

# Perbandingan Algoritma Random Forest dan XGBoost dalam Klasifikasi Status Stunting pada Balita Berdasarkan Data Antropometri

Pramagita Dewandra<sup>1</sup>, I Gede Susrama Mas Diyasa<sup>2\*</sup>, Ani Dijah Rahajoe<sup>3</sup>

<sup>1,2,3</sup> Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur  
<sup>1</sup>21081010207@student.upnjatim.ac.id

<sup>3</sup>anidijah.if@upnjatim.ac.id

\*Corresponding author email: [igsusrama.if@upnjatim.ac.id](mailto:igsusrama.if@upnjatim.ac.id)

**Abstrak**— Stunting merupakan permasalahan gizi kronis yang berdampak pada pertumbuhan dan perkembangan anak, sehingga deteksi dini menjadi sangat penting. Penelitian ini bertujuan membangun model klasifikasi status stunting pada balita menggunakan pendekatan *machine learning* berbasis data antropometri. Data yang digunakan berasal dari dataset balita di Kabupaten Jeneponto dengan total 40.071 data, meliputi jenis kelamin, usia (bulan), berat badan, dan tinggi badan. Status stunting ditentukan berdasarkan *Z-Score Height-for-Age* sesuai standar World Health Organization. Metode yang digunakan meliputi tahap *preprocessing* data, pembentukan variabel target, pembagian data, serta pembangunan model menggunakan algoritma Random Forest dan XGBoost. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score*. Hasil penelitian menunjukkan bahwa model XGBoost memiliki performa terbaik dengan *accuracy* sebesar 96,81% dan *recall* sebesar 95% pada kelas stunting, sedangkan Random Forest menghasilkan *accuracy* sebesar 93,34%. Tingginya performa model menunjukkan kemampuan dalam menangkap pola hubungan antar variabel antropometri, khususnya tinggi badan dan usia. Penelitian ini menyimpulkan bahwa *machine learning* dapat digunakan sebagai alat bantu *screening* awal stunting secara cepat. Namun, model masih memiliki keterbatasan dalam mendeteksi seluruh kasus stunting, sehingga pengembangan lebih lanjut diperlukan untuk meningkatkan akurasi dan generalisasi.  
**Kata Kunci**— Stunting, *Machine learning*, Klasifikasi, Random Forest, XGBoost, Data Antropometri.

badan, dan usia [6, 7]. Akibatnya, metode konvensional kurang sensitif dalam menangkap pola risiko halus yang sebenarnya bisa terdeteksi lebih awal melalui analisis multivariat [8].

Algoritma *ensemble* seperti Random Forest dan XGBoost menawarkan solusi dengan kemampuan mengolah kombinasi variabel secara adaptif dan menangkap pola non-linear yang luput dari metode regresi sederhana [8, 9]. Studi di berbagai negara berkembang menunjukkan bahwa model *machine learning* mampu mencapai akurasi tinggi dalam memprediksi *undernutrition* [8, 10]. Namun, mayoritas penelitian tersebut sangat bergantung pada variabel sosio-ekonomi (pendidikan ibu, pendapatan, kondisi rumah tangga) yang datanya sering kali tidak tersedia secara *real-time* di layanan kesehatan primer seperti Posyandu atau Puskesmas [8, 9].

Celah penelitian ini terletak pada ketergantungan literatur saat ini terhadap variabel eksternal yang kompleks dan kurangnya transparansi terkait risiko *label leakage* (kebocoran data) serta *overfitting*. Banyak studi melaporkan akurasi ekstrem namun menggunakan *pipeline* yang sulit direplikasi atau tidak relevan dengan konteks data minimal di lapangan [11, 12]. Penelitian ini bertujuan mengisi celah tersebut dengan mengevaluasi secara ketat kemampuan model *machine learning* dalam memprediksi status gizi hanya menggunakan fitur antropometri minimal (umur, jenis kelamin, berat, dan tinggi). Hal ini krusial untuk menghasilkan model yang *reproducible* dan praktis untuk deteksi dini stunting di fasilitas kesehatan dengan sumber daya terbatas.

## I. PENDAHULUAN

Stunting adalah kegagalan pertumbuhan anak akibat malnutrisi kronis yang didefinisikan secara klinis jika tinggi badan berada >2 standar deviasi di bawah median standar WHO [1]. Indonesia merupakan salah satu *high-burden country* dengan beban stunting tertinggi kelima di dunia, yang berdampak langsung pada kegagalan pencapaian SDG 2 (*Zero Hunger*) [2, 3]. Kegagalan intervensi stunting tidak hanya berdampak pada fisik, tetapi juga hilangnya potensi intelektual dan kerugian ekonomi nasional akibat penurunan produktivitas individu sebesar 5–53% [4, 5].

Selama ini, klasifikasi status gizi mengandalkan perhitungan *Z-Score* yang membandingkan indikator tunggal terhadap standar WHO. Pendekatan ini bersifat deterministik dan kaku karena hanya berbasis ambang batas tunggal (*threshold*), sehingga gagal memodelkan variasi alami pola pertumbuhan maupun interaksi non-linear yang kompleks antara berat badan, tinggi

## II. TINJAUAN PUSTAKA

### A. Penelitian Terdahulu

Penelitian terdahulu menjadi landasan untuk memetakan perkembangan algoritma *machine learning* dalam prediksi stunting serta mengidentifikasi celah metodologis yang ada.

1) Shen et al. (Papua Nugini) mengevaluasi berbagai model ML menggunakan data survei kesehatan nasional. Model terbaik, kombinasi LASSO + XGBoost, menghasilkan AUC 0,765. Variabel prediktor utama meliputi wilayah, status ekonomi, dan ukuran saat lahir (*birth size*). Namun, ketergantungan pada variabel demografis yang luas dan data tahun 2016–2018 membatasi aplikabilitas model pada skenario data antropometri minimal saat ini [9].

2) Anku & Duah (Ghana) menggunakan 7 model ML pada 8.564 sampel data MICS. XGBoost melaporkan performa

ekstrem dengan akurasi hingga 98% dan AUC 100% pada indikator *wasting*. Tingginya angka ini mengindikasikan adanya hubungan asosiatif yang sangat kuat antara variabel input (berat/tinggi) dengan label, namun sekaligus memicu kecurigaan adanya *overfitting* atau *label leakage* yang tidak dieksplorasi lebih lanjut oleh peneliti [13].

3) Sahamony et al. (Indonesia) melakukan studi komparatif yang menempatkan Random Forest sebagai model terbaik dengan akurasi 99,95%. Meskipun angkanya impresif, penelitian ini memiliki kelemahan pada transparansi *pipeline* data dan detail populasi. Performa yang mendekati sempurna tanpa pengujian pada data independen menimbulkan pertanyaan besar mengenai generalisasi model pada populasi eksternal yang lebih heterogen [11].

4) Sugihartono et al. (Indonesia) mencoba mengatasi ketidakseimbangan kelas (*imbalanced data*) menggunakan teknik *SMOTE* pada 10.000 sampel. Model XGBoost menghasilkan akurasi 0,8574 dan AUC 0,9311. Meskipun penggunaan data sintetis meningkatkan *recall*, terdapat risiko bahwa model menjadi terlalu optimis dan tidak merefleksikan distribusi risiko nyata di lapangan yang lebih kompleks [12]. Secara kolektif, literatur saat ini menunjukkan bahwa XGBoost dan Random Forest adalah algoritma yang paling kompetitif. Namun, terdapat tren pelaporan akurasi yang tidak realistis (di atas 98%) dan ketergantungan pada data sintetis (*SMOTE*) serta variabel sosio-ekonomi yang sulit didapat di layanan primer. Hal ini memperkuat urgensi penelitian ini untuk menguji stabilitas model hanya dengan fitur antropometri minimal serta pengawasan terhadap *label leakage*.

#### B. Landasan Teori

- 1) *Balita dan Status Gizi*: Balita (0–59 bulan) adalah kelompok rentan yang pertumbuhan fisiknya menjadi indikator utama kesehatan masyarakat. Status gizi balita ditentukan melalui *Z-Score* antropometri berdasarkan standar WHO dan Permenkes No. 2 Tahun 2020 [1, 2]. Indikator utama meliputi BB/U (*underweight*), TB/U (*stunting*), dan BB/TB (*wasting*). *Stunting* didefinisikan secara spesifik sebagai kondisi malnutrisi kronis dengan nilai *Height-for-Age Z-Score* (HAZ)  $< -2$  SD. Dalam penelitian ini, data antropometri (umur, jenis kelamin, BB, TB) digunakan sebagai fitur input untuk mengevaluasi apakah pola pertumbuhan minimal cukup untuk memetakan kategori status gizi secara akurat.
- 2) *Machine learning dan Klasifikasi*: *Machine learning* (ML) melalui pendekatan *supervised learning* memungkinkan komputer mempelajari pola non-linear dari data berlabel untuk tugas klasifikasi [5]. Penelitian ini memformulasikan penentuan status gizi sebagai masalah klasifikasi multi-kelas, memetakan fitur antropometri ke label status gizi sesuai standar WHO.
- 3) *Algoritma Ensemble*: Random Forest dan XGBoost
  - a) Random Forest (RF) adalah algoritma yang menggunakan teknik *bagging* (*Bootstrap Aggregating*) untuk membangun sekumpulan pohon keputusan yang independen. Prediksi akhir diambil melalui *majority voting* (klasifikasi) atau rata-rata (regresi). Random Forest efektif mengurangi

*overfitting* dan menangani interaksi kompleks antar variabel tanpa asumsi distribusi normal.

- b) XGBoost (Extreme Gradient Boosting) adalah Algoritma *boosting* yang membangun pohon secara bertahap untuk mengoreksi kesalahan model sebelumnya menggunakan optimasi *gradient descent* dan regularisasi untuk mencegah *overfitting*.
- 4) *Evaluasi Performa*: Performa model dievaluasi menggunakan *Confusion matrix* untuk menghitung metrik utama:

*Accuracy*: Proporsi total prediksi benar.

*Precision*: Kemampuan model menghindari false positive.

*Recall*: Kemampuan model menangkap seluruh kasus positif

*F1-score*: Rata-rata harmonik *precision* dan *recall*, memberikan gambaran objektif pada data kelas tidak seimbang.

### III. METODE PENELITIAN

#### A. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen pada bidang *data mining* dan *machine learning*. Tujuan utama penelitian adalah membangun model klasifikasi untuk memprediksi status *stunting* pada balita berdasarkan data antropometri. Model yang digunakan dalam penelitian ini adalah algoritma Random Forest dan XGBoost, yang termasuk dalam kategori *supervised learning*.

Pendekatan eksperimen dilakukan dengan membandingkan performa kedua algoritma dalam mengklasifikasikan status *stunting* berdasarkan data yang tersedia. Evaluasi dilakukan menggunakan metrik performa seperti *accuracy*, *precision*, *recall*, dan *f1-score*.

#### B. Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari dataset *Stunting and Nutritional Status of Toddler from Jeneponto Regency, South Sulawesi, Indonesia*. Dataset ini mencakup 40.071 data balita usia 0–59 bulan yang dikumpulkan di periode tahun 2021 hingga 2024 [14]. Data yang digunakan telah melalui proses *preprocessing* awal dan tersedia dalam format Microsoft Excel (.xlsx). Variabel yang digunakan dalam penelitian ini meliputi gender, age (month), weight (kg), height (cm), dan *Z-Score* Height-for-Age (H/A). Variabel target (label) dalam penelitian ini adalah status *stunting* yang ditentukan berdasarkan standar World Health Organization, yaitu balita dikategorikan *stunting* apabila nilai *Z-Score* Height-for-Age kurang dari -2. Adapun contoh dataset bisa dilihat pada Tabel 1 berikut ini.

**Tabel 1** Contoh Dataset *Stunting*

No	Gender	Age (Month)	Weight	Height	Z-Score H/A
1	0	54	13.2	97.5	-2.11
2	1	44	12	2	-2.22
3	1	57	14	97	-2.58
4	1	26	11	79	-3.11
5	0	52	13.4	96.5	-1.94
6	0	45	11	90	-287

### C. Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis mulai dari pengolahan data hingga evaluasi model. Berikut adalah tahapan yang dilakukan:

- 1) *Pengumpulan Data*: Data diimpor ke dalam lingkungan Jupyter Notebook menggunakan bahasa pemrograman Python dengan bantuan pustaka *pandas*. Pada tahap ini dilakukan eksplorasi awal terhadap struktur data, tipe data, serta identifikasi potensi permasalahan seperti *missing value* dan inkonsistensi format.
- 2) *Data Preprocessing*: Tahap *preprocessing* merupakan bagian krusial dalam penelitian ini, mengingat ditemukan adanya inkonsistensi format data, khususnya pada variabel numerik. Langkah-langkah *preprocessing* yang dilakukan meliputi:
  - a) *Pembersihan Format Data*: Data numerik seperti berat badan, tinggi badan, usia, dan *Z-Score* dikonversi dari format string menjadi numerik. Proses ini mencakup: Penghapusan karakter non-numerik (seperti "kg" dan spasi), Konversi tanda koma menjadi titik desimal, dan Ekstraksi nilai numerik menggunakan *regular expression*
  - b) *Penanganan Missing value*: Nilai yang tidak dapat dikonversi menjadi numerik akan diubah menjadi nilai kosong (NaN). Selanjutnya, data yang memiliki nilai kosong pada variabel penting dihapus untuk menjaga kualitas dataset.
  - c) *Pembentukan Variabel Target*: Variabel target "Stunting" dibuat berdasarkan nilai *Z-Score Height-for-Age* dengan aturan Stunting (1):  $Z\text{-Score} < -2$  dan Tidak stunting (0):  $Z\text{-Score} \geq -2$ .
  - d) *Encoding Data Kategorikal*: Variabel gender diubah menjadi numerik menggunakan teknik *label encoding*.
  - e) *Feature engineering*: Dibuat variabel tambahan berupa *Body Mass Index (BMI)* yang dihitung menggunakan rumus:  $BMI = \text{Weight} / (\text{Height}/100)^2$ .

Tahapan *preprocessing* menghasilkan dataset bersih dengan jumlah data sebesar 40.069 record yang siap digunakan untuk proses pemodelan.

- 3) *Pembagian Data*: Dataset dibagi menjadi dua bagian, yaitu data latih (*training*) dan data uji (*testing*) dengan perbandingan 80:20. Pembagian dilakukan menggunakan metode *stratified sampling* untuk menjaga proporsi kelas stunting tetap seimbang pada kedua subset data.
- 4) *Pencegahan Label Leakage*: Untuk mencegah terjadinya *label leakage* (kebocoran data dari fitur ke target), penelitian ini menerapkan beberapa langkah: Variabel *Z-Score Height-for-Age* tidak digunakan sebagai fitur input, melainkan hanya untuk membentuk variabel target. Proses *feature engineering* (penambahan BMI) dilakukan pada data training setelah split, bukan sebelum split, untuk menghindari kebocoran informasi dari data uji ke proses pembuatan fitur. *Stratified sampling* pada pembagian data memastikan proporsi kelas tetap, namun tidak mencampur informasi antar set.

- 5) *Pembangunan Model*: Model yang digunakan adalah dua algoritma *ensemble*, yaitu:
  - a) *Random Forest*: Model ini merupakan metode *ensemble* berbasis *decision tree* yang bekerja dengan membangun banyak pohon keputusan dan menggabungkan hasilnya untuk meningkatkan akurasi dan mengurangi *overfitting*.
  - b) *XGBoost*: Model ini merupakan algoritma *boosting* yang mengoptimalkan performa model secara iteratif dengan memperbaiki kesalahan dari model sebelumnya. XGBoost dikenal memiliki performa tinggi di berbagai klasifikasi.
- 6) *Evaluasi Model*: Evaluasi model dilakukan menggunakan *confusion matrix* dan matrik klasifikasi, yaitu: *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Fokus utama evaluasi dalam penelitian ini adalah nilai *recall* pada kelas stunting, karena kesalahan dalam mendeteksi kasus stunting (*false negative*) memiliki dampak yang lebih serius dibandingkan kesalahan lainnya.

## IV. HASIL PENELITIAN

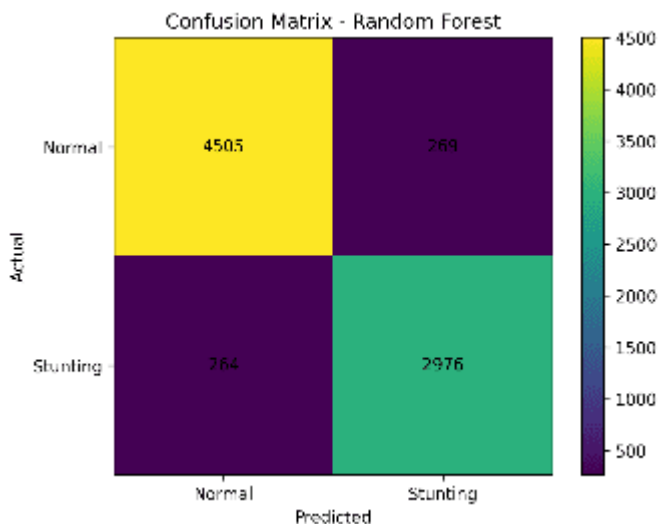
### A. Hasil Penelitian

- 1) *Hasil Preprocessing Data*: Tahap *preprocessing* menghasilkan dataset bersih sebanyak 40.069 data dari total awal 40.071 data. Proses ini mencakup pembersihan format numerik, penanganan *missing value*, serta pembentukan variabel target berdasarkan standar *World Health Organization*. Distribusi data setelah *preprocessing* menunjukkan bahwa 23.867 data termasuk kategori tidak stunting dan 16.202 data termasuk kategori stunting. Distribusi ini menunjukkan bahwa dataset relatif seimbang dan layak digunakan untuk proses klasifikasi tanpa memerlukan teknik penyeimbangan tambahan.
- 2) *Hasil Pembagian Data*: Data dibagi menjadi 2 bagian yaitu data latih: 80% dan data uji: 20%. Pembagian dilakukan menggunakan *stratified sampling* sehingga proporsi kelas tetap terjaga pada kedua subset data.
- 3) *Hasil Pengujian Model*: Model *Random Forest* menghasilkan *Accuracy*: 93.34%, *Precision*: 0.92, *Recall*: 0.92, dan *F1-score*: 0.92. Sedangkan model *XGBoost* menunjukkan performa yang lebih tinggi yaitu *Accuracy*: 96.81%, *Precision*: 0.97, *Recall*: 0.95, dan *F1-score*: 0.96. Seperti yang bisa dilihat pada Tabel 2 berikut ini.

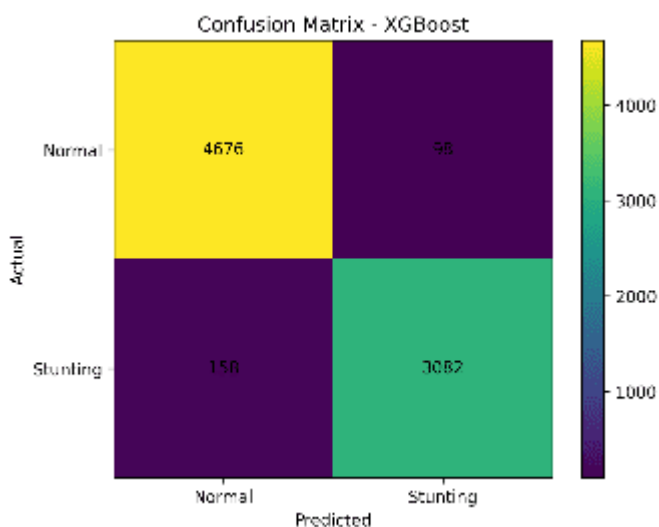
**Tabel 2** Perbandingan Performa Model

Model	Matriks Evaluasi			
	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Random Forest	93.34%	0.92	0.92	0.92
XGBoost	96.80%	0.97	0.95	0.96

- 4) *Hasil Confusion matrix*: Berdasarkan *confusion matrix*, model *Random Forest* dan *XGBoost* mampu mengklasifikasikan sebagian besar data dengan benar dan meminimalkan kesalahan klasifikasi, khususnya pada kategori stunting. Namun, masih terdapat sejumlah kecil kasus *false negative*, yaitu balita stunting yang diprediksi sebagai tidak stunting. Adapun detail bisa dilihat pada Gambar 1 dan Gambar 2 berikut ini.



Gambar 1 Grafik Confusion Matric Random Forest



Gambar 2 Grafik Confusion matrix XGBoost

## B. Pembahasan

- 1) *Analisis Performa Model*: Hasil menunjukkan bahwa model XGBoost memiliki performa yang lebih unggul dibandingkan Random Forest pada seluruh metrik evaluasi. Hal ini sejalan dengan karakteristik XGBoost yang mampu melakukan optimasi bertahap terhadap kesalahan model sebelumnya. Namun demikian, perbedaan performa antara kedua model tidak terlalu ekstrem, yang menunjukkan bahwa kedua algoritma sama-sama efektif dalam menangkap pola pada data.
- 2) *Analisis Recall pada Kasus Stunting*: Dalam konteks kesehatan, metrik yang paling penting bukanlah *accuracy*, melainkan *recall* pada kelas stunting. Model XGBoost memiliki *recall* sebesar 0.95, yang berarti sekitar 95% kasus stunting berhasil dideteksi oleh model, sementara 5% lainnya tidak terdeteksi (*false negative*). Kesalahan ini memiliki implikasi yang signifikan, karena balita yang tidak terdeteksi berisiko tidak mendapatkan intervensi yang diperlukan.
- 3) *Analisis Hubungan Fitur terhadap Target*: Model yang dibangun menggunakan variabel antropometri dasar, yaitu usia, tinggi badan, dan berat badan. Secara konseptual, status stunting ditentukan berdasarkan rasio tinggi badan terhadap usia (*Height-for-Age*), yang juga digunakan dalam perhitungan *Z-Score* oleh World Health Organization. Hal ini menunjukkan bahwa model *machine learning* pada penelitian ini pada dasarnya adalah mempelajari kembali pola yang telah diformalkan dalam standar WHO. Dengan kata lain, performa tinggi yang diperoleh bukan semata-mata karena kemampuan model menemukan pola baru, tetapi karena model berhasil merepresentasikan hubungan matematis yang sudah ada.
- 4) *Analisis Peran Feature engineering (BMI)*: Penambahan fitur BMI bertujuan untuk memperkaya informasi terkait kondisi tubuh balita. Namun, berdasarkan analisis konseptual, BMI bukan indikator utama dalam menentukan stunting, karena stunting lebih berkaitan dengan tinggi badan terhadap usia. Dengan demikian, peran BMI dalam model kemungkinan bersifat tambahan dan tidak dominan dibandingkan fitur lain seperti tinggi badan dan usia.
- 5) *Analisis Mitigasi Label Leakage*: Sebagaimana disorot dalam pendahuluan, banyak penelitian melaporkan akurasi di atas 98% tanpa transparansi terkait potensi *label leakage*. Penelitian ini secara sadar menghindari hal tersebut dengan mengecualikan *Z-score* dari fitur. Hasil yang lebih rendah (namun realistis) dari XGBoost (96,81%) justru mengindikasikan bahwa model bekerja pada data yang “bersih” dari kebocoran. Jika *Z-score* dijadikan fitur, akurasi akan mendekati 100%, namun itu tidak bermakna secara klinis. Dengan demikian, celah transparansi yang diidentifikasi telah diatasi melalui protokol yang dijelaskan di subbab Metode.
- 6) *Keterbatasan Penelitian*: Penelitian ini memiliki beberapa keterbatasan, antara lain:
  - a) Ketergantungan pada Variabel Antropometri: Model hanya menggunakan data fisik (tinggi, berat, usia), tanpa mempertimbangkan faktor lain seperti pola makan, kondisi ekonomi, atau riwayat kesehatan.
  - b) Kemiripan dengan Standar WHO: Karena target ditentukan berdasarkan *Z-Score*, model cenderung mereplikasi aturan yang sudah ada, sehingga kontribusi inovatif model menjadi terbatas.
  - c) Metode Pembagian Data: Pembagian data dilakukan secara acak (*random split*), sehingga belum menguji kemampuan model dalam menghadapi data dari periode waktu yang berbeda.
- 7) *Implikasi Penelitian*: Meskipun memiliki keterbatasan, model yang dihasilkan tetap memiliki nilai praktis, terutama dalam kondisi ketika data *Z-Score* tidak tersedia, diperlukan klasifikasi cepat di lapangan, atau ketika sistem otomatis dibutuhkan untuk *screening* awal. Model ini dapat digunakan sebagai alat bantu dalam mendeteksi potensi stunting secara cepat berdasarkan data sederhana.

## V. KESIMPULAN

Berdasarkan serangkaian eksperimen dan analisis yang telah dilakukan, beberapa kesimpulan utama yang dapat diambil adalah sebagai berikut:

1. Model XGBoost *outperforms Random Forest* dengan akurasi 96,81% dan *recall* kelas stunting sebesar 95%, menjadikannya model terbaik untuk deteksi dini stunting berbasis data antropometri (usia, jenis kelamin, berat badan, tinggi badan).
2. Penelitian ini telah mengatasi risiko *label leakage* dengan tidak menggunakan *Z-score* sebagai fitur input, sehingga hasil performa mencerminkan kemampuan generalisasi model yang realistis.
3. Meskipun model memiliki akurasi tinggi, masih terdapat sekitar 5% kasus stunting yang tidak terdeteksi (*false negative*), sehingga perlu pengembangan lebih lanjut dengan menambahkan variabel eksternal (sosial ekonomi, riwayat kesehatan) serta evaluasi berbasis waktu (*time-based split*).

## UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada seluruh pihak yang mendukung penelitian ini sehingga berhasil dilaksanakan dengan baik.

## REFERENSI

- [1] M. Mardiaty, F. Maulina, dan M. Sayuti, "HUBUNGAN INTERPRETASI WHO (WORLD HEALTH ORGANIZATION) ANTROPOMETRI Z-SCORE DAN INFEKSI KECACINGAN PADA ANAK USIA 36 – 60 BULAN DI KOTA LHKSEUMAWA," *AVERROUS J. Kedokt. Dan Kesehat. Malikussaleh*, vol. 6, no. 2, hlm. 44, Des 2020, doi: 10.29103/averrous.v6i2.3325.
- [2] C. R. Titaley, I. Ariawan, D. Hapsari, A. Muasyaroh, dan M. J. Dibley, "Determinants of the Stunting of Children Under Two Years Old in Indonesia: A Multilevel Analysis of the 2013 Indonesia Basic Health Survey," *Nutrients*, vol. 11, no. 5, hlm. 1106, Mei 2019, doi: 10.3390/nu11051106.
- [3] T. Beal, A. Tumilowicz, A. Sutrisna, D. Izwardy, dan L. M. Neufeld, "A review of child stunting determinants in INDONESIA," *Matern. Child. Nutr.*, vol. 14, no. 4, hlm. e12617, Okt 2018, doi: 10.1111/mcn.12617.
- [4] A. Komaruzaman dkk., "Achieving Zero Stunting: A Sustainable Development Goal Interlinkage Approach at District Level," *Sustainability*, vol. 15, no. 11, hlm. 8890, Mei 2023, doi: 10.3390/su15118890.
- [5] M. Shekar dkk., "Reaching the global target to reduce stunting: an investment framework," *Health Policy Plan.*, vol. 32, no. 5, hlm. 657–668, Jun 2017, doi: 10.1093/heapol/czw184.
- [6] "Use and interpretation of anthropometric indicators of nutritional status," *Bull. World Health Organ.*, vol. 64, no. 6, hlm. 929–941, 1986.
- [7] R. Roediger, D. T. Hendrixson, dan M. J. Manary, "A roadmap to reduce stunting," *Am. J. Clin. Nutr.*, vol. 112, hlm. 773S–776S, Sep 2020, doi: 10.1093/ajcn/nqaa205.
- [8] E. K. Anku dan H. O. Duah, "Predicting and identifying factors associated with *undernutrition* among children under five years in Ghana using *machine learning* algorithms," *PLOS ONE*, vol. 19, no. 2, hlm. e0296625, Feb 2024, doi: 10.1371/journal.pone.0296625.
- [9] H. Shen, H. Zhao, dan Y. Jiang, "*Machine learning* Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea," *Children*, vol. 10, no. 10, hlm. 1638, Sep 2023, doi: 10.3390/children10101638.
- [10] A. B. Zemariam dkk., "Prediction of stunting and its socioeconomic determinants among adolescent girls in Ethiopia using *machine learning* algorithms," *PLOS ONE*, vol. 20, no. 1, hlm. e0316452, Jan 2025, doi: 10.1371/journal.pone.0316452.
- [11] N. F. Sahamony, T. Terttiaavini, dan H. Rianto, "Analisis Perbandingan Kinerja Model *Machine learning* untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak: Analysis of Performance Comparison of *Machine learning* Models for Predicting Stunting Risk in Children's Growth," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, hlm. 413–422, Feb 2024, doi: 10.57152/malcom.v4i2.1210.
- [12] T. Sugihartono, "Optimizing Stunting Detection through *SMOTE* and *Machine learning*: a Comparative Study of XGBoost, Random Forest, SVM, and k-NN," *J. Appl. Data Sci.*, vol. 6, no. 1, hlm. 667–682, Jan 2024, doi: 10.47738/jads.v6i1.494.
- [13] E. K. Anku dan H. O. Duah, "Predicting and identifying factors associated with *undernutrition* among children under five years in Ghana using *machine learning* algorithms," *PLOS ONE*, vol. 19, no. 2, hlm. e0296625, Feb 2024, doi: 10.1371/journal.pone.0296625.
- [14] RS, Asmaul Husna; Purnamawati, Purnamawati; Jaya, Hendra; Arfandi, Anas; Suhardi, Iwan; Widodo, Slamet; Lonang, Syahrani (2026), "Dataset Stunting and Nutritional Status of Toddler from Jeneponto Regency, South Sulawesi, Indonesia", Mendeley Data, V4, doi: 10.17632/wzwpc9j5bx.4