

Klasifikasi Status Gizi Balita Berbasis Data Antropometri menggunakan Random Forest

Classification of Nutritional Status in Toddlers Based on Anthropometric Data Using Random Forest

Mundirin^{*1}, Idawati², Ibrahim Latief³

^{1,2,3}Prodi Sistem Informasi, Institut Sains dan Teknologi Al-Kamal

E-mail: ¹mundirin@ista.ac.id, ²idawati@ista.ac.id, ³ibrahim.latief@ista.ac.id

Received: June 20, 2025 | Revised: June 28, 2025 | Accepted: August 06, 2025

Abstrak

Stunting merupakan masalah gizi kronis yang masih menjadi tantangan utama dalam pembangunan kesehatan anak di Indonesia. Kondisi ini berdampak jangka panjang terhadap pertumbuhan fisik, perkembangan kognitif, dan produktivitas anak di masa depan. Deteksi dini status gizi balita sangat penting untuk mencegah dan menangani kasus stunting secara efektif. Penelitian ini bertujuan untuk mengembangkan model klasifikasi status gizi balita berbasis machine learning menggunakan data antropometri sederhana, yaitu umur (bulan), jenis kelamin, dan tinggi badan (cm). Dataset yang digunakan bersumber dari data histori dinas Kesehatan tahun 2022 dan Elektronik Pencatatan dan Pelaporan Gizi Berbasis Masyarakat (E-PPGBM). Dataset yang digunakan berjumlah 120.999 entri yang terbagi dalam empat kategori status gizi: normal, tinggi, stunted, dan severely stunted. Pra-pemrosesan data meliputi pengkodean label dan standardisasi fitur. Model yang digunakan adalah Random Forest Classifier, yang dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil pelatihan menunjukkan bahwa model mampu mengklasifikasikan data uji dengan akurasi sebesar 99.93%, serta nilai presisi, recall, dan F1-score per kelas Normal: F1 = 0.9998, Severely Stunted: F1 = 0.9985, Stunted: F1 = 0.9975, Tinggi: F1 = 0.9997. Analisis feature importance menunjukkan bahwa tinggi badan merupakan fitur paling dominan dalam klasifikasi. Hasil ini menunjukkan bahwa algoritma machine learning, khususnya Random Forest, efektif digunakan untuk prediksi status gizi balita dan berpotensi diintegrasikan dalam aplikasi digital untuk mendukung program penurunan stunting di Indonesia. Keterbatasan model dalam penelitian ini yaitu menggunakan fitur antropometri dasar, yaitu umur, jenis kelamin, dan tinggi badan, tanpa mempertimbangkan variabel tambahan seperti berat badan, riwayat penyakit, pola makan, status ekonomi, atau status imunisasi, yang juga dapat berpengaruh terhadap status gizi anak. Untuk meningkatkan akurasi dan relevansi model, disarankan untuk menambahkan fitur lain yang terkait dengan status gizi, seperti Berat badan, Riwayat penyakit dan asupan gizi, Indikator sosial-ekonomi dan lain-lain.

Kata kunci: machine learning; status gizi; stunting; klasifikasi; Random Forest

Abstract

Stunting is a chronic nutritional problem that remains a major challenge in improving child health in Indonesia. This condition has long-term impacts on physical growth, cognitive development, and future productivity of children. Early detection of toddlers' nutritional status is crucial for effectively preventing and addressing stunting cases. This study aims to develop a machine learning-based classification model for toddlers' nutritional status using simple anthropometric data, namely age (in months), sex, and height (in cm). The dataset used in this study was sourced from the 2022 historical records of the Health Department and the Community-Based Nutrition Recording and Reporting System (E-PPGBM), comprising 120,999 entries categorized into four nutritional status classes: normal, tall, stunted, and severely stunted. Data

324

preprocessing included label encoding and feature standardization. The model employed is the Random Forest Classifier, evaluated using accuracy, precision, recall, and F1-score metrics. The training results show that the model achieved a classification accuracy of 99.93% on the test data, with F1-scores for each class as follows: Normal = 0.9998, Severely Stunted = 0.9985, Stunted = 0.9975, Tall = 0.9997. Feature importance analysis indicates that height is the most influential feature in the classification task. These findings demonstrate that machine learning algorithms, particularly Random Forest, are effective for predicting toddlers' nutritional status and have strong potential to be integrated into digital applications that support Indonesia's stunting reduction programs. However, the model's limitation lies in its use of only basic anthropometric features—age, sex, and height—without considering additional variables such as weight, disease history, dietary patterns, socioeconomic status, or immunization history, which may also influence a child's nutritional status. To improve the model's accuracy and relevance, it is recommended to incorporate other related features, such as body weight, nutritional intake, health history, and social-economic indicators, in future research.

Keywords: machine learning; stunting; nutritional status; classification; Random Forest;

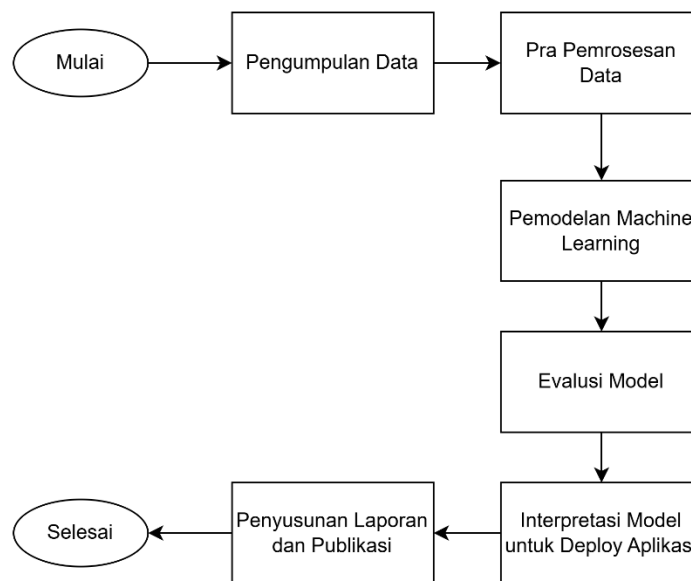
1. PENDAHULUAN

Stunting merupakan salah satu permasalahan kesehatan masyarakat yang krusial di Indonesia. Kondisi ini ditandai dengan kegagalan pertumbuhan anak yang diukur berdasarkan tinggi badan menurut umur (TB/U), yang mencerminkan kekurangan gizi kronis akibat asupan gizi yang tidak memadai dan infeksi berulang dalam jangka waktu yang lama. Berdasarkan laporan Survei Status Gizi Indonesia (SSGI) tahun 2022, prevalensi stunting di Indonesia mencapai 21,6% [1] angka yang masih jauh dari target nasional sebesar 14% pada tahun 2024 dan target global yang ditetapkan oleh World Health Organization (WHO), yaitu di bawah 20% [2]. Penelitian ini bertujuan untuk mengembangkan model klasifikasi status gizi balita berbasis machine learning menggunakan data antropometri sederhana, yaitu umur (bulan), jenis kelamin, dan tinggi badan (cm). Sebagian besar penelitian terdahulu masih menggunakan pendekatan statistik konvensional (seperti regresi logistik atau uji chi-square), yang kurang mampu menangkap pola non-linear dalam hubungan multivariat antar variabel gizi. Pendekatan tersebut mengasumsikan hubungan linier dan independensi antar variabel, yang sering kali tidak mencerminkan kompleksitas faktor penentu stunting dalam konteks nyata. Banyak studi belum memanfaatkan algoritma machine learning yang kuat seperti Random Forest, XGBoost, atau neural networks. Sebagian besar hanya menggunakan metode dasar seperti KNN atau Naïve Bayes, yang memiliki keterbatasan dalam performa ketika menangani dataset besar atau tidak seimbang. Saat ini, proses penilaian status gizi masih dilakukan secara manual dengan membandingkan data antropometri anak terhadap standar WHO. Proses ini membutuhkan waktu, sumber daya manusia yang kompeten, dan alat ukur yang akurat, yang belum tentu tersedia di seluruh wilayah, khususnya daerah terpencil. Di sinilah pendekatan berbasis teknologi kecerdasan buatan seperti machine learning dapat berperan sebagai solusi alternatif. Machine learning merupakan cabang dari kecerdasan buatan yang memungkinkan sistem komputer mempelajari pola dari data dan melakukan prediksi secara otomatis. Dengan memanfaatkan data sederhana seperti umur, jenis kelamin, dan tinggi badan, machine learning dapat membangun model klasifikasi status gizi balita secara cepat dan akurat [3][4]. Salah satu algoritma yang banyak digunakan karena kemampuannya dalam menangani data tabular adalah Random Forest, yaitu metode ensemble berbasis pohon keputusan yang terbukti akurat dan tahan terhadap overfitting [5]. Penanganan stunting tidak hanya menuntut intervensi kuratif, tetapi juga pencegahan melalui deteksi dini status gizi anak. Namun, proses klasifikasi status gizi secara manual masih bergantung pada tenaga kesehatan, peralatan antropometri, dan interpretasi terhadap standar pertumbuhan WHO, yang rentan terhadap kesalahan teknis dan keterbatasan sumber daya di daerah terpencil. Oleh karena itu, pendekatan

berbasis teknologi, khususnya pemanfaatan machine learning (ML), menjadi alternatif strategis dalam mempercepat dan meningkatkan akurasi deteksi dini.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis machine learning untuk membangun dan mengevaluasi model klasifikasi status gizi balita. Tahapan metode penelitian dibagi menjadi lima tahap utama: pengumpulan data, pra-pemrosesan data, pelatihan model, evaluasi model, dan analisis hasil. Dalam penelitian ini menggunakan model algoritma Random Forest Classifier dari pustaka scikit-learn dengan parameter `n_estimators` 100 Jumlah pohon dalam hutan (ensemble), `max_depth` None Tidak ada batasan kedalaman; pohon tumbuh sampai semua daun homogen.



Gambar 1. Skema Alur Proses Penelitian

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari dataset terbuka yang berisi data antropometri balita, terdiri atas variabel umur (dalam bulan), jenis kelamin, tinggi badan (dalam cm), dan label status gizi. Dataset ini memuat 120.999 entri dengan empat kategori status gizi: normal, tinggi, stunted, dan severely stunted.[3]. Dataset yang digunakan bersumber dari data histori dinas Kesehatan tahun 2022 dan Elektronik Pencatatan dan Pelaporan Gizi Berbasis Masyarakat (E-PPGBM).

2.2 Pra-Pemrosesan Data

Pra-pemrosesan dilakukan untuk menyiapkan data agar sesuai dengan kebutuhan algoritma machine learning. Langkah yang dilakukan meliputi:

- A. Encoding label pada variabel kategorikal, seperti jenis kelamin dan status gizi, dilakukan menggunakan LabelEncoder agar dapat dibaca oleh algoritma machine learning [3].
- B. Standardisasi fitur numerik, seperti umur dan tinggi badan, menggunakan metode Z-score (*StandardScaler*) untuk menyetarakan skala antar fitur dan menghindari bias model terhadap variabel dengan skala besar [6].

- C. Pembagian data (train-test split) dalam penelitian ini dilakukan dengan stratifikasi berdasarkan label kelas (status gizi). Pembagian data menjadi data latih dan data uji dengan rasio 80:20 menggunakan metode *train-test split*, sebagaimana umum diterapkan dalam eksperimen supervised learning [7].

2.3 Pemodelan

Algoritma yang digunakan adalah Random Forest Classifier, yaitu metode ensemble learning berbasis *bagging* yang menggabungkan banyak pohon keputusan (decision trees) untuk meningkatkan akurasi dan mengurangi risiko overfitting [8]. Random Forest dipilih karena kemampuannya menangani dataset berskala besar dan variabel input heterogen, serta telah terbukti efektif dalam berbagai aplikasi klasifikasi data kesehatan [9]. Pelatihan model dilakukan pada data latih menggunakan pustaka *scikit-learn* di Python, dengan parameter default dan random seed yang diatur agar hasil replikasi dapat dikontrol.

2.4 Evaluasi Model

Evaluasi dilakukan menggunakan metrik-metrik klasifikasi, yaitu:

- a. Akurasi: persentase prediksi yang tepat terhadap total data uji.
- b. Precision, Recall, dan F1-score: untuk mengukur keseimbangan antara prediksi positif yang benar dan kesalahan klasifikasi per kelas [10]
- c. Confusion Matrix: digunakan untuk menganalisis distribusi prediksi benar dan salah pada setiap kelas status gizi.

2.5 Analisis Fitur

Setelah pelatihan, dilakukan analisis *feature importance* untuk mengetahui fitur apa yang paling memengaruhi hasil prediksi model. Hal ini penting untuk interpretabilitas model dalam konteks implementasi kebijakan kesehatan dan sebagai dasar untuk pengembangan sistem rekomendasi [11].

3. HASIL DAN PEMBAHASAN

Model klasifikasi menggunakan algoritma Random Forest berhasil mengembangkan model prediksi status gizi balita ke dalam empat kategori: normal, tinggi, stunted, dan severely stunted dengan akurasi sebesar 99,93%. Evaluasi kinerja model menunjukkan nilai precision, recall, dan F1-score yang tinggi dan merata di seluruh kelas.

3.1 Interpretasi Confusion Matrix

Visualisasi *confusion matrix* menunjukkan bahwa model berhasil mengklasifikasikan hampir seluruh data uji secara tepat, dengan jumlah prediksi salah sangat rendah. Beberapa kesalahan klasifikasi yang terjadi umumnya adalah antara kelas Stunted dan Severely Stunted, yang secara antropometrik memiliki rentang tinggi badan yang berdekatan. Hal ini menunjukkan bahwa:

1. Model sangat sensitif dalam membedakan status gizi yang jelas (misalnya normal dan tinggi).
2. Kesalahan minor terjadi pada batas klasifikasi yang secara klinis pun sulit dipisahkan, seperti antara stunted dan severely stunted

3.2 Interpretasi Feature Importance

Analisis *feature importance* dari Random Forest mengungkapkan bahwa:

1. Tinggi badan merupakan prediktor paling dominan, yang sejalan dengan definisi status gizi berbasis Z-score tinggi badan menurut usia (TB/U) dalam standar WHO
2. Usia balita turut memberikan kontribusi penting karena status gizi dinilai relatif terhadap usia.
3. Jenis kelamin, meskipun tetap relevan, memiliki pengaruh yang lebih kecil dibandingkan

fitur lainnya

Implikasi dari temuan ini adalah bahwa data antropometri minimal (tinggi badan, usia, jenis kelamin) sudah cukup informatif untuk mendeteksi status gizi balita secara otomatis menggunakan pendekatan machine learning.

3.3 Interpretasi Klinis dan Programatik

Keakuratan model yang sangat tinggi serta kesesuaian dengan indikator gizi WHO memperkuat potensi aplikatif model ini. Dalam konteks layanan kesehatan masyarakat:

1. Model ini dapat menjadi alat bantu deteksi dini risiko stunting, terutama di daerah dengan keterbatasan tenaga ahli gizi.
2. Klasifikasi yang cepat dan akurat dapat mendukung pengambilan keputusan berbasis data, seperti penentuan balita yang perlu diintervensi dalam program pemulihan gizi
3. Prediksi status gizi berbasis machine learning dapat diintegrasikan dalam sistem informasi posyandu, puskesmas, atau aplikasi mobile yang mendukung surveilans kesehatan anak.

3.4 Sintesis Temuan

Temuan utama penelitian ini menunjukkan bahwa algoritma Random Forest tidak hanya mampu melakukan klasifikasi status gizi balita secara kuantitatif sangat akurat, tetapi juga secara kualitatif sesuai dengan logika klinis dan praktik kesehatan masyarakat. Artinya, model tidak sekadar menghafal data, tetapi mampu menggeneralisasi pola yang sesuai dengan prinsip antropometri dan pemantauan tumbuh kembang balita

Model belum sepenuhnya layak diterapkan namun Model menunjukkan potensi yang sangat baik, tetapi ada beberapa syarat penting yang harus dipenuhi sebelum implementasi di lapangan.

Syarat-Syarat Kelayakan Implementasi:

- 1) Validasi Eksternal (External Validation) Performa model perlu diuji menggunakan data baru dari sumber atau wilayah lain untuk memastikan bahwa model tidak overfit terhadap data awal.
- 2) Generalisasi Populasi Dataset saat ini harus mewakili berbagai daerah, ras, sosial ekonomi, dan kondisi kesehatan anak di Indonesia.
- 3) Antarmuka Pengguna (UI/UX) Perlu dikembangkan sistem/aplikasi berbasis model ini dengan antarmuka yang mudah digunakan oleh tenaga kesehatan non-teknis.
- 4) Integrasi Standar WHO Model harus mematuhi ambang batas klasifikasi WHO agar hasilnya dapat diterima secara klinis dan dapat dipercaya oleh petugas kesehatan.
- 5) Keamanan dan Etika Data Perlindungan data pribadi anak harus dijamin, terutama jika model digunakan dalam skala nasional atau melalui aplikasi daring.

3.5 Interpretasi Pra Pemrosesan Data

Berikut adalah proses Pra Pemrosesan Data:

- A. Dataset awal terdiri dari 120.999 entri dengan fitur: Umur (bulan), Jenis Kelamin, dan Tinggi Badan (cm), serta target Status Gizi.
- B. Label Encoding diterapkan untuk mengubah nilai kategorikal menjadi numerik:
 - 1) Jenis Kelamin: Laki-laki = 0, Perempuan = 1
 - 2) Status Gizi: normal = 0, severely stunted = 1, stunted = 2, tinggi = 3
- C. StandardScaler digunakan untuk menstandarisasi fitur Umur dan Tinggi Badan agar memiliki rata-rata 0 dan standar deviasi 1
- D. Train_test_split digunakan untuk membagi data menjadi 80% data latih dan 20% data uji.

Berikut adalah Hasil dari Pra Pemrosesan data:

- a) Distribusi data seimbang dan bersih (tidak ada missing value).
- b) Pembagian data dilakukan secara acak, menjaga proporsi kelas tetap seimbang di data latih dan data uji.

Pembahasan dalam tahap Pra-pemrosesan ini memastikan bahwa data numerik dan kategorikal disiapkan dengan baik, meminimalkan bias dalam pelatihan model. Standardisasi sangat penting karena algoritma seperti Random Forest dapat dipengaruhi oleh skala fitur saat menentukan pemisahan pada decision tree.

3.6 Interpretasi Pemodelan

Proses pemodelan dalam penelitian ini dilakukan dengan membangun algoritma Random Forest Classifier untuk memprediksi status gizi balita berdasarkan data antropometri sederhana, yaitu umur (bulan), jenis kelamin, dan tinggi badan (cm). Algoritma ini dipilih karena memiliki karakteristik yang unggul dalam menangani data klasifikasi multikategori, toleran terhadap outlier, serta mampu menangani non-linearitas pada data.

1. Pemilihan Model

Random Forest dipilih karena :

- a) Merupakan ensemble method berbasis pohon keputusan (decision tree) yang menggunakan pendekatan bagging untuk meningkatkan akurasi dan stabilitas.
- b) Secara empiris dan teoretis terbukti handal dalam mengatasi overfitting, terutama ketika jumlah fitur terbatas dan data tidak sepenuhnya linear.
- c) Memungkinkan analisis pentingnya fitur (feature importance), yang relevan dalam konteks penilaian faktor risiko status gizi.

Pemilihan algoritma ini didasarkan pada kebutuhan untuk menghasilkan klasifikasi yang akurat, stabil, dan dapat diinterpretasikan secara praktis untuk digunakan dalam pengambilan keputusan di sektor kesehatan masyarakat.

2. Parameter Model

Model dibangun dengan parameter default scikit-learn, antara lain:

- a) `n_estimators = 100` (jumlah pohon)
- b) `criterion = "gini"` (impuritas pemisahan)
- c) `max_depth = None` (tanpa Batasan kedalaman)
- d) `random_state = 42` (untuk reproduktibilitas)

Penggunaan parameter default memberikan baseline performa yang sudah kuat, dengan 100 pohon cukup untuk menghasilkan stabilitas prediksi, dan tidak dibatasinya kedalaman pohon memungkinkan model belajar dari seluruh struktur data secara optimal

3. Pelatihan dan Validasi

Model dilatih dengan 80% data latih dan diuji pada 20% data uji, serta divalidasi ulang dengan 5-fold cross-validation untuk menguji konsistensi performa antar subset data.

- a) Akurasi: 99,93%
- b) F1-score rata-rata: 0,999
- c) Stabil antar fold cross-validation

Model tidak hanya menunjukkan performa tinggi pada satu kali split data, tetapi juga konsisten pada berbagai skenario pembagian data, yang menandakan kemampuan generalisasi model sangat baik.

4. Analisis Feature Importance

Model menunjukkan bahwa:

- a) Tinggi badan adalah fitur paling dominan dalam klasifikasi status gizi

- b) Umur berkontribusi signifikan
- c) Jenis kelamin berpengaruh kecil

3.7 Interpretasi Evaluasi Model

Evaluasi model merupakan tahap krusial dalam mengukur sejauh mana model machine learning mampu melakukan klasifikasi dengan akurasi dan konsistensi yang tinggi. Pada penelitian ini, model dievaluasi menggunakan beberapa metrik utama, yaitu akurasi, precision, recall, dan F1-score, yang masing-masing memberikan gambaran berbeda terhadap performa model.

a) Akurasi (Accuracy = 99,93%)

Akurasi yang sangat tinggi menunjukkan bahwa model berhasil mengklasifikasikan hampir seluruh data uji secara benar. Dalam konteks klasifikasi status gizi balita, ini berarti bahwa model mampu mengidentifikasi secara tepat kategori status gizi sebagian besar balita, sehingga memiliki nilai praktis tinggi untuk digunakan dalam sistem deteksi dini stunting.

Akurasi mendekati 100% mengindikasikan bahwa pola hubungan antar fitur (umur, jenis kelamin, tinggi badan) yang dipelajari oleh model benar-benar representatif terhadap karakteristik populasi dalam data.

b) Precision, Recall, dan F1-Score

Tabel 3.1 Precision, Recall dan F1-Score

Kelas	Precision	Recall	F1-Score
normal	0.9996	1.00	0.9998
severely stunted	0.9977	0.9992	0.9985
stunted	0.9989	0.9968	0.9979
tinggi	1.00	0.9995	0.9997
Macro Rata-rata	0.9991	0.9989	0.9990

- a) Precision tinggi menunjukkan bahwa prediksi model untuk suatu kelas sangat tepat (minim false positive). Misalnya, jika model memprediksi seorang balita termasuk stunted, kemungkinan besar memang benar
- b) Recall tinggi menunjukkan bahwa model sangat sensitif dalam mendeteksi setiap kelas (minim false negative), artinya sangat kecil kemungkinan model melewati balita yang seharusnya diklasifikasikan sebagai stunted/severely stunted.
- c) F1-score tinggi di semua kelas menunjukkan bahwa model memiliki keseimbangan antara presisi dan sensitivitas, menjadikannya ideal untuk aplikasi di sektor kesehatan, di mana kesalahan klasifikasi berpotensi berdampak besar terhadap intervensi gizi.

3.9 Interpretasi Akurasi Model

Model klasifikasi yang dikembangkan dalam penelitian ini menggunakan algoritma Random Forest Classifier dengan data antropometri sederhana yang terdiri dari umur (bulan), jenis kelamin, dan tinggi badan (cm). Model dilatih pada dataset berjumlah 120.999 data balita yang telah dikategorikan ke dalam empat kelas status gizi, yaitu: Normal, Tinggi, Stunted, dan Severely Stunted.

Setelah proses pelatihan dan evaluasi terhadap data uji, diperoleh hasil bahwa akurasi model mencapai 99,93%, yang berarti bahwa dari seluruh prediksi yang dilakukan oleh model, hanya sekitar 0,07% yang salah klasifikasi

Secara akademik, akurasi mengukur proporsi total prediksi yang benar dibandingkan dengan keseluruhan jumlah observasi. Dengan nilai akurasi yang hampir sempurna, ini menunjukkan bahwa model memiliki kapasitas generalisasi yang sangat tinggi terhadap data baru. Hal ini mencerminkan bahwa pola-pola yang ditangkap oleh algoritma Random Forest dari data pelatihan konsisten dengan karakteristik data uji.

Performa ini diperkuat oleh hasil *confusion matrix*, yang menunjukkan bahwa:

1. Prediksi benar (true positive) mendominasi setiap kelas
2. Kesalahan klasifikasi sangat kecil, dan hanya terjadi antar kelas yang berdekatan secara klinis (misalnya antara *stunted* dan *severely stunted*)

Dengan demikian, nilai akurasi 99,93% tidak hanya mencerminkan kekuatan teknis model, tetapi juga relevansi klinisnya

3.13 Tampilan Aplikasi

Berikut adalah tampilan Aplikasi Prediksi Risiko Stunting pada Balita

Prediksi Risiko Stunting pada Balita

Masukkan data balita untuk memprediksi status gizinya.

Umur (bulan)

0 24 60

Jenis Kelamin

Laki-laki

Tinggi Badan (cm)

75.24 - +

Prediksi

Status Gizi Diprediksi: **STUNTED**

Gambar 6. Aplikasi Prediksi Risiko Stunting pada Balita

Dari hasil prediksi aplikasi menunjukkan bahwa anak balita yang berumur 24 bulan berjenis kelamin laki-laki dengan tinggi badan 75.24 cm tergolong resiko stunting.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi status gizi balita menggunakan algoritma *Random Forest* berbasis data antropometri sederhana, yaitu umur (bulan), jenis kelamin, dan tinggi badan (cm). Dataset yang digunakan terdiri dari 120.999 entri yang dikategorikan ke dalam empat kelas status gizi: normal, tinggi, stunted, dan severely stunted. Model menunjukkan performa klasifikasi yang sangat tinggi dengan, Akurasi keseluruhan sebesar 99,93%. Nilai precision, recall, dan F1-score per kelas yang konsisten tinggi ($\geq 0,997$). Tingkat kesalahan klasifikasi yang sangat rendah, terutama antar kelas yang secara fisiologis berdekatan. Hasil analisis feature importance menunjukkan bahwa tinggi badan merupakan indikator paling dominan dalam klasifikasi status gizi. Hasil analisis feature importance menunjukkan bahwa tinggi badan merupakan indikator paling dominan dalam klasifikasi status gizi. Meskipun hasil yang diperoleh sangat baik, penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Keterbatasan Fitur, data input yang digunakan hanya mencakup tiga fitur utama: umur, jenis

kelamin, dan tinggi badan. Model belum mempertimbangkan variabel lain yang juga berpengaruh terhadap status gizi, seperti berat badan, status imunisasi, pola makan, kondisi ekonomi keluarga, dan faktor lingkungan. Ketidakseimbangan Data, meskipun kinerja model tinggi, distribusi jumlah data pada masing-masing kelas tidak sepenuhnya seimbang, yang dapat memengaruhi generalisasi model pada kelas minoritas jika diterapkan pada populasi berbeda. Keterbatasan Generalisasi Konteks Wilayah, dataset bersumber dari data Dinas Kesehatan dan E-PPGBM tahun 2022, yang kemungkinan besar merepresentasikan wilayah tertentu. Oleh karena itu, validitas eksternal model perlu diuji lebih lanjut pada wilayah lain dengan kondisi demografis yang berbeda. Belum Dilakukan Uji Lapangan, model belum diuji dalam konteks implementasi nyata di layanan kesehatan masyarakat seperti posyandu atau puskesmas, sehingga efektivitas operasionalnya dalam praktik masih perlu dibuktikan.

DAFTAR PUSTAKA

- [1] Fatmawati, M., Herlambang, B. A., & Nada, N. Q. (2024). Random forest algorithm for toddler nutritional status classification website. *Journal of Applied Informatics and Computing*, 8(2), 428-433.
- [2] Gustriansyah, R., Suhandi, N., Puspasari, S., & Sanmorino, A. (2024). Machine learning method to predict the toddlers' nutritional status. *Indonesian Journal of Artificial Intelligence and Data Mining*, 7(1), 23-32.
- [3] Reza, A. A., & Rohman, M. S. (2024). Prediction stunting analysis using random forest algorithm and random search optimization. *Journal of Informatics and Telecommunication Engineering*, 7(2), 110-117.
- [4] Lestari, W. S., Saragih, Y. M., & Caroline. (2024). Comparison of deep neural networks and random forest algorithms for multiclass stunting prediction in toddlers. *Teknika*, 13(3), 412-417.
- [5] Marpaung, S. H., Handoko, E., & Anggraeni, I. R. (2024). Random forest optimization using recursive feature elimination for stunting classification. *Indonesian Journal of Artificial Intelligence and Data Mining*, 7(1), 45-54.
- [6] Dewi, Y. S., Hastuti, S., & Fatekurohman, M. (2024). Analysis of stunting in East Java, Indonesia using random forest and geographically weighted random forest regression. *Brazilian Journal of Biometrics*, 42(3), 213-224.
- [7] Hasdyna, N., Fitriani, Y., & Sari, D. P. (2024). Hybrid machine learning for stunting prevalence: Classification, prediction, and clustering in Aceh, Indonesia. *Informatics*, 11(4), 89.
- [8] Rao, B., Rashid, M., Hasan, M. G., & Thunga, G. (2025). Machine learning in predicting child malnutrition: A meta-analysis of demographic and health surveys data. *International Journal of Environmental Research and Public Health*, 22(3), 449.
- [9] Wara, T. U., Mahmud, T., & Joy, C. D. (2024). Risk factor identification and classification of malnutrition among under-five children in Bangladesh: Machine learning and statistical approach. *arXiv preprint*.
- [10] Pramana, A. A. G. Y., Zidan, H. M., Maulana, M. F., & Natan, O. (2024). ESDS: AI-powered early stunting detection and monitoring system using Edited Radius-SMOTE algorithm. *arXiv preprint*.
- [11] Nishimwe, B. A., Munyanshongore, C., & Hakizimana, J. (2023). Stunting prediction in under-five children using logistic regression and random forest: Evidence from Rwanda. *African Journal of Health Sciences*, 36(1), 45-56.
- [12] Yusuf, S., & Abdurrahman, A. (2023). Predicting child undernutrition using ensemble machine learning: A case from Indonesia. *International Journal of Health Sciences and Research*, 13(2), 87-94.
- [13] Children. (2023). Machine learning algorithms for predicting stunting among under-five children in Papua New Guinea. *Children*, 10(10), 1638.

-
- [14] Lestari, W. S., Saragih, Y. M., & Caroline. (2025). Multiclass classification for stunting prediction using deep neural networks. *Jurnal Ilmu Teknologi Komputer*, 10(2), 386–393.
 - [15] Hossain, M., Ahmed, S., & Huda, T. (2022). Predictive modeling of child malnutrition using artificial intelligence techniques: A systematic review. *Journal of Public Health Informatics*, 14(1), e123.
 - [16] Mundirin, Mundirin, and Joko Prasetiana. "The tidal station management application (PAS)." *Journal of Information System, Applied, Management, Accounting and Research* 6.1 (2022): 232-239.