

**KLASIFIKASI MAKANAN SEHAT BERDASARKAN
KANDUNGAN NUTRISI MENGGUNAKAN RANDOM
FOREST DAN GRADIENT BOOSTING**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
AGUNG HAYUSTAM
22.11.4923

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

**KLASIFIKASI MAKANAN SEHAT BERDASARKAN
KANDUNGAN NUTRISI MENGGUNAKAN RANDOM
FOREST DAN GRADIENT BOOSTING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
AGUNG HAYUSTAM
22.11.4923

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI MAKANAN SEHAT BERDASARKAN
KANDUNGAN NUTRISI MENGGUNAKAN RANDOM
FOREST DAN GRADIENT BOOSTING**

yang disusun dan diajukan oleh

Agung Hayustam

22.11.4923

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Januari 2026

Dosen Pembimbing,



Rumini, S.Kom., M.Kom.

NIK. 190302246

HALAMAN PENGESAHAN
SKRIPSI
KLASIFIKASI MAKANAN SEHAT BERDASARKAN
KANDUNGAN NUTRISI MENGGUNAKAN RANDOM
FOREST DAN GRADIENT BOOSTING

yang disusun dan diajukan oleh

Agung Hayustam

22.11.4923

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Nur Aini, A.Md., S.Kom., M.Kom.
NIK. 190302066

Majid Rahardi, S.Kom., M.Eng
NIK. 190302393

Rumini, S.Kom., M.Kom.
NIK. 190302246

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Agung Hayustam
NIM : 22.11.4923

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Makanan Sehat Berdasarkan Kandungan Nutrisi Menggunakan Algoritma Random Forest dan Gradient Boosting

Dosen Pembimbing : Rumini, S.Kom., M. Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 Januari 2026

Yang Menyatakan,



Agung Hayustam

HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan sebagai ungkapan rasa syukur dan terima kasih yang mendalam kepada berbagai pihak yang telah memberikan dukungan, doa, serta motivasi selama proses penyelesaian studi dan penyusunan skripsi ini.

Terutama kepada Ibu tercinta, yang dengan penuh kesabaran dan ketulusan senantiasa memberikan doa, kasih sayang, serta dukungan yang tiada henti. Segala pengorbanan, nasihat, dan semangat yang diberikan menjadi kekuatan utama bagi penulis dalam menyelesaikan pendidikan dan menghadapi setiap tantangan selama proses perkuliahan hingga penyusunan skripsi ini.

Penulis juga mempersembahkan skripsi ini kepada pasangan penulis, yang selalu memberikan dukungan, perhatian, dan motivasi, serta menjadi tempat berbagi cerita dan penyemangat di saat penulis mengalami kesulitan maupun kelelahan. Dukungan yang diberikan sangat berarti dalam membantu penulis tetap fokus dan konsisten hingga skripsi ini dapat diselesaikan dengan baik.

Selain itu, skripsi ini juga penulis persembahkan untuk diri penulis sendiri, sebagai bentuk penghargaan atas usaha, ketekunan, dan komitmen dalam menjalani seluruh proses akademik. Penyelesaian skripsi ini menjadi bukti bahwa dengan kerja keras, kesabaran, dan keyakinan, setiap tantangan dapat dihadapi dan diselesaikan dengan baik.

Semoga skripsi ini dapat memberikan manfaat dan menjadi langkah awal bagi penulis untuk terus belajar, dan memberikan kontribusi yang positif.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Makanan Sehat Berdasarkan Kandungan Nutrisi Menggunakan Random Forest dan Gradient Boosting” dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada program studi yang penulis tempuh.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa keberhasilan penyelesaian skripsi tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Rumini, S.Kom., M.Kom. selaku Dosen Pembimbing, yang telah meluangkan waktu, memberikan arahan, bimbingan, serta saran yang sangat berarti selama proses penyusunan skripsi ini.
2. Tim Dosen Penguji, yang telah memberikan masukan dan saran yang membangun demi penyempurnaan skripsi ini.
3. Orang tua dan keluarga, yang senantiasa memberikan doa, dukungan moral, dan motivasi sehingga penulis dapat menyelesaikan studi dengan baik.
4. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan, dukungan, dan semangat selama proses penyusunan skripsi ini.

Yogyakarta, 26 Januari 2026

Agung Hayustam

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xviii
ABSTRACT	xix
BAB I PENDAHULUAN	20
1.1 Latar Belakang	20
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	15
2.2.1 Klasifikasi	15
2.2.2 Kandungan Nutrisi Pada Makanan	15
2.2.3 Algoritma Random Forest	16
2.2.4 Algoritma Gradient Boosting	17
2.2.5 Confusion Matrix	18

2.2.6 Synthetic Minority Over-sampling Technique (SMOTE)	19
2.2.7 Feature Importance	20
BAB III METODE PENELITIAN	21
3.1 Objek Penelitian.....	21
3.2 Alur Penelitian	22
3.2.1 Akuisisi Data.....	23
3.2.2 Pra-pemrosesan Data (Preprocessing)	23
3.2.3 Pembagian Data (Split Data).....	24
3.2.4 Normalisasi Data.....	24
3.2.5 Penanganan Ketidakseimbangan Data	24
3.2.6 Pembangunan Model Klasifikasi	25
3.2.7 Evaluasi Model	25
3.2.8 Penyimpanan Model	26
3.2.9 Implementasi Sistem.....	26
3.3 Alat dan Bahan.....	26
3.3.1 Data Penelitian.....	26
3.3.2 Perangkat Keras (Hardware).....	27
3.3.3 Perangkat Lunak (Software)	28
3.4 Contoh Perhitungan Algoritma	28
3.4.1 Contoh Perhitungan Algoritma Random Forest	29
3.4.2 Contoh Perhitungan Algoritma Gradient Boosting.....	30
BAB IV HASIL DAN PEMBAHASAN	33
4.1 Analisis Dataset	33
4.2 Pra-pemrosesan Data.....	34
4.3 Pembagian Data Latih dan Data Uji	38
4.4 Normalisasi Data.....	40
4.5 Hasil Klasifikasi Menggunakan Random Forest	41
4.5.1 Random Forest Tanpa Resampling.....	41
4.5.2 Random Forest dengan Oversampling (SMOTE).....	43
4.5.3 Random Forest dengan Undersampling	45
4.5.4 Analisis Hasil Random Forest	46

4.6 Hasil Klasifikasi Menggunakan Gradient Boosting.....	47
4.6.1 Gradient Boosting Tanpa Resampling	47
4.6.2 Gradient Boosting dengan Oversampling (SMOTE).....	48
4.6.3 Gradient Boosting dengan Undersampling	50
4.7 Analisis Feature Importance	53
4.7.1 Feature Importance pada Random Forest	53
4.7.2 Feature Importance pada Gradient Boosting	54
4.7.3 Pembahasan Feature Importance	55
4.8 Pembahasan Hasil Penelitian	56
BAB V PENUTUP	62
5.1 Kesimpulan.....	62
5.2 Saran	62
REFERENSI	64
LAMPIRAN.....	68

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	11
Tabel 3. 1 Contoh Data Makanan	29
Tabel 3.2 Pohon Keputusan	30
Tabel 3. 3 Contoh Tabel Nutrisi	31
Tabel 4.1 Distribusi Data Makanan Berdasarkan Label Kesehatan	33



DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	22
Gambar 4.1 Gambar Distribusi Data Makanan Sehat dan Tidak Sehat	34
Gambar 4.2 Penyesuaian Nama Atribut ke dalam Bahasa Indonesia	35
Gambar 4.3 Analisis Struktur dan Tipe Data Dataset	36
Gambar 4.4 Analisis Statistik Deskriptif Dataset	37
Gambar 4.5 Hasil Pemeriksaan <i>Missing Value</i> pada Dataset	38
Gambar 4.6 Seleksi Fitur Nutrisi untuk Proses Klasifikasi	38
Gambar 4.7 Hasil Pembagian Data Latih dan Data Uji	39
Gambar 4.8 Normalisasi Data Menggunakan <i>StandardScaler</i>	40
Gambar 4.9 Hasil Evaluasi Model Random Forest Tanpa Resampling	42
Gambar 4.10 Hasil Confusion Matrix Random Forest Tanpa Resampling	42
Gambar 4.11 Hasil Evaluasi Model Random Forest dengan SMOTE	44
Gambar 4.12 Hasil Confusion Matrix Random Forest SMOTE	44
Gambar 4.13 Hasil Evaluasi Model Random Forest dengan Undersampling	45
Gambar 4.14 Hasil Confusion Matrix Random Forest Undersampling	46
Gambar 4.15 Hasil Evaluasi Model Gradient Boosting Tanpa Resampling	47
Gambar 4.16 Hasil Confusion Matrix Gradient Boosting Tanpa Resampling	48
Gambar 4.17 Hasil Evaluasi Model Gradient Boosting dengan SMOTE	49
Gambar 4.18 Hasil Confusion Matrix Gradient Boosting SMOTE	50
Gambar 4.19 Hasil Evaluasi Model Gradient Boosting dengan Undersampling	51
Gambar 4.20 Hasil Confusion Matrix Gradient Boosting Undersampling	52

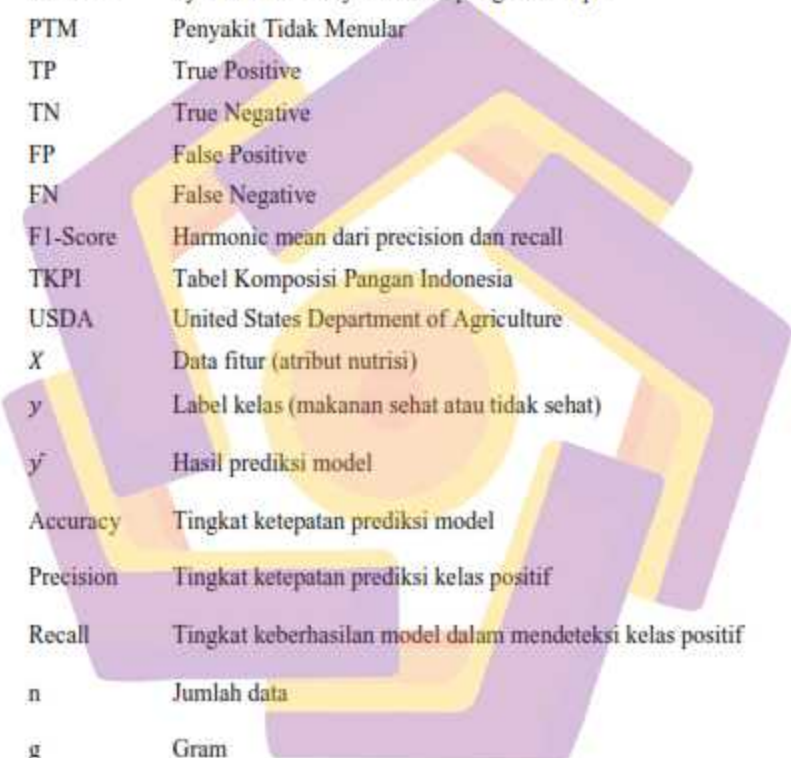
Gambar 4.21 Feature Importance pada Model Random Forest	54
Gambar 4.22 Feature Importance pada Model Gradient Boosting	55
Gambar 4.23 Tampilan Antarmuka Sistem Klasifikasi Makanan Sehat Berbasis Streamlit	57
Gambar 4.24 Contoh Hasil Prediksi Makanan tidak sehat pada Sistem	58
Gambar 4.25 Contoh Hasil Prediksi Makanan tidak sehat pada Sistem	59
Gambar 4.26 Contoh Hasil Prediksi Makanan <i>sehat</i> pada Sistem	60
Gambar 4.27 Contoh Hasil Prediksi Makanan <i>sehat</i> pada Sistem	60



DAFTAR LAMPIRAN

Lampiran 1 Kode Import Library	68
Lampiran 2 Kode Load Dataset	68
Lampiran 3 Kode Rename Kolom	68
Lampiran 4 Kode Distribusi Kelas	69
Lampiran 5 Kode Seleksi Fitur dan Label	69
Lampiran 6 Kode Split Data 80:20	69
Lampiran 7 Kode Fungsi Evaluasi	69
Lampiran 8 Kode Normalisasi Data	70
Lampiran 9 Kode Random Forest Resampling	70
Lampiran 10 Kode Gradient Boosting Resampling	70
Lampiran 11 Kode Distribusi Kelas Setelah SMOTE	70
Lampiran 12 Kode Random Forest SMOTE	71
Lampiran 13 Kode Gradient Boosting SMOTE	71
Lampiran 14 Kode Distribusi Kelas Setelah Undersampling	71
Lampiran 15 Kode Random Forest Undersampling	72
Lampiran 16 Kode Gradient Boosting Undersampling	72
Lampiran 17 Kode Feature Importance	72
Lampiran 18 Kode Simpan Model	73

DAFTAR LAMBANG DAN SINGKATAN



ML	Machine Learning
RF	Random Forest
GB	Gradient Boosting
SMOTE	Synthetic Minority Over-sampling Technique
PTM	Penyakit Tidak Menular
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
F1-Score	Harmonic mean dari precision dan recall
TKPI	Tabel Komposisi Pangan Indonesia
USDA	United States Department of Agriculture
X	Data fitur (atribut nutrisi)
y	Label kelas (makanan sehat atau tidak sehat)
$\hat{y}$	Hasil prediksi model
Accuracy	Tingkat ketepatan prediksi model
Precision	Tingkat ketepatan prediksi kelas positif
Recall	Tingkat keberhasilan model dalam mendeteksi kelas positif
n	Jumlah data
g	Gram
mg	Miligram
kcal	Kilokalori

DAFTAR ISTILAH

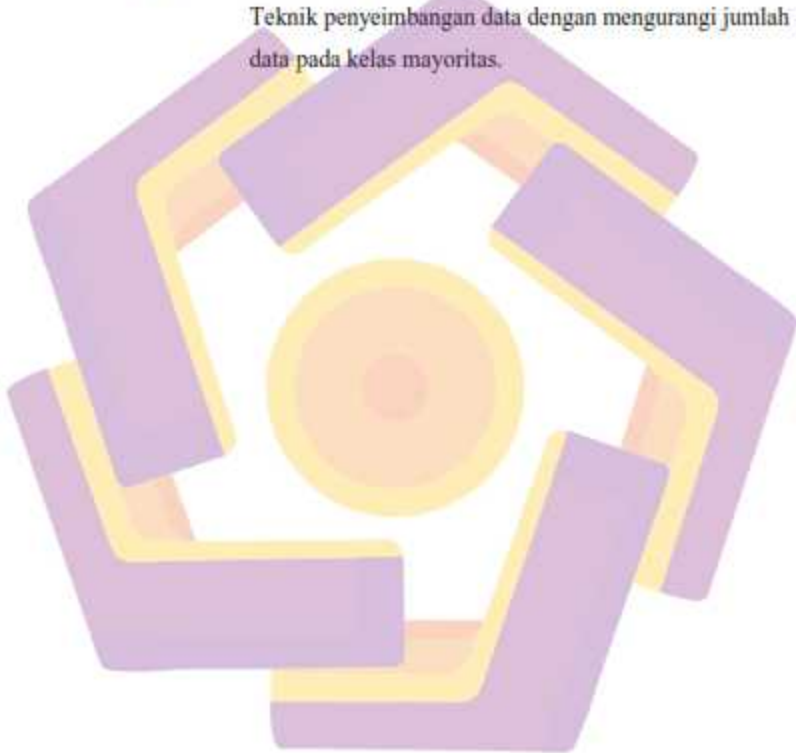
Akurasi	Tingkat ketepatan model dalam memprediksi seluruh data dengan membandingkan jumlah prediksi yang benar terhadap total data.
Algoritma	Tingkat ketepatan model dalam memprediksi seluruh data dengan membandingkan jumlah prediksi yang benar terhadap total data.
Confusion Matrix	Matriks evaluasi yang digunakan untuk menilai kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap label aktual. Confusion matrix menampilkan jumlah prediksi benar dan salah pada setiap kelas.
Dataset	Serangkaian langkah atau aturan logis yang digunakan untuk menyelesaikan suatu masalah atau melakukan perhitungan tertentu.
Data Latih	Data yang digunakan untuk melatih model agar dapat mempelajari pola dari fitur yang tersedia.
Data Uji	Data yang digunakan untuk menguji dan mengevaluasi kinerja model setelah proses pelatihan selesai.
False Positive (FP)	Jumlah data yang diprediksi sebagai kelas positif, tetapi sebenarnya termasuk ke dalam kelas negatif. Pada penelitian ini, FP menunjukkan makanan tidak sehat yang salah diprediksi sebagai makanan sehat.
False Negative (FN)	Jumlah data yang diprediksi sebagai kelas negatif, tetapi sebenarnya termasuk ke dalam kelas positif. Pada penelitian ini, FN menunjukkan makanan sehat yang salah diprediksi sebagai makanan tidak sehat.
Feature Importance	Ukuran tingkat kepentingan suatu fitur dalam mempengaruhi hasil prediksi model.

Fitur (Feature)	Atribut atau variabel input yang digunakan oleh model, seperti kandungan kalori, lemak, gula, protein, dan serat.
Gradient Boosting	Algoritma machine learning berbasis ensemble yang membangun model secara bertahap untuk memperbaiki kesalahan model sebelumnya.
Klasifikasi	Proses pengelompokan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik yang dimiliki.
Machine Learning	Cabang kecerdasan buatan yang memungkinkan sistem belajar dari data tanpa diprogram secara eksplisit.
Model	Representasi matematis yang dibangun dari data untuk melakukan prediksi atau klasifikasi.
Normalisasi Data	Proses penskalaan data agar berada pada rentang nilai tertentu sehingga setiap fitur memiliki pengaruh yang seimbang.
Precesion	Ukuran ketepatan model dalam memprediksi kelas positif dibandingkan dengan seluruh prediksi positif yang dihasilkan.
Recall	Ukuran kemampuan model dalam mengenali seluruh data yang benar-benar termasuk kelas positif.
Random Forest	Algoritma ensemble learning yang membangun banyak pohon keputusan dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi.
Resampling	Teknik penyesuaian distribusi data untuk menangani ketidakseimbangan kelas.
SMOTE	Metode oversampling yang menghasilkan data sintesis pada kelas minoritas untuk menyeimbangkan dataset.
Streamlit	Framework berbasis Python yang digunakan untuk membangun aplikasi web interaktif untuk visualisasi dan implementasi model machine learning.
True Positive (TP)	Jumlah data yang diprediksi sebagai kelas positif dan benar-benar termasuk ke dalam kelas positif. Pada

penelitian ini, TP menunjukkan jumlah makanan sehat yang diprediksi sebagai makanan sehat.

True Negative (TN) Jumlah data yang diprediksi sebagai kelas negatif dan benar-benar termasuk ke dalam kelas negatif. Pada penelitian ini, TN menunjukkan jumlah makanan tidak

Undersampling sehat yang diprediksi sebagai makanan tidak sehat. Teknik penyeimbangan data dengan mengurangi jumlah data pada kelas mayoritas.



INTISARI

Perubahan pola konsumsi masyarakat yang cenderung mengarah pada makanan tinggi kalori, gula, dan lemak meningkatkan risiko penyakit tidak menular. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasikan makanan ke dalam kategori sehat dan tidak sehat berdasarkan kandungan nutrisi menggunakan pendekatan machine learning.

Penelitian ini menggunakan algoritma Random Forest dan Gradient Boosting dengan tahapan pra-pemrosesan data, pembagian data latih dan data uji, normalisasi, serta pengujian model dengan dan tanpa teknik penyeimbangan data. Evaluasi performa model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score.

Hasil penelitian menunjukkan bahwa algoritma Random Forest tanpa resampling memberikan performa terbaik dengan akurasi 99,75% dan F1-score 98,63%, lebih unggul dibandingkan Gradient Boosting dan skenario resampling lainnya. Analisis feature importance menunjukkan bahwa kandungan gula, lemak, dan kalori merupakan faktor nutrisi yang paling berpengaruh dalam klasifikasi kesehatan makanan. Model terbaik kemudian diimplementasikan ke dalam sistem berbasis web menggunakan framework Streamlit untuk membantu pengguna dalam menginterpretasikan tingkat kesehatan makanan berdasarkan data nutrisi yang tersedia.

Kata kunci: klasifikasi makanan, machine learning, Random Forest, Gradient Boosting, nutrisi makanan.

ABSTRACT

Changes in dietary patterns characterized by high consumption of calorie-, sugar-, and fat-rich foods increase the risk of non-communicable diseases. Therefore, this study aims to classify foods into healthy and unhealthy categories based on nutritional content using a machine learning approach.

This research applies Random Forest and Gradient Boosting algorithms with data preprocessing, training-testing split, normalization, and model evaluation with and without data balancing techniques. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics.

The results show that the Random Forest model without resampling achieves the best performance, with an accuracy of 99.75% and an F1-score of 98.63%, outperforming Gradient Boosting and resampling scenarios. Feature importance analysis indicates that sugar, fat, and calorie content are the most influential nutritional factors in food health classification. The best-performing model is implemented in a web-based system using the Streamlit framework to assist users in interpreting food healthiness based on available nutritional data.

Keyword: food classification, machine learning, Random Forest, Gradient Boosting, nutritional data.

