



Classification of Nutritional Status in Children Under Five Using Machine Learning with WHO-Standard Anthropometric Z-Score Labels

Klasifikasi Status Gizi Balita Menggunakan *Machine Learning* Berdasarkan Indikator Label Z-score Antropometri Standar WHO

Revanda Putri Rahmadani^{1*}, Muhammad Arifin², Fajar Nugraha³

^{1,2,3}Program Studi Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus, Indonesia

E-Mail: ¹202253154@std.umk.ac.id, ²arifin.m@umk.ac.id, ³fajar.nugraha@umk.ac.id

Received Apr 02nd 2026; Revised Jul 05th 2026; Accepted Aug 10th 2026; Available Online Aug 21th 2026

Corresponding Author: Revanda Putri Rahmadani

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

This study applied the Cross-Industry Standard Process for Data Mining (CRISP-DM) to classify nutritional status in children under five using routine anthropometric records from Puskesmas Wergu Wetan, Kudus, with the recorded WHO weight-for-length/height category as the target. The dataset contained 4,593 records and 39 attributes. After validity screening and deduplication based on combinations of analytical variables, 4,540 records were analyzed. The target comprised six categories: severe wasting, wasting, normal, possible risk of overweight, overweight, and obesity. Predictors included sex, age in months, body weight, and recorded height. The recorded weight-for-length/height Z-score was used solely to audit label consistency and was excluded from model inputs to prevent target leakage. Six algorithms were compared using the same stratified 80:20 train-test split without resampling. XGBoost achieved the best initial performance, with 94.93% accuracy and a 94.61% weighted F1-score. After GridSearchCV hyperparameter tuning, the final model correctly classified 861 of 908 test records, yielding 94.82% accuracy, a 94.51% weighted F1-score, a 68.33% macro F1-score, and 64.35% balanced accuracy. Unlike prior studies that generally examined fewer model families or did not combine leakage prevention, class-sensitive evaluation, paired statistical testing, and deployment, this study contributes a traceable end-to-end workflow that integrates these four elements in a non-diagnostic Streamlit prototype.

Keywords: CRISP-DM, Machine Learning, Nutritional Status, Streamlit, XGBoost, Z-score

Abstrak

Penelitian ini menerapkan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk mengklasifikasikan status gizi balita berdasarkan catatan antropometri Puskesmas Wergu Wetan, dengan kategori BB/TB WHO sebagai target. Dataset memuat 4.593 catatan dan 39 atribut. Setelah penyaringan validitas dan deduplikasi berdasarkan kombinasi variabel analisis, sebanyak 4.540 catatan dianalisis. Target terdiri atas enam kategori, yaitu gizi buruk, gizi kurang, gizi baik, risiko gizi lebih, gizi lebih, dan obesitas. Prediktor mencakup jenis kelamin, usia dalam bulan, berat badan, dan tinggi badan tercatat. Nilai Z-score BB/TB hanya digunakan untuk audit konsistensi label dan kemudian dikeluarkan dari masukan model untuk mencegah kebocoran target. Enam algoritma *machine learning* dibandingkan menggunakan pembagian data latih-uji terstratifikasi 80:20 yang sama tanpa *resampling*. XGBoost menunjukkan kinerja awal terbaik pada data uji internal dengan akurasi 94,93% dan F1-score tertimbang 94,61%. Setelah penalaan hiperparameter menggunakan GridSearchCV, model akhir mengklasifikasikan 861 dari 908 data uji dengan benar serta menghasilkan akurasi 94,82%, F1-score tertimbang 94,51%, F1-score makro 68,33%, dan *balanced accuracy* 64,35%. Berbeda dari studi terdahulu yang umumnya menguji lebih sedikit model keluarga atau belum menggabungkan pencegahan kebocoran data, evaluasi peka kelas, pengujian statistik berpasangan, dan implementasi, penelitian ini berkontribusi melalui satu alur yang dapat ditelusuri serta mengintegrasikan keempat unsur dalam prototipe Streamlit nondiagnostik.

Kata Kunci: CRISP-DM, Pembelajaran Mesin, Status Gizi Balita, Streamlit, XGBoost, Z-score

1. PENDAHULUAN

Status gizi balita perlu dipantau secara akurat karena ketidaksesuaian pertumbuhan dapat memengaruhi perkembangan fisik, kemampuan kognitif, dan daya tahan tubuh anak. Pemantauan tersebut memerlukan acuan antropometri terstandar agar hasil pengukuran dapat ditafsirkan secara tepat dan konsisten. World Health



Organization (WHO) menetapkan standar pertumbuhan anak usia 0 sampai 5 tahun sebagai rujukan penilaian pertumbuhan [1]. Di Indonesia, penilaian status gizi anak mengacu pada Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 tentang Standar Antropometri Anak, yang menetapkan indeks, kategori, dan ambang batas untuk menafsirkan hasil pengukuran antropometri [2].

Z-score merupakan ukuran terstandar yang menunjukkan posisi hasil pengukuran antropometri anak terhadap median standar pertumbuhan WHO dalam satuan simpangan baku. Nilai Z-score 0 menunjukkan posisi pada median standar, nilai negatif menunjukkan posisi di bawah median, dan nilai positif menunjukkan posisi di atas median. Pada penilaian berat badan terhadap panjang atau tinggi badan, Z-score menggambarkan posisi berat badan anak terhadap standar berdasarkan ukuran panjang atau tinggi badan dan jenis kelamin yang sesuai. WHO membedakan berat badan menurut panjang badan untuk anak sejak lahir sampai usia dua tahun dan berat badan menurut tinggi badan untuk anak usia dua sampai lima tahun [3]. Di Indonesia, kategori gizi buruk, gizi kurang, gizi baik, risiko gizi lebih, gizi lebih, dan obesitas ditetapkan berdasarkan ambang batas Standar Antropometri Anak [2]. Istilah BB/TB dalam penelitian ini digunakan secara konsisten mengikuti nomenklatur kolom target pada *dataset* sumber. Penelitian tidak menghitung ulang Z-score WHO dan tidak melakukan konversi ulang antara panjang badan dan tinggi badan. Penelitian juga tidak menggunakan indikator BB/U, TB/U, atau IMT/U. Oleh karena itu, keluaran model tidak ditujukan untuk mengklasifikasikan berat badan kurang menurut umur, stunting, atau status gizi berdasarkan indeks massa tubuh menurut umur.

Penelitian terdahulu telah menerapkan berbagai pendekatan *machine learning* untuk klasifikasi status gizi balita, tetapi ruang lingkup dan rancangan evaluasinya berbeda. Yanto et al. mengembangkan pendekatan klasifikasi status gizi anak [4]. Handayani et al. berfokus pada Random Forest [5], sedangkan Erkamim et al. membandingkan tiga metode *boosting*, yaitu AdaBoost, Gradient Boosting, dan XGBoost [19]. Rifai et al. menerapkan Gradient Boosting pada indikator BB/TB dan melaporkan dampak ketidakseimbangan kelas [7]. Pangestu dan Mujiyono mengimplementasikan XGBoost berbasis web [8], sedangkan Huda dan Wulandari mengembangkan aplikasi berbasis Z-score [9]. Insany et al. membandingkan KNN dan ANN pada indeks BB/U, BB/TB, dan TB/U [10]. Abas et al. menerapkan Naive Bayes pada 700 data bayi dan balita dengan akurasi 88,14% [11], sementara Setiawan dan Triayudi membandingkan Naive Bayes dan K-Nearest Neighbor dalam aplikasi berbasis web dengan akurasi masing-masing 80,60% dan 91,79% [12]. Dibandingkan penelitian-penelitian tersebut, studi ini menggunakan enam keluarga algoritma heterogen pada prediktor dan pembagian data uji yang sama, secara sengaja mengecualikan Z-score BB/TB dari masukan model untuk mencegah *target leakage*, mengevaluasi kinerja dengan metrik agregat dan per kelas pada distribusi yang sangat tidak seimbang, serta menguji perbedaan pola kesalahan antarmodel pada catatan uji yang sama. Model akhir kemudian diterapkan dalam prototipe non-diagnostik. Dengan demikian, *research gap* penelitian bukan sekadar pemilihan algoritma, melainkan belum terintegrasinya pencegahan *leakage*, evaluasi peka kelas, pengujian statistik berpasangan, dan implementasi dalam satu alur CRISP-DM yang dapat ditelusuri [18].

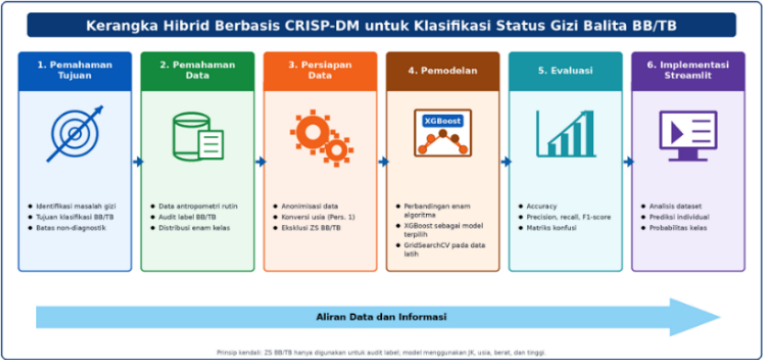
Validitas evaluasi model sangat dipengaruhi oleh pemisahan yang tepat antara prediktor dan target. Nilai Z-score BB/TB berhubungan langsung dengan aturan pembentukan kategori status gizi BB/TB. Apabila nilai tersebut dimasukkan sebagai prediktor, model dapat mengalami *target leakage* sehingga performa terlihat tinggi karena model memperoleh informasi yang berkaitan langsung dengan target, bukan hanya mempelajari pola dari variabel antropometri dasar [13], [14]. Penelitian mengenai kebocoran data menunjukkan bahwa informasi yang tidak semestinya tersedia pada tahap pelatihan atau prapemrosesan dapat menghasilkan estimasi kinerja yang terlalu optimistik dan mengurangi keandalan model [13], [14]. Oleh karena itu, Z-score BB/TB hanya digunakan untuk memeriksa konsistensi label dan tidak dilibatkan dalam pelatihan, penalaan hiperparameter, atau prediksi model.

Distribusi kelas pada data penelitian sangat tidak seimbang. Sebanyak 80,62% data akhir berada pada kelas gizi baik, sedangkan kelas gizi buruk hanya terdiri atas tiga catatan. Dalam kondisi tersebut, akurasi dan metrik tertimbang dapat didominasi oleh kelas mayoritas. Evaluasi karena itu mencakup metrik makro, metrik per kelas, *balanced accuracy*, jumlah *support*, matriks konfusi, dan pengujian statistik agar kinerja model dapat dinilai secara lebih proporsional [15], [17].

Berdasarkan kesenjangan tersebut, permasalahan penelitian ini adalah bagaimana kinerja enam algoritma *machine learning* yang berbeda ketika dibandingkan menggunakan data, prediktor, pembagian data, dan prosedur evaluasi yang sama, serta sejauh mana model dengan kinerja awal terbaik mempertahankan kinerja ketika distribusi kelas sangat tidak seimbang. Penelitian ini memiliki empat tujuan yang saling terkait: (1) membandingkan kinerja enam algoritma pada kondisi eksperimen yang sama; (2) melakukan penalaan hiperparameter XGBoost sebagai model dengan F1-score tertimbang awal tertinggi; (3) mengevaluasi hasil secara agregat, per kelas, dan statistik berpasangan; serta (4) menerapkan model akhir pada prototipe Streamlit. Kebaruan penelitian terletak pada integrasi empat unsur dalam satu alur: pencegahan *target leakage* dengan mengecualikan Z-score BB/TB dari variabel *input*; evaluasi data sangat tidak seimbang menggunakan metrik tertimbang, metrik makro, *balanced accuracy*, ROC-AUC, dan matriks konfusi; pengujian perbedaan antarmodel menggunakan uji McNemar dengan koreksi Holm-Bonferroni; serta implementasi model dalam prototipe pendukung analisis non-diagnostik yang terstruktur dan dapat ditelusuri.

2. METODE PENELITIAN

Penelitian ini menggunakan desain kuantitatif berbasis *data mining* terapan. Unit analisis adalah satu catatan pengukuran antropometri, bukan satu anak unik yang telah diverifikasi. Alur penelitian mengikuti enam tahap *CRISP-DM*, yaitu pemahaman tujuan, pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi [18]. Keseluruhan alur penelitian ditampilkan pada Gambar 1. Penelitian ini menggunakan bantuan AI Generative untuk membangkitkan gambar dan sebagai alat bantu dalam melakukan visualisasi untuk lebih jelas.



Gambar 1. Alur penelitian berdasarkan CRISP-DM

2.1. Sumber Data dan Variabel

Dataset berasal dari 4.593 catatan pelayanan rutin Puskesmas Wergu Wetan, Kabupaten Kudus, yang memuat 39 atribut. Data dipertahankan apabila jenis kelamin berisi L atau P, target BB/TB termasuk dalam enam kelas penelitian, usia dapat dikonversi menjadi bulan, serta berat dan tinggi badan dapat dibaca sebagai angka. Empat baris judul berulang dan 23 target tidak valid dihapus sehingga tersisa 4.566 catatan valid. Setelah menghapus 26 kombinasi nilai analisis yang identik, diperoleh 4.540 catatan untuk pemodelan.

Data identitas langsung, seperti NIK, nama anak, nama orang tua, alamat, RT/RW, kelurahan, dan posyandu, tidak digunakan dalam analisis. Izin administratif penggunaan data diperoleh dari Puskesmas Wergu Wetan. Variabel model mencakup jenis kelamin, usia saat ukur, berat badan, dan tinggi badan sebagai prediktor, serta status gizi BB/TB sebagai target. Keempat prediktor dipilih karena tersedia secara rutin dalam catatan pelayanan dan relevan dengan penilaian antropometri dasar. Berat badan, panjang/tinggi badan, dan jenis kelamin berkaitan langsung dengan kerangka penilaian BB/PB atau BB/TB WHO, sedangkan usia digunakan sebagai konteks perkembangan dan karakteristik pengukuran. Nilai Z-score BB/TB tidak dimasukkan sebagai prediktor karena berhubungan langsung dengan pembentukan label dan berpotensi menimbulkan *target leakage*. Dengan demikian, model ditujukan untuk mempelajari kemampuan variabel dasar dalam mereplikasi label pelayanan, bukan menggantikan perhitungan antropometri standar atau diagnosis klinis. Karakteristik setiap fitur dan variabel disajikan pada Tabel 1.

Tabel 1. Karakteristik fitur dan variabel *dataset*

Fitur/ Variabel	Tipe Data dan Transformasi	Keterangan	Peran
Jenis Kelamin (JK)	Kategorikal L/P menjadi 1/0	Jenis kelamin balita	Input
Umur (Usia Saat Ukur)	Tahun-bulan-hari menjadi total bulan	Usia balita saat pengukuran	Input
Berat Badan (Berat)	Numerik (kg)	Berat badan hasil penimbangan	Input
Tinggi Badan (Tinggi)	Numerik (cm)	Tinggi badan hasil pengukuran	Input
Status Gizi (BB/TB)	Kategorikal, enam kelas	Kategori status gizi berdasarkan label BB/TB	Target

2.2. Audit Label dan Prapemrosesan

Nilai ZS BB/TB telah tersedia pada data sumber dan tidak dihitung ulang. Audit terhadap 4.566 catatan dengan target valid menunjukkan bahwa label konsisten dengan rentang Z-score yang tercatat. Nilai ZS BB/TB hanya digunakan untuk audit konsistensi label dan tidak dimasukkan ke dalam variabel *input*, pelatihan, pencarian hiperparameter, maupun prediksi. Hubungan nilai Z-score dengan pembentukan kategori status gizi BB/TB mengacu pada standar antropometri [2]. Nilai ZS BB/TB tetap dikeluarkan dari variabel *input* untuk mencegah *target leakage* [13], [14]. Dengan demikian, model mempelajari pemetaan empat prediktor terhadap label pelayanan yang tercatat dan tidak menggantikan perhitungan Z-score WHO. Prapemrosesan meliputi

konversi usia ke dalam bulan, penggantian tanda koma desimal menjadi titik, konversi berat dan tinggi ke format numerik, serta pengkodean jenis kelamin menjadi L=1 dan P=0. Data akhir dibagi secara terstratifikasi menjadi 3.632 data latih dan 908 data uji internal dengan rasio 80:20. Pembagian berdasarkan kelas disajikan pada Tabel 2.

Tabel 2. Distribusi data latih dan data uji berdasarkan kelas

Kategori Status Gizi	Data Latih (80%)	Data Uji (20%)	Total
Gizi Baik	2.928	732	3.66
Gizi Buruk	2	1	3
Gizi Kurang	77	20	97
Gizi Lebih	153	38	191
Obesitas	78	19	97
Risiko Gizi Lebih	394	98	492
Total	3.632	908	4.54

StandardScaler dihitung hanya dari data latih dan diterapkan pada data latih serta data uji untuk KNN, SVM, dan Gaussian Naive Bayes. Decision Tree, Random Forest, dan XGBoost menggunakan fitur numerik tanpa standarisasi yang wajib. Pembagian dilakukan pada tingkat catatan karena identitas anak unik dan periode pengukuran tidak digunakan sebagai dasar pemisahan. Data uji yang sama digunakan untuk membandingkan model awal dan menilai model hasil optimasi, sehingga seluruh hasil diposisikan sebagai estimasi internal pada tingkat catatan.

2.3. Pemodelan, Optimasi Hiperparameter, dan Evaluasi

Enam algoritma dibandingkan menggunakan prediktor, target, dan pembagian data yang sama, yaitu Decision Tree, Random Forest, K-Nearest Neighbor, Support Vector Machine, Gaussian Naive Bayes, dan XGBoost. Pemilihannya dimaksudkan untuk mewakili berbagai mekanisme pembelajaran: pohon tunggal, *ensemble bagging*, kedekatan berbasis jarak, pemisahan berbasis margin, pendekatan probabilistik, dan *gradient boosting*. Perbandingan lintas keluarga ini mengurangi ketergantungan kesimpulan pada satu paradigma model dan memungkinkan penilaian yang lebih adil karena setiap algoritma menerima informasi prediktor serta data uji yang sama. Implementasi dilakukan menggunakan pustaka scikit-learn dan XGBoost. F1-score tertimbang digunakan sebagai kriteria utama perbandingan awal karena tetap memperhitungkan *precision* dan *recall* sekaligus memberikan bobot sesuai *support* kelas. Namun, karena metrik tertimbang masih dapat dipengaruhi oleh kelas mayoritas, penetapan model terbaik tidak ditafsirkan dari satu metrik saja dan tetap dibaca bersama F1-score makro, *balanced accuracy*, metrik per kelas, ROC-AUC, serta uji statistik [15]-[17]. XGBoost memperoleh F1-score tertimbang awal tertinggi sehingga dipilih untuk tahap penalaan hiperparameter.

Penalaan hiperparameter XGBoost dilakukan menggunakan GridSearchCV pada 3.632 data latih. Validasi silang menggunakan tiga lipatan, *scoring* = accuracy, *n_jobs* = -1, dan 72 kombinasi parameter atau 216 proses pelatihan-validasi. Tiga lipatan dipilih sebagai kompromi praktis agar setiap lipatan tetap mempertahankan sebanyak mungkin observasi kelas minoritas yang sangat terbatas. Ruang pencarian dirancang untuk menguji variasi kompleksitas model dan kekuatan pembelajaran secara terkontrol: *n_estimators* mengatur jumlah pohon, *max_depth* mengendalikan kedalaman dan kompleksitas pohon, *learning_rate* mengatur besarnya pembaruan pada setiap tahap *boosting*, sedangkan *subsample* dan *colsample_bytree* mengatur proporsi observasi dan fitur yang digunakan untuk membantu mengendalikan varians. Rentang yang relatif kompak dipilih agar pencarian tetap sistematis tanpa memperluas kompleksitas secara berlebihan pada *dataset* dengan jumlah kelas gizi buruk yang sangat sedikit. Akurasi dipertahankan sebagai fungsi *skor* internal sesuai konfigurasi eksperimen, tetapi tidak digunakan sebagai dasar interpretasi model akhir. Penilaian akhir tetap mempertimbangkan F1-score tertimbang, F1-score makro, *balanced accuracy*, metrik per kelas, dan matriks konfusi. Karena kelas gizi buruk hanya memiliki dua catatan pada data latih, hasil validasi silang dan penalaan ditafsirkan sebagai estimasi internal, bukan bukti kinerja eksternal yang definitif. *Grid search* digunakan untuk menelusuri kombinasi parameter secara sistematis [20]. Ruang pencarian serta nilai optimal disajikan pada Tabel 3.

Tabel 3. Ruang pencarian dan hiperparameter optimal XGBoost

Hiperparameter	Rentang Pencarian	Nilai Terpilih
<i>n_estimators</i>	[50, 100]	100
<i>max_depth</i>	[3, 5, 7]	7
<i>learning_rate</i>	[0,05; 0,10; 0,20]	0,20
<i>subsample</i>	[0,80; 1,00]	0,80
<i>colsample_bytree</i>	[0,80; 1,00]	1,00

Evaluasi mencakup akurasi, presisi, *recall*, F1-score per kelas, *macro average*, *weighted average*, *balanced accuracy*, matriks konfusi, kurva ROC *micro-average*, dan area under the curve (AUC). Kombinasi metrik digunakan karena distribusi target sangat tidak seimbang; akurasi dan *weighted average* menggambarkan kinerja keseluruhan yang dipengaruhi *support* kelas, sedangkan *macro average* dan *balanced accuracy* memberi bobot yang lebih setara pada kelas minoritas [15], [17]. Kurva ROC multikelas dihitung dengan pendekatan *one-vs-rest* pada enam model awal. Label target dibinarisasi; skor kelas diperoleh melalui *predict_proba* pada Decision Tree, Random Forest, KNN, Gaussian Naive Bayes, dan XGBoost, sedangkan SVM menggunakan *decision_function*. Nilai *micro-average* dihitung dengan menggabungkan seluruh elemen label biner dan skor kelas pada 908 data uji. ROC-AUC digunakan sebagai ukuran diskriminasi tambahan dan tidak diperlakukan sebagai pengganti metrik klasifikasi berbasis keputusan, terutama pada data tidak seimbang [16], [17]. Rumus evaluasi utama disajikan pada Persamaan (1) sampai Persamaan (7).

Kepentingan fitur dilaporkan untuk XGBoost setelah penalaan menggunakan skor *gain* ternormalisasi melalui atribut *feature_importances_* [6]. Skor ini digunakan untuk menjelaskan kontribusi relatif fitur terhadap pemisahan yang dibentuk model dan tidak ditafsirkan sebagai hubungan kausal. Perbedaan kesalahan dua model yang diuji pada catatan yang sama dianalisis menggunakan uji McNemar dua sisi karena pengujiannya berfokus pada pasangan hasil benar-salah dari observasi uji yang identik. Hipotesis nol menyatakan bahwa proporsi kesalahan kedua model setara. Karena enam model menghasilkan 15 perbandingan berpasangan, keputusan signifikansi menggunakan koreksi Holm-Bonferroni untuk mengendalikan *family-wise error rate* pada $\alpha = 0,05$ [17].

Evaluasi yang digunakan terdiri dari *Accuracy*, *Precision*, *Recall*, *F1-Score*, *Macro Average* (Rata-rata Makro), *Weighted Average* (Rata-rata Tertimbang) dan *Balanced Accuracy*, masing-masing ditunjukkan pada persamaan 1 sampai dengan 6.

$$\text{Accuracy} = \frac{\sum_{i=1}^K TP_i}{N} \quad (1)$$

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad (2)$$

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

$$F1_i = \frac{2 \times \text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (4)$$

$$M_{\text{macro}} = \frac{\sum_{i=1}^K M_i}{K} \quad (5)$$

$$M_{\text{weighted}} = \frac{\sum_{i=1}^K n_i \times M_i}{N} \quad (6)$$

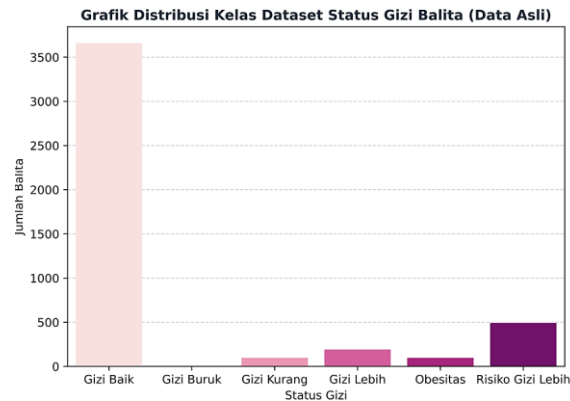
$$\text{Balanced Accuracy} = \frac{\sum_{i=1}^K \text{Recall}_i}{K} \quad (7)$$

Macro average memberi bobot yang sama pada setiap kelas, sedangkan *weighted average* memberi bobot berdasarkan jumlah data aktual pada tiap kelas. *Balanced accuracy* pada klasifikasi multikelas merupakan rata-rata *recall* seluruh kelas, sehingga nilainya sama dengan *recall* makro dalam laporan ini. Selain metrik pada Persamaan (1) sampai Persamaan (7), evaluasi model juga menggunakan *confusion matrix*, ROC-AUC, dan uji McNemar. *Confusion matrix* digunakan untuk mengidentifikasi distribusi prediksi yang benar dan pola kesalahan klasifikasi pada setiap kelas. ROC-AUC digunakan sebagai ukuran tambahan untuk mengevaluasi kemampuan diskriminasi model. Pada klasifikasi multikelas, kurva ROC dihitung menggunakan pendekatan *one-vs-rest* dengan nilai *micro-average*. Perbedaan pola kesalahan antarmodel dianalisis menggunakan uji McNemar dua sisi pada data uji yang sama. Karena enam algoritma menghasilkan 15 perbandingan berpasangan, keputusan signifikansi dikoreksi menggunakan metode Holm-Bonferroni dengan *family-wise* $\alpha = 0,05$.

3. HASIL DAN PEMBAHASAN

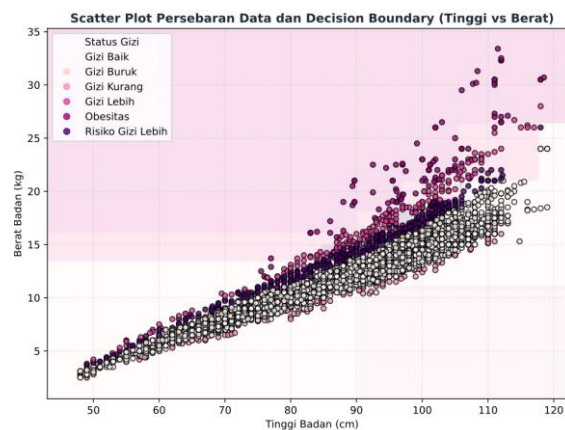
3.1. Karakteristik dan Eksplorasi Data

Dataset final memuat 4.540 catatan. Usia berada pada rentang 0-60 bulan dengan median 34,57 bulan dan IQR 17,40-47,13. Berat badan berada pada rentang 2,5-33,4 kg dengan median 12,7 kg dan IQR 9,7-15,0. Tinggi badan tercatat berada pada rentang 48-119 cm dengan median 90,5 cm dan IQR 77,5-99,0. Kelas gizi baik berjumlah 3.660 catatan atau 80,62%, sedangkan gizi buruk berjumlah tiga catatan atau 0,07%. Distribusi kelas lengkap ditampilkan pada Gambar 2.



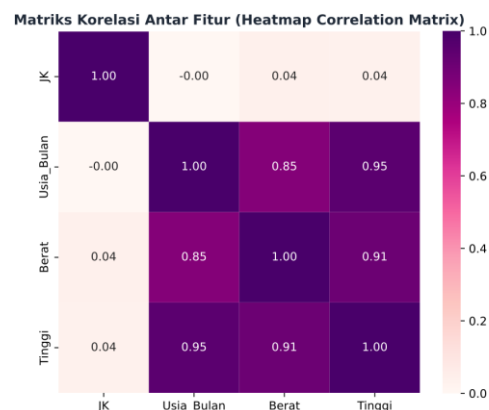
Gambar 2. Distribusi kelas status gizi pada *dataset* final (n = 4.540)

Dominasi kelas gizi baik menunjukkan ketidakseimbangan yang kuat, dengan rasio kelas terbesar terhadap kelas terkecil sebesar 1.220:1. Kondisi tersebut menjelaskan mengapa akurasi dan metrik tertimbang perlu dibaca bersama metrik makro dan kinerja per kelas. Hubungan antara tinggi badan dan berat badan memperlihatkan pola peningkatan yang jelas, tetapi sebagian kategori saling beririsan pada area batas. Untuk membantu membaca pola tersebut, *decision boundary* pada Gambar 3 dibuat sebagai visualisasi eksploratif dua dimensi yang hanya menggunakan fitur tinggi dan berat badan. Visualisasi ini bukan representasi lengkap model utama yang menggunakan empat prediktor.



Gambar 3. Persebaran tinggi badan, berat badan, dan status *decision boundary* gizi

Tumpang tindih titik pada batas antarkategori menunjukkan bahwa pemisahan visual berbasis dua fitur belum cukup untuk menilai model utama. Dalam pemodelan utama, usia dan jenis kelamin tetap digunakan bersama dengan berat dan tinggi badan untuk membentuk prediksi. Matriks korelasi menunjukkan korelasi positif tinggi antara usia dan tinggi badan (0,95), berat dan tinggi badan (0,91), serta usia dan berat badan (0,85). Korelasi jenis kelamin terhadap fitur numerik berada di sekitar nol. Hubungan antarfitur ditampilkan pada Gambar 4.



Gambar 4. Matriks korelasi antarfitur pada *dataset* final

Korelasi tinggi antarfaktor pertumbuhan merupakan pola yang wajar pada data antropometri. Nilai tersebut digunakan untuk memahami struktur data, bukan untuk menyimpulkan hubungan sebab-akibat. Model berbasis pohon tetap dapat memanfaatkan kombinasi fitur tanpa mengharuskan asumsi independensi linear.

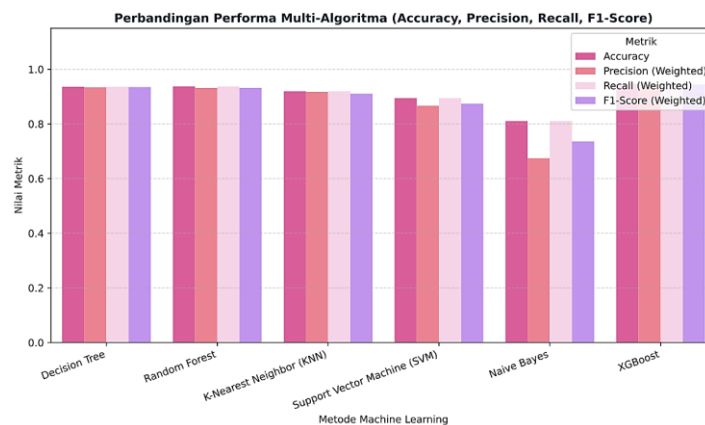
3.2. Perbandingan Kinerja Model

Hasil evaluasi enam model awal pada 908 data uji internal menunjukkan perbedaan performa agregat dan performa antarkelas. Nilai lengkap *accuracy*, *precision*, *recall*, F1-score tertimbang, serta metrik makro disajikan pada Tabel 4.

Tabel 4. Perbandingan kinerja enam algoritma pada data uji internal

Metode	Acc.	P-W	R-W	F1-W	P-M	R-M	F1-M
Decision Tree	93,61%	93,42%	93,61%	93,48%	67,68%	64,57%	65,91%
Random Forest	93,72%	93,17%	93,72%	93,22%	71,39%	60,83%	64,95%
KNN	91,96%	91,73%	91,96%	91,04%	74,18%	53,87%	58,97%
SVM	89,43%	86,66%	89,43%	87,38%	54,80%	41,74%	46,45%
Naive Bayes	81,06%	67,43%	81,06%	73,60%	19,35%	25,30%	21,84%
XGBoost	94,93%	94,63%	94,93%	94,61%	74,25%	63,42%	67,69%

XGBoost memperoleh *accuracy* 94,93% dan F1-score tertimbang 94,61%, tertinggi pada konfigurasi awal. F1-score makro XGBoost sebesar 67,69% juga menjadi nilai tertinggi, sedangkan *recall* makro tertinggi diperoleh oleh Decision Tree sebesar 64,57%. Perbandingan empat metrik agregat keenam algoritma ditampilkan pada Gambar 5.

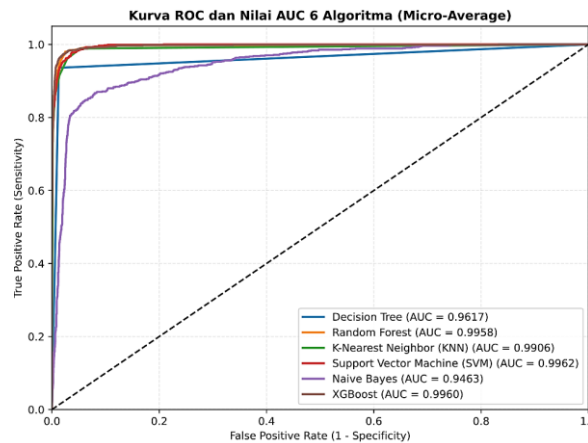


Gambar 5. Perbandingan metrik agregat enam algoritma

Perbedaan antara metrik tertimbang dan metrik makro menunjukkan pengaruh kelas mayoritas. Naive Bayes memiliki kinerja terendah, sedangkan XGBoost, Decision Tree, dan Random Forest berada pada kelompok kinerja agregat tertinggi. Secara metodologis, pola ini dapat dijelaskan karena hubungan usia, berat, tinggi, dan kategori BB/TB bersifat nonlinier dan melibatkan interaksi antarfitur, sehingga model berbasis pohon mampu membentuk batas keputusan bertingkat. XGBoost juga membangun pohon secara berurutan untuk memperbaiki kesalahan prediksi sebelumnya. Namun, hasil ini hanya berlaku pada data dan konfigurasi penelitian ini. Dibandingkan dengan penelitian terdahulu yang berfokus pada satu algoritma, kelompok *boosting*, perbandingan dua model, atau pengembangan aplikasi [5], [7]-[12], [19], penelitian ini membandingkan enam keluarga algoritma pada data uji yang sama, menggunakan metrik peka kelas, dan menambahkan pengujian statistik berpasangan.

Evaluasi probabilistik enam model awal melalui kurva ROC *one-vs-rest micro-average* menghasilkan AUC 0,9617 pada Decision Tree, 0,9958 pada Random Forest, 0,9906 pada KNN, 0,9962 pada SVM, 0,9463 pada Naive Bayes, dan 0,9960 pada XGBoost. Kurva ROC keenam algoritma ditampilkan pada Gambar 6.

SVM memperoleh AUC *micro-average* tertinggi, tetapi XGBoost tetap menjadi model dengan kinerja awal terbaik berdasarkan kriteria utama penelitian, yaitu F1-score tertimbang yang dibaca bersama metrik makro dan per kelas. Perbedaan peringkat ini tidak kontradiktif karena ROC-AUC menilai kemampuan model dalam memberi peringkat atau memisahkan skor antarkelas di berbagai ambang, sedangkan F1-score dihitung dari keputusan kelas pada konfigurasi prediksi tertentu. Pada skema *micro-average*, kontribusi keputusan dari kelas dengan jumlah besar juga dapat lebih dominan. Oleh karena itu, AUC yang sangat tinggi tidak otomatis menghasilkan F1-score tertinggi pada klasifikasi multikelas yang tidak seimbang. Temuan ini menguatkan alasan penggunaan beberapa metrik secara simultan dan menghindari penetapan model terbaik berdasarkan satu indikator saja [15]-[17].



Gambar 6. Kurva ROC dan nilai AUC enam algoritma

3.3. Kinerja Model Optimasi dan Interpretasi Model

Model XGBoost hasil optimasi memperoleh *accuracy* 94,82%, *precision* tertimbang 94,56%, *recall* tertimbang 94,82%, F1-score tertimbang 94,51%, *precision* makro 74,56%, *recall* makro 64,35%, dan F1-score makro 68,33%. Dibandingkan dengan model awal, *accuracy* turun 0,11 poin persentase dan F1-score tertimbang turun 0,10 poin, sedangkan *recall* makro meningkat 0,93 poin dan F1-score makro meningkat 0,64 poin. Pola tersebut menunjukkan adanya *trade-off* kecil antara kinerja yang didominasi oleh kelas mayoritas dan pemerataan pengenalan antarkelas: beberapa keputusan yang sebelumnya menguntungkan akurasi keseluruhan dapat bergeser sehingga sensitivitas rata-rata kelas meningkat. Karena perubahan agregatnya kecil, hasil optimasi tidak dapat diklaim meningkatkan seluruh metrik secara seragam. Nilai akhirnya lebih tepat ditafsirkan sebagai konfigurasi dengan sedikit perbaikan pada metrik makro, tetapi dengan akurasi dan F1-score tertimbang yang sedikit lebih rendah. Rincian kinerja setiap kelas disajikan pada Tabel 5.

Tabel 5. Kinerja per kelas model XGBoost hasil optimasi

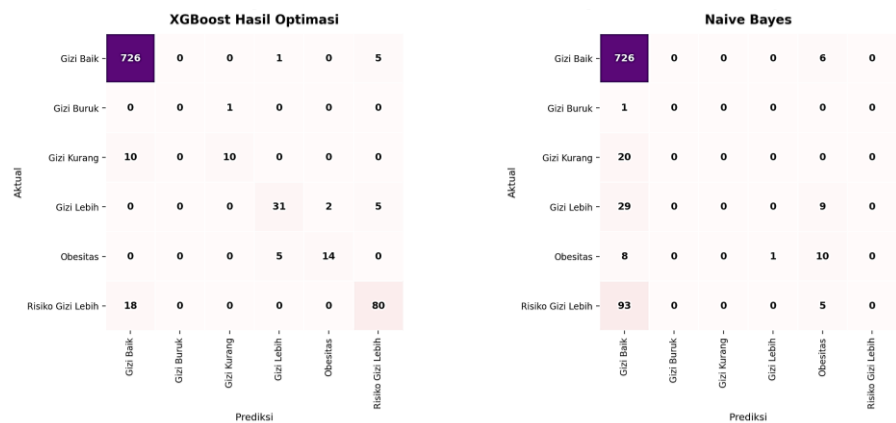
Kategori Status Gizi	Precision	Recall	F1-score	Support
Gizi Baik	96,29%	99,18%	97,71%	732
Gizi Buruk	0,00%	0,00%	0,00%	1
Gizi Kurang	90,91%	50,00%	64,52%	20
Gizi Lebih	83,78%	81,58%	82,67%	38
Obesitas	87,50%	73,68%	80,00%	19
Risiko Gizi Lebih	88,89%	81,63%	85,11%	98
Macro average	74,56%	64,35%	68,33%	908
Weighted average	94,56%	94,82%	94,51%	908

Model akhir mengklasifikasikan 861 dari 908 data uji dengan benar. *Recall* tertinggi diperoleh kelas gizi baik sebesar 99,18%. *Recall* kelas risiko gizi lebih mencapai 81,63%, gizi lebih 81,58%, obesitas 73,68%, dan gizi kurang 50,00%. *Support* kelas gizi buruk hanya satu data uji, sehingga nilai 0,00% pada kelas tersebut dibaca sesuai konteks jumlah sampel yang tersedia.

Perbandingan *confusion matrix* model XGBoost hasil optimasi dan Naive Bayes memperlihatkan perbedaan pola kesalahan. XGBoost mempertahankan lebih banyak prediksi pada diagonal utama, sedangkan Naive Bayes lebih sering mengarahkan kelas minoritas ke kelas gizi baik atau obesitas. Visualisasi perbandingan tersebut disajikan pada Gambar 7.

Matriks konfusi XGBoost pada Gambar 7 konsisten dengan Tabel 5, yaitu 861 prediksi benar dari 908 data uji. Sebagian besar kesalahan terjadi pada kelas yang berdekatan, terutama risiko gizi lebih yang diprediksi sebagai gizi baik dan gizi kurang yang diprediksi sebagai gizi baik. Pola ini sejalan dengan tumpang tindih ruang fitur yang terlihat pada visualisasi eksploratif dan dengan dominasi kelas gizi baik dalam data. Ketika contoh kelas minoritas sedikit, model memiliki informasi yang lebih terbatas untuk membentuk batas keputusan yang stabil bagi kelas tersebut. Perbandingan dengan Naive Bayes memperlihatkan bahwa XGBoost mempertahankan lebih banyak prediksi pada diagonal utama dan lebih baik pada beberapa kelas minoritas, tetapi kegagalan mengenali satu-satunya data uji gizi buruk tetap menjadi keterbatasan utama yang harus dipertimbangkan sebelum penggunaan lebih luas.

Uji McNemar berpasangan menunjukkan bahwa, setelah koreksi Holm-Bonferroni, XGBoost berbeda signifikan dari KNN, SVM, dan Naive Bayes. Perbedaannya dengan Decision Tree ($p = 0,1272$) dan Random Forest ($p = 0,0817$) tidak signifikan. Nilai p dan keputusan lengkap disajikan pada Tabel 6.



Gambar 7. Perbandingan confusion matrix XGBoost hasil optimasi dan Naive Bayes

Tabel 6. Nilai p uji McNemar berpasangan antar model

Algoritma	Decision Tree	Random Forest	KNN	SVM	Naive Bayes	XGBoost
Decision Tree	-	1,0000 (TS)	0,0872 (TS)	<0,001 (S)	<0,001 (S)	0,1272 (TS)
Random Forest	1,0000 (TS)	-	0,0177 (TS)	<0,001 (S)	<0,001 (S)	0,0817 (TS)
KNN	0,0872 (TS)	0,0177 (TS)	-	0,0021 (S)	<0,001 (S)	<0,001 (S)
SVM	<0,001 (S)	<0,001 (S)	0,0021 (S)	-	<0,001 (S)	<0,001 (S)
Naive Bayes	<0,001 (S)	<0,001 (S)	<0,001 (S)	<0,001 (S)	-	<0,001 (S)
XGBoost	0,1272 (TS)	0,0817 (TS)	<0,001 (S)	<0,001 (S)	<0,001 (S)	-

Keterangan: nilai yang ditampilkan merupakan *p-value* dua sisi; S = signifikan dan TS = tidak signifikan setelah koreksi Holm-Bonferroni pada *family-wise* $\alpha = 0,05$.

Hasil uji statistik menunjukkan bahwa keunggulan numerik XGBoost dibandingkan model pohon lain perlu ditafsirkan secara proporsional. XGBoost berbeda secara signifikan dari KNN, SVM, dan Naive Bayes setelah koreksi Holm-Bonferroni, tetapi tidak berbeda secara signifikan dari Decision Tree maupun Random Forest. Dengan demikian, XGBoost ditetapkan sebagai model dengan kinerja agregat awal tertinggi pada eksperimen ini, bukan sebagai model yang secara statistik selalu mengungguli seluruh pembandingan. Temuan tersebut juga memperjelas kontribusi analitis penelitian: peringkat berbasis metrik dilengkapi dengan pengujian pola kesalahan pada observasi yang sama, sehingga perbedaan numerik kecil tidak langsung dianggap sebagai bukti keunggulan yang pasti. Koreksi perbandingan berganda selanjutnya mengurangi risiko kesimpulan positif palsu pada 15 pasangan pengujian.

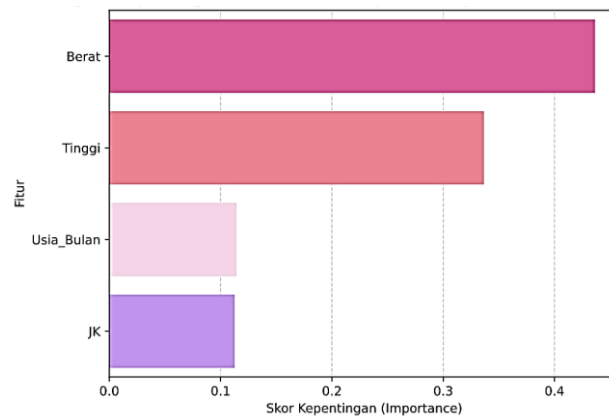
Secara praktis, model dapat dipertimbangkan sebagai alat bantu penyaringan awal dan pemeriksaan konsistensi catatan antropometri di lingkungan pelayanan, misalnya untuk membantu petugas mengidentifikasi catatan yang perlu ditinjau kembali atau diprioritaskan untuk diverifikasi. Fungsi tersebut berbeda dari diagnosis. Kinerja yang sangat terbatas pada kelas gizi buruk menegaskan bahwa keluaran model tidak boleh digunakan sebagai dasar tunggal untuk keputusan klinis dan tetap harus dikonfirmasi melalui standar antropometri resmi serta penilaian tenaga kesehatan. Implikasi metodologisnya adalah bahwa evaluasi model kesehatan dengan kelas yang sangat tidak seimbang perlu melaporkan metrik per kelas dan dukungan, bukan hanya akurasi. Implikasi pengembangannya adalah perlunya penambahan data kelas minoritas serta validasi eksternal sebelum sistem dipertimbangkan untuk digunakan secara operasional dalam skala yang lebih luas.

Pada hasil optimasi XGBoost, skor *gain* ternormalisasi menunjukkan bahwa berat badan memiliki kepentingan tertinggi, diikuti oleh tinggi badan, usia dalam bulan, dan jenis kelamin. Urutan *fitur penting* ditampilkan pada Gambar 8.

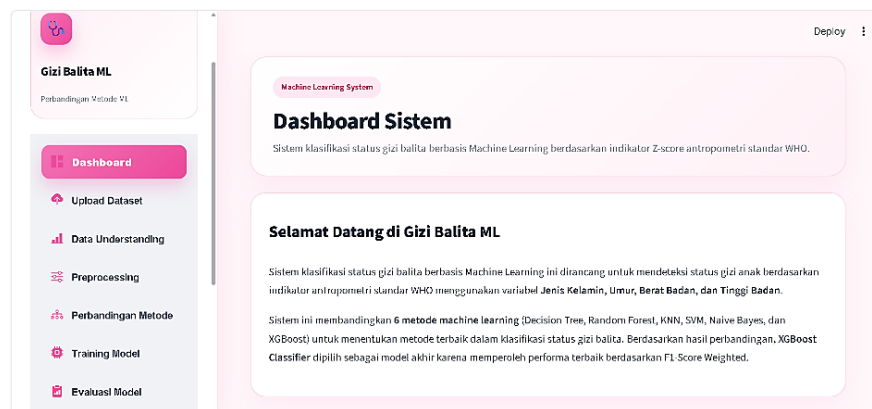
Berat dan tinggi badan memiliki kepentingan relatif tertinggi, yang konsisten dengan karakter target BB/TB dan menunjukkan bahwa model paling sering memperoleh peningkatan pemisahan dari kedua variabel antropometri tersebut. Usia dan jenis kelamin tetap memberikan informasi tambahan dalam konfigurasi model, tetapi kontribusi relatifnya lebih kecil. Skor *gain* hanya menggambarkan kontribusi fitur terhadap pemisahan yang dibentuk oleh XGBoost pada *dataset* ini; nilai tersebut tidak menunjukkan hubungan sebab-akibat, tidak membuktikan kepentingan klinis, dan dapat berubah apabila distribusi data atau konfigurasi model berubah.

3.4. Implementasi

Model XGBoost setelah penalaan diintegrasikan ke dalam prototipe Streamlit yang memuat proses unggah data, prapemrosesan, perbandingan model, evaluasi, dan klasifikasi. Pengguna memasukkan jenis kelamin, usia, berat badan, dan tinggi badan, kemudian sistem menampilkan kategori BB/TB serta probabilitas setiap kelas. Antarmuka prototipe disajikan pada Gambar 9.



Gambar 8. Feature importance XGBoost hasil optimasi berdasarkan *gain* ternormalisasi



Gambar 9. Antarmuka prototipe Streamlit untuk klasifikasi status gizi balita

Prototipe mendukung alur analisis yang terstruktur dan dapat ditelusuri. Implementasi menunjukkan bahwa pemodelan, evaluasi, dan penyajian prediksi dapat diintegrasikan dalam satu aplikasi. Sistem diposisikan sebagai pendukung analisis non-diagnostik. Dalam konteks Puskesmas, prototipe dapat digunakan untuk demonstrasi alur pengolahan dan peninjauan awal catatan, bukan untuk menggantikan penghitungan Z-score WHO atau keputusan tenaga kesehatan. Hasil prediksi tetap harus diinterpretasikan berdasarkan standar antropometri resmi dan verifikasi klinis. Penelitian lanjutan perlu memprioritaskan penambahan data kelas minoritas, pemisahan berdasarkan anak unik, validasi temporal dan eksternal, evaluasi kalibrasi probabilitas, serta studi kegunaan aplikasi sebelum prototipe dipertimbangkan untuk penggunaan operasional.

4. KESIMPULAN

Penelitian ini telah menjawab tujuan pertama dengan membandingkan enam algoritma *machine learning* pada 4.540 catatan antropometri menggunakan empat prediktor dasar, yaitu jenis kelamin, usia dalam bulan, berat badan, dan tinggi badan tercatat, dengan enam kategori status gizi BB/TB sebagai target. Nilai Z-score BB/TB hanya digunakan untuk audit konsistensi label dan dikeluarkan dari *input* model untuk mencegah *target leakage*. Desain tersebut memungkinkan perbandingan model berdasarkan informasi dasar yang tersedia tanpa memberikan variabel yang berhubungan langsung dengan aturan pembentukan target.

Tujuan kedua dan ketiga dijawab melalui pemilihan serta evaluasi XGBoost. Pada konfigurasi awal, XGBoost memperoleh akurasi 94,93% dan F1-score tertimbang 94,61%, tertinggi di antara enam model yang diuji. Setelah GridSearchCV, model akhir mengklasifikasikan 861 dari 908 data uji dengan benar dan menghasilkan akurasi 94,82%, F1-score tertimbang 94,51%, F1-score makro 68,33%, dan *balanced accuracy* 64,35%. Peningkatan kecil pada metrik makro disertai penurunan kecil pada akurasi dan F1-score tertimbang, sehingga optimasi tidak diklaim meningkatkan seluruh aspek kinerja. Uji McNemar dengan koreksi Holm-Bonferroni menunjukkan perbedaan signifikan antara XGBoost dan KNN, SVM, serta Naive Bayes, tetapi tidak menunjukkan perbedaan signifikan terhadap Decision Tree dan Random Forest. Hasil per kelas juga menegaskan bahwa ketidakseimbangan data tetap menjadi keterbatasan utama, terutama karena kelas gizi buruk hanya memiliki satu observasi pada data uji dan tidak berhasil dikenali model akhir.

Tujuan keempat dijawab melalui implementasi model akhir dalam prototipe Streamlit yang mengintegrasikan *input* data, prapemrosesan, evaluasi, dan klasifikasi dalam satu alur. Kontribusi ilmiah penelitian ini terletak pada integrasi pencegahan *target leakage*, perbandingan enam keluarga algoritma pada

data uji yang sama, evaluasi peka kelas, pengujian statistik berpasangan, dan implementasi prototipe dalam kerangka CRISP-DM yang dapat ditelusuri. Secara praktis, prototipe berpotensi mendukung penyaringan awal atau peninjauan konsistensi catatan, tetapi tidak dapat menggantikan standar antropometri WHO maupun penilaian tenaga kesehatan. Generalisasi hasil masih terbatas karena evaluasi bersifat internal pada tingkat catatan dan distribusi kelas sangat tidak seimbang. Penelitian selanjutnya perlu menambah data kelas minoritas, melakukan pemisahan dan validasi berbasis anak unik, menguji data dari periode serta fasilitas berbeda, mengevaluasi kalibrasi probabilitas, dan menilai kegunaan aplikasi dalam alur kerja pelayanan.

REFERENSI

- [1] S. D. G. Indicators, *The UNICEF-WHO- World Bank Joint Child Malnutrition Estimates (JME) standard methodology Tracking progress on SDG Indicators*, vol. 2, no. 1.
- [2] M. Djaenudin and Tupan, "Pengelolaan Data Penelitian pada Repositori Pengetahuan di Perpustakaan Khusus Lembaga Pemerintah Non Kementerian Lembaga Pemerintah Non Kementerian yang berada di bawah kordinasi Kementerian n," vol. 27, no. 3, pp. 195–211, 2020.
- [3] W. H. O. Child and G. Standards, "Weight-for-length / height GIRLS Weight-for-length / height BOYS," no. cm.
- [4] M. Yanto, F. Hadi, and S. Arlis, "Determination of children ' s nutritional status with machine learning classification analysis approach," vol. 33, no. 1, pp. 303–313, 2024, doi: 10.11591/ijeecs.v33.i1.pp303-313.
- [5] P. Handayani and A. C. Fauzan, "Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest," vol. 4, no. 6, pp. 3064–3072, 2024, doi: 10.30865/klik.v4i6.1909.
- [6] S. Chang, X. Wang, Y. Luo, and L. Jia, "Data augmentation alters feature importance in XGBoost for CVD prediction," pp. 1–11, 2025.
- [7] A. Rifai, S. Seimahaira, and E. H. Juningsih, "Klasifikasi Status Gizi Anak Berdasarkan Indikator BB / TB Menggunakan Gradient Boosting Classifier," vol. 8, no. 4, pp. 2011–2017, 2025.
- [8] A. Pangestu *et al.*, "Implementasi Algoritma XGBoost Untuk Prediksi Status Gizi Balita Berbasis Website," pp. 176–187, 2025, doi: 10.33364/algoritma/v.22-2.2390.
- [9] I. G. Huda *et al.*, "TIN : Terapan Informatika Nusantara Pengembangan Aplikasi Mobile Berbasis Z-score untuk Deteksi Dini dan Pemantauan Status Gizi Balita Menggunakan Standar Antropometri TIN : Terapan Informatika Nusantara," vol. 6, no. 6, pp. 610–621, 2025, doi: 10.47065/tin.v6i6.8553.
- [10] G. P. Insany, I. Yustiana, S. Rahmawati, T. Informatika, and U. N. Putra, "Jurnal Computer Science and Information Technology (CoSciTech)," vol. 4, no. 2, pp. 385–393, 2023.
- [11] M. I. Abas, R. Lamusu, W. E. Pranata, S. Syahrial, and I. Ibrahim, "BULLETIN OF COMPUTER SCIENCE RESEARCH Penerapan Algoritma Naive Bayes Untuk Sistem Klasifikasi Status Gizi Bayi Balita," vol. 5, no. 5, pp. 1249–1260, 2025, doi: 10.47065/bulletincsr.v5i5.508.
- [12] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K- Nearest Neighbor Berbasis Web," vol. 6, no. 2, pp. 777–785, 2022, doi: 10.30865/mib.v6i2.3566.
- [13] S. Kapoor and A. Narayanan, "Leakage and the Reproducibility Crisis in ML-based Science," no. ML, 2020.
- [14] M. A. Bouke, S. A. Zaid, and A. Abdullah, "Implications of Data Leakage in Machine Learning Preprocessing : A Multi-Domain Investigation".
- [15] J. Opitz, "A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice," vol. 12, no. 2018, pp. 820–836, 2024.
- [16] E. Richardson, R. Trevizani, J. A. Greenbaum, and H. Carter, "Article The receiver operating characteristic curve accurately assesses imbalanced *datasets* The receiver operating characteristic curve accurately assesses imbalanced *datasets*," *Patterns*, vol. 5, no. 6, p. 100994, 2024, doi: 10.1016/j.patter.2024.100994.
- [17] O. Rainio, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, pp. 1–14, 2024, doi: 10.1038/s41598-024-56706-x.
- [18] Y. A. Singgalen, "Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT)," vol. 7, pp. 1551–1562, 2023, doi: 10.30865/mib.v7i3.6461.
- [19] M. Erkamim, A. M. Tanniewa, and I. Ap, "Analisis Perbandingan Metode AdaBoost , Gradient Boosting , dan XGBoost Untuk Kalsifikasi Status Gizi Pada Balita," vol. 6, no. 3, pp. 1799–1807, 2024, doi: 10.47065/bits.v6i3.5717.
- [20] M. Arifin, W. Widowati, and F. Farikhin, "Ingénierie des Systèmes d ' Information Optimization of Hyperparameters in Machine Learning for Enhancing Predictions of Student Academic Performance," vol. 28, no. 3, pp. 575–582, 2023.