

KLASIFIKASI KELAYAKAN PRODUK PANGAN UMKM MENGGUNAKAN MACHINE LEARNING UNTUK Mendukung PROGRAM MAKAN BERGIZI GRATIS

Ainun Hidayah^{1*}, Ismail², Ghina Raudhatul Janna³, Ikra Juwita⁴, Nurkhalizah⁵

^{1,2,3,4,5} Program Studi Sistem Informasi, Universitas Lamappapoleonro

^{1*}hiddciliknun@gmail.com, ²ismail@unipol.ac.id, ³ghina4raudhatuljanna@gmail.com,

⁴iita241123@gmail.com, ⁵433017nurkhalizahh@gmail.com,

DOI: 10.46880/mtk.v12i2.6388

ABSTRACT

The Free Nutritious Meal Program requires an objective and standardized approach to evaluate the eligibility of food products supplied by Micro, Small, and Medium Enterprises (MSMEs). This study aims to develop a machine learning-based classification model using Random Forest to support the initial screening of MSME food product eligibility. A real-world dataset containing 120 MSME food product records was utilized, consisting of nutritional, economic, legality, certification, and packaging quality attributes. The data were preprocessed and divided into training and testing sets using an 80:20 ratio. The Random Forest model achieved the best classification performance, obtaining an accuracy of 96.00%, precision of 93.75%, recall of 100.00%, and F1-score of 96.77%. Feature analysis showed that PIRT license, packaging hygiene, and halal certification were the most influential factors in determining product eligibility. The proposed model provides practical support for improving the objectivity, efficiency, and documentation of MSME food product screening in the implementation of the Free Nutritious Meal Program. This study is limited by the use of a small-scale prototype dataset; therefore, future research should involve larger real-world datasets and further validation in operational environments.

Keywords: MSME food classification, Random Forest, food product eligibility, PIRT certification, Free Nutritious Meal Program

I. PENDAHULUAN

Program Makan Bergizi Gratis (MBG) merupakan salah satu program prioritas nasional yang bertujuan meningkatkan kualitas gizi masyarakat melalui penyediaan makanan yang aman, bergizi, dan berkelanjutan [1]. Keberhasilan implementasi program tersebut tidak hanya bergantung pada ketersediaan bahan pangan, tetapi juga pada kemampuan sistem rantai pasok dalam memastikan bahwa produk yang digunakan memenuhi standar keamanan, kualitas, dan kelayakan konsumsi. Dalam konteks ini, keterlibatan Usaha Mikro, Kecil, dan Menengah (UMKM) pangan menjadi aspek penting karena UMKM memiliki potensi besar dalam menyediakan produk pangan lokal sekaligus memperkuat keberlanjutan ekosistem penyediaan makanan bergizi [2].

UMKM pangan berperan strategis dalam mendukung perekonomian nasional melalui produksi berbagai jenis makanan olahan, penciptaan lapangan kerja, dan peningkatan nilai tambah produk lokal. Namun, karakteristik UMKM pangan di Indonesia masih sangat beragam, terutama dalam aspek legalitas usaha, penerapan standar keamanan pangan, informasi produk, dan pengendalian kualitas. Salah satu indikator penting dalam memastikan keamanan produk pangan rumah tangga adalah kepemilikan Sertifikat Produksi Pangan Industri Rumah Tangga (SPP-IRT), yang menunjukkan bahwa produk telah memenuhi persyaratan dasar keamanan pangan sesuai ketentuan Badan Pengawas Obat dan Makanan Republik Indonesia [3]. Selain itu, sertifikasi halal menjadi salah satu aspek penting dalam menjamin kepastian kehalalan produk pangan sesuai dengan ketentuan penyelenggaraan Jaminan Produk Halal yang diatur dalam Peraturan Pemerintah Republik Indonesia Nomor 42 Tahun 2024 [4]. Oleh karena itu, aspek legalitas,

kandungan nutrisi, harga, serta kualitas kemasan menjadi beberapa faktor yang perlu dipertimbangkan dalam proses seleksi produk pangan UMKM.

Dalam implementasinya, proses penilaian kelayakan produk pangan UMKM untuk mendukung program berskala besar seperti MBG masih menghadapi tantangan apabila dilakukan secara manual. Pemeriksaan manual yang melibatkan berbagai atribut produk berpotensi menghasilkan keputusan yang subjektif karena bergantung pada pengalaman dan interpretasi masing-masing evaluator. Selain membutuhkan waktu yang relatif lama, proses manual juga memiliki keterbatasan dalam menjaga konsistensi keputusan ketika jumlah produk yang harus dievaluasi meningkat. Kondisi tersebut menunjukkan kebutuhan terhadap sistem otomatis berbasis data yang mampu membantu proses penyaringan awal secara lebih objektif, cepat, transparan, dan terdokumentasi.

Perkembangan teknologi kecerdasan buatan, khususnya machine learning, memberikan peluang dalam membangun sistem pendukung keputusan berbasis data untuk membantu proses evaluasi produk pangan. Machine learning mampu mengenali pola dari berbagai atribut dan menghasilkan model klasifikasi berdasarkan data historis yang tersedia [5]. Dalam bidang pangan, beberapa penelitian menunjukkan bahwa pendekatan machine learning telah digunakan untuk meningkatkan proses inspeksi dan pengendalian kualitas produk. Penelitian Bastian et al. [6] menerapkan Convolutional Neural Network (CNN) untuk melakukan klasifikasi kualitas beras sebagai pendekatan dalam meningkatkan keamanan pangan. Penelitian lain juga menunjukkan bahwa algoritma berbasis machine learning mampu digunakan untuk mengidentifikasi karakteristik kualitas pangan melalui analisis data dan pola tertentu [7]. Selain itu, metode

ensemble learning seperti Random Forest memiliki kemampuan dalam meningkatkan stabilitas model klasifikasi melalui kombinasi beberapa decision tree serta menyediakan informasi mengenai tingkat kepentingan setiap fitur [8].

Pendekatan metode klasifikasi berbasis machine learning juga telah banyak digunakan dalam berbagai permasalahan prediksi dan pengambilan keputusan karena mampu menghasilkan pola keputusan berdasarkan karakteristik data yang kompleks [9], [10]. Analisis komparatif beberapa algoritma klasifikasi menunjukkan bahwa Random Forest, Decision Tree, Naive Bayes, K-Nearest Neighbor, dan Support Vector Machine memiliki karakteristik performa yang berbeda sehingga pemilihan algoritma menjadi faktor penting dalam membangun model klasifikasi yang optimal [11].

Meskipun berbagai penelitian telah menerapkan machine learning dalam bidang pangan, sebagian besar penelitian masih berfokus pada klasifikasi kualitas bahan pangan tertentu seperti beras atau identifikasi karakteristik produk berdasarkan citra digital. Penelitian tersebut belum secara khusus membahas mekanisme klasifikasi kelayakan produk pangan UMKM yang mempertimbangkan kombinasi aspek nutrisi, harga, legalitas PIRT, sertifikasi halal, dan kebersihan kemasan sebagai indikator kelayakan. Selain itu, belum banyak penelitian yang mengembangkan model klasifikasi berbasis machine learning untuk mendukung proses seleksi awal pemasok pangan dalam konteks Program Makan Bergizi Gratis. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu membantu proses evaluasi produk secara lebih objektif, konsisten, dan terdokumentasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan model klasifikasi kelayakan produk pangan UMKM menggunakan pendekatan machine learning sebagai mekanisme pendukung seleksi awal dalam Program Makan Bergizi Gratis. Penelitian ini mengevaluasi dan membandingkan beberapa algoritma klasifikasi, yaitu Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM), untuk memperoleh model dengan performa terbaik. Selain itu, penelitian ini juga menganalisis tingkat kepentingan fitur (feature importance) untuk mengidentifikasi atribut utama yang memengaruhi keputusan klasifikasi. Kontribusi penelitian ini adalah menyediakan pendekatan berbasis data yang dapat mendukung proses evaluasi produk pangan UMKM secara lebih objektif, konsisten, transparan, dan terdokumentasi.

II. KAJIAN TERKAIT

Evaluasi kelayakan produk pangan merupakan permasalahan yang melibatkan berbagai aspek, seperti kualitas nutrisi, keamanan pangan, legalitas produk, serta karakteristik produksi. Pendekatan berbasis data dan machine learning telah banyak diterapkan dalam berbagai bidang untuk membantu proses klasifikasi dan pengambilan keputusan secara objektif [5]. Penelitian terkait klasifikasi kualitas pangan menunjukkan bahwa algoritma machine learning mampu mengenali pola berdasarkan karakteristik produk. Bastian et al. [6]

menerapkan Convolutional Neural Network (CNN) untuk melakukan klasifikasi kualitas beras sebagai pendekatan dalam meningkatkan keamanan pangan. Hasil penelitian tersebut menunjukkan bahwa metode pembelajaran mesin memiliki kemampuan dalam melakukan identifikasi kualitas produk pangan berdasarkan pola data.

Selain itu, penelitian mengenai sistem pendukung keputusan dalam pemilihan pemasok juga telah memanfaatkan kombinasi metode machine learning dan metode pengambilan keputusan multikriteria [7]. Beberapa penelitian menggunakan pendekatan ensemble learning seperti Random Forest untuk meningkatkan akurasi prediksi karena mampu menggabungkan beberapa model pembelajaran sehingga lebih stabil dalam menghadapi variasi data [8].

Dalam konteks pangan, penelitian sebelumnya lebih banyak berfokus pada klasifikasi kualitas komoditas tertentu, pengenalan berbasis citra, atau pengembangan model prediksi umum [6], [7], [10]. Namun, penelitian yang secara khusus membangun model klasifikasi kelayakan produk pangan UMKM dengan mempertimbangkan aspek nutrisi, legalitas, keamanan kemasan, dan harga untuk mendukung Program Makan Bergizi Gratis masih terbatas.

Oleh karena itu, penelitian ini mengembangkan model klasifikasi berbasis machine learning dengan mempertimbangkan aspek gizi, legalitas, keamanan kemasan, dan harga produk sebagai atribut penilaian kelayakan. Model yang dikembangkan diharapkan dapat menjadi pendukung keputusan dalam proses seleksi awal produk pangan UMKM secara objektif dan berbasis data.

A. State of the Art dan Posisi Penelitian

Tabel 1. Deskripsi Atribut Dataset Produk Pangan UMKM

No	Atribut	Tipe Data	Rentang/Kelas	Keterangan
1	Protein	Numerik	gram	Kandungan protein produk
2	Karbohidrat	Numerik	gram	Kandungan karbohidrat
3	Lemak	Numerik	gram	Kandungan lemak
4	Vitamin	Numerik	nilai numerik	Kandungan vitamin
5	Kalori	Numerik	kcal	Nilai energi produk
6	Harga	Numerik	Rupiah	Harga per porsi
7	Legalitas PIRT	Kategorikal	Ada/Tidak Ada	Legalitas produk
8	Label Halal	Kategorikal	Ada/Tidak Ada	Status halal produk
9	Kebersihan Kemasan	Kategorikal	Kategori kondisi kemasan	Keamanan kemasan
10	Kategori Aktual	Target	Layak/Tidak Layak	Kelas keluaran model

Berdasarkan Tabel 1, dataset penelitian terdiri atas sembilan atribut prediktor dan satu atribut target yang

digunakan dalam proses klasifikasi kelayakan produk pangan UMKM. Atribut prediktor mencakup aspek kandungan gizi, nilai ekonomi, legalitas, serta kualitas keamanan produk yang terdiri atas protein, karbohidrat, lemak, vitamin, kalori, harga, legalitas PIRT, label halal, dan kebersihan kemasan. Seluruh atribut tersebut digunakan untuk merepresentasikan karakteristik produk pangan UMKM dalam proses pembelajaran model machine learning.

Atribut target yang digunakan adalah kategori aktual produk dengan dua kelas, yaitu Layak dan Tidak Layak. Kombinasi atribut nutrisi, legalitas, keamanan kemasan, dan harga tersebut menjadi dasar dalam membangun model klasifikasi untuk membantu proses seleksi awal produk pangan UMKM dalam mendukung Program Makan Bergizi Gratis (MBG).

III. METODE

Kerangka Dasar Penelitian

Penelitian ini termasuk dalam kategori riset terapan yang menggunakan pendekatan kuantitatif melalui eksperimen komputasional dalam lingkup data mining dan pembelajaran mesin [12]. Objek penelitian adalah produk pangan UMKM yang dinilai berdasarkan aspek legalitas, kandungan gizi, kualitas kemasan, dan harga produk. Penelitian menggunakan dataset terkontrol sebanyak 120 record produk pangan UMKM yang disusun untuk merepresentasikan kondisi produk yang layak dan tidak layak sebagai calon pendukung rantai pasok Program Makan Bergizi Gratis (MBG).

Dataset penelitian mencakup sembilan atribut prediktor dan satu atribut target. Atribut prediktor terdiri dari Legalitas PIRT, Sertifikasi Halal, Kebersihan Kemasan, Kandungan Protein, Kandungan Karbohidrat, Kandungan Lemak, Kandungan Vitamin, Kalori, dan Harga per Porsi. Variabel target adalah Kelayakan Produk, yang dikategorikan menjadi dua kelas, yaitu Layak (nilai 1) dan Tidak Layak (nilai 0). Variabel independen dalam penelitian ini adalah seluruh atribut prediktor, sedangkan variabel dependen adalah status kelayakan produk. Karakteristik setiap atribut penelitian beserta tipe data dan pengkodeannya disajikan pada Tabel 2.

Tabel 2. Statistik Deskriptif Atribut Numerik Dataset

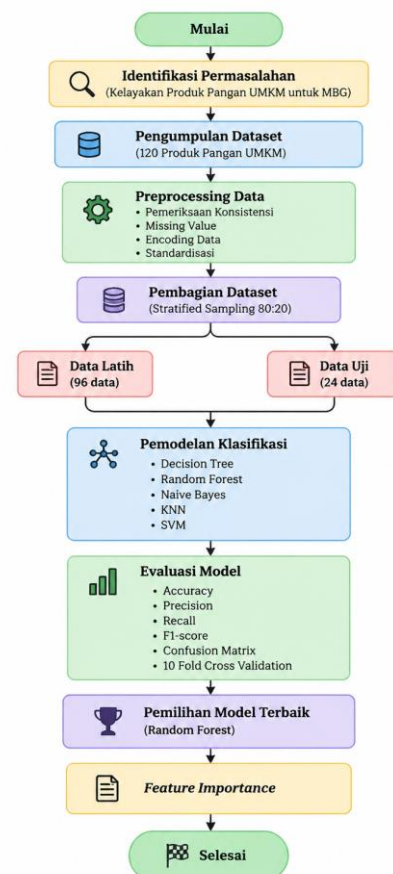
Atribut	Mean	Std. Deviasi	Minimum	Maksimum
Protein (g)	11,96	8,83	0,51	30,00
Karbohidrat (g)	35,55	11,86	8,22	59,12
Lemak (g)	16,01	5,84	5,63	29,89
Vitamin	0,13	0,09	0,00	0,30
Kalori (kkal)	418,63	40,49	308	500
Harga (Rp)	8.016,67	4.297,22	2.000	18.000

Berdasarkan Tabel 2, atribut numerik pada dataset menunjukkan variasi karakteristik produk pangan UMKM yang digunakan dalam penelitian. Kandungan protein memiliki nilai rata-rata sebesar 11,96 gram dengan standar deviasi 8,83, sedangkan kandungan karbohidrat dan lemak masing-masing memiliki rata-rata sebesar 35,55 gram dan 16,01 gram. Nilai rata-rata kalori produk sebesar 418,63 kkal dengan rentang nilai antara 308 hingga 500 kkal.

Selain aspek gizi, atribut harga menunjukkan rata-rata sebesar Rp8.016,67 per porsi dengan variasi nilai antara Rp2.000 hingga Rp18.000. Variasi tersebut menunjukkan bahwa dataset memiliki karakteristik produk yang beragam sehingga sesuai digunakan untuk menguji kemampuan algoritma machine learning dalam membedakan kategori produk layak dan tidak layak.

Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis untuk menghasilkan model klasifikasi yang mampu memprediksi kelayakan produk pangan UMKM berdasarkan atribut yang telah ditentukan. Penelitian ini mengadopsi tahapan *data mining* yang meliputi proses identifikasi permasalahan, pembacaan dataset, *preprocessing* data, pembagian dataset, pembangunan model klasifikasi, serta evaluasi model. Setiap tahapan memiliki peran yang saling berkaitan sehingga menghasilkan model yang mampu melakukan klasifikasi secara objektif dan akurat [13]. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Metodologi Penelitian Klasifikasi Kelayakan Produk Pangan UMKM Berbasis Machine Learning

Tahapan Data Mining

Penelitian ini menerapkan pendekatan data mining sebagai metode utama dalam membangun model klasifikasi kelayakan produk pangan UMKM. Tahapan data mining dimulai dari identifikasi permasalahan yang berkaitan dengan proses seleksi produk pangan UMKM untuk mendukung Program Makan Bergizi Gratis (MBG). Permasalahan tersebut berkaitan dengan perlunya proses seleksi yang lebih cepat, objektif, dan

konsisten dalam menentukan produk yang layak maupun tidak layak berdasarkan berbagai indikator penilaian [14].

Setelah permasalahan ditentukan, dilakukan penyusunan atribut penelitian berdasarkan aspek legalitas, kandungan gizi, kualitas kemasan, dan harga produk. Atribut-atribut tersebut kemudian digunakan sebagai dasar dalam penyusunan dataset penelitian. Selanjutnya dilakukan proses preprocessing untuk meningkatkan kualitas data sebelum digunakan dalam pembentukan model klasifikasi. Dataset yang telah diproses kemudian dibagi menjadi data latih dan data uji, dilanjutkan dengan pembangunan model menggunakan beberapa algoritma machine learning. Tahap terakhir adalah evaluasi performa model menggunakan beberapa metrik evaluasi sehingga dapat diperoleh algoritma dengan performa terbaik.

Import Dataset

Tahap berikutnya adalah proses pembacaan (*import*) dataset ke dalam lingkungan pemrograman Python menggunakan pustaka Pandas. Dataset yang digunakan merupakan dataset yang terdiri atas 120 record produk pangan UMKM. Dataset penelitian ini disusun secara terkontrol berdasarkan karakteristik umum produk pangan UMKM dan indikator kelayakan yang mengacu pada aspek keamanan pangan, nilai gizi, legalitas, serta kualitas produk. Dataset yang digunakan merupakan dataset nyata yang terdiri atas 120 record produk pangan UMKM. Dataset mencakup karakteristik produk berdasarkan aspek keamanan pangan, nilai gizi, legalitas, sertifikasi, kualitas kemasan, dan harga.

Penyusunan dataset dilakukan dengan menetapkan variasi nilai atribut produk yang merepresentasikan kondisi produk Layak dan Tidak Layak berdasarkan aspek legalitas, sertifikasi halal, kandungan nutrisi, kualitas kemasan, dan kesesuaian harga. Proses pelabelan dilakukan berdasarkan kombinasi kriteria tersebut sehingga menghasilkan dua kategori target, yaitu Layak (*eligible*, nilai 1) dan Tidak Layak (*ineligible*, nilai 0). Dataset penelitian terdiri atas 120 produk pangan UMKM, dengan 80 produk kategori Layak dan 40 produk kategori Tidak Layak. Perbedaan jumlah antar kelas menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) dengan kategori Layak sebagai kelas mayoritas.

Dataset terdiri atas sembilan atribut prediktor, yaitu kandungan protein, karbohidrat, lemak, vitamin, kalori, harga per porsi, legalitas PIRT, sertifikasi halal, dan kebersihan kemasan. Variabel target yang digunakan adalah kategori kelayakan produk yang terdiri atas dua kelas, yaitu Layak (*eligible*, nilai 1) dan Tidak Layak (*ineligible*, nilai 0). Setelah proses import dilakukan, pemeriksaan awal dilakukan terhadap struktur data, tipe atribut, jumlah record, serta distribusi kelas sebelum memasuki tahap preprocessing.

Preprocessing Data

Tahap preprocessing bertujuan untuk meningkatkan kualitas dataset sebelum digunakan dalam proses pembentukan model klasifikasi [15]. Pada penelitian ini dilakukan beberapa tahapan preprocessing, yaitu pemeriksaan konsistensi data, pemeriksaan missing value, encoding terhadap variabel kategorikal, dan standardisasi atribut numerik.

Pemeriksaan konsistensi dilakukan untuk memastikan bahwa seluruh atribut memiliki tipe data yang sesuai dengan karakteristik masing-masing variabel. Selanjutnya dilakukan pemeriksaan missing value untuk mengetahui apakah terdapat data yang kosong pada setiap atribut. Hasil pemeriksaan menunjukkan bahwa seluruh data telah lengkap sehingga tidak diperlukan proses penanganan data yang hilang.

Selain itu, dilakukan pemeriksaan distribusi kelas untuk mengetahui potensi ketidakseimbangan data (*class imbalance*). Dataset penelitian terdiri atas 80 produk kategori Layak dan 40 produk kategori Tidak Layak. Kondisi tersebut menunjukkan bahwa dataset memiliki distribusi kelas yang tidak seimbang (*class imbalance*), sehingga diperlukan perhatian dalam proses evaluasi model agar performa klasifikasi pada masing-masing kelas tetap dapat dianalisis secara objektif.

Tahap berikutnya adalah proses encoding terhadap atribut kategorikal, yaitu egalitas PIRT, sertifikasi halal, dan kebersihan kemasan. Proses encoding dilakukan untuk mengubah data kategorikal menjadi representasi numerik agar dapat diproses oleh algoritma machine learning.

Selanjutnya dilakukan standardisasi terhadap atribut numerik menggunakan metode *StandardScaler*. Standardisasi diterapkan terutama pada algoritma *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) karena kedua algoritma tersebut sensitif terhadap perbedaan skala atribut dalam proses perhitungan jarak dan pembentukan batas keputusan (*hyperplane*) [16]. Ringkasan hasil preprocessing disajikan pada Tabel 3.

Tabel 3. Kesimpulan Hasil Preprocessing Data

No	Tahap Preprocessing	Proses	Hasil
1	Konsistensi Data	Pemeriksaan tipe dan struktur atribut	Data konsisten
2	Missing Value	Pemeriksaan data kosong pada seluruh atribut	Tidak ditemukan missing value
3	Encoding Data	Transformasi atribut kategorikal (PIRT, halal, kebersihan kemasan, target)	Berhasil dilakukan
4	Standardisasi Data	<i>StandardScaler</i> pada atribut numerik untuk KNN dan SVM	Berhasil dilakukan

Berdasarkan Tabel 3, tahap preprocessing dilakukan untuk meningkatkan kualitas dataset sebelum proses pemodelan machine learning. Tahapan preprocessing meliputi pemeriksaan konsistensi data, pengecekan missing value, encoding atribut kategorikal, dan standardisasi atribut numerik.

Pembagian Dataset

Dataset yang telah melalui proses preprocessing kemudian dibagi menjadi data latih dan data uji menggunakan metode *stratified sampling*. Metode ini dipilih untuk mempertahankan proporsi masing-masing

kelas sehingga distribusi produk layak dan tidak layak tetap seimbang pada data latih maupun data uji.

Perbandingan pembagian data yang digunakan adalah 80% data latih dan 20% data uji. Dari total 120 record diperoleh sebanyak 96 record sebagai data latih dan 24 record sebagai data uji. Data latih digunakan untuk membangun model klasifikasi melalui proses pembelajaran algoritma, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data yang belum pernah digunakan selama proses pelatihan.

Pembagian dataset dengan mempertahankan proporsi kelas ini dilakukan untuk mengurangi potensi bias akibat perbedaan distribusi data dan menghasilkan evaluasi performa model yang lebih objektif. Dengan demikian, hasil pengujian dapat menggambarkan kemampuan masing-masing algoritma dalam melakukan klasifikasi kelayakan produk pangan UMKM.

Modeling Klasifikasi

Tahap pemodelan dilakukan menggunakan lima algoritma machine learning, yaitu Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM). Kelima algoritma tersebut dipilih karena mewakili pendekatan klasifikasi yang berbeda sehingga dapat memberikan gambaran mengenai performa masing-masing metode dalam mengklasifikasikan kelayakan produk pangan UMKM [16].

Decision Tree digunakan karena mampu menghasilkan aturan keputusan yang mudah diinterpretasikan melalui struktur pohon keputusan [17]. Random Forest merupakan metode berbasis ensemble learning yang menggabungkan beberapa decision tree sehingga mampu meningkatkan stabilitas model dan mengurangi risiko overfitting [10]. Naive Bayes menggunakan pendekatan probabilistik berdasarkan Teorema Bayes dengan asumsi independensi antaratribut. Sementara itu, KNN melakukan klasifikasi berdasarkan tingkat kedekatan antar data, sedangkan SVM membangun batas pemisah (hyperplane) optimal untuk membedakan kelas produk layak dan tidak layak [18], [19].

Seluruh model dibangun menggunakan dataset hasil preprocessing yang terdiri atas 120 record produk pangan UMKM. Data tersebut dibagi menjadi 96 data latih (80%) dan 24 data uji (20%) menggunakan metode stratified sampling untuk mempertahankan proporsi kelas Layak dan Tidak Layak pada masing-masing subset data.

Konfigurasi Model dan Hyperparameter

Pada penelitian ini, setiap algoritma dikonfigurasi menggunakan parameter utama yang berpengaruh terhadap proses klasifikasi.

Decision Tree menggunakan parameter `criterion='gini'` dan `max_depth` untuk mengatur kompleksitas struktur pohon keputusan. Random Forest menggunakan parameter `n_estimators`, `criterion`, dan `random_state` untuk meningkatkan stabilitas model melalui kombinasi beberapa pohon keputusan.

Naive Bayes menggunakan metode Gaussian Naive Bayes karena atribut penelitian terdiri atas variabel numerik yang memiliki distribusi kontinu. Algoritma ini

menggunakan parameter bawaan (default parameter) karena pendekatan probabilistiknya tidak memerlukan banyak konfigurasi.

K-Nearest Neighbor menggunakan parameter `n_neighbors` dan metode pengukuran jarak (distance metric) untuk menentukan kedekatan antar data. Support Vector Machine menggunakan kernel RBF dengan parameter `C` dan `gamma` untuk menentukan batas pemisah antar kelas.

Pemilihan parameter model dilakukan menggunakan konfigurasi algoritma yang telah ditentukan dan dievaluasi menggunakan 10-fold cross validation untuk memperoleh gambaran kestabilan performa model serta mengurangi kemungkinan bias akibat pembagian data.

Evaluasi Model

Tahap evaluasi dilakukan untuk mengetahui kemampuan masing-masing algoritma dalam melakukan klasifikasi kelayakan produk pangan UMKM. Evaluasi menggunakan data uji sebanyak 24 record yang tidak digunakan selama proses pelatihan sehingga mampu memberikan gambaran kemampuan model terhadap data baru.

Selain pengujian menggunakan data uji, penelitian ini juga menerapkan 10-fold cross validation untuk mengevaluasi konsistensi performa model pada beberapa kombinasi data pelatihan dan validasi. Pendekatan ini digunakan karena jumlah dataset penelitian relatif terbatas sehingga diperlukan metode evaluasi tambahan untuk meningkatkan reliabilitas hasil pengujian.

Selain pengujian menggunakan data uji, penelitian ini juga menerapkan 10-fold cross validation untuk mengevaluasi konsistensi performa model pada beberapa kombinasi data pelatihan dan validasi. Pendekatan ini digunakan karena jumlah dataset penelitian relatif terbatas sehingga diperlukan metode evaluasi tambahan untuk meningkatkan reliabilitas hasil pengujian.

Performa setiap algoritma dievaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk mengetahui tingkat ketepatan prediksi keseluruhan model. Precision menunjukkan tingkat ketepatan model dalam mengidentifikasi produk yang diprediksi layak. Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh produk yang benar-benar memenuhi kategori layak. Sementara itu, F1-score digunakan sebagai ukuran keseimbangan antara precision dan recall [18], [20].

Dalam konteks seleksi produk pangan UMKM untuk mendukung Program Makan Bergizi Gratis (MBG), nilai recall pada kelas Layak menjadi perhatian penting karena kesalahan mengenali produk yang sebenarnya layak (false negative) dapat menyebabkan produk potensial tidak lolos tahap seleksi. Selain itu, precision juga diperhatikan untuk mengurangi kemungkinan produk yang belum memenuhi kriteria dikategorikan sebagai produk layak (false positive).

Confusion matrix digunakan untuk menganalisis hasil klasifikasi berdasarkan empat kondisi prediksi, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Melalui

confusion matrix, dapat diketahui jenis kesalahan yang dilakukan model, terutama kesalahan false negative pada kelas Layak yang berpengaruh terhadap nilai recall.

Perhitungan metrik evaluasi dilakukan menggunakan persamaan berikut.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

$$Precision = TP/(TP + FP) \quad (2)$$

$$Recall = TP/(TP + FN) \quad (3)$$

$$F1-score = 2 \times Precision \times Recall / (Precision + Recall) \quad (4)$$

Berdasarkan hasil evaluasi, algoritma dengan performa terbaik dipilih berdasarkan kombinasi nilai accuracy, precision, recall, F1-score, serta kemampuan interpretasi model melalui analisis feature importance.

IV. HASIL DAN PEMBAHASAN

Deskripsi Dataset dan Distribusi Kelas

Bagian ini menyajikan hasil pengolahan dataset, distribusi kelas, perbandingan performa algoritma, confusion matrix, serta interpretasi atribut yang paling berpengaruh terhadap klasifikasi kelayakan produk pangan UMKM. Dataset penelitian terdiri atas 120 record produk pangan UMKM dengan dua kelas keluaran, yaitu Layak dan Tidak Layak. Sebelum proses pemodelan dilakukan, dataset dibagi menjadi data latih dan data uji menggunakan metode stratified sampling dengan perbandingan 80:20. Metode ini digunakan untuk mempertahankan proporsi masing-masing kelas pada data latih maupun data uji sehingga proses evaluasi model dapat dilakukan secara lebih objektif.

Berdasarkan hasil pembagian dataset, diperoleh sebanyak 96 record data latih (80%) dan 24 record data uji (20%). Pembagian tersebut mempertahankan distribusi kelas berdasarkan proporsi awal dataset, yaitu kategori Layak sebanyak 80 produk dan kategori Tidak Layak sebanyak 40 produk. Selanjutnya, pemeriksaan distribusi kelas dilakukan untuk mengetahui karakteristik data sebelum memasuki tahap pemodelan klasifikasi. Distribusi kelas target pada dataset penelitian disajikan pada Tabel 4.

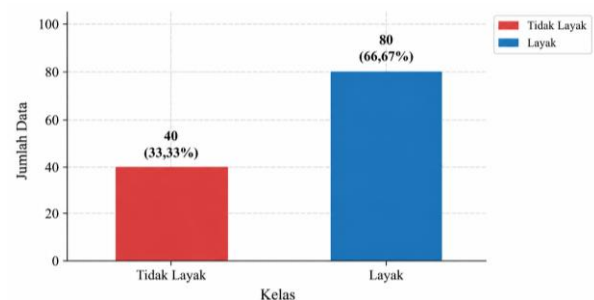
Tabel 4. Distribusi Kelas Kelayakan Produk pada Dataset

Kelas	Jumlah Data	Persentase (%)
Tidak Layak	40	33,33
Layak	80	66,67
Total	120	100,00

Berdasarkan Tabel 4, dataset penelitian terdiri atas 80 produk kategori Layak (66,67%) dan 40 produk kategori Tidak Layak (33,33%). Distribusi tersebut menunjukkan bahwa dataset memiliki perbedaan jumlah antar kelas, dengan kategori Layak sebagai kelas mayoritas. Kondisi ini menunjukkan adanya ketidakseimbangan kelas (class imbalance) dengan tingkat perbedaan proporsi yang masih dapat dianalisis

menggunakan evaluasi performa yang mempertimbangkan masing-masing kelas.

Penelitian ini tidak menerapkan teknik penyeimbangan data seperti oversampling maupun undersampling, karena jumlah data pada masing-masing kelas masih dapat digunakan untuk proses eksperimen klasifikasi. Untuk memastikan kemampuan model dalam mengenali kedua kategori, evaluasi dilakukan menggunakan beberapa metrik, yaitu accuracy, precision, recall, F1-score, serta confusion matrix. Penggunaan metrik tersebut bertujuan untuk melihat tidak hanya ketepatan prediksi secara keseluruhan, tetapi juga kemampuan model dalam mengidentifikasi produk kategori Layak dan Tidak Layak secara lebih detail.



Gambar 2. Distribusi Kelas Kelayakan Produk

Gambar 2 memperlihatkan distribusi kelas pada dataset penelitian yang terdiri atas dua kategori, yaitu Layak dan Tidak Layak. Berdasarkan visualisasi tersebut, kategori Layak memiliki jumlah data lebih banyak dibandingkan kategori Tidak Layak, yaitu sebanyak 80 produk (66,67%), sedangkan kategori Tidak Layak sebanyak 40 produk (33,33%). Kondisi ini menunjukkan bahwa dataset memiliki perbedaan proporsi antar kelas, dengan kategori Layak sebagai kelas mayoritas.

Meskipun terdapat perbedaan distribusi kelas, proses pembagian dataset dilakukan menggunakan metode stratified sampling untuk mempertahankan proporsi masing-masing kelas pada data latih dan data uji. Selain itu, evaluasi model tidak hanya menggunakan accuracy, tetapi juga mempertimbangkan precision, recall, F1-score, dan confusion matrix agar kemampuan model dalam mengenali kedua kategori dapat dianalisis secara lebih objektif.

Selain dilakukan analisis distribusi kelas dan statistik secara keseluruhan terhadap atribut numerik dataset, penelitian ini juga melakukan analisis perbandingan karakteristik atribut berdasarkan kategori kelayakan produk. Analisis ini bertujuan untuk mengetahui perbedaan pola nilai rata-rata atribut numerik antara produk kategori Layak dan Tidak Layak sebelum dilakukan proses klasifikasi menggunakan algoritma machine learning. Perbandingan karakteristik atribut tersebut dapat memberikan gambaran awal mengenai faktor numerik yang memiliki kecenderungan berbeda pada masing-masing kelas. Hasil perbandingan rata-rata atribut numerik berdasarkan kategori kelayakan produk disajikan pada Tabel 5.

Tabel 5. Rata-Rata Atribut Numerik Berdasarkan Kelas Kelayakan Produk

Atribut	Tidak Layak	Layak
Protein (g)	3,38	16,26
Karbohidrat (g)	45,35	30,65
Lemak (g)	20,29	13,88
Vitamin	0,03	0,19
Kalori (kkal)	447,60	404,15
Harga (Rp)	3.925,00	10.062,50

Berdasarkan Tabel 5, terdapat perbedaan karakteristik atribut numerik antara produk kategori Layak dan Tidak Layak. Produk kategori Layak memiliki rata-rata kandungan protein dan vitamin yang lebih tinggi dibandingkan produk kategori Tidak Layak, yaitu masing-masing sebesar 16,26 gram dan 0,19. Hal ini menunjukkan bahwa aspek kandungan nutrisi tertentu memiliki pola yang berbeda antara produk yang dikategorikan layak dan tidak layak.

Sebaliknya, produk kategori Tidak Layak memiliki rata-rata kandungan karbohidrat, lemak, dan kalori yang lebih tinggi dibandingkan produk kategori Layak. Rata-rata kandungan karbohidrat pada kategori Tidak Layak mencapai 45,35 gram, sedangkan kategori Layak sebesar 30,65 gram. Selain itu, kandungan lemak dan kalori pada kategori Tidak Layak masing-masing sebesar 20,29 gram dan 447,60 kkal, lebih tinggi dibandingkan kategori Layak.

Dari aspek harga, produk kategori Layak memiliki rata-rata harga sebesar Rp10.062,50 per porsi, sedangkan produk kategori Tidak Layak memiliki rata-rata harga sebesar Rp3.925,00 per porsi. Perbedaan karakteristik tersebut menunjukkan bahwa atribut numerik memiliki pola tertentu yang dapat membantu algoritma machine learning dalam mengenali karakteristik masing-masing kelas. Namun, keputusan akhir klasifikasi tidak hanya ditentukan oleh atribut numerik, melainkan merupakan hasil kombinasi seluruh atribut penelitian, termasuk aspek legalitas PIRT, label halal, dan kebersihan kemasan.

Hasil Pengujian Model Klasifikasi

Setelah proses preprocessing dan pembagian dataset, dilakukan pengujian terhadap lima algoritma machine learning, yaitu Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM). Evaluasi dilakukan menggunakan data uji sebanyak 24 record dengan menggunakan metrik accuracy, precision, recall, dan F1-score. Perbandingan performa masing-masing algoritma pada proses klasifikasi kelayakan produk pangan UMKM disajikan pada Tabel 6.

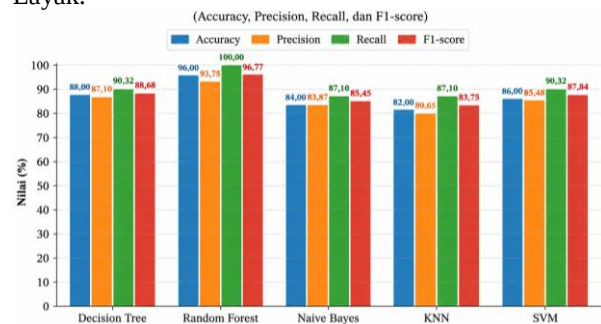
Tabel 6. Perbandingan Performa Algoritma pada Data Uji

Algoritma	Accuracy (%)	Precision (%)	Recall (%)	F1score (%)
Decision Tree	88,00	87,10	90,32	88,68
Random Forest	96,00	93,75	100,00	96,77
Naive Bayes	84,00	83,87	87,10	85,45

K-Nearest Neighbor (KNN)	82,00	80,65	87,10	83,75
Support Vector Machine (SVM)	86,00	85,48	90,32	87,84

Berdasarkan Tabel 6, Random Forest menghasilkan performa terbaik dibandingkan algoritma lainnya dengan nilai accuracy sebesar 96,00%, precision sebesar 93,75%, recall sebesar 100,00%, dan F1-score sebesar 96,77%. Nilai recall sebesar 100,00% menunjukkan bahwa seluruh produk yang termasuk kategori Layak berhasil dikenali oleh model tanpa adanya kesalahan klasifikasi pada kelas tersebut (False Negative = 0).

Decision Tree menunjukkan performa yang cukup baik dengan accuracy sebesar 88,00%, sedangkan Support Vector Machine, Naive Bayes, dan K-Nearest Neighbor menghasilkan nilai accuracy yang lebih rendah. Hasil tersebut menunjukkan bahwa pendekatan ensemble pada Random Forest mampu menghasilkan klasifikasi yang lebih stabil dibandingkan algoritma lainnya dalam membedakan produk Layak dan Tidak Layak.



Gambar 3. Perbandingan Performa Lima Algoritma Klasifikasi Berdasarkan Data Uji

Berdasarkan hasil pengujian pada data uji, Random Forest menghasilkan performa terbaik dibandingkan algoritma lainnya dengan nilai accuracy, precision, recall, dan F1-score tertinggi. Random Forest memperoleh accuracy sebesar 96,00%, precision sebesar 93,75%, recall sebesar 100,00%, dan F1-score sebesar 96,77%. Nilai recall yang mencapai 100% menunjukkan bahwa seluruh produk yang termasuk kategori layak berhasil dikenali oleh model.

Selain evaluasi menggunakan data uji, dilakukan validasi silang 10-fold untuk mengetahui konsistensi performa model pada beberapa pembagian data pelatihan dan validasi. Metode ini digunakan karena jumlah dataset penelitian relatif terbatas, sehingga diperlukan evaluasi tambahan untuk mengurangi kemungkinan bias akibat pembagian data tertentu. Hasil validasi silang 10-fold pada masing-masing algoritma disajikan pada Tabel 7.

Tabel 7. Hasil Validasi Silang 10-Fold pada Algoritma Klasifikasi

Algoritma	Accuracy (%)	Precision (%)	Recall (%)	F1-core (%)
Decision Tree	96,50	95,80	96,40	96,10
Random Forest	95,80	94,90	95,60	95,20
Naive Bayes	95,00	94,10	95,30	94,70

K-Nearest Neighbor (KNN)	90,20	88,70	90,00	89,50
Support Vector Machine (SVM)	92,80	91,90	92,50	92,40

Hasil validasi silang 10-fold pada Tabel 7 menunjukkan bahwa seluruh algoritma memiliki performa klasifikasi yang baik pada beberapa kombinasi data pelatihan dan validasi. Random Forest memperoleh rata-rata accuracy sebesar 95,8%, precision 94,9%, recall 95,6%, dan F1-score 95,2%. Nilai tersebut menunjukkan bahwa model Random Forest memiliki performa yang stabil ketika diuji pada beberapa subset data yang berbeda.

Pada hasil validasi silang, Decision Tree memperoleh nilai accuracy tertinggi sebesar 96,5% dengan F1-score sebesar 96,1%. Meskipun Decision Tree menunjukkan nilai rata-rata validasi silang yang sedikit lebih tinggi dibandingkan Random Forest, pemilihan model terbaik dalam penelitian ini tidak hanya didasarkan pada nilai validasi silang, tetapi juga mempertimbangkan hasil pengujian data uji, konsistensi performa model, serta kemampuan interpretasi melalui analisis feature importance. Berdasarkan pertimbangan tersebut, Random Forest dipilih sebagai model terbaik karena memberikan keseimbangan antara akurasi, stabilitas, dan kemampuan menjelaskan faktor yang memengaruhi keputusan klasifikasi.

Penggunaan validasi silang 10-fold menjadi penting karena dataset penelitian memiliki jumlah data yang terbatas dengan distribusi kelas yang tidak sepenuhnya seimbang. Metode ini memungkinkan evaluasi model dilakukan pada beberapa kombinasi data pelatihan dan validasi sehingga memberikan gambaran yang lebih representatif mengenai kemampuan generalisasi model dalam melakukan klasifikasi kelayakan produk pangan UMKM.

Algoritma Naive Bayes memperoleh rata-rata accuracy sebesar 95,0%, precision 94,1%, recall 95,3%, dan F1-score 94,7%. Hasil tersebut menunjukkan bahwa pendekatan probabilistik masih mampu mengenali pola klasifikasi dengan baik meskipun terdapat kemungkinan hubungan antaratribut dalam dataset.

Sementara itu, K-Nearest Neighbor (KNN) menghasilkan performa terendah dengan accuracy sebesar 90,2%, precision 88,7%, recall 90,0%, dan F1-score 89,5%. Performa tersebut menunjukkan bahwa metode berbasis jarak cukup dipengaruhi oleh karakteristik distribusi data dan variasi skala atribut numerik meskipun telah dilakukan proses standardisasi.

Support Vector Machine (SVM) memperoleh accuracy sebesar 92,8%, precision 91,9%, recall 92,5%, dan F1-score 92,4%. Hasil ini menunjukkan bahwa SVM mampu melakukan pemisahan kelas Layak dan Tidak Layak dengan baik melalui pembentukan hyperplane, namun interpretasi keputusan model masih lebih terbatas dibandingkan model berbasis pohon keputusan seperti Decision Tree dan Random Forest.

Secara keseluruhan, hasil validasi silang 10-fold memperlihatkan bahwa Random Forest memiliki performa yang konsisten dan mampu memberikan hasil klasifikasi yang baik. Kombinasi antara hasil pengujian

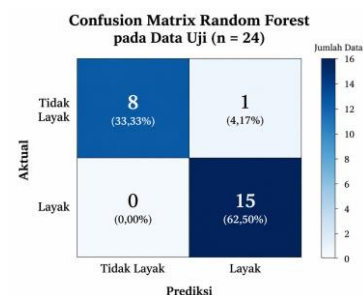
data uji, validasi silang, dan kemampuan interpretasi melalui feature importance menjadi dasar pemilihan Random Forest sebagai model terbaik untuk klasifikasi kelayakan produk pangan UMKM dalam mendukung proses seleksi awal Program Makan Bergizi Gratis (MBG).

Confusion Matrix Model Terbaik

Setelah model terbaik ditentukan berdasarkan hasil evaluasi performa, dilakukan analisis confusion matrix untuk mengetahui distribusi hasil prediksi Random Forest terhadap data uji. Confusion matrix digunakan untuk melihat jumlah prediksi benar dan kesalahan klasifikasi berdasarkan nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Hasil confusion matrix Random Forest pada 24 data uji disajikan pada Tabel 8.

Tabel 8. Confusion Matrix Random Forest

Aktual/Prediksi	Tidak Layak	Layak
Tidak Layak	8	1
Layak	0	15
Total	8	16



Gambar 4. Confusion Matrix Heatmap Model Random Forest pada Data Uji

Berdasarkan Tabel 8 dan Gambar 4, model Random Forest berhasil mengklasifikasikan 8 produk Tidak Layak dengan benar sebagai Tidak Layak (True Negative) dan 15 produk Layak dengan benar sebagai Layak (True Positive). Selain itu, terdapat 1 produk Tidak Layak yang diprediksi sebagai Layak (False Positive), sedangkan tidak terdapat produk Layak yang diprediksi sebagai Tidak Layak (False Negative = 0).

Tidak adanya nilai False Negative menunjukkan bahwa model Random Forest mampu mengenali seluruh produk yang termasuk kategori Layak pada data uji. Hal ini sejalan dengan nilai recall kelas Layak sebesar 100%, karena seluruh produk yang sebenarnya layak berhasil ditemukan oleh model. Namun, masih terdapat satu kesalahan False Positive, yaitu produk yang sebenarnya belum memenuhi kategori kelayakan tetapi diprediksi sebagai produk Layak.

Dalam konteks Program Makan Bergizi Gratis (MBG), kondisi tersebut perlu menjadi perhatian karena produk yang belum memenuhi kriteria berpotensi masuk ke tahap rekomendasi awal. Oleh karena itu, hasil prediksi model tetap perlu dikombinasikan dengan proses verifikasi dokumen legalitas, pemeriksaan kualitas pangan, dan evaluasi lapangan sebelum digunakan sebagai dasar pengambilan keputusan akhir.

Selain confusion matrix, evaluasi lebih rinci terhadap performa Random Forest dilakukan menggunakan classification report. Analisis ini digunakan untuk mengetahui kemampuan model dalam

mengenali masing-masing kelas, yaitu Layak dan Tidak Layak, berdasarkan nilai precision, recall, F1-score, dan jumlah data pada setiap kelas. Hasil classification report Random Forest pada data uji disajikan pada Tabel 9.

Tabel 9. Classification Report Random Forest pada Data Uji

Kelas	Precision (%)	Recall (%)	F1-score (%)	Support
Tidak Layak	100,00	88,89	94,12	9
Layak	93,75	100,00	96,77	15
Macro Average	96,88	94,44	95,45	24
Weighted Average	96,09	95,83	95,77	24

Berdasarkan Tabel 9, Random Forest memperoleh nilai recall sebesar 100,00% pada kelas Layak, yang menunjukkan bahwa seluruh produk yang termasuk kategori Layak berhasil dikenali oleh model. Nilai precision pada kelas Layak sebesar 93,75% menunjukkan bahwa sebagian besar produk yang diprediksi layak memang sesuai dengan kondisi sebenarnya.

Pada kelas Tidak Layak, model memperoleh precision sebesar 100,00%, yang menunjukkan bahwa seluruh produk yang diprediksi sebagai Tidak Layak benar-benar termasuk kategori tersebut. Namun, nilai recall sebesar 88,89% menunjukkan masih terdapat satu produk Tidak Layak yang salah diklasifikasikan sebagai Layak (False Positive).

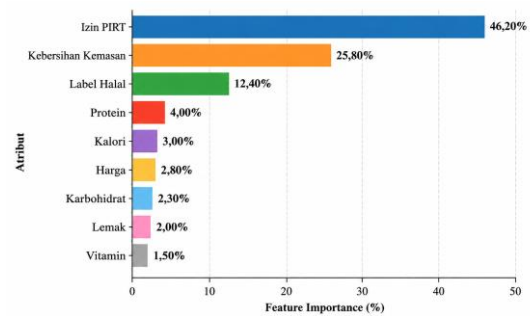
Secara keseluruhan, nilai weighted average F1-score sebesar 95,77% menunjukkan bahwa Random Forest mampu melakukan klasifikasi dengan performa yang sangat baik pada data uji. Hasil ini memperkuat pemilihan Random Forest sebagai model terbaik karena mampu mempertahankan keseimbangan antara kemampuan mengenali produk layak dan mengurangi kesalahan prediksi.

Analisis Atribut Paling Berpengaruh

Setelah diperoleh model terbaik berdasarkan hasil evaluasi performa dan validasi silang, dilakukan analisis feature importance menggunakan algoritma Random Forest. Analisis ini bertujuan untuk mengetahui kontribusi masing-masing atribut prediktor dalam proses pengambilan keputusan model pada klasifikasi kelayakan produk pangan UMKM. Nilai feature importance diperoleh dari tingkat kontribusi setiap atribut dalam membentuk struktur keputusan pada kumpulan decision tree yang digunakan oleh Random Forest. Hasil analisis feature importance Random Forest disajikan pada Tabel 10.

Tabel 10. Feature Importance Random Forest

Rank	Atribut	Feature Importance (%)
1	Legalitas PIRT	46,20
2	Kebersihan Kemasan	25,80
3	Label Halal	12,40
4	Protein	4,00
5	Kalori	3,00
6	Harga	2,80
7	Karbohidrat	2,30
8	Lemak	2,00
9	Vitamin	1,50
	Total	100,00



Gambar 5. Visualisasi Feature Importance Random Forest

Visualisasi feature importance pada Tabel 10 dan Gambar 5 menunjukkan tingkat kontribusi setiap atribut dalam menentukan keputusan klasifikasi kelayakan produk pangan UMKM menggunakan algoritma Random Forest. Hasil analisis menunjukkan bahwa Legalitas PIRT merupakan atribut dengan kontribusi terbesar yaitu sebesar 46,20%, diikuti oleh Kebersihan Kemasan sebesar 25,80%, dan Label Halal sebesar 12,40%.

Tingginya nilai kepentingan atribut Legalitas PIRT menunjukkan bahwa aspek legalitas menjadi faktor utama yang dipertimbangkan oleh model dalam membedakan produk kategori Layak dan Tidak Layak. Keberadaan legalitas PIRT dapat menjadi indikator bahwa produk pangan telah memenuhi persyaratan dasar terkait keamanan pengolahan dan kelayakan produksi pangan [3]. Dalam konteks seleksi awal produk pangan UMKM untuk mendukung Program Makan Bergizi Gratis (MBG), legalitas produk menjadi salah satu aspek penting yang perlu diperhatikan sebelum produk memasuki tahap evaluasi lebih lanjut.

Selain Legalitas PIRT, atribut Kebersihan Kemasan memiliki kontribusi terbesar kedua dengan nilai 25,80%. Hasil ini menunjukkan bahwa kondisi kebersihan dan kualitas kemasan merupakan faktor penting dalam proses klasifikasi. Kemasan yang baik dapat memberikan informasi mengenai aspek keamanan produk, mengurangi risiko kontaminasi, serta meningkatkan kepercayaan terhadap produk pangan yang akan digunakan dalam rantai penyediaan makanan [19].

Atribut Label Halal berada pada urutan ketiga dengan kontribusi sebesar 12,40%. Hal ini menunjukkan bahwa aspek sertifikasi dan informasi kehalalan produk turut menjadi pertimbangan model dalam menentukan kelayakan produk pangan UMKM. Keberadaan label halal memberikan nilai tambah terkait transparansi informasi dan kepastian standar produk bagi konsumen.

Selain tiga atribut utama tersebut, atribut numerik juga memberikan kontribusi terhadap keputusan model, yaitu Protein sebesar 4,00%, Kalori sebesar 3,00%, Harga sebesar 2,80%, Karbohidrat sebesar 2,30%, Lemak sebesar 2,00%, dan Vitamin sebesar 1,50%. Meskipun memiliki nilai kepentingan yang lebih rendah, seluruh atribut tetap digunakan secara bersama-sama oleh Random Forest dalam membentuk pola keputusan klasifikasi melalui mekanisme ensemble learning.

Secara keseluruhan, hasil feature importance menunjukkan bahwa model Random Forest lebih

banyak mempertimbangkan aspek legalitas dan keamanan produk dibandingkan atribut numerik lainnya dalam menentukan kelayakan produk pangan UMKM. Temuan ini menunjukkan bahwa produk dengan legalitas yang jelas, kemasan yang memenuhi aspek kebersihan, dan kelengkapan informasi produk memiliki kecenderungan lebih tinggi untuk diklasifikasikan sebagai produk layak.

Dalam implementasi Program Makan Bergizi Gratis (MBG), hasil analisis feature importance dapat dimanfaatkan sebagai informasi pendukung dalam menentukan prioritas pemeriksaan awal terhadap produk pangan UMKM. Namun, hasil prediksi model tetap perlu dikombinasikan dengan proses verifikasi dokumen, pemeriksaan mutu pangan, dan validasi lapangan sebelum digunakan sebagai dasar pengambilan keputusan akhir.

Kontribusi Keilmuan

Penelitian ini memberikan beberapa kontribusi dalam pengembangan model klasifikasi kelayakan produk pangan UMKM berbasis machine learning untuk mendukung proses seleksi awal pemasok dalam Program Makan Bergizi Gratis (MBG).

Pertama, dari aspek kontribusi metodologis, penelitian ini mengembangkan kerangka klasifikasi multiatribut yang mengintegrasikan berbagai karakteristik produk pangan UMKM, meliputi aspek legalitas, keamanan produk, kandungan gizi, kualitas kemasan, dan harga. Pendekatan ini memberikan kontribusi melalui penggabungan beberapa atribut produk secara simultan dalam proses klasifikasi, sehingga keputusan model tidak hanya bergantung pada satu indikator, tetapi mempertimbangkan kombinasi faktor legalitas, keamanan pangan, nutrisi, dan karakteristik ekonomi produk. Selain itu, penggunaan lima algoritma klasifikasi, yaitu Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM), memberikan perbandingan performa berbagai pendekatan machine learning dalam menyelesaikan permasalahan klasifikasi kelayakan produk pangan UMKM.

Kedua, dari aspek kontribusi praktis, penelitian ini menghasilkan model klasifikasi yang berpotensi dikembangkan sebagai komponen dalam sistem pendukung keputusan (Decision Support System/DSS) untuk membantu proses penyaringan awal produk pangan UMKM. Model yang dikembangkan dapat membantu proses evaluasi awal secara lebih objektif berdasarkan atribut produk yang tersedia, sehingga mendukung proses seleksi pemasok secara lebih cepat, konsisten, dan berbasis data. Namun, implementasi sistem secara penuh masih memerlukan pengembangan lebih lanjut dalam bentuk aplikasi atau integrasi dengan sistem informasi yang digunakan oleh pemangku kepentingan terkait.

Ketiga, dari aspek kontribusi empiris, penelitian ini memberikan temuan mengenai atribut yang paling berpengaruh terhadap keputusan klasifikasi melalui analisis feature importance pada model Random Forest. Hasil analisis menunjukkan bahwa Legalitas PIRT merupakan atribut dengan tingkat kepentingan tertinggi sebesar 46,20%, diikuti oleh Kebersihan Kemasan sebesar 25,80% dan Label Halal sebesar 12,40%.

Temuan tersebut menunjukkan bahwa aspek legalitas dan keamanan produk menjadi faktor dominan yang dipertimbangkan model dalam menentukan kategori kelayakan produk pangan UMKM. Selain itu, atribut nutrisi seperti protein, kalori, karbohidrat, lemak, dan vitamin tetap memberikan kontribusi dalam membentuk keputusan klasifikasi.

Keempat, dari aspek kontribusi operasional dan rekomendasi kebijakan, hasil penelitian dapat menjadi bahan pertimbangan dalam penyusunan mekanisme seleksi awal pemasok pangan UMKM. Indikator utama seperti legalitas PIRT, kebersihan kemasan, dan kelengkapan label produk dapat dijadikan prioritas pemeriksaan awal sebelum dilakukan proses verifikasi lebih lanjut. Pemerintah daerah, pendamping UMKM, maupun pelaksana Program Makan Bergizi Gratis dapat memanfaatkan pendekatan berbasis data ini sebagai alat bantu dalam menentukan prioritas evaluasi produk pangan UMKM.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan merupakan data riil produk pangan UMKM, tetapi jumlah data dan variasi produk masih terbatas sehingga belum sepenuhnya menggambarkan seluruh karakteristik produk pangan UMKM secara luas. Selain itu, model klasifikasi yang dikembangkan masih berada pada tahap pengujian eksperimental dan belum diuji secara langsung pada lingkungan operasional Program Makan Bergizi Gratis (MBG). Oleh karena itu, penelitian selanjutnya perlu melakukan pengumpulan data lapangan dengan jumlah sampel yang lebih besar, memperluas variasi jenis produk UMKM, menambahkan atribut pendukung seperti umur simpan, parameter mikrobiologi, dan standar keamanan pangan, serta melakukan uji implementasi lapangan melalui integrasi dengan sistem informasi daerah atau platform pendukung seleksi pemasok.

V. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi kelayakan produk pangan UMKM berbasis machine learning sebagai pendekatan pendukung keputusan dalam proses seleksi awal produk pangan untuk mendukung Program Makan Bergizi Gratis (MBG). Berdasarkan hasil pengujian menggunakan data uji, algoritma Random Forest menunjukkan performa terbaik dengan nilai accuracy sebesar 96,00%, precision sebesar 93,75%, recall sebesar 100,00%, dan F1-score sebesar 96,77%. Hasil tersebut menunjukkan bahwa model mampu melakukan klasifikasi produk Layak dan Tidak Layak dengan tingkat akurasi yang baik serta memiliki kemampuan tinggi dalam mengenali produk yang termasuk kategori layak.

Analisis feature importance pada model Random Forest menunjukkan bahwa atribut yang paling berpengaruh terhadap keputusan klasifikasi adalah Legalitas PIRT dengan kontribusi sebesar 46,20%, diikuti oleh Kebersihan Kemasan sebesar 25,80% dan Label Halal sebesar 12,40%. Temuan tersebut menunjukkan bahwa aspek legalitas dan keamanan produk menjadi faktor dominan yang dipertimbangkan model dalam menentukan kelayakan produk pangan UMKM, sementara atribut nutrisi dan harga tetap

memberikan kontribusi dalam proses pengambilan keputusan klasifikasi.

Secara praktis, model yang dikembangkan dapat menjadi dasar pengembangan Decision Support System (DSS) untuk membantu proses penyaringan awal produk pangan UMKM secara lebih objektif, konsisten, dan berbasis data. Model ini tidak dimaksudkan untuk menggantikan proses verifikasi oleh evaluator, tetapi berfungsi sebagai alat bantu dalam menentukan prioritas pemeriksaan produk sebelum dilakukan evaluasi lebih lanjut.

Penelitian ini memiliki keterbatasan karena meskipun menggunakan data riil produk pangan UMKM, jumlah data dan variasi karakteristik produk masih terbatas sehingga kemampuan generalisasi model terhadap seluruh kondisi produk pangan UMKM belum dapat dipastikan secara menyeluruh. Selain itu, model yang dikembangkan masih berada pada tahap pengujian eksperimental dan belum dilakukan implementasi serta pengujian langsung pada lingkungan operasional Program Makan Bergizi Gratis (MBG).

Untuk penelitian selanjutnya, pengembangan dapat dilakukan melalui beberapa langkah, yaitu memperluas pengumpulan data produk pangan UMKM dari berbagai wilayah dan jenis produk, melakukan pengujian lapangan terhadap model, serta menambahkan atribut pendukung seperti parameter mikrobiologi, umur simpan produk, standar keamanan pangan, kapasitas produksi, dan aspek distribusi. Selain itu, model dapat dikembangkan melalui integrasi dengan sistem informasi daerah atau platform digital pendukung Program Makan Bergizi Gratis agar proses seleksi dan monitoring produk pangan UMKM dapat dilakukan secara lebih efektif dan berkelanjutan.

REFERENSI

- [1] A. N. Ahsan and M. Alfaries, "Program pemerintah pemberian makan bergizi gratis (MBG) untuk meningkatkan sumber daya manusia unggul menuju Indonesia Emas 2045 dan implikasinya terhadap ketahanan nasional," *J. Elektrosista*, vol. 13, no. 1, pp. 28–39, 2025.
- [2] K. Syska *et al.*, "Menjaga Keamanan Pangan Berkelanjutan melalui Integrasi Gizi dan Jaring Pengaman Sosial pada Program MBG: Tinjauan Kritis," *J. Ilm. Pangan Halal*, vol. 7, no. 3, p. 424, 2025.
- [3] A. S. Krisna, A. F. Busthomi, M. Pramudita, and R. S. Utami, "Perlindungan Hukum Bagi Konsumen Terhadap Peredaran Obat Tradisional Tanpa Izin Edar Badan Pengawas Obat Dan Makanan," *Al-Zayn J. Ilmu Sos. Huk.*, vol. 4, no. 1, pp. 6008–6023, 2026.
- [4] C. L. Chasanah and S. Muzammil, "Sertifikasi Halal Badan Penyelenggara Jaminan Produk Halal (BPJPH) dalam Perspektif Yuridis," *Maslahah J. Manaj. Dan Ekon. Syariah*, vol. 3, no. 1, pp. 211–225, 2025.
- [5] M. R. Rizaldi, A. E. A. Srihanto, I. N. A. Afidah, A. Selviani, A. D. Sabilla, and D. Mahendra, "Penerapan Machine Learning Berbasis Supervised Learning Menggunakan Metode Naive BAYES UNTUK KLASIFIKASI PRESTASI SISWA," *J. Inf. Syst. Comput.*, vol. 6, no. 1, 2026.
- [6] A. Bastian *et al.*, "Penerapan Convolutional Neural Network (CNN) untuk Klasifikasi Kualitas Beras sebagai Strategi Peningkatan Keamanan Pangan di Indonesia," *TEMATIK*, vol. 12, no. 1, pp. 50–58, 2025.
- [7] O. S. Mahendra and A. P. Ramadhan, "Klasifikasi Kualitas Beras Menggunakan Convolutional Neural Network Berbasis Citra Digital," in *Seminar Nasional Teknologi & Sains*, 2026, pp. 650–655.
- [8] K. Syaban, "Evaluasi Model Ensemble Learning pada Identifikasi Faktor Risiko Diabetes Mellitus," *J. Teknol. dan Inf.*, vol. 15, no. 2, pp. 121–130, 2025.
- [9] A. B. Trisnawan and T. Susilawati, "Pemanfaatan Machine Learning untuk Peningkatan Akurasi Sistem Pendukung Keputusan Prediktif," *J. Unitek*, vol. 18, no. 2, pp. 303–312, 2025.
- [10] J. T. J. Simamora, A. S. Sihite, O. K. Hutagaol, and M. A. Sihombing, "Analisis Penerapan Algoritma Decision Tree, Random Forest, dan Naive Bayes pada Pengolahan Data untuk Sistem Cerdas," *IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 10, no. 2, pp. 641–651, 2026.
- [11] M. H. Setiono, "Komparasi Algoritma Decision Tree, Random Forest, Svm Dan K-Nn Dalam Klasifikasi Kepuasan Penumpang Maskapai Penerbangan," *INTI Nusa Mandiri*, vol. 17, no. 1, pp. 32–39, 2022.
- [12] M. Febrian, "Analisis Dan Implementasi Data Mining Untuk Memprediksi Hasil Belajar Mahasiswa Menggunakan Metode Naive Bayes Pada Fakultas Teknik Prodi Teknologi Informasi Universitas Muhammadiyah Palembang," 2026, *Universitas Bina Darma*.
- [13] B. Irawan, W. Prihartono, and N. Dienwatinuris, "Optimalisasi Decision Tree Untuk Penentuan Kelayakan Penerima Bantuan Kesejahteraan," *J. Comput. Sci. Artif. Intell.*, vol. 3, no. 06, 2026.
- [14] M. Ulfah, D. Darmansyah, and R. Rehani, "Instrumen Pengujian Produk Pembelajaran (Pengujian Validitas, Praktikalitas, Efektivitas)," *At-Tarbiyah J. Penelit. Dan Pendidik. Agama Islam*, vol. 3, no. 1, pp. 43–51, 2025.
- [15] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024.
- [16] F. D. Agustina, M. Arif, and S. Ahmad, "Systematic Literature Review atas Kinerja Algoritma KNN, Naive Bayes, dan Decision Tree pada Berbagai Studi Prediksi dan Klasifikasi," *J. Jawara Sist. Inf.*, vol. 3, no. 1, 2025.
- [17] H. Z. Rahimi, S. Defit, and J. Veri, "Implementasi decision tree dalam pengambilan keputusan untuk pemberian beasiswa," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 3, pp. 4175–4180, 2025.
- [18] A. A. Putri, "Perbandingan Metode Naive Bayes, KNN dan SVM pada Analisis Sentimen Jasa Transportasi Online= Comparison of Naive Bayes, KNN dan SVM for Sentiment Analysis of Online Transportation Services," 2023, *Universitas Hasanuddin*.
- [19] M. A. S. Ekowati, N. Hidayat, and A. Karim, "Predictive Modeling for Underweight Detection in Toddlers Using Support Vector Machine, K-Nearest Neighbors, and Decision Tree C4. 5 Algorithms," *J. Tek. Inform.*, vol. 6, no. 6, pp. 5570–5589, 2025.
- [20] S. Helmiyah and R. Pramestiawan, "Analisis Komparatif Algoritma Machine Learning dengan Metrik Akurasi, Presisi, Recall, dan F1-Score pada Dataset Kacang Kering," *J. Ilmu Komput. dan Teknol.*, vol. 6, no. 3, pp. 152–159, 2025.