Implementasi Algoritma C5.0 Pada Klasifikasi Status Gizi Balita Di Kecamatan Ponorogo

Natasya Shindo Bekti Nugrahani[1], Aditya Prapanca[2]

[1.2]Program Studi S1 Teknik Informatika, Universitas Negeri Surabaya
[1]natasyashindo.20077@mhs.unesa.ac.id
[2]adityaprapanca@unesa.ac.id

***Abstrak*— Permasalahan gizi pada balita merupakan isu kesehatan masyarakat yang masih menjadi perhatian di Indonesia. Penelitian ini bertujuan untuk mengklasifikasikan status gizi balita di Kecamatan Ponorogo dengan menggunakan algoritma *Decision Tree* C5.0. Data yang digunakan merupakan data antropometri balita yang diperoleh dari Puskesmas Ponorogo Selatan, yang meliputi berat badan, tinggi badan, usia dan jenis kelamin balita. Proses penelitian melibatkan beberapa tahapan, dimulai dengan data *preprocessing* yang terdiri dari data *cleaning*, transformasi data, *label encoding*, seleksi fitur, *resampling* dan data *splitting*, selanjutnya data processing yang digunakan untuk perhitungan entropy dan information gain, hingga evaluasi model menggunakan *Confusion matrix* dan kurva ROC. Hasil pengujian menunjukkan bahwa akurasi tertinggi yang diperoleh mencapai 97,92% pada skema pembagian 60% data training dan 40% data testing. Fitur yang paling berpengaruh dalam klasifikasi adalah berat badan, tinggi badan dan usia balita. Berdasarkan hasil tersebut, algoritma C5.0 terbukti mampu memberikan performa tinggi dalam klasifikasi status gizi dan berpotensi digunakan sebagai alat bantu dalam pemantauan tumbuh kembang anak di layanan kesehatan dasar.**

Kata Kunci – Status Gizi Balita, Algoritma C5.0, Klasifikasi, Antropometri dan *Data Mining*.

I. PENDAHULUAN

Kesehatan dan gizi anak balita sangat penting untuk menunjang pertumbuhan dan perkembangan optimal di masa depan. Anak di bawah lima tahun rentan terhadap masalah kesehatan dan gizi yang dapat berdampak jangka panjang [1]. Di Indonesia, permasalahan gizi balita masih menjadi tantangan serius. Data Survei Status Gizi Indonesia (SSGI) tahun 2021 menunjukkan 17% balita mengalami gizi kurang, 3,8% gizi lebih, dan 7,1% gizi buruk [2]. Tantangan yang dihadapi dalam upaya penurunan masalah gizi balita antara lain yaitu: pola asuh yang kurang efektif, malnutrisi jangka panjang, kurangnya pengetahuan tentang pola makan gizi seimbang, sanitasi yang buruk dan kurangnya perawatan pasca melahirkan hingga menimbulkan infeksi pada anak. Untuk menghadapi tantangan tersebut perlu solusi yang terintegrasi, yaitu termasuk pendekatan lintas sektoral oleh kader posyandu, pemberdayaan perempuan serta membuat program atau kebijakan yang mendukung pencegahan masalah gizi pada balita.

Faktor penyebab utama meliputi pola asuh yang kurang efektif, malnutrisi, rendahnya pengetahuan tentang gizi seimbang, sanitasi buruk, dan kurangnya perawatan pasca melahirkan. Oleh karena itu, diperlukan pendekatan terintegrasi yang melibatkan lintas sektor, pemberdayaan perempuan, serta program pencegahan gizi buruk pada balita. Pemantauan status gizi balita biasanya dilakukan oleh kader posyandu menggunakan metode antropometri, yaitu pengukuran berat badan dan tinggi badan [2]. Indeks yang umum digunakan antara lain berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U), dan berat badan menurut tinggi badan (BB/TB) [3].

Beberapa penelitian telah menggunakan algoritma *Decision Tree*, khususnya C4.5, untuk membantu klasifikasi status gizi balita dengan hasil akurasi yang baik [1]. Namun, algoritma C5.0 sebagai penyempurnaan C4.5 menawarkan kecepatan, akurasi, dan kemudahan interpretasi yang lebih baik, meskipun memiliki keterbatasan pada data dengan variansi kecil.

Algoritma C4.5 merupakan salah satu algoritma yang cukup dikenal dalam bidang *Data Mining* karena kemampuannya tidak hanya dalam melakukan klasifikasi, tetapi juga dalam menampilkan hubungan antar atribut. Salah satu keunggulan utama dari algoritma ini adalah fleksibilitasnya dalam mengolah data numerik maupun diskrit [4]. Selain itu, algoritma C4.5 juga mampu menangani atribut yang kosong serta missing value melalui teknik *continuous attribute splitting*, dan dapat mengurangi risiko *overfitting* dengan menggunakan metode *pruning* [5]. Meskipun demikian, algoritma ini memiliki beberapa kelemahan, antara lain kecenderungannya menghasilkan pohon keputusan yang lebih panjang dibanding algoritma lain, kurang efektif dalam menangani data berskala ordinal, serta tidak optimal dalam mengolah data berbentuk teks.

Sementara itu, algoritma C5.0 merupakan pengembangan lebih lanjut dari algoritma ID3 dan C4.5 yang dirancang oleh Ross Quinlan pada tahun 1987. Algoritma ini dinilai lebih unggul karena mampu menghasilkan akurasi yang lebih tinggi serta penggunaan memori yang lebih efisien [6]. C5.0 juga dikenal memiliki performa yang sangat cepat dalam membangun model, dengan tingkat akurasi yang lebih baik dibandingkan pendahulunya. Di samping menghasilkan model berupa pohon keputusan, algoritma ini juga membentuk model berbasis aturan, yang memudahkan dalam memahami hasil outputnya. Seperti halnya C4.5, algoritma C5.0 dapat menangani atribut numerik maupun kategorik [5]. Namun demikian, kelemahan algoritma ini masih sejalan dengan karakteristik umum pohon keputusan, yaitu prediksi yang dihasilkan menjadi kurang stabil ketika varian data sangat kecil, serta adanya ketergantungan yang tinggi terhadap informasi input, sehingga perubahan kecil pada data dapat menimbulkan perubahan besar dalam struktur pohon keputusannya.

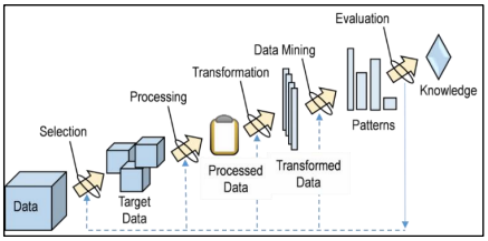
Fokus penelitian meliputi 1) Cara mengimplementasikan algoritma C5.0 dalam proses klasifikasi status gizi balita di Kecamatan Ponorogo. 2) Hasil akurasi dari algoritma C5.0 dalam klasifikasi status gizi balita di Kecamatan Ponorogo. Dan 3) Faktor-faktor yang berpengaruh dalam menentukan status gizi balita berdasarkan hasil klasifikasi. Maka penelitian untuk menentukan klasifikasi status gizi balita ini akan di lakukan menggunakan algoritma C5.0 dengan data yang didapat dari posyandu di Kecamatan Ponorogo. Algoritma C5.0 dipilih karena memiliki kinerja yang lebih baik dalam menghasilkan informasi melalui perhitungan gain dan entropy. Diharapkan dengan menerapkan algoritma tersebut dapat mengklasifikasikan status gizi balita dengan akurat.

A. *Data Mining*

Data Mining merupakan proses penggalian informasi dari sejumlah besar data untuk menemukan pola tersembunyi yang bermanfaat. Proses ini mencakup teknik statistik, *machine learning*, dan *database systems*, serta sering diaplikasikan untuk menyelesaikan permasalahan kompleks dalam berbagai bidang [6].

Tujuan utama dari *Data Mining* adalah memperoleh pengetahuan baru dari data yang ada, melalui pembangunan model yang dapat digunakan untuk memprediksi atau mendeskripsikan fenomena tertentu. Secara umum, terdapat dua pendekatan dalam *Data Mining*:

1) Metode Prediktif: digunakan untuk memperkirakan nilai atau kategori berdasarkan data historis, seperti klasifikasi, regresi, dan deteksi anomali.
2) Metode Deskriptif: bertujuan untuk mengidentifikasi pola yang tersembunyi dalam data tanpa memprediksi nilai tertentu, seperti asosiasi dan klastering.



Gbr. 1 Data Mining

Data Mining merupakan bagian dari proses yang lebih luas, yaitu *Knowledge Discovery in Database (KDD)* [7]. KDD adalah serangkaian tahap sistematis yang bertujuan mengubah data mentah menjadi informasi yang bernilai. Tahapan dalam KDD mencakup:

1) *Data Selection* adalah Proses yang dilakukan untuk mentransformasikan data mentah ke format yang sesuai untuk analisis. Proses ini terdiri dari seleksi fitur, reduksi *dimensionalitas, normalisasi* dan *subsetting data* [10].
2) *Preprocessing (Cleaning)*, sebelum proses *Data Mining* dilakukan, perlu melakukan *proses cleaning* data yakni pembersihan data dari kesalahan, duplikasi, dan ketidakkonsistenan.
3) *Transformation* adalah Proses untuk merubah skala data kedalam bentuk lain sehingga data memiliki distribusi yang diperlukan. Terdapat beberapa teknik untuk melakukan transformasi data, yaitu *smoothing, generalization, normalization, attribute construction*.
4) *Data Mining* merupakan penerapan algoritma untuk menemukan pola atau hubungan dalam data. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan.
5) *Interpretation / Evaluation* adalah penafsiran hasil dan validasi pola untuk memastikan keterkaitan dengan tujuan analisis.

Tahapan-tahapan ini menjadi dasar dalam pengembangan model *Data Mining* yang tepat sesuai dengan kebutuhan analisis dan karakteristik dataset yang digunakan [11].

B. Klasifikasi

Klasifikasi adalah proses membangun model atau fungsi yang digunakan untuk mengelompokkan data ke dalam kelas-kelas tertentu berdasarkan data latih (*training dataset*) yang telah diberi label. Model ini kemudian digunakan untuk mengklasifikasikan data uji (*testing dataset*) dan mengukur tingkat akurasinya. Beberapa algoritma yang umum digunakan dalam klasifikasi antara lain: *Decision Tree, Naive Bayes, Backpropagation, K-Nearest Neighbor*, dan pendekatan sistem *fuzzy* [1].

C. *Decision Tree*

Decision Tree merupakan salah satu algoritma dalam metode klasifikasi *Data Mining* yang berbentuk struktur pohon. Setiap node menunjukkan atribut, cabang mewakili hasil pengujian, dan daun menunjukkan kelas. Algoritma yang termasuk dalam *Decision Tree* antara lain ID3, CART, C4.5, dan C5.0, yang banyak digunakan dalam proses klasifikasi dan ekstraksi data. *Decision Tree* bekerja dengan membentuk diagram pohon keputusan, di mana setiap atribut diuji pada node dan hasilnya menentukan arah cabang hingga mencapai keputusan akhir di node daun. Algoritma ini memetakan proses klasifikasi secara visual sehingga memudahkan interpretasi [8].

D. Algoritma C5.0

Algoritma C5.0 merupakan pengembangan dari algoritma ID3 dan C4.5 yang dikenalkan oleh J. Ross Quinlan. Algoritma ini menggunakan pendekatan *Information Gain* untuk menentukan atribut terbaik dalam pemisahan data. Atribut yang memiliki nilai *Information Gain* tertinggi dipilih sebagai simpul utama dalam pohon keputusan[6]. Rumus perhitungan Information Gain:

$$Gain(S,A) = Entropy(S) - \sum c \in values(A)\frac{|si|}{|s|}Entropy(\text{Si}) \quad (1)$$

Keterangan:

- *A*: Atribut
- *S*: Jumlah total sampel
- |*Si*|*:* Jumlah sampel pada partisi ke -1
- |*S*| *:* Jumlah sampel pada *S*

Dengan perhitungan nilai entropi dari koleksi label pada atribut A dan S didefinisikan menggunakan rumus sebagai berikut:

$$Entropy\ (S) = \sum_i^c -P_i log_2 P_i \quad (2)$$

Keterangan :

- S : Sampel
- *Pi*: Proporsi dari S_i terhadap S
- c : Jumlah partisi S

E. *Confusion matrix*

Confusion matrix [9] adalah alat ukur yang digunakan untuk menghitung kinerja atau tingkat kebenaran proses klasifikasi. Dengan menggunakan *Confusion matrix* dapat menganalisa seberapa baik *classifier* dapat mengenali *record* dari kelas yang berbeda.
Matriks ini terdiri:

Table 1
Confusion matrix

		Prediksi	
		Positif	Negatif
Aktual	Positif	TP	FN
	Negatif	FP	TN

Pengukuran Kinerja Model

1) *Accuracy* untuk mengukur tingkat klasifikasi yang tepat, Persamaan akurasi menggunakan rumus:

$$Accuracy = \frac{(TP+TN)}{(TP+FN+FP+TN)}\ x\ 100\% \quad (3)$$

2) *Precision* mengukur proporsi prediksi positif yang benar, Persamaan *precision* menggunakan rumus:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

3) *Recall* mengukur proporsi data positif yang berhasil diprediksi dengan benar. [12]

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

F. Kurva ROC

Receiver Operating Characteristic (ROC) adalah grafik yang menunjukkan performa klasifikasi melalui dua metrik. Kurva ROC memiliki tujuan untuk melakukan pengamatan model pada klasifikasi yang dihasilkan menggunakan 20 perhitungan confusion matrix. Rumus dari kurva ROC adalah sebagai berikut:

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP+TN} \quad (6)$$

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{TP+FN} \quad (7)$$

Dan di dalam nilai AUC, klasifikasi data mining dapat dibagi menjadi beberapa kelompok yaitu sebagai berikut :

1) 0.90 – 1.00 = Klasifikasi Sangat Baik
2) 0.80 – 0.90 = Klasifikasi Baik
3) 0.70 – 0.80 = Klasifikasi Cukup
4) 0.60 – 0.70 = Klasifikasi Buruk
5) 0.50 – 0.60 = Klasifikasi Gagal

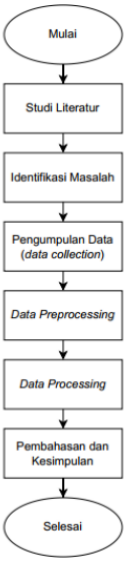
G. Populasi dan Sampel

Menurut Sugiyono [13] populasi adalah wilayah generalisasi yang terdiri dari objek atau subjek yang memiliki karakteristik tertentu yang ditetapkan oleh peneliti untuk diteliti dan ditarik kesimpulannya, Populasi dalam penelitian ini adalah Seluruh data balita usia 0–5 tahun yang tercatat di wilayah kerja Puskesmas Ponorogo Selatan dan memiliki data antropometri lengkap (berat badan, tinggi badan, usia, dan jenis kelamin), yang diperoleh dari catatan Buku Kesehatan Ibu dan Anak (KIA) melalui posyandu, Total populasi yang tersedia dalam bentuk data adalah 1.021 balita.

Sedangkan sampel adalah bagian dari populasi yang dianggap dapat mewakili keseluruhan populasi. Sebanyak 120 data balita usia 1–5 tahun yang dipilih secara acak dari total data 866 balita (hasil filter usia dari populasi awal). Sampel ini dibagi secara seimbang ke dalam empat kategori status gizi (normal, kurang, risiko lebih, sangat kurang) menggunakan teknik undersampling agar distribusi kelas merata (masing-masing 30 data per kelas).

II. Metode Penelitian

Dalam penelitian ini alur Metode Penelitian digambarkan pada Gbr. 1;



Gbr. 1 Flowchart Penelitian

A. *Analisis Kebutuhan*

Penelitian ini dilakukan untuk menentukan spesifikasi perangkat keras dan lunak yang diperlukan dalam proses klasifikasi status gizi balita menggunakan algoritma C5.0. Seluruh proses komputasi dan implementasi algoritma dijalankan melalui platform *Google Colab*, yang sekaligus berfungsi sebagai lingkungan pengujian dataset. Oleh karena itu, dibutuhkan perangkat pendukung yang memadai guna memastikan kelancaran proses penelitian.
Kebutuhan *Hardware* yang digunakan dalam penelitian ini adalah laptop dengan spesifikasi

1) Prosesor : AMD Athlon Gold 3150U with Radeon Graphics 2.40 GHz
2) RAM : 4.00 GB
3) Penyimpanan : SSD 128 GB
4) Sistem Operasi : Windows 11 Home Single Language

Spesifikasi ini dinilai cukup untuk menjalankan aplikasi berbasis *cloud* seperti *Google Colab*, yang membutuhkan performa stabil pada sisi client.

Kebutuhan *Software* yang digunakan meliputi peramban web *Google Chrome* versi 122.0.6261.112 (*build resmi*, 32 bit), yang digunakan untuk mengakses *Google Colab*. Di dalam *Google Colab*, implementasi algoritma C5.0 dilakukan dengan menggunakan bahasa pemrograman *Python*. Pemilihan *Google Colab* sebagai platform pengembangan didasarkan pada kemudahan akses, integrasi dengan berbagai pustaka *Python* untuk pemrosesan data, serta kemampuan eksekusi berbasis *cloud* yang mengurangi beban komputasi lokal.

B. *Implementasi Sistem*

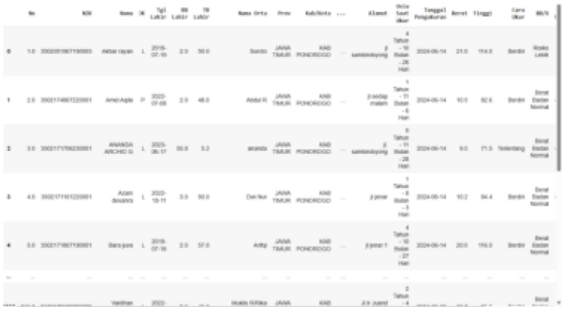
Implementasi algoritma C5.0 dalam penelitian ini dilakukan untuk melakukan klasifikasi terhadap status gizi balita berdasarkan data antropometri. Data yang digunakan mencakup berat badan, tinggi badan, usia, dan jenis kelamin balita, yang disusun dalam bentuk database. Proses implementasi dimulai dengan menyiapkan data tersebut dalam format yang sesuai, kemudian mengimpor dataset ke dalam lingkungan *Google Colab* yang menggunakan bahasa pemrograman *Python* . Selanjutnya, dilakukan pembuatan fungsi pemrograman untuk mengimplementasikan algoritma C5.0. Algoritma ini kemudian dijalankan untuk melakukan proses klasifikasi terhadap status gizi balita berdasarkan data yang telah dimasukkan. Hasil dari proses ini mencakup label klasifikasi status gizi untuk setiap entri data, serta nilai akurasi yang diperoleh dari model klasifikasi yang telah dibangun. Nilai akurasi tersebut digunakan sebagai indikator kinerja dari algoritma C5.0 dalam melakukan prediksi terhadap data yang diuji.

C. *Pengujian Sistem*

Tahap pengujian sistem bertujuan untuk mengevaluasi apakah metode klasifikasi yang digunakan telah memenuhi kriteria fungsional dan non-fungsional secara optimal. Pengujian dilakukan dengan menerapkan algoritma C5.0 terhadap dataset uji yang telah disiapkan, serta memeriksa keakuratan hasil klasifikasi yang dihasilkan. Selain itu, proses pengujian juga berfungsi untuk mengidentifikasi potensi kelemahan dalam sistem, baik dari segi efisiensi algoritma maupun integritas data. Dengan demikian, pengujian ini dapat meminimalisir kesalahan atau gangguan dalam pengoperasian sistem secara keseluruhan dan memastikan bahwa implementasi algoritma telah berjalan sesuai dengan tujuan penelitian.

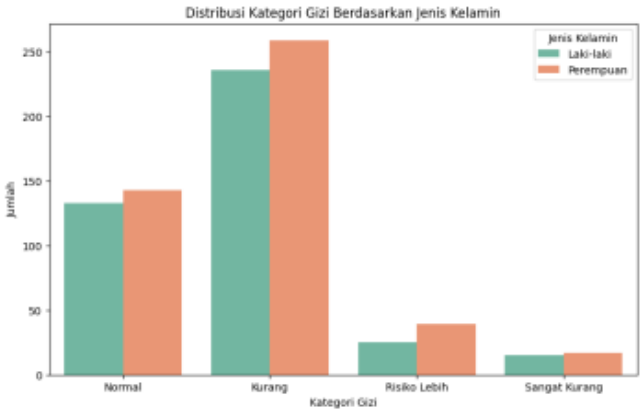
D. *Pengumpulan Data*

Pengumpulan data dalam penelitian ini dilakukan melalui *documentary research*, yaitu teknik pengambilan data sekunder yang telah tersedia sebelumnya. Data diperoleh dari catatan pengukuran berat badan dan tinggi badan balita yang tercantum dalam Buku Kesehatan Ibu dan Anak (KIA), yang dicatat secara rutin oleh petugas posyandu di bawah naungan Puskesmas Ponorogo Selatan. Data tersebut diambil setiap bulan dan mencakup lima kelurahan dalam wilayah kerja Puskesmas Ponorogo Selatan, yaitu Kelurahan Kepatihan, Purbosuman, Surodikraman, Tonatan, dan Pakunden.



Gbr. 2 Dataset

Secara keseluruhan, jumlah data balita yang tersedia mencapai 1.021 entri. Namun, dalam penelitian ini hanya digunakan sekitar 120 data balita yang telah disaring berdasarkan kriteria umur, yaitu dari usia 1 tahun hingga 5 tahun. Data yang digunakan mencakup informasi demografis dan antropometrik, seperti berat badan, tinggi badan, usia, jenis kelamin, serta status gizi balita yang ditentukan berdasarkan standar Z-score berat badan menurut umur (BB/U).



Gbr. 3 Grafik Jumlah Status Gizi setiap Kelas

III. *HASIL DAN PEMBAHASAN*

Pada bab ini disajikan hasil dari proses klasifikasi status gizi balita menggunakan algoritma C5.0 serta pembahasan terhadap hasil tersebut. Penjabaran dimulai dari tahapan pra-pemodelan (*preprocessing*), pemrosesan informasi untuk seleksi fitur penting, hingga evaluasi model. Setiap langkah dijelaskan berdasarkan temuan empiris yang telah diperoleh selama penelitian.

A. *Data Preprocessing*

Data mentah yang diperoleh awalnya terdiri atas 20 atribut, termasuk beberapa informasi yang tidak relevan untuk klasifikasi seperti nama orang tua dan tanggal lahir. Oleh karena itu, dilakukan proses pre-processing untuk menyaring hanya atribut yang dibutuhkan dalam klasifikasi status gizi. Selanjutnya, data yang telah difilter akan digunakan dalam tahap pemodelan menggunakan algoritma C5.0. Tahapan yang dilakukan dalam data *pre-processing adalah:*

1) Data Cleaning

Data cleaning dilakukan untuk menghapus atribut yang tidak relevan dan berpotensi mengganggu proses klasifikasi. Meliputi: nomor, NIK, nama, tanggal lahir, nama orang tua, alamat, provinsi, kabupaten/kota, kecamatan, desa/kelurahan, RT/RW, nama posyandu, puskesmas, cara ukur, dan tanggal pengukuran. Data yang dipertahankan sebagai berikut:

	JK	BB Lahir	TB Lahir	Usia Saat Ukur	Berat	Tinggi	Naik Berat Badan
0	L	2.0	50.0	4 Tahun - 10 Bulan - 26 Hari	21.0	114.0	T
1	P	2.0	48.0	1 Tahun - 11 Bulan - 6 Hari	10.5	82.6	T
2	L	55.8	5.2	0 Tahun - 11 Bulan - 28 Hari	9.5	71.5	N
3	L	3.0	50.0	1 Tahun - 8 Bulan - 3 Hari	10.2	84.4	T
4	L	2.0	57.0	4 Tahun - 10 Bulan - 27 Hari	20.0	116.0	T

Gbr. 4 Hasil Data Cleaning

2) *Data Transformation*

Tahap ini dilakukan untuk menyamakan format data usia, yang semula tercatat dalam format campuran (tahun, bulan, hari), menjadi format numerik dalam satuan tahun. Proses ini bertujuan agar data usia lebih konsisten dan dapat dianalisis secara komputasional.

	JK	BB Lahir	TB Lahir	Berat	Tinggi	Naik Berat Badan	Usia Tahun
0	L	2.0	50.0	21.0	114.0	T	4.904566
1	P	2.0	48.0	10.5	82.6	T	1.933105
2	L	3.0	50.0	10.2	84.4	T	1.674886
3	L	2.0	57.0	20.0	116.0	T	4.907306
4	P	3.0	49.0	11.3	81.5	N	1.952283
...	...	...	...	...	...	...	...
862	P	3.4	51.0	22.0	110.7	T	4.774658
863	P	3.3	50.0	12.0	95.0	N	3.838813
864	P	3.6	51.0	19.8	112.3	N	4.232420
865	L	2.9	46.0	10.7	85.8	T	2.347032
866	L	3.0	49.0	9.5	80.6	T	1.699543

867 rows × 7 columns

Gbr. 5 Hasil Data Transformation

Setelah transformasi, dilakukan proses penyaringan data berdasarkan rentang usia 1 hingga 5 tahun, karena usia tersebut merupakan fase krusial dalam pertumbuhan dan perkembangan balita. Hasil penyaringan menghasilkan 866 data balita dari total awal 1.021 entri.

3) *Label Encoding*

Algoritma *machine learning* hanya dapat memproses data numerik, sehingga diperlukan teknik *label encoding* untuk mengubah data kategorikal menjadi representasi numerik

Table II
Data Numerik Jenis Kelamin

Jenis Kelamin	Data Numerik
Perempuan (P)	1
Laki-Laki (L)	2

Table III
Data Numerik Naik Berat Badan

Naik Berat Badan	Data Numerik
Normal (N)	1
Turun (T)	2
Overweight (O)	3

Table IV
Data Numerik BB/U

BB/U	Data Numerik
Berat Badan Normal	0
Kurang	1
Risiko Lebih	2
Sangat Kurang	3

Perubahan kolom setelah *label encoding*:

	JK	BB Lahir	TB Lahir	Berat	Tinggi	Naik Berat Badan	Usia Tahun	ZS BB/U	BB/U
0	2	2.0	50.0	21.0	114.0	2	4	0.952941	0
1	1	2.0	48.0	10.5	82.6	2	1	0.555556	0
2	2	3.0	50.0	10.2	84.4	2	1	0.162791	0
3	2	2.0	57.0	20.0	116.0	2	4	0.717647	0
4	1	3.0	49.0	11.3	81.5	1	1	0.911111	0
...	...	...	...	...	...	...	...	...	...
862	1	3.4	51.0	22.0	110.7	2	4	1.108696	2
863	1	3.3	50.0	12.0	95.0	1	3	-0.671233	1
864	1	3.6	51.0	19.8	112.3	1	4	0.630435	0
865	2	2.9	46.0	10.7	85.8	2	2	-0.642857	1
866	2	3.0	49.0	9.5	80.6	2	1	-0.162791	1

867 rows × 9 columns

Gbr. 6 Dataset setelah Label Encoding

4) Feature Selection

Feature selection dilakukan untuk memilih fitur yang paling relevan dalam membangun model klasifikasi. Dalam penelitian ini digunakan metode Pearson Correlation Coefficient untuk mengukur hubungan linier antara masing-masing fitur dengan target variabel (BB/U). Korelasi bernilai mendekati ±1 menunjukkan hubungan yang kuat, sementara nilai mendekati 0 menunjukkan hubungan yang lemah atau tidak signifikan.

	JK	BB Lahir	TB Lahir	Berat	Tinggi	Naik Berat Badan	Usia Tahun	ZS BB/U	BB/U
JK	1.000000	0.032637	0.062756	0.075082	0.080469	-0.051924	0.012355	0.002582	0.026455
BB Lahir	0.032637	1.000000	-0.039262	-0.026649	-0.014466	-0.030058	-0.013609	-0.036580	0.008331
TB Lahir	0.062756	-0.039262	1.000000	0.103284	0.110840	0.025976	0.086713	0.049440	-0.011945
Berat	0.075082	-0.026649	0.103284	1.000000	0.839135	-0.013689	0.713016	0.580925	-0.021573
Tinggi	0.080469	-0.014466	0.110840	0.839135	1.000000	0.010697	0.896812	0.169134	0.040788
Naik Berat Badan	-0.051924	-0.030058	0.025976	-0.013689	0.010697	1.000000	0.010587	-0.047722	0.018950
Usia Tahun	0.012355	-0.013609	0.086713	0.713016	0.896812	0.010587	1.000000	-0.123457	0.138448
ZS BB/U	0.002582	-0.036580	0.049440	0.580925	0.169134	-0.047722	-0.123457	1.000000	-0.194960
BB/U	0.026455	0.008331	-0.011945	-0.021573	0.040788	0.018950	0.138448	-0.194960	1.000000

Gbr. 7 Output Correlations Pearson

Hasil korelasi menunjukkan bahwa fitur yang memiliki korelasi paling signifikan terhadap BB/U:

a) Tinggi Badan; hubungan positif lemah dengan nilai 0.040788.
b) Berat Badan; hubungan positif lemah dengan nilai 0.018950
c) Z-score BB/U; hubungan negatif lemah dengan nilai -0.194960.

5) *Resampling Data*

Resampling merupakan teknik yang digunakan untuk mengatasi permasalahan ketidakseimbangan kelas dalam dataset, yang dapat menyebabkan model cenderung bias terhadap kelas mayoritas. Dalam penelitian ini, dilakukan teknik undersampling, yaitu dengan mengurangi jumlah data dari kelas mayoritas agar seimbang dengan jumlah data pada kelas minoritas.

Dataset awal berjumlah 866 data balita dengan rentang usia 1 hingga 5 tahun. Dari data tersebut, dipilih secara acak sebanyak 120 data yang terbagi secara seimbang ke dalam empat kelas status gizi, yaitu masing-masing 30 data per kelas. Tujuan dari *resampling* ini adalah untuk memastikan bahwa setiap kelas memiliki representasi yang seimbang, sehingga proses pelatihan model dapat berjalan lebih optimal dan tidak memihak pada kelas tertentu. Visualisasi

distribusi jenis kelamin setelah dilakukan proses *resampling* ditampilkan pada Gbr. 8;

Gbr. 8 Visualisasi Distribusi Jenis

6) *Data Splitting*

Data splitting adalah proses pembagian dataset ke dalam dua subset utama, yaitu data training dan data testing. Pembagian ini bertujuan untuk melatih model pada sebagian data (training) dan mengevaluasi performanya pada data yang tidak terlihat sebelumnya (testing), guna menghindari overfitting dan meningkatkan generalisasi model.

Setelah proses *resampling*, diperoleh 120 data balita yang digunakan dalam eksperimen. Penelitian ini melakukan lima skenario pengujian dengan variasi rasio pembagian data training dan data testing sebagai berikut:

Table V
Data Splitting

Pengujian	Training (%)	Testing (%)	Jumlah Data Training	Jumlah Data Testing
1	60%	40%	72	48
2	65%	35%	78	42
3	70%	30%	84	36
4	75%	25%	90	30
5	80%	20%	96	24

Penjelasan tiap skenario adalah sebagai berikut:

1) Pengujian 1: Model dilatih dengan 72 data (60%) dan diuji dengan 48 data (40%).
2) Pengujian 2: Model dilatih dengan 78 data (65%) dan diuji dengan 42 data (35%).
3) Pengujian 3: Model dilatih dengan 84 data (70%) dan diuji dengan 36 data (30%).
4) Pengujian 4: Model dilatih dengan 90 data (75%) dan diuji dengan 30 data (25%).
5) Pengujian 5: Model dilatih dengan 96 data (80%) dan diuji dengan 24 data (20%).

Variasi rasio ini digunakan untuk mengevaluasi sejauh mana proporsi data training dan testing memengaruhi performa model klasifikasi status gizi balita.

B. *Data Processing*

Proses pengolahan data (*data processing*) dalam penelitian ini dilakukan melalui beberapa tahapan penting, salah satunya adalah pemilihan atribut menggunakan metode Information Gain. Tujuan dari tahap ini adalah untuk mengetahui atribut mana yang memberikan kontribusi paling besar terhadap proses klasifikasi status gizi balita.

Information Gain merupakan metode seleksi atribut yang digunakan untuk mengukur kualitas pemisahan data berdasarkan suatu fitur. Semakin tinggi nilai Information Gain, maka semakin besar kontribusi fitur tersebut dalam mengurangi entropi atau ketidakpastian pada data target.

Total semua *entropy* adalah 1.999999. Nilai tersebut mendekati nilai maksimum yang artinya data target terdistribusi hampir merata diantara kelas yang ada. Untuk mengurangi ketidakpastian, fitur dengan nilai *Information Gain* yang tinggi akan membantu memisahkan data dengan lebih baik sehingga dapat membuat data lebih terstruktur dan prediksi lebih akurat.

Table VI
Deskripsi Fitur Information Gain

Deskripsi Fitur Information Gain	
0	Jenis Kelamin
1	BB Lahir
2	TB Lahir
3	Berat
4	Tinggi
5	ZS BB/U
6	Naik Berat Badan
7	Usia Dalam Tahun

Menurut hasil perhitungan *Information Gain* pada Gbr. 9, diperoleh hasil bahwa fitur berat, tinggi dan usia dalam tahun memiliki bobot tinggi dan merupakan pengaruh kuat untuk memprediksi klasifikasi BB/U. Fitur tersebut akan di prioritaskan dalam proses klasifikasi status gizi balita menggunakan algoritma C5.0

```
Information Gain for Feature 0:  0.02163478998064816
Information Gain for Feature 1:  0.5531505853398888
Information Gain for Feature 2:  0.295535305538563
Information Gain for Feature 3:  1.508170413083992
Information Gain for Feature 4:  1.7270426002293164
Information Gain for Feature 5:  0.06190302729855457
Information Gain for Feature 6:  0.12795246407374528
Information Gain for Feature 7:  1.9999999956719148
```

Gbr. 9 Information Gain

C. *Pengujian Model Algoritma C5.0*

Tahap pengujian model dimulai dengan membagi dataset menjadi lima skenario pembagian data (*data splitting*), yang masing-masing menggunakan proporsi berbeda antara data *training* dan *testing*. Pada setiap skenario, model dilatih menggunakan sebagian data dan diuji menggunakan data sisanya. Proses ini diulang sebanyak lima kali untuk memperoleh hasil akurasi yang representatif dari setiap proporsi.

1) Pengujian pertama dilakukan dengan 60% data training dan 40% data testing. hasil akurasi sebesar 0.9792.

Rasio Pembagian: 60% Train - 40% Test

	precision	recall	f1-score	support
0	1.00	0.93	0.96	14
1	1.00	1.00	1.00	12
2	0.92	1.00	0.96	11
3	1.00	1.00	1.00	11
accuracy			0.98	48
macro avg	0.98	0.98	0.98	48
weighted avg	0.98	0.98	0.98	48

Akurasi: 0.9792

Gbr. 10 Hasil Pengujian Pertama

2) Pengujian kedua dilakukan dengan memasukkan 65% data training dan 35% data testing. Pengujian ini memperoleh hasil akurasi sebesar 0.9762.

Rasio Pembagian: 65% Train - 35% Test

	precision	recall	f1-score	support
0	1.00	0.92	0.96	12
1	1.00	1.00	1.00	12
2	0.90	1.00	0.95	9
3	1.00	1.00	1.00	9
accuracy			0.98	42
macro avg	0.97	0.98	0.98	42
weighted avg	0.98	0.98	0.98	42

Akurasi: 0.9762

Gbr. 11 Hasil Pengujian Kedua

3) Pengujian ketiga dilakukan dengan memasukkan 70% data training dan 30% data testing. Pengujian ini memperoleh hasil akurasi sebesar 0.9722.

Rasio Pembagian: 70% Train - 30% Test

	precision	recall	f1-score	support
0	1.00	0.90	0.95	10
1	1.00	1.00	1.00	9
2	0.90	1.00	0.95	9
3	1.00	1.00	1.00	8
accuracy			0.97	36
macro avg	0.97	0.97	0.97	36
weighted avg	0.98	0.97	0.97	36

Akurasi: 0.9722

Gbr. 12 Hasil Pengujian Ketiga

4) Pengujian keempat dilakukan dengan memasukkan 75% data training dan 25% data testing. Pengujian ini memperoleh hasil akurasi sebesar 0.9667.

Rasio Pembagian: 75% Train - 25% Test

	precision	recall	f1-score	support
0	1.00	0.88	0.93	8
1	1.00	1.00	1.00	8
2	0.89	1.00	0.94	8
3	1.00	1.00	1.00	6
accuracy			0.97	30
macro avg	0.97	0.97	0.97	30
weighted avg	0.97	0.97	0.97	30

Akurasi: 0.9667

Gbr. 13 Hasil Pengujian Keempat

5) Pengujian kelima atau pengujian terakhir dilakukan dengan memasukkan 80% data training dan 20% data testing. Pengujian ini memperoleh hasil akurasi sebesar 0.9583.

Rasio Pembagian: 80% Train - 20% Test

	precision	recall	f1-score	support
0	1.00	0.83	0.91	6
1	1.00	1.00	1.00	7
2	0.88	1.00	0.93	7
3	1.00	1.00	1.00	4
accuracy			0.96	24
macro avg	0.97	0.96	0.96	24
weighted avg	0.96	0.96	0.96	24

Akurasi: 0.9583

Gbr. 14 Hasil Pengujian Kelima

Pengujian model klasifikasi menggunakan algoritma C5.0 menghasilkan akurasi tertinggi sebesar 0.9792 atau 97,92%. Model tertinggi dihasilkan dari data splitting berupa 60% data training dan 40% data testing. 5 hasil akurasi seperti table VI menunjukan pengujian tertinggi yaitu pada pembagian 60% data training dan 40% data testing yang menghasilkan akurasi sebesar 0.9792 atau 97,92%.

Table VII
Hasil Pengujian Terbaik

Pengujian	60% - 40%	65% - 35%	70% - 30%	75% - 25%	80% - 20%
1	0,9792				
2		0,9762			
3			0,9722		
4				0,9667	
5					0,9583

D. *Evaluasi dan Validasi Hasil Algoritma C5.0*

Untuk mengukur performa dan validitas model klasifikasi yang dibangun menggunakan algoritma C5.0, dilakukan evaluasi menggunakan *Confusion matrix* dan kurva ROC (*Receiver Operating Characteristic*). Dari *Confusion matrix* diperoleh nilai-nilai evaluasi *Accuracy* , *Precision* , dan *Recall*, sedangkan kurva ROC digunakan untuk menghitung nilai AUC (*Area Under Curve*) guna mengevaluasi kemampuan model dalam membedakan antar kelas. Hasil-hasil dari pengujian ini akan dibahas lebih lanjut pada sub-bagian berikut.

1) Confusion Matrix

Confusion matrix yang digunakan dalam penelitian ini memperlihatkan empat kelas, yaitu:

Table VIII
Kelas Confusion Matrix

Kelas 0	Berat badan normal
Kelas 1	Kurang
Kelas 2	Risiko lebih
Kelas 3	Sangat kurang

Pada *confusion matrix*, sumbu Y merepresentasikan *true label* (label asli), sementara sumbu X merepresentasikan *predicted label* (label yang diprediksi oleh model). Visualisasi ini memungkinkan peneliti untuk menganalisis seberapa baik model dalam memprediksi kelas target dibandingkan dengan label yang sebenarnya.

a) *Confusion matrix* pada pengujian pertama, model berhasil memprediksi 13 data pada kelas 0 dengan benar, namun terdapat 1 kesalahan di mana data tersebut diprediksi sebagai kelas 2. Kelas lainnya

menunjukkan hasil yang cukup baik, dengan 12 data pada kelas 1, 11 data pada kelas 2, dan 11 data pada kelas 3 yang diprediksi benar. Pengujian ini menunjukkan mayoritas data diprediksi dengan benar, dengan hanya satu kesalahan klasifikasi.

b) *Confusion matrix* Pada pengujian kedua, hasilnya sedikit menurun, dengan 11 data pada kelas 0 diprediksi benar dan satu kesalahan klasifikasi ke kelas 2. Kelas 1, 2, dan 3 masing-masing memiliki 12, 9, dan 9 data yang diprediksi benar. Secara keseluruhan, hanya terdapat dua kesalahan klasifikasi pada pengujian ini.

c) *Confusion matrix* Pengujian ketiga menunjukkan hasil yang lebih konsisten, dengan 9 data pada kelas 0 diprediksi benar, dan 1 kesalahan klasifikasi ke kelas 2. Kelas 1, 2, dan 3 masing-masing memiliki 9 data yang diprediksi benar, dengan total kesalahan klasifikasi hanya satu kali.

d) *Confusion matrix* Pada pengujian keempat, kelas 0 memiliki 7 data yang diprediksi benar dan satu kesalahan klasifikasi ke kelas 2. Kelas 1, 2, dan 3 masing-masing memiliki 8 data yang diprediksi benar, dengan hanya satu kesalahan klasifikasi yang terjadi.

e) *Confusion matrix* Hasil pengujian kelima menunjukkan 5 data pada kelas 0 yang diprediksi benar dan satu kesalahan klasifikasi ke kelas 2. Kelas 1 dan 2 masing-masing memiliki 7 data yang diprediksi benar, sementara kelas 3 hanya memiliki 4 data yang diprediksi benar. Sekali lagi, pengujian ini menunjukkan mayoritas data diprediksi dengan benar, dengan hanya satu kesalahan klasifikasi.

Model klasifikasi menunjukkan performa baik dan konsisten pada semua pembagian data. Semua pembagian menghasilkan akurasi tinggi. Hal tersebut menunjukkan bahwa model mampu mempelajari pola data dengan baik. Berikut adalah tabel *Confusion matrix* dari pengujian dengan nilai tertinggi, yaitu 60% data training dan 40% data testing.

Table IX
Confusion Matrix

True Positive (TP)	47
False Positive (FP)	0
True Negative (TN)	0
False Negative (FN)	1

2) *Accuracy*

$$Accuracy = \frac{(TP+TN)}{Total\ Data}\ x\ 100\%$$

$$Accuracy = \frac{(47+0)}{48}\ x\ 100\%$$

$$Accuracy = \frac{47}{48}\ x\ 100\%$$

$$Accuracy = 0{,}9792 = 97{,}92\%$$

Berdasarkan hasil pengujian yang telah dilakukan, dapat ditarik kesimpulan bahwa presentase *Accuracy* pada pengujian algoritma C5.0 dari 40% data testing sebesar 97,92%

3) *Precision*

$$Precision = \frac{TP}{(TP+FP)}$$

$$Precision = \frac{47}{(47+0)}$$

$$= 1 = 100\%$$

Berdasarkan hasil pengujian yang telah dilakukan, dapat ditarik kesimpulan bahwa presentase *Precision* pada pengujian algoritma C5.0 dari 40% data testing sebesar 100%.

4) *Recall*

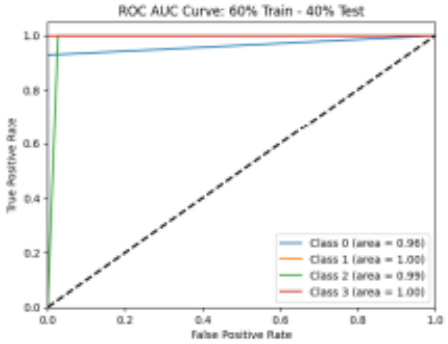
$$Recall = \frac{TP}{(TP+FN)}$$

$$Recall = \frac{47}{(47+1)} = 0{,}9792 = 97{,}92\%$$

Berdasarkan hasil pengujian yang telah dilakukan, dapat ditarik kesimpulan bahwa presentase *Recall* pada pengujian algoritma C5.0 dari 40% data testing sebesar 97,92%.

E. *Kurva ROC*

1) Kurva ROC pengujian pertama



Gbr. 15 Kurva ROC Pengujian

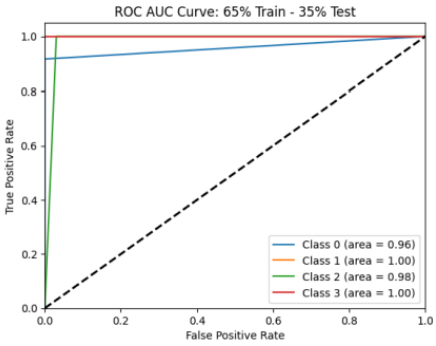
Gambar 4.24 menunjukkan kurva ROC dari pengujian pertama, yang menggunakan 60% data untuk pelatihan dan 40% data untuk pengujian. Setiap garis dalam kurva ini mewakili performa model dalam membedakan satu kelas dengan kelas lainnya. Nilai AUC untuk setiap kelas dihitung sebagai berikut:

$$AUC\ macro = \frac{AUC(class\ 0) + AUC(class\ 1) + AUC(class\ 2) + AUC(class\ 3)}{Jumlah\ Kelas}$$

$$AUC\ macro = \frac{0{,}96+1{,}00+0{,}99+1{,}00}{4} = \frac{3{,}95}{4} = 0{,}987$$

Hasil AUC macro sebesar 0.987 menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik, mendekati nilai sempurna.

2) Kurva ROC pengujian kedua



Gbr. 16 Kurva ROC Pengujian

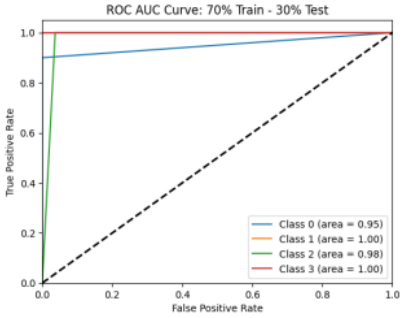
Gambar diatas menunjukkan hasil dari kurva ROC dengan 65% data training dan 35% data testing. Setiap garis mewakili performa model dalam membedakan satu kelas dengan kelas yang lain. Nilai area under the curve (AUC) digunakan untuk mengukur klasifikasi sebagai berikut :

$$AUC\ macro = \frac{AUC(class\ 0) + AUC(class\ 1) + AUC(class\ 2) + AUC(class\ 3)}{Jumlah\ Kelas}$$

$$AUC\ macro = \frac{0{,}96+1{,}00+0{,}98+1{,}00}{4} = \frac{3{,}94}{4} = 0{,}985$$

Hasil AUC macro sebesar 0,985 menunjukkan kemampuan klasifikasi yang sangat baik secara keseluruhan, yang hampir mencapai nilai maksimal.

3) Kurva ROC pengujian ketiga



Gbr. 17 Kurva ROC pengujian ketiga

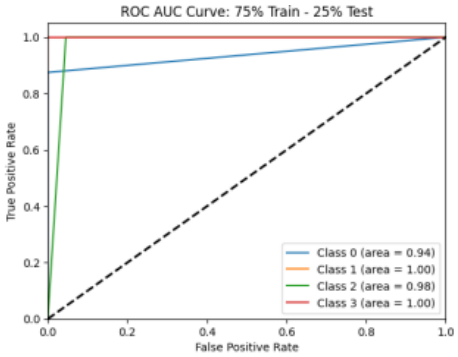
Gambar diatas menunjukkan hasil dari kurva ROC dengan 70% data training dan 30% data testing. Setiap garis mewakili performa model dalam membedakan satu kelas dengan kelas yang lain. Nilai area under the curve (AUC) digunakan untuk mengukur klasifikasi sebagai berikut :

$$AUC\ macro = \frac{AUC(class\ 0) + AUC(class\ 1) + AUC(class\ 2) + AUC(class\ 3)}{Jumlah\ Kelas}$$

$$AUC\ macro = \frac{0{,}95+1{,}00+0{,}98+1{,}00}{4} = \frac{3{,}93}{4} = 0{,}982$$

Hasil AUC macro yang mendekati nilai sempurna yaitu 0,982. menunjukkan bahwa model memiliki kemampuan klasifikasi tinggi secara keseluruhan.

4) Kurva ROC pengujian keempat



Gbr. 18 Kurva ROC pengujian keempat

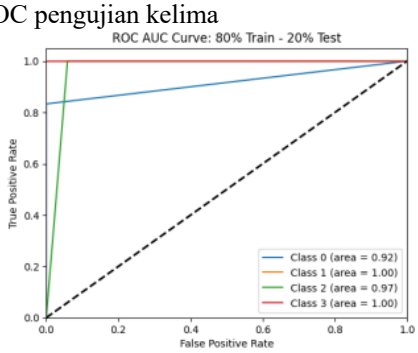
Gambar diatas menunjukkan hasil dari kurva ROC dengan 75% data training dan 25% data testing. Setiap garis mewakili performa model dalam membedakan satu kelas dengan kelas yang lain. Nilai area under the curve (AUC) digunakan untuk mengukur klasifikasi sebagai berikut :

$$AUC\ macro = \frac{AUC(class\ 0) + AUC(class\ 1) + AUC(class\ 2) + AUC(class\ 3)}{Jumlah\ Kelas}$$

$$AUC\ macro = \frac{0{,}94+1{,}00+0{,}98+1{,}00}{4} = \frac{3{,}92}{4} = 0{,}98$$

Hasil AUC macro yang mendekati nilai sempurna yaitu 0,98 menandakan model memiliki kemampuan klasifikasi yang tinggi secara keseluruhan.

5) Kurva ROC pengujian kelima



Gbr. 19 Kurva ROC pengujian kelima

Gambar diatas menunjukkan hasil dari kurva ROC dengan 80% data training dan 20% data testing. Setiap garis mewakili performa model dalam membedakan satu kelas dengan kelas yang lain. Nilai area under the curve (AUC) digunakan untuk mengukur klasifikasi sebagai berikut :

$$AUC\ macro = \frac{AUC(class\ 0) + AUC(class\ 1) + AUC(class\ 2) + AUC(class\ 3)}{Jumlah\ Kelas}$$

$$AUC\ macro = \frac{0{,}92+1{,}00+0{,}97+1{,}00}{4} = \frac{3{,}89}{4} = 0{,}972$$

Hasil AUC macro yang mendekati nilai sempurna yaitu 0,972 menandakan model memiliki kemampuan klasifikasi yang tinggi secara keseluruhan.

F. Analisis Hasil Klasifikasi Algoritma C5.0

Pengujian klasifikasi dilakukan menggunakan algoritma C5.0 dengan bahasa pemrograman *Python* pada *Google Colab*. Atribut yang digunakan dalam proses klasifikasi mencakup jenis kelamin, berat badan saat lahir, tinggi badan

saat lahir, berat badan, tinggi badan, naik berat badan, usia, dan ZS BB/U.

Proses klasifikasi dilakukan melalui lima kali pengujian dengan rasio data training dan data testing yang bervariasi. Salah satu pengujian yang menggunakan rasio 60% data training dan 40% data testing menunjukkan hasil terbaik dibandingkan dengan rasio pembagian data lainnya. Performa model pada pengujian ini dievaluasi menggunakan *Confusion matrix* yang mencakup metrik evaluasi seperti *Accuracy* , *Precision* , *Recall*, serta kurva ROC dan nilai AUC. Hasil dari *Confusion matrix* menunjukkan bahwa mayoritas data berhasil diklasifikasikan dengan benar, yang dapat dilihat dari nilai true positive (TP) yang tinggi serta tidak adanya false positive (FP) dan false negative (FN). Hasil akan disajikan dalam Tabel IX;

Table X
Hasil Pengujian Algoritma 5.0

Matrik	Hasil Pengujian	Keterangan
Accuracy	97,92%	Tergolong tinggi, menunjukkan sebagian besar data diklasifikasikan dengan benar.
Precision	100%	Nilai sempurna, tidak ada kesalahan dalam klasifikasi positif.
Recall	97,92%	Model mampu mengidentifikasi status gizi balita dengan tepat dan jarang terjadi kesalahan klasifikasi.
AUC (Area Under Curve)	Kelas 0: 0,96	AUC untuk kelas 0: 0,96, kelas 1: 1.00, kelas 2: 0,99, kelas 3: 1.00
Rata-rata AUC *Macro*	0,987	Menunjukkan kemampuan klasifikasi yang sangat baik dalam membedakan antara status gizi balita.

Secara keseluruhan, hasil pengujian ini mengindikasikan bahwa algoritma C5.0 dapat menghasilkan model klasifikasi yang sangat baik dalam memprediksi status gizi balita berdasarkan atribut yang diberikan.

IV. KESIMPULAN

Implementasi algoritma C5.0 untuk klasifikasi status gizi balita di Kecamatan Ponorogo dilakukan melalui berbagai tahapan, mulai dari pengumpulan data yang diperoleh dari Puskesmas Ponorogo Selatan, hingga proses data *preprocessing* yang mencakup pembersihan data, transformasi ke format yang sesuai, *label encoding*, seleksi fitur, *resampling* data, dan pembagian data untuk lima kali pengujian. Selanjutnya, dilakukan data *processing* untuk menghitung *entropy* dan *information gain*, pengujian model menggunakan algoritma C5.0, serta evaluasi dan validasi hasil dengan menggunakan *Confusion matrix* dan kurva ROC. Proses akhir adalah analisis hasil klasifikasi. Hasil pengujian menunjukkan bahwa akurasi tertinggi yang diperoleh dari penelitian ini mencapai 97,92% (0,9792), yang dihasilkan dari skema pembagian data sebesar 60% untuk *data training* dan 40% untuk *data testing*. Nilai ini mengindikasikan bahwa algoritma C5.0 memiliki tingkat ketepatan klasifikasi yang sangat baik.

X. REFERENSI

[1] Basilius, Y. (2020). "Klasifikasi Status Gizi Balita Menggunakan Metode Decision Tree C4.5." Universitas Sanata Dharma Yogyakarta Fakultas Sains Dan Teknologi Program Studi Informatika

[2] Kemenkes, R. (2023). Buku kesehatan ibu dan anak. In Kementerian kesehatan RI.

[3] Abidin, Z., Nurhana, E., Permata, & Ulum, F. (2023). "Analisis Perbandingan Algoritma Decision Tree C4.5 dan C5.0 pada Data Karyawan Berpotensi Promosi Jabatan". Jurnal Teknoinfo, 17(2), 567–582. https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index.

[4] Harliana, & Anggraini, D. (2023). Penerapan Algoritma Naïve Bayes Pada Klasifikasi Status Gizi Balita di Posyandu Desa Kalitengah. Jurnal Informatika Komputer, Bisnis Dan Manajemen, 21(2), 38– 45. https://doi.org/10.61805/fahma.v21i2.16

[5] Benediktus, N., & Oetama, R. S. (2020). Algoritma Klasifikasi Decision Tree C5.0 untuk Memprediksi Performa Akademik Siswa Natanael. Ultimatics : Jurnal Teknik Informatika, 12(1), 14–19

[6] Damanik, S. F., Wanto, A., & Gunawan, I. (2022). Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok. Jurnal Krisnadana, 1(2), 21–32. https://doi.org/10.58982/krisnadana.v1i2.108

[7] Devi Asri, S., & Miftahul Huda, A. (2023). "Implementation Of The C5.0 Algorithm In Classification Of Social Data" (Case Study: Eligibility of BLT Acceptance in Condong Village, Singkawang City). Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster), 12(3), 259–268 https://doi.org/10.26418/bbimst.v12i3.66693

[8] Effendi, M., Ariani, R. D., & Astuti, R. (2021). Determination of Work Schedule Based on Employee Data Classification Using the Decision Tree Algorithm C4.5 Method. Industria: Jurnal Teknologi Dan Manajemen Agroindustri, 10(3), 249–259. https://doi.org/10.21776/ub.industria.2021.010.03.6.

[9] Fajri, M., Utami, I. T., & Maruf, Muh. (2022). Perbandingan Model Pohon Klasifikasi Algoritma C4.5 dan C5.0 untuk Analisis Faktor yang Mempengaruhi Keberhasilan Lelang. Indonesian Journal of Statistics and Its Applications, 6(1), 13–22. https://doi.org/10.29244/ijsa.v6i1p13-22.

[10] Sahputra, I., Mauliza, M., & Zohra, S. F. A. (2023). Implementasi Algoritma C5.0 Pada Klasifikasi Status Gizi Ibu Hamil di Kota Lhokseumawe. METIK JURNAL, 7(1), 42–46. https://doi.org/10.47002/metik.v7i1.562

[11] Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 3(1), 15–19. https://doi.org/10.57152/malcom.v3i1.591

[12] Siti Sahara Nasution, Natalia Silalahi, & Abdul Karim. (2021). Implementasi Algoritma C5.0 Untuk Memprediksi Kondisi Kelahiran Bayi. Bulletin of Computer Science Research, 2(1), 18–24. https://doi.org/10.47065/bulletincsr.v2i1.138

[13] Sugiyono. (2020). Metodologi Penelitian Kuantitatif, Kualitatif dan R & D