

Pengembangan Aplikasi Website untuk Klasifikasi Status Gizi Balita Menggunakan Integrasi SMOTE pada Algoritma *Naive Bayes*

Damar Kuncoro Adji^{1*}, Taufik Hidayat², Djamaludin³

^{1,2,3} Teknik, Universitas Islam Syekh Yusuf, Jl. Maulana Yusuf No.10, RT.001/RW.003, Babakan, Kec. Tangerang, Kota Tangerang, Banten 15118, Indonesia
E-mail: ¹Damaradji05@gmail.com, ²thidayat@unis.ac.id, ³djamaludin@unis.ac.id

*Penulis Korespondensi

Abstrak— Status gizi balita menjadi indikator krusial dalam menilai kondisi kesehatan dan pertumbuhan anak, sehingga penelitian ini bertujuan mengembangkan sistem klasifikasi status gizi balita berbasis *machine learning* dengan menggunakan data dari Puskesmas Rajeg sebanyak 6062 entri. Proses diawali dengan pemrosesan dataset dalam format .csv, dilanjutkan dengan pembersihan data, pemilihan atribut penting, penanganan data hilang, penghapusan *outlier* menggunakan metode *interquartile range* (IQR), dan penghilangan duplikasi. Selanjutnya, data diolah melalui label *encoding*, normalisasi *MinMaxScaler*, dan dibagi menjadi data latih dan data uji dengan rasio 70:30. Untuk menangani ketidakseimbangan kelas pada atribut target (BB/TB), digunakan metode SMOTE agar distribusi kelas menjadi seimbang. Model klasifikasi yang digunakan adalah *Naive Bayes* dan dievaluasi menggunakan metrik akurasi, presisi, *recall*, serta *F1-score*. Hasil menunjukkan bahwa model tanpa SMOTE memiliki akurasi lebih tinggi (89%) dibandingkan model dengan SMOTE (63%), meskipun model dengan SMOTE lebih baik dalam mengenali kelas minoritas. Sistem ini diimplementasikan dalam bentuk aplikasi *website* prediksi status gizi yang interaktif dan mudah digunakan, sehingga memudahkan tenaga kesehatan dalam menginput data, mencari informasi, serta memantau status gizi balita secara efisien.

Kata Kunci — *Machine Learning, Naive Bayes, SMOTE, Gizi Balita, Klasifikasi.*

Abstract— The nutritional status of toddlers is a crucial indicator in assessing the health and growth of children. Therefore, this study aims to develop a machine learning-based classification system for toddlers' nutritional status using 6,062 entries of data from the Rajeg Community Health Center. The process begins with processing the dataset in .csv format, followed by data cleaning, selecting important attributes, handling missing data, removing outliers using the IQR method, and eliminating duplication. Next, the data is processed through label encoding, *MinMaxScaler* normalization, and divided into training and test data with a 70:30 ratio. To address class imbalance in the target attribute (BW/H), the SMOTE method is used to balance the class distribution. The classification model used is *Naive Bayes* and is evaluated using accuracy, precision, recall, and *F1-score* metrics. The results show that the model without SMOTE has a higher accuracy (89%) than the model with SMOTE (63%), although the model with SMOTE is better at recognizing minority classes. This system is implemented in the form of an interactive and easy-to-use nutritional status prediction application, making it easier for health workers to input data, search for information, and monitor the nutritional status of toddlers efficiently.

Keywords — *Machine Learning, Naive Bayes, SMOTE, Toddler nutrition, Classification.*

I. PENDAHULUAN

Status gizi merupakan indikator utama yang mencerminkan kondisi kesehatan dan kesejahteraan anak balita. Pada periode emas pertumbuhan (*golden age*), pemenuhan gizi yang seimbang menjadi penentu penting terhadap perkembangan fisik, kognitif, dan emosional anak. Namun, kenyataannya, berbagai permasalahan gizi masih banyak ditemukan di Indonesia, termasuk di wilayah Kabupaten Tangerang. Merujuk pada data yang dirilis oleh *World Health Organization* (WHO) pada tahun 2022 84 negara yang disurvei, Indonesia menduduki posisi kedua tertinggi dalam angka kejadian gizi buruk, yang memengaruhi lebih dari 800 ribu anak usia balita [1].

Hasil Riskesdas 2018. Sebanyak 10,2% balita di Indonesia tercatat mengalami kondisi kurang gizi, sementara 3,5% lainnya menderita gizi buruk [2]. Berdasarkan hasil Studi Status Gizi Indonesia tahun 2021, Persentase di Provinsi Banten mencapai 6,3%, dengan tingkat kasus tertinggi ditemukan di wilayah Kabupaten Tangerang [3]. Salah satu wilayah yang menghadapi masalah ini adalah Puskesmas Rajeg. Total 6.062 data balita yang tercatat di wilayah kerja Puskesmas Rajeg, ditemukan bahwa 24,34%

balita mengalami gizi lebih dan 18,47% tergolong obesitas. Angka-angka ini mengindikasikan bahwa masalah gizi, baik kekurangan maupun kelebihan, masih menjadi tantangan serius dalam sistem pelayanan kesehatan dasar [4].

Penilaian status gizi di Indonesia sendiri mengacu pada pedoman Kementerian Kesehatan yang menggunakan standar *antropometri* sesuai ketentuan WHO. Indikator yang digunakan meliputi Berat Badan menurut Tinggi Badan (BB/TB), Lingkar Lengan Atas (LILA), dan Lingkar Kepala (LiPA) untuk anak usia 0–60 bulan. Berdasarkan indikator tersebut, status gizi balita dikategorikan menjadi gizi buruk, gizi kurang, normal, gizi lebih, berisiko gizi lebih, dan obesitas.

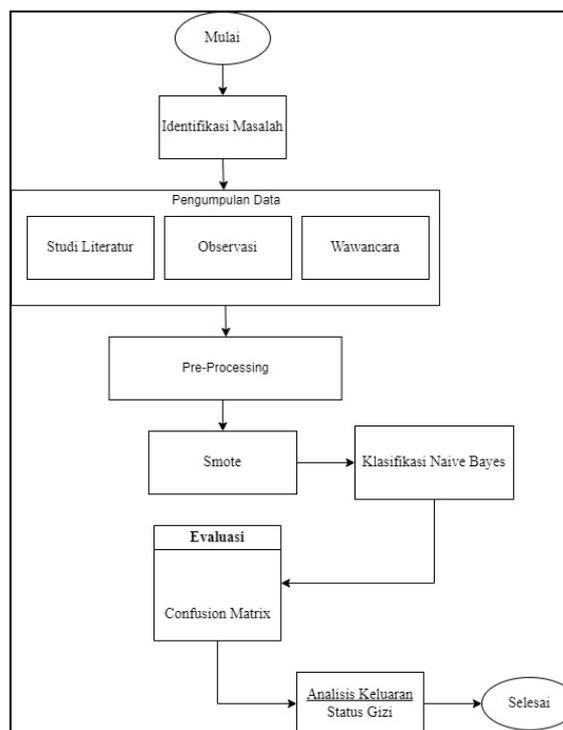
Masalah gizi pada balita dapat disebabkan oleh beragam faktor, antara lain rendahnya pemahaman orang tua tentang pentingnya pola makan yang seimbang, kondisi ekonomi yang kurang mendukung untuk memperoleh asupan makanan bergizi, serta kurang maksimalnya upaya pemantauan terhadap pertumbuhan dan perkembangan anak [5]. Selain itu, salah satu tantangan yang dihadapi dalam proses deteksi status gizi adalah ketidakseimbangan data (*imbalanced data*). Konteks data gizi, mayoritas balita dikategorikan sebagai bergizi normal, sedangkan kasus-kasus gizi buruk dan obesitas jauh lebih sedikit.

Menjawab tantangan tersebut, pendekatan berbasis kecerdasan buatan mulai dikembangkan dalam bidang kesehatan. Salah satunya adalah dengan menggunakan algoritma Naive Bayes untuk mengklasifikasi status gizi balita. Naive Bayes merupakan metode klasifikasi probabilistik yang efisien [6][7]. Karena itu, penelitian ini mengusulkan integrasi metode SMOTE (*Synthetic Minority Over-sampling Technique*) untuk mengatasi ketidakseimbangan data sebelum dilakukan klasifikasi dengan Naive Bayes[8].

Penelitian ini bertujuan untuk membangun sistem klasifikasi status gizi balita berbasis *machine learning* dengan mengintegrasikan SMOTE dan Naive Bayes, menggunakan data dari Puskesmas Rajeg sebagai studi kasus. Selain itu, penelitian ini juga mengimplementasikan sistem klasifikasi dalam bentuk aplikasi interaktif yang dapat digunakan oleh tenaga kesehatan untuk melakukan prediksi status gizi balita dengan cepat dan tepat. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam upaya peningkatan kualitas layanan kesehatan dan penanganan masalah gizi pada anak sejak dini.

II. METODE PENELITIAN

Tahapan penelitian yang dirancang sebagai panduan untuk membantu meneliti dalam menjalankan penelitian ini secara sistematis.



Gambar 1. Tahapan Penelitian

Berikut adalah penjelasan dari tahapan penelitian pada Gambar :

1. Identifikasi masalah
Tahap ini adalah awal penelitian untuk memahami dengan mendalam permasalahan yang ingin dipecahkan.
2. Pengumpulan data
Mengumpulkan data dan informasi yang relevan dengan masalah penelitian dengan metode observasi dan wawancara.
3. Preprocessing
Preprocessing data merupakan langkah awal yang penting dalam analisis data dan *machine learning* Tujuan utamanya adalah memastikan data berada dalam kondisi yang bersih, lengkap, dan siap untuk digunakan oleh model. Tahapan ini membantu meningkatkan kualitas data, mengatasi ketidaksesuaian, serta menyusun data dalam format yang sesuai.

Menurut [9], terdapat beberapa tahapan utama dalam *preprocessing*:

1. *Cleaning Data*
Proses ini bertujuan untuk mengidentifikasi dan memperbaiki data yang rusak, duplikat, tidak konsisten, atau tidak lengkap. Data yang tidak valid dapat menyebabkan bias dan menurunkan akurasi model. Oleh karena itu, pembersihan data menjadi fondasi penting untuk menjamin integritas dataset [10].
2. Penanganan *Missing Value*
Missing value adalah nilai kosong atau hilang pada satu atau beberapa variabel. Keberadaannya dapat memengaruhi hasil analisis jika tidak ditangani dengan tepat. Metode umum yang digunakan meliputi penghapusan baris kosong (jika proporsinya kecil, seperti <5%) atau imputasi menggunakan nilai rata-rata, median, maupun algoritma seperti K-NN. Namun, teknik ini harus diterapkan secara hati-hati agar tidak menimbulkan bias [11].
3. *Reducing Data*
Tahapan ini bertujuan mengurangi ukuran data tanpa menghilangkan informasi penting. Beberapa metode seperti *Principal Component Analysis* (PCA) digunakan untuk mereduksi dimensi, sementara teknik seperti sampling, clustering, dan pruning membantu memilih data yang paling relevan. Hal ini penting terutama ketika menangani big data untuk menghemat waktu dan sumber daya komputasi.
4. Transformasi Data
Transformasi data mengubah format data menjadi lebih sesuai untuk analisis. Salah satu metode yang umum digunakan adalah normalisasi menggunakan *MinMax Scaler*, yang mengubah skala fitur menjadi rentang 0-1 [12]. Normalisasi dengan *Min Max Scaler* menggunakan persamaan (1), dimana X adalah nilai asli dari data yang akan dinormalisasi, X_{min} adalah nilai terkecil dalam dataset untuk fitur tersebut, X_{max} adalah nilai terbesar dalam dataset untuk fitur tersebut dan X_{baru} adalah nilai yang telah dinormalisasi dalam rentang 0 hingga 1.

$$X_{baru} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Kualitas data yang digunakan untuk pelatihan model akan lebih baik dengan menerapkan tahapan *preprocessing* secara sistematis, sehingga meningkatkan kinerja dan akurasi dalam proses klasifikasi maupun prediksi.

4. SMOTE
SMOTE (*Synthetic Minority Over-sampling Technique*) adalah suatu teknik yang digunakan untuk mengatasi ketimpangan distribusi kelas dalam sebuah dataset, khususnya ketika jumlah data pada kelas minoritas jauh lebih sedikit dibandingkan dengan kelas mayoritas [13]. Untuk mengatasi hal ini, SMOTE menghasilkan sampel sintetis dengan cara melakukan interpolasi antara titik data yang ada pada kelas minoritas, bukan hanya menduplikasi data yang sudah ada. Proses ini memilih titik data terdekat dalam kelas minoritas dan menciptakan data baru di antara

titik-titik tersebut. Dengan cara ini, jumlah data kelas minoritas meningkat, yang membantu model untuk belajar dengan lebih baik dan mengurangi bias terhadap kelas mayoritas [14]. Hasilnya, performa model dalam mengklasifikasikan data kelas minoritas pun menjadi lebih akurat.

5. Klasifikasi Naïve Bayes

1. Salah satu tujuan utama *data mining* adalah klasifikasi, yaitu proses untuk mengungkap pola atau aturan yang ada dalam dataset berskala besar. Hal ini dicapai dengan membagi data menjadi kelompok-kelompok atau kelas-kelas yang telah didefinisikan sebelumnya [15]. Proses ini dilakukan dengan menganalisis data yang telah tersedia (data latih) guna membentuk sebuah model yang nantinya dapat digunakan untuk memprediksi atau menentukan kategori dari data yang belum dikenal sebelumnya.
2. Naive Bayes memprediksi kelas dengan probabilitas tertinggi untuk data baru. Algoritma ini dikenal sederhana, cepat, dan efektif, terutama untuk dataset besar dengan banyak atribut. Karena efisiensinya, Naive Bayes banyak digunakan dalam aplikasi seperti deteksi spam, klasifikasi teks, analisis sentimen, dan prediksi status kesehatan seperti stunting atau gizi balita [16][17]. Rumus dasar algoritma Naive Bayes adalah sebagai berikut :

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (2)$$

Pada persamaan tersebut, $P(C|X)$ merupakan probabilitas suatu kelas C apabila atribut X diketahui atau sering disebut probabilitas posterior. Nilai $P(X|C)$ menyatakan peluang kemunculan atribut X pada kelas C atau *likelihood*. Sementara itu, $P(C)$ adalah probabilitas awal dari kelas C sebelum mempertimbangkan data atribut, yang disebut *prior probability*. Adapun $P(X)$ menunjukkan probabilitas kemunculan atribut X secara keseluruhan atau *evidence*.

6. Evaluasi

Setelah proses klasifikasi selesai dilakukan evaluasi untuk menentukan 4 macam hasil performa *machine learning* yaitu akurasi, presisi, *recall*, dan *F-Score* dengan menggunakan *Confusion Matrix*.

Confusion Matrix adalah teknik evaluasi yang digunakan untuk menilai kinerja sebuah algoritma klasifikasi. Dengan menggunakan matriks ini dapat melihat sejauh mana model mampu mengelompokkan data ke dalam kelas yang sesuai secara akurat [18][19]. Tabel ini menampilkan jumlah prediksi yang sesuai dengan kelas sebenarnya (klasifikasi benar) maupun yang tidak sesuai (klasifikasi salah) [20].

1. Mengevaluasi kedekatan antara hasil prediksi model dan nilai sebenarnya, digunakanlah akurasi. Nilai ini diperoleh dengan membandingkan jumlah data yang sukses dikelompokkan dengan benar terhadap total keseluruhan data. Rumus perhitungan akurasi adalah sebagai berikut:

$$Accuracy = \frac{TP + TN}{N} \quad (3)$$

2. Metrik evaluasi, presisi mengukur proporsi data yang sebenarnya positif dari semua data yang diklasifikasikan model sebagai positif. Dengan kata lain, presisi mencerminkan ketepatan model dalam mengidentifikasi hasil yang dianggap relevan atau positif. Persamaan presisi dapat dituliskan sebagai:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

3. *Recall* adalah indikator seberapa baik suatu model dalam mengidentifikasi semua kasus positif yang relevan. Metrik ini dihitung sebagai proporsi antara jumlah data positif yang

berhasil diprediksi dengan benar oleh model, dibagi dengan total keseluruhan data yang memang berlabel positif di dalam dataset. Rumus recall adalah:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

4. *F1-Score* adalah cara untuk mengevaluasi kinerja model dengan menggabungkan presisi dan *recall*. Ini sangat berguna saat jumlah data di setiap kategori (kelas) tidak seimbang. Dengan menggunakan rata-rata harmonik, *F1-Score* memberikan gambaran yang lebih akurat tentang seberapa baik model bekerja, karena tidak hanya fokus pada satu metrik saja, melainkan mencari keseimbangan antara kemampuan model untuk mengidentifikasi hasil yang benar (presisi) dan menemukan semua hasil yang relevan (*recall*). Rumus *F1-Score* ditunjukkan sebagai berikut:

$$F1 - Score = \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

7. Pembuatan Aplikasi

Sebagai luaran utama, penelitian ini berhasil mengembangkan sebuah aplikasi dasar untuk klasifikasi status gizi balita yang dioperasikan dengan algoritma Naive Bayes dan SMOTE.

III. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan membangun sistem klasifikasi status gizi balita dengan menerapkan algoritma Naive Bayes yang dikombinasikan dengan metode SMOTE untuk mengatasi ketidakseimbangan distribusi kelas dalam data. Dataset yang digunakan berasal dari Puskesmas Rajeg dan terdiri dari 6062 data entri, mencakup sejumlah fitur seperti usia, berat badan, tinggi badan, lingkaran lengan atas (LILA), serta lingkaran kepala (LIPA). Target klasifikasi difokuskan pada status gizi berdasarkan rasio berat badan terhadap tinggi badan (BB/TB) yang telah dibagi ke dalam enam kategori, yaitu gizi buruk, gizi kurang, gizi baik, berisiko gizi lebih, gizi lebih, dan obesitas.

Memanggil Dataset

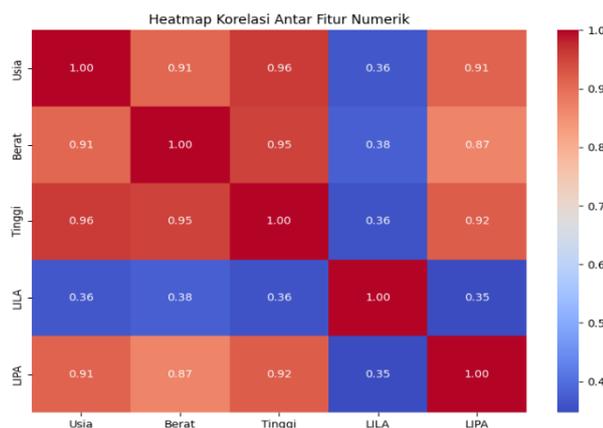
Dataset yang digunakan adalah file *.csv yang di-import menggunakan library pandas, seperti yang ditunjukkan pada Tabel 1.

Preprocessing

Proses ini dilakukan dengan berbagai tahapan seperti menangani *missing data*, duplikat data, deteksi dan penanganan *outlier*, korelasi, *encode label*, *split data*, normalisasi data, SMOTE dan model klasifikasi.

Tabel 1. Dataset

No	Usia	Berat Badan	Tinggi Badan	LILA	LIPA	BB/TB
1	50	12	88	16	48,8	Normal
2	58	12	87,1	16,2	48,6	Normal
3	10	17,7	107,8	17	50,7	Normal
4	35	17	105,6	16,8	50,5	Normal
5	14	16,2	102	16,9	50,3	Normal
6060	18	8,9	82	19	47,13	Gizi Kurang
6061	17	10,1	82	20	47,2	Normal
6062	25	10,5	86	17	48	Normal
6063	28	10,8	87	20	48,6	Normal



Gambar 2. Korelasi Antar Atribut

- 1) Penanganan *Missing Value* dan Duplikat
Ditemukan nilai kosong dan duplikat dalam data dengan atribut “Usia”, “Berat”, “Tinggi”, “LILA”, “LIPA”, “BB/TB”. Seluruh entri dengan nilai kosong dihapus, dan baris duplikat dieliminasi dengan `drop_duplicates(keep=False)`. Langkah ini paling penting untuk menjaga integritas data.
- 2) Deteksi dan Penanganan *Outlier*
Outlier pada atribut numerik diidentifikasi menggunakan metode *interquartile range (IQR)*. Data diluar batas $Q1 - 1.5IQR$ dan $Q3 + 1.5IQR$ dihapus untuk mencegah distorsi dalam proses pelatihan model.
- 3) Korelasi Antar Atribut
Analisis korelasi menunjukkan hubungan kuat antara usia, berat, dan tinggi badan. Sementara itu, atribut LILA memiliki korelasi rendah dengan atribut lainnya, mengindikasikan informasi independen yang bermanfaat bagi klasifikasi seperti ditunjukkan pada Gambar 2. Berdasarkan hasil *heatmap* korelasi antar fitur numerik, diketahui bahwa terdapat hubungan yang sangat kuat antara usia, berat, tinggi, dan LIPA dengan nilai korelasi di atas 0.87, menunjukkan bahwa penambahan usia balita umumnya diikuti oleh peningkatan berat dan tinggi badan serta lingkar pinggang. Fitur tinggi memiliki korelasi tertinggi terhadap usia (0.96) dan berat badan (0.95), menandakan keterkaitan erat antar variabel pertumbuhan. Sebaliknya, fitur LILA menunjukkan korelasi rendah terhadap semua fitur lainnya (berkisar 0.35–0.38), yang mengindikasikan bahwa LILA memberikan informasi yang lebih independen dan unik dalam dataset. Hal ini menjadikan LILA sebagai fitur penting yang dapat membantu model klasifikasi membedakan status gizi balita secara lebih akurat tanpa bias terhadap variabel lain yang saling berkaitan.
- 4) *Encode Label*
Label pada atribut target BB/TB dikodekan ke dalam bentuk numerik agar dapat diterima oleh algoritma klasifikasi. Misalnya, “Normal” = 0, “Gizi Buruk” = 1, dan seterusnya.
- 5) Pembagian Data
Data dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian menggunakan *pseudo code train_test_split* dengan stratifikasi label agar distribusi kelas tetap seimbang.
- 6) Normalisasi
Normalisasi menggunakan *MinMaxScaler* diterapkan agar skala antar fitur seragam dalam rentang 0 hingga 1, meningkatkan stabilitas model.

Penyeimbangan Data Menggunakan SMOTE

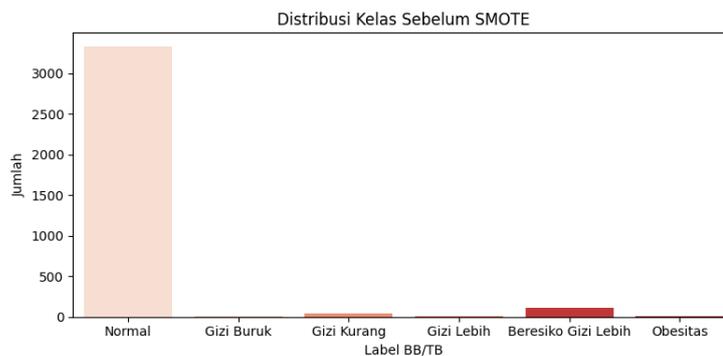
Kelas BB/TB sangat tidak seimbang. Kelas "Normal" memiliki jumlah data yang sangat besar (lebih dari 3000), sedangkan kelas lainnya seperti "Gizi Buruk", "Gizi Kurang", "Gizi Lebih", "Beresiko Gizi Lebih", dan "Obesitas" jumlahnya sangat sedikit. Ketimpangan ini dapat menyebabkan model klasifikasi bias terhadap kelas mayoritas, sehingga performa terhadap kelas minoritas cenderung rendah. Setelah dilakukan proses SMOTE, distribusi kelas menjadi seimbang. Setiap kelas kini memiliki jumlah data yang hampir sama, yaitu sekitar 3300 data. Teknik SMOTE bekerja dengan menambahkan data

sintetis pada kelas-kelas minoritas, sehingga model memiliki kesempatan belajar yang sama dari semua kelas. Ini membantu meningkatkan akurasi klasifikasi terutama pada kelas yang sebelumnya kurang terwakili.

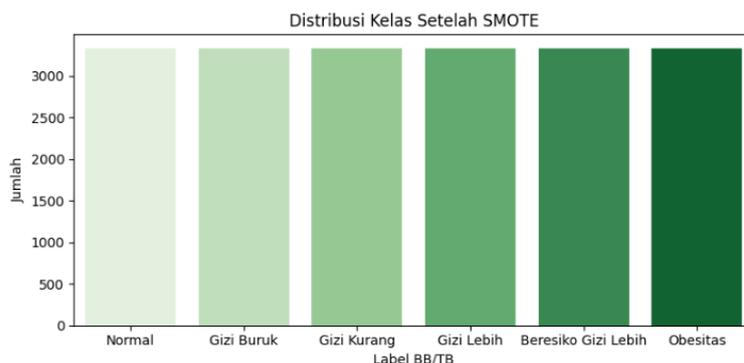
Hasil Klasifikasi Naïve Bayes

Tanpa SMOTE: Akurasi tinggi (89%), namun model kurang efektif dalam mengenali kelas minoritas seperti "Gizi Buruk" dan "Obesitas". Hal ini ditunjukkan oleh nilai *recall* dan *F1-score* yang rendah pada kelas minoritas. SMOTE: Akurasi menurun menjadi 62%, namun distribusi *recall* dan *precision* menjadi lebih seimbang di seluruh kelas. Hal ini membuktikan bahwa *SMOTE* meningkatkan sensitivitas model terhadap kelas minoritas.

Hasil perbandingan antara model Naïve Bayes tanpa dan dengan SMOTE menunjukkan bahwa model tanpa SMOTE menghasilkan performa yang lebih seimbang dengan akurasi 85%, precision 91%, recall 85%, dan F1-score 87%, menandakan bahwa model bekerja baik secara keseluruhan, meskipun cenderung bias terhadap kelas mayoritas. Sebaliknya, setelah diterapkan SMOTE untuk menyeimbangkan data, precision meningkat menjadi 93% (menandakan model lebih mampu mengenali kelas minoritas), namun akurasi dan recall menurun signifikan menjadi 62%, yang berdampak pada penurunan F1-score menjadi 73%. Hal ini menunjukkan bahwa meskipun SMOTE membantu memperhatikan kelas minoritas, model menjadi kurang stabil dalam mengenali data secara keseluruhan dan mengalami penurunan generalisasi terhadap distribusi data yang asli.



Gambar 3. Distribusi Sebelum SMOTE



Gambar 4. Distribusi Sesudah SMOTE

Tabel 2. Hasil Klasifikasi.

Model Klasifikasi	Pembagian Data	Acuracy	Precision	Recall	F1-Score
Naïve Bayes	70:30	85%	91%	85%	87%
Naïve Bayes + SMOTE	70:30	62%	93%	62%	73%

The image shows a web application interface titled "Prediksi Status Gizi Balita". It features a form with five input fields: "Usia (bulan)", "Berat (kg)", "Tinggi (cm)", "LILA (cm)", and "LIPA". Below these fields is a purple button labeled "Prediksi". The entire form is enclosed in a blue border.

Gambar 5. Tampilan Aplikasi

Implementasi Aplikasi

Gambar 5 menunjukkan antarmuka form input untuk fitur Prediksi Status Gizi Balita pada sebuah aplikasi. Form ini memungkinkan pengguna untuk memasukkan data balita secara manual, yaitu usia (dalam bulan), berat badan (kg), tinggi badan (cm), LILA (Lingkar Lengan Atas), dan LIPA (Lingkar Kepala). Setelah data diisi, pengguna dapat menekan tombol "Prediksi" untuk mendapatkan hasil klasifikasi status gizi balita berdasarkan data yang dimasukkan. Fitur ini memanfaatkan model prediktif yang telah dilatih sebelumnya dan dirancang untuk membantu tenaga kesehatan atau orang tua dalam memantau dan menilai kondisi gizi anak secara cepat dan praktis berbasis data antropometri. Desain antarmuka sederhana ini menunjukkan fokus pada kemudahan penggunaan dan kejelasan input.

IV. KESIMPULAN

Berdasarkan hasil penelitian dan implementasi yang telah dilakukan, dapat disimpulkan bahwa sistem klasifikasi status gizi balita berbasis *machine learning* berhasil dibangun dengan baik dan terstruktur. Proses dimulai dari pengumpulan data sebanyak 6062 entri dari Puskesmas Rajeg, dilanjutkan dengan eksplorasi untuk memastikan tidak ada kesalahan *input* serta memahami struktur data. Pada tahap seleksi dan pembersihan data, atribut yang tidak relevan seperti NIK dan nama dihapus, menyisakan atribut penting yaitu Usia, Berat, Tinggi, LILA, LIPA, dan BB/TB sebagai label. Tahap *preprocessing* mencakup pengecekan *missing value*, penghapusan *outlier* menggunakan metode *IQR*, penghapusan data duplikat, *encoding* label target ke bentuk numerik, pembagian data menjadi 70% data latih dan 30% data uji secara *stratified*, normalisasi dengan *MinMaxScaler*, serta penyeimbangan kelas menggunakan metode SMOTE. Model klasifikasi dibangun menggunakan algoritma *Naïve Bayes*, dan dievaluasi dalam dua kondisi: tanpa SMOTE yang menghasilkan akurasi sebesar 89% dengan *precision* 90%, *recall* 88%, dan *F1-score* 89%. Setelah SMOTE yang menghasilkan akurasi menurun menjadi 63%, namun *precision* meningkat menjadi 93% yang menunjukkan peningkatan dalam mengenali kelas minoritas. Implementasi sistem dilakukan melalui aplikasi interaktif dengan fitur seperti beranda, data lengkap, pencarian data, *preprocessing*, dan prediksi, yang memungkinkan pengguna memasukkan data dan memperoleh hasil klasifikasi status gizi balita secara cepat dan akurat. Secara keseluruhan, sistem ini dapat mendukung petugas kesehatan dalam memantau dan mendeteksi dini masalah gizi balita, sehingga berpotensi menjadi alat bantu yang efektif dalam menciptakan generasi anak yang sehat dan berkualitas.

UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah mendukung, membantu, dan berkontribusi dalam penyelesaian penelitian ini. Penghargaan khusus juga disampaikan kepada dosen pembimbing, rekan-rekan, serta individu lain yang terlibat langsung maupun tidak langsung dalam proses ini.

DAFTAR PUSTAKA

- [1] 2023 World Health Organization, "Level and trend in child malnutrition," *World Heal. Organ.*, p. 4, 2023, [Online]. Available: <https://www.who.int/publications/i/item/9789240073791>
- [2] 2023 Kementerian Kesehatan RI, *Laporan Ski Tematik 2023*. 2023.
- [3] F. Ongko, C. Willyana, M. Liman, and Z. Atzmardina, "Kasus Gizi Buruk pada An. MFA dengan Pendekatan Kedokteran Keluarga di Wilayah Kerja Puskesmas Gembong, Kecamatan Balaraja, Kabupaten Tangerang, Provinsi Banten Periode: 10 November 2022 – 12 Desember 2022," *Malahayati Nurs. J.*, vol. 5, no. 8, pp. 2518–2537, 2023, doi: 10.33024/mnj.v5i8.9311.
- [4] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, "Aplikasi Naive Bayes Classifier (NBC) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *J. Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, doi: 10.14710/j.gauss.v11i1.33991.
- [5] A. Rizal, D. Rahmawati, and S. A. Murad, "Perancangan Aplikasi Informasi Gizi Balita Pada Posyandu Edelweis Serpong 2 Berbasis Android Menggunakan Metode Extreme Programming," vol. 12, no. 1, pp. 102–114, 2024.
- [6] T. Hidayat, M. J. Siddiq, S. Jayasri, A. Suhendi, and R. Rizky, "Analisis Sentimen Opini Masyarakat Terhadap Pilkada 2024 Di Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, pp. 609–622, 2025, doi: 10.23960/jitet.v13i2.6280.
- [7] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *J. Inform.*, vol. 8, no. 2, pp. 111–122, 2021, doi: 10.31294/ji.v8i2.10494.
- [8] R. A. Nurdian, Mujib Ridwan, and Ahmad Yusuf, "Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 1, pp. 24–32, 2022, doi: 10.28932/jutisi.v8i1.4004.
- [9] D. Varma, A. Nehansh, and P. Swathy, "Data Preprocessing Toolkit : An Approach to Automate Data Preprocessing," *Interantional J. Sci. Res. Eng. Manag.*, vol. 07, no. 03, pp. 1–5, 2023, doi: 10.55041/ijsrem18270.
- [10] B. Setio and P. Prasetyaningrum, "Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means," *J. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 27–32, 2021, doi: 10.54840/jstech.v1i1.9.
- [11] K. H. Erian, P. H. Regalado, and J. M. Conrad, "Missing data handling for machine learning models," *IAES Int. J. Robot. Autom.*, vol. 10, no. 2, p. 123, 2021, doi: 10.11591/ijra.v10i2.pp123-132.
- [12] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *J. Manaj. Inform.*, vol. 11, no. 2, pp. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [13] A. Nugroho and E. Rilvani, "Penerapan Metode Oversampling SMOTE Pada Algoritma Random Forest Untuk Prediksi Kebangkrutan Perusahaan," *Techno.Com*, vol. 22, no. 1, pp. 207–214, 2023, doi: 10.33633/tc.v22i1.7527.
- [14] M. P. Pulungan, A. Purnomo, and A. Kurniasih, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, pp. 1033–1042, 2024, doi: 10.25126/jtiik.2024117989.
- [15] H. Hafizan and A. N. Putri, "Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 1, no. 2, pp. 68–72, 2020, doi: 10.30645/kesatria.v1i2.23.
- [16] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.
- [17] F. Fiddin, M. Y. Syahbarna, and M. Ridwan, "Penggunaan Supervised Learning untuk Prediksi Validitas Ulasan Negatif Aplikasi Tokopedia Berdasarkan Pengalaman Pengguna Ahli," vol. 23, pp. 409–417, 2024, doi: <https://doi.org/10.53513/jis.v23i2.10030>.
- [18] T. D. Putra, E. Utami, and M. P. Kurniawan, "Analisis Sentimen Pemilu 2024 dengan Naive Bayes Berbasis Particle Swarm Optimization (PSO)," *Explore*, vol. 13, no. 1, pp. 1–5, 2023, doi: 10.35200/ex.v11i2.13.

- [19] D. Valero-Carreras, J. Alcaraz, and M. Landete, "Comparing two SVM models through different metrics based on the confusion matrix," *Comput. Oper. Res.*, vol. 152, no. December 2022, p. 106131, 2023, doi: 10.1016/j.cor.2022.106131.
- [20] A. W. Septyanto and H. L. Hariyanto, "Perbandingan Teknik Klasifikasi Catatan Medis untuk Indeks Antropometri Status Gizi Balita," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 1, pp. 229–235, 2024, doi: 10.47233/jteksis.v6i1.1064.