai sumber supaya hanya atribut sesuai tujuan analisis yang dipakai, krusial untuk keberhasilan data mining.

2. Data Preprocessing / Cleaning

Pada tahap ini dilakukan pembersihan data dari berbagai masalah seperti nilai hilang (missing values), duplikasi, inkonsistensi, atau keberadaan noise. Tahap ini bertujuan meningkatkan kualitas data sehingga tidak memengaruhi akurasi hasil analisis. Pembersihan data dari missing values, duplikasi, inkonsistensi, dan noise yang bisa mempengaruhi kualitas dan akurasi hasil.

3. Data Transformation

Data hasil preprocessing kemudian diubah ke dalam diubah ke format yang optimal melalui normalisasi, encoding, agregasi, atau reduksi dimensi. Transformasi dapat berupa normalisasi, standarisasi, encoding variabel kategorikal, agregasi, atau reduksi dimensi. Langkah ini memastikan data berada dalam struktur yang optimal untuk diterapkan.

4. Data Mining

Merupakan tahap inti di mana algoritma atau metode tertentu diterapkan untuk mengekstraksi pola, hubungan, atau pengetahuan dari data. Teknik yang digunakan dapat berupa klasifikasi, clustering, regresi, ataupun analisis asosiasi tergantung pada tujuan penelitian.

5. Interpretation and Evaluation

Tahap terakhir melibatkan evaluasi hasil data mining menggunakan metrik atau indikator tertentu, serta interpretasi temuan agar dapat dipahami dan dimanfaatkan dalam pengambilan keputusan. Pada fase ini dinilai apakah pola yang ditemukan relevan, akurat, dan memiliki nilai praktis sesuai konteks analisis. Evaluasi hasil dan interpretasi pola untuk memastikan relevansi dan kegunaan dalam pengambilan keputusan.



Gambar 2.1 Tahapan Proses Data Mining

2.3 Klasifikasi

Klasifikasi merupakan salah satu tugas utama dalam bidang data mining dan machine learning yang termasuk dalam kategori tugas prediktif (predictive tasks). Klasifikasi adalah proses pengelompokan atau pemetaan objek data ke dalam salah

satu dari sejumlah kelas atau kategori yang telah didefinisikan sebelumnya berdasarkan karakteristik atau fitur yang dimiliki oleh objek tersebut (Han et al., 2011). Berbeda dengan clustering yang bersifat tidak terarah (unsupervised), klasifikasi termasuk dalam pendekatan supervised learning karena memerlukan dataset pelatihan yang telah dilengkapi label kelas untuk setiap recordnya. Secara formal, tugas klasifikasi dapat didefinisikan sebagai berikut: diberikan suatu dataset pelatihan $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ di mana $x_i \in \mathbb{R}^d$ merupakan vektor fitur berdimensi d dan $y_i \in \{C_1, C_2, \dots, C_k\}$ merupakan label kelas dari k kelas yang mungkin, maka tujuan klasifikasi adalah membangun fungsi atau model $f: \mathbb{R}^d \rightarrow \{C_1, C_2, \dots, C_k\}$ yang mampu memprediksi label kelas y^* untuk data baru x^* dengan akurasi setinggi mungkin.

Proses klasifikasi secara umum terbagi menjadi dua tahap utama (Witten et al., 2016):

1. Tahap Pelatihan (Training/Learning Phase)

Algoritma klasifikasi mempelajari pola hubungan antara fitur-fitur (atribut) dengan label kelas pada data pelatihan. Pada tahap ini dihasilkan model klasifikasi yang dapat berupa aturan keputusan, hyperplane, pohon keputusan, probabilitas posterior, atau sekumpulan prototipe (seperti pada instance-based learning).

2. Tahap Pengujian dan Penerapan (Testing/Classification Phase)

Model yang telah dibangun digunakan untuk memprediksi kelas pada data uji atau data baru yang belum memiliki label. Performanya kemudian dievaluasi menggunakan metrik seperti akurasi, presisi, recall, F1-score, atau confusion matrix.

Beberapa paradigma utama dalam klasifikasi meliputi :

1. Probabilistik yaitu mengestimasi probabilitas posterior $P(C_k|x)$
2. Berbasis jarak yaitu mengukur kedekatan data baru dengan data pelatihan
3. Berbasis hyperplane yaitu mencari batas pemisah linear atau non-linear di ruang fitur
4. Berbasis pohon yaitu membagi ruang fitur secara rekursif berdasarkan nilai atribut
5. Berbasis neural network yaitu meniru jaringan saraf biologis dengan lapisan neuron dan fungsi aktivasi.

2.4 Dataset Penelitian

Dataset penelitian yang digunakan dalam studi ini berasal dari Open Food Facts, yaitu basis data nutrisi terbuka berskala global yang dikembangkan secara kolaboratif sejak 2012 oleh komunitas sukarelawan dan organisasi independen. Basis data ini berisi lebih dari 4 juta produk makanan dan minuman kemasan dari sekitar 182 negara, dengan masing-masing produk dilengkapi lebih dari 200 atribut yang mencakup informasi komposisi gizi (seperti nilai energi, lemak, karbohidrat, gula, protein, dan garam per 100 gram/ml), daftar bahan baku, kategori produk, label sertifikasi, asal-usul, serta metadata kemasan seperti foto dan kode barcode (Open Food Fact, 2025). Dataset ini dipilih karena cakupannya yang sangat luas dan bersifat open-source di bawah lisensi Open Database License (ODbL), sehingga memungkinkan akses gratis dan penggunaan ulang untuk tujuan penelitian, termasuk proses klasifikasi kandungan gula menggunakan pembelajaran mesin. Open Food Facts diperbarui harian oleh lebih dari 25.000 kontributor melalui aplikasi mobile dan web, sehingga menyediakan data terkini.

2.5 Pra-Pemrosesan Data

Data Preprocessing bertujuan untuk mengubah data mentah menjadi data yang berkualitas sehingga data layak untuk diolah pada tahapan selanjutnya. Tahapan ini dilakukan pada data mentah untuk menghilangkan data yang bermasalah atau inkonsisten. Data yang bermasalah seperti data yang noise atau yang mengandung error (Alghifari & Juardi, 2021).

tahapan tahapan dalam pra-pemrosesan data adalah :

1. Data Cleaning

Tahap pertama yang perlu dilakukan ketika akan preprocessing data adalah data cleaning atau membersihkan data. Artinya, data mentah yang telah diperoleh perlu diseleksi kembali. Kemudian, hapus atau hilangkan data-data yang tidak lengkap, tidak relevan, dan tidak akurat. Dengan melakukan tahap ini, Anda akan menghindari kesalahpahaman ketika menganalisis data tersebut.

2. Data Integration

Karena data preprocessing akan menggabungkan beberapa data dalam suatu dataset, maka kita harus mengecek data-data yang datang dari berbagai sumber tersebut supaya memiliki format yang sama.

3. Transformasi Data

Proses berikutnya yang harus dilakukan adalah transformasi data. Seperti yang telah dijelaskan di atas, data akan diambil dari berbagai sumber yang kemungkinan memiliki perbedaan format. Kita harus menyamakan seluruh data yang terkumpul supaya dapat mempermudah proses analisis data.

4. Mengurangi Data

Tahap terakhir yang perlu dilakukan adalah mengurangi jumlah data (data reduction). Maksudnya adalah kita harus mengurangi sampel data yang diambil, tetapi dengan catatan, tidak akan mengubah hasil analisis data. Ada tiga teknik yang bisa diterapkan saat melakukan pengurangan data, yakni dimensionality reduction (pengurangan dimensi), numerosity reduction (pengurangan jumlah), dan data compression (kompresi data).

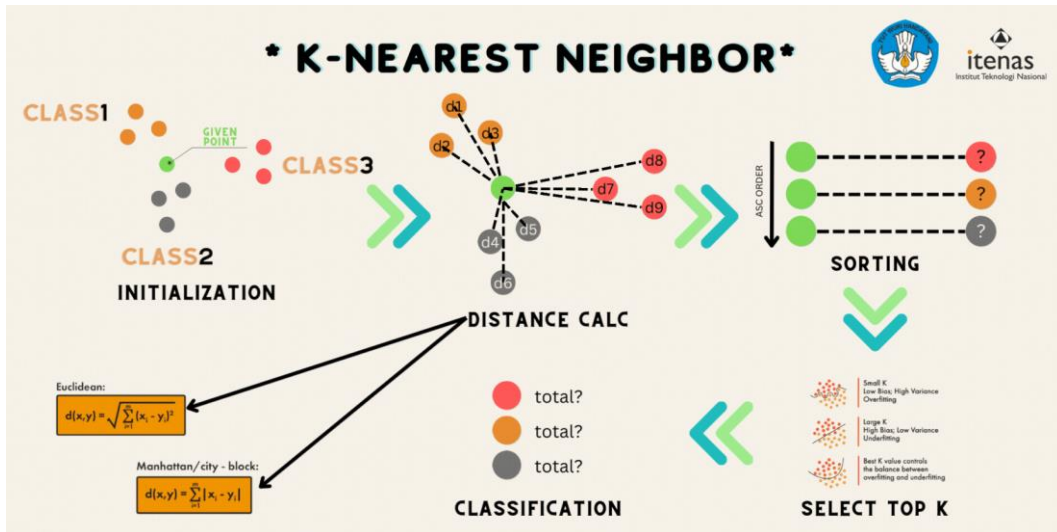
2.6 Algoritma K-Nearest Neighbors (KNN)

Algoritma K-Nearest Neighbors (KNN) merupakan salah satu metode supervised learning yang digunakan untuk tugas klasifikasi maupun regresi. KNN bekerja berdasarkan prinsip similarity-based learning, yaitu menentukan kelas atau nilai suatu data baru dengan melihat sejumlah tetangga terdekat (nearest neighbors) dalam ruang fitur dengan mempertimbangkan keserupaan value and weight (Khairunnisa & Martiano, 2025). Kedekatan antar data dihitung menggunakan ukuran jarak tertentu, seperti Euclidean Distance, Manhattan Distance, atau Minkowski Distance. Karena tidak memerlukan proses pelatihan model secara eksplisit, KNN dikategorikan sebagai metode lazy learner yang melakukan perhitungan hanya ketika proses prediksi berlangsung. Sebagai salah satu metode yang banyak digunakan dalam berbagai tugas klasifikasi, KNN termasuk ke dalam algoritma klasifikasi populer sebagaimana disebutkan oleh (Ramadhani et al., 2023), yang menempatkan KNN bersama algoritma umum lainnya seperti Naive Bayes, Decision Tree, dan SVM dalam kelompok metode klasifikasi yang sering digunakan.

2.6.1 Langkah Kerja Algoritma K-NN

Proses kerja algoritma K-Nearest Neighbors (KNN) berlangsung melalui beberapa tahapan sistematis yang bertujuan menentukan kelas atau nilai prediksi berdasarkan kedekatan data baru dengan data yang telah diketahui labelnya. Secara umum, langkah kerja KNN meliputi:

1. Menentukan nilai k , yaitu jumlah tetangga terdekat yang akan digunakan dalam proses klasifikasi atau regresi. Nilai k dipilih melalui pertimbangan empiris atau teknik validasi seperti *cross-validation*.
2. Menghitung jarak antara data baru dan seluruh data pada dataset pelatihan menggunakan ukuran jarak tertentu, seperti *Euclidean Distance*, *Manhattan Distance*, atau *Minkowski Distance*. Perhitungan jarak ini menjadi dasar penentuan kedekatan antar objek dalam ruang fitur.
3. Mengurutkan hasil perhitungan jarak dari yang terkecil hingga terbesar untuk mengidentifikasi k data dengan nilai jarak paling kecil yang dianggap sebagai tetangga terdekat.
4. Melakukan pengambilan keputusan berdasarkan mayoritas kelas (untuk klasifikasi) atau rata-rata nilai (untuk regresi) dari k tetangga terdekat tersebut.
5. Menetapkan hasil prediksi sebagai kelas atau nilai akhir yang diberikan kepada data baru sesuai hasil perhitungan dan agregasi pada langkah sebelumnya.



Gambar 2.2 Cara Kerja Algoritma K-NN

2.6.2 Rumus jarak Euclidean Distance

Jarak Euclidean merupakan salah satu ukuran jarak yang paling sering digunakan dalam algoritma K-Nearest Neighbors (KNN) karena mampu merepresentasikan kedekatan geometris antar objek dalam ruang fitur. Ukuran ini mendasarkan perhitungannya pada jarak garis lurus (straight-line distance) antara dua titik dalam ruang berdimensi banyak (n-dimensional feature space).

Secara matematis, jarak Euclidean antara dua vektor data

$$x = (x_1, x_2, \dots, x_n),$$

$$y = (y_1, y_2, \dots, y_n)$$

didefinisikan sebagai persamaan:

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

Dibawah ini adalah table penjelasan dari persamaan rumus Euclidean diatas.

Tabel 2.1 Penjelasan Rumus Persamaan Euclidean Distance

Simbol	Keterangan	Penjelasan Konseptual
x	Vektor data pertama yang merepresentasikan satu objek dalam ruang fitur (x_1, x_2, \dots, x_n) .	Menggambarkan posisi objek pertama di dalam ruang multidimensi berdasarkan nilai fitur yang dimilikinya.
y	Vektor data kedua yang dibandingkan terhadap x dan memiliki struktur fitur yang sama (y_1, y_2, \dots, y_n) .	Menunjukkan posisi objek pembanding dalam ruang yang sama untuk dianalisis jaraknya dari objek pertama.
n	Jumlah fitur atau atribut yang menentukan dimensi ruang analisis.	Mewakili banyaknya dimensi yang digunakan untuk mengukur perbedaan antar objek.
x_i	Nilai atribut ke- i pada objek x .	Koordinat objek pertama pada dimensi ke- i .
y_i	Nilai atribut ke- i pada objek y .	Koordinat objek pembanding pada dimensi ke- i .
$(x_i - y_i)^2$	Kuadrat selisih nilai atribut ke- i dari kedua objek.	Besaran yang menunjukkan kontribusi perbedaan pada dimensi

		tersebut terhadap total jarak.
$\Sigma = 1n$	Operasi penjumlahan atas setiap komponen selisih kuadrat dari seluruh dimensi.	Menggabungkan seluruh kontribusi jarak dari tiap dimensi menjadi satu nilai kumulatif.
$\sqrt{\quad}$	Operasi akar kuadrat untuk memperoleh jarak Euclidean akhir.	Menghasilkan nilai jarak geometris yang merepresentasikan kedekatan dua objek dalam ruang fitur.

Rumus ini menghitung jarak lurus (straight-line) Euclidean, yang sensitif terhadap skala fitur sehingga memerlukan normalisasi data sebelumnya untuk hasil optimal (Asri Mulyani et al., 2025).

2.6.3 Kelebihan dan kelemahan Algoritma KNN

Algoritma KNN memiliki beberapa kelebihan yang menjadikannya populer dalam berbagai aplikasi klasifikasi. Pertama, algoritma ini sangat mudah dipahami dan diimplementasikan karena tidak memerlukan proses pelatihan yang kompleks (Wijaya et al., 2019). Kedua, KNN bersifat non-parametrik sehingga tidak mengasumsikan distribusi data tertentu dan mampu menangani data yang bersifat non-linear dengan baik (Isnain et al., 2021). Ketiga, algoritma ini fleksibel karena dapat digunakan baik untuk tugas klasifikasi maupun regresi, serta bersifat lazy learning sehingga mudah beradaptasi dengan penambahan data baru tanpa pelatihan ulang secara keseluruhan.

Namun, KNN juga memiliki beberapa kelemahan signifikan. Komputasi menjadi sangat mahal pada dataset besar karena jarak harus dihitung terhadap seluruh data pelatihan pada setiap prediksi (Utami et al., 2021). Selain itu, algoritma ini sensitif terhadap noise dan outlier, terutama ketika nilai k kecil, serta rentan mengalami overfitting atau underfitting jika nilai k tidak dipilih dengan tepat.

2.7 Metodologi Penelitian

Metodologi penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional, karena penelitian berfokus pada pengukuran dan evaluasi performa algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan tingkat kandungan gula berdasarkan data numerik nilai gizi. Jenis penelitian yang digunakan adalah penelitian kuantitatif eksplanatori, yang bertujuan menjelaskan hubungan antaratribut nutrisi melalui pengujian model secara sistematis. Data yang digunakan merupakan data sekunder berupa informasi nilai gizi produk makanan dan minuman kemasan, yang kemudian dianalisis menggunakan teknik statistik prediktif untuk mengukur efektivitas model. Pemilihan metodologi ini didasarkan pada kebutuhan akan hasil yang objektif, terukur, dan dapat direplikasi, sehingga evaluasi akurasi, presisi, recall, dan F1-score dapat digunakan sebagai dasar penilaian kinerja model secara ilmiah.

2.8 Penelitian Terdahulu

Penelitian mengenai klasifikasi kandungan gula pada produk makanan dan minuman kemasan dengan menggunakan algoritma K-Nearest Neighbor (KNN) telah dilakukan dalam berbagai bidang, khususnya dalam aplikasi klasifikasi data terkait kesehatan dan makanan.

Tabel 2.2 Penelitian Terdahulu

No	Judul Penelitian	Penulis / Tahun	Metode	Hasil Temuan
1	Penggunaan Algoritma K-Nearest Neighbor Untuk Klasifikasi Data Diabetes Pada Wanita	(Shelin Sahira et al., 2024)	K-Nearest Neighbor (KNN)	Akurasi 79.93%, presisi kelas 78.19%, recall kelas hingga 96.45%. Menunjukkan kekuatan KNN untuk klasifikasi data yang berhubungan dengan kadar gula darah.
2	Penerapan Algoritma K-Nearest Neighbor (KNN) dalam klasifikasi penyakit diabetes melitus berdasarkan gejala klinis	(Dwi Fasnuari et al., 2022)	K-Nearest Neighbor (KNN)	Akurasi hingga 93%, proses normalisasi data dan penggunaan jarak Euclidean dengan nilai K optimal. Acuan dalam klasifikasi data kandungan gula pada produk makanan dan minuman kemasan.
3	Sistem Klasifikasi Penyakit Diabetes Mellitus dengan K-Nearest Neighbor	(Yunita, 2016)	K-Nearest Neighbor (KNN)	Tingkat keberhasilan klasifikasi sampai 90%. Menjelaskan klasifikasi variabel input dan target dalam konteks kesehatan gula darah.
4	Penggunaan Algoritma K-Nearest Neighbor (KNN) untuk	(Davies et al., 2022)	K-Nearest Neighbor (KNN) dengan pemilihan	KNN berhasil memprediksi kandungan gula tambahan dengan

	Prediksi Kandungan Gula Tambahan pada Produk Makanan Kemasan		hyperparameter optimal (jumlah tetangga dan jarak Manhattan)	validitas tinggi pada dataset besar produk kemasan dari database nutrisi Amerika, menunjukkan potensi penggunaan algoritma ini untuk monitor dan intervensi pengurangan gula dalam produk makanan/minuman kemasan.
5	Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Klasifikasi Kualitas dan Kandungan Manis pada Produk Pertanian (Buah)	(Iwan Sudipa et al., 2024)	K-Nearest Neighbor (KNN) dengan validasi silang 5-fold untuk klasifikasi kualitas buah berdasarkan atribut manis, ukuran, dan atribut lain terkait	KNN menunjukkan akurasi, presisi, dan recall yang tinggi, menandakan kemampuan model untuk mengklasifikasikan kandungan manis secara efektif, yang relevan untuk aplikasi klasifikasi kandungan gula dalam produk makanan/minuman kemasan.

BAB III

ANALISA DAN PERANCANGAN SISTEM

3.1 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.

Berdasarkan uraian tersebut, permasalahan utama yang mendasari penelitian ini dapat dirumuskan sebagai berikut:

1. Rendahnya tingkat pemahaman masyarakat terhadap informasi nilai gizi pada label pangan kemasan, terutama terkait kandungan gula.

2. Tidak adanya alat bantu yang sederhana dan mudah digunakan oleh konsumen untuk menilai kategori kandungan gula suatu produk secara cepat.
3. Diperlukan sistem klasifikasi berbasis kecerdasan buatan yang mampu menghasilkan informasi kategori gula yang mudah dipahami sebagai pendukung pengambilan keputusan konsumen.

Penelitian ini dilakukan untuk merespons kebutuhan tersebut melalui pengembangan model klasifikasi kandungan gula berbasis algoritma K-Nearest Neighbor (KNN) yang diimplementasikan ke dalam bentuk sistem sederhana berbasis web. Sistem ini diharapkan dapat membantu meningkatkan literasi gizi masyarakat dan mendukung perilaku konsumsi yang lebih sehat.

3.2 Analisis Kebutuhan Sistem

Analisis kebutuhan sistem dilakukan untuk memastikan bahwa aplikasi yang dikembangkan dapat menjawab permasalahan utama, yaitu memberikan klasifikasi tingkat kandungan gula yang akurat, cepat, dan mudah dipahami oleh pengguna awam sekalipun. Kebutuhan sistem dibagi menjadi dua kelompok utama, yaitu kebutuhan fungsional dan kebutuhan non-fungsional.

3.2.1 Kebutuhan Fungsional

Kebutuhan fungsional merupakan spesifikasi perilaku dan fungsi utama sistem yang harus tersedia agar proses klasifikasi kandungan gula dapat berjalan dengan baik. Sistem yang dikembangkan dalam penelitian ini dirancang untuk menerima input nilai nutrisi dari pengguna dan menghasilkan kategori kadar gula menggunakan algoritma K-Nearest Neighbor (KNN) berdasarkan dataset internal

yang telah dipersiapkan sebelumnya. Adapun kebutuhan fungsional sistem adalah sebagai berikut:

1. Sistem dapat memuat dan menyimpan dataset internal sebagai basis pengetahuan KNN

Sistem harus memiliki mekanisme untuk menyimpan dataset internal yang telah melalui proses kurasi dan pembersihan sebelumnya. Dataset tersebut digunakan sebagai data latih yang menjadi acuan dalam perhitungan jarak pada algoritma KNN. Pengguna tidak dapat mengubah atau mengunggah dataset baru untuk menjaga konsistensi model.

2. Sistem dapat menerima input nilai nutrisi produk dari pengguna.

Sistem menyediakan formulir input yang memungkinkan pengguna memasukkan nilai komponen nutrisi seperti gula (100g atau ml), energi, lemak, karbohidrat, dan protein. Data ini berfungsi sebagai data uji yang akan diklasifikasikan oleh sistem.

3. Sistem dapat melakukan preprocessing terhadap dataset internal.

Sistem harus mampu melakukan tahapan pra-pemrosesan (preprocessing) secara otomatis, mencakup pembersihan data, penanganan nilai hilang, normalisasi/standardisasi, dan pemilihan fitur. Tahapan ini dilakukan sebelum model digunakan sehingga dataset internal berada dalam kondisi yang siap untuk proses klasifikasi.

4. Sistem dapat menerapkan algoritma K-Nearest Neighbor (KNN).

Sistem harus dapat menjalankan algoritma KNN untuk menghitung jarak antara input pengguna dengan seluruh data latih menggunakan Euclidean

Distance, memilih nilai k yang telah ditentukan, serta menentukan kelas berdasarkan hasil voting mayoritas dari tetangga terdekat.

5. Sistem dapat mengklasifikasikan produk ke dalam kategori tingkat kandungan gula.

Berdasarkan hasil perhitungan KNN, sistem harus mampu mengelompokkan produk ke dalam salah satu kategori kandungan gula, yaitu: rendah, sedang, atau tinggi. Mekanisme penentuan kategori harus konsisten dengan standar pembagian kelas yang ditetapkan pada penelitian.

6. Sistem dapat menampilkan hasil klasifikasi secara jelas kepada pengguna.

Sistem menampilkan hasil klasifikasi dalam bentuk informasi yang mudah dipahami, mencakup kategori gula produk serta penjelasan singkat mengenai makna kategori tersebut. Hasil ditampilkan secara langsung setelah pengguna memasukkan data nutrisi.

7. Sistem dapat menampilkan evaluasi model berdasarkan dataset internal.

Untuk keperluan penelitian dan analisis performa model, sistem harus dapat menghitung dan menampilkan metrik evaluasi seperti akurasi, presisi, recall, f1-score, serta confusion matrix. Evaluasi dilakukan terhadap dataset internal menggunakan metode pembagian data atau cross-validation.

8. Sistem menyediakan antarmuka berbasis web yang mudah digunakan.

Sistem harus memiliki antarmuka web yang sederhana, responsif, dan mudah digunakan oleh pengguna umum. Antarmuka ini mencakup halaman input nilai nutrisi, serta halaman hasil klasifikasi.

3.2.2 Kebutuhan Non-Fungsional

Kebutuhan non-fungsional menggambarkan karakteristik kualitas yang harus dimiliki sistem agar dapat beroperasi secara optimal, handal, dan mudah digunakan. Adapun kebutuhan non-fungsional dalam penelitian ini adalah sebagai berikut:

1. Usability (Kemudahan Penggunaan)

Sistem harus memiliki antarmuka yang sederhana, intuitif, dan mudah dipahami oleh pengguna, termasuk mereka yang tidak memiliki latar belakang teknis. Elemen navigasi harus jelas, dan alur penggunaan—mulai dari pengisian nilai nutrisi hingga mendapatkan hasil klasifikasi—harus dapat dilakukan tanpa memerlukan panduan kompleks. Penyajian hasil klasifikasi juga harus dirancang secara informatif dan tidak membingungkan agar pengguna dapat memahami informasi tingkat kandungan gula dengan cepat.

2. Performance (Kinerja Sistem)

Sistem harus mampu memproses input pengguna dan menghasilkan hasil klasifikasi dalam waktu singkat. Efisiensi perhitungan sangat penting, terutama pada algoritma KNN yang melakukan perhitungan jarak terhadap seluruh data latih. Selain itu, sistem harus mampu menangani dataset internal dalam jumlah yang cukup besar tanpa menyebabkan penurunan performa atau waktu tanggap yang berlebih. Optimasi seperti normalisasi data dan pemilihan fitur harus mendukung kinerja sistem tetap stabil.

3. Reliability (Keandalan Sistem)

Sistem harus memberikan hasil klasifikasi yang konsisten ketika menerima input yang sama. Keandalan sistem mencakup stabilitas saat dijalankan, kemampuan menangani variasi input pengguna, serta ketahanan terhadap

kegagalan atau error internal. Sistem juga harus mampu menjaga integritas dataset internal agar model KNN dapat bekerja secara akurat dan tidak berubah akibat kesalahan input atau proses.

4. Maintainability

Sistem harus dirancang dengan struktur kode dan arsitektur yang mudah dipelihara dan dikembangkan. Dokumentasi kode perlu disediakan agar perubahan pada dataset internal, pengaturan nilai k , atau penambahan fitur lain dapat dilakukan tanpa mengganggu fungsi utama sistem. Kemudahan pemeliharaan juga mencakup kemampuan memperbarui algoritma atau menambah metode klasifikasi baru apabila dibutuhkan pada penelitian selanjutnya.

5. Security

Sistem harus mampu melindungi integritas data dengan mencegah perubahan tidak sah terhadap dataset internal maupun komponen pemrosesan algoritma. Selain itu, sistem harus melakukan validasi terhadap seluruh input pengguna untuk mencegah kesalahan format, data tidak realistis, atau input berbahaya. Mekanisme keamanan ini penting untuk memastikan sistem tetap stabil dan tidak mengalami gangguan yang dapat memengaruhi hasil klasifikasi.

3.3 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).

3.3.1 Sumber Data

Data yang digunakan dalam penelitian ini berasal dari Open Food Facts, yaitu basis data nutrisi internasional yang berisi informasi nilai gizi dan komposisi produk makanan serta minuman kemasan.

Untuk memastikan efisiensi pemrosesan dan kesesuaian dengan kebutuhan klasifikasi, penelitian ini tidak menggunakan seluruh dataset, melainkan mengambil irisan data (subset) sebanyak ± 1.000 entri yang memenuhi kriteria:

1. Memiliki informasi komposisi gizi yang lengkap, terutama `sugar_100g`, `energy_100g`, `fat_100g`, `carbohydrates_100g`, dan `proteins_100g`.
2. Merupakan produk makanan atau minuman kemasan yang relevan.
3. Data telah melalui proses penyaringan untuk menghilangkan nilai yang tidak valid atau ekstrem.

Subset ini kemudian digunakan sebagai dataset internal untuk proses pelatihan dan klasifikasi menggunakan algoritma KNN.

3.3.2 Deskripsi Variabel

Dataset yang digunakan dalam penelitian mengandung beberapa variabel nutrisi yang berfungsi sebagai fitur (features) dalam proses klasifikasi. Tabel berikut menggambarkan variabel utama yang digunakan:

Tabel 3.1 Variabel Utama Dataset

Variabel	Tipe Data	Penjelasan
<code>sugar_100g</code>	numerik	Kandungan gula total dalam 100 gram produk. Menjadi indikator utama dalam klasifikasi

energi_100g	numerik	Nilai energi dalam kilokalori per 100 gram
carbohydrates_100g	numerik	Total karbohidrat per 100 gram
fat_100g	numerik	Total lemak per 100 gram
proteins_100g	numerik	Kandungan protein dalam 100 gram
category_label	kategorikal	Label hasil klasifikasi (rendah, sedang, tinggi) yang ditentukan menggunakan standar rumus perhitungan gula oleh WHO.

Variabel nutrisi tersebut dipilih karena memiliki hubungan yang kuat dengan profil gizi produk dan relevan dalam pengukuran kandungan gula menurut standar kesehatan.

3.3.3 Label Kategori

Label kategori digunakan sebagai target (target class) dalam model klasifikasi. Penelitian ini tidak menggunakan label kategori bawaan dari Open Food Facts karena:

1. Open Food Facts tidak menyediakan kategori kadar gula secara langsung.
2. Setiap produk memiliki variasi data yang luas sehingga diperlukan standar kategorisasi yang seragam.
3. WHO memiliki pedoman angka rekomendasi konsumsi gula yang dapat dijadikan acuan ilmiah.

Oleh karena itu, label kategori ditentukan sendiri (self-labeling) menggunakan rumus dan standar WHO mengenai batas konsumsi gula harian. Proses pelabelan dilakukan berdasarkan nilai *sugar_100g* dengan mempertimbangkan batas konsumsi gula harian WHO (≤ 50 gram/hari, idealnya ≤ 25 gram/hari), Kategori ditentukan sebagai berikut :

1. Rendah Gula

Produk dengan kandungan gula $\leq X$ gram per 100g, sesuai ambang batas rendah menurut perhitungan WHO.

2. Sedang Gula

Produk dengan kandungan gula dalam rentang $> X$ hingga $\leq Y$ gram per 100g.

3. Tinggi Gula

Produk dengan kandungan gula $> Y$ gram per 100g, yang melebihi batas aman konsumsi menurut standar WHO.

3.4 Pre Processing Data

Preprocessing data merupakan tahapan penting untuk memastikan bahwa dataset berada dalam kondisi yang optimal sebelum digunakan dalam proses klasifikasi. Mengingat data yang diperoleh dari Open Food Facts memiliki bentuk yang tidak seragam, banyak nilai hilang, dan tingkat kelengkapan yang bervariasi, langkah preprocessing diperlukan untuk meningkatkan kualitas data dan mendukung performa algoritma K-Nearest Neighbor (KNN) yang sensitif terhadap kualitas dan skala data. Tahapan preprocessing yang dilakukan dalam penelitian ini meliputi data cleaning, data transformation, feature selection, dan splitting data.

3.4.1 Data Cleaning

Tahap data cleaning dilakukan untuk menghilangkan atau memperbaiki data yang tidak valid, tidak lengkap, atau tidak konsisten. Mengingat struktur dataset Open Food Facts sangat masif dan heterogen, proses pembersihan data menjadi langkah fundamental.

Adapun proses data cleaning yang dilakukan meliputi:

1. Menghapus Menghapus entri dengan nilai nutrisi yang tidak lengkap, terutama untuk variabel utama seperti `sugar_100g`, `energy_100g`, `carbohydrates_100g`, `fat_100g`, dan `proteins_100g`. Data dengan nilai kosong (missing values) tidak dapat digunakan dalam perhitungan jarak pada KNN.
2. Menghapus data dengan nilai ekstrem atau tidak realistis, seperti nilai gula negatif, energi nol, atau nutrisi melebihi batas wajar menurut standar pangan.
3. Menghilangkan duplikasi data, baik berdasarkan nama produk maupun identitas unik produk yang terdapat dalam dataset.
4. Menyeragamkan format data, misalnya memastikan satuan nutrisi berada dalam format per 100 gram atau per 100 mililiter sesuai standar yang berlaku dalam Open Food Facts.

3.4.2 Data Transformation

Pada tahap ini, data yang telah dibersihkan perlu ditransformasikan agar siap untuk diproses menggunakan algoritma KNN. Transformasi dilakukan untuk memastikan seluruh fitur memiliki skala yang seragam, mengingat KNN sangat dipengaruhi oleh perbedaan skala antar fitur.

Adapun bentuk transformasi yang dilakukan meliputi:

1. Normalisasi / Standarisasi Fitur

Normalisasi menggunakan metode Min-Max Scaling atau standarisasi Z-score dilakukan untuk menyetarakan rentang nilai semua fitur.

Contoh:

- a. energy_100g dapat memiliki rentang puluhan hingga ratusan
- b. sugar_100g memiliki rentang 0–100

Tanpa normalisasi, fitur dengan rentang lebih besar akan mendominasi perhitungan jarak.

2. Konversi nilai nutrisi tertentu agar seragam, jika terdapat data dalam bentuk per serving.

3. Penentuan label kategori (Encoding Target Class)

Berdasarkan perhitungan manual menggunakan standar WHO, setiap produk diberi label:

- a. Rendah gula
- b. Sedang gula
- c. Tinggi gula

3.4.3 Feature Selection

Feature selection dilakukan untuk memilih fitur paling relevan yang berpengaruh terhadap klasifikasi kadar gula. Pemilihan fitur penting karena mengurangi beban komputasi KNN, Menghindari noise dari fitur yang tidak relevan, dan Meningkatkan akurasi model. Fitur-fitur yang dipilih berdasarkan analisis nutrisi dan relevansi terhadap kandungan gula adalah:

1. sugar_100g → fitur utama yang menggambarkan kadar gula aktual.

2. `carbohydrates_100g` → berkorelasi dengan kandungan gula karena gula merupakan bagian dari karbohidrat.
3. `energy_100g` → menunjukkan kontribusi kalori produk; gula merupakan salah satu sumber kalori.
4. `fat_100g` → digunakan untuk menyeimbangkan klasifikasi karena beberapa produk rendah gula tetapi tinggi lemak.
5. `proteins_100g` → digunakan untuk memperkaya dimensi fitur dalam ruang vektor.

3.4.4 Splitting Data

Sebelum dilakukan klasifikasi, dataset internal perlu dibagi menjadi dua bagian utama:

1. Data latih (training data) = Digunakan sebagai acuan bagi algoritma KNN untuk menghitung jarak. Persentase umum yang digunakan adalah 80% dari total dataset.
2. Data uji (testing data) = Digunakan untuk mengukur performa model sebelum digunakan pada input pengguna. Sebanyak 20% dari total dataset biasanya digunakan sebagai data uji.

Pembagian data ini dilakukan secara acak (randomized splitting) agar representasi setiap kategori (rendah, sedang, tinggi) seimbang dalam kedua subset tersebut.

Metode pembagian data yang digunakan dapat berupa:

- a. Train-test split (80:20) → digunakan untuk evaluasi standar
- b. K-Fold Cross Validation → untuk evaluasi lebih stabil dan reliabel

Hasil splitting ini menjadi dasar untuk menghitung akurasi, presisi, recall, dan f1-score dari model KNN.

3.5 Proses Klasifikasi Menggunakan Algoritma K-Nearest Neighbor

Setelah data melewati tahapan cleaning, transformasi, seleksi fitur, dan pembagian train-test, proses klasifikasi tingkat kandungan gula dilakukan dengan algoritma K-Nearest Neighbor (KNN).

Langkah-langkah klasifikasi dengan KNN pada penelitian ini adalah sebagai berikut:

1. Menentukan nilai K optimal

Pemilihan nilai K dilakukan melalui serangkaian eksperimen dengan menggunakan $K = 3, 5, 7, 9,$ dan 11 . Evaluasi dilakukan menggunakan metode 10-fold cross validation pada data pelatihan untuk memperoleh nilai K yang memberikan performa terbaik. Berdasarkan hasil pengujian, $K = 5$ menghasilkan akurasi tertinggi sehingga ditetapkan sebagai nilai K optimal pada penelitian ini.

2. Normalisasi data

Semua fitur numerik (energy-kcal_100g, fat_100g, saturated-fat_100g, carbohydrates_100g, sugars_100g, fiber_100g, proteins_100g, salt_100g) dinormalisasi menggunakan Min-Max Scaling ke rentang $[0,1]$ agar tidak ada fitur yang mendominasi karena skala yang berbeda.

3. Perhitungan jarak

Ketika ada data baru (produk yang akan diprediksi), sistem menghitung jarak Euclidean terhadap seluruh data latih menggunakan rumus:

$$d(x, y) = \sqrt{[(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_8 - y_8)^2]}$$

Setiap data baru dibandingkan dengan data latih yang telah dikumpulkan dan diklasifikasikan sebelumnya. Data latih tersebut telah diberi label kategori

kadar gula berdasarkan informasi nutrisi pada produk makanan dan minuman kemasan. Proses perhitungan jarak dilakukan dengan menghitung selisih nilai masing-masing fitur, kemudian mengkuadratkannya untuk menghilangkan nilai negatif dan memperbesar perbedaan yang signifikan. Selanjutnya, seluruh nilai kuadrat dijumlahkan untuk memperoleh squared distance, dan akar kuadratnya diambil untuk mendapatkan jarak Euclidean sesungguhnya. Nilai jarak ini digunakan sebagai dasar dalam menentukan kedekatan antara data baru dan data latih.

4. Pengurutan dan pemilihan tetangga

Setelah seluruh jarak dihitung, sistem melakukan pengurutan berdasarkan jarak terkecil dan mengambil $K = 5$ data latih terdekat untuk digunakan sebagai tetangga dalam proses klasifikasi.

5. Voting mayoritas

Penentuan kelas dilakukan dengan menerapkan mekanisme majority voting. Kelas yang paling banyak muncul di antara lima tetangga terdekat ditetapkan sebagai hasil prediksi. Contoh:

- a. 3 tetangga kelas “Tinggi”
- b. 1 tetangga kelas “Sedang”
- c. 1 tetangga kelas “Rendah”

Hasil prediksi: Tinggi (confidence 60%)

6. Output akhir

Sistem mengembalikan label kategori gula:

- a. Rendah (≤ 5 g)
- b. Sedang ($> 5 - \leq 22,5$ g)

c. Tinggi ($> 22,5 - \leq 40$ g)

disertai persentase confidence dan rekomendasi konsumsi.

Proses ini diimplementasikan menggunakan library scikit-learn (KNeighborsClassifier) dan diintegrasikan ke dalam backend prototipe web, sehingga pengguna cukup memasukkan komposisi nutrisi produk, dan sistem secara otomatis menjalankan perhitungan serta menghasilkan kategori kadar gula beserta tingkat kepercayaannya.

3.6 Kerangka Pemikiran

Kerangka pemikiran merupakan alur logis yang menggambarkan hubungan antara konsep, teori, dan variabel yang digunakan dalam penelitian. Kerangka ini menjelaskan bagaimana proses analisis dimulai dari identifikasi masalah hingga menghasilkan solusi yang ditetapkan. Pada penelitian ini, kerangka pemikiran disusun untuk menunjukkan tahapan konseptual dalam membangun sistem klasifikasi kandungan gula pada produk makanan dan minuman kemasan menggunakan algoritma K-Nearest Neighbors (KNN).

3.6.1 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.

3.6.2 Diagram Kerangka Pemikiran (Flow Diagram)



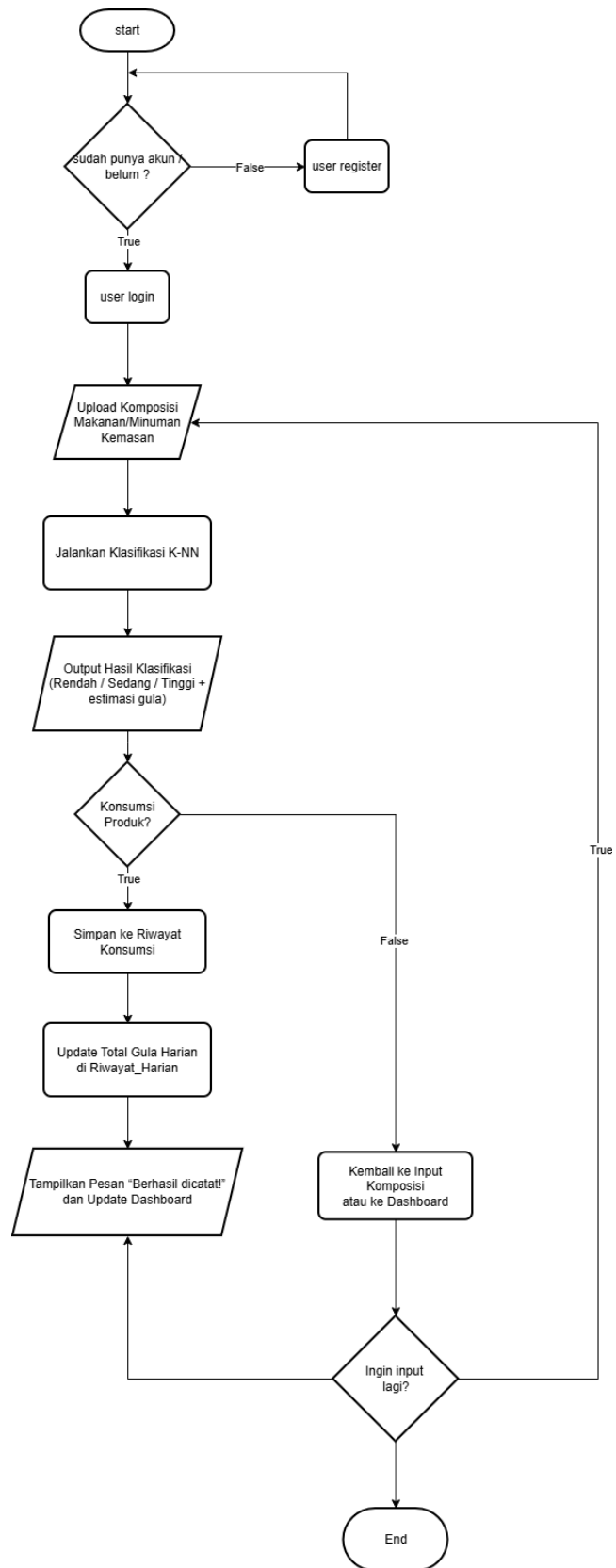
Gambar 3.1 Diagram Kerangka Pemikiran

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.

3.7 Perancangan Sistem

Perancangan sistem menggambarkan alur kerja dan struktur aplikasi klasifikasi kadar gula, meliputi pembuatan flowchart proses input hingga hasil KNN, perancangan ERD untuk struktur basis data, serta use case diagram yang menunjukkan interaksi pengguna dalam mendeteksi kadar gula dan mengelola riwayat konsumsi.

3.7.1 Flowchart Sistem

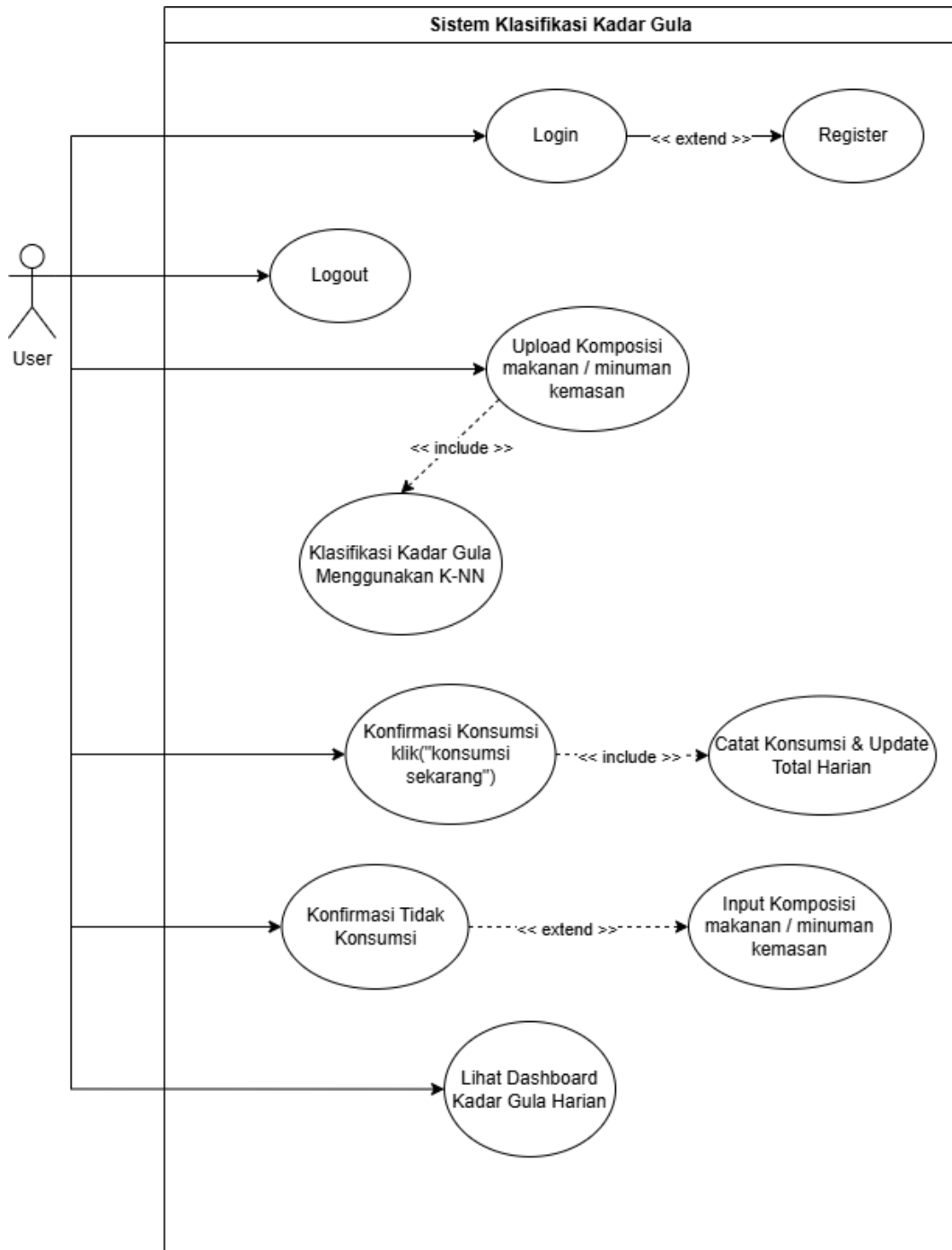


Gambar 3.2 Flowchart Sistem

Flowchart sistem ini menggambarkan alur kerja aplikasi klasifikasi kandungan gula yang berfungsi membantu pengguna memantau konsumsi gula harian. Proses dimulai ketika pengguna membuka aplikasi, kemudian sistem memeriksa kepemilikan akun untuk menentukan apakah pengguna perlu melakukan registrasi atau dapat langsung login. Setelah berhasil masuk, pengguna diarahkan untuk menginput data komposisi gizi produk, seperti kandungan gula, karbohidrat, dan komponen gizi lainnya. Data tersebut kemudian dianalisis oleh algoritma K-Nearest Neighbors (K-NN) dengan membandingkannya terhadap data pelatihan, sehingga menghasilkan klasifikasi kadar gula produk dalam kategori rendah, sedang, atau tinggi, disertai estimasi jumlah gula berdasarkan komposisi yang diberikan.

Setelah hasil klasifikasi ditampilkan, pengguna menentukan apakah produk tersebut dikonsumsi. Jika dikonsumsi, sistem menyimpan data ke entitas Riwayat Konsumsi dan memperbarui total gula harian melalui Riwayat Harian, kemudian menampilkan notifikasi bahwa pencatatan berhasil dilakukan serta memperbarui dashboard pengguna. Jika pengguna memilih tidak mengonsumsi produk, maka data tidak disimpan dan sistem mengarahkan kembali ke menu input atau dashboard. Pada tahap akhir, pengguna diberi opsi untuk kembali memasukkan produk lainnya atau mengakhiri penggunaan aplikasi. Secara keseluruhan, flowchart ini menunjukkan alur yang terstruktur antara proses input, analisis klasifikasi, pengambilan keputusan, dan pembaruan riwayat, sehingga pemantauan konsumsi gula dapat dilakukan secara sistematis dan akurat.

3.7.2 Diagram Use Case



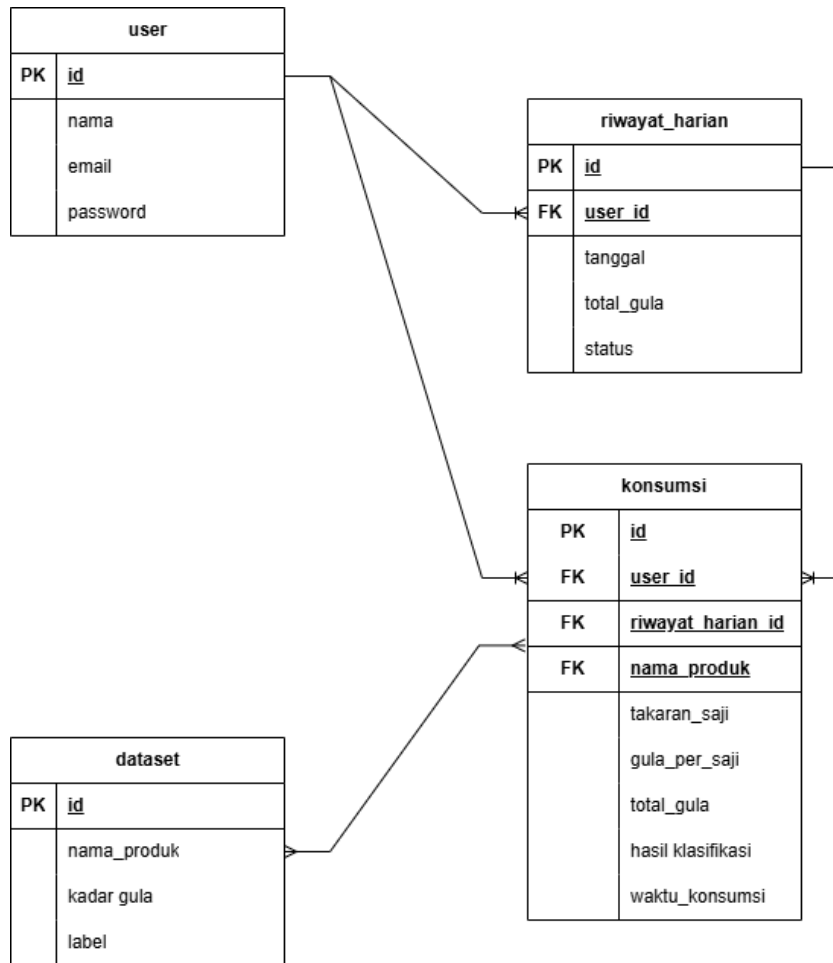
Gambar 3.3 Diagram Use Case

Diagram use case ini menggambarkan interaksi antara pengguna dan sistem klasifikasi kadar gula. Pengguna memulai dengan melakukan login, dan apabila belum memiliki akun, proses dapat diperluas dengan melakukan registrasi terlebih dahulu. Setelah masuk ke dalam sistem, pengguna dapat melakukan input

komposisi gizi produk makanan atau minuman kemasan. Aktivitas ini secara otomatis mencakup proses klasifikasi kadar gula menggunakan algoritma K-NN untuk menghasilkan kategori gula produk.

Setelah mendapatkan hasil klasifikasi, pengguna dapat memilih untuk mengonfirmasi bahwa produk tersebut dikonsumsi. Jika dikonfirmasi, sistem akan mencatat konsumsi tersebut dan memperbarui total gula harian pengguna. Sebaliknya, jika pengguna memilih tidak mengonsumsi produk, sistem memberikan opsi untuk kembali melakukan input komposisi produk lainnya. Selain itu, pengguna juga dapat melihat dashboard yang menampilkan total kadar gula harian serta riwayat konsumsi. Sistem menyediakan fitur logout apabila pengguna ingin keluar dari aplikasi. Secara keseluruhan, diagram ini menunjukkan alur fungsional utama dalam membantu pengguna memantau dan mengelola konsumsi gula hariannya.

3.7.3 Entity Relationship Diagram



Gambar 3.4 Entity Relationship Diagram (ERD)

Entity Relationship Diagram (ERD) ini menggambarkan struktur basis data yang digunakan dalam sistem klasifikasi kandungan gula. Terdapat empat entitas utama, yaitu user, riwayat_harian, konsumsi, dan dataset. Entitas user menyimpan informasi pengguna berupa nama, email, dan password, serta menjadi acuan bagi entitas lainnya melalui atribut id sebagai primary key. Setiap pengguna memiliki relasi ke entitas riwayat_harian, yang mencatat total gula yang dikonsumsi per hari, lengkap dengan tanggal dan status harian. Entitas dataset berisi informasi produk seperti nama produk, kadar gula, dan label hasil klasifikasi. Data pada entitas ini digunakan sebagai referensi ketika pengguna memasukkan komposisi produk. Relasi lebih lanjut terlihat pada entitas konsumsi, yang berfungsi mencatat detail

produk yang dikonsumsi pengguna, termasuk takaran saji, gula per saji, total gula, waktu konsumsi, serta hasil klasifikasi yang dihasilkan oleh sistem. Entitas konsumsi memiliki foreign key yang terhubung ke user, riwayat_harian, dan dataset, sehingga membentuk hubungan yang saling terintegrasi untuk memantau dan menghitung konsumsi gula harian secara otomatis. Struktur ERD ini memastikan bahwa seluruh proses pencatatan konsumsi dan perhitungan total gula harian dapat berjalan dengan tersusun dan konsisten.

3.8 Implementasi Prototype Web

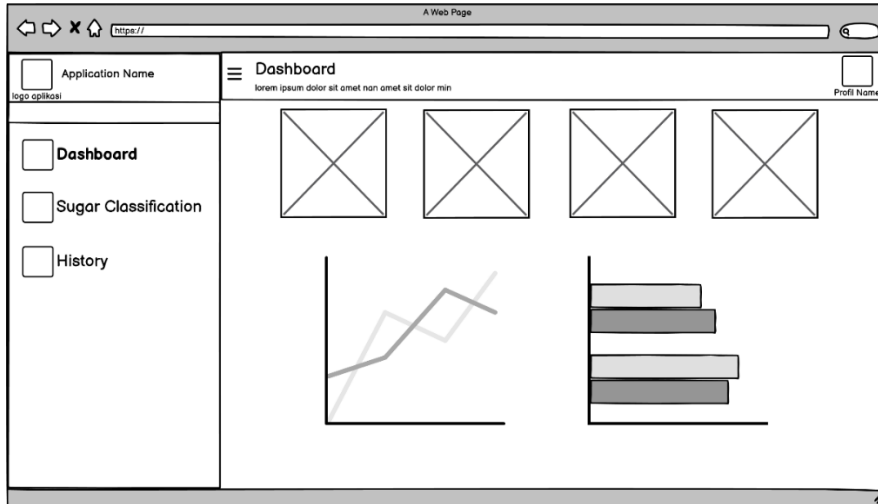
Prototipe web dibangun dalam bentuk wireframe untuk mewujudkan sistem klasifikasi kandungan gula berbasis komposisi pada kemasan secara end-to-end. Implementasi ini mencakup pengembangan antarmuka pengguna yang responsif, integrasi model machine learning, serta penyimpanan riwayat scan pengguna.

3.8.1 Desain Antarmuka

Desain antarmuka ini hanya disajikan sebagai gambaran rancangan awal melalui wireframe (mockup) menggunakan gaya sederhana. Tampilan prototipe ini bersifat konseptual dan fokus pada struktur serta alur penggunaan, bukan pada detail visual akhir. Terdapat tiga halaman utama:

1. Dashboard

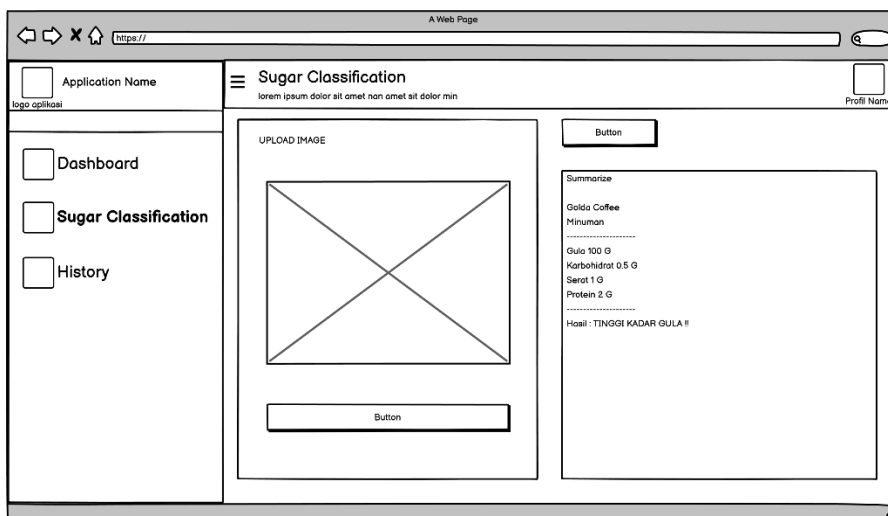
Sebagai halaman beranda yang menampilkan statistik ringkas pengguna (jumlah produk yang pernah discan, rata-rata tingkat gula harian/mingguan,



Gambar 3.5 Dashboard Page

2. Halaman Sugar Classification

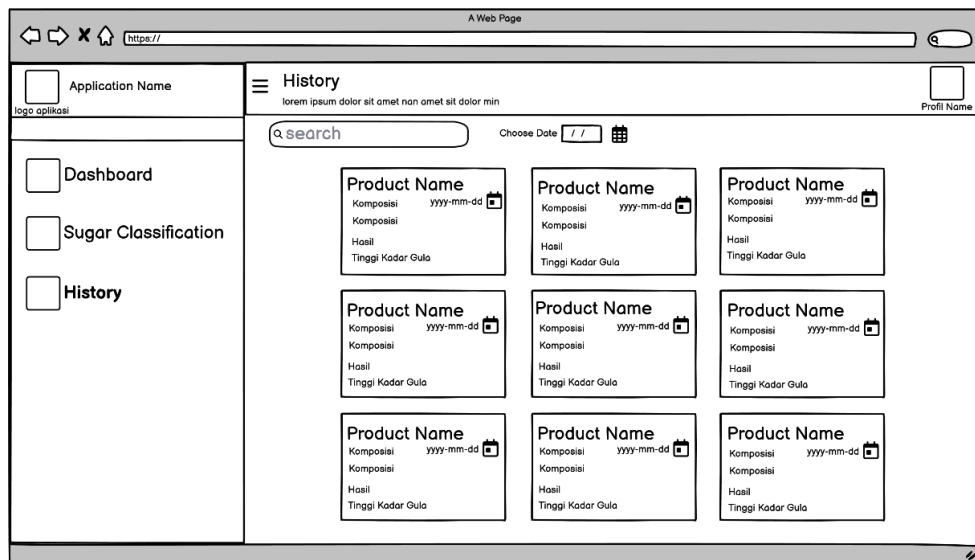
Halaman yang menampilkan hasil analisis secara real-time, mencakup foto produk terdeteksi, tabel nutrisi yang terekstrak, skor gula dengan indikator warna (hijau–merah), visualisasi radar chart nutrisi, serta rekomendasi konsumsi dalam bahasa sederhana.



Gambar 3.6 Sugar Classification Page

3. History

Halaman yang menyajikan daftar riwayat scan dalam bentuk kartu yang dapat diurutkan berdasarkan tanggal, disertai thumbnail produk, label tingkat gula, dan tombol untuk melihat detail kembali.



Gambar 3.7 History Page

BAB IV

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1 Kebutuhan Sistem

Kebutuhan sistem pada penelitian ini mencakup spesifikasi perangkat keras dan perangkat lunak yang digunakan dalam proses perancangan dan implementasi sistem. Spesifikasi ini menjadi acuan dalam memastikan sistem dapat berjalan dengan optimal sesuai dengan fungsinya.

4.1.1 Perangkat Keras

Perangkat keras yang digunakan dalam pengembangan dan pengujian sistem adalah sebagai berikut:

1. Perangkat : Asus VivoBook
2. Ram : 4 GB
3. Processor : Intel Core I3-1005G1(gen 10th)
4. Penyimpanan : SSD 500GB
5. Sistem Operasi : Windows 11 Home

4.1.2 Perangkat Lunak Sistem

Perangkat lunak yang digunakan dalam pengembangan sistem ini terdiri dari komponen pengembangan web, manajemen basis data, pemrosesan machine learning, serta alat bantu pengembangan. Seluruh komponen dirancang agar saling terintegrasi dalam mendukung proses deteksi objek, ekstraksi teks, dan klasifikasi kandungan gula produk.

1. Framework dan Web Server
 - a. Laravel 12

Sistem dibangun berbasis website menggunakan framework Laravel versi 12 yang menerapkan arsitektur Model-View-Controller (MVC). Laravel dipilih karena menyediakan fitur routing, middleware, serta ORM Eloquent yang mempermudah pengelolaan logika bisnis dan interaksi basis data.

b. Laragon

Lingkungan server lokal menggunakan Laragon versi 6.0 (Full Version) dengan konfigurasi sebagai berikut:

- 1) PHP versi 8.5.1
- 2) Node.JS versi 24.13.1
- 3) Nginx versi 1.22.0

2. Basis Data dan Manajemen Database

a. SQLite

Basis data yang digunakan dalam sistem adalah SQLite, yang bersifat lightweight dan tidak memerlukan server terpisah. SQLite dipilih karena sesuai untuk kebutuhan penelitian serta efisien dalam implementasi sistem skala prototipe.

b. TablePlus

Manajemen database dilakukan menggunakan TablePlus, yang berfungsi sebagai antarmuka grafis untuk melakukan manipulasi tabel, pengujian query, dan monitoring struktur database. Interaksi antara Laravel dan SQLite dilakukan melalui ORM Eloquent.

3. Machine Learning

a. Python

Model klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN) yang dibangun menggunakan bahasa pemrograman Python. Python dipilih karena memiliki dukungan pustaka machine learning yang lengkap dan stabil.

4. Deteksi Objek dan OCR

a. YOLO v8 (You Only Look Once)

Dikarenakan fokus penelitian ini bukan pada proses deteksi, peneliti memutuskan untuk menggunakan model YOLOv8 yang dikembangkan oleh Ultralytics sebagai pendekatan deteksi area tabel nutrisi. Model yang digunakan dalam penelitian ini diperoleh dari proyek Nutrition-Table-Detection yang tersedia pada platform *Roboflow Universe* dengan nama proyek *Nutrition-Table-Detection* oleh pengguna *allagui* (allagui, 2025). Model tersebut telah dilatih sebelumnya (pre-trained) untuk mendeteksi area tabel nutrisi pada gambar produk. Pada penelitian ini, model dijalankan secara lokal tanpa proses pelatihan ulang penuh (retraining), dan digunakan untuk menghasilkan *bounding box* pada area tabel nutrisi. Hasil deteksi tersebut kemudian diproses lebih lanjut pada tahap ekstraksi teks (OCR).

b. Optical Character Recognition (OCR)

Tahap ekstraksi teks dilakukan menggunakan pustaka **EasyOCR**, sebuah framework OCR berbasis *deep learning* yang dibangun di atas modul **PyTorch**. Pustaka ini menggunakan arsitektur *Convolutional Neural Network* (CNN) untuk pengenalan karakter dan *Long Short-Term Memory* (LSTM) untuk pemrosesan urutan teks.

Pendekatan ini dipilih karena EasyOCR mendukung deteksi teks dalam berbagai bahasa dan mampu menangani gambar dengan kualitas menengah melalui prapemrosesan internal. Selain itu, penggunaan EasyOCR memungkinkan pemrosesan data dilakukan secara lokal (on-premise) tanpa ketergantungan pada koneksi API eksternal, sehingga lebih efisien dalam hal biaya dan privasi data.

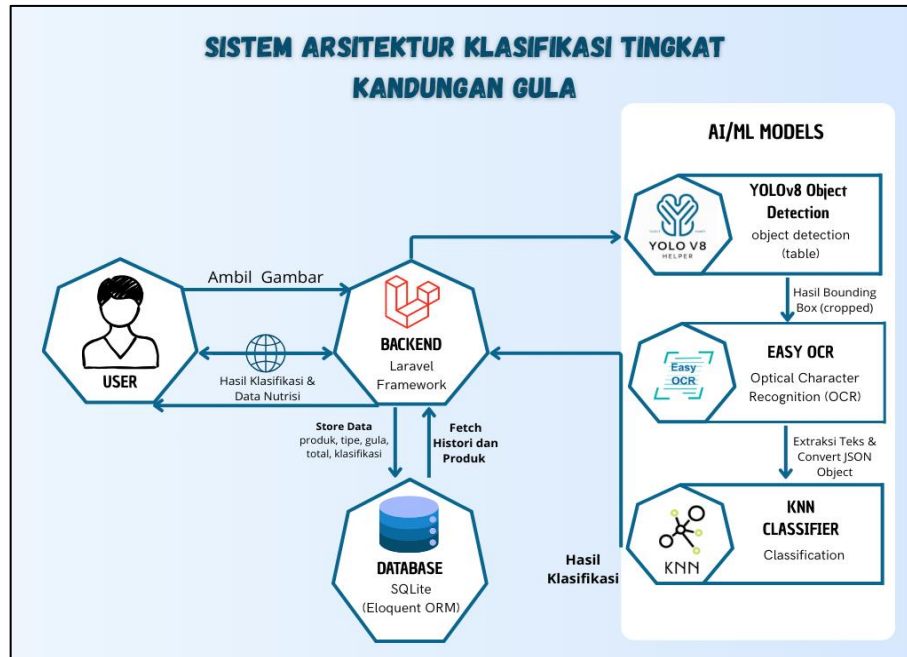
Lingkungan Pengembangan

Seluruh proses pengembangan dilakukan menggunakan Visual Studio Code sebagai Integrated Development Environment (IDE). Visual Studio Code mendukung berbagai ekstensi untuk Laravel, Python, serta manajemen API sehingga mempermudah proses integrasi dan debugging sistem.

4.2 Implementasi Sistem

4.2.1 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 4.1 Diagram Arsitektur Sistem

Berdasarkan diagram arsitektur yang ditunjukkan pada Gambar 4.1, sistem terdiri dari empat komponen utama, yaitu User, Backend (Laravel), AI/ML Models, dan Database dengan detail sebagai berikut :

1. Komponen User (Client)

User berinteraksi dengan sistem melalui antarmuka berbasis web menggunakan browser. Pada tahap ini, pengguna melakukan proses pengambilan atau pengunggahan gambar produk yang mengandung informasi tabel nutrisi.

a. Input utama dari user

- 1) Gambar produk
- 2) Konfirmasi konsumsi

b. Output yang diterima user

- 1) Hasil klasifikasi
- 2) Informasi Nilai Nutrsi hasil ekstraksi
- 3) Riayat produk yang dikonsumsi dan history

2. Backend

Backend sistem di bangun diatas framework Laravel yang berperan sebagai pusat kontrol dan store data menggunakan controller dan model untuk berinteraksi dengan database sql menggunakan *Eloquent ORM*. Dimana Laravel menangani :

- a. Penerimaan request dari user berupa gambar.
- b. Penyimpanan sementara file gambar pada server
- c. Pengiriman gambar ke server untuk melalui proses crop area oleh model deteksi objek YOLOv8
- d. Pengelolaan komunikasi untuk memberikan request sesuai api key ke layanan OCR Gemini
- e. Pengiriman data nutrisi hasil ekstraksi berupa JSON object format, ke model klasifikasi KNN.
- f. Penyimpanan produk ke dalam database(opsional apabila di konsumsi)
- g. Pengambilan dan penyajian data produk dikonsumsi beserta histori ke pada user

3. AI/ML Models

Komponen AI/ML di bagi menjadi 3 bagian sebagai berikut :

- a. YOLOv8 Object Detection
- b. Gemini API OCR
- c. KNN Classifier

4. Database SQLite

Database SQLite digunakan untuk menyimpan informasi produk berupa nama produk, tipe produk, menyimpan nilai nutrisi hasil ekstraksi ocr, hasil

klasifikasi dan total gula konsumsi pengguna. Interaksi antara Laravel dan database dilakukan menggunakan Eloquent ORM yang mempermudah manipulasi data tanpa query SQL manual.

Tabel 4.1 Ringkasan Komponen Sistem

No	Komponen	Teknologi	Fungsi Utama
1.	User	Web Browser	Mengunggah gambar dan menerima hasil klasifikasi
2.	Backend	Laravel	Mengatur alur proses dan integrasi sistem
3.	Deteksi Objek	YOLO	Mendeteksi area tabel nutrisi pada gambar
4.	OCR	Gemini API	Mengekstraksi teks dari tabel nutrisi
5.	Klasifikasi	K-Nearest Neighbor (KNN)	Mengklasifikasikan tingkat kandungan gula
6.	Database	SQLite	Menyimpan data produk dan hasil klasifikasi

4.2.2 Implementasi Modul Deteksi Objek (YOLOv8)

Modul deteksi objek pada penelitian ini bertujuan untuk mengidentifikasi area tabel informasi nilai gizi pada kemasan produk sebelum dilakukan proses ekstraksi teks. Proses deteksi ini diperlukan untuk membatasi area pemrosesan hanya pada bagian yang relevan sehingga dapat meningkatkan akurasi tahap OCR.

4.2.2.1 Metode Pelatihan

Pendekatan yang digunakan adalah transfer learning, yaitu memanfaatkan model YOLOv8 pretrained, kemudian dilakukan pelatihan ulang (fine-tuning).

Tahapan pelatihan yang dilakukan adalah sebagai berikut:

- a. Mengunduh konfigurasi model YOLOv8 dalam format .yaml dari Roboflow.
- b. Menyesuaikan parameter jumlah kelas sesuai kebutuhan deteksi (kelas: nutrition table).
- c. Melakukan proses pelatihan ulang menggunakan dataset yang telah dianotasi.
- d. Proses training dilakukan selama lebih dari 50 epoch hingga model mencapai konvergensi.
- e. Model dengan performa terbaik secara otomatis disimpan dalam file *best.pt*.

4.2.3 Implementasi Modul OCR

4.2.3.1 Teknologi yang Digunakan

Pada penelitian ini, modul Optical Character Recognition (OCR) dibangun menggunakan library EasyOCR berbasis deep learning. EasyOCR dipilih karena mampu mengenali teks pada citra dengan tingkat akurasi yang baik serta mendukung berbagai variasi bentuk huruf, ukuran teks, dan kondisi pencahayaan. EasyOCR bekerja dengan pendekatan convolutional neural network (CNN) dan recurrent neural network (RNN) untuk mendeteksi serta mengenali karakter dari citra.

4.2.3.2 Mekanisme Ekstraksi Informasi

Proses ekstraksi informasi nutrisi menggunakan OCR dilakukan melalui beberapa tahapan sebagai berikut:

1. Preprocessing Citra

Sebelum dilakukan pembacaan teks, citra tabel nutrisi diproses untuk meningkatkan kualitas visual dan keterbacaan karakter. Tahapan preprocessing meliputi:

- a. Upscaling citra sebesar $2\times$ menggunakan interpolasi Lanczos untuk mempertajam detail teks.
- b. Konversi citra ke grayscale untuk menyederhanakan informasi warna.
- c. Denoising menggunakan metode Non-local Means Denoising untuk mengurangi noise.
- d. Penerapan sharpening filter untuk mempertegas karakter yang buram.
- e. Fallback ke adaptive threshold apabila proses OCR awal tidak menghasilkan teks.

Tahapan ini bertujuan untuk meningkatkan akurasi pengenalan teks, terutama pada gambar dengan pencahayaan kurang optimal atau teks yang kecil.

2. Pembacaan Teks Menggunakan EasyOCR

Setelah preprocessing, citra diproses oleh EasyOCR untuk mendeteksi dan mengenali teks. Output yang dihasilkan berupa:

- a. Bounding box lokasi teks
- b. Teks hasil pembacaan
- c. Nilai probabilitas (confidence score)

Teks dengan probabilitas rendah difilter untuk mengurangi kesalahan pembacaan. Selanjutnya, hasil teks diurutkan berdasarkan posisi vertikal dan horizontal agar membentuk struktur baris seperti pada tabel nutrisi asli. Hasil akhir tahap ini berupa teks mentah (raw text) dari label nutrisi.

3. Parsing Informasi Nutrisi

Teks mentah yang dihasilkan kemudian diproses menggunakan pendekatan berbasis keyword mapping dan regular expression untuk mengekstraksi nilai numerik nutrisi.

Sistem mencari kata kunci seperti:

- a. Calories / Energi
- b. Total fat / Lemak total
- c. Total carbohydrate / Karbohidrat total
- d. Sugars / Gula
- e. Protein
- f. Sodium / Garam

Untuk mengatasi kesalahan pembacaan OCR (misalnya typo atau karakter yang tidak terbaca sempurna), sistem menggunakan pendekatan pencarian fleksibel terhadap variasi kata kunci. Nilai numerik diekstraksi menggunakan pola regular expression yang mendeteksi angka desimal maupun bilangan bulat.

4. Validasi Data Hasil OCR

Data yang berhasil diekstraksi tidak langsung digunakan, tetapi melalui proses validasi untuk memastikan nilai yang diperoleh masuk akal secara nutrisi.

Validasi dilakukan dengan pendekatan threshold logis, misalnya:

- a. Kalori berada dalam rentang yang wajar untuk produk kemasan.
- b. Nilai gula tidak melebihi total karbohidrat.
- c. Nilai protein dan lemak berada dalam rentang realistis.

Jika nilai yang terbaca berada di luar rentang logis, maka dianggap sebagai noise hasil kesalahan OCR dan diabaikan. Pendekatan ini membantu meningkatkan keandalan sistem dalam menghadapi kesalahan pembacaan akibat kualitas gambar yang kurang baik atau tata letak label yang tidak standar.

4.2.4 Implementasi Model Klasifikasi KNN

4.2.4.1 Sampling dan Persiapan Dataset

Tahap ini menjelaskan proses persiapan dataset yang digunakan dalam penelitian. Dataset bersumber dari *Open Food Facts*, yaitu basis data terbuka yang menyediakan informasi komposisi nutrisi berbagai produk makanan dan minuman dari berbagai negara. Dataset mentah yang digunakan berjumlah lebih dari 2 juta baris sehingga diperlukan proses pembersihan dan seleksi untuk menghasilkan data yang valid dan layak digunakan dalam pelatihan model klasifikasi kadar gula.

Tahapan ini mencakup proses pembersihan data mentah(Data Cleaning), pelabelan(Labelling), rekayasa fitur(Feature Engineering), serta balanced sampling hingga diperoleh dataset final yang siap digunakan dalam pelatihan model.

1. Data Cleaning

Pada tahap pertama dilakukan proses pembersihan dataset mentah menggunakan metode chunking untuk mengatasi keterbatasan memori saat memproses data berukuran besar. Dataset dibaca secara bertahap dengan ukuran 150.000 baris per chunk.

A. Seleksi Kolom

Hanya atribut nutrisi utama yang digunakan, yaitu:

- 1) energy-kcal_100g
- 2) fat_100g
- 3) saturated-fat_100g
- 4) carbohydrates_100g
- 5) proteins_100g
- 6) fiber_100g

- 7) salt_100g
- 8) sugars_100g

B. Pembersihan Data (Validasi dan Filtering)

Beberapa tahapan validasi dilakukan untuk memastikan kualitas data:

1) Konversi Numerik dan Penghapusan Missing Value

Seluruh kolom dikonversi ke format numerik dan baris yang mengandung nilai NaN dihapus.

2) Validasi Rentang Nutrisi (0–100 gram)

Nilai lemak, lemak jenuh, karbohidrat, protein, serat, garam, dan gula difilter agar berada dalam rentang logis 0–100 gram per 100 gram produk.

3) Validasi Rentang Energi (0–900 kkal)

Nilai energi dibatasi pada rentang 0–900 kkal per 100 gram untuk menghindari anomali data ekstrem.

4) Relasi Logis Gula dan Karbohidrat

$\text{sugars_100g} \leq \text{carbohydrates_100g}$, Tidak diperbolehkan karbohidrat = 0 tetapi gula > 0

5) Relasi Logis Lemak Jenuh

$\text{saturated-fat_100g} \leq \text{fat_100g}$

6) Penghapusan Baris Kosong Total Nutrisi

Baris dengan total nutrisi = 0 dihapus.

2. Labelling Dataset

Setelah dataset bersih diperoleh, dilakukan proses pelabelan berdasarkan kandungan gula per 100 gram (sugars_100g).

A. Skema Labeling

Pelabelan dilakukan berdasarkan kategori kadar gula per 100 gram sebagai berikut:

- 1) **Low**
- 2) **Medium**
- 3) **High**

3. Feature Engineering

Untuk memperkaya representasi karakteristik dominasi gula dalam produk, ditambahkan beberapa fitur turunan. Untuk menghindari pembagian dengan nol, digunakan nilai epsilon kecil ($1e-6$) pada penyebut. Untuk memperkaya representasi data dan membantu model mengenali pola kadar gula secara lebih komprehensif, dilakukan penambahan tiga fitur baru:

Statistik Deskriptif:					
	energy-kcal_100g	carbohydrates_100g	sugar_per_carbs_ratio	sugar_per_kcal_ratio	sugar_dominance_score
count	5001.000000	5001.000000	5001.000000	5.001000e+03	5.001000e+03
mean	283.867476	34.097321	0.500134	8.953980e+04	3.581622e+04
std	178.874536	26.590132	0.349273	2.793461e+06	1.117385e+06
min	0.000000	0.000000	0.000000	0.000000e+00	0.000000e+00
25%	114.000000	10.200000	0.153846	4.904632e-02	1.235078e-01
50%	286.000000	28.200000	0.500000	2.000000e-01	4.033062e-01
75%	417.000000	56.862745	0.833333	4.333333e-01	6.622428e-01
max	900.000000	100.000000	1.000000	1.176471e+08	4.705882e+07

Gambar 4.2 Feature Engineering

A. sugar_per_carbs_ratio

Mengukur proporsi gula terhadap total karbohidrat:

$$sugar_per_carbs_ratio = \frac{sugars_100g}{carbohydrates_100g + e}$$

Fitur ini menunjukkan dominasi gula dalam komposisi karbohidrat suatu produk.

B. `sugar_per_kcal_ratio`

$$\text{sugar_per_kcal_ratio} = \frac{\text{sugars_100g} \times 4}{\text{energy-kcal_100g} + e}$$

Karena 1 gram gula setara ± 4 kkal, fitur ini merepresentasikan densitas energi yang berasal dari gula.

C. `net_carbs`

Menghitung jumlah karbohidrat yang dapat dicerna tubuh

$$\text{net_carbs} = \text{carbohydrates_100g} - \text{fiber_100g}$$

Fitur ini relevan karena serat tidak berkontribusi langsung terhadap peningkatan kadar gula darah.

Untuk menghindari kesalahan pembagian dengan nol, ditambahkan nilai epsilon kecil ($1e-6$) pada penyebut saat perhitungan rasio.

D. Sugar Dominance Score

Fitur tambahan yang menjadi pembeda utama penelitian ini adalah skor dominasi gula, Fitur ini dirancang untuk merepresentasikan dominasi gula secara komprehensif, baik dari sisi komposisi karbohidrat maupun kontribusi energi. Skor dominasi gula yang dihitung sebagai kombinasi berbobot:

4. Pencegahan Data Leakage

Setelah proses feature engineering selesai, kolom `sugars_100g` dihapus dari dataset sebelum pelatihan model. Langkah ini dilakukan untuk mencegah data leakage, karena variabel tersebut merupakan dasar pembentukan label. Jika tetap digunakan, model akan memperoleh informasi jawaban secara langsung sehingga menghasilkan evaluasi yang tidak valid.

5. Balanced Sampling

Setelah labeling dan penambahan fitur, dilakukan proses balancing untuk memastikan distribusi kelas merata. Setiap kategori (Low, Medium, High) diambil sebanyak maksimal 1.667 data sehingga total dataset menjadi 5.001 baris. Dataset kemudian diacak (shuffle) menggunakan random state tetap untuk menjaga reproduisibilitas hasil.

```
Menghapus kolom sugars_100g (mencegah data leakage)...
Melakukan Balanced Sampling...
Distribusi label setelah balancing:
label
Low      1667
Medium   1667
High     1667
Name: count, dtype: int64
```

Gambar 4.3 Balanced Sampling

Balancing dilakukan karena algoritma KNN bersifat sensitif terhadap distribusi data. Jika salah satu kelas mendominasi, maka proses voting pada tetangga terdekat akan cenderung menghasilkan bias terhadap kelas mayoritas.

6. Karakteristik Dataset Akhir

Dataset akhir yang digunakan dalam penelitian ini merupakan hasil proses pembersihan data (*data cleaning*), pelabelan, rekayasa fitur (*feature engineering*), serta penyeimbangan kelas (*balanced sampling*).

```
• Total data : 5,001 baris
• Total fitur : 6 kolom (5 fitur + 1 label)

✦ Distribusi Kelas:
- Low : 33.33%
- Medium : 33.33%
- High : 33.33%

🔵 Validasi Tipe Data Fitur:
✅ Seluruh fitur dalam bentuk numerik
```

Gambar 4.4 Deskripsi Dataset

Dataset tersebut memiliki karakteristik sebagai berikut:

- a. Total Data: 5.001 baris.
- b. Total atribut: 6 kolom (5 fitur independen dan 1 variabel target).
- c. Distribusi kelas: relatif seimbang, dengan proporsi $\pm 33,3\%$ untuk masing-masing kategori.
- d. Tipe data fitur: seluruh fitur berbentuk numerik sehingga dapat langsung diproses oleh algoritma klasifikasi berbasis jarak.

Keseimbangan distribusi kelas dilakukan untuk menghindari bias model terhadap kelas tertentu serta meningkatkan stabilitas performa algoritma klasifikasi. Fitur yang digunakan terdiri atas:

```
◆ Info Lengkap Dataset:
<class 'pandas.DataFrame'>
RangeIndex: 5001 entries, 0 to 5000
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   energy-kcal_100g      5001 non-null   float64
1   carbohydrates_100g    5001 non-null   float64
2   sugar_per_carbs_ratio 5001 non-null   float64
3   sugar_per_kcal_ratio  5001 non-null   float64
4   sugar_dominance_score 5001 non-null   float64
5   label                 5001 non-null   str
dtypes: float64(5), str(1)
```

Gambar 4.5 Fitur Dataset

A. Fitur Nutrisi Dasar (Raw Features)

Fitur nutrisi dasar merepresentasikan kandungan zat gizi per 100 gram produk pangan. Seluruh nilai dinyatakan dalam satuan gram (g) atau kilokalori (kcal) per 100 gram. Fitur yang digunakan meliputi:

- a. energy-kcal_100g
- b. carbohydrates_100g

Fitur-fitur ini menggambarkan komposisi makronutrien utama yang secara langsung maupun tidak langsung berkaitan dengan kadar gula suatu produk.

B. Fitur Rekayasa (Engineered Features)

Selain fitur nutrisi dasar, penelitian ini menambahkan beberapa fitur turunan hasil rekayasa untuk memperkaya representasi karakteristik gula dalam produk. Fitur rekayasa yang digunakan adalah:

- a. *sugar_per_carbs_ratio*: Menggambarkan proporsi gula terhadap total karbohidrat.
- b. *sugar_per_kcal_ratio*: Merepresentasikan kontribusi energi dari gula terhadap total energi produk
- c. *sugar_dominance_score*: Menggambarkan dominasi gula terhadap suatu produk.

C. Fitur Target (Label)

Variabel target dalam penelitian ini adalah kategori kadar gula produk yang dibagi menjadi tiga kelas, yaitu:

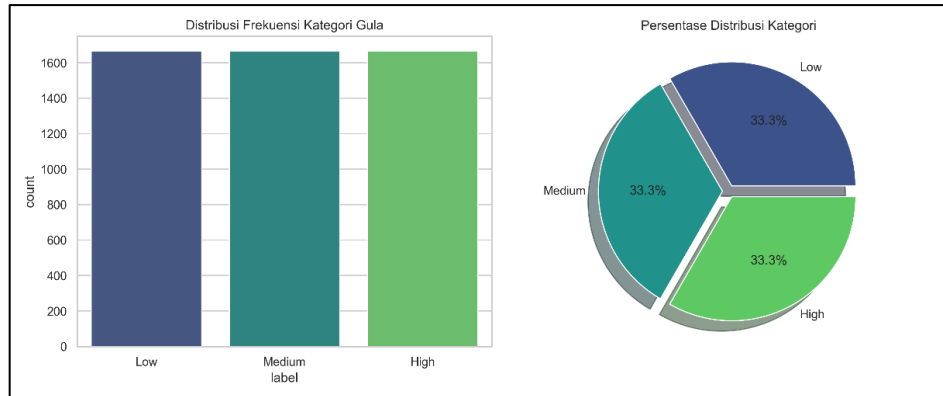
- a. Low
- b. Medium
- c. High

7. Exploratory Dataset Analysis(EDA)

Analisis distribusi kelas dilakukan menggunakan grafik count plot dan pie chart. Hasil visualisasi menunjukkan bahwa ketiga kelas memiliki jumlah data yang relatif seimbang, masing-masing sekitar $\pm 33,3\%$.

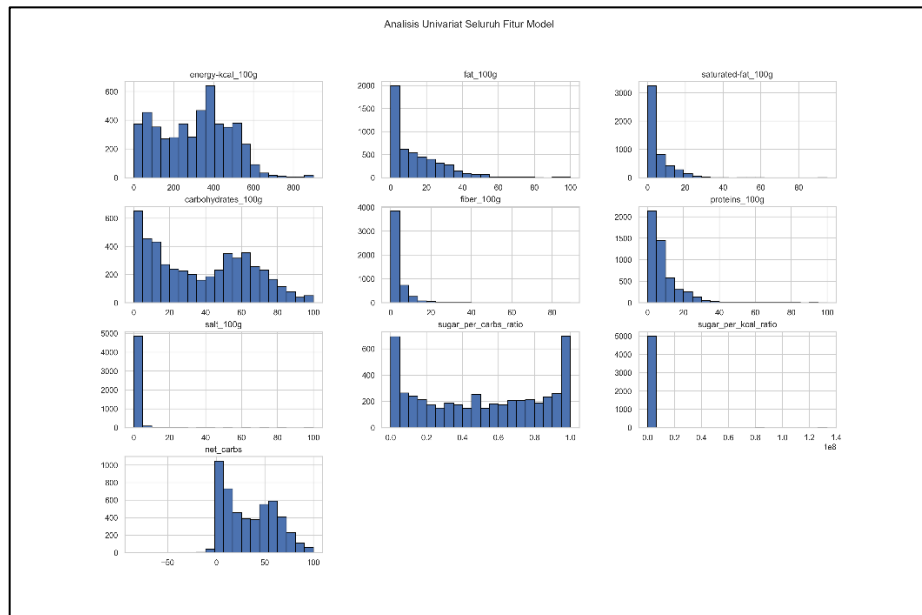
Keseimbangan distribusi ini penting karena algoritma KNN bersifat sensitif terhadap dominasi kelas tertentu. Dengan distribusi yang merata, proses voting pada tetangga terdekat tidak akan bias terhadap satu kategori tertentu.

A. Distribusi Kelas



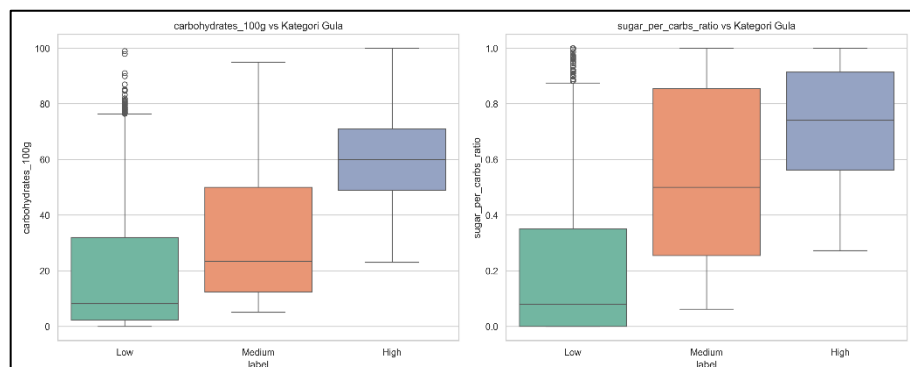
Gambar 4.6 Distribusi Kelas

B. Analisis Distribusi Fitur



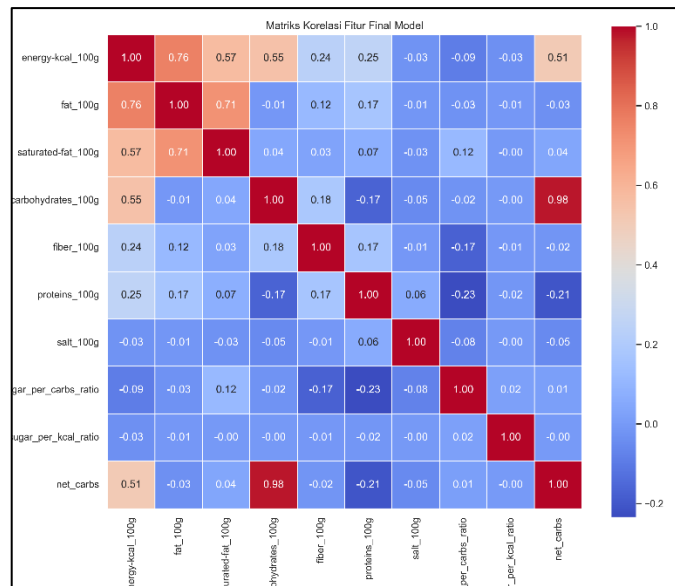
Gambar 4.7 Distribusi Fitur

C. Hubungan Fitur dengan Kategori Gula



Gambar 4.8 Hubungan Fitur Terhadap Gula

D. Korelasi Antarfitur



Gambar 4.9 Korelasi Antarfitur

4.2.4.2 Processing Data

Tahap *processing data* dilakukan untuk mempersiapkan dataset sebelum digunakan dalam proses pelatihan model klasifikasi. Proses ini meliputi pemisahan variabel fitur dan label, transformasi label ke bentuk numerik, serta pembagian data menjadi data latih dan data uji. Dataset yang digunakan terdiri dari sejumlah fitur numerik dan satu variabel target (label) yang merepresentasikan kategori kadar gula. Seluruh fitur telah berada dalam bentuk numerik sehingga tidak diperlukan proses transformasi tambahan pada variabel independen.

1. Validasi Dataset dan Fitur

Setelah dataset dimuat, dilakukan pemeriksaan terhadap, Dimensi dataset (jumlah baris dan kolom), Nama kolom, Keberadaan missing value. Selain itu, dilakukan validasi untuk memastikan bahwa fitur yang digunakan dalam penelitian sesuai dengan rancangan pada tahap feature engineering. Fitur yang divalidasi meliputi:

- a. energy-kcal_100g
- b. carbohydrates_100g
- c. sugar_per_carbs_ratio
- d. sugar_per_kcal_ratio
- e. sugar_dominance_score

Proses validasi ini bertujuan untuk memastikan konsistensi antara desain metodologi penelitian dan implementasi teknis pada dataset yang digunakan.

2. Pemisahan Fitur dan Label

Dataset dipisahkan menjadi dua bagian utama yaitu :

- a. Variabel independen (X), yaitu seluruh kolom fitur
- b. Variabel dependen (y), yaitu kolom label

Pemisahan ini bertujuan untuk membedakan atribut yang digunakan sebagai dasar prediksi dan kelas yang akan diprediksi oleh model.

3. Transformasi Label (Label Encoding)

```
2 LABEL_ENCODING
-----
Mapping label → numerik:
High → 0
Low → 1
Medium → 2

Distribusi label setelah encoding:
1    1667
2    1667
0    1667
Name: count, dtype: int64
```

Gambar 4.10 Label Encoder

Karena algoritma K-Nearest Neighbor (KNN) memerlukan representasi numerik pada variabel target, maka dilakukan proses Label Encoding. Proses ini mengubah label kategori menjadi bilangan numerik secara otomatis berdasarkan urutan alfabet kelas.

Transformasi ini hanya diterapkan pada variabel target dan tidak memengaruhi karakteristik fitur karena seluruh variabel independen sudah berbentuk numerik. Proses encoding dilakukan menggunakan metode Label Encoder dari pustaka scikit-learn.

4. Pembagian Data (Stratified Train-Test Split)

```
-----  
3 STRATIFIED TRAIN-TEST SPLIT (80:20)  
-----  
Jumlah data training : 4000  
Jumlah data testing  : 1001  
  
Distribusi label pada TRAINING:  
0    1334  
1    1333  
2    1333  
Name: count, dtype: int64  
  
Distribusi label pada TESTING:  
2     334  
1     334  
0     333  
Name: count, dtype: int64
```

Gambar 4.11 Train-Test Split Data

Dataset kemudian dibagi menjadi dua bagian, yaitu:

- a. 80% data training
- b. 20% data testing

Pembagian dilakukan menggunakan metode Stratified Train-Test Split. Teknik stratifikasi digunakan untuk memastikan bahwa proporsi masing-masing kelas pada data training dan data testing tetap seimbang sesuai distribusi awal dataset.

Pembagian dilakukan menggunakan fungsi `train_test_split` dari pustaka scikit-learn, dengan parameter:

- a. `test_size = 0.2`

- b. `random_state = 42`
- c. `stratify = y_encoded`

Hal ini penting untuk menjaga representativitas distribusi kelas dan menghindari bias pada proses evaluasi model. Parameter `random_state=42` digunakan untuk memastikan bahwa proses pembagian data bersifat konsisten dan dapat direproduksi pada eksperimen selanjutnya. Hasil pembagian data kemudian disimpan dalam bentuk berkas (`split_data.pkl`) untuk menjaga konsistensi dan reproduktibilitas penelitian. Selain itu, objek *Label Encoder* juga disimpan (`label_encoder.pkl`) untuk memastikan konsistensi transformasi label pada tahap implementasi dan evaluasi.

4.2.4.3 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.

1. Normalisasi Data dan Pipeline

Karena KNN menghitung jarak antar data, perbedaan skala antar fitur dapat menyebabkan fitur dengan rentang nilai besar mendominasi perhitungan jarak. Oleh karena itu dilakukan normalisasi menggunakan metode Min-Max Scaling. Metode ini mentransformasikan nilai fitur ke dalam rentang:

$$0 \leq x \leq 1$$

Normalisasi dilakukan menggunakan kelas `MinMaxScaler` dari pustaka *scikit-learn*. Untuk mencegah data leakage, proses normalisasi dan pelatihan model diintegrasikan dalam sebuah Pipeline, yang terdiri dari:

- a. MinMaxScaler
- b. KNeighborsClassifier

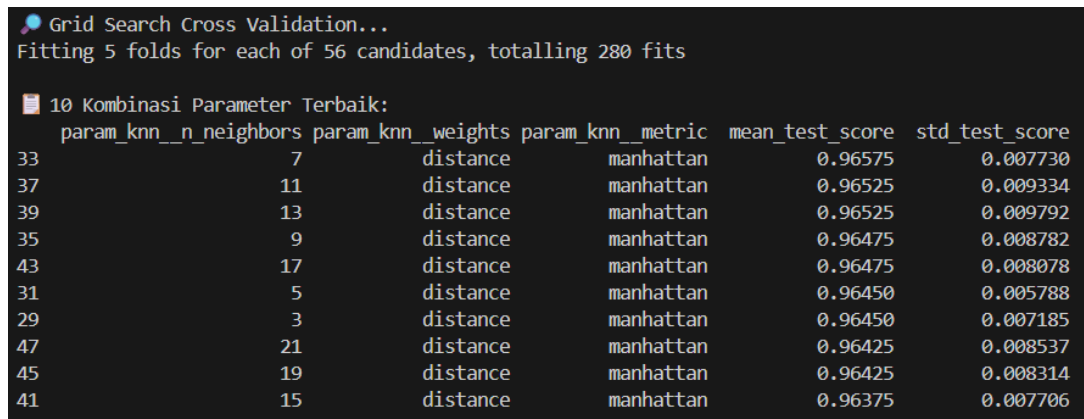
Penggunaan Pipeline memastikan bahwa proses scaling hanya dipelajari dari data training pada setiap lipatan validasi silang (*cross validation*), sehingga tidak terjadi kebocoran informasi dari data testing.

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 4.12 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.

2. Optimasi Hyperparameter



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

10 Kombinasi Parameter Terbaik:

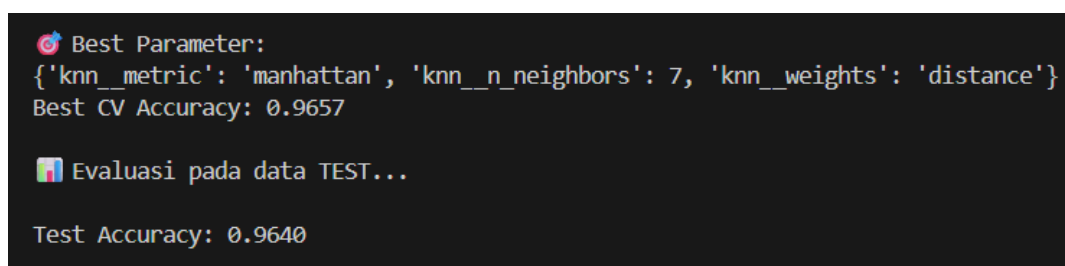
	param_knn_n_neighbors	param_knn_weights	param_knn_metric	mean_test_score	std_test_score
33	7	distance	manhattan	0.96575	0.007730
37	11	distance	manhattan	0.96525	0.009334
39	13	distance	manhattan	0.96525	0.009792
35	9	distance	manhattan	0.96475	0.008782
43	17	distance	manhattan	0.96475	0.008078
31	5	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 4.13 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 4.14 Hasil Optimasi

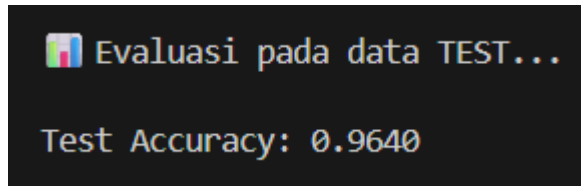
Pada penelitian ini, dua metode perhitungan jarak digunakan, yaitu *Euclidean Distance* dan *Manhattan Distance*. *Euclidean Distance* menghitung jarak garis lurus antar dua titik dalam ruang multidimensi, sedangkan *Manhattan Distance* menghitung jarak berdasarkan jumlah selisih absolut tiap dimensi. 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.

3. Evaluasi dan Visualisasi Hasil Optimasi

A. Accuracy

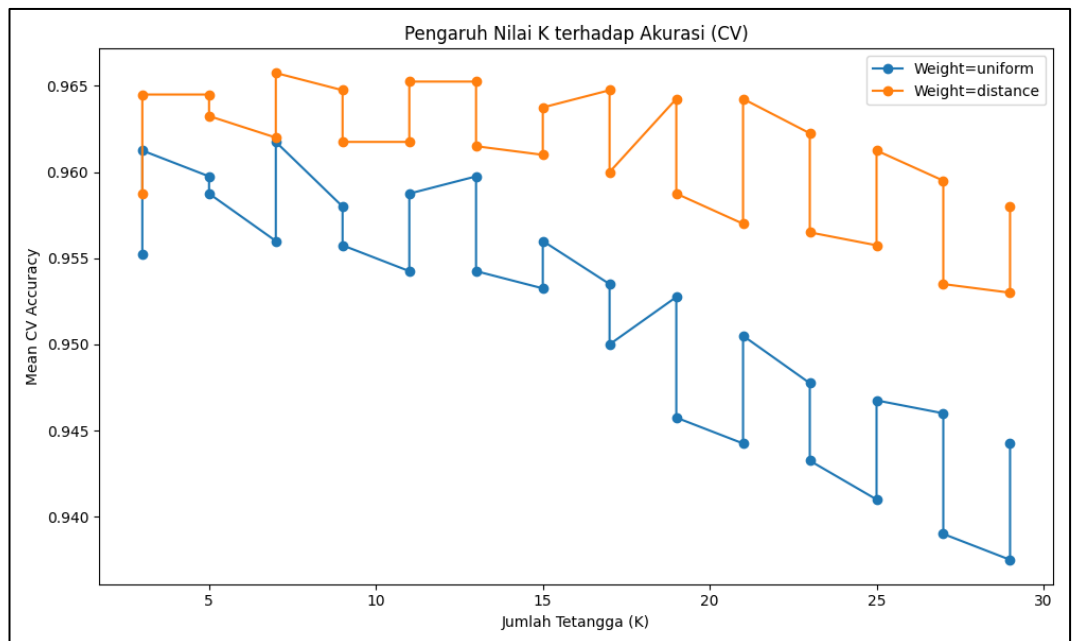
Evaluasi model dilakukan menggunakan metrik Accuracy untuk mengukur sejauh mana model mampu mengklasifikasikan data dengan benar pada dataset yang belum pernah dilihat sebelumnya (*Test Set*). Berdasarkan hasil pengujian yang dilakukan, diperoleh nilai sebagai berikut:



Gambar 4.15 Accuracy Testing

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.



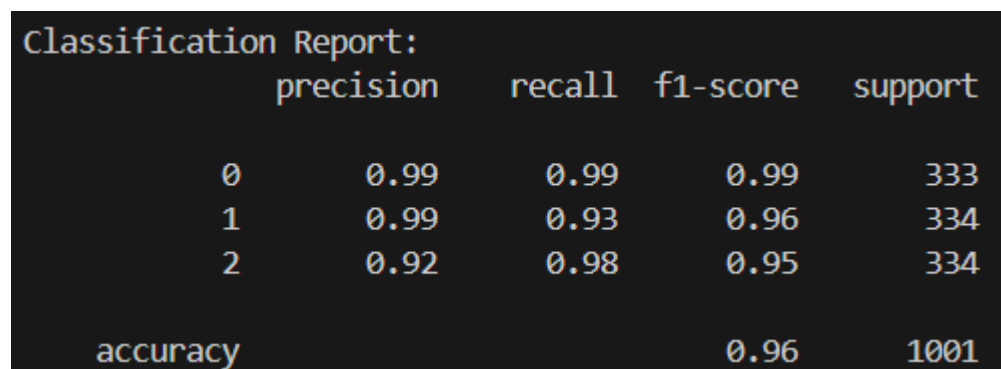
Gambar 4.16 Pengaruh Nilai K Terhadap Akurasi

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.

B. Classification Report

Berdasarkan hasil evaluasi, model K-Nearest Neighbor (KNN) menunjukkan performa yang sangat baik dengan tingkat akurasi keseluruhan mencapai 96%, yang menegaskan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan yang tinggi.



```
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
```

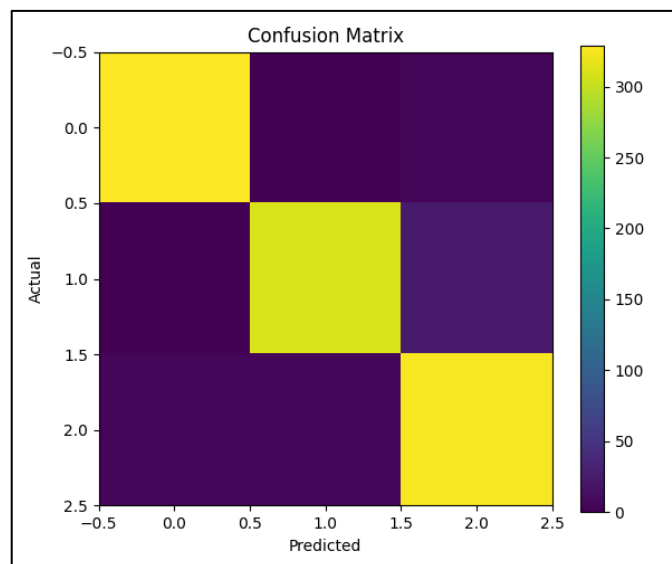
Gambar 4.17 Evaluasi Model

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.

C. 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 4.18 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.

4.2.4.4 Perhitungan Manual K-Nearest Neighbor

Pada bagian ini dilakukan simulasi perhitungan algoritma K-Nearest Neighbor (KNN) secara manual untuk menunjukkan mekanisme kerja model dalam melakukan klasifikasi tingkat kandungan gula produk. Algoritma KNN bekerja dengan menghitung jarak antara data uji dengan seluruh data latih, kemudian memilih K data dengan jarak terdekat untuk menentukan kelas berdasarkan mayoritas tetangga terdekat.

1. Data Latih

Tabel 4.2 Data Latih

Data	Energy	Fat	SatFat	Carbs	Fiber	Protein	Salt	Sugar/Carbs	Sugar/Cal	Net/Carbs	Label
D1	500	26.67	11.67	63.33	6.67	6.67	1.33	0.0526	0.0267	56.67	Low
D2	112	4.50	3.20	14.80	0.00	2.90	0.19	0.8243	0.4357	14.80	Medium
D3	467	20.00	1.67	63.30	3.33	10.00	1.92	0.1580	0.0857	59.97	Medium

2. Data Uji

Ambil 1 data uji yaitu :

Tabel 4.3 Data Uji

Energy	Fat	Satfat	Carbs	Fiber	Protein	Salt	Sugar/carbs	Sugar/kcal	Net carbs
450	22	5	60	4	8	1.5	0.10	0.06	56

3. Rumus Perhitungan Jarak

Metode jarak yang digunakan adalah Euclidean Distance:

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

4. Perhitungan Jarak ke Masing-Masing Data

a. Jarak ke D1

$$d = \sqrt{(450 - 500)^2 + (22 - 26.67)^2 + (5 - 11.67)^2 + (60 - 63.33)^2 + (4 - 6.67)^2 + (8 - 6.67)^2 + (1.5 - 1.33)^2 + (0.10 - 0.0526)^2 + (0.06 - 0.0267)^2 + (56 - 56.67)^2}$$

Hasil perhitungan:

$$d_{D1} \approx 51.02$$

b. Jarak ke D2

$$d_{D2} \approx 351.88$$

c. Jarak ke D3

$$d_{D3} \approx 24.66$$

5. Pengurutan Jarak

Tabel 4.4 Pengurutan Jarak

Data	Jarak	Label
D3	24.66	Medium
D1	51.02	Low
D2	351.88	Medium

6. Penentuan Nilai K

Misalkan digunakan: $K = 3$

Tetangga terdekat:

- a. D3 → Medium
- b. D1 → Low
- c. D2 → Medium

Jumlah kelas:

- 1) Medium = 2
- 2) Low = 1

Karena kelas Medium memiliki jumlah terbanyak, maka hasil klasifikasi untuk data uji adalah Medium

4.2.4.5 Uji Coba Proses Prediksi

Tahap proses prediksi dilakukan untuk menguji kemampuan model dalam mengklasifikasikan data baru yang belum pernah digunakan dalam proses pelatihan maupun pengujian. Model terbaik yang telah disimpan (`knn_final_model.pkl`) dimuat kembali bersama dengan objek Label Encoder (`label_encoder.pkl`).

Prediksi dilakukan menggunakan dua metode utama:

- a. `predict()` untuk menentukan kelas hasil klasifikasi.
- b. `predict_proba()` untuk memperoleh probabilitas masing-masing kelas.

Hasil prediksi kemudian dikembalikan ke bentuk label asli menggunakan metode `inverse_transform()`.

Berdasarkan pengujian terhadap tiga contoh produk, diperoleh hasil sebagai berikut:

- a. Produk 1 dengan kandungan gula 4 gram menghasilkan probabilitas

```
=== ANALISIS Produk A ===
Gula      : 4.0 g
Karbohidrat : 20.0 g
Energi    : 120.0 kkal
-----
Rasio Gula/Karbo      : 0.2000 (20.00%)
Rasio Energi dari Gula : 0.1333 (13.33%)
-----
HASIL KATEGORI : Low
```

Gambar 4. 19 Probabilitas Produk 1

- b. Produk 2 dengan kandungan gula 8 gram menghasilkan probabilitas

```
=== ANALISIS Produk B ===
Gula      : 8.0 g
Karbohidrat : 9.0 g
Energi    : 150.0 kkal
-----
Rasio Gula/Karbo      : 0.8889 (88.89%)
Rasio Energi dari Gula : 0.2133 (21.33%)
-----
HASIL KATEGORI : Medium
```

Gambar 4. 20 Probabilitas Produk 2

- c. Produk 3 dengan kandungan gula 25 gram menghasilkan probabilitas

```
=== ANALISIS Produk C ===  
Gula      : 25.0 g  
Karbohidrat : 30.0 g  
Energi    : 200.0 kkal  
-----  
Rasio Gula/Karbo      : 0.8333 (83.33%)  
Rasio Energi dari Gula : 0.5000 (50.00%)  
-----  
HASIL KATEGORI : High
```

Gambar 4. 21 Probabilitas Produk 3

4.2.5 Integrasi OCR dan YOLO ke Model

Tahap ini bertujuan untuk mengintegrasikan sistem deteksi area label nutrisi dan ekstraksi informasi nutrisi dari citra kemasan produk ke dalam model klasifikasi K-Nearest Neighbor (KNN) yang telah dilatih sebelumnya. Integrasi ini memungkinkan sistem melakukan klasifikasi kadar gula secara otomatis berdasarkan gambar label nutrisi produk. Sistem dibangun menggunakan framework Flask sebagai backend server yang mengelola proses deteksi objek, ekstraksi teks, pemrosesan fitur, hingga proses prediksi model.

4.2.5.1 Deteksi Area Tabel Nutrisi Menggunakan YOLO

Tahap pertama dalam pemrosesan citra adalah mendeteksi lokasi tabel nutrisi pada kemasan produk. Proses ini menggunakan model YOLO (You Only Look Once) yang telah dilatih sebelumnya dan disimpan dalam file best.pt.

Model YOLO bekerja dengan:

- a. Mendeteksi bounding box area tabel nutrisi.
- b. Menghitung confidence score untuk setiap deteksi.
- c. Memilih bounding box dengan nilai confidence tertinggi.
- d. Melakukan cropping pada area tersebut untuk diproses lebih lanjut.
- e. Parameter deteksi yang digunakan:
- f. Confidence threshold: 0.2
- g. IoU threshold: 0.5

Hasil dari tahap ini adalah:

- a. Citra hasil cropping tabel nutrisi
- b. Koordinat bounding box (x, y, width, height)
- c. Nilai confidence deteksi

Jika tidak ditemukan objek, sistem akan langsung memproses seluruh gambar tanpa cropping.

4.2.5.2 Ekstraksi Informasi Nutrisi Menggunakan EasyOCR

Citra hasil deteksi kemudian diproses menggunakan EasyOCR untuk mengekstraksi teks dari tabel nutrisi. Sebelum proses OCR dilakukan, citra melalui beberapa tahap preprocessing untuk meningkatkan akurasi pembacaan teks:

- a. Upscaling citra sebesar $2\times$ menggunakan interpolasi Lanczos.
- b. Konversi ke grayscale.
- c. Denoising menggunakan Non-local Means Denoising.
- d. Sharpening filter untuk mempertegas karakter.
- e. OCR tahap pertama menggunakan beam search decoder.
- f. Jika gagal, dilakukan fallback menggunakan adaptive threshold.

Teks hasil OCR kemudian:

- a. Diurutkan berdasarkan posisi vertikal dan horizontal.
- b. Digabungkan menjadi baris-baris teks.
- c. Difilter berdasarkan nilai probabilitas minimal 0.10.

Nilai nutrisi yang diekstraksi kemudian diproses untuk mendapatkan nilai numerik menggunakan fungsi ekstraksi angka berbasis regular expression.

Beberapa tahapan validasi dilakukan, antara lain:

- a. Konversi satuan miligram (mg) ke gram (g) untuk sodium.
- b. Estimasi nilai garam dari sodium jika field garam tidak tersedia (garam \approx sodium \times 2.5).
- c. Penggunaan nilai median dataset sebagai fallback untuk fitur yang hilang, yaitu:

- 1) Fiber = 1.8
 - 2) Protein = 6.0
 - 3) Salt = 0.47
- d. Pembatasan nilai tidak logis (misalnya gula lebih besar dari total karbohidrat).

4.2.5.2 Parsing dan Validasi Data Nutrisi

Teks hasil OCR kemudian diproses menggunakan metode berbasis regular expression untuk mengekstraksi nilai numerik nutrisi.

Sistem menggunakan pendekatan:

- a. Keyword mapping (misalnya: “gula”, “sugar”, typo hasil OCR seperti “guia”).
- b. Windowing (mencari angka pada baris sebelum dan sesudah).
- c. Penghapusan noise seperti persentase AKG.
- d. Ekstraksi seluruh angka dalam area pencarian.

Selanjutnya dilakukan validasi berbasis threshold logis, antara lain:

- a. Kalori: 50–200 kkal
- b. Protein: 0.5–5 gram
- c. Karbohidrat: 10–40 gram
- d. Gula: 5–30 gram
- e. Lemak: 0–10 gram
- f. Sodium: 5–500 mg

Jika angka berada di luar rentang logis, maka dianggap sebagai noise dan diabaikan.

Pendekatan ini digunakan untuk mengatasi kesalahan pembacaan OCR seperti:

- a. Angka AKG terbaca sebagai nilai nutrisi

- b. Salah baca 130 menjadi 13
- c. Tertukar antar baris akibat label miring

4.2.5.3 Perhitungan Fitur Turunan (Feature Engineering)

Setelah nilai nutrisi diperoleh, sistem menghitung fitur turunan yang sesuai dengan proses pelatihan model KNN.

Fitur yang digunakan adalah:

- 1. Energy per 100g (kkal)
- 2. Carbohydrates per 100g (gram)
- 3. Sugar per Carbohydrate Ratio

$$\frac{Gula}{Karbohidrat}$$

- 4. Sugar per Kcal Ratio

$$\frac{Gula}{Energi}$$

- 5. Sugar Dominance Score

$$\frac{Gula \times 4}{Energi}$$

4.2.5.5 Penyusunan Fitur Sesuai Struktur Model

Data fitur disusun dalam bentuk DataFrame dan disesuaikan dengan struktur model pelatihan menggunakan atribut: `knn_model.feature_names_in_`. Langkah ini penting karena algoritma KNN sensitif terhadap urutan kolom fitur. Kesalahan urutan dapat menyebabkan prediksi tidak akurat. Jika terdapat fitur yang tidak tersedia, sistem mengisi nilai default 0 agar struktur tetap konsisten.

4.2.5.6 Proses Prediksi KNN

Setelah fitur lengkap disusun, sistem melakukan prediksi menggunakan model KNN yang telah dilatih sebelumnya.

Metode yang digunakan:

1. `predict()` → Menghasilkan kelas numerik (0, 1, atau 2)
2. Label encoder → Mengonversi indeks menjadi label kategori

Mapping kelas berdasarkan hasil training:

- a. Index 1 → Low Sugar (Grade A)
- b. Index 2 → Medium Sugar (Grade B)
- c. Index 0 → High Sugar (Grade C)

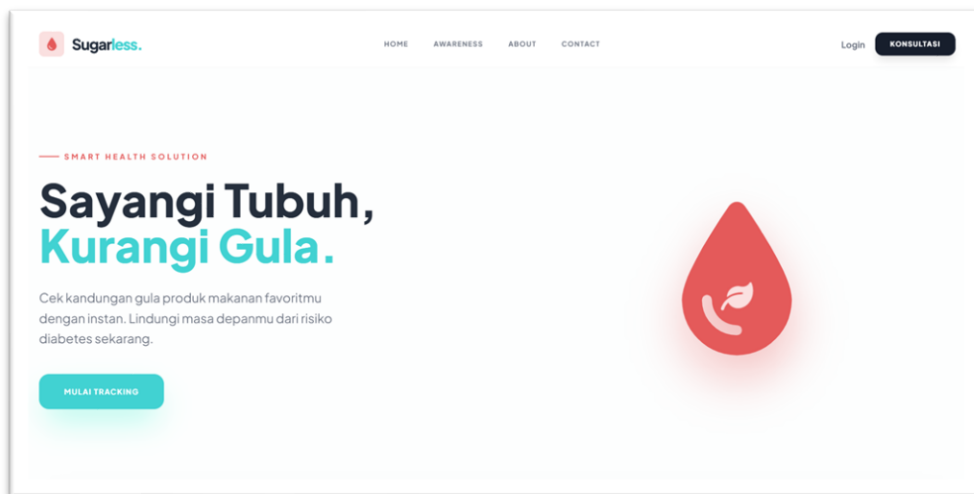
Sistem kemudian mengembalikan hasil dalam format JSON yang berisi:

1. Nilai nutrisi hasil OCR
2. Fitur yang digunakan model
3. Kategori kadar gula (Low / Medium / High)
4. Grade (A/B/C)
5. Confidence deteksi YOLO
6. Waktu pemrosesan.

4.2.6 Implementasi Antarmuka Sistem

4.2.6.1 Halaman Home

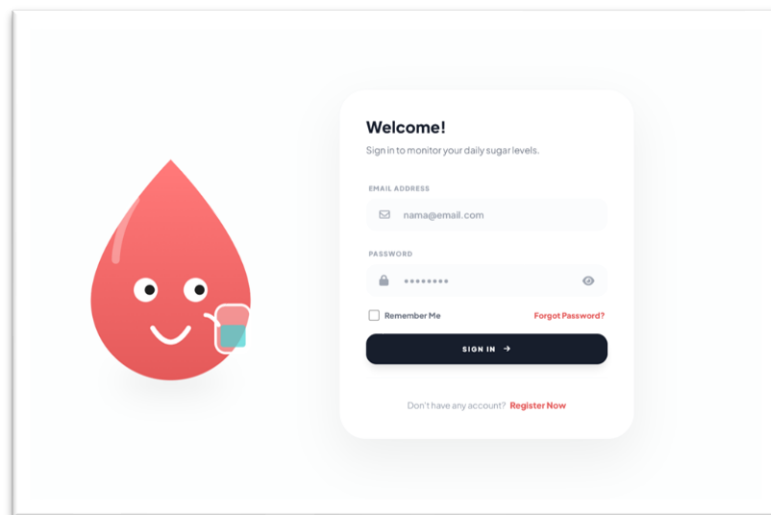
Halaman ini merupakan Landing Page aplikasi “Sugarless” dengan desain bersih dan modern berlatar putih, dilengkapi navigasi utama seperti Home, Awareness, About, Contact, serta tombol “Login” dan “KONSULTASI”. Bagian tengah menampilkan judul persuasif “Sayangi Tubuh, Kurangi Gula.” sebagai pesan pencegahan diabetes, disertai sub-teks fitur pengecekan kandungan gula dan tombol “MULAI TRACKING” berwarna toska sebagai Call to Action. Ikon tetesan darah berwarna merah di sisi kanan memperkuat identitas aplikasi kesehatan, sementara keseluruhan antarmuka dirancang intuitif untuk meningkatkan kesadaran pengguna terhadap pola konsumsi gula.



Gambar 4.22 Halaman Home

4.2.6.1 Halaman Form Login

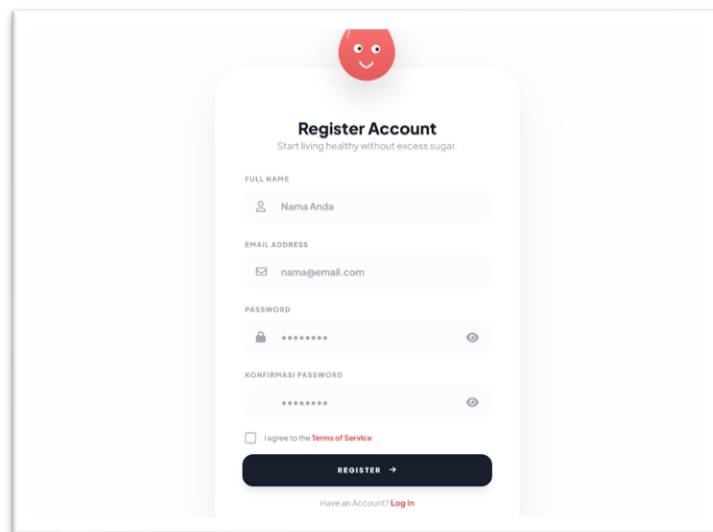
Halaman ini merupakan tampilan Login dengan desain bersih dan minimalis bernuansa putih dengan aksen merah, menampilkan judul “Welcome!” dan sub-teks yang menjelaskan fungsi sistem untuk memantau kadar gula harian. Di sisi kiri terdapat maskot tetesan darah tersenyum yang memberi kesan ramah, sementara form menyediakan kolom Email Address dan Password lengkap dengan ikon, placeholder, serta fitur ikon mata untuk menampilkan atau menyembunyikan kata sandi. Halaman ini juga dilengkapi opsi “Remember Me”, tautan “Lupa Password?” untuk pemulihan akun, tombol “SIGN IN” sebagai akses masuk, dan pilihan “Daftar gratis” bagi pengguna yang belum memiliki akun.



Gambar 4.23 Halaman Form Login

4.2.6.2 Halaman Form Register

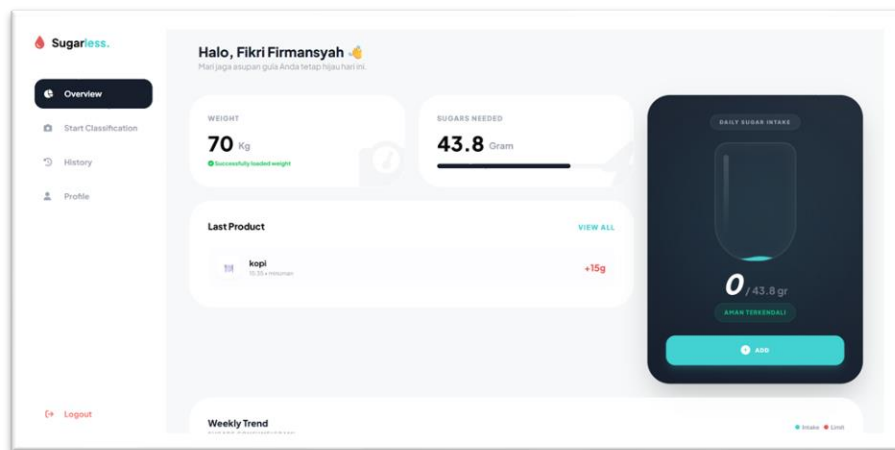
Halaman ini merupakan formulir pendaftaran akun baru (Register Account) dengan desain kartu putih terpusat dan aksesoris maskot tetesan darah di bagian atas. Form ini menyediakan empat kolom input utama, yaitu Full Name, Email Address, Password, dan Konfirmasi Password, di mana masing-masing kolom dilengkapi dengan ikon pendukung dan fitur *visibility* pada kolom kata sandi. Di bagian bawah input, terdapat *checkbox* persetujuan "Terms of Service" dan tombol utama "REGISTER" berwarna gelap untuk memproses pembuatan akun. Halaman ini juga menyediakan opsi "Log In" di bagian paling bawah untuk memudahkan pengguna yang sudah memiliki akun berpindah ke halaman masuk secara langsung.

The image shows a mobile-style registration form titled "Register Account" with the subtitle "Start living healthy without excess sugar". At the top center is a red circular maskot character with a smiling face. The form contains four input fields: "FULL NAME" with a person icon and placeholder "Nama Anda"; "EMAIL ADDRESS" with an envelope icon and placeholder "nama@email.com"; "PASSWORD" with a lock icon, placeholder "*****", and a visibility toggle; and "KONFIRMASI PASSWORD" with a lock icon, placeholder "*****", and a visibility toggle. Below the fields is a checkbox labeled "I agree to the Terms of Service". At the bottom is a dark blue button with the text "REGISTER" and a right-pointing arrow. Below the button is a link: "Have an Account? Log In".

Gambar 4.24 Halaman Form Register

4.2.6.7 Halaman Dashboard

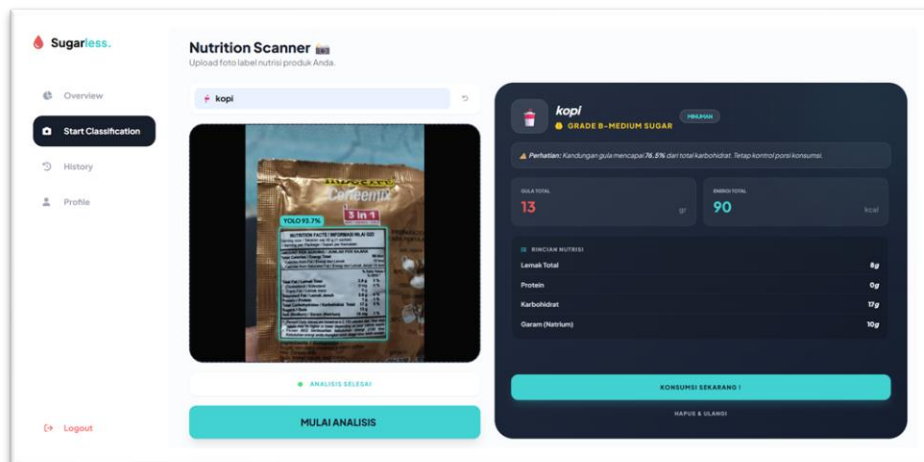
Halaman ini merupakan panel kendali utama pengguna yang menyajikan ringkasan data kesehatan secara visual. Di bagian kiri terdapat sidebar navigasi (Overview, Start Classification, History, Profile), sementara bagian utama menampilkan data personal seperti berat badan (Weight) dan batas konsumsi gula harian (Sugars Needed). Terdapat pula daftar riwayat produk terakhir yang dikonsumsi (Last Product) dan indikator progres konsumsi gula harian (Daily Sugar Intake) yang interaktif untuk memastikan asupan gula pengguna tetap dalam batas aman.



Gambar 4.25 Halaman Dashboard

4.2.6.8 Halaman Klasifikasi

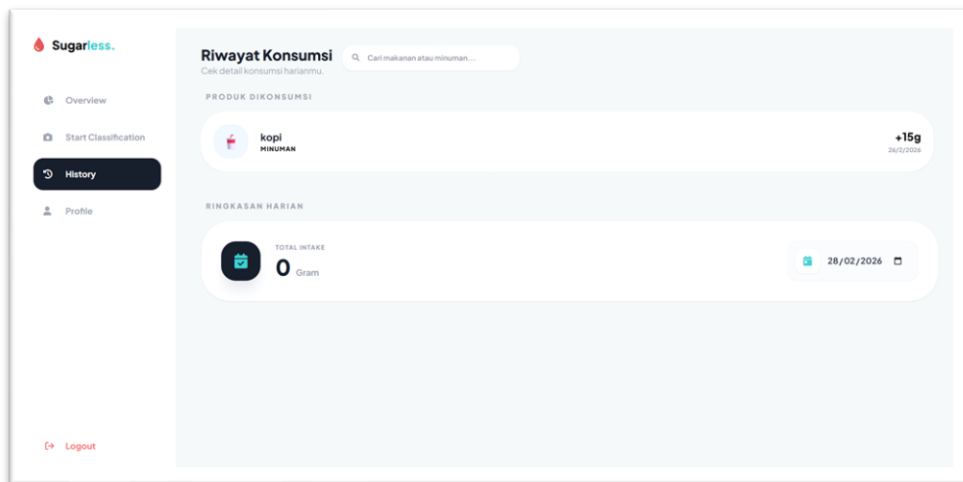
Halaman ini merupakan fitur utama aplikasi untuk mendeteksi kandungan nutrisi melalui kamera. Tampilan terbagi menjadi dua bagian. Jendela pemindai di sisi kiri untuk menangkap label nutrisi produk, dan panel hasil analisis di sisi kanan. Panel hasil menampilkan klasifikasi produk (misalnya: "GRADE HIGH – MEDIUM SUGAR") beserta rincian jumlah gula, energi, karbohidrat, dan lemak. Panel hasil analisis akan keluar apabila proses ekstraksi dan klasifikasi telah selesai dan server mengembalikan kembali berupa nilai nutrisi yang dimuat didalam antarmuka tabel dinamis. Pengguna dapat memilih untuk menekan tombol "KONSUMSI SEKARANG !" untuk mencatat konsumsi atau "HAPUS & ULANGI" untuk memindai produk lain.



Gambar 4.26 Halaman Klasifikasi

4.2.6.9 Halaman History

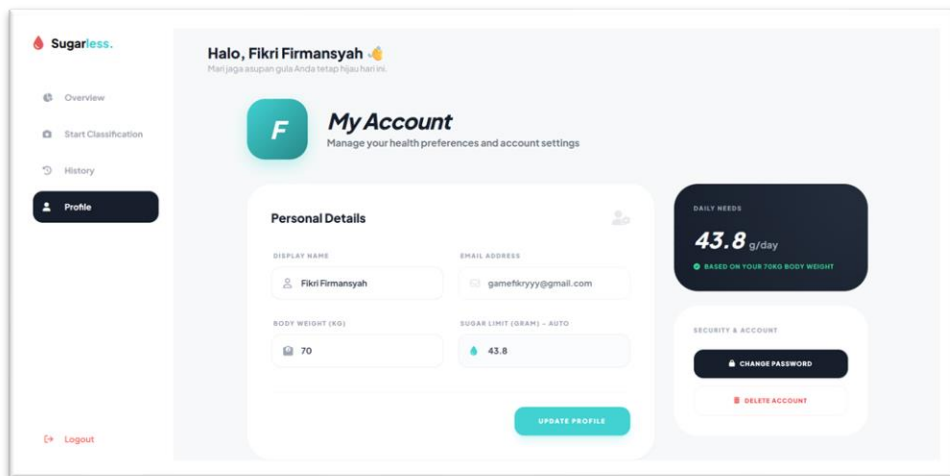
Halaman ini merupakan antarmuka yang menampilkan rekaman aktivitas konsumsi pengguna secara kronologis. Pada bagian atas, terdapat kolom pencarian untuk memudahkan pengguna menemukan jenis makanan atau minuman tertentu yang telah dicatat sebelumnya. Bagian utama halaman terbagi menjadi dua panel kartu: “PRODUK DIKONSUMSI” yang merinci nama item, kategori (misalnya: minuman), tanggal input, serta jumlah akumulasi gula (misalnya: +15g), dan panel “RINGKASAN HARIAN” yang menampilkan total asupan (Total Intake) pada tanggal tertentu. Halaman ini juga dilengkapi dengan fitur kalender interaktif untuk memantau riwayat kesehatan pada hari-hari sebelumnya, guna mendukung konsistensi pengguna dalam menjaga diet rendah gula.



Gambar 4.27 Halaman History

4.2.6.10 Halaman Profile

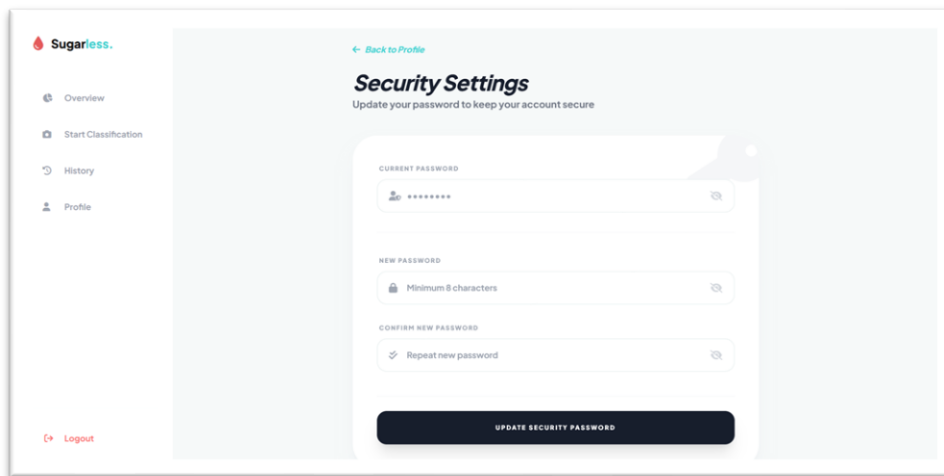
Halaman ini merupakan antarmuka manajemen akun yang memungkinkan pengguna mengelola informasi personal dan preferensi kesehatan. Di bagian atas ditampilkan identitas pengguna berupa inisial profil dan judul “My Account”, sementara form utama memuat kolom Personal Details seperti Display Name, Email Address, Body Weight (KG), serta Sugar Limit (Gram) yang dihitung otomatis. Di sisi kanan terdapat ringkasan Daily Needs yang menampilkan batas konsumsi gula harian berdasarkan berat badan, serta panel Security & Account dengan opsi “CHANGE PASSWORD” dan “DELETE ACCOUNT”. Perubahan data dapat disimpan melalui tombol “UPDATE PROFILE” berwarna toska di bagian bawah formulir.



Gambar 4. 28 Halaman Profil

4.2.6.11 Halaman Change Password

Halaman ini merupakan sub-menu dari profil yang berfungsi untuk memperbarui kredensial keamanan pengguna. Dengan tampilan kartu putih yang bersih, halaman ini menyediakan tiga kolom input utama: Current Password untuk verifikasi identitas awal, New Password dengan instruksi minimal 8 karakter, serta Confirm New Password untuk memastikan kecocokan kata sandi baru. Setiap kolom dilengkapi dengan ikon indikator dan fitur *visibility* untuk meminimalisir kesalahan pengetikan. Pengguna dapat menyelesaikan perubahan dengan menekan tombol “UPDATE SECURITY PASSWORD” yang berwarna gelap. Terdapat pula navigasi “Back to Profile” di bagian atas untuk memudahkan pengguna kembali ke halaman pengaturan akun sebelumnya.



Gambar 4.29 Halaman Change Password

4.3 Pengujian Sistem

Pengujian sistem ini dilakukan untuk memastikan bahwa setiap fitur dalam aplikasi Sugarless berfungsi sesuai dengan kebutuhan fungsional yang telah dirancang sebelumnya. Metode yang digunakan dalam tahap ini adalah Blackbox Testing. Blackbox Testing (Pengujian Kotak Hitam) merupakan metode pengujian perangkat lunak yang berfokus pada sisi fungsionalitas tanpa harus melihat atau memahami struktur kode internal aplikasi. Dalam pengujian ini, sistem diperlakukan sebagai "kotak hitam" di mana penguji hanya memberikan input dan mengamati output yang dihasilkan untuk memastikan apakah sudah sesuai dengan hasil yang diharapkan.

Penggunaan metode ini pada aplikasi Sugarless bertujuan untuk:

1. Memvalidasi Alur Kerja: Memastikan alur navigasi dari Login hingga proses analisis produk berjalan lancar bagi pengguna.
2. Keakuratan Output: Memastikan data nutrisi yang diekstraksi melalui OCR dan hasil klasifikasi tingkat gula ditampilkan dengan benar pada antarmuka pengguna.
3. Deteksi Kesalahan Input: Menjamin bahwa sistem mampu menangani kesalahan input, seperti format email yang salah atau kolom yang dikosongkan.

Dengan pendekatan ini, pengujian dapat dilakukan dari sudut pandang pengguna akhir untuk menjamin pengalaman pengguna (user experience) yang optimal sebelum aplikasi diimplementasikan secara penuh.

4.3.1 Blackbox Testing Halaman Register

Tabel 4.5 Blackbox Halaman Register

No	Pengujian	Test Case	Hasil Yang Diharapkan	Hasil Pengujian	Status
1	Validasi Input Lengkap	Mengisi semua <i>field</i> (Nama, Email, Password, Berat Badan) dengan data valid	Sistem menerima input dan mengaktifkan tombol Register .	Sesuai harapan	Valid
2	Fitur <i>Show/Hide</i> Password	Mengklik ikon mata pada kolom kata sandi	Karakter password berubah dari mode tersembunyi (<i>bullet</i>) menjadi teks terbaca, dan sebaliknya.	Sesuai harapan	Valid
3	Cek Duplikasi Email	Menginput alamat email yang sudah terdaftar dalam database	Sistem menampilkan pesan error "Email sudah digunakan" dan pendaftaran ditolak.	Sesuai harapan	Valid
4	Navigasi <i>Terms & Conditions</i>	Mengklik tautan "Syarat dan Ketentuan" pada form registrasi	Sistem menampilkan page Syarat dan Ketentuan penggunaan aplikasi.	Sesuai harapan	Valid
5	Eksekusi Registrasi	Mengklik tombol Register setelah semua data valid diisi	Data pengguna baru tersimpan ke database dan sistem mengirimkan email verifikasi.	Sesuai harapan	Valid

4.3.2 Blackbox Testing Halaman Login

Tabel 4.6 Blackbox Testing Halaman Login

No	Pengujian	Test Case	Hasil Yang Diharapkan	Hasil Pengujian	Status
1	Login Berhasil	Mengisi Email dan Password yang valid lalu klik tombol SIGN IN	Sistem berhasil masuk ke Dashboard/Halaman Utama.	Sesuai Harapan	Valid
2	Login Gagal	Mengosongkan Email dan Password	Muncul pesan peringatan bahwa	Sesuai Harapan	Valid

	(Data Kosong)	lalu klik tombol SIGN IN	form tidak boleh kosong.		
3	Validasi Format Email	Mengisi Email tanpa karakter "@" (contoh: namagmail.com)	Sistem menolak input dan menampilkan pesan "Format email tidak valid".	Sesuai Harapan	Valid
4	Fitur <i>Show Password</i>	Mengklik ikon mata pada kolom Password	Teks password yang disamarkan (titik-titik) berubah menjadi teks asli.	Sesuai Harapan	Valid
5	Navigasi Registrasi	Mengklik link Register Now di bagian bawah	Sistem mengarahkan pengguna ke halaman pendaftaran akun baru.	Sesuai Harapan	Valid
6	Lupa Password	Mengklik link Forgot Password?	Sistem mengarahkan pengguna ke halaman pemulihan kata sandi.	Sesuai Harapan	Valid

4.3.3 Blackbox Testing Halaman Dashboard

Tabel 4.7 Blackbox Testing Halaman Dashboard

No	Pengujian	Test Case	Hasil Yang Diharapkan	Hasil Pengujian	Status
1		Memeriksa teks "Successfully loaded weight" pada kartu Weight	Sistem berhasil mengambil dan menampilkan data berat badan pengguna secara <i>real-time</i> .	Sesuai Harapan	Valid
2	Fitur Tambah Intake	Mengklik tombol + ADD pada kartu Daily Sugar Intake	Sistem memunculkan page klasifikasi untuk menambah jumlah konsumsi gula harian.	Sesuai Harapan	Valid
3		Mengklik link VIEW ALL pada bagian Last Product	Sistem menampilkan seluruh daftar produk yang pernah dikonsumsi di halaman History.	Sesuai Harapan	Valid

4		indikator "AMAN TERKENDALI" di kartu intake	Status berubah (misal menjadi merah/peringatan) jika intake mendekati atau melebihi limit yang ditetapkan melalui proporsi berat badan	Sesuai Harapan	Valid
---	--	---	--	----------------	-------

4.3.4 Blackbox Testing Halaman Klasifikasi

Tabel 4.8 Blackbox Testing Halaman Klasifikasi

No	Pengujian	Test Case	Hasil Yang Diharapkan	Hasil Pengujian	Status
1	Pemilihan Jenis Produk	Mengklik opsi Makanan atau Minuman sebelum unggah	Sistem menandai kategori produk yang dipilih untuk menyesuaikan logika klasifikasi.	Sesuai Harapan	Valid
2	Unggah Foto Label Nutrisi	Mengklik area Klik untuk Pilih Foto dan memilih gambar label produk	Gambar berhasil terunggah dan ditampilkan pada area pratinjau scanner.	Sesuai Harapan	Valid
3	Deteksi Objek (YOLOv8), OCR dan Klasifikasi KNN	Mengklik tombol MULAI ANALISIS	Sistem menjalankan YOLOv8 untuk mendeteksi area informasi nilai gizi (muncul <i>bounding box</i> hijau). Serta mengekstraksi nilai nutrisi yang terdeteksi OCR. Lalu sistem memproses hasil nutrisi ke model KNN, hasil di keluarkan dan dikembalikan ke frontend oleh sistem.	Sesuai Harapan	Valid
4	Output berhasil di tampilkan	Tabel result muncul	Panel hasil analisis muncul di sisi kanan, menampilkan Nama Produk, Grade Gula, dan Rincian Nutrisi (Lemak, Karbo, dll).	Sesuai Harapan	Valid
5	Akurasi Data (OCR)	Membandingkan data pada label fisik dengan teks di tabel Rincian Nutrisi	Nilai angka (gram/kkal) yang diekstraksi oleh OCR, dapat di	Sesuai Harapan	Valid

			perbaiki ketika salah angka karna ekstraksi OCR		
6	Simpan Riwayat Konsumsi	Mengklik tombol KONSUMSI SEKARANG!	Data produk berhasil disimpan ke dalam sistem dan ditambahkan ke <i>Daily Sugar Intake</i> di halaman Overview.	Sesuai Harapan	Valid
7	Reset Analisis	Mengklik tombol HAPUS & ULANGI	Panel hasil analisis tertutup dan form kembali ke keadaan awal (kosong) untuk pemindaian baru	Sesuai Harapan	Valid

4.3.5 Blackbox Testing Halaman Histori

Tabel 4.9 Blackbox Testing Halaman Histori

No	Pengujian	Test Case	Hasil Yang Diharapkan	Hasil Pengujian	Status
1	Fitur Pencarian Produk	Mengetikkan kata kunci "kopi" pada kolom <i>Cari makanan atau minuman...</i>	Sistem memfilter daftar dan hanya menampilkan produk dengan nama "kopi".	Sesuai Harapan	Valid
2	Filter Berdasarkan Tanggal	Memilih tanggal tertentu (misal: 04/03/2026) pada <i>datepicker</i> di bagian Ringkasan Harian	Sistem memperbarui tampilan daftar produk sesuai dengan tanggal yang dipilih.	Sesuai Harapan	Valid
3	Akurasi Total Intake	Memeriksa nilai Total Intake di bagian bawah halaman	Sistem mengambil seluruh gram gula dari produk yang dikonsumsi pada hari tersebut dengan benar.	Sesuai Harapan	Valid
4	Menampilkan seluruh produk	Mengambil semua produk yang tersimpan	Sistem menampilkan kategori produk sesuai dengan detail data yang diinput.	Sesuai Harapan	Valid

4.3.6 Blackbox Testing Halaman Profil

Tabel 4.10 Blackbox Testing Halaman Profil

No	Pengujian	Test Case	Hasil Yang Diharapkan	Hasil Pengujian	Status
1	Update Nama	Mengubah data pada kolom <i>Display Name</i> lalu klik “UPDATE PROFILE”	Sistem berhasil menyimpan perubahan dan menampilkan notifikasi sukses.	Sesuai Harapan	Valid
2	Perhitungan Otomatis Batas Gula	Mengubah nilai pada <i>Body Weight (KG)</i> (misal dari 70 ke 80)	Nilai pada <i>Sugar Limit (Gram) - Auto</i> dan kartu <i>Daily Needs</i> otomatis diperbarui berdasarkan berat badan baru.	Sesuai Harapan	Valid
3	Validasi Input Berat Badan	Menginput nilai non-numerik atau angka negatif pada kolom berat badan	Sistem menolak input dan menampilkan pesan peringatan validasi.	Sesuai Harapan	Valid
4	Fitur Ganti Password	Mengklik tombol “CHANGE PASSWORD”	Sistem menampilkan modal atau mengarahkan ke halaman untuk mengubah kata sandi lama ke baru.	Sesuai Harapan	Valid
5	Keamanan Hapus Akun	Mengklik tombol “DELETE ACCOUNT”	Sistem menampilkan dialog konfirmasi (<i>confirmation box</i>) untuk mencegah penghapusan akun yang tidak sengaja.	Sesuai Harapan	Valid

4.3.7 Blackbox Testing Navigasi Sidebar

Tabel 4.11 Blackbox Testing Navigasi Sidebar

No	Pengujian	Test Case	Hasil Yang Diharapkan	Hasil Pengujian	Status
1	Sidebar Overview	Ketika di klik menampilkan halaman dashboard	Berhasil menampilkan halaman dashboard	Sesuai Harapan	Valid
2	Sidebar Start Classification	Ketika di klik menampilkan halaman klasifikasi	Berhasil menampilkan halaman klasifikasi	Sesuai Harapan	Valid
3	Sidebar History	Ketika di klik menampilkan halaman Histori	Berhasil menampilkan halaman histori	Sesuai Harapan	Valid
4	Sidebar Profile	Ketika di klik menampilkan halaman Profil	Berhasil menampilkan halaman profil	Sesuai Harapan	Valid
5	Sidebar Logout	Ketika di klik memunculkan konfirmasi ulang untuk Logout	Berhasil logout sesi di hapus	Sesuai Harapan	Valid

BAB V

PENUTUP

5.1 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:

1. 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.
2. 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.
3. 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.
4. 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.

5. Implementasi sistem dalam bentuk prototipe berbasis web telah berhasil merepresentasikan proses klasifikasi secara terintegrasi, mulai dari input data, perhitungan KNN, hingga penyajian hasil kategori gula beserta confidence dan rekomendasi konsumsi. Sistem ini dapat menjadi alat bantu edukatif dalam meningkatkan literasi gizi masyarakat.

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.

5.2 Saran

Berdasarkan hasil penelitian dan implementasi sistem yang telah dilakukan, terdapat beberapa saran untuk pengembangan lebih lanjut, yaitu:

1. Penelitian selanjutnya dapat melakukan perbandingan algoritma K-Nearest Neighbor (KNN) dengan metode klasifikasi lain seperti Decision Tree, Random Forest, atau Support Vector Machine (SVM) untuk mengetahui performa yang paling optimal dalam klasifikasi kandungan gula.

2. Dataset yang digunakan dapat diperluas dengan jumlah data yang lebih besar dan lebih beragam agar model memiliki kemampuan generalisasi yang lebih baik serta dapat menangani variasi produk yang lebih luas.
3. Teknologi Optical Character Recognition (OCR) yang digunakan dalam penelitian ini masih terbatas pada gambar dengan kualitas yang baik (tidak buram, pencahayaan cukup, dan posisi sejajar). Pengembangan selanjutnya disarankan menggunakan model OCR yang lebih robust atau dilakukan pelatihan ulang (retraining) agar sistem lebih tahan terhadap variasi kualitas gambar.
4. Sistem deteksi objek dan OCR pada penelitian ini tidak menggunakan metode deep learning yang lebih kompleks karena keterbatasan sumber daya perangkat keras dan komputasi. Pada pengembangan berikutnya, penggunaan model yang lebih berat dengan dukungan spesifikasi perangkat yang lebih tinggi atau layanan cloud computing dapat dipertimbangkan untuk meningkatkan akurasi dan stabilitas sistem.
5. Implementasi saat ini masih terbatas pada mekanisme unggah (upload) gambar dan belum mendukung pengambilan gambar secara real-time melalui kamera. Penelitian selanjutnya dapat mengembangkan fitur pemindaian langsung (live camera scanning) agar sistem lebih praktis dan mendekati implementasi aplikasi mobile yang sesungguhnya.

DAFTAR PUSTAKA

- 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>
- BPOM RI. (2021). *Edukasi Nutrisi Seimbang Edukasi Nutrisi Seimbang Edisi: Masyarakat Umum Edisi: Masyarakat Umum.*
- 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>

Diabetes Atlas. (2024).

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>
- Khairunnisa, K., & Martiano, M. (2025). Classification of Defects in Arabica Coffee Beans Using the KNN (K-Nearest Neighbor) Method in the Gayo Farming Cooperative, Bebeseb District, Central Aceh. *IJATCoS: Indonesian Journal of Applied Technology, Computer and Science*, 2(1), 15–24. <https://jurnal.umsu.ac.id/index.php/Ijatcos/article/view/23300>
- Open Food Fact. (2025, April 30). *Reusing Open Food Facts Data - Open Food Facts wiki*. https://wiki.openfoodfacts.org/Reusing_Open_Food_Facts_Data
- 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 Mining Clustering Dalam Mengelompokkan 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>
- 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 techniques* Witten, 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 Klasifikasi penyakit Diabetes....(Fitri)*.
- allagui. (2025, 7 3). *Roboflow Universe*. Retrieved from <https://universe.roboflow.com/allagui/nutrition-table-detection/dataset/3>

ORIGINALITY REPORT

22%
SIMILARITY INDEX

19%
INTERNET SOURCES

11%
PUBLICATIONS

11%
STUDENT PAPERS

PRIMARY SOURCES

1	repository.umsu.ac.id Internet Source	5%
2	mie.binus.ac.id Internet Source	1%
3	docplayer.info Internet Source	1%
4	Submitted to SDM Universitas Gadjah Mada Student Paper	1%
5	jurnal.tiga-mutiara.com Internet Source	1%
6	123dok.com Internet Source	<1%
7	Submitted to UPN Veteran Yogyakarta Student Paper	<1%
8	repository.machung.ac.id Internet Source	<1%
9	Submitted to Universitas Sumatera Utara Student Paper	<1%
10	Submitted to Dongguk University Student Paper	<1%
11	Misbahu Surur, Nugroho Adhi Santoso, Bayu Aji Santoso. "Klasifikasi Keterlambatan Pembayaran SPP Santri Menggunakan Algoritma K-Nearest Neighbor di Pesantren Al	<1%

Fajar Tegal", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025

Publication

12	repository.usni.ac.id Internet Source	<1 %
13	repository.upr.ac.id Internet Source	<1 %
14	ejurnal.seminar-id.com Internet Source	<1 %
15	ejournal.uniramalang.ac.id Internet Source	<1 %
16	moam.info Internet Source	<1 %
17	Submitted to Universitas Muhammadiyah Palembang Student Paper	<1 %
18	eprints.amikom.ac.id Internet Source	<1 %
19	repository.ub.ac.id Internet Source	<1 %
20	M. Althaf Kiram, Eva Darnila, Ilham Sahputra. "Machine Learning Klasifikasi Penyakit Jiwa Menggunakan Metode K-Nearest Neighbor Berbasis Web", Jurnal Ners, 2025 Publication	<1 %
21	adinjava.blogspot.com Internet Source	<1 %
22	Submitted to Universitas Brawijaya Student Paper	<1 %
23	Submitted to Universitas Tarumanagara Student Paper	<1 %

24	www.tribunnews.com Internet Source	<1 %
25	pt.scribd.com Internet Source	<1 %
26	Tatya Hanum Pramudita, M. Zainal Arifin. "Hyperparameter Optimization of Light Gradient Boosting Machine for Microcirculation Detection Wearable Data", Indonesian Journal of Innovation Studies, 2026 Publication	<1 %
27	eprints.umpo.ac.id Internet Source	<1 %
28	repository.unifa.ac.id Internet Source	<1 %
29	www.scribd.com Internet Source	<1 %
30	iainpurwokerto.ac.id Internet Source	<1 %
31	I Gede Haga Olas Tyamarta -, I Nyoman Sukajaya -, Putu Kartika Dewi -. "DIAGNOSIS PENYAKIT DEMAM BERDARAH DENGUE MENERAPKAN ALGORITMA K-NEAREST NEIGHBOR", Jurnal Informatika dan Teknik Elektro Terapan, 2026 Publication	<1 %
32	repository.ubharajaya.ac.id Internet Source	<1 %
33	Submitted to City University Student Paper	<1 %

34	Zaenul Amri, Muhammad Rodi, M. Nurul Wathani, Amir Bagja, Zulkipli. "Prediksi Diabetes Menggunakan Algoritma K-Nearest (KNN) Teknik SMOTE-ENN", Infotek: Jurnal Informatika dan Teknologi, 2025 Publication	<1 %
35	fikti.umsu.ac.id Internet Source	<1 %
36	Submitted to iGroup Student Paper	<1 %
37	journal.unhas.ac.id Internet Source	<1 %
38	journals.gesociety.org Internet Source	<1 %
39	repository.dinamika.ac.id Internet Source	<1 %
40	journal.aiska-university.ac.id Internet Source	<1 %
41	journal.aptii.or.id Internet Source	<1 %
42	jurnalfti.unmer.ac.id Internet Source	<1 %
43	repositori.uma.ac.id Internet Source	<1 %
44	Lukman Hakim, Agus Waluyo. "Tingkat religiusitas, pengetahuan produk, dan sikap konsumen terhadap pembelian makanan dan minuman halal", Journal of Halal Industry Studies, 2023 Publication	<1 %
45	Submitted to Universitas Budi Luhur	

<1 %

46

Submitted to University of San Diego

Student Paper

<1 %

47

eprints.akakom.ac.id

Internet Source

<1 %

48

repo.palcomtech.ac.id

Internet Source

<1 %

49

repository.unissula.ac.id

Internet Source

<1 %

50

www.researchgate.net

Internet Source

<1 %

51

Submitted to Police Academy – University of
Police Science

Student Paper

<1 %

52

Submitted to Telkom University

Student Paper

<1 %

53

ejurnal.stmik-budidarma.ac.id

Internet Source

<1 %

54

gentongmas.com

Internet Source

<1 %

55

journal.mediapublikasi.id

Internet Source

<1 %

56

Submitted to Universitas Pancasila

Student Paper

<1 %

57

ejournal.catursakti.ac.id

Internet Source

<1 %

58

Murtiwiayati Murtiwiayati, Rischa Fitriawhalia
Arini, Leli Safitri, Santi Widiyanti, Iwan Setiadi.
"Implementasi Algoritma K-Nearest

<1 %

Neighbors untuk Klasifikasi Sentimen pada
Ulasan Pakaian Anak di Platform TikTok
Shop", Jurnal Minfo Polgan, 2025

Publication

-
- | | | |
|----|---|------|
| 59 | Septian Dwi Saputra, Jadiaman Parhusip.
"PENGUKURAN USABILITY APLIKASI WEB
MENGUNAKAN SUS (SYSTEM USABILITY
SCALE) DAN PENGUJIAN BLACK BOX PADA
WEBSITE E-LEARNING (GEN-IT) KABUPATEN
KATINGAN", Jurnal Informatika dan Teknik
Elektro Terapan, 2026
Publication | <1 % |
| 60 | Submitted to Sultan Agung Islamic University
Student Paper | <1 % |
| 61 | Submitted to Universitas Kristen Wira Wacana
Sumba
Student Paper | <1 % |
| 62 | Submitted to Universitas Pamulang
Student Paper | <1 % |
| 63 | Submitted to Universitas Pendidikan
Indonesia
Student Paper | <1 % |
| 64 | Submitted to Universitas Siliwangi
Student Paper | <1 % |
| 65 | sistemasi.ftik.unisi.ac.id
Internet Source | <1 % |
| 66 | Anindya Khrisna Wardhani, Rano Indradi
Sudra, Ega Nugraha, Astrid Novita Putri.
"Optimasi Nilai K pada Algoritma K-Nearest
Neighbor untuk Klasifikasi Kesehatan Janin",
Journal of Computer Science and Technology
(JCS-TECH), 2025
Publication | <1 % |
-

67	Betrisandi Betrisandi, Rahmat Thaib. "Comparison of NBC and KNN in Classifying Stunting in Children in Rural Areas", Jambura Journal of Electrical and Electronics Engineering, 2026 Publication	<1 %
68	Fahreza Ananda Kusuma. "PEMODELAN KLASIFIKASI ANEMIA APLASTIK MENGGUNAKAN TEKNIK OVERSAMPLING DAN K-NEAREST NEIGHBORS", Jurnal Informatika dan Teknik Elektro Terapan, 2024 Publication	<1 %
69	Submitted to Instituto Madrileno de Formacion Student Paper	<1 %
70	Submitted to Sriwijaya University Student Paper	<1 %
71	Submitted to Universitas Airlangga-1 Student Paper	<1 %
72	Submitted to Universitas Indraprasta PGRI Student Paper	<1 %
73	Submitted to Universitas Malikussaleh Student Paper	<1 %
74	Submitted to Universitas Sangga Buana YPKP Student Paper	<1 %
75	jurnal.polbeng.ac.id Internet Source	<1 %
76	jurnal.tau.ac.id Internet Source	<1 %
77	opac.uad.ac.id Internet Source	<1 %

78	repo.undiksha.ac.id Internet Source	<1 %
79	www.obatjantungstroke.com Internet Source	<1 %
80	Budi Santoso, Nira Rusanti, Azwar Riza Habibi, Vivi Aida Fitria. "Implementasi Smart Class Berbasis IoT di Institut Teknologi dan Bisnis Asia Malang", Jurnal Ilmiah Teknologi Informasi Asia, 2020 Publication	<1 %
81	Submitted to Hankuk University of Foreign Studies Student Paper	<1 %
82	Kurnia Ramadhan Putra, Bahy Tsany Rizqullah. "Perbandingan Algoritma KNN dan SVM Dalam Memprediksi Metaverse Crypto Token", Teknologi, 2025 Publication	<1 %
83	Submitted to Universitas Islam Riau Student Paper	<1 %
84	Submitted to Universitas Muhammadiyah Purwokerto Student Paper	<1 %
85	Submitted to Universitas Muslim Indonesia Student Paper	<1 %
86	Submitted to Universitas Putera Batam Student Paper	<1 %
87	dspace.uui.ac.id Internet Source	<1 %
88	sisfotenika.stmikpontianak.ac.id Internet Source	<1 %

89 Alifqi Radjavani, Theopilus Bayu Sasongko. "Analisa Perbandingan Algoritma CNN dan LSTM untuk Klasifikasi Pesan Cyberbullying pada Twitter", *The Indonesian Journal of Computer Science*, 2023
Publication

90 Hidayat Hidayat, Musliadi KH, Indar Kusmanto, Munawirah Kadir, Kristian Kristian. "Klasifikasi Mahasiswa Calon Penerima Beasiswa KIP Menggunakan Algoritma Naive Bayes di Universitas Tomakaka Mamuju", *JURNAL FASILKOM*, 2025
Publication

91 Muswardi Muswardi, Joko Widodo. "PkM: Sosialisasi Deep Learning untuk Optimalisasi Kualitas dan Efisiensi pada Pembelajaran Teaching Factory di Era Industri 4.0", *Jurnal Pengabdian UntukMu NegeRI*, 2026
Publication

92 Siti Fadhilah, Dian Monalisa Rusliani, Arum Nuryati, Bima Suryantara. "Deteksi Dini Faktor Risiko Penyakit Tidak Menular pada Remaja Karang Taruna: Skrining dan Upaya Pencegahan", *Room of Civil Society Development*, 2025
Publication

93 Tata Arya Cahyaaty, Herlawati Herlawati, Andy Achmad Hendhar Setiawan. "Model Prediksi Kondisi Kesehatan dari Data Medical Check-Up Menggunakan K-Nearest Neighbors dan Decision Tree", *Journal of Students' Research in Computer Science*, 2025
Publication

94 Submitted to University of Hertfordshire

<1 %

95

ejournal.sisfokomtek.org

Internet Source

<1 %

96

garuda.kemdikbud.go.id

Internet Source

<1 %

97

proceeding.unpkediri.ac.id

Internet Source

<1 %

98

repository.itk.ac.id

Internet Source

<1 %

99

www.biomed.uninet.edu

Internet Source

<1 %

100

Achmad Fauzihan Bagus Sajiwo, Basuki Rahmat, Achmad Junaidi. "KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARAN (ISPU) MENGGUNAKAN ALGORITMA XGBOOST DENGAN TEKNIK IMBALANCED DATA (SMOTE)", Jurnal Informatika dan Teknik Elektro Terapan, 2024

Publication

<1 %

101

Ahmad Fauzi, Nurlaelatul Maulidah, Riki Supriyadi, Hiya Nalatissifa, Sri Diantika. "Prediksi Harga Properti Di Indonesia Menggunakan Algoritma Random Forest", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025

Publication

<1 %

102

brainly.co.id

Internet Source

<1 %

103

generic.ilkom.unsri.ac.id

Internet Source

<1 %

jurnal.univrab.ac.id

104	Internet Source	<1 %
105	media.neliti.com Internet Source	<1 %
106	pbsi-upr.id Internet Source	<1 %
107	repository.its.ac.id Internet Source	<1 %
108	repository.pelitabangsa.ac.id:8080 Internet Source	<1 %
109	repository.unair.ac.id Internet Source	<1 %
110	stt-wastukencana.ac.id Internet Source	<1 %
111	Ahmad Rizki Saputra, Rina - Wati, Tri - Susilowati. "KLASIFIKASI STATUS KELAYAKAN PENERIMA BLT-DD DI PEKON SUKA BANJAR II UJUNG REMBUN MENGGUNAKAN NAÏVE BAYES", <i>Technologia : Jurnal Ilmiah</i> , 2026 Publication	<1 %
112	Erika Tampubolon, Ruhul Amin. "Penerapan Algoritma K Nearest Neighbor Untuk Prediksi Akurasi Penyakit Diabetes", <i>Informatics and Digital Expert (INDEX)</i> , 2024 Publication	<1 %
113	Submitted to Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur II Student Paper	<1 %
114	I Gusti Made Ngurah Ari Bhawanaputra, I Gede Iwan Sudipa, Ni Putu Suci Meinarni, I Gusti Ayu Agung Mas Aristamy et al. "SVM-Based Pediatric Disease Classification Model	<1 %

from the Balinese Lontar Usada Rare
Manuscript", sinkron, 2026

Publication

115 Ihya Bahrul Alam, Hasbi Firmansyah, Wahyu Asriyani. "Prediksi Klasifikasi Diabetes Berbasis Faktor Risiko Kesehatan Menggunakan Gradient Boosted Trees", Jurnal Dinamika Informatika, 2026
Publication

116 Mohammad Firos Anggarda, Nurul Fahmi Arief Hakim. "Pengembangan Sistem Prediksi Waktu Penyiraman Optimal pada Perkebunan: Pendekatan Machine Learning untuk Peningkatan Produktivitas Pertanian", JURNAL BUDIDAYA PERTANIAN, 2023
Publication

117 Sandi Badiwibowo Atim, Muhammad Afdhaluddin. "Komparasi Perceptron dan Regresi Logistik pada Klasifikasi Data Breast Cancer Wisconsin", Jurnal Komputasi, 2026
Publication

118 ajayjaelani45.blogspot.com
Internet Source

119 ejurnal.umri.ac.id
Internet Source

120 eprints.undip.ac.id
Internet Source

121 etheses.uin-malang.ac.id
Internet Source

122 garuda.ristekdikti.go.id
Internet Source

123 jurnalmahasiswa.stiesia.ac.id
Internet Source

124	publikasiilmiah.ums.ac.id Internet Source	<1 %
125	repositori.buddhidharma.ac.id Internet Source	<1 %
126	repository.ipb.ac.id Internet Source	<1 %
127	scholar.archive.org Internet Source	<1 %
128	scholar.unand.ac.id Internet Source	<1 %
129	www.ijert.org Internet Source	<1 %
130	Achmad Hakim Qoirul Haq, Harminto Mulyo, Adi Sucipto. "Optimasi Kinerja Algoritma K-Nearest Neighbor melalui Metode Random Forest untuk Klasifikasi Penyakit Ginjal", Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi), 2026 Publication	<1 %
131	Arkananta Emier, Daffa Arief, Aqila Adam, Jubel Hiero Oktovan Lumban Gaol et al. "Penerapan Aplikasi Kasir Berbasis Web Untuk Mengoptimalkan Layanan Transaksi Pada UMKM", RIGGS: Journal of Artificial Intelligence and Digital Business, 2026 Publication	<1 %
132	Danu Warta, Pramono Pramono, Joni Maulindar. "A Solo Iconic Culinary Recommendation System Using Content-Based Filtering Method", Jurnal Teknologi Terpadu, 2025 Publication	<1 %

133 Eunike Charina Ibrena Tarigan, Angela Wita Simanullang, Daniel S. Simbolon, Sardo Sipayung. "Clustering Penjualan Toko Retail Menggunakan Algoritma K-Means dalam Proses Penambangan Data", remik, 2026
Publication

<1%

134 Miftakhul Rahman, Mantri Kromo Fandith Fili, Wardianto. "Analisis Hasil Rekapitulasi Pilkada Daerah Khusus Jakarta (DKJ) 2024 Menggunakan Metode Support Vector Machine", JIcode: Jurnal Informatika dan Komputer, 2025
Publication

<1%

135 Nanda Fadillah Fattah. "PENERAPAN DATA MINING UNTUK KLASIFIKASI KUALITAS AIR DENGAN ALGORITMA SUPPORT VECTOR MACHINE PADA DINAS LINGKUNGAN HIDUP DAN PERTANAHAN PROVINSI SUMSEL", PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer, 2024
Publication

<1%

136 Solly Aryza. "DESIGN ROBOT OTOMATIS PENYIRAM TANAMAN BERBASISKAN ARTIFICIAL NEURAL NETWORK", INA-Rxiv, 2018
Publication

<1%

137 Submitted to Universitas Kristen Duta Wacana
Student Paper

<1%

138 Zian Asti Dwiyantri, Cahyo Prianto. "Prediksi Cuaca Kota Jakarta Menggunakan Metode Random Forest", Jurnal Tekno Insentif, 2023
Publication

<1%

affiliateprogram1001.blogspot.com

139	Internet Source	<1 %
140	de.scribd.com Internet Source	<1 %
141	eprints.walisongo.ac.id Internet Source	<1 %
142	eprints3.upgris.ac.id Internet Source	<1 %
143	id.123dok.com Internet Source	<1 %
144	journal.ugm.ac.id Internet Source	<1 %
145	jurnal.itg.ac.id Internet Source	<1 %
146	link.springer.com Internet Source	<1 %
147	mypublikasi.com Internet Source	<1 %
148	repo.darmajaya.ac.id Internet Source	<1 %
149	repositori.usu.ac.id Internet Source	<1 %
150	repository.unisi.ac.id Internet Source	<1 %
151	senter.ee.uinsgd.ac.id Internet Source	<1 %
152	text-id.123dok.com Internet Source	<1 %
153	www.coursehero.com Internet Source	<1 %

154

www.ijirset.com

Internet Source

<1 %

155

Gent Imeraj, Erida Muka, Maaruf Ali, Ali Osman Topal, Mevlida Zenuni, Kristjan Marcinaj. "Cognitive Health Reflection to Keto Diet Attitude with Machine Learning Clustering", 2021 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA), 2021

Publication

<1 %

156

Lulu'ul Badriyah, Abdullah Syafei. "Persepsi dan Perilaku Membaca Label Pangan dan Informasi Gizi pada Siswa SMK Wijaya Kusuma", Jurnal Ilmu Kesehatan Masyarakat, 2019

Publication

<1 %

157

Maria Anjelina Domu Tukan, Alfian Nara Weking, Dominikus Boli Watomakin. "Implementasi Algoritma K-Nearest Neighbors untuk Rekomendasi Diet Tinggi Serat dalam Mencegah Penyakit Jantung", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025

Publication

<1 %

158

Rony Heri Irawan, Rizki Dwi Febrian, Risa Helilintar. "Implementation of the K-Nearest Neighbor Algorithm Determination of Fish Feed Dosage", Nusantara of Engineering (NOE), 2025

Publication

<1 %

159

Endi Permana, Joko Susilo. "Optimasi Prediksi Jumlah Kontainer Aktual di Kapal Menggunakan Random Forest dan XGBoost

<1 %

dengan Hyperparameter Tuning", Jurnal Informatika dan Bisnis, 2026

Publication

160 Tursina Tursina, Hafiz Muhardi, Dian Aulia Sari. "Diagnosis Tahapan Pengguna Narkoba Menggunakan Metode K-Nearest Neighbor", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2020 <1%

Publication

161 Yun Muntarti, Alfi Zahra Muharramah, Fiqrah Idhul Dwi, Yabrina Angelika Rantelili, Ilham Julian Effendi. "Klasifikasi citra asap kebakaran hutan untuk monitoring lingkungan berbasis CNN", INFOTECH : Jurnal Informatika & Teknologi, 2025 <1%

Publication

Exclude quotes Off

Exclude matches Off

Exclude bibliography On

Submissions

My Queue 1

Archives

My Assigned

16056 **fikri Firmansyah et al.**
Sistem Klasifikasi Kandungan Gula pada Produk Makanan dan Minuman Kemasan Berbasis Algoritma K-Nearest Neighbor (KNN)

Search

New Submission

Help

Production

Platform &
workflow by
OJS / PKP



JURNAL MINFO POLGAN

JURNAL & PENELITIAN MANAJEMEN INFORMATIKA

Sekretariat: Jl. Veteran No. 193, Manunggal, Medan - Indonesia

<https://jmp.polgan.ac.id> — Email: jmp@polgan.ac.id — Phone: 082363800909

SURAT PENERIMAAN PUBLIKASI ARTIKEL ILMIAH

No.16056/SK-JMP/POLGAN/IV/2026

Kepada Yth,

Fikri Firmansyah, Indah Purnama Sari

Salam hangat,

Dengan hormat, kami sampaikan bahwa setelah melalui proses tinjauan sejawat (peer review), artikel Anda yang berjudul: **"Sistem Klasifikasi Kandungan Gula pada Produk Makanan dan Minuman Kemasan Berbasis Algoritma K-Nearest Neighbor (KNN) "** telah **DITERIMA** dan dinyatakan layak untuk diterbitkan dalam **Jurnal Minfo Polgan** pada **Volume 15 Nomor 1, April 2026**. Naskah daring akan tersedia pada tautan <https://jurnal.polgan.ac.id/index.php/jmp/article/view/16056>

Kami mengucapkan terima kasih atas kontribusi ilmiah Anda kepada jurnal ini. Kami berharap Anda semua sukses dan terus berpartisipasi dalam publikasi artikel-artikel berkualitas di masa mendatang.

Medan, 09 April 2026

Pemimpin Redaksi



Surya Guntur, M.Kom

Note:

Validasi Keaslian LoA melalui link berikut: <https://jurnal.polgan.ac.id/index.php/jmp/loa-validation>

Dengan memeriksa ID Artikel berikut: **16056**

JURNAL MINFO POLGAN (JMP)

e-ISSN : [2797-3298](https://doi.org/10.2797-3298) | p-ISSN : [2089-9424](https://doi.org/10.2089-9424)

Indexasi : [SINTA 3](#), [Garuda](#), [Crossref](#), [Google Scholar](#), [Dimensions](#)



UMSU
Unggul | Cerdas | Terpercaya

Bila menjeleh surat ini agar disebarkan
kepada para mahasiswa

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

<https://fikti.umsu.ac.id>

fikti@umsu.ac.id

[umsumedan](https://www.facebook.com/umsumedan)

[umsumedan](https://www.instagram.com/umsumedan)

[umsumedan](https://www.tiktok.com/@umsumedan)

[umsumedan](https://www.youtube.com/umsumedan)

Berita Acara Pembimbingan Proposal

Nama Mahasiswa : Fikri Firmansyah
Program Studi : Sistem Informasi
NPM : 2209010131
Judul Penelitian : Sistem Klasifikasi Kandungan Gula pada Produk Makanan dan Minuman Kemasan Berbasis Algoritma K-Nearest Neighbor (KNN)
Nama Dosen Pembimbing : Indah Purnama Sari, ST., M.Kom.

Tanggal Bimbingan	Hasil Evaluasi	Paraf Dosen
21-11-2025	Revisi Bab 1 (Latar Belakang)	
24-11-2025	Revisi bab 1 (Rumusan masalah) & sesuaikan jumlah rumusan masalah dengan tujuan penelitian	
25-11-2025	Revisi bab 1 (rumusan masalah dalam bentuk pertanyaan dan jawaban dari rumusan masalah di cantumkan di latar belakang)	
01-12-2025	ACC Bab 1 dan Revisi Bab 2 (Tambahkan penelitian terdahulu minimal 5)	
04-12-2025	Revisi bab 2 (Landasan teori terlalu banyak)	
09-12-2025	ACC Bab 2 dan Revisi Bab 3 (Flowchart)	
12-12-2025	ACC Sempro	





UMSU

Unggul | Cerdas | Terpercaya

Bila menjaahit surat ini agar disetujui
nama dan tanggalnya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

<https://fikti.umsu.ac.id>

fikti@umsu.ac.id

[umsumedan](https://www.facebook.com/umsumedan)

[umsumedan](https://www.instagram.com/umsumedan)

[umsumedan](https://www.linkedin.com/company/umsumedan)

[umsumedan](https://www.youtube.com/channel/UC...)

27-2-2026	Revisi BAB IV dan V	
2-3-2026	Pengujian Sistem	
4-3-2026	ACC Sidang	

Medan, 03 Maret 2026

Diketahui oleh :

Ketua Program Studi

(Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom)

Disetujui oleh :

Dosen Pembimbing

(Indah Purnama Sari, S.T., M.Kom)





UMSU
Unggul | Cerdas | Terpercaya

Bila membuat surat ini agar disebutkan nomor dan tanggalnya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi Unggul Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 174/SK/BAN-PT/Ak.Ppi/PT/11/2024

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax (061) 6625474 - 6631003

<http://www.umsu.ac.id>

ftki@umsu.ac.id

[f umsumedan](#)

[@ umsumedan](#)

[umsumedan](#)

[umsumedan](#)

FORMULIR PERBAIKAN UJIAN SKRIPSI

Pada hari ini, Jumat 13 Maret 2026 telah dilaksanakan Ujian Skripsi bagi mahasiswa Fakultas Ilmu Komputer Dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara. Sbb:

Nama Mahasiswa : Fikri Firmansyah
NPM : 2209010131
Program Studi : Sistem Informasi
Judul Proposal : Sistem Klasifikasi Kandungan Gula Pada Produk Makanan Dan Minuman Kemasan Berbasis Algoritma K-Nearest Neighbor (KNN)

Materi/Point yang Diperbaiki :

		Paraf
Mahardika Prawira Tanjung, S.Kom.,M.Kom.	Ubah Ubah grafik weekly trend menjadi daily trend	<i>AA</i>
Yohanni Syahra, S.Si.,M.Kom.	- PERBAIKI JUDUL PADA COVER - BATA LATAH PADA KATA PEMASUKAN - PERBAIKI YANG MISS PADA PROGRAMNYA.	<i>[Signature]</i>
Indah Purnama Sari, S.T., M.Kom	- Nihil	<i>[Signature]</i>

Berita acara ini **ditandatangani** setelah skripsi diperbaiki sesuai petunjuk/arahan dari Pembimbing dan Penguji/Pembahas.