

Clustering Status Gizi Menggunakan Algoritma K-Means Dengan Pendekatan CRISP-DM

Delia Wulan Rahmadhani^{1*}, Amelia Yusnita², Aisyah Fajriantini³

¹ Program Studi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda, Indonesia

² Program Studi Sistem Informasi, STMIK Widya Cipta Dharma, Samarinda, Indonesia

³ Program Studi Teknik Informatika, STMIK Widya Cipta Dharma, Samarinda, Indonesia

Email: ¹2243089@wicida.ac.id, ²amelia@wicida.ac.id, ³aisyah@wicida.ac.id

(* : 2243089@wicida.ac.id)

Abstrak - Status gizi balita merupakan indikator penting dalam menilai tingkat kesehatan masyarakat. Permasalahan gizi seperti gizi kurang masih sering terjadi dan memerlukan analisis yang efektif untuk mengidentifikasi pola data secara cepat dan akurat. Penelitian ini bertujuan untuk mengelompokkan status gizi balita di wilayah Posyandu Puskesmas Bukuan menggunakan algoritma K-Means dengan pendekatan CRISP-DM. Permasalahan utama dalam penelitian ini adalah pengolahan data status gizi yang masih dilakukan secara manual sehingga kurang efektif dalam mengidentifikasi pola dan kelompok risiko secara cepat dan akurat. Dataset yang digunakan berjumlah 2145 data balita yang berasal dari 12 Posyandu, dengan atribut utama berupa berat badan, tinggi badan, usia, dan jenis kelamin. Proses penelitian dilakukan melalui tahapan CRISP-DM yang meliputi *business understanding*, *data understanding*, *data preparation*, *modeling*, dan *evaluation*, tanpa implementasi *deployment* karena penelitian berfokus pada analisis data. Proses clustering dilakukan menggunakan algoritma K-Means dengan penentuan jumlah cluster optimal melalui metode Elbow yang menghasilkan tiga cluster. Evaluasi model menggunakan *Silhouette Score* menghasilkan nilai sebesar 0,629 yang menunjukkan kualitas clustering dalam kategori baik. Hasil penelitian menunjukkan bahwa data balita dapat dikelompokkan ke dalam tiga kategori status gizi, yaitu gizi kurang, gizi cukup (normal), dan gizi lebih berdasarkan karakteristik nilai centroid. Distribusi data menunjukkan bahwa kategori gizi cukup mendominasi, namun masih terdapat jumlah signifikan pada kategori gizi kurang. Dengan demikian, penerapan algoritma K-Means mampu memberikan informasi yang lebih terstruktur dan akurat dalam mengidentifikasi kondisi status gizi balita serta dapat digunakan sebagai dasar dalam pengambilan keputusan program kesehatan masyarakat berbasis data.

Kata Kunci: K-Means, Clustering, Status Gizi, Balita, CRISP-DM, Posyandu, Data Mining

The nutritional status of toddlers is a key indicator for assessing public health levels. Nutritional problems such as undernutrition remain common and require effective analysis to identify data patterns quickly and accurately. This study aims to cluster the nutritional status of toddlers in the Bukuan Community Health Center (Puskesmas) Posyandu area using the K-Means algorithm with the CRISP-DM approach. The main challenge in this study is that the processing of nutritional status data is still done manually, making it less effective in quickly and accurately identifying patterns and risk groups. The dataset consists of 2,145 records of toddlers from 12 Posyandu, with primary attributes including weight, height, age, and gender. The research process was conducted through the CRISP-DM stages, which include business understanding, data understanding, data preparation, modeling, and evaluation, without deployment implementation since the study focused on data analysis. The clustering process was performed using the K-Means algorithm, with the optimal number of clusters determined via the Elbow method, resulting in three clusters. Model evaluation using the Silhouette Score yielded a value of 0.629, indicating that the clustering quality falls into the “good” category. The results show that data on toddlers can be grouped into three nutritional status categories: under-nutrition, adequate nutrition (normal), and over-nutrition, based on centroid values. The data distribution indicates that the adequate nutrition category dominates, though there remains a significant number of cases in the under-nutrition category. Thus, the application of the K-Means algorithm provides more structured and accurate information for identifying the nutritional status of toddlers and can serve as a basis for data-driven decision-making in public health programs.

Keywords: K-Means, Clustering, Nutritional Status, Toddlers, CRISP-DM, Posyandu, Data Mining

1. PENDAHULUAN

Status gizi balita merupakan salah satu indikator utama dalam menilai kualitas kesehatan masyarakat serta keberhasilan pembangunan suatu negara. Permasalahan gizi seperti *stunting*, *wasting*, dan *underweight* masih menjadi tantangan global, khususnya di negara berkembang. Kondisi ini tidak hanya berdampak pada pertumbuhan fisik anak, tetapi juga memengaruhi perkembangan kognitif, kesehatan jangka panjang, serta produktivitas di masa depan [1]. Di Indonesia, permasalahan status gizi balita masih menjadi isu yang signifikan dan memerlukan perhatian serius. Berdasarkan laporan Survei Kesehatan Indonesia tahun 2023, status gizi balita masih menjadi fokus utama dalam pembangunan kesehatan nasional karena berkaitan erat dengan kualitas sumber daya manusia di masa mendatang [2]. Selain itu, data Riskesdas menunjukkan bahwa prevalensi stunting dan gizi kurang masih cukup tinggi di berbagai wilayah Indonesia, yang menunjukkan adanya ketimpangan dalam pemenuhan kebutuhan gizi masyarakat [3].

Secara teoritis, status gizi balita dapat diukur menggunakan indikator antropometri seperti berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U), dan berat badan menurut tinggi badan (BB/TB). Indikator ini digunakan untuk menilai kondisi gizi secara objektif dan mendeteksi gangguan pertumbuhan sejak dini [4]. Pengukuran antropometri dinilai efektif karena sederhana, mudah dilakukan, serta mampu memberikan gambaran kondisi gizi secara akurat berdasarkan standar WHO.

Pada tingkat pelayanan kesehatan masyarakat, Posyandu memiliki peran penting dalam melakukan pemantauan status gizi balita melalui kegiatan penimbangan dan pencatatan rutin. Namun, pengolahan data yang masih dilakukan secara manual menyebabkan keterbatasan dalam mengidentifikasi pola dan kelompok risiko secara cepat dan akurat [5]. Hal ini berdampak pada kurang optimalnya proses pengambilan keputusan dalam menentukan intervensi gizi yang tepat sasaran. Selain itu, status gizi balita dipengaruhi oleh berbagai faktor, seperti pola asuh, kondisi sosial ekonomi, serta tingkat pengetahuan orang tua terhadap pemenuhan gizi anak. Ketidakseimbangan asupan gizi dalam lima tahun pertama kehidupan dapat menyebabkan gangguan pertumbuhan yang bersifat permanen dan berdampak pada kualitas hidup anak di masa depan [6]. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu mengidentifikasi kondisi gizi secara lebih komprehensif.

Seiring dengan perkembangan teknologi informasi, metode data mining menjadi salah satu solusi dalam pengolahan data kesehatan. Salah satu teknik yang banyak digunakan adalah clustering dengan algoritma K-Means, yang mampu mengelompokkan data berdasarkan kemiripan karakteristik secara efisien [7]. Algoritma ini bekerja dengan membagi data ke dalam beberapa cluster sehingga data dalam satu kelompok memiliki tingkat kesamaan yang tinggi dibandingkan dengan kelompok lainnya.

Penelitian mengenai pengelompokan status gizi balita menggunakan algoritma K-Means telah banyak dilakukan. [8] menerapkan algoritma K-Means dengan pendekatan Soft System Methodology (SSM) pada data Posyandu Kencana Kota Bekasi. Hasil penelitian menunjukkan bahwa data balita berhasil dikelompokkan ke dalam lima kategori, yaitu gizi buruk, gizi kurang, gizi baik, gizi lebih, dan obesitas. Pendekatan tersebut dinilai mampu membantu perancangan program intervensi kesehatan yang lebih tepat sasaran. Selanjutnya, [9] menerapkan algoritma K-Means pada data gizi balita di UPTD Puskesmas Bumi Agung. Penelitian tersebut menghasilkan sistem yang mampu melakukan pencarian, penyimpanan, dan pengelompokan data gizi balita secara otomatis sehingga membantu petugas kesehatan dalam mengelola data yang sebelumnya dilakukan secara manual. [10] melakukan analisis clustering data gizi balita pada Puskesmas 1 Ulu Palembang menggunakan algoritma K-Means dengan pendekatan Knowledge Discovery in Database (KDD). Hasil penelitian menunjukkan terbentuknya tiga cluster status gizi yang dapat digunakan sebagai dasar dalam penyusunan program intervensi gizi yang lebih tepat sasaran. [11] menerapkan algoritma K-Means untuk mengelompokkan kondisi gizi balita pada Posyandu. Hasil penelitian menunjukkan bahwa metode K-Means mampu mengidentifikasi kelompok status gizi balita secara efektif dan membantu proses pengambilan keputusan berbasis data dalam bidang kesehatan masyarakat. [12] menerapkan algoritma K-Means untuk memprediksi status gizi balita pada tiga puskesmas di Kecamatan Simanindo menggunakan 955 data balita. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan balita ke dalam empat kategori status gizi, yaitu gizi sangat kurang, gizi kurang, gizi normal, dan risiko lebih. Penelitian ini membuktikan bahwa metode K-Means dapat digunakan sebagai alat bantu dalam pemantauan dan analisis status gizi balita secara efektif.

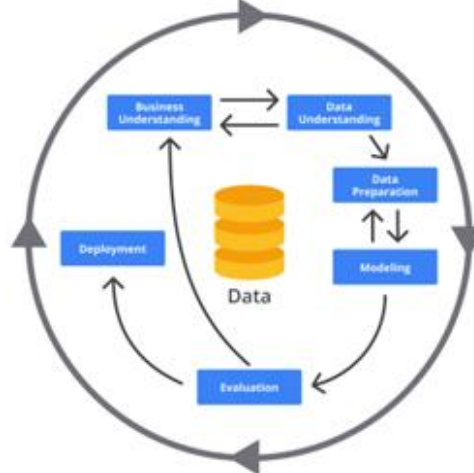
Penelitian terbaru [13] menerapkan algoritma K-Means pada 300 data balita di UPT Puskesmas Barong Tongkok. Hasil penelitian berhasil membentuk tiga kelompok status gizi, yaitu underweight, gizi baik, dan gizi lebih. Selain itu, penelitian tersebut membandingkan beberapa metode perhitungan jarak dan menemukan bahwa Minkowski Distance menghasilkan nilai Sum Squared Error (SSE) terbaik sehingga mampu menghasilkan pengelompokan yang lebih optimal.

Meskipun demikian, sebagian besar penelitian sebelumnya hanya berfokus pada penerapan algoritma clustering atau pengembangan sistem berbasis K-Means tanpa menggunakan kerangka kerja data mining yang sistematis. Beberapa penelitian telah menerapkan pendekatan KDD maupun Soft System Methodology, namun masih terbatas penelitian yang mengintegrasikan algoritma K-Means dengan metodologi CRISP-DM pada data status gizi balita di tingkat Posyandu. Padahal, CRISP-DM menyediakan tahapan analisis yang terstruktur mulai dari business understanding, data understanding, data preparation, modeling, hingga evaluation sehingga mampu meningkatkan kualitas proses analisis dan interpretasi hasil. Oleh karena itu, penelitian ini menerapkan algoritma K-Means dengan pendekatan CRISP-DM untuk mengelompokkan status gizi balita pada Posyandu wilayah Puskesmas Bukuan sebagai dasar pendukung pengambilan keputusan program kesehatan berbasis data. Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk melakukan clustering status gizi balita menggunakan algoritma K-Means di Posyandu wilayah Puskesmas Bukuan dengan pendekatan CRISP-DM. Penelitian ini diharapkan dapat menghasilkan pengelompokan data yang akurat serta memberikan informasi yang dapat digunakan sebagai dasar dalam pengambilan keputusan dan perencanaan program kesehatan masyarakat secara lebih efektif dan berbasis data.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Penelitian ini menggunakan metode *data mining* dengan pendekatan *Cross Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM merupakan model standar dalam proses analisis data yang terdiri dari beberapa tahapan sistematis, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Namun, dalam penelitian ini tahapan hanya dilakukan hingga *evaluation* tanpa implementasi *deployment*, karena fokus penelitian adalah analisis dan pengelompokan data status gizi balita.



Gambar 1 Alur Penelitian

Adapun tahapan penelitian berdasarkan metode CRISP-DM dapat dirinci sebagai berikut [14]:

Tahap Business Understanding merupakan tahap awal dalam penelitian yang bertujuan untuk memahami permasalahan terkait status gizi balita serta menentukan tujuan yang ingin dicapai. Pada tahap ini dilakukan identifikasi terhadap pentingnya pemantauan status gizi balita sebagai upaya mendukung kesehatan dan pertumbuhan anak. Permasalahan yang sering muncul adalah sulitnya mengidentifikasi kelompok balita berdasarkan karakteristik gizi secara cepat dan tepat. Oleh karena itu, penelitian ini bertujuan menerapkan metode clustering untuk mengelompokkan balita berdasarkan kesamaan karakteristik tertentu sehingga dapat membantu tenaga kesehatan dalam melakukan pemantauan dan pengambilan keputusan terkait status gizi.

Tahap Data Understanding dilakukan untuk memahami data yang akan digunakan dalam proses analisis. Pada tahap ini dilakukan pengumpulan dataset balita dari sumber yang tersedia serta identifikasi atribut yang relevan untuk penelitian. Atribut yang digunakan meliputi berat badan, tinggi badan, usia, jenis kelamin, dan posyandu. Selain memahami struktur data, tahap ini juga mencakup pemeriksaan kualitas data, identifikasi nilai yang hilang (*missing values*), data duplikat, serta analisis deskriptif awal untuk mengetahui karakteristik dan distribusi data yang akan digunakan dalam proses pengelompokan.

Tahap Data Preparation merupakan proses persiapan data sebelum diterapkan ke dalam algoritma clustering. Kegiatan yang dilakukan meliputi pembersihan data (*data cleaning*) untuk mengatasi data yang tidak lengkap, duplikat, maupun data yang tidak valid. Selanjutnya dilakukan transformasi data, seperti mengonversi usia balita ke dalam satuan bulan agar memiliki format yang seragam. Variabel kategorikal seperti jenis kelamin dan posyandu juga dilakukan proses encoding sehingga dapat diolah oleh algoritma K-Means. Selain itu, dilakukan normalisasi atau standarisasi data numerik untuk mengurangi pengaruh perbedaan skala antar atribut sehingga proses pengelompokan dapat menghasilkan cluster yang lebih optimal.

Tahap Modeling merupakan proses penerapan algoritma K-Means untuk mengelompokkan data balita berdasarkan tingkat kemiripan karakteristik yang dimiliki. Algoritma K-Means bekerja dengan membagi data ke dalam sejumlah cluster berdasarkan jarak terdekat terhadap centroid. Sebelum proses clustering dilakukan, jumlah cluster yang optimal ditentukan menggunakan metode Elbow, yaitu dengan menganalisis nilai Within Cluster Sum of Squares (WCSS) pada beberapa alternatif jumlah cluster. Setelah jumlah cluster optimal diperoleh, algoritma K-Means dijalankan untuk menghasilkan kelompok-kelompok balita yang memiliki karakteristik status gizi yang serupa.

Tahap Evaluation dilakukan untuk menilai kualitas hasil pengelompokan yang dihasilkan oleh algoritma K-Means. Evaluasi dilakukan menggunakan metrik Silhouette Score yang mengukur tingkat kesamaan data dalam satu cluster dibandingkan dengan cluster lainnya. Semakin tinggi nilai Silhouette Score, semakin baik kualitas pengelompokan yang dihasilkan. Selain itu, dilakukan interpretasi terhadap nilai centroid pada setiap cluster untuk memahami karakteristik masing-masing kelompok. Hasil interpretasi tersebut kemudian digunakan untuk mengidentifikasi kategori status gizi

balita, seperti kelompok dengan status gizi baik, berisiko, atau memerlukan perhatian khusus, sehingga hasil penelitian dapat memberikan informasi yang bermanfaat bagi pihak terkait dalam pengambilan keputusan.

2.2 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data balita yang diperoleh dari 12 Posyandu di wilayah kerja Puskesmas Bukuan. Data tersebut merupakan hasil pencatatan kegiatan Posyandu yang mencakup informasi antropometri dan identitas dasar balita. Atribut yang digunakan dalam penelitian ini meliputi berat badan, tinggi badan, usia saat pengukuran, jenis kelamin, dan Posyandu. Atribut tersebut dipilih karena merupakan indikator utama dalam penilaian status gizi balita.

Sebelum dilakukan proses clustering, data terlebih dahulu melalui tahap preprocessing yang meliputi pembersihan data, konversi usia ke dalam satuan bulan, serta transformasi data kategorikal menjadi numerik. Proses ini bertujuan untuk memastikan data dapat diproses dengan baik oleh algoritma K-Means. Keberadaan 12 Posyandu dalam dataset memungkinkan analisis distribusi status gizi berdasarkan wilayah, sehingga hasil penelitian dapat memberikan gambaran kondisi gizi balita pada masing-masing Posyandu.

2.3 Algoritma K-Means

K-Means merupakan algoritma pengklasteran yang akan mengelompokkan data dengan melihat pada nilai pada centroid (titik pusat cluster). K-Means menggunakan fungsi jarak pada pengklasteran data terhadap centroid [15]. Berikut ini metode dalam algoritma K-Means:

Proses K-Means dimulai dengan menentukan jumlah cluster (k), kemudian dilanjutkan dengan pemilihan centroid awal secara acak. Selanjutnya, algoritma menghitung jarak antara setiap data dengan centroid menggunakan fungsi jarak. Data akan dikelompokkan ke dalam cluster dengan jarak terdekat, kemudian centroid diperbarui berdasarkan rata-rata nilai data dalam cluster tersebut. Proses ini dilakukan secara iteratif hingga tidak terjadi perubahan cluster atau mencapai kondisi konvergen. Perhitungan jarak antara data dan centroid dalam penelitian ini menggunakan metode *Euclidean Distance* yang dirumuskan sebagai berikut [16]:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Gambar 2 Rumus menghitung jarak data ke centroid pada K-Means

Rumus tersebut digunakan untuk mengukur tingkat kedekatan antara data dengan centroid, sehingga dapat menentukan cluster yang paling sesuai.

Sedangkan perhitungan centroid dilakukan dengan menghitung nilai rata-rata dari seluruh data dalam suatu cluster, yang dirumuskan sebagai berikut [17],[18]:

$$C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$$

Gambar 3 Rumus menghitung centroid pada setiap cluster

Nilai centroid yang diperoleh akan digunakan kembali dalam proses iterasi hingga posisi centroid stabil. Dalam penelitian ini, hasil clustering kemudian diinterpretasikan menjadi kategori status gizi berdasarkan karakteristik masing-masing cluster.

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Pada tahap ini dilakukan identifikasi permasalahan terkait kondisi status gizi balita di wilayah Posyandu Puskesmas Bukuan. Permasalahan utama yang ditemukan adalah proses analisis status gizi yang masih dilakukan secara manual, sehingga kurang efektif dalam mengelompokkan data dalam jumlah besar dan berpotensi menghasilkan informasi yang kurang optimal.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengelompokkan data balita ke dalam beberapa kategori status gizi menggunakan pendekatan *data mining* dengan algoritma K-Means. Hasil pengelompokan diharapkan dapat membantu dalam memberikan gambaran kondisi gizi balita serta mendukung pengambilan keputusan di tingkat Posyandu.

3.2 Data Understanding

Tabel 1. Dataset Awal

No	NIK	Nama	JK	Tgl Lahir	BB Lahir	...	Detail
1	4311229623	HELSA NOVIYANTI	P	03/11/2022	3	...	
2	4712227947	NATHANIA FELICIA	P	07/12/2022	3	...	
3	4806237201	YEMIMA D	P	08/06/2023	3	...	
...	
5	6611218738	NUR ANINDHYA. N	P	26/11/2021	3	...	

Pada Tabel 1. menunjukkan dataset awal yang digunakan dalam penelitian, yang terdiri dari 2145 data dengan 36 atribut yang mencakup informasi identitas balita, data antropometri, serta data administratif dari 12 Posyandu di wilayah Puskesmas Bukuan.

	Berat	Tinggi	Usia_Saat_Ukur	JK
0	15.0	99.9	2 Tahun - 10 Bulan - 0 Hari	P
1	11.9	90.1	2 Tahun - 8 Bulan - 27 Hari	P
2	12.0	84.0	2 Tahun - 2 Bulan - 26 Hari	P
3	13.8	103.2	4 Tahun - 1 Bulan - 4 Hari	P
4	18.3	104.2	4 Tahun - 11 Bulan - 23 Hari	P

Gambar 5 Seleksi atribut

Gambar 3. menunjukkan hasil seleksi atribut yang digunakan dalam penelitian. Dari total 36 atribut, hanya dipilih beberapa atribut utama, yaitu berat badan, tinggi badan, usia saat ukur, dan jenis kelamin. Pemilihan atribut ini didasarkan pada indikator antropometri yang umum digunakan dalam penilaian status gizi balita.

3.3 Data Preparation

	Berat	Tinggi	Usia_Saat_Ukur	JK
0	15.0	99.9	34	1
1	11.9	90.1	32	1
2	12.0	84.0	26	1
3	13.8	103.2	49	1
4	18.3	104.2	59	1

Gambar 6 Transformasi data

Gambar 4. menunjukkan hasil transformasi data, khususnya pada variabel usia yang sebelumnya berbentuk teks (tahun, bulan, hari) menjadi satuan bulan. Transformasi ini dilakukan untuk memastikan bahwa seluruh data berada dalam bentuk numerik sehingga dapat diproses oleh algoritma K-Means. Selain itu, variabel jenis kelamin juga diubah ke dalam bentuk numerik menggunakan teknik encoding.

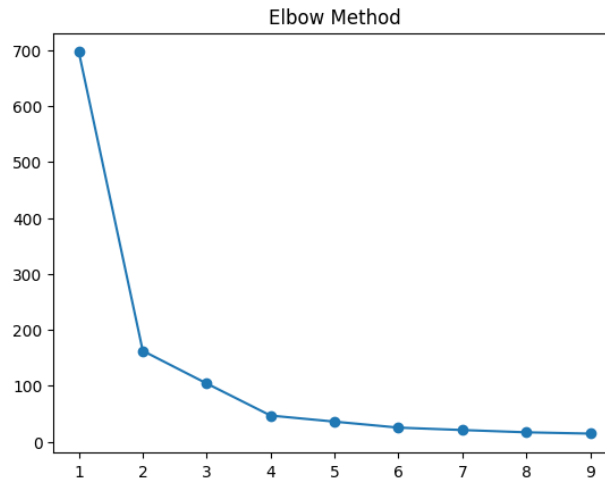
```
fitur = fitur.dropna()
print("\nJumlah data setelah cleaning:", len(fitur))
```

Jumlah data setelah cleaning: 2145

Gambar 7 Data cleaning

Gambar 5. menunjukkan hasil proses pembersihan data (*data cleaning*). Hasil cleaning menunjukkan bahwa data yang digunakan dalam penelitian berjumlah 2145 data, yang berarti seluruh data telah memenuhi kriteria untuk dianalisis.

3.4 Modeling



Gambar 8 Elbow method

Gambar 6. menunjukkan grafik Elbow Method yang digunakan untuk menentukan jumlah cluster optimal. Grafik tersebut memperlihatkan hubungan antara jumlah cluster (k) dengan nilai inertia. Berdasarkan grafik, terlihat bahwa penurunan inertia yang signifikan terjadi dari $k=1$ hingga $k=3$, setelah itu penurunan mulai melandai. Hal ini menunjukkan bahwa titik siku (*elbow*) berada pada $k=3$, sehingga jumlah cluster optimal yang digunakan dalam penelitian ini adalah tiga cluster. Pemilihan jumlah cluster ini juga diperkuat oleh konteks penelitian, yaitu pengelompokan status gizi yang secara umum terbagi menjadi tiga kategori.

K = 2, Silhouette Score = 0.7001187395292758
 K = 3, Silhouette Score = 0.6294141016160586
 K = 4, Silhouette Score = 0.5911219549166972
 K = 5, Silhouette Score = 0.5507551464129639

Gambar 9 Silhouette score

Gambar 7. menunjukkan hasil evaluasi clustering menggunakan metode *Silhouette Score*. Nilai yang diperoleh sebesar 0.629 menunjukkan bahwa kualitas clustering berada dalam kategori baik. Nilai ini mengindikasikan bahwa sebagian besar data berada pada cluster yang tepat dan memiliki jarak yang cukup jelas dengan cluster lainnya, meskipun masih terdapat sedikit overlap.

Cluster	Berat	Tinggi	Keterangan
0	2 10.639163	80.523317	Gizi Kurang
1	0 12.172792	88.456032	Gizi Normal
2	1 15.036979	98.155347	Gizi Lebih

Gambar 10 Hasil Centroid

Gambar 8. menunjukkan nilai centroid dari masing-masing cluster yang dihasilkan oleh algoritma K-Means. Setiap centroid merepresentasikan nilai rata-rata atribut data dalam satu cluster, yang terdiri dari berat badan dan tinggi

badan. Berdasarkan hasil yang diperoleh, cluster dengan nilai rata-rata berat badan dan tinggi badan paling rendah, yaitu sebesar 10.63 kg dan 80.52 cm, dikategorikan sebagai gizi kurang. Selanjutnya, cluster dengan nilai rata-rata sedang, yaitu 12.17 kg dan 88.45 cm, dikategorikan sebagai gizi normal. Sementara itu, cluster dengan nilai rata-rata tertinggi, yaitu 15.03 kg dan 98.15 cm, dikategorikan sebagai gizi lebih. Interpretasi ini menunjukkan bahwa pengelompokan yang dihasilkan oleh algoritma K-Means telah sesuai dengan karakteristik data antropometri, dimana perbedaan nilai berat dan tinggi badan menjadi dasar utama dalam menentukan kategori status gizi pada masing-masing cluster.

	Berat	Tinggi	Usia_Saat_Ukur	JK	Cluster
0	15.0	99.9	34	1	0
1	11.9	90.1	32	1	0
2	12.0	84.0	26	1	0
3	13.8	103.2	49	1	0
4	18.3	104.2	59	1	0

Gambar 11 Label Cluster pada Data

Gambar 9. menampilkan contoh lima data pertama yang telah melalui proses clustering menggunakan algoritma K-Means. Setiap data terdiri dari atribut berat badan, tinggi badan, usia saat pengukuran, jenis kelamin, serta label cluster yang dihasilkan. Kolom Cluster menunjukkan hasil pengelompokan data berdasarkan kemiripan karakteristik. Pada contoh tersebut, seluruh data termasuk ke dalam cluster 0, yang menandakan bahwa data-data tersebut memiliki karakteristik yang serupa berdasarkan atribut yang digunakan dalam proses clustering. Hasil ini menunjukkan bahwa algoritma K-Means telah berhasil mengelompokkan data secara otomatis tanpa label awal, dimana setiap data ditempatkan ke dalam cluster yang memiliki kedekatan nilai atribut, khususnya berat badan dan tinggi badan sebagai indikator utama dalam analisis status gizi.

3.5 Evaluation

Cluster	Status_Gizi	count
0	Gizi Kurang	1003
2	Gizi Lebih	610
1	Gizi Normal	532

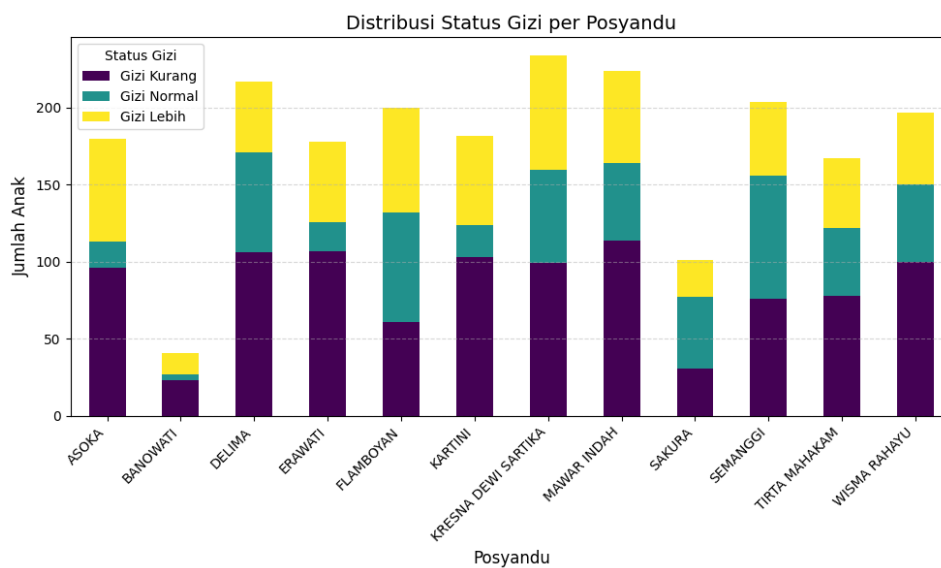
Gambar 12 Distribusi gizi

Gambar 10. menunjukkan bahwa kategori gizi cukup memiliki jumlah terbanyak yaitu 1003 data, diikuti oleh kurang gizi sebanyak 610 data, dan gizi baik sebanyak 532 data. Hasil ini menunjukkan bahwa sebagian besar balita berada pada kondisi gizi cukup, namun jumlah balita dengan kurang gizi masih cukup signifikan sehingga memerlukan perhatian lebih.

Status_Gizi	Gizi Kurang	Gizi Lebih	Gizi Normal
Posyandu			
ASOKA	96	67	17
BANOWATI	23	14	4
DELIMA	106	46	65
ERAWATI	107	52	19
FLAMBOYAN	61	68	71
KARTINI	103	58	21
KRESNA DEWI SARTIKA	99	74	61
MAWAR INDAH	114	60	50
SAKURA	31	24	46
SEMANGGI	76	48	80
TIRTA MAHAKAM	78	45	44
WISMA RAHAYU	100	47	50

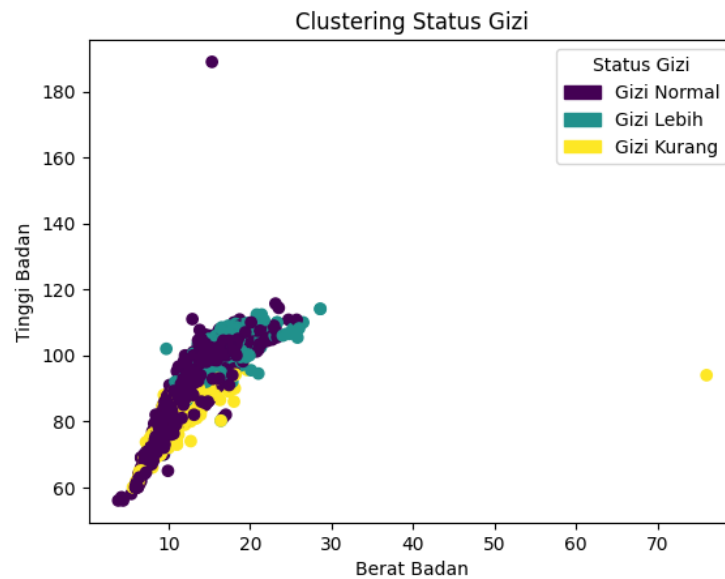
Gambar 13 Distribusi per Posyandu

Gambar 11. menunjukkan distribusi status gizi pada masing-masing Posyandu. Terlihat adanya variasi jumlah balita pada setiap kategori gizi di tiap Posyandu. Beberapa Posyandu memiliki dominasi pada kategori gizi kurang, sementara lainnya menunjukkan distribusi yang lebih seimbang. Hal ini mengindikasikan adanya perbedaan kondisi status gizi antar wilayah Posyandu.



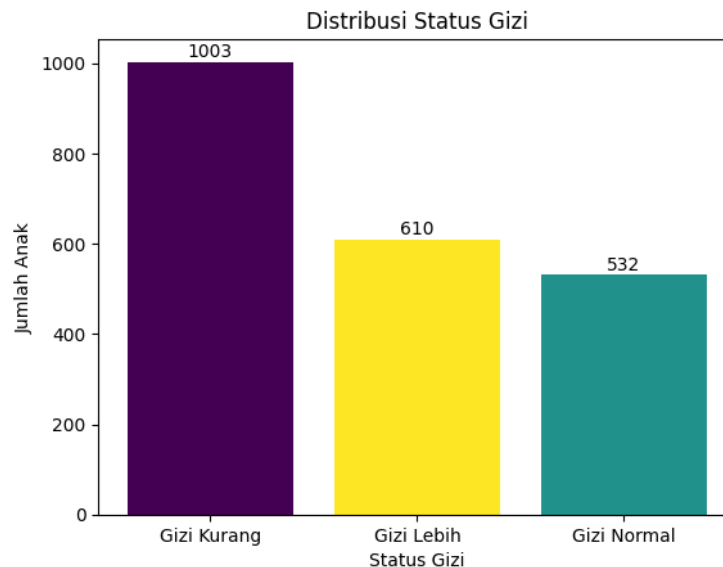
Gambar 14 Tabel Distribusi Status Gizi per Posyandu

Gambar 12. menampilkan jumlah balita pada setiap kategori status gizi berdasarkan Posyandu. Data ini memberikan informasi rinci terkait sebaran status gizi pada masing-masing wilayah. Sebagai contoh, Posyandu Mawar Indah memiliki jumlah balita gizi kurang yang cukup tinggi dibandingkan kategori lainnya, sedangkan Posyandu Sakura menunjukkan jumlah gizi normal yang relatif lebih dominan.



Gambar 15 *Scatter plot* Cluster

Gambar 13. menunjukkan visualisasi hasil clustering dalam bentuk *Scatter plot*. Setiap titik pada grafik merepresentasikan satu data balita, sedangkan perbedaan warna menunjukkan keanggotaan cluster yang dihasilkan oleh algoritma K-Means. Berdasarkan visualisasi tersebut, terlihat bahwa data terbagi menjadi tiga kelompok cluster yang relatif terpisah. Sebagian besar titik dalam setiap cluster cenderung berdekatan, yang menunjukkan bahwa data dalam cluster memiliki karakteristik yang serupa. Hal ini menandakan bahwa algoritma K-Means berhasil mengelompokkan data berdasarkan kemiripan atribut yang digunakan, yaitu berat badan, tinggi badan, usia, dan jenis kelamin.



Gambar 16 Grafik distribusi gizi

Gambar 14. menunjukkan distribusi jumlah data berdasarkan kategori status gizi. Hasil menunjukkan bahwa kategori gizi cukup memiliki jumlah terbanyak, diikuti oleh kurang gizi dan gizi baik. Distribusi ini memberikan gambaran umum kondisi gizi balita di wilayah penelitian, yang menunjukkan bahwa sebagian besar balita berada dalam kondisi cukup, namun masih terdapat jumlah yang signifikan pada kategori kurang gizi.

4. KESIMPULAN

Berdasarkan hasil penelitian, penerapan metode data mining menggunakan algoritma K-Means dengan pendekatan CRISP-DM mampu mengelompokkan data status gizi balita secara efektif dan sistematis. Proses analisis yang dilakukan mulai dari tahap pemahaman masalah hingga evaluasi menghasilkan model clustering yang optimal dengan jumlah tiga cluster berdasarkan metode Elbow. Hasil evaluasi menggunakan *Silhouette Score* sebesar 0,629 menunjukkan bahwa kualitas pengelompokan berada dalam kategori baik, sehingga dapat dikatakan bahwa model mampu memisahkan data antar cluster dengan cukup jelas. Berdasarkan nilai centroid, cluster yang terbentuk berhasil diinterpretasikan menjadi tiga kategori status gizi, yaitu gizi kurang, gizi cukup, dan gizi lebih. Distribusi hasil clustering menunjukkan bahwa sebagian besar balita berada pada kategori gizi cukup, namun jumlah balita dengan gizi kurang masih tergolong signifikan dan memerlukan perhatian lebih. Selain itu, terdapat variasi distribusi status gizi antar Posyandu yang mengindikasikan adanya perbedaan kondisi kesehatan di setiap wilayah. Secara keseluruhan, penelitian ini membuktikan bahwa penggunaan algoritma K-Means dapat membantu dalam mengidentifikasi pola status gizi balita secara lebih akurat dibandingkan metode manual. Hasil penelitian ini diharapkan dapat menjadi dasar dalam pengambilan keputusan serta perencanaan intervensi program kesehatan yang lebih tepat sasaran dan berbasis data di tingkat Posyandu maupun Puskesmas.

REFERENCES

1. Ipmawati, J., & Unggara, I. (2024). Analisis Status Gizi Anak Menggunakan Metode Klastering Pada Dataset Anthropometri. *Bit-Tech*, 7(2), 494–504. <https://doi.org/10.32877/Bt.V7i2.1869>
2. Kementerian Kesehatan Republik Indonesia, "Survei Kesehatan Indonesia (Ski) 2023 Dalam Angka," Jakarta: Badan Kebijakan Pembangunan Kesehatan, 2023.
3. Vicky Pranandika Wijaksana. (2025). Penerapan Metode K-Means Clustering Status Gizi Balita Di Upt Puskesmas Barong Tongkok. *Data Sciences Indonesia (Dsi)*, 5(1), 156–165. <https://doi.org/10.47709/Dsi.V5i1.6517>
4. Fadillah, I., Gizi, A., & Banda Aceh Dosen Jurusan Gizi Poltekkes Kemenkes Aceh, R. R. (N.D.). *Penilaian Status Gizi Dan Pertumbuhan Balita : Standar Baru Antropometri Who-2006 Multicentre Growth Reference Study (Mgrs)*.
5. Roza, Y. B., Defit, S., & Arlis, S. (2025). Analisis Cluster Algoritma K-Means Untuk Pengelompokan Kondisi Gizi Balita Pada Posyandu. *Bulletin Of Computer Science Research*, 5(5), 1182–1187. <https://doi.org/10.47065/Bulletincsr.V5i5.752>
6. Komang Sri Julyantari, N., Komang Budiarta, I., Made Dewi Kansa Putri, N., Teknologi Dan Bisnis Stikom Bali, I., Komputer, S., Informasi Dan Komputer, F., Digital, B., Bisnis Dan Vokasi, F., & Stikom Bali, I. (2021). Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih) Implementation Of K-Means For Clustering The Nutritional Status Of Toddlers (Banjar Titih Case Study). *Jurnal Janitra Informatika Dan Sistem Informasi*, 1(2), 92–101. <https://doi.org/10.25008/Janitra>
7. Mustaqim, K., Al Raffi, G., Nita Sawalia Fitri, & Riza, N. (2025). Penerapan Metode K-Means Untuk Klasifikasi Tingkat Literasi Terhadap Penyebaran Hoaks Di Jawa Barat. *Data Sciences Indonesia (DSI)*, 5(2), 1–9. <https://doi.org/10.47709/dsi.v5i2.5664>
8. Rahmatul Husna, & Verry Riyanto. (2024). Klasterisasi Status Gizi Balita Menggunakan Algoritma K-Means Melalui Pendekatan Soft System Methodology. *Jurnal Ticom: Technology of Information and Communication*, 13(1), 1–9. <https://doi.org/10.70309/ticom.v13i1.120>
9. Febriansyah, F. (2024). Penerapan Algoritma K-Means Clustering Data Gizi Balita Pada Uptd Puskesmas Bumi Agung. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3). <https://doi.org/10.23960/jitet.v12i3>
10. Edwardo, E., & Damayanti, N. R. (2025). Analisis Clustering Data Gizi Pada Puskesmas 1 Ulu Menggunakan Metode Algoritma K-Means. *Jurnal Algoritme*, 5(2), 209–217. <https://doi.org/10.35957/algoritme.v5i2.10707>
11. Bayu Lokananta, R., Yuana, H., & Dwi Puspitasari, W. (2024). Implementasi Algoritma K-Means Terhadap Pengelompokan Status Gizi Balita (Studi Kasus : Posyandu Melati Vii). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(5), 3585–3592. <https://doi.org/10.36040/jati.v7i5.7377>
12. Napitu, S., & Hutabarat, H. D. M. (2024). Implementasi Algoritma K-Means untuk Prediksi Status Gizi Balita pada Tiga Puskesmas di Kecamatan Simanindo. *Jurnal Ilmu Komputer Dan Teknologi*, 5(3), 8–15. <https://doi.org/10.35960/ikomti.v5i3>
13. Vicky Pranandika Wijaksana. (2025). Penerapan Metode K-Means Clustering Status Gizi Balita Di UPT Puskesmas Barong Tongkok. *Data Sciences Indonesia (DSI)*, 5(1), 156–165. <https://doi.org/10.47709/dsi.v5i1.6517>
14. Putri Jelita, H., & Ibnu Saad, M., Wahyuni (2025). Penerapan Algoritma Naïve Bayes Dalam Analisis Sentimen Masyarakat Terhadap STMIK Widya Cipta Dharma. *Bulletin of Information Technology (BIT)*, 6(2), 148–160. <https://doi.org/10.47065/bit.v5i2.2029>

15. Fatkhudin, A., Khambali, A., Artanto, F. A., Putra Zade, N. A., & Muhammadiyah Pekajangan Pekalongan, U. (2023). Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika). *Jurnal Minfo Polgan*, 12(2). <https://doi.org/10.33395/jmp.v12i2.12494>
16. Wardhana. A.E, Subroto. I.M.I, & Prasetyowati. S.A.D. (2018). Optimizing Group Discussion Generation Using K-Means Clustering And Fair Distribution. *Journal of Telematics and Informatics (JTI)*, 6.
17. Sintawati, I. D., Widiarina, W., & Mariskhana, K. (2024). Analysis of Malnutrition Status in Toddlers Using the K-MEANS Algorithm Case Study in DKI Jakarta Province. *Sinkron*, 8(4), 2318–2324. <https://doi.org/10.33395/sinkron.v8i4.14087>
18. Yulia, B., Fahik, L., Djahi, B. S., Rumlaklak, N. D., & Komputer, J. I. (2018). Data Mining Untuk Klasifikasi Status Gizi Desa Di Kabupaten Malaka Menggunakan Metode K-Nearest Neighbor. *J-Icon*, 6(1), 1–7.
19. R. Sari and A. Wahyudi, Penerapan Algoritma K-Means untuk Pengelompokan Status Gizi Balita pada Posyandu, *Jurnal Informatika dan Teknologi Komputer*, vol. 4, no. 2, pp. 115–123, 2023, doi: 10.31294/jtk.v4i2.1456.
20. H. Ekawati, Wahyuni, and N. R. Sari, Penerapan Taksonomi Bloom dan Krathwohl's pada Aplikasi Rubrik Penilaian Hasil Belajar Siswa di Samarinda untuk Aspek Afektif, *Jurnal Ilmiah Matrik*, vol. 23, no. 2, pp. 189–200, Aug. 2021.