Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

Klasifikasi Status Gizi Bayi sebagai Upaya Deteksi Dini Masalah Gizi Menggunakan Algoritma Naive Bayes

Infant Nutritional Status Classification for Early Detection of Nutritional Problems Using Naive Bayes Algorithm

Muhammad Rafif AR, Fathoni*, Ken Ditha Tania, Allsela Meiriza, Ali Ibrahim

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia
Email: [1]09031382328127@student.unsri.ac.id, [2,*]fathoni@unsri.ac.id, [3]kenditha@ilkom.unsri.ac.id,
[4]allsela_meiriza@yahoo.co.id, [5]aliibrahim210784@gmail.com
Email Penulis Korespondensi: fathoni@unsri.ac.id
Received: **06/03/2026**, *Revised*: **18/04/2026**, *Accepted*: **04/09/2026**, *Published*: **08/09/2026**

Abstrak−Permasalahan dalam mengklasifikasikan status gizi bayi masih menjadi tantangan dalam pemanfaatan data kesehatan, terutama dalam menghasilkan model prediksi yang akurat dan terpercaya guna mendukung pengambilan keputusan di tingkat layanan kesehatan dasar. Penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes dalam proses klasifikasi status gizi bayi dengan memanfaatkan indikator antropometri, yaitu berat badan menurut umur (BB/U) dan tinggi badan menurut umur (TB/U). Data yang digunakan bersumber dari laporan kegiatan Posyandu di Puskesmas Citra Medika selama periode 2023 hingga 2025, dengan jumlah total 10.654 data bayi. Distribusi kelas pada dataset meliputi kategori normal, gizi baik, gizi kurang, gizi buruk, gizi lebih, dan risiko gizi lebih. Proporsi antar kelas yang relatif seimbang menyebabkan penelitian ini tidak memerlukan teknik penyeimbangan data seperti oversampling. Tahapan penelitian mencakup proses pembersihan data (data cleaning), penyeragaman kategori, transformasi label kelas, serta pembagian dataset menggunakan metode stratified sampling dengan komposisi 80% data latih dan 20% data uji. Hasil evaluasi menunjukkan bahwa model Naïve Bayes mampu mencapai tingkat akurasi sebesar 82%. Nilai precision tercatat sebesar 0,92 untuk kelas normal, 0,39 untuk kelas gizi bermasalah, dan 0,14 untuk kelas overweight. Kemudian nilai recall masing-masing sebesar 0,89 pada kelas normal, 0,63 pada kelas gizi bermasalah, dan 0,08 pada kelas overweight. Nilai F1-score yang diperoleh adalah 0,83 untuk kelas normal, 0,48 untuk kelas gizi bermasalah, dan 0,11 untuk kelas overweight. Hasil tersebut menunjukkan bahwa model memiliki performa yang cukup baik dalam mengklasifikasikan status gizi bayi berdasarkan data antropometri. Jika algoritma Naïve Bayes dapat dianggap efektif untuk digunakan dalam klasifikasi kondisi gizi bayi berbasis indikator BB/U dan TB/U. Penelitian ini diharapkan dapat berkontribusi dalam pengembangan sistem berbasis data guna meningkatkan kualitas analisis serta mendukung pengambilan keputusan pada layanan kesehatan tingkat dasar.

Kata Kunci: Naïve Bayes; Klasifikasi Gizi Bayi; Indikator Antropometri; Berat Badan Menurut Umur (BB/U); Tinggi Badan menurut Umur (TB/U); Data Mining

Abstract−The classification of infant nutritional status remains a significant challenge in the utilization of health data, particularly in developing predictive models that are accurate and reliable for supporting decision-making at the primary healthcare level. This study aims to implement the Naïve Bayes algorithm to classify infant nutritional status based on anthropometric indicators, namely weight-for-age (W/A) and height-for-age (H/A). The data used in this study were obtained from Posyandu activity reports at Citra Medika Public Health Center covering the period from 2023 to 2025, with a total of 10,654 infant records. The class distribution in the dataset includes normal, well-nourished, undernourished, severely undernourished, overnourished, and at risk of overnutrition categories. Since the class proportions are relatively balanced, no oversampling technique was applied. The research process involved several stages, including data cleaning, category normalization, class label transformation, and dataset splitting using stratified sampling with a composition of 80% training data and 20% testing data. The evaluation results indicate that the Naïve Bayes model achieved an accuracy of 82%. The precision values were 0.92 for the normal class, 0.39 for the malnutrition class, and 0.14 for the overweight class. The recall values were 0.89 for the normal class, 0.63 for the malnutrition class, and 0.08 for the overweight class. Meanwhile, the F1-scores were 0.83 for the normal class, 0.48 for the malnutrition class, and 0.11 for the overweight class. These findings suggest that the model demonstrates fairly good performance in classifying infant nutritional status based on anthropometric data. Therefore, the Naïve Bayes algorithm can be considered effective for classifying infant nutritional conditions using W/A and H/A indicators. This study is expected to contribute to the development of data-driven systems to enhance analytical quality and support decision-making in primary healthcare services.

Keywords: Naïve Bayes; Infant Nutrition Classification; Anthropometric Indicators; Weight-for-Age (W/A); Height-for-Age (H/A); Data Mining

1. PENDAHULUAN

Permasalahan gizi pada bayi dan balita merupakan isu krusial dalam kesehatan masyarakat yang memerlukan penanganan berbasis data secara cepat dan akurat. Berdasarkan laporan World Health Organization tahun 2021, sekitar 22% anak di dunia mengalami stunting, sementara 6,7% mengalami wasting, yang menunjukkan masih tingginya prevalensi gangguan pada anak usia dini dengan gizinya [1]. Di Indonesia, kondisi ini juga menjadi perhatian serius, sebagaimana dilaporkan oleh UNICEF (2023) bahwa masih terdapat proporsi signifikan anak dengan gizi kurang dan gizi buruk [2]. Tingginya angka tersebut menunjukkan urgensi pengembangan sistem yang mampu melakukan analisis kondisi gizi secara otomatis untuk mendukung deteksi dan penanganan yang lebih cepat.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

Namun, dalam praktiknya, proses penentuan kondisi gizi bayi di tingkat layanan kesehatan dasar masih banyak dilakukan secara manual dengan membandingkan parameter antropometri seperti berat badan, tinggi badan, dan usia terhadap standar tertentu. Proses ini tidak hanya memerlukan waktu yang relatif lama, tetapi juga berpotensi menimbulkan kesalahan dalam interpretasi data, terutama ketika jumlah data yang dianalisis cukup besar. Oleh karena itu, diperlukan pendekatan berbasis komputasi yang mampu mengotomatisasi proses klasifikasi kondisi gizi secara lebih efisien dan konsisten.

Kemajuan teknologi informasi, terutama pada bidang data mining dan machine learning, membuka peluang yang luas untuk mengatasi permasalahan tersebut. Salah satu metode yang sering dimanfaatkan dalam klasifikasi data medis adalah algoritma Naïve Bayes, karena keunggulannya dalam mengolah data dengan jumlah terbatas, memiliki proses komputasi yang efisien, serta menunjukkan kinerja yang cukup baik pada berbagai penelitian klasifikasi [15]. Pada bidang kesehatan, algoritma ini dinilai sesuai karena mampu merepresentasikan probabilitas suatu kejadian berdasarkan atribut-atribut sederhana, seperti data antropometri.

Sejumlah penelitian sebelumnya telah menerapkan metode machine learning untuk klasifikasi kondisi gizi. Penelitian oleh Putri et al. (2020) [9] menggunakan Naïve Bayes berbasis nilai Z-score dan indeks antropometri, namun masih terbatas pada atribut tertentu dan belum mengeksplorasi integrasi data dari berbagai periode. Selanjutnya, Saputra et al. (2023) [10] menerapkan Naïve Bayes untuk klasifikasi kondisi kesehatan balita, tetapi dataset yang digunakan relatif kecil sehingga berpotensi memengaruhi generalisasi model. Penelitian oleh Widhari et al. (2024) [5] membandingkan Naïve Bayes dan K-NN dalam diagnosis stunting, namun fokus penelitian lebih pada perbandingan algoritma daripada optimalisasi proses klasifikasi berbasis data terintegrasi. Selain itu, Miranda et al. (2024) [7] mengembangkan model prediksi berbasis machine learning untuk data tidak seimbang, tetapi memerlukan teknik penyeimbangan data yang menambah kompleksitas proses.

Berdasarkan tinjauan tersebut, terdapat beberapa kelemahan pada penelitian sebelumnya, yaitu: (1) keterbatasan jumlah dan cakupan dataset, (2) penggunaan atribut yang belum terintegrasi secara optimal, (3) fokus pada perbandingan metode tanpa optimalisasi implementasi pada data nyata layanan kesehatan, serta (4) kebutuhan teknik tambahan seperti oversampling yang meningkatkan kompleksitas model. Oleh karena itu, penelitian ini mengusulkan pendekatan yang berbeda dengan memanfaatkan dataset terintegrasi dari beberapa periode (2023–2025), menggunakan atribut antropometri utama (BB/U dan TB/U), serta menerapkan algoritma Naïve Bayes tanpa teknik oversampling karena distribusi data yang relatif seimbang.

Kontribusi utama penelitian ini untuk pengembangan model klasifikasi status gizi bayi berbasis algoritma Naïve Bayes yang bersifat sederhana, memiliki proses komputasi yang cepat, serta tetap mampu memberikan tingkat akurasi yang baik dengan memanfaatkan data antropometri terintegrasi (BB/U dan TB/U). Model yang dihasilkan diharapkan dapat mendukung tenaga kesehatan dalam melakukan deteksi dini terhadap permasalahan gizi secara lebih efektif dan efisien di fasilitas layanan kesehatan tingkat pertama.

Penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes dalam proses pengelompokan kondisi gizi bayi berdasarkan data antropometri yang tersedia. Melalui pendekatan tersebut, diharapkan model yang dibangun tidak hanya mudah diimplementasikan, tetapi juga mampu menghasilkan kinerja yang optimal. Kemudian hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan di sektor kesehatan, khususnya dalam meningkatkan ketepatan dan efisiensi analisis terhadap kondisi gizi bayi pada layanan kesehatan primer.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan yang terstruktur dan sistematis untuk menghasilkan model klasifikasi yang mampu mengidentifikasi status gizi bayi secara akurat berdasarkan data antropometri. Secara umum, proses penelitian meliputi studi literatur, pengumpulan serta integrasi data, tahap praproses (preprocessing), transformasi data, pembagian dataset, proses pelatihan model menggunakan algoritma Naïve Bayes, hingga tahap evaluasi kinerja model. Seluruh tahapan tersebut dirancang secara runtut guna memastikan bahwa proses pengolahan data dan pengembangan model berjalan secara optimal serta menghasilkan output yang dapat dipertanggungjawabkan secara ilmiah.

Sebagai dasar konseptual, penelitian ini diawali dengan kajian literatur terhadap berbagai publikasi ilmiah terbaru yang membahas penerapan machine learning di bidang kesehatan, khususnya terkait klasifikasi status gizi bayi dan balita. Kajian ini bertujuan untuk mengidentifikasi metode yang relevan, memahami keunggulan dan keterbatasan masing-masing algoritma, serta menemukan peluang pengembangan penelitian lebih lanjut. Berdasarkan hasil kajian tersebut, algoritma Naïve Bayes dipilih karena memiliki kelebihan dalam hal kesederhanaan model, efisiensi dalam komputasi, serta kemampuan yang cukup baik dalam menangani berbagai kasus klasifikasi pada data medis.

Tahapan penelitian selanjutnya direpresentasikan dalam bentuk diagram alur (flowchart) yang menggambarkan keterkaitan antar proses secara menyeluruh, mulai dari tahap pengumpulan data hingga evaluasi model. Diagram tersebut memberikan gambaran yang jelas mengenai urutan langkah penelitian serta hubungan antar setiap tahapan yang dilakukan.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

Mengacu pada diagram alur tersebut, proses penelitian diawali dengan pengumpulan data yang bersumber dari laporan kegiatan Posyandu di Puskesmas Citra Medika selama periode 2023 hingga 2025. Data yang diperoleh kemudian digabungkan menjadi satu kesatuan dataset untuk selanjutnya diproses pada tahap praproses. Pada tahap ini dilakukan pembersihan data guna memastikan kualitas dataset, yang mencakup penanganan data yang tidak lengkap, penghapusan data duplikat, serta penyesuaian format data. Setelah itu, dilakukan tahap transformasi data agar sesuai dengan kebutuhan algoritma, termasuk proses encoding pada atribut kategorikal dan penyesuaian label kelas.

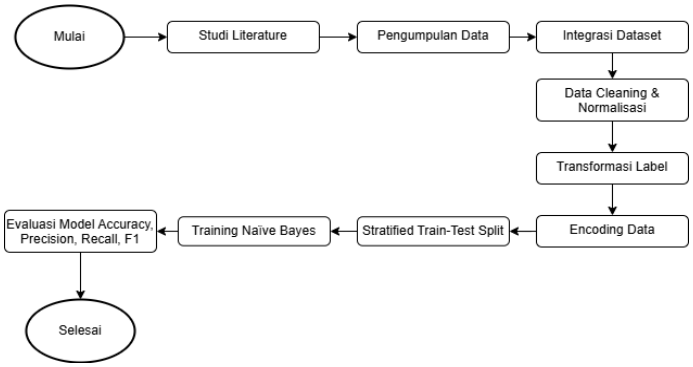
Tabel 1. Pembagian Dataset

Jenis Data	Persentase
Data Latih	80%
Data Uji	20%

Pembagian dataset dilakukan dengan teknik stratified sampling dengan rasio 80% untuk data pelatihan dan 20% untuk data pengujian. Teknik ini digunakan agar komposisi setiap kelas tetap terjaga secara seimbang pada kedua bagian data tersebut. Setelah seluruh data melalui tahap praproses, langkah berikutnya adalah memisahkan dataset menjadi dua bagian, yaitu data latih dan data uji, dengan tetap mempertahankan proporsi distribusi kelas.

Selanjutnya, pembangunan model dilakukan dengan menerapkan algoritma Gaussian Naïve Bayes yang menggunakan pendekatan probabilistik serta mengasumsikan bahwa setiap atribut saling independen. Setelah model selesai dilatih, dilakukan pengujian kinerja dengan menggunakan metrik evaluasi seperti accuracy, precision, recall, dan F1-score untuk mengetahui tingkat ketepatan model dalam mengklasifikasikan kondisi gizi bayi.

Melalui tahapan penelitian yang disusun secara sistematis sebagaimana tergambar pada diagram alur, penelitian ini diharapkan mampu menghasilkan model klasifikasi yang tidak hanya memiliki tingkat ketepatan yang baik, tetapi juga efisien serta praktis untuk diterapkan dalam sistem pendukung keputusan di bidang kesehatan, khususnya pada layanan kesehatan dasar. Diagram alur tahapan penelitian tersebut ditampilkan pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

2.1 Studi Literatur

Tahapan studi literatur dilakukan dengan mengkaji 16 jurnal ilmiah yang diterbitkan pada rentang tahun 2020 hingga 2025, yang berfokus pada klasifikasi status gizi bayi menggunakan pendekatan machine learning. Hasil penelaahan menunjukkan bahwa mayoritas penelitian memanfaatkan algoritma seperti Naïve Bayes, Decision Tree, dan Support Vector Machine, dengan variabel utama berupa indikator antropometri, seperti weight-for-age (W/A), height-for-age (H/A), serta weight-for-height (W/H).

Meskipun begitu pada beberapa penelitian terdahulu masih menunjukkan sejumlah keterbatasan, di antaranya penggunaan dataset dengan jumlah yang relatif kecil, pemilihan atribut yang belum optimal, serta proses evaluasi model yang belum dilakukan secara menyeluruh. Berdasarkan hasil analisis tersebut, penelitian ini memfokuskan pada penggunaan variabel utama BB/U dan TB/U serta memanfaatkan dataset yang telah terintegrasi guna meningkatkan kinerja klasifikasi sekaligus mengatasi kekurangan yang ditemukan pada studi sebelumnya.

2.2 Akuisisi dan Integrasi Data

Tahap selanjutnya adalah proses akuisisi data yang bersumber dari laporan Posyandu Puskesmas Citra Medika periode tahun 2023 hingga 2025 dalam format multi-sheet Microsoft Excel. Setiap lembar kerja merepresentasikan periode pencatatan yang berbeda, sehingga dilakukan proses integrasi untuk menggabungkan seluruh sheet menjadi satu dataset terpadu. Hasil integrasi menghasilkan dataset dengan total sebanyak N = 16 dataset/jurnal. Integrasi diperkukan dan diharapkan untuk mendapat data yang komprehensif dan berkesinambungan antar tahun sehingga analisis yang dilakukan dapat mencerminkan kondisi riil secara longitudinal.

2.3 Preprocessing dan Data Cleaning

Setelah integrasi, dilakukan tahap *preprocessing* dan pembersihan data untuk memastikan kualitas dataset. Tahapan ini mencakup penghapusan data yang memiliki nilai kosong atau tidak lengkap pada variabel-variabel utama, seperti

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

jenis kelamin (JK), BB/U (*weight-for-age*), TB/U (*height-for-age*), serta BB/TB (*weight-for-height*). Selain itu, dilakukan standarisasi format jenis kelamin dan normalisasi kategori status gizi agar konsisten. Tahapan ini bertujuan untuk mengurangi *noise*, mencegah duplikasi label, serta meningkatkan akurasi model klasifikasi.

2.4 Transformasi Label

Pada tahap transformasi data, kategori status gizi bayi yang diperoleh dari indikator berat badan terhadap tinggi badan (BB/TB) pada awalnya memiliki beberapa variasi label yang cukup beragam. Variasi label tersebut dapat menyebabkan meningkatnya kompleksitas dalam proses klasifikasi karena model harus membedakan banyak kategori yang memiliki karakteristik yang relatif mirip. Untuk mengatasi permasalahan tersebut, dalam penelitian ini dilakukan proses penyederhanaan kategori dengan mengelompokkan label status gizi ke dalam tiga kelas utama, yaitu NORMAL, GIZI BERMASALAH, dan OVERWEIGHT. Proses pengelompokan ini dilakukan dengan mempertimbangkan kesamaan karakteristik kondisi gizi pada masing-masing kategori sehingga tetap mampu merepresentasikan kondisi kesehatan bayi secara akurat. Penyederhanaan kelas juga membantu model dalam mempelajari pola hubungan antara atribut data dengan kategori status gizi secara lebih jelas dan terstruktur.

Selain untuk mengurangi kompleksitas klasifikasi, proses transformasi kategori ini juga bertujuan untuk mengatasi permasalahan ketidakseimbangan kelas (*class imbalance*) yang sering terjadi pada dataset kesehatan. Dalam banyak kasus, jumlah data pada kategori tertentu seperti kondisi gizi normal cenderung jauh lebih banyak dibandingkan dengan kategori lainnya. Ketidakseimbangan distribusi data tersebut dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas sehingga performa prediksi terhadap kelas minoritas menjadi kurang optimal. Dengan melakukan pengelompokan label menjadi tiga kategori utama, distribusi data menjadi lebih seimbang dan proses pembelajaran model dapat berlangsung dengan lebih efektif. Hasil dari tahap transformasi ini menghasilkan dataset yang lebih sederhana, terstruktur, serta siap digunakan pada tahap pengolahan data selanjutnya dalam proses pembangunan model klasifikasi.

2.5 Encoding Data

Pada tahap *encoding*, dilakukan konversi terhadap variabel penelitian yang masih berupa data kategorikal. Jenis data ini umumnya ditampilkan dalam bentuk teks atau label, sehingga tidak dapat langsung diproses oleh algoritma machine learning yang membutuhkan input numerik. Karena itu, penelitian ini menerapkan teknik *label encoding* untuk mengubah setiap kategori menjadi bentuk angka yang dapat dikenali oleh model klasifikasi. Langkah ini penting karena memungkinkan algoritma Gaussian Naïve Bayes melakukan perhitungan berbasis probabilitas secara matematis menggunakan nilai numerik hasil proses encoding. Seluruh variabel kategorikal dalam dataset diolah dengan cara yang teratur dan seragam agar tidak terjadi perbedaan makna antar data. Keteraturan dalam proses transformasi ini diperlukan untuk mempertahankan hubungan antar variabel, sehingga meskipun data telah dikonversi ke dalam bentuk angka, informasi dari setiap kategori tetap terjaga. Dengan demikian, data menjadi lebih siap dan optimal untuk digunakan dalam proses pelatihan model pada tahap berikutnya.

2.6 Pembagian Data

Jika dilihat setelah pada tahap transformasi dan encoding selesai dilakukan, dataset yang telah diproses kemudian dipisahkan menjadi dua kelompok utama, yaitu data pelatihan (*training set*) dan data pengujian (*testing set*). Pada penelitian ini, komposisi yang digunakan adalah 80% untuk data pelatihan dan 20% untuk data pengujian. Tujuan dari pembagian ini adalah agar model dapat belajar dari pola yang terdapat pada data pelatihan, kemudian digunakan untuk menguji kemampuannya dalam memprediksi data baru yang belum pernah diproses sebelumnya. Dengan pemisahan ini, penilaian performa model dapat dilakukan secara lebih objektif dan menghasilkan hasil evaluasi yang lebih valid. Pembagian data tersebut dilakukan menggunakan metode *stratified sampling*, yaitu teknik yang mempertahankan proporsi setiap kelas tetap seimbang pada kedua bagian data. Dengan cara ini, setiap kategori status gizi tetap memiliki representasi yang proporsional pada data pelatihan maupun data pengujian. Hal ini penting untuk menghindari ketidakseimbangan data yang dapat menimbulkan bias dalam proses pembelajaran model, terutama ketika terdapat kelas dengan jumlah data yang lebih dominan. Melalui penerapan *stratified sampling*, model yang dibangun diharapkan mampu menangkap karakteristik data dengan lebih tepat, sehingga hasil prediksi menjadi lebih konsisten dan dapat diandalkan.

2.7 Pelatihan Model

Penelitian ini melanjutkan ke tahap pelatihan model klasifikasi dengan memanfaatkan algoritma Gaussian Naïve Bayes. Metode ini bekerja berdasarkan pendekatan probabilistik yang mengacu pada Teorema Bayes, di mana setiap variabel diasumsikan tidak saling memengaruhi satu sama lain. Walaupun asumsi tersebut cukup sederhana, pendekatan ini tetap sering digunakan karena mampu memberikan hasil yang stabil pada berbagai jenis data, termasuk data kesehatan. Selain itu, algoritma ini juga dikenal ringan secara komputasi sehingga proses pelatihan dapat dilakukan dengan cepat.

Pada tahap ini, data latih digunakan untuk membentuk pola distribusi probabilitas dari setiap variabel terhadap kategori hasil yang telah ditentukan. Dari proses tersebut, model belajar mengenali keterkaitan antara atribut seperti berat badan, tinggi badan, dan usia bayi dengan status gizi yang sesuai. Setelah model selesai dibangun, hasil pelatihan

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

tersebut kemudian dimanfaatkan untuk melakukan prediksi terhadap data baru yang belum pernah diproses sebelumnya berdasarkan pola yang telah dipelajari. Dengan demikian, model diharapkan mampu menghasilkan klasifikasi status gizi bayi secara lebih efektif dan konsisten.

2.8 Evaluasi Model

Pada tahap yang terakhir yaitu melakukan sebuah pengujian kinerj pada model klasidikasi yang sudah dibangun. Namun pada pengujian tersebut data yang diuji sebelum dipisahkan dari data yang ada pada pelatihan. Kemudian tujuan dari proses yang dibentuk untuk mengetahui bagaimana model tersebut dapat diberikan hasil prediksi data yang belum dikenali sebelumnya. Kemudian dengan data uji yang sifatnya independen dan pada hasil evaluasi juga terdapat penelitian yang sebenarnya netral dan cerminkan kemampuan sebenarnya dari model untuk klasifikasikan status pada gizi bayi. Kemudian juga

Tahap akhir dalam penelitian ini adalah melakukan pengujian kinerja terhadap model klasifikasi yang telah dibangun. Pengujian ini menggunakan data uji yang sebelumnya telah dipisahkan dari data pelatihan. Tujuan dari proses ini adalah untuk mengetahui sejauh mana model mampu memberikan hasil prediksi pada data yang belum pernah dikenali sebelumnya. Dengan menggunakan data uji yang bersifat independen, hasil evaluasi dapat memberikan penilaian yang lebih netral dan mencerminkan kemampuan sebenarnya dari model dalam mengklasifikasikan status gizi bayi.

Penilaian kinerja model dilakukan dengan menggunakan beberapa metrik klasifikasi, antara lain Accuracy, Precision, Recall, dan F1-Score. Selain itu, penelitian ini juga memanfaatkan Confusion Matrix untuk memberikan gambaran yang lebih rinci mengenai hubungan antara hasil prediksi model dan data aktual pada setiap kelas. Dengan bantuan Confusion Matrix, jumlah prediksi yang tepat maupun yang keliru pada tiap kategori status gizi dapat terlihat dengan jelas, sehingga memudahkan dalam menganalisis kesalahan klasifikasi yang terjadi. Hasil evaluasi tersebut kemudian digunakan sebagai acuan untuk menilai sejauh mana tingkat ketepatan dan kestabilan model dalam melakukan klasifikasi status gizi bayi. Di samping itu, hasil ini juga dapat dijadikan dasar dalam pengembangan model yang lebih optimal pada penelitian selanjutnya.

2.9 Metode Penyelesaian Masalah

Metode *Naïve Bayes* merupakan algoritma klasifikasi berbasis teorema Bayes yang mengasumsikan independensi antar fitur [5] [8] [14]. Probabilitas posterior suatu kelas dihitung berdasarkan probabilitas prior dan likelihood masing-masing fitur [15] [16]. Pendekatan *Gaussian Naïve Bayes* digunakan karena mampu memodelkan distribusi data kontinu dengan asumsi distribusi normal [8] [11]. Evaluasi performa model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score yang umum digunakan dalam klasifikasi data kesehatan [19].

Tabel 2. Struktur Dataset Penelitian

Atribut	Tipe Data	Peran
Jenis Kelamin	Kategorikal	L / P
BB/U	Kategorikal	Status berat badan
TB/U	Kategorikal	Status tinggi badan
BB/TB	Kategorikal	Target Klasifikasi

Berdasarkan Tabel 1, variabel BB/U dan TB/U digunakan sebagai fitur prediktor, sedangkan BB/TB berperan sebagai variabel target dalam proses klasifikasi.

2.10 Transformasi Label

Tabel 3. Transformasi Label Status Gizi

Label Awal	Label Hasil Transformasi
Gizi Baik	NORMAL
Gizi Kurang	GIZI BERMASALAH
Gizi Buruk	GIZI BERMASALAH
Gizi Lebih	NORMAL
Risiko Gizi Lebih	NORMAL
Obesitas	OVERWEIGHT

Tabel 3 menunjukkan proses transformasi label status gizi dari kategori awal menjadi tiga kelas utama yang lebih sederhana untuk analisis, yaitu NORMAL, GIZI BERMASALAH, dan OVERWEIGHT. Kategori Gizi Baik, Gizi Lebih, dan Risiko Gizi Lebih digabungkan menjadi kelas NORMAL karena masih berada dalam rentang kondisi gizi yang tidak bermasalah. Sementara itu, Gizi Kurang dan Gizi Buruk digabungkan menjadi GIZI BERMASALAH karena sama-sama menunjukkan adanya kekurangan asupan gizi yang berisiko terhadap kesehatan anak. Kemudian kategori Obesitas dipertahankan sebagai OVERWEIGHT karena menunjukkan kondisi kelebihan berat badan yang berbeda dari kategori lainnya. Transformasi ini dilakukan untuk menyederhanakan proses klasifikasi sehingga model

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

lebih mudah dilatih, lebih stabil, dan mampu menghasilkan prediksi yang lebih konsisten dalam sistem analisis status gizi.

2.11 Konsep Dasar Naïve Bayes

Algoritma *Naïve Bayes* didasarkan pada Teorema Bayes yang menyatakan bahwa probabilitas posterior suatu kelas dapat dihitung berdasarkan probabilitas prior dan likelihood fitur terhadap kelas tersebut [5]. Secara matematis dirumuskan pada persamaan (1):

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

P(C|X) menyatakan probabilitas posterior, yaitu peluang suatu data termasuk ke dalam kelas C setelah mempertimbangkan fitur X. Nilai P(X|C)P(X|C)P(X|C) merupakan likelihood, yaitu probabilitas kemunculan fitur X pada kelas C, sedangkan P(C)P(C)P(C) adalah probabilitas prior yang menunjukkan peluang awal suatu kelas sebelum mempertimbangkan data. Adapun P(X)P(X)P(X) disebut evidence, yaitu probabilitas keseluruhan dari data yang diamati. Dalam proses klasifikasi, kelas dengan nilai probabilitas posterior tertinggi akan dipilih sebagai hasil prediksi akhir.

2.l1.1 Distribusi Gaussian

Pada pendekatan *Gaussian Naïve Bayes*, setiap fitur diasumsikan mengikuti distribusi normal [6]. Fungsi distribusi Gaussian dirumuskan pada persamaan (2):

$$P(x_i|C) = \frac{1}{\sqrt{2\pi\sigma 2y^2}} exp^{\frac{(x\_i-\mu)y^2}{2\sigma^2}} \quad (2)$$

Pada persamaan tersebut, μ menyatakan nilai rata-rata (mean) dari fitur pada kelas tertentu, σ merupakan standar deviasi yang menggambarkan tingkat penyebaran data, dan x_i adalah nilai fitur ke-i yang akan dihitung probabilitasnya. Pendekatan ini memungkinkan model untuk mengestimasi likelihood berdasarkan parameter statistik yang diperoleh dari data latih.

2.11.2 Tahapan Metode Gaussian Naïve Bayes

Tahapan penerapan metode *Gaussian Naïve Bayes* dalam penelitian ini dilakukan secara sistematis dimulai dengan menghitung probabilitas *prior* untuk setiap kelas berdasarkan distribusi data latih. Selanjutnya, dilakukan perhitungan parameter statistik berupa nilai *mean* (μ) dan standar deviasi (σ) untuk setiap fitur pada masing-masing kelas. Parameter tersebut digunakan untuk menghitung nilai *likelihood* menggunakan Persamaan (2), yang merepresentasikan distribusi Gaussian dari setiap fitur terhadap kelas tertentu. Setelah nilai *likelihood* diperoleh, probabilitas *posterior* dihitung menggunakan Persamaan (1) berdasarkan Teorema Bayes. Tahap akhir dalam proses klasifikasi adalah menentukan kelas dengan nilai probabilitas *posterior* tertinggi sebagai hasil prediksi akhir model.

2.12 Evaluasi Model

Untuk mengukur kinerja model, digunakan metrik evaluasi klasifikasi. *Accuracy, Precision, Recall* dan *F1-Score* dihitung menggunakan Persamaan (3) (4) (5) (6):

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

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

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

$$f1\ score = 2x \frac{PrecisionxRecall}{Precision+Recall} \quad (6)$$

Pada Persamaan (3), TP (True Positive) menyatakan jumlah data yang diklasifikasikan dengan benar sebagai kelas positif, TN (True Negative) adalah jumlah data yang diklasifikasikan dengan benar sebagai kelas negatif, FP (False Positive) menunjukkan jumlah data yang salah diprediksi sebagai kelas positif, sedangkan FN (False Negative) merupakan jumlah data yang salah diprediksi sebagai kelas negatif. Perhitungan ini digunakan untuk mengukur tingkat ketepatan model secara keseluruhan dalam melakukan klasifikasi.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Implementasi Metode

Dataset akhir yang digunakan dalam penelitian ini berjumlah 10.433 data setelah melalui proses pembersihan dan validasi. Dataset dibagi menjadi 80% data latih dan 20% data uji menggunakan metode *stratified sampling* untuk

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

menjaga proporsi distribusi kelas [12]. Model *Gaussian Naïve Bayes* dilatih menggunakan data latih untuk mempelajari parameter probabilistik setiap kelas sesuai pendekatan Bayesian [5], [17].

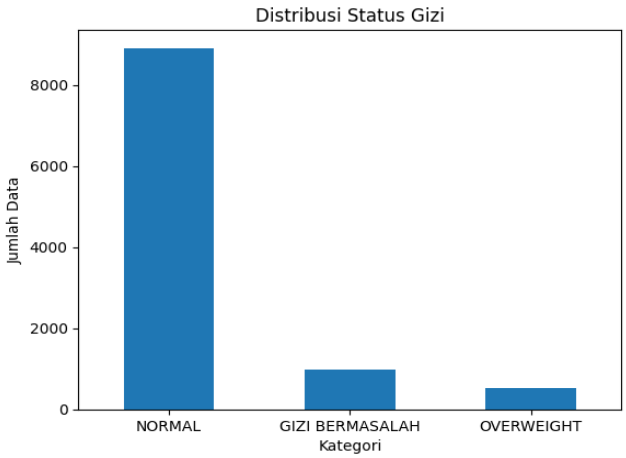
3.1.1 Hasil Preprocessing Data

Tahap awal implementasi adalah memastikan data telah bersih dari nilai kosong serta memiliki konsistensi label. Setelah dilakukan proses data cleaning dan transformasi label, diperoleh dataset akhir dengan tiga kelas utama, yaitu NORMAL, GIZI BERMASALAH, dan OVERWEIGHT.

Tabel 4. Distribusi Akhir Status Gizi

Kategori	Jumlah	Persentase
NORMAL	8.913	85,43%
GIZI BERMASALAH	991	9,50%
OVERWEIGHT	529	5,07%
Total	10.433	100%

Berdasarkan Tabel 4, terlihat bahwa dataset memiliki distribusi yang tidak seimbang (*imbalanced dataset*), di mana kelas NORMAL mendominasi sebesar 85,43% dari total data. Sementara itu, kelas GIZI BERMASALAH dan OVERWEIGHT memiliki proporsi yang jauh lebih kecil. Meskipun distribusi data pada Tabel 4 menunjukkan ketimpangan yang cukup tinggi antara kelas mayoritas (±85%) dan kelas minoritas (±5%), penelitian ini tidak menerapkan teknik penyeimbangan data seperti SMOTE. Hal ini dikarenakan tujuan utama penelitian adalah mengevaluasi performa model Gaussian Naïve Bayes pada kondisi data riil tanpa modifikasi distribusi, sehingga hasil yang diperoleh lebih merepresentasikan kondisi sebenarnya di lapangan. Selain itu, penerapan teknik oversampling seperti SMOTE berpotensi menghasilkan data sintetis yang dapat menyebabkan overfitting, terutama pada algoritma probabilistik seperti Naïve Bayes yang sensitif terhadap distribusi data. Distribusi tersebut divisualisasikan dalam bentuk diagram batang sebagaimana ditunjukkan pada Gambar 1.



Gambar 2. Diagram Batang Distribusi Kelas Status Gizi

Gambar 2 menunjukkan distribusi kelas status gizi bayi dalam bentuk diagram batang, yang memperlihatkan bahwa kelas NORMAL memiliki jumlah data yang jauh lebih dominan dibandingkan kelas lainnya seperti GIZI BERMASALAH dan OVERWEIGHT. Ketimpangan distribusi ini mengindikasikan adanya kondisi class imbalance pada dataset, di mana sebagian besar data terkonsentrasi pada satu kelas mayoritas. Kondisi tersebut berpotensi mempengaruhi performa model klasifikasi, karena model cenderung lebih mudah mempelajari pola pada kelas yang dominan dibandingkan dengan kelas minoritas yang jumlah datanya lebih sedikit. Kemudian distribusi kelas ini menjadi faktor penting yang perlu diperhatikan dalam proses evaluasi model.

3.2 Hasil Pelatihan Model

Dataset dibagi menggunakan metode *stratified sampling* dengan komposisi 80% data latih dan 20% data uji untuk menjaga proporsi distribusi kelas pada masing-masing subset [12]. Dengan total 10.433 data akhir, diperoleh pembagian sebagai berikut:

a. Data latih ≈ 8.346 data
b. Data uji ≈ 2.087 data

Metode *stratified sampling* digunakan untuk memastikan distribusi kelas pada data latih dan data uji tetap proporsional terhadap distribusi dataset awal, sehingga model tidak mengalami bias distribusi saat proses pelatihan maupun pengujian [12]. Model *Gaussian Naïve Bayes* dilatih menggunakan data latih untuk menghitung parameter statistik berupa nilai mean dan standar deviasi pada masing-masing kelas target [5] [14]. Parameter tersebut digunakan untuk membentuk distribusi probabilistik setiap fitur terhadap kelas tertentu berdasarkan asumsi distribusi normal

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

(Gaussian) [8] [11]. Proses pelatihan ini menghasilkan model yang mampu mempelajari pola distribusi fitur BB/U dan TB/U terhadap kelas target BB/TB secara sistematis.

3.3 Hasil Evaluasi Model

Evaluasi model dilakukan menggunakan data uji sebanyak 2.087 data yang tidak terlibat dalam proses pelatihan. Pengujian dilakukan untuk mengetahui kemampuan generalisasi model dalam mengklasifikasikan status gizi bayi pada data baru. Evaluasi performa dilakukan menggunakan beberapa metrik klasifikasi yang umum digunakan dalam penelitian *machine learning* pada bidang kesehatan, yaitu accuracy, precision, recall, dan F1-score [19]. Penggunaan beberapa metrik ini bertujuan untuk memberikan gambaran performa model secara lebih komprehensif, terutama pada dataset dengan distribusi kelas yang tidak seimbang (*imbalanced dataset*) [10], [19].

3.3.1 Accuracy

Berdasarkan hasil pengujian terhadap 2.087 data uji, model memperoleh nilai *accuracy* sebesar 82,27%. Nilai *accuracy* tersebut menunjukkan proporsi prediksi yang benar dibandingkan dengan keseluruhan data uji. Secara matematis, *accuracy* dihitung sebagai perbandingan antara jumlah prediksi benar (*True Positive* dan *True Negative*) dengan total prediksi. Dengan dataset yang bersifat tidak seimbang, nilai accuracy cenderung dipengaruhi oleh dominasi kelas mayoritas, yaitu kelas NORMAL yang mencakup 85,43% dari total dataset. Fenomena ini sesuai dengan temuan pada penelitian sebelumnya yang menyatakan bahwa metrik *accuracy* dapat memberikan gambaran yang terlalu optimistis pada kondisi *imbalanced dataset* [10], [11].

Pada model klasifikasi dengan data yang tidak seimbang, model biasanya lebih sering mengenali kelas yang paling banyak, seperti kategori NORMAL. Akibatnya, nilai accuracy bisa terlihat tinggi karena banyak data memang berasal dari kelas tersebut, tetapi belum tentu model mampu mendeteksi kelas lain seperti GIZI BERMASALAH atau OVERWEIGHT dengan baik. Hal ini menunjukkan bahwa accuracy saja belum cukup untuk menilai kinerja model secara keseluruhan.

Kemudian juga diperlukan metrik lain seperti precision, recall, dan F1-score agar penilaian performa model menjadi lebih lengkap. Metrik-metrik ini membantu melihat kemampuan model dalam mengklasifikasikan setiap kelas secara lebih detail, terutama pada kelas yang jumlah datanya lebih sedikit.

Pada penelitian ini, nilai accuracy menunjukkan bahwa model cukup baik dalam mengenali kelas mayoritas. Namun, agar hasilnya lebih meyakinkan dan seimbang, tetap diperlukan evaluasi menggunakan metrik lainnya. Kemudian dilihat dapat diketahui apakah model benar-benar mampu mengklasifikasikan semua kategori status gizi bayi dengan baik, bukan hanya pada satu kelas saja.

3.3.2 Precision, Recall, dan F1-Score

Tabel 5. Hasil Evaluasi Model per Kelas

Kelas	Precision	Recall	F1-Score
NORMAL	0.92	0.89	1783
GIZI BERMASALAH	0.39	0.63	198
OVERWEIGHT	0.14	0.08	106

Precision adalah ukuran yang digunakan untuk menilai ketepatan prediksi positif dari sebuah model. Nilai ini menunjukkan seberapa banyak hasil yang diprediksi sebagai suatu kelas benar-benar merupakan kondisi yang sesuai dengan kenyataan [11].

Berdasarkan hasil evaluasi, nilai precision untuk kelas NORMAL relatif tinggi dibandingkan kelas GIZI BERMASALAH dan OVERWEIGHT. Hal ini menunjukkan bahwa model cukup akurat dalam mengidentifikasi kelas mayoritas. Namun, untuk kelas minoritas, nilai precision cenderung lebih rendah akibat jumlah data yang lebih sedikit dan kemungkinan terjadinya kesalahan klasifikasi. Kondisi ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa model klasifikasi pada dataset kesehatan dengan distribusi tidak seimbang sering mengalami penurunan performa pada kelas minoritas [20]. Oleh karena itu, interpretasi precision harus dilakukan dengan mempertimbangkan distribusi data yang ada.

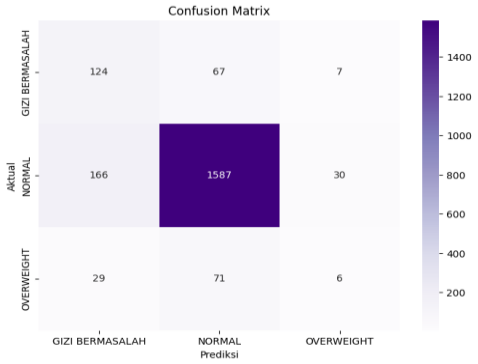
Hasil evaluasi menunjukkan bahwa performa model cenderung lebih baik pada kelas mayoritas (NORMAL) dibandingkan kelas minoritas (GIZI BERMASALAH dan OVERWEIGHT). Nilai recall yang lebih tinggi pada kelas NORMAL mengindikasikan bahwa model memiliki sensitivitas yang lebih baik dalam mengenali data pada kelas tersebut, sementara pada kelas minoritas masih terdapat sejumlah data yang tidak terdeteksi dengan baik.

Selain itu, nilai precision yang tinggi pada kelas NORMAL menunjukkan bahwa sebagian besar prediksi pada kelas ini tergolong tepat, sedangkan precision pada kelas minoritas relatif lebih rendah, yang menandakan masih adanya kesalahan klasifikasi. Kondisi ini dipengaruhi oleh ketidakseimbangan distribusi data, di mana kelas NORMAL mendominasi dataset sehingga model lebih optimal dalam mempelajari pola pada kelas tersebut.

Nilai F1-score secara keseluruhan menunjukkan bahwa model memiliki performa yang cukup baik, namun masih terdapat ketidakseimbangan kinerja antar kelas. Hal ini mengindikasikan bahwa model masih perlu ditingkatkan, terutama dalam meningkatkan kemampuan deteksi pada kelas minoritas agar hasil klasifikasi menjadi lebih akurat dan merata.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

3.3.3 Confusion Matrix



Gambar 3. Confusion Matrix Gaussian Naïve Bayes

Pada Confusion Matrix di Gambar 3 terlihat bagaimana perbandingan dari label aktual dan label prediksi pada 2.087 data yang telah diuji. Pada elemen diagonal utama matriks yang tunjukkan bagainana sejumlah prediksi yang benar pada masing-masing kelas. Namun pada elemen di luar diagonal menunjukkan adanya kesalahan pada tahap klasifikas [11].

Berdasarkan hasil pengujian, sebagian besar prediksi benar terkonsentrasi pada kelas NORMAL, yang sejalan dengan distribusi dataset yang didominasi oleh kelas tersebut. Kesalahan klasifikasi paling banyak terjadi pada kelas GIZI BERMASALAH yang diprediksi sebagai NORMAL. Hal tersebut terlihat bagaimana adanya indikasi kemiripan karakteristik yang fitur antara kedua kelas yang dilakuakan. Sehingga mendapatkan hasil bahwa model yang alami kesulitan bisa bedakan batas dari distribusinya. Konteks seperti ini sebenarnya sering terjadi pada dataset dengan distribusi kelas yang tidak seimbang [20] [19].

Selain itu, pada kelas OVERWEIGHT, beberapa data juga terklasifikasi sebagai NORMAL. Kondisi ini menunjukkan bahwa fitur antropometri yang digunakan mungkin memiliki tumpang tindih nilai pada beberapa kategori, sehingga asumsi independensi fitur dalam algoritma *Naïve Bayes* menjadi salah satu faktor yang mempengaruhi hasil klasifikasi [5], [8], [17].

Pada analisis confusion juga memberikan informasi bahwa mode ini memang punya performa yang baik jika dilihat dari accuracy. Namun masih perlu perbaikan dari segi sensitivitas terhadap kelas minoritas. Maka dari itu pengembannya akan dilihat lebih lanjut dengan menambahkan fitur dan numerika tambahan. Kemudian juga bisa menerapkan bagaimana teknik penyeimbangan data untuk tingkatkan kemampuan deteksi dari kelas risiko. Seperti yang direkomendasikan pada penelitian klasifikasi kesehatan yang telah dilakukan sebelumnya [20] [19].

4. KESIMPULAN

Dalam penelitian yang dilakukan, memberikan implementasi dari medoe Gaussian Naive Bayes sebagai klasifikasi kondisi pada gizi bayi. Kemudian berdasarkan dari indikator antropometri yang menghasilkan evaluasi performa yang cukup baik dengan akurasi 82,7%. Didukung oleh nilai precision, recall dan F1-score yang tunjukan kemampuan pada model saat identfikasi pola data secara efektif. Tetapi model ini juga masih memperlihatkan keterbatasan pada klasifikasi kelas minoritas akibat tidak seimbangnya distribusi data, maka dari itu performa pada kategi tertentu tidak optimal. Metode Gaussian Naïve Bayes terlihat cukup baik dan efisien sebagai langkah awal dalam membangun sistem pendukung keputusan di bidang kesehatan, khususnya untuk menganalisis kondisi gizi bayi. Metode ini memudahkan pengolahan data antropometri seperti BB/U dan TB/U sehingga proses penentuan status gizi dapat dilakukan lebih cepat dan lebih sederhana dibandingkan cara manual. Kontribusi penelitian ini menunjukkan bahwa algoritma Naïve Bayes dapat dimanfaatkan untuk mengklasifikasikan status gizi bayi berdasarkan data antropometri (BB/U dan TB/U) secara efektif dan efisien. Kemudian yang diberikan juga memiliki tingkat akurasi yang cukup baik dengan proses yang sederhana.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: M.R.AR., F, K.D.T., dan A.I.;
Metodologi: M.R.AR., F., A.M., dan A.I.;
Perangkat Lunak: M.R.AR., dan K.D.T.;
Validasi: M.R.AR., F., K.D.T., A.M., dan A.I.;
Analisis Formal: M.R.AR., dan A.M.;
Investigasi: M.R.AR., F., K.D.T., dan A.I.;
Sumber Daya: M.R.AR., F., K.D.T., A.M., dan A.I.;

Kurasi Data: M.R.AR., dan F.;
Penulisan Draf Awal: M.R.AR., dan A.M.;
Penulisan, Peninjauan, dan Penyuntingan: M.R.AR., F., K.D.T., dan A.I.;
Visualisasi: M.R.AR., F., K.D.T., A.M., dan A.I.;
Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

5.2 Ketersediaan Data

Data yang disajikan dalam penelitian ini tersedia berdasarkan permintaan kepada penulis pertama.

5.3 Pendanaan

Penelitian ini tidak menerima pendanaan eksternal.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (Not applicable).

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku (Not applicable).

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Para penulis menggunakan AI Generatif untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

REFERENCES

[1] World Health Organization, “Levels and trends in child malnutrition,” World Health Organization (WHO), 2021. [Online]. Available: https://www.who.int/publications/i/item/9789240025257
[2] UNICEF, “Nutrition, For Every Child Global Annual Results Report 2022,” *UNICEF: For Every Child*, 2023, [Online]. Available: https://www.unicef.org/documents/nutrition-annual-results-2022
[3] H. A. Santoso, N. S. Dewi, S. C. Haw, A. D. Pambudi, and S. A. Wulandari, “Enhancing nutritional status prediction through attention-based deep learning and explainable AI,” *Intelligence-Based Medicine*, vol. 11, p. 100255, 2025, doi: 10.1016/j.ibmed.2025.100255.
[4] M. L. Arqam, A. A. Firdaus, A. M. Atmojo, and G. Z. Saputri, “Integrating education-based interventions and machine learning for stunting prevention: A case study in East Lombok, Indonesia,” *Dialogues in Health.*, vol. 8, no. 6 , p. 100264, 2026, doi: 10.1016/j.dialog.2025.100264.
[5] W. Widhari, A. Triayudi, and R. T. K. Sari, “Implementation of Naïve Bayes and K-NN Algorithms in Diagnosing Stunting in Children,” *SAGA: Journal of Technology and Information System.*, vol. 2, no. 1, pp. 164–174, 2024, doi: 10.58905/saga.v2i1.242.
[6] J. I. Molina and E. J. Malaikosa, “Application of the Naïve Bayes Algorithm and Radipminer Application to Determine the Nutritional Status of Toddlers,” *International Journal of Advanced Technology and Social Sciences.*, vol. 2, no. 3, pp. 389–402, 2024, doi: 10.59890/ijatss.v2i3.1565.
[7] E. Miranda, M. Aryuni, A. Y. Zakiyyah, Y. E. Kurniawati, A. V. D. Sano, and M. Kumbangsila, “An early prediction model for toddler nutrition based on machine learning from imbalanced data,” *Procedia Computer Science.*, vol. 245, pp. 263–271, 2024, doi: 10.1016/j.procs.2024.10.251.
[8] M. I. Abas, R. Lamusu, W.E. Pranata, I. Ibrahim, W. Hasyim, and V. Kiayi, “Penerapan Algoritma Naive Bayes Untuk Sistem Klasifikasi Status Gizi Bayi Balita,” *Bulletin of Computer Science Research.*, vol. 5, no. 5, pp. 1249–1260, 2025, doi: 10.47065/bulletincsr.v5i5.508.
[9] T. E. Putri, R. T. Subagio, Kusnadi, and P. Sobiki, “Classification System of Toddler Nutrition Status using Naïve Bayes Classifier Based on Z- Score Value and Anthropometry Index,” *Journal of Physics: Conference Series.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012005.
[10] Y. Saputra, N. Khoirunisa, and S. Arinal Haqq, “Classification of Health and Nutritional Status of Toddlers Using the Naïve Bayes Classification,” *CoreID Journal.*, vol. 1, no. 2, pp. 49–57, 2023, doi: 10.60005/coreid.v1i2.8.
[11] R. A. Subhan, Asriyanik, and F. F. Az Zahra, “Klasifikasi Status Gizi Balita Berdasarkan Tinggi Badan Dan Berat Badan Menggunakan Metode Naive Bayes Di Puskesmas Limusnunggal,” *SYNERGY: Jurnal Ilmiah Multidisiplin.*, vol. 2, no. 1, pp. 51–61, 2024, doi: https://doi.org/10.59585/jimad.
[12] R. Hariyanto, M. Z. Sarwani, and Y. N. Aprilia, “Prediction of Stunting Nutritional Status in Toddlers Using Naïve Bayes Classifier Algorithm,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 2, pp. 1111–1117, 2025, doi: 10.29100/jipi.v10i2.5930.
[13] T. Mabukela, P. K. Chelule, and P. Modjadji, “Nutrition and Development of Children in Foundational Learning Spaces in Johannesburg: A Cross-Sectional Study of Dietary Diversity and Nutritional Status,” *Applied Sciences (Switzerland).*, vol. 15, no. 23, pp. 1–18, 2025, doi: 10.3390/app152312385.
[14] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*. Mexico: Litografica Ingramex, 2024. [Online].

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 799−809
DOI 10.47065/bits.v8i2.9490
https://ejurnal.seminar-id.com/index.php/bits

Available: https://books.google.co.id/books?id=vfDiEAAAQBAJ&printsec=copyright&hl=id#v=onepage&q&f=false
[15] G. James, D. Witten, T. Hastie, and O. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, 2nd ed. Springer Nature, 2021.
[16] O. Simeone, *A Brief Introduction to Machine Learning for Engineers*. Foundations and Trends, 2018. [Online]. Available: http://arxiv.org/abs/1709.02840
[17] I. Dewancker, M. McCourt, and S. Clark, "Bayesian Optimization for Machine Learning : A Practical Guidebook," pp. 1–15, 2016, [Online]. Available: http://arxiv.org/abs/1612.04858
[18] T. Herlambang and V. Asy'ari, "Comparison Of Naive Bayes And K-Nearest Neighbor Models For Identifying The Highest Prevalence Of Stunting Cases In East Java," *Barekeng Jurnal Ilmu Matematika dan Terapan.*, vol. 18, no. 4, pp. 2153–2164, 2024, doi: 10.30598/barekengvol18iss4pp2153-2164.
[19] A. Rozaq and A. J. Purnomo, "Classification of Stunting Status in Toddlers Using Naive Bayes Method in the City of Madiun Based on Website," *Jurnal Techno Nusa Mandiri*, vol. 19, no. 2, pp. 69–76, 2022, doi: 10.33480/techno.v19i2.3337.
[20] M. I. Kamil and A. P. Wibowo, "Implementation of the Naive Bayes Classifier Algorithm for Classifying Toddler Nutritional Status," *Journal of Applied Informatics and Computing.*, vol. 8, no. 2, pp. 421–427, 2024, doi: 10.30871/jaic.v8i2.8669.