Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

EVALUASI KOMPARATIF *MULTINOMIAL LOGISTIC REGRESSION, RANDOM FOREST,* DAN *LIGHTGBM* PADA KLASIFIKASI STATUS GIZI BALITA BERBASIS DATA ANTROPOMETRI TIDAK SEIMBANG

Erip Suratno[1], Adhi Kusnadi[2]
[1,2] *Universitas Nusa Putra*
erip.suratno_mif24@nusaputra.ac.id

ABSTRAK

Klasifikasi status gizi balita berbasis antropometri mendukung pemantauan kesehatan secara cepat dan konsisten. Namun, distribusi kelas yang tidak seimbang membuat akurasi konvensional tidak memadai sebagai metrik evaluasi tunggal. Penelitian ini bertujuan membandingkan kinerja pipeline *"Multinomial Logistic Regression","Random Forest"*, dan *"LightGBM"* dalam klasifikasi multikelas status gizi balita pada dataset antropometri yang tidak seimbang. Penelitian kuantitatif ini menggunakan data sekunder dari Posyandu Desa Ujunggenteng, UPTD Puskesmas Ciracap tahun 2025 sebanyak 3.716 sampel balita dengan fitur jenis kelamin, usia, berat badan, tinggi badan, serta label status gizi. Evaluasi model dilakukan melalui *"Stratified Train-Test Split"*, "*Stratified K-Fold Cross Validation"*,"GridSearchCV", serta penanganan ketidakseimbangan data menggunakan "class weight" dan "SMOTENC, dengan fokus utama pada metrik "i Hasil pengujian menunjukkan bahwa pipeline "LGBM_SMOTENC" memberikan performa terbaik dengan *"Best CV F1 Macro"* sebesar 0,829, "Test F1 Macro" 0,844, i 0,956, dan *"Test Balanced Accuracy"* 0,858. Secara rinci, *"F1-score"* per kelas mencapai 0,800 untuk kurang gizi, 0,976 untuk gizi baik, 0,831 untuk risiko gizi lebih, dan 0,769 untuk gizi lebih. Dengan demikian, "LightGBM" disertai *"SMOTENC"* terbukti paling konsisten dan efektif, menegaskan bahwa evaluasi pada data tidak seimbang wajib mempertimbangkan *"F1 Macro"* dan *"Balanced Accuracy*" di samping akurasi.
Kata Kunci: Status Gizi Balita, Antropometri, Data Tidak Seimbang, F1 Macro, LightGBM

ABSTRACT

Classification of the nutritional status of under-age children based on anthropometry supports quick and consistent health monitoring. However, unbalanced class distribution and conventional accuracy are insufficient as single evaluation metrics. This research aims to compare the performance of "Multinomial Logistic Regression", "Random Forest", and "LightGBM" pipelines in the multiclass classification of nutritional status on unbalanced anthropometric datasets. This quantitative research uses secondary data from Posyandu Desa Ujunggenteng, UPTD Puskesmas Ciracap in 2025 as many as 3,716 samples of children under five with the characteristics of gender, age, weight, height, and nutritional status labels. The evaluation model is performed through "Stratified Train-Test Split", "Stratified K-Fold Cross Validation", "GridSearchCV", as well as the handling of data imbalance using "class weight" and "SMOTENC, with the main focus on the metric "i The test results show that the pipeline "LGBM_SMOTENC" provides the best performance with "Best CV F1 Macro" of 0.829 "Macro", 0.829 0.956, and "Balanced Accuracy Test" 0.858. In detail, the "F1-score" of each class reached 0.800 for malnutrition, 0.976 for good nutrition, 0.831 for the risk of overnutrition, and 0.769 for overnutrition. Thus, "LightGBM" accompanied by "SMOTENC" proved to be the most effective and efficient, insisting that evaluation on unbalanced data must consider "F1 Macro" and "Balanced Accuracy" in addition to accuracy.
Keywords: *Toddler Nutritional Status, Anthropometry, Imbalanced Data, F1 Macro, LightGBM**

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

Pendahuluan

Status gizi balita merupakan indikator penting dalam menilai kualitas pertumbuhan, kesehatan, dan risiko gangguan perkembangan anak. Pada kelompok usia dini, gangguan gizi tidak hanya berdampak pada kondisi fisik jangka pendek, tetapi juga dapat memengaruhi kemampuan kognitif, kerentanan terhadap penyakit, serta produktivitas pada masa dewasa. Pemantauan status gizi secara rutin menjadi bagian penting dari sistem kesehatan masyarakat, terutama di wilayah dengan beban gizi ganda, yaitu ketika masalah kekurangan gizi dan kelebihan gizi muncul dalam populasi yang sama.

Antropometri masih menjadi pendekatan yang luas digunakan dalam penilaian status gizi anak karena relatif murah, tidak invasif, mudah diperoleh, dan dapat diterapkan pada skala layanan primer. Variabel seperti jenis kelamin, usia, berat badan, dan tinggi badan menjadi dasar penting dalam berbagai skema penilaian pertumbuhan anak. Kombinasi ukuran antropometri dapat menghasilkan kategori status gizi yang berbeda dan memiliki implikasi klinis maupun epidemiologis yang tidak sama [1].

Perkembangan machine learning membuka peluang untuk membangun model klasifikasi yang mampu mengenali pola dari data antropometri dan data kesehatan anak. Penggunaan data Bangladesh Demographic and Health Survey menunjukkan bahwa Random Forest memiliki performa lebih baik dibandingkan beberapa algoritma lain dalam prediksi malnutrisi anak di bawah lima tahun [2]. Pada konteks Ethiopia, pendekatan klasifikasi berbasis Random Forest dilaporkan efektif untuk undernutrition anak [3]. Pendekatan berbasis machine learning telah digunakan untuk menganalisis undernutrition anak berdasarkan data demografis dan kesehatan) [4]. Algoritma Boruta telah digunakan untuk menyeleksi determinan multidimensi malnutrisi anak di bawah lima tahun di Pakistan [5]. Perbandingan machine learning dengan logistic regression tradisional telah digunakan untuk mengevaluasi performa prediksi dan identifikasi faktor risiko luaran gizi anak di Pakistan [6]. Pemodelan multinomial logistic regression dan proportional odds regression telah digunakan untuk menganalisis determinan malnutrisi anak di bawah lima tahun pada data Bangladesh [7].

Klasifikasi status gizi balita memiliki tantangan metodologis yang tidak sederhana. Salah satu tantangan utama adalah distribusi kelas yang tidak seimbang. Dalam data status gizi, kelas normal atau gizi baik umumnya memiliki jumlah dominan, sedangkan kelas minoritas seperti kurang gizi atau gizi lebih jauh lebih kecil. Ketidakseimbangan kelas dapat membuat model bias terhadap kelas mayoritas, sehingga metrik evaluasi perlu memperhatikan performa kelas minoritas [8] Pada dataset penelitian ini, kondisi tersebut terlihat dari dominasi kelas gizi baik dan sangat kecilnya jumlah kelas gizi lebih, sehingga F1 Macro lebih aman digunakan sebagai metrik utama daripada akurasi.

Masalah ketidakseimbangan kelas juga ditemukan dalam penelitian ini. Dataset yang digunakan terdiri dari 3.716 data balita dengan empat kelas status gizi. Kelas gizi baik mendominasi 86,65% data, sedangkan kelas gizi lebih hanya 0,83%. Kondisi tersebut menuntut penggunaan metrik evaluasi yang lebih adil terhadap seluruh kelas. Penelitian ini menggunakan F1 Macro sebagai metrik utama, karena metrik tersebut menghitung rata-rata performa setiap kelas tanpa membobotinya berdasarkan jumlah sampel. Balanced Accuracy, Precision Macro, Recall Macro, F1 Weighted, dan ROC AUC OVR digunakan sebagai metrik pendukung.

Penelitian ini membandingkan tiga keluarga model, yaitu Multinomial Logistic Regression, Random Forest, dan LightGBM. Multinomial Logistic Regression dipilih sebagai baseline linear yang interpretatif dan umum digunakan dalam analisis klasifikasi multikelas. Random Forest dipilih karena mampu menangkap hubungan non-linear dan sering menunjukkan performa baik pada data berbentuk tabel.

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

LightGBM dipilih sebagai model gradient boosting modern yang efisien dan banyak digunakan dalam klasifikasi berbasis data terstruktur. Agar perbandingan tidak hanya mencerminkan perbedaan algoritma, tetapi juga strategi penanganan ketidakseimbangan kelas, eksperimen dirancang pada level pipeline. Random Forest dan LightGBM dievaluasi menggunakan varian class weight dan SMOTENC, sedangkan MLR digunakan sebagai baseline class-weighted.

Kontribusi penelitian ini adalah sebagai berikut. Pertama, penelitian ini mengevaluasi tiga keluarga model machine learning untuk klasifikasi multikelas status gizi balita menggunakan fitur antropometri dasar. Kedua, penelitian ini menempatkan ketidakseimbangan kelas sebagai isu metodologis utama, sehingga evaluasi tidak hanya bertumpu pada akurasi. Ketiga, penelitian ini membandingkan pipeline class weight dan SMOTENC untuk melihat pengaruh strategi imbalance terhadap performa model. Keempat, penelitian ini membandingkan performa validasi silang dan test set untuk menentukan model final secara lebih hati-hati. Kelima, penelitian ini menyajikan analisis per kelas agar performa pada kelas minoritas dapat terlihat secara eksplisit.

Materi dan Metode

1. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif berbasis data sekunder dengan desain eksperimen komputasional yang bersifat evaluatif-komparatif. Pendekatan yang digunakan adalah *supervised learning*, karena penelitian ini memanfaatkan fitur prediktor berupa jenis kelamin, usia, berat badan, dan tinggi badan untuk mengklasifikasikan label status gizi yang telah tersedia pada *dataset*. Rancangan penelitian bersifat komparatif karena membandingkan performa beberapa *pipeline* klasifikasi multikelas, yaitu *Multinomial Logistic Regression*, *Random Forest*, dan *LightGBM*, termasuk strategi penanganan ketidakseimbangan kelas melalui *class weight* dan *SMOTENC*. Penelitian ini tidak memberikan intervensi kepada subjek, tetapi berfokus pada eksperimen pemodelan dan evaluasi performa algoritma terhadap data sekunder. Alur penelitian meliputi pemeriksaan *dataset*, prapemrosesan data, pembagian data latih dan data uji, pelatihan model, *hyperparameter tuning*, validasi silang, evaluasi pada *test set*, serta pemilihan *pipeline final* berdasarkan metrik utama *F1 Macro*.

2. Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah data balita pada Posyandu Desa Ujunggenteng, UPTD Puskesmas Ciracap, tahun 2025 yang memiliki atribut antropometri dasar berupa jenis kelamin, usia, berat badan, tinggi badan, dan label status gizi. Jenis data yang digunakan adalah data sekunder. Sumber data sekaligus instrumen data penelitian adalah arsip *dataset* Posyandu Desa Ujunggenteng yang berasal dari UPTD Puskesmas Ciracap. Populasi penelitian dibatasi pada data sekunder yang tersedia dari sumber tersebut, bukan seluruh populasi balita di luar cakupan data Posyandu Desa Ujunggenteng. Informasi mengenai prosedur pencatatan data lapangan dan kriteria inklusi-eksklusi klinis tidak dianalisis lebih lanjut dalam penelitian ini, sehingga interpretasi populasi diarahkan pada populasi operasional sesuai data yang diterima dan dianalisis.

Teknik pengambilan sampel dalam konteks analisis data adalah *total sampling* terhadap seluruh data sekunder yang tersedia dan memenuhi struktur *dataset* penelitian. Sampel penelitian adalah 3.716 baris data balita, dengan unit analisis berupa satu baris data balita. *Dataset* terdiri dari variabel jk, usia (bulan), berat (kg), tinggi (cm), dan status gizi. Berdasarkan pemeriksaan *dataset*, tidak ditemukan *missing value* maupun duplikasi baris. Rentang usia balita pada data adalah 0-59 bulan, rentang berat badan 2,5-19,6 kg, dan rentang tinggi badan 48-114 cm. Variabel

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

jk dikodekan sebagai 1 untuk laki-laki dan 2 untuk perempuan, dengan distribusi laki-laki sebanyak 2.078 data (55,92%) dan perempuan sebanyak 1.638 data (44,08%). Label status gizi dikodekan sebagai 1 = kurang gizi; 2 = gizi baik; 3 = risiko gizi lebih; dan 4 = gizi lebih. Distribusi status gizi terdiri dari kurang gizi sebanyak 74 data (1,99%); gizi baik sebanyak 3.220 data (86,65%); risiko gizi lebih sebanyak 391 data (10,52%); dan gizi lebih sebanyak 31 data (0,83%). Untuk evaluasi model, sampel dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan *stratified train-test split*, sehingga diperoleh 2.972 data latih dan 744 data uji.

Dataset

Dataset yang digunakan berjumlah 3.716 baris dan terdiri dari lima kolom, yaitu jk, usia (bulan), berat (kg), tinggi (cm), dan status gizi. Empat kolom pertama digunakan sebagai fitur prediktor, sedangkan kolom status gizi digunakan sebagai target klasifikasi. Tidak ditemukan *missing value* maupun duplikasi baris pada *dataset.* Kolom status gizi diperlakukan sebagai label operasional yang sudah tersedia dalam arsip data Posyandu Desa Ujunggenteng dari UPTD Puskesmas Ciracap. Penelitian ini tidak melakukan rekonstruksi ulang label berdasarkan e-PPGBM, standar Kemenkes, atau standar *WHO* berbasis *z-score*; model dilatih dan dievaluasi untuk memprediksi label arsip tersebut berdasarkan fitur jenis kelamin, usia, berat badan, dan tinggi badan yang tersedia.

Tabel 1 menunjukkan struktur variabel yang digunakan dalam penelitian. Variabel jk, usia, berat badan, dan tinggi badan berperan sebagai fitur prediktor yang digunakan oleh model untuk mengklasifikasikan status gizi balita. Variabel jk merupakan fitur kategorikal numerik dengan kode 1 untuk laki-laki dan 2 untuk perempuan, sedangkan usia, berat badan, dan tinggi badan merupakan fitur numerik yang masing-masing dinyatakan dalam bulan, kilogram, dan sentimeter. Variabel Status Gizi berperan sebagai target klasifikasi multikelas yang terdiri dari empat kategori, yaitu 1 = kurang gizi; 2 = gizi baik; 3 = risiko gizi lebih; dan 4 = gizi lebih. Dengan struktur tersebut, penelitian ini menggunakan data antropometri dasar dan jenis kelamin sebagai masukan model untuk memprediksi kelas status gizi balita.

Tabel 1. Fitur dan target penelitian

Variabel	Peran	Tipe	Keterangan
JK	Fitur	Kategorikal numerik	Jenis kelamin: 1 = laki-laki; 2 = perempuan
Usia (bulan)	Fitur	Numerik	Usia balita dalam bulan
Berat (kg)	Fitur	Numerik	Berat badan balita dalam kilogram
Tinggi (cm)	Fitur	Numerik	Tinggi badan balita dalam sentimeter
Status Gizi	Target	Kategorikal multikelas	Kelas status gizi: 1 = kurang gizi; 2 = gizi baik; 3 = risiko gizi lebih; 4 = gizi lebih

Tabel 2 menunjukkan distribusi kelas status gizi pada *dataset* penelitian. Dari total 3.716 data balita, kelas gizi baik merupakan kelas mayoritas dengan 3.220 data atau 86,65% dari keseluruhan *dataset*. Sebaliknya, kelas gizi lebih merupakan kelas paling minoritas dengan 31 data atau 0,83%, diikuti kelas kurang gizi sebanyak 74 data atau 1,99%. Kelas risiko gizi lebih berjumlah 391 data atau 10,52%. Distribusi ini menunjukkan adanya ketidakseimbangan kelas yang kuat, karena sebagian besar data terkonsentrasi pada kelas gizi baik, sedangkan kelas lainnya memiliki jumlah data jauh lebih kecil. Kondisi tersebut menjadi dasar penting untuk menggunakan metrik evaluasi seperti *F1 Macro* dan *Balanced Accuracy*, agar performa model tidak hanya dinilai dari kelas mayoritas.

Tabel 2. Distribusi kelas status gizi

Kelas	Nama 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

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

Prapemrosesan Data

Prapemrosesan dilakukan untuk memastikan data dapat digunakan secara konsisten dalam *pipeline machine learning*. Nilai jenis kelamin dinormalisasi ke dalam kode numerik yang *valid*. Variabel usia, berat, tinggi, dan target dikonversi ke tipe numerik. Data diperiksa untuk mendeteksi *missing value*, duplikasi, label target yang tidak valid, serta nilai numerik negatif. Karena *dataset* tidak memiliki *missing value*, proses imputasi tetap dimasukkan ke dalam *pipeline* sebagai langkah pengamanan agar model tetap *robust* apabila digunakan pada data baru.

Pada *Multinomial Logistic Regression*, fitur numerik distandardisasi menggunakan *StandardScaler* karena model linear sensitif terhadap skala fitur. Pada *Random Forest* dan *LightGBM*, standardisasi tidak diwajibkan karena model berbasis pohon relatif tidak sensitif terhadap skala numerik.

Untuk *pipeline* yang menggunakan *SMOTENC*, *oversampling* ditempatkan di dalam *pipeline* pelatihan dan dijalankan hanya pada data latih pada setiap proses validasi silang. Penempatan ini penting untuk mencegah *data leakage*, karena data validasi dan data uji tidak boleh ikut digunakan dalam pembentukan sampel sintetis.

Pembagian Data

Dataset dibagi menjadi data latih dan data uji dengan proporsi 80:20. Pembagian dilakukan menggunakan *stratified train-test split* agar proporsi setiap kelas pada data latih dan data uji tetap mengikuti distribusi kelas asal. Penggunaan stratifikasi penting karena kelas minoritas memiliki jumlah sampel yang sangat kecil.

Tabel 3 menunjukkan distribusi data latih dan data uji setelah dilakukan pembagian *dataset* menggunakan pendekatan *stratified train-test split*. Secara umum, proporsi setiap kelas pada data latih dan data uji relatif konsisten dengan distribusi kelas pada *dataset* awal. Kelas gizi baik tetap menjadi kelas mayoritas, yaitu 2.575 data pada data latih dan 645 data pada data uji, dengan proporsi masing-masing 86,64% dan 86,69%. Kelas minoritas juga tetap terwakili pada kedua *subset*, yaitu kurang gizi sebanyak 59 data latih dan 15 data uji, risiko gizi lebih sebanyak 313 data latih dan 78 data uji, serta gizi lebih sebanyak 25 data latih dan 6 data uji. Kesesuaian proporsi ini menunjukkan bahwa proses stratifikasi berhasil menjaga representasi kelas pada data latih dan data uji.

Tabel 3. Distribusi Data Latih Dan Data Uji

Kelas	Nama Kelas	*Train*	*Test*	*Train* (%)	*Test* (%)
1	Kurang Gizi	59	15	1,99	2,02
2	Gizi Baik	2.575	645	86,64	86,69
3	Risiko Gizi Lebih	313	78	10,53	10,48
4	Gizi Lebih	25	6	0,84	0,81

Model yang Dibandingkan

Multinomial Logistic Regression

Multinomial Logistic Regression digunakan sebagai model linear baseline untuk klasifikasi multikelas. Model ini sesuai untuk membandingkan apakah hubungan linear antara fitur antropometri dan status gizi sudah cukup memadai. Pipeline model terdiri dari imputasi, standardisasi fitur numerik, dan klasifikasi multinomial. Class weight digunakan untuk membantu model memperhatikan kelas minoritas.

Random Forest

Random Forest merupakan metode ensemble berbasis banyak pohon keputusan yang digunakan untuk menangkap pola non-linear pada data berbentuk tabel. Dalam penelitian ini, Random Forest digunakan sebagai pembanding karena model tersebut sering kompetitif pada studi prediksi status gizi anak. Pada studi undernutrition anak di Ethiopia, Random Forest dipilih sebagai model machine learning terbaik berdasarkan beberapa evaluasi performa (Fenta et al., 2021)[3].

Random Forest dievaluasi dalam dua pipeline. Pipeline pertama menggunakan class weight untuk memberikan penalti lebih besar terhadap kesalahan pada kelas minoritas. Pipeline kedua menggunakan SMOTENC pada data latih

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

untuk menangani ketidakseimbangan kelas dengan mempertimbangkan keberadaan fitur kategorikal. Evaluasi empiris pada data tidak seimbang menunjukkan bahwa pengaruh metode sampling dapat berbeda antar kombinasi dataset dan algoritma [8].). Analisis teoretis yang lebih baru menunjukkan bahwa sampel sintetis SMOTE tidak selalu mengikuti distribusi asli kelas minoritas) [9]. Penggunaan SMOTENC hanya diterapkan pada data latih di dalam pipeline untuk menghindari kebocoran informasi ke data uji.

LightGBM

LightGBM merupakan model gradient boosting berbasis pohon yang efisien untuk data berbentuk tabel. Model ini dipilih karena mampu membangun learner secara bertahap dan sering menunjukkan performa tinggi pada masalah klasifikasi. Pada prediksi malnutrisi akut anak di Afrika Timur, LightGBM digunakan sebagai salah satu algoritma pembanding bersama Random Forest, XGBoost, CatBoost, dan beberapa model lain [10].

Pada penelitian ini, LightGBM dievaluasi dalam dua pipeline, yaitu LightGBM dengan class weight dan LightGBM dengan SMOTENC. Varian SMOTENC disertakan untuk menguji apakah peningkatan representasi kelas minoritas pada data latih dapat memperbaiki F1 Macro dibandingkan pembobotan kelas saja.

Skema Pipeline Eksperimen

Eksperimen dirancang sebagai evaluasi pipeline, bukan perbandingan algoritma mentah. Skema ini digunakan karena performa model pada dataset tidak seimbang dipengaruhi oleh algoritma dan strategi penanganan kelas minoritas.

Tabel 4 menunjukkan lima pipeline eksperimen yang dibandingkan dalam penelitian. MLR_CW digunakan sebagai baseline linear dengan model Multinomial Logistic Regression dan strategi class weight untuk menangani ketidakseimbangan kelas. Random Forest dievaluasi dalam dua skema, yaitu RF_CW dengan pembobotan kelas dan RF_SMOTENC dengan oversampling data latih menggunakan SMOTENC. LightGBM juga dievaluasi dalam dua skema, yaitu LGBM_CW dengan class weight dan LGBM_SMOTENC dengan SMOTENC. Evaluasi tidak hanya membandingkan model, tetapi juga menilai apakah pendekatan pembobotan kelas atau oversampling lebih efektif dalam meningkatkan performa klasifikasi pada dataset status gizi yang tidak seimbang.

Tabel 4. Pipeline Eksperimen

Pipeline	**Model**	**Strategi *Imbalance***	**Keterangan**
MLR_CW	*Multinomial Logistic Regression*	*Class weight*	*Baseline linear* dengan pembobotan kelas
RF_CW	*Random Forest*	*Class weight*	*Ensemble bagging* dengan pembobotan kelas
RF_SMOTENC	*Random Forest*	*SMOTENC*	*Ensemble bagging* dengan *oversampling* data latih
LGBM_CW	*LightGBM*	*Class weight*	*Gradient boosting* dengan pembobotan kelas
LGBM_SMOTENC	*LightGBM*	*SMOTENC*	*Gradient boosting* dengan *oversampling* data latih

***Tuning Hyperparameter* dan Validasi Silang**

Tuning hyperparameter dilakukan menggunakan *GridSearchCV* dengan *grid standard* dan *Stratified K-Fold Cross Validation*. Stratifikasi pada validasi silang digunakan untuk menjaga proporsi kelas pada setiap *fold*. Metrik seleksi utama pada *GridSearchCV* adalah *F1 Macro*. Pemilihan *F1 Macro* bertujuan agar performa model pada kelas minoritas ikut menentukan proses pemilihan *hyperparameter*, bukan hanya performa pada kelas mayoritas.

Validasi silang dilakukan dengan lima *fold*, pengacakan data, dan *random state* tetap agar eksperimen dapat direplikasi. *GridSearchCV* diterapkan pada data latih, sedangkan data uji

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

dipertahankan sebagai evaluasi akhir setelah kombinasi *hyperparameter* terbaik diperoleh. Untuk *SMOTENC,* indeks fitur kategorikal ditetapkan pada variabel jk setelah tahap imputasi, sehingga pembentukan sampel sintetis tetap memperlakukan jenis kelamin sebagai fitur kategorikal.

Lingkungan Analisis

Analisis dilakukan dalam *Python* dengan pustaka *machine learning* yang sesuai dengan *pipeline* penelitian. *Scikit-learn* digunakan untuk pembagian data, prapemrosesan, validasi silang, *GridSearchCV, Multinomial Logistic Regression*, dan *Random Forest. Imbalanced-learn* digunakan untuk implementasi *SMOTENC,* sedangkan *LightGBM* digunakan untuk model *gradient boosting*. Eksperimen final dijalankan dengan *scikit-learn 1.8.0, imbalanced-learn 0.14.1,* dan *LightGBM 4.6.0. Seed* atau *random_state* yang digunakan pada pembagian data, validasi silang, model, dan *SMOTENC* adalah 42.

Metrik Evaluasi

Metrik evaluasi yang digunakan meliputi *Accuracy, Balanced Accuracy, Precision Macro, Recall Macro, F1 Macro, Precision Weighted, Recall Weighted, F1 Weighted,* dan *ROC AUC OVR. F1 Macro* digunakan sebagai metrik utama karena menghitung rata-rata *F1-score* setiap kelas secara setara. Pada *dataset* tidak seimbang, metrik ini lebih sesuai dibandingkan akurasi karena tidak didominasi oleh kelas mayoritas.

Balanced Accuracy digunakan sebagai metrik pendukung karena menghitung rata-rata *recall* dari setiap kelas. *Precision* dan *recall* per kelas juga dianalisis untuk melihat *trade-off* antara kemampuan model menemukan suatu kelas dan ketepatan prediksi pada kelas tersebut.

Hasil dan Pembahasan

Hasil

Berdasarkan desain penelitian kuantitatif eksperimen komputasional, hasil penelitian disajikan dalam bentuk analisis univariat dan analisis bivariat-komparatif. Analisis univariat digunakan untuk menggambarkan karakteristik *dataset* dan distribusi kelas status gizi. Analisis bivariat-komparatif digunakan untuk membandingkan *pipeline*/model terhadap metrik evaluasi pada validasi silang dan *test set*. Analisis multivariat inferensial tidak dilakukan karena penelitian ini tidak bertujuan menguji pengaruh independen masing-masing variabel terhadap status gizi, penggunaan beberapa variabel prediktor secara bersamaan diwujudkan dalam pemodelan *supervised learning.*

Analisis Univariat : Karakteristik *Dataset* dan Distribusi Kelas

Analisis univariat menunjukkan bahwa *dataset* terdiri dari 3.716 data balita. Berdasarkan jenis kelamin, data laki-laki berjumlah 2.078 data (55,92%) dan perempuan berjumlah 1.638 data (44,08%). Rentang usia balita adalah 0-59 bulan; rentang berat badan 2,5-19,6 kg; dan rentang tinggi badan 48-114 cm. Distribusi status gizi menunjukkan bahwa *dataset* sangat didominasi oleh kelas gizi baik. Dari 3.716 data, sebanyak 3.220 data (86,65%) berada pada kelas gizi baik; 391 data (10,52%) pada kelas risiko gizi lebih; 74 data (1,99%) pada kelas kurang gizi; dan 31 data (0,83%) pada kelas gizi lebih.

Distribusi tersebut menunjukkan ketidakseimbangan kelas yang kuat. Kelas gizi baik menjadi kelas mayoritas, sedangkan kelas gizi lebih menjadi kelas paling minoritas. Pada data uji, kelas gizi lebih hanya berjumlah 6 data. Kondisi ini menyebabkan evaluasi kelas minoritas menjadi sensitif terhadap perubahan kecil pada jumlah prediksi benar atau salah. Interpretasi hasil penelitian tidak difokuskan hanya pada akurasi, tetapi juga pada *F1 Macro, Balanced Accuracy,* dan performa per kelas.

Analisis Bivariat-Komparatif : Hasil Validasi Silang

Analisis bivariat-komparatif pada tahap validasi silang dilakukan dengan membandingkan setiap *pipeline*/model terhadap nilai *F1 Macro* sebagai metrik seleksi utama. Tabel 5 menunjukkan hasil validasi silang setiap *pipeline* berdasarkan metrik

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

utama *F1 Macro*. *LGBM_SMOTENC* memperoleh *Best CV F1 Macro* tertinggi sebesar 0,829; diikuti *RF_SMOTENC* sebesar 0,802; *LGBM_CW* sebesar 0,775; *RF_CW* sebesar 0,768; dan *MLR_CW* sebesar 0,724. Hasil ini menunjukkan bahwa strategi *oversampling* berbasis *SMOTENC* memberikan peningkatan yang jelas pada *LightGBM* dan tetap kompetitif pada *Random Forest*. *LGBM_CW* memiliki *Train-CV Gap* terbesar sebesar 0,213; yang menunjukkan adanya selisih cukup besar antara performa data latih dan validasi. *LGBM_SMOTENC* memberikan kombinasi terbaik antara nilai *F1 Macro* validasi yang tinggi dan selisih *train-validasi* yang masih lebih terkendali dibandingkan beberapa *pipeline*.

Tabel 5. Hasil Validasi Silang Antar *Pipeline*

Pipeline	*Best CV F1 Macro*	*CV F1 Macro Std*	*Mean Train F1 Macro*	*Train-CV Gap*
Mlr_Cw	0,724	0,032	0,731	0,007
Rf_Cw	0,768	0,066	0,966	0,198
Rf_Smotenc	0,802	0,046	0,956	0,155
Lgbm_Cw	0,775	0,093	0,988	0,213
Lgbm_Smotenc	0,829	0,054	0,992	0,163

Analisis Bivariat-Komparatif : Hasil Evaluasi pada *Test Set*

Analisis bivariat-komparatif pada *test set* dilakukan dengan membandingkan *pipeline*/model terhadap *Accuracy*, *Balanced Accuracy*, *Precision Macro*, *Recall Macro*, *F1 Macro*, dan *F1 Weighted*. Tabel 6 menunjukkan bahwa *LGBM_SMOTENC* memiliki performa paling kuat berdasarkan kombinasi metrik utama, yaitu *Accuracy* sebesar 0,956; *F1 Macro* sebesar 0,844; *F1 Weighted* sebesar 0,956; dan *Balanced Accuracy* sebesar 0,858. *MLR_CW* memperoleh *Balanced Accuracy* dan *Recall Macro* tertinggi sebesar 0,911; tetapi *Precision Macro*-nya rendah sebesar 0,597 sehingga *F1 Macro*-nya hanya 0,676. Temuan ini menunjukkan bahwa *MLR_CW* lebih sensitif dalam mengenali kelas minoritas, tetapi menghasilkan lebih banyak prediksi positif yang kurang tepat. Dengan mempertimbangkan *F1 Macro*, *Precision Macro*, akurasi, dan evaluasi per kelas, *LGBM_SMOTENC* menjadi *pipeline* paling layak dipilih pada *dataset* tidak seimbang.

Tabel 6. Komparasi Metrik *Test Set*

Pipeline	*Accuracy*	*Balanced Accuracy*	*Precision Macro*	*Recall Macro*	*F1 Macro*	*F1 Weighted*
Mlr_Cw	0,883	0,911	0,597	0,911	0,676	0,904
RF_Czw	0,950	0,754	0,829	0,754	0,785	0,950
Rf_Smotenc	0,946	0,761	0,817	0,761	0,783	0,945
Lgbm_Cw	0,953	0,808	0,840	0,808	0,822	0,953
Lgbm_Smotenc	0,956	0,858	0,833	0,858	0,844	0,956

Evaluasi Multikelas : Analisis Per Kelas

Evaluasi per kelas merupakan rata-rata performa dari seluruh kelas. Tanpa evaluasi per kelas, performa model pada kelas minoritas dapat tertutupi oleh performa kelas mayoritas.

Tabel 7 menunjukkan perbandingan nilai *F1-score* setiap *pipeline* pada masing-masing kelas status gizi. *LGBM_SMOTENC* memperoleh nilai tertinggi pada seluruh kelas, yaitu 0,800 pada kelas kurang gizi; 0,976 pada gizi baik; 0,831 pada risiko gizi lebih; dan 0,769 pada gizi lebih. Hasil ini menunjukkan bahwa *LGBM_SMOTENC* tidak hanya kuat pada kelas mayoritas, tetapi juga memberikan performa yang lebih baik pada kelas minoritas dibandingkan *pipeline* lainnya. *LGBM_CW* menjadi pembanding terdekat, terutama pada kelas risiko gizi lebih dan gizi lebih, tetapi nilainya masih berada di bawah *LGBM_SMOTENC*. *MLR_CW* memiliki performa paling rendah pada kelas kurang gizi dengan nilai 0,385; meskipun cukup baik pada kelas risiko gizi lebih. Secara keseluruhan, hasil per kelas memperkuat bahwa kombinasi *LightGBM* dan *SMOTENC* memberikan keseimbangan performa terbaik antar kelas.

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

Tabel 7. Komparasi *F1-score* per kelas

Pipeline	Kurang Gizi	Gizi Baik	Risiko Gizi Lebih	Gizi Lebih
MLR_CW	0,385	0,932	0,798	0,588
RF_CW	0,759	0,974	0,808	0,600
RF_SMOTENC	0,774	0,972	0,784	0,600
LGBM_CW	0,759	0,975	0,827	0,727
LGBM_SMOTENC	0,800	0,976	0,831	0,769

Evaluasi Multikelas : *Confusion Matrix*

Confusion matrix menunjukkan pola kesalahan yang berbeda antar *pipeline*. Tabel 8 menunjukkan bahwa *MLR_CW* mampu mengenali seluruh data kurang gizi dan sebagian besar kelas risiko gizi lebih, tetapi menghasilkan banyak kesalahan pada kelas gizi baik, yaitu 48 data diprediksi sebagai kurang gizi dan 31 data diprediksi sebagai risiko gizi lebih. Pola ini menunjukkan bahwa *MLR_CW* cenderung sensitif terhadap kelas minoritas, tetapi kurang presisi pada kelas mayoritas. Pada *LGBM_SMOTENC*, 12 dari 15 data kurang gizi; 630 dari 645 data gizi baik; 64 dari 78 data risiko gizi lebih; dan 5 dari 6 data gizi lebih berhasil diklasifikasikan dengan benar.

Kesalahan terbesar *LGBM_SMOTENC* terjadi pada kelas risiko gizi lebih yang diprediksi sebagai gizi baik sebanyak 13 data. Temuan ini memperkuat bahwa *LGBM_SMOTENC* memiliki pola prediksi yang lebih seimbang antara kelas mayoritas dan kelas minoritas.

Tabel 8. *Confusion matrix pipeline*

Pipeline	Model	*Imbalance Strategy*	Aktual	Prediksi Kurang Gizi	Prediksi Gizi Baik	Prediksi Risiko Gizi Lebih	Prediksi Gizi Lebih
MLR_CW	*MLR*	*class_weight*	Kurang gizi	15	0	0	0
MLR_CW	*MLR*	*class_weight*	Gizi baik	48	564	31	2
MLR_CW	*MLR*	*class_weight*	Risiko gizi lebih	0	1	73	4
MLR_CW	*MLR*	*class_weight*	Gizi Lebih	0	0	1	5
RF_CW	*Random Forest*	*class_weight*	Kurang Gizi	11	4	0	0
RF_CW	*Random Forest*	*class_weight*	Gizi Baik	3	630	12	0
RF_CW	*Random Forest*	*class_weight*	Risiko Gizi Lebih	0	14	63	1
RF_CW	*Random Forest*	*class_weight*	Gizi Lebih	0	0	3	3
RF_SMOTENC	*Random Forest*	*SMOTENC*	Kurang Gizi	12	3	0	0
RF_SMOTENC	*Random Forest*	*SMOTENC*	Gizi Baik	4	629	12	0
RF_SMOTENC	*Random Forest*	*SMOTENC*	Risiko Gizi Lebih	0	17	60	1
RF_SMOTENC	*Random Forest*	*SMOTENC*	Gizi Lebih	0	0	3	3
LGBM_CW	*LightGBM*	*class_weight*	Kurang Gizi	11	4	0	0
LGBM_CW	*LightGBM*	*class_weight*	Gizi Baik	3	627	15	0
LGBM_CW	*LightGBM*	*class_weight*	Risiko Gizi Lebih	0	10	67	1
LGBM_CW	*LightGBM*	*class_weight*	Gizi Lebih	0	0	2	4
LGBM_SMOTENC	*LightGBM*	*SMOTENC*	Kurang Gizi	12	3	0	0
LGBM_SMOTENC	*LightGBM*	*SMOTENC*	Gizi Baik	3	630	11	1
LGBM_SMOTENC	*LightGBM*	*SMOTENC*	Risiko Gizi Lebih	0	13	64	1
LGBM_SMOTENC	*LightGBM*	*SMOTENC*	Gizi Lebih	0	0	1	5

Dalam konteks pemantauan status gizi, kesalahan prediksi pada kelas minoritas perlu mendapat perhatian khusus karena dapat memengaruhi ketepatan identifikasi awal kelompok yang membutuhkan pemantauan lebih lanjut. Model final tidak cukup dipilih berdasarkan jumlah prediksi benar secara keseluruhan, tetapi juga perlu mempertimbangkan kestabilan performa terhadap seluruh kelas.

Pembahasan

Pemilihan Model Final

Berdasarkan hasil validasi silang dan *test set*, *LGBM_SMOTENC* dipilih sebagai *pipeline* final karena memperoleh *Best CV F1 Macro* tertinggi sebesar 0,829 dan *Test F1 Macro* tertinggi sebesar 0,844.

SENABISTEKES
Seminar Nasional Bisnis Teknologi Dan Kesehatan
Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:158-169

Meskipun *MLR_CW* memiliki *Balanced Accuracy* tertinggi, *LGBM_SMOTENC* menunjukkan keseimbangan yang lebih baik berdasarkan kombinasi *F1 Macro*, *Precision Macro*, akurasi, evaluasi per kelas, dan konsistensi hasil validasi silang dengan *test set*. Dengan demikian, pemilihan *LGBM_SMOTENC* tidak hanya didasarkan pada satu metrik, tetapi pada kestabilan performa model dalam menghadapi distribusi kelas yang tidak seimbang.

Keunggulan *LGBM_SMOTENC* menunjukkan bahwa kombinasi model berbasis *gradient boosting* dan *SMOTENC* dapat membantu meningkatkan representasi kelas minoritas pada proses pelatihan. Pada *dataset* status gizi yang didominasi oleh kelas gizi baik, strategi ini penting karena model yang hanya mengoptimalkan akurasi berisiko mengabaikan kelas dengan jumlah data kecil. Hasil penelitian ini menegaskan bahwa evaluasi model pada data tidak seimbang perlu mempertimbangkan algoritma, strategi penanganan kelas minoritas, dan metrik evaluasi secara bersamaan.

Implikasi Penelitian

Secara praktis, model *machine learning* dalam penelitian ini dapat diposisikan sebagai alat bantu *skrining* awal status gizi balita berbasis data antropometri dasar. Namun, model tidak dapat menggantikan tenaga kesehatan atau standar klinis dalam penentuan status gizi. Hasil prediksi tetap perlu diverifikasi oleh petugas kesehatan, terutama pada kelas minoritas yang membutuhkan perhatian dalam pemantauan gizi.

Secara metodologis, penelitian ini menunjukkan bahwa akurasi saja tidak cukup untuk mengevaluasi model pada *dataset* tidak seimbang. Pelaporan metrik makro, *Balanced Accuracy*, *confusion matrix*, dan performa per kelas membuat evaluasi model lebih transparan. Pendekatan ini juga membantu mengurangi risiko klaim performa yang terlalu optimistis akibat dominasi kelas mayoritas.

Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Jumlah data pada kelas minoritas masih kecil, terutama kelas gizi lebih yang hanya memiliki 31 data secara keseluruhan dan 6 data pada *test set*. Kondisi ini membuat performa kelas minoritas perlu ditafsirkan secara hati-hati karena perubahan satu atau dua prediksi dapat memengaruhi nilai *precision*, *recall*, dan *F1-score*.

Selain itu, fitur yang digunakan masih terbatas pada variabel antropometri dasar, sehingga model belum mempertimbangkan faktor klinis, sosial-ekonomi, sanitasi, riwayat penyakit, dan asupan makanan. Label status gizi juga digunakan sebagaimana tersedia pada *dataset*, sehingga penelitian ini tidak melakukan rekonstruksi label berdasarkan *z-score* antropometri. Evaluasi akhir masih menggunakan satu pembagian *stratified train-test split*, sehingga validasi eksternal atau *repeated stratified cross-validation* diperlukan untuk memperkuat bukti generalisasi.

Kesimpulan

Penelitian ini menunjukkan bahwa *pipeline LightGBM* dengan *SMOTENC* menjadi model paling konsisten untuk klasifikasi multikelas status gizi balita berbasis data antropometri pada *dataset* tidak seimbang. Pemilihan model pada konteks ini tidak cukup hanya mengandalkan akurasi, tetapi perlu mempertimbangkan *F1 Macro*, *Precision Macro*, *Balanced Accuracy*, performa kelas minoritas, dan konsistensi hasil validasi silang dengan *test set*. Model

final lebih tepat digunakan sebagai alat bantu *skrining* awal atau pendukung keputusan, bukan sebagai pengganti verifikasi tenaga kesehatan.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Rumah Jurnal Ilmiah Universitas Muhammadiyah Muara Bungo atas dukungan dan kesempatan yang diberikan sehingga kegiatan penulis bisa berjalan dengan baik. Penulis juga menyampaikan apresiasi kepada seluruh Pengurus Posyandu Desa Ujunggenteng, UPTD Puskesmas Ciracap, tahun 2025, yang sudah memberika izn dalam melaksanakan penelitian ini sehingga sampai artikel ini diterbitkan.

Daftar Pustaka

[1] Ferreira, H. D. S. (2020). Anthropometric assessment of children's nutritional status: A new approach based on an adaptation of Waterlow's classification. *BMC Pediatrics*, *20*(1), 1–11. doi:10.1186/s12887-020-1940-6

[2] [2] Talukder, A., & Ahammed, B. (2020). Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. *Nutrition*, *78*, 110861. doi:10.1016/j.nut.2020.110861

[3] [3] Fenta, H. M., Zewotir, T., & Muluneh, E. K. (2021). A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones. *BMC Medical Informatics and Decision Making*, *21*(1), 1–12. doi:10.1186/s12911-021-01652-1

[4] [4] Bitew, F. H., Sparks, C. S., & Nyarko, S. H. (2022). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. *Public Health Nutrition*, *25*(2), 269–280. doi:10.1017/S1368980021004262

[5] [5] Saleem, J., Zakar, R., Butt, M. S., Aadil, R. M., Ali, Z., Bukhari, G. M. J., … Fischer, F. (2024). Application of the Boruta algorithm to assess the multidimensional determinants of malnutrition among children under five years living in southern Punjab, Pakistan. *BMC Public Health*, *24*(1), 1–10. doi:10.1186/s12889-024-17701-z

[6] [[6] Shahid, M., Yahya, M. A., Song, J., Naveed, H. M., Yuksel, S., Dincer, H., & Ali, M. (2025). Machine learning vs. traditional logistic regression: predictive performance and risk factor identification for child nutritional outcome in Pakistan. *BMC Public Health*, *26*(1), 667. doi:10.1186/s12889-025-25621-9

[7] [7] Mim, M. S., Sajib, A. H., & Nipa, J. F. (2026). Determinants of malnutrition among under-five children in Bangladesh: a cross-sectional analytical study comparing multinomial logistic and proportional odds regression models using MICS 2019 data. *Journal of Health, Population and Nutrition*, *45*(1), 50. doi:10.1186/s41043-025-01201-w

[8] [8] Kim, M., & Hwang, K. B. (2022). An empirical evaluation of sampling methods for the classification of imbalanced data. *PLoS ONE*, *17*(7 July), 1–22. doi:10.1371/journal.pone.0271260

[9] [9] Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*, *113*(7), 4903–4923. doi:10.1007/s10994-022-06296-4

[10] [10] Feleke, H. G., Simegn, M. B., Gebreegziabher, Z. A., Mazengia, E. M., & Tilahun, W. M. (2026). Application of machine learning algorithm for predicting acute malnutrition among under 5 children in east Africa using recent DHS. *Scientific Reports*, *16*(1), 16242. doi:10.1038/s41598-026-46944-6