



## Algoritma Random Forest, Decision Tree dan XGboost Untuk Klasifikasi Stunting Pada Balita

Dhika Malita Puspita<sup>1\*</sup>, Kartika Imam Santoso<sup>2</sup>, Andri Triyono<sup>3</sup>, Eko Supriyadi<sup>4</sup>, Agus S Nugroho<sup>5</sup>, Edi Widodo<sup>6</sup>

<sup>1</sup>Program Studi Ilmu Komputer, Universitas An Nuur

Jalan Gajah Mada No. 7, Purwodadi, Kab Grobogan 58112, e-mail: dhika.malita11@gmail.com

<sup>2</sup>Program Studi Ilmu Komputer, Universitas An Nuur

Jalan Gajah Mada No. 7, Purwodadi, Kab Grobogan 58112, e-mail: kartikaimams@gmail.com

<sup>3</sup>Program Studi Ilmu Komputer, Universitas An Nuur

Jalan Gajah Mada No. 7, Purwodadi, Kab Grobogan 58112, e-mail: andritriyono1@gmail.com

<sup>4</sup>Program Studi Ilmu Komputer, Universitas An Nuur

Jalan Gajah Mada No. 7, Purwodadi, Kab Grobogan 58112, e-mail: ekalaya56@gmail.com

<sup>5</sup>Program Studi Ilmu Komputer, Universitas An Nuur

Jalan Gajah Mada No. 7, Purwodadi, Kab Grobogan 58112, e-mail: nugro333@gmail.com

<sup>6</sup>Universitas Semarang

Jl. Sukarno Hatta Semarang, e-mail: ediwidodo@usm.ac.id

### ARTICLE INFO

#### *History of the article :*

Received 13 Juni 2025

Received in revised form 27 Juni 2025

Accepted 12 Juli 2025

Available online 14 Juli 2025

#### **Keywords:**

*Comparing, Decision Tree, Random orest, Stunting, XGBoost*

#### **\* Correspondence:**

Telepon:

+62 857 2692 5616

E-mail: [dhika.malita11@gmail.com](mailto:dhika.malita11@gmail.com)

### ABSTRACT

At toddler age, they require special attention, as this is the period when the brain develops up to 80%.

Stunting is a form of chronic malnutrition that affects a child's growth and development. According to WHO standards, it is characterised by height that is below or lower than that of peers. This condition has negative impacts on cognitive development and overall health. Identifying toddlers at risk of stunting early is crucial to minimising the negative effects that could impact their quality of life in the future. Traditional methods are less effective in predicting stunting because they often overlook the complex factors that influence an infant's nutritional status. This study aims to compare three algorithms and identify the most effective one for analysing infant stunting data. The method used involves comparing the results of the Random Forest, Decision Tree, and Extreme Gradient Boost (XGBoost) algorithms. The results obtained show that the Random Forest algorithm achieved the highest accuracy at 99.72%, Extreme Gradient Boost (XGBoost) at 99.58%, and Decision Tree at 98.87%.

## 1. INTRODUCTION

Usia balita merupakan masa krusial, karena pada masa tersebut balita mengalami pertumbuhan fisik dan psikomotorik yang signifikan. Balita memerlukan perhatian khusus karena pada masa tersebut otak mereka berkembang sekitar 80%. Balita perlu memiliki pola makan yang

seimbang selama fase ini untuk memastikan mereka memperoleh kesehatan gizi yang baik [1] [2]. Stunting merupakan masalah gizi kronis, dan salah satu bentuk kekurangan gizi jangka panjang yang terjadi pada waktu masa pertumbuhan dan perkembangan anak. Stunting ditandai dengan tinggi badan yang tidak sesuai atau kurang dibandingkan anak seusianya berdasarkan standar dari WHO. Kondisi ini dapat berdampak buruk pada perkembangan kognitif dan kesehatan anak selanjutnya [3] [4].

Masalah stunting pada anak balita masih menjadi masalah yang signifikan di negara-negara belum maju. Menurut Kementerian Kesehatan, angka stunting pada balita Indonesia dilaporkan sebesar 21,6% pada tahun 2022. Meskipun angka tersebut mengalami penurunan dari tahun-tahun sebelumnya, pendekatan strategis tetap penting untuk memenuhi tujuan pemerintah untuk menurunkan prevalensi stunting naik menjadi 14% pada tahun 2024 [5]. Mengidentifikasi anak kecil yang berisiko mengalami stunting pada tahap awal penting sekali untuk mengurangi dampak buruk yang dapat mempengaruhi kualitas hidup mereka di masa mendatang. Di lapangan petugas kesehatan sering kali mengandakan proses manual untuk mengklasifikasikan dan menilai status gizi balita yang mengalami stunting. Metode tradisional yang masih manual tersebut masih cenderung kurang efektif untuk memprediksi stunting. Metode tersebut sering kali tidak konsisten tiap petugas dalam menilai serta mengabaikan faktor-faktor yang mempengaruhi status gizi balita. Untuk mengatasi masalah tersebut bisa dengan memanfaatkan teknologi berbasis data dengan Data Mining. Metode ini bisa menjanjikan dan lebih akurat untuk meningkatkan ketepatan mengidentifikasi balita yang berisiko mengalami stunting [6].

Data Mining merupakan teknik komputasi yang bertujuan untuk mengekstraksi informasi penting dan mengenali pola-pola tertentu dalam kumpulan data besar. Berbagai teknik tersedia untuk Data mining termasuk klasifikasi, yang merupakan proses untuk mengkategorikan sekumpulan data ke dalam sejumlah kelas tertentu serta memberi label pada setiap kelas [7].

Metode yang dapat digunakan untuk klasifikasi diantaranya yaitu Decision Tree, Random Forest, dan XGBoost. Random Forest merupakan teknik Machine Learning dengan menggabungkan pohon keputusan untuk menghasilkan hasil prediksi, Extreme Gradient Boosting atau XGBoost merupakan algoritma Machine Learning yang menggabungkan pohon keputusan dengan kerangka kerja Gradient Boosting untuk meningkatkan kinerja [8] [9] [10]. Algoritma ini merupakan kemajuan dan peningkatan dari Gradient Boosting yang menawarkan beberapa manfaat, salah satunya adalah kemampuannya dalam mencegah overfitting. XGBoost menggunakan teknik boosting tingkat lanjut untuk mengelola data yang besar dan kompleks [11].

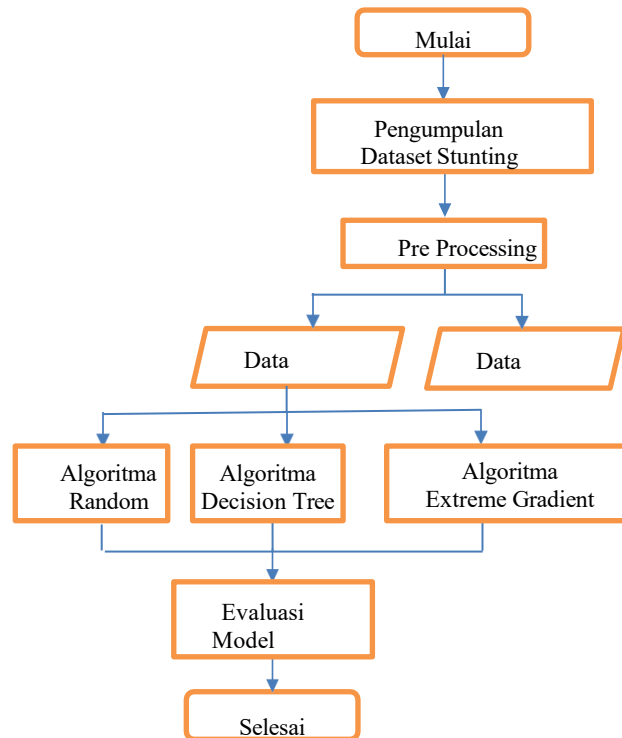
Penelitian tentang stunting dan metode Support Vector Machine (SVM) menghasilkan akurasi sebesar 80% [12]. Kombinasi Algoritman untuk memprediksi Resiko Stunting di Kota Dumai menunjukkan Decision Tree lebih unggul dari SVM, Algoritma Decision Tree dan Support Vector Machine (SVM) menghasilkan akurasi 96,15% dengan Decision Tree dan 62,48% menggunakan SVM [13]. Penelitian lainnya membandingkan Algoritma Random Forest, Decision Tree, dan XGBoost untuk mengklasifikasi Penyakit Infeksi Gigi dan Mulut menunjukkan hasil akurasi tertinggi di dapat oleh XGboost sebesar 97,2% , sedangkan untuk Decision Tree dan Random Forest sebesar 88,9% dan 93,1 % [14].

Berdasarkan penelitian yang telah dilakukan, tujuan dari penelitian ini adalah untuk Analisis Algoritma Decision Tree, Random Forest, dan Extreme Gradient Boosting dalam Klasifikasi Stunting untuk mendorong ketepatan Identifikasi Klasifikasi Stunting.

## 2. RESEARCH METHODS

### a. Diagram Urutan Langkah Pada Penelitian

Langkah-langkah yang harus dilalui pada penelitian ini sangat penting agar dapat memastikan penelitian ini dapat berjalan dengan lancar. Metode penelitian dan urutan langkah pada Penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Diagram urutan langkah Penelitian

## b. Pengumpulan Dataset

Data Stunting yang digunakan dalam Penelitian ini merupakan dataset publik yang tersedia di Platform Kaggle dengan nama “ Stunting Toddler Detection” [6]. Dataset berjumlah 10.000 record yang bisa dilihat pada Gambar 2. Data tersebut terdiri dari 4 atribut atau kolom yaitu no baris, Umur (bulan), Jenis Kelamin dan Tinggi badan. Status Tinggi badan Balita di klasifikasikan menjadi empat kategori yaitu Tinggi, Normal, Pendek, dan Sangat Pendek.

Data awal yang digunakan dalam penelitian ini berjumlah 10.000 record yang bisa dilihat pada Tabel 1.

Open in [Turbo Prep](#) [Auto Model](#) Filter (10,000 / 10,000 examples): [all](#)

| Row No. | Status Gizi       | Umur (bulan) | Jenis Kelamin | Tinggi Badan... |
|---------|-------------------|--------------|---------------|-----------------|
| 9998    | severely stunt... | 4            | perempuan     | 53.900          |
| 9999    | normal            | 4            | perempuan     | 62.700          |
| 9990    | severely stunt... | 4            | perempuan     | 53.100          |
| 9991    | stunted           | 4            | perempuan     | 56.800          |
| 9992    | normal            | 4            | perempuan     | 59.300          |
| 9993    | tinggi            | 4            | perempuan     | 68.800          |
| 9994    | severely stunt... | 4            | perempuan     | 51.900          |
| 9995    | normal            | 4            | perempuan     | 60.500          |
| 9996    | normal            | 4            | perempuan     | 65.500          |
| 9997    | tinggi            | 4            | perempuan     | 70.200          |
| 9998    | normal            | 4            | perempuan     | 58.800          |
| 9999    | normal            | 4            | perempuan     | 64.900          |
| 10000   | severely stunt... | 5            | laki-laki     | 54.400          |

ExampleSet (10,000 examples, 1 special attribute, 3 regular attributes)

Gambar 2. Data Awal

Tabel 1. Data Awal

| No    | Status Gizi      | Umur (Bulan) | Jenis Kelamin | Tinggi Badan |
|-------|------------------|--------------|---------------|--------------|
| 1     | Stunted          | 0            | Laki-laki     | 44,592       |
| 2     | Tinggi           | 0            | Laki-laki     | 56,705       |
| 3     | Normal           | 0            | Laki-laki     | 46,863       |
| 4     | Normal           | 0            | Laki-laki     | 47,508       |
| 5     | Severely stunted | 0            | Laki-laki     | 42,743       |
| 6     | Stunted          | 0            | Laki-laki     | 44,258       |
| 7     | Tinggi           | 0            | Laki-laki     | 59,573       |
| 8     | Severely stunted | 0            | Laki-laki     | 42,702       |
| 9     | Stunted          | 0            | Laki-laki     | 45,252       |
| 10    | Tinggi           | 0            | Laki-laki     | 57,202       |
| ..... |                  |              |               |              |
| 9999  | Normal           | 4            | Perempuan     | 64,900       |
| 10000 | Severely stunted | 5            | Laki-laki     | 54,400       |

### c. Preprocessing

Preprocessing dilakukan untuk mempersiapkan data sebelum dilakukan pemodelan. Beberapa preprocessing yang dilakukan pada penelitian yang dilakukan ini yaitu Data Cleaning diantaranya menangani missing value dan menghapus data duplikat. Preprocessing merupakan tahap penting agar dapat memastikan kualitas data yang akan dipergunakan dalam proses pemodelan [15]. Dalam penelitian ini, preprocessing juga meliputi normalisasi dan standardisasi fitur numerik untuk meningkatkan performa algoritma machine learning yang bisa dilihat pada Gambar 3. Pada tahap ini dilakukan cleaning diantaranya missing value dan menghapus data duplikat. Dari 10.000 record data awal yang digunakan terdapat data duplikat yang di hapus sehingga data yang siap di olah yaitu 7081 record yang ditampilkan pada Tabel 2.

Tahapan Pre-Processing :

- 1) Data Cleaning
- 2) Penanganan Missing Value
- 3) Penghapusan Data Duplikat (-2.919)
- 4) Normalisasi Data Numerik

| Row No. | Status Gizi      | Umur (bulan) | Jenis Kelamin | Tinggi Badan... |
|---------|------------------|--------------|---------------|-----------------|
| 7067    | stunted          | 4            | perempuan     | 57.800          |
| 7068    | tinggi           | 4            | perempuan     | 73.500          |
| 7069    | severely stunted | 4            | perempuan     | 53.300          |
| 7070    | tinggi           | 4            | perempuan     | 69.300          |
| 7071    | normal           | 4            | perempuan     | 66.800          |
| 7072    | severely stunted | 4            | perempuan     | 54              |
| 7073    | severely stunted | 4            | perempuan     | 50.400          |
| 7074    | normal           | 4            | perempuan     | 62.100          |
| 7075    | normal           | 4            | perempuan     | 62.200          |
| 7076    | severely stunted | 4            | perempuan     | 55              |
| 7077    | tinggi           | 4            | perempuan     | 71.800          |
| 7078    | stunted          | 4            | perempuan     | 57.500          |
| 7079    | severely stunted | 4            | perempuan     | 53.200          |
| 7080    | normal           | 4            | perempuan     | 58.300          |
| 7081    | severely stunted | 5            | laki-laki     | 54.400          |

ExampleSet (7,081 examples, 1 special attribute, 3 regular attributes)

Gambar 3. Pre Processing

Tabel 2. Data Setelah Pre Processing

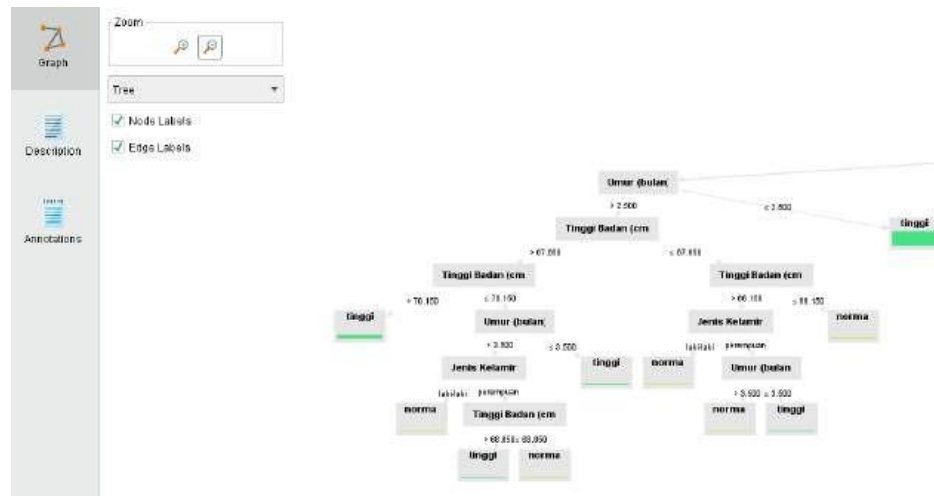
| No    | Status Gizi      | Umur (bulan) | Jenis Kelamin | Tinggi Badan |
|-------|------------------|--------------|---------------|--------------|
| 1     | Stunted          | 0            | Laki-laki     | 44,592       |
| 2     | Tinggi           | 0            | Laki-laki     | 56,705       |
| 3     | Normal           | 0            | Laki-laki     | 46,863       |
| 4     | Normal           | 0            | Laki-laki     | 47,508       |
| 5     | Severely stunted | 0            | Laki-laki     | 42,743       |
| 6     | Stunted          | 0            | Laki-laki     | 44,258       |
| 7     | Tinggi           | 0            | Laki-laki     | 59,573       |
| 8     | Severely stunted | 0            | Laki-laki     | 42,702       |
| 9     | Stunted          | 0            | Laki-laki     | 45,252       |
| 10    | Tinggi           | 0            | Laki-laki     | 57,202       |
| ..... |                  |              |               |              |
| 7080  | Normal           | 4            | Perempuan     | 58,300       |
| 7081  | Severely stunted | 5            | Laki-laki     | 54,400       |

**d. Klasifikasi dengan Algoritma Decision Tree, Random Forest, dan XGBoost**

Pemodelan dilakukan dengan tiga algoritma yaitu Algoritma Random Forest, Decision Tree dan Extreme Gradient Boosting (XGBoost), kemudian membandingkan dari ketiga algoritma tersebut manakah yang memiliki akurasi tertinggi. Setiap algoritma memiliki karakteristik dan kelebihan tersendiri dalam mengklasifikasikan data [6] [15].

Random Forest menggunakan ensemble dari multiple decision tree untuk mendapatkan hasil prediksi yang lebih stabil dan akurat. Setiap pohon dalam Random Forest dilatih pada subset data yang berbeda dengan teknik bootstrap sampling, dan pemilihan fitur secara acak pada setiap node split, sehingga mengurangi overfitting dan meningkatkan generalisasi model [16]. Decision Tree membangun model prediksi dalam bentuk struktur pohon yang intuitif dan mudah diinterpretasi yang bisa dilihat pada Gambar 3. Algoritma ini bekerja dengan membagi data berdasarkan fitur yang paling informatif pada setiap langkah, menciptakan jalur keputusan dari root node ke leaf node yang merepresentasikan klasifikasi [14]. XGBoost merupakan implementasi optimized dari gradient boosting yang membangun pohon keputusan secara sekuensial, dimana setiap pohon baru berusaha memperbaiki kesalahan dari pohon sebelumnya. XGBoost menambahkan regularisasi dan pengoptimalan paralel yang membuat algoritma ini sangat efisien dan berkinerja tinggi.

Dataset yang siap digunakan di bagi menjadi Data Training dan juga Data Testing dengan proporsi Data Training sebesar 70% , dan Data Testing 30%. Data Training dipergunakan untuk membentuk model, sedangkan Data Testing digunakan untuk menguji model. Model yang digunakan yaitu Decision Tree, Random Forest, dan XGBoost.



Gambar 4. Hasil Pohon Keputusan

### e. Evaluasi

Tingkat evaluasi digunakan untuk mengukur kinerja model yang digunakan. Studi ini menggunakan Confusion Matrix dengan akurasi, precision, dan recall. Hasil Evaluasi pada Algoritma Random Forest menggunakan Confusion Matrix menunjukkan akurasi sebesar 99,72 % dengan precision 99,69% dan recall 99,68%. Hasil evaluasi pada Algoritma XGBoost menunjukkan akurasi 99,58% dengan precision 99,48% dan recall 99,32%, sedangkan hasil evaluasi menggunakan algoritma Decision Tree menunjukkan akurasi 98,87%, dengan precision 98,35% dan recall 97,34%.

## 3. RESULT

### a. Hasil Evaluasi Algoritma *Random Forest* menggunakan *Confusion Matrix*.

Pada Gambar 5. Ditampilkan accurasi dengan Algoritma Random Forest menunjukkan accurasi sebesar 99, 72 %.

ExampleSet (Remove Duplicates)

Random Forest Model (Random Forest)

Result History

Criterion

accuracy

weighted mean recall

weighted mean precisi...

Table View

Plot View

accuracy: 99.72%

|                        | true stunted | true tinggi | true normal | true severely stunted | class precision |
|------------------------|--------------|-------------|-------------|-----------------------|-----------------|
| pred. stunted          | 165          | 0           | 1           | 0                     | 99.40%          |
| pred. tinggi           | 0            | 683         | 1           | 0                     | 99.85%          |
| pred. normal           | 1            | 3           | 837         | 0                     | 99.52%          |
| pred. severely stun... | 0            | 0           | 0           | 433                   | 100.00%         |
| class recall           | 99.40%       | 99.56%      | 99.76%      | 100.00%               |                 |

Gambar 5. Hasil Klasifikasi Random Forest

### b. Hasil Evaluasi Algoritma Decision Tree menggunakan Confusion Matrix.

Pada Gambar 5. Ditampilkan accurasi dengan Algoritma Decision Tree menunjukkan accurasi sebesar 98, 87 %.

PerformanceVector (Performance) ExampleSet (Remove Duplicates) Tree (Decision Tree)

Table View Plot View

accuracy: 98.87%

|                        | true stunted | true tinggi | true normal | true severely stunted | class precision |
|------------------------|--------------|-------------|-------------|-----------------------|-----------------|
| pred. stunted          | 150          | 0           | 5           | 0                     | 96.77%          |
| pred. tinggi           | 0            | 684         | 1           | 0                     | 99.85%          |
| pred. normal           | 6            | 2           | 833         | 0                     | 99.05%          |
| pred. severely stun... | 10           | 0           | 0           | 433                   | 97.74%          |
| class recall           | 90.36%       | 99.71%      | 99.28%      | 100.00%               |                 |

Gambar 6. Hasil Klasifikasi Decision Tree

#### 4. Hasil Evaluasi Algoritma Extreme Gradient Boosting(XGBoost) menggunakan Confusion Matrix

PerformanceVector (Performance) ExampleSet (Remove Duplicates) XGBoost (XGBoost)

Table View Plot View

accuracy: 99.58%

|                        | true stunted | true tinggi | true normal | true severely stunted | class precision |
|------------------------|--------------|-------------|-------------|-----------------------|-----------------|
| pred. stunted          | 163          | 0           | 1           | 1                     | 98.79%          |
| pred. tinggi           | 0            | 883         | 1           | 0                     | 99.85%          |
| pred. normal           | 3            | 3           | 837         | 0                     | 99.29%          |
| pred. severely stun... | 0            | 0           | 0           | 432                   | 100.00%         |
| class recall           | 98.19%       | 99.56%      | 99.76%      | 99.77%                |                 |

Gambar 7. Hasil Klasifikasi *Extreme Gradient Boosting(XGBoost)*

Tabel 3. Perbandingan hasil analisis dari tiga Algoritma

| Algoritma     | Akurasi | Precision | Recall | F1-Score | Ranking |
|---------------|---------|-----------|--------|----------|---------|
| Random Forest | 99,72%  | 99,69%    | 99,68% | 99,68%   | 1       |
| XGBoost       | 99,58%  | 99,48%    | 99,32% | 99,40%   | 2       |
| Decision Tree | 98,87%  | 98,35%    | 97,34% | 97,84%   | 3       |

Penelitian ini berhasil menganalisis dan membandingkan tiga algoritma machine learning yang digunakan untuk klasifikasi stunting pada anak kecil / balita dengan hasil sebagai berikut:

- Random Forest mencapai performa terbaik dengan akurasi 99,72%, precision 99,69%, dan recall 99,68%, menjadikannya algoritma paling efektif untuk klasifikasi stunting dalam penelitian ini.
- XGBoost menunjukkan performa yang sangat kompetitif dengan akurasi 99,58%, hanya selisih 0,14% dari Random Forest, membuktikan kemampuannya dalam menangani data klasifikasi stunting.
- Decision Tree, meskipun memiliki akurasi terendah (98,87%), tetap menunjukkan performa yang baik dan memiliki keunggulan dalam interpretabilitas model.
- Semua algoritma menunjukkan performa yang sangat baik (akurasi >98%), mengindikasikan bahwa ketiga fitur yang digunakan (umur, jenis kelamin, tinggi badan) cukup representatif untuk klasifikasi stunting.

## 5. DISCUSSION

Penelitian ini berhasil menghasilkan beberapa kontribusi penting dalam bidang deteksi stunting menggunakan pendekatan machine learning. Pertama, penelitian ini menyediakan benchmark performa algoritma machine learning pada dataset stunting yang dapat menjadi referensi bagi penelitian-penelitian selanjutnya di bidang ini. Hasil benchmark ini memberikan gambaran objektif mengenai kemampuan berbagai algoritma dalam mengidentifikasi kasus stunting, sehingga peneliti lain dapat menggunakan hasil ini sebagai dasar perbandingan untuk pengembangan metode yang lebih baik.

Temuan penelitian menunjukkan bahwa sistem deteksi otomatis dapat dikembangkan menggunakan algoritma Random Forest untuk membantu tenaga kesehatan dalam melakukan screening stunting. Implementasi sistem ini memiliki potensi besar untuk meningkatkan efisiensi diagnosa dengan mengurangi ketergantungan pada penilaian manual yang selama ini memerlukan waktu yang cukup lama dan rentan terhadap human error. Dengan adanya sistem otomatis ini, tenaga kesehatan dapat melakukan screening dengan lebih cepat dan akurat, sehingga dapat mengalokasikan waktu mereka untuk aktivitas pelayanan kesehatan lainnya yang memerlukan intervensi langsung.

Keunggulan lain dari sistem yang dikembangkan adalah skalabilitasnya yang memungkinkan penerapan pada program kesehatan masyarakat untuk monitoring stunting secara massal. Hal ini sangat relevan dengan kebutuhan Indonesia dalam mengatasi masalah stunting yang memerlukan monitoring berkelanjutan pada populasi yang besar. Sistem ini dapat diintegrasikan ke dalam program posyandu, puskesmas, atau sistem kesehatan digital lainnya untuk memfasilitasi deteksi dini dan intervensi yang tepat waktu.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan yang perlu diakui dan menjadi agenda untuk penelitian selanjutnya. Keterbatasan utama adalah penggunaan fitur yang masih terbatas, yaitu hanya menggunakan 3 atribut dasar dalam analisis. Penelitian belum mencakup faktor-faktor penting lainnya seperti kondisi sosio-ekonomi keluarga, riwayat kesehatan anak, dan status nutrisi yang secara empiris diketahui memiliki pengaruh signifikan terhadap kejadian stunting. Inklusi faktor-faktor ini dalam penelitian mendatang diharapkan dapat meningkatkan akurasi dan reliabilitas model prediksi.

Keterbatasan kedua adalah penggunaan dataset tunggal yang bersumber dari satu dataset publik. Hal ini menimbulkan kekhawatiran mengenai generalisasi hasil penelitian, karena karakteristik populasi dari dataset tersebut mungkin tidak sepenuhnya merepresentasikan keragaman populasi Indonesia yang memiliki variasi geografis, budaya, dan sosio-ekonomi yang luas. Penggunaan multiple dataset dari berbagai wilayah dan karakteristik populasi yang berbeda akan memberikan validasi yang lebih kuat terhadap robustness model yang dikembangkan.

Selain itu, penelitian ini belum melakukan validasi eksternal pada dataset independen dari sumber yang berbeda. Validasi eksternal merupakan langkah penting untuk memastikan bahwa model yang dikembangkan dapat bekerja dengan baik tidak hanya pada data training dan testing, tetapi juga pada data real-world yang mungkin memiliki karakteristik yang berbeda. Ketiadaan validasi eksternal ini menjadi keterbatasan yang perlu diatasi dalam penelitian lanjutan untuk memastikan aplikabilitas praktis dari sistem yang dikembangkan.

## 6. CONCLUSIONS AND RECOMMENDATIONS

Penelitian ini berhasil menunjukkan efektivitas algoritma machine learning dalam mengklasifikasikan status stunting pada balita menggunakan fitur demografis dan antropometrik dasar. Algoritma Random Forest mencapai performa tertinggi dengan akurasi 99,72%, diikuti XGBoost sebesar 99,58%, dan Decision Tree sebesar 98,87%. Keunggulan Random Forest dapat

dikaitkan dengan pendekatan ensemble yang efektif menangani noise dan outlier dalam data sambil mengurangi overfitting melalui kombinasi multiple decision tree yang independen. Hasil ini menunjukkan bahwa sistem klasifikasi otomatis dapat secara signifikan meningkatkan efisiensi dan akurasi deteksi stunting dibandingkan dengan metode penilaian manual tradisional, memberikan dukungan berharga bagi tenaga kesehatan dalam identifikasi dini anak-anak yang berisiko.

Penelitian selanjutnya sebaiknya fokus pada perluasan dataset untuk mencakup pengukuran antropometrik tambahan (berat badan, lingkaran kepala), riwayat kesehatan, dan faktor sosio-ekonomi guna menciptakan model prediksi yang lebih komprehensif. Implementasi teknik Feature Engineering lanjutan dan optimasi Hyperparameter disarankan untuk lebih meningkatkan performa model pada semua algoritma. Selain itu, pengembangan aplikasi praktis yang mengintegrasikan model terbaik (Random Forest) ke dalam sistem informasi kesehatan akan memfasilitasi penilaian risiko stunting secara real-time di lingkungan klinis. Para pembuat kebijakan dan praktisi kesehatan hendaknya mempertimbangkan adopsi alat berbasis machine learning ini sebagai metode skrining komplementer untuk meningkatkan program intervensi dini dan pada akhirnya mengurangi prevalensi stunting sesuai dengan target kesehatan nasional.

## REFERENCES

- [1] T. Hardiani and R. N. Putri, "Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Stunting Pada Balita," *Digital Transformation Technology (Digitech)*, vol. 4, no. 1, pp. 621-627, 2024.
- [2] A. Istiqomah, K. M. S, R. A. Amali and S. Tiawati, "Peran Gizi Terhadap Pertumbuhan Dan Perkembangan Balita," *Antigen: Jurnal Kesehatan Masyarakat dan Ilmu Gizi*, vol. 2, no. 2, pp. 67-74, 2024.
- [3] A. Daracantika, A. Ainin and B. Besral, "Systematic Literature Review: Pengaruh Negatif Stunting terhadap Perkembangan Kognitif Anak," *Jurnal Biostatistik, Kependudukan dan Informatika Kesehatan (BIKFOKES)*, vol. 1, no. 2, pp. 124-135, 2020.
- [4] N. K. T. Apriastini, N. P. T. Adnyani, P. O. Selvyani and K. H. Setiawan, "STUNTING:FAKTOR RISIKO, DIAGNOSIS, TATALAKSANA, DAN PROGNOSIS," *Ganesha Medicina*, vol. 4, no. 1, pp. 17-23, 2024.
- [5] S. L. Munira, "https://ayosehat.kemkes.go.id/pub/files/files46531.\_MATERI\_KABKPK\_SOS\_SSGI.pdf," 03 022023. [Online]. Available: [https://ayosehat.kemkes.go.id/pub/files/files46531.\\_MATERI\\_KABKPK\\_SOS\\_SSGI.pdf](https://ayosehat.kemkes.go.id/pub/files/files46531._MATERI_KABKPK_SOS_SSGI.pdf).
- [6] Z. I. Bimawan, T. Astuti and P. Arsi, "Comparison of Random Forest, K-Nearest Neighbor, Decision Tree, and XGBoost Algorithms for Detecting Stunting in Toddlers," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 6, pp. 1599-1607, 2024.
- [7] D. Papakyriakou and I. S. Barbounakis, "Data Mining Methods: A Review," *International Journal of Computer Applications*, vol. 183, no. 48, pp. 5-19, 2022.
- [8] H. A. Salman, A. Kalakech and A. Steiti, "Random Forest Algorithm Overview," *The Babylonian Journal of Machine Learning (BJML)*, vol. 2024, p. 69–79, 2024.
- [9] H. A. Abdulqader and A. M. Abdulazez, "A Review on Decision Tree Algorithm in Healthcare Applications," *Indonesian Journal of Computer Science*, vol. 13, no. 3, pp. 3863-3881, 2024.
- [10] Y. Pristyanto, Z. Mukarabiman and A. F. Nugraha, "Extreme Gradient Boosting Algorithm to Improve Machine Learning Model Performance on Multiclass Imbalanced Dataset," *JOIV : International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 710-715, 2023.

- 
- [11] S. E. H. Yulianti, O. Soesanto and Y. Sukmawaty, "Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit," *Journal of Mathematics: Theory and Applications (JOMTA)*, vol. 4, no. 1, pp. 21-26, 2022.
  - [12] Romlah, S. Faisal and J. I. Rahmat3, "Prediksi Risiko Angka Stunting Pada Balita Menggunakan Algoritma Support Vector Machine," *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, vol. 7, no. 2, pp. 837-842, 2025.
  - [13] A. I. Putri, Y. Syarif, P. Jayadi, F. Arrazak and F. N. Salisah, "Implementasi Algoritma Decision Tree dan Support Vector Machine (SVM) untuk Prediksi Risiko Stunting pada Keluarga," *MALCOM : Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 349-357, 2023.
  - [14] B. A. S. Aji, Y. Setiawan and S. D. Anggraini, "Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan XGBoost untuk Klasifikasi Penyakit Infeksi Gigi dan Mulut," *INTEGER: Journal of Information Technology*, vol. 10, no. 1, pp. 135-148, 2025.
  - [15] O. N. Chilyabanyama, R. Chilengi, M. Simuyandi, C. C. Chisenga, M. Chirwa, K. Hamusonde, R. K. Saroj, N. T. Iqbal, I. Ngaruye and S. Bosomprah, "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," *Children*, vol. 9, no. 7, pp. 1-11, 2022.
  - [16] P. Handayani, A. C. Fauzan and Harliana, "Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 6, pp. 3064-3072, 2024.
  - [17] F. Fadmadika, H. H. Handayani, T. A. Mudzakir and J. Indra, "engaruh smote terhadap performa algoritma random forest dan algoritma gradient boosting dalam memprediksi penyakit stroke," *Jurnal TEKINKOM*, vol. 7, no. 2, pp. 837-846, 2024.
  - [18] A. I. Putri and Y. Syarif, "Implementasi Algoritma Decision Tree dan Support Vector Machine (SVM) untuk Prediksi Risiko Stunting pada Keluarga," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 349-357, 2023.