PENERAPAN DATA MINING UNTUK PREDIKSI STUNTING PADA BALITA MENGGUNAKAN ALGORITMA C4.5

Nurul Qisthi[1*], Dian Kasoni[2], Liesnaningsih[3], Nofitri Heriyani[4]

[1]Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Poltek Cirebon
[2]Program Studi Teknik Informatika, STMIK Antar Bangsa
[3,4]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang
nurul.qisthi@stikompoltek.ac.id[1*], dhekalearning@gmail.com[2], liesnaningsih@ft-umt.ac.id[3], [4]nofitri.heriyani@ft-umt.ac.id
Penulis Korespondensi: nurul.qisthi@stikompoltek.ac.id

ABSTRAK

Stunting pada balita merupakan masalah kesehatan serius dengan dampak jangka panjang pada perkembangan fisik dan kognitif anak. Tingginya angka *stunting* saat ini menunjukkan perlunya upaya intensif dan inovatif untuk penanganannya, termasuk melalui pendekatan berbasis data dan teknologi. Identifikasi risiko *stunting* umumnya dilakukan secara manual melalui pengukuran antropometri, yang memerlukan waktu, tenaga ahli, dan rentan terhadap kesalahan. Penelitian ini bertujuan untuk menerapkan pendekatan *data mining* menggunakan algoritma C4.5 dalam prediksi risiko *stunting* pada balita. Algoritma C4.5 dipilih karena kemampuannya menangani data kategori dan numerik serta menghasilkan model klasifikasi berbentuk pohon keputusan yang mudah diinterpretasi. C4.5 memanfaatkan informasi gain untuk memilih fitur paling relevan, sehingga efektif dalam membagi data dan mengklasifikasikan status gizi balita. Penelitian ini menghasilkan model prediksi *stunting* yang memanfaatkan fitur-fitur seperti tinggi badan, umur, dan jenis kelamin dengan algoritma C4.5. Model ini menunjukkan akurasi tinggi sebesar 98.88%, dengan nilai precision, recall, dan F1-*score* yang konsisten di setiap kelas. Selain itu, nilai ROC-AUC sebesar 98.98% menunjukkan kemampuan model yang baik dalam membedakan kelas status gizi yang berbeda. Tingginya akurasi ini mengindikasikan bahwa algoritma C4.5 dapat diandalkan untuk mendeteksi risiko *stunting* secara efektif, sehingga dapat mendukung tenaga kesehatan dan pembuat kebijakan dalam melakukan intervensi yang lebih tepat dan akurat.

Kata Kunci: C4.5, *Data Mining*, Prediksi, Status Gizi, *Stunting*

PENDAHULUAN

Stunting atau kondisi gagal tumbuh pada balita merupakan masalah kesehatan yang serius dan berdampak jangka panjang terhadap perkembangan fisik serta kognitif anak. Menurut data UNICEF tahun 2023, sekitar 149 juta anak di bawah usia lima tahun di seluruh dunia mengalami *stunting* (Rahman et al., 2023). Di Indonesia, prevalensi *stunting* masih berada pada angka 24,4% pada tahun 2022, berdasarkan data dari Kementerian Kesehatan, yang masih jauh dari target nasional sebesar 14% pada tahun 2024 (Octavia et al., 2023). Tingginya angka *stunting* ini menyoroti perlunya upaya intensif dan inovatif untuk menanganinya, termasuk melalui pendekatan berbasis data dan teknologi. Identifikasi dan prediksi *stunting* saat ini sebagian besar dilakukan secara manual melalui pengukuran antropometri seperti tinggi dan berat badan. Meskipun efektif, pendekatan manual ini memiliki keterbatasan, seperti ketergantungan pada tenaga ahli, membutuhkan waktu, serta rentan terhadap kesalahan manusia, terutama di daerah yang memiliki sumber daya terbatas. Dengan banyaknya variabel yang terlibat, analisis manual semakin sulit dilakukan untuk menentukan risiko *stunting* dengan cepat dan akurat (Nurmalasari et al., 2024).

Pendekatan *data mining* atau penambangan data muncul sebagai solusi yang potensial untuk mengatasi keterbatasan dalam identifikasi manual (Nurmalasari et al., 2024). *Data mining* adalah proses analisis data yang bertujuan menemukan pola atau hubungan tersembunyi dari dataset yang besar, yang memungkinkan pengambilan keputusan yang lebih cepat dan berdasarkan data yang lebih kaya (Borman

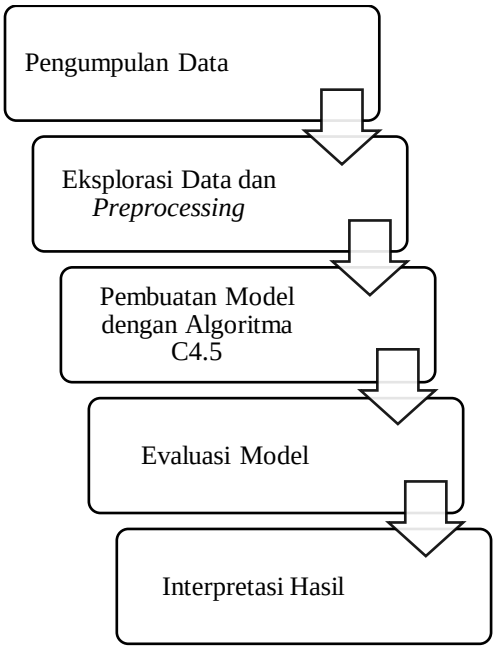
& Wati, 2020). Di antara berbagai algoritma *data mining*, algoritma C4.5 adalah salah satu metode pohon keputusan yang sangat populer dan cocok untuk masalah klasifikasi (Matondang et al., 2021). Algoritma ini bekerja dengan memecah dataset berdasarkan kriteria gain ratio, yang mampu memilih fitur terbaik untuk klasifikasi dengan mempertimbangkan faktor informasi (Fathurrohman et al., 2024). Kelebihan utama C4.5 adalah kemampuannya untuk menangani data kategori dan numerik, serta mengatasi data yang hilang (Maulana et al., 2022).

Algoritma C4.5 dipilih dalam penelitian ini karena mampu menghasilkan model klasifikasi yang mudah diinterpretasikan dalam bentuk pohon keputusan, yang mempermudah pemahaman bagi tenaga kesehatan dan pengambil kebijakan. Selain itu, algoritma ini memiliki mekanisme pruning untuk mengurangi kompleksitas pohon, yang dapat mengurangi risiko *overfitting* dan meningkatkan akurasi prediksi pada data baru (Arfandi et al., 2021). Dengan kelebihan ini, algoritma C4.5 dapat diandalkan untuk mengidentifikasi faktor-faktor risiko utama *stunting* dan memprediksi risiko *stunting* pada balita dengan tingkat akurasi yang lebih tinggi dibandingkan metode manual.

Tujuan dari penelitian ini adalah mengembangkan model prediksi *stunting* pada balita menggunakan algoritma C4.5, yang dapat memproses data terkait faktor risiko *stunting* dengan efisien dan menghasilkan model klasifikasi yang mudah diinterpretasi. Penelitian ini diharapkan dapat memberikan kontribusi berupa model prediktif yang akurat untuk mendukung pengambilan keputusan dalam intervensi *stunting*, serta membantu pihak kesehatan dalam mengidentifikasi balita yang memiliki risiko tinggi *stunting* sehingga tindakan pencegahan dan penanganan dapat dilakukan lebih tepat waktu.

METODOLOGI PENELITIAN

Metodologi penelitian ini dirancang untuk memberikan suatu kerangka atau pendekatan yang digunakan untuk mengumpulkan, menganalisis, dan menginterpretasikan data untuk menjawab pertanyaan penelitian (Ahmad et al., 2021). Dalam metodologi penelitian memuat tahapan yang sistematis dalam mencapai tujuan penelitian (Sah et al., 2022). Langkah-langkah dalam mencapai tujuan tersebut disusun dalam Gambar 1.



Gambar 1. Langkah-Langkah Penelitian

Merujuk pada ilustrasi yang ditampilkan dalam Gambar 1, berikut ini dijelaskan secara rinci langkah-langkah yang dilakukan dalam proses pengembangan sistem:

Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diambil dari website Kaggle dengan judul "*Stunting Toddler Detection*" (Pradana, 2024). Dataset ini menggunakan pendekatan z-*score* untuk menentukan status *stunting*, yang sesuai dengan pedoman dari *World Health Organization* (WHO). Pendekatan ini memberikan penilaian status gizi berdasarkan pengukuran antropometri, khususnya tinggi badan, yang distandardisasi sesuai usia dan jenis kelamin. Dataset ini terdiri dari sekitar 121.000 baris data, yang memuat informasi rinci tentang balita, seperti usia, jenis kelamin, tinggi badan, dan status gizi. Status gizi balita dikategorikan ke dalam empat kelompok: Normal, *Severely Stunted*, *Stunted*, dan Tinggi. Dengan data ini, penelitian dapat berfokus pada deteksi *stunting*, menggunakan algoritma klasifikasi untuk mengidentifikasi pola yang

memengaruhi status gizi balita dan mendukung intervensi kesehatan yang lebih efektif.

Eksplorasi Data dan *Preprocessing*

Tahap eksplorasi data dan *preprocessing* mencakup beberapa langkah penting untuk mempersiapkan data agar siap digunakan dalam pelatihan model. Pertama, eksplorasi data dilakukan dengan memeriksa karakteristik dasar dari dataset, seperti bentuk data, tipe variabel, jumlah entri, dan mengidentifikasi nilai kosong atau duplikat. Visualisasi distribusi variabel seperti umur dan tinggi badan dilakukan untuk memahami sebaran data dan pola yang mungkin terdapat di dalamnya.

Setelah itu, *preprocessing* data dimulai dengan *encoding variabel* kategori, seperti "Jenis Kelamin" dan "Status Gizi," agar dapat digunakan dalam model pembelajaran mesin. Variabel prediktor dan target kemudian dipisahkan, di mana variabel-variabel prediktor mencakup umur, jenis kelamin, dan tinggi badan, sementara targetnya adalah status gizi balita. Standarisasi fitur dilakukan untuk memastikan semua variabel memiliki skala yang sebanding, sehingga mempermudah algoritma dalam mengolah data.

Terakhir, *dataset* dibagi menjadi data latih dan data uji dengan rasio, 80:20. Hal ini dilakukan untuk memastikan model memiliki cukup data untuk belajar dan mengidentifikasi pola yang ada pada dataset, sementara sebagian data lainnya digunakan untuk menguji kinerja model (Salsabil et al., 2024). Pembagian data ini penting untuk mengukur seberapa baik model dapat melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

Pembuatan Model dengan Algoritma C4.5

Pada tahap ini dibanngun model prediksi berbasis *Decision Tree* dengan kriteria *entropy*, yang selaras dengan algoritma C4.5. Algoritma C4.5 adalah metode pohon keputusan yang dikembangkan oleh Ross Quinlan sebagai penyempurnaan dari algoritma ID3 (Sari et al., 2021). C4.5 bekerja dengan membangun pohon keputusan berdasarkan data latih yang ada, dengan tujuan mengklasifikasikan data baru ke dalam kelas yang sesuai (Prasetyo & Wahyono, 2024). Algoritma ini memilih atribut terbaik pada setiap *node* dalam pohon berdasarkan nilai *gain ratio*, yaitu ukuran yang digunakan untuk menilai seberapa baik atribut tertentu memisahkan data ke dalam kelas yang berbeda (Handayani et al., 2021).

Untuk menghitung *gain ratio*, C4.5 pertama-tama menghitung *information gain* dari setiap atribut, yang diukur menggunakan *entropy*. *Entropy* $E(S)$ dari dataset S adalah ukuran ketidakpastian atau keberagaman dalam kelas target dan dihitung dengan persamaan (1).

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

di mana p_i adalah proporsi data dalam kelas i.

Information gain dari suatu atribut kemudian dihitung sebagai pengurangan *entropy* setelah data dipartisi berdasarkan atribut tersebut. Namun, karena *information gai*n cenderung memberikan preferensi pada atribut dengan banyak nilai, C4.5 memperbaikinya dengan menggunakan *split information* dan menghitung *gain ratio* menggunakan persamaan (2).

$$Gain\ Ratio = \frac{Information\ Gain}{Split\ Information} \quad (2)$$

dengan *split information* dihitung dengan persamaan (3).

$$Split\ Information = -\sum_{j=1}^{v} \frac{|s_j|}{s} \log_2\left(\frac{|s_j|}{s}\right) \quad (3)$$

di mana s_j adalah *subset* dari s yang dihasilkan oleh suatu nilai dari atribut yang dipartisi, dan y adalah jumlah nilai unik dalam atribut tersebut. *Gain ratio* ini mengukur efisiensi pemisahan data dengan mempertimbangkan jumlah partisi, sehingga menghasilkan pemisahan yang lebih informatif.

Evaluasi Model

Evaluasi model bertujuan untuk menilai seberapa baik model memprediksi status gizi balita. Pada tahap ini, model diuji menggunakan data uji yang belum pernah dilihat selama pelatihan, sehingga dapat menggambarkan kinerja model pada data baru (Borman et al., 2021). Evaluasi dilakukan melalui confusion *matrix*, yaitu tabel yang dirancang untuk membandingkan hasil prediksi model dengan label asli dari data (Borman et al., 2022). Dari *confusion matrix* ini, diperoleh berbagai metrik evaluasi, seperti akurasi, *precision*, *recall*,

dan F1-*score*, yang digunakan untuk mengukur performa model dalam menghadapi berbagai jenis kesalahan klasifikasi.

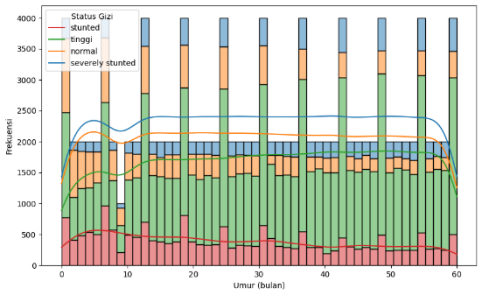
Selain itu, ROC-AUC *score* dihitung untuk mengukur seberapa baik model membedakan antara kelas-kelas yang ada dalam status gizi balita. Setiap metrik ini memberikan perspektif yang berbeda tentang performa model, yang berguna untuk menentukan keandalan prediksi dalam aplikasi praktis. Jika metrik menunjukkan performa yang baik, model ini dapat dianggap cukup andal untuk diaplikasikan lebih lanjut.

Interpretasi Hasil

Tahap Interpretasi Hasil mencakup analisis kinerja model, penyajian metrik evaluasi seperti akurasi, *precision*, *recall*, F1-*score*, serta visualisasi hasil dalam bentuk pohon keputusan, *confusion matrix*, dan kurva ROC. Selain itu, pada tahap ini akan dijabarkan analisis dan pemahaman terhadap kinerja model, yang mencakup identifikasi pola klasifikasi, efektivitas model dalam mendeteksi *stunting*, dan kesimpulan dari model yang dibangun.

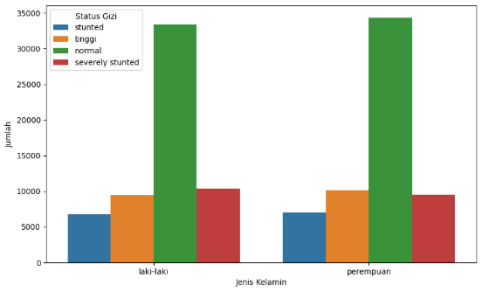
HASIL DAN PEMBAHASAN

Proses prediksi menggunakan algoritma C4.5 dimulai dengan membaca dataset berjudul "*Stunting Toddler Detection*" yang diambil dari website Kaggle (Pradana, 2024). Dataset ini, dalam file bernama "data_balita.csv," mencakup data penting terkait balita, seperti usia, jenis kelamin, tinggi badan, dan status gizi. Setelah membaca dataset, langkah selanjutnya adalah memeriksa data untuk mendeteksi nilai kosong dan menghapus duplikasi data agar kualitas data terjaga. Dalam mempersiapkan data untuk algoritma pembelajaran mesin, variabel kategori seperti "Jenis Kelamin" dan "Status Gizi" dikonversi menjadi format numerik menggunakan teknik *encoding*. Proses *preprocessing* dilanjutkan dengan standarisasi fitur menggunakan "*StandardScaler*" untuk memastikan bahwa setiap variabel berada dalam skala yang seimbang. Visualisasi data juga dilakukan untuk mendapatkan gambaran lebih mendalam tentang distribusi fitur-fitur utama, seperti distribusi umur berdasarkan status gizi yang divisualisasikan pada Gambar 2.



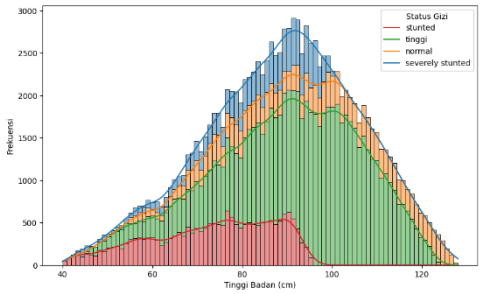
Gambar 2. Pola Sebaran Data Umur Terhadap Status Gizi

Gambar 2 menunjukkan distribusi status gizi pada balita berdasarkan umur dalam rentang 0 hingga 60 bulan. Status gizi dikategorikan menjadi "*Stunted*," "Tinggi," "Normal," dan "*Severely Stunted*," yang masing-masing ditampilkan dengan warna berbeda. Grafik ini ini memberikan Gambaran pola-pola tertentu yang dihasilkan dari data umur terhadap status gizi anak yang dapat menjadi pertimbangan dalam melakukan prediksi. Pola distribusi data berikutnya yang perlu dianalisis adalah hubungan antara jenis kelamin dan status gizi. Grafik distribusi dataset yang menunjukkan kaitan antara jenis kelamin dan status gizi dapat dilihat pada Gambar 3.



Gambar 3. Pola Sebaran Data Jenis Kelamin Terhadap Status Gizi

Gambar 3 menunjukkan distribusi status gizi berdasarkan jenis kelamin anak-anak dalam dataset. Grafik tersebut memberikan gambaran tentang perbedaan atau kesamaan distribusi status gizi antara jenis kelamin laki-laki dan perempuan. Selain itu, penting juga untuk memahami distribusi data tinggi badan berdasarkan status gizi guna mengidentifikasi pola-pola yang ada dalam data tersebut. Visualisasi distribusi ini ditampilkan pada Gambar 4.

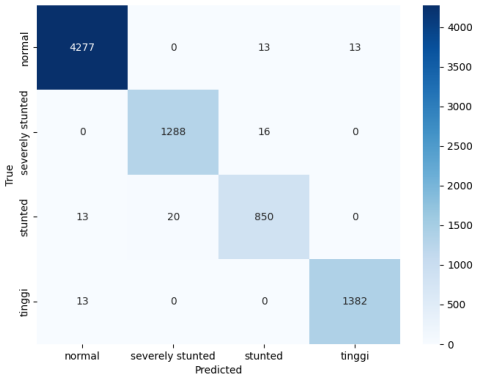


Gambar 4. Pola Sebaran Data Tinggi Badan Terhadap Status Gizi

Gambar 4 menunjukkan distribusi tinggi badan pada balita berdasarkan status gizi. Grafik ini memperlihatkan bahwa balita dengan status gizi normal memiliki tinggi badan yang tersebar luas pada rentang sekitar 60 hingga 100 cm, dengan frekuensi yang tinggi di antara 80 hingga 100 cm. Kategori *stunted* dan *severely stunted* lebih banyak berada pada tinggi badan yang lebih rendah, sedangkan kategori tinggi mendominasi pada tinggi badan di atas 100 cm. Pola ini membantu memahami hubungan antara status gizi dan tinggi badan pada balita.

Proses selanjutnya yaitu membangun model prediksi menggunakan C4.5 yang digunakan untuk pelatihan dan pengujian. Namun sebelumnya *dataset* yang digunakan dibagi menjadi data latih dan data uji dengan rasio 80:20, memastikan bahwa model dapat diuji pada data yang belum pernah dilihat selama pelatihan untuk mengevaluasi performa pada data baru. Proses pelatihan model menggunakan algoritma C4.5 dimulai dengan pemilihan fitur yang paling efektif untuk memisahkan data berdasarkan kelas target. C4.5 melakukan pemisahan ini dengan menggunakan konsep information gain yang diperoleh melalui entropy. *Entropy* mengukur tingkat ketidakpastian atau impurity dalam data, dan information gain menghitung seberapa besar penurunan ketidakpastian setelah data dipisahkan berdasarkan suatu fitur. Fitur dengan information gain tertinggi dipilih sebagai *node* (atau akar) pertama dari pohon keputusan.

Setelah model pelatihan selesai dibangun, langkah selanjutnya adalah menguji model untuk menilai kinerjanya dalam melakukan prediksi. Pengujian dilakukan menggunakan *confusion matrix*, yang menunjukkan jumlah prediksi benar dan salah pada setiap kelas. *Confusion matrix* ini divisualisasikan dalam bentuk heatmap untuk memudahkan interpretasi. Hasil *confusion matrix* dari model prediksi yang dikembangkan ditampilkan pada Gambar 5.



Gambar 5. Hasil *Confusion Matrix*

Berdasarkan confusion *matrix* pada Gambar 5, kemudian didapatkan parameter uji seperti presisi dan recall yang ditunjukkan pada Tabel 1.

Tabel 1. Nilah *Precision* dan *Recall*

Kelas	*Precision*	*Recall*
0	0.9940	0.9940
1	0.9847	0.9877
2	0.9670	0.9626
3	0.9907	0.9907

0: Normal; 1: *Severely Stunted*; 2: *Stunted*; 3: Tinggi

Selain parameter diatas diperoeleh juga nilai F1-*Score*, akurasi dan ROC-AUC. Nilai yang dihasilkan tersebut disusun pada Tabel 2.

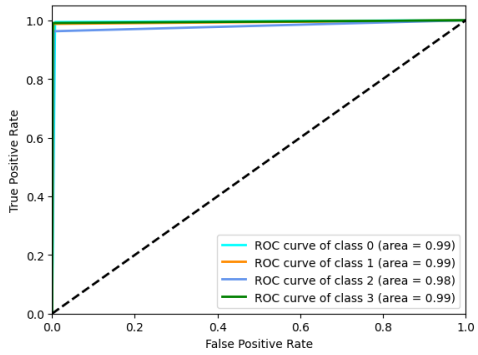
Tabel 2. Nilai F1-*Sore*, *Accuracy* dan ROC-AUC

Kelas	F1-*Score*	*Accuracy*	ROC-AUC
0	0.9940	0.9888	0.9898
1	0.9862		
2	0.9648		
3	0.9907		

0: Normal; 1: *Severely Stunted*; 2: *Stunted*; 3: Tinggi

Tabel 2 menampilkan skor ROC-AUC, yang digunakan untuk menilai kemampuan model dalam membedakan setiap kelas. Kurva ROC untuk masing-

masing kelas divisualisasikan pada Gambar 6, memperlihatkan performa klasifikasi model terhadap setiap kategori status gizi.



Gambar 6. Kurva ROC Untuk Setiap Kelas

Evaluasi model prediksi *stunting* pada balita menggunakan algoritma C4.5 menunjukkan hasil yang sangat baik berdasarkan nilai *precision*, *recall*, F1-*score*, *accuracy*, dan ROC-AUC yang ditampilkan dalam Tabel 1 dan Tabel 2. Pada Tabel 1, nilai *precision* dan *recall* untuk setiap kelas (Normal, *Severely Stunted*, *Stunted*, dan Tinggi) cukup tinggi, dengan kelas "Normal" memiliki *precision* dan *recall* tertinggi sebesar 0.9940, menunjukkan kemampuan model dalam memprediksi kelas ini dengan sangat akurat. Kelas "*Severely Stunted*" dan "Tinggi" juga memiliki *precision* dan *recall* di atas 0.9800, menunjukkan bahwa model dapat membedakan kelas-kelas tersebut dengan baik. Kelas "*Stunted*" memiliki *precision* dan *recall* yang sedikit lebih rendah, tetapi tetap berada di kisaran yang baik.

Pada Tabel 2, F1-*score* untuk setiap kelas hampir mendekati nilai *precision* dan *recall*, yang mengindikasikan keseimbangan antara keduanya. Secara keseluruhan, model ini mencapai tingkat akurasi sebesar 98.88%, yang berarti model mampu mengklasifikasikan status gizi balita dengan tepat pada sebagian besar data uji. Nilai ROC-AUC yang mencapai 0.9898 juga memperlihatkan bahwa model memiliki performa yang baik dalam membedakan setiap kelas, dengan kurva ROC yang hampir mendekati garis sempurna.

Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa algoritma C4.5 efektif dalam mengklasifikasikan status gizi balita berdasarkan data yang tersedia. Algoritma C4.5 memainkan peran penting dalam klasifikasi status gizi balita pada penelitian ini. Dengan kemampuan membangun pohon keputusan berdasarkan informasi gain, algoritma ini dapat memilih fitur-fitur yang paling relevan dan membagi data secara efektif untuk membedakan status gizi anak balita. C4.5 bekerja dengan memilih fitur yang memberikan pemisahan terbaik, seperti tinggi badan, umur, dan jenis kelamin, sehingga menghasilkan klasifikasi yang akurat.

KESIMPULAN

Penelitian ini telah berhasil menerapkan algoritma C4.5 untuk mengklasifikasikan status gizi balita, khususnya dalam mendeteksi risiko *stunting*. Dengan menggunakan pohon keputusan yang memprioritaskan pemisahan berdasarkan *information gain*, algoritma C4.5 memanfaatkan fitur-fitur seperti tinggi badan, umur, dan jenis kelamin untuk mengelompokkan data ke dalam kategori status gizi yang sesuai. Hasil evaluasi menunjukkan bahwa model ini mencapai tingkat akurasi yang tinggi, yaitu 98.88%, dengan nilai *precision*, *recall*, dan F1-*score* yang kuat di setiap kelas. Selain itu, nilai ROC-AUC sebesar 98.98% menunjukkan kemampuan model dalam membedakan kelas-kelas status gizi yang berbeda dengan baik. Tingginya akurasi ini menekankan keandalan algoritma C4.5 dalam mengenali pola yang terkait dengan kondisi gizi anak, menjadikannya alat yang bermanfaat bagi tenaga kesehatan dan pembuat kebijakan untuk mendeteksi *stunting* secara efisien dan akurat. Untuk penelitian selanjutnya, ada beberapa saran yang dapat dipertimbangkan. Beberapa kelas menunjukkan nilai *precision* dan *recall* yang sedikit lebih rendah dibandingkan kelas lainnya; menambah jumlah data untuk kelas yang lebih sulit diklasifikasi dapat membantu meningkatkan performa. Selain itu, penerapan metode *ensemble* seperti *Random Forest* atau *Boosting* dapat membantu mengurangi *overfitting* dan meningkatkan akurasi model secara keseluruhan.

DAFTAR PUSTAKA

Ahmad, I., Rahmanto, Y., Pratama, D., & Borman, R. I. (2021). Development of Augmented Reality Application for

Introducing Tangible Cultural Heritages at The Lampung Museum Using The Multimedia Development Life Cycle. *ILKOM Jurnal Ilmiah*, *13*(2), 187–194.

Arfandi, A., Windarto, A. P., & Saragih, lham S. (2021). Penerapan Data Mining Klasifikasi Pada Calon Pelanggan Baru Indihome dengan C.45. *Journal of Informatics, Electrical and Electronics Engineering*, *1*(1), 31–38.

Borman, R. I., Napianto, R., Nugroho, N., Pasha, D., Rahmanto, Y., & Yudoutomo, Y. E. P. (2021). Implementation of PCA and KNN Algorithms in the Classification of Indonesian Medicinal Plants. *International Conference on Computer Science, Information Technology and Electrical Engineering (ICOMITEE)*, 46–50.

Borman, R. I., Rossi, F., Alamsyah, D., Nuraini, R., & Jusman, Y. (2022). Classification of Medicinal Wild Plants Using Radial Basis Function Neural Network with Least Mean Square. *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*.

Borman, R. I., & Wati, M. (2020). Penerapan Data Maining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes. *Jurnal Ilmiah Fakultas Ilmu Komputer*, *9*(1), 25–34.

Fathurrohman, R., S, M. F., Cahyono, S. M., Abdillah, D. F., & Ramadhan, V. (2024). Data Mining Pada Klasifikasi Jamur Menggunakan Algoritma C.45 Berdasarkan Karakteristik Morfologi Mushroom. *Simpatik: Jurnal Sistem Informasi Dan Informatika*, *4*(1), 29–38.

Handayani, N., Wahyono, H., Trianto, J., & Permana, D. S. (2021). Prediksi Tingkat Risiko Kredit dengan Data Mining Menggunakan Algoritma Decision Tree C.45. *JURIKOM (Jurnal Riset Komputer)*, *8*(6), 198–204. https://doi.org/10.30865/jurikom.v8i6.3643

Matondang, M. R., Lubis, M. R., & Tambunan, H. S. (2021). Analisis Data mining dengan Metode C.45 pada Klasifikasi Kenaikan Rata-Rata Volume Perikanan Tangkap. *BRAHMANA: Jurnal Penerapan Kecerdasan Buatan*, *2*(2), 74–81.

Maulana, Y., Winanjaya, R., & Rizki, F. (2022). Penerapan Data Mining dengan Algoritma C4.5 Dalam Memprediksi Penjualan Tempe. *Bulletin of Computer Science Research*, *2*(2), 53–58. https://doi.org/10.47065/bulletincsr.v2i2.163

Nurmalasari, E., Aryanti, U., & Rachman, T. T. (2024). Penerapan Data Mining Algoritma Apriori untuk Menemukan Pola Hubungan Status Gizi Balita. *INTERNAL (Information System Journal)*, *6*(2), 156–166.

Octavia, Y. T., Siahaan, J. M., & Barus, E. (2023). Upaya Percepatan Penurunan Stunting (Gizi Buruk dan Pola Asuh) Pada Balita yang Beresiko Stunting. *Journal Abdimas Mutiara*, *5*(1), 131–140.

Pradana, R. P. (2024). *Stunting Toddler Detection*. Kaggle. https://www.kaggle.com/datasets/rendiputra/stunting-balita-detection-121k-rows/

Prasetyo, A., & Wahyono, A. (2024). Implementasi Algoritma C.45 dalam Klasifikasi Kondisi Ekonomi Warga Kabupaten Boyolali. *Generation Journal*, *8*(2), 121–131.

Rahman, H., Rahmah, M., & Saribulan, N. (2023). Upaya Penanganan Stunting di Indonesia: Analisis Bibliometrik dan Analisis Konten. *Jurnal Ilmu Pemerintahan Suara Khatulistiwa (JIPSK)*, *VIII*(01), 44–59.

Sah, A., Jusmawati, J., Nurhayati, S., Tonggiroh, M., & Bonay, S. (2022). Sistem Informasi Manajemen Pada Puskesmas Kota Jayapura Berbasis Web. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *4*(3), 212–

220.

Salsabil, M., Lutvi, N., & Eviyanti, A. (2024). Implementasi Data Mining dalam Melakukan Prediksi Penyakit Diabetes Menggunakan Metode Random Forest dan Xgboost. *Jurnal Ilmiah KOMPUTASI*, *23*(1), 51–58.

Sari, M., Kom, S., & Kom, M. (2021). Analisis Tingkat Kepuasan Konsumen Rumah Nutrisi Di Banjarmasin Sebagai Implementasi Penerapan Data Mining Algoritma C.45. *Technologia*, *12*(1), 49–52.