



Implementasi Algoritma C4.5 Pada Klasifikasi Status Gizi Balita Menggunakan Framework Flask

Hafizh Prayoga¹, Rakhmat Kurniawan¹

¹Program Studi Sistem Informasi, Universitas Islam Negeri Sumatera Utara, Indonesia.

Artikel Info

Kata Kunci:

Algoritma C4.5;
Data Mining;
Flask;
Klasifikasi Status Gizi;
Pohon Keputusan;
Stunting.

Keywords:

*C4.5 Algorithm;
Classification Nutritional Status;
Data Mining;
Decision Tree;
Flask;
Stunting.*

Riwayat Artikel:

Submitted: 20 Juni 2025

Accepted: 31 Juli 2025

Published: 31 Juli 2025

Abstrak: Pemantauan status gizi balita berperan krusial dalam mencegah stunting melalui intervensi dini yang tepat sasaran. Namun, keterbatasan metode manual dalam klasifikasi data gizi menyebabkan keterlambatan dan risiko kesalahan pengambilan keputusan. Penelitian ini mengembangkan sistem klasifikasi status gizi balita berbasis web menggunakan algoritma C4.5 yang mampu mengidentifikasi pola hubungan antar atribut untuk menghasilkan prediksi status gizi secara otomatis dan akurat. Sistem dibangun dengan *framework Flask* guna memastikan aksesibilitas dan efisiensi pemrosesan. Data yang digunakan berasal dari Dinas Kesehatan Kota Pematangsiantar tahun 2022-2024 sebanyak 697 data. Proses penelitian meliputi pengumpulan data, *preprocessing* mencakup pembersihan dan transformasi data, *split* data dengan rasio 70:30, serta pelatihan model, dan evaluasi menggunakan *confusion matrix*. Hasil menunjukkan akurasi model mencapai 95% pada data uji, menandakan kinerja yang sangat baik dalam menggeneralisasi data baru. Sistem yang dikembangkan dilengkapi fitur visualisasi pohon keputusan, evaluasi model, pembaruan dataset, serta monitoring prevalensi gizi per kecamatan. Inovasi ini mendukung pengambilan keputusan berbasis data yang strategis merumuskan program-program yang menunjang pencegahan ataupun intervensi yang tepat.

Abstract: Monitoring the nutritional status of children under five plays a crucial role in preventing stunting through targeted early intervention. However, the limitations of manual methods in nutrition data classification cause delays and the risk of decision-making errors. This research develops a web-based toddler nutritional status classification system using the C4.5 algorithm that is able to identify patterns of relationships to produce nutritional status predictions automatically and accurately. The system is built with the Flask framework to ensure accessibility and processing efficiency. The data used comes from the Pematangsiantar City Health Office for 2022-2024, totaling 697 data. The research process includes data collection, preprocessing including data cleaning and transformation, Split the data into a 70:30 ratio and model training, and evaluation using confusion matrix. The results showed that the model accuracy reached 95% on the test data, indicating excellent performance in generalizing new data. The developed system features decision tree visualization, model evaluation, dataset updates, and monitoring of nutrition prevalence per sub-district. This innovation supports strategic data-based decision-making to formulate programs that support prevention or appropriate intervention.

Corresponding Author:

Hafizh Prayoga

Email: hafizprayoga23@gmail.com

PENDAHULUAN

Perkembangan teknologi informasi (TI) saat ini telah memberikan dampak yang signifikan di berbagai sektor kesehatan (El-Hasnony et al., 2022). Digitalisasi sistem kesehatan memungkinkan pengolahan data secara lebih cepat dan akurat, mempercepat proses pengambilan keputusan klinis. Salah satu pemanfaatan TI dalam sektor kesehatan adalah penerapan teknik *Data Mining* (Ozcan & Peker, 2023) dan *Machine Learning* (Rastogi & Bansal, 2023) yang telah terbukti mendukung klasifikasi (Alshammari, 2024), prediksi (Febrian et al., 2025), dan analisis data medis (Anthira & Suendri, 2024) (Jang & Nemoto, 2024). Penerapan algoritma tersebut berkontribusi dalam meningkatkan efisiensi dan efektivitas layanan kesehatan.

Di tingkat daerah, Dinas Kesehatan Kota Pematang Siantar berperan strategis dalam merumuskan dan melaksanakan berbagai kebijakan serta program kesehatan masyarakat yang komprehensif, dalam mengupayakan pencegahan stunting. Salah satu aspek penting dalam upaya pencegahannya adalah pemantauan status gizi balita, yang merupakan indikator utama dalam menilai pertumbuhan dan perkembangan anak (Madiko et al., 2023). Status gizi yang buruk pada balita dapat berdampak negatif terhadap perkembangan fisik, mental, dan kualitas hidup anak dalam jangka panjang (Yuningsih, 2022). Oleh karena itu, pemantauan status gizi sangat diperlukan agar dapat dilakukan intervensi yang tepat sejak dini.

Namun saat ini, proses mengklasifikasikan status gizi balita dimulai dengan pengukuran balita yang dilakukan oleh Posyandu di tiap kelurahan yang berada di Kota Pematang Siantar kemudian dicatat secara manual di kertas dan dikirimkan ke Dinas Kesehatan Pematang Siantar untuk dikonfirmasi, diverifikasi, divalidasi, dan dianalisis. Setelah diperoleh data prevalensi stunting, barulah dilakukan intervensi yang diperlukan, yang didukung oleh PEMKO Pematang Siantar dan dinas-dinas terkait. Tentunya, hal ini menyebabkan proses pemantauan dan intervensi menjadi kurang efektif karena memerlukan waktu yang cukup lama dalam pengolahan data, serta berisiko terjadi kesalahan dalam klasifikasi status gizi akibat keterbatasan metode manual yang digunakan. Selain itu, seiring bertambahnya volume data balita yang terus bertambah naik membuat proses klasifikasi menjadi semakin lambat, sehingga menghambat pengambilan keputusan yang cepat dan tepat dalam upaya pencegahan stunting.

Untuk mengatasi permasalahan ini, diperlukan sistem berbasis web yang dapat mengotomatisasi proses klasifikasi status gizi balita dengan akurasi tinggi dan waktu pemrosesan yang lebih cepat. Salah satu pendekatan yang potensial adalah penerapan algoritma *C4.5* dalam sistem klasifikasi, mengingat kemampuannya dalam menghasilkan aturan keputusan yang transparan dan akurat. Implementasi algoritma *C4.5* melalui framework *Flask* diharapkan mampu meningkatkan objektivitas dan kecepatan dalam proses klasifikasi status gizi.

Berbagai studi sebelumnya menunjukkan bahwa Algoritma *C4.5* telah banyak digunakan untuk menghasilkan klasifikasi yang akurat di berbagai bidang. Implementasi *C4.5* menunjukkan model yang efektif untuk digunakan sebagai dasar sistem klasifikasi penerima Program Keluarga Harapan (PKH) pada penelitian yang dilakukan di Dinas Sosial Kabupaten Manokwari (Daud et al., 2025). Dalam dunia perbankan, algoritma ini digunakan untuk mengklasifikasi nasabah berdasarkan karakteristik tertentu (Limabri et al., 2023). Di sektor pendidikan, algoritma ini membantu proses penjurusan siswa sekolah menengah atas agar minat dan bakat peserta didik tersalurkan secara efektif (Eirlangga & Syaputra, 2022), serta mengkategorikan peserta didik berdasarkan faktor relevan lainnya. (F Amalia, 2023), untuk mengukur tingkat akurasi kepuasan mahasiswa terhadap pelayanan administrasi pada Fakultas Sains dan Teknologi UIN Sumatera Utara selama masa pandemi Covid-19 (Suendri, 2020), dan membantu menentukan klasifikasi ketepatan kelulusan siswa SMK guna mendukung peningkatan tingkat kelulusan (Lestari & Dina, 2023). Selain itu, algoritma ini juga diterapkan di bidang perikanan untuk mengidentifikasi hubungan antara atribut yang memengaruhi hasil produksi (Nurrajiati et al., 2024) dan juga di bidang pertanian, penggunaan algoritma *C4.5* dalam memprediksi jumlah produksi kelapa sawit dapat memberikan hasil yang akurat dengan tingkat akurasi yang tinggi. Berdasarkan studi-studi tersebut, dapat disimpulkan bahwa *C4.5* mampu mengidentifikasi pola dan hubungan antaratribut

dengan baik, sehingga dapat diterapkan dalam proses pengambilan keputusan yang lebih sistematis dan berbasis data.

Dalam konteks klasifikasi status gizi balita, sejumlah algoritma lain juga telah digunakan, seperti k-Nearest Neighbors (KNN) (Bachri & Bhakti, 2021) dan Naïve Bayes (Rozaq & Purnomo, 2022). Studi yang dilakukan Bachri menunjukkan bahwa KNN dapat memberikan akurasi yang cukup baik dalam klasifikasi status gizi, namun algoritma ini sangat bergantung pada pemilihan parameter k dan tidak menghasilkan aturan yang mudah diinterpretasikan. Naïve Bayes efisien namun lemah jika data latih sangat sedikit dan lemah dalam menangani ketergantungan antaratribut seperti terdapat transformasi data numerik menjadi data kategorik sehingga atribut yang tidak terwakili dalam data latih dapat mempengaruhi tingkat akurasi. Sebaliknya, algoritma C4.5 menawarkan keunggulan yang menonjol, terutama dalam hal interpretabilitas dan efektivitasnya dalam menangani data kategorikal maupun numerik. C4.5 membangun model dalam bentuk pohon keputusan yang mudah dipahami oleh pengguna non-teknis, seperti tenaga kesehatan di lapangan..

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem klasifikasi status gizi balita berbasis web menggunakan algoritma C4.5. Sistem ini akan menggunakan algoritma C4.5, yang berfungsi untuk membangun pohon keputusan berdasarkan atribut-atribut relevan seperti berat badan, tinggi badan, usia, dan jenis kelamin balita. Algoritma C4.5 akan digunakan untuk mengolah data, menentukan aturan klasifikasi, serta menghasilkan prediksi status gizi balita secara otomatis dan akurat. Sistem ini akan dikembangkan menggunakan framework *Flask* untuk memastikan kemudahan akses dan efisiensi dalam pemrosesan data. Sistem ini diharapkan dapat membantu petugas kesehatan dalam mengidentifikasi status gizi balita secara lebih mudah, cepat, dan akurat, sekaligus mendukung upaya pencegahan dan penanganan masalah gizi di kota Pematang Siantar secara lebih optimal.

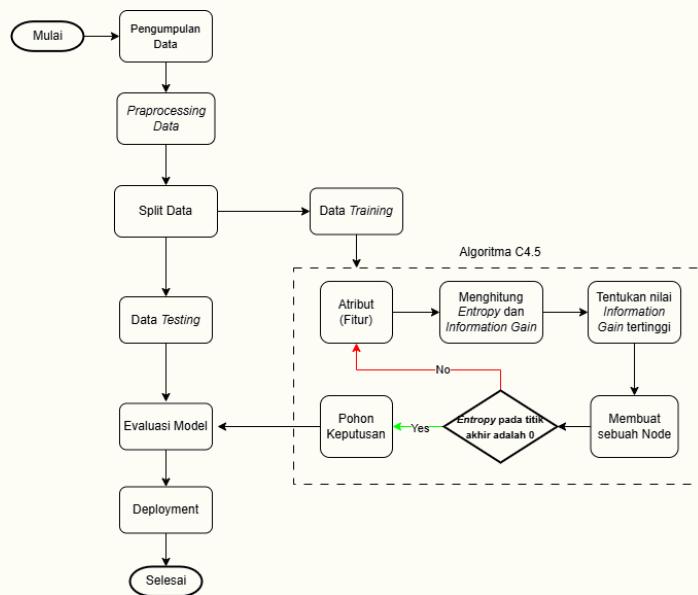
METODE

Klasifikasi Status Gizi Balita

Balita merupakan anak yang berada pada rentang usia 0 hingga 59 bulan, yaitu masa yang ditandai dengan laju pertumbuhan dan perkembangan yang sangat cepat dan asupan nutrisi selama periode *golden age* ini dapat memberikan efek jangka panjang pada kesehatan anak di masa depan (Munawaroh et al., 2022). Melalui klasifikasi ini, dapat diketahui apakah seorang balita mendapatkan gizi yang cukup, mengalami kekurangan gizi, atau gizi buruk. Mengklasifikasikan status gizi balita melibatkan pengkategorian anak-anak berdasarkan kondisi gizi mereka, yang dinilai melalui faktor-faktor seperti berat badan, tinggi badan, dan parameter lain yang relevan (Sinaga et al., 2023). Klasifikasi ini sangat penting karena memberikan informasi yang berharga untuk upaya pencegahan dan intervensi gizi.

Tahapan Penelitian

Penelitian ini memiliki beberapa tahapan yang bisa dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berikut adalah penjelasan untuk setiap tahapan dalam penelitian ini :

1. Pengumpulan Data

Populasi dalam penelitian ini mencakup seluruh data balita yang terdokumentasi di Dinas Kesehatan Kota Pematang Siantar selama 2022-2024 adalah sebanyak 697 baris data. Data tersebut mencakup atribut usia, berat badan (dalam kg), tinggi badan (dalam cm), jenis kelamin, dan status gizi (gizi baik, gizi kurang, atau gizi buruk).

Metode pengumpulan data dilakukan melalui observasi, wawancara, serta studi pustaka. Observasi dilakukan pada Dinas Kesehatan Pematang Siantar untuk melihat secara langsung bagaimana proses untuk mendapatkan data klasifikasi status gizi balita. Kemudian melakukan wawancara terkait petugas yang terlibat dalam proses yang dilakukan untuk menggali kebutuhan, tujuan dan hasil dari sistem yang akan dibangun serta pengumpulan data yang dibutuhkan untuk membangun prediksi klasifikasi. Studi pustaka dilakukan untuk mendapatkan referensi dari berbagai sumber, seperti buku, artikel jurnal, skripsi dan sebagainya, untuk mengetahui seberapa relevan, dan keterkaitan dalam membangun sebuah sistem.

2. Preprocessing Data

Tahap ini merupakan langkah awal dalam proses pengolahan data, di mana data yang akan digunakan perlu melalui proses pembersihan terlebih dahulu untuk memastikan kualitas dan konsistensinya. Data yang dibersihkan meliputi data yang tidak lengkap (*missing values*), hilang, atau tidak valid (*outlier*). Hal ini bertujuan untuk menghapus *noise* serta data yang tidak konsisten atau kurang relevan (Han et al., 2022). Selanjutnya transformasi data untuk mempermudah dalam proses untuk mengetahui hasil analisis data. Hasil dari *preprocessing data*, akan dijadikan sebagai *Dataset*.

3. Split Data

Sebelum membangun model, *dataset* dipisah secara acak menjadi dua bagian dengan proporsi 70:30 dengan tujuan data *training* untuk melatih dan membangun model dan data *testing* untuk mengevaluasi kinerja model yang telah dibangun terhadap data baru yang belum pernah dilatih sebelumnya, dimana 70% dari total *dataset* untuk data *training* dan 30% dari total *dataset* untuk data *testing*.

4. Pembangunan Model dan Evaluasi

Dari data *training* model klasifikasi dibangun menggunakan algoritma *C4.5*, yang merupakan pengembangan dari *ID3* dan memiliki kemampuan menangani atribut numerik dan atribut kategorikal, serta menghasilkan model pohon keputusan yang dapat dijelaskan (*interpretable*).

Setelah model berhasil dibangun kemudian model dievaluasi menggunakan *confusion matrix* untuk melihat distribusi prediksi benar dan salah pada masing-masing kelas. Informasi dari *confusion matrix* dapat digunakan untuk menghitung *accuracy*, *precision*, *recall*, dan skor f-1, dengan keempat metrik tersebut digunakan untuk mengukur kinerja model secara kuantitatif.

5. Deployment

Flask adalah sebuah *Framework web* ringan yang ditulis dalam bahasa *Python* (Irmayanti, 2023). Model *C4.5* untuk klasifikasi status gizi balita yang telah dilatih dan dievaluasi kemudian diimplementasikan ke dalam sistem berbasis web yang dibangun menggunakan *framework Flask*. Dengan tujuan agar dapat diakses dan dimanfaatkan oleh tenaga kesehatan atau pihak terkait untuk mempermudah dalam melakukan klasifikasi status gizi balita secara cepat dan akurat

Algoritma C4.5

Algoritma *C4.5* adalah algoritma klasifikasi berbasis pohon keputusan yang dikembangkan oleh J. Ross Quinlan, yang merupakan pengembangan dari *ID3*. Algoritma *C4.5*, sebagai salah satu implementasi *decision tree*, memilih atribut berdasarkan nilai *gain* tertinggi dan tingkat *entropy*, sehingga menghasilkan pohon keputusan yang optimal (Adinda, 2022) untuk mendukung pengambilan keputusan secara akurat dan efisien.

Dari data pelatihan pohon keputusan akan dibangun dengan menghitung *entropy* dan *information gain*. Pemilihan atribut mana yang akan menjadi simpul ditentukan dengan nilai *gain* tertinggi. Proses ini berlanjut secara iterative hingga semua entropy pada titik akhir pohon keputusan mencapai nol (Rofiani et al., 2024). *Entropy* dan *Information Gain* didefinisikan sebagai berikut :

Persamaan 1, *Entropy* digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut.

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i \log 2p_i$$

(Pratama et al., 2023) (1)

dimana :

S : Himpunan Kasus

N : Jumlah Partisi S

pi : Proporsi dari Si terhadap S

Persamaan 2, *Information gain* digunakan untuk menghitung seberapa efektif sebuah atribut dalam mengklasifikasikan data.

$$\text{Gain}(S.A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

(Pratama et al., 2023) (2)

dimana :

S : Himpunan Kasus

A : Atribut

Si : Jumlah Kasus pada Partisi ke-i

N : Jumlah Partisi Atribut A

|Si| : Jumlah Kasus dalam S

Diketahui dataset *D* dan atribut kontinu *a*, misalkan kita mengamati *n* nilai *a* dalam *D* dan menyusun mereka dalam urutan terkecil ke terbesar $\{a^1, a^2, \dots, a^n\}$. Dengan titik pemisah *t*, *D* dibagi menjadi subset *D_t^-* dan *D_t^+*. Dimana *D_t^-* mencakup sampel dengan nilai *t* dan kurang dari *t*, dan *D_t^+* mencakup sampel dengan nilai lebih besar dari *t*.

$$T_a = \left\{ \frac{a^i + a^{i+1}}{2}, 1 \leq i \leq n - 1 \right\}$$

(Purwaningrum & Agustina, 2024) (3)

Evaluasi Model

Klasifikasi status gizi balita ini diklasifikasikan ke dalam tiga kelas: gizi baik, gizi kurang dan gizi buruk. Sehingga penggunaan *confusion matrix* sangat tepat untuk kasus klasifikasi multiclass, karena mampu menunjukkan seberapa baik model dalam mengklasifikasikan data ke salah satu dari ketiga kelas tersebut, serta memungkinkan pengukuran kinerja model secara menyeluruh (Angkoso & Irmayansyah, 2023). *Confusion matrix* merupakan teknik evaluasi kinerja model klasifikasi, yang dilakukan dengan membandingkan hasil prediksi model terhadap data aktual untuk mengukur tingkat akurasi dan performa klasifikasinya (Rajoub, 2020) (Amalda et al., 2022). **Tabel 1** mengilustrasikan struktur dari *confusion matrix*.

Tabel 1. *Confusion Matrix*

	Aktual Positif	Aktual Negatif
Diprediksi Positif	TP	FP
Diprediksi Negatif	FN	TN

Seperti yang ditunjukkan pada Tabel 1, Nilai *True Positive* (TP) menunjukkan jumlah data positif yang berhasil diklasifikasikan dengan benar. Sebaliknya, nilai *False Negative* (FN) menggambarkan data yang sebenarnya positif, namun salah diklasifikasikan sebagai negatif. Sementara itu, *True Negative* (TN) merupakan jumlah data negatif yang teridentifikasi dengan tepat, dan *False Positive* (FP) adalah data yang sebenarnya negatif tetapi salah diklasifikasikan sebagai positif.

Informasi dari *confusion matrix* dapat dimanfaatkan untuk menghitung metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan skor f-1, yang tercantum di tabel 2. (Kastawan et al., 2018)

Tabel 2. Metrik Ukuran Kinerja Model

Metrik	Deskripsi	Formula
<i>Accuracy</i>	Tingkat prediksi yang benar dari model	$TP + TN / Jumlah Set Data$(3)
<i>Precision</i>	Kemampuan model untuk mengidentifikasi hanya data yang benar-benar termasuk ke dalam kelas target (menghindari <i>false positive</i>).	$TP / (TP + FP)$ (4)
<i>Recall</i>	Kemampuan model untuk menemukan semua data yang termasuk ke dalam suatu kelas target (menghindari <i>false negative</i>).	$TP / (TP + FN)$ (5)
Skor f-1	Menunjukkan seberapa baik tingkat <i>recall</i> dan <i>precision</i>	$2 * (Precision * Recall) / (Precision + Recall)$ (6)

Metrik pada Tabel 2 dapat digunakan sebagai evaluasi kinerja model. Semakin tinggi nilai metrik, semakin baik model yang dibangun.

HASIL DAN PEMBAHASAN

Pemodelan Algoritma C4.5

Total data yang didapatkan dari Dinas Kesehatan Pematangsiantar dari Tahun 2022-2024 adalah sebanyak 697 baris data.

Tabel 3. Data Mentah

No	Nama Balita	Usia	Tinggi Badan	Berat Badan	Jenis Kelamin	Kecamatan	Status Gizi
1	Chritstine T.S	27 bulan	72.2	6.4	Perempuan	Siantar Marimbun	Gizi buruk
2	Egy S.S	2 Tahun - 2 Bulan	0	0	Laki-laki	Siantar Marimbun	Gizi Buruk
3	M .Arkanza A.	4 Tahun - 0 Bulan	80	10.5	Laki-laki	Siantar Timur	Gizi baik
4	Gabriel	3 Tahun - 1 Bulan	0	11	Perempuan	Siantar Marimbun	?
...
697	Nayla Khaliqa D.	1 Tahun - 0 Bulan	67.0	5.5	Perempuan	Siantar Timur	Gizi buruk

Keseluruhan data pada Tabel 3 akan dilakukan *Praprocessing Data* yaitu praktik menyiapkan data mentah untuk agar dapat dipahami dan digunakan untuk analisis. Hasil data yang telah melewati tahapan *praprocessing* kemudian disebut *Dataset*. Proses ini terdiri dari tahap *cleaning data* dan transformasi data.

Tabel 4. *Dataset*

No	Jenis Kelamin	Usia	Berat Badan	Tinggi Badan	Status Gizi
1	P	27	6.4	72.2	Gizi buruk
2	P	39	7.1	76.0	Gizi kurang
3	P	38	7.7	72.2	Gizi baik
4	L	48	10.5	80.0	Gizi baik
...
656	L	40	7.0	73.5	Gizi buruk

Proses menangani nilai yang hilang atau tidak lengkap dalam dataset dikarakterisasikan dalam tahap ini sebagai *cleaning data*. Baris yang memiliki nilai yang hilang akan dihapus dari *dataset*. Proses transformasi yaitu mengubah format data usia "Tahun-Bulan" menjadi usia "Bulan". Atribut "nama balita" dan "kecamatan" tidak digunakan dalam proses pelatihan model karena tidak memberikan kontribusi signifikan terhadap proses klasifikasi status gizi. Kedua atribut ini bersifat identifikasi dan tidak memiliki pengaruh terhadap output klasifikasi, sehingga dihapus pada tahap *praprocessing data*.

Berdasarkan Tabel 4. Terdapat 5 kolom yang akan menjadi atribut. Kelima kolom tersebut adalah: Usia, tinggi badan, jenis kelamin, berat badan, dan status gizi. Selanjutnya, deskripsi atribut-atribut *dataset* disajikan pada Tabel 5.

Tabel 5. Deskripsi Atribut *Dataset*

Atribut	Deskripsi atribut	Tipe Data
JK (X1)	Jenis Kelamin Balita (Laki-Laki (L) atau Perempuan (P))	Kategorikal
Usia (X2)	Usia balita (dalam bulan)	Numerik
anak_tb (X3)	Tinggi badan balita (cm)	Numerik
anak_bb (X4)	Berat badan balita (kg)	Numerik
Status Gizi (Y)	Label kelas yang menunjukkan status gizi balita tersebut	Kategorikal

Seperi yang ditunjukkan pada Tabel 5, atribut *input* yang digunakan berkontribusi terhadap klasifikasi status gizi, yaitu Jenis Kelamin (X1), Usia (X2), Tinggi Badan (X3), Berat Badan (X4).

Selanjutnya *Dataset* dipisah secara acak menjadi data *training* dan data *testing* dengan rasio 70:30, *Dataset* dengan rasio 70:30 menghasilkan 459 baris data *training*. Dengan menggunakan **Persamaan 3**, menghasilkan ambang batas untuk setiap Atribut numerik :

$$T_{Usia} = \{2.5, \dots, 58\}, T_{Berat-Badan} = \{3.3, \dots, 17.5\}, T_{Tinggi-Badan} = \{51, \dots, 108.95\}.$$

Perhitungan *Entropy* dan *Information Gain* untuk memilih simpul akar disajikan pada **Tabel 6**.

Tabel 6. Pemilihan Simpul Akar

Atribut	Total Sample	Gizi	Gizi	Gizi	Entropy	<i>Information Gain</i>
		Baik	Kurang	Buruk		
JK	459	188	123	148	1,56307	
	L	292	121	75	1,55797438	0,0008053
Usia	P	167	67	48	1,56975794	
	≤ 2.5	2	1	0	1	0,0020111
anak_bb	> 2.5	457	187	123	1,56351122	

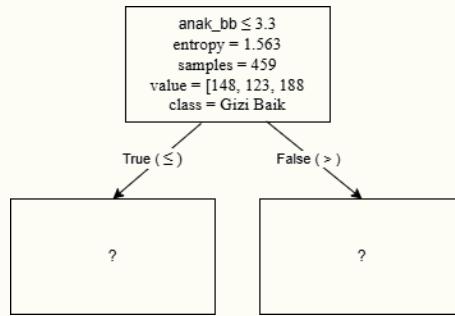
anak_tb	≤ 25	394	157	102	1,563135708	0,00881
	> 25	65	31	21	1,500453	
anak_bb
	≤ 58.5	458	187	123	1,563663076	0,002810528
anak_bb	> 58.5	1	1	0	0	
	≤ 3.3	45	1	0	0,15374218033	0,160065363
anak_bb	> 3.3	414	187	123	1,538790632	

anak_tb	≤ 8.2	343	111	94	1,56699804	0,08950957
	> 8.2	116	77	29	1,197263	
anak_tb
	≤ 17.5	458	187	123	1,563663076	0,002810528
anak_tb	> 17.5	1	1	0	0	
	≤ 51	1	1	0	0	0,002810528
anak_tb	> 51	458	187	123	1,563663076	

anak_tb	≤ 70	168	102	37	1,355302	0,0735493014
	> 70	291	86	86	1,5670030005	
anak_tb
	$\leq 108,95$	458	187	123	1,5636631	0,00281053
anak_tb	$> 108,95$	1	1	0	0	

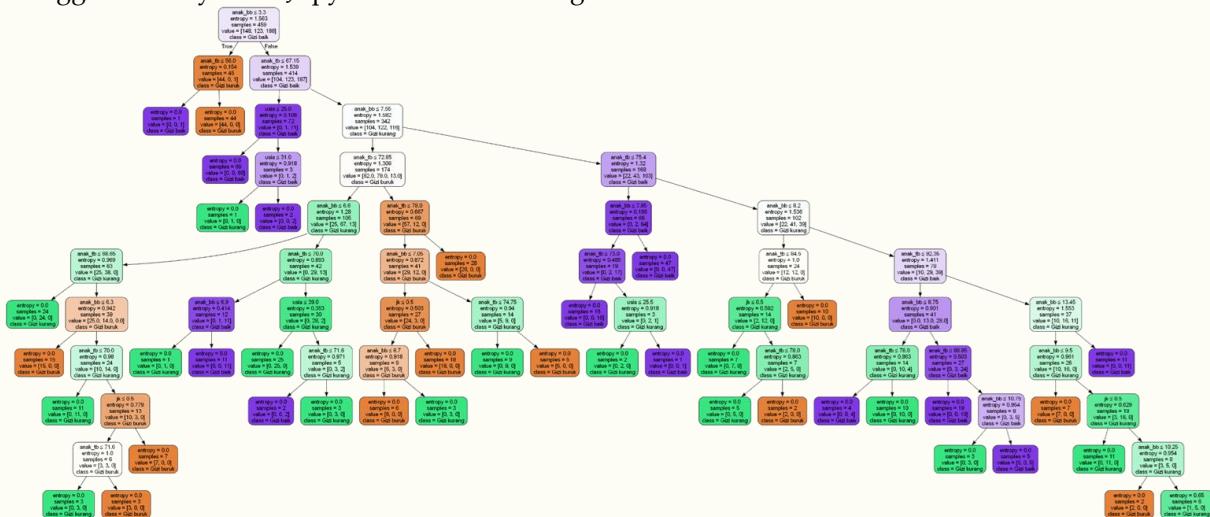
Seperti yang disajikan pada **Tabel 6**, proses awal membangun pohon keputusan dimulai dengan perhitungan nilai *entropy* dan *information gain* untuk masing-masing atribut. *Entropy* mengukur ketidakmurnian data, sedangkan *information gain* menunjukkan seberapa besar suatu atribut dapat mengurangi ketidakpastian. Untuk atribut kategorikal seperti jenis kelamin, entropy dihitung berdasarkan kategori (L/P). Sementara pada atribut numerik seperti usia, berat badan (anak_bb), dan tinggi badan (anak_tb), entropy dihitung dari subset yang terbentuk berdasarkan nilai ambang tertentu.

Dari hasil perhitungan, nilai *information gain* tertinggi diperoleh pada atribut Berat Badan (anak_bb) dengan ambang batas $\leq 3,3$ sehingga atribut ini dipilih sebagai simpul akar (*root node*). Simpul akar merupakan titik awal pohon keputusan dan menjadi dasar percabangan berikutnya. Pemilihan atribut ini penting karena simpul akar yang tepat dapat secara signifikan meningkatkan efisiensi dan akurasi proses klasifikasi. Oleh karena itu pohon keputusan yang terbentuk adalah seperti Gambar 2 berikut :



Gambar 2. Bentuk Sementara Pohon Keputusan

Pada Gambar 2, terlihat bahwa *node* akar "anak_bb ≤ 3.3 " berisi 469 data pelatihan dengan distribusi kelas Gizi buruk sejumlah 148 kasus, Gizi kurang sejumlah 123 kasus dan Gizi baik sejumlah 188 kasus. Karena kelas mayoritas pada *node* ini adalah Gizi baik, maka *node* ini diberi label "Gizi baik". Kalkulasi dari ketiga kelas tersebut menghasilkan *entropy* yang tinggi yaitu 1.563, mengindikasikan bahwa *node* belum mencapai kemurnian. Oleh karena itu proses pemilihan atribut dilakukan secara iteratif hingga mencapai *node* daun dengan *entropy* 0. Hasil pohon keputusan dari model yang dibangun menggunakan Python "Jupyter Notebook" sebagai berikut :



Gambar 3. Hasil Pohon Keputusan

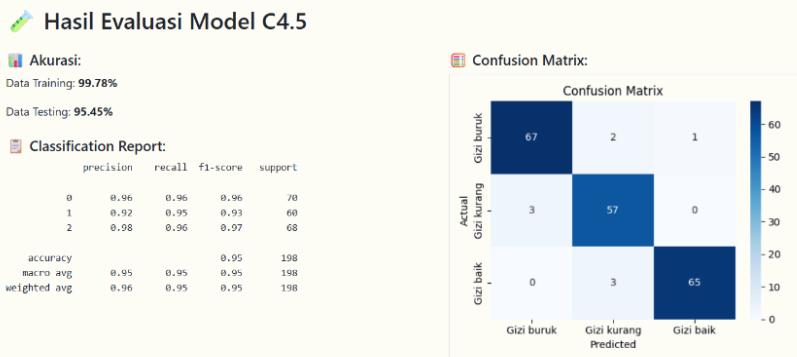
Setelah melewati tahap perhitungan keseluruhan simpul (*node*), maka akan terbentuk sebuah pohon keputusan yang terlihat pada gambar 3. Penelusuran pohon keputusan dimulai dari simpul akar dengan kondisi berat badan (*anak_bb*) $\leq 3,3$. Jika kondisi tersebut terpenuhi, maka proses klasifikasi mengikuti cabang kiri (*True*), sedangkan jika tidak, proses dilanjutkan ke cabang kanan (*False*). Setiap simpul (*node*) dalam pohon mencerminkan keputusan yang dibuat berdasarkan atribut dan nilai ambang tertentu. Proses ini berlangsung secara bertahap hingga mencapai simpul akhir (*leaf node*), yang berisi hasil klasifikasi akhir. Pola-pola keputusan yang terbentuk pada pohon mencerminkan hubungan antara atribut yang digunakan untuk mengklasifikasikan data balita ke dalam salah satu dari tiga kelas status gizi, yaitu gizi baik, kurang, dan buruk. Struktur pohon keputusan ini tidak hanya menggambarkan alur pengambilan keputusan, tetapi juga menyediakan interpretasi yang jelas bagi pengguna non-teknis, seperti tenaga kesehatan, dalam memahami logika prediksi yang dilakukan oleh sistem.

Pohon keputusan menghasilkan 80 *nodes*. Dimana setiap *node* memiliki warna yang mengindikasikan kelas target untuk mempermudah interpretasi, seperti Hijau mewakili kelas Gizi baik, Ungu mewakili kelas Gizi buruk dan Jingga mewakili kelas Gizi kurang dan Putih mempunyai makna bahwa *node* tersebut terdapat campuran kelas atau *node* belum murni. Semakin pekat warna *node*, semakin tinggi kemurniannya (*entropy* mendekati 0), artinya data dalam *node* tersebut dominan berasal

dari satu kelas. Sebaliknya, warna yang lebih terang menunjukkan bahwa *node* tersebut masih terdiri dari campuran beberapa kelas.

Evaluasi Model C4.5

Evaluasi performa model C4.5 dilakukan untuk mengukur sejauh mana model mampu mengklasifikasikan status gizi secara akurat berdasarkan data yang tersedia. Pengujian dilakukan menggunakan data latih dan data uji, serta dinilai melalui sejumlah metrik evaluasi umum seperti *confusion matrix*, dan *classification report* (*accuracy*, *precision*, *recall*, serta *f1-score*).



Gambar 4. Hasil Evaluasi Model C4.5

Berdasarkan evaluasi dengan teknik *confusion matrix multiclass classification* yang ditampilkan pada Gambar 4. Evaluasi dilakukan terhadap 198 data uji, yang terbagi ke dalam tiga kelas : Gizi buruk (label 0), Gizi kurang (label 1) dan Gizi baik (label 2). Nilai *True Positive* (TP), *False Negative* (FN), *False Positive* (FP), *True Negative* (TN) dihitung untuk masing masing kelas. Sebagai contoh, untuk kelas Gizi Buruk, model menghasilkan nilai **TP** sebesar 67, yang berarti terdapat 67 data yang benar-benar termasuk dalam kelas tersebut dan berhasil diprediksi dengan tepat. Sementara itu, nilai **FN** sebesar 3 menunjukkan bahwa 3 data dari kelas Gizi Buruk salah diklasifikasikan sebagai kelas lain, yaitu dua data diprediksi sebagai Gizi Kurang dan satu data sebagai Gizi Baik. Nilai **FP** sebesar 3 menunjukkan bahwa terdapat tiga data dari kelas lain yang salah diprediksi sebagai Gizi Buruk. Adapun nilai **TN** sebesar 125, yang menunjukkan bahwa 125 data dari kelas non-Gizi Buruk berhasil diklasifikasikan dengan benar sebagai bukan Gizi Buruk. Nilai *False Negative* (FN) dan *False Positive* (FP) relatif rendah, menunjukkan kesalahan klasifikasi yang minim.

Model C4.5 menunjukkan kinerja yang sangat baik dengan ukuran *dataset* cukup besar (459 data latih, 198 data uji), sehingga hasilnya relatif stabil. Model memiliki tingkat akurasi 99% pada data latih, sehingga mengindikasikan kemampuan model mempelajari pola data secara optimal dan tingkat akurasi sebesar 95% pada data uji, yang mencerminkan bahwa model mengklasifikasikan data yang belum pernah dilihat dengan tingkat kesalahan yang sangat rendah. Model tidak mengalami *overfitting* yang signifikan terlihat dari selisih kecil antara akurasi data latih dan data uji yang menandakan generalisasi yang baik juga mempertahankan performanya dengan stabil.

Hasil evaluasi lebih lanjut ditunjukkan melalui nilai *precision*, *recall*, dan *f1-score* dan nilai *macro average* dan nilai *weighted average* untuk ketiga metrik tersebut secara konsisten berada pada angka 0,95 atau lebih tinggi, yang menunjukkan bahwa model tidak bias terhadap kelas tertentu, dan performa klasifikasinya cukup seimbang.

Deployment

Setelah dilakukan membangun model klasifikasi *C4.5* dan melakukan tahap evaluasi menggunakan data uji. Tahap selanjutnya adalah mengimplementasikan kedalam sistem berbasis web yang dibangun menggunakan *framework Flask*. Berikut adalah halaman antarmuka yang dapat dilihat di bawah ini.



Gambar 5. Halaman Login

Saat sistem diakses, halaman pertama yang ditampilkan adalah halaman login. Di halaman ini, pengguna diminta untuk memasukkan Username dan Password yang telah didaftarkan sebelumnya. Sistem ini mempunyai dua (2) *role* akses, yaitu *role* user dan *role* admin, dimana admin mempunyai hak akses semua halaman. Dan user tidak memiliki akses pada halaman perbarui dataset dan juga halaman prevalensi gizi.

Gambar 6. Halaman Data Balita

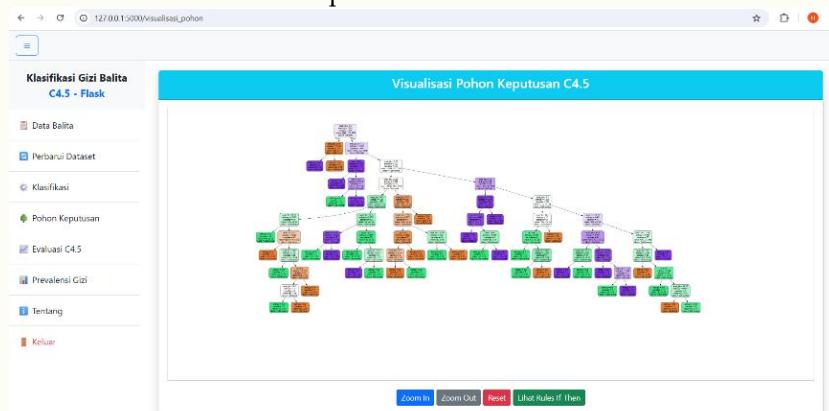
Setelah login berhasil, pengguna akan diarahkan ke halaman data balita sebagaimana ditampilkan pada Gambar 6. Di halaman ini, pengguna dapat melihat seluruh data balita yang telah dimasukkan sebelumnya, serta terdapat tombol aksi "hapus" di setiap baris yang berfungsi untuk menghapus data pada baris tersebut.

Gambar 7. Halaman Klasifikasi Gizi Balita

Gambar 7. menampilkan halaman Klasifikasi Gizi Balita. Pengguna memasukkan data balita yaitu nama, jenis kelamin, usia dalam bulan, berat badan dan tinggi badan anak dan asal kecamatan. Setelah data dimasukkan sistem akan mengambil dan memproses atribut jenis kelamin, usia, berat badan dan tinggi badan lalu sistem akan menampilkan hasil klasifikasi status gizi nya. Data yang dimasukkan akan tersimpan kedalam tabel data balita.

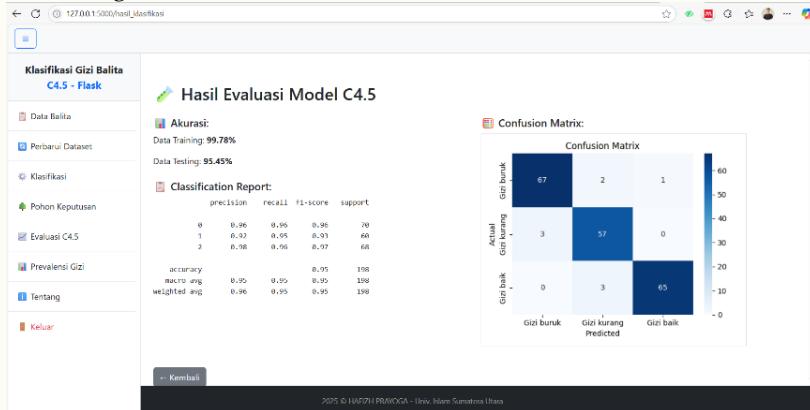
Gambar 8. Halaman Perbarui Dataset

Gambar 8 menampilkan halaman perbarui dataset. Pengguna dapat melakukan pelatihan ulang model C4.5 dengan memilih data-data yang tampil didalam tabel halaman ini lalu klik tombol perbarui dataset. Pada halaman ini terdapat kondisi dimana hanya data balita yang belum pernah dimasukkan kedalam database tabel "dataset" akan tampil di halaman ini.



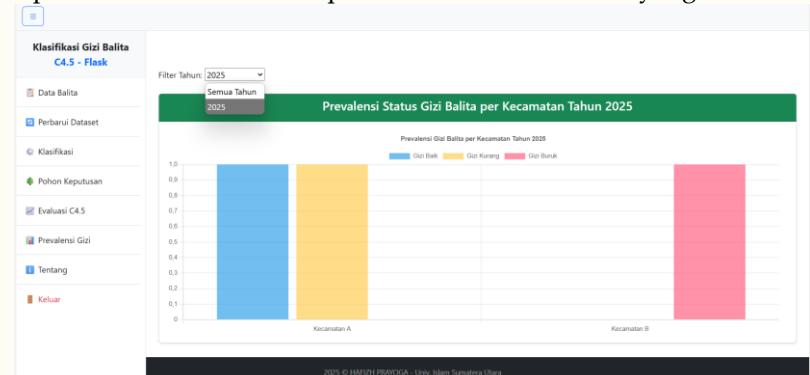
Gambar 9. Halaman Pohon Keputusan

Pengguna dapat melihat pohon keputusan dari model C4.5 yang telah dibangun seperti yang terlihat Gambar 9. Pada halaman ini terdapat beberapa tombol interaksi untuk zoom in, zoom out gambar dan reset ukuran gambar pohon keputusan. Tombol lihat Rules If-then sistem akan mengubah pohon keputusan menjadi format rule/aturan jika-maka. Sebagai contoh, Jika bb anak ≤ 3.3 dan tb anak \leq maka status gizi adalah gizi baik.



Gambar 10. Halaman Evaluasi Model

Gambar 10 menampilkan halaman evaluasi performa dari model C4.5 yang telah dibangun.



Gambar 11. Halaman Prevalensi Status Gizi Balita

Pada halaman ini, pengguna dapat melihat distribusi data status gizi balita berdasarkan kecamatan melalui diagram batang seperti yang ditampilkan pada Gambar 11. Selain itu, pengguna juga dapat memfilter sebaran data berdasarkan tahun tertentu atau memilih untuk menampilkan gabungan data dari semua tahun.

KESIMPULAN

Hasil implementasi algoritma C4.5 yang membentuk pohon keputusan memberikan wawasan mengenai faktor-faktor yang memengaruhi status gizi balita. Dalam hal ini, atribut yang paling berpengaruh dalam proses klasifikasi adalah berat badan dan tinggi badan. Penelitian ini berhasil merancang dan mengimplementasikan sistem klasifikasi status gizi balita berbasis web menggunakan algoritma C4.5 dan *framework Flask*, dengan menggunakan dataset dari Dinas Kesehatan Kota Pematangsiantar. Model menunjukkan performa memuaskan, dengan akurasi mencapai 99% pada data latih dan 95% pada data uji. Di tambah lagi, skor precision, recall, dan *f1-score* secara konsisten tinggi ($\geq 0,95$), menunjukkan bahwa model tidak bias antar kelas dan andal dalam mendeteksi tiga kelas status gizi (baik, kurang, buruk). Implementasi sistem berbasis web mempermudah petugas kesehatan dalam memasukkan data kemudian melakukan klasifikasi otomatis menggunakan model C4.5, melihat aturan pohon keputusan, dan meninjau visualisasi data prevalensi stunting per kecamatan. Fitur otentifikasi berbasis peran (admin dan user) memastikan keamanan dan kemudahan akses sesuai kebutuhan

pengguna. Sehingga sistem ini cukup andal untuk digunakan dalam mendukung keputusan intervensi gizi oleh tenaga kesehatan. Dengan klasifikasi yang cepat dan akurat, langkah penanganan balita berisiko stunting dapat dilakukan lebih tepat waktu dan terarah. Meskipun model menunjukkan performa yang sangat baik, masih terdapat ruang pengembangan, yaitu menambahkan variabel lain seperti lingkar kepala, pola makan anak, riwayat penyakit, status imunisasi, riwayat kesehatan Ibu, serta pemberian ASI yang mungkin relevan dan berpengaruh terhadap status gizi balita. Penelitian lanjutan juga dapat membandingkan algoritma C4.5 dengan metode klasifikasi modern seperti *Gradient Boosting*, atau *Neural Network* untuk mengevaluasi apakah ada peningkatan performa yang signifikan dalam konteks data gizi balita. Penelitian selanjutnya juga dapat mengkaji sejauh mana sistem klasifikasi ini dapat mempengaruhi proses pengambilan keputusan di sektor kesehatan, khususnya dalam hal penyaluran bantuan gizi, penentuan prioritas intervensi, serta pemantauan program penanggulangan stunting.

DAFTAR PUSTAKA

- Adinda, P. R. (2022). Pohon Keputusan C4. 5 Algoritma Untuk Klasifikasi Program Bantuan Belajar. *Jurnal Portal Data*, 2(9), 1–13.
- Alshammari, T. S. (2024). Applying Machine Learning Algorithms for the Classification of Sleep Disorders. *IEEE Access*, 12, 36110–36121. <https://doi.org/10.1109/ACCESS.2024.3374408>
- Amalda, R. N., Millah, N., & Fitria, I. (2022). Implementasi Algoritma C5.0 Dalam Menganalisa Kelayakan Penerima Keringanan UKT Mahasiswa ITK. *Teorema: Teori Dan Riset Matematika*, 7(1), 101. <https://doi.org/10.25157/teorema.v7i1.6692>
- Angkoso, B., & Irmayansyah, I. (2023). Penerapan Algoritma C4.5 Untuk Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan (SPP) Santri. *TeknoIS : Jurnal Ilmiah Teknologi Informasi Dan Sains*, 13(1), 13–23. <https://doi.org/10.36350/jbs.v13i1.166>
- Anthira, N., & Suendri. (2024). Penerapan Data Mining Pada Klasifikasi Gangguan Jiwa Menggunakan Algoritma C5.0 Di RSJ. Mahoni Kota Medan. *Jurnal Teknika*, 18(x), 571–582.
- Bachri, O. S., & Bhakti, R. M. H. (2021). Penentuan Status Stunting pada Anak dengan Menggunakan Algoritma KNN. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, 3(02), 130–137. <https://doi.org/10.46772/intech.v3i02.533>
- Daud, M., Juita, R., Suhendra, & Dwi, C. (2025). *Penerapan Metode Algoritma C4 . 5 Untuk Klasifikasi Kelayakan Penerima Program Bantuan Pada Dinas Sosial Kabupaten Manokwari*. 5(1), 271–278.
- Eirlangga, Y. S., & Syaputra, A. E. (2022). Klasifikasi Penjurusan pada Sekolah Menengah Atas (SMA) dengan Metode Algoritma C4.5. *Jurnal Informasi Dan Teknologi*, 4(3), 160–165. <https://doi.org/10.37034/jidt.v4i3.235>
- El-Hasnony, I. M., Elzeki, O. M., Alshehri, A., & Salem, H. (2022). Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction. *Sensors*, 22(3), 1184. <https://doi.org/10.3390/s22031184>
- F Amalia, M. P. (2023). Klasifikasi Peserta Didik Menggunakan Algoritma C4.5. *Jurnal Citra Widya Edukasi*, Vo. 15 No.
- Hastono, A. F. D., Vitianingsih, A. V., Pamudi, P., Maukar, A. L., & Wati, S. F. A. (2025). Diabetes Mellitus Disease Prediction Using Logistic Regression (LR) and Support Vector Machine (SVM) Methods. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(1), 54-64. <https://doi.org/10.51454/decode.v5i1.1039>
- Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques*. Morgan Kaufmann. <https://books.google.co.id/books?id=NR1oEAAAQBAJ>

- Irmayanti. (2023). Perancangan Sistem Informasi Penyewaan Thermoking Pada PT . Moderen Prima Transportasi Menggunakan Python Dengan Framework Flask. *JuSTICe*, 1(1), 24–34.
- Jang, J.-H., & Nemoto, M. (2024). A Study of Factors Influencing Happiness in Korea: Topic Modelling and Neural Network Analysis. *Data and Metadata*, 3, 238. <https://doi.org/10.56294/dm2024238>
- Kastawan, P. W., Wiharta, D. M., & Sudarma, M. (2018). Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil. *Majalah Ilmiah Teknologi Elektro*, 17(3), 371. <https://doi.org/10.24843/MITE.2018.v17i03.P11>
- Lestari, S., & Dina, B. (2023). Klasifikasi Ketepatan Kelulusan Siswa Pada Smk Yadika 9 Bintara Jaya Kota Bekasi Menggunakan Algoritma C4.5. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 4(3), 1618–1632. <https://doi.org/10.35870/jimik.v4i3.389>
- Limabri, R. I. O., Putrawansyah, F., & Arif, A. (2023). Penerapan Data Mining Untuk Mengklasifikasi Nasabah Bank Sumsel Babel Menggunakan Algoritma C4. 5. *Escaf*, 1101–1108.
- Madiko, S., Ilham, R., & Mojdo, D. (2023). Hubungan Status Gizi Balita Dengan Kejadian Stunting Di Wilayah Kerja Puskesmas Kota Timur. *Jurnal Ilmu Kedokteran Dan Kesehatan Indonesia*, 3, 155–164. <https://doi.org/10.55606/jikki.v3i1.1201>
- Munawaroh, H., Nada, N. K., Hasjiandito, A., Faisal, V. I. A., Heldanita, H., Anjarsari, I., & Fauziddin, M. (2022). Peranan Orang Tua Dalam Pemenuhan Gizi Seimbang Sebagai Upaya Pencegahan Stunting Pada Anak Usia 4-5 Tahun. *Sentra Cendekia*, 3(2), 47. <https://doi.org/10.31331/sencenivet.v3i2.2149>
- Nurrajiati, D., Irawan, B., Bahtiar, A., & Tohidi, E. (2024). Implementasi Algoritma C4.5 Pada Produksi Perikanan Di Kecamatan Ciaurbeuti. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 238–244. <https://doi.org/10.36040/jati.v8i1.8208>
- Ozcan, M., & Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, 3, 100130. <https://doi.org/10.1016/j.health.2022.100130>
- Pratama, D. A., Mutaqin, I. R., & Manuela, K. R. (2023). Analisis Terjadinya Kanker Paru-Paru Pada Pasien Menggunakan Decision Tree: Penerapan Algoritma C4.5 Dan RapidMiner Untuk Menentukan Risiko Kanker Pada Gejala Pasien. *Jtmei*, 2, 156–170.
- Purwaningrum, D. A. S. K., & Agustina, D. (2024). Implementation Of Machine Learning Algorithm C4.5 In Classification Of Patients With Type 2 Diabetes Mellitus. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 18(1), 0193–0204. <https://doi.org/10.30598/barekengvol18iss1pp0193-0204>
- Rajoub, B. (2020). Supervised and unsupervised learning. In *Biomedical Signal Processing and Artificial Intelligence in Healthcare* (pp. 51–89). Elsevier. <https://doi.org/10.1016/B978-0-12-818946-7.00003-2>
- Rastogi, R., & Bansal, M. (2023). Diabetes prediction model using data mining techniques. *Measurement: Sensors*, 25, 100605. <https://doi.org/10.1016/j.measen.2022.100605>
- Rofiani, R., Oktaviani, L., Vernanda, D., & Hendriawan, T. (2024). Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5. *Jurnal Tekno Kompak*, 18(1), 126. <https://doi.org/10.33365/jtk.v18i1.3525>
- Rozaq, A., & Purnomo, A. J. (2022). Classification Of Stunting Status In Toddlers Using Naive Bayes Method In The City Of Madiun Based On Website. *Jurnal Techno Nusa Mandiri*, 19(2), 69–76. <https://doi.org/10.33480/techno.v19i2.3337>
- Sinaga, E. S., Rasyid, I. A., Mubarok, M. R., Sudharma, N. I., & Nolia, H. (2023). Pemantauan Konsumsi Pemberian Makanan Tambahan (PMT) Dalam Meningkatkan Berat Badan Balita Dengan Masalah Gizi. *ABDI MOESTOPO: Jurnal Pengabdian Pada Masyarakat*, 6(1), 1–8.

<https://doi.org/10.32509/abdimoestopo.v5i2.2236>

Suendri, S. (2020). *Analisis Tingkat Kepuasan Mahasiswa Terhadap Layanan Administrasi Fakultas Sains Dan Teknologi UIN Sumatera Utara Selama Masa Pandemi Covid-19*. http://repository.uinsu.ac.id/12304/1/Suendri_Penelitian_2020.pdf

Yuningsih, Y. (2022). Hubungan Status Gizi dengan Stunting pada Balita. *Oksitosin: Jurnal Ilmiah Kebidanan*, 9(2), 102–109. <https://doi.org/10.35316/oksitosin.v9i2.1845>