



Klasifikasi Status Gizi pada Balita Menggunakan Metode light Gradient Boosting Machine (LightGBM) dengan Optimasi Hyperparameter GridSearchCV

Muhammad Yustaf Lana Badriul Hegar¹, Ani Dijah Rahajoe², Muhammad Muharom Al Haromainy³

^{1,2,3}Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia

E-mail: 21081010289@student.upnjatim.ac.id, anidijah.if@upnjatim.ac.id,
muhammad.muharrom.if@upnjatim.ac.id

Article Info	Abstract
Article History Received: 2025-10-07 Revised: 2025-11-13 Published: 2025-12-02 Keywords: <i>Classification; Toddler Nutritional Status; LightGBM; GridSearchCV.</i>	The nutritional status of toddlers is an important indicator in determining the health of a country because it influences the growth and development of children in the future. The problem of nutritional status classification is often faced with data complexity and imbalanced class distribution that can reduce the performance of the prediction model. This study aims to build a classification model for toddlers' nutritional status using the Light Gradient Boosting Machine (LightGBM) method and improve its performance through hyperparameter optimization using GridSearchCV. The research data went through several preprocessing stages including data cleaning, transformation, data division, and model evaluation using metrics of accuracy, precision, recall, f1-score, and confusion matrix. The results showed that the LightGBM model before optimization showed an accuracy of 91.53% and after optimization produced an accuracy value of 92.23% with improved performance in most target classes. These results indicate that the application of LightGBM with GridSearchCV is able to provide quite good classification performance and can be recommended as an approach in a decision support system related to toddlers' nutritional status.
Artikel Info	Abstrak
Sejarah Artikel Diterima: 2025-10-07 Direvisi: 2025-11-13 Dipublikasi: 2025-12-02 Kata kunci: <i>Klasifikasi; Status Gizi Balita; LightGBM; GridSearchCV.</i>	Status gizi balita adalah salah satu indikator penting dalam menentukan ukuran kesehatan suatu negara karena berpengaruh terhadap tumbuh kembang anak dimasa depan. Permasalahan klasifikasi status gizi sering dihadapkan pada kompleksitas data dan ketidakseimbangan distribusi kelas yang dapat menurunkan performa model prediksi. Penelitian ini bertujuan untuk membangun model klasifikasi status gizi balita menggunakan metode Light Gradient Boosting Machine (LightGBM) dan meningkatkan kinerjanya melalui optimasi hyperparameter menggunakan GridSearchCV. Data penelitian ini melalui beberapa tahapan preprocessing yang mencakup pembersihan data, transformasi, pembagian data, serta evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, f1-score, dan confusion matrix. Hasil penelitian menunjukkan bahwa model LightGBM sebelum dilakukan optimasi menunjukan akurasi 91,53% dan setelah dilakukan optimasi menghasilkan nilai akurasi 92,23% dengan peningkatan performa pada sebagian besar kelas target. Hasil ini menunjukkan bahwa penerapan LightGBM dengan GridSearchCV mampu memberikan performa klasifikasi yang cukup baik dan dapat direkomendasikan sebagai salah satu pendekatan dalam sistem pendukung keputusan terkait status gizi balita.

I. PENDAHULUAN

Salah satu upaya untuk menentukan kesehatan masyarakat disetiap wilayah adalah dengan memperbaiki gizi masyarakat, gizi yang seimbang dapat meningkatkan ketahanan tubuh, kecerdasan, dan menjadikan pertumbuhan yang normal (Setianingsih, Anggarini, and Arifin 2024). Gizi yang baik pada anak dibawah lima tahun sangat berpengaruh terhadap pertumbuhan normal, perkembangan fisik, dan kecerdasan pada anak (Kota, Sebagai, and Pencegahan 2025). Perolehan gizi pada anak usia dibawah lima tahun perlu diperhatikan dalam menjaga kesehatan anak. Karena masa balita merupakan periode perkembangan yang rentan

gizi. balita merupakan kelompok masyarakat yang rentan gizi. pada kelompok tersebut mengalami siklus pertumbuhan dan perkembangan yang membutuhkan zat-zat gizi yang lebih besar dari kelompok umur lainnya. permasalahan mengenai masalah gizi ini sering terjadi pada anak dibawah lima tahun (Salang et al. 2025).

Status gizi balita adalah salah satu dari banyaknya permasalahan kesehatan yang terjadi di Indonesia. Negara dengan jumlah penduduk yang banyak, Indonesia menghadapi dinamika masalah gizi buruk (Rusliani, Hidayani, and Sulistyoningsih 2022). Gizi buruk pada anak secara langsung dipengaruhi oleh makanan, sedangkan secara tidak langsung disebabkan

oleh penyakit infeksi, baik yang berkaitan dengan faktor gen, sosial ekonomi, demografi, maupun perilaku. Penanganan yang lambat atau kurang tepat terhadap kondisi gizi anak dapat berdampak negatif terhadap pertumbuhan dan kualitas hidup anak di masa depan (Nasution et al. 2025). pemantauan status gizi secara rutin di fasilitas layanan dasar, seperti Puskesmas, posyandu menjadi sangat penting. Puskesmas berperan strategis sebagai pusat pelayanan kesehatan masyarakat, terutama dan anak di bawah lima tahun. Salah satu kegiatan rutinnya adalah penimbangan dan pencatatan perkembangan balita. Pengukuran status gizi berlandaskan atas standart WHO (World Healt Organization) yang sudah ditetapkan pada keputusan menteri kesehatan Nomor 1995/Menkes/SK/XII/2020 mengenai Standar Antropometri penilaian status gizi anak. Dapat diukur berdasarkan tiga index, yaitu berat badan menurut umur (BB/UU), tinggi badan menurut umur (TB/U), dan berat badan menurut tinggi badan (BB/TB), dan berbagai indikator lainnya (Qurrotunnisa and Ratnasari 2025)

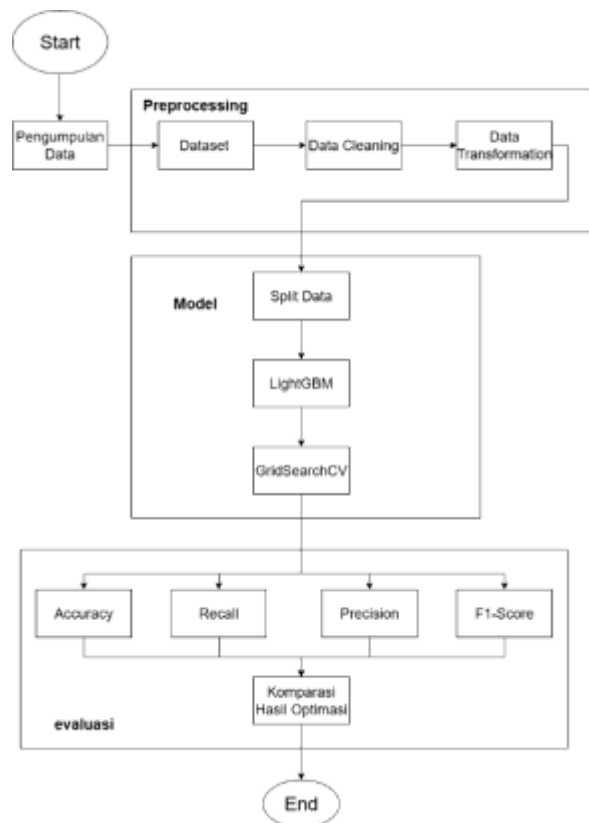
Penelitian ini menggunakan teknologi machine learning untuk menjadi Solusi inovatif untuk klasifikasi status gizi balita (Wantoro et al. n.d.). Pendekatan yang dilakukan adalah penggunaan algoritma perkembangan Gradien Boosting Decission Tree yaitu LightGBM yang telah terbukti memiliki kinerja tinggi dalam analisis data (Baihaqi et al. 2023). Kemudian dilakukan juga teknik optimasi supaya memberikan hasil akurasi yang lebih akurat. Teknik optimasi yang digunakan yaitu GridSearchCV yaitu suatu pendekatan yang menyediakan pencarian nilai yang optimal dengan menggabungkan Grid Search dan Cross-Validation. Cara kerja metode ini dengan membagi data menjadi beberapa bagian dan melakukan validasi silang pada setiap kombinasi parameter model (Aprilliandhika and Fauzi Abdulloh 2024).

Penelitian ini berfokus pada klasifikasi model LightGBM menggunakan GridSearchCV untuk optimasi dengan kombinasi parameter yang paling terbaik. berbagai penelitian sebelumnya yang dapat dijadikan referensi dalam penelitian ini. Bhavana Srinivasan, melakukan analisis risiko transaksi blockchain dalam lingkungan metaverse. LightGBM menggunakan metode histogram-based yang memproses LightGBM bekerja dengan metode histogram-based, yaitu mengelompokkan nilai fitur ke dalam beberapa bin daripada memproses semua nilai asli. Setiap fitur diubah menjadi histogram yang berisi jumlah data dan total gradien tiap bin. hasil

penelitian menunjukkan akurasi LightGBM mencapai akurasi sebesar 91.2% dan nilai AUC-ROC sebesar 0.94 (Srinivasan 2025).

Penelitian selanjutnya yang dilakukan Yusril Isra Mahendra dkk. Peningkatan kinerja model LightGBM dengan optimasi GridSearchCV yaitu dengan membandingkan kinerja XGBoost dan LightGBM menggunakan teknik pendekatan GridSearchCV untuk optimasi hyperparameter. Hasil dari pengujian menunjukkan bahwa kedua algoritma mencapai akurasi model LightGBM memperoleh akurasi sebesar 97,39%, teknik ini menunjukkan performa stabil, efisien dalam waktu pelatihan, dan hemat memori (Mahendra and Putra 2024).

II. METODE PENELITIAN



Gambar 1. Alur Penelitian

1. Pengumpulan Data

Pada proses pengumpulan data dapat dilaksanakan dengan berbagai cara dan metode yang berbeda. Pada penelitian ini menggunakan data yang didapatkan di Puskesmas Kecamatan Kanor kabupaten Bojonegoro. data yang diperoleh berupa dataset status gizi pada balita dengan jumlah 5017 data, yang mencakup variabel-variabel utama seperti usia, berat badan, tinggi badan, status gizi dan tambahan indikator lainnya. data yang dikumpulkan

dengan tetap memperhatikan aspek etika penelitian termasuk mendapatkan izin dari pihak terkait sebelum digunakan untuk dilakukan analisis lebih lanjut. Pada dataset mencakup atribut-atribut sebagai berikut :

Tabel 1. Atribut Dataset

Id Atribut	Atribut	Tipe Data
1	Jenis Kelamin	String (L/P)
2	Umur	Integer (Bulan)
3	Berat Badan	Float (Kg)
4	Tinggi Badan	Float (Cm)
5	Status Gizi	String (Kategori)

Data ini diperoleh dari puskesmas yang melakukan pencatatan rutin terhadap pertumbuhan dan perkembangan anak balita. Dataset ini mencakup berbagai atribut usia, berat badan, tinggi badan, jenis kelamin, serta status gizi yang ditentukan berdasarkan standar antropometri WHO.

2. Dataset

Dataset yang diperoleh kemudian diproses untuk memastikan kelengkapan dan konsistensi data sebelum dilakukan tahap analisis. Tahapan preprocessing seperti pembersihan data, penanganan nilai yang hilang serta normalisasi variabel dilakukan untuk memastikan kualitas dataset yang optimal untuk penelitian ini.

3. Data Cleaning

Pada tahap pembersihan data dilakukan proses pembersihan dataset yaitu menghapus kolom yang tidak relevan dari dataset karena tidak memberikan kontribusi terhadap proses analisis karena tidak mengandung informasi yang bisa dipelajari oleh machine learning. Penanganan terhadap nilai yang hilang keberadaan data yang tidak lengkap dapat mempengaruhi akurasi model. Di tentukan pendekatan terbaik dengan menghapus baris yang mengandung nilai kosong dan mengisinya menggunakan nilai mean atau median. Data duplikat diidentifikasi dan dihapus untuk memastikan bahwa dataset yang digunakan benar-benar menampilkan informasi yang relevan.

4. Data Tranformation

Pada tahap ini, data yang bersifat kategorikal seperti data dalam bentuk string atau atribut non-numerik perlu diubah menjadi format angka numerik supaya dapat dipahami oleh model machine learning. Proses ini penting karena sebagian besar

algoritma pembelajaran mesin hanya dapat bekerja dengan data numerik, sehingga data kategorikal harus dikonversi terlebih dahulu menggunakan metode seperti label encoding atau one-hot encoding. Setelah ditahap pembersihan data selesai diproses, tahap selanjutnya adalah proses diskritisasi pada dataset. Diskritisasi merupakan proses mengubah data yang bersifat kategorikal ke dalam bentuk kategori atau interval tertentu agar lebih mudah dianalisis dan digunakan dalam model. Dalam penelitian ini, variabel jenis kelamin dan status gizi akan didiskritisasi untuk memudahkan proses klasifikasi dan interpretasi data. Proses ini juga membantu model dalam mengurangi kompleksitas data, meningkatkan kinerja algoritma, serta menghasilkan pola yang lebih jelas dan terukur dalam proses pembelajaran mesin.

5. Split Data

Pada tahap pembagian data, dilakukan proses pembagian data yang bertujuan untuk memisahkan dataset menjadi bagian-bagian yang berbeda agar model machine learning dapat dilatih dan diuji secara efektif. Pembagian data ini penting untuk memastikan bahwa model tidak hanya memahami dari data yang sama yang digunakan untuk evaluasi, sehingga hasil prediksi lebih valid dan tidak bias. Langkah pertama dalam proses ini adalah membagi dataset menjadi 80% untuk data training dan 20% untuk data testing.

Data training digunakan model untuk mempelajari pola, hubungan, dan karakteristik data, sedangkan data testing digunakan untuk mengevaluasi kinerja model pada data yang sebelumnya belum pernah dilihat. Dengan metode pembagian ini, model dapat diuji secara objektif untuk mengukur akurasi, presisi, recall, dan metrik evaluasi, sehingga dapat menilai seberapa baik model dapat melakukan prediksi pada data baru. Proses pembagian data ini merupakan langkah penting dalam pra-pemodelan karena mempengaruhi kemampuan model untuk menggeneralisasi dan memberikan hasil prediksi yang andal dan konsisten.

6. Model LightGBM

Model yang digunakan dalam penelitian ini adalah LightGBM. Untuk mendapatkan performa optimal, model ini dioptimasi menggunakan GridSearchCV, sehingga

parameter seperti learning rate, max depth, dan n_estimators dapat disesuaikan secara otomatis untuk mencapai akurasi terbaik pada dataset.

7. GridSearchCV

Pada tahap ini, GridSearchCV digunakan untuk melakukan optimasi hyperparameter pada algoritma LightGBM dengan cara melakukan pencarian sistematis terhadap kombinasi parameter yang telah ditentukan. GridSearchCV akan menguji setiap kombinasi parameter tersebut menggunakan cross-validation, yang artinya membagi data menjadi beberapa subset, melatih model pada sebagian subset, dan menguji pada subset lainnya untuk memastikan hasil yang stabil dan tidak bias. Setelah semua kombinasi diuji, Teknik gridSearchCV akan melakukan seleksi kombinasi hyperparameter yang terbaik berdasarkan metrik evaluasi yang digunakan, sehingga model yang dihasilkan memiliki akurasi tinggi dan kemampuan generalisasi yang baik pada data baru. Dengan cara ini, LightGBM dapat bekerja secara optimal dan menghasilkan prediksi yang lebih akurat dan konsisten.

III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

1. Hasil Preprocessing

```
Bentuk Data Awal: (5117, 17)

Informasi Missing Value:
no          0
date       176
sex        101
age_months  0
weight_kg   101
height_cm   102
address     144
bmi         0
breastfed   0
recent_illness  0
immunization_complete  0
parent_education  0
household_income_cat  0
sanitation   0
meal_freq    0
nutrition_score  0
gizi_label   0
dtype: int64
```

Gambar 2. Bentuk Data Awal

Pada gambar 2 ditunjukkan dataset awal memiliki ukuran 5.117 baris dan 17 variabel, yang merepresentasikan data status gizi balita beserta informasi karakteristik terkait. Berdasarkan informasi missing value, terlihat bahwa sebagian besar variabel memiliki data yang lengkap,

namun terdapat beberapa variabel yang mengandung nilai hilang (*missing values*). Variabel date memiliki 176 data yang hilang, variabel sex kehilangan 101 data, weight_kg sebanyak 101 data, height_cm sebanyak 102 data, dan variabel address memiliki 144 data yang kosong. Sementara itu, variabel lainnya seperti age_months, bmi, breastfed, recent_illness, immunization_complete, parent_education, household_income_cat, sanitation, meal_freq, nutrition_score, serta gizi_label tidak memiliki nilai yang hilang dan dinyatakan lengkap. Kondisi seperti ini menunjukkan bahwa sebelum dilakukan proses analisis atau pemodelan machine learning lebih lanjut, dataset ini memerlukan tahap data cleaning terutama pengelolaan nilai yang hilang pada beberapa variabel. Selain itu, variabel target gizi_label tersedia secara lengkap sehingga dapat digunakan untuk proses klasifikasi status gizi balita pada tahap penelitian selanjutnya.

```
--- Data Cleaning: Langkah yang dilakukan ---
Menghapus kolom tidak relevan: ['no', 'date', 'address', 'nutrition_score']

Shape sebelum drop kolom: (5117, 17)
Shape setelah drop kolom: (5117, 13)

Jumlah baris duplikat terdeteksi: 100
Baris dihapus karena duplikat: 100
Shape setelah hapus duplikat: (5017, 13)
```

Gambar 3. Hapus Data Duplikat

Berdasarkan gambar 4, ditunjukkan tahapan pembersihan pada data untuk meningkatkan kualitas data sebelum digunakan dalam proses analisis dan pemodelan. Langkah pertama merupakan proses penghapusan kolom yang dianggap tidak relevan dengan kebutuhan penelitian, yaitu kolom no, date, address, nutrition_score. Setelah penghapusan, jumlah kolom dalam dataset berkurang dari 17 menjadi 13, yang berarti hanya atribut penting saja yang dipertahankan untuk analisis lebih lanjut. Selanjutnya dilakukan proses deteksi dan penghapusan data duplikat supaya tidak ada baris data yang berulang yang dapat memengaruhi hasil analisis. Dari hasil deteksi ditemukan sebanyak 100 baris data duplikat, dan semuanya dihapus sehingga jumlah baris data berkurang dari 5117 menjadi 5017. Dengan demikian, dataset akhir setelah proses pembersihan terdiri dari 5017 baris dan 13 kolom yang telah bebas dari data tidak relevan dan duplikasi.

```
Missing value sebelum imputasi:
sex          100
age_months   0
weight_kg     0
height_cm     0
bmi           0
breastfed     0
recent_illness 0
immunization_complete 0
parent_education 0
household_income_cat 0
sanitation    0
meal_freq     0
gizi_label    0
dtype: int64

Missing value setelah imputasi:
sex          0
age_months   0
weight_kg     0
height_cm     0
bmi           0
breastfed     0
recent_illness 0
immunization_complete 0
parent_education 0
household_income_cat 0
sanitation    0
meal_freq     0
gizi_label    0
dtype: int64
Jumlah data akhir setelah cleaning: (5017, 13)
```

Gambar 4. Missing Value

Pada gambar 4, menjelaskan tahap proses data cleaning juga mencakup penanganan terhadap nilai yang hilang pada beberapa atribut. Sebelum dilakukan imputasi, terdapat beberapa kolom yang memiliki nilai kosong, yaitu sex, age_months, weight_kg, dan height_cm, masing-masing sebanyak 100 data yang tidak terisi. Keberadaan nilai kosong ini dapat menyebabkan hasil analisis atau pemodelan menjadi tidak akurat, sehingga perlu dilakukan proses imputasi untuk menggantinya dengan nilai yang sesuai. Setelah proses imputasi dilakukan, semua kolom menunjukkan jumlah missing value sebesar 0, menandakan bahwa data telah lengkap dan tidak ada lagi nilai yang hilang. Dengan demikian, dataset menjadi lebih bersih dan siap digunakan untuk tahap analisis berikutnya. Jumlah data akhir setelah seluruh proses cleaning tetap sebanyak 5017 baris dan 13 kolom, menunjukkan bahwa tidak ada penghapusan data tambahan, melainkan hanya pengisian nilai kosong agar dataset menjadi konsisten dan valid untuk pemodelan.

```
Distribusi Target:
gizi_label
Gizi Baik      3740
Gizi Kurang    1247
Gizi Buruk      80
Gizi Lebih     50
```

Gambar 5. Distribusi Label Target

Pada gambar 5 menjelaskan distribusi awal data pada variabel target gizi_label menunjukkan adanya ketidakseimbangan kelas (class imbalance) yang cukup signifikan. Berdasarkan hasil analisis, kelas Gizi Baik mendominasi dengan jumlah 3.740 data, diikuti oleh Gizi Kurang sebanyak 1.247 data. Sementara itu, kelas Gizi Buruk dan Gizi Lebih hanya memiliki 80 dan 50 data. Kondisi ini menunjukkan bahwa mayoritas sampel berasal dari satu kelas dominan, sedangkan kelas lainnya berada pada kategori minoritas. Ketidakseimbangan ini dapat menimbulkan bias pada model klasifikasi karena algoritma mengarah untuk mempelajari pola dari kelas mayoritas dan mengabaikan kelas minoritas. Dampaknya, akurasi prediksi terhadap kelas minoritas menjadi rendah karena kontribusinya terhadap fungsi pembelajaran sangat kecil.

2. Splitting Data

proses pembagian dataset dilakukan untuk memisahkan data yang digunakan dalam tahap pelatihan model dan data yang digunakan untuk evaluasi performa model. Pada penelitian ini digunakan proporsi pembagian data sebesar 80% untuk data latih (training data) dan 20% untuk data uji (testing data). Pembagian ini bertujuan untuk menghindari overfitting, yaitu kondisi di mana model terlalu menyesuaikan diri dengan data pelatihan sehingga tidak mampu memberikan prediksi yang baik pada data baru. Proporsi 80:20 ini merupakan salah satu metode pembagian data yang umum digunakan dalam machine learning karena bisa memberikan keseimbangan yang baik antara kecukupan data untuk pelatihan dan evaluasi performa model secara objektif.

3. Uji Coba Model LGBM Tanpa Optimasi

	precision	recall	f1-score	support
Gizi Baik	0.95	0.95	0.95	733
Gizi Buruk	1.00	0.25	0.40	16
Gizi Kurang	0.82	0.86	0.84	245
Gizi Lebih	1.00	0.50	0.67	10
accuracy			0.92	1004
macro avg	0.94	0.64	0.71	1004
weighted avg	0.92	0.92	0.91	1004

Gambar 6. Clasification Report Tanpa Optimasi

Pada gambar 7 menunjukan hasil classification report dari model LGBM

sebelum dilakukan optimasi (baseline). Hasil evaluasi menunjukkan bahwa model memiliki performa yang sangat baik pada kelas Gizi Baik, dengan nilai precision 0.95, recall 0.95, dan f1-score 0.95. Hal ini karena kelas tersebut memiliki jumlah data terbanyak sehingga model mampu belajar lebih baik pada kelas ini. Untuk kelas Gizi Kurang, nilai precision 0.82 dan recall 0.86 menunjukkan bahwa model juga cukup mampu mengenali kelas ini secara baik. Namun, performa model menurun pada kelas Gizi Buruk dan Gizi Lebih, masing-masing hanya memperoleh recall 0.25 dan 0.50. Hal ini menunjukkan bahwa model masih kesulitan dalam mendeteksi kelas minoritas yang jumlah datanya sangat sedikit. Nilai macro average f1-score 0.71 menunjukkan bahwa performa model belum merata pada seluruh kelas, sedangkan weighted average f1-score 0.91 menunjukkan bahwa model secara keseluruhan masih cukup baik karena mempertimbangkan dominasi kelas mayoritas.

Baseline - Testing Accuracy : 0.9153

Gambar 7. Hasil Akurasi

Gambar 8 menampilkan nilai akurasi model baseline, yaitu sebesar 0.9153 atau 91.53%. Nilai akurasi ini menunjukkan bahwa model mampu melakukan prediksi dengan benar terhadap 91.53% data uji. Ini menandakan bahwa performa model cukup baik untuk tahap awal sebelum dilakukan proses optimasi. Namun, meskipun akurasinya tinggi, perlu diperhatikan bahwa karena dataset awal bersifat tidak seimbang, nilai akurasi saja belum cukup menggambarkan kualitas model secara menyeluruh. Oleh karena itu, evaluasi tambahan seperti precision, recall, dan f1-score tetap diperlukan agar performa model pada kelas minoritas dapat dinilai secara lebih tepat.

4. Uji Coba Model LGBM dengan Optimasi

	precision	recall	f1-score	support
Gizi Baik	0.96	0.95	0.95	733
Gizi Buruk	0.75	0.38	0.50	16
Gizi Kurang	0.82	0.89	0.86	245
Gizi Lebih	1.00	0.50	0.67	10
accuracy			0.92	1004
macro avg	0.88	0.68	0.74	1004
weighted avg	0.92	0.92	0.92	1004

Gambar 8. Clasification Report Dengan Optimasi

Berdasarkan gambar 9 menunjukkan hasil evaluasi kinerja model Light Gradient Boosting Machine (LightGBM) setelah dilakukan optimasi hyperparameter menggunakan GridSearchCV, diperoleh classification report. Model menunjukkan kinerja yang sangat baik pada kelas Gizi Baik dengan nilai precision sebesar 0.96, recall 0.95, dan f1-score 0.95, yang mencerminkan bahwa model mampu mengklasifikasikan kelas mayoritas secara akurat. Kinerja model juga cukup baik pada kelas Gizi Kurang dengan nilai f1-score mencapai 0.86, menunjukkan tingkat keseimbangan yang baik antara kemampuan model dalam mengenali data aktual dan menghasilkan prediksi yang benar. Namun demikian, performa model pada kelas Gizi Buruk dan Gizi Lebih masih menunjukkan keterbatasan, masing-masing dengan nilai recall 0.38 dan 0.50. Hal ini mengindikasikan bahwa model masih kurang optimal dalam mendeteksi data aktual pada kelas minoritas yang jumlahnya lebih sedikit dalam dataset. Hal ini diperkuat oleh nilai macro average f1-score sebesar 0.74, yang menunjukkan performa rata-rata antar kelas masih dipengaruhi oleh ketidakseimbangan data, sementara weighted average f1-score sebesar 0.92 menggambarkan bahwa performa keseluruhan model tetap tinggi stabil karena mempertimbangkan proporsi setiap kelas dalam perhitungan.

Optimasi - Testing Accuracy : 0.9223

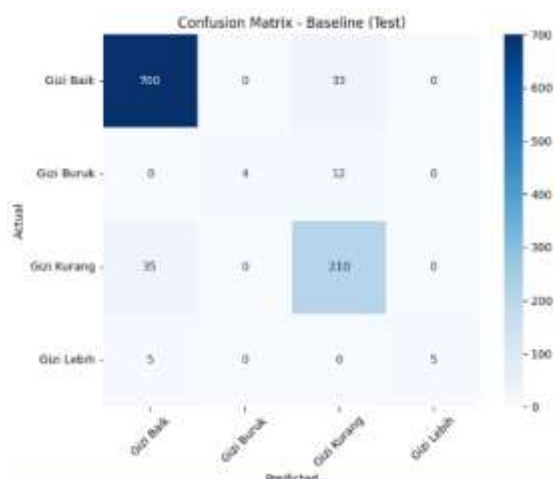
Gambar 9. Hasil Akurasi Optimasi

Nilai akurasi pengujian yang ditunjukkan pada Gambar 10 menunjukkan bahwa model LightGBM yang telah dilakukan teknik optimasi menghasilkan tingkat akurasi sebesar 92.23%. Nilai ini menunjukkan bahwa model mampu melakukan prediksi dengan tingkat ketepatan yang tinggi terhadap data uji. Peningkatan performa ini tidak terlepas dari penerapan teknik optimasi hyperparameter GridSearchCV yang mampu memilih kombinasi parameter terbaik untuk meningkatkan kemampuan generalisasi model. Akurasi yang dicapai berada pada kategori sangat baik dan memenuhi standar evaluasi dalam penelitian klasifikasi, khususnya di bidang kesehatan masyarakat yang mengharuskan

ketepatan tinggi dalam proses pengambilan keputusan. Hasil ini membuktikan bahwa LightGBM merupakan algoritma yang efektif dalam menyelesaikan permasalahan klasifikasi status gizi balita, terutama pada dataset berukuran besar dan memiliki kompleksitas variabel yang tinggi.

B. Pembahasan

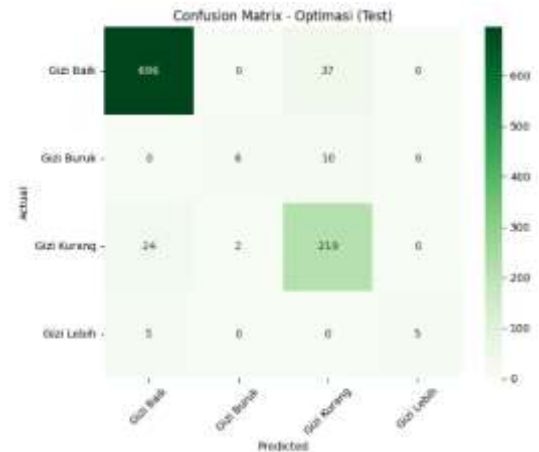
Pada pembahasan mengenai penelitian ini, hasil yang diperoleh menunjukkan bahwa proses analisis data dan pemilihan metode klasifikasi telah memberikan kontribusi yang signifikan terhadap peningkatan performa model. Temuan ini mengindikasikan bahwa setiap tahapan—mulai dari pengolahan data, pemilihan fitur, hingga pengujian model dengan berbagai variasi parameter berperan penting dalam menghasilkan prediksi yang lebih akurat.



Gambar 10. Confusion Matrik Tanpa Optimasi

Pada gambar 11 menjelaskan Confusion matrix sebelum optimasi menunjukkan bahwa model LightGBM baseline masih memiliki beberapa keterbatasan dalam mengklasifikasikan data pada kelas minoritas. Pada kelas Gizi Baik, model berhasil memprediksi dengan sangat baik dengan jumlah prediksi benar sebesar 700 data, namun masih terdapat 33 data yang salah diklasifikasikan sebagai Gizi Kurang. Kinerja model pada kelas Gizi Buruk terlihat kurang optimal, karena hanya 4 data yang berhasil diprediksi dengan benar, sedangkan 12 data salah diklasifikasikan sebagai Gizi Kurang. Demikian pula pada kelas Gizi Kurang, model mampu memprediksi 210 data dengan benar, namun masih terjadi kesalahan prediksi terhadap 35 data yang diklasifikasikan sebagai Gizi Baik. Pada kelas

Gizi Lebih, performa model masih terbatas dengan hanya 5 prediksi benar. Secara keseluruhan, confusion matrix ini menunjukkan bahwa model baseline cenderung bias terhadap kelas mayoritas (Gizi Baik), sementara kelas minoritas seperti Gizi Buruk dan Gizi Lebih mengalami misclassification yang cukup signifikan akibat ketidakseimbangan data.



Gambar 11. Confusion Matrik Dengan Optimasi

Confusion matrix setelah optimasi memperlihatkan adanya peningkatan kinerja model dalam melakukan klasifikasi terutama pada kelas minoritas. Jumlah prediksi benar pada kelas Gizi Baik sedikit berkurang menjadi 696 data, namun kesalahan prediksi pada kelas ini menurun terutama terhadap kelas Gizi Kurang. Peningkatan signifikan terlihat pada kelas Gizi Buruk, di mana jumlah prediksi benar meningkat dari 4 menjadi 6 data, dan kesalahan klasifikasi berkurang. Kelas Gizi Kurang menunjukkan peningkatan yang jelas, dengan 219 prediksi benar dibandingkan hanya 210 sebelumnya, serta kesalahan prediksi terhadap kelas lain menurun. Pada kelas Gizi Lebih, performa model tetap stabil dengan mempertahankan 5 prediksi benar dan tidak terjadi kesalahan klasifikasi yang berarti. Hasil ini menunjukkan bahwa proses optimasi menggunakan GridSearchCV berhasil meningkatkan generalisasi model dan memperbaiki distribusi prediksi terutama pada kelas dengan jumlah data yang lebih sedikit. Secara keseluruhan, model yang telah dioptimasi menunjukkan kemampuan klasifikasi yang lebih proporsional dan tidak terlalu bias terhadap kelas mayoritas.

IV. SIMPULAN DAN SARAN

A. Simpulan

Berdasarkan hasil penelitian yang telah dilakukan, Berdasarkan hasil penelitian, algoritma Light Gradient Boosting Machine (LightGBM) yang dioptimalkan menggunakan GridSearchCV terbukti mampu meningkatkan performa klasifikasi status gizi balita. Model memberikan akurasi akhir sebesar 92,23%, disertai peningkatan nilai precision, recall, dan F1-score dibandingkan model baseline. Confusion matrix juga menunjukkan peningkatan jumlah prediksi benar pada sebagian besar kelas. Meskipun demikian, performa model pada kelas dengan jumlah data rendah seperti Gizi Buruk dan Gizi Lebih masih belum optimal sehingga masih ditemukan kesalahan klasifikasi. Secara praktis, model ini berpotensi digunakan sebagai sistem pendukung keputusan di Puskesmas atau fasilitas kesehatan lainnya, terutama untuk membantu proses skrining dini dan identifikasi risiko gizi balita secara lebih cepat dan objektif. Temuan penelitian ini juga dapat menjadi dasar pengembangan sistem monitoring gizi yang lebih terintegrasi dan adaptif di masa mendatang.

B. Saran

Berdasarkan hasil penelitian ini, terdapat beberapa saran yang dapat digunakan untuk pengembangan pada penelitian selanjutnya. Pertama, perlu dilakukan penyeimbangan distribusi data antar kelas secara lebih optimal, misalnya dengan kombinasi teknik SMOTE untuk mengurangi misclassification pada kelas minoritas. Kedua, penelitian selanjutnya disarankan mencoba algoritma pembandingan lain seperti XGBoost, CatBoost, Random Forest, atau model berbasis deep learning untuk mengetahui model terbaik dalam memprediksi status gizi balita. Ketiga, perlu dilakukan feature engineering lebih lanjut dan analisis feature importance untuk memastikan variabel prediktor yang digunakan benar-benar relevan dan memberikan pengaruh signifikan terhadap model. Terakhir, disarankan untuk memperluas dataset agar model lebih mampu melakukan generalisasi serta diuji pada data dari wilayah atau tahun yang berbeda sehingga hasil penelitian lebih aplikatif dalam dunia nyata.

DAFTAR RUJUKAN

- Aprilliandhika, Wahyu, and Ferian Fauzi Abdulloh. 2024. "Comparison of K-Nearest Neighbor and Support Vector Machine Algorithm Optimization With Grid Search Cv on Stroke Prediction Komparasi Optimasi Algoritma K-Nearest Neighbor Dan Support Vector Machine Dengan Grid Search Cv Pada Prediksi Stroke." 5(4): 991–1000.
<https://doi.org/10.52436/1.jutif.2024.5.4.1951>.
- Baihaqi, Wiga Maulana, Ika Romadoni Yunita, Aulia Shafira Tri Damayanti, and Luthfi Akhaerunnisa. 2023. "Analysis of Classification Algorithm Performance on User Review Sentiment of the Muamalat DIN Application." *CogITo Smart Journal* 9(2): 241–51.
[doi:10.31154/cogito.v9i2.511.241-251](https://doi.org/10.31154/cogito.v9i2.511.241-251).
- Kota, Daycare, Bogor Sebagai, and Upaya Pencegahan. 2025. "Jurnal Pengabdian Kepada Masyarakat Bidang Ilmu Kesehatan Dan Kedokteran." 1(2): 30–35.
- Mahendra, Yusril Isra, and Ricky Eka Putra. 2024. "Penerapan Algoritma Gradient Boosted Decision Tree (GBDT) Untuk Klasifikasi Serangan DDoS." *JINACS: (Journal of Informatics and Computer Science)* ISSN 06: 158–66.
- Nasution, Nur Hayati, Abdul Hakim Dalimunthe, Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Al, and Washliyah Labuhanbatu. 2025. "Sistem Pendukung Keputusan Penentuan Gizi Balita Di Posyandu Dengan Metode AHP." 4(2): 275–81.
- Qurrotunnisa, Lulu Hammada, and Diah Ratnasari. 2025. "FAKTOR-FAKTOR YANG MEMPENGARUHI STATUS GIZI BALITA USIA 12 – 24 BULAN DI DESA WANASARI." 6(02): 96–101.
- Rusliani, Novie, Wuri Ratna Hidayani, and Hariyani Sulistyoningih. 2022. "Literature Review: Faktor-Faktor Yang Berhubungan Dengan Kejadian Stunting Pada Balita." *Buletin Ilmu Kebidanan dan Keperawatan* 1(01): 32–40. [doi:10.56741/bikk.v1i01.39](https://doi.org/10.56741/bikk.v1i01.39).

- Salang, Margrita, Anna Henny Talahatu, Marselinus Laga Nur, Ilmu Kesehatan Masyarakat, and Universitas Nusa Cendana. 2025. "Hubungan Asupan Gizi Dan Riwayat Sakit Dengan Status Gizi Balita Di Wilayah Kerja Puskesmas Feapopi." 4(1): 140–52. doi:10.55123/sehatmas.v4i1.3861.
- Setianingsih, Ajeng, Inggit Meliana Anggarini, and Lisa Yulinanda Arifin. 2024. "Behavior of Utilization of Posyandu by Mothers Who Have Babies and Toddlers Posyandu Wijaya Kusuma Puskesmas Bahagia." *Jurnal Ilmiah Kesehatan* 17(1): 93–107. doi:10.37012/jik.v17i1.2171.
- Srinivasan, Bhavana. 2025. "Comparative Analysis of LightGBM and XGBoost for Predictive Risk Assessment in Blockchain Transactions within the Metaverse." *Journal of Current Research in Blockchain* 2(1): 1–12. doi:10.47738/jcrb.v2i1.23.
- Wantoro, Agus, Aviv Fitria Yulia, Dwi Yana Ayu, Syazili Mustofa, Jurusan Teknik Informatika, Universitas Aisyah Pringsewu, Jurusan Pendidikan Dokter, Fakultas Kedokteran, and Universitas Lampung. "EVALUASI KINERJA ALGORITMA MACHINE LEARNING (ML)." (MI): 311–16.