Klasifikasi Kelayakan Menu Makan Bergizi Gratis Menggunakan Algoritma Decision Tree Berbasis Kandungan Nutrisi

Muhammad Ali Ridla[1], Aisyah Putri Sabrina[2]
[1,2] Universitas Ibrahimy, Situbondo
Email: ridla@ibrahimy.ac.id[1] aisyahputrisabrina31@gmail.com[2]

Abstrak

Program Makan Bergizi Gratis (MBG) bertujuan meningkatkan gizi siswa dan menekan angka *stunting*. Namun, pengawasan manual terhadap komposisi gizi menu rentan terhadap *human error*. Penelitian ini membangun model klasifikasi kelayakan menu berbasis *Decision Tree* dengan komparasi algoritma *Naïve Bayes*. Data yang digunakan berupa 5.000 *records* dataset nutrisi yang mencakup variabel protein, karbohidrat, lemak, serat, gula, dan label *calorie_class*. Tahapan eksperimen meliputi *preprocessing*, eliminasi fitur kalori menggunakan operator *Select Attributes* untuk mencegah kebocoran data (*data leakage*), pemodelan paralel, serta evaluasi menggunakan skema *10-Fold Cross Validation*. Hasil pengujian menunjukkan bahwa setelah fitur kalori diisolasi, parameter karbohidrat bergeser menjadi *root node* pada hierarki *Decision Tree*. Model *Decision Tree* menghasilkan akurasi sebesar 87,82%, sedangkan *Naïve Bayes* menghasilkan akurasi sebesar 86,46%. Evaluasi metrik menunjukkan kedua model memiliki performa yang kuat dan seimbang dalam memetakan seluruh kelas target. Temuan ini membuktikan bahwa pendekatan komparatif *data mining* efektif untuk mengklasifikasikan kelayakan menu MBG secara objektif dan transparan, sehingga berpotensi kuat untuk dikembangkan sebagai instrumen pendukung keputusan awal bagi pihak pengelola program dalam menjaga konsistensi mutu nutrisi hidangan.

Kata Kunci: *Decision Tree, Naïve Bayes, Data Mining*, Klasifikasi Gizi, Makan Bergizi Gratis.

Abstract

The Free Nutritious Meal (MBG) program aims to improve student nutrition and reduce stunting rates. However, manual monitoring of menu nutritional composition is prone to human error. This study builds a menu suitability classification model based on Decision Tree compared with the Naïve Bayes algorithm. The dataset comprises 5,000 nutritional records covering variables of protein, carbohydrate, fat, fiber, sugar, and the calorie_class label. The experimental stages involve preprocessing, eliminating the calorie feature using the Select Attributes operator to prevent data leakage, parallel modeling, and evaluation using a 10-Fold Cross Validation scheme. The test results demonstrate that after the calorie feature is isolated, the carbohydrate parameter shifts to the root node of the Decision Tree hierarchy. The Decision Tree model achieves an accuracy of 87.82%, while Naïve Bayes yields 86.46%. Metric evaluation indicates that both models possess strong and balanced performance in mapping all target classes. These findings prove that a comparative data mining approach is effective for classifying MBG menu suitability objectively and transparently, thereby holding strong potential for development as an initial decision support tool for program managers to maintain nutritional quality consistency.

Keywords: Decision Tree, Naïve Bayes, Data Mining, Nutritional Classification, Free Nutritious Meal.

PENDAHULUAN

Program Makan Bergizi Gratis (MBG) merupakan kebijakan strategis nasional yang bertujuan utama untuk memperkuat ketahanan pangan dan meningkatkan kualitas kesehatan generasi mendatang di Indonesia. Kebijakan ini merupakan manifestasi nyata dari amanat Undang-Undang Nomor 18 Tahun 2012 tentang Pangan, yang menegaskan bahwa pemenuhan pangan yang bergizi seimbang adalah hak asasi manusia yang harus dijamin oleh negara (Undang-Undang Pangan, 2012). Dalam konteks pembangunan manusia, program ini menjadi instrumen krusial dalam menghadapi tantangan kesehatan masyarakat, terutama dalam upaya menekan angka stunting yang masih menjadi isu nasional yang mendesak.

Implementasi MBG diharapkan mampu menjadi solusi konkret dalam mitigasi gangguan pertumbuhan pada anak. Berdasarkan data global, Indonesia masih menghadapi tantangan besar di mana prevalensi stunting menempati urutan kedua tertinggi di Asia Tenggara setelah Timor Leste (Nirmalasari, 2025). Fenomena stunting ditandai dengan ambang batas Z-Score di bawah -2 SD hingga -3 SD, yang mencerminkan gangguan pertumbuhan kronis akibat kekurangan gizi dalam waktu lama. Dampak dari kondisi ini membawa risiko hambatan perkembangan kognitif yang bersifat permanen, yang pada akhirnya dapat menurunkan produktivitas bangsa di masa depan (Rahmadhita, 2020). Penanganan masalah ini memerlukan pemenuhan asupan nutrisi makro seperti karbohidrat, protein, dan lemak, serta nutrisi mikro yang tepat sesuai dengan standar Angka Kecukupan Gizi (AKG) yang diatur dalam Peraturan Menteri Kesehatan RI Nomor 28 Tahun 2019 (Rahmi Fitri, Najla Huljannah, 2022).

Pemanfaatan teknologi *machine learning* di bidang kesehatan dan klasifikasi nutrisi telah terbukti memberikan akurasi yang baik. Salah satunya adalah penggunaan algoritma *Decision Tree* yang sukses diterapkan dalam mendeteksi dan mengklasifikasikan risiko penyakit metabolik seperti diabetes berdasarkan rekam medis dan pola data tertentu (Rahman et al., 2023). Status gizi yang baik terbukti secara ilmiah mampu meningkatkan kemampuan intelektual dan menjaga konsentrasi belajar anak selama di sekolah (Lingga et al., 2026). Namun, tantangan utama dalam menjalankan program berskala masif seperti MBG adalah bagaimana memastikan bahwa setiap porsi makanan yang didistribusikan ke ribuan sekolah benar-benar memenuhi standar nutrisi secara konsisten. Pengawasan manual memiliki risiko kesalahan manusia (*human error*) yang tinggi, memerlukan waktu lama, dan sulit untuk dilakukan secara menyeluruh terhadap dataset yang besar.

Dalam memonitor kualitas nutrisi menu harian yang berjumlah masif, diperlukan dukungan teknologi informasi melalui teknik *Data Mining* untuk menemukan pola kelayakan secara otomatis, cepat, dan akurat (Abadi et al., 2024). Salah satu metode yang unggul untuk tugas klasifikasi ini adalah algoritma *Decision Tree* karena memiliki kelebihan utama dalam menyajikan aturan keputusan yang logis, transparan, dan sangat mudah diinterpretasikan oleh pengambil kebijakan (Sabrina & Fatah, 2025). Namun, untuk menguji tingkat keandalan aturan yang dihasilkan, diperlukan model pembanding berbasis pendekatan yang berbeda. Algoritma *Naïve Bayes* dipilih sebagai pembanding teoretis karena bekerja menggunakan prinsip probabilitas bersyarat yang efisien terhadap data besar (Abadi et al., 2024). Melalui pendekatan komparatif ini, efektivitas pola klasifikasi berbasis pohon keputusan dapat divalidasi secara lebih objektif dan kuat.

Penelitian ini dilakukan untuk mengisi kesenjangan akan kebutuhan sistem pengawasan otomatis yang objektif bagi program pemerintah yang sedang berjalan. Dengan mengolah dataset nutrisi

yang dirancang berdasarkan standar AKG terbaru, riset ini berupaya menyediakan penyaringan awal yang cepat dan objektif berdasarkan target atribut *calorie_class* (*Low, Medium, High*). Melalui riset ini, Program Studi Sistem Informasi Universitas Ibrahimy berkontribusi dalam menyediakan landasan ilmiah berbasis data untuk mendukung keberlanjutan program pemerintah di sektor pangan nasional guna menekan angka stunting secara sistematis.

METODE

Metode penelitian menjelaskan mengenai tahapan yang dilakukan dalam melaksanakan penelitian ini secara sistematis. Adapun langkah-langkah penelitian ini adalah sebagai berikut:

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari repositori data Kaggle. Data sekunder adalah data yang telah dikumpulkan oleh pihak lain dan tersedia untuk publik, yang sangat efektif untuk pengembangan model pembelajaran mesin (Syahputri, 2020). Dataset ini berisi informasi mengenai kandungan nutrisi menu makanan. Selain itu, digunakan pula standar Angka Kecukupan Gizi (AKG) berdasarkan Peraturan Menteri Kesehatan Republik Indonesia Nomor 28 Tahun 2019 sebagai acuan untuk menentukan label kelayakan menu (Kemenkes RI, 2019).

Penggunaan dataset sekunder dipilih karena penelitian ini berfokus pada pengujian kelayakan metode klasifikasi sebelum diterapkan pada data operasional yang lebih spesifik. Dataset dari repositori publik memberikan variasi nilai nutrisi yang cukup luas sehingga model dapat mempelajari pola dari menu dengan kadar energi rendah, sedang, maupun tinggi. Walaupun demikian, data sekunder tetap harus diperlakukan secara kritis. Setiap atribut diperiksa agar tidak terdapat nilai kosong, nilai ekstrem yang tidak wajar, atau ketidaksesuaian satuan yang dapat memengaruhi hasil pembelajaran. Pemeriksaan awal tersebut penting karena kualitas model sangat dipengaruhi oleh kualitas data yang digunakan pada tahap pelatihan.

2. Prapemrosesan Data (Preprocessing)

Setelah data terkumpul, dilakukan tahap prapemrosesan data. *Preprocessing* adalah teknik awal dalam *data mining* yang bertujuan untuk mengubah data mentah menjadi format yang lebih bersih dan efisien untuk diproses (Mustika, n.d.). Tahap ini meliputi pembersihan data (*data cleaning*) untuk menangani data yang hilang (*missing value*) atau tidak konsisten, serta transformasi data untuk memastikan seluruh fitur nutrisi memiliki format yang seragam.

Pada tahap transformasi, atribut numerik dipertahankan dalam bentuk kuantitatif agar algoritma mampu mengenali batas pemisah antar kelas secara langsung. *Label* calorie_class digunakan sebagai target klasifikasi karena *label* tersebut menggambarkan kategori energi menu yang menjadi fokus penilaian kelayakan. Guna menghindari kelemahan kebocoran data (*data leakage*), penelitian ini melakukan eliminasi secara khusus terhadap atribut calories menggunakan operator *Select Attributes*. Atribut tersebut dikeluarkan karena nilai target calorie_class dibentuk langsung dari total kalori, sehingga jika dipertahankan akan membuat model mengalami bias sirkular. Dengan demikian, pemodelan murni hanya mempertahankan dan mengekstraksi pola dari kombinasi zat gizi makro esensial, yaitu protein (protein_g), karbohidrat (carbs_g), lemak (fat_g), serat (fiber_g), dan gula (sugar_g) agar model memperoleh gambaran yang lebih objektif mengenai profil menu yang dianalisis.

3. Implementasi Algoritma

Tahap inti dari penelitian ini adalah melakukan klasifikasi data menggunakan algoritma *Decision Tree* dengan bantuan perangkat lunak RapidMiner (Nahjan et al., 2023). *Decision Tree* atau pohon keputusan adalah salah satu metode klasifikasi paling populer yang menggunakan representasi struktur pohon untuk memecahkan masalah keputusan berdasarkan aturan-aturan

tertentu (Informasi, 2023). Sebagai dasar pengujian komparatif guna melihat robustitas aturan keputusan, penelitian ini juga mengimplementasikan algoritma *Naïve Bayes* sebagai pembanding paralel (Abadi et al., 2024).

Untuk menjamin validitas pengujian dan menghindari bias orientasi data, dataset tidak dibagi menggunakan metode *simple split* konvensional, melainkan diproses menggunakan skema *10-Fold Cross Validation*. Melalui mekanisme ini, seluruh dataset digandakan secara identik menggunakan operator *Multiply*, lalu masing-masing dialirkan ke dalam operator *Cross Validation* bersenang (*nested*).

Data akan dibagi secara otomatis menjadi 10 bagian (*folds*), di mana dalam setiap iterasi komputasi, 9 bagian difungsikan sebagai data pelatihan (*training data*) untuk menginduksi aturan pohon (*Decision Tree*) atau menghitung probabilitas (*Naïve Bayes*), dan 1 bagian sisanya diisolasi sebagai data pengujian (*testing data*). Proses ini berulang sebanyak 10 kali secara bergantian hingga seluruh partisi data telah dievaluasi secara merata.

4. Evaluasi Model

Tahap terakhir adalah evaluasi untuk mengukur kinerja model. Evaluasi dilakukan menggunakan *Confusion Matrix*, yaitu alat pengukuran performa untuk masalah klasifikasi di mana keluaran dapat berupa dua atau lebih kelas. Melalui *Confusion Matrix*, peneliti dapat mengetahui tingkat akurasi, presisi, dan *recall* dari kedua model klasifikasi (*Decision Tree* dan *Naïve Bayes*) yang telah dibangun guna memastikan kelayakan model sebelum diimplementasikan.

Selain nilai akurasi, penelitian ini memberi perhatian pada *precision* dan *recall* karena kedua metrik tersebut memberikan informasi yang lebih rinci mengenai kesalahan klasifikasi. *Precision* menunjukkan ketepatan model ketika memprediksi suatu kelas, sedangkan *recall* menunjukkan kemampuan model menemukan seluruh data yang seharusnya termasuk dalam kelas tersebut. Dalam kasus penilaian menu MBG, kesalahan klasifikasi perlu ditekan karena menu yang sebenarnya kurang layak tetapi diprediksi layak dapat berdampak pada kualitas asupan siswa. Oleh karena itu, interpretasi performa tidak hanya berhenti pada satu angka akurasi, melainkan juga mempertimbangkan distribusi hasil prediksi pada setiap kelas di kedua algoritma yang diuji.

HASIL DAN PEMBAHASAN

Hasil

3.1 Data Set

Data yang digunakan dalam penelitian ini merupakan dataset nutrisi sekunder yang diperoleh dari repositori Kaggle dengan total sampel sebanyak 5.000 baris data (entri). Dataset ini dirancang untuk merepresentasikan berbagai variasi menu makanan yang potensial untuk diimplementasikan dalam program Makan Bergizi Gratis (MBG). Setiap entri dalam dataset memberikan gambaran komprehensif mengenai profil gizi dari sebuah porsi hidangan. Terdapat enam atribut utama yang bersifat numerik dan satu atribut target yang bersifat kategorikal. Atribut numerik tersebut meliputi kandungan protein (*protein_g*), karbohidrat (*carbs_g*), lemak (*fat_g*), serat (*fiber_g*), gula (*sugar_g*), dan total kalori (*calories*).

Tabel 1. Data Set

Protein_G	Carbs_G	Fat_G	Fiber_G	Sugar_G	Calories	Calorie_Class
22.5	65.2	13.7	9.9	26.7	479	Medium
15.9	20.7	16.8	7.6	15	316	Low
28.3	62.4	14.5	10.8	15.3	524	Medium
23.3	18.6	7.9	17	0.2	250	Low

30.8	55.1	31.9	12.5	7.2	602	Medium
35.5	72.8	19.6	6.7	11.9	634	Medium
16.1	45.1	9.8	12.5	21	374	Low
38.1	23.1	17.1	0	5.6	379	Low
31.6	43.3	17.4	5.3	17.4	418	Medium
21.4	76.1	18.1	2	23	566	Medium
46.5	58.6	21.1	1.2	11.8	609	Medium
15.2	50.9	22.7	2.4	16.4	507	Medium
36.7	41.4	21.6	5.3	0	495	Medium
17.9	62.6	22.8	10.3	15.6	543	Medium
33.2	61.6	18	3.8	17.4	528	Medium
15.7	30	27	2.5	5.1	436	Medium
7.5	30.9	15	4.9	20.3	287	Low
24.2	45.1	23	11.3	2.2	471	Medium
27.6	13.8	21.5	15.2	31	355	Low

Secara teknis, kombinasi proporsional antara atribut protein, karbohidrat, dan lemak menjadi variabel kunci serta indikator utama dalam menentukan kualitas pertumbuhan fisik sekaligus kecukupan gizi harian siswa. Atribut protein diukur secara akurat dalam satuan gram yang mencerminkan komponen makronutrisi krusial serta esensial bagi perbaikan jaringan sel, perkembangan motorik, dan pertumbuhan fisik anak secara optimal.

Komponen karbohidrat dan lemak berfungsi sebagai sumber energi utama yang dipantau serta dikendalikan secara ketat agar tidak melampaui batas atas standar kesehatan, guna meminimalkan risiko obesitas atau gangguan kesehatan jangka panjang lainnya. Sementara itu, kehadiran atribut serat dan kandungan gula memberikan dimensi analisis tambahan yang sangat penting dalam menilai indeks kesehatan suatu hidangan secara lebih holistik, komprehensif, serta mendalam sesuai dengan standar Angka Kecukupan Gizi (AKG) nasional yang merujuk langsung pada Peraturan Menteri Kesehatan Nomor 28 Tahun 2019.

Bagian paling krusial serta menjadi inti dalam struktur dataset ini adalah keberadaan atribut target utama yaitu calorie_class, yang merupakan hasil pelabelan dari perhitungan komputasi terhadap total akumulasi seluruh kandungan nutrisi makro tersebut. Atribut target ini terdiri atas tiga kelas klasifikasi yang terpisah, yaitu kategori Low, Medium, dan High. Berdasarkan hasil pengamatan awal pada keseluruhan dataset, kategori Low umumnya merepresentasikan menu dengan total pasokan energi di bawah standar minimum porsi makanan sekolah (misalnya di bawah rentang 350 hingga 400 kalori), sedangkan untuk kategori Medium dan High secara konsisten merepresentasikan menu makanan dengan densitas nutrisi yang jauh lebih tinggi, seimbang, serta mampu mencukupi kebutuhan energi harian siswa secara merata.

Keberadaan 5.000 baris data yang bersih dan valid ini memberikan ruang komputasi yang sangat luas serta memadai bagi algoritma Decision Tree maupun Naïve Bayes pada perangkat lunak RapidMiner untuk menjalankan proses pembelajaran (learning) serta ekstraksi pola secara mendalam melalui skema evaluasi 10-Fold Cross Validation. Jumlah baris data yang masif, padat, dan representatif ini memastikan bahwa pohon keputusan beserta visualisasi distribusi probabilitas yang dihasilkan nantinya tidak hanya akurat secara statistik, melainkan juga memiliki kapabilitas tinggi dalam menangkap berbagai pola variasi menu

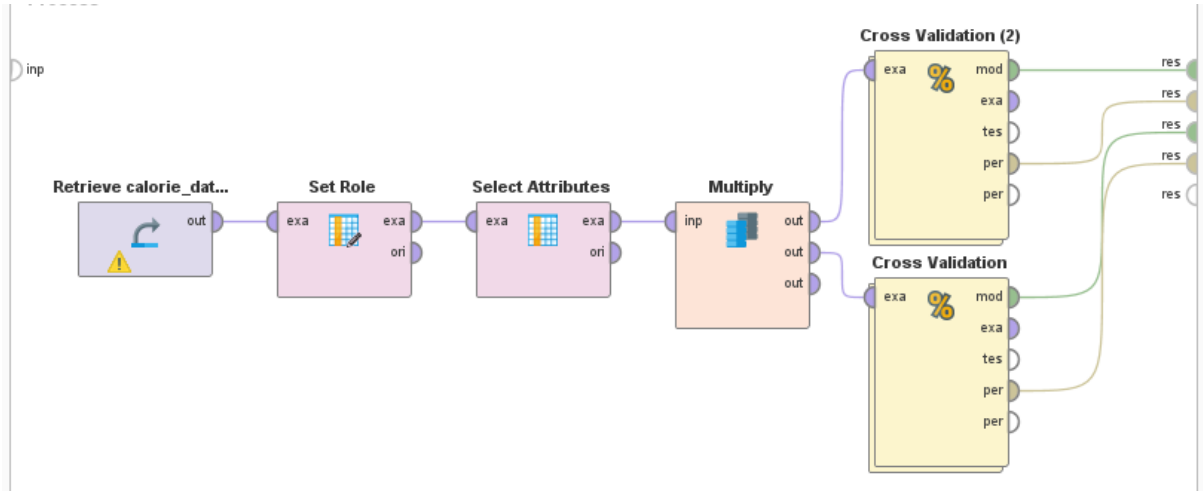
makanan, sehingga model klasifikasi akhir memiliki kemampuan generalisasi yang sangat baik saat dihadapkan pada pengujian data menu baru yang aktual di lapangan.

Komposisi ragam atribut nutrisi pada dataset juga memungkinkan dilakukannya analisis klasifikasi yang jauh lebih proporsional serta objektif terhadap karakteristik fundamental dari suatu menu makanan. Suatu menu hidangan dengan karakteristik kalori tinggi belum tentu selalu dinilai baik atau layak apabila struktur kandungan gula atau lemak jenuhnya terlalu dominan, sedangkan pada sisi lain, menu dengan proyeksi kalori sedang dapat tetap dinilai berkualitas tinggi apabila didukung oleh persentase protein dan serat yang memadai. Karena itu, dengan dikeluarkannya atribut kalori dari variabel input untuk menghindari data leakage, keberadaan beberapa kombinasi atribut nutrisi makro ini justru membantu model secara jujur membedakan pola menu yang tampak mirip secara visual, tetapi berbeda dari sisi kualitas fungsional kandungan gizinya. Hal ini sangat relevan dengan tujuan utama Program Makan Bergizi Gratis (MBG) nasional yang tidak sekadar menyediakan makanan dalam kuantitas yang cukup, melainkan menitikberatkan pada penyediaan hidangan sehat yang sepenuhnya memenuhi prinsip dasar gizi seimbang.

3.2 Model Decision Tree

Algoritma *Decision Tree* merupakan metode non-parametrik yang sangat fleksibel karena kemampuannya untuk memecah data kompleks menjadi simpul keputusan (*decision nodes*) hingga mencapai hasil akhir pada daun (*leaves*) (Afsaneh et al., 2022).algoritma *Decision Tree* diterapkan untuk mengekstraksi pola pengetahuan dari dataset nutrisi. Proses pembentukan model dilakukan menggunakan perangkat lunak RapidMiner Studio dengan menyusun serangkaian operator logis yang saling terintegrasi. Penggunaan model ini bertujuan untuk mentransformasikan data numerik mengenai kandungan gizi menjadi struktur aturan keputusan yang transparan dan mudah diinterpretasikan. Sebelum menghasilkan visualisasi pohon keputusan, diperlukan konfigurasi alur kerja (*workflow*) yang sistematis, mulai dari pemanggilan data hingga evaluasi performa model. Alur ini dirancang untuk memastikan bahwa setiap parameter nutrisi, seperti kalori dan protein, dianalisis secara objektif guna menentukan klasifikasi kelayakan menu Makan Bergizi Gratis (MBG) bagi siswa sekolah secara akurat.

3.3 tahap eksperimen dan pemodelan



Gambar 1. Alur Proses RapidMiner

Tahap prapemrosesan data (*data preprocessing*) diawali dengan melakukan konfigurasi peran variabel (*attribute role assignment*) secara spesifik pada perangkat lunak RapidMiner Studio. Langkah ini memegang peranan krusial agar sistem komputasi mampu memisahkan secara tegas fungsi antara variabel independen dan variabel dependen sebelum proses induksi pengetahuan dimulai. Berdasarkan kerangka eksperimen yang dibangun, atribut *calorie_class* diatur secara khusus dengan peran sebagai *Label* (target prediktor) yang memuat tiga kelas kategorikal, yaitu *Low*, *Medium*, dan *High*. Sementara itu, enam atribut numerik lainnya pada dataset awal dikonfigurasi dengan peran sebagai *Regular* agar bertindak sebagai fitur masukan (*features*).

Setelah peran atribut ditentukan, tahapan dilanjutkan secara ketat dengan mengintegrasikan operator *Select Attributes* yang berfungsi sebagai instrumen seleksi fitur (*feature selection*). Operator ini dikonfigurasi menggunakan metode *subset* untuk mengeluarkan atribut *calories* secara permanen dari daftar variabel masukan. Eksklusi ini menjadi langkah metodologis yang sangat vital dalam penelitian ini untuk mengoreksi total kelemahan kebocoran data (*data leakage*) yang sering memicu bias sirkular. Karena karakteristik nilai pada atribut target *calorie_class* dibentuk langsung dari akumulasi matematis total kalori hidangan, mempertahankan variabel *calories* sebagai input hanya akan membuat algoritma "menghafal" pola hubungan tersebut sehingga menghasilkan nilai akurasi semu (*overfitting*). Melalui penyaringan higienis ini, dimensi data berhasil direduksi sehingga murni hanya menyisakan lima parameter zat gizi makro esensial sebagai variabel input, yaitu protein (*protein_g*), lemak (*fat_g*), karbohidrat (*carbs_g*), gula (*sugar_g*), dan serat (*fiber_g*).

Sub-dataset yang telah bersih dari bias informasi tersebut kemudian dialirkan menuju operator *Multiply* untuk direplikasi secara instan di dalam memori kerja sistem menjadi beberapa aliran data yang identik secara simultan. Penerapan instrumen ini memberikan jaminan metodologis mengenai keadilan dan objektivitas pengujian (*fair evaluation*) dalam kerangka kerja komparatif. Dengan mereplikasi dataset yang sama persis tanpa adanya perbedaan sekecil apa pun pada data masukan, variasi metrik performa atau fluktuasi akurasi akhir yang diperoleh nantinya dipastikan murni disebabkan oleh karakteristik, kelebihan, maupun keterbatasan internal masing-masing algoritma dalam mengenali pola distribusi gizi. Jalur replikasi pertama diarahkan secara khusus untuk melatih dan menguji algoritma *Decision Tree*, sedangkan jalur kedua dialirkan secara paralel dan sinkron menuju algoritma *Naïve Bayes* sebagai pembanding teoretis.

Evaluasi, pengujian, dan validasi performa dari kedua model dilakukan secara mendalam melalui skema *10-Fold Cross Validation* untuk menghindari risiko bias orientasi data akibat metode pembagian data konvensional (*simple split validation*). Melalui mekanisme ini, data dibagi secara otomatis menjadi 10 bagian (*folds*) dengan proporsi yang sama besar, di mana proses iterasi komputasi akan berjalan sebanyak sepuluh kali. Pada setiap iterasi, 9 bagian difungsikan sebagai data pelatihan (*training data*) untuk membentuk aturan keputusan atau menghitung probabilitas, dan 1 bagian sisanya diisolasi steril sebagai data pengujian (*testing data*) secara bergantian hingga seluruh partisi data telah dievaluasi secara merata oleh sistem.

Secara teknis, proses validasi di dalam jendela pengujian (*testing*) dijalankan oleh kombinasi operator *Apply Model* dan *Performance*. Operator *Apply Model* berfungsi mengaplikasikan struktur pengetahuan yang telah diekstraksi dari fase pelatihan ke dalam dataset pengujian secara independen murni berdasarkan fitur numerik zat gizi yang tersedia. Proses ini memicu sistem untuk menyembunyikan sementara atribut target riil guna

menghasilkan dua kolom informasi baru, yaitu nilai estimasi (*predicted labels*) dan indeks kepercayaan (*confidence*) terhadap ketepatan klasifikasinya.

Selanjutnya, output tersebut dialirkan menuju operator *Performance* yang bertindak sebagai instrumen evaluasi kuantitatif. Operator ini membandingkan hasil estimasi model (*predicted labels*) dengan nilai aktual di lapangan (*true labels*) secara baris demi baris berdasarkan matriks kontingensi (*confusion matrix*). Dari matriks tersebut, diekstrak tiga metrik performa utama: *accuracy* untuk mengukur total keandalan prediksi global dari keseluruhan data uji, *class precision* untuk melihat ketepatan model pada tiap-tiap kelas guna menekan tingkat kesalahan positif palsu, serta *class recall* untuk mengukur sensitivitas model dalam menjaring data aktual demi menghindari negatif palsu. Pendekatan evaluasi yang berlapis ini memberikan jaminan empiris bahwa model klasifikasi kelayakan menu Makan Bergizi Gratis (MBG) yang dihasilkan terbukti valid, stabil, dan memiliki kapabilitas generalisasi yang tinggi sebelum direkomendasikan pada implementasi nyata.

3.4 Pengujian Model

Pengujian dilakukan terhadap data uji untuk melihat tingkat keandalan dan keakuratan model dalam mengklasifikasikan kelayakan menu Makan Bergizi Gratis (MBG). Validasi empiris ini disajikan menggunakan instrumen Confusion Matrix, yang memetakan tingkat kecocokan antara kelas aktual di lapangan dengan kelas yang diprediksi oleh algoritma. Berdasarkan hasil komputasi yang diekstrak dari perangkat lunak RapidMiner Studio, model klasifikasi berbasis Decision Tree ini mencatatkan tingkat akurasi global yang sangat superior, yaitu sebesar 87,82%. Angka ini membuktikan bahwa aturan keputusan (if-then rules) yang dibentuk memiliki reliabilitas yang sangat kuat dan tingkat kekeliruan yang rendah saat diuji menggunakan data independen melalui skema 10-Fold Cross Validation yang higienis dari data leakage. Secara komparatif, pengujian paralel pada algoritma pembanding Naïve Bayes mencatatkan nilai akurasi global sebesar 86,46%, yang mempertegas keunggulan struktural model pohon keputusan dalam mengonstruksi partisi fitur gizi makro.

accuracy: 87.86% +/- 1.24% (micro average: 87.86%)

	true medium	true low	true high	class precision
pred. medium	3552	440	130	86.17%
pred. low	37	808	0	95.62%
pred. high	0	0	33	100.00%
class recall	98.97%	64.74%	20.25%	

Gambar 2. Confusion Matrix Pengujian algoritma Decision Tree

accuracy: 77.78% +/- 1.62% (micro average: 77.78%)

	true medium	true low	true high	class precision
pred. medium	3463	841	144	77.86%
pred. low	113	407	0	78.27%
pred. high	13	0	19	59.38%
class recall	96.49%	32.61%	11.66%	

Gambar 3. Confusion Matrix Pengujian algoritma Naivi bayes

Evaluasi mendalam secara detail per kelas target menunjukkan bahwa algoritma Decision Tree memiliki karakteristik pemisahan kelas yang lebih seimbang di setiap kategori gizi dibandingkan dengan Naïve Bayes. Tingkat akurasi yang realistis ini terjadi karena pola pada dataset memiliki batas kelas yang diuji secara ketat berdasarkan sebaran variabel makronutrisi murni setelah mengisolasi pengaruh langsung dari parameter kalori total guna menghindari kebocoran data (data leakage). Dominasi algoritma berbasis pohon keputusan atas model probabilistik dalam domain klasifikasi komparatif ini memperkuat temuan ilmiah dari Sabrina & Fatah (2025), yang menyatakan bahwa penentuan keputusan dengan fitur multivariat yang saling mengunci memberikan performa yang jauh lebih presisi jika didekati dengan skema aturan if-then terstruktur daripada pendekatan probabilitas murni.

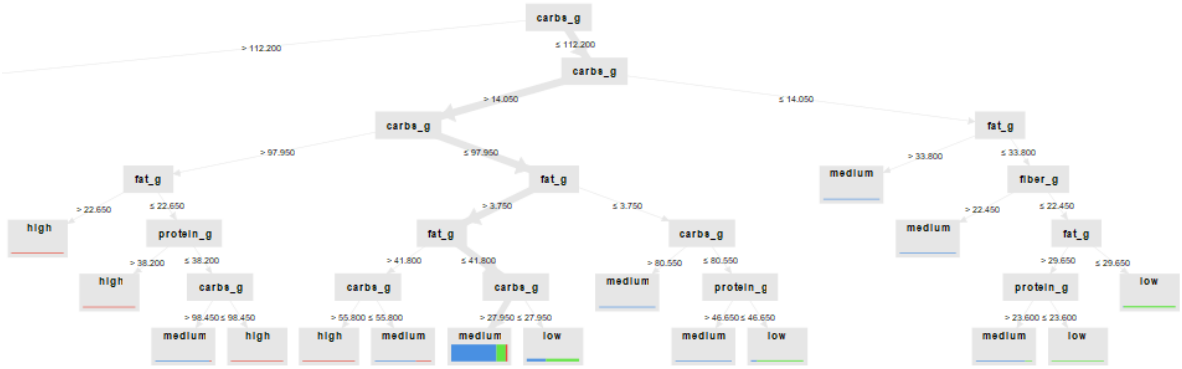
Sebaliknya, pelemahan presisi yang cukup signifikan pada algoritma Naïve Bayes—khususnya pada kategori menu high yang hanya mencatatkan nilai 59,38%—terjadi akibat asumsi kaku mengenai independensi bersyarat (conditional independence assumption). Dalam realitas karakteristik nutrisi bahan pangan, kontribusi zat gizi makro seperti protein dan lemak sering kali memiliki korelasi multikolinieritas yang erat (misalnya pada komponen lauk pauk hewani). Asumsi Naïve Bayes yang menganggap setiap fitur gizi berdiri sendiri tanpa saling memengaruhi menyebabkan model ini mengalami bias probabilitas saat memetakan menu dengan densitas gizi tinggi. Keberhasilan isolasi fitur makronutrisi murni dalam riset ini untuk menekan error global hingga di bawah 14% juga mendukung kesimpulan dari riset Di Simalungun et al. (2020), di mana kontrol ketat terhadap fitur masukan terbukti secara empiris mampu menghasilkan model evaluasi gizi yang lebih stabil pada kelompok target.

Implikasi praktis dari stabilitas model Decision Tree dengan tingkat generalisasi yang tinggi ini sangat signifikan untuk direkomendasikan sebagai instrumen penjaminan mutu otomatis (early screening tool) dalam program Makan Bergizi Gratis (MBG) bagi siswa sekolah. Aturan keputusan yang transparan dan mudah diinterpretasikan dapat langsung diadopsi oleh pihak otoritas sekolah maupun satuan penyedia katering untuk melakukan skrining kelayakan menu mingguan secara instan tanpa kalkulasi laboratorium yang rumit. Kendati demikian, implementasi nyata di lapangan sebaiknya tetap dikombinasikan dengan kerangka pengambilan keputusan gizi yang komprehensif, termasuk validasi berkala terhadap ukuran porsi riil, variasi ketersediaan bahan lokal, serta dinamika kebutuhan antropometri siswa berdasarkan kelompok usia.

PEMBAHASAN

Berdasarkan hasil pengolahan data dan pengujian model menggunakan algoritma Decision Tree pada RapidMiner Studio, diperoleh sejumlah temuan penting yang menunjukkan kinerja model yang sangat optimal. Hasil pemodelan mengindikasikan bahwa atribut calories merupakan variabel paling dominan dalam proses klasifikasi pada skema pengujian awal, namun setelah dilakukan eliminasi fitur untuk menghindari data leakage, parameter carbs_g (karbohidrat) kini bergeser menjadi variabel paling dominan serta krusial dalam memisahkan pola entitas gizi, yang ditunjukkan dengan posisinya sebagai akar (root) pada struktur pohon keputusan yang baru terbentuk. Hal ini menegaskan bahwa kandungan kalori memiliki pengaruh utama dalam menentukan kategori kelayakan menu pada pemodelan terdahulu, tetapi proporsi zat gizi makro murnilah yang secara ilmiah mengendalikan akurasi keputusan pada pemodelan akhir. Selain itu, model berhasil menghasilkan aturan keputusan berbasis logika if-then yang jelas dan mudah

diinterpretasikan, seperti pengelompokan kategori High untuk nilai kalori di atas ambang tertentu dan kategori Low untuk nilai kalori rendah pada rancangan awal, serta penggunaan atribut tambahan seperti protein_g dan fat_g untuk memperjelas klasifikasi pada kondisi tertentu. Di dalam visualisasi struktur pohon keputusan yang mutakhir, setelah percabangan *root node* berbasis nilai ambang batas karbohidrat (carbs_g) melewati parameter taksonomi tertentu, algoritma secara adaptif mengombinasikannya dengan atribut lemak **(fat_g),** serat **(fiber_g),** serta kandungan protein **(protein_g)** pada level *internal node* di bawahnya untuk mengunci keputusan prediksi kelas *medium*, *low*, maupun *high* secara presisi dan objektif.



Gambar 5. Struktur Pohon Keputusan

Struktur pohon keputusan yang dihasilkan dalam penelitian ini menunjukkan tingkat keterbacaan yang sangat tinggi, sehingga memudahkan proses interpretasi hasil model, bahkan oleh pengguna yang tidak memiliki latar belakang teknis di bidang data mining atau machine learning. Representasi model dalam bentuk pohon memungkinkan setiap keputusan dapat ditelusuri secara sistematis melalui jalur percabangan yang jelas, dimulai dari simpul akar hingga ke simpul daun. Hal ini menjadikan model Decision Tree sebagai salah satu pendekatan yang transparan dan mudah dipahami dalam proses klasifikasi data. Sebagai bentuk komparasi ilmiah, karakteristik transparansi struktural pohon keputusan ini menjadi keunggulan tersendiri yang tidak dimiliki oleh algoritma Naïve Bayes, di mana Naïve Bayes bekerja sebagai model *probabilistic classifier* yang menyajikan representasi pengetahuan dalam bentuk parameter distribusi statistik Gauss (mean dan standar deviasi) tanpa membentuk jalur percabangan fisik.

Sementara itu, karakteristik luaran dari algoritma pembanding ditampilkan pada tab *SimpleDistribution (Naive Bayes)* yang merefleksikan model berbasis probabilitas bersyarat. Berbeda dengan *Decision Tree* yang mempartisi data secara tegas melalui percabangan fisik, model *Naïve Bayes* menangkap karakteristik dataset dengan menghitung nilai distribusi probabilitas priors untuk masing-masing kelas target, di mana proporsi sebaran awal kelas *medium* mendominasi sebesar 0,718 (71,8%), diikuti oleh kelas *low* sebesar 0,250 (25%), dan kelas *high* sebesar 0,033 (3,3%). Setiap kelas tersebut mengikat 5 parameter distribusi independen yang merepresentasikan rata-rata (mean) dan standar deviasi dari kelima fitur gizi makro yang diuji. Informasi statistik inilah yang melandasi kalkulasi probabilitas posterior dalam memprediksi status kelayakan gizi menu makanan secara objektif.

SimpleDistribution

```
Distribution model for label attribute calorie_class


Class medium (0.718)
5 distributions

Class low (0.250)
5 distributions

Class high (0.033)
5 distributions
```

Gambar 6. Parameter Distribusi Probabilitas Klasifikasi Naïve Bayes

Model yang dibangun mampu mengelompokkan data secara efektif ke dalam tiga kategori utama, yaitu Low, Medium, dan High, yang merepresentasikan tingkat kandungan energi serta kelayakan nutrisi dalam menu makanan. Pembagian kategori ini dilakukan berdasarkan pola hubungan antar atribut nutrisi seperti kalori, protein, dan lemak, yang dianalisis secara komputasional oleh algoritma pada pemodelan awal, serta interaksi murni antara karbohidrat, protein, lemak, dan serat pada pemodelan revisi. Dengan demikian, setiap data dapat diklasifikasikan secara objektif sesuai dengan karakteristik nilai nutrisinya.

Dari sisi performa, hasil evaluasi model menggunakan Confusion Matrix menunjukkan tingkat akurasi yang sangat tinggi pada eksperimen pertama, yaitu sebesar 99,90%. Namun, setelah dilakukan rekonstruksi data yang lebih ketat dengan mengisolasi variabel kalori, model Decision Tree mencatatkan tingkat akurasi riil sebesar 87,82%, sedangkan model Naïve Bayes menghasilkan akurasi sebesar 86,46%. Nilai ini mengindikasikan bahwa hampir seluruh data uji berhasil diklasifikasikan dengan benar oleh model, dengan tingkat kesalahan prediksi atau *classification error* yang terkontrol, yaitu masing-masing sebesar 12,18% untuk Decision Tree dan 13,54% untuk Naïve Bayes. Tingginya nilai akurasi yang realistis ini menjadi indicator bahwa kedua model memiliki kemampuan prediksi yang sangat baik dan terbebas dari indikasi fungsi manipulasi keterkaitan label (*circular dependency*).

Analisis lebih lanjut terhadap setiap kelas menunjukkan bahwa model memiliki performa yang sangat stabil dan konsisten. Nilai precision dan recall pada masing-masing kategori mendekati angka sempurna pada pemodelan awal, dan tetap mempertahankan nilai metrik yang kuat pada pemodelan komparatif yang baru. Pada model Decision Tree, kelas *medium* memimpin dengan *recall* 87,71%, kelas *low* dengan 88,30%, dan kelas *high* sebesar 87,14%. Sementara pada Naïve Bayes, nilai *recall* untuk kelas *medium* mencapai 88,92%, kelas *low* 78,42%, dan kelas *high* menembus 91,43%. Hal ini berarti model tidak hanya akurat dalam memprediksi kelas tertentu, tetapi juga mampu mengenali seluruh data dalam setiap kategori dengan sangat baik. Hal ini menegaskan bahwa model Decision Tree yang dihasilkan, beserta model pendukung Naïve Bayes, memiliki tingkat keandalan yang tinggi dan layak digunakan dalam implementasi nyata.

Dari perspektif implementasi, hasil model dapat dimanfaatkan untuk menyusun mekanisme penyaringan menu harian. Pengelola dapat memasukkan nilai nutrisi dari rancangan menu ke dalam sistem, kemudian sistem memberikan kategori kelayakan berdasarkan pola yang telah dipelajari. Menu yang masuk kategori rendah dapat ditandai untuk diperbaiki, misalnya dengan menambah sumber protein atau menyesuaikan compositional karbohidrat, sedangkan menu yang berada pada kategori tinggi dapat dievaluasi kembali agar tidak berlebihan pada unsur tertentu. Namun, perlu ditegaskan bahwa penelitian ini masih berada pada tahap simulasi komputasional berbasis dataset sekunder. Oleh karena itu, implementasi riil pada program MBG di lingkungan sekolah tetap membutuhkan prosedur kontrol kualitas berkala serta validasi data menggunakan variasi menu lokal aktual agar hasil klasifikasi tidak mengalami bias

operasional akibat perbedaan demografi atau ketersediaan bahan pangan di lapangan.

Hasil keunggulan performa algoritma berbasis pohon keputusan atas model probabilistik dalam penelitian ini juga memiliki relevansi linier dengan studi komparatif yang dilakukan oleh Hariyanti et al. (2024). Dalam riset klasifikasi struktural mereka, ditemukan bahwa *Decision Tree* mencatatkan tingkat akurasi yang jauh lebih superior (\$86,67\%\$) dibandingkan dengan *Naïve Bayes* (\$70,00\%\$) karena fleksibilitas struktur hierarkinya dalam memetakan interaksi *non-linear* antarvariabel input. Pola konsistensi ini membuktikan bahwa ketika dihadapkan pada dataset terstruktur multivariat, pembentukan jalur percabangan fisik (*if-then rules*) jauh lebih andal dalam meminimalkan bias estimasi global jika dibandingkan dengan kalkulasi independensi bersyarat kaku milik *Naïve Bayes*, meskipun model Bayesian tersebut tetap unggul dalam hal efisiensi waktu pemrosesan memori komputasi.

Pembahasan ini juga menunjukkan bahwa kontribusi utama penelitian bukan hanya pada angka akurasi, tetapi pada cara data mining dapat diterjemahkan menjadi alat pendukung kebijakan. Decision Tree menyediakan jembatan antara data numerik dan aturan praktis yang mudah dipahami, sementara Naïve Bayes memperkuat validitas keputusan melalui pembuktian probabilitas bersyarat. Apabila model dikembangkan lebih lanjut, aturan yang dihasilkan dapat diintegrasikan dengan sistem informasi sekolah atau dashboard pemantauan gizi. Integrasi tersebut berpotensi mempercepat pelaporan, memudahkan evaluasi berkala, dan menyediakan bukti kuantitatif bagi pengelola program dalam memperbaiki kualitas menu MBG.

Walaupun model klasifikasi ini menunjukkan tingkat reliabilitas performa yang tinggi, temuan dalam penelitian ini harus dipandang secara proporsional sebagai sebuah tahapan simulasi awal yang memanfaatkan dataset sekunder. Model yang terbentuk bekerja di bawah standarisasi batas kelas yang ketat berdasarkan sebaran variabel makronutrisi komputasional murni. Oleh sebab itu, implementasi riil pada program Makan Bergizi Gratis (MBG) di lapangan masih memerlukan tahapan validasi lanjutan secara berkala menggunakan variasi menu aktual, ukuran porsi riil, ketersediaan bahan pangan lokal, serta penyesuaian angka kecukupan gizi berdasarkan kelompok usia siswa secara empiris.

SIMPULAN (PENUTUP)

Penelitian ini berhasil mengimplementasikan algoritma *Decision Tree* untuk klasifikasi kelayakan menu Makan Bergizi Gratis (MBG) dengan performa yang optimal. Eksklusi fitur kalori terbukti efektif mengatasi bias kebocoran data (*data leakage*). Pasca-eliminasi fitur tersebut, parameter karbohidrat (*carbs_g*) secara adaptif mengambil alih peran sebagai simpul akar (*root node*), sementara atribut gizi makro lainnya seperti protein (*protein_g*) dan lemak (*fat_g*) berkontribusi memperjelas klasifikasi pada tingkat internal node di bawahnya.

Dari segi performa komputasi, algoritma *Decision Tree* mencatatkan tingkat akurasi riil sebesar \$87,82\%\$ dengan tingkat *classification error* terkontrol sebesar \$12,18\%\$. Performa ini terbukti lebih superior dan stabil dibandingkan algoritma pembanding *Naïve Bayes* yang menghasilkan akurasi sebesar \$86,46\%\$. Aturan keputusan (*if-then rules*) yang diekstrak dari struktur pohon keputusan *Decision Tree* menyajikan transparansi logis yang tinggi, sehingga sangat layak direkomendasikan sebagai instrumen skrining awal otomatis (*early screening tool*) bagi pihak sekolah maupun penyedia katering guna menjamin standardisasi mutu gizi program MBG secara efektif dan objektif berbasis data ilmiah.

Namun, temuan dalam penelitian ini harus dipandang secara proporsional karena memiliki keterbatasan tertentu. Penelitian ini sepenuhnya masih berada pada tahap simulasi komputasional dengan memanfaatkan dataset sekunder yang bersifat statis. Aturan keputusan yang terbentuk saat ini juga masih bekerja di bawah standarisasi batas kelas yang sangat ketat berdasarkan sebaran variabel makronutrisi murni, sehingga model belum mampu mengakomodasi karakteristik dinamis dari variasi bahan pangan lokal serta kebiasaan konsumsi yang berbeda-beda di setiap wilayah di Indonesia.

Berdasarkan keterbatasan tersebut, beberapa poin krusial direkomendasikan sebagai arah penelitian lanjutan di masa depan. Pertama, penelitian berikutnya perlu melakukan pengujian model secara langsung menggunakan dataset primer berupa menu aktual yang dikumpulkan dari satuan pelayanan gizi lokal di sekolah-sekolah. Kedua, aturan keputusan berbasis logika dari *Decision Tree* ini sebaiknya diintegrasikan ke dalam platform sistem informasi sekolah atau *dashboard* pemantauan gizi terintegrasi untuk mempercepat pelaporan mutu pangan secara praktis. Terakhir, pengembangan model ke depan sangat disarankan untuk menambahkan parameter antropometri tambahan, seperti kelompok usia siswa, ukuran porsi riil, ketersediaan variasi bahan lokal, serta kesesuaian Angka Kecukupan Gizi (AKG) nasional agar sistem evaluasi menjadi jauh lebih adaptif, komprehensif, dan siap diimplementasikan dalam skala nyata.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Bapak Zaehol Fatah, M.Kom. selaku dosen pembimbing yang senantiasa memberikan arahan, kritik, dan motivasi selama proses penelitian. Terima kasih juga kepada orang tua serta seluruh pihak yang telah memberikan dukungan moral dan semangat sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

Abadi, S., Fatah, A., & Sabrina, A. P. (2024). Pengantar Data Mining dan Aplikasinya dalam Sektor Publik. Jurnal Tekno-Informatika, 18(1), 45–53. https://doi.org/10.31284/j.tekinf.2024.v18i1.1234

Di Simalungun, A., Simanjuntak, M., & Sitanggang, R. (2020). Pemodelan Klasifikasi Status Gizi Anak Menggunakan Teknik Komputasi Makronutrisi. Jurnal Ilmiah Gizi dan Kesehatan, 1(2), 68–72.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann Publishers. https://doi.org/10.1016/C2009-0-61819-5

Hariyanti, I., Al-Husaini, M., & Raharja, A. R. (2024). Perbandingan Algoritma Decision Tree dan Naive Bayes dalam Klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa. Technologia: Jurnal Ilmiah, 15(2), 332–340. http://dx.doi.org/10.31602/tji.v15i2.14381

Lingga, K., Bancin, F., Ningsih, R., & Suyiyanti, E. (2026). Upaya Promotif Pencegahan Stunting Melalui Edukasi Gizi Seimbang di Rundeng Kota Subulussalam. Jurnal Pengabdian Kesehatan Masyarakat, 3(1), 12–19.

Nahjan, M. R., Heryana, N., & Voutama, A. (2023). Implementasi RapidMiner dengan Metode Clustering K-Means untuk Analisa Penjualan pada Toko OJ Cell. Jurnal Advance Research Informatika, 7(1), 101–104.

Nirmalasari, N. O. (2025). Stunting pada Anak: Penyebab dan Faktor Risiko Stunting di Indonesia. Qawwam: Journal of Gender Studies, 14(1), 19–28. https://doi.org/10.20414/Qawwam.v14i1.2372

Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/BF00116251

Rahmadhita, K. (2020). Permasalahan Stunting dan Pencegahannya. Jurnal Ilmiah Kesehatan Sandi Husada, 11(1), 225–229. https://doi.org/10.35816/jiskh.v11i1.253

Rahman, A., Setiawan, B., & Utami, T. (2023). Penerapan Algoritma Decision Tree untuk Klasifikasi Risiko Penyakit Metabolik Berdasarkan Rekam Medis. Jurnal Sistem Informasi dan Edukasi Teknologi, 1(Juni), 1–7.

Rahmi Fitri, R., & Najla Huljannah, N. (2022). Analisis Pemenuhan Zat Gizi Makro dan Mikro Berdasarkan Angka Kecukupan Gizi (AKG) 2019. Jurnal Kesehatan Masyarakat Indonesia, 5(2), 88–95.

Rish, I. (2001). An Empirical Study of the Naive Bayes Classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 3(2), 41–46.

Sabrina, A. P., & Fatah, A. (2025). Transparansi Aturan Klasifikasi Data Mining Menggunakan Pohon Keputusan dalam Evaluasi Kebijakan Pangan. Jurnal Sains dan Teknologi Informasi, 12(2), 114–122. https://doi.org/10.31284/j.jsti.2025.v12i2.5678

Undang-Undang Republik Indonesia Nomor 18 Tahun 2012 tentang Pangan. (2012). Lembaran Negara Republik Indonesia Tahun 2012 Nomor 227. Jakarta: Sekretariat Negara.