

Evaluasi Komparatif Algoritma *Machine Learning* Menggunakan *SMOTENC* pada Klasifikasi Status Gizi Balita Berbasis Data Antropometri

Erip Suratno¹, Adhi Kusnadi², Zeldi Suryady³

^{1,2,3} Program Magister Informatika, Universitas Nusa Putra, Sukabumi, Indonesia
erip.suratno_mif24@nusaputra.ac.id

Abstrak

Ketidakeimbangan distribusi kelas pada data status gizi balita dapat membuat model klasifikasi tampak akurat, tetapi kurang mampu mengenali kelas minoritas. Penelitian ini mengevaluasi *Support Vector Machine*, *K-Nearest Neighbors*, *Naive Bayes*, *Decision Tree*, dan *Random Forest* untuk klasifikasi multikelas status gizi balita berbasis data antropometri. Dataset berjumlah 3.716 data Posyandu Desa Ujunggenteng tahun 2025 dengan fitur jenis kelamin, usia, berat badan, dan tinggi badan serta empat kelas target : kurang gizi, gizi baik, risiko gizi lebih, dan gizi lebih. Label diperlakukan sebagai label operasional dalam dataset, bukan hasil penghitungan ulang *z-score WHO*/Kementerian Kesehatan. Eksperimen menggunakan pembagian data latih-uji terstratifikasi 80:20, *5-fold Stratified K-Fold Cross-Validation*, pencarian kandidat parameter terbatas, *preprocessing* numerik-kategorik, dan *SMOTENC* hanya pada data latih atau *fold* pelatihan. Hasil *cross-validation* menunjukkan *SVM* memperoleh *CV F1 Macro* tertinggi sebesar 0,829578. Pada data uji, *Random Forest* memperoleh *F1 Macro* tertinggi sebesar 0,827132, sedangkan *KNN* memperoleh akurasi tertinggi sebesar 0,956989 dan *balanced accuracy* tertinggi sebesar 0,883393. Pemilihan model perlu mempertimbangkan *F1 Macro*, *balanced accuracy*, *confusion matrix*, dan performa per kelas. Berdasarkan *cross-validation*, *SVM* ditetapkan sebagai model final, sedangkan *Random Forest* dilaporkan terbaik pada data uji berdasarkan *F1 Macro*.

Kata kunci: status gizi balita; *machine learning*; *SMOTENC*; klasifikasi multikelas; data antropometri

Abstract

Class imbalance in under-five nutritional status data can make classification models appear accurate while limiting minority-class recognition. This study evaluated Support Vector Machine, K-Nearest Neighbors, Naive Bayes, Decision Tree, and Random Forest for multiclass classification of under-five nutritional status using anthropometric data. The dataset contained 3,716 Posyandu Desa Ujunggenteng records from 2025, with sex, age, weight, and height as features and four target classes: undernutrition, good nutritional status, risk of overnutrition, and overnutrition. Labels were treated as operational records, not as recalculated WHO/Indonesian Ministry of Health z-score standards. The experiment used an 80:20 stratified train-test split, 5-fold Stratified K-Fold Cross-Validation, limited candidate-based parameter search, numerical-categorical preprocessing, and SMOTENC only on training data or training folds. Cross-validation showed that SVM achieved the highest CV Macro F1-score of 0.829578. On the test set, Random Forest obtained the highest Macro F1-score of 0.827132, while KNN achieved the highest accuracy of 0.956989 and balanced accuracy of 0.883393. Model selection should consider Macro F1-score, balanced accuracy, confusion matrix, and class-wise performance. Based on cross-validation, SVM was selected as the final model, while Random Forest was reported as the best test-set model based on Macro F1-score.

Keywords: *under-five nutritional status; machine learning; SMOTENC; multiclass classification; anthropometric data*

I. PENDAHULUAN

Pemantauan status gizi balita merupakan proses penting karena gangguan pertumbuhan pada usia

dini dapat memengaruhi kualitas kesehatan anak dalam jangka panjang. Penguatan pemantauan kesehatan anak secara digital dapat membantu penyajian indikator pertumbuhan secara lebih cepat, terstruktur, dan konsisten [1]. Namun, validitas

penilaian status gizi tetap bergantung pada kualitas pengukuran antropometri, kelengkapan data, dan ketelitian pencatatan indikator pertumbuhan [2]. Pada *dataset* Posyandu yang digunakan dalam penelitian ini, permasalahan utama tidak hanya berkaitan dengan ketersediaan data, tetapi juga dengan distribusi kelas yang sangat tidak seimbang. *Dataset* penelitian berjumlah 3.716 data, dengan kelas gizi baik mencapai 3.220 data atau 86,65%, sedangkan kelas gizi lebih hanya 31 data atau 0,83%. Ketimpangan tersebut dapat menyebabkan model klasifikasi tampak memiliki akurasi tinggi karena dominasi kelas mayoritas, tetapi kurang mampu mengenali kelas minoritas yang penting dalam proses *skrining* status gizi.

Pemanfaatan *machine learning* dalam kajian status gizi anak telah banyak dilakukan dengan fokus, data, dan target klasifikasi yang berbeda. Algoritma *machine learning* telah digunakan untuk memprediksi *undernutrition* pada anak usia di bawah lima tahun berdasarkan karakteristik kesehatan dan data anak [3]. Pendekatan klasifikasi berbasis *machine learning* juga telah digunakan untuk mengidentifikasi *stunting* pada anak usia di bawah lima tahun dengan membandingkan performa beberapa model klasifikasi [4]. Prediksi *stunting* menggunakan teknik *machine learning* menunjukkan bahwa performa model dapat dipengaruhi oleh karakteristik fitur, target, dan skema validasi yang digunakan [5]. Dalam konteks Indonesia, *benchmarking* algoritma *machine learning* telah digunakan untuk mengevaluasi risiko *stunting*, sehingga menunjukkan pentingnya perbandingan model dalam masalah kesehatan anak [6]. Pendekatan *ensemble machine learning* telah digunakan untuk meningkatkan prediksi status gizi atau *stunting* pada anak [7], [8]. Pendekatan *deep learning* dan *explainable machine learning* juga digunakan untuk memprediksi status gizi serta mengidentifikasi faktor yang berkaitan dengan *undernutrition* pada anak [9], [10], [11]. Pada sisi metodologis, ketidakseimbangan distribusi kelas dapat memengaruhi proses klasifikasi karena model cenderung lebih mudah mengenali kelas mayoritas dibandingkan kelas minoritas [12]. Teknik *oversampling* berbasis *SMOTE* relevan untuk mengurangi ketimpangan kelas pada data latih, tetapi pembentukan sampel sintesis tetap perlu dikendalikan agar tidak menghasilkan evaluasi yang terlalu optimistis atau menurunkan kualitas klasifikasi [13]. Pada data status gizi balita yang tidak seimbang, *RF*, *XGBoost*, dan *SVM* telah dibandingkan dengan penerapan *SMOTE*, dan *RF* pada data hasil *SMOTE* dilaporkan memperoleh kinerja terbaik [14]. Klasifikasi status gizi balita berbasis usia, *BMI*, berat badan, dan panjang atau

tinggi badan pernah dievaluasi dengan *NB*, *LDA*, *DT*, *KNN*, *RF*, dan *SVM*, dengan *RF* sebagai metode yang paling direkomendasikan [15]. Prediksi *stunting* multikelas berbasis data antropometri telah dikaji melalui kerangka *AutoML* yang membandingkan beberapa model dengan dan tanpa *SMOTE*, sehingga relevan sebagai pembanding metodologis untuk masalah kelas tidak seimbang [16]. Klasifikasi *stunting* pada anak di bawah lima tahun telah dievaluasi menggunakan *XGBoost*, *Logistic Regression*, *Random Forest*, *Gradient Boosting*, dan *KNN* [17]. Pendekatan *hybrid* untuk prevalensi *stunting* telah menggabungkan *SVM*, regresi linear, dan *K-Medoids* teroptimasi dalam satu kerangka analisis [18]. Perbandingan *XGBoost* dan *Random Forest* pada deteksi *stunting* menunjukkan bahwa kedua model digunakan untuk menangani masalah klasifikasi *stunting* dengan data tidak seimbang dan data seimbang [19]. Perbandingan *SVM* dan *Random Forest* dengan optimasi *Grid Search* menunjukkan peningkatan kinerja *SVM* dan kinerja tinggi *Random Forest* pada prediksi *stunting* [20].

Berdasarkan penelitian terdahulu, *machine learning* telah banyak digunakan untuk memprediksi *undernutrition*, *stunting*, dan status gizi anak. Namun, sebagian besar penelitian masih berfokus pada target klasifikasi tertentu, seperti *stunting* atau *undernutrition*, sehingga evaluasi komparatif untuk klasifikasi multikelas status gizi balita berbasis data antropometri Posyandu lokal dengan ketidakseimbangan kelas ekstrem masih perlu diperkuat. Kebaruan penelitian ini terletak pada evaluasi komparatif lima algoritma *machine learning*, yaitu *SVM*, *KNN*, *Naive Bayes*, *Decision Tree*, dan *Random Forest*, dengan penerapan *SMOTENC* secara terkontrol hanya pada data latih atau *fold* pelatihan untuk mencegah *data leakage*. Penelitian ini bertujuan membandingkan kinerja kelima algoritma dalam klasifikasi multikelas status gizi balita berdasarkan skema evaluasi yang konsisten. Kontribusi penelitian ini terletak pada penyusunan *pipeline* eksperimen yang dapat ditelusuri, penanganan ketidakseimbangan kelas secara terkontrol, serta pemisahan interpretasi antara model final berdasarkan validasi silang dan model dengan performa terbaik pada data uji.

II. METODE PENELITIAN

A. Dataset dan Variabel

Data yang digunakan dalam penelitian ini berasal dari *dataset* antropometri balita pada Posyandu Desa Ujunggenteng tahun 2025. Total *dataset* berjumlah 3.716 baris data, tidak memiliki *missing value* awal, tidak memiliki duplikasi, dan tidak ada baris yang

dibuang pada tahap validasi. Fitur masukan terdiri atas jenis kelamin, usia dalam bulan, berat badan dalam kilogram, dan tinggi badan dalam sentimeter. Variabel target adalah status gizi balita yang terdiri atas empat kelas, yaitu 1 = kurang gizi, 2 = gizi baik, 3 = risiko gizi lebih, dan 4 = gizi lebih. Penggunaan data antropometri sebagai dasar pemodelan perlu memperhatikan kualitas pengukuran dan konsistensi pencatatan karena kualitas data antropometri berpengaruh terhadap keandalan pemantauan status gizi anak [2].

Tabel 1 menyajikan struktur data yang digunakan dalam eksperimen terdiri atas jenis kelamin, usia, berat badan, tinggi badan, dan status gizi. Tabel ini berfungsi untuk menunjukkan bentuk data mentah yang menjadi dasar pemodelan, bukan sebagai hasil prediksi model.

Tabel 1. Data antropometri balita

No	Jenis kelamin	Usia (bulan)	Berat (kg)	Tinggi (cm)	Status gizi
1	1	59	15	102	2
2	1	57	14	100	2
3	1	56	15,4	104	2
4	2	56	17,2	104	2
5	2	58	14,4	100	2
...
3.714	1	0	4,2	53	2
3.715	1	0	2,8	49	2
3.716	2	0	3,8	50	3

B. Penentuan Label dan Perlindungan Data

Label status gizi pada penelitian ini berasal dari variabel status_gizi yang telah tersedia dalam *dataset* Posyandu. Penelitian ini tidak menghitung ulang nilai *z-score* antropometri berdasarkan standar WHO/Kementerian Kesehatan, sehingga keluaran model dipahami sebagai klasifikasi terhadap label operasional yang tercatat dalam *dataset*. Hasil klasifikasi tidak dimaksudkan untuk menggantikan penentuan status gizi berbasis standar resmi, melainkan digunakan untuk mengevaluasi kemampuan algoritma *machine learning* dalam mempelajari pola klasifikasi dari data yang tersedia.

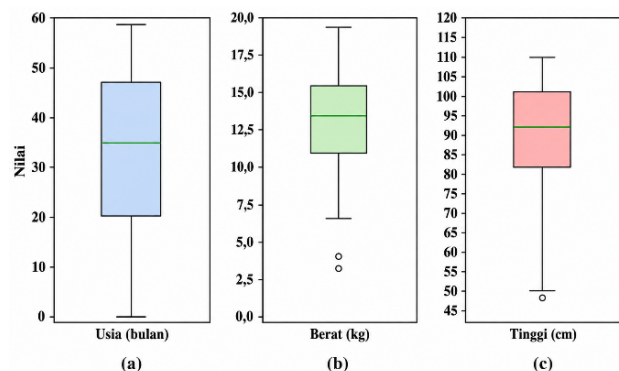
Dataset eksperimen yang diproses dalam pemodelan hanya memuat atribut jenis kelamin, usia, berat badan, tinggi badan, dan status gizi. Atribut identitas pribadi, seperti nama, alamat, nomor induk, atau nomor identitas, tidak digunakan dalam *pipeline* eksperimen. Mengingat data balita termasuk data sensitif, hasil penelitian disajikan dalam bentuk agregat dan tidak menampilkan data individu. Penggunaan data untuk implementasi lanjutan tetap memerlukan izin dari pengelola data serta kepatuhan terhadap ketentuan perlindungan data pribadi yang berlaku.

C. Distribusi Kelas

Tabel 2 menyajikan distribusi kelas dengan ketidakseimbangan yang tinggi. Kelas gizi baik mendominasi 3.220 data atau 86,65%, sedangkan kelas gizi lebih hanya 31 data atau 0,83%. Rasio antara kelas mayoritas dan kelas minoritas mencapai sekitar 104:1. Evaluasi model tidak cukup hanya menggunakan akurasi keseluruhan karena model dapat terlihat memiliki performa tinggi akibat dominasi kelas mayoritas. Pada data tidak seimbang, proses klasifikasi perlu memperhatikan risiko dominasi kelas mayoritas agar kinerja model pada kelas minoritas tidak terabaikan [12].

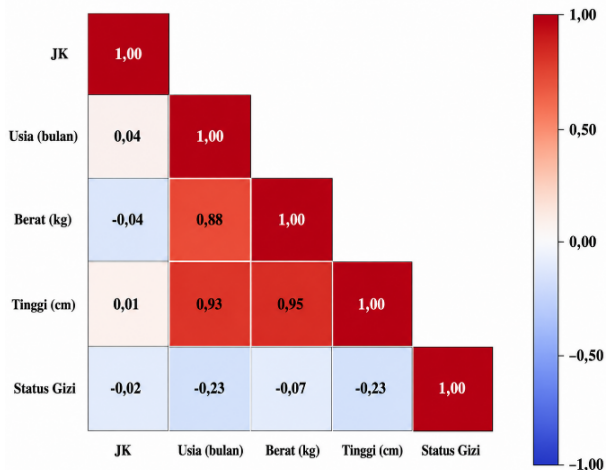
Tabel 2. Distribusi kelas status gizi balita

Kode	Kelas	Jumlah	Persentase (%)
1	Kurang gizi	74	1,99
2	Gizi baik	3.220	86,65
3	Risiko gizi lebih	391	10,52
4	Gizi lebih	31	0,83



Gambar 1. Visualisasi *boxplot* untuk fitur numerik: (a) Usia (bulan), (b) Berat (kg), dan (c) Tinggi (cm)

Gambar 1 menunjukkan sebaran variabel usia, berat badan, dan tinggi badan balita. Variabel usia menunjukkan rentang distribusi yang cukup luas, sedangkan berat badan dan tinggi badan cenderung terkonsentrasi pada rentang tertentu. Pada variabel berat badan dan tinggi badan ditemukan beberapa nilai pencilan di bagian bawah distribusi. Nilai pencilan tersebut tetap perlu diperhatikan karena dapat merepresentasikan kondisi antropometri balita yang berbeda dari mayoritas data dan berpotensi berpengaruh terhadap proses klasifikasi status gizi.



Gambar 2. Heatmap korelasi spearman antarvariabel

Gambar 2 menunjukkan heatmap korelasi Spearman antarvariabel pada dataset. Fitur antropometri menunjukkan korelasi positif yang kuat, terutama antara berat badan dan tinggi badan sebesar 0,945, usia dan tinggi badan sebesar 0,929, serta usia dan berat badan sebesar 0,885. Temuan ini menunjukkan bahwa usia, berat badan, dan tinggi badan saling berkaitan erat dalam struktur data yang digunakan. Sebaliknya, jenis kelamin memiliki korelasi yang sangat rendah terhadap variabel lainnya. Variabel status gizi menunjukkan korelasi negatif lemah terhadap usia dan tinggi badan, serta korelasi negatif sangat lemah terhadap berat badan. Hasil korelasi ini digunakan sebagai analisis eksploratif awal untuk memahami hubungan antarvariabel dan tidak ditafsirkan sebagai hubungan kausal.

D. Pipeline Eksperimen, SMOTENC, dan Validasi

Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan stratifikasi target dan random state 42. Pembagian ini menghasilkan 2.972 data latih dan 744 data uji. Validasi model dilakukan menggunakan 5-fold Stratified K-Fold Cross-Validation agar proporsi kelas pada setiap fold tetap terjaga. Seluruh kandidat model dievaluasi menggunakan skema pipeline yang sama sehingga perbandingan performa tidak dipengaruhi oleh perbedaan perlakuan preprocessing, oversampling, atau pembagian data. Pemisahan data latih dan data uji, validasi silang, feature scaling, dan pipeline merupakan tahapan penting dalam evaluasi model machine learning agar proses pelatihan dan pengujian berjalan lebih terkontrol [21].

SMOTENC diterapkan hanya pada data latih atau fold pelatihan untuk menangani ketidakseimbangan kelas tanpa mencampurkan informasi dari validation fold maupun data uji ke dalam proses pembentukan

sampel sintetis. Implementasi program menggunakan *imblearn.pipeline.Pipeline* dengan urutan *SMOTENC*, *preprocessing*, dan *classifier*. *Validation fold* dan data uji tidak di-resampling agar hasil evaluasi tetap mencerminkan performa model pada distribusi data yang tidak diubah. Penanganan *imbalance* dalam penelitian ini mengacu pada prinsip *oversampling* berbasis *SMOTE*, yaitu pembentukan sampel sintetis untuk membantu mengurangi ketimpangan distribusi kelas pada data latih [12]. Penggunaan teknik *oversampling* pada data tidak seimbang perlu dikendalikan agar tidak menghasilkan evaluasi yang terlalu optimistis atau menurunkan kualitas klasifikasi [13].

Berdasarkan urutan fitur input sebelum preprocessing, variabel jenis kelamin berada pada indeks 0, sehingga parameter *categorical_features* indeks 0 digunakan pada *SMOTENC*. Nilai *k_neighbors=3* digunakan karena jumlah sampel pada kelas minoritas sangat terbatas, terutama kelas gizi lebih yang hanya memiliki 25 data pada data latih. Setelah proses *oversampling* pada data latih atau fold pelatihan, fitur numerik diproses menggunakan *SimpleImputer* dengan strategi median dan *StandardScaler*, sedangkan fitur kategorik diproses menggunakan *SimpleImputer* dengan strategi *most_frequent*. Dataset yang digunakan telah divalidasi tidak memiliki missing value awal, sehingga penempatan imputer berfungsi sebagai pengaman konsistensi pipeline. Jika pipeline diterapkan pada dataset lain yang memiliki missing value, proses imputasi sebelum *SMOTENC* perlu dipertimbangkan agar pembentukan sampel sintetis tetap valid.

Tabel 3 menyajikan pembagian data latih dan data uji dilakukan secara stratified sehingga proporsi kelas tetap terjaga. Kelas gizi lebih memiliki support yang sangat kecil, yaitu 25 data pada data latih dan 6 data pada data uji.

Tabel 3. Distribusi kelas data latih dan data uji

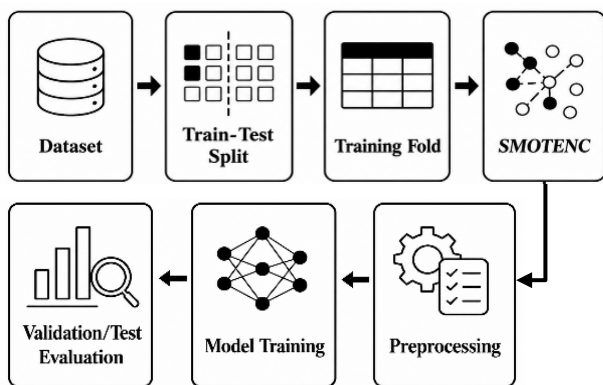
Kode	Kelas	Train	Test
1	Kurang gizi	59	15
2	Gizi baik	2.575	645
3	Risiko gizi lebih	313	78
4	Gizi lebih	25	6

Tabel 4 menyajikan skema eksperimen penelitian mencakup dataset, fitur, target, pembagian data, validasi, penanganan imbalance, parameter SMOTENC, aturan seleksi, dan runtime utama. Tabel ini berfungsi memastikan eksperimen dapat ditelusuri ulang sesuai program.

Tabel 4. Skema eksperimen penelitian

Komponen	Implementasi
Dataset	DATASET ML.xlsx, 3.716 data akhir

Fitur	JK, usia (bulan), berat (kg), tinggi (cm)
Target	Status gizi dengan 4 kelas
Pembagian data	<i>Train-test split</i> 80:20 dengan <i>stratify=y</i>
Validasi	<i>5-fold Stratified K-Fold Cross-Validation</i>
<i>Random state</i>	42
Penanganan <i>imbalance</i>	<i>SMOTENC</i> hanya pada data latih/ <i>fold</i> pelatihan
Parameter <i>SMOTENC</i>	<i>categorical_features</i> indeks 0, <i>k_neighbors=3</i> , <i>random_state=42</i>
Aturan seleksi	<i>CV F1 Macro</i> -> <i>Balanced Accuracy</i> -> <i>Accuracy</i>
<i>Runtime</i> utama	<i>Python 3.12.13</i> , <i>scikit-learn 1.6.1</i> , <i>imbalanced-learn 0.14.1</i>



Gambar 3. Alur *pipeline* eksperimen dan kontrol *data leakage*

Gambar 3 menunjukkan alur *pipeline* eksperimen dalam dua baris yang menggambarkan tahapan pembagian data, penerapan *SMOTENC* pada data latih atau *training fold*, *preprocessing*, pelatihan model, dan evaluasi pada *validation fold* serta data uji. *SMOTENC* diterapkan hanya pada data latih atau *training fold*, sedangkan *validation fold* dan data uji tidak di-*resampling*. Dengan skema ini, proses evaluasi tetap mencerminkan distribusi data yang tidak diubah dan risiko *data leakage* dapat diminimalkan.

E. Algoritma dan Konfigurasi Eksperimen

Lima algoritma yang dievaluasi dalam penelitian ini adalah *SVM*, *KNN*, *Naive Bayes*, *Decision Tree*, dan *Random Forest*. Pemilihan algoritma dilakukan untuk membandingkan performa model berbasis margin, kedekatan tetangga, probabilistik, pohon keputusan, dan *ensemble* pada klasifikasi multikelas status gizi balita. Prinsip evaluasi beberapa model dalam proyek *machine learning* perlu dilakukan dengan skema validasi yang konsisten agar perbandingan performa tidak dipengaruhi oleh perbedaan perlakuan eksperimen [21].

Konfigurasi model diperoleh melalui pencarian kandidat parameter terbatas yang dinyatakan eksplisit dalam program. Setiap kandidat dijalankan menggunakan *pipeline* yang sama, dievaluasi dengan *5-fold Stratified K-Fold Cross-Validation*, kemudian diranking berdasarkan *CV F1 Macro*, *CV balanced accuracy*, dan *CV accuracy*. Data uji tidak digunakan dalam proses pemilihan kandidat parameter agar evaluasi *holdout* tetap independen.

Ruang pencarian parameter pada penelitian ini dibatasi untuk menjaga keterulangan eksperimen dan menyesuaikan dengan tujuan utama penelitian, yaitu evaluasi komparatif lima algoritma pada skema *preprocessing* dan penanganan *imbalance* yang sama, bukan pencarian hiperparameter secara menyeluruh.

Ruang kandidat yang diuji meliputi *KNN* dengan nilai *n_neighbors* 3, 5, dan 7. *SVM* dengan *kernel RBF* pada nilai *C* sebesar 1,0, 3,0 dan 10,0 serta *kernel linear* pada nilai *C* sebesar 1,0. *Naive Bayes* dengan *var_smoothing* $1e-09$ sampai $1e-06$. *Decision Tree* dengan *criterion* *gini* dan *entropy* pada *max_depth* 5 dan 7, serta *Random Forest* dengan kombinasi *n_estimators* 200, 300, dan 500, *max_depth* *None*, 10, dan 12, serta *class_weight balanced_subsample*.

Tabel 5 menyajikan konfigurasi parameter terbaik tiap model berdasarkan evaluasi *cross-validation* dalam ruang kandidat yang telah ditetapkan. Prioritas *F1 Macro* digunakan karena *dataset* memiliki distribusi kelas yang sangat tidak seimbang, sehingga setiap kelas perlu diberi kontribusi evaluasi yang setara. *SVM* memperoleh *CV F1 Macro* tertinggi sebesar 0,829578 dengan konfigurasi *kernel RBF*, *C=10,0*, *gamma=scale*, dan *class_weight=balanced*.

Tabel 5. Konfigurasi parameter terbaik dari lima model pada evaluasi *cross-validation*

Model	Konfigurasi Terbaik	<i>CV F1 Macro</i>
<i>SVM</i>	<i>kernel=rbf</i> ; <i>C=10,0</i> ; <i>gamma=scale</i> ; <i>class_weight=balanced</i>	0,829578
<i>Random Forest</i>	<i>n_estimators=300</i> ; <i>max_depth=None</i> ; <i>min_samples_split=2</i> ; <i>min_samples_leaf=1</i> ; <i>class_weight=balanced_subsample</i>	0,804431
<i>KNN</i>	<i>n_neighbors=3</i> ; <i>weights=uniform</i> ; <i>metric=minkowski</i> ; <i>p=2</i>	0,797241
<i>Decision Tree</i>	<i>criterion=gini</i> ; <i>max_depth=7</i> ; <i>min_samples_split=10</i> ; <i>min_samples_leaf=5</i> ; <i>class_weight=balanced</i>	0,636993
<i>Naive Bayes</i>	<i>var_smoothing=1e-09</i>	0,312745

F. Metrik Evaluasi

Evaluasi model menggunakan *accuracy*, *precision*, *recall*, *F1-score*, *F1 Macro*, *balanced accuracy*, rerata *cross-validation*, dan simpangan baku *cross-validation*. Penggunaan metrik evaluasi, validasi silang, pemisahan data latih dan data uji, *pipeline* transformasi, serta *feature scaling* mengacu pada konsep evaluasi model *machine learning* yang menekankan pentingnya pemisahan data dan pengukuran performa secara sistematis [21]. Pada data tidak seimbang, penggunaan metrik selain akurasi perlu dilakukan karena akurasi keseluruhan dapat dipengaruhi oleh dominasi kelas mayoritas [12].

Notasi yang digunakan dalam (1) sampai (8) didefinisikan sebagai berikut. K menyatakan jumlah kelas, M menyatakan jumlah *fold* pada *cross-validation*, N menyatakan jumlah seluruh data uji, TP_k menyatakan *True Positive* pada kelas ke- k , FP_k menyatakan *False Positive* pada kelas ke- k , FN_k menyatakan *False Negative* pada kelas ke- k , s_i menyatakan skor evaluasi pada *fold* ke- i , dan \bar{s} menyatakan rerata skor *cross-validation*.

$$Accuracy = \frac{\sum_{k=1}^K TP_k}{N} \quad (1)$$

Persamaan (1), *accuracy* mengukur proporsi seluruh data yang berhasil diklasifikasikan dengan benar terhadap jumlah seluruh data uji.

$$Precision_k = \frac{TP_k}{TP_k + FP_k} \quad (2)$$

Persamaan (2), *precision* kelas ke- k mengukur ketepatan model ketika memprediksi suatu data sebagai kelas ke- k . Nilai *precision* yang rendah menunjukkan bahwa masih terdapat data dari kelas lain yang ikut diprediksi sebagai kelas tersebut.

$$Recall_k = \frac{TP_k}{TP_k + FN_k} \quad (3)$$

Persamaan (3), *recall* kelas ke- k mengukur kemampuan model dalam mengenali seluruh data aktual pada kelas ke- k .

$$F1_k = 2 \times \frac{Precision_k \times Recall_k}{Precision_k + Recall_k} \quad (4)$$

Persamaan (4), *F1-score* kelas ke- k merupakan rata-rata harmonik antara *precision* dan *recall*.

$$F1\ Macro = \frac{1}{K} \sum_{k=1}^K F1_k \quad (5)$$

Persamaan (5), *F1 Macro* menghitung rata-rata *F1-score* dari seluruh kelas dengan bobot yang sama.

$$Balanced\ Accuracy = \frac{1}{K} \sum_{k=1}^K Recall_k \quad (6)$$

Persamaan (6), *balanced accuracy* menghitung rata-rata *recall* dari seluruh kelas.

$$MeanCV = \bar{s} = \frac{1}{M} \sum_{i=1}^M s_i \quad (7)$$

Persamaan (7), *MeanCV* menghitung rerata skor evaluasi dari seluruh *fold* pada proses *cross-validation*.

$$StdCV = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (s_i - \bar{s})^2} \quad (8)$$

Persamaan (8), *StdCV* menunjukkan simpangan baku skor evaluasi antar-*fold*.

III. HASIL DAN PEMBAHASAN

A. Hasil Cross-Validation

Hasil *cross-validation* menunjukkan bahwa *SVM* memperoleh nilai tertinggi pada metrik seleksi utama. Aturan seleksi sejak awal memprioritaskan *CV F1 Macro*, *CV balanced accuracy*, dan *CV accuracy*.

Tabel 6 menyajikan *SVM* memperoleh *CV F1 Macro* tertinggi sebesar 0,829578 dengan simpangan baku 0,066709 dan *CV balanced accuracy* sebesar 0,908484 dengan simpangan baku 0,070782. *CV accuracy* sebesar 0,954244 dengan simpangan baku 0,006563. Tabel ini menjadi dasar utama pemilihan *SVM* sebagai model final penelitian berdasarkan aturan seleksi *cross-validation*. *Random Forest* menempati urutan kedua pada *CV F1 Macro*, disusul *KNN*, *Decision Tree*, dan *Naive Bayes*.

Tabel 6. Ringkasan hasil 5 *fold cross-validation*

Model	CV Accuracy	CV Bal. Acc.	CV Prec. Macro	CV F1 Macro
<i>SVM</i>	0,954244 ± 0,006563	0,908484 ± 0,070782	0,793050 ± 0,062581	0,829578 ± 0,066709
<i>Random Forest</i>	0,952896 ± 0,008694	0,840563 ± 0,047067	0,787037 ± 0,052341	0,804431 ± 0,042924
<i>KNN</i>	0,944487 ± 0,009603	0,860700 ± 0,048358	0,770718 ± 0,064636	0,797241 ± 0,052063
<i>Decision Tree</i>	0,863744 ± 0,021516	0,767002 ± 0,041163	0,607385 ± 0,074426	0,636993 ± 0,067112
<i>Naive Bayes</i>	0,565273 ± 0,015364	0,618338 ± 0,040104	0,326124 ± 0,005854	0,312745 ± 0,008359

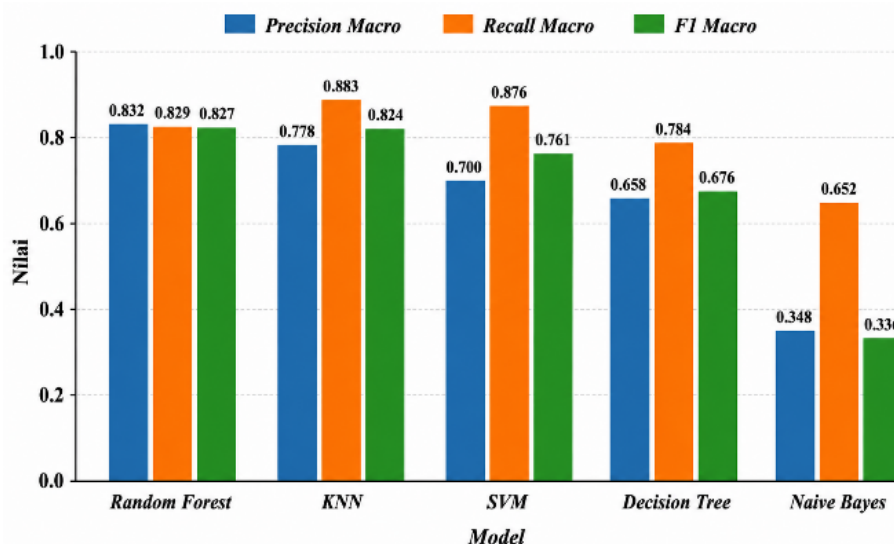
B. Hasil Evaluasi pada Data Uji

Pada data uji, pola hasil tidak sepenuhnya sama dengan *cross-validation*. Tabel 7 menyajikan *Random Forest* memperoleh *F1 Macro* tertinggi sebesar 0,827132 dan *precision macro* tertinggi sebesar 0,831591. *KNN* memperoleh *accuracy*

tertinggi sebesar 0,956989 dan *balanced accuracy* tertinggi sebesar 0,883393. *SVM* memperoleh *AUC Macro* tertinggi sebesar 0,981852 tetapi *F1 Macro* data ujinya sebesar 0,761414 lebih rendah dibandingkan *Random Forest* dan *KNN*.

Tabel 7. Ringkasan evaluasi model pada data uji

Model	Accuracy	Bal. Acc.	Prec. Macro	Recall Macro	F1 Macro	AUC Macro	Kappa
<i>Random Forest</i>	0,952957	0,829055	0,831591	0,829055	0,827132	0,980992	0,801650
<i>KNN</i>	0,956989	0,883393	0,778416	0,883393	0,824300	0,956529	0,830774
<i>SVM</i>	0,943548	0,876431	0,699712	0,876431	0,761414	0,981852	0,793947
<i>Decision Tree</i>	0,872312	0,783617	0,657627	0,783617	0,675782	0,915334	0,578345
<i>Naive Bayes</i>	0,586022	0,652445	0,348318	0,652445	0,336199	0,806354	0,195077



Gambar 4. Perbandingan *precision*, *recall*, dan *F1 macro* pada data uji

Gambar 4 menunjukkan perbandingan *precision macro*, *recall macro*, dan *F1 Macro* pada data uji untuk seluruh model yang dievaluasi. Visualisasi ini memperlihatkan bahwa *KNN* memiliki *recall macro* dan *balanced accuracy* yang tinggi, *Random Forest* memiliki *F1 Macro* dan *precision macro* tertinggi, sedangkan *SVM* tetap menunjukkan *recall macro* tinggi tetapi *precision macro* lebih rendah.

C. Analisis Per Kelas

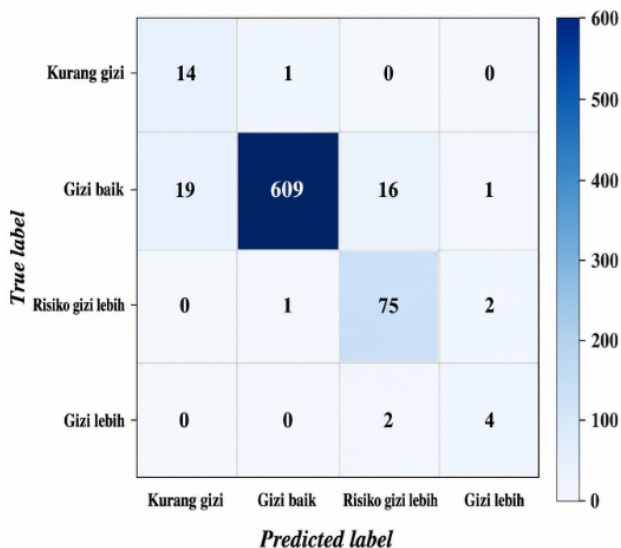
Analisis per kelas dilakukan untuk mengevaluasi kinerja tiga model kompetitif pada setiap kelas status gizi. Analisis ini penting karena performa pada kelas minoritas tidak cukup dinilai hanya melalui metrik agregat, tetapi juga perlu ditinjau berdasarkan distribusi kesalahan prediksi pada *confusion matrix*. Evaluasi per kelas dapat memberikan gambaran yang lebih rinci mengenai kemampuan model dalam mengenali setiap kategori status gizi, terutama pada kelas dengan jumlah data yang terbatas.

Tabel 8 menunjukkan bahwa *Random Forest* memperoleh *F1-score* sebesar 0,787879 pada kelas kurang gizi, 0,975194 pada kelas gizi baik, 0,818182 pada kelas risiko gizi lebih, dan 0,727273 pada kelas gizi lebih. Hasil ini menunjukkan bahwa *Random Forest* memiliki performa yang relatif lebih seimbang, terutama pada kelas minoritas seperti kurang gizi dan gizi lebih. Sementara itu, *SVM* memiliki *recall* tinggi pada kelas kurang gizi, tetapi *precision* lebih rendah. Kondisi tersebut menunjukkan bahwa *SVM* lebih sensitif dalam mengenali data aktual kurang gizi, namun cenderung menghasilkan lebih banyak prediksi kurang gizi yang berasal dari kelas lain. Pada kelas dengan jumlah *support* sangat kecil, perubahan satu prediksi benar atau salah dapat memengaruhi nilai *recall* dan *F1-score* secara signifikan.

Tabel 8. Kinerja per kelas pada tiga model kompetitif

Model	Kelas	Precision	Recall	F1	Support
KNN	Kurang gizi	0,571429	0,800000	0,666667	15
KNN	Gizi baik	0,988871	0,964341	0,976452	645
KNN	Risiko gizi lebih	0,839080	0,935897	0,884848	78
KNN	Gizi lebih	0,714286	0,833333	0,769231	6
SVM	Kurang gizi	0,424242	0,933333	0,583333	15
SVM	Gizi baik	0,996727	0,944186	0,969745	645
SVM	Risiko gizi lebih	0,806452	0,961538	0,877193	78
SVM	Gizi lebih	0,571429	0,666667	0,615385	6
Random Forest	Kurang gizi	0,722222	0,866667	0,787879	15
Random Forest	Gizi baik	0,975194	0,975194	0,975194	645
Random Forest	Risiko gizi lebih	0,828947	0,807692	0,818182	78
Random Forest	Gizi lebih	0,800000	0,666667	0,727273	6

D. Confusion Matrix Model Final



Gambar 5. Confusion matrix model final SVM

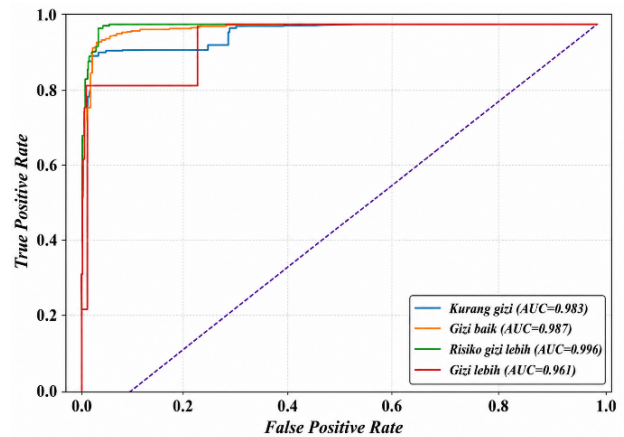
Gambar 5 menunjukkan *confusion matrix* model final SVM. Baris menunjukkan label aktual dan kolom menunjukkan label prediksi, sehingga gambar ini memperlihatkan jumlah prediksi benar dan salah pada setiap kelas status gizi.

E. AUC dan ROC Model Final

Tabel 9 menyajikan nilai *AUC* per kelas pada model final. SVM memperoleh *AUC Macro* sebesar 0,981852. Nilai ini menunjukkan kemampuan separabilitas model pada skema *one-vs-rest*, bukan menunjukkan bahwa model tersebut paling baik pada seluruh metrik klasifikasi akhir. Interpretasi kelas gizi lebih tetap perlu hati-hati karena *support* data uji hanya 6 data.

Tabel 9. AUC per kelas pada model final SVM

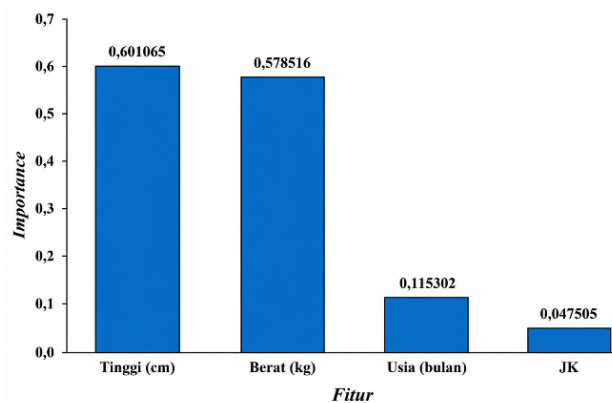
Kode	Kelas	AUC
1	Kurang gizi	0,982990
2	Gizi baik	0,987127
3	Risiko gizi lebih	0,996362
4	Gizi lebih	0,960930



Gambar 6. Kurva ROC multikelas model final SVM

Gambar 6 menunjukkan visualisasi kurva *ROC* multikelas model final SVM. Kurva ini memperlihatkan hubungan antara *true positive rate* dan *false positive rate* pada masing-masing kelas dalam skema *one-vs-rest*. Dalam konteks klasifikasi multikelas tidak seimbang, *AUC* digunakan sebagai metrik pendukung, sedangkan *F1 Macro*, *balanced accuracy*, *confusion matrix*, dan performa per kelas tetap menjadi dasar interpretasi utama.

F. Interpretabilitas Model Final



Gambar 7. Permutation importance model SVM

Gambar 7 menunjukkan urutan *permutation importance* pada model final *SVM*. Gambar ini memperjelas bahwa tinggi badan dan berat badan menjadi fitur paling dominan, sedangkan usia dan JK memiliki kontribusi lebih kecil pada performa prediksi. Fitur tinggi badan memiliki nilai *permutation importance* tertinggi sebesar 0,601065 diikuti berat badan sebesar 0,578516. Kedua fitur antropometri tersebut paling berpengaruh terhadap performa prediksi pada *pipeline* model final *SVM*.

G. Penetapan Model Terpilih

Berdasarkan aturan seleksi yang telah ditetapkan sebelumnya, model final penelitian dipilih berdasarkan *CV F1 Macro*, kemudian *CV balanced accuracy*, dan *CV accuracy*. Dengan aturan ini, *SVM* ditetapkan sebagai model final penelitian. *Random Forest* tetap penting dilaporkan karena memperoleh *F1 Macro* tertinggi pada data uji. Perbedaan ini bukan kontradiksi, tetapi menunjukkan dua sudut evaluasi: performa rata-rata pada validasi silang dan performa *holdout* pada data uji.

Tabel 10 menyajikan ringkasan keputusan model final. *SVM* ditetapkan sebagai model final penelitian karena unggul pada *CV F1 Macro*, sedangkan *Random Forest* dilaporkan sebagai model terbaik pada *F1 Macro* data uji. Tabel ini mencegah penafsiran ganda antara dasar seleksi penelitian dan performa *holdout*.

Tabel 10. Ringkasan keputusan model final

Aspek Keputusan	Hasil
Model final penelitian	<i>SVM</i>
Model top pada <i>test set</i>	<i>Random Forest</i>
Dasar seleksi	<i>CV F1 Macro -> Balanced Accuracy -> Accuracy</i>
Model terbaik pada data uji berdasarkan <i>F1 Macro</i>	<i>Random Forest, F1 Macro = 0,827132</i>
Model dengan <i>AUC Macro</i> tertinggi pada data uji	<i>SVM, AUC Macro = 0,981852</i>

IV. KESIMPULAN

Penelitian ini mengevaluasi lima algoritma *machine learning* menggunakan *SMOTENC* untuk klasifikasi multikelas status gizi balita berbasis data antropometri pada *dataset* 3.716 data dengan distribusi kelas yang tidak seimbang. Label status gizi diperlakukan sebagai label operasional yang tercatat pada *dataset* Posyandu. Hasil *cross-validation* menunjukkan bahwa *SVM* menjadi model final penelitian karena memperoleh *CV F1 Macro* tertinggi sebesar $0,829578 \pm 0,066709$ dan *balanced accuracy* sebesar $0,908484 \pm 0,070782$. Pada data uji, *Random Forest* memperoleh *F1 Macro* tertinggi sebesar 0,827132, sedangkan *KNN* memperoleh

akurasi tertinggi sebesar 0,956989 dan *balanced accuracy* tertinggi sebesar 0,883393. Hasil ini menunjukkan bahwa pemilihan model pada data tidak seimbang tidak dapat hanya didasarkan pada akurasi keseluruhan, tetapi perlu mempertimbangkan *F1 Macro*, *balanced accuracy*, *AUC Macro*, *confusion matrix*, dan performa per kelas. Kesimpulan penelitian ini hanya berlaku pada *dataset* Posyandu yang digunakan dan belum dapat digeneralisasi tanpa validasi eksternal. Penelitian selanjutnya disarankan memperluas jumlah data kelas minoritas, memverifikasi label status gizi terhadap standar resmi WHO/Kementerian Kesehatan, melakukan validasi pada *dataset* dari wilayah lain, serta membandingkan strategi penanganan *imbalance* lain seperti *ADASYN*, *Borderline-SMOTE*, *cost-sensitive learning*, dan skenario tanpa *oversampling*.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Program Magister Informatika Universitas Nusa Putra atas dukungan akademik dalam pelaksanaan penelitian ini. Penulis juga menyampaikan terima kasih kepada pihak pengelola data yang telah mendukung ketersediaan data agregat untuk kebutuhan penelitian.

REFERENSI

- [1] P. Mendon, M. Witsch, M. Becker, A. Adamski, and M. Vaillant, "Facilitating comprehensive child health monitoring within REDCap - an open-source code for real-time Z-score assessments," *BMC Med. Res. Methodol.*, vol. 24, no. 1, 2024, doi: 10.1186/s12874-024-02405-0.
- [2] R. Oliveira-Santos *et al.*, "Composite anthropometric data quality index for children under the age of 5 on the Brazilian National Food and Nutrition Surveillance System, 2019–2021," *Popul. Health Metr.*, vol. 23, no. 1, 2025, doi: 10.1186/s12963-025-00371-3.
- [3] F. H. Bitew, C. S. Sparks, and S. H. Nyarko, "Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia," *Public Health Nutr.*, vol. 25, no. 2, pp. 269–280, 2022, doi: 10.1017/S1368980021004262.
- [4] O. N. Chilyabanyama *et al.*, "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," *Children*, vol. 9, no. 7, 2022, doi: 10.3390/children9071082.
- [5] S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, "Prediction of Stunting among Under-5 Children in Rwanda Using Machine Learning Techniques," *J. Prev. Med. Public Heal.*, vol. 56, no. 1, pp. 41–49, 2023, doi: 10.3961/jpmph.22.388.

- [6] N. Novalina, I. A. A. Tarigan, F. K. Kameela, and M. Rizkinia, "Benchmarking machine learning algorithm for stunting risk prediction in Indonesia," *Bull. Electr. Eng. Informatics*, vol. 14, no. 3, pp. 2252–2263, 2025, doi: 10.11591/eei.v14i3.8997.
- [7] M. K. Ayele, G. A. Baye, S. H. Yesuf, A. A. Engda, and E. T. Mitiku, "Predicting stunting status among under five children in ethiopia using ensemble machine learning algorithms," *Sci. Rep.*, vol. 15, no. 1, pp. 1–11, 2025, doi: 10.1038/s41598-025-03206-1.
- [8] M. N. A. Khan and R. M. Yunus, "A hybrid ensemble approach to accelerate the classification accuracy for predicting malnutrition among under-five children in sub-Saharan African countries," *Nutrition*, vol. 108, p. 111947, 2023, doi: <https://doi.org/10.1016/j.nut.2022.111947>.
- [9] M. M. Islam, N. M. Shoukot Jahan Kibria, S. Kumar, D. C. Roy, and M. R. Karim, "Prediction of undernutrition and identification of its influencing predictors among under-five children in Bangladesh using explainable machine learning algorithms," *PLoS One*, vol. 19, no. 12, pp. 1–22, 2024, doi: 10.1371/journal.pone.0315393.
- [10] G. B. Begashaw, T. Zewotir, and H. M. Fenta, "A deep learning approach for classifying and predicting children's nutritional status in Ethiopia using LSTM-FC neural networks," *BioData Min.*, vol. 18, no. 1, 2025, doi: 10.1186/s13040-025-00425-0.
- [11] E. K. Anku and H. O. Duah, "Predicting and identifying factors associated with undernutrition among children under five years in Ghana using machine learning algorithms," *PLoS One*, vol. 19, no. 2 February, pp. 1–16, 2024, doi: 10.1371/journal.pone.0296625.
- [12] S. Wang, Y. Dai, J. Shen, and J. Xuan, "Research on expansion and classification of imbalanced data based on SMOTE algorithm," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-03430-5.
- [13] J. Li, Q. Zhu, Q. Wu, and Z. Fan, "A novel oversampling technique for class-imbalanced learning based on SMOTE and natural neighbors," *Inf. Sci. (Ny)*, vol. 565, pp. 438–455, 2021, doi: <https://doi.org/10.1016/j.ins.2021.03.041>.
- [14] E. Miranda, M. Aryuni, A. Y. Zakiyah, Y. E. Kurniawati, A. V. D. Sano, and M. Kumbangsila, "An early prediction model for toddler nutrition based on machine learning from imbalanced data," *Procedia Comput. Sci.*, vol. 245, pp. 263–271, 2024, doi: 10.1016/j.procs.2024.10.251.
- [15] R. Gustriansyah, N. Suhandi, S. Puspasari, and A. Sanmorino, "Machine Learning Method to Predict the Toddlers' Nutritional Status," *J. Infotel*, vol. 16, no. 1, pp. 32–43, 2024, doi: 10.20895/infotel.v15i4.988.
- [16] A. Subekti, "Comparative Analysis of Automated Machine Learning Methods for Multiclass Stunting Prediction Using Anthropometric Data," vol. 10, no. 2, pp. 991–1002, 2026, doi: 10.33395/sinkron.v10i2.15886.
- [17] A. Hendy *et al.*, "Supervised machine learning for classification and prediction of stunting among under-five Egyptian children," *BMC Pediatr.*, vol. 25, no. 1, 2025, doi: 10.1186/s12887-025-06138-x.
- [18] N. Hasdyna, R. K. Dinata, Rahmi, and T. I. Fajri, "Hybrid Machine Learning for Stunting Prevalence: A Novel Comprehensive Approach to Its Classification, Prediction, and Clustering Optimization in Aceh, Indonesia," *Informatics*, vol. 11, no. 4, 2024, doi: 10.3390/informatics11040089.
- [19] M. Anggito Herlambang Hadisuwarno *et al.*, "Komparasi performa model machine learning algoritma XGBoost dan Random Forest pada studi kasus mendeteksi stunting," *AITI J. Teknol. Inf.*, vol. 22, no. 2, pp. 266–278, 2025, doi: 10.24246/aiti.v22i2.266-278.
- [20] M. I. Elim and E. Utami, "Performance Comparison of Child Stunting Prediction Support Vector Machine vs Random Forest with Grid Search Optimization," *J. Tek. Inform.*, vol. 6, no. 5, pp. 5305–5319, 2025, doi: 10.52436/1.jutif.2025.6.5.5285.
- [21] A. Géron, *Hands-on TensorFlow Machine Learning Keras & TensorFlow*. Sebastopol, CA, USA: O'Reilly Media, 2019.