

Penerapan Data Mining Untuk Analisis Kesehatan Balita Menggunakan Metode *K-Means* dan *Decision Tree* (Studi Kasus Puskesmas Paal Lima)

Rike Limia Budiarti¹, Sri Mulyati², Ahmad Zaidan³

¹Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Nurdin Hamzah, Jambi, Indonesia

²³Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Nurdin Hamzah, Jambi, Indonesia

Email: ¹rikelimiabudiarti@unh.ac.id, ²srimulyati@unh.ac.id, ³ahmadzaidan140303@gmail.com

Article Information

Article history

Received 08 September 2025

Revised 08 November 2025

Accepted 11 November 2025

Available 30 November 2025

Keywords

Data Mining

Health Toddlers

Nutritional Status

K-Means

Decision Tree

Corresponding Author:

Rike Limia Budiarti,
Program Studi Sistem Informasi,
Fakultas Ilmu Komputer,
Universitas Nurdin Hamzah, Jam
rikelimiabudiarti@unh.ac.id

Abstract

Advances in information technology provide opportunities in health data analysis, particularly in the nutritional status of toddlers. This study applies data mining methods using the K-Means and Decision Tree algorithms to quickly and accurately group and classify the nutritional status of toddlers. K-Means is used to group data based on age, weight, and height into categories of good, poor, and very poor nutrition. The clustering results are then fed into the Decision Tree, which generates data-based nutritional intervention recommendations. This study used a quantitative approach with Orange Tools and a dataset of toddlers from the Paal Lima Community Health Center in Jambi City. The test results showed a classification accuracy of 92.3% and a 30% increase in nutritional analysis efficiency compared to the manual method. Thus, the application of data mining has proven to be effective in supporting community health centers in decision-making and improving toddler health services.

Keywords: Data Mining, Health, Toddlers, Nutritional Status, K-Means, Decision Tree

Abstrak

Perkembangan teknologi informasi memberikan peluang dalam analisis data kesehatan, khususnya status gizi balita. Penelitian ini menerapkan metode *data mining* menggunakan algoritma *K-Means* dan *Decision Tree* untuk mengelompokkan serta mengklasifikasikan status gizi balita secara cepat dan akurat. *K-Means* digunakan untuk mengelompokkan data berdasarkan usia, berat badan, dan tinggi badan ke dalam kategori gizi baik, kurang, dan sangat kurang. Hasil *clustering* kemudian menjadi masukan bagi *Decision Tree* yang menghasilkan rekomendasi intervensi gizi berbasis data. Penelitian ini menggunakan pendekatan kuantitatif dengan *Tools Orange* dan dataset balita dari Puskesmas Paal Lima Kota Jambi. Hasil pengujian menunjukkan akurasi klasifikasi sebesar 92,3% dan peningkatan efisiensi analisis gizi sebesar 30% dibandingkan metode manual. Dengan demikian, penerapan *data mining* terbukti efektif mendukung puskesmas dalam pengambilan keputusan dan peningkatan layanan kesehatan balita.

Kata Kunci: Data Mining, Kesehatan, Balita, Status Gizi, K-Means, *Decision Tree*

Copyright©2025 Rike Limia Budiarti, Sri Mulyati, and Ahmad Zaidan

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



1. Pendahuluan

Perkembangan teknologi informasi saat ini telah membawa dampak signifikan dalam berbagai sektor, termasuk bidang kesehatan. Salah satu pemanfaatan yang berkembang pesat adalah penggunaan data mining untuk menganalisis data kesehatan secara lebih efisien dan mendalam. Di fasilitas layanan kesehatan tingkat pertama seperti puskesmas, pencatatan data balita umumnya dilakukan melalui sistem Rekam Medis Elektronik (RME), namun sistem ini masih bersifat administratif tanpa dukungan analitik yang mampu memberikan informasi diagnostik atau rekomendasi intervensi secara langsung.

Dalam konteks kesehatan balita, analisis status gizi secara cepat dan akurat sangat penting untuk mendeteksi masalah kesehatan sejak dini dan mencegah kondisi gizi sangat kurang. Permasalahan yang dihadapi saat ini adalah minimnya sistem yang mampu mengelola dan menganalisis data kesehatan balita secara otomatis dan akurat. Ketergantungan pada analisis manual menyebabkan keterlambatan dalam pengambilan keputusan, terlebih dengan terus meningkatnya jumlah data yang harus diolah setiap tahunnya.

Dalam penelitian ini, sampel data diperoleh dari beberapa sumber, seperti Puskesmas dan Posyandu, yang berperan sebagai acuan dalam proses pengelompokan status gizi balita, yaitu Gizi Baik, Gizi Kurang, dan Gizi Sangat Kurang. Metode K-Means digunakan karena bersifat *unsupervised learning* yang mampu mengelompokkan data balita berdasarkan kesamaan karakteristik seperti usia, berat badan, dan tinggi badan tanpa memerlukan label kelas terlebih dahulu. Hasil pengelompokan ini kemudian menjadi dasar untuk penerapan metode *Decision Tree*, yang bersifat *supervised learning*, guna melakukan klasifikasi status gizi dan memberikan rekomendasi solusi intervensi yang sesuai bagi setiap kategori gizi yang dihasilkan oleh proses clustering. Kombinasi kedua metode ini dipilih karena K-Means dapat menemukan pola tersembunyi dalam data secara otomatis, sementara *Decision Tree* mampu menjelaskan pola tersebut secara logis dan mudah dipahami dalam bentuk aturan keputusan. Dengan demikian, sistem yang dihasilkan dapat mengidentifikasi kondisi gizi balita secara lebih terstruktur serta menjadi dasar dalam penentuan langkah intervensi gizi yang tepat untuk masing-masing kategori status gizi tersebut.

Berdasarkan beberapa uraian diatas maka peneliti tertarik untuk melakukan penelitian yang berjudul **“PENERAPAN DATA MINING UNTUK ANALISIS KESEHATAN BALITA MENGGUNAKAN METODE K-MEANS DAN DECISION TREE”**.

1.1 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan sebelumnya, maka dapat dirumuskan permasalahan dalam penelitian ini yaitu : “ Bagaimana penerapan data mining untuk analisis kesehatan balita menggunakan metode *K-Means* dan menggunakan metode *Decision Tree*?”.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah mengimplementasikan data mining menggunakan metode *K-Means* dan *Decision Tree* untuk membangun serta memberikan kontribusi dalam analisis dan klasifikasi status gizi balita, serta mengembangkan sistem analisis

yang ditujukan untuk memberikan rekomendasi intervensi gizi yang tepat. Sistem ini bertujuan untuk meningkatkan efisiensi pemantauan status gizi balita guna pencegahan dini serta membantu tenaga kesehatan dalam menganalisis, mengelompokkan, dan menentukan solusi terhadap data gizi balita secara lebih efektif.

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Memaksimalkan data RME untuk dikelola guna mengurangi risiko kesalahan analisis yang disebabkan oleh subjektivitas atau ketidakkonsistenan dalam pengolahan data.
2. Menyediakan sistem berbasis data mining yang dapat digunakan sebagai alat bantu dalam pengambilan keputusan terkait intervensi gizi balita.
3. Meningkatkan efektivitas pemantauan kesehatan gizi balita di bawah 5 tahun dengan sistem yang lebih terstruktur dan berbasis data.

2. Kajian Terdahulu

2.1 Penelitian Terdahulu

Berdasarkan penelitian sebelumnya, klasifikasi status gizi balita menjadi salah satu fokus utama dalam bidang kesehatan masyarakat, terutama dalam mendukung pemantauan dan pencegahan masalah gizi sejak dini. Salah satu penelitian yang dilakukan menunjukkan bahwa metode *Polynomial Regression* dapat digunakan untuk memprediksi status gizi anak balita berdasarkan variabel usia, tinggi badan (TB), dan berat badan (BB). Hasil prediksi menunjukkan bahwa jumlah anak dengan status gizi tertentu pada tahun 2019 mencapai 290.806, pada 2020 sebesar 68.176, dan pada 2021 sebesar 122.239. Meskipun metode ini dapat memodelkan hubungan antara variabel *Input* dan output, prediksi yang dihasilkan sangat bergantung pada kualitas data yang digunakan. Oleh karena itu, penelitian lebih lanjut diperlukan untuk meningkatkan validasi model agar dapat digunakan secara lebih akurat dalam pemantauan status gizi anak balita [1].

Sementara itu, penelitian lain menggunakan metode *Support Vector Machine (SVM)* dengan kernel *Radial Basis Function (RBF)* untuk mengklasifikasikan status gizi bayi berdasarkan berat badan, tinggi badan, serta indeks status gizi (BB/U, TB/U, BB/TB). Data yang digunakan dalam penelitian ini berasal dari Posyandu Kecamatan Bangun Purba dengan jumlah 1732 data bayi setelah proses pembersihan. Model ini mampu mengklasifikasikan bayi ke dalam 6 kategori status gizi, yaitu Gizi Baik (284 bayi), Gizi Buruk (15 bayi), Gizi Kurang (23 bayi), Gizi Lebih (8 bayi), Obesitas (6 bayi), dan Risiko Gizi Lebih (11 bayi). Dengan pembagian 80% data untuk training dan 20% untuk testing, model menghasilkan akurasi sebesar 87.6%, dengan nilai F1-score 0.865, precision 0.871, dan recall 0.876. Dari 347 data uji, model berhasil mengklasifikasikan 304 data dengan benar, sedangkan 43 data lainnya diklasifikasikan kurang tepat [2].

2.2 Balita

Masa balita adalah periode yang penting dalam tumbuh kembang anak, masa balita dapat menentukan tahap perkembangan anak di masa datang [3]. Balita adalah salah satu kelompok yang sensitif terhadap masalah kesehatan, terutama masalah gizi kurang atau buruk.

Kesehatan seorang balita sangat berhubungan dengan gizi yang terserap di dalam tubuh. Anak balita kerap menderita akibat kekurangan gizi sebab masa balita adalah masa pertumbuhan sehingga memerlukan gizi yang baik [4].

2.3 Gizi

Gizi merupakan faktor terpenting untuk tercapainya tumbuh dan berkembangnya balita secara optimal. Gizi yang tepat sangat mendukung dalam fase periode emas pertumbuhan [5]. Gizi anak balita merujuk pada asupan nutrisi yang diperlukan oleh anak usia 1 hingga 5 tahun untuk mendukung pertumbuhan, perkembangan, dan kesehatan mereka. Pada usia ini, anak mengalami fase pertumbuhan yang sangat pesat, sehingga kebutuhan gizi mereka menjadi sangat penting. Gizi adalah ilmu yang mempelajari segala sesuatu tentang makanan dan hubungannya dengan kesehatan optimal [6].

2.4 Status Gizi

Status gizi merupakan bentuk dari penilaian terhadap tubuh yang terjadi dan disebabkan karena keseimbangan antara pemasukkan zat gizi dengan kebutuhan tubuh. Keseimbangan itu dapat di lihat berdasarkan model pertumbuhan, seperti panjang tungkai, tinggi badan atau panjang badan, berat badan, lingkaran lengan, dan lingkaran kepala. Jika keseimbangan tersebut berubah menjadi kurang baik, yaitu situasi di mana berat badan lebih rendah daripada berat sesuai dengan normalnya menurut usianya disebut gizi kurang [7].

2.5 Data Mining

Data mining adalah sebuah metode penambangan data yang menemukan pengetahuan dari data yang informasinya masih tersembunyi. Jumlah data yang disimpan oleh suatu organisasi bertambah dengan pesat. Jumlah data meningkat dari tahun ke tahun dan mungkin ada pembayaran dalam mengungkapkan informasi tersembunyi di balik data yang ada [8]. Data Mining adalah metode untuk menganalisis pola dan karakteristik di masa depan serta untuk mengumpulkan informasi tak terduga yang belum pernah terlihat sebelumnya dari database yang besar [9].

2.6 Algoritma K-Means

Algoritma *K-Means* merupakan algoritma klusterisasi yang mengelompokkan data berdasarkan titik pusat kluster (*centroid*) terdekat dengan data. Tujuan *K-Means* adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu kluster dan meminimalkan kemiripan data antara kluster. ukuran kemiripan yang digunakan dalam kluster adalah fungsi jarak. sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data titik centroid [10]. *K-Means Clustering* adalah salah satu teknik clustering yang paling umum digunakan dalam analisis data. Algoritma ini bekerja dengan cara membagi data ke dalam K kelompok yang telah ditentukan sebelumnya, di mana K merupakan jumlah kelompok yang diinginkan [11].

Tahapan *Clustering* perhitungan pada *K-Means* :

1. Menentukan (K) (nilainya bebas) Sebagai jumlah cluster yang ingin dibentuk.

2. Membangkitkan nilai random untuk pusat cluster awal (*centroid*) sebanyak (K) atau menggunakan rumus *MAX*, *Average* dan *MIN*.
3. Menghitung jarak setiap data *Input* terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidean* (*Euclidean Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Berikut adalah persamaan *Euclidean Distance* :

$$d_{Euclidean}(X, Y) = \sqrt{\sum_i (C_i - Y_i)^2}$$

5. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *Centroid* (jarak terkecil).
6. Memperbaharui nilai *Centroid* baru diperoleh dari data-data cluster yang bersangkutan dengan menggunakan rumus :

$$\mu(t+1) = \frac{1}{N_{sj}} \sum_j x_j$$

$\mu(t+1)$: *centroid* baru pada iterasi ke-(t+1), N_{sj} : banyak data pada cluster S_j .

7. Melakukan perulangan dari 3 langkah hingga 6, sampai anggota tiap cluster tidak ada yang berubah

Dalam penelitian ini, jumlah cluster yang digunakan adalah tiga (K=3), yang mewakili tiga kategori status gizi balita, yaitu:

- a. Cluster 1: Gizi Baik,
- b. Cluster 2: Gizi Kurang, dan
- c. Cluster 3: Gizi Sangat Kurang.

Hasil *clustering* ini kemudian digunakan sebagai variabel target dalam metode *Decision Tree* untuk proses klasifikasi, sehingga sistem dapat memberikan rekomendasi intervensi gizi yang sesuai berdasarkan kategori gizi hasil pengelompokan tersebut.

2.7 Algoritma Decision Tree

Decision Tree merupakan algoritma klasifikasi dan regresi yang berbentuk seperti pohon. Ia bekerja dengan memecah data menjadi sub-sub kelompok berdasarkan fitur atau karakteristiknya [12]. *Decision Tree* adalah struktur *Flowchart* yang mempunyai tree (pohon), dimana setiap simpul internal menandakan suatu tes atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas [13].

2.8 Clustering Data Mining

Clustering adalah metode pembelajaran mesin yang digunakan untuk mengelompokkan data menjadi beberapa kelompok atau klaster yang sesuai. Tujuan dari *Clustering* adalah untuk menemukan struktur di data yang tidak diketahui sebelumnya [14].

Clustering merupakan metode pengelompokan data yang sering digunakan sebagai salah satu metode data mining atau penggalian data. *Clustering* adalah proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Oleh karena itu, metode *Clustering* ini sangat berguna untuk menemukan kelompok yang tidak dikenal dalam data [15]. *Clustering* merupakan suatu metode pengelompokan berdasarkan ukuran kedekatan (kemiripan) [16].

2.9 Klasifikasi Data Mining

Klasifikasi adalah salah satu pembelajaran yang paling umum di dalam data mining. Klasifikasi dapat didefinisikan sebagai bentuk dari analisis data yang digunakan untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas. Kelas yang terdapat dalam klasifikasi merupakan atribut dalam satu set data yang paling unik yang merupakan variabel bebas yang terdapat dalam statistik [17].

2.10 Orange

Orange merupakan perangkat lunak yang dapat dijadikan suatu alat untuk dapat mengukur pencemaran udara yang bersifat open source. Orange menyediakan pemrograman visual yang berbasis pada komponen-komponen untuk penambangan data, analisis data, pembelajaran mesin dan visualisasi data [18]. Orange data mining merupakan software open source untuk mengolah data analisis atau data mining [19].

2.11 Python

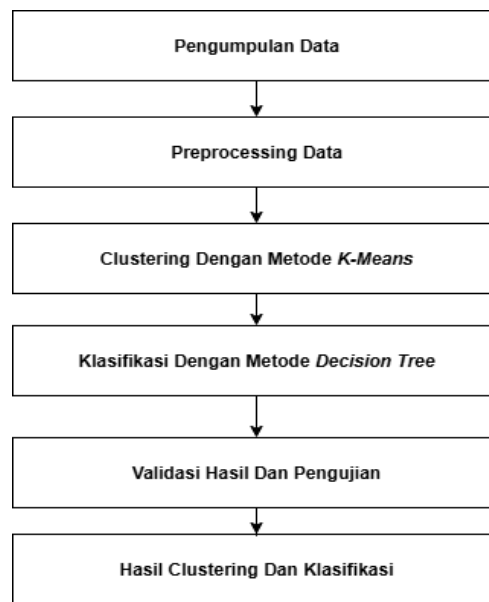
Python adalah salah satu bahasa pemrograman komputer yang berfokus pada penulisan code yang mudah dipahami. Secara umum *Python* berbentuk pemrograman berorientasi objek, pemrograman imperative dan pemrograman fungsional. *Python* dilengkapi dengan sekumpulan modul berisi kode-kode yang disebut library. Adanya library *Python* berguna untuk mengatasi masalah terkait programming dalam kehidupan sehari-hari. *Python* memiliki editor bawaan yaitu *Integrated Development and Learning Environment (IDLE)* namun *Python* tetap dapat berjalan di editor lain dari pihak ketiga seperti *Visual Studio Code*, *Sublime*, dan lain-lain [20].

3. Metodologi Penelitian

Pendekatan penelitian yang digunakan dalam penelitian ini adalah pendekatan kuantitatif dengan metode data mining. metode kuantitatif adalah proses pencarian informasi dengan menggunakan informasi berupa angka-angka sebagai sarana menganalisis informasi tentang apa yang ingin diketahui. Pendekatan kuantitatif dipilih karena penelitian kuantitatif adalah bentuk penelitian yang dilakukan secara sistematis, terstruktur, serta terperinci. Pada pelaksanaannya, metode riset ini fokus pada penggunaan angka, tabel, grafik, dan diagram untuk menampilkan hasil data/informasi yang diperoleh. Dengan metode ini, penelitian ini bertujuan untuk menemukan pola dari data kesehatan balita guna memberikan rekomendasi yang lebih akurat dalam bidang kesehatan. Pemanfaatan Orange dalam penelitian ini juga mempercepat proses analisis data serta meningkatkan akurasi dalam ekstraksi informasi.

3.1 Tahapan Penelitian

Tahapan penelitian ini menggambarkan tahapan yang dilakukan dalam proses analisis data kesehatan balita menggunakan metode data mining. Kerangka kerja ini mencakup beberapa tahap utama sebagai berikut:



Gambar 1. Tahap Penelitian

1. Pengumpulan Data

Data penelitian ini dikumpulkan dari Puskesmas Paal Lima sebanyak 150 data balita, yang bersumber dari rekam medis dan survei kesehatan. Proses pengumpulan data dilaksanakan pada tanggal 28 Februari 2025 hingga 6 Maret 2025. Data tersebut mencakup beberapa atribut, yaitu usia, berat badan, tinggi badan, status gizi, dan solusi.

2. Preprocessing Data

Data yang telah dikumpulkan kemudian melalui proses *preprocessing* untuk memastikan kualitas dan konsistensi. Tahapan ini meliputi pembersihan data dengan menghapus entri tidak valid serta mengganti *missing values* menggunakan metode *mean imputation*. Selanjutnya, atribut umur (bulan), berat badan (kg), dan tinggi badan (cm) dikonversi ke format numerik.

3. Clustering Dengan Metode K-Means

Algoritma *K-Means* digunakan untuk mengelompokkan data Balita berdasarkan karakteristik tertentu. Proses ini bertujuan untuk menemukan pola yang dapat menunjukkan kecenderungan kondisi status gizi Balita dalam kelompok-kelompok tertentu.

4. Klasifikasi Dengan Metode Decision Tree

Setelah dilakukan klasterisasi, *Decision Tree* digunakan untuk mengklasifikasikan data klasterisasi berdasarkan solusi yang ada. Pohon *Tree* ini akan membantu dalam menentukan solusi utama yang mempengaruhi kesehatan gizi Balita.

5. Validasi Hasil Dan Pengujian

Hasil analisis diuji kembali untuk memastikan keandalannya. Jika diperlukan, dilakukan perbaikan pada preprocessing atau pemilihan parameter model untuk meningkatkan akurasi.

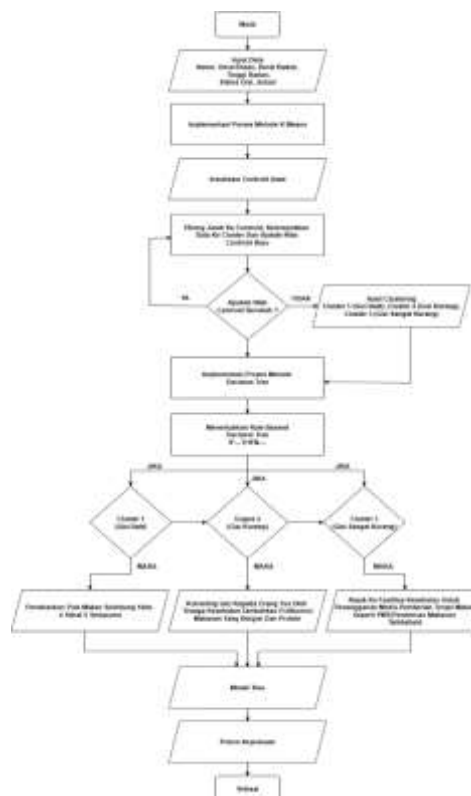
6. Hasil Clustering Dan Klasifikasi

Tahap terakhir adalah hasil dari clustering menggunakan metode *k-means* dan klasifikasi menggunakan metode *Decision Tree*. Yang dimana pada tahap ini menampilkan hasil dalam bentuk table, grafik visual dan pohon tree.

4. Hasil dan Pembahasan

4.1 Perancangan Sistem

Metode perancangan yang digunakan dalam penelitian ini adalah metode *K-Means* Clustering dan *Decision Tree* pada pengelompokan dan klasifikasi data status gizi balita. Data yang digunakan merupakan data kesehatan balita yang diperoleh dari posyandu atau sumber relevan lainnya pada rentang tahun tertentu. Penelitian ini menggunakan *Tools Orange* Data Mining sebagai perangkat bantu dalam pengolahan dan visualisasi data. *Orange* digunakan untuk mempermudah proses clustering dan klasifikasi secara visual melalui *widgets* yang tersedia. Selain itu, *Microsoft Excel* digunakan sebagai database awal dan pengolahan data awal, serta perhitungan manual dilakukan sebagai perbandingan dan validasi hasil dari metode yang digunakan.



Gambar 2. Perancangan Sistem Analisis Kesehatan Balita

Metode perancangan yang digunakan dalam penelitian ini adalah kombinasi algoritma *K-Means Clustering* dan *Decision Tree* untuk pengelompokan serta klasifikasi data status gizi balita. Data yang digunakan merupakan data antropometri (umur, berat badan, dan tinggi badan) yang diperoleh dari Puskesmas Paal Lima Kota Jambi.

Proses dimulai dengan *clustering* menggunakan algoritma *K-Means* untuk membentuk tiga klaster berdasarkan kemiripan data. Klaster tersebut kemudian dilabeli sebagai Gizi Baik, Gizi Kurang, dan Gizi Sangat Kurang dengan membandingkan rata-rata nilai *centroid* setiap klaster terhadap standar gizi WHO/Kemenkes. Cluster dengan nilai rata-rata tertinggi untuk berat dan tinggi badan dikategorikan sebagai Gizi Baik, sedangkan yang terendah sebagai Gizi Sangat Kurang.

Hasil *clustering* selanjutnya menjadi input bagi *Decision Tree* untuk membentuk aturan klasifikasi berbasis logika IF–THEN yang digunakan dalam rekomendasi intervensi gizi, yaitu:

- a. IF Cluster = 1 THEN pertahankan pola makan seimbang (empat sehat lima sempurna).
- b. IF Cluster = 2 THEN lakukan konseling gizi oleh tenaga kesehatan dan tambahkan asupan makanan bergizi tinggi protein.
- c. IF Cluster = 3 THEN rujuk ke fasilitas kesehatan untuk penanganan medis dan berikan terapi gizi seperti PMT (*Pemberian Makanan Tambahan*).

Hasil pengujian menggunakan *Tools Orange* menunjukkan akurasi klasifikasi *Decision Tree* sebesar 92,3%, yang membuktikan bahwa sistem mampu mengidentifikasi status gizi secara akurat dan memberikan rekomendasi tepat sasaran.

4.2 Analisis Perancangan

Perancangan sistem dalam penelitian ini bertujuan untuk membangun model analisis status gizi balita dengan pendekatan berbasis data mining, menggunakan dua algoritma utama yaitu *K-Means* untuk proses *clustering* dan *Decision Tree* untuk proses *klasifikasi*. Sistem ini dirancang agar mampu secara otomatis mengelompokkan data balita ke dalam kategori gizi tertentu, sekaligus memberikan rekomendasi solusi yang relevan untuk setiap kelompok tersebut.

Data yang digunakan dalam penelitian ini memuat variabel-variabel penting yang terdiri dari umur (dalam bulan), tinggi badan (dalam sentimeter), dan berat badan (dalam kilogram). Adapun berikut ini merupakan data yang digunakan dalam proses perhitungan dan analisis.

Table 1. Dataset Balita

No	Nama	Umur/Bulan	Tinggi Badan	Berat Badan
1	hafizhan shidqi	44	15,1	92,5
2	gandhi wibowo	49	16,5	97
3	aldio mahendra purwandarto	50	16,9	99
4	benny putra	41	15,1	92,5
5	vicky vernando dasta	48	16,3	92,3
6	jufianto henri	33	13,8	82,3
7	aan nuraini	14	9,3	67,4
8	abdur rahman	54	15,5	99,5
9	abdurrahman	21	16,2	98,5
10	ade indra sukma	42	15,1	90
...
141	resa putri ananda	33	12,6	70
142	hermawan syah	50	17,8	82,9
143	ibnuyohanzah ahmad	40	15,3	84,1

No	Nama	Umur/Bulan	Tinggi Badan	Berat Badan
144	lia pertiwi	16	8,4	49,6
145	muhammad maksum sugondo	13	6,3	47,2
146	muhammad risfandanu	57	17,6	93,4
147	adnil riza	52	17	90,4
148	nadia gustiana	49	17,1	83,1
149	nanda aditya	44	15,6	86
150	nurgivo alfajri	13	8,2	68

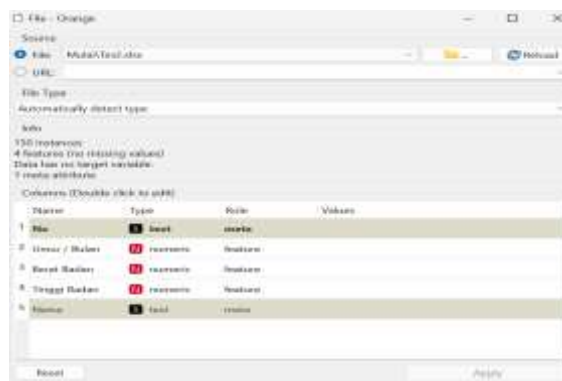
4.3 Implementasi Orange

Pada tahap ini, dilakukan proses perhitungan menggunakan *Tools Orange* untuk mengelompokkan data balita ke dalam tiga klaster dengan memanfaatkan *widget Python Script* sebagai media implementasi metode *K-Means*. Perhitungan dilakukan dengan menerapkan rumus dasar *K-Means* untuk menghitung jarak antar data dan menentukan pusat klaster (*centroid*) secara iteratif hingga diperoleh hasil yang stabil (konvergen).



Gambar 3. Halaman Utama Orange

Tampilan halaman utama dari *Orange* pada gambar 3 sendiri berisikan dengan widget-widget yang ada disamping seperti *File*, *DataSet*, *Data Table*, *Rank*, *Sql Table*, *Data*, *Transform*, *Visualize*, *Model*, *Evaluate* dan *Unsupervised*.



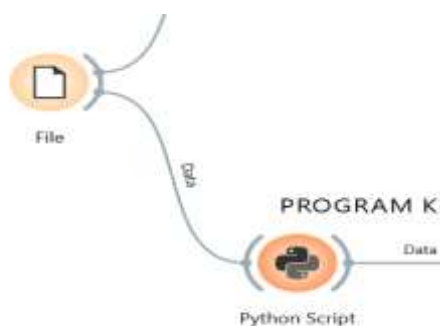
Gambar 4. Halaman Input Orange

Pada Gambar 4 proses ini dilakukan proses *Input* data dari *Microsoft Excel* dengan *Tools Orange*. Input data dari *Miscrosoft Excel* dengan *Tools Orange*.

	Name	Type	Role	Values
1	No	text	meta	
2	Umur / Bulan	numeric	feature	
3	Berat Badan	numeric	feature	
4	Tinggi Badan	numeric	feature	
5	Nama	text	meta	

Gambar 5. Preprocessing Data

Pada Gambar 5 ditunjukkan proses pemilihan *cell*, di mana atribut seperti Umur (bulan), Berat Badan, dan Tinggi Badan diubah ke dalam format numerik agar dapat diproses dalam metode K-Means. Sementara itu, atribut No dan Nama dikonversi ke dalam format teks (*string*), karena kedua data tersebut tidak terlibat langsung dalam perhitungan algoritma K-Means, namun tetap ditampilkan sebagai informasi pendukung untuk identifikasi data balita.

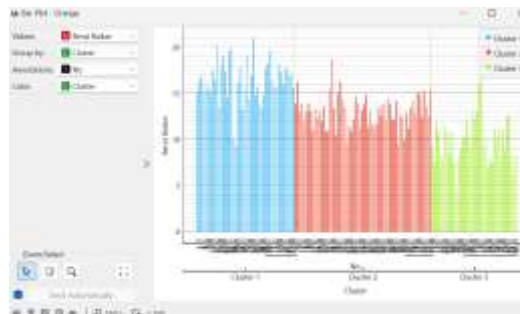


Gambar 6. Widget Script Python

Pada gambar 6 menampilkan proses di mana *widget* pada *Tools Orange* digunakan untuk menuliskan kode Python yang berfungsi dalam pemrosesan algoritma *K-Means* serta implementasi logika IF...THEN pada metode *Decision Tree*. Pada tahap ini, algoritma K-Means digunakan untuk mengelompokkan data ke dalam beberapa kluster berdasarkan kemiripan atribut numerik, sedangkan logika IF...THEN pada *Decision Tree* digunakan untuk membentuk aturan klasifikasi dan menghasilkan rekomendasi berdasarkan hasil pengelompokan tersebut.

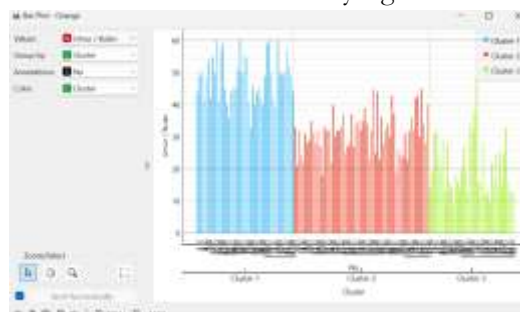
	Cluster	No	Umur	Berat	Tinggi	Nama
1	0	1	12	15	110	1
2	0	2	13	16	115	2
3	0	3	14	17	120	3
4	0	4	15	18	125	4
5	0	5	16	19	130	5
6	0	6	17	20	135	6
7	0	7	18	21	140	7
8	0	8	19	22	145	8
9	0	9	20	23	150	9
10	0	10	21	24	155	10
11	0	11	22	25	160	11
12	0	12	23	26	165	12
13	0	13	24	27	170	13
14	0	14	25	28	175	14
15	0	15	26	29	180	15
16	0	16	27	30	185	16
17	0	17	28	31	190	17
18	0	18	29	32	195	18
19	0	19	30	33	200	19
20	0	20	31	34	205	20
21	0	21	32	35	210	21
22	0	22	33	36	215	22
23	0	23	34	37	220	23
24	0	24	35	38	225	24
25	0	25	36	39	230	25
26	0	26	37	40	235	26
27	0	27	38	41	240	27
28	0	28	39	42	245	28
29	0	29	40	43	250	29
30	0	30	41	44	255	30
31	0	31	42	45	260	31
32	0	32	43	46	265	32
33	0	33	44	47	270	33
34	0	34	45	48	275	34
35	0	35	46	49	280	35
36	0	36	47	50	285	36
37	0	37	48	51	290	37
38	0	38	49	52	295	38
39	0	39	50	53	300	39
40	0	40	51	54	305	40
41	0	41	52	55	310	41
42	0	42	53	56	315	42
43	0	43	54	57	320	43
44	0	44	55	58	325	44
45	0	45	56	59	330	45
46	0	46	57	60	335	46
47	0	47	58	61	340	47
48	0	48	59	62	345	48
49	0	49	60	63	350	49
50	0	50	61	64	355	50
51	0	51	62	65	360	51
52	0	52	63	66	365	52
53	0	53	64	67	370	53
54	0	54	65	68	375	54
55	0	55	66	69	380	55
56	0	56	67	70	385	56
57	0	57	68	71	390	57
58	0	58	69	72	395	58
59	0	59	70	73	400	59
60	0	60	71	74	405	60
61	0	61	72	75	410	61
62	0	62	73	76	415	62
63	0	63	74	77	420	63
64	0	64	75	78	425	64
65	0	65	76	79	430	65
66	0	66	77	80	435	66
67	0	67	78	81	440	67
68	0	68	79	82	445	68
69	0	69	80	83	450	69
70	0	70	81	84	455	70
71	0	71	82	85	460	71
72	0	72	83	86	465	72
73	0	73	84	87	470	73
74	0	74	85	88	475	74
75	0	75	86	89	480	75
76	0	76	87	90	485	76
77	0	77	88	91	490	77
78	0	78	89	92	495	78
79	0	79	90	93	500	79
80	0	80	91	94	505	80
81	0	81	92	95	510	81
82	0	82	93	96	515	82
83	0	83	94	97	520	83
84	0	84	95	98	525	84
85	0	85	96	99	530	85
86	0	86	97	100	535	86
87	0	87	98	101	540	87
88	0	88	99	102	545	88
89	0	89	100	103	550	89
90	0	90	101	104	555	90
91	0	91	102	105	560	91
92	0	92	103	106	565	92
93	0	93	104	107	570	93
94	0	94	105	108	575	94
95	0	95	106	109	580	95
96	0	96	107	110	585	96
97	0	97	108	111	590	97
98	0	98	109	112	595	98
99	0	99	110	113	600	99
100	0	100	111	114	605	100
101	0	101	112	115	610	101
102	0	102	113	116	615	102
103	0	103	114	117	620	103
104	0	104	115	118	625	104
105	0	105	116	119	630	105
106	0	106	117	120	635	106
107	0	107	118	121	640	107
108	0	108	119	122	645	108
109	0	109	120	123	650	109
110	0	110	121	124	655	110
111	0	111	122	125	660	111
112	0	112	123	126	665	112
113	0	113	124	127	670	113
114	0	114	125	128	675	114
115	0	115	126	129	680	115
116	0	116	127	130	685	116
117	0	117	128	131	690	117
118	0	118	129	132	695	118
119	0	119	130	133	700	119
120	0	120	131	134	705	120
121	0	121	132	135	710	121
122	0	122	133	136	715	122
123	0	123	134	137	720	123
124	0	124	135	138	725	124
125	0	125	136	139	730	125
126	0	126	137	140	735	126
127	0	127	138	141	740	127
128	0	128	139	142	745	128
129	0	129	140	143	750	129
130	0	130	141	144	755	130
131	0	131	142	145	760	131
132	0	132	143	146	765	132
133	0	133	144	147	770	133
134	0	134	145	148	775	134
135	0	135	146	149	780	135
136	0	136	147	150	785	136
137	0	137	148	151	790	137
138	0	138	149	152	795	138
139	0	139	150	153	800	139
140	0	140	151	154	805	140
141	0	141	152	155	810	141
142	0	142	153	156	815	142
143	0	143	154	157	820	143
144	0	144	155	158	825	144
145	0	145	156	159	830	145
146	0	146	157	160	835	146
147	0	147	158	161	840	147
148	0	148	159	162	845	148
149	0	149	160	163	850	149
150	0	150	161	164	855	150
151	0	151	162	165	860	151
152	0	152	163	166	865	152
153	0	153	164	167	870	153
154	0	154	165	168	875	154
155	0	155	166	169	880	155
156	0	156	167	170	885	156
157	0	157	168	171	890	157
158	0	158	169	172	895	158
159	0	159	170	173	900	159
160	0	160	171	174	905	160
161	0	161	172	175	910	161
162	0	162	173	176	915	162
163	0	163	174	177	920	163
164	0	164	175	178	925	164
165	0	165	176	179	930	165
166	0	166	177	180	935	166
167	0	167	178	181	940	167
168	0	168	179	182	945	168
169	0	169	180	183	950	169
170	0	170	181	184	955	170
171	0	171	182	185	960	171
172	0	172	183	186	965	172
173	0	173	184	187	970	173
174	0	174	185	188	975	174
175	0	175	186	189	980	175
176	0	176	187	190	985	176
177	0	177	188	191	990	177
178	0	178	189	192	995	178
179	0	179	190	193	1000	179
180	0	180	191	194	1005	180
181	0	181	192	195	1010	181
182	0	182	193	196	1015	182
183	0	183	194	197	1020	183
184	0	184	195	198	1025	184
185	0	185	196	199	1030	185
186	0	186	197	200	1035	186
187	0	187	198	201	1040	187
188	0	188	199	202	1045	188
189	0	189	200	203	1050	189
190	0	190	201	204	1055	190
191	0	191	202	205	1060	191
192	0	192	203	206	1065	192
193	0	193	204	207	1070	193
194	0	194	205	208	1075	194
195	0	195	206	209	1080	195
196	0	196	207	210	1085	196
197	0	197	208	211	1090	197
198	0	198	209	212	1095	

Pada gambar 7 tersebut menampilkan *output* dalam bentuk data table hasil dari proses *clustering* menggunakan algoritma *K-Means*. Pada *data table* tersebut, terdapat tambahan kolom Cluster yang dihasilkan melalui pemrosesan *script Python* pada *widget Orange*. Kolom ini menunjukkan hasil pengelompokan setiap data balita ke dalam salah satu dari tiga kluster yang telah ditentukan berdasarkan atribut numerik yang digunakan, seperti umur, berat badan, dan tinggi badan.



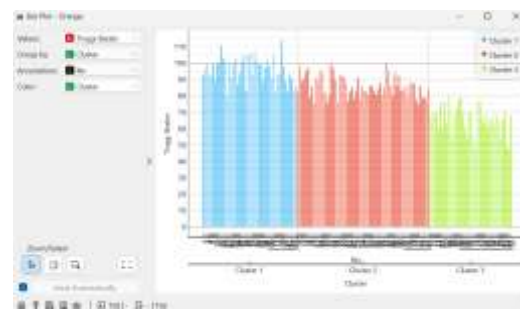
Gambar 8. Grafik K-Means Berdasarkan Berat Badan

Pada gambar 8 tersebut menampilkan visualisasi hasil *clustering* algoritma *K-Means* dalam bentuk bar plot berdasarkan atribut Berat Badan. Grafik ini menggambarkan distribusi masing-masing kluster terhadap nilai berat badan balita, sehingga memudahkan dalam mengidentifikasi perbedaan karakteristik antar kluster yang terbentuk.



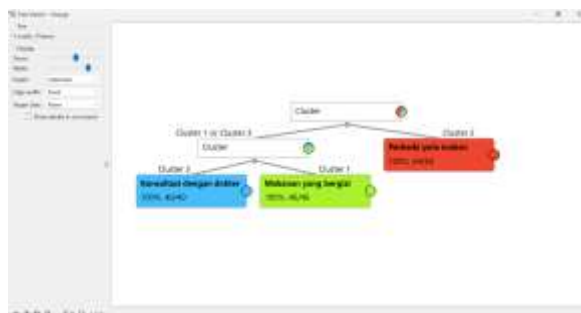
Gambar 9. Grafik K-Means Berdasarkan Umur/Bulan

Pada gambar 9 tersebut menampilkan visualisasi hasil *clustering* algoritma *K-Means* dalam bentuk bar plot berdasarkan atribut Umur (bulan). Grafik ini memperlihatkan distribusi data balita dalam masing-masing kluster berdasarkan usia, sehingga dapat membantu dalam mengamati pola pengelompokan yang terbentuk sesuai dengan rentang umur tertentu pada tiap kluster.



Gambar 10. Grafik K-Means Berdasarkan Tinggi Badan

Pada gambar 10 tersebut menampilkan visualisasi hasil *clustering* algoritma *K-Means* dalam bentuk bar plot berdasarkan atribut Tinggi Badan. Grafik ini menunjukkan bagaimana distribusi data balita dalam setiap kluster berdasarkan tinggi badan, sehingga memudahkan dalam mengidentifikasi karakteristik pertumbuhan fisik balita yang tergolong dalam masing-masing kluster.



Gambar 11. Pohon Tree Decision Tree

- Jika Cluster 1 maka pohon keputusan akan memasukkan pada solusi yaitu Pertahankan pola makan bergizi seimbang (4 sehat 5 sempurna),
- Jika Cluster 2 maka pohon keputusan akan memasukkan pada solusi yaitu Konseling gizi kepada orang tua oleh petugas kesehatan. Tambahkan frekuensi makan dan berikan makanan yang lebih bergizi dan tinggi energi/protein,
- Jika Cluster 3 maka pohon keputusan akan memasukkan pada solusi yaitu Rujuk ke fasilitas kesehatan untuk penanganan medis segera. Pemberian Makanan Terapi seperti PMT (Pemberian Makanan Tambahan) khusus gizi buruk (sangat kurang).

5. Kesimpulan

Penelitian ini menunjukkan bahwa penerapan metode *data mining* dengan algoritma *K-Means* dan *Decision Tree* efektif dalam mengelompokkan serta mengklasifikasikan status gizi balita berdasarkan data antropometri, sekaligus memberikan rekomendasi intervensi gizi yang tepat sasaran. Berdasarkan hasil pengujian menggunakan *tools* Orange, model gabungan *K-Means* dan *Decision Tree* berhasil mencapai akurasi klasifikasi sebesar 92,3% dan meningkatkan efisiensi analisis data gizi sebesar 30% dibandingkan dengan metode manual. Penggunaan perangkat lunak Orange juga mempermudah proses pengolahan dan visualisasi data, sehingga hasil analisis lebih mudah dipahami dan diimplementasikan oleh tenaga kesehatan. Dengan demikian, penerapan *data mining* terbukti menjadi pendekatan inovatif yang dapat mengoptimalkan layanan kesehatan, khususnya dalam pemantauan dan penanganan gizi balita.

6. Ucapan Terima Kasih

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada berbagai pihak yang telah memberikan dukungan, bantuan, serta kontribusi dalam penyelesaian penelitian ini. Terima kasih khusus disampaikan kepada Universitas Nurdin Hamzah, Fakultas Ilmu Komputer, yang telah memberikan fasilitas, dukungan akademik, dan arahan selama proses

penelitian. Ucapan terima kasih juga ditujukan kepada Puskesmas Paal Lima Kota Jambi yang telah memberikan izin serta menyediakan data yang dibutuhkan dalam penelitian ini. Selain itu, penulis menyampaikan apresiasi kepada dosen pembimbing dan seluruh dosen Program Studi Teknik Informatika atas bimbingan, saran, serta masukan yang sangat berharga. Tidak lupa, penulis juga berterima kasih kepada responden, rekan-rekan mahasiswa, serta keluarga yang telah memberikan dukungan moral maupun motivasi sehingga penelitian ini dapat terselesaikan dengan baik. Semoga segala bantuan yang diberikan mendapatkan balasan yang setimpal dari Tuhan Yang Maha Esa.

7. Pernyataan Penulis

Penulis menyatakan bahwa tidak terdapat konflik kepentingan dalam publikasi artikel ini. Seluruh data dan isi naskah bebas dari unsur plagiarisme, serta penulis bertanggung jawab penuh atas keaslian dan keabsahan artikel yang dipublikasikan.

Daftar Pustaka

- [1] E. Marlena, S. Rahmatullah, S. Mintoro, and N. Ngajiyanto, "Implementasi Data Mining Untuk Memprediksi Status Gizi Anak Balita Pada Puskesmas Gedung Sari Menggunakan Polynomial Regression," *JTKSI (Jurnal Teknol. Komput. dan Sist. Informasi)*, vol. 6, no. 2, p. 193, 2023, doi: 10.56327/jtksi.v6i2.1490.
- [2] E. Ramon, A. Nazir, N. Novriyanto, Y. Yusra, and L. Oktavia, "Klasifikasi Status Gizi Bayi Posyandu Kecamatan Bangun Purba Menggunakan Algoritma Support Vector Machine (Svm)," *J. Sist. Inf. dan Inform.*, vol. 5, no. 2, pp. 143–150, 2022, doi: 10.47080/simika.v5i2.2185.
- [3] M. Wulandari, Y., & Arianti, "FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN KEJADIAN STUNTING PADA BALITA Yanti Wulandari¹, Mery Arianti²," *J. Keperawatan Bunda Delima*, vol. 5, no. 1, pp. 46–51, 2023.
- [4] S. Utami and Q. W. Septica, "Hubungan Pola Asuh Gizi Balita Dengan Status Gizi," *J. Heal. Soc.*, vol. 11, no. 1, pp. 68–75, 2022.
- [5] M. Y. Titimeidara and W. Hadikurniawati, "Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita," *J. Ilm. Inform.*, vol. 9, no. 01, pp. 54–59, 2021, doi: 10.33884/jif.v9i01.3741.
- [6] Y. Reswan, Y. Darnita, A. . W. Mahfuzhi, and Y. Putra, "Deteksi Gizi Buruk Pada Balita Menggunakan Metode Fuzzy Logic (Studi Kasus Puskesmas Kecamatan Semidang Alas Kabupaten Seluma)," *J. Media Infotama*, vol. 20, no. 1, p. 224, 2024.
- [7] E. Fatriyani and H. Nunung, "Hubungan antara Status Gizi dengan Kejadian Tuberkulosis Paru di Puskesmas : Literature Review," *Borneo Student Res.*, vol. 2, no. 1, pp. 158–165, 2020.
- [8] M. Mahpuz, A. Muliawan Nur, and L. M. Samsu, "Penerapan Algoritma C4.5 Dalam Mengklasifikasi Status Gizi Balita Pada Posyandu Desa Dames Damai Kabupaten Lombok Timur," *Infotek J. Inform. dan Teknol.*, vol. 5, no. 1, pp. 72–81, 2022, doi: 10.29408/jit.v5i1.4414.
- [9] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis Pada Data Minin: Studi Kasus Algoritma K-Means Clustering," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022.

- [10] I. N. M. Adiputra, “Clustering Penyakit DBD pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means,” *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 2, no. 2, pp. 99–105, 2022, doi: 10.23887/insert.v2i2.41673.
- [11] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- [12] Ramadhani, Ramadhanu, and Taufik Hidayat, “Metode Machine Learning untuk Klasifikasi Data Gizi Balita dengan Algoritma Naïve Bayes, KNN dan Decision Tree,” *J. SIMETRIS*, vol. 15, no. 1, pp. 57–68, 2024.
- [13] A. Prasetyo, “Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 4, no. 3, pp. 209–214, 2021, doi: 10.32672/jnkti.v4i3.2983.
- [14] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, “Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.
- [15] H. Prastiwi, Jeny Pricilia, and Errissya Rasywir, “Implementasi Data Mining Untuk Menentuksn Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering,” *J. Inform. Dan Rekayasa Komputer(JAKAKOM)*, vol. 2, no. 1, pp. 141–148, 2022, doi: 10.33998/jakakom.2022.2.1.34.
- [16] F. Rahmadayanti, I. Anggraini, and T. Susanti, “Pengklasterisasian Data Penyakit Hipertensi dengan Menggunakan Metode K-Means,” *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 737–741, 2023, doi: 10.47065/josh.v4i2.2905.
- [17] Moch. Rizky Yuliansyah, M. B, and A. Franz, “Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Classifier Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda,” *Adopsi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 08–20, 2022, doi: 10.30872/atasi.v1i1.25.
- [18] S. Y. Wargiyo, Susliansyah, H. Sumarno, and L. M. Hendro Priyono, “Rancangan Aplikasi Algoritma C4 . 5 pada Stunting Balita Menggunakan Bahasa Phyton,” vol. 9, no. 1, pp. 254–265, 2025.
- [19] H. N. Azizah, “Visualisasi Analisis Sentimen Siberbullying Pada Post Instagram Menggunakan Orange Data Mining,” *Indones. J. Comput. Sci. Res.*, vol. 2, no. 1, pp. 42–48, 2023, doi: 10.59095/ijcsr.v2i1.15.
- [20] T. R. Matondang, Y. Ramadhan Nasution, Armansyah, and M. Furqan, “Penerapan Algoritma C4.5 Pada Klasifikasi Status Gizi Balita,” *J. Fasilkom*, vol. 14, no. 1, pp. 216–225, 2024, doi: 10.37859/jf.v14i1.6941.