Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 5 Iss. 4 October 2025, pp: 1498-1507
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Classification of Processed Food Menu Compositions Against Toddler Nutrition Standards Using Random Forest

Klasifikasi Komposisi Menu Makanan Olahan Terhadap Standar Gizi Balita Menggunakan Random Forest

Rahmi[1], Diva Nabila Herisnan[2*], Suandi Daulay[3] , Rahmaddeni[4]

[1,3]Program Studi Sistem Informasi, Sekolah Tinggi Teknologi Pekanbaru, Indonesia
[2,4]Program Studi Teknik Informatika,Universitas Sains dan Teknologi Indonesia,Indonesia

E-Mail: [1]rahmikairuddin@gmail.com, [2]divanabillapku123@gmail.com,
[3]suwandidaulay90@gmail.com [4]rahmaddeni@usti.ac.id

Received Aug 05th 2025; Revised Oct 20th 2025; Accepted Oct 31th 2025; Available Online Nov 04th 2025
Corresponding Author: Diva Nabila Herisnan

Abstract

Increasing public awareness of balanced nutrition in early childhood is crucial for preventing malnutrition and its associated health complications. This study presents a data mining approach to classify the nutritional content of processed food menus based on dietary standards for toddlers. The methodology employs the Random Forest algorithm to categorize menus into three classes: balanced, unbalanced, and excessive. A dataset comprising the nutritional values of menus, validated against the Recommended Dietary Allowance (RDA) for children aged 1–5 years, was used for model training and validation. The research process included data preprocessing, feature selection, and normalization. The model's performance was rigorously evaluated using metrics such as accuracy, precision, recall, and F1-score. The Random Forest model achieved no Table performance with an accuracy of 90%. Out of 136 menus, the model classified 9 as balanced, 59 as unbalanced, and 68 as excessive. These results underscore the potential of the Random Forest algorithm as a robust and efficient tool for nutritional surveillance in toddlers.

Keyword: Data Mining Classification, Nutrition, Processed Food Menu, Random Forest, Toddler Nutritional Standards.

Abstrak

Peningkatan kesadaran masyarakat akan pentingnya asupan gizi seimbang, khususnya pada anak usia dini, menjadi aspek krusial dalam upaya pencegahan malnutrisi dan masalah kesehatan terkait. Penelitian ini bertujuan untuk mengklasifikasikan komposisi menu makanan olahan terhadap standar gizi balita menggunakan pendekatan *data mining* dengan algoritma Random Forest. *Dataset* yang digunakan memuat kandungan nutrisi menu yang divalidasi terhadap standar Angka Kecukupan Gizi (AKG) untuk anak usia 1–5 tahun. Klasifikasi dilakukan ke dalam tiga kategori: seimbang, tidak seimbang, dan berlebihan. Penelitian melibatkan tahapan *preprocessing data*, *feature selection*, normalisasi, serta pelatihan model menggunakan Random Forest. Evaluasi menggunakan akurasi, presisi, *recall*, serta *f1-score*. Hasil pengujian diperoleh algoritma bahwa Random Forest menghasilkan kinerja terbaik dengan akurasi 90%. Dari 136 menu, 9 diklasifikasikan sebagai seimbang, 59 tidak seimbang, dan 68 berlebihan. Penelitian ini membuktikan jika algoritma Random Forest bisa dijadikan alat yang efektif dalam pemantauan gizi balita.

Kata Kunci: Klasifikasi *Data Mining*, Menu Makanan Olahan, Nutrisi, Random Forest, Standar Gizi Balita

1. PENDAHULUAN

Indonesia masih menghadapi tantangan besar dalam upaya pemenuhan gizi anak. Indonesia tercatat sebagai negara dengan jumlah orang yang kekurangan nutrisi tertinggi di Asia Tenggara, dengan sekitar 17,7 juta orang mengalami masalah gizi buruk [1]. Selain itu, Indonesia juga masih bergelut dengan tiga beban malnutrisi, yaitu *stunting*, *wasting*, dan kekurangan gizi mikro [2]. Pemerintah dan masyarakat telah melakukan berusaha untuk mendeteksi status gizi balita melalui Posyandu dan instansi kesehatan lainnya. Berdasarkan hasil Studi Status Gizi Indonesia (SSGI), prevalensi status gizi balita memang menunjukkan penurunan, namun angka permasalahan masih tergolong tinggi. Pada tahun 2021, prevalensi *underweight* sebesar 17% naik dari

16,3% pada 2019; prevalensi *stunting* sebesar 24,4% menurun dari 27,7%; dan *wasting* sebesar 7,1% sedikit menurun dari 7,4% pada tahun 2019 [3].

Standar gizi anak merujuk pada kebutuhan energi dan zat gizi berdasarkan usia, jenis kelamin, dan aktivitas fisik yang ditetapkan oleh Kementerian Kesehatan RI atau standar internasional seperti *World Health Organization* (WHO) dan *Food and Agriculture Organization* (FAO). Di Indonesia, standar tersebut tertuang dalam Angka Kecukupan Gizi (AKG) dari Kemenkes Replubik Indonesia (RI), yang mencakup rekomendasi energi (kkal), protein (gram), serta vitamin dan mineral penting sesuai usia. Standar ini digunakan sebagai acuan dalam pemantauan pertumbuhan anak melalui *antropometri* seperti berat badan dan tinggi badan terhadap umur [4].

Salah satu faktor utama penyebab permasalahan ini adalah pola makan yang tidak memadai, di mana 2 dari 5 anak di bawah usia lima tahun tidak menerima jumlah kelompok makanan yang direkomendasikan[5]. Pemantauan kepatuhan menu terhadap standar gizi seringkali memerlukan tenaga ahli gizi dan proses manual yang memakan waktu. Hal ini menunjukkan pentingnya pemantauan serta analisis terhadap kepatuhan nutrisi dalam menu makanan anak sebagai langkah strategis dalam menanggulangi malnutrisi [6].

Salah satu pendekatan yang menjanjikan dalam analisis kepatuhan terhadap standar gizi adalah penerapan *data mining*. *Data mining* merupakan suatu proses eksplorasi dan analisis data sesuai dengan tujuan menemukan pola berguna, hubungan tertutup, dan rincian informasi yang ditemukan dari kumpulan data yang sangat besar. Proses ini dilakukan dengan menggunakan berbagai teknik dan algoritma untuk menemukan tren, mengklasifikasikan data, dan membuat prediksi yang berguna untuk pengambilan keputusan [7]. Algoritma klasifikasi pada *Data mining* dapat mengolah dan menganalisis sejumlah besar data kompleks, serta mengidentifikasi pola yang tidak terdeteksi dengan mudah oleh metode konvensional [8]. Salah satu algoritma yang digunakan adalah random forest. Algoritma ini menggunakan banyak pohon keputusan yang dibangun secara acak, dan hasilnya adalah kombinasi prediksi dari seluruh pohon. Cara kerja algoritma ini adalah dengan membuat beberapa pohon keputusan dan menyatukannya untuk menghasilkan prediksi yang lebih akurat dan stabil [9]. Tahap awal penelitian ini mencakup pemberian *labeling* pada data gizi anak dengan kategori seimbang, tidak seimbang, dan berlebihan sesuai dengan standar gizi yang ada. Selain itu, *feature selection* juga digunakan untuk memilih fitur paling relevan yang dapat meningkatkan akurasi model. Dengan pendekatan ini, *data mining* dapat memberikan prediksi yang lebih akurat dalam mengidentifikasi kepatuhan gizi serta memungkinkan perancangan intervensi yang lebih efektif.

Penelitian sebelumnya juga menunjukkan pentingnya klasifikasi status gizi menggunakan algoritma *data mining*. Misalnya, penelitian menggunakan algoritma C4.5 untuk mengklasifikasikan status gizi balita dan berhasil mengidentifikasi prevalensi *stunting* dengan tingkat akurasi tinggi [3]. Penelitian dengan mengembangkan model deteksi *stunting* berbasis web menggunakan algoritma Random Forest dengan akurasi 85%, performa stabil melalui *K-Fold Cross Validation*[10]. Selanjutnya, Peneliti menerapkan algoritma Random Forest dan memperoleh akurasi klasifikasi hingga 88,6%[11]. Penelitian memanfaatkan algoritma Decision Tree CHAID dengan hasil akurasi yang tinggi [12], sedangkan penelitian lainnya menggunakan SVM dengan akurasi 91,91% [13].

Berdasarkan permasalahan dan penelitian terdahulu yang menerapkan algoritma Random Forest untuk deteksi *stunting* atau klasifikasi status gizi secara umum, penelitian ini difokuskan untuk membangun model klasifikasi komposisi menu makanan olahan terhadap standar gizi balita dengan menggunakan algoritma Random Forest. Penelitian ini diharapkan sebagai kontribusi untuk meningkatkan kesadaran diri masyarakat terhadap pentingnya gizi seimbang, mendukung pengembangan sistem pemantauan gizi berbasis *data mining*, serta menjadi referensi dalam penelitian lanjutan di bidang teknologi dan kesehatan anak.

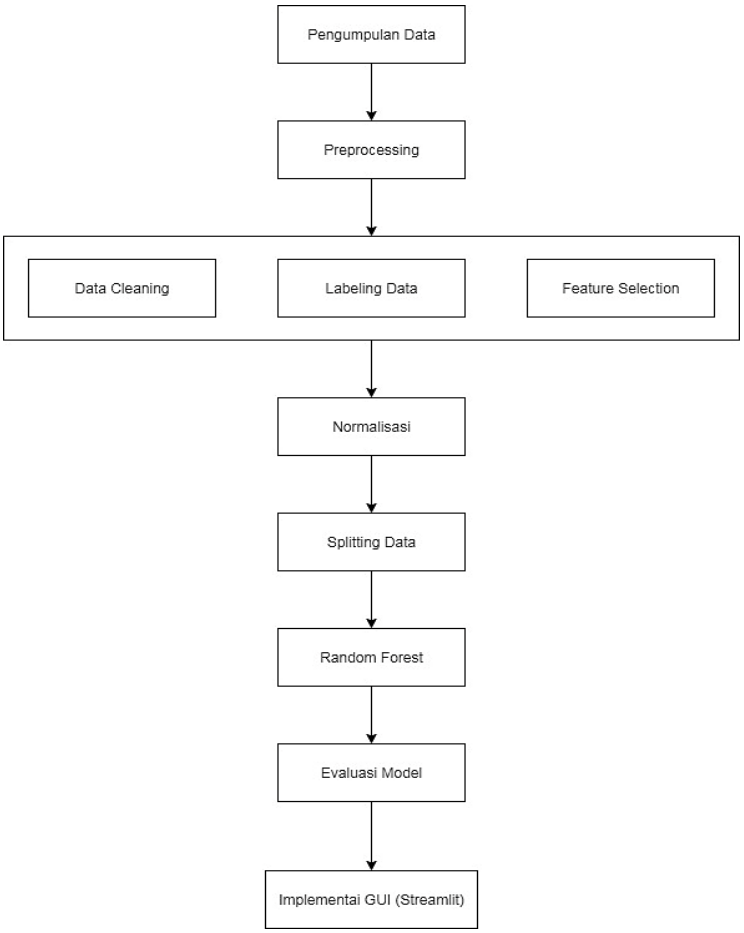
2. METODOLOGI PENELITIAN

Metodologi penelitian ini dibagi dalam enam tahapan (Gambar 1) . beberapa penjelasan dari metodologi penelitian yaitu :

1. Pengumpulan Data, Data makanan berupa informasi nilai nutrisi pada menu makanan olahan balita dikumpulkan dari *dataset* publik yang tersedia di situs Kaggle
2. *Preprocessing*, *Dataset* yang akan digunakan akan diolah menjadi data bersih yang siap untuk diuji.
3. Normalisasi, Normalisasi dilakukan agar seluruh atribut numerik berada dalam rentang nilai yang sama, sehingga proses pembelajaran model dapat berjalan secara optimal
4. *Splitting Data*, dilakukan untuk melakukan pelatihan model pada subset data tertentu, serta menilai performa prediksinya dengan memanfaatkan data yang sebelumnya tidak pernah diakses.
5. Evaluasi Model, melakukan pengujian menggunakan *dataset* pengujian untuk mengevaluasi performa model untuk menentukan apakah model yang diuji mampu mencapai akurasi yang memadai untuk penelitian.
6. Implementasi GUI, penelitian menggunakan Streamlit sebagai framework untuk membangun antarmuka grafis pengguna (GUI) yang interaktif dan berbasis web.

2.1. Pengumpulan Data

Penelitian ini memperoleh data dari dua sumber. Pertama, *dataset* publik dari Kaggle dengan judul *Food Nutrition Data*. Dari total 1.226 baris data, dipilih 136 baris dengan 21 fitur gizi, antara lain energi (kkal), protein (g), lemak (g), karbohidrat (g), serat pangan (g), natrium (mg), kalsium (mg), zat besi (mg), seng (mg), dan fitur nutrisi lainnya. *Dataset* kedua berisi standar kebutuhan gizi balita sebagai batas minimal dan maksimal, mengacu pada AKG 2019 [14].



Gambar 1. Alur Penelitian

2.2. *Preprocessing*

Tahap *preprocessing* dilakukan untuk memastikan data siap diproses dengan baik. Tahapan *Preprocessing* meliputi:

1. Pembersihan Data: menghapus data yang salah, tidak lengkap, tidak relevan, atau memiliki *missing value*. Jika ada nilai hilang, diganti dengan median agar tetap representatif.
2. Pemberian Label (*Labeling*): setiap data diberi label sesuai kategori "kekurangan", "seimbang", dan "berlebihan" berdasarkan ambang batas nutrisi yang sudah ditentukan.
3. Seleksi Fitur (*Feature Selection*): digunakan metode *embedded* untuk pilih fitur mana yang paling sebanding dan menghilangkan fitur yang kurang berpengaruh agar model lebih efisien.

2.3. Normalisasi Data

Dalam penelitian ini, normalisasi dilakukan agar seluruh atribut numerik berada dalam rentang nilai yang sama, sehingga proses pembelajaran model dapat berjalan secara optimal. Metode yang digunakan yaitu *Min-Max Normalization*, di mana nilai dari setiap fitur akan dikonversi ke dalam rentang 0 hingga 1[15].

2.4. *Splitting Data*

Splitting data atau pemisahan data adalah tahapan penting dalam pemodelan yang bertujuan untuk membagi *dataset* menjadi dua bagian utama, yaitu data latih *(training data)* dan data uji *(testing data)*, agar model yang dibangun dapat dievaluasi secara objektif tanpa bias terhadap data pelatihan [16]. Untuk membagi data latihan dan data uji, validasi *holdout* dan *cross-k-fold* bisa digunakan. Tujuannya adalah agar setiap data

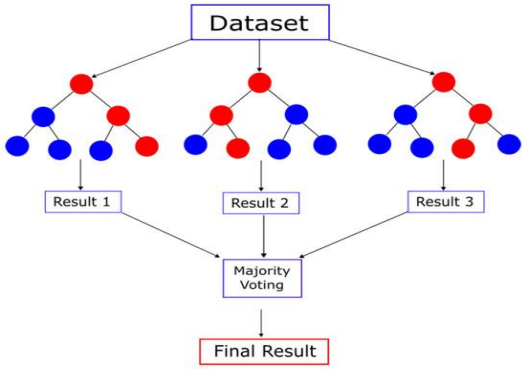
memiliki kesempatan untuk dilatih dan diuji [17]. Dalam penelitian ini, digunakan tiga variasi rasio pembagian data, yaitu: 80% :20%, 70% :30%, 50% :50%.

2.5. Random Forest

Random Forest adalah algoritma *ensemble* yang menggabungkan banyak pohon keputusan guna memperoleh prediksi yang lebih akurat dan stabil sekaligus mengurangi potensi *overfitting* [18]. Penelitian menunjukkan bahwa Random Forest mencapai akurasi 95% dalam klasifikasi status gizi balita, menjadikannya algoritma yang efektif dalam konteks ini [19]. Pada Random Forest, pembentukan pohon keputuan diawali dengan pemilihan atribut akar berdasarkan nilai gain yang diperoleh dari perhitungan entropi. Proses ini kemudian di ulan pada tiap atribut hingga semua instance berada dalam kelas yang sama [11]. Berikut ini merupakan rumus dari Random Forest (RF) [20] :

$$Y = mode\left(h_1(\propto), h_2(\propto), \dots, h_n(\propto)\right) \quad (1)$$

1. Mengambil *subset* acak dari data pelatihan
2. Mengambil *subset* acak dari fitur (*fitur selection*)
3. Membangun banyak pohon keputusan
4. Melakukan *voting* mayoritas untuk menentukan label akhir (lihat Gambar 2)



Gambar 2. Random Forest

2.6. Evaluasi Model

Untuk menganalisi kinerja model algoritma *Machine Learning* penulis menggunakan metode *Confusion Matrix*. *Confusion Matrix* adalah metode yang dipakai untuk menilai tingkat akurasi dari suatu model klasifikasi. Presisi/*confidence* dihitung dari perbandingan jumlah prediksi positif yang benar dengan total kasus yang diprediksi positif. *Recall* atau sensitivitas adalah proporsi kasus positif sebenarnya yang di prediksi positif secara benar [21].

Tabel 1. *Confusion Matrix*

Sebenarnya	Prediksi	
	Positive	*Negative*
Positive	*True Positive (TP)*	*False Negative(FN)*
Negative	*False Positive (FP)*	*True Negative(TN)*

Tabel 1 merupakan tabel *confusion matrix* dengan keterangan sebagai berikut: *True Positive (TP)* = Jumlah data yang benar diklasifikasikan sebagai positif oleh model; *True Negative(TN)* = Jumlah data yang benar diklasifikasikan sebagai negatif oleh model; *False Negative(FN)* = Jumlah data yang salah diklasifikasikan sebagai negatif oleh model; dan *False Positive (FP)* = Jumlah data yang salah diklasifikasikan sebagai positif oleh model.

2.7. *Literature Review*

Dalam beberapa tahun terakhir, berbagai penelitian telah menunjukkan bahwa algoritma Random Forest banyak digunakan dalam klasifikasi status gizi maupun bidang kesehatan karena kemampuannya menghasilkan akurasi tinggi dan mengurangi *overfitting*. Penerapan Random Forest pada klasifikasi status gizi balita memberikan akurasi hingga 88,6%, menunjukkan performa yang stabil dalam mengolah data kompleks [11]. Model serupa juga diterapkan pada sistem deteksi *stunting* berbasis web dengan tingkat akurasi mencapai 85% [10] . Selain itu, kombinasi Support Vector Machine (SVM) dan Random Forest terbukti mampu meningkatkan hasil klasifikasi status gizi balita dengan akurasi hingga 91% [19].

Penelitian terbaru mengembangkan sistem klasifikasi status gizi balita berbasis web menggunakan algoritma Random Forest dan memperoleh akurasi sangat tinggi mencapai 99,91% pada *dataset* berjumlah lebih dari 120.999 data anak. Hasil ini membuktikan bahwa algoritma Random Forest sangat efektif diterapkan dalam sistem berbasis aplikasi untuk pemantauan status gizi secara *real-time* [22].

Analisis lebih lanjut terhadap faktor risiko *stunting* juga dilakukan dalam penelitian [23], yang menggunakan metode Random Forest untuk mengidentifikasi variabel paling berpengaruh terhadap kondisi gizi balita. Hasil penelitian tersebut menunjukkan akurasi sebesar 97%, dengan variabel utama yang memengaruhi adalah berat lahir dan usia pengukuran, yang mengonfirmasi keunggulan Random Forest dalam menentukan fitur penting.

Berdasarkan hasil-hasil penelitian tersebut, dapat disimpulkan bahwa algoritma Random Forest terbukti efektif dan adaptif dalam berbagai konteks analisis gizi. Oleh karena itu, penelitian ini memilih Random Forest sebagai model utama untuk mengklasifikasikan komposisi menu makanan olahan terhadap standar gizi balita berdasarkan AKG 2019, dengan harapan dapat mencapai hasil klasifikasi yang akurat, stabil, dan relevan secara praktis.

2.8. Implementasi GUI

Pada proses ini, penelitian menggunakan *Streamlit* sebagai *framework* untuk membangun GUI yang interaktif dan berbasis web. Aplikasi ini dirancang untuk menampilkan seluruh alur klasifikasi secara *end-to-end*, mulai dari unggah data, pelatihan model, evaluasi, hingga hasil akhir klasifikasi.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Penelitian ini memanfaatkan dua jenis data, yaitu data komposisi menu makanan olahan serta data standar kebutuhan gizi balita usia 1-5 tahun yang merujuk AKG 2019. (Lihat Tabel 2 dan Tabel 3)

Tabel 2. *Dataset* Komposisi Menu Makanan Olahan

Menu	*Energy (kkal)*	*Protein (g)*	*Fat (g)*	*Carbohydrates (g)*	*Dietary Fiber (g)*	...	*Zinc (mg)*
nasi tim + Perkedel Tempe Wortel	390	6.5	12.5	59	5.5	...	0.4
nasi tim ayam tomat wortel	200	6	4	38	2	...	0.6
nasi tim bayam	418	2.1	0.2	21.8	0.5	...	0.3
nasi tim daging	649	5.4	4.8	21.5	0.2	...	0.9
nasi tim wortel kentang	439	2	0.2	23.2	0.7	...	0.3
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
Nasi tim ikan	320	8	6	65	4	...	0.4

Tabel 3. Data Standar Kebutuhan Gizi Balita

Nutrisi	Minimum	Mximum
Energy (kkal)	338	420
Protein(g)	5	8
Fat(g)	11	15
Carbohydrates(g)	54	66
Dietary Fiber (g)	4.8	6
Sodium (mg)	300	350

3.2. *Preprocessing*

Dalam tahap ini, beberapa langkah biasanya dilakukan untuk membersihkan dan mempersiapkan data sebelum pemrosesan lebih lanjut. Proses pelabelan dilakukan dengan membandingkan nilai setiap nutrisi pada menu terhadap batas minimal dan maksimal dari standar AKG balita. Jika nilai berada di bawah batas minimal maka diberi label “kekurangan”, jika berada di dalam rentang maka “seimbang”, dan jika melebihi batas maksimal maka “berlebihan”. Tabel 4 adalah hasil dari data yang telah di *preprocessing*.

3.3. Normalisasi Data

Proses ini mengubah skala nilai pada atribut numerik agar berada dalam rentang yang sama, sehingga proses pelatihan model menjadi lebih optimal. Metode yang diterapkan yaitu *Min-Max Normalization*, yang menyesuaikan nilai fitur agar berada dalam kisaran 0 hingga 1. Hasil normalisasi dapat dilihat pada Tabel 5.

3.4. *Splitting Data*

Setelah proses normalisasi data, tahap selanjutnya yaitu memisahkan data ke dalam dua kelompok, data latih (*training set*) dan data uji (*testing set*). Pemisahan ini bertujuan untuk mengukur kemampuan model dlam mengenali data baru yang belum pernah ditemui, sehingga dapat dievaluasi tingkat generalisasinya. Dalam penelitian ini, pemisah dilakukan dengan memanfaatkan fungsi *train_test_split* dari pustaka *Scikit-Learn.* Gambar 3 merupakan codingan dari *solitting data.*

Tabel 4. Hasil *Preprocessing*

Menu	*Energy (kkal)*	*Protein (g)*	*Fat (g)*	*Carbohydrates (g)*	*Dietary Fiber (g)*	*Label*
nasi tim + Perkedel Tempe Wortel	390	6.5	12.5	59	5.5	Seimbang
nasi tim ayam tomat wortel	200	6	4	38	2	Tidak Seimbang
nasi tim bayam	418	2.1	0.2	21.8	0.5	Tidak Seimbang
nasi tim daging	649	5.4	4.8	21.5	0.2	Berlebihan
nasi tim wortel kentang	439	2	0.2	23.2	0.7	Berlebihan
.	.	.	.	.	.	
.	.	.	.	.	.	
.	.	.	.	.	.	
Nasi tim ikan	320	8	6	65	4	Tidak Seimbang

Tabel 5. Hasil Normalisasi Data

Energy (kkal)	*Protein(g)*	*Fat(g)*	*Carbohydrates(g)*	*Dietary Fiber (g)*
0.300175	0.239382	0.261053	0.750636	0.647059
0.134380	0.220077	0.082105	0.483461	0.235294
0.324607	0.069498	0.082105	0.277354	0.058824
0.526178	0.196911	0.098947	0.277354	0.023529
0.342932	0.065637	0.002105	0.295165	0.082353

```
for split in split_options:
    print(f"\n Rasio Split: {1 - split:.0%} : {split:.0%}")
    X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=split, stratify=y, random_state=42)
```

Gambar 3. *Splitting Data*

3.5. Evaluasi Model

Tahap ini akan menggunakan python yang memberikan hasil seperti *accuracy*, *recall*, *precision*, dan *f1-score*. Berdasarkan model random forest yang telah dibangun dan dilatih pada tahap permodelan (setelah proses *splitting data*) , model ini kemudian dievaluasi performanya terhadap setiap *splitting data* yang diujikan dan disajikan pada Tabel 6.

Tabel 6. Evaluasi Model

Splitting Data	Akurasi	*Precision*	*Recall*	*F1- score*
80 : 20	89%	93%	93%	92%
70 : 30	85%	83%	90%	85%
50 : 50	90%	93%	82%	85%

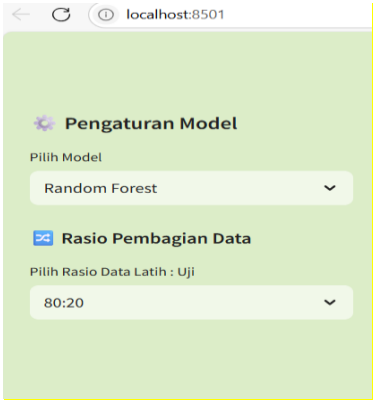
Nilai akurasi, presisi, *recall*, dan *f1-score* diperoleh melalui hasil pengujian menggunakan *Confusion Matrix* pada data uji. Proses ini bertujuan untuk mengukur sejauh mana model mampu mengklasifikasikan data baru secara tepat, serta memastikan bahwa hasil yang diperoleh tidak hanya bergantung pada data pelatihan. Berdasarkan hasil pada tabel tersebut, diketahui bahwa Random Forest memberikan performa paling optimal saat digunakan dengan rasio pembagian data 50:50, dengan akurasi mencapai 90%, *precision* 93%, *recall* 82%, dan *f1-score* 85%. Ini menjadikan Random Forest sebagai model terbaik dalam penelitian ini.

3.6. Implementasi GUI (Stramlit)

Proses ini mempermudah pengguna dalam melakukan klasifikasi menu makanan, sistem dikembangkan dalam bentuk aplikasi berbasis web menggunakan framework Streamlit. Aplikasi ini menyediakan antarmuka interaktif yang intuitif dan ramah pengguna.

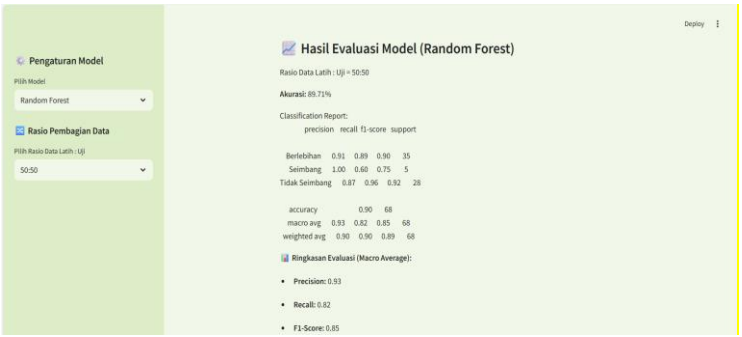
Gambar 4. Tampilan Unggah File

Gambar 4 menunjukkan antarmuka awal aplikasi, di mana pengguna dapat mengunggah dua file penting, yaitu data menu makanan dan standar gizi balita dalam format Excel. Kedua file ini menjadi input utama yang akan diproses oleh sistem untuk dilakukan klasifikasi.



Gambar 5. Tampilan Pengaturan Model dan Rasio Pembagian Data

Gambar 5 menampilkan pengaturan algoritma dan rasio pembagian data. Setelah file diunggah, pengguna dapat menentukan rasio pembagian data untuk pelatihan dan pengujian (80:20, 70:30, atau 50:50). Pilihan ini memberikan fleksibilitas bagi pengguna untuk mencoba berbagai konfigurasi dan membandingkan performa model.



Gambar 6. Tampilan Evaluasi Model

Gambar 6 menampilkan metrik evaluasi model, seperti akurasi, *precision*, *recall*, dan *f1-score*. Metrik ini membantu pengguna memahami seberapa baik model bekerja dalam mengklasifikasikan data yang diberikan. Berikut tabel evaluasi model pada Tabel 7

Tabel 7. Evaluasi Model

Splitting Data	Akurasi	*Precision*	*Recall*	*F1- score*
80 : 20	89%	93%	93%	92%
70 : 30	85%	83%	90%	85%
50 : 50	90%	93%	82%	85%

Gambar 7. Tampilan *Confusion Matrix*

Gambar 7 menampilkan tampilan *confusion matrix* yang dihasilkan oleh aplikasi klasifikasi berbasis Streamlit. Setelah pengguna memilih algoritma dan rasio pembagian data, sistem secara otomatis melakukan pelatihan model dan menampilkan *confusion matrix* sebagai visualisasi hasil klasifikasi. Hasil confusion matrix, bisa dilihat pada Tabel 8.

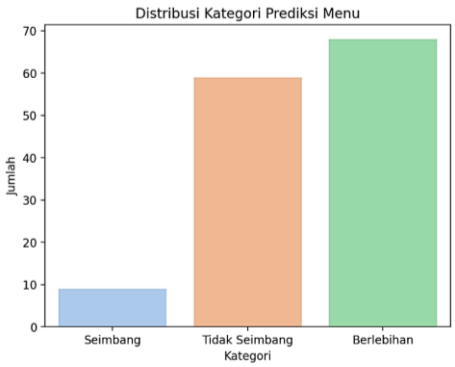
Tabel 8. Hasil *Confusion Matrix*

Aktual/ Prediksi	Seimbang	Tidak Seimbang	Berlebihan
Seimbang	3	0	2
Tidak Seimbang	0	27	1
Berlebihan	0	4	31

Tabel 9. Hasil Prediksi Data

No	Menu	Label	Prediksi
0	Nasi tim +Perkedel tempe wortel	Seimbang	Seimbang
18	Bubur kentang krim ayam	Seimbang	Seimbang
20	Mie kuah kari sayuran	Seimbang	Seimbang
…	…	…	…
64	Bubur manado (Tinutuan)	Seimbang	Seimbang

Tabel 9 menampilkan tabel hasil klasifikasi menu makanan dan ringkasan jumlah kategori prediksi berdasarkan *input* yang telah diunggah dan model yang dipilih. Setiap menu akan diberikan label “Seimbang”, “Tidak Seimbang”, atau “Berlebihan” sesuai dengan hasil prediksi model *machine learning*.



Gambar 8. Tampilan *Bar Chart*

Gambar 8 menampilkan hasil klasifikasi dalam bentuk grafik batang (*bar chart*) yang memperlihatkan distribusi jumlah menu makanan pada setiap kategori: Seimbang, Tidak, Seimbang, dan Berlebihan apabila sudah mengatur pengaturan model sesuai dengan rasio yang diinginkan. Dari 136 menu makanan terdapat 9 menu yang termasuk kategori seimbang, 59 menu tidak seimbang, dan 68 menu berlebihan

4. DISKUSI

Hasil penelitian menunjukkan bahwa algoritma Random Forest mampu mengklasifikasikan komposisi menu makanan olahan terhadap standar gizi balita dengan akurasi 90%. Tingginya performa ini disebabkan oleh kemampuan Random Forest dalam menggabungkan banyak pohon keputusan untuk menghasilkan prediksi yang stabil dan mengurangi risiko *overfitting*. Proses *feature selection* dan normalisasi yang dilakukan

sebelumnya juga berkontribusi dalam meningkatkan akurasi karena memastikan setiap fitur memiliki skala yang seimbang dan relevan terhadap model.

Hasil ini sejalan dengan penelitian [11] serta [19] yang menunjukkan efektivitas algoritma Random Forest dalam klasifikasi status gizi dengan tingkat akurasi di atas 85%. Namun, penelitian ini memiliki perbedaan fokus, yaitu mengklasifikasikan komposisi menu makanan berdasarkan standar AKG 2019, bukan berdasarkan data *antropometri* seperti pada penelitian terdahulu. Perbedaan konteks ini menunjukkan bahwa Random Forest tidak hanya efektif untuk deteksi status gizi, tetapi juga relevan untuk analisis kepatuhan gizi makanan olahan terhadap standar nutrisi balita.

Penelitian ini berkontribusi pada pengembangan sistem pemantauan gizi berbasis *data mining* yang dapat membantu orang tua dan tenaga kesehatan dalam menilai keseimbangan nutrisi menu anak secara cepat dan objektif. Kelebihan penelitian ini terletak pada penerapan model yang praktis dan dapat diintegrasikan ke dalam aplikasi berbasis web. Namun, keterbatasannya adalah jumlah *dataset* yang terbatas dan variasi nutrisi yang belum mencakup seluruh zat gizi mikro penting. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas jumlah data, menambahkan parameter mikronutrien, dan mengeksplorasi algoritma lain untuk meningkatkan akurasi model.

5. KESIMPULAN

Penelitian ini menunjukkan penerapan pendekatan *data mining* dengan algoritma Random Forest untuk mengklasifikasikan menu makanan olahan terhadap standar gizi balita. Melalui tahap *preprocessing* yang meliputi *data cleaning*, *labeling*, *feature selection*, dan normalisasi, sebanyak 136 menu berhasil dianalisis. Hasil klasifikasi menunjukkan bahwa hanya 9 menu yang termasuk kategori seimbang, 59 menu tidak seimbang, dan 68 menu berlebihan. Algoritma Random Forest menghasilkan performa terbaik dengan akurasi sebesar 90%, serta nilai presisi, *recall*, dan *f1-score* yang konsisten pada setiap kategori.

Hasil penelitian ini membuktikan bahwa algoritma Random Forest dapat dimanfaatkan secara efektif sebagai alat bantu dalam pemantauan gizi balita. Implementasi model dalam aplikasi berbasis Streamlit memberikan solusi praktis dan efisien bagi orang tua, pengasuh, maupun tenaga kesehatan untuk mengevaluasi keseimbangan gizi menu makanan balita. Penelitian selanjutnya disarankan untuk memperluas *dataset* serta menambahkan parameter gizi lain guna meningkatkan akurasi dan penerapan model dalam konteks yang lebih luas.

REFERENCES

[1] P. S. Pribadi, N. E. Aristina, and K. Sianipar, "Faktor Yang Berhubungan Dengan Status Gizi Balita Menurut BB/U Puspita," *Ensiklopedia J.*, vol. 6, no. 2, pp. 76–82, 2024.

[2] L. Lolita *et al.*, "Upaya Pencegahan dan Penanganan Stunting, Wasting, Underweight pada Satuan Pendidikan Anak Usia Dini," *J. Surya Masy.*, vol. 6, no. 2, p. 167, 2024, doi: 10.26714/jsm.6.2.2024.167-173.

[3] N. O. Syaqhasdy, Z. Martha, N. Amalita, and D. Fitria, "Classification of Nutrition Problems for Indonesian Toddler With Decision Tree Algorithm C4.5," *UNP J. Stat. Data Sci.*, vol. 1, no. 5, pp. 413–419, 2023, doi: 10.24036/ujsds/vol1-iss5/98.

[4] M. Mujayanto and E. R. Pratiwi, "The Correlation between Macronutrient Intake and Physical Activity with Overnutrition among Fifth-Grade Students at Banjarbendo State Elementary," *Amerta Nutr.*, vol. 8, no. 2SP, pp. 31–40, 2024, doi: 10.20473/amnt.v8i2SP.2024.31-40.

[5] M. H. A. Syahroni, N. Astuti, V. Indrawati, and R. Ismawati, "Faktor-Faktor yang Mempengaruhi Kebiasaaan Makan Anak Usia Pasekolah (4-6 Tahun) Ditinjau dari Capaian Gizi Seimbang," *J. Tata Boga*, vol. 10, no. 1, pp. 12–22, 2021, [Online]. Available: https://ejournal.unesa.ac.id/index.php/jurnal-tata-boga/

[6] M. Nur, I. Bahsur, S. Raodhah, S. Alam, and Z. Fadhilah, "Hubungan Kepatuhan Ibu Berkunjung Ke Posyandu Dengan Status Gizi Balita Di Kelurahan Mawang Kecamatan Somba Opu Kabupaten Gowa Tahun 2022," vol. 10, no. 2277, pp. 16–17, 2022.

[7] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, doi: 10.58602/jima-ilkom.v3i1.26.

[8] Arifin Yusuf Permana, Hari Noer Fazri, M.Fakhrizal Nur Athoilah, Mohammad Robi, and Ricky Firmansyah, "Penerapan Data Mining Dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *J. Ilm. Tek. Inform. dan Komun.*, vol. 3, no. 2, pp. 27–41, 2023, doi: 10.55606/juitik.v3i2.472.

[9] Citra Nursihah, Zaehol Fatah, and Rizki Hidayaturrochman, "Penerapan Data Mining Untuk Penilaian Tes Hadrah Di Pesantren Salafiyah Syafi'Iyah Menggunakan Metode Random Forest," *J. Ris. Tek. Komput.*, vol. 2, no. 1, pp. 40–45, 2025, doi: 10.69714/jfe1c445.

[10] S. N. Syafa Iswahyudi and R. Eka Putra, "Sistem Deteksi Stunting pada Balita Berbasis Web Menggunakan Metode Random Forest," *J. Informatics Comput. Sci.*, vol. 6, no. 03, pp. 755–764, 2025,

doi: 10.26740/jinacs.v6n03.p755-764.
[11] P. Handayani and A. Charis Fauzan, "Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest Putri," *Media Online)*, vol. 4, no. 6, pp. 3064–3072, 2024, doi: 10.30865/klik.v4i6.1909.
[12] S. Larissa, Z. Lewoema, and P. T. Prasetyaningrum, "Implementasi Data Mining Pada Klasifikasi Status Gizi Bayi Dengan Metode Decision Tree CHAID (Studi Kasus : Puskesmas Godean 1 Yogyakarta)," vol. 5, no. 1, pp. 61–74, 2024, doi: 10.51519/journalita.v5i1.538.
[13] A. W. Septyanto, H. L. Hariyanto, and H. Permatasari, "Utilizing The SVM Approach For Toddler Nutrition Status Classification Based On Anthropometric Measurements," vol. 10, no. 4, 2023.
[14] I. Rahmadi, D. T. Mareta, and D. Fithriyani, "Tingkat Kecukupan Energi dan Zat Gizi Makro Mahasiswa Tahun ke-3 Program Studi Teknologi Pangan ITERA," *J. Sci. Technol. Virtual Sci.*, vol. 1, no. 1, pp. 44–50, 2021.
[15] P. P. Allorerung, A. Erna, M. Bagussahrir, and S. Alam, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, 2024, doi: 10.14421/jiska.2024.9.3.178-191.
[16] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
[17] R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning," *J. Sains dan Inform.*, vol. 9, no. April, pp. 19–28, 2023, doi: 10.34128/jsi.v9i1.622.
[18] R. Irfannandhy, L. B. Handoko, and N. Ariyanto, "Analisis Performa Model Random Forest dan CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes," *Edumatic J. Pendidik. Inform.*, vol. 8, no. 2, pp. 714–723, 2024, doi: 10.29408/edumatic.v8i2.27990.
[19] F. Widyawati *et al.*, "Classification Of Toddler Nutritional Status Using Support Vector Machine And Random Forest Techniques With Optimal," vol. 5, no. 6, pp. 1893–1904, 2024.
[20] D. Lopez-Bernal, D. Balderas, P. Ponce, M. Rojas, and A. Molina, "Implications of Artificial Intelligence Algorithms in the Diagnosis and Treatment of Motor Neuron Diseases—A Review," *Life*, vol. 13, no. 4, 2023, doi: 10.3390/life13041031.
[21] H. N. Irmanda and Ria Astriratma, "Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 915–922, 2020, doi: 10.29207/resti.v4i5.2313.
[22] M. Fatmawati, B. A. Herlambang, and N. Q. Nada, "Random Forest Algorithm for Toddler Nutritional Status Classification Website," *J. Appl. Informatics Comput.*, vol. 8, no. 2, pp. 428–433, 2024, doi: 10.30871/jaic.v8i2.8463.
[23] C. P. W. Kase and S. Y. J. Prasetyo, "Analisis Faktor Risiko Stunting pada Balita di Desa Kesetnana Menggunakan Metode Random Forest," *J. Indones. Manaj. Inform. dan Komun.*, vol. 6, no. 3, pp. 1556–1566, 2025, doi: 10.63447/jimik.v6i3.1449.