



JURNAL STATISTIKA DAN SAINS DATA

Volume 3 Edisi 2, April, 2026, 164-172

<https://journal.student.uny.ac.id/index.php/jssd>

<https://dx.doi.org/10.21831/jssd.v3i2.24754>

Penerapan algoritma *Decision Tree* untuk klasifikasi status *stunting* pada balita di Indonesia

Miftakhiyah Fazza Baita, Universitas Muhammadiyah Semarang

Siti Nurhalisa*, Universitas Muhammadiyah Semarang

M. Al Haris, Universitas Muhammadiyah Semarang

Saeful Amri, Universitas Muhammadiyah Semarang

*e-mail: stnurhalisaa12@gmail.com

Abstrak. *Stunting* merupakan permasalahan kesehatan masyarakat yang serius di Indonesia. Kondisi ini terjadi akibat kekurangan gizi kronis yang berlangsung dalam jangka waktu lama, sehingga menyebabkan anak balita memiliki tinggi badan lebih rendah dibandingkan standar usianya. Selain menghambat pertumbuhan fisik, *stunting* juga berdampak pada perkembangan kognitif dan produktivitas anak di masa depan. Berdasarkan Survei Status Gizi Indonesia (SSGI) tahun 2022, prevalensi *stunting* di Indonesia mencapai 21,6%, melebihi ambang batas yang ditetapkan WHO, yaitu kurang dari 20%. Penelitian ini bertujuan untuk membangun model klasifikasi status *stunting* pada balita menggunakan algoritma *Decision Tree* C4.5 Hasil penelitian menunjukkan bahwa model *decision tree* C4.5 yang dibangun menggunakan 10.000 data balita dari dataset mampu mencapai akurasi 57,80%, *precision* 79,93%, dan *recall* 30,29%, dengan atribut tinggi badan dan umur sebagai pemisah utama. Hasil ini menunjukkan bahwa *decision tree* dapat digunakan sebagai model awal deteksi *stunting*, meski perlu peningkatan performa melalui teknik *balancing* data.

Kata kunci: *Stunting, Decision tree, klasifikasi, balita, gizi anak.*

Abstract. *Stunting is a serious public health problem in Indonesia. This condition is caused by chronic nutritional deficiencies over a prolonged period, resulting in children having a lower height than the standard for their age. In addition to inhibiting physical growth, stunting also affects cognitive development and future productivity. According to the 2022 Indonesian Nutritional Status Survey (SSGI), the prevalence of stunting in Indonesia reached 21.6%, exceeding the threshold set by the World Health Organization (WHO), which is below 20%. This study aims to develop a classification model for toddler stunting status using the C4.5 Decision Tree algorithm. The results showed that the C4.5 Decision Tree model achieved an accuracy of 57.80%, a precision of 79.93%, and a recall of 30.29%, with height and age identified as the main splitting attributes. These findings indicate that the Decision Tree algorithm can be used as an initial model for stunting detection, although performance improvements through data balancing techniques are still required.*

Keywords: *Stunting, Decision tree, classification, toddlers, child nutrition.*

PENDAHULUAN

Kekurangan gizi kronis tetap menjadi isu multidimensi yang signifikan di negara-negara berkembang, termasuk Indonesia, dan berimplikasi langsung terhadap pembangunan kualitas sumber daya manusia. Salah satu manifestasi klinis dari kondisi tersebut adalah *stunting*, yakni gangguan pertumbuhan linear pada anak balita yang diakibatkan oleh defisiensi nutrisi dalam jangka waktu panjang, ditandai dengan tinggi badan yang berada di bawah standar menurut usia (Aurima et al., 2021). *Stunting* bukan sekadar persoalan keterlambatan pertumbuhan fisik, tetapi juga berdampak luas dan mendalam terhadap berbagai aspek perkembangan anak. Anak-anak yang mengalami *stunting* berisiko mengalami hambatan dalam perkembangan otak, penurunan kemampuan kognitif, keterbatasan dalam proses pembelajaran, dan rendahnya produktivitas di masa depan. Dalam jangka panjang, *stunting* dapat menyebabkan penurunan kualitas sumber daya manusia, berkurangnya kapasitas ekonomi individu, hingga memperburuk ketimpangan sosial dan ekonomi di masyarakat (Kemenkes RI, 2022). Oleh karena itu, *stunting* telah menjadi isu strategis nasional yang membutuhkan perhatian lintas sektor serta intervensi yang menyeluruh dan terarah.

Menurut data dari *World Health Organization* (WHO), *stunting* masih menjadi salah satu permasalahan gizi utama di dunia dan dialami oleh sekitar 149 juta anak balita. Kondisi ini menjadikan *stunting* sebagai indikator penting dalam menilai tingkat kesehatan masyarakat serta kualitas pembangunan suatu negara. Secara global, jumlah balita yang mengalami *stunting* masih tergolong tinggi, yaitu sekitar 149,2 juta pada tahun 2020, 148,1 juta pada tahun 2022, dan 150,2 juta pada tahun 2024, yang menunjukkan bahwa masalah ini masih menjadi tantangan kesehatan masyarakat dunia. Di Indonesia, prevalensi *stunting* berdasarkan Survei Status Gizi Indonesia (SSGI) menunjukkan tren penurunan dari 24,4% pada tahun 2021 menjadi 21,6% pada tahun 2022, walaupun menunjukkan tren penurunan dibandingkan tahun-tahun sebelumnya, angka tersebut masih melebihi batas maksimal yang ditetapkan WHO, yaitu di bawah 20% (J et al., 2022).

Penanggulangan *stunting* telah menjadi salah satu fokus utama program nasional di Indonesia. Pemerintah menargetkan penurunan angka *stunting* hingga mencapai prevalensi 14% pada tahun 2024 sebagai bagian dari Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020–2024 (Bappenas, 2020). Target ini sejalan dengan komitmen global dalam Tujuan Pembangunan Berkelanjutan (*Sustainable Development Goals/SDGs*), khususnya pada tujuan ke-2 yaitu mengakhiri kelaparan dan segala bentuk malnutrisi. Untuk mewujudkan target tersebut, dibutuhkan strategi yang tidak hanya bersifat kuratif, tetapi juga preventif, salah satunya melalui pemanfaatan teknologi informasi untuk mendeteksi risiko *stunting* secara dini.

Dalam bidang kesehatan masyarakat, pendekatan *data mining* dan *machine learning* semakin banyak digunakan untuk menganalisis data dalam jumlah besar secara efisien dan akurat. Teknologi ini telah dimanfaatkan dalam berbagai aplikasi, mulai dari diagnosis penyakit, prediksi risiko pasien, hingga klasifikasi status gizi anak. Salah satu metode yang populer dalam klasifikasi adalah algoritma *decision tree*. Metode ini bekerja dengan membagi data ke dalam beberapa subset berdasarkan atribut tertentu, membentuk struktur pohon bercabang yang menyajikan informasi dalam format yang mudah dipahami bahkan oleh pihak non-teknis (S. Rasoul Safavian, 1991).

Keunggulan utama dari algoritma *decision tree*, khususnya tipe C4.5, adalah kemampuannya dalam menangani data numerik maupun kategorikal tanpa perlu melakukan proses normalisasi data terlebih dahulu (Quinlan, 1993). Selain itu, algoritma ini mampu menghasilkan model yang dapat divisualisasikan secara intuitif, sehingga mempermudah interpretasi hasil bagi pengambil kebijakan atau tenaga kesehatan. Penelitian oleh (Nirmalasari, 2020) menunjukkan bahwa algoritma *decision tree* efektif dalam mengklasifikasikan

permasalahan gizi pada anak balita di Indonesia, dengan tingkat akurasi yang cukup tinggi dibandingkan metode lainnya. Hal ini menjadikan *decision tree* sebagai salah satu pendekatan yang relevan untuk mendukung pengambilan keputusan dalam intervensi gizi masyarakat.

Dengan kemampuannya yang fleksibel dan interpretatif, *decision tree* telah diterapkan dalam berbagai sektor, seperti keuangan, pemasaran, pendidikan, serta kesehatan masyarakat. Dalam konteks penanganan *stunting*, algoritma ini dapat digunakan untuk membangun sistem klasifikasi yang mampu mengidentifikasi balita yang berisiko tinggi mengalami *stunting* berdasarkan atribut seperti usia, jenis kelamin, berat badan, tinggi badan, serta riwayat kesehatan keluarga. Hasil klasifikasi ini dapat membantu petugas kesehatan dalam merancang program intervensi gizi yang lebih terarah dan efisien.

Sejumlah penelitian sebelumnya juga telah menunjukkan efektivitas algoritma *decision tree*, termasuk tipe C4.5 dalam klasifikasi status gizi balita. Misalnya, penelitian oleh (Prasetio, 2021) menunjukkan bahwa algoritma ini dapat mengklasifikasikan status gizi dengan akurasi yang tinggi menggunakan parameter seperti usia, berat badan, dan tinggi badan. Menurut (Mulyono et al., 2023), mengembangkan sistem berbasis web yang memanfaatkan algoritma C4.5 dengan akurasi mencapai 93,62%. Penelitian lain oleh (Hanif & Muntari, 2024), (Hakim et al., 2024), serta (Bulkisah et al., 2024) menunjukkan keberhasilan *decision tree* dalam prediksi *stunting* di berbagai daerah dengan akurasi berkisar antara 88% hingga 99,18%. Perbandingan dengan metode lain seperti Naïve Bayes dan SVM juga dilakukan, dan hasilnya menunjukkan bahwa *decision tree* seringkali memberikan performa klasifikasi terbaik (Hanif & Muntari, 2024). Studi-studi ini memperkuat landasan metodologis penelitian ini dan mendorong eksplorasi lebih lanjut terhadap penerapannya pada data skala nasional.

Dengan mempertimbangkan urgensi penanganan *stunting* dan potensi pemanfaatan teknologi informasi, penelitian ini berfokus pada penerapan algoritma *decision tree* untuk mengklasifikasikan status *stunting* pada balita di Indonesia. Penelitian ini bertujuan untuk membangun model klasifikasi status *stunting* menggunakan data sekunder dari platform *Kaggle*, yang mencakup atribut seperti usia, jenis kelamin, berat badan, tinggi badan, dan status *stunting*. Kebaruan dari penelitian ini terletak pada penerapan algoritma *decision tree* dalam konteks klasifikasi *stunting* berbasis data terbuka yang belum banyak dieksplorasi, serta potensi penerapan model yang dihasilkan sebagai alat bantu pengambilan keputusan dalam program intervensi gizi masyarakat. Diharapkan, hasil penelitian ini dapat memberikan kontribusi nyata dalam upaya deteksi dini dan penanganan *stunting* secara lebih efektif, efisien, dan berbasis data.

METODE

Sumber Data

Data penelitian ini bersumber dari dataset *Stunting Balita Detection* yang tersedia di platform *Kaggle*, dikembangkan oleh Rendi Putra. Dataset tersedia dalam format file *.csv* dengan total 120.999 data balita, yang terdiri atas 33.684 balita dengan status *stunting* dan 87.315 balita dengan status tidak *stunting*. Data ini memuat empat atribut utama, yaitu umur (dalam bulan), jenis kelamin, tinggi badan (dalam cm), dan status gizi. Rincian atribut yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Atribut Data *Stunting* Balita

No.	Nama Variabel	Jenis Data	Keterangan
1.	Umur (bulan)	Numerik	Usia balita dalam satuan bulan
2.	Jenis Kelamin	Kategorikal	Laki-laki atau Perempuan
3.	Tinggi Badan (cm)	Numerik	Tinggi badan balita dalam satuan cm
4.	Status Gizi	Kategorikal	Target klasifikasi (<i>Stunted</i> , <i>normal</i> , dll.)

Algoritma *Decision tree*

Algoritma decision tree merupakan salah satu metode klasifikasi berbasis supervisi yang digunakan untuk membangun model keputusan dalam bentuk struktur pohon bercabang. Metode ini bekerja dengan membagi dataset ke dalam subset-subset berdasarkan atribut-atribut tertentu hingga mencapai kondisi klasifikasi yang optimal (Hanif & Muntiari, 2024).

Pada penelitian ini, digunakan algoritma decision tree tipe C4.5 untuk membangun model klasifikasi status stunting pada balita. Algoritma C4.5 dipilih karena memiliki keunggulan dalam menangani atribut bertipe numerik maupun kategorikal, mampu mengelola data yang memiliki missing value, serta menghasilkan model yang intuitif dan mudah diinterpretasikan. Selain itu, C4.5 melakukan proses pruning untuk mengurangi kompleksitas model dan menghindari overfitting, sehingga meningkatkan generalisasi terhadap data baru. Proses pembentukan pohon keputusan dalam algoritma C4.5 dilakukan melalui beberapa tahap perhitungan, sebagai berikut:

1. *Entropy dataset*

Entropy mengukur tingkat ketidakpastian (*randomness*) dalam dataset:

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2 (p_i) \quad (1)$$

S = dataset pada node tertentu

n = jumlah kelas (misal: *stunting*, tidak *stunting*)

p_i = proporsi data pada kelas ke- i di node tersebut.

2. *Entropy setiap class pada atribut*

Untuk setiap atribut A , bagi data ke dalam subset S_v berdasarkan nilai-nilai atribut A .

Hitung *entropy* masing-masing subset:

$$Entropy(S) = \sum_{i=1}^n \frac{|S_i|}{|S|} \cdot Entropy(S_i) \quad (2)$$

S = Himpunan data

A = semua nilai unik dari atribut A

N = Jumlah partisi atribut A

$|S_i|$ = Jumlah data pada partisi ke- i

$|S|$ = Jumlah data pada S

3. *Gain untuk setiap atribut*

Information Gain mengukur seberapa besar pengurangan entropy jika data dibagi berdasarkan atribut tertentu:

$$Gain(A) = Entropy(S) - Entropy_A(S) \quad (3)$$

4. *Information split Info*

Split Information digunakan untuk mengukur seberapa banyak dan seberapa merata data dibagi oleh atribut:

$$SplitInfo(S, A) = - \sum_{i=1}^C \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (4)$$

S = Himpunan data

A = semua nilai unik dari atribut A

N = Jumlah partisi atribut A

$|S_i|$ = Jumlah data pada partisi ke- i

$|S|$ = Jumlah data pada S

5. *Information Gain Ratio (khusus C4.5)*

Gain Ratio digunakan untuk menormalkan nilai *Gain* terhadap *Split Info*, agar pemilihan atribut tidak berat sebelah terhadap atribut dengan banyak nilai. *Gain ratio* dihitung dengan cara:

$$Gain\ Ratio(A) = \frac{Gain(S,A)}{SplitInfo(A)} \quad (5)$$

S = Himpunan data

A = semua nilai unik dari atribut A

Atribut dengan *Information Gain* tertinggi dipilih sebagai pemisah pada node tersebut.

Tahapan Implementasi

Tahapan Implementasi klasifikasi status stunting pada balita dilakukan menggunakan perangkat lunak RapidMiner Studio, dengan tahapan analisis data yang dilakukan adalah sebagai berikut:

1. Data dikumpulkan dengan mengunduh dataset *Stunting Balita Detection* dari platform *Kaggle*. Dataset ini berisi informasi mengenai umur, jenis kelamin, tinggi badan, dan status gizi balita.
2. Selanjutnya, pada tahap pra-pemrosesan, dataset diimpor ke dalam RapidMiner, kemudian dilakukan pemeriksaan data, penghapusan data duplikat, serta pengecekan *missing value*. Atribut kategorik, seperti jenis kelamin dan status gizi, kemudian disesuaikan tipenya menggunakan operator *Nominal to Numerical* dan *Set Role* untuk menentukan atribut target klasifikasi (Safavian & Landgrebe, 1991)
3. Dataset kemudian dibagi menjadi dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji. Pembagian dilakukan secara acak menggunakan teknik *random splitting* untuk menjaga proporsi data yang representatif.
4. Model klasifikasi dibangun menggunakan operator *Decision Tree* pada *software* RapidMiner dengan algoritma C4.5. Proses pembentukan pohon keputusan dilakukan berdasarkan nilai *Gain Ratio* tertinggi yang diperoleh dari perhitungan *entropy* dan *information gain*, sehingga menghasilkan struktur pohon keputusan yang optimal (Quinlan, 1993).
5. Model yang telah dibangun dievaluasi menggunakan data uji dengan evaluasi yang dilakukan berdasarkan nilai akurasi, *precision*, *recall*, dan *F1-score*. Selain itu, *confusion matrix* digunakan untuk melihat distribusi hasil klasifikasi pada masing-masing kelas.
6. Struktur pohon keputusan yang dihasilkan divisualisasikan menggunakan fasilitas *Tree View* pada RapidMiner untuk mempermudah interpretasi pola klasifikasi dan aturan keputusan yang terbentuk.
7. Tahap terakhir dilakukan dengan menginterpretasikan hasil klasifikasi dan performa model untuk mengidentifikasi atribut yang berpengaruh terhadap status stunting balita serta menilai efektivitas algoritma *Decision Tree* C4.5 dalam proses klasifikasi.

HASIL DAN PEMBAHASAN

Hasil

Dataset yang digunakan dalam penelitian ini merupakan hasil pengambilan sebanyak 10.000 baris pertama dari keseluruhan data yang tersedia, yaitu 120.999 data balita. Pengambilan data ini dilakukan karena keterbatasan lisensi perangkat lunak *RapidMiner Studio* versi gratis yang hanya mendukung pengolahan maksimal 10.000 baris data. Meskipun demikian, sampel 10.000 data tersebut tetap dianggap representatif karena tetap mencerminkan

variasi status gizi yang ada. Distribusi status gizi pada 10.000 data ini dapat dilihat pada Tabel 2.

Tabel 2. Distribusi Kategori Status Gizi Balita dalam Dataset

No.	Kategori Status Gizi	Jumlah Balita
1.	<i>Severely Stunted</i>	3.206
2.	<i>Stunted</i>	987
3.	Normal	4.461
4.	Tinggi	1.346
Total		10.000

Berdasarkan **Tabel 2**, kategori balita dengan status gizi *Normal* merupakan kelas terbanyak dengan jumlah 4.461 data, diikuti oleh kategori *Severely Stunted* sebanyak 3.206 data, *Tinggi* sebanyak 1.346 data, dan *Stunted* sebanyak 987 data. Distribusi tersebut menunjukkan adanya ketidakseimbangan kelas yang berpotensi memengaruhi hasil klasifikasi.

Model klasifikasi dibangun menggunakan algoritma Decision Tree C4.5 dengan proporsi data latih sebesar 80% dan data uji sebesar 20%. Pemilihan atribut pada setiap node dilakukan berdasarkan nilai *Gain Ratio* tertinggi. Hasil klasifikasi model ditampilkan dalam bentuk *confusion matrix* pada Tabel 3.

Tabel 3. *Confusion Matrix* Hasil Klasifikasi Decision tree

	<i>True Stunted</i>	<i>True Tinggi</i>	<i>True Normal</i>	<i>True Severely Stunted</i>	<i>Class Precision</i>
<i>Pred. Stunted</i>	11	0	0	6	64,71%
<i>Pred. Tinggi</i>	0	42	2	0	95,45%
<i>Pred. Normal</i>	215	309	1088	312	56,55%
<i>Pred. Severely Stunted</i>	0	0	0	15	100,00%
<i>Class Recall</i>	4,7%	11,97%	99,82%	4,50%	

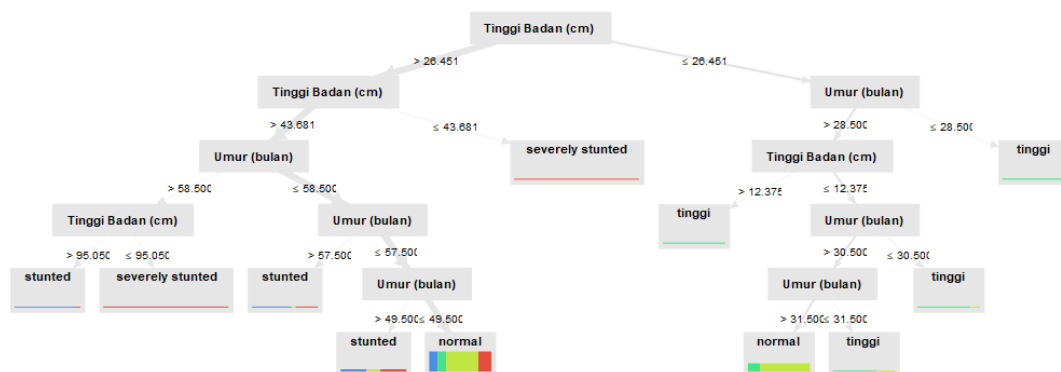
Berdasarkan *confusion matrix*, model mampu mengklasifikasikan kelas *Normal* dengan sangat baik, ditunjukkan oleh nilai *class recall* sebesar 99,82%. Sebaliknya, kelas *Stunted* dan *Severely Stunted* memiliki nilai *class recall* yang rendah, masing-masing sebesar 4,70% dan 4,50%. Hasil Evaluasi model ditampilkan pada Tabel 4.

Tabel 4. Hasil Evaluasi *Confusion Matrix*

No.	Metrik Evaluasi	Nilai
1.	Akurasi	57,80 %
2.	<i>Precision</i>	79,93 %
3.	<i>Recall</i>	30,29 %
4.	<i>F1-score</i>	43,96 %

Berdasarkan Tabel 4, Model *Decision Tree* C4.5 menghasilkan akurasi sebesar 57,80%, *precision* sebesar 79,93%, *recall* sebesar 30,29%, dan *F1-score* sebesar 43,96%. Selain itu, struktur pohon keputusan yang terbentuk menunjukkan bahwa atribut tinggi badan menjadi akar pohon, sedangkan atribut umur berperan sebagai pemisah lanjutan dalam proses klasifikasi status gizi balita. Hasil penelitian kemudian divisualisasikan dalam bentuk struktur pohon

menggunakan *tree view* pada *RapidMiner* pada **Gambar 2**. Pohon memperlihatkan bahwa atribut tinggi badan menjadi akar pohon, diikuti oleh umur sebagai cabang selanjutnya. Beberapa aturan klasifikasi yang terbentuk memperlihatkan pemisahan yang logis, namun banyak percabangan yang sangat spesifik terhadap data latih. Hal ini mengindikasikan adanya *overfitting*.



Gambar 2. Visualisasi Pohon Keputusan

Pembahasan

Berdasarkan pohon keputusan yang terbentuk, variabel tinggi badan (cm) merupakan atribut yang paling dominan dalam menentukan status gizi balita karena digunakan sebagai pemisah utama pada akar pohon. Pemisahan awal terjadi pada tinggi badan 26,451 cm yang membagi data ke dalam dua kelompok besar. Pada kelompok dengan tinggi badan $\leq 26,451$ cm, klasifikasi cenderung mengarah pada status tinggi dan normal dengan umur sebagai variabel pendukung. Sebaliknya, pada kelompok dengan tinggi badan $> 26,451$ cm, klasifikasi lebih banyak mengarah pada status severely stunted, stunted, dan normal melalui beberapa percabangan lanjutan berdasarkan kombinasi tinggi badan dan umur balita.

Secara keseluruhan, pohon keputusan menunjukkan bahwa kombinasi tinggi badan dan umur mampu membedakan status gizi balita ke dalam kategori tinggi, normal, stunted, dan severely stunted. Semakin tinggi posisi percabangan pada pohon, semakin besar pengaruh variabel tersebut terhadap proses klasifikasi. Banyaknya percabangan dan batas pemisah yang spesifik menunjukkan bahwa model menangkap karakteristik data latih secara rinci untuk meningkatkan ketepatan klasifikasi, sehingga pola hubungan antara tinggi badan, umur, dan status gizi balita dapat dijelaskan dengan baik oleh model yang terbentuk.

Berdasarkan hasil evaluasi, model menghasilkan akurasi sebesar 57,80%, *precision* sebesar 79,93%, *recall* sebesar 30,29%, dan F1-score sebesar 43,96%. Nilai *precision* yang relatif tinggi menunjukkan bahwa sebagian besar hasil prediksi yang diberikan model sudah tepat. Namun, nilai *recall* yang rendah menunjukkan bahwa model masih mengalami kesulitan dalam mengenali kelas minoritas, khususnya kategori *Stunted* dan *Severely Stunted*. Kondisi ini diduga dipengaruhi oleh distribusi data yang tidak seimbang, di mana kategori Normal memiliki jumlah data yang jauh lebih besar dibandingkan kategori lainnya sehingga model cenderung lebih baik dalam mengenali kelas mayoritas.

Hasil klasifikasi menunjukkan bahwa algoritma *Decision Tree C4.5* mampu membentuk aturan klasifikasi yang mudah diinterpretasikan, meskipun performanya dipengaruhi oleh karakteristik data, seperti jumlah data, distribusi kelas, dan tahapan prapemrosesan. Hanif &

Muntiari, (2024) dalam analisisnya menghasilkan akurasi sebesar 99,00% dengan menerapkan teknik *balancing* data, sedangkan (Prasetio, 2021) memperoleh akurasi sebesar 88,00% tanpa teknik *balancing*. Temuan tersebut menunjukkan bahwa kinerja model klasifikasi tidak hanya dipengaruhi oleh penggunaan teknik *balancing*, tetapi juga oleh karakteristik dataset dan proses pengolahan data yang diterapkan pada masing-masing penelitian.

Pada penelitian ini, model dibangun tanpa penerapan teknik *balancing* data. Meskipun demikian, hasil yang diperoleh menunjukkan bahwa algoritma Decision Tree C4.5 tetap mampu mengidentifikasi pola dalam data dan menghasilkan aturan klasifikasi yang dapat dijelaskan secara jelas. Temuan ini menunjukkan bahwa metode C4.5 memiliki potensi untuk digunakan dalam klasifikasi status stunting balita, sementara penerapan teknik *balancing* dapat dipertimbangkan pada penelitian selanjutnya sebagai salah satu alternatif untuk meningkatkan kemampuan model dalam mengenali seluruh kelas secara lebih merata.

SIMPULAN

Model klasifikasi status stunting pada balita menggunakan algoritma *Decision Tree* C4.5 berhasil dibangun menggunakan 10.000 data balita yang berasal dari dataset *Stunting Balita Detection* pada platform *Kaggle*. Berdasarkan hasil evaluasi, model menghasilkan akurasi sebesar 57,80%, *precision* sebesar 79,93%, *recall* sebesar 30,29%, dan *F1-score* sebesar 43,96%. Hasil tersebut menunjukkan bahwa model mampu melakukan klasifikasi status stunting, meskipun masih mengalami keterbatasan dalam mengenali kelas minoritas.

Struktur pohon keputusan yang terbentuk menunjukkan bahwa atribut tinggi badan dan umur merupakan faktor yang paling berpengaruh dalam proses klasifikasi status gizi balita. Secara keseluruhan, algoritma Decision Tree C4.5 dapat digunakan sebagai model awal untuk klasifikasi status stunting pada balita. Namun, penelitian selanjutnya disarankan menerapkan teknik *balancing* data atau pengembangan metode klasifikasi lainnya untuk meningkatkan kemampuan model dalam mendeteksi kelas minoritas dan memperbaiki performa klasifikasi.

DAFTAR PUSTAKA

- Aurima, J., Susaldi, S., Agustina, N., Masturoh, A., Rahmawati, R., & Tresiana Monika Madhe, M. (2021). Faktor-Faktor yang Berhubungan dengan Kejadian Stunting pada Balita di Indonesia. *Open Access Jakarta Journal of Health Sciences*, 1(2), 43–48. <https://doi.org/10.53801/oajjhs.v1i3.23>
- Bulkisah, S. B., Astuti, R., & Bahtiar, A. (2024). Implementasi Data Mining Algoritma Decision Tree untuk Klasifikasi Status Gizi Balita di Kecamatan Ciledug. *Jurnal Ilmiah Informatika Komputer*, 29(1), 1–12. <https://doi.org/10.35760/ik.2024.v29i1.10346>
- Hakim, R. X., Putrawansyah, F., & Syahri, R. (2024). Penerapan Algoritma C4.5 Untuk Prediksi Anak Stunting Di Kota Pagar Alam. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 2469–2478. <https://doi.org/10.36040/jati.v8i2.9301>
- Hanif, K. H., & Muntiari, N. R. (2024). Penerapan Algoritma Decision Tree, Svm, Naive Bayes Dalam Deteksi Stunting Pada Balita. *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 8(1), 105–109. <https://doi.org/10.46880/jmika.vol8no1.pp105-109>
- J, R. F., Huljannah, N., & Rochmah, T. N. (2022). Stunting Prevention Program in Indonesia: A SYSTEMATIC REVIEW. *Media Gizi Indonesia*, 17(3), 281–292. <https://doi.org/10.20473/mgi.v17i3.281-292>
- Kemenkes RI. (2022). Hasil Survei Status Gizi Indonesia (SSGI) 2022. *Kemenkes*, 1–150.

- Mulyono, M., Budianita, E., Nazir, A., & Syafria, F. (2023). Klasifikasi Status Stunting Balita Menggunakan Metode Naïve Bayes Gaussian Berbasis Web. *Jurnal Informatika Universitas Pamulang*, 8(3), 399–406. <https://doi.org/10.32493/informatika.v8i3.33399>
- Nirmalasari, N. O. (2020). Stunting Pada Anak : Penyebab dan Faktor Risiko Stunting di Indonesia. *Qawwam: Journal For Gender Mainstreaming*, 14(1), 19–28. <https://doi.org/10.20414/Qawwam.v14i1.2372>
- Prasetio, A. (2021). Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 4(3), 209–214. <https://doi.org/10.32672/jnkti.v4i3.2983>
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.
- S. Rasoul Safavian, D. L. (1991). A survey of decision tree classifier methodology - *Systems, Man and Cybernetics*, *IEEE Transactions on*. 21(3), 660–674. <https://doi.org/10.9790/5933-0612104113>
- Safavian, S., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3), 660–674. <https://doi.org/978-1558602380>