

Pembuatan Profil Status Gizi Balita Berbasis Pembelajaran Mesin Menggunakan Indikator Antropometri

Machine Learning-Based Nutritional Status Profiling of Children Under Five Using Anthropometric Indicators

Dianda Rifaldi^{1*}, Galih Pramuja Inngam Fanani², Alya Masitha³, Setiawan Ardi Wijaya⁴, Irwandi Rizki Putra⁵

¹ Informatika, Fakultas Teknologi Informasi, Universitas Riau Indonesia, Rengat, Indonesia

^{2,3} Sistem dan Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas 'Aisyiyah Surakarta, Indonesia

⁴ Sistem Informasi, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau, Indonesia

⁵ Sistem Informasi, Fakultas Teknologi Informasi, Universitas Riau Indonesia, Rengat, Indonesia

*E-mail: diandarifaldi@gmail.com

fanani@unisa-surakarta.ac.id; alya.masitha@unisa-surakarta.ac.id; setiawanardiwijaya@umri.ac.id; irwandirizkiputra2@gmail.com

Article History

Submitted : July 01, 2026
Revised : July 20, 2026
Accepted : July 28, 2026
Available Online : July 31, 2026
Published Regularly : July 31, 2026

Kata Kunci: Antropometri, Balita, Pembelajaran Mesin, Profil Status Gizi

Keywords: Anthropometry, Toddlers, Machine Learning, Nutritional Status Profile

Contact



Author

diandarifaldi@gmail.com

ABSTRAK

Stunting dan permasalahan gizi balita masih menjadi tantangan utama kesehatan masyarakat di Indonesia yang memerlukan pemantauan secara presisi dan berbasis data. Penelitian ini bertujuan untuk membangun profil status gizi balita menggunakan pendekatan pembelajaran mesin berbasis indikator antropometri. Metodologi diawali dengan audit kualitas data dan eksplorasi data analitis Exploratory Data Analysis terhadap dataset yang mencakup 120.999 observasi balita tanpa nilai yang hilang (missing value). Dataset yang terdiri dari atribut umur, tinggi badan, dan jenis kelamin ini dibagi secara stratified (80:20), kemudian distandardisasi menggunakan StandardScaler dan OneHotEncoder. Empat algoritma klasifikasi (Logistic Regression, Decision Tree, Gradient Boosting, dan Random Forest) dievaluasi secara komparatif. Hasil eksperimen menunjukkan bahwa Random Forest mencatatkan performa terbaik dengan tingkat akurasi sebesar 98,94%, Precision makro 98,74%, Recall makro 98,45%, dan F1-Score makro 98,59%. Keandalan dan konsistensi model ini dikonfirmasi melalui 5-Fold Stratified Cross-Validation yang menghasilkan rata-rata akurasi sebesar 99,04% dengan standar deviasi yang sangat rendah (0,0009). Selanjutnya, analisis feature importance menghasilkan tinggi badan 60% dan umur 39% merupakan indikator antropometri yang paling dominan dalam menentukan profil status gizi balita. Penelitian ini membuktikan bahwa algoritma Random Forest berbasis indikator antropometri layak diterapkan sebagai sistem pendukung keputusan dalam pemantauan pertumbuhan dan deteksi dini masalah gizi anak.

ABSTRACT

Stunting and nutritional issues among children under five remain major public health challenges in Indonesia, requiring precise, data-driven monitoring. This study aims to establish nutritional status profiles for children under five using a machine learning approach based on anthropometric indicators. The methodology began with a data quality audit and Exploratory Data Analysis (EDA) on a dataset comprising 120,999 observations with no missing values. The dataset consisting of age, height, and gender attributes was split using a stratified approach (80:20) and subsequently standardized using StandardScaler and OneHotEncoder. Four classification algorithms (Logistic Regression, Decision Tree, Gradient Boosting, and Random Forest) were evaluated comparatively. Experimental results demonstrated that Random Forest achieved the best performance, with an accuracy of 98.94%, a macro-averaged Precision of 98.74%, a macro-averaged Recall of 98.45%, and a macro-averaged F1-Score of 98.59%. The model's reliability and consistency were confirmed via 5-Fold Stratified Cross-Validation, yielding an average accuracy of 99.04% with a very low standard deviation (0.0009). Furthermore, feature importance analysis revealed that height (60%) and age (39%) were the most dominant anthropometric indicators in determining the nutritional status profiles of children under five. This study demonstrates that the Random Forest algorithm, based on anthropometric indicators, is suitable for implementation as a decision support system for growth monitoring and the early detection of nutritional issues in children.

1. Pendahuluan

Stunting merupakan salah satu permasalahan gizi kronis yang masih menjadi tantangan dalam pembangunan kesehatan masyarakat, khususnya pada kelompok anak usia balita [1]. Kondisi ini ditandai dengan tinggi atau panjang badan anak yang lebih rendah dibandingkan standar pertumbuhan anak seusianya akibat dari kekurangan gizi yang berlangsung dalam jangka waktu lama [2]. Stunting tidak hanya berkaitan dengan kondisi fisik anak, tetapi juga memberikan dampak terhadap perkembangan kognitif, kemampuan belajar, produktivitas, serta kualitas kesehatan anak dalam jangka panjang [3].

Permasalahan stunting banyak ditemukan di negara-negara berkembang, termasuk Indonesia yang memiliki karakteristik permasalahan gizi yang cukup kompleks [4]. Terjadinya stunting dapat dipengaruhi oleh berbagai faktor, seperti: asupan makanan yang tidak mencukupi kebutuhan gizi, infeksi berulang, kondisi kesehatan ibu selama kehamilan, pola asuh, sanitasi dan akses terhadap air bersih, serta kondisi sosial dan ekonomi keluarga [5].

Salah satu periode yang sangat menentukan dalam pencegahan stunting adalah 1.000 hari pertama kehidupan, yaitu sejak masa kehamilan hingga anak berusia dua tahun [6]. Pada periode tersebut, kebutuhan zat gizi sangat tinggi untuk mendukung pertumbuhan dan perkembangan organ tubuh, sehingga kekurangan gizi yang terjadi secara kronis dapat meningkatkan risiko terjadinya gangguan pertumbuhan [7]. Oleh karena itu, upaya pencegahan dan penanganan stunting perlu dilakukan secara melalui pemenuhan kebutuhan gizi, peningkatan kualitas pelayanan kesehatan ibu dan anak, pemantauan pertumbuhan secara berkala, serta peningkatan pengetahuan keluarga mengenai praktik pemberian makanan dan pengasuhan yang sesuai [8].

Pertumbuhan dan status gizi balita merupakan salah satu indikator penting dalam menggambarkan kondisi kesehatan dan kualitas perkembangan anak [9]. Masa balita merupakan periode yang sangat penting karena terjadi proses pertumbuhan fisik dan perkembangan yang

berlangsung secara cepat, sehingga kebutuhan energi, protein, vitamin, mineral dan berbagai zat gizi lainnya harus terpenuhi secara optimal [5].

Pemenuhan kebutuhan gizi yang baik dapat mendukung pertambahan berat badan, peningkatan tinggi badan, perkembangan otak, pembentukan sistem kekebalan tubuh, serta kemampuan anak dalam melakukan aktivitas sehari-hari [10]. Sebaliknya, ketidakseimbangan antara kebutuhan dan asupan zat gizi dapat menyebabkan berbagai permasalahan gizi, baik berupa kekurangan gizi maupun kelebihan gizi yang dapat menghambat proses pertumbuhan dan perkembangan anak [11]. Oleh sebab itu, pemantauan pertumbuhan balita perlu dilakukan secara rutin dengan memperhatikan beberapa indikator antropometri, seperti berat badan menurut umur, panjang atau tinggi badan menurut umur, serta berat badan menurut panjang atau tinggi badan.

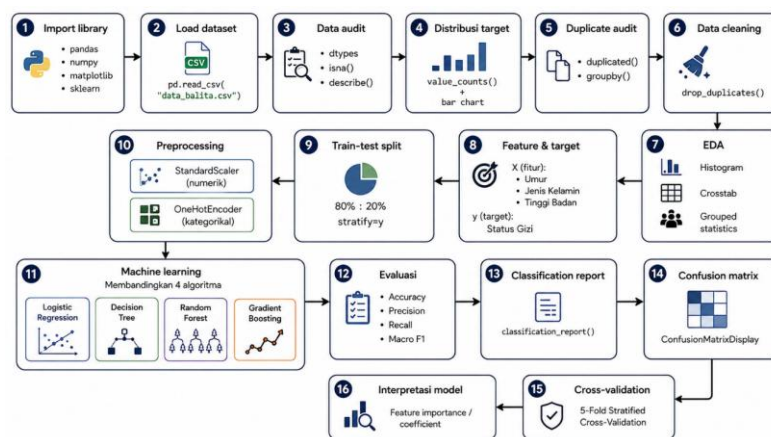
Hasil pemantauan tersebut dapat digunakan untuk mengidentifikasi lebih awal adanya penyimpangan pertumbuhan sehingga intervensi dapat diberikan sebelum kondisi berkembang menjadi masalah gizi yang lebih serius [12]. Selain faktor asupan makanan, pertumbuhan gizi balita juga dipengaruhi oleh pola pemberian makan, riwayat penyakit infeksi, kondisi lingkungan, sanitasi, pola pengasuhan, serta kondisi sosial ekonomi keluarga [13]. Dengan demikian, pemantauan pertumbuhan dan pemenuhan gizi balita secara berkelanjutan menjadi bagian penting dalam upaya menjaga kesehatan anak sekaligus mencegah terjadinya stunting dan permasalahan gizi lainnya.

Pembelajaran mesin dapat menjadi salah satu pendekatan yang potensial dalam mengidentifikasi pola status gizi berdasarkan karakteristik pertumbuhan anak khususnya kebutuhan terhadap pemantauan status gizi balita yang lebih sistematis dan berbasis data dengan pemanfaatan teknologi informasi [14]. Indikator antropometri seperti umur, jenis kelamin, berat badan dan tinggi badan merupakan variabel penting yang dapat digunakan untuk menggambarkan kondisi pertumbuhan serta mengidentifikasi status gizi balita [15].

Pada penelitian ini, pembelajaran mesin diterapkan untuk membangun profil status gizi balita berbasis indikator antropometri dengan membandingkan empat algoritma klasifikasi, yaitu Logistic Regression, Decision Tree, Random Forest dan Gradient Boosting. Keempat algoritma tersebut dipilih karena memiliki karakteristik pendekatan yang berbeda dalam membangun model klasifikasi, sehingga memungkinkan dilakukan evaluasi komparatif untuk mengetahui algoritma yang paling sesuai dalam mengidentifikasi status gizi balita. Dengan pendekatan tersebut, penelitian ini diharapkan tidak hanya menghasilkan model klasifikasi dengan performa yang baik, tetapi juga memberikan profil dan pemahaman yang lebih komprehensif mengenai pola status gizi balita berdasarkan indikator antropometri, sehingga dapat menjadi dasar pendukung dalam proses pemantauan pertumbuhan dan identifikasi awal permasalahan gizi pada anak.

2. Metode Penelitian

Penelitian ini mengusulkan pendekatan untuk pembuatan profil status gizi balita berbasis pembelajaran mesin menggunakan indikator antropometri sebagai upaya untuk mengidentifikasi dan menggambarkan pola status gizi berdasarkan karakteristik pertumbuhan anak. Pendekatan yang dikembangkan memanfaatkan data antropometri dan karakteristik dasar balita sebagai masukan dalam proses analisis dan pemodelan, sehingga dapat diperoleh gambaran mengenai distribusi serta klasifikasi status gizi secara lebih sistematis dan berbasis data untuk memperoleh hasil analisis yang komprehensif. Secara keseluruhan, metodologi yang digunakan dalam penelitian ini digambarkan pada Gambar 1 sebagai alur penelitian yang menunjukkan tahapan mulai dari data masukan hingga proses pemodelan dan evaluasi hasil.



Gambar 1. Kerangka Kerja Penelitian

Gambar 1 menunjukkan kerangka kerja penelitian yang terdiri atas tahapan persiapan data, eksplorasi dan pembersihan data, pembentukan fitur, pemodelan, evaluasi, dan interpretasi. Tahap persiapan meliputi *import library* dan *load dataset*, kemudian dilanjutkan dengan data audit untuk memeriksa struktur, tipe data, nilai hilang, dan statistik deskriptif. Selanjutnya dilakukan analisis distribusi target, pemeriksaan duplikasi dan konflik label, serta *data cleaning* untuk memperoleh dataset yang lebih valid untuk dianalisis. Tahap berikutnya adalah Exploratory Data Analysis (EDA) untuk memahami pola data berdasarkan umur, jenis kelamin, tinggi badan, dan status gizi. Selanjutnya ditentukan Umur, Jenis Kelamin dan Tinggi Badan sebagai fitur, sedangkan Status Gizi sebagai target. Dataset kemudian dibagi menjadi data training dan testing dengan proporsi 80:20 secara stratified, dilanjutkan preprocessing menggunakan StandardScaler untuk fitur numerik dan OneHotEncoder untuk fitur kategorikal. Setelah itu, dibangun empat model klasifikasi, yaitu Logistic Regression, Decision Tree, Random Forest, dan Gradient Boosting. Tahap akhir meliputi evaluasi dan interpretasi model menggunakan Accuracy, Precision, Recall, dan Macro F1-score, serta *classification report* untuk melihat performa pada setiap kelas. Model terbaik dianalisis menggunakan confusion matrix untuk mengidentifikasi pola kesalahan klasifikasi dan 5-Fold Stratified Cross-Validation untuk menguji kestabilan performanya. Selanjutnya dilakukan interpretasi model melalui feature importance atau *coefficient* untuk mengetahui kontribusi fitur terhadap klasifikasi status gizi balita.

2.1. Dataset Stunting

Dataset yang digunakan dalam penelitian ini bersumber dari kaggle, dataset "Stunting Toddler (Balita) Detection" merupakan kumpulan data berdasarkan rumus z-score penentuan stunting menurut WHO (*World Health Organization*), yang berfokus pada deteksi stunting pada balita. Dataset ini terdiri dari 121.000 baris data yang merinci informasi mengenai umur, jenis kelamin, tinggi badan, dan status gizi balita. Fokus pemodelan dibatasi pada profil status gizi berbasis pertumbuhan panjang atau tinggi badan.

2.2. Data Quality Audit

Data quality audit dilakukan sebelum tahapan preprocessing data untuk menilai kelengkapan data dan validitas data [16]. Audit ini mencakup identifikasi nilai yang hilang, data duplikat, nilai yang tidak valid atau berada di luar rentang yang wajar, representasi kategori yang tidak konsisten, serta potensi outliers. Statistik deskriptif untuk mengidentifikasi observasi yang tidak wajar dan menilai distribusi variabel target. Setiap masalah kualitas data yang teridentifikasi didokumentasikan beserta perbaikan yang sesuai. Rekaman yang memuat observasi tidak valid atau duplikat dihapus jika tidak dapat diperbaiki, sedangkan nilai numerik yang hilang diatasi menggunakan strategi imputasi.

2.3. Data Preprocessing

Prosedur lainnya adalah Preprocessing data, yang dilakukan untuk mempersiapkan dataset

bagi pemodelan pembelajaran mesin [17]. Missing values pada atribut numerik ditangani menggunakan imputasi median, sedangkan variabel kategorikal diubah menjadi numerik melalui **label encoding** [18]. Atribut pengenalan yang tidak memberikan informasi prediktif dikeluarkan dari proses pemodelan. Fitur numerik kemudian distandardisasi menggunakan **StandardScaler** untuk memastikan bahwa variabel dengan skala pengukuran berbeda memberikan kontribusi yang sebanding terhadap proses pembelajaran [19]. Dataset dibagi menjadi subset pelatihan dan pengujian. Dataset pelatihan dan pengujian yang dihasilkan kemudian digunakan untuk pengembangan model serta evaluasi kinerja.

2.4. Exploratory Data Analysis

EDA dilakukan untuk memperoleh pemahaman awal mengenai distribusi, karakteristik, dan hubungan antarvariabel dalam dataset [20]. Statistik deskriptif dihitung untuk variabel numerik, yang mencakup nilai rata-rata, median, simpangan baku, serta nilai minimum dan maksimum [21]. Distribusi variabel kategorikal dan variabel target untuk mengidentifikasi potensi ketidakseimbangan kelas. Teknik visualisasi seperti histogram, boxplot, diagram batang, dan heatmap. Penelitian mengevaluasi karakteristik distribusi variabel antropometri yang digunakan sebagai prediktor, yaitu umur dan tinggi badan balita. Distribusi umur (bulan) divisualisasikan menggunakan histogram dengan 30 *bins* untuk mengidentifikasi rentang usia dan konsentrasi observasi pada dataset, sedangkan distribusi tinggi badan (cm) divisualisasikan menggunakan histogram dengan 40 *bins* untuk mengamati pola penyebaran ukuran tinggi badan. Eksplorasi ini penting karena variasi umur dan tinggi badan merupakan karakteristik antropometri yang berkaitan dengan penilaian status pertumbuhan balita. Hasil visualisasi digunakan sebagai pemeriksaan awal terhadap karakteristik distribusi data dan untuk mengidentifikasi kemungkinan pola atau nilai yang tidak umum sebelum data digunakan dalam proses *preprocessing* dan klasifikasi menggunakan algoritma *machine learning*.

2.5. Machine Learning Models

Penelitian ini membandingkan empat algoritma *machine learning* untuk melakukan klasifikasi status gizi balita, yaitu Logistic Regression, Decision Tree, Random Forest, dan Gradient Boosting [22]. Keempat algoritma dipilih untuk merepresentasikan karakteristik pendekatan pemodelan yang berbeda, mulai dari model linear hingga algoritma berbasis *decision tree* dan *ensemble learning*. Logistic Regression digunakan sebagai model baseline untuk mengevaluasi kemampuan pendekatan linear dalam memetakan hubungan antara fitur antropometri dan kategori status gizi. Decision Tree digunakan untuk menangkap pola non-linear melalui pembentukan aturan keputusan berdasarkan fitur input. Selanjutnya, Random Forest digunakan sebagai pendekatan *bagging ensemble* yang menggabungkan sejumlah pohon keputusan untuk menghasilkan prediksi yang lebih robust, sedangkan Gradient Boosting digunakan sebagai pendekatan *boosting* yang membangun model secara bertahap dengan memperbaiki kesalahan prediksi dari model sebelumnya. Seluruh model dilatih menggunakan fitur Umur, Jenis Kelamin dan Tinggi Badan dengan Status Gizi sebagai variabel target. Performa masing-masing model kemudian dibandingkan menggunakan Accuracy, Precision, Recall, dan Macro F1-score, dilanjutkan dengan *classification report*, *confusion matrix*, dan 5-Fold Stratified Cross-Validation untuk memperoleh evaluasi performa dan kestabilan model. Pada library scikit-learn versi standar, modul GradientBoostingClassifier tidak memiliki parameter bawaan `class_weight='balanced'` secara langsung seperti *Random Forest* atau *Decision Tree*.

2.5.1. Logistic Regression

Logistic Regression digunakan sebagai salah satu model klasifikasi untuk memprediksi status gizi balita berdasarkan variabel Umur, Jenis Kelamin dan Tinggi Badan [23]. Dataset terlebih dahulu dibagi menjadi data training dan testing dengan proporsi 80:20 menggunakan *stratified sampling* berdasarkan variabel target untuk mempertahankan proporsi masing-masing kelas. Variabel numerik, yaitu umur dan tinggi badan, distandardisasi menggunakan **StandardScaler** dengan persamaan (1)

$$z = (x - \mu)/\sigma \quad (1)$$

sedangkan variabel Jenis Kelamin dikonversi menjadi representasi numerik menggunakan **OneHotEncoder**.

Model Logistic Regression digunakan untuk menghitung probabilitas keanggotaan setiap observasi terhadap masing-masing kategori Status Gizi menggunakan fungsi Softmax, yang secara matematis seperti persamaan (2) dengan persamaan (3)

$$\frac{P(y = k|X) = e^{zk}}{\sum_{j=1}^K e^{zj}} \quad (2)$$

$$zk = \beta_{k0} + \sum_{j=1}^p \beta_{kj} X_j \quad (3)$$

Kelas dengan probabilitas terbesar ditetapkan sebagai hasil prediksi model. Untuk mengatasi kemungkinan ketidakseimbangan distribusi kelas, penelitian menerapkan parameter `class_weight="balanced"` yang memberikan bobot kelas berdasarkan $wk = \frac{N}{(KN_k)}$, sehingga kelas dengan jumlah observasi lebih sedikit memperoleh bobot yang lebih besar dalam proses pembelajaran. Model dikonfigurasi dengan `max_iter=2000` untuk memberikan jumlah iterasi yang memadai dalam proses optimasi, sedangkan performa model selanjutnya dievaluasi menggunakan Accuracy, Precision, Recall, dan Macro F1-score.

2.5.2. Decision Tree

Decision Tree digunakan sebagai salah satu model klasifikasi untuk memprediksi Status Gizi balita berdasarkan variabel Umur, Jenis Kelamin dan Tinggi Badan [24]. Algoritma ini membentuk struktur pohon keputusan dengan melakukan pemisahan data secara rekursif berdasarkan fitur dan nilai *threshold* yang mampu menghasilkan pemisahan kelas secara optimal. DecisionTreeClassifier menggunakan kriteria *Gini impurity* sebagai fungsi pengukuran ketidakmurnian node, dimana pada persamaan (4) menjelaskan pengukuran impurity, sedangkan rumus kedua menjelaskan dasar pemilihan split pada pembentukan pohon. (4) dan persamaan (5)

$$Ginit(t) = 1 - \sum_{k=1}^K P_k^2 \quad (4)$$

$$\Delta Gini = Gini(parent) - \left(\frac{N_L}{N} Gini(L) + \frac{N_R}{N} Gini(R) \right) \quad (5)$$

Pemilihan *split* dilakukan dengan mempertimbangkan pengurangan impurity antara node induk dan node hasil pemisahan, sehingga model dapat membentuk aturan keputusan yang membedakan kategori Status Gizi. Penelitian ini menetapkan `max_depth=8` untuk membatasi kedalaman maksimum pohon dan mengendalikan kompleksitas model guna mengurangi risiko *overfitting*. Selain itu, parameter `class_weight="balanced"` diterapkan untuk memberikan bobot yang lebih besar terhadap kelas dengan jumlah observasi lebih sedikit, sehingga proses pembelajaran tidak terlalu didominasi oleh kelas mayoritas. Parameter `random_state` digunakan untuk memastikan hasil pemodelan dapat direproduksi. Model Decision Tree selanjutnya digunakan untuk menghasilkan prediksi Status Gizi pada data pengujian dan dibandingkan dengan algoritma Logistic Regression, Random Forest, dan Gradient Boosting berdasarkan Accuracy, Precision, Recall, dan Macro F1-score.

2.5.3. Random Forest

Random Forest digunakan sebagai algoritma *ensemble learning* untuk mengklasifikasikan Status Gizi balita berdasarkan variabel Umur, Jenis Kelamin, dan Tinggi Badan [25]. Model dibangun dengan menggabungkan 300 Decision Tree melalui mekanisme *bootstrap aggregation* dan pemilihan subset fitur secara acak, sehingga setiap pohon memiliki variasi data dan struktur keputusan yang berbeda. Parameter `n_estimators=300` menetapkan jumlah pohon yang digunakan dalam ensemble, sedangkan `max_depth=None` memungkinkan setiap pohon berkembang tanpa

batas kedalaman yang ditentukan secara eksplisit hingga kondisi penghentian terpenuhi. Mengatasi ketidakseimbangan kelas Status Gizi, digunakan parameter `class_weight="balanced"`, sehingga kelas dengan jumlah observasi lebih sedikit memperoleh bobot yang lebih besar selama proses pembelajaran, penelitian ini menggunakan persamaan (6) untuk prediksi akhir melalui majority voting dan persamaan (7) untuk probabilitas kelas

$$\hat{y} = \text{mode}\{T_1(X), T_2(X), \dots, T_{300}(X)\} \quad (6)$$

$$P(y = k|X) = \frac{1}{300} \sum_{b=1}^{300} I(T_b(X) = k) \quad (7)$$

2.5.4. Gradient Boosting

Gradient Boosting digunakan sebagai salah satu algoritma *ensemble learning* untuk melakukan klasifikasi Status Gizi balita berdasarkan variabel Umur, Jenis Kelamin dan Tinggi Badan [26]. Berbeda dengan Random Forest yang membangun sejumlah pohon keputusan secara relatif independen, Gradient Boosting membangun model secara bertahap (*sequential*), di mana setiap *decision tree* baru diarahkan untuk memperbaiki kesalahan prediksi yang dihasilkan oleh model sebelumnya. Implementasi penelitian menggunakan GradientBoostingClassifier dengan parameter `random_state` untuk memastikan konsistensi data dalam eksperimen. Parameter lainnya menggunakan konfigurasi default, yaitu 100 *boosting stages* (`n_estimators=100`), *learning rate* sebesar 0,1, dan kedalaman maksimum *base decision tree* sebesar 3. Tidak seperti Logistic Regression, Decision Tree, dan Random Forest yang pada penelitian ini menggunakan `class_weight="balanced"`, implementasi Gradient Boosting tidak menerapkan pembobotan kelas tersebut. Model selanjutnya digunakan untuk memprediksi Status Gizi pada data pengujian dan dibandingkan dengan model lainnya berdasarkan Accuracy, Precision, Recall, dan Macro F1-score, serta dianalisis menggunakan *classification report*, *confusion matrix*, dan 5-Fold Stratified Cross-Validation.

2.7. Nutritional Status Profiling Framework

Pembuatan profil status gizi yang diusulkan ini dirancang untuk mengidentifikasi dan mengkarakterisasi status gizi anak balita secara sistematis berdasarkan indikator antropometri menggunakan teknik pembelajaran mesin. Kerangka kerja ini terdiri dari tahapan-tahapan berurutan, yaitu audit data, *cleaning data*, Exploratory Data Analysis (EDA), penentuan fitur dan target, pembagian data, pra-pemrosesan, pemodelan pembelajaran mesin, evaluasi kinerja, *cross validation* dan interpretasi model. Kumpulan data awalnya diperiksa untuk menentukan karakteristik struktural, tipe data, statistik deskriptif, *missing values* dan distribusi kelas dari variabel status gizi. Selanjutnya, audit kualitas data dilakukan dengan mengidentifikasi rekaman duplikat persis serta memeriksa kombinasi duplikat dari variabel prediktor khususnya usia, jenis kelamin dan tinggi badan yang terkait dengan label status gizi berbeda. Langkah ini untuk observasi identik yang muncul baik pada subset pelatihan maupun pengujian dapat menghasilkan kinerja model. Setelah penilaian kualitas data, analisis data eksploratif dilakukan untuk mengkarakterisasi distribusi usia dan tinggi badan serta mengamati pola status gizi berdasarkan jenis kelamin dan kelompok usia yang telah ditentukan. Usia dikelompokkan ke dalam enam interval, yaitu 0–5, 6–11, 12–23, 24–35, 36–47, dan 48–60 bulan yang memungkinkan pengamatan distribusi status gizi pada berbagai tahapan usia dini. Analisis deskriptif ini memberikan profil awal kumpulan data dan menetapkan konteks empiris untuk pemodelan pembelajaran mesin selanjutnya.

Pada tahap pemodelan prediktif, variabel masukan terdiri dari tiga prediktor antropometri dan demografi, yaitu usia, jenis kelamin dan tinggi badan. Sedangkan, status gizi ditetapkan sebagai variabel target. Kumpulan data yang telah dibersihkan dibagi menjadi subset pelatihan dan pengujian menggunakan rasio 80:20 dengan pengambilan sampel menggunakan **stratified sampling** untuk mempertahankan distribusi kelas asli pada kedua subset tersebut. Variabel numerik, yang terdiri dari usia dan tinggi badan, distandardisasi menggunakan **StandardScaler**,

sedangkan variabel kategorikal jenis kelamin ditransformasikan menggunakan **one-hot encoding**. Operasi pra-pemrosesan ini diimplementasikan dalam sebuah **pipeline** dan diterapkan secara konsisten pada berbagai model pembelajaran mesin untuk menghindari ketidakkonsistenan antara data pelatihan dan data evaluasi. Empat algoritma klasifikasi dievaluasi secara komparatif, yaitu Regresi Logistik, Decision Tree, Random Forest dan Gradient Boosting. Regresi Logistik digunakan sebagai *baseline* klasifikasi linear, Pohon Keputusan mewakili pendekatan nonlinear berbasis aturan, Random Forest menyediakan ansambel pohon keputusan yang didasarkan pada teknik *bagging*, sedangkan Gradient Boosting mewakili pendekatan boosting sekuensial. Pembobotan kelas diterapkan pada Regresi Logistik, Pohon Keputusan, dan Random Forest menggunakan strategi penyeimbangan *balanced strategy* untuk mengurangi potensi pengaruh ketimpangan representasi kelas, sementara Gradient Boosting diimplementasikan sebagai model boosting dasar yang tersedia dalam library *scikit-learn*. Setiap model dilatih menggunakan data pelatihan yang sama dan dievaluasi pada set pengujian menggunakan metrik akurasi, presisi *macro-averaged precision*, *macro-averaged recall*, dan *macro-averaged F1-score*. Metrik rata-rata makro ditekankan karena memberikan bobot yang sama bagi setiap kelas status gizi sehingga lebih informatif ketika distribusi kelas tidak seimbang sempurna.

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil dan pembahasan penelitian mengenai profil status gizi balita berbasis pembelajaran mesin menggunakan indikator antropometri. Analisis dilakukan secara bertahap untuk memberikan gambaran mengenai karakteristik data, kualitas data, pola distribusi status gizi, serta kemampuan model pembelajaran mesin dalam melakukan klasifikasi status gizi balita. Pembahasan diawali dengan karakteristik dataset dan hasil audit kualitas serta preprocessing untuk memastikan bahwa data yang digunakan dalam pemodelan memiliki struktur dan kualitas yang memadai. Selanjutnya, dilakukan analisis eksploratif terhadap distribusi status gizi berdasarkan karakteristik antropometri dan demografis untuk mengidentifikasi pola pertumbuhan serta perbedaan status gizi pada kelompok balita. Tahap berikutnya membahas hasil klasifikasi menggunakan empat algoritma pembelajaran mesin, yaitu Logistic Regression, Decision Tree, Random Forest dan Gradient Boosting lalu dibandingkan berdasarkan metrik performa untuk mengidentifikasi model dengan kemampuan klasifikasi terbaik. Evaluasi lebih lanjut dilakukan melalui analisis *confusion matrix* dan performa pada masing-masing kelas untuk mengetahui pola kesalahan klasifikasi serta kemampuan model dalam membedakan kategori status gizi. Model terbaik kemudian dianalisis melalui *feature importance* dan interpretasi model untuk mengidentifikasi kontribusi indikator antropometri dalam membentuk profil status gizi balita. Hasil penelitian selanjutnya dibandingkan dengan penelitian terdahulu untuk melihat kesesuaian maupun perbedaan performa dan pendekatan yang digunakan.

3.1. Karakteristik dataset dan distribusi Kelas

Dataset yang digunakan dalam penelitian ini terdiri atas 120.999 observasi yang merepresentasikan karakteristik pertumbuhan dan status gizi balita. Dataset mencakup empat variabel utama, yaitu Umur, Jenis Kelamin, Tinggi Badan, dan Status Gizi. Variabel umur dan tinggi badan merupakan variabel numerik, sedangkan jenis kelamin dan status gizi merupakan variabel kategorikal. Berdasarkan pemeriksaan tipe data, variabel Umur memiliki tipe *int64*, Tinggi Badan bertipe *float64*, sedangkan Jenis Kelamin dan Status Gizi bertipe *object*. Struktur tersebut menunjukkan bahwa dataset telah memiliki pemisahan karakteristik data numerik dan kategorikal yang jelas sehingga dapat digunakan sebagai dasar dalam proses analisis dan pemodelan klasifikasi, untuk lebih jelasnya dapat dilihat pada Tabel 1.

Tabel 1. Statistik Deskriptif

Atribut	count	unique	top	freq	mean	std	min	25%	50%	75%	Max
Umur	120999	NaN	NaN	NaN	30.1738	17.5751	0.0	15.0	30.0	45.0	60.0
Jenis Kelamin	120999	2	Perempuan	61002	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Tinggi Badan	120999	NaN	NaN	NaN	88.6554	17.3009	40.01	77.0	89.8	101.2	128.0
Status Gizi	120999	4	67755	67755	NaN	NaN	NaN	NaN	NaN	NaN	NaN

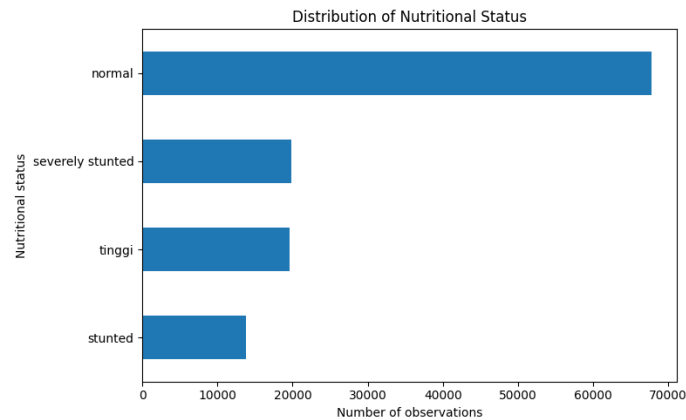
Tabel 1 merupakan hasil pemeriksaan kualitas data menunjukkan bahwa tidak terdapat nilai yang hilang (*missing values*) pada seluruh variabel. Masing-masing variabel memiliki jumlah observasi lengkap sebanyak 120.999 data dengan persentase *missing value* sebesar 0,0%. Kondisi ini menunjukkan bahwa dataset memiliki kelengkapan data yang baik dan tidak memerlukan proses imputasi untuk menangani nilai yang hilang. Secara deskriptif, variabel Umur memiliki rentang nilai antara 0 hingga 60 bulan, dengan nilai rata-rata sebesar 30,17 bulan, standar deviasi 17,58, median 30 bulan, serta kuartil pertama dan ketiga masing-masing sebesar 15 dan 45 bulan. Distribusi tersebut menunjukkan bahwa cakupan usia dalam dataset berada pada rentang balita secara umum, dengan persebaran umur yang relatif merata di sekitar nilai tengah. Nilai minimum 0 bulan menunjukkan bahwa dataset juga mencakup bayi pada usia awal kehidupan. Sementara itu, variabel Tinggi Badan (cm) memiliki nilai minimum 40,01 cm dan maksimum 128 cm, dengan rata-rata 88,66 cm, standar deviasi 17,30 cm, serta median 89,8 cm. Nilai kuartil pertama sebesar 77 cm dan kuartil ketiga sebesar 101,2 cm menunjukkan adanya variasi tinggi badan yang cukup besar sesuai dengan rentang usia yang terdapat dalam dataset. Variasi kedua variabel numerik tersebut penting karena umur dan tinggi badan merupakan karakteristik antropometri yang berkaitan erat dengan penentuan status pertumbuhan dan gizi balita. Pada variabel kategorikal Jenis Kelamin, terdapat dua kategori, dengan kategori *perempuan* menjadi kategori yang paling dominan sebanyak 61.002 observasi. Berdasarkan total dataset, jumlah tersebut menunjukkan bahwa sekitar 50,42% observasi merupakan balita perempuan, sedangkan sekitar 49,58% merupakan balita laki-laki. Selanjutnya, variabel Status Gizi memiliki empat kategori, dengan kategori *normal* sebagai kelas yang paling banyak ditemukan, yaitu sebanyak 67.755 observasi atau sekitar 55,99% dari keseluruhan data. Dominasi kelas *normal* menunjukkan bahwa distribusi target tidak sepenuhnya seimbang dan terdapat perbedaan proporsi antar kelas status gizi. Distribusi frekuensi status gizi balita berdasarkan indikator antropometri pada dataset awal sebelum dilakukan pemrosesan lanjutan. Secara visual, data menunjukkan ketidakseimbangan kelas (*class imbalance*) yang signifikan, di mana kategori status gizi normal mendominasi secara masif dengan jumlah mencapai kurang lebih 68.000 observasi balita.

3.2. Data Quality Audit

Dataset awal memiliki 120.999 observasi dan 4 variabel, yaitu Umur, Jenis Kelamin, Tinggi Badan dan Status Gizi. Ketiga variabel pertama digunakan sebagai karakteristik atau prediktor, sedangkan Status Gizi digunakan sebagai variabel target. Struktur tipe data juga telah sesuai dengan karakteristik masing-masing variabel: Umur bertipe integer (int64), Tinggi Badan bertipe numerik (float64), sedangkan Jenis Kelamin dan Status Gizi bertipe kategorikal (object). Dari sisi kelengkapan data, seluruh variabel memiliki 0 missing value atau 0,00%, sehingga tidak diperlukan imputasi untuk menangani data yang hilang. Hal ini menunjukkan bahwa dataset awal memiliki kelengkapan data yang baik untuk digunakan dalam tahapan analisis selanjutnya.

3.3. Distribusi Status Gizi

Analisis karakteristik dataset merupakan tahapan krusial untuk memahami eksplorasi awal struktur data sebelum dilakukan pemodelan *machine learning*. Pemahaman mengenai sebaran variabel target sangat diperlukan untuk mengidentifikasi potensi kendala statistik, seperti ketidakseimbangan kelas (*class imbalance*) yang dapat memengaruhi performa algoritma klasifikasi. Gambar 1 menyajikan distribusi frekuensi status gizi balita berdasarkan indikator antropometri pada dataset awal penelitian.

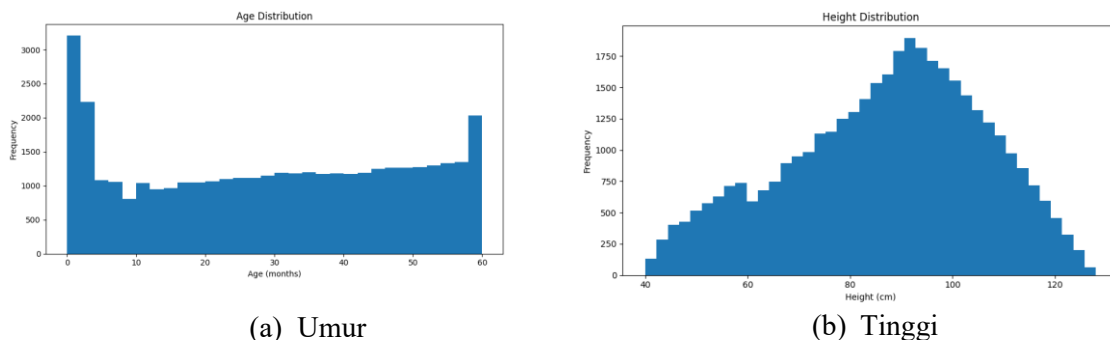


Gambar 1. Distribusi Berdasarkan Status Gizi

Berdasarkan visualisasi pada Gambar 1, data menunjukkan ketidakseimbangan kelas yang cukup signifikan. Kategori status gizi normal mendominasi dataset secara masif dengan jumlah mencapai sekitar 68.000 observasi balita. Sementara itu, kategori severely stunted (sangat pendek) dan tinggi berada pada kisaran 20.000 observasi, dan kategori stunted (pendek) menjadi kelas minoritas dengan frekuensi terendah, yaitu sekitar 14.000 observasi.

3.4. Distribusi Umur dan Tinggi Badan

Distribusi umur dan tinggi badan merupakan langkah awal yang krusial untuk memahami pola persebaran fisik balita yang mendasari penentuan profil status gizi berbasis indikator antropometri. Evaluasi terhadap bentuk distribusi ini penting untuk mendeteksi rentang nilai dominan, *outliers*, serta kecenderungan pemusatan data sebelum dimasukkan ke dalam algoritma *machine learning*. Gambar 2 menampilkan histogram sebaran data balita berdasarkan dua variabel antropometri utama, yaitu umur dan tinggi badan.

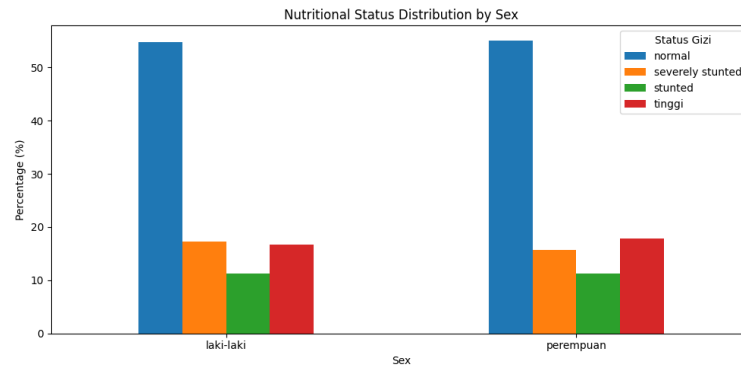


Gambar 2. (a), (b) Distribusi Berdasarkan Umur dan Tinggi

Berdasarkan visualisasi pada Gambar 2(a), distribusi umur balita mencakup rentang 0 hingga 60 bulan (0–5 tahun) dengan frekuensi tertinggi terkonsentrasi pada kelompok usia awal lahir (0–3 bulan) yang mencapai lebih dari 3.000 observasi, serta lonjakan frekuensi kembali terlihat pada kelompok usia mendekati 60 bulan (sekitar 2.000 observasi). Sementara itu, untuk kelompok umur 10 hingga 55 bulan, distribusi data terbagi secara relatif merata pada kisaran 1.000 hingga 1.300 observasi per segmen umur. Di sisi lain, Gambar 2(b) menunjukkan distribusi tinggi badan balita pada rentang 40 cm hingga 128 cm yang membentuk pola mendekati distribusi normal (*bell-shaped curve*). Frekuensi puncak berada pada kisaran tinggi badan 90 cm hingga 95 cm dengan jumlah observasi mendekati 1.900 balita, sedangkan frekuensi terendah berada pada batas ekstrem kiri (< 45 cm) dan ekstrem kanan (> 120 cm). Karakteristik distribusi ini mengonfirmasi bahwa dataset memiliki variasi antropometri yang representatif untuk membedakan kategori status gizi (*severely stunted*, *stunted*, *normal*, dan *tinggi*) dalam proses klasifikasi.

3.5. Nutrisional Status berdasarkan Jenis Kelamin atau Kelompok Umur

Analisis demografis berdasarkan faktor jenis kelamin menjadi aspek krusial dalam identifikasi profil status gizi balita, mengingat pola pertumbuhan fisik dapat menunjukkan karakteristik yang berbeda antara balita laki-laki dan perempuan. Evaluasi terhadap proporsi ini penting untuk memastikan bahwa dataset tidak memiliki bias gender yang ekstrem sebelum diintegrasikan ke dalam pemodelan *machine learning*. Gambar 3 menyajikan persentase distribusi status gizi balita yang dikelompokkan berdasarkan jenis kelamin.



Gambar 3. Distribusi Berdasarkan Jenis Kelamin

Gambar 3 merupakan distribusi status gizi antara kelompok balita laki-laki dan perempuan menunjukkan pola proporsi yang sangat identik secara keseluruhan. Pada kedua kelompok gender, kategori status gizi normal mendominasi secara signifikan dengan proporsi tertinggi mencapai lebih dari 50% (sekitar 55%). Kategori severely stunted berada pada urutan kedua dengan persentase berkisar antara 16% hingga 17%, di mana kelompok laki-laki sedikit lebih tinggi dibandingkan perempuan. Sebaliknya, persentase kategori tinggi pada balita perempuan sedikit lebih dominan (sekitar 18%) dibandingkan balita laki-laki (sekitar 16%). Sementara itu, kategori stunted mencatatkan proporsi terendah pada kedua jenis kelamin, yaitu berada pada kisaran 11%. Konsistensi pola distribusi ini mengindikasikan bahwa indikator jenis kelamin memiliki korelasi yang seimbang terhadap variasi status gizi antropometri dalam dataset.

3.6. Perbandingan Performa Model Pembelajaran Mesin

Perbandingan performa dilakukan untuk mengevaluasi kemampuan masing-masing model pembelajaran mesin dalam mengklasifikasikan status gizi balita berdasarkan indikator antropometri yang digunakan. Evaluasi tidak hanya mempertimbangkan akurasi keseluruhan, tetapi juga menggunakan *precision*, *recall*, dan *F1-score* makro untuk memperoleh gambaran performa yang lebih komprehensif pada setiap kelas. Penggunaan metrik makro penting karena memberikan bobot yang sama terhadap setiap kelas, sehingga performa model pada kelas dengan jumlah sampel yang lebih sedikit tetap diperhitungkan dalam evaluasi. Hasil perbandingan performa Random Forest, Gradient Boosting, Decision Tree, dan Logistic Regression berdasarkan keempat metrik tersebut disajikan pada Tabel 2.

Tabel 2. Perbandingan Performa Model Pembelajaran Mesin

Model	Accuracy	Precision makro	Recall makro	F1 makro
Random Forest	0.9894	0.9874	0.9845	0.9859
Gradient Boosting	0.9320	0.9331	0.8778	0.9000
Decision Tree	0.8268	0.7725	0.8387	0.7914
Logistic Regression	0.6569	0.6123	0.6817	0.6250

Berdasarkan Tabel 2, Random Forest menunjukkan performa terbaik pada seluruh metrik evaluasi, dengan *accuracy* sebesar 0.9894, *precision* makro sebesar 0.9874, *recall* makro sebesar 0.9845, dan *F1-score* makro sebesar 0.9859. Nilai tersebut menunjukkan bahwa Random Forest

mampu menghasilkan prediksi yang sangat baik dalam mengenali kelas yang sebenarnya maupun dalam meminimalkan kesalahan klasifikasi antar kelas. Selisih yang relatif kecil antara nilai *precision*, *recall*, dan *F1-score* juga mengindikasikan bahwa model memiliki performa yang seimbang pada berbagai kelas status gizi. Secara keseluruhan, hasil evaluasi memperlihatkan adanya perbedaan performa yang cukup jelas antar algoritma. Random Forest memberikan peningkatan *accuracy* sekitar 5,74% dibandingkan Gradient Boosting, 16,26% dibandingkan Decision Tree, dan 33,25% dibandingkan Logistic Regression. Keunggulan yang konsisten pada *precision*, *recall* dan *F1-score* makro memperkuat bahwa Random Forest merupakan model yang paling sesuai untuk tugas klasifikasi pada dataset penelitian ini. Oleh karena itu, Random Forest dipilih sebagai model dengan performa paling optimal dalam pembentukan profil status gizi balita pada penelitian ini.

3.7. 5-Fold Stratified Cross-Validation

Evaluasi performa model dilakukan melalui *confusion matrix* untuk memperoleh gambaran yang lebih rinci mengenai pola prediksi pada setiap kelas status gizi. *Confusion matrix* memungkinkan identifikasi jumlah prediksi yang benar (*true positive*) maupun kesalahan klasifikasi (*false positive* dan *false negative*) pada masing-masing kelas. Analisis ini penting karena nilai *accuracy* yang tinggi belum tentu menunjukkan bahwa model mampu mengenali seluruh kelas secara seimbang. Selain itu, untuk memastikan bahwa performa model tidak bergantung pada satu pembagian data tertentu, penelitian ini menggunakan 5-Fold Stratified Cross-Validation. Metode tersebut membagi dataset menjadi lima subset (*fold*) dengan mempertahankan proporsi kelas pada setiap subset, kemudian model dilatih dan diuji sebanyak lima kali dengan kombinasi subset yang berbeda. Nilai rata-rata dan standar deviasi dari metrik evaluasi pada kelima *fold* digunakan untuk menilai tingkat performa sekaligus konsistensi model. Hasil evaluasi menggunakan 5-Fold Stratified Cross-Validation disajikan pada Tabel 3.

Tabel 3. 5-Fold Stratified Cross-Validation

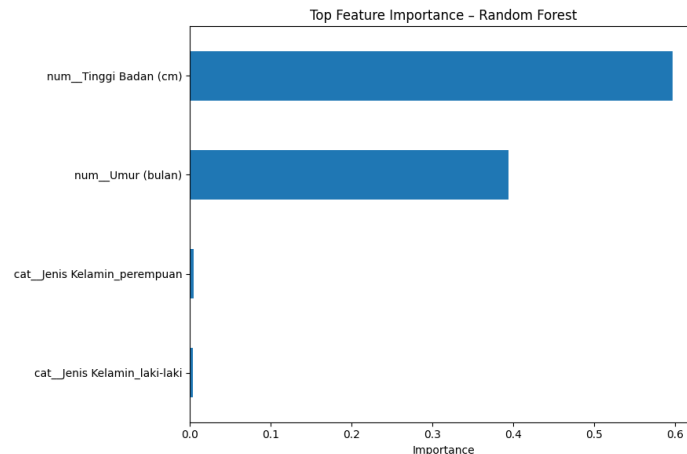
Model	Accuracy Mean	Accuracy std	Precision Macro Mean	Recall Macro Mean	F1 Macro Mean	F1 Macro std
Random Forest	0.9904	0.0009	0.9872	0.9865	0.9869	0.0011
Gradient Boosting	0.9335	0.0040	0.9369	0.8780	0.9012	0.0081
Decision Tree	0.8231	0.0102	0.7679	0.8338	0.7871	0.0092
Logistic Regression	0.6487	0.0052	0.6036	0.6718	0.6158	0.0060

Berdasarkan Tabel 3, Random Forest kembali menunjukkan performa paling unggul dan paling konsisten dibandingkan model lainnya. Model ini memperoleh rata-rata *accuracy* sebesar 0.9904 dengan standar deviasi hanya 0.0009. Nilai *precision* makro sebesar 0.9872, *recall* makro sebesar 0.9865, dan *F1-score* makro sebesar 0.9869 menunjukkan bahwa model mampu mempertahankan performa yang tinggi pada seluruh kelas. Standar deviasi *F1-score* makro yang hanya sebesar 0.0011 juga menunjukkan variasi performa yang sangat kecil antar-*fold*. Hasil dari 5-Fold Stratified Cross-Validation secara keseluruhan menghasilkan temuan pada evaluasi sebelumnya bahwa Random Forest merupakan model dengan performa paling optimal. Selain menghasilkan rata-rata *accuracy*, *precision*, *recall*, dan *F1-score* makro tertinggi, model ini juga memiliki standar deviasi paling kecil, khususnya pada *accuracy* dan *F1-score*. Perbedaan tersebut menunjukkan bahwa performa Random Forest relatif tidak berubah meskipun model dievaluasi menggunakan lima kombinasi pembagian data yang berbeda.

Temuan dari *confusion matrix* dan 5-Fold Stratified Cross-Validation selanjutnya memberikan bukti yang lebih kuat mengenai kemampuan Random Forest dalam melakukan klasifikasi status gizi balita. Jika *confusion matrix* digunakan untuk mengidentifikasi pola kesalahan klasifikasi pada tingkat kelas, maka *cross-validation* digunakan untuk menguji apakah performa tersebut konsisten terhadap variasi data. Kombinasi kedua pendekatan evaluasi tersebut menunjukkan bahwa pemilihan Random Forest tidak hanya didasarkan pada satu nilai *accuracy*, tetapi mempertimbangkan ketepatan klasifikasi, kemampuan mendeteksi setiap kelas, keseimbangan performa, serta stabilitas model secara keseluruhan.

3.8. Feature Importance and Nutritional Profiling

Feature importance merupakan aspek mendasar dalam klasifikasi *machine learning* untuk mengidentifikasi indikator yang paling berpengaruh dalam pembentukan profil status gizi balita. Evaluasi ini memberikan transparansi terhadap model *Random Forest* mengenai bagaimana setiap variabel antropometri dan demografi berkontribusi dalam memprediksi kategori status gizi secara akurat. Gambar 4 menampilkan hasil analisis *Top Feature Importance* dari algoritma *Random Forest* yang menggambarkan kontribusi relatif dari masing-masing fitur.



Gambar 4. Feature Importance

Gambar 4 merupakan variabel numerik Tinggi Badan menjadi fitur yang paling dominan dengan nilai kontribusi tertinggi mendekati 0.60 (60%). Fitur Umur menempati posisi kedua paling berpengaruh dengan nilai kontribusi sekitar 0.39 (39%). Sebaliknya, fitur kategorikal berbasis gender, yaitu Jenis Kelamin_perempuan dan Jenis Kelamin_laki-laki, memberikan kontribusi yang sangat minimal (kurang dari 0.01 atau 1%) terhadap keputusan pemodelan. Temuan ini menegaskan bahwa kombinasi indikator fisik numerik khususnya tinggi badan dan umur merupakan penentu utama dalam pembuatan profil status gizi balita, sedangkan variabel jenis kelamin berfungsi sebagai fitur pendukung sekunder dalam model prediktif berbasis *Random Forest*.

4. Kesimpulan

Penelitian ini berhasil menyusun profil status gizi balita berbasis pembelajaran mesin menggunakan indikator antropometri melalui alur eksperimen yang terstruktur dan komprehensif. Tahap eksplorasi awal pada dataset yang mencakup 120.999 observasi menunjukkan kualitas data yang sangat baik tanpa adanya nilai yang hilang (*missing value*), meskipun teridentifikasi adanya ketidakseimbangan kelas (*class imbalance*) di mana kategori normal mendominasi sebesar 55,99%. Dari evaluasi empat algoritma pembelajaran mesin (*Logistic Regression*, *Decision Tree*, *Gradient Boosting*, dan *Random Forest*), model *Random Forest* terbukti memiliki performa paling optimal dan seimbang dalam memetakan profil status gizi balita. Model ini mencapai tingkat akurasi sebesar 98,94% dengan *Precision* makro 98,74%, *Recall* makro 98,45%, dan *F1-Score* makro 98,59%. Tingkat keandalan dan konsistensi pemodelan ini semakin diperkuat oleh hasil *5-Fold Stratified Cross-Validation* yang menghasilkan rata-rata akurasi sebesar 99,04% dengan variasi yang sangat minimal (standar deviasi 0,0009). Lebih lanjut, analisis *feature importance* mengonfirmasi bahwa dalam pembuatan profil status gizi balita, indikator antropometri numerik Tinggi Badan sebesar 60% dan Umur sebesar 39% merupakan variabel penentu paling dominan, sedangkan jenis kelamin hanya berfungsi sebagai indikator pendukung sekunder. Dengan demikian, model *Random Forest* yang dikembangkan terbukti unggul, stabil, serta transparan secara interpretasi untuk diterapkan dalam pembuatan profil status gizi balita berbasis pembelajaran mesin menggunakan indikator antropometri.

Daftar Pustaka

- [1] R. Desmita, R. S. Siregar, V. Ivanda, Y. Hanoselina, and R. F. Helmi, "Peran Tenaga Kesehatan dalam Pemberdayaan Keluarga untuk Pencegahan Stunting," *Cult. Educ. Technol.*, vol. 2, no. 2, pp. 177–190, 2025, doi: 10.31004/ctr.v2i1.123.
- [2] Rakhmalia Imeldawati, "Dampak Terjadinya Stunting terhadap Perkembangan Kognitif Anak : Literature Review," *J. Med. Nusantara*, vol. 3, no. 1, pp. 101–107, 2025, doi: 10.59680/medika.v3i1.1632.
- [3] A. Mulyana, Hendra, and Retnoningsih, "Implikasi Stunting Terhadap Kemampuan Kognitif Dan Fisik Anak Usia Dini Di TK At-Taqwa Teta Lambitu," *J. Ilm. Pendidik. Dasar*, vol. 10, no. 2, pp. 122–133, 2025.
- [4] S. Monica, T. C. Maigoda, and A. Krisnasary, "Literature review : Intervensi stunting di negara berkembang," *Nutr. Heal. Insights*, vol. 1, no. 2, pp. 87–97, 2024, [Online]. Available: <https://journal.dapupublishing.com/NaHI/article/view/20>.
- [5] D. Sartika, M. Munawarah, and M. I. S, "Pengaruh konsumsi makanan bergizi pada balita terhadap stunting," *J. Nurs. Pract. Educ.*, vol. 5, no. 1, pp. 1–9, 2024, doi: 10.34305/jnpe.v5i1.1370.
- [6] F. A. Nursifa, I. M. Hafiz, S. Khoirunnisa, H. Ridwan, and P. Haryeti, "Pemenuhan Gizi pada Seribu Hari Pertama Kehidupan Terhadap Kejadian Stunting di Indonesia : Literature Review," *J. Kesehat. Indra Husada*, vol. 13, no. 1, pp. 82–89, 2025.
- [7] 'Aliah Istiqomah, Kristin Masmur S, Ribby Aurellia Amali, and Sulis Tiawati, "Peran Gizi Terhadap Pertumbuhan Dan Perkembangan Balita," *Antigen J. Kesehat. Masy. dan Ilmu Gizi*, vol. 2, no. 2, pp. 67–74, 2024, doi: 10.57213/antigen.v2i2.260.
- [8] T. T. M. Ginting and A. Zebua, "Sosialisasi Pencegahan Stunting di Kelurahan Sei Mati Kecamatan Medan Maimun: Upaya Peningkatan Kesehatan Ibu dan Anak," *J. Pengabd. Masy. Bhinneka*, vol. 3, no. 1, pp. 27–31, 2024, doi: 10.58266/jpmb.v3i1.97.
- [9] Y. P. Dea Tazkia Aulia, Suganda Tanuwidjaja, "Gambaran Status Gizi pada Balita di Puskesmas Jatiwangi Tahun 2024," *Med. Sci.*, p. 8, 2024, [Online]. Available: <https://doi.org/10.29313/bcsms.v6i1.22623>.
- [10] Regiska Falensya, "Pentingnya Gizi Seimbang dalam Mencegah Stunting pada Anak Usia Dini," *J. Innov. Educ.*, vol. 3, no. 4, pp. 08–21, 2025, doi: 10.59841/inoved.v3i4.3549.
- [11] C. Fauziyah, R. Hasibuan, A. A. Sholikhah, M. S. Jenar, and A. Zahrotunnisa, "Dampak Kurangnya Asupan Gizi pada Perkembangan Anak di Masa Pandemi," *JIIIP - J. Ilm. Ilmu Pendidik.*, vol. 8, no. 7, pp. 7684–7691, 2025, doi: 10.54371/jiip.v8i7.8742.
- [12] Dedeng Nurkholik Sidik Permana, Nilawaty Uly, and Andi Alim, "Efektivitas Berbagai Intervensi Gizi dalam Penanggulangan Stunting pada Anak: Tinjauan Literatur," *Borneo Nurs. J.*, vol. 7, no. 2, pp. 401–411, 2025, doi: 10.61878/bnj.v7i2.102.
- [13] A. Emilda, Alchalidi, and M. Sukmadewi, "Faktor-faktor yang mempengaruhi status gizi balita berdasarkan BB/U," *J. Kesehat. Ibu dan Anak*, vol. 6, no. 2, pp. 13–20, 2014.
- [14] S. Mau, Five Bunda De Putri Dopo, F. Tedy, and P. F. M. Tengangatu, "Implementasi Algoritma Naive Bayes pada Aplikasi Web untuk Klasifikasi Status Gizi Balita," *J. Inform. Polinema*, vol. 12, no. 2, pp. 333–340, 2026, doi: 10.33795/jip.v12i2.8898.
- [15] J. Ipmawati and I. Unggara, "Analisis Status Gizi Anak Menggunakan Metode Klastering

- pada Dataset Anthropometri,” *bit-Tech*, vol. 7, no. 2, pp. 494–504, 2024, doi: 10.32877/bt.v7i2.1869.
- [16] P. Fahmi Ruziq and Mk. M. Rhifky Wayahdi, “Implementasi Algoritma K-Nearest Neighbor (K-Nn) Untuk Klasifikasi Data Kesehatan,” 2026, [Online]. Available: www.media.hadlacr.com.
 - [17] D. Sartika, F. Elfaladonna, and A. Octarina, “Optimizing Data Preprocessing on Stunting Datasets: Identifying Relevant Attributes for Machine Learning Analysis,” pp. 481–501, 2025, doi: 10.2991/978-94-6463-678-9_45.
 - [18] A. Masitha, S. Lonang, and J. M. Reski, “Machine Learning Approach for Heart Failure Patient Classification Using K-Nearest Neighbors Algorithm,” *Methods Sci. Technol. Stud.*, vol. 1, no. 2, pp. 81–88, 2025, doi: 10.64539/msts.v1i2.2025.44.
 - [19] D. Rifaldi, Abdul Fadlil, and Herman, “Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet ‘Mental Health,’” *Decod. J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, pp. 161–171, 2023, doi: 10.51454/decode.v3i2.131.
 - [20] M. Hudaya, W. Nor, M. Yulastina, and M. Nordiansyah, “Exploring Stunting in South Kalimantan Province Using R Programming-Based Data Visualization,” *Kesmas J. Kesehat. Masy. Nas.*, vol. 20, no. 3, pp. 213–221, 2025, doi: 10.7454/kesmas.v20i3.2267.
 - [21] A. Heryati, T. Terttiaivini, D. Marcelina, and H. Romli, “Optimization of Stunting Risk Prediction Using a Hybrid Genetic-Machine Learning Model,” *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 2, pp. 807–815, 2025, [Online]. Available: <https://ejurnal.pnl.ac.id/JAISE/article/view/6988>.
 - [22] R. Ratnasari, A. J. Wahidin, and T. H. Andika, “Deteksi Dini Stunting Pada Anak Berdasarkan Indikator Antropometri dengan Menggunakan Algoritma Machine Learning,” *J. Algoritm.*, vol. 21, no. 2, pp. 378–387, 2024, doi: 10.33364/algoritma/v.21-2.2122.
 - [23] N. I. E. Harahap, L. H. Hasibuan, and S. Musthofa, “Model Klasifikasi Tinggi Badan Balita Stunting Menggunakan Regresi Logistik Biner,” *JOSTECH J. Sci. Technol.*, vol. 5, no. 2, pp. 257–270, 2025, doi: 10.15548/jostech.v5i2.12420.
 - [24] F. MILLANO, S. Kurniawan, and R. Gustriansyah, “Penerapan Metode Decision Tree Dalam Klasifikasi Status Gizi Balita,” *Komputa J. Ilm. Komput. dan Inform.*, vol. 14, no. 2, pp. 53–62, 2025, doi: 10.34010/komputa.v14i2.16468.
 - [25] MAULINA, “Analisis Jejak Stunting Dengan Prediksi Kejadian Pada Anak Balita Menggunakan Metode Random Forest,” 2024.
 - [26] R. Y. Pratama and A. Baita, “Prediksi Stunting pada Anak Balita Menggunakan Algoritma Extreme Gradient Boosting dan Bayesian Optimization,” *J. Technol. Informatics*, vol. 7, no. 2, pp. 175–189, 2025, doi: 10.37802/joti.v7i2.1174.