Jurnal Komputer Antartika
Volume 1 Nomor 1 2026 Halaman 1-6
ISSN (Online) 2988-5272
DOI: https://doi.org/10.70052/jka.v4i1.1277

Deteksi Dini Stunting pada Balita Menggunakan Data Mining dengan Algoritma C4.5

Muhamad Regen[1], Yogi Nur Alamsyah[2], Rendi Simon Lesmana[3*], Asima Rodame Tampubolon[4], Diana Nur Hafifah[5], Annida Purnamawati[6]

[1,2,3,4,5,6] Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98, Senen, Jakarta Pusat 10450, Indonesia

e-mail korespondensi: rendisimon98@gmail.com

Submit: 07-01-2026 | Revisi: 15-01-2026 | Terima: 23-01-2026 | Terbit online: 02-02-2026

Abstrak - Masalah stunting, yang cukup serius, mempengaruhi pertumbuhan dan perkembangan anak di Indonesia. Angka prevalensinya masih tinggi, mencapai 148 juta balita. Dalam penelitian ini, kami berusaha mengembangkan model deteksi dini stunting dengan memanfaatkan algoritma pohon keputusan C4.5. Dataset besar yang kami gunakan berisi 120.999 catatan, mencakup atribut usia, tinggi badan, dan jenis kelamin. Metode yang kami pilih adalah pendekatan eksperimental kuantitatif dengan teknik penambangan data. Model ini dievaluasi menggunakan 10-fold cross-validation untuk memastikan akurasi dan generalisasi. Hasilnya, model C4.5 mencapai akurasi 99,87%, dengan presisi dan recall yang sangat tinggi, serta interpretabilitas yang baik. Ini membuatnya sangat cocok untuk diterapkan dalam sistem kesehatan masyarakat. Temuan ini menegaskan pentingnya tinggi badan sebagai indikator utama dalam mendeteksi stunting, dan memberikan dasar untuk integrasi model dalam inisiatif kesehatan digital di Indonesia. Kami juga merekomendasikan penggabungan atribut sosial ekonomi dan lingkungan untuk analisis yang lebih komprehensif di masa depan.

Kata Kunci : Stunting, Deteksi dini, Algoritma C4.5, Kesehatan masyarakat, Machine learning

Abstract *- Stunting is a serious issue affecting the growth and development of children in Indonesia, with a prevalence still high, reaching 148 million children under five. This study aims to develop an early detection model for stunting using the C4.5 decision tree algorithm, utilizing a large dataset containing 120,999 records that include attributes of age, height, and gender. The research method used is a quantitative experimental approach with data mining techniques, where the model was evaluated using 10-fold cross-validation to ensure accuracy and generalizability. The results show that the C4.5 model achieves 99.87% accuracy, with very high precision and recall, and good interpretability, making it suitable for implementation in public health systems. These findings emphasize the importance of height as a key indicator in detecting stunting and provide a basis for model integration in digital health initiatives in Indonesia. This study recommends incorporating socioeconomic and environmental attributes for more comprehensive analysis in the future.*

Keywords : Stunting, Early detection, C4.5 algorithm, Public health, Machine learning

1. Pendahuluan

Stunting adalah gangguan yang mempengaruhi pertumbuhan dan perkembangan anak, yang disebabkan oleh kekurangan gizi yang berlangsung dalam jangka waktu lama serta infeksi yang terjadi berulang kali. Kondisi ini ditandai dengan ukuran tinggi atau panjang badan anak yang berada di bawah standar yang ditetapkan [1]. Pada penelitian lain yang dilakukan oleh soliman dkk [2] juga menerangkan bahwa Stunting merupakan suatu proses yang dapat mempengaruhi perkembangan anak sejak dalam kandungan hingga anak berusia tiga atau empat tahun, dimana gizi ibu dan anak merupakan faktor penentu pertumbuhan yang penting.

Menurut hasil Survei Status Gizi Indonesia (SSGI) pada tahun 2024, target penurunan angka stunting untuk tahun 2025 ditetapkan sebesar 18,8%, yang menjadi tantangan besar bagi negara Indonesia dan sejalan dengan kebijakan nasional percepatan penurunan stunting melalui RAN PASTI (Rencana Aksi Nasional Percepatan Penurunan Angka Stunting) [3]. Secara global, sekitar 148 juta anak balita mengalami stunting, dan Indonesia masih berada di peringkat sepuluh besar negara dengan prevalensi stunting tertinggi [4].

Selain itu, dampak negatif yang dialami oleh balita di Indonesia akibat kekurangan gizi dalam waktu yang lama dapat mengganggu pertumbuhan fisik, perkembangan kognitif, serta kemampuan motorik mereka, yang pada gilirannya akan mempengaruhi kesehatan dan produktivitas mereka di masa dewasa[5]. Oleh karena itu,

pendekatan yang berbasis pada status gizi dilakukan dengan menggunakan beberapa indikator, seperti berat badan bayi, panjang badan bayi, dan status gizi secara keseluruhan.

Metode tradisional yang biasa digunakan untuk menilai status gizi umumnya mengandalkan grafik pertumbuhan dan penilaian klinis secara manual, yang sangat rentan terhadap kesalahan dari operator dan tidak efisien dalam hal waktu. Sebaliknya, dengan penerapan teknik machine learning, penilaian dapat dilakukan secara otomatis dan konsisten, serta mampu mengklasifikasikan status gizi dengan tingkat akurasi yang lebih tinggi[6][7].

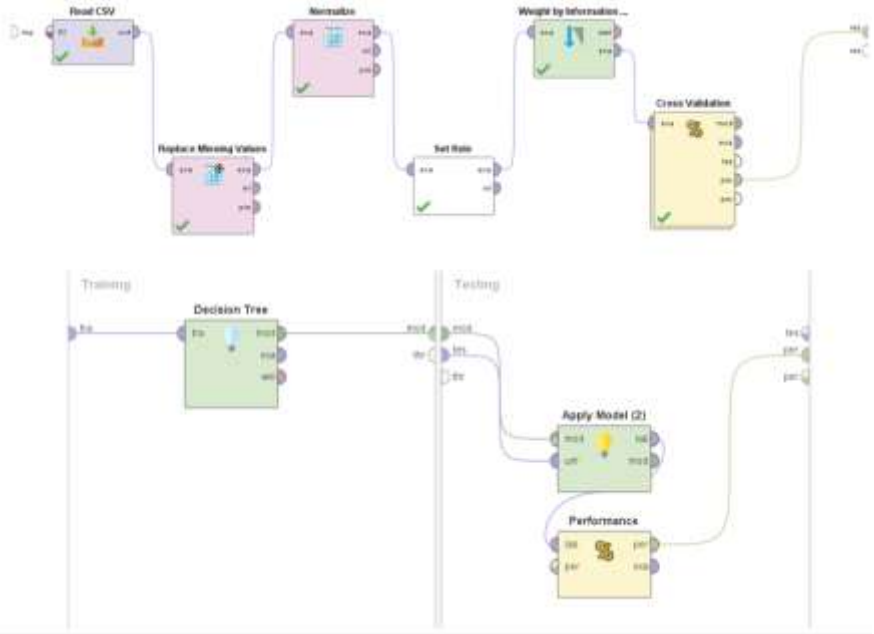
Penelitian sebelumnya yang dilakukan oleh Revaldo, Ferry, dan Riduan[8] menggunakan algoritma C4.5 untuk memprediksi anak yang mengalami stunting dengan tingkat akurasi mencapai 84,00%. Hal serupa juga dilakukan oleh Alamsyah[9] yang berhasil mencapai akurasi 96,00%. Pada penelitian lain yang dilakukan oleh Brekmans Darkel dkk[10] yang mengimplementasikan model algoritma C4.5 untuk klasifikasi status stunting di Kabupaten Sikka mencapai tingkat akurasi 92,62%. Selanjutnya Namun, penelitian-penelitian tersebut masih menggunakan kumpulan data yang relatif kecil dan atribut yang terbatas.

Berbeda dengan penelitian sebelumnya yang hanya menggunakan dataset berskala kecil dan klasifikasi biner, penelitian ini memperkenalkan sebuah dataset berskala besar yang terdiri dari 120.999 catatan dan melakukan klasifikasi multi-kelas terhadap status gizi (normal, stunting, stunting berat, tinggi). Penerapan algoritma pohon keputusan C4.5 menekankan pada akurasi prediktif serta interpretabilitas model, sehingga berpotensi untuk diterapkan dalam bidang kesehatan masyarakat.

2. Metode Penelitian

2.1 Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimental kuantitatif yang didasarkan pada teknik penambangan data. Seluruh proses dilaksanakan dengan memanfaatkan RapidMiner Studio karena menyediakan antarmuka yang lebih mudah digunakan dan menyarankan algoritma terbaik berdasarkan kinerja dan waktu[11][12], seperti yang ditunjukkan pada Gambar 1.



Gambar. 1 Alur kerja penelitian di RapidMiner

Secara ringkas, alur kerja penelitian ditunjukkan pada Gambar 1, sedangkan detail tahapan pengolahan data, pemodelan, dan evaluasi dijelaskan pada Bab 3.

Untuk mengurangi risiko overfitting dan memastikan ketahanan model, digunakan 10-fold cross-validation[13]. Dataset dibagi menjadi sepuluh subset yang sama, setiap subset digunakan satu kali untuk pengujian, sementara sembilan subset lainnya digunakan untuk pelatihan. Proses ini diulang hingga setiap subset berfungsi sebagai lipatan pengujian.

2.2 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle dengan judul “Deteksi Stunting Balita – 121K Baris” dan mencakup 120.999 catatan balita yang memiliki empat atribut utama, yang dirangkum dalam Tabel 1.

Tabel 1. Deskripsi Dataset

Atribut	Tipe	Deskripsi	Missing Value
Usia	Numerik	Usia dalam bulan	Tidak ada
Tinggi	Numerik	Tinggi dalam cm	Tidak ada
Jenis Kelamin	Kategorikal	Laki-laki atau perempuan	Tidak ada
Status Gizi	Kategorikal (Label)	Normal, Stunting, stunting berat, tinggi	Tidak ada

Tidak ditemukan nilai yang hilang, yang memastikan kelengkapan dan keandalan data. Distribusi data berdasarkan kategori status gizi ditampilkan dalam Tabel 2.

Tabel 2. Distribusi Kelas Status Gizi

Status Gizi	Jumlah Data	Persentase (%)
Normal	68.455	56.6
Tinggi	22.990	19.0
Stunting	16.245	13.4
Stunting berat	13.319	11.0
Total	120.999	100

Berdasarkan dataset pada tabel 1 dan tabel 2, tahapan pengolahan data, pemodelan, dan evaluasi disajikan pada bagian Hasil dan Pembahasan berikut.

3. Hasil dan Pembahasan

3.1 Preprocessing Data

Tahap preprocessing bertujuan untuk memastikan bahwa dataset dalam kondisi bersih, terstandarisasi, dan siap untuk proses pemodelan[14]. Proses ini melibatkan langkah-langkah berikut:

1. Normalisasi: Atribut numerik, seperti usia dan tinggi, dinormalisasi dengan menggunakan transformasi Z-score agar skala menjadi konsisten
2. Penetapan Peran Label: "Status Gizi" ditentukan sebagai variabel yang menjadi fokus.
3. Validasi Silang: Selain itu, metode 10-fold cross-validation diterapkan selama penyetelan model untuk menghindari overfitting dan meningkatkan generalisasi.

3.2 Pembobotan Fitur (Information Gain)

Penilaian terhadap pentingnya fitur dilakukan dengan menggunakan metode Information Gain. Hasil yang ditampilkan dalam Tabel 3 menunjukkan bahwa atribut tinggi badan merupakan faktor yang paling berpengaruh dalam memprediksi status gizi seseorang.

Tabel. 3 Hasil Pembobotan Fitur

Atribut	Information Gain
Tinggi	0.1675
Usia	0.0246
Jenis Kelamin	0.0004

Dominasi tinggi badan ini sejalan dengan standar pertumbuhan WHO, di mana keterlambatan pertumbuhan linear menjadi indikator utama dari stunting [4].

3.3 Pemodelan Algoritma C4.5

Algoritma pohon keputusan C4.5 dipilih karena kemampuannya dalam memberikan interpretasi yang jelas serta kinerja klasifikasi yang kuat pada dataset kesehatan[15]. Parameter yang digunakan ditentukan melalui proses penyetelan iteratif untuk mencapai keseimbangan antara akurasi dan kompleksitas model, yang meliputi:

1. Kriteria: Rasio Informasi Gain
2. Minimum gain: 0,01
3. Maximal Depth: 20
4. Confidence Factor: 0,25

Algoritma ini secara rekursif membagi data berdasarkan pengurangan entropi, sehingga menghasilkan aturan keputusan yang transparan dan cocok untuk interpretasi oleh para ahli. Gambar 2 menunjukkan visualisasi pohon keputusan yang dihasilkan dalam RapidMiner.

3.4 Evaluasi Model

Kinerja model dievaluasi dengan menggunakan metode 10-fold cross-validation, yang menghasilkan metrik yang konsisten di seluruh lipatan. Penilaian kinerja model dilakukan dengan mengukur akurasi, presisi, recall, dan F1-score, yang diperoleh dari confusion matrix yang ditampilkan pada Tabel 4.

Gambar. 2 visualisasi pohon keputusan yang dihasilkan dalam RapidMiner

Tabel. 4 Confusion Matrix

True Class	stunting	tinggi	normal	Stunting berat
stunting	13.745	0	21	24
tinggi	0	19.536	22	0
normal	36	24	67.712	0
Stunting berat	34	0	0	19.845

Dari tabel 4, diperoleh Metrik Evaluasi akurasi = 99,87%, presisi = 0.998, recall = 0.997, dan F1-Score = 0.997. Hasil ini menunjukkan kinerja klasifikasi yang sangat baik berdasarkan dataset yang diberikan, yang mengindikasikan kemampuan prediktif yang kuat serta potensi generalisasi untuk deteksi dini stunting pada balita.

3.5 Performa Model

Model C4.5 berhasil mencapai tingkat akurasi yang sangat tinggi, yaitu 99,87%, yang lebih baik dibandingkan dengan penelitian sebelumnya yang melaporkan akurasi di bawah 96%. Peningkatan ini dapat dijelaskan oleh penggunaan dataset yang besar dengan distribusi kelas yang relatif representatif, terdiri dari 120.999 catatan, serta proses pra-pemrosesan yang sistematis, termasuk normalisasi dan penimbangan atribut. Untuk memastikan generalisasi dan mencegah *overfitting*, model ini dievaluasi menggunakan metode *10-fold cross-validation*. Metrik kinerja yang lebih rinci disajikan dalam Tabel 5.

Tabel. 5 Metrik Kinerja

Metric	*Value*	*Interpretation*
Accuracy	99.87%	*Overall correctness of classification*
Precision (weighted mean)	99.83%	*Reliability of positive class predictions*
Recall (weighted mean)	99.80%	*Sensitivity to actual class occurrences*
F1-Score	99.81%	*Balance between precision and recall*
Kappa	0.998%	*Agreement beyond chance (almost perfect)*

Tingginya konsistensi antara presisi dan recall menunjukkan bahwa model ini berfungsi dengan baik di semua kelas status gizi. Pada bagian 3.4, ditampilkan hasil confusion matrix yang diperoleh dari cross-validation, di mana sebagian besar kesalahan klasifikasi sangat minim (kurang dari 0,2%), yang menunjukkan pemisahan kelas yang sangat baik. Secara keseluruhan, hasil ini menunjukkan bahwa algoritma C4.5 mampu melakukan generalisasi dengan efektif, meskipun hanya menggunakan atribut yang sedikit (usia, tinggi badan, jenis kelamin). Tingkat akurasi yang tinggi ini kemungkinan mencerminkan adanya korelasi linier yang kuat antara tinggi badan dan status gizi, yang sejalan dengan standar referensi pertumbuhan WHO.

3.6 Signifikansi Atribut

Analisis bobot atribut yang dijelaskan dalam Bagian 3.2 menunjukkan bahwa tinggi badan dan usia merupakan prediktor paling berpengaruh terhadap status gizi, yang sejalan dengan standar pertumbuhan WHO. Secara khusus, tinggi badan memiliki nilai Information Gain sebesar 0.1675, yang jauh lebih tinggi dibandingkan dengan usia (0.0246) dan jenis kelamin (0.0004). Temuan ini menunjukkan bahwa keterlambatan pertumbuhan linear yang diwakili oleh tinggi badan anak relatif terhadap usia merupakan fitur paling penting dalam mendeteksi stunting. Kontribusi minimal dari jenis kelamin menunjukkan bahwa stunting lebih banyak dipengaruhi oleh faktor fisiologis dan nutrisi ketimbang perbedaan biologis berdasarkan jenis kelamin.

3.7 Interpretabilitas Model

Salah satu keuntungan utama dari algoritma C4.5 adalah tingkat interpretabilitasnya yang tinggi, karena algoritma ini menghasilkan aturan keputusan yang dapat dibaca oleh manusia. Aturan-aturan tersebut dapat dengan mudah diterapkan dalam sistem kesehatan atau digunakan oleh petugas lapangan untuk melakukan penyaringan secara cepat. Sebagai contoh, versi sederhana dari aturan keputusan yang dihasilkan adalah *IF height < 85 cm AND age > 24 months THEN status = Severely Stunted.*

Visualisasi pohon keputusan yang disederhanakan ini (Gambar 3) menunjukkan pola klasifikasi yang bersifat hierarkis, di mana tinggi badan berfungsi sebagai atribut pemisah utama, diikuti oleh usia sebagai penentu sekunder. Keterbacaan dari aturan-aturan ini meningkatkan penerapan model dalam konteks kesehatan masyarakat seperti Posyandu atau Puskesmas, di mana alat digital dapat membantu tenaga kesehatan dalam memberikan peringatan dini terhadap kemungkinan kasus stunting. Selain itu, struktur transparan dari model C4.5 memastikan adanya penjelasan, yang merupakan aspek penting dalam sistem pendukung keputusan di bidang kesehatan.

3.8. Diskusi

Metrik yang hampir sempurna dari model ini mungkin dipengaruhi oleh beberapa faktor, antara lain:

1. Struktur dataset yang relatif sederhana dengan batas kelas yang jelas antara tinggi badan dan status gizi.
2. Ukuran dataset yang besar, yaitu 120.999 catatan, yang mendukung pembelajaran statistik yang stabil.
3. Penggunaan validasi silang yang dapat mengurangi overfitting dan meningkatkan kemampuan generalisasi. Namun demikian, tingginya nilai akurasi yang diperoleh juga mengindikasikan perlunya validasi eksternal menggunakan data dunia nyata atau data lintas wilayah untuk memastikan kinerja model tetap konsisten di luar dataset penelitian.

Meskipun kinerjanya sangat baik, penelitian di masa depan sebaiknya mempertimbangkan:

1. Mengintegrasikan atribut sosial ekonomi dan lingkungan (misalnya, pendidikan orang tua, pendapatan, sanitasi) untuk meningkatkan realisme prediktif.
2. Mengintegrasikan model ini ke dalam sistem deteksi dini stunting digital untuk pekerja kesehatan masyarakat.
3. Membandingkan C4.5 dengan metode ensemble seperti Random Forest atau XGBoost untuk menilai skalabilitas.

4. Kesimpulan

Penelitian ini menunjukkan bahwa algoritma pohon keputusan C4.5 memiliki kinerja yang menjanjikan untuk mendeteksi stunting pada balita secara dini. Dengan memanfaatkan tiga atribut utama usia, tinggi badan, dan jenis kelamin, model mencapai akurasi sebesar 99,87% serta menghasilkan aturan keputusan yang interpretatif dan konsisten, sehingga berpotensi diintegrasikan ke dalam sistem kesehatan digital yang mendukung inisiatif RAN PASTI (Rencana Aksi Nasional Percepatan Penurunan Angka Stunting Indonesia**)** di Indonesia. Ke depan, penelitian dapat diperluas dengan penggabungan fitur sosial ekonomi, pola makan, dan lingkungan untuk meningkatkan realisme prediksi, disertai validasi eksternal maupun longitudinal, serta penerapan model dalam aplikasi kesehatan berbasis mobile atau web untuk pemantauan waktu nyata dan pencegahan stunting secara nasional.

Referensi

[1] Pemerintah Republik Indonesia, “Peraturan Presiden Republik Indonesia Nomor 72 Tahun 2021 Tentang Percepatan Penurunan Stunting,” 2021.

[2] A. Soliman *et al.*, “Early and long-term consequences of nutritional stunting: From childhood to adulthood,” *Acta Biomedica*, vol. 92, no. 1, Mar. 2021, doi: 10.23750/abm.v92i1.11346.

[3] Kementerian Kesehatan Republik Indonesia, “SSGI 2024: Prevalensi Stunting Nasional Turun Menjadi 19,8%.” Accessed: Nov. 11, 2025. [Online]. Available: https://kemkes.go.id/id/ssgi-2024-prevalensi-stunting-nasional-turun-menjadi-198

[4] UNICEF, “Climate Change and Nutrition in Indonesia. A review of the evidence for policy and programme strengthening. United Nations Children’s Fund.,” Jakarta, Indonesia, 2024.

[5] Pemerintah Republik Indonesia, "Strategi Nasional Percepatan Pencegahan dan Penurunan Stunting 2025-2029," 2025.
[6] J. Han, M. Kamber, and J. Pei, *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*. Elsevier, 2012.
[7] J. R. Quinlan, "Improved Use of Continuous Attributes in C4.5," 1996.
[8] R. xsanal Hakim *et al.*, "Penerapan Algoritma C4.5 Untuk Prediksi Anak Stunting Di Kota Pagar Alam," 2024.
[9] P. C. Algoritma, P. Klasifikasi Status Gizi Balita di Posyandu Desa Sukalilah Cibatu Kabupaten Garut Jawa Barat Sri Lestari, R. Amanda Amalia, and S. Lestari, "Penerapan Algoritma C.45 Pada Klasifikasi Status Gizi Balita di Posyandu Desa Sukalilah Cibatu Kabupaten Garut Jawa Barat," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 177–182, 2023, doi: 10.55338/saintek.v5i1.1375.
[10] Y. M. Brekmans Darkel, L. Ermilinda, G. Kurniawan Al Yulianto, and C. Fransiska Pacolinus, "Implementasi Model Algoritma C4.5 Untuk Klasifikasi Status Stunting Di Kabupaten Sikka Implementation of the C4.5 Algorithm Model for Classification of Stunting Status in Sikka Regency," 2024.
[11] N. Baharun, N. F. M. Razi, S. Masrom, N. A. M. Yusri, and A. S. A. Rahman, "Auto Modellingfor Machine Learning: A Comparison Implementation between Rapid Miner and Python," *International Journal of Emerging Technology and Advanced Engineering*, vol. 12, no. 5, pp. 15–27, May 2022, doi: 10.46338/ijetae0522_03.
[12] L. Kovács and H. Ghous, "Efficiency comparison of Python and RapidMiner," *Multidiszciplináris Tudományok*, vol. 10, no. 3, pp. 212–220, 2020, doi: 10.35925/j.multi.2020.3.26.
[13] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, "The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction," *Clean Eng Technol*, vol. 15, Aug. 2023, doi: 10.1016/j.clet.2023.100664.
[14] A. Mosavi, "Multiple Criteria Decision-Making Preprocessing Using Data Mining Tools," *IJCSI International Journal of Computer Science Issues*, vol. 7, no. 1, 2010, [Online]. Available: www.IJCSI.org
[15] Fakhruddin Fakhruddin and Sefrika Entas, "Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Prediksi Kualitas Tidur pada Kesehatan," *Vitamin : Jurnal ilmu Kesehatan Umum*, vol. 3, no. 4, pp. 217–234, Sep. 2025, doi: 10.61132/vitamin.v3i4.1773.